Probability Metrics and the Stability of Stochastic Models

SVETLOZAR T. RACHEV
Department of Statistics and Applied Probability
University of California
Santa Barbara
USA

JOHN WILEY & SONS
Chichester . New York . Brisbane . Toronto . Singapore

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Jacaranda Wiley Ltd, G.P.O. Box 859, Brisbane,
Queensland 4001, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd, 37 Jalan Pemimpin 05-04,
Block B, Union Industrial Building, Singapore 2057

Library of Congress Cataloging-in-Publication Data:

Rachev, S. T. (Svetlozar Todorov)
Probability metrics / Svetlozar T. Rachev.
p. cm. — (Wiley series in probability and mathematical statistics. Applied probability and statistics section)
Includes bibliographical references and index.
ISBN 0 471 92877 1
1. Limit theorems (Probability theory) 2. Metric spaces.
I. Title. II. Series.
QA273.67.R33 1991
519.2—dc20 90-43733
CIP

British Library Cataloguing in Publication Data:

Rachev, Svetlozar T.
Probability metrics.
1. Probabilities & statistical mathematics
I. Title
519.2

ISBN 0 471 92877 1

Typeset by Techset Composition Ltd, Salisbury, UK.
Printed in Great Britain by Courier International, Tiptree, Essex

To my children

Borjana and Vladimir

Contents

Preface

The study of limit theorems and a number of other questions in probability theory makes it necessary to introduce functionals, defined either on classes of probability distributions or on classes of random elements, and evaluating their nearness in one or another probabilistic sense. Thus various metrics have appeared, among which are the well known Kolmogorov (uniform) metric, L^p metrics, the Prokhorov metric, the metric of convergence in probability (Ky Fan metric) and others. The use of metrics in many problems in probability theory is connected with the following fundamental question:

'Is the proposed stochastic model a satisfactory approximation to the real model, and if so, within what limits?' To answer this question, an investigation of the qualitative and quantitative stability of the stochastic model is required. Analysis of quantitative stability assumes the use of metrics as measures of comparability. The main idea of the *method of metric distances* (MMD), developed by V. M. Zolotarev and his students to solve stability problems, is reduced to the following two aspects.

Problem 1 (Choice of ideal metrics). Find the most appropriate (ideal) metrics for the stability problem under consideration. Then, solve the problem in terms of these ideal metrics.

Problem 2 (Comparison of metrics). If it is required to write the solution of the stability problem in terms of other metrics, one must solve the problem of comparison of these metrics with the chosen (ideal) metrics.

Unlike Problem 1, Problem 2 does not depend on the specific stochastic model under consideration. Thus, the independent solution of the second problem allows its re-use in any particular situation. In addition, it enables us to use a variety of metric relationships without making any effort in different kinds of stability problems. Moreover, following the stated two-stage approach, we get a clear comprehension of the specific regularities which form the stability effect.

In probability theory, metrics have been used for a long time, although one usually exploits a very limited class of metrics. Also, some ideas of the MMD have been used for a long time in approximation theory and functional analysis. In view of the variety of stability problems, there are no regular selection rules determining the 'ideal' metric for the given problem. Therefore, the development of the MMD demands the creation of a *theory of probability metrics* (TPM).

The term 'probability metric' means simply a semimetric in a space of random variables (taking values in some separable metric space). In probability theory, sample spaces are usually not fixed and one is interested in those metrics whose values depend on the joint distributions of the pairs of random variables being considered. Each such metric can be considered just as given by a function defined on the set of probability measures on the Cartesian square of the sample space. Complications connected with the question of existence of pairs of random variables on a given space with given probability laws can be easily avoided. Although such a function is not a metric on a space of probability distributions (it is not a function of pairs of measures), small values of it say that the measure is concentrated near the diagonal. Therefore its marginal distributions are close to each other. Fixing these marginal distributions, one can find the infimum of the values of our function on the class of all measures with the given marginals. Such an infimum is a metric on the class of probability distributions and in some concrete cases (for example, for the L_1 distance in the space of random variables—Kantorovich's theorem; for the Ky Fan metric—Strassen–Dudley's theorem; for the indicator metric—Dobrushin's theorem) were found earlier (giving, respectively, the Kantorovich (or Wasserstein) metric, the Prokhorov metric and the total variation distance).

The necessary classification of the set of probability metrics (*p. metrics*) is naturally carried out from the point of view of metric structure and generating topologies. That is why the following two research directions arise.

Direction 1. Description of the basic structures of p. metrics.

Direction 2. Analysis of the topologies in the space of probability measures, generated by different types of p. metrics. This analysis can be carried out with the help of convergence criteria for different metrics.

At the same time, more specialized research directions arise. Namely,

Direction 3. Characterization of the ideal metrics for the given problem.

Direction 4. Investigations of the main relationships between different type of p. metrics.

In this book, all four directions are considered as well as applications to different problems of probability theory. Much attention is paid to the possibility of giving equivalent definitions of p. metrics (for example, in direct and dual terms, in terms of the Hausdorff metric for sets, etc). Indeed, in concrete applications of p. metrics, the use of different equivalent variants of the definitions in different steps of the proof is often a decisive factor.

One of the main classes of metrics considered is the class of *minimal metrics*, the idea of which goes back to the work of Kantorovich in the 1940s on the transportation problems in linear programming. Such metrics have been found independently by many authors in several parts of probability theory (Markov processes, statistical physics, etc.). They are connected with the widely known method of 'coupling.' Then it is natural to evaluate the distance between variables with metrics of the indicated type. Distances for such metrics are hard to compute, but it is easy to give upper bounds, attained by at least one joint probability distribution. Another useful class of metrics studied in this book is the class of 'ideal' metrics having satisfied the following properties: (1) $\mu(P_c, Q_c) \leqslant |c|^r \mu(P, Q)$ for all $c \in [-C, C], c \neq 0$, where $P_c(A) := P((1/c)A)$ for any Borel set A on a Banach space U, and (2) $\mu(P_1 * Q, P_2 * Q) \leqslant \mu(P_1, P_2)$, where $*$ denotes the convolution. This class is convenient for the study of functionals of sums of independent random variables, giving nearest bounds of the distance to limit distributions.

The presentation here is given in a general form, although specific cases are considered as they arise in the process of finding supplementary bounds, or in applications to important special cases.

The MMD given herein is illustrated in some concrete problems. First, there are problems of the type of the Glivenko–Cantelli theorem on the convergence of the empirical measures. Originally, this kind of problem was different in view of the various possible natural modes of convergence and was considered as a problem needing an *ad hoc* approach. Here it is shown that from a general point of view such results turn out to be obvious consequences of the SLLN and general properties of metrics. Analogously, we considered a generalization of the Prokhorov theorem on convergence of random polygons to the Wiener process in the case where one considers the question of convergence of distributions of unbounded functionals. Also considered are applications to the rate of convergence for sums and maxima of random variables and convolution of random motions. Special sections are devoted to application of TPM to the solution of stability problems in queueing theory, risk theory, quality usage, and others.

I would like to thank a number of people who have directly contributed to this project. V. M. Zolotarev introduced me to the subject of theory of probability metrics and stability of stochastic models. He has been a constant source of intellectual guidance and inspiration. R. M. Shortt, L. Rüschendorf, L. de Haan, J. Yukich, E. Omey, L. Baxter, P. Todorovich, M. Taksar and J. Beirlant worked with me on papers which formed the foundation of this book. R. M. Dudley's lecture notes from Aarhus University have been a rich source of ideas. In addition, I want to thank H. Robbins, S. Cambanis, G. Simons, H. Kellerer, S. Resnick and G. Samorodnitski for many helpful discussions during the last three years. My warm thanks and appreciation go to Ruth Bahr, Michelle Bebb and Lee Trimble for their expert typing of this manuscript. Finally, I must thank the editorial and production department of Wiley for their support, patience and superb final product.

While trying to correct all the mistakes in my manuscript, I realize that all of them have not been found. I will be very happy to learn of any mistakes found by the reader together with any comments one might have†.

Svetlozar T. Rachev

† Research supported by NATO Scientific Affairs Division Grant CRG 900798 and Grant from UC Regents, University of California, Santa Barbara.

PART I

General Topics in the Theory of Probability Metrics

CHAPTER 1

Main Directions in the Theory of Probability Metric

In the period of the formation of probability theory as a mathematical science, the first limit theorems (the Law of Large Numbers, De Moivre and Laplace Central Limit Theorems) were obtained (cf. Feller 1970, Chap. VI, 4, and Chap. VII, 3). In applications of these limit theorems, a given probability distribution is often approximated by some limiting distribution. In such cases, the convergence rate problem arises. As mentioned in the Preface, the convergence rate problem requires the concept of a metric as a measure of comparability.

Even in the case of probability distributions on the real line, several different types of metrics (Lévy metric, Kolmogorov metric, L_p-metric) are often used to estimate the closeness between distributions. Since the early thirties, the demands of various applications resulted in the creation of new, more complicated probability models. Both the theory of random processes and the theory of distributions on functional spaces were extensively developed. In connection with limit theorems in general spaces, Fortet and Mourier (1953), Prokhorov (1956), Kantorovich and Rubinstein (1958), and Dudley (1966a) suggested a series of new metrics on spaces of distributions. Certainly, the study of metric properties is not confined to probability theory; such investigations have occurred in other mathematical areas. In fact, some metrics on spaces of measures (e.g. the Kantorovich–Rubinstein metric, total variation norm, and L_p-metrics) play an important role in functional analysis (see Dunford and Schwartz 1988, Kantorovich and Akilov 1984). In probability theory, an even greater variety of such metrics arises in completely natural ways. This variety and suitability can be partly explained by the fact that the stochastic problem under consideration often dictates the choice of an appropriate metric. Frequently, this choice will determine the investigation's success.

Questions concerning the bounds within which stochastic models can be applied (as in all probabilistic limit theorems) can only be answered by investigation of qualitative and quantitative stability. Such stability is very often convenient to express in terms of a metric. This was the case with Zolotarev's Method of Metric Distances (MMD) and the Theory of Probability Metrics (TPM) (see Zolotarev 1976a–d, 1977a,b, 1983a,b, Zolotarev and Rachev 1985).

Figure 1.1.1 summarizes the problems concerning MMD and TPM.

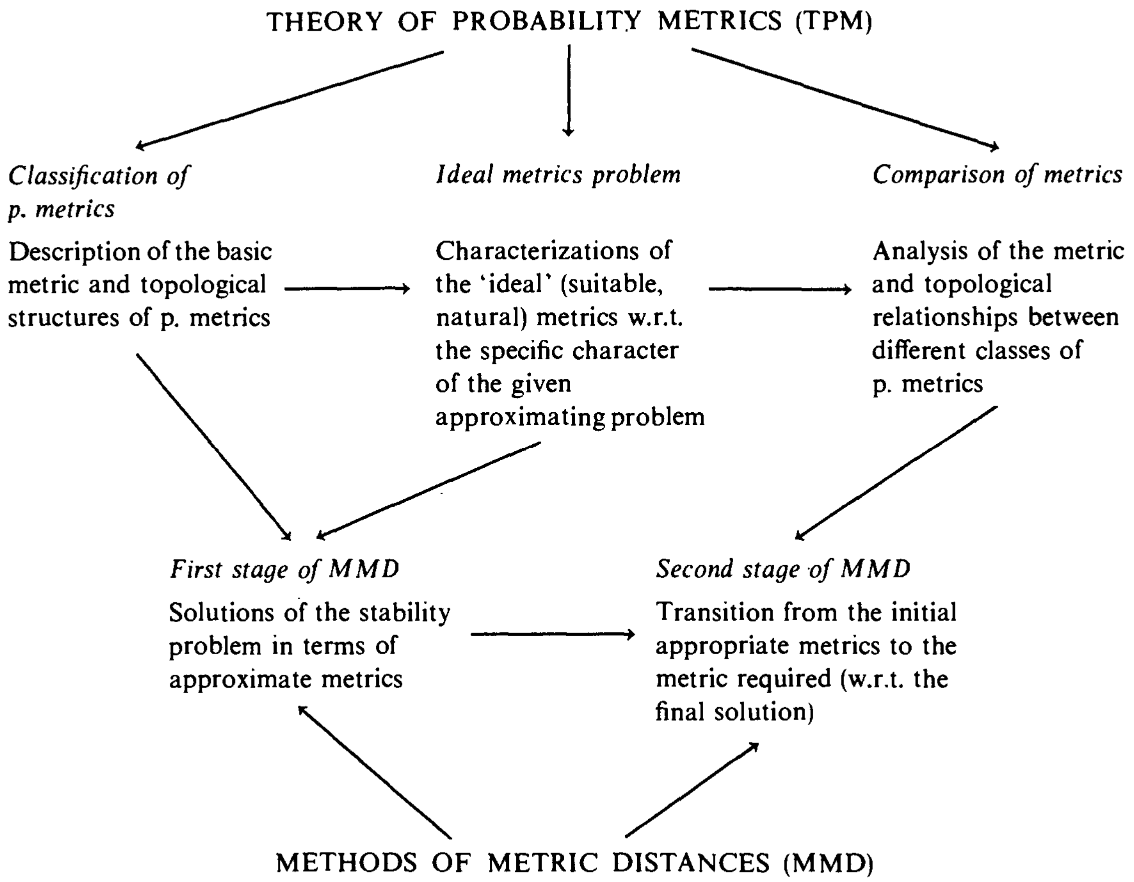


Figure 1.1.1 Theory of the probability metrics as a necessary tool to investigate the method of metric distances. (From Rachev and Shortt, 1990. Reproduced by permission of the American Mathematical Society.)

CHAPTER 2

Probability Distances and Probability Metrics: Definitions

2.1 SOME EXAMPLES OF METRICS IN PROBABILITY THEORY

Below is a list of various metrics commonly found in probability and statistics.

1. *The engineer's metric*

$$\mathbf{EN}(X, Y) := |\mathbb{E}(X) - \mathbb{E}(Y)| \quad X, Y \in \mathfrak{X}^1 \tag{2.1.1}$$

where $\mathfrak{X}^p$ is the space of all real-valued random variables (r.v.s) with $\mathbb{E}|X|^p < \infty$.

2. *The uniform (or Kolmogorov) metric*

$$\boldsymbol{\rho}(X, Y) := \sup\{|F_X(x) - F_Y(x)| : x \in \mathbb{R}\} \quad X, Y \in \mathfrak{X} = \mathfrak{X}(\mathbb{R}) \tag{2.1.2}$$

where F_X is the distribution function (d.f.) of X, $\mathbb{R} = (-\infty, +\infty)$, and $\mathfrak{X}$ is the space of all real-valued r.v.s.

3. *The Lévy metric*

$$\mathbf{L}(X, Y) := \inf\{\varepsilon > 0 : F_X(x - \varepsilon) - \varepsilon \leqslant F_Y(x) \leqslant F_X(x + \varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}. \tag{2.1.3}$$

Remark 2.1.1. We see that $\boldsymbol{\rho}$ and $\mathbf{L}$ may actually be considered as metrics on the space of all distribution functions. However, this cannot be done for $\mathbf{EN}$ simply because $\mathbf{EN}(X, Y) = 0$ does not imply the coincidence of F_X and F_Y, while $\boldsymbol{\rho}(X, Y) = 0 \Leftrightarrow \mathbf{L}(X, Y) = 0 \Leftrightarrow F_X = F_Y$. The Lévy metric metrizes weak convergence (convergence in distribution) in the space $\mathscr{F}$, whereas $\boldsymbol{\rho}$ is often applied in the CLT, cf. Hennequin and Tortrat (1965).

4. *The Kantorovich metric*

$$\boldsymbol{\kappa}(X, Y) = \int_{\mathbb{R}} |F_X(x) - F_Y(x)| \, dx \quad X, Y \in \mathfrak{X}^1.$$

5. *The L_p-metrics between distribution functions*

$$\boldsymbol{\theta}_p(X, Y) := \left(\int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|^p \, \mathrm{d}t\right)^{1/p} \quad p \geqslant 1 \quad X, Y \in \mathfrak{X}^1. \tag{2.1.4}$$

Remark 2.1.2. Clearly, $\boldsymbol{\kappa} = \boldsymbol{\theta}_1$. Moreover, we can extend the definition of $\boldsymbol{\theta}_p$ when $p = \infty$ by setting $\boldsymbol{\theta}_\infty = \boldsymbol{\rho}$. One reason for this extension is the following dual representation for $1 \leqslant p \leqslant \infty$

$$\boldsymbol{\theta}_p(X, Y) = \sup_{f \in \mathscr{F}_p} |\mathbb{E}f(X) - \mathbb{E}f(Y)|, \quad X, Y \in \mathfrak{X}^1$$

where $\mathscr{F}_p$ is the class of all measurable functions f with $\|f\|_q < 1$. Here, $\|f\|_q (1/p + 1/q = 1)$ is defined, as usual, by

$$\|f\|_q := \begin{cases} \left(\int |f|^q\right)^{1/q} & 1 \leqslant q < \infty \\ \operatorname*{ess\,sup}_{\mathbb{R}} |f| & q = \infty. \end{cases}$$

(The proof of the above representation is given by Dudley (1989), p. 333, for the case $p = 1$.)

6. *The Ky Fan metrics*

$$\mathbf{K}(X, Y) := \inf\{\varepsilon > 0 : \Pr(|X - Y| > \varepsilon) < \varepsilon\} \quad X, Y \in \mathfrak{X} \tag{2.1.5}$$

and

$$\mathbf{K}^*(X, Y) := \mathbb{E}\frac{|X - Y|}{1 + |X - Y|}. \tag{2.1.6}$$

Both metrics metrize convergence in probability on $\mathfrak{X} = \mathfrak{X}(\mathbb{R})$, the space of real random variables (Lukacs 1968, Chapter 3, Dudley 1976, Theorem 3.5).

7. *The L_p-metric*

$$\mathscr{L}_p(X, Y) := \{\mathbb{E}|X - Y|^p\}^{1/p} \quad p \geqslant 1 \quad X, Y \in \mathfrak{X}^p. \tag{2.1.7}$$

Remark 2.1.3. Define

$$m^p(X) := \{\mathbb{E}|X|^p\}^{1/p} \quad p \geqslant 1 \quad X \in \mathfrak{X}^p \tag{2.1.8}$$

and

$$\mathbf{MOM}_p(X, Y) := |m^p(X) - m^p(Y)| \quad p \geqslant 1 \quad X, Y \in \mathfrak{X}^p. \tag{2.1.9}$$

Then we have, for $X_0, X_1, \ldots$ in $\mathfrak{X}^p$

$$\mathscr{L}_p(X_n, X_0) \to 0 \Leftrightarrow \begin{cases} \mathbf{K}(X_n, X_0) \to 0 \\ \mathbf{MOM}_p(X_n, X_0) \to 0 \end{cases} \tag{2.1.10}$$

see, for example, Lukacs (1968), Chapter 3.

All of the (semi-)metrics on subsets of $\mathfrak{X}$ mentioned above may be divided into three main groups: primary, simple, and compound (semi-)metrics. A metric μ is *primary* if $\mu(X, Y) = 0$ implies that certain moment characteristics of X and Y agree. As examples, we have **EN** (2.1.1) and $\mathbf{MOM}_p$ (2.1.9). For these metrics

$$\begin{aligned} \mathbf{EN}(X, Y) = 0 &\Leftrightarrow \mathbb{E}X = \mathbb{E}Y \\ \mathbf{MOM}_p(X, Y) = 0 &\Leftrightarrow m^p(X) = m^p(Y). \end{aligned} \tag{2.1.11}$$

A metric μ is *simple* if

$$\mu(X, Y) = 0 \Leftrightarrow F_X = F_Y. \tag{2.1.12}$$

Examples are $\boldsymbol{\rho}$ (2.1.2), **L** (2.1.3), and $\boldsymbol{\theta}_p$ (2.1.4). The third group, the *compound* (semi-)metrics have the property

$$\mu(X, Y) = 0 \Leftrightarrow \Pr(X = Y) = 1. \tag{2.1.13}$$

Some examples are **K** (2.1.5), $\mathbf{K}^*$ (2.1.6), and $\mathscr{L}_p$ (2.1.7).

Later on, precise definitions of these classes will be given, and a study made of the relationships between them. Now we shall begin with a common definition of probability metric which will include the types mentioned above.

2.2 METRIC AND SEMIMETRIC SPACES; DISTANCE AND SEMIDISTANCE SPACES

First of all, let us recall the notions of metric and semimetric space. Generalizations of these notions will be needed in TPM.

Definition 2.2.1. A set $S := (S, \rho)$ is said to be a *metric* space with the metric ρ if ρ is a mapping from the product $S \times S$ to $[0, \infty)$ having the following properties for each $x, y, z \in S$

(1) *Identity property:* $\rho(x, y) = 0 \Leftrightarrow x = y$;
(2) *Symmetry:* $\rho(x, y) = \rho(y, x)$;
(3) *Triangle inequality:* $\rho(x, y) \leqslant \rho(x, z) + \rho(z, y)$.

Some well known examples of metric space are the following.

(a) *The n-dimensional vector space* $\mathbb{R}^n$ endowed with the metric $\rho(x, y) =$

$\|x - y\|_p$, where

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{\min(1,1/p)} \qquad x = (x_1, \ldots, x_n) \in \mathbb{R}^n \quad 0 < p < \infty$$

$$\|x\|_\infty := \sup_{1 \leqslant i \leqslant n} |x_i|.$$

(b) *The Hausdorff metric between closed sets*

$$r(C_1, C_3) = \max\left\{\sup_{x_1 \in C_1} \inf_{x_2 \in C_2} \rho(x_1, x_2), \sup_{x_2 \in C_2} \inf_{x_1 \in C_1} \rho(x_1, x_2)\right\}$$

where C_is are closed sets in a bounded metric space (S, ρ) (Hausdorff 1949).

(c) *The H-metric.* Let $D(\mathbb{R})$ be the space of all bounded function $f: \mathbb{R} \to \mathbb{R}$, continuous from the right and having limits from the left, $f(x-) = \lim_{t \uparrow x} f(t)$. For any $f \in D(\mathbb{R})$ define the graph Γ_f as the union of the sets $\{(x, y): x \in \mathbb{R}, y = f(x)\}$ and $\{(x, y): x \in \mathbb{R}, y = f(x-)\}$. *The H-metric $H(f, g)$ in $D(\mathbb{R})$ is defined* by the Hausdorff distance between the corresponding graphs, $H(f, g) := r(\Gamma_f, \Gamma_g)$. Note that in the space $\mathscr{F}(\mathbb{R})$ of distribution functions, H metrizes the same convergence as the *Skorokhod metric*:

$$s(F, G) = \inf\Big\{\varepsilon > 0\text{: there exists a strictly increasing continuous function } \lambda: \mathbb{R} \to \mathbb{R}, \text{ such that } \lambda(\mathbb{R}) = \mathbb{R}, \sup_{t \in \mathbb{R}} |\lambda(t) - t| < \varepsilon \text{ and } \sup_{t \in \mathbb{R}} |F(\lambda(t)) - G(t)| < \varepsilon\Big\}.$$

Moreover, H-convergence in $\mathscr{F}$ implies convergence in distributions (the weak convergence). Clearly, $\boldsymbol{\rho}$-convergence (see (2.1.2)) implies H-convergence. (A more detailed analysis of the metric H will be given in Section 4.1.)

If the identity property in Definition 2.2.1 is weakened by changing (1) to

$$(1^*) \qquad x = y \Rightarrow \boldsymbol{\rho}(x, y) = 0,$$

then S is said to be a *semimetric space* (or *pseudometric space*) and ρ a *semimetric* (or *pseudometric*) in S. For example, the Hausdorff metric r is only a semimetric in the space of all Borel subsets of a bounded metric space (S, ρ).

Obviously, in the space of real numbers **EN** (see 2.1.1) is the usual uniform metric on the real line $\mathbb{R}$, i.e. $\mathbf{EN}(a, b) := |a - b|$, $a, b \in \mathbb{R}$. For $p \geqslant 0$, define $\mathscr{F}^p$ as the space of all distribution functions F with $\int_{\infty}^{0} F(x)^p \, dx + \int_0^\infty (1 - F(x))^p \, dx < \infty$. The distribution function space $\mathscr{F} = \mathscr{F}^0$ can be considered as a metric space with metrics $\boldsymbol{\rho}$ and $\mathbf{L}$, while $\boldsymbol{\theta}_p (1 \leqslant p < \infty)$ is a metric in $\mathscr{F}^p$. The Ky-Fan metrics (see (2.1.5), (2.1.6)) [resp. $\mathscr{L}_p$-metric (see 2.1.7)] may be viewed as semimetrics in $\mathfrak{X}$ (resp. $\mathfrak{X}^p$) as well as metrics in the space of all Pr-equivalence classes

$$\tilde{X} := \{Y \in \mathfrak{X}: \Pr(Y = X) = 1\} \quad \forall X \in \mathfrak{X} \text{ [resp. } \mathfrak{X}^p]. \tag{2.2.1}$$

EN, $\mathbf{MOM}_p$, $\boldsymbol{\theta}_p$, $\mathscr{L}_p$ can take infinite values in $\mathfrak{X}$ so we shall assume, in the next generalization of the notion of metric, that ρ may take infinite values; at the same time we shall extend also the notion of triangle inequality.

Definition 2.2.2. The set S is called a *distance space* with distance ρ and parameter $\mathbb{K} = \mathbb{K}_\rho$ if ρ is a function from $S \times S$ to $[0, \infty]$, $\mathbb{K} \geqslant 1$ and for each $x, y, z \in S$ the identity property (1) and the symmetry property (2) hold as well as the following version of the triangle inequality: (3*) (*Triangle inequality with parameter* $\mathbb{K}$)

$$\rho(x, y) \leqslant \mathbb{K}[\rho(x, z) + \rho(z, y)]. \tag{2.2.2}$$

If, in addition, the identity property (1) is changed to (1*) then S is called a semidistance space and ρ is called a semidistance (with parameter $\mathbb{K}_\rho$).

Here and in the following we shall distinguish the notions 'metric' and 'distance', using 'metric' only in the case of 'distance with parameter $\mathbb{K} = 1$, taking finite or infinite values'.

Remark 2.2.1. It is not difficult to check that each distance ρ generates a topology in S with a basis of open sets $B(a, r) := \{x \in S; \rho(x, a) < r\}$, $a \in S$, $r > 0$. We know, of course, that every metric space is normal and that every separable metric space has a countable basis. In much the same way, it is easily shown that the same is true for distance space. Hence, by Urysohn's Metrization Theorem (Dunford and Schwartz 1988, I.6.19), every separable distance space is metrizable.

Actually, distance spaces have been used in functional analysis for a long time as is seen by the following examples.

Example 2.2.1. Let $\mathscr{H}$ by the class of all nondecreasing continuous functions H from $[0, \infty)$ onto $[0, \infty)$ which vanish at the origin and satisfy Orlicz's condition

$$K_H := \sup_{t>0} \frac{H(2t)}{H(t)} < \infty. \tag{2.2.3}$$

Then $\tilde{\rho} := H(\rho)$ is a distance in S for each metric ρ in S and $\mathbb{K}_{\tilde{\rho}} = K_H$.

Example 2.2.2. (*Birnbaum–Orlicz distance space*, Birnbaum and Orlicz (1931), and Dunford and Schwartz (1988), p. 400.)

The *Birnbaum–Orlicz space* $L^H(H \in \mathscr{H})$ consists of all integrable functions on [0, 1] endowed with *Birnbaum–Orlicz distance*

$$\rho_H(f_1, f_2) := \int_0^1 H(|f_1(x) - f_2(x)|)\,dx. \tag{2.2.4}$$

Obviously, $\mathbb{K}_{\rho_H} = K_H$.

Example 2.2.3. Similarly to (2.2.4) Kruglov (1973) introduced the following distance in the space of distribution functions

$$\mathbf{Kr}(F, G) = \int \phi(F(x) - G(x))\,dx \tag{2.2.5}$$

where the function ϕ satisfies the following conditions

(a) ϕ is even and strictly increasing on $[0, \infty)$, $\phi(0) = 0$,
(b) for any x and y and some fixed $A \geqslant 1$

$$\phi(x + y) \leqslant A(\phi(x) + \phi(y)). \tag{2.2.6}$$

Obviously, $\mathbb{K}_{\mathbf{Kr}} = A$.

2.3 DEFINITIONS OF PROBABILITY DISTANCE AND PROBABILITY METRIC

Let U be a separable metric space (s.m.s.) with metric d, $U^k = \underbrace{U \times \cdots \times U}_{k \text{ times}}$ the k-fold Cartesian product of U and $\mathscr{P}_k = \mathscr{P}_k(U)$ the space of all probability measures defined on the σ algebra $\mathscr{B}_k = \mathscr{B}_k(U)$ of Borel subsets of U^k. We shall use the terms 'probability measure' and 'law' interchangeably. For any set $\{\alpha, \beta, \ldots, \gamma\} \subseteq \{1, 2, \ldots, k\}$ and for any $P \in \mathscr{P}_k$ let us define the marginal of P on the coordinates $\alpha, \beta, \ldots, \gamma$ by $T_{\alpha, \beta, \ldots, \gamma}P$. For example, for any Borel subsets A and B of U, $T_1P(A) = P(A \times U \times \cdots \times U)$, $T_{1,3}P(A \times B) = P(A \times U \times B \times \cdots \times U)$. Let $\mathbb{B}$ be the operator in U^2 defined by $\mathbb{B}(x, y) := (y, x)$ $(x, y \in U)$. All metrics $\mu(X, Y)$ cited in Section 2.1 (see (2.1.1)–(2.1.9)) are completely determined by the joint distributions $\Pr_{X,Y}(\Pr_{X,Y} \in \mathscr{P}_2(\mathbb{R}))$ of the random variables $X, Y \in \mathfrak{X}(\mathbb{R})$. In the next definition we shall introduce the notion of probability distance and thus we shall describe the primary, simple, and compound metrics in a uniform way. Moreover, the space where the r.v.s X and Y take values will be extended to U, an arbitrary s.m.s.

Definition 2.3.1. A mapping μ defined on $\mathscr{P}_2$ and taking values in the extended interval $[0, \infty]$ is said to be a *probability semidistance with parameter* $\mathbb{K} :=$

$\mathbb{K}_\mu \geqslant 1$ (or briefly, *p. semidistance*) in $\mathscr{P}_2$, if it possesses the three properties listed below

(1) **ID** (*Identity Property*). If $P \in \mathscr{P}_2$ and $P(\bigcup_{x \in U} \{(x, x)\}) = 1$ then $\mu(P) = 0$

(2) **SYM** (*Symmetry*). If $P \in \mathscr{P}_2$ then $\mu(P \circ \mathbb{B}^{-1}) = \mu(P)$

(3) **TI** (*Triangle Inequality*). If P_{13}, P_{12}, $P_{23} \in \mathscr{P}_2$ and there exists a law $Q \in \mathscr{P}_3$ such that the following 'consistency' condition holds:

$$T_{13}Q = P_{13} \quad T_{12}Q = P_{12} \quad T_{23}Q = P_{23}, \tag{2.3.1}$$

then

$$\mu(P_{13}) \leqslant \mathbb{K}[\mu(P_{12}) + \mu(P_{23})].$$

If $\mathbb{K} = 1$ then μ is said to be a *probability semimetric*. If we strengthen the condition ID to
$\widetilde{\mathbf{ID}}$: If $P \in P_2$, then

$$P(\bigcup \{(x, x): x \in U\}) = 1 \Leftrightarrow \mu(P) = 0,$$

then we say that μ is a *probability distance with parameter* $\mathbb{K} = \mathbb{K}_\mu \geqslant 1$ (or briefly, *p. distance*).

Definition 2.3.1 acquires a visual form in terms of random variables, namely, let $\mathfrak{X} := \mathfrak{X}(U)$ be the set of all r.v.s on a given probability space $(\Omega, \mathscr{A}, \Pr)$ taking values in $(U, \mathscr{B}_1)$. By $\mathscr{L}\mathfrak{X}_2 := \mathscr{L}\mathfrak{X}_2(U) := \mathscr{L}\mathfrak{X}_2(U; \Omega, \mathscr{A}, \Pr)$ we denote the space of all joint distributions $\Pr_{X,Y}$ generated by the pairs $X, Y \in \mathfrak{X}$. Since $\mathscr{L}\mathfrak{X}_2 \subseteq \mathscr{P}_2$, then the notion of p. (semi-)distance is naturally defined on $\mathscr{L}\mathfrak{X}_2$. Considering μ on the subset $\mathscr{L}\mathfrak{X}_2$, we shall put

$$\mu(X, Y) := \mu(\Pr_{X,Y})$$

and call μ a *p. semidistance on* $\mathfrak{X}$. If μ is a p. distance, then we use the phrase *p. distance* on $\mathfrak{X}$. Each p. semidistance [resp. distance] μ on $\mathfrak{X}$ is a semidistance [resp. distance] on $\mathfrak{X}$ in the sense of Definition 2.2.2. Then the relationships **ID**, $\widetilde{\mathbf{ID}}$, **SYM**, and **TI** have simple 'metrical' interpretations:

ID$^{(*)}$ $\quad \Pr(X = Y) = 1 \Rightarrow \mu(X, Y) = 0$

$\widetilde{\mathbf{ID}}^{(*)}$ $\quad \Pr(X = Y) = 1 \Leftrightarrow \mu(X, Y) = 0$

SYM$^{(*)}$ $\quad \mu(X, Y) = \mu(Y, X)$

TI$^{(*)}$ $\quad \mu(X, Z) \leqslant \mathbb{K}[\mu(X, Z) + \mu(Z, Y)]$.

Definition 2.3.2. A mapping $\mu: \mathscr{L}\mathfrak{X}_2 \to [0, \infty]$ is said to be a *probability semidistance* in $\mathfrak{X}$ [resp. *distance*] *with parameter* $\mathbb{K} := \mathbb{K}_\mu \geqslant 1$, if $\mu(X, Y) = \mu(\Pr_{X,Y})$ satisfies the properties **ID**$^{(*)}$ [resp. $\widetilde{\mathbf{ID}}^{(*)}$], **SYM**$^{(*)}$ and **TI**$^{(*)}$ for all r.v.s $X, Y, Z \in \mathfrak{X}(U)$.

Example 2.3.1. Let $H \in \mathscr{H}$ (see Example 2.2.1) and (U, d) be a s.m.s. Then $\mathscr{L}_H(X, Y) = \mathbb{E}H(d(X, Y))$ is a p. distance in $\mathfrak{X}(U)$. Clearly, $\mathscr{L}_H$ is finite in the subspace of all X with finite moment $\mathbb{E}H(d(X, a))$ for some $a \in U$. The Kruglov's distance $\mathbf{Kr}(X, Y) := \mathbf{Kr}(F_X, F_Y)$ is a p. semidistance in $\mathfrak{X}(\mathbb{R})$.

Examples of p. metrics in $\mathfrak{X}(U)$ are the Ky Fan metric

$$\mathbf{K}(X, Y) := \inf\{\varepsilon > 0 \colon \Pr(d(X, Y) > \varepsilon) < \varepsilon\} \quad (X, Y \in \mathfrak{X}(U) \tag{2.3.2}$$

and the $\mathscr{L}_p$*-metrics* $(0 \leqslant p \leqslant \infty)$

$$\mathscr{L}_p(X, Y) := \{\mathbb{E}d^p(X, Y)\}^{\min(1, 1/p)} \quad 0 < p < \infty, \tag{2.3.3}$$

$$\mathscr{L}_\infty(X, Y) := \operatorname{ess\,sup} d(X, Y) := \inf\{\varepsilon > 0 \colon \Pr(d(X, Y) > \varepsilon) = 0\} \tag{2.3.4}$$

$$\mathscr{L}_0(X, Y) := \mathbb{E}I\{X, Y\} := \Pr(X, Y). \tag{2.3.5}$$

The engineer's metric **EN**, Kolmogorov metric $\boldsymbol{\rho}$, Kantorovitch metric $\boldsymbol{\kappa}$, and the Lévy metric **L** (see Section 2.1) are p. semimetrics in $\mathfrak{X}(\mathbb{R})$.

Remark 2.3.1. Unlike Definition 2.3.2, Definition 2.3.1 is free of the choice of the initial probability space, and depends only on the structure of the metric space U. The main reason for considering not arbitrary but separable metric spaces (U, d) is that we need the measurability of the metric d in order to connect the metric structure of U with that of $\mathfrak{X}(U)$. In particular, the measurability of d enables us to handle in a well defined way, p. metrics such as the Ky Fan metric **K** and $\mathscr{L}_p$-metrics. Note that $\mathscr{L}_0$ does not depend on the metric d, so one can define $\mathscr{L}_0$ on $\mathfrak{X}(U)$, where U is an arbitrary measurable space, while in (2.3.2)–(2.3.4) we need $d(X, Y)$ to be a random variable. Thus the natural class of spaces appropriate to our investigation is the class of s.m.s.

2.4 UNIVERSALLY MEASURABLE SEPARABLE METRIC SPACES

What follows is an exposition of some basic results regarding universally measurable separable metric spaces (u.m.s.m.s.). As we shall see, the notion of u.m.s.m.s. plays an important role in TPM.

Definition 2.4.1. Let P be a Borel probability measure on a metric space (U, d). We say that P is *tight* if for each $\varepsilon > 0$, there is a compact $K \subseteq U$ with $P(K) \geqslant 1 - \varepsilon$. See Dudley (1989), Section 11.5.

Definition 2.4.2. A s.m.s. (U, d) is *universally measurable* (u.m.) if every Borel probability measure on U is tight.

Definition 2.4.3. A s.m.s. (U, d) is *Polish* if it is topologically complete (i.e. there is a topologically equivalent metric e such that (U, e) is complete). Here the topological equivalence of d and e simply means that for any $x, x_1, x_2, \ldots$ in U

$$d(x_n, x) \to 0 \Leftrightarrow e(x_n, x) \to 0.$$

Theorem 2.4.1. Every Borel subset of a Polish space is u.m.

Proof. See Billingsley (1968) Theorem 1.4, Cohn (1980) Proposition 8.1.10, and Dudley (1989), p. 391.

Remark 2.4.1. Theorem 2.4.1 provides us with many examples of u.m. spaces, but does not exhaust this class. The topological characterization of u.m. s.m.s. is a well known open problem (see Billingsley 1968, Appendix III, p. 234).

In his famous paper on measure theory, Lebesgue (1905) claimed that the projection of any Borel subset of $\mathbb{R}^2$ onto $\mathbb{R}$ is a Borel set. As noted by Souslin and his teacher Lusin (1930), this is in fact not true. As a result of the investigations surrounding this discovery, a theory of such projections (the so-called 'analytic' or 'Souslin' sets) was developed. Although not a Borel set, such a projection was shown to be Lebesgue-measurable, in fact u.m. This train of thought leads to the following definition.

Definition 2.4.4. Let S be a Polish space and suppose that f is a measurable function mapping S onto a separable metric space U. In this case, we say that U is *analytic.*

Theorem 2.4.2. Every analytic s.m.s. is u.m.

Proof. See Cohn's (1980), Theorem 8.6.13, p. 294 and Dudley (1989), Theorem 13.2.6.

Example 2.4.1. Let $\mathbb{Q}$ be the set of rational numbers with the usual topology. Since $\mathbb{Q}$ is a Borel subset of the Polish space R, then $\mathbb{Q}$ is u.m., however, $\mathbb{Q}$ is not itself a Polish space.

Example 2.4.2. In any uncountable Polish space, there are analytic (hence u.m.) non-Borel sets. See Cohn's (1980) Corollary 8.2.17 and Dudley (1989), Proposition 13.2.5.

Example 2.4.3. Let $C[0, 1]$ be the space of continuous functions f: $[0, 1] \to \mathbb{R}$ under the uniform norm. Let $E \subseteq C[0, 1]$ be the set of f which fail to be differentiable at some $t \in [0, 1]$. Then a theorem of Mazurkiewicz (1936) says that E is an analytic, non-Borel subset of $C[0, 1]$. In particular, E is u.m.

Recall again the notion of *Hausdorff metric* $r := r_\rho$ in the space of all subsets of a given metric space (S, ρ)

$$r(A, B) = \max\left\{\sup_{x\in A}\inf_{y\in B}\rho(x, y), \sup_{y\in B}\inf_{x\in A}\rho(x, y)\right\}$$
$$= \inf\{\varepsilon > 0 \colon A^\varepsilon \supseteq B, B^\varepsilon \supseteq A\} \tag{2.4.1}$$

where A^ε is the open ε-neighborhood of A, $A^\varepsilon = \{x \colon d(x, A) < \varepsilon\}$.

As we noticed in the space 2^S of all subsets $A \neq \varnothing$ of S, the Hausdorff distance r is actually only a semidistance. However, in the space $\mathscr{C} = \mathscr{C}(S)$ of all closed non-empty subsets, r is a metric (see Definition 2.2.1) and takes on both finite and infinite values, and if S is a bounded set then r is a finite metric on $\mathscr{C}$.

Theorem 2.4.3. Let (S, ρ) be a metric space, and let $(\mathscr{C}(S), r)$ be the space described above. If (S, ρ) is separable [resp. complete; resp. totally bounded], then $(\mathscr{C}(S), r)$ is separable [resp. complete; resp. totally bounded].

Proof. See Hausdorff (1944), Section 29; Kuratowski (1969), Sections 21 and 23.

Example 2.4.4. Let $S = [0, 1]$ and let ρ be the usual metric on S. Let $\mathscr{R}$ be the set of all finite complex-valued Borel measures m on S such that the Fourier transform

$$\hat{m}(t) = \int_0^1 \exp(\mathrm{i}ut)m(\mathrm{d}u)$$

vanishes at $t = \pm\infty$. Let $\mathscr{M}$ be the class of sets $E \in \mathscr{C}(S)$ such that there is some $m \in \mathscr{R}$ concentrated on E. Then $\mathscr{M}$ is an analytic, non-Borel subset of $(\mathscr{C}(S), r_\rho)$, see Kaufman (1984).

We seek a characterization of u.m. s.m.s. in terms of their Borel structure.

Definition 2.4.5. A measurable space M with σ-algebra $\mathscr{M}$ is *standard* if there is a topology $\mathscr{T}$ on M such that $(M, \mathscr{T})$ is a compact metric space and the Borel σ-algebra generated by $\mathscr{T}$ coincides with $\mathscr{M}$.

A s.m.s. is standard if it is a Borel subset of its completion (see Dudley 1989, p. 347). Obviously, every Borel subset of a Polish space is standard.

Definition 2.4.6. Say that two s.m.s. U and V are called Borel-isomorphic if there is a one–one correspondence f of U onto V such that $B \in \mathscr{B}(U)$ if and only if $f(B) \in \mathscr{B}(V)$.

Theorem 2.4.4. Two standard s.m.s. are Borel-isomorphic if and only if they have the same cardinality.

Proof. See Cohn (1980), Theorem 8.3.6 and Dudley (1989), Theorem 13.1.1.

Theorem 2.4.5. Let U be a separable metric space. The following are equivalent:

(1) U is u.m.

(2) For each Borel probability m on U, there is a standard set $S \in \mathscr{B}(U)$ such that $m(S) = 1$.

Proof. $1 \Rightarrow 2$: Let m be a law on U. Choose compact $K_n \subseteq U$ with $m(K_n) \geqslant 1 - 1/n$. Put $S = \bigcup_{n \geqslant 1} K_n$. Then S is σ-compact, and hence standard. So $m(S) = 1$, as desired.

$2 \Leftarrow 1$: Let m be a law on U. Choose a standard set $S \in \mathscr{B}(U)$ with $m(S) = 1$. Let $\bar{U}$ be the completion of U. Then S is Borel in its completion $\bar{S}$, which is closed in $\bar{U}$. Thus, S is Borel in $\bar{U}$. It follows from Theorem 2.4.1 that

$$1 = m(S) = \sup\{m(K)\colon K \text{ compact}\}.$$

Thus, every law m on U is tight, so that U is u.m. QED

Corollary 2.4.1. Let (U, d) and (V, e) be Borel-isomorphic separable metric spaces. If (U, d) is u.m., then so is (V, e).

Proof. Suppose that m is a law on V. Define a law n on U by $n(A) = m(f(A))$ where $f\colon U \to V$ is a Borel-isomorphism. Since U is u.m. there is a standard set $S \subseteq U$ with $n(S) = 1$. Then $f(S)$ is a standard subset of V with $m(f(S)) = 1$. Thus, by Theorem 2.4.5, V is u.m. QED

The following result, which is in essence due to Blackwell (1956), will be used in an important way later on (cf. the basic theorem of Section 3.2, Theorem 3.2.1).

Theorem 2.4.6. Let U be a u.m. separable metric space and suppose that Pr is a probability measure on U. If $\mathscr{A}$ is a countably generated sub-σ-algebra of $\mathscr{B}(U)$, then there is a real-valued function $P(B|x)$, $B \in \mathscr{B}(U)$, $x \in U$ such that

(1) for each fixed $B \in \mathscr{B}(U)$, the mapping $x \to P(B|x)$ is an $\mathscr{A}$-measurable function on U;

(2) for each fixed $x \in U$, the set function $B \to P(B|x)$ is a law on U;

(3) for each $A \in \mathscr{A}$ and $B \in \mathscr{B}(U)$, we have $\int_A P(B|x)\, \Pr(dx) = \Pr(A \cap B)$;

(4) there is a set $N \in \mathscr{A}$ with $\Pr(N) = 0$ such that $P(B|x) = 1$ whenever $x \in U - N$.

Proof. Choose a sequence $F_1, F_2, \ldots$ of sets in $\mathscr{B}(U)$ which generates $\mathscr{B}(U)$ and is such that a subsequence generates $\mathscr{A}$. We shall prove that there exists a metric e on U such that (U, d) and (U, e) are Borel-isomorphic and for which the sets $F_1, F_2, \ldots$ are *clopen*, i.e., open and closed.

Claim 1. If (U, d) is a s.m.s. and $A_1, A_2, \ldots$ is a sequence of Borel subsets of U, then there is some metric e on U such that

(i) (U, e) is a separable metric space isometric with a closed subset of $\mathbb{R}$;
(ii) $A_1, A_2, \ldots$ are clopen subsets of (U, e);
(iii) (U, d) and (U, e) are Borel-isomorphic (see Definition 2.4.6).

Proof of claim. Let $B_1, B_2, \ldots$ be a countable base for the topology of (U, d). Define sets $C_1, C_2, \ldots$ by $C_{2n-1} = A_n$ and $C_{2n} = B_n$ $(n = 1, 2, \ldots)$ and $f\colon U \to \mathbb{R}$ by $f(x) = \sum_{n=1}^{\infty} 2I_{C_n}(x)/3^n$. Then f is a Borel-isomorphism of (U, d) onto $f(U) \subseteq K$, where K is the Cantor set,

$$K := \left\{ \sum_{n=1}^{\infty} \alpha_n/3^n : \alpha_n\text{s take value 0 or 2} \right\}.$$

Define the metric e by $e(x, y) = |f(x) - f(y)|$, so that (U, e) is isometric with $f(U) \subseteq K$. Then $A_n = f^{-1}\{x \in K;\ x(n) = 2\}$, where $x(n)$ is the nth digit in the ternary expansion of $x \in K$. Thus, A_n is clopen in (U, e), as required.

Now (U, e) is (Corollary 2.4.1) u.m., so there are compact sets $K_1 \subseteq K_2 \subseteq \ldots$ with $\Pr(K_n) \to 1$. Let $\mathscr{G}_1$ and $\mathscr{G}_2$ be the (countable) algebras generated by the sequences $F_1, F_2, \ldots$ and $F_1, F_2, \ldots, K_1, K_2, \ldots$, respectively. Then define $P_1(B \mid x)$ so that (1) and (3) are satisfied for $B \in \mathscr{G}_2$. Since $\mathscr{G}_2$ is countable, there is some set $N \in \mathscr{A}$ with $\Pr(N) = 0$ and such that for $x \in N$,

(a) $P_1(\cdot \mid x)$ is a finitely additive probability on $\mathscr{G}_2$;
(b) $P_1(A \mid x) = 1$ for $A \in \mathscr{A} \cap \mathscr{G}_2$ and $x \in A$;
(c) $P_1(K_n \mid x) \to 1$ as $n \to \infty$.

Claim 2. For $x \in N$, the set function $B \to P_1(B \mid x)$ is countably additive on $\mathscr{G}_1$.

Proof of claim: Suppose that $H_1, H_2, \ldots$ are disjoint sets in $\mathscr{G}_1$ whose union is U. Since the H_n are clopen and the K_n are compact in (U, e), there is, for each n, some $M = M(n)$ such that $K_n \subseteq H_1 \cup H_2 \cup \cdots \cup H_M$. Finite additivity of $P_1(x, \cdot)$ on $\mathscr{G}_2$ yields, for $x \notin N$, $P_1(K_n \mid x) \leqslant \sum_{i=1}^{M} P_1(H_i \mid x) \leqslant \sum_{i=1}^{\infty} P_1(H_i \mid x)$. Let $n \to \infty$ and apply (c) to obtain $\sum_{i=1}^{\infty} P_1(H_i \mid x) = 1$, as required.

In view of the claim, for each $x \in N$, we define $B \to P(B \mid x)$ as the unique countably additive extension of P_1 from $\mathscr{G}_1$ to $\mathscr{B}(U)$. For $x \in N$, put $P(B \mid x) = \Pr(B)$. Clearly, (2) holds. Now the class of sets in $\mathscr{B}(U)$ for which (1) and (3) hold is a monotone class containing $\mathscr{G}_1$, and so coincides with $\mathscr{B}(U)$.

Claim 3. Condition (4) holds.

Proof of claim. Suppose that $A \in \mathscr{A}$ and $x \in A - N$. Let A_0 be the $\mathscr{A}$-atom containing x. Then $A_0 \subseteq A$ and there is a sequence $A_1, A_2, \ldots$ in $\mathscr{G}_1$ such that

$A_0 = A_1 \cap A_2 \cap \cdots$. From (b), $P(A_n|x) = 1$ for $n \geqslant 1$, so that $P(A_0|x) = 1$, as desired. QED

Corollary 2.4.2. Let U and V be u.m. s.m.s. and let Pr be a law on $U \times V$. Then there is a function $P: \mathscr{B}(V) \times U \to \mathbb{R}$ such that

(1) for each fixed $B \in \mathscr{B}(V)$, the mapping $x \to P(B\,|\,x)$ is measurable on U;
(2) for each fixed $x \in U$, the set function $B \to P(B|x)$ is a law on V;
(3) for each $A \in \mathscr{B}(U)$ and $B \in \mathscr{B}(V)$, we have

$$\int_A P(B|x)P_1(\mathrm{d}x) = \Pr(A \cap B)$$

where P_1 is the marginal of Pr on U.

Proof. Apply the preceding theorem with $\mathscr{A}$ the σ-algebra of rectangles $A \times U$ for $A \in \mathscr{B}(U)$. QED

2.5 THE EQUIVALENCE OF THE NOTIONS OF p. (SEMI-)DISTANCE ON $\mathscr{P}_2$ AND ON $\mathfrak{X}$

As we have seen in Section 2.3, every p. (semi-)distance on $\mathscr{P}_2$ induces (by restriction) a p. (semi-)distance on $\mathfrak{X}$. It remains to be seen whether every p. (semi-)distance on $\mathfrak{X}$ arises in this way. This will certainly be the case whenever

$$\mathscr{L}\mathfrak{X}_2(U, (\Omega, \mathscr{A}, \Pr)) = \mathscr{P}_2(U). \tag{2.5.1}$$

Note that the left member depends not only on the structure of (U, d) but also on the underlying probability space.

In this section we will prove the following facts.

(i) There is some probability space $(\Omega, \mathscr{A}, \Pr)$ such that (2.5.1) holds for every separable metric space U.

(ii) If U is a separable metric space, then (2.5.1) holds for every non-atomic probability space $(\Omega, \mathscr{A}, \Pr)$ if and only if U is universally measurable.

We need a few preliminaries.

Definition 2.5.1 (see Loeve 1963, p. 99; Dudley 1989, p. 82). If $(\Omega, \mathscr{A}, \Pr)$ is a probability space, we say that $A \in \mathscr{A}$ is an *atom* if $\Pr(A) > 0$ and $\Pr(B) = 0$ or $\Pr(A)$ for each measurable $B \subseteq A$. A probability space is *non-atomic* if it has no atoms.

Lemma 2.5.1 (Berkes and Phillip 1979). Let v be a law on a complete s.m.s. (U, d) and suppose that $(\Omega, \mathscr{A}, \Pr)$ is a non-atomic probability space. Then there is a U-valued random variable X with distribution $\mathscr{L}(X) = v$.

Proof. Denote by d^* the following metric on U^2: $d^*(x, y) := d(x_1, x_2) + d(y_1, y_2)$ for $x = (x_1, y_1)$ and $y = (x_2, y_2)$. For each k, there is a partition of U^2 comprising non-empty Borel sets $\{A_{ik}: i = 1, 2, \ldots\}$ with $\operatorname{diam}(A_{ik}) < 1/k$ and such that A_{ik} is a subset of some $A_{j,k-1}$.

Since $(\Omega, \mathscr{A}, \Pr)$ is non-atomic, we see that for each $\mathscr{C} \in \mathscr{A}$ and for each sequence p_i of non-negative numbers such that $p_1 + p_2 + \cdots = \Pr(\mathscr{C})$, there exists a partitioning $\mathscr{C}_1, \mathscr{C}_2, \ldots$ of $\mathscr{C}$ such that $\Pr(\mathscr{C}_i) = p_i$, $i = 1, 2, \ldots$ (see e.g. Loeve 1963, p. 99).

Therefore, there exist partitions $\{B_{ik}: i = 1, 2, \ldots\} \subseteq \mathscr{A}$, $k = 1, 2, \ldots$ such that $B_{ik} \subseteq B_{jk-1}$ for some $j = j(i)$ and $\Pr(B_{ik}) = v(A_{ik})$ for all i, k. For each pair (i, j), let us pick a point $x_{ik} \in A_{ik}$ and define U^2-valued $X_k(\omega) = x_{ik}$ for $\omega \in B_{ik}$. Then $d^*(X_{k+m}(\omega), X_k(\omega)) < 1/k$, $m = 1, 2, \ldots$ and since (U^2, d^*) is a complete space, then there exists the limit $X(\omega) = \lim_{k\to\infty} X_k(\omega)$. Thus

$$d^*(X(\omega), X_k(\omega)) \leqslant \lim_{m\to\infty} [d^*(X_{k+m}(\omega), X(\omega)) + d^*(X_{k+m}(\omega), X_k(\omega))] \leqslant \frac{1}{k}.$$

Let $P_k := \Pr_{X_k}$ and $P^* := \Pr_X$. Further, our aim is to show that $P^* = v$. For each closed subset $A \subseteq U$

$$P_k(A) = \Pr(X_k \in A) \leqslant \Pr(X \in A^{1/k}) = P^*(A^{1/k}) \leqslant P_k(A^{2/k}) \tag{2.5.2}$$

where $A^{1/k}$ is the open $1/k$-neighborhood of A. On the other hand,

$$\begin{aligned} P_k(A) &= \sum \{P_k(x_{ik}): x_{ik} \in A\} = \sum \{\Pr(B_{ik}): x_{ik} \in A\} \\ &= \sum \{v(A_{ik}): x_{ik} \in A\} \leqslant \sum \{v(A_{ik} \cap A^{1/k}): x_{ik} \in A\} \\ &\leqslant v(A^{1/k}) \leqslant \sum \{v(A_{ik}): x_{ik} \in A^{2/k}\} \leqslant P_k(A^{2/k}). \end{aligned} \tag{2.5.3}$$

Further, we can estimate the value $P_k(A^{2/k})$ in the same way as in (2.5.2) and (2.5.3) and thus, we get the inequalities

$$P^*(A^{1/k}) \leqslant P_k(A^{2/k}) \leqslant P^*(A^{2/k}) \tag{2.5.4}$$

$$v(A^{1/k}) \leqslant P_k(A^{2/k}) \leqslant v(A^{3/k}). \tag{2.5.5}$$

Since $v(A^{1/k})$ tends to $v(A)$ with $k \to \infty$ for each closed set A and analogously $P^*(A^{1/k}) \to P^*(A)$ as $k \to \infty$, then by (2.5.4) and (2.5.5) we obtain the equalities

$$P^*(A) = \lim_{k\to\infty} P_k(A^{2/k}) = v(A)$$

for each closed A and hence, $P^* = v$. QED

Theorem 2.5.1. There is a probability space $(\Omega, \mathscr{A}, \Pr)$ such that for every separable metric space U and every Borel probability μ on U, there is a random variable $X: \Omega \to U$ with $\mathscr{L}(X) = \mu$.

Proof. Define $(\Omega, \mathscr{A}, \Pr)$ as the measure-theoretic (von Neumann) product (see Hewitt and Stromberg (1965), Theorems 22.7 and 22.8, pp. 432–433) of the probability spaces $(C, \mathscr{B}(C), v)$, where C is some non-empty subset of $\mathbb{R}$ with Borel σ-algebra $\mathscr{B}(C)$, and v is some Borel probability on $(C, \mathscr{B}(C))$.

Now, given a separable metric space U, there is some set $C \subseteq \mathbb{R}$ Borel-isomorphic with U (cf. Claim 1 in Theorem 2.4.6). Let $f: C \to U$ supply the isomorphism. If μ is a Borel probability on U, let v be a probability on C such that $f(v) := vf^{-1} = \mu$. Define $X: \Omega \to U$ as $X = f \circ \pi$, where $\pi: \Omega \to C$ is a projection onto the factor $(C, \mathscr{B}(C), v)$. Then $\mathscr{L}(X) = \mu$, as desired.

Remark 2.5.1. The result above establishes the claim (i) made at the beginning of the section. It provides one way of ensuring (2.5.1): simply insist that all r.v.s be defined on a 'super-probability space' as in Theorem 2.5.1. We make this assumption throughout the sequel.

The next theorem extends the Berkes and Phillips's Lemma 2.5.1 to the case of u.m. s.m.s. U.

Theorem 2.5.2. Let U be a separable metric space. The following are equivalent.

(1) U is u.m.

(2) If $(\Omega, \mathscr{A}, \Pr)$ is a non-atomic probability space, then for every Borel probability P on U, there is a random variable $X: \Omega \to U$ with law $\mathscr{L}(X) = P$.

Proof. $1 \Rightarrow 2$: Since U is u.m. there is some standard set $S \in \mathscr{B}(U)$ with $P(S) = 1$ (Theorem 2.4.5). Now there is a Borel-isomorphism f mapping S onto a Borel subset B of $\mathbb{R}$ (Theorem 2.4.4). Then $f(P) := P \circ f^{-1}$ is a Borel probability on $\mathbb{R}$. Thus, there is a random variable $g: \Omega \to \mathbb{R}$ with $\mathscr{L}(g) = f(P)$ and $g(\Omega) \subseteq B$ (Lemma 2.5.1 with $(U, d) = (\mathbb{R}, |\cdot|)$). We may assume that $g(\Omega) \subseteq B$ since $\Pr(g^{-1}(B)) = 1$. Define $x: \Omega \to U$ by $x(\omega) = f^{-1}(g(\omega))$. Then $\mathscr{L}(X) = v$, as claimed.

$2 \Rightarrow 1$: Now suppose that v is a Borel probability on U. Consider a random variable $X: \Omega \to U$ on the (non-atomic) probability space $((0, 1), \mathscr{B}(0, 1), \lambda)$ with $\mathscr{L}(X) = v$. Then range (X) is an analytic subset of U with $v^*(\text{range}(X)) = 1$. Since range(X) is u.m. (Theorem 2.4.2), there is some standard set $S \subseteq \text{range}(X)$ with $P(S) = 1$. This follows from Theorem 2.4.5. The same theorem shows that U is u.m. QED

Remark 2.5.2. If U is u.m. s.m.s., we operate under the assumption that all U-valued r.v.s are defined on a non-atomic probability space. Then (2.5.1) will be valid.

CHAPTER 3

Primary, Simple and Compound p. Distances. Minimal and Maximal Distances and Norms

In this chapter we shall give a more detailed analysis of the notions of p. distance and p. metric in order to provide the first (crude) classification of p. semidistances.

3.1 PRIMARY DISTANCES AND PRIMARY METRICS

Let $h: \mathscr{P}_1 \to \mathbb{R}^J$ be a mapping, where $\mathscr{P}_1 = \mathscr{P}_1(U)$ is the set of Borel probability measures (laws) for some s.m.s. (U, d) and J is some index set. This function h induces a partition of $\mathscr{P}_2 = \mathscr{P}_2(U)$ (the set of laws on U^2) into equivalence classes for the relation

$$P \overset{h}{\sim} Q \Leftrightarrow h(P_1) = h(Q_1) \text{ and } h(P_2) = h(Q_2) \quad P_i := T_i P,\ Q_i := T_i Q \quad (3.1.1)$$

where P_i and Q_i ($i = 1, 2$) are the ith marginals of P and Q, respectively. Let μ be a p. semidistance on $\mathscr{P}_2$ with parameter $\mathbb{K}_\mu$ (Definition 2.3.1), such that μ is constant on the equivalence classes of $\sim$, i.e.

$$P \overset{h}{\sim} Q \Leftrightarrow \mu(P) = \mu(Q). \quad (3.1.2)$$

Definition 3.1.1. If the p. semidistance $\mu = \mu_h$ satisfies Relation (3.1.2), then we call μ a *primary distance* (*with parameter* $\mathbb{K}_\mu$). If $\mathbb{K}_\mu = 1$ and μ assumes only finite values, we say that μ is a *primary metric.*

Obviously, by Relation (3.1.2), any primary distance is completely determined by the pair of marginal characteristics (hP_1, hP_2). In case of primary distance μ we shall write $\mu(hP_1, hP_2) := \mu(P)$ and hence μ may be viewed as a distance in the image space $h(\mathscr{P}_1) \subseteq R^J$, i.e., the following metric properties hold

ID$^{(1)}$ $hP_1 = hP_2 \Leftrightarrow \mu(hP_1, hP_2) = 0$
SYM$^{(1)}$ $\mu(hP_1, hP_2) = \mu(hP_2, hP_1)$
TI$^{(1)}$ If the following marginal conditions are fulfilled

$$a = h(T_1 P^{(1)}) = h(T_1 P^{(2)}) \quad b = h(T_2 P^{(2)}) = h(T_1 P^{(3)}) \quad c = h(T_2 P^{(1)}) = h(T_2 P^{(3)})$$

for some law $P^{(1)}, P^{(2)}, P^{(3)} \in \mathscr{P}_2$ then $\mu(a, c) \leqslant \mathbb{K}_\mu[\mu(a, b) + \mu(b, c)]$.

The notion of primary semidistance μ_h becomes easier to interpret assuming that a probability space $(\Omega, \mathscr{A}, \Pr)$ with property (2.5.1) is fixed (see Remark 2.5.1). In this case μ_h is a usual distance (see Definition 2.2.1) in the space

$$h(\mathfrak{X}) := \{hX := h \Pr_X, \text{ where } X \in \mathfrak{X}(U)\} \tag{3.1.3}$$

and thus, the metric properties of $\mu := \mu_h$ take the simplest form (cf. Definition 2.2.2):

ID$^{(1^*)}$ $hX = hY \Leftrightarrow \mu(hX, hY) = 0$
SYM$^{(2^*)}$ $\mu(hX, hY) = \mu(hY, hX)$,
TI$^{(3^*)}$ $\mu(hX, hZ) \leqslant \mathbb{K}_\mu[\mu(hX, hY) + \mu(hY, hZ)]$.

Further, we shall consider several examples of primary distances and metrics.

Example 3.1.1. Primary minimal distances. Each p. semidistance μ and each mapping $h\mathscr{P}_1 \to \mathbb{R}^J$ determine a functional $\tilde{\mu}_h \to \tilde{\mu}_h$: $h(\mathscr{P}_1) \times h(\mathscr{P}_1) \to [0, \infty]$ defined by the following equality

$$\tilde{\mu}_h(\bar{a}_1, \bar{a}_2) := \inf\{\mu(P): hP_i \equiv \bar{a}_i, i = 1, 2\} \tag{3.1.4}$$

(where P_i are the marginals of P) for any pair $(\bar{a}_1, \bar{a}_2) \in h(\mathscr{P}_1) \times h(\mathscr{P}_1)$.

Further, we shall prove (see Chapter 5) that $\tilde{\mu}_h$ is a primary distance for different special functions h and spaces U.

Definition 3.1.2. The functional $\tilde{\mu}_h$ is called a *primary h-minimal distance* with respect to the p. semidistance μ.

Open problem 3.1.1. In general it is not true that the metric properties of a p. distance μ imply that $\tilde{\mu}$ is a distance. The following two examples illustrate this fact (see further Chapter 9):

(a) Let $U = \mathbb{R}$, $d(x, y) = |x - y|$. Consider the p. metric

$$\mu(X, Y) = \mathscr{L}_0(X, Y) = \Pr(X \neq Y) \quad X, Y \in \mathfrak{X}(\mathbb{R})$$

and the mapping $h: \mathfrak{X}(\mathbb{R}) \to [0, \infty]$ given by $hX = \mathbb{E}|X|$. Then (see further Section 9.1)

$$\tilde{\mu}_h(a, b) = \inf\{\Pr(X \neq Y): \mathbb{E}|X| = a, \mathbb{E}|Y| = b\} = 0$$

for all $a \geqslant 0$ and $b \geqslant 0$. Hence in this case the metric properties of μ imply only semimetric properties for $\tilde{\mu}_h$.

(b) Now let μ be defined as in (a) but $h\colon \mathfrak{X}(R) \to [0, \infty] \times [0, \infty]$ be defined by $hX = (\mathbb{E}|X|, \mathbb{E}X^2)$. Then

$$\mu_h((a_1, a_2), (b_1, b_2)) = \inf\{\Pr(X \neq Y)\colon \mathbb{E}|X| = a_1, \mathbb{E}X^2 = a_2, \mathbb{E}|Y| = b_1, \mathbb{E}Y^2 = b_2\} \quad (3.1.5)$$

where $\tilde{\mu}_h$ is not even p. semidistance since the triangle inequality $TI^{(3^*)}$ is not valid.

With respect to this, the following open problem arises: *under which condition on the space U, p. distance μ on $\mathfrak{X}(U)$ and transformation $h\colon \mathfrak{X}(U) \to R^J$ the primary h-minimal distance $\tilde{\mu}_h$ is a primary p. distance in $h(\mathfrak{X})$?*

As we shall see later on (Section 9.1), all further examples 3.12 to 3.15 of primary distances are special cases of primary h-minimal distances.

Example 3.1.2. Let $H \in \mathscr{H}$ (see Example 2.2.1) and $\bar{0}$ be a fixed point of a s.m.s. (U, d). For each $P \in \mathscr{P}_2$ with marginals $P_i = T_iP$, let m_1P, m_2P denote the '*marginal moments of order $p > 0$*',

$$m_iP := m_i^{(p)}P := \left(\int_U d^p(x, \bar{0})P_i(\mathrm{d}x)\right)^{p'} \quad p > 0 \quad p' := \min(1, 1/p).$$

Then

$$\mathscr{M}_{H,p}(P) := \mathscr{M}_{H,p}(m_1P, m_2P) := H(|m_1P - m_2P|) \quad (3.1.6)$$

is a primary distance. One can also consider $\mathscr{M}_{H,p}$ as a distance in the space

$$m^{(p)}(\mathscr{P}_1) := \left\{m^{(p)}P := \left(\int_U d^p(x, a)P(\mathrm{d}x)\right)^{p'} < \infty,\ P \in \mathscr{P}(U)\right\} \quad (3.1.7)$$

of moments $m^{(p)}P$ of order $p > 0$. If $H(t) = t$ then

$$\mathscr{M}(P) := \mathscr{M}_{H,1}(P) = \left|\int_U d(x, \bar{0})(\mathrm{P}_1 - P_2)(\mathrm{d}x)\right|$$

is a primary metric in $m^{(p)}(\mathscr{P}_1)$.

Example 3.1.3. Let $g\colon [0, \infty] \to \mathbb{R}$ and $H \in \mathscr{H}$. Then

$$\mathscr{M}(g)_{H,p}(m_1P, m_2P) := H(|g(m_1P) - g(m_2P)|) \quad (3.1.8)$$

is a primary distance in $g \circ m(\mathscr{P}_1)$ and

$$\mathscr{M}(g)(m_1P, m_2P) := |g(m_1P) - g(m_2P)| \quad (3.1.9)$$

is a primary metric.

If U is a Banach space with norm $\|\cdot\|$ then we define the primary distance $\mathcal{M}_{H,p}(g)$ as follows

$$\mathcal{M}_{H,p}(g)(m^{(p)}X, m^{(p)}Y) := H(|g(m^{(p)}X) - g(m^{(p)}Y)|) \tag{3.1.10}$$

where (cf. (2.1.8)) $m^{(p)}X$ is the '*p-th moment (norm) of X*'

$$m^{(p)}X := \{\mathbb{E}\|X\|^p\}^{p'}.$$

By Equation (3.1.9), $\mathcal{M}_{H,p}(g)$ may be viewed as a distance (see Definition 2.2.2) in the space

$$g \circ m(\mathfrak{X}) := \{g \circ m(X) := g(\{\mathbb{E}\|X\|^p\}^{p'}),\ X \in \mathfrak{X}\} \quad p' = \min(1, p^{-1}),\ \mathfrak{X} = \mathfrak{X}(U) \tag{3.1.11}$$

of moments $g \circ m(X)$. If U is the real line $\mathbb{R}$ and $g(t) = H(t) = t (t \geqslant 0)$ then $\mathcal{M}_{H,p}(g)(m^{(p)}X, m^{(p)}Y)$ is the usual deviation between moments $m^{(p)}X$ and $m^{(p)}Y$ (see (2.1.9)).

Example 3.1.4. Let J be an index set (with arbitrary cardinality), g_i $(i \in J)$ be real functions on $[0, \infty]$ and for each $P \in \mathscr{P}(U)$ define the set

$$hP := \{g_i(mP),\ i \in J\} \tag{3.1.12}$$

Further, for each $P \in \mathscr{P}_2(U)$ let us consider hP_1 and hP_2 where P_i's are the marginals of P. Then

$$\boldsymbol{\Omega}(hP_1, hP_2) = \begin{cases} 0 & \text{if } hP_1 \equiv hP_2 \\ 1 & \text{otherwise} \end{cases} \tag{3.1.13}$$

is a primary metric.

Example 3.1.5. Let U be the n-dimensional Euclidean space $\mathbb{R}^n$, $H \in \mathscr{H}$. Define the '*engineer distance*'

$$\mathbf{EN}(X, Y; H) := H\left(\left|\sum_{i=1}^{n} (\mathbb{E}X_i - \mathbb{E}Y_i)\right|\right) \tag{3.1.14}$$

where $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$ belong to the subset $\tilde{\mathfrak{X}}(\mathbb{R}^n) \subseteq \mathfrak{X}(\mathbb{R}^n)$ of all n-dimensional random vectors that have integrable components. Then $\mathbf{EN}(\cdot, \cdot; H)$ is a p. semidistance in $\tilde{\mathfrak{X}}(\mathbb{R}^n)$. Analogously, the '$L_p$-*engineer metric*'

$$\mathbf{EN}(X, Y; p) := \left[\sum_{i=1}^{n} |\mathbb{E}X_i - \mathbb{E}Y_i|^p\right]^{\min(1, 1/p)}, \quad p > 0 \tag{3.1.15}$$

is a primary metric in $\tilde{\mathfrak{X}}(\mathbb{R}^n)$. In the case $p = 1$ and $n = 1$, the metric $\mathbf{EN}(\cdot, \cdot; p)$ coincides with the engineer metric in $\mathfrak{X}(\mathbb{R})$ (see (2.1.1)).

3.2 SIMPLE DISTANCES AND METRICS; CO-MINIMAL FUNCTIONALS AND MINIMAL NORMS

Clearly, any primary distance $\mu(P)$ ($P \in \mathscr{P}_2$) is completely determined by the pair of marginal distributions $P_i = T_i P$ ($i = 1, 2$), since the equality $P_1 = P_2$ implies $hP_1 = hP_2$ (see Relations (3.1.1), (3.1.2) and Definition 3.1.1). On the other hand, if the mapping h is 'rich enough' then the opposite implication

$$hP_1 = hP_2 \Rightarrow P_1 = P_2$$

takes place. The simplest example of such 'rich' h: $\mathscr{P}(U) \to R^J$ is given by the equalities

$$h(P) := \{P(C),\ C \in \mathscr{C},\ P \in \mathscr{P}(U)\} \tag{3.2.1}$$

where $J \equiv \mathscr{C}$ is the family of all closed non-empty subsets $C \subseteq U$. Another example is

$$h(P) = \left\{Pf := \int_U f \, \mathrm{d}P : f \in C^b(U)\right\} \quad P \in \mathscr{P}(U)$$

where $C^b(U)$ is the set of all bounded continuous functions on U. Keeping in mind these two examples we shall define the notion of 'simple' distance as a particular case of primary distance with h given by Equality (3.2.1).

Definition 3.2.1. The p. semidistance μ is said to be a *simple semidistance in* $\mathscr{P} = \mathscr{P}(U)$, if for each $P \in \mathscr{P}_2$

$$\mu(P) = 0 \Leftarrow T_1 P = T_2 P.$$

If, in addition, μ is a p. semimetric, then μ will be called a *simple semimetric.* If the converse implication ($\Rightarrow$) also holds, we say that μ is *simple distance.* If, in addition, μ is a p. semimetric, then μ will be called a *simple metric.*

Since the values of the simple distance $\mu(P)$ depend only on the pair marginals P_1, P_2 we shall consider μ as a functional on $\mathscr{P}_1 \times \mathscr{P}_2$ and we shall use the notation

$$\mu(P_1, P_2) := \mu(P_1 \times P_2) \qquad (P_1, P_2 \in \mathscr{P}_1)$$

where $P_1 \times P_2$ means the measure product of laws P_1 and P_2. In this case the metric properties of μ take the form (cf. Definition 2.3.1) (for each $P_1, P_2, P_3 \in \mathscr{P}$):

ID$^{(2)}$ $\quad P_1 = P_2 \Leftrightarrow \mu(P_1, P_2) = 0$

SYM$^{(2)}$ $\quad \mu(P_1, P_2) = \mu(P_2, P_1)$

TI$^{(2)}$ $\quad \mu(P_1, P_3) \leqslant \mathbb{K}_\mu(\mu(P_1, P_2) + \mu(P_2, P_3)).$

Hence, the space $\mathscr{P}$ of laws P with a simple distance μ is a distance space (see Definition 2.2.2). Clearly each primary distance is a simple semidistance in $\mathscr{P}$.

The Kolmogorov metric $\boldsymbol{\rho}$ (2.1.2), the Lévy metric $\mathbf{L}$ (2.1.3) and the $\boldsymbol{\theta}_p$-metrics (2.1.4) are simple metrics in $\mathscr{P}(\mathbb{R})$.

Let us consider a few more examples of simple metrics which we shall use later on.

Example 3.2.1. Minimal distances.

Definition 3.2.2. For a given p. semidistance $\hat{\mu}$ on $\mathscr{P}_2$ the functional $\tilde{\mu}$ on $\mathscr{P} \times \mathscr{P}$ defined by the equality

$$\hat{\mu}(P_1, P_2) := \inf\{\mu(P);\ T_i P = P_i,\ i = 1, 2\} \quad P_1, P_2 \in \mathscr{P} \tag{3.2.2}$$

is said to be (simple) *minimal* (w.r.t. μ) *distance.*

As we showed in Section 2.5.1, for a 'rich enough' probability space, the space $\mathscr{P}_2$ of all laws on U^2 coincides with the set of joint distributions $\mathrm{Pr}_{X,Y}$ of U-valued r.v.s. Thus always $\mu(P) = \mu(\mathrm{Pr}_{X,Y})$ for some $X, Y \in \mathfrak{X}(U)$ and therefore Equation (3.2.2) can be rewritten as follows

$$\hat{\mu}(P_1, P_2) = \inf\{\mu(X, Y) \colon \mathrm{Pr}_X = P_1,\ \mathrm{Pr}_Y = P_2\}.$$

The last is the Zolotarev definition of a minimal metric (Zolotarev, 1976b).

In the next theorem we shall consider the conditions on U that guarantee $\hat{\mu}$ to be a simple metric.

We use the notation $\xrightarrow{\mathrm{w}}$ 'to mean weak convergence of laws' (cf., Billingsley, 1968).

Theorem 3.2.1. Let U be a u.m. s.m.s. (see Definition 2.4.2) and let μ be a p. semidistance with parameter $\mathbb{K}_\mu$. Then $\hat{\mu}$ is a simple semidistance with parameter $\mathbb{K}_{\hat{\mu}} = \mathbb{K}_\mu$. Moreover, if μ is a p. distance satisfying the following 'continuity' condition

$$\left.\begin{array}{l} P^{(n)} \in \mathscr{P}^2 \quad P^{(n)} \xrightarrow{\mathrm{w}} P \in \mathscr{P}^2 \\ \mu(P^{(n)}) \to 0 \end{array}\right\} \Rightarrow \mu(P) = 0.$$

Then $\hat{\mu}$ is a simple distance with parameter $\mathbb{K}_{\hat{\mu}} = \mathbb{K}_\mu$.

Remark 3.2.1. The continuity condition is not restrictive; in fact, all p. distances we are going to use satisfy this condition.

Remark 3.2.2. Clearly, if μ is a p. semimetric then, by the above theorem, $\hat{\mu}$ is a simple semimetric.

Proof. $\mathbf{ID}^{(2)}$: If $P_1 \in \mathscr{P}_1$ then we let $X \in \mathfrak{X}(U)$ have the distribution P_1. Then, by $\mathbf{ID}^{(*)}$ (Definition 2.3.2),

$$\hat{\mu}(P_1, P_1) \leqslant \mu(\mathrm{Pr}_{(X,X)}) = 0.$$

Suppose now that μ is a p. distance and the continuity condition holds. If $\hat{\mu}(P_1, P_2) = 0$ then there exists a sequence of laws $P^{(n)} \in \mathscr{P}_2$ with fixed marginals $T_i P^{(n)} = P_i$ $(i = 1, 2)$ such that $\mu(P^{(n)}) \to 0$ as $n \to \infty$. Since P_i is a tight measure then the sequence $\{P^n, n \geqslant 1\}$ is uniformly tight, i.e., for any $\varepsilon > 0$ there exists a compact $K_\varepsilon \subseteq U^2$ such that $P^{(n)}(K_\varepsilon) \geqslant 1 - \varepsilon$ for all $n \geqslant 1$ (cf. Dudley, 1989, Section 11.5). Using Prokhorov compactness criteria (see, for instance, Billingsley 1968, Theorem 6.1) we choose a subsequence $P^{(n')}$ that weakly tends to a law $P \in \mathscr{P}_2$, hence, $T_i P = P_i$ and $\mu(P) = 0$. Since μ is a p. distance P is concentrated on the diagonal $x = y$ and thus $P_1 = P_2$ as desired.

SYM$^{(2)}$: Obvious.

TI$^{(2)}$: Let P_1, P_2, $P_3 \in \mathscr{P}_1$. For any $\varepsilon > 0$ define a law $P_{12} \in \mathscr{P}_2$ with marginals $T_i P_{12} = P_i$ $(i = 1, 2)$ and a law $P_{23} \in \mathscr{P}_2$ with $T_i P_{23} = P_{i+1}$ $(i = 1, 2)$ such that $\hat{\mu}(P_1, P_2) \geqslant \mu(P_{12}) - \varepsilon$ and $\hat{\mu}(P_2, P_3) \geqslant \mu(P_{23}) - \varepsilon$. Since U is a u.m. s.m.s. then there exist Markov kernels $P'(A/z)$ and $P''(A/z)$ defined by the equalities

$$P_{12}(A_1 \times A_2) := \int_{A_2} P'(A_1/z) P_2(\mathrm{d}z) \tag{3.2.3}$$

$$P_{23}(A_2 \times A_3) := \int_{A_2} P''(A_3/z) P_2(\mathrm{d}z) \tag{3.2.4}$$

for all A_1, A_2, $A_3 \in \mathscr{B}_1$ (see Corollary 2.4.2). Then define a set function Q on the algebra $\mathscr{A}$ of finite unions of Borel rectangles $A_1 \times A_2 \times A_3$ by the equation

$$Q(A_1 \times A_2 \times A_3) := \int_{A_2} P'(A_1/z) P''(A_3/z) P_2(\mathrm{d}z). \tag{3.2.5}$$

It is easily checked that Q is countably additive on $\mathscr{A}$ and therefore extends to a law on U^3. We use 'Q' to represent this extension also. The law Q has the projections $T_{12}Q = P_{12}, T_{23}Q = P_{23}$. Since μ is a p. semidistance with parameter $\mathbb{K} = \mathbb{K}_\mu$ we have

$$\begin{aligned} \mu(P_1, P_3) \leqslant \mu(T_{13}Q) &\leqslant \mathbb{K}[\mu(P_{12}) + \mu(P_{13})] \\ &\leqslant \mathbb{K}[\hat{\mu}(P_1, P_2) + \hat{\mu}(P_2, P_3)] + 2\mathbb{K}\varepsilon. \end{aligned}$$

Letting $\varepsilon \to 0$ we complete the proof of **TI**$^{(2)}$. QED

As will be shown later (Part 2), all simple distances in the next examples are actually simple minimal $\hat{\mu}$ distances with respect to p. distances μ that will be introduced in Section 3.3 (see further examples 3.3.1 to 3.3.3).

Example 3.2.2. (Kantorovich metric and Kantorovich distance). In Section 2.1, we introduced the Kantorovich metric $\boldsymbol{\kappa}$ and its 'dual' representation

$$\boldsymbol{\kappa}(P_1, P_2) = \int_{-\infty}^{\infty} |F_1(x) - F_2(x)|\mathrm{d}x$$
$$= \sup\left\{\left|\int_{\mathbb{R}} f\,\mathrm{d}(P_1 - P_2)\right| : f\colon \mathbb{R} \to \mathbb{R},\ f' \text{ exists a.e. and } |f'| \leq 1 \text{ a.e.}\right\}$$

where P_is are laws on $\mathbb{R}$ with d.f.s F_i and finite first absolute moment. From the above representation it also follows that

$$\boldsymbol{\kappa}(P_1, P_2) = \sup\left\{\left|\int_{\mathbb{R}} f\,\mathrm{d}(P_1 - P_2)\right| : f\colon \mathbb{R} \to \mathbb{R},\ f \text{ is } (1, 1)\text{-Lipschitz, i.e., } |f(x) - f(y)| \leqslant |x - y|\ \forall x, y \in \mathbb{R}\right\}.$$

In this example we shall extend the definition of the above simple p. metric of the set $\mathscr{P}(U)$ of all laws on a s.m.s. (U, d). For any $\alpha \in (0, \infty)$ and $\beta \in [0, 1]$ define the Lipschitz functions class

$$\mathrm{Lip}_{\alpha\beta}(U) := \{f\colon U \to \mathbb{R}\colon |f(x) - f(y)| \leqslant \alpha\, \mathrm{d}^{\beta}(x, y)\ \forall x, y \in U\} \tag{3.2.6}$$

with the convention

$$d^0(x, y) := \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases} \tag{3.2.7}$$

Denote the set of all bounded functions $f \in \mathrm{Lip}_{\alpha\beta}(U)$ by $\mathrm{Lip}^b_{\alpha\beta}(U)$. Let $\mathscr{G}_H(U)$ be the class of all pairs (f, g) of functions that belong to the set

$$\mathrm{Lip}^b(U) := \bigcup_{\alpha > 0} \mathrm{Lip}_{\alpha,1}(U) \tag{3.2.8}$$

and satisfy the inequality

$$f(x) + g(y) \leqslant H(d(x, y))\ \forall x, y \in U \tag{3.2.9}$$

where H is a convex function from $\mathscr{H}$. Recall that $H \in \mathscr{H}$ if H is a nondecreasing continuous function from $[0, \infty)$ onto $[0, \infty)$, vanishes at the origin and $K_H := \sup_{t>0} H(2t)/H(t) < \infty$. For any two laws P_1 and P_2 on a s.m.s. (U, d) define

$$\ell_H(P_1, P_2) := \sup\left\{\int_U f\,\mathrm{d}P_1 + \int_U g\ \mathrm{d}P_2\colon (f, g) \in \mathscr{G}_H(U)\right\}. \tag{3.2.10}$$

We shall prove further that ℓ_H is a simple distance with $\mathbb{K}_{\ell_H} = K_H$ in the space of all laws P with finite 'H-moment', $\int H(\mathrm{d}(x, a))P(\mathrm{d}x) < \infty$. The proof is based on the representation of ℓ_H as a minimal distance $\ell_H = \hat{\mathscr{L}}_H$ (Corollary 5.2.2) with respect to a p. distance (with $\mathbb{K}_{\mathscr{L}_H} = K_H)\ell_H(P) =$

$\int_{U^2} H(d(x, y))P(dx, dy)$ and then an appeal to Theorem 3.2.1 proves that ℓ_H is a simple p. distance if (U, d) is a universally measurable s.m.s. In the case $H(t) = t^p$ $(1 < p < \infty)$ define

$$\ell_p(P_1, P_2) := \ell_H(P_1, P_2)^{1/p} \quad 1 < p < \infty. \tag{3.2.11}$$

In addition, for $p \in [0, 1]$ and $p = \infty$, denote

$$\ell_p(P_1, P_2) := \sup\left\{\left|\int_U f\, d(P_1 - P_2)\right| : f \in \mathrm{Lip}^b_{1,p}(U)\right\} \quad p \in (0, 1] \quad P_1, P_2 \in \mathscr{P}(U) \tag{3.2.12}$$

$$\ell_0(P_1, P_2) := \left\{\left|\int_U f\, d(P_1 - P_2)\right| : f \in \mathrm{Lip}_{1,0}(U)\right\} \tag{3.2.13}$$

$$= \boldsymbol{\sigma}(P_1, P_2) := \sup_{A \in \mathscr{B}_1} |P_1(A) - P_2(A)|$$

$$\ell_\infty(P_1, P_2) := \inf\{\varepsilon > 0 : P_1(A) \leqslant P_2(A^\varepsilon)\ \forall A \in \mathscr{B}_1\} \tag{3.2.14}$$

where, as above, $\mathscr{B}_1 = \mathscr{B}(U)$ is the Borel σ-algebra on a s.m.s. (U, d), and $A^\varepsilon := \{x : d(x, A) < \varepsilon\}$.

For any $0 \leqslant p \leqslant 1$, $p = \infty$, ℓ_p is a simple metric in $\mathscr{P}(U)$ which follows immediately from the definition. To prove that ℓ_∞ is a p. metric (taking possibly infinite values) one can use the equality

$$\sup_{A \in \mathscr{B}_1} [P_1(A) - P_2(A^\varepsilon)] = \sup_{A \in \mathscr{B}_1} [P_2(A) - P_1(A^\varepsilon)].$$

The equality $\ell_0 = \boldsymbol{\sigma}$ in Equation (3.2.13) follows from the fact that both metrics are minimal with respect to one and the same p. distance $\mathscr{L}_0(P) = P((x, y) : x \neq y)$, see further Corollary 6.1.1 and Corollary 7.4.2. We shall prove also (see Corollary 7.3.2) that $\ell_H = \hat{\mathscr{L}}_H$, as a minimal distance w.r.t. ℓ_H defined above, admits the Birnbaum–Orlicz representation (see Example 2.2.2)

$$\ell_H(P_1, P_2) = \ell_H(F_1, F_2) := \int_0^1 H(|F_1^{-1}(t) - F_2^{-1}(t)|)\, dt \tag{3.2.15}$$

in the case of $U = \mathbb{R}$ and $d(x, y) = |x - y|$. In Equation (3.2.15),

$$F_i^{-1}(t) := \sup\{x : F_i(x) \leqslant t\} \tag{3.2.16}$$

is the (generalized) *inverse* of the d.f. F_i determined by P_i $(i = 1, 2)$. Letting $H(t) = t$ we claim that

$$\ell_1(P_1, P_2) = \int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|\, dt$$

$$= \boldsymbol{\kappa}(P_1, P_2) := \int_{-\infty}^{\infty} |F_1(x) - F_2(x)|\, dx \quad P_i \in \mathscr{P}(\mathbb{R}) \quad i = 1, 2. \tag{3.2.17}$$

Remark 3.2.3. Here and in the sequel, for any simple semidistance μ on $\mathscr{P}(\mathbb{R}^n)$ we shall use the following notations interchangeably

$$\mu = \mu(P_1, P_2)\ \forall P_1, P_2 \in \mathscr{P}(\mathbb{R}^n)$$
$$\mu = \mu(X_1, X_2) := \mu(\Pr_{X_1}, \Pr_{X_2})\ \forall X_1, X_2 \in \mathfrak{X}(\mathbb{R}^n)$$
$$\mu = \mu(F_1, F_2) := \mu(P_1, P_2)\ \forall F_1, F_2 \in \mathscr{F}(\mathbb{R}^n)$$

where $\Pr_{X_i}$ is the distribution of X_i, F_i is the d.f. of P_i and $\mathscr{F}(\mathbb{R}^n)$ stands for the class of d.f.s on $\mathbb{R}^n$.

The ℓ_1-metric (3.2.17) is known as the *average metric* in $\mathscr{F}(\mathbb{R})$ as well as the *first difference pseudomoment*, and it is also denoted by $\boldsymbol{\kappa}$ (see Zolotarev, 1976b). A great contribution in the investigation of ℓ_1-metric properties was made by Kantorovich (1942, 1948), Kantorovich and Akilov (1984, 4, Chap. VIII). That is the reason the metric ℓ_1 is called the *Kantorovich metric.* Considering ℓ_H as a generalization ℓ_1, we shall call ℓ_H the *Kantorovich distance.*

Example 3.2.3 (Prokhorov metric and Prokhorov distance). Prokhorov (1956) introduced his famous metric

$$\pi(P_1, P_2) := \inf\{\varepsilon > 0\colon P_1(C) \leqslant P_2(C^\varepsilon) + \varepsilon,\ P_2(C) \leqslant P_1(C^\varepsilon) + \varepsilon\ \forall C \in \mathscr{C}\} \tag{3.2.18}$$

where $\mathscr{C} := \mathscr{C}(U)$ is the set of all nonempty closed subsets of a Polish space U and

$$C^\varepsilon := \{x\colon d(x, C) < \varepsilon\}. \tag{3.2.19}$$

The metric π admits the following representations: for any laws P_1 and P_2 on a s.m.s. (U, d)

$$\begin{aligned}\pi(P_1, P_2) &= \inf\{\varepsilon > 0\colon P_1(C) \leqslant P_2(C^\varepsilon) + \varepsilon \text{ for any } C \in \mathscr{C}\}\\ &= \inf\{\varepsilon > 0\colon P_1(C) \leqslant P_2(C^{\varepsilon]}) + \varepsilon, \text{ for any } C \in \mathscr{C}\}\\ &= \inf\{\varepsilon > 0\colon P_1(A) \leqslant P_2(A^\varepsilon) + \varepsilon, \text{ for any } A \in \mathscr{B}_1\}\end{aligned} \quad (3.2.20$$

where

$$C^{\varepsilon]} := \{x\colon d(x, C) \leqslant \varepsilon\} \tag{3.2.21}$$

is the ε-closed neighborhood of C (see, for example, Theorem 8.1, Dudley, 1976).

Let us introduce a *parametric version of the Prokhorov metric*

$$\pi_\lambda(P_1, P_2) := \inf\{\varepsilon > 0\colon P_1(C) \leqslant P_2(C^{\lambda\varepsilon}) + \varepsilon \text{ for any } C \in \mathscr{C}\}. \tag{3.2.22}$$

The next lemma gives the main relationship between the Prokhorov-type metrics and the metrics ℓ_0 and ℓ_∞ defined by Equalities (3.2.13) and (3.2.14).

Lemma 3.2.1. For any $P_1, P_2 \in \mathscr{P}(U)$

$$\lim_{\lambda \to 0} \pi_\lambda(P_1, P_2) = \boldsymbol{\sigma}(P_1, P_2) = \ell_0(P_1, P_2) \tag{3.2.23}$$

$$\lim_{\lambda \to 0} \lambda \pi_\lambda(P_1, P_2) = \ell_\infty(P_1, P_2).$$

Proof. For any fixed $\lambda > 0$ the function $A_\varepsilon(\lambda) := \sup\{P_1(C) - P_2(C^{\lambda\varepsilon}): C \in \mathscr{C}\}$, $\lambda \geqslant 0$ is non-increasing on $\varepsilon > 0$, hence

$$\pi_\lambda(P_1, P_2) = \inf\{\varepsilon > 0: A_\varepsilon(\lambda) \leqslant \varepsilon\} = \max_{\varepsilon > 0} \min(\varepsilon, A_\varepsilon(\lambda)).$$

For any fixed $\varepsilon > 0$, $A_\varepsilon(\cdot)$ is non-increasing and

$$\lim_{\lambda \downarrow 0} A_\varepsilon(\lambda) = A_\varepsilon(0) = \sup_{C \in \mathscr{C}}(P_1(C) - P_2(C)) = \sup_{A \in \mathscr{B}(U)} (P_1(A) - P_2(A))$$

$$= \sup_{A \in \mathscr{B}(U)} |P_1(A) - P_2(A)| =: \boldsymbol{\sigma}(P_1, P_2)$$

Thus

$$\lim_{\lambda \to 0} \pi_\lambda(P_1, P_2) = \max_{\varepsilon > 0} \min\left(\varepsilon, \lim_{\lambda \to 0} A_\varepsilon(\lambda)\right) = \max_{\varepsilon > 0} \min(\varepsilon, \boldsymbol{\sigma}(P_1, P_2)) = \boldsymbol{\sigma}(P_1, P_2).$$

Analogously, as $\lambda \to \infty$

$$\begin{aligned}\lambda\pi_\lambda(P_1, P_2) &= \inf\{\lambda\varepsilon > 0: A_\varepsilon(\lambda) \leqslant \varepsilon\} \\ &= \inf\{\varepsilon > 0: A_\varepsilon(1) \leqslant \varepsilon/\lambda\} \to \inf\{\varepsilon > 0: A_\varepsilon(1) \leqslant 0\} \\ &= \ell_\infty(P_1, P_2).\end{aligned}$$

QED

As a generalization of π_λ we define the *Prokhorov distance*

$$\pi_H(P_1, P_2) := \inf\{H(\varepsilon) > 0: P_1(A^\varepsilon) \leqslant P_2(A) + H(\varepsilon), \quad \forall A \in \mathscr{B}_1\} \tag{3.2.24}$$

for any strictly increasing function $H \in \mathscr{H}$. From Equation (3.2.24),

$$\pi_H(P_1, P_2) = \inf\{\delta > 0: P_1(A) \leqslant P_2(A^{H^{-1}(\delta)}) + \delta \text{ for any } A \in \mathscr{B}_1\} \tag{3.2.25}$$

and it is easy to check that π_H is a simple distance with $\mathbb{K}_{\pi_H} = K_H$ (see Condition (2.2.3)). The metric π_λ is a special case of π_H with $H(t) = t/\lambda$.

Example 3.2.4 (Birnbaum–Orlicz distance ($\boldsymbol{\theta}_H$) and $\boldsymbol{\theta}_p$-metric in $\mathscr{P}(\mathbb{R})$). Let $U = \mathbb{R}$, $\mathrm{d}(x, y) = |x - y|$. Following Example 2.2.2 we define the *Birnbaum–Orlicz average distance*

$$\boldsymbol{\theta}_H(F_1, F_2) := \int_{-\infty}^{\infty} H(|F_1(t) - F_2(t)|)\,\mathrm{d}t \quad H \in \mathscr{H} \quad F_i \in \mathscr{F}(\mathbb{R}) \quad i = 1, 2 \tag{3.2.26}$$

and the *Birnbaum–Orlicz uniform distance*

$$\rho_H(F_1, F_2) := H(\rho(F_1, F_2)) = \sup_{x \in \mathbb{R}} H(|F_1(x) - F_2(x)|). \tag{3.2.27}$$

The $\boldsymbol{\theta}_p$-*metric* ($p > 0$)

$$\boldsymbol{\theta}_p(F_1, F_2) := \left\{ \int_{-\infty}^{\infty} |F_1(t) - F_2(t)|^p \, dt \right\}^{p'} \quad p' := \min(1, 1/p) \tag{3.2.28}$$

is a special case of $\boldsymbol{\theta}_H$ with appropriate normalization that makes $\boldsymbol{\theta}_p$ p. metric taking finite and infinite values in the distribution functions space $\mathscr{F} := \mathscr{F}(\mathbb{R})$. In case $p = \infty$ we denote $\boldsymbol{\theta}_\infty$ to be the Kolmogorov metric

$$\boldsymbol{\theta}_\infty(F_1, F_2) := \rho(F_1, F_2) := \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)|. \tag{3.2.29}$$

In the case $p = 0$ we put

$$\boldsymbol{\theta}_0(F_1, F_2) := \int_{-\infty}^{\infty} I\{t: F_1(t) \neq F_2(t)\} \, dt = \mathrm{Leb}\{F_1 \neq F_2\}.$$

Here as in the following, $I(A)$ is the indicator of the set A.

Example 3.2.5 (Co-minimal metrics). As we have seen in Section 3.1 each primary distance $\mu(P) = \mu(h_1 P, h_2 P)$ ($P \in \mathscr{P}_2$) determines a semidistance (see Definition 2.2.2) in the space of equivalence classes

$$\{P \in \mathscr{P}_2 : h_1 P = a,\ h_2 P = b\} \quad a, b \in R^J. \tag{3.2.30}$$

Analogously, the minimal distance

$$\begin{aligned}\hat{\mu}(P) &:= \hat{\mu}(T_1 P, T_2 P) \\ &:= \inf\{\mu(\tilde{P}) : \tilde{P} \in \mathscr{P}_2(U),\ \tilde{P} \text{ and } P \text{ have one and the same marginals,} \\ &\qquad T_i \tilde{P} = T_i P,\ i = 1, 2\},\ P \in \mathscr{P}_2(U)\end{aligned}$$

may be viewed as a semidistance in the space of classes of equivalence

$$\{P \in \mathscr{P}_2 : T_1 P = P_1,\ T_2 P = P_2\} \quad P_1, P_2 \in \mathscr{P}_1. \tag{3.2.31}$$

Obviously, the partitioning (3.2.31) is more refined than Equation (3.2.30) and hence each primary semidistance is a simple semidistance. Thus

{*the class of primary distances* (Definition 3.1.1)}
⊂ {*the class of simple semidistances* (Definition 3.2.1)}
⊂ {*the class of all p. semidistances* (Definition 2.3.1)}.

Open problem 3.2.1. A basic open problem in TPM is to find a good classification of the set of all p. semidistances. Does there exist a 'Mendeleyev periodic table' of p. semidistances?

One can get a classification of probability semidistances considering more and more refined partitions of $\mathscr{P}_2$. For instance, one can use a partition finer than Equation (3.2.31), generated by

$$\{P \in \mathscr{P}_2: T_1 P = P_1,\ T_2 P = P_2,\ P \in \mathscr{PC}_t\} \quad t \in T \tag{3.2.32}$$

where P_1 and P_2 are laws in $\mathscr{P}_1$ and $\mathscr{PC}_t (t \in T)$ are subsets of $\mathscr{P}_2$, whose union covers $\mathscr{P}_2$. As an example of the set $\mathscr{PC}_t$ one could consider

$$\mathscr{PC}_t = \left\{P \in \mathscr{P}_2: \int_U f_i \, \mathrm{d}P \leqslant b_i,\ i \in J\right\} \quad t = (J, \bar{b}, \bar{f}) \tag{3.2.33}$$

where J is an index set, $\bar{b} := (b_i, i \in J)$ is a set of reals and $\bar{f} = \{f_i, i \in J\}$ is a family of bounded continuous functions on U^2 (Kemperman 1983, Levin and Rachev 1990).

Another useful example of a set $\mathscr{PC}_t$ is constructed using a given probability metric $\nu(P)(P \in \mathscr{P}_2)$ and has the form

$$\mathscr{PC}_t = \{P \in \mathscr{P}_2: \nu(P) \leqslant t\} \tag{3.2.34}$$

where $t \in [0, \infty]$ is a fixed number.

Open problem 3.2.2. Under which conditions is the functional

$$\mu(P_1, P_2; \mathscr{PC}_t) := \inf\{\mu(P): P \in \mathscr{P}_2,\ T_i P = P_i (i = 1, 2),\ P \in \mathscr{PC}_t\} (P_1, P_2 \in \mathscr{P}_2)$$

a simple semidistance (resp., semimetric) w.r.t. the given p. distance (resp. metric) μ?

Further, we shall examine this problem in the special case of (3.2.34) (see Theorem 3.2.2). Analogously, one can investigate the case of $\mathscr{PC}_t = \{P \in \mathscr{P}_2: \nu_i(P) \leqslant \alpha_i,\ i = 1, 2, \dots\}$ $(t = (\alpha_1, \alpha_2, \dots))$ for fixed p. metrics ν_i and $\alpha_i \in [0, \infty]$.

Following the main idea of obtaining primary and simple distances by means of minimization procedures of certain types (see Definitions 3.1.2 and 3.2.2) we shall give the notion of 'co-minimal distance'.

For a given compound semidistances μ and ν with parameters $\mathbb{K}_\mu$ and $\mathbb{K}_\nu$, respectively, and for each $\alpha > 0$ denote

$$\mu\nu(P_1, P_2, \alpha) = \inf\{\mu(P): P \in \mathscr{P}_2,\ T_1 P = P_1,\ T_2 P = P_2,\ \nu(P) \leqslant \alpha\} \quad P_1, P_2 \in \mathscr{P}_1 \tag{3.2.35}$$

(cf. Equation (3.2.32) and (3.2.34)).

Definition 3.2.4. The functional $\mu\nu(P_1, P_2, \alpha)$ $(P_1, P_2 \in \mathscr{P}_1, \alpha > 0)$ will be called the *co-minimal (metric) functional w.r.t. the p. distances μ and ν* (see Fig. 3.2.1)

As we will see in the next theorem, the functional $\mu\nu(\cdot,\cdot,\alpha)$ has some metric properties but nevertheless it is not a p. distance, however, $\mu\nu(\cdot,\cdot,\alpha)$ induces p. semidistances as follows.

Let $\mu\nu$ be the so-called *co-minimal distance*

$$\mu\nu(P_1, P_2) = \inf\{\alpha > 0;\ \mu\nu(P_1\ P_2, \alpha) < \alpha\} \tag{3.2.36}$$

(see Fig. 3.2.1) and let

$$\overline{\mu\nu}(P_1, P_2) = \limsup_{\alpha \to 0} \alpha\mu\nu(P_1, P_2, \alpha).$$

Then the following theorem is true.

Theorem 3.2.2. Let U be an u.m. s.m.s. and μ be a p. distance satisfying the 'continuity' condition in Theorem 3.2.1. Then, for any p. distance ν,

(a) $\mu\nu(\cdot,\cdot,\alpha)$ satisfies the following metric properties

ID$^{(3)}$: $\mu\nu(P_1, P_2, \alpha) = 0 \Leftrightarrow P_1 = P_2$

SYM$^{(3)}$: $\mu\nu(P_1, P_2, \alpha) = \mu\nu(P_2, P_1, \alpha)$

TI$^{(3)}$: $\mu\nu(P_1, P_3, \mathbb{K}_\nu(\alpha + \beta)) \leqslant \mathbb{K}_\mu(\mu\nu(P_1, P_2, \alpha) + \mu\nu(P_2, P_3, \beta))$

for any $P_1, P_2, P_3 \in \mathscr{P}_1$, $\alpha \geqslant 0$, $\beta \geqslant 0$.

(b) $\mu\nu$ is a simple distance with parameter $\mathbb{K}_{\mu\nu} = \max[\mathbb{K}_\mu, \mathbb{K}_\nu]$. In particular, if μ and ν are p. metrics then $\mu\nu$ is a simple metric.

(c) $\overline{\mu\nu}$ is a simple semidistance with parameter $\mathbb{K}_{\overline{\mu\nu}} = 2\mathbb{K}_\mu\mathbb{K}_\nu$.

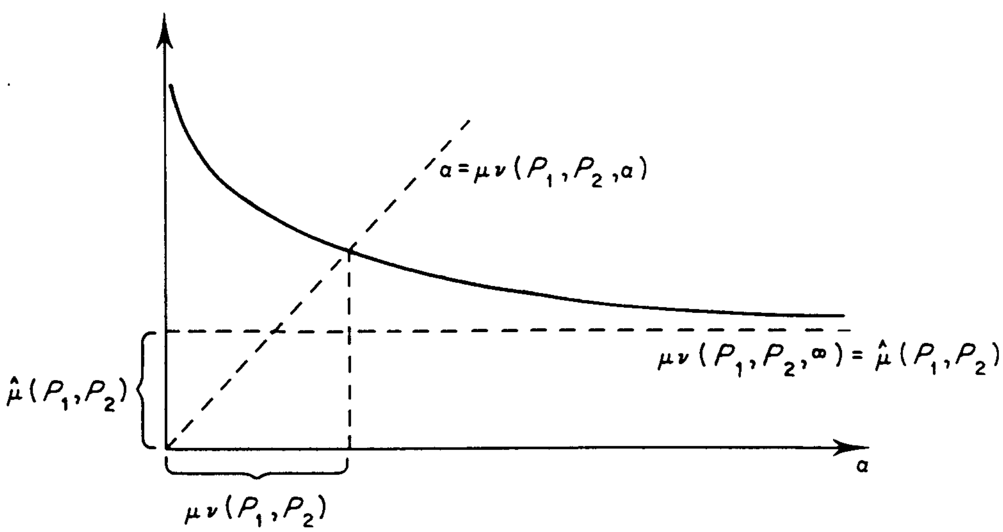


Figure 3.2.1 Co-minimal distance $\mu\nu(P_1, P_2)$. (From Rachev and Shortt, 1990. Reproduced by permission of the American Mathematical Society.)

Proof.

(a) By Theorem 3.2.1, and Fig. 3.2.1, $\mu\nu(P_1, P_2, \alpha) = 0 \Rightarrow \hat{\mu}(P_1, P_2) = 0 \Rightarrow P_1 = P_2$ as well as if $P_1 \in \mathscr{P}_1$ and X is a r.v. with distribution P_1 then $\mu\nu(P_1, P_1; \alpha) \leqslant \mu(\Pr_{X,X}) = 0$. So, $\mathbf{ID}^{(3)}$ is valid. Let us prove $\mathbf{TI}^{(3)}$. For each $P_1, P_2, P_3 \in \mathscr{P}_1$ $\alpha \geqslant 0$, $\beta \geqslant 0$ and $\varepsilon \geqslant 0$ define laws $P_{12} \in \mathscr{P}_2$ and $P_{23} \in \mathscr{P}_2$ such that $T_iP_{12} = P_i, T_iP_{23} = P_{i+1} (i = 1, 2)$, $\nu(P_{12}) \leqslant \alpha$, $\nu(P_{23}) \leqslant \beta$ and $\mu\nu(P_1, P_2, \alpha) \geqslant \mu(P_{12}) - \varepsilon$, $\mu\nu(P_2, P_3, \alpha) \geqslant \mu(P_{23}) - \varepsilon$. Define a law $Q \in \mathscr{P}_3$ by Equation (3.2.5). Then Q has bivariate marginals $T_{12}Q = P_{12}$ and $T_{23}Q = P_{23}$, hence, $\nu(T_{13}Q) \leqslant \mathbb{K}_\nu[\nu(P_{12}) + \nu(P_{23})] \leqslant \mathbb{K}_\nu(\alpha + \beta)$ and

$$\begin{aligned}\mu\nu(P_1, P_3, \mathbb{K}_\nu(\alpha + \beta)) &\leqslant \mu(T_{13}Q) \leqslant \mathbb{K}_\mu[\mu(P_{12}) + \mu(P_{23})] \\ &\leqslant \mathbb{K}_\mu[\mu\nu(P_1, P_2, \alpha) + \mu\nu(P_2, P_3, \beta) + 2\varepsilon].\end{aligned}$$

Letting $\varepsilon \to 0$, we get $\mathbf{TI}^{(3)}$.

(b) If $\mu\nu(P_1, P_2) < \alpha$ and $\mu\nu(P_2, P_3) < \beta$, then there exists P_{12} (resp. P_{23}) with marginals P_1 and P_2 (resp. P_2 and P_3) such that $\mu(P_{12}) < \alpha$, $\nu(P_{12}) < \alpha$, $\mu(P_{23}) < \beta$. In a similar way, as in (a) we conclude that $\mu\nu(P_1, P_3, \mathbb{K}_\nu(\alpha + \beta)) \leqslant \mathbb{K}_\mu(\alpha + \beta)$, thus, $\mu\nu(P_1, P_2) \leqslant \max(\mathbb{K}_\mu, \mathbb{K}_\nu)(\alpha + \beta)$.

(c) Follows from (a) with $\alpha = \beta$. QED

Example 3.2.6. (Minimal norms). Each co-minimal distance $\mu\nu$ is greater than the minimal distance $\hat{\mu}$ (see Fig. 3.3.1). We now consider examples of simple metrics $\mathring{\mu}$ corresponding to given p. distances μ that have (like $\mu\nu$) a 'minimal' structure but $\mathring{\mu} \leqslant \hat{\mu}$.

Let $\mathscr{M}_k$ be the set of all finite non-negative measures on the Borel σ-algebra $\mathscr{B}_k = \mathscr{B}(U^k)$ (U is a s.m.s.). Let $\mathscr{M}_0$ denote the space of all finite signed measures ν on $\mathscr{B}_1$ with total mass $m(U) = 0$. Denote by $\mathscr{CS}(U^2)$ the set of all continuous, symmetric and non-negative functions on U^2. Define the functionals

$$\mu_c(m) := \int_{U^2} c(x, y) m(\mathrm{d}x, \mathrm{d}y), \ m \in \mathscr{M}_2 \quad c \in \mathscr{CS}(U^2) \tag{3.2.37}$$

and

$$\mathring{\mu}_c(\nu) := \inf\{\mu_c(m) \colon T_1 m - T_2 m = \nu\} \quad \nu \in \mathscr{M}_0 \tag{3.2.38}$$

where $T_i m$ means the ith marginal measure of m.

Lemma 3.2.2. For any $c \in \mathscr{CS}(U^2)$ the functional $\mathring{\mu}_c$ is a seminorm in the space $\mathscr{M}_0$.

Proof. Obviously, $\mathring{\mu}_c \geqslant 0$. For any positive constant a we have

$$\begin{aligned}\mathring{\mu}_c(a\nu) &= \inf\{\mu_c(m) \colon T_1(1/a)m - T_2(1/a)m = \nu\} \\ &= a \inf\{\mu_c((1/a)m) \colon T_1(1/a)m - T_2(1/a)m = \nu\} \\ &= a\mathring{\mu}_c(\nu).\end{aligned}$$

If $a \leqslant 0$ and $\tilde{m}(A \times B) := m(B \times A)(A, B \in \mathscr{B}_1)$ then by the symmetry of c we get

$$\begin{aligned}\mu_c(av) &= \inf\{\mu_c(m)\colon T_2(-1/a)m - T_1(-1/a)m = v\}\\ &= \inf\{\mu_c(\tilde{m})\colon T_1(-1/a)\tilde{m} - T_2(-1/a)\tilde{m} = v\}\\ &= |a|\mathring{\mu}_c(v).\end{aligned}$$

Let us prove now that $\mathring{\mu}_c$ is a subadditive function. Let $v_1, v_2 \in \mathscr{M}$. For m_1, $m_2 \in \mathscr{M}_2$ with $T_1 m_i - T_2 m_i = v_i$ $(i = 1, 2)$, let $m = m_1 + m_2$. Then we have $\mu_c(m) = \mu_c(m_1) + \mu_c(m_2)$ and $T_1 m - T_2 m = v_1 + v_2$, hence, $\mathring{\mu}_c(v_1 + v_2) \leqslant \mathring{\mu}_c(v_1) + \mathring{\mu}_c(v_2)$. QED

In the next theorem we give a sufficient condition for

$$\mathring{\mu}_c(P_1, P_2) := \mathring{\mu}_c(P_1 - P_2) \quad P_1, P_2 \in \mathscr{P}_1 \tag{3.2.39}$$

to be a simple metric in $\mathscr{P}_1$. In the proof we shall make use of the *Zolotarev's semimetric* $\zeta_{\mathscr{F}}$. Namely, for a given class $\mathscr{F}$ of bounded continuous function $f\colon U \to \mathbb{R}$, we define

$$\zeta_{\mathscr{F}}(P_1, P_2) = \sup_{f \in \mathscr{F}} \left| \int_U f \, \mathrm{d}(P_1 - P_2) \right| \quad P_i \in \mathscr{P}(U).$$

Clearly, $\zeta_{\mathscr{F}}$ is a simple semimetric. Moreover, if the class $\mathscr{F}$ is 'rich enough' to preserve the implication $\zeta_{\mathscr{F}}(P_1, P_2) = 0 \Rightarrow P_1 = P_2$, we have the $\zeta_{\mathscr{F}}$ is a simple metric.

Theorem 3.2.3.

(i) For any $c \in \mathscr{C}\mathscr{S}(U^2)$, $\mathring{\mu}_c(P_1, P_2)$ defined by Equality (3.2.39) is a semimetric in $\mathscr{P}_1$.

(ii) Further, if the class $\mathscr{F}_c := \{f\colon U \to \mathbb{R},\ |f(x) - f(y)| \leqslant c(x, y)\ \forall x, y \in U\}$ contains the class $\mathscr{G}$ of all functions

$$f(x) := f_{k,C}(x) := \max\{0, 1/k - d(x, C)\} \quad x \in U$$

(k is an integer greater than some fixed k_0, C is a closed non-empty set) then $\mathring{\mu}_c$ is a simple metric in $\mathscr{P}_1$.

Proof.

(i) The proof follows immediately from Lemma 3.2.2 and the definition of semimetric (see Definition 2.2.1).

(ii) For any $m \in \mathscr{M}_2$ such that $T_1 m - T_2 m = P_1 - P_2$ and for any $f \in \mathscr{F}_c$ we have

$$\begin{aligned}\left| \int_U f \, \mathrm{d}(P_1 - P_2) \right| &= \left| \int_{U^2} f(x) - f(y) m(\mathrm{d}x, \mathrm{d}y) \right|\\ &\leqslant \int_{U^2} |f(x) - f(y)| m(\mathrm{d}x, \mathrm{d}y) \leqslant \mu_c(m)\end{aligned}$$

hence, the Zolatarev's metric $\zeta_{\mathscr{F}_c}(P_1, P_2)$ is a lower bound for $\mathring{\mu}_c(P_1, P_2)$. On the other hand, by assumption, $\zeta_{\mathscr{F}_c} \geqslant \zeta_{\mathscr{G}}$. Thus assuming $\mathring{\mu}_c(P_1, P_2) = 0$ we get $0 \leqslant \zeta_{\mathscr{G}}(P_1, P_2) \leqslant \zeta_{\mathscr{F}_c}(P_1, P_2) \leqslant \mathring{\mu}_c(P_1, P_2) = 0$. Next, for any closed nonempty set C we have

$$P_1(C) \leqslant k \int_U f_{k,C}\,\mathrm{d}P_1 \leqslant k\zeta_{\mathscr{G}}(P_1, P_2) + k \int_U f_{k,C}\,\mathrm{d}P_2 \leqslant P_2(C^{1/k}).$$

Letting $k \to \infty$ we get $P_1(C) \leqslant P_2(C)$ and hence, by the symmetry, $P_1 = P_2$.

QED

Remark 3.2.1. Obviously $\mathscr{F}_d \supseteq \mathscr{G}$ and hence $\mathring{\mu}_d$ is a simple metric in $\mathscr{P}_1$, however, if $p > 1$ then $\mathring{\mu}_{d^p}$ is not a metric in $\mathscr{P}_1$ as is shown in the following example. Let $U = [0, 1]$, $d(x, y) = |x - y|$. Let P_1 be a law concentrated on the origin and P_2 a law concentrated on 1. For any $n = 1, 2, \ldots$ consider a measure $m^{(n)} \in \mathscr{M}_2$ with total mass $m^{(n)}(U^2) = 2n+1$ and

$$m^{(n)}\left(\left\{\frac{i}{n}, \frac{i}{n}\right\}\right) = 1 \quad i = 0, \ldots, n$$

$$m^{(n)}\left(\left\{\frac{i}{n}, \frac{(i+1)}{n}\right\}\right) = 1 \quad i = 0, \ldots, n-1$$

(see Fig. 3.2.2). Then obviously, $T_1 m^{(n)} - T_2 m^{(n)} = P_1 - P_2$ and

$$\int_{U \times U} |x - y|^p m^{(n)}(\mathrm{d}x, \mathrm{d}y) = \sum_{i=0}^{n-1} \left(\frac{1}{n}\right)^p = n^{1-p}$$

hence, if $p > 1$ then

$$\mathring{\mu}_d(P_1, P_2) \leqslant \inf_{n>0} \int_{U^2} |x - y|^p m^{(n)}(\mathrm{d}x, \mathrm{d}y) = 0$$

and thus $\mathring{\mu}_{d^p}(P_1, P_2) = 0$.

Definition 3.2.5. The simple semimetric $\mathring{\mu}_c$ (see Equality (3.2.39)) is said to be the *minimal norm w.r.t. the functional* μ_c.

Obviously, for any $c \in \mathscr{C}\mathscr{S}$,

$$\mathring{\mu}_c(P_1, P_2) \leqslant \hat{\mu}_c(P_1, P_2) := \inf\{\mu_c(P) : P \in \mathscr{P}_2,\ T_i P = P_i,\ i = 1, 2\} \quad P_1, P_2 \in \mathscr{P}_2 \tag{3.2.40}$$

hence, for each minimal metric of the type $\hat{\mu}_c$ we can construct an estimate from below by means of $\mathring{\mu}_c$, but what is more important, $\mathring{\mu}_c$ is a *simple semimetric even though* μ_c *is not a probability semidistance.* For instance, let

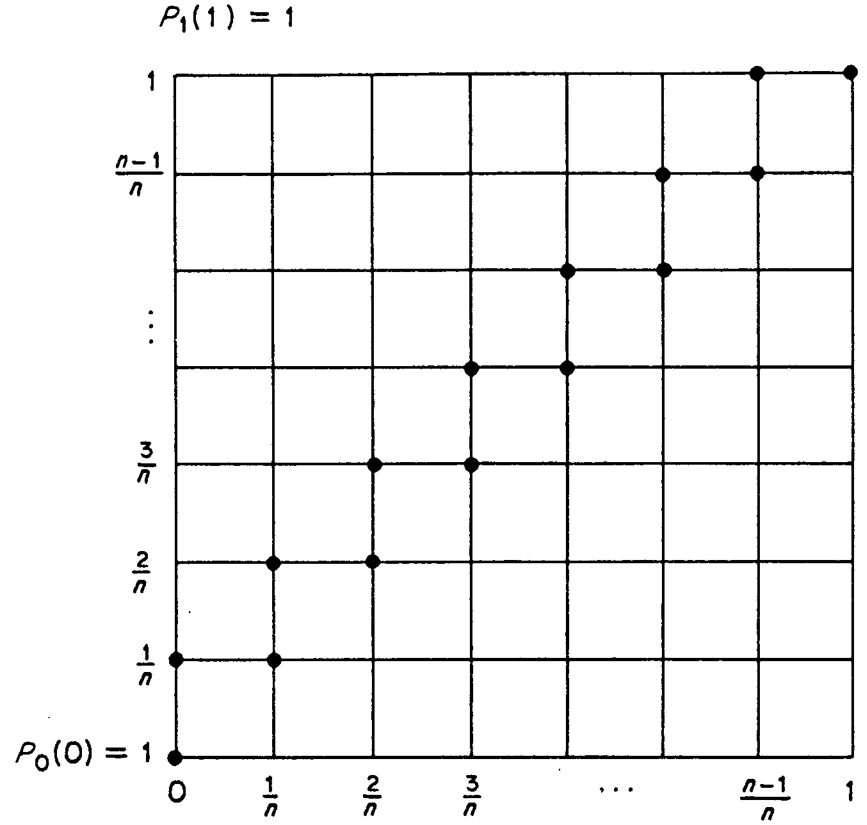


Figure 3.2.2 Support of the measure $m^{(n)}$ with marginals P_1 and P_2. (From Rachev and Shortt, 1990. Reproduced by permission of the American Mathematical Society.)

$c_h(x, y) := d(x, y)h(\max(d(x, a), d(y, a)))$, where h is a nondecreasing non-negative continuous function on $[\alpha, \infty)$ for some $\alpha > 0$. Then, as in Theorem 3.2.3, we conclude that $\zeta_{c_h} \leqslant \mathring{\mu}_{c_h}$ and $\zeta_{c_h}(P_1, P_2) = 0 \Rightarrow P_1 = P_2$. Thus $\mathring{\mu}_{c_h}$ is a simple metric, while if $h(t) = t^p (p > 1) \mu_{c_h}$ is not a p. distance. Further (Section 5.3), we shall prove that $\mathring{\mu}$ admits the dual formula: for any laws P_1 and P_2 on a s.m.s. (U, d), with finite $\int d(x, a)h(d(x, a))(P_1 + P_2)(dx)$,

$$\mathring{\mu}_{c_h}(P_1, P_2) = \sup\left\{\left|\int_U f\, d(P_1 - P_2)\right| : f: U \to \mathbb{R},\right.$$
$$\left. |f(x) - f(y)| \leqslant c_h(x, y) \quad \forall x, y \in U\right\} \tag{3.2.41}$$

From Equality (3.2.41) it follows that if $U = \mathbb{R}$ and $d(x, y) = |x - y|$, then $\mathring{\mu}_c$ may be represented explicitly as an average metric with weight $h(\cdot - a)$ between d.f.s

$$\mathring{\mu}_{c_h}(P_1, P_2) = \mathring{\mu}_{c_h}(F_1, F_2) := \int_{-\infty}^{\infty} |F_1(x) - F_2(x)| h(|x - a|)\, dx \tag{3.2.42}$$

where F_i is the d.f. of P_i (see further Section 5.4).

3.3 COMPOUND DISTANCES AND MOMENT FUNCTIONS

We continue the classification of probability distances. Recall some basic examples of p. metrics on a s.m.s. (U, d):

(a) The *moment metric* (see Example 3.1.2):

$$\mathcal{M}(X, Y) = |\mathbb{E}d(X, a) - \mathbb{E}d(Y, a)| \quad X, Y \in \mathfrak{X}(U)$$

($\mathcal{M}$ is a primary metric in the space $\mathfrak{X}(U)$ of U-valued r.v.s).

(b) The Kantorovich metric (see Example 3.2.2):

$$\boldsymbol{\kappa}(X, Y) = \sup\{|\mathbb{E}f(X) - \mathbb{E}f(Y)|: f\colon U \to \mathbb{R} \text{ bounded}, \\ |f(x) - f(y)| \leqslant \mathrm{d}(x, y) \quad \forall x \text{ and } y \in U\}$$

($\boldsymbol{\kappa}$ is a simple metric in $\mathfrak{X}(U)$).

(c) The L_1-metric (see (2.3.3)):

$$\mathcal{L}_1(X, Y) = \mathbb{E}d(X, Y) \qquad X, Y \in \mathfrak{X}(U).$$

The $\mathcal{L}_1$-metric is a p. metric in $\mathfrak{X}(U)$ (Definition 2.3.2). Since the value of $\mathcal{L}_1(X, Y)$ depends on the joint distribution of the pair (X, Y) we shall call $\mathcal{L}_1$ a compound metric.

Definition 3.3.1. A *compound distance* (resp., metric) is any probability distance μ (resp., metric). See Definitions 2.3.1 and 2.3.2.

Remark 3.3.1. In many papers on probability metrics, 'compound' metric stands for a metric which is not simple, however, all 'non-simple' metrics that have been used in these papers are in fact 'compound' in the sense of Definition 3.3.1. The problem of classification of p. metrics which are neither compound (in the sense of Definition 3.3.1) nor simple is open (see Open problems 3.2.1 and 3.2.2).

Let us consider some examples of compound distances and metrics.

Example 3.3.1. (Average compound distances). Let (U, d) be a s.m.s. and $H \in \mathcal{H}$ (see Example 2.2.1). Then

$$\mathcal{L}_H(P) := \int_{U^2} H(\mathrm{d}(x, y))P(\mathrm{d}x, \mathrm{d}y) \quad P \in \mathcal{P}_2 \tag{3.3.1}$$

is a compound distance with parameter K_H (see (2.2.3)), and will be called *H-average compound distance.*

If $H(t) = t^p$, $p > 0$ and $p' = \min(1, 1/p)$ then

$$\mathcal{L}_p(P) := [\mathcal{L}_H(P)]^{p'} \quad P \in \mathcal{P}_2 \tag{3.3.2}$$

is a compound metric in

$$\mathscr{P}_2^{(p)}(U) := \left\{P \in \mathscr{P}_2 : \int_{U^2} d^p(x, a)[P(\mathrm{d}x, \mathrm{d}y) + P(\mathrm{d}y, \mathrm{d}x)] < \infty\right\}.$$

In the space

$$\mathfrak{X}^{(p)}(U) := \{X \in \mathfrak{X}(U) : \mathbb{E}d^p(X, a) < \infty\}$$

the metric $\mathscr{L}_p$ is the usual *p-average metric*

$$\mathscr{L}_p(X, Y) := [\mathbb{E}d^p(X, Y)]^{p'} \quad 0 < p < \infty. \tag{3.3.3}$$

In the limit cases $p = 0$, $p = \infty$ we shall define the compound metrics

$$\mathscr{L}_0(P) := P\left(\bigcup_{x \neq y} (x, y)\right) \quad P \in \mathscr{P}_2 \tag{3.3.4}$$

and

$$\mathscr{L}_\infty(P) := \inf\{\varepsilon > 0 : P(d(x, y) > \varepsilon) = 0\} \quad P \in \mathscr{P}_2 \tag{3.3.5}$$

that on $\mathfrak{X}$ have the forms

$$\mathscr{L}_0(X, Y) := \mathbb{E}I\{X \neq Y\} = \Pr(X \neq Y) \quad X, Y \in \mathfrak{X} \tag{3.3.6}$$

and

$$\mathscr{L}_\infty(X, Y) := \operatorname{ess\,sup} d(X, Y) := \inf\{\varepsilon > 0 : \Pr(d(X, Y) > \varepsilon) = 0\}. \tag{3.3.7}$$

Example 3.3.2. (Ky Fan distance and Ky Fan metric). The Ky Fan metric $\mathbf{K}$ in $\mathfrak{X}(\mathbb{R})$ was defined by Equality (2.1.5) and we shall extend that definition considering the space $\mathscr{P}_{12}(U)$ for a s.m.s. (U, d). We define the Ky Fan metric in $\mathscr{P}_2(U)$ as follows

$$\mathbf{K}(P) := \inf\{\varepsilon > 0 : P(d(x, y) > \varepsilon) < \varepsilon\} \quad P \in \mathscr{P}_2$$

and on $\mathfrak{X}(U)$ by $\mathbf{K}(X, Y) = \mathbf{K}(\Pr_{X,Y})$. In this way $\mathbf{K}$ takes the form of the *distance in probability* in $\mathfrak{X} = \mathfrak{X}(U)$

$$\mathbf{K}(X, Y) := \inf\{\varepsilon > 0 : \Pr(d(X, Y) > \varepsilon) < \varepsilon\} \quad X, Y \in \mathfrak{X}. \tag{3.3.8}$$

A possible extension of the metric structure of $\mathbf{K}$ is the *Ky Fan distance*:

$$\mathbf{KF}_H(P) := \inf\{\varepsilon > 0 : P(H(d(x, y)) > \varepsilon) < \varepsilon\} \tag{3.3.9}$$

for each $H \in \mathscr{H}$. It is easy to check that $\mathbf{KF}_H$ is a compound distance with parameter $\mathbb{K}_{\mathbf{KF}} := K_H$ (see (2.2.3)). A particular case of the Ky Fan distance is the *parametric family of Ky Fan metrics*

$$\mathbf{K}_\lambda(P) := \inf\{\varepsilon > 0 : P(d(x, y) > \lambda\varepsilon) < \varepsilon\}. \tag{3.3.10}$$

For each $\lambda > 0$

$$\mathbf{K}_\lambda(X, Y) := \inf\{\varepsilon > 0\colon \Pr(d(X, Y) > \lambda\varepsilon) < \varepsilon\} \quad X, Y \in \mathfrak{X}$$

metrizes the convergence 'in probability' in $\mathfrak{X}(U)$, i.e.

$$\mathbf{K}_\lambda(X_n, Y) \to 0 \Leftrightarrow \Pr(d(X_n, Y) > \varepsilon) \to 0 \text{ for any } \varepsilon > 0.$$

In the limit cases,

$$\lim_{\lambda \to 0} \mathbf{K}_\lambda = \mathscr{L}_0 \quad \lim_{\lambda \to \infty} \lambda\mathbf{K}_\lambda = \mathscr{L}_\infty \tag{3.3.11}$$

we get, however, average compound metrics (see Equalities (3.3.4)–(3.3.7)) that induce convergence, stronger than convergence in probability, i.e.,

$$\mathscr{L}_0(X_n, Y) \to 0 \underset{\not\Leftarrow}{\Rightarrow} X_n \to Y \text{ 'in probability'}$$

and

$$\mathscr{L}_\infty(X_n, Y) \to 0 \underset{\not\Leftarrow}{\Rightarrow} X_n \to Y \text{ 'in probability'.}$$

Example 3.3.3 (Birnbaum–Orlicz compound distances). Let $U = \mathbb{R}$, $d(x, y) = |x - y|$. For each $p \in [0, \infty]$ consider the following compound metrics in $\mathfrak{X}(\mathbb{R})$

$$\mathbf{\Theta}_p(X_1, X_2) := \left[\int_{-\infty}^{\infty} \tau_p(t; X_1, X_2)\mathrm{d}t\right]^{p'} \quad 0 < p < \infty \quad p' := \min(1, 1/p) \tag{3.3.12}$$

$$\mathbf{\Theta}_\infty(X_1, X_2) := \sup_{t \in \mathbb{R}} \tau(t; X_1, X_2) \tag{3.3.13}$$

$$\mathbf{\Theta}_0(X_1, X_2) := \int_{-\infty}^{\infty} I\{t\colon \tau(t; X_1, X_2) \neq 0\}\, \mathrm{d}t$$

where

$$\tau(t; X_1, X_2) := \Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1). \tag{3.3.14}$$

If $H \in \mathscr{H}$ then

$$\mathbf{\Theta}_H(X_1, X_2) := \int_{-\infty}^{\infty} H(\tau(t; X_1, X_2))\, \mathrm{d}t \tag{3.3.15}$$

is a compound distance with $\mathbb{K}_{\mathbf{\Theta}_H} = K_H$. The functional $\mathbf{\Theta}_H$ will be called a *Birnbaum–Orlicz compound average distance* and

$$\mathbf{R}_H(X_1, X_2) := H(\mathbf{\Theta}_\infty(X_1, X_2)) = \sup_{t \in \mathbb{R}} H(\tau(t; X_1, X_2)) \tag{3.3.16}$$

will be called a *Birnbaum–Orlicz compound uniform distance.*

Each example 3.3.i. ($i = 1, 2, 3$) is closely related to the corresponding example 3.2.i. In fact, we shall prove (Corollary 5.2.2) that ℓ_H (see Equation 3.2.10) is a minimal distance (see Definition 3.2.2) w.r.t. $\mathscr{L}_H$ for any convex $H \in \mathscr{H}$, i.e.

$$\ell_H = \hat{\mathscr{L}}_H. \tag{3.3.17}$$

Analogously, the simple metrics ℓ_p (see (3.2.11)–(3.2.14)), the Prokhorov metric π_λ (see (3.2.22)), and the Prokhorov distance π_H (see (3.2.24)) are minimal with respect to the $\mathscr{L}_p$-metric, Ky Fan metric $\mathbf{K}_\lambda$ and Ky Fan distance $\mathbf{KF}_H$, i.e.

$$\ell_p = \hat{\mathscr{L}}_p (p \in [0, \infty]) \quad \pi_\lambda = \hat{\mathbf{K}}_\lambda (\lambda > 0) \quad \pi_H = \widehat{\mathbf{KF}}_H. \tag{3.3.18}$$

Finally, the Birnbaum–Orlicz metric and distance $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_H$ (see Equations (3.2.28) and (3.2.26)) and the Birnbaum–Orlicz uniform distance $\boldsymbol{\rho}_H$ (see Equation (3.2.27)) are minimal with respect to their 'compound versions' $\boldsymbol{\Theta}_p$, $\boldsymbol{\Theta}_H$ and $\mathbf{R}_H$, i.e.

$$\boldsymbol{\theta}_p = \hat{\boldsymbol{\Theta}}_p (p \in [0, \infty]) \quad \boldsymbol{\theta}_H = \hat{\boldsymbol{\Theta}}_H \quad \rho_H = \hat{\mathbf{R}}_H. \tag{3.3.19}$$

The equalities (3.3.17) to (3.3.19) represent the main relationships between simple and compound distances (resp., metrics) and serve as a framework for TPM (see Fig. 1.1.1, *comparison of metrics*).

Analogous relationships exist between primary and compound distances. For example, we shall prove (Section 9) that the primary distance

$$\mathscr{M}_{H,1}(\alpha, \beta) = H(|\alpha - \beta|) \tag{3.3.20}$$

(see (3.1.5)) is a primary minimal distance (see Definition 3.1.2) w.r.t. the p. distance $H(\mathscr{L}_1)$ ($H \in \mathscr{H}$), i.e.,

$$\mathscr{M}_{H,1}(\alpha, \beta) := \inf\left\{H(\mathscr{L}_1(P)) \colon \int_{U^2} d(x, a)P(\mathrm{d}x, \mathrm{d}y) = \alpha, \int_{U^2} d(a, y)P(\mathrm{d}x, \mathrm{d}y) = \beta\right\}. \tag{3.3.21}$$

Since a compound metric μ may take infinite values we have to determine a concept of μ-boundedness. With that aim in view we introduce the notion of a 'moment function' which differs from the notion of simple distance in the 'identity' property only (cf. Definition 3.2.1 and $\mathbf{ID}^{(2)}$, $\mathbf{TI}^{(2)}$).

Definition 3.3.2. A mapping $\mathbb{M}\colon \mathscr{P}_1 \times \mathscr{P}_1 \to [0, \infty]$ is said to be a *moment function* (with parameter $\mathbb{K}_{\mathbb{M}} \geqslant 1$) if it possesses the following properties for all $P_1, P_2, P_3 \in \mathscr{P}_1$.

$\mathbf{SYM}^{(4)}$: $\mathbb{M}(P_1, P_2) = \mathbb{M}(P_2, P_1)$,

$\mathbf{TI}^{(4)}$: $\mathbb{M}(P_1, P_3) \leqslant \mathbb{K}_{\mathbb{M}}[\mathbb{M}(P_1, P_2) + \mathbb{M}(P_2, P_3)]$.

We shall use moment functions as upper bounds for p. distances μ. As an example, we shall now consider μ to be the p. average distance (see

Equalities (3.3.2) and (3.3.3))

$$\mathscr{L}_p(P) := \left[\int_{U \times U} d^p(x, y) P(\mathrm{d}x, \mathrm{d}y)\right]^{p'} \quad p > 0 \quad p' := \min(1, 1/p) \quad P \in \mathscr{P}_2. \quad (3.3.22)$$

For any $p > 0$ and $a \in U$ define the moment function

$$\Lambda_{p,a}(P_1, P_2) := \left[\int_U d^p(x, a) P_1(\mathrm{d}x)\right]^{p'} + \left[\int_U d^p(x, a) P_2(\mathrm{d}x)\right]^{p'}. \quad (3.3.23)$$

By the Minkovski inequality we get our first (rough) upper bound for the value $\mathscr{L}_p(P)$ under the convention that the marginals $T_i P = P_i$ $(i = 1, 2)$ are known

$$\mathscr{L}_p(P) \leqslant \Lambda_{p,a}(P_1, P_2). \quad (3.3.24)$$

Obviously, by the Inequality (3.3.24), we can get a more refined estimate

$$\mathscr{L}_p(P) \leqslant \Lambda_p(P_1, P_2) \quad (3.3.25)$$

where

$$\Lambda_p(P_1, P_2) := \inf_{a \in U} \Lambda_{p,a}(P_1, P_2). \quad (3.3.26)$$

Further, we shall consider the following question.

Problem 3.3.1. What is the best possible inequality of the type

$$\mathscr{L}_p(P) \leqslant \check{\mathscr{L}}_p(P_1, P_2), \quad (3.3.27)$$

where $\check{\mathscr{L}}_p$ is a functional that depends on the marginals $P_i = T_i P$ $(i = 1, 2)$ only?

Remark 3.3.1. Suppose (X, Y) is a pair of *dependent* random variables taking on values in s.m.s. (U, d). Knowing only the marginal distributions $P_1 = \Pr_X$ and $P_2 = \Pr_Y$, what is the best possible improvement of the 'triangle inequality' bound

$$\mathscr{L}_1(X, Y) := \mathbb{E}d(X, Y) \leqslant \mathbb{E}d(X, a) + \mathbb{E}d(Y, a). \quad (3.3.28)$$

The answer is simple: the best possible upper bound for $\mathbb{E}d(X, Y)$ is given by

$$\check{\mathscr{L}}_1(P_1, P_2) := \sup\{\mathscr{L}_1(X_1, X_2) \colon \Pr_{X_i} = P_i, i = 1, 2\}. \quad (3.3.29)$$

More difficult is to determine dual and explicit representations for $\check{\mathscr{L}}_1$ similar to those of the minimal metric $\hat{\mathscr{L}}_1$ (the Kantorovich metric). We shall discuss this problem in Section 8.1.

More generally, for any compound semidistance $\mu(P)$ $(P \in \mathscr{P}_2)$ let us define the functional

$$\check{\mu}(P_1, P_2) := \sup\{\mu(P) \colon T_i P = P_i,\ i = 1, 2\} \quad P_1, P_2 \in \mathscr{P}_1. \tag{3.3.30}$$

Definition 3.3.3. The functional $\check{\mu}: \mathscr{P}_1 \times \mathscr{P}_1 \to [0, \infty]$ will be called *maximal distance* w.r.t. the given compound semidistance μ.

Note that, by definition, a maximal distance need not be a distance. We prove the following theorem.

Theorem 3.3.1. If (U, d) is an u.m. s.m.s. and μ is a compound distance with parameter $\mathbb{K}_\mu$ then $\check{\mu}$ is a moment function and $\mathbb{K}_{\check{\mu}} = \mathbb{K}_\mu$. Moreover, the following stronger version of the $\mathbf{TI}^{(4)}$ is valid

$$\check{\mu}(P_1, P_3) \leqslant \mathbb{K}_\mu[\hat{\mu}(P_1, P_2) + \check{\mu}(P_2, P_3)] \quad P_1, P_2, P_3 \in \mathscr{P}_1 \tag{3.3.31}$$

where $\hat{\mu}$ is the minimal metric w.r.t. μ.

Proof. We shall prove Inequality (3.3.31) only. For each $\varepsilon > 0$ define laws $P_{12}, P_{13} \in \mathscr{P}_2$ such that

$$T_1 P_{12} = P_1 \quad T_2 P_{12} = P_2 \quad T_1 P_{13} = P_1 \quad T_2 P_{13} = P_3$$

and

$$\hat{\mu}(P_1, P_2) \geqslant \mu(P_{12}) - \varepsilon, \quad \check{\mu}(P_1, P_3) \leqslant \mu(P_{13}) + \varepsilon.$$

As in Theorem 3.2.1, let us define a law $Q \in \mathscr{P}_3$ (cf. (3.2.5)) having marginals $T_{12}Q = P_{12}$, $T_{13}Q = P_{13}$. By Definitions 2.3.1, 3.2.2 and 3.3.3 we have

$$\begin{aligned}\check{\mu}(P_1, P_3) &\leqslant \mu(T_{13}Q) + \varepsilon \leqslant \mathbb{K}_\mu[\mu(P_{12}) + \mu(P_{23})] + \varepsilon \\ &\leqslant \mathbb{K}_\mu[\hat{\mu}(P_1, P_2) + \varepsilon + \check{\mu}(P_2, P_3)] + \varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$ we get Equation (3.3.31). QED

Definition 3.3.4. The moment functions $\check{\mu}$ will be called *a maximal distance with parameter* $\mathbb{K}_{\check{\mu}} = \mathbb{K}_\mu$ and if $\mathbb{K}_\mu = 1$, then $\check{\mu}$ will be called *maximal metric.*

As before, we note that a maximal distance (resp. metric) may fail to be distance (resp. metric). (The **ID** property may fail.)

Corollary 3.3.1. If (U, d) is an u.m. s.m.s. and μ is a compound metric on $\mathscr{P}_2$ then

$$|\check{\mu}(P_1, P_3) - \check{\mu}(P_2, P_3)| \leqslant \hat{\mu}(P_1, P_2) \tag{3.3.32}$$

for all $P_1, P_2, P_3 \in \mathscr{P}_1$.

Remark 3.3.2. By the triangle inequality $\mathbf{TI}^{(3)}$ we have

$$|\check{\mu}(P_1, P_3) - \check{\mu}(P_2, P_3)| \leqslant \check{\mu}(P_1, P_2). \tag{3.3.33}$$

Inequality (3.3.32) thus gives us refinement of the triangle inequality for maximal metrics.

We shall further investigate the following problem, which is related to a description of the minimal and maximal distances.

Problem 3.3.2. If c is a non-negative continuous function on U^2 and

$$\mu_c(P) := \int_{U^2} c(x, y) P(\mathrm{d}x, \mathrm{d}x) \quad P \in \mathscr{P}_2 \tag{3.3.34}$$

then what are the best possible inequalities of the type

$$\phi(P_1, P_2) \leqslant \mu_c(P) \leqslant \psi(P_1, P_2) \tag{3.3.35}$$

when the marginals $T_i P = P_i$, $i = 1, 2$ are fixed?

If $c(x, y) = H(d(x, y))$, $H \in \mathscr{H}$ then $\mu_c = \mathscr{L}_H$ (see Equation (3.3.1)) and the best possible lower and upper bounds for $\mathscr{L}_H(P)$ (with fixed $P_i = T_i P$ $(i = 1, 2)$) are given by the minimal distance $\phi(P_1, P_2) = \hat{\mathscr{L}}_H(P_1, P_2)$ and the maximal distance $\psi(P_1, P_2) = \check{\mathscr{L}}_H(P_1, P_2)$. For more general functions c the dual and explicit representations of $\hat{\mu}_c$ and $\hat{\mu}_c$ will be discussed later (Chapter 8).

Remark 3.3.3. In particular, for any convex non-negative function ψ on $\mathbb{R}$ and $c(x, y) = \psi(x - y)$ $(x, y \in \mathbb{R})$, the functionals of $\hat{\mathscr{L}}_H$ and $\mathscr{L}_H$ have the following explicit forms

$$\hat{\mathscr{L}}_H(P_1, P_2) := \int_0^1 H(F_1^{-1}(t) - F_2^{-1}(t))\,\mathrm{d}t$$

$$\check{\mathscr{L}}_H(P_1, P_2) := \int_0^1 H(F_1^{-1}(t) - F_2^{-1}(1 - t))\,\mathrm{d}t$$

where F_i^{-1} is the generalized inverse function (3.2.16) w.r.t. the d.f. F_i (see further Section 8.1).

Another example of a moment function that is an upper bound for $\mathscr{L}_H (H \in \mathscr{H})$ is given by

$$\Lambda_{H,\mathbf{0}}(P_1, P_2) := K_H \int_U H(d(x, \mathbf{0}))(P_1 + P_2)(\mathrm{d}x) \tag{3.3.36}$$

where $\mathbf{0}$ is a fixed point of U. In fact, since $H \in \mathscr{H}$ then $H(d(x, y)) \leqslant K_H[H(d(x, \mathbf{0})) + H(d(y, \mathbf{0}))]$ for all $x, y \in U$ and hence

$$\mathscr{L}_H(P) \leqslant \bar{\Lambda}_{H,\mathbf{0}}(P_1, P_2). \tag{3.3.37}$$

One can easily improve Inequality (3.3.37) by the following inequality

$$\mathscr{L}_H(P) \leqslant \bar{\Lambda}_H(P_1, P_2) := \inf_{a \in U} \bar{\Lambda}_{H,a}(P_1, P_2). \tag{3.3.38}$$

The upper bounds $\bar{\Lambda}_{H,a}$, $\bar{\Lambda}_H$ of $\mathscr{L}_H$ depend on the sum $P_1 + P_2$ only, hence, if P is an unknown law in $\mathscr{P}_2$ and we know only the sum of marginals $P_1 + P_2 = T_1P + T_2P$, then the best improvement of Inequality (3.3.38) is given by

$$\mathscr{L}_H(P) \leqslant \mathscr{L}_H^{(s)}(P_1 + P_2) \tag{3.3.39}$$

where

$$\mathscr{L}_H^{(s)}(P_1 + P_2) := \sup\{\mathscr{L}_H(P)\colon T_1P + T_2P = P_1 + P_2\}. \tag{3.3.40}$$

Remark 3.3.9. Following Remark 3.3.1, we have that if (X, Y) is a pair of dependent U-valued r.v.s, and we know only the sum of distributions $\Pr_X + \Pr_Y$, then $\mathscr{L}_1^{(s)}(\Pr_X + \Pr_Y)$ is the best possible improvement of the triangle inequality (3.3.28). Further (Section 8.1), we shall prove that in the particular case $U = \mathbb{R}$, $d(x, y) = |x - y|$, and $p \geqslant 1$

$$\mathscr{L}_p^{(s)}(P_1 + P_2) = \left(\int_0^1 |V^{-1}(t) - V^{-1}(1 - t)|^p \, dt\right)^{1/p}$$

where V^{-1} is the generalized inverse (see Equation (3.2.16)) of $V(t) = \frac{1}{2}(F_1(t) + F_2(t))$, $t \in \mathbb{R}$ and F_i is the d.f. of P_i $(i = 1, 2)$.

For more general cases we shall use the following definition.

Definition 3.3.5. For any compound distance μ, the functional

$$\overset{(s)}{\mu}(P_1 + P_2) := \sup\{\mu(P)\colon T_1P + T_2P = P_1 + P_2\}$$

will be called the *μ-upper bound with marginal sum fixed.*

Let us consider another possible improvement of Minkovski's inequality (3.3.24). Suppose we need to estimate from above (in the best possible way) the value $\mathscr{L}_p(X, Y)(p > 0)$ having available only the moments

$$m_p(X) := [\mathbb{E}d^p(X, \mathbf{0})]^{p'} \quad p' := \min(1, 1/p) \tag{3.3.41}$$

and $m_p(Y)$. Then the problem consists in evaluating the quantity

$$\psi_p(a_1, a_2) := \sup\left\{\mathscr{L}_p(P)\colon P \in \mathscr{P}(U), \left(\int_U d^p(x, \mathbf{0}) T_i P(dx)\right)^{p'} = a_i, i = 1, 2\right\}$$
$$p' = \min(1, 1/p)$$

for each $a_1 \geqslant 0$ and $a_2 \geqslant 0$.

Obviously, ψ_p is a moment function. Further, (see Section 9.1) we shall obtain an explicit representation of $\psi_p(a_1, a_2)$.

Definition 3.3.6. For any p. distance μ, the function

$$\overset{(m,p)}{\mu}(a_1, a_2) := \sup\left\{\mu(P)\colon P \in \mathscr{P}_2(U), \left(\int_U d^p(x, \mathbf{0}) T_i P(\mathrm{d}x)\right)^{p'} = a_i,\ i = 1, 2\right\}$$

where $a_1 \geqslant 0$, $a_2 \geqslant 0$, $p > 0$ is said to be the *μ-upper bound with fixed pth marginal moments* a_1 and a_2.

Hence, $\overset{(m,1)}{\mathscr{L}}(a_1, a_2)$ is the best possible improvement of the triangle inequality (3.3.28) when we know only the 'marginal' moments

$$a_1 = \mathbb{E}d(X, \mathbf{0}) \quad a_2 = \mathbb{E}d(Y, \mathbf{0}).$$

We shall investigate improvements of inequalities of the type

$$\mathbb{E}d(X, \mathbf{0}) - \mathbb{E}d(Y, \mathbf{0}) \leqslant \mathbb{E}d(X, Y) \leqslant \mathbb{E}d(X, \mathbf{0}) + \mathbb{E}d(Y, \mathbf{0})$$

for dependent r.v.s X and Y. We make the following definition.

Definition 3.3.7. For any p. distance μ,

(i) the functional

$$\underset{(m,p)}{\mu}(a_1, a_2) := \inf\left\{\mu(P)\colon P \in \mathscr{P}(U), \left[\int_U d^p(x, \mathbf{0}) T_i P(\mathrm{d}x)\right]^{p'} = a_i,\ i = 1, 2\right\}$$

where $a_1 \geqslant 0$, $a_2 \geqslant 0$, $p > 0$ is said to be the *μ-lower bound with fixed marginal pth moments* a_1 and a_2;

(ii) the functional

$$\bar{\mu}(a_1 + a_2; m, p) := \sup\left\{\mu(P)\colon P \in \mathscr{P}_2(U), \left[\int_U d^p(x, \mathbf{0}) T_1 P(\mathrm{d}x)\right]^{p'} + \left[\int_U d^p(x, \mathbf{0}) T_2 P(\mathrm{d}x)\right]^{p'} = a_1 + a_2\right\}$$

where $a_1 \geqslant 0$, $a_2 \geqslant 0$, $p > 0$ is said to be the *μ-upper bound with fixed sum of marginal pth moments* $a_1 + a_2$;

(iii) the functional

$$\underline{\mu}(a_1 - a_2; m, p) := \inf\left\{\mu(P)\colon P \in \mathscr{P}_2(U), \left[\int_U d^p(x, \mathbf{0}) T_1 P(\mathrm{d}x)\right]^{p'} - \left[\int_U d^p(x, \mathbf{0}) T_2 P(\mathrm{d}x)\right]^{p'} = a_1 - a_2\right\}$$

where $a_1 \geqslant 0$, $a_2 \geqslant 0$, $p > 0$ is said to be the *μ-lower bound with fixed difference* of marginal p. moments $a_1 - a_2$.

Knowing explicit formulae for $\overset{(m,p)}{\mu}$ and $\underset{(m,p)}{\mu}$ (see Section 9.1), we can easily determine $\bar{\mu}(a_1 + a_2; m, p)$ and $\bar{\mu}(a_1 - a_2; m, p)$ by using the representations

$$\bar{\mu}(a; m, p) = \sup\left\{\overset{(m,p)}{\mu}(a_1, a_2): a_1 \geqslant 0, a_2 \geqslant 0, a_1 + a_2 = a\right\}$$

and

$$\underline{\mu}(a; m, p\} = \inf\left\{\underset{(m,p)}{\mu}(a_1, a_2): a_1 \geqslant 0, a_2 \geqslant 0, a_1 - a_2 = a\right\}.$$

Let us summarize the bounds for μ we have obtained up to now. For any compound distance μ (see Fig. 3.3.1), the maximal distance $\check{\mu}$ (see Definition 3.3.9) is not greater than the moment distance

$$\overset{(m,p)}{\mu}(a_1, a_2) := \sup\left\{\mu(P_1, P_2): \left[\int_U d^p(x, 0)P_i(\mathrm{d}x)\right]^{p'} = a_i, i = 1, 2\right\}. \tag{3.3.42}$$

As we have seen, all compound distances μ can be estimated from above by means of $\check{\mu}$, $\overset{(s)}{\mu}$, $\overset{(m,p)}{\mu}$, $\mu(\cdot; m, p)$ and in addition, the following inequality holds

$$\mu \leqslant \check{\mu} \leqslant \overset{(s)}{\mu} \leqslant \bar{\mu}(\cdot; m, p), \check{\mu} \leqslant \overset{(m,p)}{\mu}. \tag{3.3.43}$$

The p. distance μ can be estimated from below by means of the minimal metric $\hat{\mu}$ (see Definition 3.2.2), the co-minimal metric $\mu\nu$ (see Definition 3.2.4), the primary minimal distance $\tilde{\mu}_h$ (see Definition 3.1.2), as well as for such μ as $\mu = \mu_c$ (see Equation (3.2.40)) by means of minimal norms $\mathring{\mu}_c$ (see Definition 3.2.5).

Thus

$$\underline{\mu}(\cdot; m, p) \leqslant \tilde{\mu}_h \leqslant \hat{\mu} \leqslant \mu\nu \leqslant \mu, \mathring{\mu}_c \leqslant \mu_c \tag{3.3.44}$$

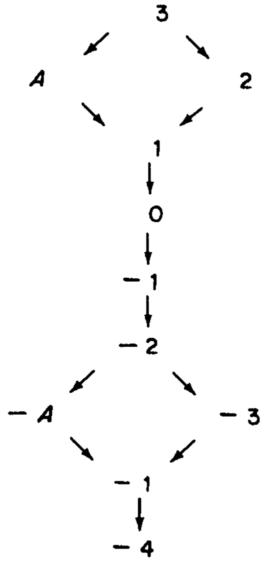
and moreover, we can compute the values of $\tilde{\mu}_h$ by using the values of the minimal distances μ, since

$$\tilde{\mu}_h(a_1, a_2) = (\tilde{\hat{\mu}})_h(a_1, a_2) := \inf\{\hat{\mu}(P_1, P_2): hP_i = a_i, i = 1, 2\} \tag{3.3.45}$$

Also, if $c(x, y) = H(\mathrm{d}(x, y)), H \in \mathscr{H}$, then μ_c is a p. distance and

$$\mathring{\mu}_c \leqslant \hat{\mu}_c \leqslant \mu. \tag{3.3.46}$$

The inequalities (3.3.42)–(3.3.46) are represented in the following scheme



$3 \to \mu$—upper bound with fixed sum of marginal moments— $\mu(\,\cdot\,; m, p)$

$2 \to \mu$—upper bound with fixed marginal moments—$\overset{(m,p)}{\mu}$

$A \to \mu$—upper bound with marginal sum fixed—$\overset{(s)}{\mu}$

$1 \to$ maximal distance $\check{\mu}$

$0 \to$ compound distance μ

$-1 \to$ co-minimal distance $\mu\nu$

$-2 \to$ minimal distance $\hat{\mu}$

$-A \to$ minimal norm $\mathring{\mu}$ (for $\mu = \mu_c$)

$-3 \to$ primary minimal distance μ_h (in particular, lower bound with fixed marginal moments $\underset{(m,p)}{\mu}$)

$-4 \to \mu$—lower bound with fixed difference of marginal moments—$\underline{\mu}\,(\,\cdot\,; m, p)$

Figure 3.3.1 The lower and upper bounds for values $\mu(P)(P \in \mathscr{P}_2)$ of a s compound distance μ when different kinds of marginal characteristics of P are fixed. (From Rachev and Shortt, 1990. Reproduced by permission of the American Mathematical Society.)

The functionals labeled -4, -3, -2, -1, 0, 1, 2, 3 are listed in order of numerical size, however, those labeled A and $-A$ do not fit into this scheme, e.g. the functional $\overset{(s)}{\mu}$ labeled by (A) dominates $\check{\mu}$ (labeled (1)), but $\overset{(s)}{\mu}$ and $\overset{(m,p)}{\mu}$ are not generally comparable. The left margin of Fig. 3.3.1 indicates the ordering of these functionals.

As an example illustrating the list of bounds in Fig. 3.3.1 let us consider the case $p = 1$ and $\mu(X, Y) = \mathbb{E}d(X, Y)$. Then for a fixed point $\mathbf{0} \in U$

$$(*) \qquad \mu(a_1 + a_2; m, 1) = \sup\{\mathbb{E}d(X, Y)\colon \mathbb{E}d(X, \mathbf{0}) + \mathbb{E}d(Y, \mathbf{0}) = a_1 + a_2\} \qquad a_1 + a_1 \geqslant 0 \tag{3.3.47}$$

$$(**) \qquad \overset{(m,1)}{\mu}(a_1, a_2) = \sup\{\mathbb{E}d(X, Y)\colon \mathbb{E}d(X, \mathbf{0}) = a_1,\ \mathbb{E}d(Y, \mathbf{0}) = a_2\}, \qquad a_1 \geqslant 0 \quad a_2 \geqslant 0 \tag{3.3.48}$$

$$(***) \qquad \overset{(s)}{\mu}(P_1 + P_2) = \sup\{\mathbb{E}d(X, Y)\colon \mathrm{Pr}_X + \mathrm{Pr}_Y = P_1 + P_2\} \quad P_1, P_2 \in \mathscr{P}_1 \tag{3.3.49}$$

$$(****) \quad \check{\mu}(P_1, P_2) = \sup\{\mathbb{E}d(X, Y)\colon \mathrm{Pr}_X = P_1, \mathrm{Pr}_Y = P_2\} \quad P_1, P_2 \in \mathscr{P}_1 \tag{3.3.50}$$

and each of these functionals gives the best possible refinement of the inequality

$$\mathbb{E}d(X, Y) \leqslant \mathbb{E}d(X, \mathbf{0}) + \mathbb{E}d(Y, \mathbf{0})$$

under the respective conditions

(*) $$\mathbb{E}d(X, \mathbf{0}) + \mathbb{E}d(Y, \mathbf{0}) = a_1 + a_2$$

(**) $$\mathbb{E}d(X, \mathbf{0}) = a_1 \quad \mathbb{E}d(Y, \mathbf{0}) = a_2$$

(***) $$\mathrm{Pr}_X + \mathrm{Pr}_Y = P_1 + P_2$$

(****) $$\mathrm{Pr}_X = P_1, \mathrm{Pr}_Y = P_2.$$

Analogously, the functionals

(i) $$\underline{\mu}(a_1 - a_2; m, 1) = \inf\{\mathbb{E}d(X, Y): \mathbb{E}d(X, \mathbf{0}) - \mathbb{E}d(Y, \mathbf{0}) = a_1 - a_2\} \quad a_1, a_2 \in \mathbb{R} \tag{3.3.51}$$

(ii) $$\underset{(m, 1)}{\mu}(a_1, a_2) = \inf\{\mathbb{E}d(X, Y): \mathbb{E}d(X, \mathbf{0}) = a_1, \mathbb{E}d(Y, \mathbf{0}) = a_2\} \quad a_1 \geqslant 0 \quad a_2 \geqslant 0 \tag{3.3.52}$$

(iii) $$\mathring{\mu}(P_1, P_2) = \inf\{\alpha\mathbb{E}d(X, Y): \text{for some } \alpha > 0, X \in \mathfrak{X}, Y \in \mathfrak{X} \text{ such that} \tag{3.3.53}$$
$$\alpha(\mathrm{Pr}_X - \mathrm{Pr}_Y) = P_1 - P_2\} \quad P_1, P_2 \in \mathscr{P}_2$$

(iv) $$\hat{\mu}(P_1, P_2) = \inf\{\mathbb{E}d(X, Y): \mathrm{Pr}_X = P_1, \mathrm{Pr}_Y = P_2\} \quad P_1, P_2 \in \mathscr{P}_1 \tag{3.3.54}$$

(v) $$\mu\nu(P_1, P_2, \alpha) = \inf\{\mathbb{E}d(X, Y): \mathrm{Pr}_X = P_1, \mathrm{Pr}_Y = P_2, \nu(X, Y) < \alpha\} \tag{3.3.55}$$
$$(P_1, P_2 \in \mathscr{P}_1, \nu \text{ is a p. distance in } \mathfrak{X}(U))$$

describe the best possible refinement of the inequality

$$\mathbb{E}d(X, Y) \geqslant \mathbb{E}d(X, \mathbf{0}) - \mathbb{E}d(Y, \mathbf{0})$$

under the respective conditions,

(i) $$\mathbb{E}d(X, \mathbf{0}) - \mathbb{E}d(Y, \mathbf{0}) = a_1 - a_2$$

(ii) $$\mathbb{E}d(X, \mathbf{0}) = a_1 \quad \mathbb{E}d(Y, \mathbf{0}) = a_2$$

(iii) $$\alpha(\mathrm{Pr}_X - \mathrm{Pr}_Y) = P_1 - P_2 \quad \text{for some } \alpha > 0$$

(iv) $$\mathrm{Pr}_X = P_1 \quad \mathrm{Pr}_Y = P_2$$

(v) $$\mathrm{Pr}_X = P_1 \quad \mathrm{Pr}_Y = P_2 \quad \nu(X, Y) < \alpha.$$

Remark 3.3.4. If $\mu(X, Y) = \mathbb{E}d(X, Y)$, then $\mathring{\mu} = \hat{\mu}$ (see further Theorem 6.1.1), hence, in this case,

$$\mathring{\mu}(P_1, P_2) = \inf\{\mathbb{E}d(X, Y): \mathrm{Pr}_X - \mathrm{Pr}_Y = P_1 - P_2\}. \tag{3.3.56}$$

CHAPTER 4

A Structural Classification of the Probability Distances

Chapter 3 was devoted to a classification of p. (semi-)distances $\mu(P)$ ($P \in \mathscr{P}_2$) with respect to various partitionings of the set $\mathscr{P}_2$ into classes $\mathscr{P}\mathscr{C}$ such that $\mu(P)$ takes a constant value on each $\mathscr{P}\mathscr{C}$. For instance, if $\mathscr{P}\mathscr{C} := \mathscr{P}\mathscr{C}(P_1, P_2) := \{P \in \mathscr{P}_2\colon T_1P = P_1,\ T_2P = P_2\}$, P_1, $P_2 \in \mathscr{P}_1$ and $\mu(P') = \mu(P'')$ for each P', $P'' \in \mathscr{P}\mathscr{C}$ then μ was said to be a simple semidistance. Analogously, if

$$\mathscr{P}\mathscr{C} := \mathscr{P}\mathscr{C}(\bar{a}_1, \bar{a}_2) := \{P \in \mathscr{P}_2\colon h_1P = \bar{a}_1,\ h_2P = \bar{a}_2\}$$

(cf. (3.1.2) and Definition 3.1.1) and $\mu(P') = \mu(P'')$ as P', $P'' \in \mathscr{P}\mathscr{C}(\bar{a}_1, a_2)$ then μ was said to be a primary distance.

In the present section, we shall classify the probability semidistances on the basis of their metric structure. For example, a p. metric which admits a representation as a Hausdorff metric (cf. (2.4.1)) will be called a metric with Hausdorff structure. See, for instance, the H-metric introduced in Section 2.2. On the other hand, some simple p. distances μ can be represented as $\zeta_{\mathscr{F}}$-metrics, namely,

$$\mu(P_1, P_2) = \zeta_{\mathscr{F}}(P_1, P_2) := \sup_{f \in \mathscr{F}} \left| \int f \mathrm{d}(P_1 - P_2) \right| \quad P_i \in \mathscr{P} \subset \mathscr{P}(U)$$

where $\mathscr{F}$ is a class of functions on a s.m.s. U which are P-integrable for any $P \in \mathscr{P}$. In this case, μ is said to be p. metric with ζ-structure. Examples of such μ are the Kantorovich metric ℓ_1 (3.2.12), the total variation metric $\underline{\sigma}$ (3.2.13), the Kolmogorov metric $\boldsymbol{\rho}$ (2.1.2), and the $\boldsymbol{\theta}_p$-metric (Remark 2.1.2).

4.1 HAUSDORFF STRUCTURE OF p. SEMIDISTANCES

The definition of Hausdorff p. semidistance structure (briefly, h-structure) is based on the notion of *Hausdorff semimetric* in the space of all subsets of a given metric space (S, ρ):

$$\begin{aligned} r(A, B) &= \inf\{\varepsilon > 0\colon A^\varepsilon \supseteq B,\ B^\varepsilon \supseteq A\} \\ &= \max\{\inf\{\varepsilon > 0\colon A^\varepsilon \supseteq B\}, \inf\{\varepsilon > 0\colon B^\varepsilon \supseteq A\}] \qquad (4.1.1) \end{aligned}$$

where A^ε is the open ε-neighborhood of A.

From Definition (4.1.1) follows immediately the second Hausdorff semi-distance representation

$$r(A, B) := \max(r', r'') \tag{4.1.2}$$

where

$$r' := \sup_{x \in A} \inf_{y \in B} \rho(x, y)$$

and

$$r'' := \sup_{y \in B} \inf_{x \in A} \rho(x, y).$$

As an example of a p. metric with representation closed to that of Equality (4.1.2) let us consider the following *parametric version of the Lévy metric* for $\lambda > 0$, $X, Y \in \mathfrak{X}(\mathbb{R})$ (see Fig. 4.1.1)

$$\mathbf{L}_\lambda(X, Y) := \mathbf{L}_\lambda(F_X, F_Y) := \inf\{\varepsilon > 0\colon F_X(x - \lambda\varepsilon) - \varepsilon \leqslant F_Y(x) \leqslant F_X(x + \lambda\varepsilon) + \varepsilon \quad \forall x \in \mathbb{R}\}. \tag{4.1.3}$$

Obviously, $\mathbf{L}_\lambda$ is a simple metric in $\mathfrak{X}(\mathbb{R})$ for any $\lambda > 0$, and $\mathbf{L} := \mathbf{L}_1$ is the usual Lévy metric (cf. (2.1.3)). Moreover, it is not difficult to check that $\mathbf{L}_\lambda(F, G)$ is a metric in the space $\mathscr{F}$ of all d.f.s. Considering $\mathbf{L}_\lambda$ as a function of

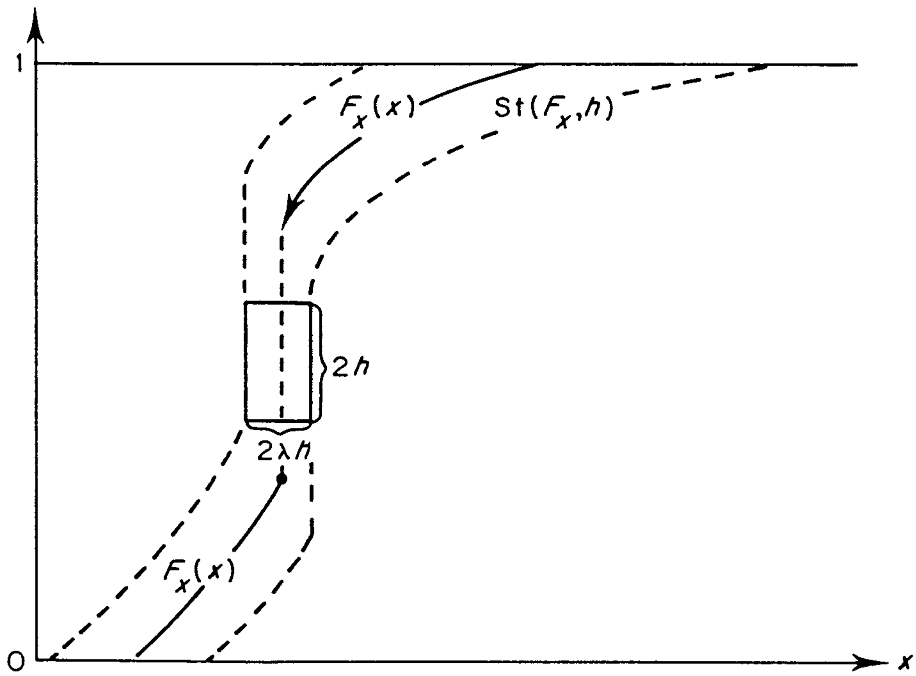


Figure 4.1.1 $\mathrm{St}(F_X, h)$ is the strip in which the graph of F_Y has to be positioned in order that the inequality $L_\lambda(X, Y) \leqslant h$ obtains.

λ we see that $\mathbf{L}_\lambda$ is non-increasing on $(0, \infty)$ and the following limit relations hold

$$\lim_{\lambda \to 0} \mathbf{L}_\lambda(F, G) = \boldsymbol{\rho}(F, G) \quad F, G \in \mathscr{F} \tag{4.1.4}$$

and

$$\lim_{\lambda \to 0} \lambda \mathbf{L}_\lambda(F, G) = \mathbf{W}(F, G). \tag{4.1.5}$$

In Equality (4.1.4), $\boldsymbol{\rho}$ is the *Kolmogorov metric* (see (2.1.2)) in $\mathscr{F}$

$$\boldsymbol{\rho}(F, G) := \sup_{x \in \mathbb{R}} |F(x) - G(x)|. \tag{4.1.6}$$

In Equality (4.1.5), $\mathbf{W}(F, G)$ is the *uniform metric between the inverse functions* F^{-1}, G^{-1}

$$\mathbf{W}(F, G) := \sup_{0 < t < 1} |F^{-1}(t) - G^{-1}(t)| \tag{4.1.7}$$

where F^{-1} is the generalized inverse of F

$$F^{-1}(t) := \sup\{x \colon F(x) < t\}. \tag{4.1.8}$$

The Equality (4.1.4) follows from (4.1.3) (see Fig. 4.1.1). Likewise, (4.1.5) is handled by the equalities

$$\begin{aligned} \lim_{\lambda \to \infty} \lambda \mathbf{L}_\lambda(F, G) &= \inf\{\delta > 0 \colon F(x) \leqslant G(x + \delta), G(x) \leqslant F(x + \delta) \quad \forall x \in \mathbb{R}\} \\ &= \mathbf{W}(F, G). \end{aligned}$$

Another way to prove Equation (4.1.5) is to use the representation of $\mathbf{L}_\lambda(F, G)$ in terms of the inverse functions F^{-1} and G^{-1}

$$\begin{aligned} \mathbf{L}_\lambda(F, G) &= \inf\{\varepsilon > 0 \colon F_X^{-1}(t - \varepsilon) - \lambda\varepsilon \leqslant F_Y^{-1}(t), F_Y^{-1}(t - \varepsilon) - \lambda\varepsilon \leqslant F_X^{-1}(t) \\ &\qquad \forall \varepsilon \leqslant t \leqslant 1\} \\ &= \frac{1}{\lambda} \inf\left\{\delta > 0 \colon F_X^{-1}\left(t - \frac{1}{\lambda}\delta\right) - \delta \leqslant F_Y^{-1}(t), F_Y^{-1}\left(t - \frac{1}{\lambda}\delta\right) - \delta \leqslant F_X^{-1}(t) \right. \\ &\qquad \left. \forall \frac{1}{\lambda}\delta \leqslant t \leqslant 1\right\}. \end{aligned}$$

We shall prove further (Corollary 7.3.1 and (7.4.15)) that $\mathbf{W}$ coincides with the ℓ_∞-metric

$$\ell_\infty(F_1, F_2) := \ell_\infty(P_1, P_2) := \inf\{\varepsilon > 0 \colon P_1(A) \leqslant P_2(A^\varepsilon) \quad \forall A \subset \mathbb{R}\}$$

where P_i is the law determined by F_i. The equality $\mathbf{W} = \ell_\infty$ illustrates, together with Equality (4.1.5), the main relationship between the Lévy metric and ℓ_∞.

Let us define the Hausdorff metric between two bounded functions on the real line $\mathbb{R}$. Let $dm_\lambda(\lambda > 0)$ be the Minkovski metric on the plane $\mathbb{R}^2$, i.e. for each $A = (x_1, y_1)$ and $B = (x_2, y_2)$ we have $dm_\lambda(A, B) := \max\{(1/\lambda)|x_1 - x_2|, |y_1 - y_2|\}$. *The Hausdorff metric* $r_\lambda(\lambda > 0)$ *in the set* $\mathscr{C}(\mathbb{R}^2)$ (of all closed non-empty sets $G \subseteq \mathbb{R}^2$) is defined as follows: for $G_1 \subseteq \mathbb{R}^2$ and $G_2 \subseteq \mathbb{R}^2$

$$r_\lambda(G_1, G_2) := \max\left\{\sup_{A \in G_1} \inf_{B \in G_2} dm_\lambda(A, B), \sup_{B \in G_2} \inf_{A \in G_1} dm_\lambda(A, B)\right\}. \tag{4.1.9}$$

We shall say that r_λ is generated by the metric dm_λ as in Equality (4.1.2) the Hausdorff distance r was generated by ρ. Let $f \in D(\mathbb{R})$, the set of all bounded right-continuous functions on $\mathbb{R}$ having limits $f(x-)$ from the left. The set

$$\bar{f} = \{(x, y): x \in \mathbb{R} \text{ and either } f(x-) \leqslant y \leqslant f(x) \text{ or } f(x) \leqslant y \leqslant f(x-)\}$$

is called the *completed graph* of the function f.

Remark 4.1.1. Obviously, the completed graph $\bar{F}$ of a d.f. $F \in \mathscr{F}$ is given by

$$\bar{F} := \{(x, y): x \in \mathbb{R}, F(x-) \leqslant y \leqslant F(x)\}. \tag{4.1.10}$$

Using Equality (4.1.9) we define the Hausdorff metric $r_\lambda = r_\lambda(\bar{f}, \bar{g})$ in the space of completed graphs of bounded, right-continuous functions.

Definition 4.1.1. The metric

$$r_\lambda(f, g) := r_\lambda(\bar{f}, \bar{g}) \quad f, g \in D(\mathbb{R}) \tag{4.1.11}$$

is said to be the *Hausdorff metric in* $D(\mathbb{R})$.

Lemma 4.1.1 (Sendov, 1969). For any $f, g \in D(\mathbb{R})$

$$r_\lambda(f, g) = \max\left\{\sup_{x \in \mathbb{R}} \inf_{(x_2, y_2) \in \bar{g}} dm_\lambda((x, f(x)), (x_2, y_2)), \sup_{x \in \mathbb{R}} \inf_{(x_1, y_1) \in \bar{f}} dm_\lambda((x_1, y_1), (x, g(x))\right\}.$$

Proof. It is sufficient to prove that if for each $x_0 \in \mathbb{R}$ there exist points $(x_1, y_1) \in \bar{f}$, $(x_2, y_2) \in \bar{g}$ such that $\max\{(1/\lambda)|x_0 - x_1|, |g(x_0) - y_1|\} \leqslant \delta$, $\max\{(1/\lambda)|x_0 - x_2|, |f(x_0) - y_2|\} \leqslant \delta$ then $r_\lambda(f, g) \leqslant \delta$. Suppose the contrary is true. Then there exists a point (x_0, y_0) in the completed graph of one of the two functions, say $f(x)$, such that in the rectangle $|x - x_0| \leqslant \lambda\delta$, $|y - y_0| \leqslant \delta$ there is no point of the completed graph $\bar{g}$. Writing

$$y_0' = \min_{(x_0, y) \in \bar{f}} y, \quad y_0'' = \max_{(x_0, y) \in \bar{f}} y$$

we then have $y_0' \leqslant y_0 < y_0''$. From the definition of (x_0, y_0') and (x_0, y_0'') it follows that there exist two sequences $\{x_n'\}$ and $\{x_n''\}$ in $\mathbb{R}$, converging to x_0, such that $\lim_{n \to \infty} f(x_n') = y_0'$, $\lim_{n \to \infty} f(x_n'') = y_0''$. Then from the hypothesis and the fact that $\bar{g}$ is a closed set, it follows that there exist two points $(x_1, y_1), (x_2, y_2) \in \bar{g}$ for which $x_1, x_2 \in [x_0 - \lambda\delta, x_0 + \lambda\delta]$, $y_1 \leqslant y_0'$, $y_2 \geqslant y_0''$. This contradicts our assumptions since by the definition of the completed graph $\bar{g}$, there exists $\tilde{x}_0 \in [x_0 - \lambda\delta, x_0 + \lambda\delta]$ such that $(\tilde{x}_0, y_0) \in \bar{g}$. QED

Remark 4.1.2. Before proceeding to the proof of the fact that the Lévy metric is a special case of the Hausdorff metric (Theorem 4.1.1), we shall mention the following two properties of the metric $r_\lambda(f, g)$ that can be considered as generalizations of well known properties of the Lévy metric.

Property 4.1.1. Let ρ be the uniform distance in $D(\mathbb{R})$, i.e., $\rho(f, g) := \sup_{u \in \mathbb{R}} |f(u) - g(u)|$, and let $\omega_f(\delta) := \sup\{|f(u) - f(u')|: |u - u'| < \delta\}$, $f \in C_b(\mathbb{R})$, $\delta > 0$ be the modulus of f-continuity. Then

$$r_\lambda(f, g) \leqslant \rho(f, g) \leqslant r_\lambda(f, g) + \min(\omega_f(\lambda r_\lambda(f, g)), \omega_g(\lambda r_\lambda(f, g))) \tag{4.1.12}$$

Proof. If $r_\lambda(f, g) = \sup_{a \in \bar{f}} \inf_{b \in \bar{g}} dm_\lambda(a, b)$, then following the proof of Lemma 4.1.1 we have

$$\begin{aligned} r_\lambda(f, g) &= \sup_{x \in \mathbb{R}} \inf_{(x_2, y_2) \in \bar{g}} dm_\lambda((x, f(x)), (x_2, y_2)) \\ &\leqslant \sup_{x \in \mathbb{R}} \inf_{y \in \mathbb{R}} \max\left\{\frac{1}{\lambda}|x - y|, |f(x) - g(y)|\right\} \leqslant \rho(f, g). \end{aligned}$$

For any $x \in \mathbb{R}$ there exists $(y_0, z_0) \in \bar{g}$ such that

$$r_\lambda(f, g) \geqslant \inf_{(y, z) \in \bar{g}} dm_\lambda((x, f(x)), (y, z)) = \max\left(\frac{1}{\lambda}|x - y_0|, |f(x) - z_0|\right).$$

Hence

$$\begin{aligned} |f(x) - g(x)| &\leqslant |f(x) - z_0| + |g(x) - z_0| \\ &\leqslant r(f, g) + \max(|g(x) - g(y_0-)|, |g(x) - g(y_0)|) \\ &\leqslant r(f, g) + \omega_g(\lambda r_\lambda(f, g)). \end{aligned}$$

QED

As a consequence of Inequalities (4.1.12) we obtain the following property.

Property 4.1.2. Let $\{f_n(x), n = 1, 2, \ldots\}$ be a sequence in $D(\mathbb{R})$, and let $f(x)$ be a continuous bounded function on the line. The sequence $\{f_n\}$ converges uniformly on $\mathbb{R}$ to $f(x)$, if and only if $\lim_{n \to \infty} r_\lambda(f_n, f) = 0$.

Theorem 4.1.1. For all $F, G \in \mathscr{F}$ and $\lambda > 0$

$$\mathbf{L}_\lambda(F, G) = r_\lambda(F, G). \tag{4.1.13}$$

Proof. Consider the completed graphs $\bar{F}$ and $\bar{G}$ of the d.f.s F and G and denote by P and Q the points where they intersect the line $(1/\lambda)x + y = u$, where u can be any real number. Then

$$\mathbf{L}_\lambda(F, G) = \max_{u \in \mathbb{R}} |PQ|(1 + \lambda^2)^{-1/2} \tag{4.1.14}$$

where $|PQ|$ is the length of the segment joining the points P and Q. (The proof of Equation (4.1.14) is quite analogous to that given in Hennequin and Tortrat (1965), Chapter 19, for the case $\lambda = 1$.) We shall show that $r_\lambda(F, G) \leqslant \mathbf{L}_\lambda(F, G)$ by applying Lemma 4.1.1.

Choose a point $x_0 \in \mathbb{R}$. The line $(1/\lambda)x + y = (1/\lambda)x_0 + F(x_0)$ intersects $\bar{F}$ and $\bar{G}$ at the points $P(x_0, F(x_0))$ and $Q(x_1, y_1)$. It follows from Equation (4.1.14) that $|F(x_0) - y_1| \leqslant \mathbf{L}_\lambda(F, G)$ and $(1/\lambda)|x_0 - x_1| \leqslant \mathbf{L}_\lambda(F, G)$. Permuting F and G, we find that for some $(x_2, y_2) \in \bar{F}$

$$\max\left[\frac{1}{\lambda}|x_0 - x_2|, |G(x_0) - y_2|\right] \leqslant \mathbf{L}_\lambda(F, G).$$

By Lemma 4.1.1, this means that $r_\lambda(F, G) \leqslant \mathbf{L}_\lambda(F, G)$.

Now let us show the reverse inequality. Assume otherwise, i.e. assume $\mathbf{L}_\lambda(F, G) > r_\lambda(F, G)$. Let $P_0(x', y')$ and $Q_0(x'', y'')$ be points such that

$$\mathbf{L}_\lambda(F, G) = \frac{|P_0Q_0|}{(1 + \lambda^2)^{1/2}} > r_\lambda(F, G).$$

Suppose that $x' < x''$. Since the points P_0 and Q_0 lie on some $(1/\lambda)x + y = u_0$, and, say $u_0 > 0$, we have $y' > y''$. By the definition of the metric $r_\lambda(F, G)$ and our assumptions, it follows that

$$\frac{|P_0Q_0|}{(1 + \lambda^2)^{1/2}} > \max_{A \in \bar{F}} \min_{B \in \bar{G}} dm_\lambda(A, B)$$

see Equation (4.1.9). Since $P_0 \in \bar{F}$, there exists a point $B_0(x^*, y^*) \in \bar{G}$ such that

$$\frac{|P_0Q_0|}{(1 + \lambda^2)^{1/2}} > \min_{B \in \bar{G}} dm_\lambda(P_0, B) = dm_\lambda(P_0, B_0).$$

Thus

$$dm_\lambda(P_0, B_0) = \max\left[\frac{1}{\lambda}|x' - x^*|, |y' - y^*|\right] < |P_0Q_0|(1 + \lambda^2)^{-1/2}. \tag{4.1.15}$$

Suppose that $x' \geqslant x^*$. Then $x^* \leqslant x' < x''$. The function G is non-decreasing, so $y^* \leqslant y''$, i.e.

$$y' - y^* \geqslant y' - y'' = \frac{|P_0Q_0|}{(1 + \lambda^2)^{1/2}}$$

which is impossible by virtue of (4.1.15). If $x' < x^*$ then

$$0 < \frac{1}{\lambda}(x^* - x') < \frac{|P_0 Q_0|}{(1+\lambda^2)^{1/2}} = \frac{1}{\lambda}(x'' - x').$$

Then $x^* < x''$ and $y^* \leqslant y''$, which is impossible, as we have proved. Thus $\mathbf{L}_\lambda(F, G) \leqslant r_\lambda(F, G)$. QED

In order to cover other p. metrics by means of the Hausdorff metric structure the following generalization of the notion of Hausdorff metric r is needed. Let $\mathcal{FS}$ be the space of all real-valued functions $F_A\colon A \to \mathbb{R}$ where A is a subset of the metric space (S, ρ).

Definition 4.1.2. Let $f = f_A$ and $g = g_B$ be elements of $\mathcal{FS}$. The quantity

$$\tilde{r}_\lambda(f, g) := \max(\tilde{r}'_\lambda(f, g), \tilde{r}'_\lambda(g, f)) \tag{4.1.16}$$

where

$$\tilde{r}'_\lambda(f, g) := \sup_{x \in A} \inf_{y \in B} \max\left\{\frac{1}{\lambda}\rho(x, y), f(x) - g(y)\right\}$$

is called the *Hausdorff semimetric* between the functions f_A, g_B.

Obviously, if $f(x) = g(y) =$ constant for all $x \in A$, $y \in B$ then $\tilde{r}_\lambda(f, g) = r(A, B)$ (cf. Equation (4.1.2)). Note that $\tilde{r}_\lambda$ is a metric in the space of all upper semi-continuous functions with closed domains.

The next two theorems are simple consequences of a more general theorem, Theorem 4.2.1.

Theorem 4.1.2. The Lévy metric $\mathbf{L}_\lambda$ (4.1.3) admits the following representation in terms of metric $\tilde{r}$ (4.1.16)

$$\mathbf{L}_\lambda(X, Y) = \tilde{r}_\lambda(f_A, g_B) \tag{4.1.17}$$

where $f_A = F_X$, $g_B = F_Y$, $A \equiv B \equiv \mathbb{R}$, $\rho(x, y) = |x - y|$.

The Lévy metric $\mathbf{L}_\lambda$, thus, has two representations in terms of r_λ and in terms of $\tilde{r}_\lambda$. Concerning the Prokhorov metric π_λ (3.2.22) only a representation in terms of $\tilde{r}_\lambda$ is known. Namely, let $\mathcal{S} = \mathcal{C}((U, d))$ be the space of all closed non-empty subsets of a metric space (U, d) and let r be the Hausdorff distance (4.1.1) in $\mathcal{S}$. Any law $P \in \mathcal{P}_1(U)$ can be considered as a function on the metric space $(\mathcal{S}, r)$ because P is determined uniquely on $\mathcal{S}$, namely

$$P(A) := \sup\{P(C)\colon C \in \mathcal{S}, C \subseteq A\} \text{ for any } A \in \mathcal{B}_1.$$

Define a metric $\tilde{r}_\lambda(P_1, P_2)(P_1, P_2 \in \mathcal{P}(U))$ by putting $A = B = \mathcal{S}$ and $\rho = r$ in Equality (4.1.16).

Theorem 4.1.3. For any $\lambda > 0$, the Prokhorov metric π_λ takes the form

$$\pi_\lambda(P_1, P_2) = \tilde{r}_\lambda(P_1, P_2) \qquad (P_1, P_2 \in \mathscr{P}_1(U))$$

where $U = (U, d)$ is assumed to be arbitrary metric space.

Remark 4.1.3. By Theorem 4.1.3, for all $P_1, P_2 \in \mathscr{P}_1$ we have the following Hausdorff representation of the Prokhorov metric π_λ, $\lambda > 0$

$$\pi_\lambda(P_1, P_2) := \max\left\{\sup_{A \in \mathscr{B}_1} \inf_{B \in \mathscr{B}_1} \max\left[\frac{1}{\lambda} r(A, B), P_1(A) - P_2(B)\right],\right.$$

$$\left.\sup_{B \in \mathscr{B}_1} \inf_{A \in \mathscr{B}_1} \max\left[\frac{1}{\lambda} r(A, B), P_2(B) - P_1(A)\right]\right\}. \tag{4.1.18}$$

Problem 4.1.1. Is it possible to represent the Prokhorov metric π_λ by means of r_λ or to find a p. metric with r_λ-structure that metrizes the weak convergence in $\mathscr{P}(U)$ for a s.m.s. U?

Remark 4.1.4. We can use the Hausdorff representation (4.1.18) of $\pi = \pi_1$ in order to extend the definition of Prokhorov metric over the set $\Phi(U)$ that strictly contains the set $\mathscr{P}(U)$ of all probability laws on an arbitrary metric space (U, d), namely, let $\Phi(U)$ be the family of all set functions $\phi: (\mathscr{S}, r) \to [0, 1]$ which are continuous from above, i.e., for any sequence $\{C_n\}_{n \geq 0}$ of closed subsets of U

$$r(C_n, C_0) \to 0 \Rightarrow \overline{\lim_{n\to\infty}}\, \phi(C_n) \leq \phi(C_0).$$

Clearly each law $P \in \Phi(U)$. We extend the Prokhorov metric over $\Phi(U)$ by simply setting

$$\pi(\phi_1, \phi_2) = \max\left\{\sup_{C_1 \in \mathscr{S}} \inf_{C_2 \in \mathscr{S}} \max[r(C_1, C_2), \phi_1(C_1) - \phi_2(C_2)],\right.$$

$$\left.\sup_{C_2 \in \mathscr{S}} \inf_{C_1 \in \mathscr{S}} \max[r(C_1, C_2), \phi_2(C_2) - \phi_1(C_1)]\right\}.$$

For $\phi_i = P_i \in \mathscr{P}(U)$ the above formula gives

$$\pi(P_1, P_2) = \inf\{\varepsilon > 0: P_1(C) \leq P_2(C^\varepsilon) + \varepsilon,\ P_2(C) \leq P_1(C^\varepsilon) + \varepsilon, \quad \forall C \in \mathscr{S}\}$$

i.e., the usual Prokhorov metric (see Theorem 4.2.1 for details).

The next step is to extend the notion of weak convergence. We shall use the analogue of the Hausdorff topological convergence of sequences of sets. For a sequence $\{\phi_n\} \subset \Phi(U)$, define the *upper topological limit* $\bar{\phi} = \overline{\ell t}\phi_n$ by

$$\bar{\phi}(C) := \sup\left\{\overline{\lim_{n\to\infty}}\, \phi_n(C_n) \colon C_n \in \mathscr{S}, r(C_n, C) \to 0\right\}.$$

Analogously, define the *lower topological limit* $\underline{\phi} = \underline{\ell t}\phi_n$ by

$$\underline{\phi}(C) := \sup\left\{\underline{\lim}_{n\to\infty}\, \phi_n(C_n) \colon C_n \in \mathscr{S}, r(C_n, C) \to 0\right\}.$$

If $\overline{\ell t}\phi_n = \underline{\ell t}\phi_n$, $\{\phi_n\}$ is said to be *topologically convergent* and $\phi := \ell t\phi_n := \overline{\ell t}\phi_n$ is said to be the *topological limit* of $\{\phi_n\}$. One can see that $\phi = \ell t\phi_n \in \Phi(U)$.

For any metric space (U, d) the following hold:

(a) Suppose P_n and P are laws on U. If $P = \ell tP_n$, then $P_n \xrightarrow{w} P$. Conversely, if (U, d) is a s.m.s. then the weak convergence $P_n \xrightarrow{w} P$ yields the topological convergence, $P = \ell tP_n$.

(b) If $\pi(\phi_n, \phi) \to 0$ for $\{\phi_n\} \subset \Phi(U)$, then $\phi = \ell t\phi_n$.

(c) If $\{\phi_n\}$ is fundamental (Cauchy) with respect to π, then ϕ_n is topologically convergent.

(d) If (U, d) is a compact set, then the π-convergence and the topological convergence coincide in $\Phi(U)$.

(e) If (U, d) is a complete metric space, then the metric space $(\phi(U), \pi)$ is also complete.

(f) If (U, d) is totally bounded, then $(\Phi(U), \pi)$ is also totally bounded.

(g) If (U, d) is a compact space, then $(\Phi(U), \pi)$ is also a compact metric space.

The extension $\Phi(U)$ of the set of laws $\mathscr{P}(U)$ seems to enjoy properties that are basic in the application of the notions of weak convergence and Prokhorov metric. Note also that in a s.m.s. (U, d) if $\{P_n\} \subset \mathscr{P}(U)$ is π-fundamental then clearly $\{P_n\}$ may not be weakly convergent, however, by (c) $\{P_n\}$ has a topological limit, $\phi = \ell tP_n \in \Phi(U)$.

Next, taking into account Definition 4.1.2, we shall define the Hausdorff structure of p. semidistances.

Without loss of generality (cf. Section 2.5), we assume that any p. semidistance $\mu(P)$, $P \in \mathscr{P}_2(U)$ has a representation in terms of pairs of U-valued random variables X, $Y \in \mathfrak{X} := \mathfrak{X}(U)$

$$\mu(P) = \mu(\Pr_{X,Y}) = \mu(X, Y).$$

Let $\mathscr{B}_0 \subseteq \mathscr{B}(U)$ and let the function $\phi\colon \mathfrak{X}^2 \times \mathscr{B}_0^2 \to [0, \infty]$ satisfy the relations

(a) if $\Pr(X = Y) = 1$ then $\phi(X, Y; A, B) = 0$ for all $A, B \in \mathscr{B}_0$;

(b) there exists a constant $K_\phi \geqslant 1$ such that for all A, B, $C \in \mathscr{B}_0$ and r.v. X, Y, Z

$$\phi(X, Z; A, B) \leqslant K_\phi[\phi(X, Y; A, C) + \phi(Y, Z; C, B)].$$

Definition 4.1.3. Let μ be p. semidistance. The representation of μ in the following form

$$\mu(X, Y) = h_{\lambda,\phi,\mathscr{B}_0}(X, Y) := \max\{h'_{\lambda,\phi,\mathscr{B}_0}(X, Y), h'_{\lambda,\phi,\mathscr{B}_0}(Y, X)\} \tag{4.1.19}$$

where

$$h'_{\lambda,\phi,\mathscr{B}_0}(X, Y) = \sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0} \max\left\{\frac{1}{\lambda} r(A, B), \phi(X, Y; A, B)\right\} \tag{4.1.20}$$

is called the *Hausdorff structure* of μ, or simply *h-structure*.

In (4.1.20), $r(A, B)$ is the Hausdorff semimetric in the Borel σ-algebra $\mathscr{B}((U, d))$ (see (4.1.1) with $\rho \equiv d$), λ is a positive number, $B_0 \subseteq \mathscr{B}(U)$ and ϕ satisfies conditions (a) and (b).

Using the properties (a) and (b) we easily obtain the following lemma.

Lemma 4.1.2. Each μ in the form (4.1.19) is a p. semidistance in $\mathfrak{X}$ with a parameter $\mathbb{K}_\mu = K_\phi$.

In the limit cases $\lambda \to 0$, $\lambda \to \infty$ the Hausdorff structure turns into a 'uniform' structure. More precisely, the following limit relations hold.

Lemma 4.1.3. Let μ have Hausdorff structure (4.1.19), then, as $\lambda \to 0$, $\mu(X, Y) = h_{\lambda,\phi,\mathscr{B}_0}(X, Y)$ has a limit which is defined to be

$$h_{0,\phi,\mathscr{B}_0}(X, Y) = \max\left\{\sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0} \phi(X, Y; A, B), \sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0} \phi(Y, X; A, B)\right\}.$$

As $\lambda \to \infty$ the limit

$$\lim_{\lambda \to \infty} \lambda h_{\lambda,\phi,\mathscr{B}_0}(X, Y) = h_{\infty,\phi,\mathscr{B}_0}(X, Y) \tag{4.1.21}$$

exists and is defined to be

$$\max\left\{\sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0, \phi(X, Y; A, B) = 0} r(A, B), \sup_{A \in \mathscr{B}_0} \inf_{A \in \mathscr{B}_0, \phi(Y, X; A, B) = 0} r(A, B)\right\}.$$

Remark 4.1.5. Since $\lim_{\lambda \to \infty} h_{\lambda,\phi,\mathscr{B}_0}(X, Y) = 0$ we normalized the quantity $h_{\lambda,\phi,\mathscr{B}_0}(X, Y)$, multiplying it by λ so that $\lambda \to \infty$ yields a non-trivial limit $h_{\infty,\phi,\mathscr{B}_0}(X, Y)$.

Proof. We shall prove Equality (4.1.21) only. Namely, for each X, $Y \in \mathfrak{X}$

$$
\begin{aligned}
&\lim_{\lambda \to \infty} \lambda h'_{\lambda, \phi, \mathscr{B}_0}(X, Y) \\
&\quad = \lim_{\lambda \to 0} \sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0} \max\left\{ r(A, B), \frac{1}{\lambda} \phi(X, Y; A, B) \right\} \\
&\quad = \lim_{\lambda \to 0} \inf\left\{ \varepsilon > 0: \inf_{B \in \mathscr{B}_0, r(A, B) < \varepsilon} \frac{1}{\lambda} \phi(X, Y; A, B) < \varepsilon \quad \text{for all } A \in \mathscr{B}_0 \right\} \\
&\quad = \inf\left\{ \varepsilon > 0: \inf_{B \in B_0, r(A, B) < \varepsilon} \phi(X, Y; A, B) = 0 \quad \text{for all } A \in \mathscr{B}_0 \right\} \\
&\quad = \sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0, \phi(X, Y; A, B) = 0} r(A, B).
\end{aligned}
$$

Now, by Equality (4.1.19), we claim Equality (4.1.21). QED

Let us consider some examples of p. semidistances with Hausdorff structure.

Example 4.1.1 (Universal Hausdorff representation). Each p. semidistance μ has the trivial form $h_{\lambda, \phi, \mathscr{B}_0} = \mu$ where the set $\mathscr{B}_0$ is a singleton, say $\mathscr{B}_0 \equiv \{A_0\}$, and $\phi(X, Y; A_0, A_0) = \mu(X, Y)$.

Example 4.1.2 (Hausdorff structure of the Prokhorov metric π_λ). The Prokhorov metric (3.2.22) admits a Hausdorff structure representation $h_{\lambda, \phi, \mathscr{B}_0} = \mu$ (see Representations (4.1.18) and (4.1.19)) where $\mathscr{B}_0$ is either the class $\mathscr{C}$ of all non-empty closed subsets of U or $\mathscr{B}_0 \equiv \mathscr{B}(U)$ and $\phi(X, Y; A, B) = \Pr(X \in A) - \Pr(Y \in B)$, A, $B \in \mathscr{B}(U)$. As $\lambda \to 0$ and $\lambda \to \infty$ (cf. Lemma 3.2.1) we obtain the limits

$$h_{0, \phi, \mathscr{B}_0} = \boldsymbol{\sigma} \quad \text{(distance in variation)}$$

and

$$h_{\infty, \phi, \mathscr{B}_0} = \ell_\infty .$$

Example 4.1.3 (Lévy metric $\mathbf{L}_\lambda (\lambda > 0)$ in the space $\mathscr{P}_1(\mathbb{R}^n)$). Let $\mathscr{F}(\mathbb{R}^n)$ be the space of all right continuous d.f.s F on $\mathbb{R}^n$. We extend the definition of the Lévy metric ($\mathbf{L}_\lambda$, $\lambda > 0$) in $\mathscr{F}(\mathbb{R}^1)$ (see Definition (4.1.3)) considering the multivariate case $\mathbf{L}_\lambda$ in $\mathscr{F}(\mathbb{R}^n)$

$$\mathbf{L}_\lambda(P_1, P_2) := \mathbf{L}_\lambda(F_1, F_2) := \inf\{\varepsilon > 0: F_1(x - \lambda\varepsilon\mathbf{e}) - \varepsilon \leqslant F_2(x) \leqslant F_1(x + \lambda\varepsilon\mathbf{e}) + \varepsilon \quad \forall x \in \mathbb{R}^n\} \tag{4.1.22}$$

where F_i is the d.f. of P_i ($i = 1, 2$) and $\mathbf{e} = 1, 1, \ldots, 1)$ is the unit vector in $\mathbb{R}^n$.

The Hausdorff representation of $\mathbf{L}_\lambda$ is handled by Representation (4.1.19) where $\mathscr{B}_0$ is the set of all multivariate intervals $(-\infty, x](x \in \mathbb{R}^n)$ and

$$\phi(X, Y; (-\infty, x], (-\infty, y]) := F_1(x) - F_2(y)$$

i.e., for r.v.s X and Y with d.f.s F_1 and F_2, respectively,

$$\mathbf{L}_\lambda(X, Y) = \mathbf{L}_\lambda(F_1, F_2) := \max\left\{\sup_{x \in \mathbb{R}^n} \inf_{y \in \mathbb{R}^n} \max\left[\frac{1}{\lambda}\|x - y\|_\infty, F_1(x) - F_2(y)\right],\right.$$

$$\left.\sup_{y \in \mathbb{R}^n} \inf_{x \in \mathbb{R}^n} \max\left[\frac{1}{\lambda}\|x - y\|_\infty, F_2(y) - F_1(x)\right]\right\} \tag{4.1.23}$$

for all $F_1, F_2 \in \mathscr{F}(\mathbb{R}^n)$ where $\|\cdot\|$ stands for the Minkovski norm in $\mathbb{R}^n$, $\|(x_1, \ldots, x_n)\|_\infty := \max_{1 \leqslant i \leqslant n} |x_i|$. Letting $\lambda \to 0$ in Definition (4.1.23) we get the *Kolmogorov distance* in $\mathscr{F}(\mathbb{R}^n)$:

$$\lim_{\lambda \to 0} \mathbf{L}_\lambda(F_1, F_2) = \boldsymbol{\rho}(F_1, F_2) := \sup_{x \in \mathbb{R}^n} |F_1(x) - F_2(x)| \tag{4.1.24}$$

The limit of $\lambda \mathbf{L}_\lambda$ as $\lambda \to \infty$ is given by (4.1.21), i.e.

$$\begin{aligned}
&\lim_{\lambda \to \infty} \lambda \mathbf{L}_\lambda(F_1, F_2) \\
&\quad = \inf\{\varepsilon > 0: \inf[F_1(x) - F_2(y): y \in \mathbb{R}^n, \|x - y\|_\infty \leqslant \varepsilon] = 0, \\
&\qquad\qquad \inf[F_2(x) - F_1(y): x \in \mathbb{R}^n, \|x - y\|_\infty \leqslant \varepsilon] = 0 \quad \forall x \in \mathbb{R}^n\} \\
&\quad = \mathbf{W}(F_1, F_2) := \inf\{\varepsilon > 0: F_1(x) \leqslant F_2(x + \varepsilon \mathbf{e}), F_2(x) \leqslant F_1(x + \varepsilon \mathbf{e}) \\
&\qquad\qquad \forall x \in \mathbb{R}^n\}.
\end{aligned} \tag{4.1.25}$$

Problem 4.1.2. If $n = 1$ then

$$\lim_{\lambda \to \infty} \lambda \mathbf{L}_\lambda(P_1, P_2) = \ell_\infty(P_1, P_2), P_1, P_2 \in \mathscr{P}(\mathbb{R}^n) \tag{4.1.26}$$

where $\ell_\infty(P_1, P_2) := \inf\{\varepsilon > 0: P_1(A) \leqslant P_2(A^\varepsilon)$ for all borel subsets of $\mathbb{R}^n\}$ (see (4.1.5) and further Corollary 7.3.2 and (7.4.15)). Is it true that the equality (4.1.26) is valid for any integer n?

Example 4.1.4 (Lévy p. distance $\mathbf{L}_{\lambda,H}$, $\lambda > 0$, $H \in \mathscr{H}$). The Lévy metric $\mathbf{L}_\lambda$ (4.1.22) can be rewritten in the form

$$\mathbf{L}_\lambda(F_1, F_2) := \inf\{\varepsilon > 0: (F_1(x) - F_2(x - \lambda\varepsilon\mathbf{e}))_+ < \varepsilon, (F_2(x) - F_1(x - \lambda\varepsilon\mathbf{e}))_+ < \varepsilon$$
$$\forall x \in \mathbb{R}^n\} \quad (\cdot)_+ := \max(\cdot, 0)$$

which can be viewed as a special case $(H(t) = t)$ of the *Lévy p. distance*

$\mathbf{L}_{\lambda,H}(\lambda > 0, H \in \mathscr{H})$ defined as follows

$$\mathbf{L}_{\lambda,H}(F_1, F_2) := \inf\{\varepsilon > 0\colon \tilde{H}(F_1(x) - F_2(x + \lambda\varepsilon\mathbf{e})) < \varepsilon,\ \tilde{H}(F_2(x) - F_1(x + \lambda\varepsilon\mathbf{e})) < \varepsilon \quad \forall x \in \mathbb{R}^n\} \tag{4.1.27}$$

where

$$\tilde{H}(t) := \begin{cases} H(t) & t \geqslant 0 \\ 0 & t \leqslant 0. \end{cases}$$

$\mathbf{L}_{\lambda,H}$ admits a Hausdorff representation of the following type

$$\mathbf{L}_{\lambda,H}(F_1, F_2) = \max\left\{\sup_{x\in\mathbb{R}^n}\inf_{y\in\mathbb{R}^n}\max\left[\frac{1}{\lambda}\|x - y\|, \tilde{H}(F_1(x) - F_2(y))\right],\ \sup_{y\in\mathbb{R}^n}\inf_{x\in\mathbb{R}^n}\max\left[\frac{1}{\lambda}\|x - y\|, \tilde{H}(F_2(y) - F_1(x))\right]\right\}. \tag{4.1.28}$$

The last representation of $\mathbf{L}_{\lambda,H}$ shows that $\mathbf{L}_{\lambda,H}$ is a simple distance with parameter $\mathbb{K}_{\mathbf{L}_{\lambda,H}} := K_H$ (see (2.2.3)). Also, from (4.1.28) as $\lambda \to 0$, we get the *Kolmogorov p. distance*

$$\lim_{\lambda\to 0}\mathbf{L}_{\lambda,H}(F_1, F_2) = H(\rho(F_1, F_2)) = \boldsymbol{\rho}_H(F_1, F_2) := \sup_{x\in\mathbb{R}^n} H(|F_1(x) - F_2(x)|). \tag{4.1.29}$$

Analogously, letting $\lambda \to \infty$ in (4.1.28), we have

$$\lim_{\lambda\to\infty}\lambda\mathbf{L}_{\lambda,H}(F_1, F_2) = \mathbf{W}(F_1, F_2). \tag{4.1.30}$$

We prove Equality (4.1.30) by arguments provided in the limit relation (4.1.25).

Example 4.1.5 (Hausdorff metric on $\mathscr{F}(\mathbb{R})$ and $\mathscr{P}(U)$). The Lévy metric in $\mathscr{F} := \mathscr{F}(\mathbb{R})$ (4.1.22) has a Hausdorff structure (see (4.1.23)), however, the function

$$\tilde{D}((x, F_1(x)), (y, F_2(y))) := \max\left\{\frac{1}{\lambda}|x - y|, F_1(x) - F_2(y)\right\}$$

is not a metric in the space $\mathbb{R} \times [0, 1]$ and hence (4.1.23) is not a 'pure' Hausdorff metric (see (4.1.2)). In the next definition we shall change the semimetric $\tilde{D}$ with the Minkovski metric dm_λ in $\mathbb{R} \times [0, 1]$

$$dm_\lambda((x, F_1(x)), (y, F_2(y))) := \max\left\{\frac{1}{\lambda}|x - y|, |F_1(x) - F_2(y)|\right\}. \tag{4.1.31}$$

By means of Equality (4.1.31) we define the Hausdorff metric in $\mathscr{F}(\mathbb{R}^n)$ as follows.

Definition 4.1.4. The metric

$$\mathbf{H}_\lambda(F, G) := \max\left\{\sup_{x\in\mathbb{R}^n}\inf_{y\in\mathbb{R}^n} dm_\lambda((x, F(x)), (y, G(y))), \right.$$

$$\left. \sup_{y\in\mathbb{R}^n}\inf_{x\in\mathbb{R}^n} dm_\lambda((x, F(x)), (y, G(y)))\right\} \quad F, G \in \mathscr{F}^n \tag{4.1.32}$$

is said to be the *Hausdorff metric with parameter* λ (or simply, H_λ-metric) in distribution functions space $\mathscr{F}$.

Lemma 4.1.4. (a) For any $\lambda > 0$, H_λ is a metric in $\mathscr{F}$.

(b) H_λ is a non-increasing function of λ and the following holds:

$$\lim_{\lambda\to 0}\mathbf{H}_\lambda(F, G) = \boldsymbol{\rho}(F, G) \tag{4.1.33}$$

and

$$\lim_{\lambda\to\infty}\lambda\mathbf{H}_\lambda(F, G) = \tilde{\mathbf{W}}(F, G)$$

$$:= \inf\{\varepsilon > 0\colon (F_1(x) - F_2(x+\varepsilon))_+ = 0, (F_2(x-\varepsilon) - F_1(x))_+ = 0 \quad \forall x \in \mathbb{R}\}. \tag{4.1.34}$$

(c) If F and G are continuous d.f.s then $\mathbf{H}_\lambda(F, G) = \mathbf{L}_\lambda(F, G)$.

Proof. (a) By means of the Minkowski metric

$$dm_\lambda((x_1, y_1), (x_2, y_2)) := \max\left\{\frac{1}{\lambda}|x_1 - x_2|, |y_1 - y_2|\right\}$$

in the space $D := \mathbb{R} \times [0, 1]$, define the Hausdorff (semi)metric in the space 2^D of all subsets $B \subseteq D$

$$h_\lambda(B_1, B_2) := \max\left\{\sup_{b_1\in B_1}\inf_{b_2\in B_2} dm_\lambda(b_1, b_2), \sup_{b_2\in B_2}\inf_{b_1\in B_1} dm_\lambda(b_1, b_2)\right\}.$$

In the Hausdorff representation (4.1.11) of the Lévy metric the main role was played by the notion of the completed graph $\bar{F}$ of a d.f. F. Here, we need the notion of the closed graph Γ_F of a d.f. F defined as follows

$$\Gamma_F := \left(\bigcup_{x\in\mathbb{R}}(x, F(x))\right) \cup \left(\bigcup_{x\in\mathbb{R}}(x, F(x-0))\right) \tag{4.1.35}$$

i.e., the closed graph Γ_F is handled by adding the points $(x, F(x-))$ to the graph of F, where x denotes points of F-discontinuity, cf. Fig. 4.1.1 and Fig. 4.1.2.

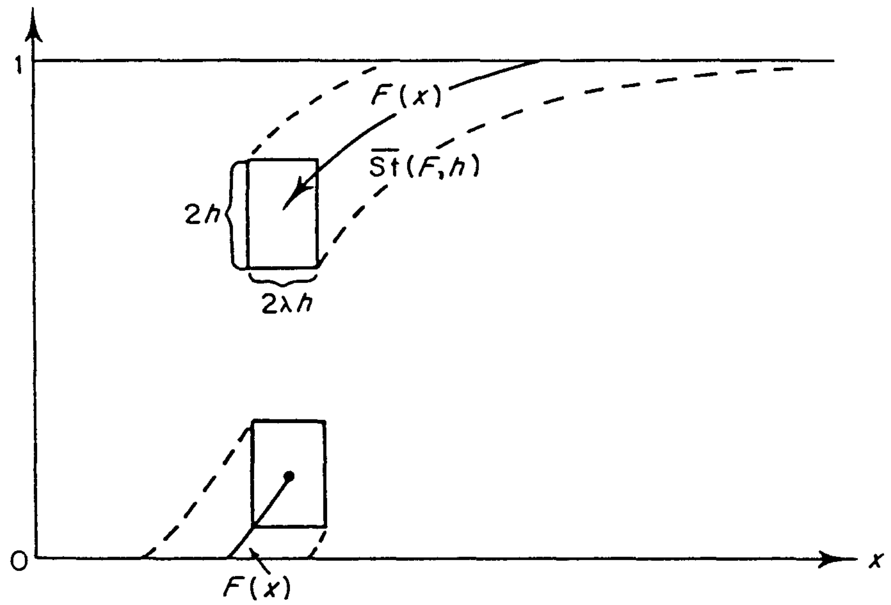


Figure 4.1.2 $\overline{\mathrm{St}}(F, h)$ is the strip into which the graph of the d.f. G has to be located in order that $H_\lambda(F, G) \leqslant h$ for $F, G \in \mathscr{F}'$.

Obviously, $H_\lambda(F, G) = h_\lambda(\Gamma_F, \Gamma_G)$. Moreover, if the closed graphs of F and G coincide, then $F(x) = G(x)$ for all continuity points x of F and G. Since F and G are right-continuous, then $\Gamma_F \equiv \Gamma_G \Leftrightarrow F \equiv G$.

(b) The limit relation (4.1.33) is a consequence of (4.1.24) and

$$\mathbf{L}_\lambda(F_1, F_2) \leqslant \mathbf{H}_\lambda(F_1, F_2) \leqslant \boldsymbol{\rho}(F_1, F_2) \quad F_1, F_2 \in \mathscr{F}. \tag{4.1.36}$$

Analogously to (4.1.25) we claim that

$$\begin{aligned}\lim_{\lambda\to 0} \lambda \mathbf{H}_\lambda(F, G) = \inf\{\varepsilon > 0\colon &\inf\{|F_1(x) - F_2(y)|\colon y \in \mathbb{R}, |x - y| \leqslant \varepsilon\} = 0,\\ &\inf\{|F_2(x) - F_1(y)|\colon y \in \mathbb{R}, |x - y| \leqslant \varepsilon\} = 0 \quad \forall x \in \mathbb{R}\}\\ = \tilde{\mathbf{W}}(F, G).&\end{aligned}$$

(c) See Fig. 4.1.1 and Fig. 4.1.2.

Remark 4.1.6. Further we need the following notations: for two metrics ρ_1 and ρ_2 on a set S, $\rho_1 \overset{\text{top}}{\leqslant} \rho_2$ means that ρ_2-convergence implies ρ_1-convergence, and $\rho_1 \overset{\text{top}}{<} \rho_2$ means $\rho_1 \overset{\text{top}}{\leqslant} \rho_2$ but not $\rho_2 \overset{\text{top}}{\leqslant} \rho_1$. Finally $\rho_1 \overset{\text{top}}{\sim} \rho_2$ means that $\rho_1 \overset{\text{top}}{\leqslant} \rho_2$ and $\rho_2 \overset{\text{top}}{\leqslant} \rho_1$. By (4.1.36) it follows that

$$\mathbf{L}_\lambda \overset{\text{top}}{\leqslant} \mathbf{H}_\lambda \overset{\text{top}}{\leqslant} \boldsymbol{\rho}. \tag{4.1.37}$$

Moreover, the following simple examples show that

$$\mathbf{L}_\lambda \overset{\text{top}}{<} \mathbf{H}_\lambda \overset{\text{top}}{<} \boldsymbol{\rho}.$$

Example 1. Let

$$F_n(x) = \begin{cases} 0 & x < \dfrac{1}{n} \\ 1 & x \geqslant \dfrac{1}{n} \end{cases}, \qquad F_0(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geqslant 1. \end{cases}$$

Then $\mathbf{\rho}(F_n, F) = 1$, $\mathbf{H}_\lambda(F_n, F) = 1/\lambda n \to 0$ as $n \to \infty$.

Example 2. Let

$$\phi_n(x) = \begin{cases} 0 & x < 0 \\ \dfrac{1}{2} & 0 \leqslant x < \dfrac{1}{n} \\ 1 & x \geqslant \dfrac{1}{n}. \end{cases}$$

Then

$$\mathbf{L}_\lambda(\phi_n, F_0) = \min\left(1, \frac{1}{\lambda}\right) n^{-1} \to 0 \qquad \text{as } n \to \infty$$

but

$$H_\lambda(\phi_n, F_0) \geqslant \inf_{y \in \mathbb{R}} \max\left\{\frac{1}{\lambda}\left|\frac{1}{2n} - y\right|, \left|\phi_n\left(\frac{1}{2n}\right) - F_0(y)\right|\right\} \geqslant \frac{1}{2}$$

for any $n = 1, 2, \ldots$.

Remark 4.1.7. For any $0 < \lambda < \infty$, $\mathbf{H}_\lambda$ metrizes one and the same topology. We characterize the $\mathbf{H}$-topology ($\mathbf{H} := \mathbf{H}_1$) by the following compactness criterion. Recall that a subset $\mathscr{A}$ of a metric space (S, ρ) is said to be *ρ-relatively compact* if any sequence in $\mathscr{A}$ has a ρ-convergent subsequence. Define the Skorokhod–Billingsley metric in the space $\mathscr{F}$ of distribution functions on $\mathbb{R}$

$$\mathbf{SB}(F, G) = \inf_{\lambda \in \Lambda} \max\left\{\sup_{s \neq t}\left|\log\frac{\lambda(s) - \lambda(t)}{s - t}\right|, \sup_{t \in \mathbb{R}}|\lambda(t) - t|, \sup_{t \in \mathbb{R}}|F(t) - G(\lambda(t))|\right\}$$

where Λ is the class of all strictly increasing continuous functions λ from $\mathbb{R}$ onto $\mathbb{R}$. The metrics $\mathbf{H}$ and $\mathbf{SB}$ generate one and the same topology in $\mathscr{F}$, the metric space $(\mathscr{F}, \mathbf{H})$ is not complete while $(\mathscr{F}, \mathbf{SB})$ is complete. To show that $\mathbf{H}$ is not a complete metric, observe that ϕ_n introduced in Example 2 is $\mathbf{H}$-fundamental but not $\mathbf{H}$-convergent. The proof that $(\mathscr{F}, \mathbf{SB})$ is complete is the same as the proof that $D[0, 1]$ is complete with the Skorokhod–Billingsley metric d_0 (see Billingsley 1968, Theorem 14.2). The equivalence of **H**- and **SB**-topology is a consequence of the compactness criterion given below. Consider the following moduli of **H**-continuity (cf. Billingsley 1968, pp. 110, 118)

$$\text{(i)} \qquad \omega'_F(\delta) := \inf_{\{t_0,\ldots,t_r\}} \max_{0\leqslant i\leqslant r} [F(t_i-) - F(t_{i-1})] \quad F \in \mathscr{F},\ \delta \in (0, 1)$$

where the infimum is taken over all $\{t_0, t_1, \ldots, t_r\}$ satisfying the conditions: $-\infty = t_0 < t_1 < \cdots < t_r = \infty$, $t_i - t_{i-1} > \delta$, $i = 1, \ldots, r$;

$$\text{(i)} \ \omega''_F := \sup_{x\in\mathbb{R}} \min\{F(x + \delta/2) - F(x), F(x) - F(x - \delta/2)\} \quad F \in \mathscr{F},\ \delta \in (0, 1).$$

For any $f \in \mathscr{F}$, $\lim_{\delta\to\infty} \omega'_F(\delta) = 0$ and $\omega''_F(\delta) \leqslant \omega'_F(2\delta)$ (cf. Billingsley 1968, §14). Let $\mathscr{A} \subset \mathscr{F}$. Then the following are equivalent (Rachev 1984a; Kakosyan *et al.*, 1988, Section 2.5):

(a) $\mathscr{A}$ is **H**-relatively compact,
(b) $\mathscr{A}$ is **SB**-relative compact,
(c) $\lim\limits_{\delta\to\infty} \sup\limits_{F\in\mathscr{A}} \omega'_F(\delta) = 0$,
(d) $\mathscr{A}$ is weakly compact (i.e., **L**-relative compact) and $\lim\limits_{\delta\to\infty} \sup\limits_{F\in\mathscr{A}} \omega''_F(\delta) = 0$.

Moreover, for F, $G \in \mathscr{F}$ and $\delta > 0$ the following hold

$$\mathbf{H}(F, G) \leqslant \mathbf{SB}(F, G)$$

$$\omega'_G(\delta) \leqslant \omega'_F(\delta + 2\mathbf{H}(F, G)) + 4\mathbf{H}(F, G)$$

$$\mathbf{H}(F, G) \leqslant \max\{\omega''_F(4\mathbf{L}(F, G)), \omega''_G(4\mathbf{L}(G, G)\}\mathbf{L}(F, G).$$

Next, let (U, d) be a metric space and define the following analog of **H**-metrics

$$\pi\mathbf{H}_\lambda(P_1, P_2) := \max\left\{\sup_{A\in\mathscr{B}_1} \inf_{B\in\mathscr{B}_1} \max\left[\frac{1}{\lambda} r(A, B), |P_1(A) - P_2(B)|\right]\right.$$
$$\left.\sup_{B\in\mathscr{B}_1} \inf_{A\in\mathscr{B}_1} \max\left[\frac{1}{\lambda} r(A, B), |P_1(A) - P_2(B)|\right]\right\} \tag{4.1.38}$$

for any laws $P_1, P_2 \in \mathscr{P}(U)$.

Lemma 4.1.5. (a) For any $\lambda > 0$, the functional $\pi\mathbf{H}_\lambda$ on $\mathscr{P}_1 \times \mathscr{P}_1$ is a metric in $\mathscr{P}_1 = \mathscr{P}(U)$.

(b) $\pi\mathbf{H}_\lambda$ is a non-increasing function of λ and the following holds

$$\lim_{\lambda\to 0} \pi\mathbf{H}_\lambda(P_1, P_2) = \boldsymbol{\sigma}(P_1, P_2) := \sup_{A\in\mathscr{B}_1} |P_1(A) - P_2(A)| \quad P_1, P_2 \in \mathscr{P}_1 \tag{4.1.39}$$

$$\lim_{\lambda\to\infty} \lambda\pi\mathbf{H}_\lambda(P_1, P_2) = \pi\mathbf{H}_\infty(P_1, P_2)$$

$$:= \inf\{\varepsilon > 0: \inf[|P_1(A) - P_2(B)|: B \in \mathscr{B}_1, r(A, B) < \varepsilon] = 0,$$

$$\inf[|P_2(A) - P_1(B)|: B \in \mathscr{B}_1, r(A, B) < \varepsilon] = 0 \quad \forall A \in \mathscr{B}_1\}. \tag{4.1.40}$$

(c) $\pi\mathbf{H}_\lambda$ is 'between' the Prokhorov metric π_λ (4.1.18) and the total variation metric σ, i.e.

$$\pi_\lambda \leqslant \pi\mathbf{H}_\lambda \leqslant \sigma \tag{4.1.41}$$

and

$$\pi_\lambda \overset{\text{top}}{<} \pi\mathbf{H}_\lambda \overset{\text{top}}{<} \sigma. \tag{4.1.42}$$

Proof. Let us prove only (4.1.40). We have

$$\pi\mathbf{H}_\lambda(P_1, P_2) = \inf\{\varepsilon > 0: \inf[|P_1(A) - P_2(B)|: B \in \mathscr{B}_1, r(A, B) < \lambda\varepsilon] < \varepsilon,$$

$$\inf[|P_2(A) - P_1(B)|: B \in \mathscr{B}_1, r(A, B) < \lambda\varepsilon] < \varepsilon \quad \forall A \in \mathscr{B}_1\}. \tag{4.1.43}$$

Further multiplying the two sides of (4.1.43) by λ and letting $\lambda \to \infty$ we get (4.1.40). QED

4.2. Λ-STRUCTURE OF p. SEMIDISTANCES

The p. semidistance structure Λ in $\mathscr{X} = \mathscr{X}(U)$ is defined by means of a non-negative function v on $\mathscr{X} \times \mathscr{X} \times [0, \infty)$ that satisfies the relationships: for all $X, Y, Z \in \mathfrak{X}$,

(a) If $\Pr(X = Y) = 1$ then $v(X, Y; t) = 0 \quad \forall t \geqslant 0$
(b) $v(X, Y; t) = v(Y, X; t)$
(c) If $t' < t''$ then $v(X, Y; t') \geqslant v(X, Y; t'')$
(d) For some $K_v \geqslant 1$, $v(X, Z; t' + t'') \leqslant K_v[v(X, Y; t') + v(Y, Z; t'')]$.

If $v(X, Y; t)$ is completely determined by the marginals $P_1 = \Pr_X$, $P_2 = \Pr_Y$, we shall use the notation $v(P_1, P_2; t)$ instead of $v(X, Y; t)$. For the case $K_v = 1$, the following definition is due to Zolotarev (1976b).

Definition 4.2.1. The p. semidistance μ has a Λ-structure if it admits a Λ-representation, i.e.,

$$\mu(X, Y) = \Lambda_{\lambda,v}(X, Y) := \inf\{\varepsilon > 0: v(X, Y; \lambda\varepsilon) < \varepsilon\} \tag{4.2.1}$$

for some $\lambda > 0$ and v satisfying (a) to (d).

Obviously, if μ has a Λ-representation (4.2.1), then μ is a p. semidistance with $\mathbb{K}_\mu = K_v$. In Example 4.1.1 it was shown that each p. semidistance has a Hausdorff representation $h_{\lambda,\phi,\mathscr{B}_0}$. In the next theorem we shall prove that each p. semidistance μ with Hausdorff structure (see Definition 4.1.3) also has a

Λ-representation. Hence, in particular, each p. semidistance has a Λ-structure as well as a Hausdorff structure.

Theorem 4.2.1. Suppose a p. semidistance μ admits the Hausdorff representation $\mu = h_{\lambda,\phi,\mathscr{B}_0}$ (4.1.19). Then μ enjoys also a Λ-representation

$$h_{\lambda,\phi,\mathscr{B}_0}(X, Y) = \Lambda_{\lambda,\nu}(X, Y) \tag{4.2.2}$$

where

$$\nu(X, Y; t) := \max\left\{\sup_{A \in \mathscr{B}_0} \inf_{B \in A(t)} \phi(X, Y; A, B), \sup_{A \in B_0} \inf_{B \in A(t)} \phi(Y, X; A, B)\right\}$$

and $A(t)$ is the collection of all elements B of $\mathscr{B}_0$ such that the Hausdorff semimetric $r(A, B)$ is not greater than t.

Proof. Let $\Lambda_{\lambda,\nu}(X, Y) < \varepsilon$. Then for each $A \in \mathscr{B}_0$, there exists a set $B \in A(\lambda\varepsilon)$ such that $\phi(X, Y; A, B) < \varepsilon$, i.e.

$$\sup_{A \in \mathscr{B}_0} \inf_{B \in \mathscr{B}_0} \max\left\{\frac{1}{\lambda} r(A, B), \phi(X, Y; A, B)\right\} < \varepsilon.$$

By symmetry, it follows that $h_{\lambda,\phi,\mathscr{B}_0}(X, Y) < \varepsilon$. If conversely, $h_{\lambda,\phi,\mathscr{B}_0}(X, Y) < \varepsilon$, then for each $A \in \mathscr{B}_0$ there exists $B \in \mathscr{B}_0$ such that $r(A, B) < \lambda\varepsilon$ and $\phi(X, Y; A, B) < \varepsilon$. Thus

$$\sup_{A \in \mathscr{B}_0} \inf_{B \in A(\lambda\varepsilon)} \phi(X, Y; A, B) < \varepsilon. \qquad \text{QED}$$

Example 4.2.1 (Λ-structure of the Lévy metric and the Lévy distance). Recall the definition of the Lévy metric in $\mathscr{P}(\mathbb{R}^n)$ (see (4.1.22))

$$\mathbf{L}_\lambda(P_1, P_2) := \inf\left\{\varepsilon > 0: \sup_{x \in \mathbb{R}^n}(F_1(x) - F_2(x + \lambda\varepsilon\mathbf{e})) \leqslant \varepsilon \right.$$
$$\left. \text{and } \sup_{x \in \mathbb{R}^n}(F_2(x) - F_1(x + \lambda\varepsilon\mathbf{e})) \leqslant \varepsilon\right\}$$

where obviously F_i is the d.f. of P_i. By Definition 4.2.1, $\mathbf{L}_\lambda$ has a Λ-representation

$$\mathbf{L}_\lambda(P_1, P_2) = \Lambda_{\lambda,\nu}(P_1, P_2) \quad \lambda > 0$$

where

$$\nu(P_1, P_2; t) := \sup_{x \in \mathbb{R}^n} \max\{(F_1(x) - F_2(x + \lambda t\mathbf{e})), (F_2(x) - F_1(x + \lambda t\mathbf{e}))\}$$

and F_i is the d.f. of P_i. With an appeal to Theorem 4.2.1, for any $F_1, F_2 \in \mathscr{F}(\mathbb{R}^n)$, we have that the metric h defined below admits a Λ-representation:

$$h(F_1, F_2) := \max\left\{\sup_{x\in\mathbb{R}^n}\inf_{y\in\mathbb{R}^n}\max\left\{\frac{1}{\lambda}\|x-y\|_\infty, F_1(x)-F_2(y)\right\},\right.$$
$$\left.\sup_{x\in\mathbb{R}^n}\inf_{y\in\mathbb{R}^n}\max\left\{\frac{1}{\lambda}\|x-y\|_\infty, F_2(x)-F_1(y)\right\}\right\}$$
$$= \Lambda_{\lambda,\nu}(P_1, P_2)$$

where

$$\nu(P_1, P_2; t) = \max\left\{\sup_{x\in\mathbb{R}^n}\inf_{y:\,\|x-y\|_\infty\leqslant t}(F_1(x)-F_2(y)),\right.$$
$$\left.\sup_{x\in\mathbb{R}^n}\inf_{y:\,\|x-y\|_\infty\leqslant t}(F_2(x)-F_1(y))\right\}.$$

By virtue of the Λ-representation of the $\mathbf{L}_\lambda$ we conclude that $h(F_1, F_2) = \mathbf{L}_\lambda(F_1, F_2)$ which proves (4.1.23) and Theorem 4.1.2.

Analogously, consider the Lévy distance $\mathbf{L}_{\lambda,H}$ (4.1.27) and apply Theorem 4.2.1 with

$$\nu(X, Y; \lambda t) = \nu(P_1, P_2; \lambda t)$$
$$:= H\left(\sup_{x\in\mathbb{R}^n}\max\{F_1(x)-F_2(x+\lambda t\mathbf{e}\}, \quad \{F_2(x)-F_1(x+\lambda t\mathbf{e}\}\right)$$

to prove the Hausdorff representation of $\mathbf{L}_{\lambda,H}$ (4.1.28).

Example 4.2.2 (Λ-structure of the Prokhorov metric π_λ). Let

$$\nu(P_1, P_2; \varepsilon) := \sup_{A\in\mathscr{B}(U)}\max\{P_1(A)-P_2(A^\varepsilon), P_2(A)-P_1(A^\varepsilon)\}$$
$$= \sup_{A\in\mathscr{B}(U)}\{P_1(A)-P_2(A^\varepsilon)\}$$

(see Dudley 1976; Theorem 8.1). Then $\Lambda_{\lambda,\nu}(P_1, P_2)$ is the Λ-representation of the Prokhorov metric $\pi_\lambda(P_1, P_2)$ (cf. (3.2.22)). In this way, Theorem 4.1.3 and the equality (4.1.18) are corollaries of Theorem 4.2.1.

For each $\lambda > 0$ the Prokhorov metric π_λ induces the weak convergence in $\mathscr{P}_1$, thus,

$$\pi_\lambda(P_n, P) \to 0 \Leftrightarrow P_n \xrightarrow{w} P.$$

Remark 4.2.1. As is well known, the weak convergence $P_n \xrightarrow{w} P$ means that

$$\int_U f\,dP_n \to \int_U f\,dP \tag{4.2.3}$$

for each continuous and bounded function f on (U, d). The Prokhorov metric π (3.2.20) metrizes the weak convergence in $\mathscr{P}(U)$ where U is a s.m.s. (see Prokhorov, 1956; Dudley, 1989; Theorem 11.3.3). The next definition was essentially used by Dudley (1966a), Ranga Rao (1962), Bhattacharya and Ranga Rao (1976).

Definition 4.2.2. Let G be a non-negative continuous function on U and $\mathscr{P}_G$ be the set of laws P such that $\int_U G\,\mathrm{d}P < \infty$. The joint convergence

$$P_n \xrightarrow{w} P \quad \int_U G\,\mathrm{d}P_n \to \int_U G\,\mathrm{d}P \qquad (P_n, P \in \mathscr{P}_G) \tag{4.2.4}$$

will be called *G-weak convergence in* $\mathscr{P}_G$.

As in Prokhorov (1956) one can show that the *G-weighted Prokhorov metric*

$$\pi_{\lambda, G}(P_1, P_2) := \inf\{\varepsilon > 0: \lambda_1(A) \leqslant \lambda_2(A^{\lambda\varepsilon}) + \varepsilon,\ \lambda_2(A) \leqslant \lambda_1(A^{\lambda\varepsilon}) + \varepsilon \quad \forall A \in \mathscr{B}(U)\} \tag{4.2.5}$$

where $\lambda_i(A) := \int_A (1 + G(x))P_i(\mathrm{d}x)$, metrizes the G-weak convergence in $\mathscr{P}_G$ where U is a s.m.s., see further Theorem 10.1.2 for details.

The metric $\pi_{\lambda, G}$ admits Λ-representation with

$$\nu(P_1, P_2; \varepsilon) := \sup_{A \in \mathscr{B}(U)} \max\{\lambda_1(A) - \lambda_2(A^\varepsilon), \lambda_2(A) - \lambda_1(A^\varepsilon)\}.$$

Example 4.2.3. (Λ-structure of the Ky Fan metric and Ky Fan distance). The Λ-structure of the Ky Fan metric $\mathbf{K}_\lambda$ (see (3.3.10)) and the Ky Fan distance $\mathbf{KF}_H$ (see 3.3.9)) is handled by assuming that in (4.2.1), $\nu(X, Y; \lambda t) := \Pr(\mathrm{d}(X, Y) > \lambda t)$ and $\nu(X, Y; t) := \Pr(H(d(X, Y)) > t)$, respectively.

4.3 ζ-STRUCTURE OF p. SEMIDISTANCES

In Example 3.2.6 we considered the notion of a minimal norm $\mathring{\mu}_c$

$$\mathring{\mu}_c(P_1, P_2) := \inf\left\{\int_{U^2} c\,\mathrm{d}m: m \in \mathscr{M}_2,\ T_1 m - T_2 m = P_1 - P_2\right\} \tag{4.3.1}$$

where $U = (U, d)$ is a s.m.s. and c is a non-negative, continuous symmetric function on U^2.

Let $\mathscr{F}_{c,1}$ be the space of all bounded $(c, 1)$-Lipschitz functions $f: U \to \mathbb{R}$, i.e.

$$\|f\|_{cL} := \sup_{c(x,y) \neq 0} \frac{|f(x) - f(y)|}{c(x, y)} \leqslant 1. \tag{4.3.2}$$

Remark 4.3.1. If c is a metric in U, then $\mathscr{F}_{c,1}$ is the space of all functions with Lipschitz constant $\leqslant 1$, w.r.t. c. Note that, if c is not a metric, the set $\mathscr{F}_{c,1}$ might be a very 'poor' one. For instance, if $U = \mathbb{R}$, $c(x, y) = |x - y|^p$ $(p > 1)$ then $\mathscr{F}_{c,1}$ contains only the constant functions.

By (4.3.2), we have that for each non-negative measure m on U^2 which marginals $T_i m$, $i = 1, 2$ satisfy $T_1 m - T_2 m = P_1 - P_2$, and for each $f \in \mathscr{F}_{c,1}$, the following inequalities hold:

$$\left|\int_U f(x)(P_1 - P_2)(\mathrm{d}x)\right| = \left|\int_{U^2} (f(x) - f(y))m(\mathrm{d}x, \mathrm{d}y)\right|$$

$$\leqslant \|f\|_{cL} \int_{U^2} c(x, y)m(\mathrm{d}x, \mathrm{d}y) \leqslant \int_{U^2} c(x, y)m(\mathrm{d}x, \mathrm{d}y).$$

The minimal norm $\mathring{\mu}_c$ then has the following estimate from below

$$\zeta(P_1, P_2; \mathscr{F}_c) \leqslant \mathring{\mu}_c(P_1, P_2) \tag{4.3.3}$$

where

$$\zeta(P_1, P_2; \mathscr{F}_{c,1}) := \sup\left\{\left|\int_{U^2} f\,\mathrm{d}(P_1 - P_2)\right| : f \in \mathscr{F}_{c,1}\right\}. \tag{4.3.4}$$

Further, Sections 5.3 and 6.1, we shall prove that for some c (as, for example, $c = d$) we have equality in (4.3.3).

Let $C^b(U)$ be the set of all bounded continuous functions on U. Then, for each subset $\mathfrak{F}$ of $C^b(U)$, the functional

$$\zeta_{\mathfrak{F}}(P_1, P_2) := \zeta(P_1, P_2; \mathfrak{F}) := \sup_{f \in \mathfrak{F}} \left|\int_U f\,\mathrm{d}(P_1 - P_2)\right| \tag{4.3.5}$$

on $\mathscr{P}_1 \times \mathscr{P}_2$ defines a simple p. semimetric in $\mathscr{P}_1$. The metric $\zeta_{\mathfrak{F}}$ was introduced by Zolotarev (1976) and it is called the *Zolotarev* $\zeta_{\mathfrak{F}}$*-metric* (or briefly $\zeta_{\mathfrak{F}}$*-metric*).

Definition 4.3.1. A simple semimetric μ having the $\zeta_{\mathfrak{F}}$*-representation*

$$\mu(P_1, P_2) = \zeta_{\mathfrak{F}}(P_1, P_2) \tag{4.3.6}$$

for some $\mathscr{F} \subseteq C^b(U)$, is called semimetric with ζ*-structure.*

Remark 4.3.2. In the space $\mathfrak{X} = \mathfrak{X}(U)$ of all U-valued r.v.s, the $\zeta_{\mathfrak{F}}$-metric $(\mathfrak{F} \subseteq C^b(U))$ is defined by

$$\zeta_{\mathfrak{F}}(X, Y) := \zeta_{\mathfrak{F}}(\Pr_X, \Pr_Y) := \sup_{f \in \mathfrak{F}} |\mathbb{E}f(X) - \mathbb{E}f(Y)|. \tag{4.3.7}$$

Simple metrics with ζ-structure are well known in probability theory. Let us consider some examples of such metrics.

Example 4.3.1 (engineer metric). Let $U = \mathbb{R}$ and $\mathfrak{X}^{(1)}$ be the set of all real valued r.v.s X with finite first absolute moment, i.e. $\mathbb{E}|X| < \infty$. In the set $\mathfrak{X}^{(1)}$ the engineer metric $\mathbf{EN}(X, Y) := |\mathbb{E}X - \mathbb{E}Y|$ admits the $\zeta_{\mathscr{F}}$-representation, where $\mathscr{F}$ is a collection of functions

$$f_N(x) = \begin{cases} -N & x < N \\ x & |x| \leqslant N \\ N & x > N, N = 1, 2, \ldots \end{cases} \tag{4.3.8}$$

Example 4.3.2 (Kolmogorov metric and $\mathscr{L}_p$-metric in the distribution function space). Let $\mathscr{F} = \mathscr{F}(\mathbb{R})$ be the space of all d.f.s on $\mathbb{R}$. The Kolmogorov metric $\boldsymbol{\rho}(F_1, F_2) := \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)|$ in $\mathscr{F}$ has $\zeta_{\mathfrak{F}}$-structure. In fact

$$\boldsymbol{\rho}(F_1, F_2) = \|F_1 - F_2\|_\infty = \sup\left\{\left|\int_{-\infty}^{\infty} u(x)(F_1(x) - F_2(x))\,\mathrm{d}x\right| : \|u\|_1 \leqslant 1\right\}. \tag{4.3.9}$$

Here and subsequently $\|\cdot\|_p$ $(1 \leqslant p \leqslant \infty)$ stands for the $\mathscr{L}^p$-norm

$$\|u\|_p := \left\{\int_{-\infty}^{\infty} |u(x)|^p\,\mathrm{d}x\right\}^{1/p} \qquad 1 \leqslant p < \infty$$

$$\|u\|_\infty := \operatorname*{ess\,sup}_{x \in \mathbb{R}} |u(x)|.$$

Further, let us denote, by $\mathfrak{F}(p)$, the space of all (Lebesgue) a.e. differentiable functions f such that the derivative f' has $\mathscr{L}^p$-norm $\|f'\|_p \leqslant 1$, hence, integrating by parts the right-hand side of (4.3.9) we obtain a ζ-representation of the uniform metric $\boldsymbol{\rho}$

$$\boldsymbol{\rho}(F_1, F_2) := \sup_{f \in \mathfrak{F}(1)} \left|\int_{-\infty}^{\infty} f(x)\,\mathrm{d}(F_1(x) - F_2(x))\right| = \zeta(F_1, F_2; \mathfrak{F}(1)). \tag{4.3.10}$$

Analogously, we have a $\zeta_{\mathfrak{F}(q)}$-representation for $\boldsymbol{\theta}_p$-metric $(p \geqslant 1)$ (see (3.2.28))

$$\begin{aligned} \boldsymbol{\theta}_p(F_1, F_2) &:= \|F_1 - F_2\|_p \\ &= \sup\left\{\left|\int_{-\infty}^{\infty} u(x)(F_1(x) - F_2(x))\,\mathrm{d}x\right| : \|u\|_q \leqslant 1\right\} \\ &= \zeta(F_1, F_2; \mathscr{F}(q)). \end{aligned} \tag{4.3.11}$$

Next, we shall examine some n-dimensional analogs of (4.3.9) and (4.3.10) by investigating the ζ-structure of (weighted) mean and uniform metrics in the space $\mathscr{F}^n = \mathscr{F}(\mathbb{R}^n)$ of all distribution functions $F(x)$, $x \in \mathbb{R}^n$.

Let $g(x)$ be a positive continuous function on $\mathbb{R}^n$ and let $p \in [1, \infty]$. Define the distances

$$\boldsymbol{\theta}_p(F, G; g) = \left(\int_{\mathbb{R}^n} |F(x) - G(x)|^p g(x)^p \, \mathrm{d}x\right)^{1/p} \quad p \in [1, \infty] \tag{4.3.12}$$

$$\boldsymbol{\theta}_\infty(F, G; g) = \sup\{g(x)|F(x) - G(x)|: x \in \mathbb{R}^n\}. \tag{4.3.13}$$

Remark 4.3.3. In (4.3.2), for $n \geqslant 2$, the weight function $g(x)$ must vanish for all x with $\|x\|_\infty = \max_{1 \leq i \leq n} = \infty$ in order to provide finite values of $\boldsymbol{\theta}_p$.

Let $A_{n,p}$ be the class of real functions f on $\mathbb{R}^n$ having a.e. the derivatives $\mathrm{D}^n f$, where

$$(\mathrm{D}^k f)(x) := \frac{\mathrm{d}^k f}{\mathrm{d}x_1 \cdots \mathrm{d}x_k} \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n \quad k = 1, 2, \ldots, n \tag{4.3.14}$$

and

$$\int_{\mathbb{R}^n} \left|\frac{\mathrm{D}^n f(x)}{g(x)}\right|^q \mathrm{d}x \leq 1 \quad \frac{1}{p} + \frac{1}{q} = 1 \quad \text{if } p > 1 \tag{4.3.15}$$

and

$$|\mathrm{D}^n f(x)| \leqslant g(x) \text{ a.e. if } p = 1.$$

Denote by $g^*(x)$ a continuous function on $\mathbb{R}^n$ such that for some point $a = (a_1, \ldots, a_n)$ the function $g^*(x)$ is non-decreasing (respectively, non-increasing) in the variables x_i if $x_i \geqslant a_i$ (resp., $x_i \leqslant a_i$), $i = 1, \ldots, n$, and $g^* \geqslant g$.

Theorem 4.3.1. Suppose that $p \in [1, \infty]$ and the functions $F, G \in \mathscr{F}^n$ satisfy the following conditions:

(1) $\boldsymbol{\theta}_p(F, G; g) < \infty$.

(2) The derivative $\mathrm{D}^{n-1}(F - G)$ exists a.e. and for any $k = 1, \ldots, n$ the limit relation

$$\lim_{x_k \to \pm\infty} |x_k|^{1/p} g^*(x) |\mathrm{D}^{k-1}(F - G)(x)| = 0 \quad x = (x_1, \ldots, x_n) \tag{4.3.16}$$

holds a.e. for $x_j \in \mathbb{R}^1$, $j \neq k$, $j = 1, \ldots, n$. Then

$$\boldsymbol{\theta}_p(F, G; g) = \zeta(F, G; A_{n,p}). \tag{4.3.17}$$

Proof. As in Equalities (4.3.9) to (4.3.11) we use the duality between $\mathscr{L}^p$ and $\mathscr{L}^q$ spaces. Integrating by parts, and using the tail condition (4.3.16) we get (4.3.17). QED

In the case $n = 1$, we get the following ζ-representation for the mean and uniform metrics with a weight.

Corollary 4.3.1. If $p \in [1, \infty]$, F, $G \in \mathscr{F}^1$, and

$$\lim_{x \to \pm\infty} |x|^{1/p} g^*(x) |F(x) - G(x)| = 0 \tag{4.3.18}$$

then

$$\boldsymbol{\theta}_p(F, G; g) = \zeta(F, G; A_{1,p}). \tag{4.3.19}$$

As a consequence of Theorem 4.3.1 we shall investigate further upper estimates of some classes of ζ-metrics with the help of metrics of type $\boldsymbol{\theta}_p(\cdot, \cdot; g)$. This is connected with the problem of characterizing 'uniform classes' with respect to $\boldsymbol{\theta}_p(\cdot, \cdot; g)$-convergence.

Definition 4.3.2. If μ is a metric on $\mathscr{F}^n$, then a class A of measurable functions on $\mathbb{R}^n$ is called a *uniform class with respect to μ-convergence* (briefly, a μ-u.c.) if for any F_n ($n = 1, 2, \ldots$) and $F \in \mathscr{F}^n$ the condition $\mu(F_n, F) \to 0$ ($n \to \infty$) implies that $\zeta_A(F_n, F) \to 0$ ($n \to \infty$).

Bhattacharya and Ranga Rao (1976), Kantorovich and Rubinstein (1958), Billingsley (1968) and Dudley (1976) have studied uniform classes w.r.t. weak convergence. It is clear that $A_{n,p}$ is an $\boldsymbol{\theta}(\cdot, \cdot; g)$-u.c. in the set of distribution functions satisfying (1) and (2) of Theorem 4.3.1.

Let $\mathscr{G}_{n,p}$ be the class of all functions in $A_{n,p}$ such that for any tuple $I = (1, \ldots, k)$, $1 \leqslant k \leqslant n-1$, we have

$$\mathrm{D}_I^k f(x^I) = 0 \text{ a.e.} \quad x^I \in \mathbb{R}^n, \quad x_i^I = \begin{cases} x_i & \text{if } i \in I \\ +\infty & \text{if } i \notin I. \end{cases}$$

Any function in $A_{n,p}$ constant outside a compact set obviously belongs to the class $\mathscr{G}_{n,p}$. Now we can omit the restriction (4.3.16) to get:

Corollary 4.3.2. For any F, $G \in \mathscr{F}^n$

$$\zeta(F, G; \mathscr{G}_{n,p}) \leqslant \theta_p(F, G; g) \quad p \in [1, \infty]. \tag{4.3.20}$$

In the case of uniform metric

$$\boldsymbol{\rho}_n(F, G) := \sup_{x \in \mathbb{R}^n} |F(x) - G(x)| = \boldsymbol{\theta}_\infty(F, G; 1) \tag{4.3.21}$$

we get the following refinement of Corollary 4.3.2. Denote by B_n the set of all real functions on $\mathbb{R}^n$ having a.e. the derivatives $\mathrm{D}^n f$ such that for any $I = (i_1, \ldots, i_k)$, $1 \leqslant k \leqslant n$, $1 \leqslant i_1 < \cdots < i_k \leqslant n$,

$$\int_{\mathbb{R}^k} |\mathrm{D}_I^k f(x^I)| \, \mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_k} \leqslant 1.$$

Denote by $F_I(x_1, \ldots, x_i) = F(x^I)$ the marginal distribution of $F \in \mathscr{F}^n$ on the first k coordinates.

Corollary 4.3.3. For any F, $G \in \mathscr{F}^n$

$$\zeta(F, G; B_n) \leqslant \sum_{\substack{I=(i,\ldots,k)\\ 1 \leqslant k \leqslant n}} \rho_k(F_I, G_I). \tag{4.3.22}$$

The obvious inequality (see (4.3.22))

$$\zeta(F, G; B_n) \leqslant n\rho_n(F, G) \tag{4.3.23}$$

implies that B_n is ρ_n-u.c.

Open problem 4.3.1. Investigating the uniform estimates of the rate of convergence in the multi-dimensional central limit theorem the authors (see, for instance, Sazonov (1981), Senatov (1980)) consider the following metric

$$\rho(P, Q; \mathscr{CB}) = \sup\{|P(A) - Q(A)|: A \in \mathscr{CB},\ P, Q \in \mathscr{P}(\mathbb{R}^n)\} \tag{4.3.24}$$

where $\mathscr{CB}$ denotes the set of all convex Borel subsets of $\mathbb{R}^n$. The metric $\rho(\cdot, \cdot; \mathscr{CB})$ may be viewed as a generalization of the notion of uniform metric ρ on $\mathscr{P}(\mathbb{R}^1)$; that is why $\rho(\cdot, \cdot; \mathscr{CB})$ is called the 'uniform metric' in $\mathscr{P}(\mathbb{R}^n)$. However, using the ζ-representation (4.3.10) of the Kolmogorov metric ρ on $\mathscr{P}(\mathbb{R}^1)$ it is possible to extend the notion of uniform metric in a way which is different from (4.3.24). Namely, define the '*uniform* $\rho\zeta$-*metric*' in $\mathscr{P}(\mathbb{R}^n)$ as follows

$$\rho\zeta(P, Q) := \zeta(P, Q; A_{n,1}(1)) \tag{4.3.25}$$

where $A_{n,1}(1)$ is the class of real functions f on $\mathbb{R}^n$ having a.e. the derivatives $\mathrm{D}^n f$ and

$$\int_{\mathbb{R}^n} |\mathrm{D}^n f(x)|\, \mathrm{d}x \leqslant 1. \tag{4.3.26}$$

What kind of quantitative relationships exist between the metric ρ_n, $\rho(\cdot, \cdot; \mathscr{CB})$ and $\rho\zeta$ (see (4.3.21), (4.3.24), (4.3.25)? Such relationships would yield the rate of convergence for the CLT in terms of $\rho\zeta$.

Example 4.3.3. (ζ-metrics that metrize G-*weak convergence).* In Example 4.2.2 we have considered Λ-metric that metrizes G-weak convergence in $\mathscr{P}_G \subseteq \mathscr{P}(U)$ (see Definition 4.2.2 and (4.2.5)). Now, we shall be interested in ζ-metrics generating G-weak convergence in $\mathscr{P}_G$. Let $\mathbb{F} = \mathbb{F}(G)$ be the class of real-valued functions f on a s.m.s. U such that the following conditions hold:

(i) $\mathbb{F}$ is an equicontinuous class, i.e.,

$$\lim_{d(y,x)\to 0} \sup_{f \in \mathbb{F}} |f(x) - f(y)| = 0$$

(ii) $$\sup_{f\in\mathbb{F}}|f(x)| \leqslant G(x) \quad \forall x\in U$$

(iii) $\alpha G\in\mathbb{F}$ for some constant $\alpha\neq 0$;

(iv) For each non-empty closed set $C\subseteq U$ and for each integer k, the function

$$f_{k,C}(x):=\max\{0,\ 1/k-d(x,C)\}$$

belongs to $\mathbb{F}$.

Note that if $\mathbb{F}$ satisfies (i) and (ii) only, then $\mathbb{F}$ is $\pi_{\lambda,G}$-u.c. (see Definition 4.3.2 and (4.2.5)), i.e. G-weak convergence implies $\zeta_{\mathbb{F}}$-convergence (cf. Bhattacharya and Ranga Rao 1976, Rao 1962). The next theorem determines the cases in which $\zeta_{\mathbb{F}}$-convergence is equivalent to G-weak convergence.

Theorem 4.3.2. If $\mathbb{F}=\mathbb{F}(G)$ satisfies (i)–(iv) then $\zeta_{\mathbb{F}}$ metrizes the G-weak convergence in $\mathscr{P}_G$.

In fact, we shall prove a more general result (cf. further Section 10.1, Theorem 10.1.2).

Let us consider some particular cases of the classes $\mathbb{F}(G)$.

Case A. Let c be a fixed point of U, a and b positive constants, and $h\colon[0,\infty]\to[0,\infty]$ a non-decreasing function, $h(0)=0$, $h(\infty)\leqslant\infty$. Define the class $S=S(a,b,h)$ of all functions $f\colon U\to\mathbb{R}$ such that

$$\|f\|_\infty:=\sup_{x\in U}|f(x)|\leqslant a \tag{4.3.27}$$

and

$$\mathrm{Lip}_h(f):=\sup_{x\neq y,\,x,y\in U}\frac{|f(x)-f(y)|}{d(x,y)\max\{1,h(d(x,c)),h(d(y,c))\}}\leqslant b. \tag{4.3.28}$$

Corollary 4.3.4. (a) If $0<a<\infty$, $0<b<\infty$, then $\zeta_{S(a,b,h)}$ metrizes the weak convergence in $\mathscr{P}(U)$.

(b) If $a=\infty$, $b<\infty$ and

$$\sup_{t\neq s}\frac{|t\max\{1,h(t)\}-s\max\{1,h(s)\}|}{|t-s|\max\{1,h(t),h(s)\}}<\infty \tag{4.3.29}$$

then $\zeta_{S(a,b,h)}$ metrizes the G-weak convergence with

$$G(x)=d(x,c)\max\{1,h(d(x,c))\}.$$

Case B. Fortet and Mourier (1953) investigated the following two $\zeta_{\mathbb{F}}$-metrics.

(a) $\zeta(\cdot,\cdot;\mathscr{G}^p)$ $(p\geqslant 1)$, where the class $\mathscr{G}^p$ is defined as follows. For each function $f: U \to \mathbb{R}$ let

$$L(f, t) := \sup\left\{\frac{f(x) - f(y)}{d(x, y)} : x \neq y, d(x, c) \leqslant t, d(y, c) \leqslant t\right\} \tag{4.3.30}$$

and

$$M(f) := \sup_{t>0} \frac{L(f, t)}{\max(1, t^{p-1})}. \tag{4.3.31}$$

Then

$$\mathscr{G}^p := \{f: U \to \mathbb{R}, M(f) \leqslant 1\}. \tag{4.3.32}$$

(b) $\zeta(\cdot, \cdot; \bar{\mathscr{G}}^p)$ where

$$\bar{\mathscr{G}}^p := \{f \in \mathscr{G}^p, \|f\|_\infty \leqslant 1\}. \tag{4.3.33}$$

Lemma 4.3.1. Let $h_p(t) = t^{p-1}$ $(p \geqslant 1, t \geqslant 0)$. Then

$$\zeta(P, Q; \mathscr{G}^p) = \zeta(P, Q; S(\infty, 1, h_p)) \tag{4.3.34}$$

and

$$\zeta(P, Q; \bar{\mathscr{G}}^p) = \zeta(P, Q; S(1, 1, h_p)). \tag{4.3.35}$$

Proof. It is enough to check that $\mathrm{Lip}_{h_p}(f) = M(f)$. Actually, let $x \neq y$ and $t_0 := \max\{d(x, c), d(y, c)\}$. Then $t_0 > 0$ and $|f(x) - f(y)| \leqslant L(f, t_0)\ d(x, y) \leqslant M(f)\max(1, t_0^{p-1})d(x, y)$, hence, $\mathrm{Lip}_{h_p}(f) \leqslant M(f)$. Conversely, for each $t > 0$ $L(f, t) \leqslant \mathrm{Lip}_{h_p}(f)\max(1, t^{p-1})$ and thus $M(f) \leqslant \mathrm{Lip}_{h_p}(f)$. QED

Corollary 4.3.4 and Lemma 4.3.1 imply the following corollary.

Corollary 4.3.5. Let (U, d) be a s.m.s. Then

(i) $\zeta(\cdot, \cdot; \bar{\mathscr{G}}^p)$ metrizes the weak convergence in $\mathscr{P}(U)$,
(ii) In the set

$$\mathscr{P}^{(p)}(U) := \left\{P \in \mathscr{P}(U), \int_U d^p(x, c)P(\mathrm{d}x) < \infty\right\} \tag{4.3.36}$$

the $\zeta(\cdot, \cdot; \mathscr{G}^p)$-convergence is equivalent to the G-weak convergence with $G(x) = d^p(x, c)$.

Case C. Dudley (1966a, 1976) considered $\boldsymbol{\beta}$-metric in $\mathscr{P}(U)$ that is defined as $\zeta_{\mathbb{F}}$-metric with

$$\mathbb{F} := \left\{f: U \to \mathbb{R}, \|f\|_\infty + \sup_{x, y \in U, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} \leqslant 1\right\}. \tag{4.3.37}$$

Corollary 4.3.6 (Dudley 1966a). The *Dudley metric* $\boldsymbol{\beta} := \zeta_{\mathbb{F}}$ defined by (4.3.7) and (4.3.37) metrizes the weak convergence in $\mathscr{P}(U)$.

Proof. Using Corollary 4.3.5 (i) with $p = 1$ and the inequality

$$\tfrac{1}{2}\zeta(P, Q; \mathscr{G}^1) \leqslant \boldsymbol{\beta}(P, Q) \leqslant \zeta(P, Q; \mathscr{G}^1) \quad P, Q \in \mathscr{P}(U) \tag{4.3.38}$$

we claim that $\boldsymbol{\beta}$ induces weak convergence in $\mathscr{P}(U)$. QED

Case D. The Kantorovich metric ℓ_1 (see (3.2.12), (3.2.17)) admits the ζ-representation $\zeta(\cdot, \cdot; \mathscr{G}^1)$, as well as $\ell_p (0 < p \leqslant 1)$ (see (3.2.12)) has the form

$$\ell_p(P_1, P_2) = \zeta(P_1, P_2; \mathscr{G}^1) \quad P_1, P_2 \in \mathscr{P}(U) \quad U = (U, d^p). \tag{4.3.39}$$

In the right-hand side of (4.3.39), U is a s.m.s. with metric d^p, i.e., in the definition of $\zeta(\cdot, \cdot; \mathscr{G}^1)$ (see (4.3.30), (4.3.34)) we replace the metric d with d^p.

Now let us touch on some special cases of (4.3.39).

(a) Let U be a separable normed space with norm $\|\cdot\|$ and $Q: U \to U$ be a function on U such that the metric $d_Q(x, y) = \|Q(x) - Q(y)\|$ metrizes the space U as a s.m.s. For instance, if Q is a homeomorphism of U, i.e. Q is a one-to-one function and both Q and Q^{-1} are continuous, then (U, d_Q) is a s.m.s. Further, let $p = 1$ and $d = d_Q$ in (4.3.39). Then

$$\boldsymbol{\kappa}_Q(P_1, P_2) := \ell_1(P_1, P_2) = \sup\left\{\left|\int_U f \, \mathrm{d}(P_1 - P_2)\right| : f: U \to \mathbb{R},\right.$$
$$\left. |f(x) - f(y)| \leqslant d_Q(x, y) \quad \forall x, y \in U\right\} \tag{4.3.40}$$

is called a *Q-difference pseudomoment in* $\mathscr{P}(U)$.

If U is a separable normed space and Q is a homeomorphism of U then (noting our earlier discussions in Theorem 2.5.1 and Example 3.2.2), in the space $\mathfrak{X}(U)$ of U-valued r.v.s, $\kappa_Q(X, Y) := \kappa_Q(\Pr_X, \Pr_Y)$ is the minimal metric w.r.t. the *compound Q-difference pseudomoment*

$$\boldsymbol{\tau}_Q(X, Y) := \mathbb{E} d_Q(X, Y) \tag{4.3.41}$$

and

$$\boldsymbol{\kappa}_Q(X, Y) = \hat{\boldsymbol{\tau}}_Q(X, Y) = \sup\{|\mathbb{E}[f(Q(X)) - f(Q(Y))]| : f: U \to \mathbb{R},$$
$$|f(x) - f(y)| \leqslant \|x - y\| \quad \forall x, y \in U\}. \tag{4.3.42}$$

In the particular case $U = \mathbb{R}$, $\|x\| = |x|$,

$$Q(x) := \int_0^x q(u) \, \mathrm{d}u \quad q(u) \geqslant 0, u \in \mathbb{R}, x \in \mathbb{R}$$

the metric κ_Q has the following explicit representation

$$\boldsymbol{\kappa}_Q(P_1, P_2) := \boldsymbol{\kappa}_Q(F_1, F_2) := \int_{-\infty}^{\infty} q(x)|F_1(x) - F_2(x)|\,\mathrm{d}x. \tag{4.3.43}$$

If, in (4.3.40), $Q(x) = x\|x\|^{s-1}$ for some $s > 0$, then $x_s := x_Q$ is called *s-difference pseudomoment* (see Zolotarev 1976b, 1977b, 1978, Hall 1981).

(b) By (4.3.39), we have that

$$\ell_p(P_1, P_2) := \sup\left\{\left|\int_U f\,\mathrm{d}(P_1 - P_2)\right| : f\colon U \to \mathbb{R},\ |f(x) - f(y)| \leqslant d^p(x, y), x, y \in U\right\} \tag{4.3.44}$$

for any $p \in (0, 1)$. Hence, letting $p \to 0$ and defining the indicator metric

$$i(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

we get

$$\begin{aligned}
&\lim_{p \to 0} \ell_p(P_1, P_2) \\
&\quad = \sup\left\{\left|\int_U f\,\mathrm{d}(P_1 - P_2)\right| : f\colon U \to \mathbb{R},\ |f(x) - f(y)| \leqslant i(x, y) \quad \forall x, y \in U\right\} \\
&\quad = \boldsymbol{\sigma}(P_1, P_2) = \ell_0(P_1, P_2)
\end{aligned}$$

where $\boldsymbol{\sigma}$ (resp., ℓ_0) is the total variation metric (see (3.2.13)).

Examples 4.3.1 to 4.3.3 show that the ζ-structure encompasses the simple metrics ℓ_p that are minimal with respect to the compound metric $\mathscr{L}_p$ (see (3.3.18) for $p \in [0, 1]$). If however $p > 1$ then $\ell_p = \hat{\mathscr{L}}_p$ (see Equalities (3.3.18), (3.3.3), (3.2.11)) has a form different from the ζ-representation, namely

$$\ell_p(P_1, P_2) = \sup\left\{\left[\int_U f\,\mathrm{d}P_1 + \int_U g\,\mathrm{d}P_2\right]^{1/p} : (f, g) \in \mathscr{G}_p\right\} \tag{4.3.46}$$

where $\mathscr{G}_p$ is the class of all pairs (f, g) of Lipschitz bounded functions f, $g \in \mathrm{Lip}^b(U)$ (cf. (3.2.8)) that satisfy the inequality

$$f(x) + g(y) \leqslant d^p(x, y) \quad x, y \in U. \tag{4.3.47}$$

The following lemma shows that $\ell_p = \hat{\mathscr{L}}_p$ $(p > 1)$ has no ζ-representation.

Lemma 4.3.2. (Neveu and Dudley 1980). If a s.m.s. (U, d) has more than one point and the minimal metric $\hat{\mathscr{L}}_p$ $(p \geqslant 1)$ has a ζ-representation (4.3.5), then $p = 1$.

Proof. Assuming that $\hat{\mathscr{L}}_p$ has $\zeta_{\mathbb{F}}$-representation for a certain class $\mathbb{F} \subseteq C^b(U)$ then

$$\sup_{f \in \mathbb{F}} \left\{ \left| \int_U f \, \mathrm{d}(P_1 - P_2) \right| \right\} = \hat{\mathscr{L}}_p(P_1, P_2) \quad \forall P_1, P_2 \in \mathscr{P}_1(U) \tag{4.3.48}$$

If in (4.3.48), the law P_1 is concentrated at the point x and P_2 is concentrated at y, then $\sup\{|f(x) - f(y)| : f \in \mathbb{F}\} \leqslant d(x, y)$. Thus, $\mathbb{F}$ is contained in the Lipschitz class

$$\begin{aligned} \mathrm{Lip}^b_{1,1} &= \mathrm{Lip}^b_{1,1}(U) \\ &:= \{f\colon U \to \mathbb{R},\ f \text{ bounded},\ |f(x) - f(y)| \leqslant d(x, y) \quad \forall x, y \in U\}. \end{aligned} \tag{4.3.49}$$

For each law $P \in \mathscr{P}_2$ with marginals P_1 and P_2

$$\begin{aligned} \hat{\mathscr{L}}_p(P_1, P_2) &\leqslant \sup_{f \in \mathrm{Lip}^b_{1,1}} \left| \int_U f \, \mathrm{d}(P_1 - P_2) \right| \\ &\leqslant \sup_{f \in \mathrm{Lip}^b_{1,1}} \int_{U^2} |f(x) - f(y)| P(\mathrm{d}x, \mathrm{d}y) \leqslant \mathscr{L}_1(P). \end{aligned}$$

Next, we can pass to the minimal metric $\hat{\mathscr{L}}_1$ in the right-hand side of the above inequality and then claim $\hat{\mathscr{L}}_p = \hat{\mathscr{L}}_1$. In particular, by the Minkovski inequality we have

$$\left\{ \int_U d^p(x, a) P_1(\mathrm{d}x) \right\}^{1/p} - \left\{ \int_U d^p(x, a) P_2(\mathrm{d}x) \right\}^{1/p} \leqslant \hat{\mathscr{L}}_p(P_1, P_2) = \hat{\mathscr{L}}_1(P_1, P_2). \tag{4.3.50}$$

Assuming that there exists $b \in U$ such that $d(a, b) > 0$, let us consider the laws P_1, P_2 with $P_1(\{a\}) = r \in (0, 1)$, $P_1(\{b\}) = 1 - r$, $P_2(\{a\}) = 1$, then, the ζ-representation of $\hat{\mathscr{L}}_1 = \ell_1$ (see (3.2.12), (3.3.18)),

$$\begin{aligned} \hat{\mathscr{L}}_1(P_1, P_2) &= \sup_{f \in \mathrm{Lip}^b_{1,1}} |rf(a) + (1 - r) f(b) - f(a)| \\ &= (1 - r) \sup_{f \in \mathrm{Lip}^b_{1,1}} |f(a) - f(b)| \leqslant (1 - r) d(a, b) \end{aligned}$$

and hence

$$(1 - r) d(a, b) \geqslant \hat{\mathscr{L}}_1(P_1, P_2) \geqslant \{d^p(b, a)(1 - r)\}^{1/p} = (1 - r)^{1/p} d(a, b)$$

i.e., $p = 1$. QED

Remark 4.3.4. A. Szulga made a conjecture that $\hat{\mathscr{L}}_p(p > 1)$ has a dual form close to that of ζ-metric, namely

$$\hat{\mathscr{L}}_p(P_1, P_2) = \mathbf{AS}_p(P_1, P_2) \quad P_1, P_2 \in \mathscr{P}^{(p)}(U). \tag{4.3.51}$$

In (4.3.49), the class $\mathscr{P}^{(p)}(U)$ consists of all laws P with finite 'pth moment', $\int d^p(x, a)P(\mathrm{d}x) < \infty$ and

$$\mathrm{AS}_p(P_1, P_2) := \sup_{f \in \mathrm{Lip}^b_{1,1}} \left| \left\{ \int_U |f|^p \,\mathrm{d}P_1 \right\}^{1/p} - \left\{ \int_U |f|^p \,\mathrm{d}P_2 \right\}^{1/p} \right|. \tag{4.3.52}$$

By the Minkovski inequality it follows easily that

$$AS_p \leqslant \hat{\mathscr{L}}_p. \tag{4.3.53}$$

Moreover, the following lemma shows that the Szulga's conjecture is partially true in the sense that $\hat{\mathscr{L}}_p \overset{\text{top}}{\sim} \mathrm{AS}_p$.

Lemma 4.3.3. In the space $\mathscr{P}^{(p)}(U)$ the metrics AS_p and $\hat{\mathscr{L}}_p$ generate one and the same topology.

Proof. It is known that (see further Section 8.2, Corollary 8.2.1) $\hat{\mathscr{L}}_p$ metrizes G_p-weak convergence in $\mathscr{P}^{(p)}(U)$ (cf. Definition 4.2.2), where $G_p(x) = d^p(x, a)$. Hence, by (4.3.51) it is sufficient to prove that AS_p-convergence implies, G_p-weak convergence. In fact, since $G_1 \in \mathrm{Lip}_{1,1}$ then

$$\mathrm{AS}_p(P_n, P) \to 0 \Rightarrow \int_U d^p(x, a)P_n(\mathrm{d}x) \to \int_U d^p(x, a)P(\mathrm{d}x). \tag{4.3.54}$$

Further, for each closed non-empty set C and $\varepsilon > 0$ let

$$f_C(x) := \max\left(0, 1 - \frac{1}{\varepsilon} d(x, C)\right).$$

Then $f_C \in \mathrm{Lip}_{1/\varepsilon, 1}(U)$ (see (3.2.6)) and

$$\begin{aligned} P_n^{1/p}(C) &\leqslant \left\{ \int_U f_C^p \,\mathrm{d}P_n \right\}^{1/p} \\ &\leqslant \left\{ \int_U f_C^p \,\mathrm{d}P \right\}^{1/p} + \frac{1}{\varepsilon} \mathrm{AS}_p(P_n, P) \\ &\leqslant \{P(C^\varepsilon)\}^{1/p} + \frac{1}{\varepsilon} \mathrm{AS}_p(P_n, P) \end{aligned}$$

which implies

$$\mathrm{AS}_p(P_n, P) \to 0 \Rightarrow P_n \xrightarrow{w} P \tag{4.3.55}$$

as desired. QED

By Lemma 4.3.2 it follows, in particular, that there exist simple metrics that have no $\zeta_{\mathbb{F}}$-representation. In the case of $\hat{\mathscr{L}}_p$-metric, however, we can find a

$\zeta_{\mathbb{F}}$-metric which is topologically equivalent to $\hat{\mathscr{L}}_p$. Namely

$$\hat{\mathscr{L}}_p \overset{\text{top}}{\sim} \zeta_{\mathscr{G}^p} \tag{4.3.56}$$

(see (4.3.6), (4.3.34) and Corollary 4.3.5(ii)). Also, it is not difficult to see that the Prokhorov metric $\boldsymbol{\pi}$ (see (3.2.20)) has no $\zeta_{\mathbb{F}}$-representation, even in the case $U = \mathbb{R}$, $d(x, y) = |x - y|$. In fact, assume that

$$\boldsymbol{\pi}(P, Q) = \zeta_{\mathbb{F}}(P, Q) \quad \forall P, Q \in \mathscr{P}(\mathbb{R}). \tag{4.3.57}$$

Denoting the measure concentrated at the point x by P_x we have

$$\boldsymbol{\pi}(P_x, P_y) = \min(1, |x - y|) \leqslant |x - y| \tag{4.3.58}$$

and hence, by (4.3.56)

$$|x - y| \geqslant \boldsymbol{\pi}(P_x, P_y) = \sup_{f \in \mathbb{F}} |f(x) - f(y)|$$

hence,

$$\boldsymbol{\pi}(P, Q) \leqslant \sup\left\{\left|\int f \,\mathrm{d}(F - G)\right| : f\colon U \to \mathbb{R}, f\text{-bounded}, \; |f(x) - f(y)| \leqslant |x - y|,\ x, y \in \mathbb{R}\right\}$$

$$\leqslant \int_{-\infty}^{\infty} |F(x) - G(x)| \,\mathrm{d}x =: \boldsymbol{\kappa}(F, G)$$

where F is the d.f. of P and G is the d.f. of Q. Obviously $\boldsymbol{\pi}(P, Q) \geqslant \mathbf{L}(F, G)$ where $\mathbf{L}$ is the Lévy metric in the distribution function space $\mathscr{F}$ (cf. (4.1.3)). Hence, the equality (4.3.57) implies

$$\mathbf{L}(F, G) \leqslant \boldsymbol{\kappa}(F, G) \quad \forall F, G \in \mathscr{F}. \tag{4.3.59}$$

Let $1 > \varepsilon > 0$ and

$$F_\varepsilon(x) = \begin{cases} 0 & x \leqslant 0 \\ 1 - \varepsilon & 0 < x \leqslant \varepsilon, \\ 1 & x > \varepsilon, \end{cases} \qquad G_\varepsilon(x) = \begin{cases} 0 & x \leqslant 0 \\ 1 & x > 0. \end{cases}$$

Then the equalities

$$\boldsymbol{\kappa}(F, G) = \varepsilon^2 = \mathbf{L}^2(F, G)$$

contradict (4.3.59), hence, $\boldsymbol{\pi}$ does not admit a ζ-representation.

Although there is no ζ-representation for the Prokhorov metric $\boldsymbol{\pi}$, nevertheless $\boldsymbol{\pi}$ is topologically equivalent to various ζ-metrics. To see this, simply note that both $\boldsymbol{\pi}$ and certain ζ-metrics metrize weak convergence (Corollary 4.3.4a).

Therefore the following question arises: is there a simple metric μ such that

$$\mu \overset{\text{top}}{\sim} \zeta_{\mathbb{F}}$$

fails for any set $\mathbb{F} \subseteq C^b(U)$? The following lemma gives an affirmative answer to this question where $\mu = \pi\mathbf{H}$ (see Equations (4.1.38) and (4.1.43)) and if $U = \mathbb{R}$, $d(x, y) = |x - y|$ then one can take $\mu = \mathbf{H}$ (see (4.1.33) and Fig. 4.1.2).

Lemma 4.3.4. Let $\lambda > 0$ and let (U, d) be a metric space containing a non-constant sequence $a_1, a_2, \ldots \to a \in U$;

(i) If (U, d) is a s.m.s then there is no set $\mathbb{F} \subseteq C^b(U)$ such that $\pi\mathbf{H}_\lambda \overset{\text{top}}{\sim} \zeta_{\mathbb{F}}$.

(ii) If $U = \mathbb{R}$, $d(x, y) = |x - y|$, then there is no set $\mathbb{F} \subseteq C^b(U)$ such that $\mathbf{H}_\lambda \overset{\text{top}}{\sim} \zeta_{\mathbb{F}}$.

Proof. We shall consider only the case (i) with $\lambda = 1$. Choose the laws P_n and P as follows: $P(\{a\}) = 1$, $P_n(\{a\}) = P_n(\{a_n\}) = \frac{1}{2}$. Then for each $B \in \mathscr{B}_1$, the measure P takes values 0 or 1 and thus

$$\pi\mathbf{H}_1(P_n, P) \geqslant \inf_{B \in \mathscr{B}_1} \max\{d(a_n, B), |P_n(a_n) - P(B)|\} \geqslant \tfrac{1}{2}. \tag{4.3.60}$$

Assuming that $\pi\mathbf{H}_1 \overset{\text{top}}{\sim} \zeta_{\mathbb{F}}$ we have, by (4.3.60), that

$$\begin{aligned} 0 &< \limsup_{n \to \infty} \zeta_{\mathbb{F}}(P_n, P) \\ &= \limsup_{n \to \infty} |\tfrac{1}{2} f(a) + \tfrac{1}{2} f(a_n) - f(a)| \\ &= \tfrac{1}{2} \limsup_{n \to \infty} |f(a) - f(a_n)|. \end{aligned} \tag{4.3.61}$$

Further, let $Q_n(\{a_n\}) = 1$. Then $\pi\mathbf{H}_1(Q_n, P) \to 0$ and hence

$$\begin{aligned} 0 &= \limsup_{n \to \infty} \zeta_{\mathbb{F}}(Q_n, P) \\ &= \limsup_{n \to \infty} |f(a) - f(a_n)|. \end{aligned} \tag{4.3.62}$$

The relationships (4.3.61) and (4.3.62) give the necessary contradiction. QED

Lemma 4.3.4 claims that the ζ-structure of simple metrics does not describe all possible topologies arising from simple metrics. Next, we shall extend the notion of ζ-structure in order to encompass all simple p. semidistances as well as all compound p. semidistances. To this end, note first that for the compound metric $\mathscr{L}_p(X, Y)$ ($p \geqslant 1$, $X, Y \in \mathfrak{X}(\mathbb{R})$) (see (3.3.3) with $d(x, y) = |x - y|$, $U = \mathbb{R}$) we have the dual representation (Neveu, 1965)

$$\mathscr{L}_p(X, Y) = \sup\{|\mathbb{E}(XZ - YZ)|: Z \in \mathfrak{X}(\mathbb{R}), \mathscr{L}_q(Z, 0) \leqslant 1\},$$
$$1 \leqslant p \leqslant \infty \quad 1/p + 1/q = 1. \quad (4.3.63)$$

The next definition generalizes the notion of ζ-structure as well as the metric structure of $\hat{\mathscr{L}}_H$-distances (see (3.2.10), (3.3.17)) and $\mathscr{L}_p$-metrics (see (3.3.3)).

Definition 4.3.3. We say that a p. semidistance μ admits a *$\bar{\zeta}$-structure* if μ can be written in the following way

$$\mu(X, Y) = \bar{\zeta}(X, Y; \bar{\mathbb{F}}(X, Y)) = \sup_{f \in \bar{\mathbb{F}}(X, Y)} \mathbb{E}f \qquad (4.3.64)$$

where $\mathbb{F}(X, Y)$ is a class of integrable functions $f: \Omega \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given on a probability space $(\Omega, \mathscr{A}, \Pr)$.

In general $\bar{\zeta}$ is not a p. semidistance, but each p. semidistance has a $\bar{\zeta}$-representation. Actually, for each p. semidistance μ, the equality (4.3.64) is valid where $\bar{\mathbb{F}}(X, Y)$ contains only a constant function $\mu(X, Y)$.

Let us consider some examples of $\bar{\zeta}$-structures of p. semidistances.

Example 4.3.4. $\mathscr{L}_H$ (see (3.3.1)) has a trivial $\bar{\zeta}$-representation, where $\bar{\mathbb{F}}(X, Y)$ contains only the function $H(d(X, Y))$.

Example 4.3.5. $\mathscr{L}_p$ on $\mathfrak{X}(\mathbb{R})$ (see (4.3.63)) enjoys a non-trivial $\bar{\zeta}$-representation where

$$\bar{\mathbb{F}}(X, Y) = \{f_Z(X, Y) = XZ - YZ: \mathscr{L}_q(Z, 0) \leqslant 1\}.$$

Example 4.3.6. The simple distance ℓ_H (see (3.2.10)) has a $\bar{\zeta}$-representation, where

$$\bar{\mathbb{F}}(X, Y) = \{f(X, Y) = f_1(X) + f_2(Y), (f_1, f_2) \in \mathscr{G}_H(U)\}$$

for each $X, Y \in \mathfrak{X}$.

Example 4.3.7. $\zeta_{\mathbb{F}}$-structure of the simple metrics is a particular case of $\bar{\zeta}$-structure with

$$\bar{\mathbb{F}}(X, Y) = \{f(X, Y) = f(X) - f(Y): f \in \mathbb{F}\} \cup \{f(X, Y) = f(Y) - f(X): f \in \mathbb{F}\}.$$

We completed the investigation of the three universal metric structures (h, Λ, and $\bar{\zeta}$). The reason we call them universal is that each p. semidistance μ has h-, Λ- and $\bar{\zeta}$-representation simultaneously. Thus, depending on the specific problem under consideration, one can use one or another p. semidistance representation.

PART II

Relations Between Compound, Simple and Primary Distances

In this part we shall concern ourselves with the study of the dual and explicit representations of minimal distances and norms as well as the topologies which these metric structures induce in the space of probability measures.

CHAPTER 5

Monge–Kantorovich Mass Transference Problem. Minimal Distances and Minimal Norms

5.1 STATEMENT OF MONGE–KANTOROVICH PROBLEM

This section should be viewed as an introduction to the Monge–Kantorovich problem (MKP) and its related probability semidistances. There are six known versions of the MKP.

(I) *Monge transportation problem.* In 1781, Monge formulated the following problem in studying the most efficient way of transporting soil.

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

In other words, two sets S_1 and S_2 are the supports of two masses μ_1, μ_2 with equal total weight $\mu_1(S_1) = \mu_2(S_2)$. The 'initial' mass μ_1 is to be transported from S_1 to S_2 so that the result is the 'final' mass μ_2. The transportation should be realized in such a way as to minimize the total labor involved.

(II) *Kantorovich's mass transference problem.* In the Monge problem let A and B be initial and final volumes. For any set $a \subset A$ and $b \subset B$ let $P(a, b)$ be the fraction of volume of A that was transferred from a into b. Note that $P(a, B)$ is equal to the ratio of volumes of a and A and $P(A, b)$ is equal to the ratio of volumes of b and B, respectively.

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

$$\int_{U \times U} c(x, y) P(\mathrm{d}x, \mathrm{d}y). \tag{5.1.1}$$

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

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

$$\mathscr{A}_c(P_1, P_2) = \inf\left\{\int_{U \times U} c(x, y) P(\mathrm{d}x, \mathrm{d}y) \colon P \in \mathscr{P}^{(P_1, P_2)}\right\} \tag{5.1.2}$$

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

We shall call the functional (5.1.2) *Kantorovich's functional*, (resp. *Kantorovich metric*) if $c = d$ (cf. Example 3.2.2, (3.3.18), (3.3.54)).

The measures P_1 and P_2 may be viewed as the initial and final distribution of mass and $\mathscr{P}^{(P_1, P_2)}$ as the space of admissible transference plans. If the infimum in (5.1.2) is realized for some measure $P^* \in \mathscr{P}^{(P_1, P_2)}$ then P^* is said to be the *optimal transference plan.* The function $c(x, y)$ can be interpreted as the cost of transferring the mass from x to y.

Remark 5.1.1. Kantorovich's formulation differs from the Monge problem in that the class $\mathscr{P}^{(P_1, P_2)}$ is broader than the class of one-to-one transference plans in Monge's sense. Sudakov (1976) showed that if measures P_1 and P_2 are given on a bounded subset of a finite-dimensional Banach space and are absolutely continuous with respect to Lebesgue measure, then there exists an optimal one-to-one transference plan.

Remark 5.1.2. Another example of the MKP is assigning army recruits to jobs to be filled. The flock of recruits has a certain distribution of parameters such as education, previous training, and physical conditions. The distribution of parameters which are necessary to fill all the jobs might not necessarily coincide with one of the contingents. There is a certain cost involved in training of an individual for a specific job depending on the job requirements and individual parameters, thus the problem of assigning recruits to the job and training them so that the total cost is minimal can be viewed as a particular case of the MKP.

Comparing the definition of $\mathscr{A}_c(P_1, P_2)$ with Definition 3.2.2 (cf. (3.2.2)) of minimal distance $\hat{\mu}$ we see that

$$\mathscr{A}_c = \hat{\mu} \tag{5.1.3}$$

for any compound distance μ of the form

$$\mu(P) = \mu_c(P) = \int_{U \times U} c(x, y)P(\mathrm{d}x, \mathrm{d}y) \quad P \in \mathscr{P}_2. \tag{5.1.4}$$

(Recall that $\mathscr{P}_k$ is the set of all Borel probability measures on the Cartesian product U^k.) If $\mu(P) = \mathscr{L}_H(P) := \int H(d(x, y))P(\mathrm{d}x, \mathrm{d}y)$, $H \in \mathscr{H}$, $P \in \mathscr{P}_2$ is the H-average compound distance (see (3.3.1)), then $\mathscr{A}_c = \hat{\mathscr{L}}_H$. This example seems to be the most important for the point of view of the theory of probability metrics. For this reason we will devote special attention to the mass transportation problem with cost function $c(x, y) = H(d(x, y))$.

(III) *Kantorovich–Rubinstein–Kemperman problem of multistaged shipping.* In 1957, Kantorovich and Rubinstein studied the problem of transferring masses in cases where transits are permitted. Rather than shipping a mass from a certain subset of U to another subset of U in one step, the shipment is made in n stages. Namely, we ship $A = A_1$ to volume A_2, then A_2 to $A_2, \ldots, A_{n-1}$ to $A_n = B$. Let $\gamma_n(a_1, a_2, a_3, \ldots, a_n)$ be the measure equal to the total mass which was removed from the set a_1 and on its way to a_n passed the sets $a_2, a_3, \ldots, a_{n-1}$. If $c(x, y)$ is the unit cost of transportation from x to y then the total cost under such a transportation plan is

$$\int_{U \times U} c(x, y)\gamma_n(\mathrm{d}x \times \mathrm{d}y \times U^{n-2}) + \sum_{i=2}^{n-2} \int_{U \times U} c(x, y)\gamma_n(U^{i-1} \times \mathrm{d}x \times \mathrm{d}y \times U^{n-i-1})$$
$$+ \int_{U \times U} c(x, y)\gamma_n(U^{n-2} \times \mathrm{d}x \times \mathrm{d}y) =: \int_{U \times U} c(x, y)\Gamma_n(\mathrm{d}x \times \mathrm{d}y). \tag{5.1.5}$$

A more sophisticated plan consists of a sequence of transportation subplans γ_n, $n = 2, 3, \ldots$, due to Kemperman (1983). Each subplan γ_n need not transfer the whole mass from A to B, rather only a certain part of it. However, combined they complete the transshipment of mass, that is,

$$P_1(A) = \sum_{n=2}^{\infty} \gamma_n(A \times U^{n-1}) \tag{5.1.6}$$

and

$$P_2(B) = \sum_{n=2}^{\infty} \gamma_n(U^{n-1} \times B). \tag{5.1.7}$$

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

$$\int_{U \times U} c(x, y)Q(\mathrm{d}x, \mathrm{d}y) \tag{5.1.8}$$

where

$$Q(A \times B) = \sum_{n=2}^{\infty} \Gamma_n(A \times B) \tag{5.1.9}$$

and Γ_n is defined by (5.1.5),

$$\Gamma_n(A, B) := \gamma_n(A \times B \times U^{n-2}) + \sum_{i=2}^{n-2} \gamma_n(U^{i-1} + A \times B \times U^{n-i-1}) + \gamma_n(U^{n-2} \times A \times B).$$

Note that now Q is not necessarily a probability measure. The marginals of Q are equal to

$$Q_1(A) = \sum_{n=2}^{\infty} \left(\gamma_n(A \times U^{n-1}) + \sum_{i=1}^{n-2} \gamma_n(U^i \times A \times U^{n-i-1}) \right) \tag{5.1.10}$$

and

$$Q_2(B) = \sum_{n=2}^{\infty} \left(\gamma_n(U^{n-1} \times B) + \sum_{i=1}^{n-2} \gamma_n(U^i \times B \times U^{n-i-1}) \right) \tag{5.1.11}$$

respectively. Combining Equalities (5.1.6), (5.1.7) and (5.1.10), (5.1.11), we obtain

$$Q_1(A) - P_1(A) = Q_2(A) - P_2(A) = \sum_{n=3}^{\infty} \sum_{i=1}^{n-2} \gamma_n(U^i \times A \times U^{n-1-i}) \tag{5.1.12}$$

for any $A \in \mathscr{B}(U)$. Denote the space of all translocations of masses (without transits permitted) by $\mathscr{P}^{(P_1, P_2)}$ (cf. (5.1.2)). Under the *translocations of masses with transits permitted* we will understand the finite Borel measure Q on $\mathscr{B}(U \times U)$ such that

$$Q(A \times U) - Q(U \times A) = P_1(A) - P_2(A) \tag{5.1.13}$$

for any $A \in \mathscr{B}(U)$. Denote the space of all Q satisfying (5.1.13) by $\mathscr{Q}^{(P_1, P_2)}$. Let a continuous non-negative function $c(x, y)$ be given that represents the cost of transferring a unit mass from x to y. The total cost of transferring the given mass distributions P_1 and P_2 is given by

$$\mu_c(P) := \int_{U \times U} c(x, y) P(\mathrm{d}x, \mathrm{d}y) \quad \text{if } P \in \mathscr{P}^{(P_1, P_2)} \tag{5.1.14}$$

(cf. (5.1.2)) or

$$\mu_c(Q) := \int_{U \times U} c(x, y) Q(\mathrm{d}x, \mathrm{d}y) \quad \text{if } Q \in \mathscr{Q}^{(P_1, P_2)} \tag{5.1.15}$$

hence, if μ_c is a p. distance then the minimal distance

$$\hat{\mu}_c(P_1, P_2) = \inf\{\mu_c(P): P \in \mathscr{P}^{(P_1, P_2)}\} \tag{5.1.16}$$

may be viewed as the minimal translocation cost while the minimal norm (cf. Definition 3.2.5)

$$\mathring{\mu}_c(P_1, P_2) = \inf\{\mu_c(Q): Q \in \mathscr{Q}^{(P_1, P_2)}\} \tag{5.1.17}$$

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

The problem of calculating the exact value of $\hat{\mu}_c$ (for general c) is known as the *Kantorovich problem* and $\hat{\mu}_c$ is called the *Kantorovich functional* (see Equality (5.1.2)). Similarly, the problem of evaluating $\mathring{\mu}_c$ is known as the *Kantorovich–Rubinstein problem* and $\mathring{\mu}_c$ is said to be the *Kantorovich–Rubinstein functional.* Some authors refer to $\mathring{\mu}_c$ as the *Wasserstein norm* if $c = d$. In Example 3.2.6 we defined $\mathring{\mu}_c$ as the *minimal norm.*

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

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

(see Lemma 4.3.1, (4.3.35)) and the minimal norm $\mathring{\mu}_c$:

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

where the cost function is given by $c(x, y) = d(x, y)\max[1, d(x, a), d^{p-1}(y, a)]$, $p \geqslant 1$ (see further Theorem 5.3.3).

Open problem 5.1.1. The last equality provides a representation of the Fortet–Mourier metric in terms of the minimal norm $\mathring{\mu}_c$. It is interesting to find a similar representation but in terms of a minimal metric $\hat{\mu}$. On the real line ($U = \mathbb{R}$, $d(x, y) = |x - y|$) one can solve this problem as follows

$$\zeta(P, Q; \mathscr{G}^p) = \int_{\mathbb{R}} \max(1, |x - a|^{p-1})|(P - Q)(-\infty, x]|\,\mathrm{d}x$$
$$= \inf\left\{\int_{\mathbb{R}} (\Pr(X \leqslant t < Y) + \Pr(Y \leqslant t < X))\max(1, |t - a|^{p-1})\,\mathrm{d}t,\right.$$
$$\left. X, Y \in \mathfrak{X}(\mathbb{R})\colon \Pr_X = P, \Pr_Y = Q\right\}$$

(see further Theorems 5.4.1 and 6.5.1). Thus, in this particular case, $\zeta(P, Q; \mathscr{G}^p) = \hat{\mu}_c(P, Q)$ where the cost function c is given by

$$c(x, y) = \int \left(I\{x \leqslant t < y\} + I\{y \leqslant t < x\} \right) \max(1, |t - a|^{p-1})\, dt.$$

However, if U is a s.m.s. then the problem of determining a minimal metric $\hat{\mu}$ such that $\zeta(\cdot, \cdot; \mathscr{G}^p) = \hat{\mu}$ is still open. Note that we can define a minimal metric, namely $\mathscr{L}_p = (\hat{\mu}_d p)^{1/p}$ (see (4.3.54)) that metrizes the same topology as $\zeta(\cdot, \cdot; \mathscr{G}^p)$.

Example 5.1.1. Kantorovich functionals and the problem of classification. In the multivariate statistical analysis the problem of classification is well known (see, for example, Anderson 1984, Chapter 6). Let us give one popular example of an alternative problem of classification.

The army recruits are given a battery of tests to determine their fitness for different jobs: the scores are a set of measurements $x \in U$, where (U, d) is a s.m.s., for example, $U = \mathbb{R}^k$, $d(x, y) = \|y - x\|$. The distribution of scores is given by the measure P_1,

$$P_1(A) = \frac{\text{Number of the recruits with scores in } A}{\text{Total number of the recruits}}.$$

On the other hand, the needs of the army can be expressed by a probability measure P_2 on U which represents the desired distribution of scores for the jobs needed to be filled in. The problem is to choose an optimal classification (or assignment) of the recruits to the jobs. A classification can be specified by choosing a bounded measure Q on $\mathscr{B}(U \times U)$. If classification satisfies the balancing conditions

$$Q(A \times U) = P_1(A) \quad Q(U \times B) = P_2(B) \tag{5.1.18}$$

then we view the quantity of recruits with scores $x \in A$ which are classified as satisfying (after retraining) the needs of the jobs which require scores $y \in B$. If we think that the training procedure might be a multistaged one, in which the same individual gradually changes his scores (and fitness for different jobs respectively) in a sequence of n retraining stages then the measure Q satisfies the balancing conditions

$$Q(A \times U) - Q(U \times A) = P_1(A) - P_2(A). \tag{5.1.19}$$

The interpretation of $Q(A \times B)$ is the combined number of GIs at all stages who had scores x in A and who were trained to fit the jobs which require scores y in B. Let $c_0(x, y)$ be the cost of training the person with score x to fit the job which requires score y. Consider the joint cost $c(x, y) = c_0(x, y) + c_0(y, x)$. (Non-symmetric cost functions will be considered in Section 7.3; see

Theorem 7.3.2.) The obvious assumption on c is that

$$c(x, x) = 0. \tag{5.1.20}$$

Moreover we can assume that

$$d(x', y') \leqslant d(x'', y'') \Rightarrow c(x', y') \leqslant c(x'', y'') \tag{5.1.21}$$

i.e., the cost $c(x, y)$ increases with $d(x, y)$. In particular (5.1.21) implies

$$d(x', a) < d(x'', a) \Rightarrow c(x', a) < c(x'', a) \tag{5.1.22}$$

$$d(a, y') < d(a, y'') \Rightarrow c(a, y') < c(a, y'') \tag{5.1.23}$$

for a fixed point $a \in U$ which one can consider as the 'center' of recruit possibilities and the army's needs. The implications (5.1.21)–(5.1.23) imply that one reasonable form of c is given by

$$c(x, y) = d(x, y)\max(h(d(x, a)), h(d(y, a))) \tag{5.1.24}$$

where h is a continuous non-decreasing function on $[0, \infty)$, $h(0) \geqslant 0$, $h(x) > 0$ for $x > 0$. Another natural choice of c might be

$$c(x, y) = H(d(x, y)) \tag{5.1.25}$$

where $H \in \mathscr{H}$ (see Examples 2.2.1 and 3.3.1). Fixing the cost function c we conclude that the total cost involved in using the classification Q is calculated by the integral

$$TC(Q) = \int_{U \times U} c(x, y)Q(\mathrm{d}x, \mathrm{d}y). \tag{5.1.26}$$

The following problems therefore arise:

Problem 5.1.1. Considering the set of classifications $\mathscr{P}^{(P_1, P_2)}$ we seek to characterize the optimal $P^* \in \mathscr{P}^{(P_1, P_2)}$ (if P^* exists) for which

$$TC(P^*) = \inf\{TC(P): P \in \mathscr{P}^{(P_1, P_2)}\} \tag{5.1.27}$$

as well as to evaluate the bound

$$\hat{\mu}_c(P_1, P_2) = \inf\{TC(P): P \in \mathscr{P}^{(P_1, P_2)}\}. \tag{5.1.28}$$

Problem 5.1.2. Considering the set of classifications $\mathscr{Q}^{(P_1, P_2)}$ we seek to characterize the optimal $Q^* \in \mathscr{Q}^{(P_1, P_2)}$ (if Q^* exists) for which

$$TC(Q^*) = \inf\{TC(Q): Q \in \mathscr{Q}^{(P_1, P_2)}\} \tag{5.1.29}$$

as well as to evaluate the bound

$$\mathring{\mu}_c(P_1, P_2) = \inf\{TC(Q): Q \in \mathscr{Q}^{(P_1, P_2)}\}.$$

Problem 5.1.3. What kind of quantitative relationships exist between $\hat{\mu}_c$ and $\mathring{\mu}_c$?

In the next three sections, we shall attempt to give some answers to Problems 5.1.1 to 5.1.3.

Example 5.1.2. Kantorovich functionals and the problem of the best allocation policy. Karatzas (1984) (see also the general discussion in Whittle (1982) pp. 210–211) considers d 'medical treatments' (or 'projects' or 'investigations') with the state of the jth of them (at time $t \geqslant 0$) denoted by $x_j(t)$. At each instant of time t, it is allowed to use only one medical treatment denoted by $i(t)$, which then evolves according to some Markovian rule; meanwhile, the states of all other projects remain frozen.

Now we will consider the situation when one is allowed to use a combination of different medical treatments (say, for brevity, medicines) denoted by $M_1, \ldots, M_d$. Let $d = 2$ and (U, d) be a s.m.s. The space U may be viewed as the space of patient's parameters. Assume that for $i = 1, 2$ and for any Borel set $A \in \mathscr{B}(U)$ the exact quantity $P_i(A)$ of the medicine M_i (which should be prescribed to the patient with parameters A is known. Normalizing the total quantity $P_i(U)$ which can be prescribed by 1, we can consider P_i as a probability measure on $\mathscr{B}(U)$. Our aim is to handle an optimal policy of treatments with medicines M_1, M_2. Such a treatment should be a combination of the medicine M_1 and M_2 varying on different sets $A \subset U$.

A policy can be specified by choosing a bounded measure Q on $\mathscr{B}(U \times U)$ and the quantity of medicines M_i on the case of 'patient parameter's interval' A_i, $i = 1, 2$, by following the policy Q. The policy may satisfy the balancing condition:

$$Q(A \times U) = P_1(A) \quad Q(U \times A) = P_2(A) \quad A \in \mathscr{B}(U) \tag{5.1.30}$$

i.e., $Q \in \mathscr{P}^{(P_1, P_2)}$ or (in the case of a multistaged treatment)

$$Q(A \times U) - Q(U \times A) = P_1(A) - P_2(A) \quad A \in \mathscr{B}(U) \tag{5.1.31}$$

i.e., $Q \in \mathscr{Q}^{(P_1, P_2)}$. Let $c(x_1, x_2)$ be the cost of treating the patient with instant parameters x_i with medicines M_i, $i = 1, 2$. The $\hat{\mu}$ and $\mathring{\mu}$ (see Equation (5.1.16) and (5.1.17)) represent the minimal total costs under the balancing conditions (5.1.30) and (5.1.31) respectively. In this context Problems 5.1.1 to 5.1.3 are of interest.

(IV) *Gini's index of dissimilarity.* Already at the beginning of this century, the following question arose among probabilists: What is the proper way to measure the degree of difference between two random quantities (see the review article (Kruskal 1958))? Specific contributions to the solution of this problem, which is closely related to Kantorovich's problem 5.1.2, were made by Gini, Hoeffding, Fréchet and by their successors. In 1914 Gini introduced the concept of 'simple index of dissimilarity', which coincides with Kantorovich's metric $\mathscr{A}_d(U = \mathbb{R}^1, d(x, y) = |x - y|)$. Namely, Gini studied the functional

$$\mathscr{K}(F_1, F_2) = \inf\left\{\int_{\mathbb{R}^2} |x - y| \, \mathrm{d}F(x, y) \colon F \in \mathscr{F}(F_1, F_2)\right\} \tag{5.1.32}$$

in the space $\mathscr{F}$ of one-dimensional distribution functions (d.f.) F_1 and F_2. In (5.1.32) $\mathscr{F}(F_1, F_2)$ is the class of all bivariate d.f. F with fixed marginal distributions $F_1(x) = F(x, \infty)$ and $F_2(x) = F(\infty, x)$, $x \in \mathbb{R}^1$ (see Equation (3.3.54)). Gini and his students devoted a great deal of study to the properties of the sample measure of discrepancy, Glivenko's theorem and goodness-of-fit tests in terms of $\mathscr{K}$. Of especial importance in these investigations was the question of finding explicit expressions for this measure of discrepancy and its generalizations. Thus in 1943, Salvemini showed that

$$\mathscr{K}(F_1, F_2) = \int_{-\infty}^{\infty} |F_1(x) - F_2(x)| \, \mathrm{d}x \tag{5.1.33}$$

in the class of discrete d.f. and Dall'Aglio in 1956, extended to all of $\mathscr{F}$. This formula was proved and generalized in many ways (cf. Example 3.3.3, (3.3.19) and further Section 7.3).

(V) *Ornstein metric.* Let (U, d) be a s.m.s. and let $d_{n,a}$, $\alpha \in [0, \infty]$, be the analog of the Hamming metric on U^n (Gray, 1988, p. 48), namely

$$d_{n,\alpha}(x, y) = \frac{1}{n}\left(\sum_{i=1}^{n} d^{\alpha}(x_i, y_i)\right)^{\alpha'}, \quad x = (x_1, \ldots, x_n) \in U^n,$$

$$y = (y_1, \ldots, y_n) \in U^n, \quad 0 < \alpha < \infty, \quad \alpha' = \min(1, 1/\alpha),$$

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

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

For any Borel probability measures P and Q on U^n, define the following analog of the Kantorovich metric:

$$\mathbf{D}_{n,\alpha}(P, Q) = \inf\left\{\int_{U^{2n}} d_{n,\alpha} \, \mathrm{d}\hat{P} \colon \hat{P} \in \mathscr{P}^{(P,Q)}\right\}. \tag{5.1.34}$$

The simple p. metric $D_{n,o}$ is known among specialists in the theory of dynamical systems and coding theory as Ornstein's $\bar{d}$-metric, while $D_{n,1}$ is called the $\bar{\rho}$-distance. In information theory, the Kantorovich metric $D_{1,1}$ is known as the Wassertein (sometimes Lévy–Wassertein) metric. We shall show that

$$\mathbf{D}_{n,\alpha}(P, Q) = \sup\left\{\left|\int f\,\mathrm{d}(P-Q)\right| : f\colon U^n \to \mathbb{R}^1,\ L_{n,\alpha}(f) \leqslant 1\right\} \tag{5.1.35}$$

$$L_{n,\alpha}(f) = \sup\{|f(x) - f(y)|/d_{n,\alpha}(x, y), \quad x \neq y,\ y \in U^n\}$$

for all $\alpha \in [0, \infty)$ (see Corollary 6.1.1 for the case $0 < \alpha < \infty$ and Corollary 7.4.2 for the case $\alpha = 0$).

(VI) *Multi-dimensional Kantorovich problem.* We now generalize the preceding problems as follows.

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

$$A_c(\tilde{P}) = \inf\left\{\int_{U^N} c\,\mathrm{d}P\colon P \in \mathfrak{P}(\tilde{P})\right\} \tag{5.1.36}$$

where c is a given continuous function on U^N.

This transportation problem of infinite-dimensional linear programming is of interest in its own right in problems of stability of stochastic models (see Kalashnikov and Rachev 1988, Chapters 3 and 6). This is related to the fact that if $\{P_1^{(i)}, \ldots, P_N^{(i)}\}$, $i = 1, 2$, are two sets of probability measures on (U, d) and $P^{(i)} := P_1^{(i)} \times \cdots \times P_N^{(i)}$ are their products, then the value of the Kantorovich functional

$$\mathscr{A}_{c^*}(P^{(1)}, P^{(2)}) = \inf\left\{\int_{U^{2N}} c^*\,\mathrm{d}\hat{P}\colon \hat{P} \in \mathscr{P}^{(P^{(1)}, P^{(2)})}\right\} \tag{5.1.37}$$

with cost function c^* given by

$$c^*(x_1, \ldots, x_N, y_1, \ldots, y_N) := \phi(c_1(x_1, y_1), \ldots, c_N(x_N, y_N)) \tag{5.1.38}$$
$$x_i, y_i \in U, \quad i = 1, \ldots, N$$

where ϕ is some non-decreasing non-negative continuous function on $\mathbb{R}^N$, coincides with

$$A_{c^*}(P_1^{(1)}, \ldots, P_N^{(1)}, P_1^{(2)}, \ldots, P_N^{(2)})$$
$$= \inf\left\{\int_{U^{2N}} c^*\,\mathrm{d}P\colon P \in \mathfrak{P}(P_1^{(1)}, \ldots, P_N^{(1)}, P_1^{(2)}, \ldots, P_N^{(2)}\right\}. \tag{5.1.39}$$

See further Theorem 7.1.3.

5.2 MULTI-DIMENSIONAL KANTOROVICH THEOREM

In this section we shall prove the duality theorem for the multi-dimensional Kantorovich problem (see Equation (5.1.36)).

For brevity, $\mathscr{P}$ will denote the space $\mathscr{P}_U$ of all Borel probability measures on a s.m.s. (U, d). Let $N = 2, 3, \dots$ and let $\|\mathbf{b}\|(\mathbf{b} \in \mathbb{R}^m,\ m = \binom{N}{2})$ *be a monotone seminorm* $\|\cdot\|$, i.e., $\|\cdot\|$ is a seminorm in $\mathbb{R}^m$ with the following property: if $0 < b_i' \leqslant b_i''$, $i = 1, \dots, m$, then $\|\mathbf{b}'\| \leqslant \|\mathbf{b}''\|$. For example,

$$\|\mathbf{b}\|_p := \left(\sum_{i=1}^m |b_i|^p\right)^{1/p}, \qquad \|\mathbf{b}\|_\infty := \max\{|b_i|: i = 1, \dots, m\}$$

$$\|\mathbf{b}\| := \left|\sum_{i=1}^m b_i\right| \quad \text{and} \quad \|\mathbf{b}\| := \left(\left|\sum_{i=1}^k b_i\right|^p + \left|\sum_{i=k+1}^m b_i\right|^p\right)^{1/p},\ p \geqslant 1.$$

For any $x = (x_1, \dots, x_N) \in U^N$, let

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

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

$$A_D(\tilde{P}) := \inf\left\{\int_{U^N} D\, \mathrm{d}P: P \in \mathfrak{P}(\tilde{P})\right\} \tag{5.2.1}$$

where $D(x) := H(\mathscr{D}(x))$, $x \in U^N$ and $H \in \mathscr{H}^* = \{H \in \mathscr{H}$ (see Example 2.2.1), H convex$\}$.

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

$$\mathrm{Lip}_{1,\alpha}(U_0) := \{f: U \to \mathbb{R}^1: |f(x) - f(y)| \leqslant \alpha d(x, y) \quad \forall x, y \in U_0 \text{ and } \sup\{|f(x)|: x \in U_0\} < \infty\}.$$

Define the class

$$\mathfrak{G}(U_0) = \left\{\mathbf{f} = (f_1, \dots, f_N): \sum_{i=1}^N f_i(x_i) \leqslant D(x_1, \dots, x_N) \text{ for } x_i \in U_0,\ f_i \in \mathrm{Lip}(U_0),\ i = 1, \dots, N\right\}$$

and for any class $\mathfrak{A}$ of vectors $\mathbf{f} = (f_1, \dots, f_N)$ of measurable functions, let

$$\mathbb{K}(\tilde{P}; \mathfrak{A}) = \sup\left\{\sum_{i=1}^N \int_U f_i\, \mathrm{d}P_i: f \in \mathfrak{A}\right\} \tag{5.2.2}$$

assuming that $P_i \in \mathscr{P}^H$ and f_i is P-integrable.

Lemma 5.2.1.

$$A_D(\tilde{P}) \geqslant \mathbb{K}(\tilde{P}; \mathfrak{G}(U)). \tag{5.2.3}$$

Proof. Let $\mathfrak{f} = (f_1, \ldots, f_N) \in \mathfrak{G}(U)$ and $P \in \mathfrak{B}(\tilde{P})$ where as in (5.1.36) $\mathscr{B}(\tilde{P})$ is the set of all laws on U^N with fixed projections P_i on the ith coordinates, $i = 1, \ldots, N$. Then

$$\sum_{i=1}^{N} \int_U f_i(x_i) P_i(\mathrm{d}x_i) = \int_{U^N} \sum_{i=1}^{N} f_i(x_i) P(\mathrm{d}x_1, \ldots, \mathrm{d}x_N)$$
$$\leqslant \int_{U^N} D \, \mathrm{d}P.$$

The last inequality together with (5.2.1) and (5.2.2) completes the proof (5.2.3).
QED

The next theorem (an extension of Kantorovich's (1940) theorem to the multi-dimensional case) shows that exact equality holds in (5.2.3).

Theorem 5.2.1. For any s.m.s. (U, d) *and for any set* $\tilde{P} = (P_1, \ldots, P_N)$, $P_i \in \mathscr{P}^H$, $i = 1, \ldots, N$,

$$A_D(\tilde{P}) = \mathbb{K}(\tilde{P}; \mathfrak{G}(U)). \tag{5.2.4}$$

If the set $\tilde{P}$ *consists of tight measures, then the infimum is attained in* (5.2.1).

Proof. (I) Suppose first that d is a bounded metric in U and let

$$\rho_i(x_i, y_i) = \sup\{|D(x_1, \ldots, x_N) - D(y_1, \ldots, y_N)| : x_j = y_j \in U, j = 1, \ldots, N, j \neq i\} \tag{5.2.5}$$

for $x_i, y_i \in U$, $i = 1, \ldots, N$. Since H is a convex function and d is bounded, $\rho_1, \ldots, \rho_N$ are bounded metrics. Let $U_0 \subseteq U$ and let $\mathfrak{G}'(U_0)$ be the space of all collections $\mathfrak{f} = (f_1, \ldots, f_N)$ of measurable functions on U_0 such that $f_1(x_1) + \cdots + f_N(x_N) \leqslant D(x_1, \ldots, x_N)$, $x_1, \ldots, x_N \in U_0$. Let $\mathfrak{G}''(U_0)$ be a subset of $\mathfrak{G}'(U_0)$ of vectors $\mathfrak{f}$ for which $|f_i(x) - f_i(y)| \leqslant \rho_i(x, y)$, $x, y \in U_0$, $i = 1, \ldots, N$. Observe that $\mathfrak{G}'' \subset \mathfrak{G} \subset \mathfrak{G}'$. We wish to show that if $P_i(U_0) = 1$, $i = 1, \ldots, N$, then

$$\mathbb{K}(\tilde{P}; \mathfrak{G}'(U_0)) = \mathbb{K}(\tilde{P}; \mathfrak{G}''(U)). \tag{5.2.6}$$

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

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

Since D is continuous, it follows that f_j^* are upper semi-continuous, hence Borel measurable. Also, $f_1^*(x_1)+\cdots+f_N^*(x_N)\leqslant D(x_1,\ldots,x_N)\quad \forall x_1,\ldots,x_N\in U$. Furthermore, for any $x_1, y_1\in U$

$$\begin{aligned}f_1^*(x_1)-f_1^*(y_1)&=\inf\{D(x_1,\ldots,x_N)-f_2(x_2)-\cdots-f_N(x_N): x_2,\ldots,x_N\in U_0\}\\&\quad+\sup\{f_2(y_2)+\cdots+f_N(y_N)-D(y_1,\ldots,y_N): y_2,\ldots,y_N\in U_0\}\\&\leqslant \sup\{D(x_1,y_2,\ldots,y_N)-D(y_1,\ldots,y_N): y_2,\ldots,y_N\in U_0\}\\&\leqslant \rho_1(x_1,y_1).\end{aligned}$$

A similar argument proves that the collection $\mathbf{f}^*=(f_1^*,\ldots,f_N^*)$ belongs to the set $\mathfrak{G}''(U)$. Given $x_1\in U_0$, we have $f(x_1)\leqslant D(x_1,x_2,\ldots,x_N)-f_2(x_2)-\cdots-f_N(x_N)$ for all $x_2,\ldots,x_N\in U_0$. Thus, $f(x_1)\leqslant f^*(x_1)$. Also, if $x_2\in U_0$, then

$$\begin{aligned}f_2^*(x_2)=&\inf_{x_1\in U,\,x_3,\ldots,x_N\in U_0}\Big\{D(x_1,\ldots,x_N)\\&-\inf_{y_2,\ldots,y_N\in U_0}[D(x_1,y_2,\ldots,y_N)-f_2(y_2)-\cdots-f_N(y_N)]\\&-f_3(x_3)-\cdots-f_N(x_N)\Big\}\\\geqslant&\inf_{x_1\in U,\,x_3,\ldots,x_N\in U_0}\{D(x_1,\ldots,x_N)-D(x_1,x_2,\ldots,x_N)+f_2(x_2)\\&+\cdots+f_N(x_N)-f_3(x_3)-\cdots-f_N(x_N)\}\\=&f_2(x_2).\end{aligned}$$

Similarly, $f_i^*(x_j)\geqslant f_i(x_i)$ for all $i=1,\ldots,N$ and $x_i\in U_0$. Hence,

$$\sum_{i=1}^N\int f_i\,\mathrm{d}P_i\leqslant\sum_{i=1}^N\int f_i^*\,\mathrm{d}P_i$$

which implies the inequality

$$\mathbb{K}(\tilde{P};\mathfrak{G}'(U_0))\leqslant\mathbb{K}(\tilde{P};\mathfrak{G}''(U))\tag{5.2.7}$$

from which (5.2.6) clearly follows.

Case 1. Let U be a finite space with the elements $u_1,\ldots,u_n$. From the duality principle in linear programming, we have (see, for example, Bazaraa and Jarvis, 1977, Chapter 6)

$$\begin{aligned}A_D(\tilde{P})=\inf\Bigg\{&\sum_{i_1=1}^n\cdots\sum_{i_N=1}^n D(u_{i_1},\ldots,u_{i_N})\pi(i_1,\ldots,i_N):\\&\pi(i_1,\ldots,i_N)\geqslant 0,\ \sum_{i_j:j\neq k}\pi(i_1,\ldots,i_N)=P_k(u_{i_k}),\ k=1,\ldots,N\Bigg\}\\=\sup\Bigg\{&\sum_{i=1}^n\sum_{j=1}^N f_j(u_i)P_j(u_i):\ \sum_{j=1}^N f_j(\tilde{u}_j)\leqslant D(\tilde{u}_1,\ldots,\tilde{u}_N),\ \tilde{u}_1,\ldots,\tilde{u}_N\in U\Bigg\}\\=\ &\mathbb{K}(\tilde{P};\mathfrak{G}'(U)).\end{aligned}$$

Therefore (5.2.7) implies the chain of inequalities

$$\mathbb{K}(\tilde{P}; \mathfrak{G}(U)) \geqslant \mathbb{K}(\tilde{P}; \mathfrak{G}''(\mathfrak{U})) \geqslant \mathbb{K}(\tilde{P}; \mathfrak{G}'(U)) \geqslant A_D(\tilde{P})$$

from which (5.2.4) follows by virtue of (5.2.3).

Case 2. Let U be a compact set. For any $n = 1, 2, \dots$, choose disjoint non-empty Borel sets $A_1, \dots, A_{m_n}$ of diameter less than $1/n$ whose union is U. Define a mapping $h_n: U \to U_n = \{u_1, \dots, u_{m_n}\}$ such that $h_n(A_i) = u_i$, $i = 1, \dots, m_n$. According to (5.2.6) we have for the collection $\tilde{P}_n = (P_1 \circ h_n^{-1}, \dots, P_N \circ h_n^{-1})$ the relation

$$\mathbb{K}(\tilde{P}_n; \mathfrak{G}'(U_n)) = \sup\left\{ \sum_{i=1}^{N} \int_U f_i(h_n(u)) P_i(\mathrm{d}u) : \mathfrak{f} \in \mathfrak{G}'(U) \right\}. \tag{5.2.8}$$

If $\mathfrak{f} \in \mathfrak{G}'(U_n)$, then $\sum_{i=1}^{N} f_i(h_n(\tilde{u}_i)) \leqslant D(h_n(\tilde{u}_1), \dots, h_n(\tilde{u}_N)) \leqslant D(\tilde{u}_1, \dots, \tilde{u}_N) + K/n$, where the constant K is independent of n and $\tilde{u}_1, \dots, \tilde{u}_N \in U$. Hence, from (5.2.8) we have

$$\mathbb{K}(\tilde{P}_n; \mathfrak{G}'(U_n)) \leqslant \mathbb{K}(\tilde{P}; \mathfrak{G}'(U)) + K/n. \tag{5.2.9}$$

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

$$\int_{U^N} D \, \mathrm{d}P^{(n)} = \mathbb{K}(\tilde{P}_n; \mathfrak{G}'(U_n)). \tag{5.2.10}$$

Since $P_i \circ h_n^{-1}$ converges weakly to P_i, $i = 1, \dots, N$, the sequence $\{P^{(n)}, n = 1, 2, \dots\}$ is weakly compact (Billingsley, 1968, Section 6, App. III). Let P^* be a weak limit of it. From the estimate (5.2.9) and Equality (5.2.10) it follows that

$$\int_{U^N} D \, \mathrm{d}P^* \leqslant \mathbb{K}(\tilde{P}; \mathfrak{G}'(U))$$

which together with Lemma 5.2.1 implies (5.2.4).

Case 3. Let (U, d) be a bounded s.m.s. Since $\int_U H(d(x, a)) P_i(\mathrm{d}x) < \infty$, the convexity of H and (5.2.5) imply that $\int_U \rho_i(x, a)\, P_i(\mathrm{d}x) < \infty$, $i = 1, \dots, N$. Let the P_i be tight measures (see Definition 2.4.1). Then for each $n = 1, 2, \dots$ there exists a compact set K_n such that

$$\sup_{1 \leqslant i \leqslant N} \int_{U \setminus K_n} (1 + \rho_i(x, a)) P_i(\mathrm{d}x) < \frac{1}{n}. \tag{5.2.11}$$

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

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

where

$$\delta_\alpha(A) := \begin{cases} 1 & \alpha \in A \\ 0 & \alpha \notin A. \end{cases}$$

By Equation (5.2.6),

$$\begin{aligned} \mathbb{K}(\tilde{P}_n; \mathfrak{G}'(K_n \cup \{a\})) &= \mathbb{K}(\tilde{P}_n; \mathfrak{G}''(U)) \\ &\leqslant \sup\left\{ \sum_{i=1}^{N} \int_U f_i(x) P_i(\mathrm{d}x) + \int_{U \setminus K_n} \rho_i(x, a) P_i(\mathrm{d}x) : \mathbf{f} \in \mathfrak{G}(U) \right\} \\ &\leqslant \mathbb{K}(\tilde{P}; \mathfrak{G}(U)) + N/n. \end{aligned} \tag{5.2.12}$$

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

$$\int_{U^N} D \, \mathrm{d}P^{(n)} \leqslant \mathbb{K}(\tilde{P}_n; \mathfrak{G}'(K_n \cup \{a\})). \tag{5.2.13}$$

Similarly to Case 2 we then obtain (5.2.4) from relations (5.2.12) and (5.2.13).

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

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

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

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

where the constant $c_{\mathbf{m}}$ is chosen so that

$$\mu_{\mathbf{m}}(A_{m_1} \times \cdots \times A_{m_N}) = \bar{\mu}(A_{m_1} \times \cdots \times A_{m_N}).$$

Let $\mu_\varepsilon = \Sigma_{\mathbf{m}} \mu_{\mathbf{m}}$. Then for any $B \in \mathscr{B}(U)$

$$\begin{aligned} \mu_\varepsilon(B \times U^{N-1}) &= \sum_{\mathbf{m}} c_{\mathbf{m}} P_{1,m_1}(B) P_{2,m_2}(U) \cdots P_{N,m_N}(U) \\ &= \sum_{\mathbf{m}} c_{\mathbf{m}} P_1(B \cap A_{m_1}) P_2(A_{m_2}) \cdots P_N(A_{m_N}) \\ &= \sum_{\mathbf{m}}{}' \frac{\bar{\mu}(\bar{A}_{m_1} \times \cdots \times \bar{A}_{m_N})}{P_{1,m_1}(A_{m_1}) \ldots P_{N,m_N}(A_{m_N})} \times P_1(B \cap A_{m_1}) P_2(A_{m_2}) \ldots P_N(A_{m_N}) \end{aligned}$$

where $\sum'_{\mathbf{m}}$ indicates summation over all $\mathbf{m}$ such that $P_{j,m_j}(A_{m_j}) > 0$ for all $j = 1, \ldots, m_N$. Note that if $P_{1,m_1}(A_{m_1}) > 0$, we have

$$\sum_{m_2,\ldots,m_N} \frac{\bar{\mu}(\bar{A}_{m_1} \times \cdots \times \bar{A}_{m_N})}{P_{1,m_1}(A_{m_1})} = \bar{\mu}(\bar{A}_{m_1} \times U^{N-1})/P_1,m_1(A_{m_1})$$

$$= \bar{P}_1(\bar{A}_{m_1})/P_{1,m_1}(A_{m_1}) = 1.$$

This, together with analogous calculations for $\mu_\varepsilon(U^k \times B \times U^{N-k-1})$, $k = 1, 2, \ldots, N-1$, shows that $\mu_\varepsilon \in \mathfrak{P}(\tilde{P})$, hence, to each positive ε,

$$\begin{aligned}
&\mu_\varepsilon(\mathscr{D}(y_1, \ldots, y_N) > \alpha + 2\varepsilon\|\mathbf{e}\|) \\
&\quad \leqslant \sum \{\mu_m(A_{m_1} \times \cdots \times A_{m_n}) \colon \mathscr{D}(x_1, \ldots, x_N) > \alpha + \varepsilon\|\mathbf{e}\|\} \\
&\quad \leqslant \bar{\mu}(\mathscr{D}(y_1, \ldots, y_N) > \alpha)
\end{aligned}$$

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

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

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

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

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

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

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

$$\int_{U^N} D_n \, \mathrm{d}P^{(n)} = \mathbb{K}(\tilde{P}; \mathfrak{G}(U, d_n)). \tag{5.2.14}$$

Since $P^{(n)}$, $n = 1, 2, \ldots$, is a uniformly tight sequence, passing on to a subsequence if necessary, we may assume that $P^{(n)}$ converges weakly to $P^{(0)} \in \mathfrak{P}(\tilde{P})$. By Skorokhod–Dudley's theorem (see Theorem 11.7.1, Dudley 1989), there exist a probability space (Ω, μ) and a sequence $\{X_k, k = 0, 1, \ldots\}$ of N-dimensional random vectors defined on (Ω, μ) and assuming values on U^N. Moreover, for any $k = 0, 1, \ldots$, the vector X_k has distribution $P^{(k)}$ and the sequence $X_1, X_2, \ldots$ converges μ-almost everywhere to X_0. According to (5.2.14) and the Fatou lemma

$$\liminf_{n\to\infty} \mathbb{K}(\tilde{P}; \mathfrak{G}(U, d_n)) = \liminf_{n\to\infty} \mathbb{E}_\mu D_n(X_n) \geqslant \mathbb{E}_\mu \liminf_{n\to\infty} D_n(X_n)$$

$$\geqslant \mathbb{E}_\mu D(X_0) - \mathbb{E}_\mu \limsup_{n\to\infty} |D_n(X_n) - D(X_0)|$$

where

$$|D_n(X_n) - D(X_0)| \leqslant |D_n(X_n) - D_n(X_0)| + |D_n(X_0) - D(X_0)| \to 0 \ \mu\text{-a.e. as } n \to \infty$$

and

$$\mathbb{E}_\mu \limsup_{n\to\infty} (D_n(X_n) + D(X_0)) \leqslant \text{const} \times \sum_{i=1}^{N} \int_U H(d(x, a)) P_i(\mathrm{d}x) < \infty.$$

Hence

$$\mathbb{K}(\tilde{P}; \mathfrak{G}(U)) \geqslant \lim_{k\to\infty} \mathbb{K}(\tilde{P}; \mathfrak{G}(U, d_k)) \geqslant A_D(\tilde{P})$$

which by virtue of (5.2.3) implies (5.2.4). If $P_1, \ldots, P_N$ are not necessarily tight, one can use arguments similar to those in Case 3 of Part I and prove (5.2.4), which completes the proof of the theorem. QED

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

$$\mathfrak{K}_H(\tilde{P}, \tilde{Q}) = \inf\left\{\int_{\mathfrak{U}\times\mathfrak{U}} \Delta(\mathbf{x}, \mathbf{y}) P(\mathrm{d}\mathbf{x}, \mathrm{d}\mathbf{y}) : P \in \mathfrak{P}(\tilde{P}, \tilde{Q})\right\} \tag{5.2.15}$$

where $\mathfrak{P}(\tilde{P}, \tilde{Q})$ is the space of all probability measures on $\mathfrak{A} \times \mathfrak{A}$ with fixed one-dimensional marginal distributions $P_1, \ldots, P_n, Q_1, \ldots, Q_n$. Further (see Chapter 7) we shall consider more examples of minimal functionals of the type (5.2.15) (the so-called K-minimal metrics).

Case $N = 2$. Dual representation of the Kantorovich functional $\mathscr{A}_c(P_1, P_2)$. $\mathscr{L}_H = \ell_H$. Let $\mathfrak{C}$ be the class of all functions $c(x, y) = H(d(x, y))$, $x, y \in U$, where the function H belongs to the class $\mathscr{H}$ of all non-decreasing continuous functions on $[0, \infty)$ for which $H(0) = 0$ and which satisfy Orlicz' condition

$$K_H = \sup\{H(2t)/H(t): t > 0\} < \infty. \tag{5.2.16}$$

We also recall that $\mathscr{H}^*$ is the subset of all convex functions in $\mathscr{H}$ and let $\mathfrak{C}^*$ be the set of all $c(x, y) = H(d(x, y))$, $H \in \mathscr{H}^*$.

Corollary 5.2.1. Let (U, d) be a s.m.s. and P_1, P_2 be Borel probability measures on U. Let $c \in \mathfrak{C}^*$ and $\mathscr{A}_c(P_1, P_2)$ be given by (5.1.2). Let $\mathrm{Lip}_{1,\alpha}(U) := \{f: U \to \mathbb{R}: |f(x) - f(y)| \leqslant \alpha d(x, y),\ x, y \in U)$,

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

and

$$\mathscr{B}_c(P_1, P_2) = \sup\left\{\int_U f\,\mathrm{d}P_1 + \int_U g\,\mathrm{d}P_2: (f, g) \in \mathrm{Lip}^c(U)\right\}.$$

If $\int_U c(x, a)(P_1 + P_2)(\mathrm{d}x) < \infty$ for some $a \in U$, then

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

Moreover, if P_1 and P_2 are tight measures, then there exists an optimal measure $P^* \in \mathscr{P}^{(P_1, P_2)}$ for which the infimum in (5.1.2) is attained.

The corollary implies that if $\mathfrak{A}$ is a class of pairs (f, g) of P_1- (resp. P_2-) integrable functions satisfying $f(x) + g(y) \leqslant c(x, y)$ for all $x, y \in U$ and $\mathfrak{A} \supset [\mathrm{Lip}^c(U)]^{\times 2}$, then the Kantorovich functional (5.1.2) admits the following dual representation

$$\mathscr{A}_c(P_1, P_2) = \sup\left\{\int_U f\,\mathrm{d}P_1 + \int_U g\,\mathrm{d}P_2: (f, g) \in \mathfrak{A}\right\}.$$

The equality $\mathscr{A}_c = \mathscr{B}_c$ furnishes the main relationship between the *H-average distance* $\mathscr{L}_H(X, Y) = \mathbb{E}H(d(X, Y))$ (3.3.1), (resp. *p-average metric* $\mathscr{L}_p(X, Y) = [\mathbb{E}d^p(X, Y)]^{1/p}$, $p \in (1, \infty)$, (3.3.3)) and the *Kantorovich distance* ℓ_H (resp. *ℓ_p-metric*, see (3.2.11)).

Corollary 5.2.2 (i) If (U, d) is a s.m.s., $H \in \mathscr{H}^*$ and

$$P_1, P_2 \in \mathscr{P}^H(U) := \left\{ P \in \mathscr{P}(U) \colon \int_U H(d(x, a))P(\mathrm{d}x) < \infty \right\}$$

then

$$\ell_H(P_1, P_2) = \hat{\mathscr{L}}_H(P_1, P_2) := \inf\{\mathscr{L}_H(X_1, X_2) \colon X_i \in \mathfrak{X}(U), \Pr_{X_i} = P_i, i = 1, 2\}. \tag{5.2.17}$$

More, if U is u.m. s.m.s., then ℓ_H is a simple distance in $\mathscr{P}^H(U)$ with parameter $\mathbb{K}_{\ell_H} = K_H$, i.e., for any P_1, P_2 and $P_3 \in \mathscr{P}^H(U)$, $\ell_H(P_1, P_2) \leqslant K_H(\ell_H(P_1, P_3) + \ell_H(P_3, P_2))$. In this case the infimum in (5.2.17) is attained.

(ii) If $1 < p < \infty$, (U, d) is a s.m.s. and

$$P_1, P_2 \in \mathscr{P}^{(p)}(U) := \left\{ P \in \mathscr{P}(U) \colon \int_U d^p(x, a)P(\mathrm{d}x) < \infty \right\}$$

then

$$\ell_p(P_1, P_2) = \hat{\mathscr{L}}_p(P_1, P_2). \tag{5.2.18}$$

In the space $\mathscr{P}^p(U)$, ℓ_p is a simple metric, provided U is u.m.s. m.s..

Proof. See Theorem 3.2.1, Corollary 5.2.1, and Remark 2.5.1. QED

5.3 DUAL REPRESENTATION OF THE MINIMAL NORMS $\mathring{\mu}_c$; A GENERALIZATION OF THE KANTOROVICH–RUBINSTEIN THEOREM

The Kantorovich–Rubinstein duality theorem has a long and colorful history, originating in the 1958 work of Kantorovich and Rubinstein on the mass transport problem. For a detailed survey, we refer the reader to the article of Kemperman (1983). Given probabilities P_1 and P_2 on a space U and a measurable cost function $c(x, y)$ on $U \times U$, satisfying some integrability conditions, let us consider the *Kantorovich–Rubinstein functional*

$$\mathring{\mu}_c(P_1, P_2) := \inf \int c(x, y)\, \mathrm{d}b(x, y) \tag{5.3.1}$$

where the infimum is over all finite measures b on $U \times U$ with marginal difference $b_1 - b_2 = P_1 - P_2$, where $b_i = T_i b$ is the ith projection of b (cf. (5.1.17)). ($\mathring{\mu}_c$ is sometimes called the Wasserstein functional; in Example 3.2.6 we defined $\mathring{\mu}_c$ as a *minimal norm.*)

Duality theorem for $\mathring{\mu}_c$ is of the general form

$$\mathring{\mu}_c(P_1, P_2) = \sup \int_U f \,\mathrm{d}(P_1 - P_2) \tag{5.3.2}$$

with the supremum taken over a class of $f: U \to \mathbb{R}$ satisfying the 'Lipschitz' condition $f(x) - f(y) \leqslant c(x, y)$. When the probabilities in question have a finite support, this becomes a dual representation of the minimal cost in a network flow problem (see Chapter 9, Bazaraa and Jarvis 1977 and Section 9.8, Berge and Chouila-Houri 1965).

Results for (5.3.2) were obtained by Kantorovich and Rubinstein (1958) with cost function $c(x, y) = d(x, y)$ where (U, d) is a compact metric space. Levin and Milyutin (1979) proved the dual relation (5.3.2) for U a compact space and for $c(x, y)$ an arbitrary continuous cost function. Dudley (1976) (Theorem 20.1) proved (5.3.2) for s.m.s. U and $c = d$. Following the proofs of Kantorovich and Rubinstein (1958) and Dudley (1976), we shall show (5.3.2) for cost functions $c(x, y)$ which are not necessarily metrics. The supremum in (5.3.2) is shown to be attained for some optimal function f.

Let (U, d) be a separable metric space. Suppose that $c: U \times U \to [0, \infty)$ and $\lambda: U \to [0, \infty)$ are measurable functions such that

(C1) $c(x, y) = 0$ iff $x = y$;
(C2) $c(x, y) = c(y, x)$ for x, y in U;
(C3) $c(x, y) \leqslant \lambda(x) + \lambda(y)$ for $x, y \in U$;
(C4) λ maps bounded sets to bounded sets;
(C5) $\sup\{c(x, y): x, y \in B(a; R), d(x, y) \leqslant \delta\}$ tends to 0 as $\delta \to 0$ for each $a \in U$ and $R > 0$. Here, $B(a; R) := \{x \in U: d(x, a) < R\}$.

We give two examples of function c satisfying (C1)–(C5) which are related to our discussion in Section 5.1 (cf. Examples 5.1.1 and 5.1.2):

(i) $c(x, y) = H(d(x, y))$, $H \in \mathscr{H}$ (see (5.2.16);
(ii) $c(x, y) = d(x, y)\max(1, h(d(x, a)), h(d(y, a)))$, where $h: [0, \infty) \to [0, \infty)$ is a continuous non-decreasing function.

Given a real-valued function $f: U \to \mathbb{R}$, we define

$$\|f\|_c := \sup\{|f(x) - f(y)|/c(x, y): x \neq y\} \tag{5.3.3}$$

and set

$$\mathbb{L} := \{f: \|f\|_c < +\infty\}. \tag{5.3.4}$$

It is easy to see that $\|\cdot\|_c$ is a seminorm on the linear space $\mathbb{L}$. Notice that for $f \in \mathbb{L}$ we have $|f(x) - f(y)| \leqslant \|f\|_c c(x, y)$ $\forall x, y \in U$. It follows from Condition (C5) on c that each function in L is continuous and hence measurable. Note also that $\|f\|_c = 0$ if and only if f is constant. Define $\mathbb{L}_0$ to be the quotient of $\mathbb{L}$ modulo the constant functions. Then $\|\cdot\|_c$ is naturally defined on $\mathbb{L}_0$, and $(\mathbb{L}_0, \|\cdot\|_c)$ is a normed linear space (Fortet and Mourier 1953).

Now suppose that $\mathscr{M} = \mathscr{M}_\lambda(U)$ denotes the linear space of all finite signed measures m on U such that

$$m(U) = 0 \quad \text{and} \quad \int \lambda \, \mathrm{d}|m| \leqslant \infty. \tag{5.3.5}$$

Here $|m| := m^+ + m^-$, where $m = m^+ - m^-$ is the Jordan decomposition of m.

For each $m \in \mathcal{M}$, let $\mathbb{B}(m)$ be the set of all finite measures b on $U \times U$ such that

$$b(A \times U) - b(U \times A) = m(A) \tag{5.3.6}$$

for each Borel $A \subseteq U$. Note that $\mathbb{B}(m)$ is always non-empty, since it contains $(m^+ \times m^-)/m^+(U)$. Here, $m^+ \times m^-$ denotes the product measure $m^+ \times m^-(A) = m^+(A)m^-(A)$. $A \in \mathscr{B}(U)$. Define a function $m \to \|m\|_w$ on $\mathcal{M}$ by

$$\|m\|_w := \inf\left\{\int c(x, y)b(\mathrm{d}x, \mathrm{d}y) \colon b \in \mathbb{B}(m)\right\}. \tag{5.3.7}$$

We have

$$\begin{aligned} \|m\|_w &\leqslant \int c(x, y)(m^+ \times m^-)(\mathrm{d}x, \mathrm{d}y)/m^+(U) \\ &\leqslant \int \lambda(x) m^+(\mathrm{d}x) + \int \lambda(y) m^-(\mathrm{d}y) \\ &= \int \lambda \, \mathrm{d}|m| < \infty. \end{aligned} \tag{5.3.8}$$

For $c(x, y) = d(x, y)$, $\|m\|_w$ is sometimes called the *Kantorovich-Rubinstein* or *Wasserstein norm* of m (see also (3.2.38) and Definition 3.2.5).

We shall demonstrate that for probabilities P and Q on U with $P - Q \in \mathcal{M}$, we have

$$\|P - Q\|_w = \sup\left\{\left|\int f \, \mathrm{d}(P - Q)\right| \colon \|f\|_c \leqslant 1\right\}. \tag{5.3.9}$$

which furnished (5.3.2) with cost function c satisfying (C1) to (C5). When $c(x, y) = d(x, y)$ and $\lambda(x) = d(x, a)$, a some fixed point of U, this is a straightforward generalization of the classical Kantorovich–Rubinstein duality theorem (see Dudley 1976, Lecture 20).

First note that $\|\cdot\|_w$ is a semi-norm on $\mathcal{M}$ (see Lemma 3.2.2). Now given $m \in \mathcal{M}, f \in \mathbb{L}$, and a fixed $a \in U$, we have

$$\begin{aligned} |f(x) &\leqslant |f(x) - f(a)| + |f(a)| \leqslant \|f\|_c c(x, a) + |f(a)| \\ &\leqslant \|f\|_c(\lambda(x) + \lambda(a)) + |f(a)| = K_1\lambda(x) + K_2 \quad \forall x \in U \end{aligned}$$

for constants $K_1, K_2 \geqslant 0$. Thus, each $f \in \mathbb{L}$ is $|m|$-integrable and induces a linear form $\phi_f\colon \mathcal{M} \to \mathbb{R}$ defined by

$$\phi_f(m) := \int f \, \mathrm{d}m. \tag{5.3.10}$$

Note that if f and g differ by a constant, then $\phi_f = \phi_g$. Given $b \in \mathbb{B}(m)$, we have

$$\begin{aligned}|\phi_f(m)| &= \left|\int f \, \mathrm{d}m\right| = \left|\int (f(x) - f(y)) b(\mathrm{d}x, \mathrm{d}y)\right| \\ &\leqslant \int |f(x) - f(y)| b(\mathrm{d}x, \mathrm{d}y) \leqslant \|f\|_c \int c(x, y) b(\mathrm{d}x, \mathrm{d}y).\end{aligned}$$

Taking the infimum over all $b \in \mathbb{B}(m)$, this yields $|\phi_f(m)| \leqslant \|f\|_c \|m\|_w$, so that ϕ_f is a continuous linear functional with dual norm $\|\phi_f\|_w^*$ such that

$$\|\phi_f\|_w^* \leqslant \|f\|_c. \tag{5.3.11}$$

Thus, we may define a continuous linear transformation

$$(\mathbb{L}_0, \|\cdot\|_c) \xrightarrow{D} (\mathcal{M}^*, \|\cdot\|_w^*) \tag{5.3.12}$$

by $D(f) = \phi_f$.

Lemma 5.3.1. The map D is an isometry, i.e. $\|f\|_c = \|\phi_f\|_w^*$.

Proof. Given $x \in U$ denote the point mass at x by δ_x. Note first that if $m_{xy} := \delta_x - \delta_y$ for some $x, y \in U$, then

$$\|m_{xy}\|_w \leqslant \int c(u, t)(\delta_x \times \delta_y)(\mathrm{d}u, \mathrm{d}t) = c(x, y).$$

Then for each $f \in \mathbb{L}$,

$$\begin{aligned}\|f\|_c &= \sup\{|f(x) - f(y)|/c(x, y) \colon x \neq y\} = \sup\{|\phi_f(m_{xy})|/c(x, y) \colon x \neq y\} \\ &\leqslant \|\phi_f\|_w^* \sup\{\|m_{xy}\|_w/c(x, y) \colon x \neq y\} \leqslant \|\phi_f\|_w^*\end{aligned}$$

so that $\|f\|_c = \|\phi_f\|_w^*$ by (5.3.11). QED

We now set about proving that the map D is subjective and hence an isometric isomorphism of Banach spaces. Recall that an isometric isomorphism between two normed linear spaces $\mathbb{A}_1$ and $\mathbb{A}_2$ is a one-to-one continuous linear map $T\colon \mathbb{A}_1 \to \mathbb{A}_2$ with $T\mathbb{A}_1 = \mathbb{A}_2$ and $\|Tx\|_{\mathbb{A}_2} = \|x\|_{\mathbb{A}_1}$ (see Dunford and Schwartz 1988, p. 65).

We need some preliminary facts. Let $\mathcal{M}_0$ be the set of signed measures of the form $m = m_1 - m_2$, where m_1 and m_2 are finite measures on U with bounded support such that $m_1(U) = m_2(U)$. Condition (C4) on λ implies that $\mathcal{M}_0 \subseteq \mathcal{M}$.

Lemma 5.3.2. $\mathcal{M}_0$ is a dense subspace of $(\mathcal{M}, \|\cdot\|_w)$.

Proof. Given $m \in \mathcal{M}(m \neq 0)$, fix $a \in U$ and put

$$B_n = B(a, n) := \{x \in U: d(x, a) < n\}$$

for $n = 1, 2, \ldots$. For all sufficiently large n, we have $m^+(B_n)m^-(B_n) > 0$. For such n, let us denote

$$m_n(A) := m^+(U)\left[\frac{m^+(A \cap B_n)}{m^+(B_n)} - \frac{m^-(A \cap B_n)}{m^-(B_n)}\right]$$

$$\delta_n := \frac{m^-(U)}{m^-(B_n)} - 1 \quad \varepsilon_n := \frac{m^+(U)}{m^+(B_n)} - 1.$$

Then δ_n, $\varepsilon_n \to 0$ as $n \to \infty$. Also

$$(m - m_n)(A) = m(A \backslash B_n) - \varepsilon_n m^+(A \cap B_n) + \delta_n m^-(A \cap B_n)$$

Define finite measures μ_n and ν_n on U by

$$\mu_n(A) := m^+(A \backslash B_n) + \delta_n m^-(A \cap B_n)$$
$$\nu_n(A) := m^-(A \backslash B_n) + \varepsilon_n m^+(A \cap B_n).$$

Then $m - m_n = \mu_n - \nu_n$. Moreover, μ_n and ν_n are absolutely continuous with respect to $|m|$. Letting P, N be the supports of m^+ and m^- in the Jordan–Hahn decomposition for m, we determine the Radon–Nikodym derivatives

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}|m|}(x) = \begin{cases} 1 & x \in P \backslash B_n \\ \delta_n & x \in N \cap B_n \\ 0 & \text{otherwise} \end{cases} \quad \frac{\mathrm{d}\nu_n}{\mathrm{d}|m|}(y) = \begin{cases} 1 & y \in N \backslash B_n \\ \varepsilon_n & y \in P \cap B_n \\ 0 & \text{otherwise.} \end{cases}$$

Then the measure $b_n = (\mu_n \times \nu_n)/\mu_n(U)$ belongs to $\mathbb{B}(m - m_n)$. Noting that

$$\begin{aligned} \nu_n(U) = \mu_n(U) &= m^+(U \backslash B_n) + \delta_n m^-(B_n) \\ &= m^+(U \backslash B_n) + (m^-(U) - m^-(B_n)) \\ &= |m|(U \backslash B_n) = |m|(B_n^c) \end{aligned}$$

we write the Radon–Nikodym derivative

$$f_n(x, y) := \frac{\mathrm{d}b_n}{\mathrm{d}(|m| \times |m|)}(x, y) := \frac{1}{|m|(B_n^c)} \frac{\mathrm{d}\mu_n}{\mathrm{d}|m|}(x) \frac{\mathrm{d}\nu_n}{\mathrm{d}|m|}(y).$$

Then we claim the following.

Claim. The function $g(x, y) = \sup_n f_n(x, y)c(x, y)$ is $|m| \times |m|$-integrable.

Proof of Claim. We show that g is integrable over various subsets of $U \times U$.

(i) g is integrable over $P \times N$: We suppose that $x \in P$ and $y \in N$. Then

$$g(x, y) \leqslant \sum_{n=1}^{\infty} \frac{c(x, y)}{|m|(B_n^c)} I_{C_n}(x, y)$$

where $C_n = (B_n^c \times B_n^c) - (B_{n+1}^c \times B_{n+1}^c)$ and $I_{(\cdot)}$ is the indicator of $(\cdot)$. So

$$\begin{aligned}\int_{P \times N} g \, \mathrm{d}|m| \times |m| &\leqslant \sum_{n=1}^{\infty} \frac{1}{|m|(B_n^c)} \int_{C_n} (\lambda(x) + \lambda(y))|m| \times |m|(\mathrm{d}x, \mathrm{d}y) \\ &\leqslant \sum_{n=1}^{\infty} \frac{2}{|m|(B_n^c)} \int_{(B_n^c - B_{n+1}^c) \times B_n^c} \lambda(x)|m| \times |m|(\mathrm{d}x, \mathrm{d}y) \\ &= 2 \sum_{n=1}^{\infty} \int_{B_n^c - B_{n+1}^c} \lambda(x)|m| \, \mathrm{d}x \\ &= 2 \int_{B_1} \lambda \, \mathrm{d}|m| < +\infty.\end{aligned}$$

(ii) $g(x, y) \leqslant Kc(x, y)$ for some $K \geqslant 0$ on $P \times P$: We suppose $x, y \in P$. Then

$$\begin{aligned}g(x, y) &\leqslant \sup_n \frac{\varepsilon_n c(x, y)}{|m|(B_n^c)} = \sup_n \frac{c(x, y)}{m^+(B_n)} (m^+(U) - m^+(B_n)) \frac{1}{|m|(B_n^c)} \\ &= \sup_n \frac{m + (B_n^c)}{|m|(B_n^c)} \frac{c(x, y)}{m^+(B_n)} \leqslant \frac{c(x, y)}{m^+(B_1)}.\end{aligned}$$

Very similar arguments serve to demonstrate

(iii) $g(x, y) \leqslant Kc(x, y)$ for some $K \geqslant 0$ on $N \times N$.

(iv) $g(x, y) \leqslant Kc(x, y)$ for some $K \geqslant 0$ on $N \times P$.

Combining (i)–(iv) establishes the claim.

Now $f_n(x, y) \to 0$ as $n \to \infty$ $\forall x, y \in U$. In view of the claim, Lebesgue's dominated convergence theorem implies that

$$\begin{aligned}\|m - m_n\|_w &\leqslant \int c(x, y) b_n(\mathrm{d}x, \mathrm{d}y) \\ &= \int c(x, y) f_n(x, y)(|m| \times |m|)(\mathrm{d}x, \mathrm{d}y) \to 0\end{aligned}$$

as $n \to \infty$. QED

Call a signed measure on U *simple* if it is a finite linear combination of signed measures of the form $\delta_x - \delta_y$. $\mathscr{M}$ contains all the simple measures. In the next lemma we shall use the Strassen–Dudley theorem.

Theorem 5.3.1. Suppose that (U, d) is a s.m.s. and that $P_n \to P$ weakly in $\mathscr{P}(U)$. Then for each $\varepsilon, \delta > 0$ there is some N such that whenever $n \geqslant N$, there is some law b_n on $U \times U$ with marginals P_n and P such that

$$b_n\{(x, y): d(x, y) > \delta) < \varepsilon. \tag{5.3.13}$$

Proof. Further (Corollary 7.4.2; see also Dudley 1989, Theorem 11.6.2), we shall prove that the Prokhorov metric π is minimal with respect to the Ky Fan metric $\mathbf{K}$. In other words

$$\pi(P_1, P_2) = \inf\{\mathbf{K}(P): P \in \mathscr{P}(U \times U), P(\cdot \times U) = P_1(\cdot), P(U \times \cdot) = P_2(\cdot)\}$$

where $\mathbf{K}(P) = \inf\{\varepsilon > 0: P((x, y): d(x, y) > \varepsilon) < \varepsilon\}$. Since π metrizes the weak topology in $\mathscr{P}(U)$ (Dudley 1989), the above equality yields (5.3.13). QED

Lemma 5.3.3. The simple measures are dense in $(\mathscr{M}, \|\cdot\|_w)$.

Proof. In view of Lemmas 3.2.2 and 5.3.2, it is no loss of generality to assume that $m = P - Q$, where P and Q are laws on U supported on a bounded set $U_0 \subseteq U$. Then there are laws $P_n \xrightarrow{w} P$, $Q_n \xrightarrow{w} Q$ such that for each n, we have $P_n(U_0) = Q_n(U_0) = 1$ and $P_n - Q_n$ is simple (see, for example, the Glivenko–Cantelli–Varadarajan theorem (Dudley 1989)). To prove the lemma, it is enough to show that $\|P_n - P\|_w \to 0$ as $n \to \infty$.

Given $\varepsilon > 0$, use the boundedness of U_0 and Condition (C5) on c to find $\delta > 0$ such that $c(x, y) < \varepsilon/2$ whenever $x, y \in U_0$ with $d(x, y) \leqslant \delta$. Put $K = \sup\{\lambda(x): x \in U_0\}$. By Theorem 5.3.1, for all large n, there is a law b_n on $U \times U$ with marginals P_n and P such that $b_n\{(x, y): d(x, y) > \delta\} < \delta\} < \varepsilon/4K$. Put $A = \{(x, y): d(x, y) > \delta\}$. Then

$$\begin{aligned}\|P_n - P\|_w &\leqslant \int c(x, y) b_n(\mathrm{d}x, \mathrm{d}y) = \int_A c(x, y) b_n(\mathrm{d}x, \mathrm{d}y) + \int_{U/A} c(x, y) b_n(\mathrm{d}x, \mathrm{d}y)\\ &\leqslant \int_A (\lambda(x) + \lambda(y)) b_n(\mathrm{d}x, \mathrm{d}y) + \varepsilon/2 \leqslant 2K b_n(A) + \varepsilon/2 < \varepsilon\end{aligned}$$

for all large n. QED

Lemma 5.3.4. The linear transformation D (5.3.12) is an isometric isomorphism of $(\mathbb{L}_0, \|\cdot\|_c)$ onto $(\mathscr{M}^*, \|\cdot\|_w^*)$.

Proof. Suppose that $\phi: \mathscr{M} \to \mathbb{R}$ is a continuous linear functional on $\mathscr{M}$. Fix $a \in U$ and define $f: U \to \mathbb{R}$ by $f(x) = \phi(\delta_x - \delta_a)$. For any $x, y \in U$

$$|f(x) - f(y)| = |\phi(\delta_x - \delta_y)| \leqslant \|\phi\|_w^* \|\delta_x - \delta_y\|_w \leqslant \|\phi\|_w^* c(x, y)$$

so that $\|f\|_c \leqslant \|\phi\|_w^* < \infty$. We see that $\phi(m) = \phi_f(m)$ for $m = \delta_x - \delta_y$ and hence for all simple $m \in \mathscr{M}$. Lemma 5.3.3 implies that $\phi(m) = \phi_f(m)$ for all $m \in \mathscr{M}$. Thus $\phi = D(f)$.

We have shown that D is subjective. Earlier results now apply to complete the argument. QED

Now we consider the adjoint of the transformation D. As usual, the Hahn–Banach theorem applies to show that $(\mathbb{L}_0^*, \|\cdot\|_c^*) \xleftarrow{D^*} (M^{**}, \|\cdot\|_w^{**})$ is an isometric isomorphism, see Theorem II 3.19, Dunford and Schwartz (1988). Let $(\mathcal{M}^{**}, \|\cdot\|_w^{**}) \xleftarrow{T} (\mathcal{M}, \|\cdot\|_w)$ be the natural isometric isomorphism of $\mathcal{M}$ into its second conjugate $\mathcal{M}^{**}$. Then $(\mathbb{L}_0{}^*, \|\cdot\|_c^*) \xleftarrow{D^* \circ T} (\mathcal{M}, \|\cdot\|_w)$ is an isometry. A routine diagram shows that $\|m\|_w = \sup\{\int f \, \mathrm{d}m \colon \|f\|_c \leqslant 1\}$.

We summarize by stating the following, the main result of this section.

Theorem 5.3.2. Let m be a measure in $\mathcal{M}$. Then

$$\|m\|_w = \sup\left\{\int f \, \mathrm{d}m \colon f(x) - f(y) \leqslant c(x, y)\right\}.$$

We now show that the supremum in Theorem 5.3.2 is attained for some optimal f.

Theorem 5.3.3. Let m be a measure in $\mathcal{M}$. Then there is some $f \in \mathbb{L}$ with $\|f\|_c = 1$ such that $\|m\|_w = \int f \, \mathrm{d}m$.

Proof. Using the Hahn–Banach theorem, choose a linear functional ϕ in $\mathcal{M}^*$ with $\|\phi\|^* = 1$ and such that $\phi(m) = \|m\|_w$. By Lemma 5.3.4, we have $\phi = \phi_f$ for some $f \in \mathbb{L}$ with $\|f\|_c = \|\phi\|^* = 1$. QED

Given probability measures P_1, P_2 on U, define the *minimal norm*

$$\mathring{\mu}_c(P_1, P_2) = \inf\left\{\int c(x, y) b(\mathrm{d}x, \mathrm{d}y) \colon b \in \mathbb{B}(P_1 - P_2)\right\} \tag{5.3.14}$$

(see (5.3.6) and Example 3.2.6). Let $\mathscr{P}_\lambda(U)$ be the set of all laws P on U such that λ is P-integrable. Then $\mathring{\mu}_c(P_1, P_2)$ defines a semimetric on $\mathscr{P}_\lambda(U)$ (see Remark 3.2.1). In the next section, we shall analyze the explicit representations and the topological properties of these semimetrics.

It should also be noted that if X and Y are random variables taking values in U, then it is natural to define

$$\mathring{\mu}_c(X, Y) = \mathring{\mu}_c(\Pr_X, \Pr_Y)$$

where $\Pr_X$ is the law of X. We shall freely use both notations in the next section.

Example 5.3.1. Suppose that $c(x, y) = d(x, y)$ and put $\lambda(x) = d(x, a)$ for some

$a \in U$. Then Conditions (C1) to (C5) are satisfied, and Theorem 5.3.2 yields

$$\inf\left\{\int d(x, y)b(\mathrm{d}x, \mathrm{d}y): b \in \mathbb{B}(P_1 - P_2)\right\}$$
$$= \sup\left\{\int f\, \mathrm{d}(P_1 - P_2): \|f\|_L \leqslant 1\right\} \tag{5.3.15}$$

where $P_1, P_2 \in \mathscr{P}_\lambda(U)$ and $\|f\|_L$ is the Lipschitz norm of f. In this case, $\mathring{\mu}_c(P_1, P_2)$ is a metric in $\mathscr{P}_\lambda(U)$. This, the classic situation, has been much studied; see Kantorovich and Rubinstein (1988) and Dudley (1976), Lecture 20. In particular, (5.3.15) gives us the dual representations of $\mathring{\mu}(P_1, P_2)$ given by (3.3.53)

$$\mathring{\mu}(P_1, P_2) = \inf\{\alpha\mathbb{E}d(X, Y): \text{for some } \alpha > 0,\ X \in \mathfrak{X}(U),\ Y \in \mathfrak{X}(U),$$
$$\text{such that } \alpha(\Pr_X - \Pr_Y) = P_1 - P_2\}$$
$$= \sup\left\{\left|\int_U f\, \mathrm{d}(P_1 - P_2)\right|: \|f\|_L \leqslant 1\right\}. \tag{5.3.16}$$

5.4 APPLICATION: EXPLICIT REPRESENTATIONS FOR A CLASS OF MINIMAL NORMS

Throughout this section we take $U = \mathbb{R}$, $d(x, y) = |x - y|$ and define $c: \mathbb{R} \times \mathbb{R} \to [0, +\infty)$ by

$$c(x, y) = |x - y| \max(h(|x - a|), h(|y - a|)) \tag{5.4.1}$$

where a is a fixed point of $\mathbb{R}$ and $h: [0, \infty) \to [0, \infty)$ is a continuous non-decreasing function such that $h(x) > 0$ for $x > 0$. Note that the cost function in Example 5.1.1 (cf. (5.1.24)) has precisely the same form as (5.4.1). Define $\lambda: \mathbb{R} \to [0, \infty)$ by

$$\lambda(x) = 2|x|h(|x - a|).$$

It is not difficult to verify that c and λ satisfy Conditions (C1) to (C5) specified in Section 5.3. As in Section 5.3, the normed space $(\mathbb{L}_0, \|\cdot\|_c)$ and the set $\mathscr{M}$, comprising all finite signed measures m on $\mathbb{R}$ such that $m(U) = 0$ and $\int \lambda\, \mathrm{d}|m| < +\infty$ are to be investigated.

We consider random variables X and Y in $\mathfrak{X} = \mathfrak{X}(\mathbb{R})$ with $\mathbb{E}(\lambda(X)) + \mathbb{E}(\lambda(Y)) < \infty$. Then $m = \Pr_X - \Pr_Y$ is an element of $\mathscr{M}$, and Theorem 5.3.2 implies the dual representation of $\mathring{\mu}_c$:

$$\mathring{\mu}_c(X, Y) = \inf\{\alpha\mathbb{E}(c(X', Y')): X', Y' \in \mathfrak{X},\ \alpha > 0,\ \alpha(\Pr_{X'} - \Pr_{Y'}) = m\}$$
$$= \sup\left\{\left|\int_{\mathbb{R}} f\, \mathrm{d}m\right|: |f(x) - f(y)| \leqslant c(x, y),\ \forall x, y \in \mathbb{R}\right\}. \tag{5.4.2}$$

An explicit representation is given in the following theorem.

Theorem 5.4.1. Suppose c is given by (5.4.1) and $X, Y \in \mathfrak{X}$ with $\mathbb{E}(\lambda(X)) + \mathbb{E}(\lambda(Y)) < \infty$ then

$$\mathring{\mu}_c(X, Y) = \int_{-\infty}^{\infty} h(|x - a|)|F_X(x) - F_Y(x)|\,\mathrm{d}x. \tag{5.4.3}$$

Proof. We begin by proving the theorem in the special case where X and Y are bounded. Suppose that $|X| \leqslant N$ and $|Y| \leqslant N$ for some N. Application of Theorem 5.3.2 with $U := U_N := [-N, N]$ yields $\mathring{\mu}_c(X, Y) = \sup\{|\int f\,dm|: f\colon U_n \to \mathbb{R}, |f(x) - f(y)| \leqslant c(x, y), \forall x,y \in U_N\}$, where $m = \Pr_X - \Pr_Y$. It is easy to check that if $|f(x) - f(y)| \leqslant c(x, y)$ as above, then f is absolutely continuous on any compact interval. Thus, f is differentiable a.e. on $[-N, N]$, and $|f'(x)| \leqslant h(|x - a|)$ wherever f' exists. Therefore

$$\mathring{\mu}_c(X, Y) \leqslant \sup\left\{\left|\int_{-\infty}^{\infty} (F_X(x) - F_Y(x))f'(x)\mathrm{d}x\right| : f\colon U_N \to R,\ |f'(x)| \leqslant h(|x-a|) \text{ a.e.}\right\}$$
$$\leqslant \int_{-\infty}^{\infty} h(|x - a|)|F_X(x) - F_Y(y)|\,\mathrm{d}x$$

using integration by parts.

On the other hand, if f is absolutely continuous with $|f'(x)| \leqslant h(|x - a|)$ a.e., then $|f(x) - f(y)| = |\int_x^y f'(t)\,\mathrm{d}t| \leqslant |x - y|\max(h(|x - a|), h(|y - a|)) = c(x, y)$. Define $f_*\colon \mathbb{R} \to \mathbb{R}$ by

$$f'_*(x) = h(|x - a|)\operatorname{sgn}(F_X(x) - F_Y(x)) \text{ a.e.}$$

Then

$$\mathring{\mu}_c(X, Y) = \sup\left\{\left|\int F_X(x) - F_Y(x))f'(x)\,\mathrm{d}x\right| : |f'(x)| \leqslant h(|x - a|) \text{ a.e.}\right\}$$
$$\geqslant \left|\int (F_X(x) - F_Y(x))f'_*(x)\,\mathrm{d}x\right|$$
$$= \int h(|x - a|)|F_X(x) - F_Y(x)|\,\mathrm{d}x.$$

We have shown that whenever X and Y are bounded random variables then (5.4.3) holds. Now define $H\colon \mathbb{R} \to \mathbb{R}$ by

$$H(t) = \int_0^t h(|x - a|)\,\mathrm{d}x. \tag{5.4.4}$$

For $t \geqslant 0$, $H(t) \leqslant h(|a|)|a| + |t - a|h(|t - a|)$ so that $\mathbb{E}(\lambda(X)) + \mathbb{E}(\lambda(Y)) < \infty$ implies that $\mathbb{E}|H(X)| + \mathbb{E}|H(Y)| < \infty$. Under this assumption integrating by parts we obtain

$$\mathbb{E}|H(X)| = \int_0^{\infty} h(|x - a|)(1 - F_X(x))\,\mathrm{d}x + \int_{-\infty}^{0} h(|x - a|)F_X(x)\,\mathrm{d}x.$$

An analogous equality holds for the variable Y. These imply that

$$\int_{-\infty}^{\infty} h(|x-a|)|F_X(x)-F_Y(x)|\,\mathrm{d}x<\infty.$$

For $n \geqslant 1$, define random variables X_n, Y_n by

$$X_n=\begin{cases} n & \text{if } X>n \\ X & \text{if } -n\leqslant X\leqslant n \\ -n & \text{if } X<-n \end{cases} \quad Y_n=\begin{cases} n & \text{if } Y>n \\ Y & \text{if } -n\leqslant Y\leqslant n \\ -n & \text{if } Y<-n \end{cases}.$$

Then $X_n \to X$, $Y_n \to Y$ in distribution, and for $n \geqslant |a|$

$$\mathring{\mu}_c(X_n, X) \leqslant \mathbb{E}c(X_n, X) \leqslant \mathbb{E}(|X|I\{|X|\geqslant n\}h(|X-a|))$$

which tends to 0 as $n\to\infty(\mathbb{E}(\lambda(X))<\infty)$. Similarly, $\mathring{\mu}_c(Y_n, Y)\to 0$. Then $\mathring{\mu}_c(X_n, Y_n)\to\mathring{\mu}_c(X, Y)$ as $n\to\infty$. Also, we have

$$|F_{X_n}(x)-F_{Y_n}(x)|=\begin{cases} |F_X(x)-F_Y(x)| & \text{for } -n\leqslant x<n \\ 0 & \text{otherwise} \end{cases}.$$

Applying dominated convergence, we see that as $n\to\infty$

$$\int h(|x-a|)|F_{X_n}(x)-F_{Y_n}(x)|\,\mathrm{d}x \to \int h(|x-a|)|F_X(x)-F_Y(x)|\,\mathrm{d}x.$$

Combining this with $\mathring{\mu}_c(X_n, Y_n)\to\mathring{\mu}_c(X, Y)$ and the result for bounded random variables yields

$$\mathring{\mu}_c(X, Y)=\int_{-\infty}^{\infty} h(|x-a|)|F_X(x)-F_Y(x)|\,\mathrm{d}x. \qquad \text{QED}$$

For $h(x)=1$, this yields a well known formula presented in Dudley (1976), Theorem 20.10. We also note the following formulation, which is not hard to derive from the strict monotonicity of H (see (5.4.4)).

Corollary 5.4.1. Suppose c is given by (5.4.1) and $X, Y\in\mathfrak{X}$ with $\mathbb{E}(\lambda(X))+\mathbb{E}(\lambda(Y))<\infty$ and put $P=\mathrm{Pr}_X$, $Q=\mathrm{Pr}_Y$. Then

$$\mathring{\mu}_c(P, Q)=\int_{-\infty}^{\infty}|F_{H(X)}(x)-F_{H(Y)}(x)|\,\mathrm{d}x \tag{5.4.5}$$

where H is given by (5.4.4).

For $h(x)=1$, we see that $H(t)=t$ and that $\mathring{\mu}_c$ gives the Kantorovich metric (cf. Section 2.1).

Corollary 5.4.2. In this context, $\mathring{\mu}_c(P_1, P_2)$ defines a metric on $\mathscr{P}_\lambda(\mathbb{R}):=\{P\colon \int_{\mathbb{R}}\lambda\,\mathrm{d}P<\infty\}$.

CHAPTER 6

Quantitative Relationships between Minimal Distances and Minimal Norms

In Chapter 5 we discussed Problems 5.1.1 and 5.1.2 of evaluating the minimal distances $\hat{\mu}_c$ and minimal norms $\mathring{\mu}_c$. In this chapter we shall concern ourselves with Problem 5.1.3: what kind of relationships exist between $\hat{\mu}_c$ and $\mathring{\mu}_c$?

6.1 KANTOROVICH METRIC $\hat{\mu}_c$ IS EQUAL TO KANTOROVICH–RUBINSTEIN NORM $\mathring{\mu}_c$ IF AND ONLY IF THE COST FUNCTION c IS A METRIC

Levin (1975) proved that if U is a compact, $c(x, x) = 0$, $c(x, y) \geqslant 0$, and $c(x, y) + c(y, x) > 0$ for $x \neq y$, then $\hat{\mu}_c = \mathring{\mu}_c$ if and only if $c(x, y) + c(y, x)$ is a metric is U. In the case of a s.m.s. U we have the following version of Levin's result.

Theorem 6.1.1. (Neveu and Dudley 1980) Suppose U is a s.m.s. and $c \in \mathfrak{C}^*$ (cf. Corollary 5.2.1). Then

$$\hat{\mu}_c(P_1, P_2) = \mathring{\mu}_c(P_1, P_2) \tag{6.1.1}$$

for all P_1 and P_2 with

$$\int_U c(x, a)(P_1 + P_2)(\mathrm{d}x) < \infty \tag{6.1.2}$$

if and only if c is a metric.

Proof. Suppose (6.1.1) holds and put $P_1 = \delta_x$ and $P_2 = \delta_y$ for $x, y \in U$. Then the set $\mathscr{P}^{(P_1, P_2)}$ of all laws in $U \times U$ with marginals P_1 and P_2 contains only

$P_1 \times P_2 = \delta_{(x,y)}$ and by Theorem 5.3.2,

$$\begin{aligned}\hat{\mu}_c(P_1, P_2) = c(x, y) = \mathring{\mu}_c(P_1, P_2) &= \sup\left\{\int f \, \mathrm{d}(P_1 - P_2) \colon \|f\|_c \leqslant 1\right\} \\ &= \sup\{|f(x) - f(y)| \colon \|f\|_c \leqslant 1\} \\ &\leqslant \sup\{|f(x) - f(z)| + |f(z) - f(y)| \colon \|f\|_c \leqslant 1\} \\ &\leqslant c(x, z) + c(z, y).\end{aligned}$$

By assumption $c \in \mathfrak{C}^*$ and hence the triangle inequality implies that c is a metric in U.

Now define $\mathscr{G}(U)$ as the set of all pairs (f, g) of continuous functions $f\colon U \to \mathbb{R}$ and $g\colon U \to \mathbb{R}$ such that $f(x) + g(y) \leqslant c(x, y)$ $\forall x, y \in U$. Let $\mathscr{G}_B(U)$ be the set of all pairs $(f, g) \in \mathscr{G}(U)$ with f and g bounded.

Now suppose that $c(x, y)$ is a metric and that $(f, g) \in \mathscr{G}_B(U)$. Define $h(x) = \inf\{c(x, y) - g(y) \colon y \in U\}$. As the infimum of a family of continuous functions, h is upper semi-continuous. For each $x \in U$, we have $f(x) \leqslant h(x) \leqslant -g(x)$. Then

$$\begin{aligned}h(x) - h(x') &= \inf_u (c(x, u) - g(u)) - \inf_v (c(x', v) - g(v)) \\ &\leqslant \sup_v (g(v) - c(x', v) + c(x, v) - g(v)) \\ &= \sup_v (c(x, v) - c(x', v)) \leqslant c(x, x')\end{aligned}$$

so that $\|h\|_c \leqslant 1$. Then for P_1, P_2 satisfying (6.1.2) we have

$$\int f \, \mathrm{d}P_1 + \int g \, \mathrm{d}P_2 \leqslant \int h \, \mathrm{d}(P_1 - P_2)$$

so that (according to Corollary 5.2.1 and Theorem 5.3.2) we have

$$\begin{aligned}\hat{\mu}_c(P_1, P_2) &= \sup\left\{\int f \, \mathrm{d}P_1 + \int g \, \mathrm{d}P_2 \colon (f, g) \in \mathscr{G}_B(U)\right\} \\ &\leqslant \sup\left\{\int h \, \mathrm{d}(P_1 - P_2) \colon \|h\|_c \leqslant 1\right\} = \mathring{\mu}_c(P_1, P_2).\end{aligned}$$

Thus $\hat{\mu}_c(P_1, P_2) = \mathring{\mu}_c(P_1, P_2)$. QED

Corollary 6.1.1. Let (U, d) be a s.m.s. and $a \in U$. Then

$$\hat{\mu}_d(P_1, P_2) = \mathring{\mu}_d(P_1, P_2) = \sup\left\{\int f \, \mathrm{d}(P_1 - P_2) \colon \|f\|_{\mathrm{L}} \leqslant 1\right\} \tag{6.1.3}$$

whenever

$$\int d(x, a)P_i(\mathrm{d}x) < \infty \quad i = 1, 2. \tag{6.1.4}$$

The supremum is attained for some optimal f_0 with $\|f_0\|_L := \sup_{x \neq y}\{|f(x) - f(y)|/d(x, y)\}$.

If P_1 and P_2 are tight, there are some $b_0 \in \mathscr{P}^{(P_1, P_2)}$ and $f_0\colon U \to \mathbb{R}$ with $\|f_0\|_L \leqslant 1$ such that

$$\hat{\mu}_d(P_1, P_2) = \int d(x, y)b_0(\mathrm{d}x, \mathrm{d}y) = \int f_0 \,\mathrm{d}(P_1 - P_2)$$

where $f_0(x) - f_0(y) = d(x, y)$ for b_0-a.e. (x, y) in $U \times U$.

Proof. Put $c(x, y) = d(x, y)$. Application of the theorem proves the first sentence. The second (existence of f_0) follows from Theorem 5.3.3.

For each $n \geqslant 1$, choose $b_n \in \mathscr{P}^{(P_1, P_2)}$ with

$$\int d(x, y)b_n(\mathrm{d}x, \mathrm{d}y) < \hat{\mu}_d(P_1, P_2) + \frac{1}{n}.$$

If P_1 and P_2 are tight then by Corollary 5.2.1, there exists $b_0 \in \mathscr{P}^{(P_1, P_2)}$ such that

$$\hat{\mu}_d(P_1, P_2) = \int d(x, y)b_0(\mathrm{d}x, \mathrm{d}y)$$

i.e., that b_0 is optimal. Integrating both sides of $f_0(x) - f_0(y) \leqslant d(x, y)$ with respect to b_0 yields $\int f_0 \mathrm{d}(P_1 - P_2) \leqslant \int d(x, y)b_0(\mathrm{d}x, \mathrm{d}y)$. However, we know that we have equality of these integrals. This implies that $f_0(x) - f_0(y) = d(x, y)$ b_0-a.e.

QED

6.2 INEQUALITIES BETWEEN $\hat{\mu}_c, \mathring{\mu}_c$

In the previous section we have studied conditions under which $\hat{\mu}_c = \mathring{\mu}_c$. In general $\mathring{\mu}_c \neq \hat{\mu}_c$. For example, if $U = \mathbb{R}$, $d(x, y) = |x - y|$,

$$c(x, y) = d(x, y)\max(1, d^{p-1}(x, a), d^{p-1}(y, a)) \quad p \geqslant 1 \tag{6.2.1}$$

then for any laws $P_i (i = 1,2)$ on $\mathscr{B}(\mathbb{R})$ with distribution functions (d.f.s) F_i, we have the following explicit expressions

$$\hat{\mu}_c(P_1, P_2) = \int_0^1 c(F_1^{-1}(t), F_2^{-1}(t))\,\mathrm{d}t \tag{6.2.2}$$

where F_i^{-1} is the function inverse to the d.f. F_i, see further Theorem 7.3.2. On the other hand

$$\mathring{\mu}_c(P_1, P_2) = \int_{-\infty}^{\infty} |F_1(x) - F_2(x)| \max(1, |x - a|^{p-1})\,\mathrm{d}x \tag{6.2.3}$$

(see Theorem 5.4.1). However, in the space $\mathscr{M}_p = \mathscr{M}_p(U)$ ($U = (U, d)$ is s.m.s.) of all Borel probability measures P with finite $\int d^p(x, a)P(\mathrm{d}x)$, the functionals $\hat{\mu}_c$ and $\mathring{\mu}_c$ (where c is given by (6.2.1)) metrize one and the same topology. Namely, the following $\hat{\mu}_c$- and $\mathring{\mu}_c$-convergence criterion will be proved.

Theorem 6.2.1. Let (U, d) be a s.m.s., c be given by (6.2.1) and $P, P_n \in \mathscr{M}_p$ $(n = 1, 2, \dots)$. Then the following are equivalent

(I) $$\hat{\mu}_c(P_n, P) \to 0$$

(II) $$\mathring{\mu}_c(P_n, P) \to 0$$

(III) P_n converges weakly to P ($P_n \xrightarrow{\mathrm{w}} P$) and

$$\lim_{N \to \infty} \sup_n \int d^p(x, a) I\{d(x, a) > N\} P_n(\mathrm{d}x) = 0,$$

(IV) $$P_n \xrightarrow{\mathrm{w}} P \text{ and } \int d^p(x, a)P_n(\mathrm{d}x) \to \int d^p(x, a)P(\mathrm{d}x).$$

(The assertion of the theorem is an immediate consequence of Theorems 6.2.2 to 6.2.5 below and the more general Theorem 6.3.1.)

Theorem 6.2.1 is a qualitative $\hat{\mu}_c(\mathring{\mu}_c)$-convergence criterion. One can rewrite (III) as

$$\pi(P_n, P) \to 0 \quad \text{and} \quad \lim_{\varepsilon \to 0} \sup_n \omega(\varepsilon; P_n; \lambda) = 0$$

where π is the Prokhorov metric

$$\pi(P, Q) := \inf\{\varepsilon > 0\colon P(A) \leqslant Q(A^\varepsilon) + \varepsilon \quad \forall A \in \mathscr{B}(U)\}$$
$$(A^\varepsilon := \{x\colon d(x, A) < \varepsilon\}) \tag{6.2.4}$$

(see Examples 3.2.3 and 4.2.2) and $\omega(\varepsilon; P; \lambda)$ is the following modulus of λ-integrability,

$$\omega(\varepsilon; P; \lambda) := \int \lambda(x) I\left\{d(x, a) > \frac{1}{\varepsilon}\right\} P(\mathrm{d}x) \tag{6.2.5}$$

where $\lambda(x) := \max(d(x, a), d^p(x, a))$. Analogously, (IV) is equivalent to

$$\pi(P_n, P) \to 0 \quad \text{and} \quad D(P_n, P; \lambda) \to 0 \tag{IV*}$$

where

$$D(P, Q; \lambda) := \left| \int \lambda(x)(P - Q)(\mathrm{d}x) \right|. \tag{6.2.6}$$

In this section we investigate quantitative relationships between $\hat{\mu}_c$, $\mathring{\mu}_c$, π, ω and D in terms of inequalities between these functionals. These relationships yield convergence and compactness criteria in the space of measures w.r.t. the Kantorovich type functionals $\hat{\mu}_c$ and $\mathring{\mu}_c$ (c.f. Examples 3.2.2 and 3.2.6) as well as the $\mathring{\mu}_c$-completeness of the space of measures.

In the following we assume that the cost function c has the form considered in Example 5.1.1:

$$c(x, y) = d(x, y)k_0(d(x, a), d(y, a)) \quad x, y \in U \tag{6.2.7}$$

where $k_0(t, s)$ is a symmetric continuous function non-decreasing on both arguments $t \geqslant 0$, $s \geqslant 0$, and satisfying the following conditions:

(C1)
$$\alpha := \sup_{s \neq t} \frac{|K(t) - K(s)|}{|t - s| k_0(t, s)} < \infty$$

where $K(t) := tk_0(t, t)$, $t \geqslant 0$;

(C2)
$$\beta := k(0) > 0$$

where $k(t) = k_0(t, t)$ $t \geqslant 0$; and

(C3)
$$\gamma := \sup_{t \geqslant 0, s \geqslant 0} \frac{k_0(2t, 2s)}{k_0(t, s)} < \infty.$$

If c is given by (6.2.1) then c admits the form (6.2.7) with $k_0(t, s) = \max(1, t^{p-1}, s^{p-1})$ and in this case $\alpha = p$, $\beta = 1$, $\gamma = 2^{p-1}$. Further, let $\mathscr{P}_\lambda = \mathscr{P}_\lambda(U)$ be the space of all probability measures on the s.m.s. (U, d) with finite λ-moment

$$\mathscr{P}_\lambda(U) = \left\{ P \in \mathscr{P}(U) : \int_U \lambda(x) P(dx) < \infty \right\} \tag{6.2.8}$$

where $\lambda(x) = K(d(x, a))$ and a is a fixed point of U.

In Theorems 6.2.2 to 6.2.5 we assume that $P_1 \in \mathscr{P}_\lambda$, $P_2 \in \mathscr{P}_\lambda$, $\varepsilon > 0$, and denote $\hat{\mu}_c := \hat{\mu}_c(P_1, P_2)$ (cf. (5.1.16)), $\mathring{\mu}_c := \mathring{\mu}_c(P_1, P_2)$ (cf. (5.1.17)), $\pi := \pi(P_1, P_2)$,

$$\omega_i(\varepsilon) := \omega(\varepsilon; P_i; \lambda) := \int \lambda(x) I\{d(x, a) > 1/\varepsilon\} P_i(\mathrm{d}x), \; P_i \in \mathscr{P}_\lambda$$

$$D := D(P_1, P_2; \lambda) := \left| \int \lambda \mathrm{d}(P_1 - P_2) \right|$$

and the function c satisfies conditions C1 to C3. We begin with an estimate of $\hat{\mu}_c$ from above in terms of π and $\omega_i(\varepsilon)$.

Theorem. 6.2.2.

$$\hat{\mu}_c \leqslant \pi[4K(1/\varepsilon) + \omega_1(1) + \omega_2(1) + 2k(1)] + 5\omega_1(\varepsilon) + 5\omega_2(\varepsilon). \tag{6.2.9}$$

Proof. Recall $\mathscr{P}^{(P_1,P_2)}$ to be the space of all laws P on $U \times U$ with prescribed marginals P_1 and P_2. Let $\mathbf{K} = \mathbf{K}_1$ be the Ky Fan metric with parameter 1 (see Example 3.3.2)

$$\mathbf{K}(P) := \inf\{\delta > 0\colon P(d(x, y) > \delta) < \delta\} \quad P \in \mathscr{P}_\lambda(U). \tag{6.2.10}$$

Claim 1. For any $N > 0$ and for any measure P on U^2 with marginals P_1 and P_2, i.e. $P \in \mathscr{P}^{(P_1,P_2)}$, we have

$$\int_{U\times U} c(x, y)P(\mathrm{d}x, \mathrm{d}y) \leq \mathbf{K}(P)\left[4K(N) + \int_U k(d(x, a))(P_1 + P_2)(dx)\right]$$
$$+ 5\omega_1(1/N) + 5\omega_2(1/N). \tag{6.2.11}$$

Proof of claim. Suppose $\mathbf{K}(P) < \sigma \leqslant 1$, $P \in \mathscr{P}^{(P_1,P_2)}$. Then by (6.2.7) and (C3),

$$\int c(x, y)P(\mathrm{d}x, \mathrm{d}y) \leqslant \int d(x, y)k(\max\{d(x, a), d(y, a)\})P(\mathrm{d}x, \mathrm{d}y)$$
$$\leqslant I_1 + I_2$$

where

$$I_1 := \int_{U\times U} d(x, y)k(d(x, a))P(\mathrm{d}x, \mathrm{d}y)$$

and

$$I_2 := \int_{U\times U} d(x, y)k(d(y, a))P(\mathrm{d}x, \mathrm{d}y).$$

Let us estimate I_1

$$I_1 := \int d(x, y)k(d(x, a))[I\{d(x, y) < \delta\} + I\{d(x, y) \geqslant \delta\}]P(\mathrm{d}x, \mathrm{d}y)$$
$$\leqslant \delta \int k(d(x, a))P(\mathrm{d}x, \mathrm{d}y) + \int d(x, y)k(d(x, a))I\{d(x, y) \geqslant \delta\}P(\mathrm{d}x, \mathrm{d}y)$$
$$\leqslant I_{11} + I_{12} + I_{13} \tag{6.2.12}$$

where

$$I_{11} := \delta \int_U k(d(x, a))[I\{d(x, a) \geqslant 1\} + I\{d(x, a) \leqslant 1\}]P_1(\mathrm{d}x)$$

$$I_{12} := \int_{U\times U} d(x, a)k(d(x, a))I\{d(x, y) \geqslant \delta\}P(\mathrm{d}x, \mathrm{d}y)$$

and

$$I_{13} := \int d(y, a)k(d(x, a))I\{d(x, y) \geq \delta\}P(\mathrm{d}x, \mathrm{d}y).$$

Obviously, by $\lambda(x) := K(d(x, a))$, $I_{11} \leq \delta \int k(d(x, a))I\{d(x, a) \geq 1\}P_1(\mathrm{d}x) + \delta k(1) \leq \delta\omega_1(1) + \delta k(1)$. Further

$$\begin{aligned}I_{12} &= \int K(d(x, a))I\{d(x, y) \geq \delta\}[I\{d(x, a) > N\} + I\{d(x, a) \leq N\}]P(\mathrm{d}x, \mathrm{d}y)\\ &\leq \int_U \lambda(x)I\{d(x, a) > N\}P_1(\mathrm{d}x) + K(N)\int_{U\times U} I\{d(x, y) \geq \delta\}P(\mathrm{d}x, \mathrm{d}y)\\ &\leq \omega_1(1/N) + K(N)\delta.\end{aligned}$$

Now let us estimate the last term in the estimate (6.2.12)

$$\begin{aligned}I_{13} = \int_{U\times U} & d(y, a)k(d(x, a))I\{d(x, y) \geq \delta\}[I\{d(x, a) \geq d(y, a) > N\}\\ &+ I\{d(y, a) > d(x, a) > N\} + I\{d(x, a) > N, d(y, a) \leq N\}\\ &+ I\{d(x, a) \leq N, d(y, a) > N] + I\{d(x, a) \leq N, d(y, a) \leq N\}]P(\mathrm{d}x, \mathrm{d}y)\\ \leq \int_{U\times U} & \lambda(x)I\{d(x, a) \geq d(y, a) > N\}P(\mathrm{d}x, \mathrm{d}y)\\ &+ \int_{U\times U} \lambda(y)I\{d(y, a) \geq d(x, a) \geq N\}P(\mathrm{d}x, \mathrm{d}y)\\ &+ \int_U \lambda(x)I\{d(x, a) > N\}P_1(\mathrm{d}x) + \int_U \lambda(y)I\{d(y, a) > N\}P_2(\mathrm{d}y)\\ &+ K(N)\int_{U\times U} I\{d(x, y) \geq \delta\}P(\mathrm{d}x, \mathrm{d}y)\\ \leq 2\omega_1(1/N) & + 2\omega_2(1/N) + K(N)\delta.\end{aligned}$$

Summarizing the above estimates we obtain $I_1 \leq \delta\omega_1(1) + \delta k(1) + 3\omega_1(1/N) + 2\omega_2(1/N) + 2K(N)\delta$. By symmetry we have $I_2 \leq \delta\omega_2(1) + \delta k(1) + 3\omega_2(1/N) + 2\omega_1(1/N) + 2K(N)\delta$. Therefore the last two estimates imply

$$\begin{aligned}\int c(x, y)P(\mathrm{d}x, \mathrm{d}y) &\leq I_1 + I_2\\ &\leq \delta(\omega_1(1) + \omega_2(1) + 2k(1) + 4K(N)) + 5\omega_1(1/N) + 5\omega_2(1/N).\end{aligned}$$

Letting $\delta \to \mathbf{K}(P)$ we obtain (6.2.11) which proves the claim.

Claim 2 (Strassen–Dudley Theorem).

$$\inf\{\mathbf{K}(P): P \in \mathscr{P}^{(P_1,P_2)}\} = \pi(P_1, P_2). \tag{6.2.13}$$

Indication: Dudley (1989), Section 11.6 (see further Corollary 7.4.2).
Claims 1 and 2 complete the proof of the theorem. QED

The next theorem shows that $\hat{\mu}_c$- and $\mathring{\mu}_c$-convergence implies the weak convergence of measures.

Theorem 6.2.3.

$$\beta\pi^2 \leqslant \mathring{\mu}_c \leqslant \hat{\mu}_c. \tag{6.2.14}$$

Proof. Obviously, for any continuous non-negative function c,

$$\mathring{\mu}_c \leqslant \hat{\mu}_c \tag{6.2.15}$$

and

$$\mathring{\mu}_c \geqslant \zeta_c \tag{6.2.16}$$

where ζ_c is the Zolatarev simple metric with ζ-structure (see Definition 4.3.1)

$$\begin{aligned}\zeta_c &:= \zeta_c(P_1, P_2)\\ &:= \sup\left\{\left|\int_U f\,\mathrm{d}(P_1 - P_2)\right| : f\colon U \to \mathbb{R},\ |f(x) - f(y)| \leqslant c(x, y) \quad \forall x, y \in U\right\}.\end{aligned} \tag{6.2.17}$$

Now, using assumption (C2) we have that $c(x, y) \geqslant \beta d(x, y)$ and hence, $\zeta_c \geqslant \beta\zeta_d$. Thus, by (6.2.16)

$$\mathring{\mu}_c \geqslant \beta\zeta_d. \tag{6.2.18}$$

Claim 3.

$$\zeta_d \geqslant \pi^2. \tag{6.2.19}$$

Proof of claim 3. Using the dual representation of $\hat{\mu}_d$ (see (6.1.3)) we are led to

$$\hat{\mu}_d = \zeta_d \tag{6.2.20}$$

which in view of the inequality

$$\int d(x, y)P(\mathrm{d}x, \mathrm{d}y) \geqslant \mathbf{K}^2(P) \quad \text{for any } P \in \mathscr{P}^{(P_1,P_2)} \tag{6.2.21}$$

establishes (6.2.19). The proof of the claim is now completed.

The desired inequalities (6.2.14) are the consequence of (6.2.15), (6.2.16), (6.2.18) and Claim 3. QED

The next theorem establishes the uniform λ-integrability

$$\lim_{\varepsilon \to 0} \sup_n \omega(\varepsilon, P_n, \lambda) = 0$$

of the sequence of measures $P_n \in \mathscr{P}_\lambda$ $\mathring{\mu}_c$-converging to a measure $P \in \mathscr{P}_\lambda$.

Theorem 6.2.4.

$$\omega_1(\varepsilon/2) \leqslant \alpha(2\gamma + 1)\mathring{\mu}_c + 2(\gamma + 1)\omega_2(\varepsilon). \tag{6.2.22}$$

Proof. For any $N > 0$, by the triangle inequality, we have

$$\omega_1(1/2N) := \int \lambda(x) I\{d(x, a) > 2N\} P_1(\mathrm{d}x) \leqslant \mathscr{T}_1 + \mathscr{T}_2 \tag{6.2.23}$$

where

$$\mathscr{T}_1 := \left| \int \lambda(x) I\{d(x, a) > 2N\}(P_1 - P_2)(\mathrm{d}x) \right|$$

and

$$\mathscr{T}_2 := \int \lambda(x) I\{d(x, a) > N\} P_2(\mathrm{d}x) = \omega_2(1/N).$$

Claim 1.

$$\mathscr{T}_1 \leqslant \alpha\mathring{\mu}_c + K(2N) \int I\{d(x, a) > 2N\}(P_1 + P_2)(\mathrm{d}x). \tag{6.2.24}$$

Proof of claim 1. Denote $f_N(x) := (1/\alpha)\max(\lambda(x), K(2N))$. Since $\lambda(x) = K(d(x, a)) = d(x, a)k_0(d(x, a), d(x, a))$ then by (C1),

$$\begin{aligned} |f_N(x) - f_N(y)| &\leqslant (1/\alpha)|\lambda(x) - \lambda(y)| \\ &\leqslant |d(x, a) - d(y, a)| k_0(d(x, a), d(y, a)) \leqslant c(x, y) \end{aligned}$$

for any $x, y \in U$. Thus the inequalities

$$\left| \int_U f_N(x)(P_1 - P_2)(\mathrm{d}x) \right| \leqslant \zeta_c(P_1, P_2) \leqslant \mathring{\mu}_c(P_1, P_2) \tag{6.2.25}$$

follow from (6.2.16) and (6.2.17). Since $\alpha f_N(x) = \max(K(d(x, a)), K(2N))$ and

(6.2.25) holds, then

$$\begin{aligned}
\mathscr{T}_1 < & \left| \int_U K(d(x, a)) I\{d(x, a) > 2N\}(P_1 - P_2)(\mathrm{d}x) \right. \\
& \left. - \int_U K(2N) I\{d(x, a) \leqslant 2N\}(P_1 - P_2)(\mathrm{d}x) \right| \\
& + K(2N) \left| \int_U I\{d(x, a) \leqslant 2N\}(P_1 - P_2)(\mathrm{d}x) \right| \\
= & \left| \int_U \alpha f_N(x)(P_1 - P_2)(\mathrm{d}x) \right| + K(2N) \left| \int_U I\{d(x, a) > 2N\}(P_1 - P_2)(\mathrm{d}x) \right| \\
\leqslant & \ \alpha \mathring{\mu}_c + K(2N) \int I\{d(x, a) > 2N\}(P_1 + P_2)(\mathrm{d}x)
\end{aligned}$$

which proves the claim.

Claim 2.

$$A(P_1) := K(2N) \int_U I\{d(x, a) > 2N\} P_1(\mathrm{d}x) \leqslant 2\alpha\gamma\mathring{\mu}_c + 2\gamma\omega_2(1/N). \tag{6.2.26}$$

Proof of claim 2. As in the proof of Claim 1 we choose an appropriate Lipschitz function. Namely, write

$$g_N(x) = (1/(2\alpha\gamma)) \min\{K(2N), K(2d(x, O(a, N)))\}$$

where $O(a, N) := \{x : d(x, a) \leqslant N\}$. Using (C1) and (C3),

$$|g_N(x) - g_N(y)| \leqslant (1/(2\alpha\gamma)) |K(2d(x, O(a, N))) - K(2d(y, O(a, N)))|$$

(by C1)

$$\leqslant (1/\gamma) |d(x, O(a, N)) - d(y, O(a, N))| k_0(2d(x, O(a, N)), 2d(y, O(a, N)))$$

(by C3)

$$\leqslant d(x, y) k_0(d(x, O(a, N)), d(y, O(a, N))) \leqslant c(x, y).$$

Hence

$$\left| \int g_N(P_1 - P_2)(\mathrm{d}x) \right| \leqslant \zeta_c \leqslant \mathring{\mu}_c. \tag{6.2.27}$$

Using (6.2.27) and the implications

$$d(x, a) > 2N \Rightarrow d(x, O(a, N)) > N \Rightarrow K(2d(x, O(a, N))) \geqslant K(2N)$$

we obtain the following chain of inequalities:

$$\begin{aligned}
A(P_1) &\leqslant 2\alpha\gamma \int g_N(x)\, P_1(\mathrm{d}x) \\
&\leqslant 2\alpha\gamma \left| \int g_N(x)(P_1 - P_2)(\mathrm{d}x) \right| + 2\alpha\gamma \int_U g_N(x) P_2(\mathrm{d}x) \\
&\leqslant 2\alpha\gamma\mathring{\mu}_c + \int K(2d(x, O(a, N))) I\{d(x, a) \geqslant N\} P_2(dx) \qquad (6.2.28)
\end{aligned}$$

$$\left(\text{by C3, } \frac{K(2t)}{K(t)} = \frac{2tk_0(2t, 2t)}{tk_0(t, t)} \leqslant 2\gamma\right)$$

$$\begin{aligned}
&\leqslant 2\alpha\gamma\mathring{\mu}_c + 2\gamma \int K(d(x, O(a, N))) I\{d(x, a) \geqslant N\} P_2(\mathrm{d}x) \\
&\leqslant 2\alpha\gamma\mathring{\mu}_c + 2\gamma\omega_2(1/N)
\end{aligned}$$

which proves the claim.

For $A(P_2)$ (see (6.2.26)) we have the following estimate

$$A(P_2) \leqslant \int_U K(d(x, a)) I\{d(x, a) > 2N\} P_2(\mathrm{d}x) \leqslant \omega_2(1/N). \qquad (6.2.29)$$

Summarizing (6.2.23), (6.2.24), (6.2.26) and (6.2.29) we obtain

$$\omega_1(1/2N) \leqslant \alpha\mathring{\mu}_c + A(P_1) + A(P_2) + \omega_2(1/N) \leqslant (\alpha + 2\alpha\gamma)\mathring{\mu}_c + (2\gamma + 2)\omega_2(1/N)$$

for any $N > 0$ as desired. QED

The next theorem shows that $\mathring{\mu}_c$-convergence implies convergence of the λ-moments.

Theorem 6.2.5.

$$D \leqslant \alpha\mathring{\mu}_c. \qquad (6.2.30)$$

Proof. By C1, for any finite non-negative measure Q with marginals P_1 and P_2, we have

$$\begin{aligned}
D := \left| \int_U \lambda(x)(P_1 - P_2)(\mathrm{d}x) \right| &= \left| \int_{U \times U} \lambda(x) - \lambda(y) Q(\mathrm{d}x, \mathrm{d}y) \right| \\
&\leqslant \int_{U \times U} \alpha |d(x, a) - d(y, a)| k_0(d(x, a), d(y, a)) Q(\mathrm{d}x, \mathrm{d}y) \\
&\leqslant \alpha \int_{U \times U} c(x, y) Q(\mathrm{d}x, \mathrm{d}y)
\end{aligned}$$

which completes the proof of (6.2.30). QED

The inequalities (6.2.9), (6.2.14), (6.2.22) and (6.2.30) described in Theorems 6.2.2 to 6.2.5 imply criteria for convergence, compactness and uniformity in the spaces of probability measures $(\mathscr{P}(U), \hat{\mu}_c)$ and $(\mathscr{P}(U), \mathring{\mu}_c)$ (cf. also the next section). Moreover, the estimates obtained for $\hat{\mu}_c$ and $\mathring{\mu}_c$ may be viewed as quantitative results demanding conditions that are necessary and sufficient for $\hat{\mu}_c$- and $\mathring{\mu}_c$-convergence. Note that in general quantitative results require assumptions additional to the set of necessary and sufficient conditions implying the qualitative results. The classic example is the CLT where the uniform convergence of the normalized sum of i.i.d. r.v.s can be at any low rate assuming only the existence of the second moment.

6.3 CONVERGENCE, COMPACTNESS AND COMPLETENESS IN $(\mathscr{P}(U), \hat{\mu}_c)$ AND $(\mathscr{P}(U), \mathring{\mu}_c)$

In this section we assume that the cost function c satisfies conditions C1 to C3 in the previous section and $\lambda(x) = K(d(x, a))$. We begin with the criterion for $\hat{\mu}_c$- and $\mathring{\mu}$-convergence.

Theorem 6.3.1. If P_n and $P \in \mathscr{P}_\lambda(U)$, then the following statements are equivalent

(A) $$\hat{\mu}_c(P_n, P) \to 0;$$

(B) $$\mathring{\mu}_c(P_n, P) \to 0;$$

(C) $P_n \xrightarrow{w} P$ (P_n converges weakly to P) and $\int \lambda \, \mathrm{d}(P_n - P) \to 0$ as $n \to \infty$;

(D) $$P_n \xrightarrow{w} P \text{ and } \lim_{\varepsilon \to 0} \sup_n \omega_n(\varepsilon) = 0$$

where $\omega_n(\varepsilon) := \omega(\varepsilon; P_n; \lambda) = \int \lambda(x) I\{d(x, a) > 1/\varepsilon\} P_n(\mathrm{d}x)$.

Proof. From the inequality (6.2.14) it is apparent that $A \Rightarrow B$ and $B \Rightarrow P_n \xrightarrow{w} P$. Using (6.2.30) we obtain that B implies $\int \lambda \mathrm{d}(P_n - P) \to 0$ and thus $B \Rightarrow C$. Now, let C hold.

Claim 1. $C \Rightarrow D$.

Proof of claim 1. Choose a sequence $\varepsilon_1 > \varepsilon_2 > \cdots \to 0$ such that $P(d(x, a) = 1/\varepsilon_n) = 0$ for any $n = 1, 2, \ldots$. Then for fixed n

$$\int \lambda(x) I\{d(x, a) \leqslant 1/\varepsilon_n\}(P_k - P)(\mathrm{d}x) \to 0 \quad \text{as } k \to \infty$$

by Billingsley (1968) (see Theorem 5.1). Since $P \in \mathscr{P}_\lambda$, $\omega(\varepsilon_n) := \omega(\varepsilon_n; P; c) \to 0$ as

$n \to \infty$ and hence

$$\limsup_{k\to\infty} \omega_k(\varepsilon_n) \leqslant \limsup_{k\to\infty} \left| \int \lambda(x) I\{d(x, a) > 1/\varepsilon_n\}(P_k - P)(\mathrm{d}x) \right| + \omega(\varepsilon_n)$$

$$\leqslant \limsup_{k\to\infty} \left| \int \lambda(x)(P_k - P)(\mathrm{d}x) \right|$$

$$+ \limsup_{k\to\infty} \left| \int \lambda(x) I\{d(x, a) \leqslant 1/\varepsilon_n\}(P_k - P)(\mathrm{d}x) \right|$$

$$+\omega(\varepsilon_n) \to 0 \quad \text{as } n \to \infty.$$

The last inequality and $P_k \in \mathscr{P}_\lambda$ imply $\lim_{\varepsilon\to 0} \sup_n \omega_n(\varepsilon) = 0$ and hence D holds. The claim is proved.

Claim 2. $D \Rightarrow A$.

Proof of claim 2. By Theorem 6.2.2,

$$\hat{\mu}_c(P_n, P) \leqslant \pi(P_n, P)[4K(1/\varepsilon_n) + \omega_n(1) + \omega(1) + 2k(1)] + 5\omega_n(\varepsilon_n) + 5\omega(\varepsilon_n)$$

where ω_n, and ω are defined as in Claim 1 and moreover $\varepsilon_n > 0$ is such that

$$4K(1/\varepsilon_n) + \sup_{n\geqslant 1} \omega_n(1) + \omega(1) + 2k(1) \leqslant (\pi(P_n, P))^{-1/2}.$$

Hence, using the last two inequalities we obtain

$$\hat{\mu}_c(P_n, P) \leqslant \sqrt{\pi}(P_n, P) + 5 \sup_{n\geqslant 1} \omega_n(\varepsilon_n) + 5\omega(\varepsilon_n)$$

and hence $D \Rightarrow A$, as we claimed. QED

The Kantorovich–Rubinstein functional $\mathring{\mu}_c$ is a metric in $\mathscr{P}_\lambda(U)$ while $\hat{\mu}_c$ is not a metric except for the case $c = d$ (see the discussion in the previous section). The next theorem establishes criterion for $\mathring{\mu}_c$-relative compactness of sets of measures. Recall that a set $\mathscr{A} \subset \mathscr{P}_\lambda$ is said to be *$\mathring{\mu}_c$-relatively compact* if any sequence of measures in $\mathscr{A}$ has a $\mathring{\mu}_c$-convergent subsequence and the limit belongs to $\mathscr{P}_\lambda$. Recall that the set $\mathscr{A} \subset \mathscr{P}(U)$ is *weakly compact* if $\mathscr{A}$ is *π-relatively compact*, i.e. any sequence of measures in $\mathscr{A}$ has a weakly (π-) convergent subsequence.

Theorem 6.3.2. The set $\mathscr{A} \subset \mathscr{P}_\lambda$ is $\mathring{\mu}_c$-relatively compact if and only if $\mathscr{A}$ is weakly compact and

$$\lim_{\varepsilon\to 0} \sup_{P\in\mathscr{A}} \omega(\varepsilon; P; \lambda) = 0. \tag{6.3.1}$$

Proof. 'If' part: If $\mathscr{A}$ is weakly compact, (6.3.1) holds and $\{P_n\}_{n\geqslant 1} \subset \mathscr{A}$, then we can choose a subsequence $\{P_{n'}\} \subset \{P_n\}$ which converges weakly to a probability measure P.

Claim 1. $P \in \mathscr{P}_\lambda$.

Proof of claim 1. Let $0 < \alpha_1 < \alpha_2 < \cdots$, $\lim \alpha_n = \infty$ be such a sequence that $P(d(x, a) = \alpha_n) = 0$ for any $n \geqslant 1$. Then, by Theorem 5.1 (Billingsley 1968) and (6.3.1)

$$\begin{aligned}\int \lambda(x) I\{d(x, a) \leqslant \alpha_{n'}\} P(\mathrm{d}x) &= \lim_{n\to\infty} \int \lambda(x) I\{d(x, a) \leqslant \alpha_{n'}\} P_{n'}(\mathrm{d}x) \\ &\leqslant \liminf_{n\to\infty} \int \lambda(x) P_{n'}(\mathrm{d}x) < \infty\end{aligned}$$

which proves the claim.

Claim 2.

$$\mathring{\mu}_c(P_{n'}, P) \to 0.$$

Proof of claim 2. Using Theorem 6.2.2, Claim 1 and (6.3.1) we have, for any $\delta > 0$,

$$\begin{aligned}\mathring{\mu}_c(P_{n'}, P) \leqslant \hat{\mu}_c(P_{n'}, P) &\leqslant \pi(P_{n'}, P)[4K(1/\varepsilon) + \omega_1(1) + \omega_2(1) + 2K(1)] \\ &\quad + 5 \sup_{n'} \omega(P_{n'}; \varepsilon; \lambda) + \omega(P; \varepsilon; \lambda) \\ &\leqslant \pi(P_{n'}, P)[4K(1/\varepsilon) + \omega_1(1) + \omega_2(1) + 2K(1)] + \delta\end{aligned}$$

if $\varepsilon = \varepsilon(\delta)$ is small enough. Hence, by $\pi(P_{n'}, P) \to 0$, we can choose $N = N(\delta)$ such that $\mathring{\mu}_c(P_{n'}, P) < 2\delta$ for any $n' \geqslant N$ as desired.

Claim 1 and Claim 2 establish the 'if' part of the theorem.

'Only if' part: If $\mathscr{A}$ is $\mathring{\mu}_c$-relatively compact and $\{P_n\} \subset \mathscr{A}$ then there exists a subsequence $\{P_{n'}\} \subset \{P_n\}$ that is convergent w.r.t. $\mathring{\mu}_c$ and let P be the limit. Hence, by Theorem 6.2.3, $\mathring{\mu}_c(P_n, P) \geqslant \beta\pi^2(P_n, P) \to 0$ which demonstrates that the set $\mathscr{A}$ is weakly compact.

Further, if (6.3.1) is not valid, then there exists $\delta > 0$ and a sequence $\{P_n\}$ such that

$$\omega(1/n; P_n; \lambda) > \delta \ \forall n \geqslant 1. \tag{6.3.2}$$

Let $\{P_{n'}\}$ be a $\mathring{\mu}_c$-convergent subsequence of $\{P_n\}$ and $P \in \mathscr{P}_\lambda$ be the corresponding limit. By Theorem 6.2.4, $\omega(1/n'; P_{n'}; \lambda) \leqslant (2\gamma + 2)(\alpha\mathring{\mu}(P_{n'}, P) + \omega(1/n'; P; \lambda)) \to 0$ as $n' \to \infty$ which is in contradiction with (6.3.2). QED

In the light of Theorem 6.3.1 we can now interpret Theorem 6.3.2 as a criterion for $\hat{\mu}_c$-relative compactness of sets of measures in $\mathscr{P}_\lambda$ by simply changing $\mathring{\mu}_c$ with $\hat{\mu}_c$ in the formation of the last theorem.

The well known Prokhorov theorem says that *if (U, d) is a complete s.m.s., then the set of all laws on U is complete w.r.t the Prokhorov metric π* (see, for example, Hennequin and Tortrat, 1965; Dudley, 1989, Theorem 11.5.5). The next theorem is an analog of the Prokhorov theorem for the metric space $(\mathscr{P}_\lambda, \mathring{\mu}_c)$.

Theorem 6.3.3. If (U, d) is a complete s.m.s., then $(\mathscr{P}_\lambda(U), \mathring{\mu}_c)$ is also complete.

Proof. If $\{P_n\}$ is a $\mathring{\mu}_c$-fundamental sequence, then by Theorem 6.2.3, $\{P_n\}$ is also π-fundamental and hence there exists the weak limit $P \in \mathscr{P}(U)$.

Claim 1.
$$P \in \mathscr{P}_\lambda.$$

Proof of claim 1. Let $\varepsilon > 0$ and $\mathring{\mu}(P_n, P_m) \leqslant \varepsilon$ for any $n, m \geqslant n_\varepsilon$. Then, by Theorem 6.2.5 $|\int \lambda(x)(P_n - P_{n_\varepsilon})(dx)| \leqslant \alpha\varepsilon$ for any $n > n_\varepsilon$, hence,

$$\sup_{n \geqslant n_\varepsilon} \int \lambda(x) P_n(dx) < \alpha\varepsilon + \int \lambda(x) P_{n_\varepsilon}(dx) < \infty.$$

Choose the sequence $0 < \alpha_1 < \alpha_2 < \cdots$, $\lim_{k\to\infty} \alpha_k = \infty$, such that $P(d(x, a) = \alpha_k) = 0$ for any $k \geqslant 1$. Then

$$\int \lambda(x) I\{d(x, a) \leqslant \alpha_k\} P(dx) = \lim_{n\to\infty} \int \lambda(x) I\{d(x, a) \leqslant \alpha_k\} P_n(dx)$$
$$\leqslant \liminf_{n\to\infty} \int \lambda(x) P_n(dx) \leqslant \sup_{n \geqslant n_\varepsilon} \int_U \lambda(x) P_n(dx) < \infty.$$

Letting $k \to \infty$ the assertion follows.

Claim 2.
$$\mathring{\mu}_c(P_n, P) \to 0.$$

Proof of claim 2. Since $\mathring{\mu}(P_n, P_{n_\varepsilon}) \leqslant \varepsilon$ for any $n \geqslant n_\varepsilon$, then, by Theorem 6.2.4,

$$\sup_{n \geqslant n_\varepsilon} \omega(\delta; P_n; \lambda) \leqslant 2(\gamma + 1)(\alpha\varepsilon + \omega(2\delta; P_{n_\varepsilon}; \lambda))$$

for any $\delta > 0$. The last inequality and Theorem 6.2.2 yield

$$\begin{aligned} \mathring{\mu}_c(P_n, P) \leqslant \hat{\mu}_c(P_n, P) &\leqslant \pi(P_n, P)[4K(1/\delta) \\ &\quad + \sup_{n \geqslant n_\varepsilon} \omega(1; P_n; \lambda) + \omega(1; P; \lambda) + 2K(1)] \\ &\quad + 10(\gamma + 1)(\alpha\varepsilon + \omega(2\delta; P_{n_\varepsilon}; \lambda) + 5\omega(\delta; P_{n_\varepsilon}; \lambda)) \end{aligned} \tag{6.3.3}$$

for any $n \geqslant \eta_\varepsilon$ and $\delta > 0$. Next, choose $\delta_n = \delta_{n,\varepsilon} > 0$ such that $\delta_n \to 0$ as $n \to \infty$ and

$$4K(1/\delta_n) + \sup_{n \geqslant n_\varepsilon} \omega(1; P_n; \lambda) + \omega(1; P; \lambda) + 2k(1) \leqslant \frac{1}{(\pi(P_n, P))^{1/2}} \tag{6.3.4}$$

Combining (6.3.3) and (6.3.4) we have that $\mathring{\mu}_c(P_n, P) \leqslant \text{const.}\ \varepsilon$ for n large enough, which proves the claim. QED

6.4 $\mathring{\mu}_c$- AND $\hat{\mu}_c$-UNIFORMITY

In the previous section we saw that $\mathring{\mu}_c$ and $\hat{\mu}_c$ induce one and the same convergence in $\mathscr{P}_\lambda$. Here we would like to analyze the uniformity of $\mathring{\mu}_c$- and $\hat{\mu}_c$-convergence. Namely, if for any $P_n, Q_n \in \mathscr{P}_\lambda$, the equivalence

$$\mathring{\mu}_c(P_n, Q_n) \to 0 \Leftrightarrow \hat{\mu}_c(P_n, Q_n) \to 0 \quad n \to \infty \tag{6.4.11}$$

holds. Obviously, $\Leftarrow$ holds, by $\mathring{\mu}_c \leqslant \hat{\mu}_c$. So, if

$$\hat{\mu}_c(P, Q) \leqslant \phi(\mathring{\mu}_c(P, Q)) \;\; P, Q \in \mathscr{P}_\lambda \tag{6.4.2}$$

for a continuous non-decreasing function, $\phi(0) = 0$, then (6.4.1) holds.

Remark 6.4.1. Given two metrics, say μ and ν, in the space of measures, the equivalence of μ- and ν-convergence does not imply the existence of a continuous non-decreasing function ϕ vanishing at 0 and such that $\mu \leqslant \phi(\nu)$. For example, both the Lévy metric $\mathbf{L}$ (see (4.1.3)) and the Prokhorov metric $\boldsymbol{\pi}$ (see (3.2.18)) metrize the weak convergence in the space $\mathscr{P}(\mathbb{R})$. Suppose there exists ϕ such that

$$\boldsymbol{\pi}(X, Y) \leqslant \phi(\mathbf{L}(X, Y)) \tag{6.4.3}$$

for any real-valued r.v.s X and Y. (Recall our notation $\mu(X, Y) := \mu(\Pr_X, \Pr_Y)$ for any metric μ in the space of measures.) Then, by (4.1.4) and (3.2.23),

$$\mathbf{L}(X/\lambda, Y/\lambda) = \mathbf{L}_\lambda(X, Y) \to \boldsymbol{\rho}(X, Y) \quad \text{as } \lambda \to 0 \tag{6.4.4}$$

and

$$\boldsymbol{\pi}(X/\lambda, Y/\lambda) = \boldsymbol{\pi}_\lambda(X, Y) \to \boldsymbol{\sigma}(X, Y) \quad \text{as } \lambda \to 0 \tag{6.4.5}$$

where $\boldsymbol{\rho}$ is the Kolmogorov metric (see (4.1.6)) and $\boldsymbol{\sigma}$ is the total variation metric (3.2.13). Thus, (6.4.3) to (6.4.5) imply that $\boldsymbol{\sigma}(X, Y) \leqslant \phi(\boldsymbol{\rho}(X, Y))$. The last inequality simply is, however, not true because in general $\boldsymbol{\rho}$-convergence does not yield $\boldsymbol{\sigma}$-convergence. (For example, if X_n is a random variable taking values k/n, $k = 1, \ldots, n$ with probability $1/n$, then $\boldsymbol{\rho}(X_n, Y) \to 0$ where Y is a (0,1)-uniformly distributed random variable. On the other hand, $\boldsymbol{\sigma}(X_n, Y) = 1$.)

We are going to prove (6.4.2) for the special but important case when $\mathring{\mu}_c$ is the Fortet–Mourier metric on $\mathscr{P}_\lambda(\mathbb{R})$, i.e. $\mathring{\mu}_c(P, Q) = \zeta(P, Q; \mathscr{G}^p)$ (see (4.3.34)); in other words, for any $P, Q \in \mathscr{P}_\lambda$,

$$\mathring{\mu}_c(P, Q) = \sup\left\{\int f\,\mathrm{d}(P - Q): f: \mathbb{R} \to \mathbb{R}, |f(x) - f(y)| \leqslant c(x, y)\ \forall x, y \in \mathbb{R}\right\}$$

where

$$c(x, y) = |x - y| \max(1, |x|^{p-1}, |y|^{p-1}) \quad p \geqslant 1. \tag{6.4.6}$$

Since $\lambda(x) := 2\max(|x|, |x|^p)$, then $\mathscr{P}_\lambda(\mathbb{R})$ is the space of all laws on $\mathbb{R}$, with finite pth absolute moment.

Theorem 6.4.1. If c is given by (6.4.6) then

$$\hat{\mu}_c(P, Q) \leqslant p\mathring{\mu}_c(P, Q) \quad \forall P, Q \in \mathscr{P}_\lambda(\mathbb{R}). \tag{6.4.7}$$

Proof. Denote $h(t) = \max(1, |t|^{p-1})$, $t \in \mathbb{R}$ and $H(x) = \int_0^x h(t)\,\mathrm{d}t$, $x \in \mathbb{R}$. Let X and Y be real-valued random variables on a non-atomic probability space $(\Omega, \mathscr{A}, \Pr)$ with distributions P and Q respectively. Theorem 5.4.1 gives us explicit representation of $\mathring{\mu}_c$. Namely

$$\mathring{\mu}_c(P, Q) = \int_{-\infty}^{\infty} h(t)|F_X(t) - F_Y(t)|\,\mathrm{d}t \tag{6.4.8}$$

and thus,

$$\mathring{\mu}_c(P, Q) = \int_{-\infty}^{\infty} |F_{H(X)}(x) - F_{H(Y)}(x)|\,\mathrm{d}x. \tag{6.4.9}$$

Claim 1. Let X and Y be real-valued r.v.s with distributions P and Q, respectively. Then

$$\mathring{\mu}_c(P, Q) = \inf\{\mathbb{E}|H(\tilde{X}) - H(\tilde{Y})|: F_{\tilde{X}} = F_X, F_{\tilde{Y}} = F_Y\}. \tag{6.4.10}$$

Proof of claim 1. Using the equality $\hat{\mu}_d = \mathring{\mu}_d$ (see (6.1.3)) and (5.4.5) with $H(t) = t$ we have that

$$\mathring{\mu}_d(F, G) = \hat{\mu}_d(F, G) = \inf\{\mathbb{E}|X' - Y'|: F_{X'} = F, F_{Y'} = G\} = \int_{-\infty}^{\infty} |F(x) - G(x)|\,\mathrm{d}x \tag{6.4.11}$$

for any d.f.s F and G. Hence, by (6.4.9)

$$\begin{aligned}\mathring{\mu}_c(P, Q) &= \inf\{\mathbb{E}|X' - Y'|: F_{X'} = F_{H(X)}, F_{Y'} = F_{H(Y)}\}\\ &= \inf\{\mathbb{E}|H(\tilde{X}) - H(\tilde{Y})|: F_{\tilde{X}} = F_{X'}, F_{\tilde{Y}} = F_Y\}\end{aligned}$$

which proves the claim.

Next we use Theorem 2.5.2 which claims that on a non-atomic probability space, the class of all joint distributions $\Pr_{X,Y}$ coincides with the class of all probability Borel measures on $\mathbb{R}^2$. This implies

$$\hat{\mu}_c(P, Q) = \inf\{\mathbb{E}c(\tilde{X}, \tilde{Y}) \colon F_{\tilde{X}} = F_X, F_{\tilde{Y}} = F_Y\}. \tag{6.4.12}$$

Claim 2. For any $x, y \in \mathbb{R}$, $c(x, y) \leqslant p|H(x) - H(y)|$.

Proof of claim 2. (a) Let $y > x > 0$. Then

$$\begin{aligned} c(x, y) &= (y - x)h(y) = yh(y) - xh(y) \leqslant yh(y) - xh(x) \\ &\leqslant (H(y) - H(x)) \sup_{y > x > 0} \frac{yh(y) - xh(x)}{H(y) - H(x)}. \end{aligned}$$

Since $H(t)$ is a strictly increasing continuous function, then

$$B := \sup_{y > x > 0} \frac{yh(y) - xh(x)}{H(y) - H(x)} = \sup_{t > s > 0} \frac{f(t) - f(s)}{t - s}$$

where $f(t) := H^{-1}(t)h(H^{-1}(t))$ and H^{-1} is the function inverse to H, hence, $B = \operatorname{ess\,sup}_t |f'(t)| \leqslant p$.

(b) Let $y > 0 > x > -y$. Then $c(x, y) = |x - y|h(y) = (y + (-x))h(y) = yh(y) + (-x)h(|x|) + ((-x)h(y) - (-x)h(|x|)) \leqslant yh(y) + (-x)h(|x|)$. Since

$$th(t) = \begin{cases} t & \text{if } t \leqslant 1 \\ t^p & \text{if } t \geqslant 1 \end{cases} \qquad H(t) = \begin{cases} t & \text{if } 0 < t \leqslant 1 \\ \dfrac{p - 1}{p} + \dfrac{1}{p} t^p & \text{if } t \geqslant 1 \end{cases}$$

then $yh(y) + (-x)h(|x|) \leqslant p(H(y) + H(-x)) = p(H(y) - H(x))$. By symmetry, the other cases are reduced to (a) or (b). The claim is shown. Now, (6.4.7) is a consequence of Claims 1, 2, and (6.4.12). QED

6.5 GENERALIZED KANTOROVICH AND KANTOROVICH–RUBINSTEIN FUNCTIONALS

In this section we shall consider a generalization of the Kantorovich type functionals $\hat{\mu}_c$ and $\mathring{\mu}_c$ (see (5.1.16) and (5.1.17)).

Let $U = (U, d)$ be a s.m.s. and $\mathscr{M}(U \times U)$ be the space of all non-negative Borel measures on the Cartesian product $U \times U$. For any probability measures P_1 and P_2 define the sets $\mathscr{P}^{(P_1, P_2)}$ and $\mathscr{Q}^{(P_1, P_2)}$ as in Section 5.1 (cf. (5.1.2) and (5.1.13)).

Let $\Lambda \colon \mathscr{M}(U \times U) \to [0, \infty]$ satisfy the conditions

(i) $$\Lambda(\alpha P) = \alpha \Lambda(P) \quad \forall \alpha \geqslant 0.$$

(ii) $$\Lambda(P + Q) \leqslant \Lambda(P) + \Lambda(Q) \quad \forall P \text{ and } Q \text{ in } \mathscr{M}(U \times U).$$

We introduce the *generalized Kantorovich functional*

$$\hat{\Lambda}(P_1, P_2) := \inf\{\Lambda(P)\colon P \in \mathscr{P}^{(P_1, P_2)}\} \tag{6.5.1}$$

and the *generalized Kantorovich–Rubinstein functional*

$$\mathring{\Lambda}(P_1, P_2) := \inf\{\Lambda(P)\colon P \in \mathscr{Q}^{(P_1, P_2)}\}. \tag{6.5.2}$$

Example 6.5.1. The Kantorovich metric (see Example 3.2.2)

$$\ell_1(P_1, P_2) := \sup\left\{\left|\int f\,\mathrm{d}(P_1 - P_2)\right|\colon f\colon U \to \mathbb{R}, \left|f(x) - f(y)\right| \leqslant d(x, y),\ x, y \in U\right\}$$

in the space of measures P with finite 'first moment', $\int d(x, a)P(\mathrm{d}x) < \infty$, has the dual representations $\ell_1(P_1, P_2) = \mathring{\Lambda}(P_1\ P_2) = \hat{\Lambda}(P_1, P_2)$, where

$$\Lambda(P) := \Lambda_1(P) := \int_{U \times U} d(x, y)P(\mathrm{d}x, \mathrm{d}y). \tag{6.5.3}$$

Example 6.5.2. Let $U = \mathbb{R}$, $d(x, y) = |x - y|$. Then

$$\ell_1(P_1, P_2) = \int_{\mathbb{R}} |F_1(t) - F_2(t)|\,\mathrm{d}t$$

where F_i is the d.f. of P_i and

$$\begin{aligned}\Lambda_1(P) &= \int_{\mathbb{R}} (\Pr(X \leqslant t < Y) + \Pr(Y \leqslant t < X))\,\mathrm{d}t \\ &= \int_{\mathbb{R}} \Pr(X \leqslant t) + \Pr(Y \leqslant t) - 2\Pr(\max(X, Y) \leqslant t)\,\mathrm{d}t \\ &= \mathbb{E}(2\max(X, Y) - X - Y) = \mathbb{E}|X - Y|\end{aligned}$$

for r.v.s X and Y with $\Pr_{X,Y} = P$. We generalize (6.5.3) as follows: for any $1 \leqslant p \leqslant \infty$, define

$$\Lambda(P) := \Lambda_p(P) := \begin{cases} \left\{\int_{\mathbb{R}}\left[\int_{\mathbb{R}^2} c_t(x, y)P(\mathrm{d}x, \mathrm{d}y)\right]^p \lambda(\mathrm{d}t)\right\}^{1/p} & 1 \leqslant p < \infty \\ \operatorname{ess\,sup}_\lambda \int_{\mathbb{R}^2} c_t(x, y)P(\mathrm{d}x, \mathrm{d}y) & \\ \qquad := \inf\left\{\varepsilon > 0\colon \lambda\left\{t\colon \int_{\mathbb{R}^2} c_t\,\mathrm{d}P > \varepsilon\right\} = 0\right\} & p = \infty \end{cases} \tag{6.5.4}$$

where $c_t(t \in \mathbb{R})$ is the following semimetric in $\mathbb{R}$

$$c_t(x, y) := I\{x \leqslant t \leqslant y\} + I\{y \leqslant t < x\}\ \forall x, y, t \in \mathbb{R} \tag{6.5.5}$$

and $\lambda(\cdot)$ is a non-negative measure on $\mathbb{R}$. In the space $\mathfrak{X} = \mathfrak{X}(\mathbb{R})$ of all real-valued r.v.s on a non-atomic probability space $(\Omega, \mathscr{A}, \Pr)$, the minimal metric with respect to Λ is given by

$$\hat{\Lambda}_p(P_1, P_2) = \begin{cases} \inf\left\{\left[\int_{\mathbb{R}} \phi_t^p(X, Y)\lambda(\mathrm{d}t)\right]^{1/p} : X, Y \in \mathfrak{X}, \Pr_X = P_1, \Pr_Y = P_2\right\} & 1 \leqslant p < \infty \\ \inf\left\{\sup_{t\in\mathbb{R}} \phi_t(X, Y) : X, Y \in \mathfrak{X}, \Pr_X = P_1, \Pr_Y = P_2\right\} & p = \infty. \end{cases} \tag{6.5.6}$$

Similarly, the minimal norm with respect to Λ is

$$\mathring{\Lambda}_p(P_1, P_2) = \begin{cases} \inf\left\{\alpha\left[\int_{\mathbb{R}} \phi_t^p(X, Y)\lambda(\mathrm{d}t)\right]^{1/p} : \alpha > 0, \quad X, Y \in \mathfrak{X}, \quad \alpha(\Pr_X - \Pr_Y) = P_1 - P_2\right\} & \text{if } p < \infty \\ \inf\left\{\alpha \sup_\lambda \phi_t(X, Y) : \alpha > 0, \quad X, Y \in \mathfrak{X}, \quad \alpha(\Pr_X - \Pr_Y) = P_1 - P_2\right\} & \text{if } p = \infty \end{cases} \tag{6.5.7}$$

where in (6.5.6) and (6.5.7)

$$\phi_t(X, Y) := \Pr(X \leqslant t < Y) + \Pr(Y \leqslant t < X). \tag{6.5.8}$$

The next theorem gives the explicit form of $\hat{\Lambda}_p$ and $\mathring{\Lambda}_p$.

Theorem 6.5.1. Let F_i be the d.f. of P_i $(i = 1, 2)$. Then

$$\hat{\Lambda}_p(P_1, P_2) = \mathring{\Lambda}_p(P_1, P_2) = \boldsymbol{\lambda}_p(F_1, F_2) \tag{6.5.9}$$

where

$$\lambda_p(F_1, F_2) = \begin{cases} \left(\int_{\mathbb{R}} |F_1(t) - F_2(t)|^p \lambda(\mathrm{d}t)\right)^{1/p} & 1 \leqslant p < \infty \\ \operatorname{ess\,sup}_\lambda |F_1 - F_2| = \inf\{\varepsilon > 0 : \lambda(t : |F_1(t) - F_2(t)| > \varepsilon) = 0\} & p = \infty. \end{cases} \tag{6.5.10}$$

Claim 1. $\boldsymbol{\lambda}_p(F_1, F_2) \leqslant \mathring{\Lambda}_p(P_1, P_2)$.

Proof of claim 1. Let $P \in \mathscr{Q}^{(P_1, P_2)}$. Then in view of Remark 2.5.2, there exist $\alpha > 0$, $X \in \mathfrak{X}$, $Y \in \mathfrak{X}$ such that $\alpha \Pr_{X,Y} = P$ and $\alpha(F_X - F_Y) = F_1 - F_2$, thus

$$\begin{aligned} |F_1(x) - F_2(x)| &= \alpha |F_X(t) - F_Y(t)| \\ &= \alpha[\max(F_X(t) - F_Y(t), 0) + \max(F_Y(t) - F_X(t), 0)] \\ &\leqslant \alpha \phi_t(X, Y). \end{aligned} \tag{6.5.11}$$

By (6.5.7) and (6.5.11), it follows that $\lambda_p(F_1, F_2) \leqslant \mathring{\Lambda}_p(P_1, P_2)$ as desired.

Further

$$\mathring{\Lambda}_p(P_1, P_2) \leqslant \hat{\Lambda}_p(P_1, P_2) \tag{6.5.12}$$

by the representations (6.5.6) and (6.5.7).

Claim 2.

$$\hat{\Lambda}_p(P_1, P_2) \leqslant \lambda_p(F_1, F_2).$$

Proof of claim 2. Let $\tilde{X} := F_1^{-1}(V)$, $\tilde{Y} := F_2^{-1}(V)$ where F_i^{-1} is the generalized inverse to the d.f. F_i (cf. (3.2.16)) and V is a (0,1)-uniformly distributed r.v. Then $F_{\tilde{X},\tilde{Y}}(t, s) = \min(F_1(t), F_2(s))$ for all $t, s \in \mathbb{R}$. Hence, $\phi_t(\tilde{X}, \tilde{Y}) = |F_1(t) - F_2(t)|$ which proves the claim by using (6.5.6) and (6.5.7).

Combining Claims 1, 2 and (6.5.12), we obtain (6.5.9). QED

Problem 6.5.1. In general dual and explicit solutions of $\hat{\Lambda}_p$ and $\mathring{\Lambda}_p$ in (6.5.1) and (6.5.2) are not known.

CHAPTER 7

K-minimal Metrics

7.1 DEFINITION; GENERAL PROPERTIES

As we have seen in the previous two chapters the notion of minimal distance

$$\hat{\mu}(P_1, P_2) = \inf\{\mu(P): P \in \mathscr{P}(U^2),\ T_i P = P_i,\ i = 1, 2\} \quad P_1, P_2 \in \mathscr{P}(U) \quad (7.1.1)$$

represents the main relationship between compound and simple distances (cf. the general discussion in Section 3.2). In view of the multi-dimensional Kantorovich problem (cf. Section 5.1, VI) we have been interested in the n-dimensional analog of the notion of minimal metrics. Namely, we have defined the following distance between n-dimensional vectors of probability measures (cf. (5.2.15))

$$\mathfrak{R}(\tilde{P}, \tilde{Q}) = \inf\left\{\int_{U^n \times U^n} \Delta(x, y)P(\mathrm{d}x, \mathrm{d}y): P \in \mathfrak{P}(\tilde{P}, \tilde{Q})\right\} \quad (7.1.2)$$

where $\tilde{P} = (P_1, \ldots, P_n)$, $\tilde{Q} = (Q_1, \ldots, Q_n)$, $P_i, Q_i \in \mathscr{P}(U)$, $\Delta(x, y)$ is a distance in the Cartesian product U^n and $\mathfrak{P}(\tilde{P}, \tilde{Q})$ is the space of all probability measures on U^{2n} with fixed one-dimensional marginals $P_1, \ldots, P_n, Q_1, \ldots, Q_n$. In the sixties, H. G. Kellerer investigated the multi-dimensional marginal problem. His results on this topic were the major source for the famous Strassen (1965) work on minimal probabilistic functionals. In this section we shall study the properties of metrics in the space of vectors $\tilde{\mathscr{P}}$ which have representation similar to that of $\mathfrak{R}$.

Definition 7.1.1. Let μ be a p. distance in $\mathscr{P}_2(U^n)$ (U is a s.m.s.). For any two vectors $\tilde{P}_i = (P_i^{(1)}, \ldots, P_i^{(n)})$, $i = 1, 2$ of probability measures $P_i^{(j)} \in \mathscr{P}_1(U)$ define the *K-minimal distance*

$$\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2) = \inf\{\mu(P): P \in \mathfrak{P}(\tilde{P}_1, \tilde{P}_2)\} \quad (7.1.3)$$

where $\mathfrak{P}(\tilde{P}_1, \tilde{P}_2) = \{P \in \mathscr{P}_2(U^N): T_j P = P_1^{(j)},\ T_{j+n} P = P_2^{(j)}, j = 1, \ldots, n\}$.

Obviously $\overset{1}{\mu} = \hat{\mu}$. One of the main reasons to study K-minimal metrics is based on the simple observation that in most cases the minimal metric between the product measures $\hat{\mu}(P_1^{(1)} \times \cdots \times P_1^{(n)},\ P_2^{(1)} \times \cdots \times P_2^{(n)})$ coincides with $\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2)$. Surprisingly it is much easier to find explicit representations for $\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2)(\tilde{P}_i \in \mathscr{P}(U)^n)$ than for $\hat{\mu}(P_1, P_2)$ $(P_i \in \mathscr{P}(U^n))$. Some general relations between compound, minimal and K-minimal distances are given in the next four theorems. Recall that for any $P \in \mathscr{P}(U^k)$, $k \geqslant 2$, the law $T_{\alpha_1,\ldots,\alpha_m} P \in \mathscr{P}(U^m)$ $(1 \leqslant m \leqslant k)$ is the marginal distribution of P on the coordinates $\alpha_1 < \alpha_2 < \cdots < \alpha_m$.

Theorem 7.1.1. Let ψ be a right semi-continuous (r.s.c.) function on $(0, \infty)$ and $\phi(t_1, \ldots, t_n)$ non-decreasing function in each argument $t_i \geqslant 0$, $i = 1, \ldots, n$. Suppose that a p. distance μ on $\mathscr{P}_2(U^n)$ and p. distances $\mu_1, \ldots, \mu_n$ on $\mathscr{P}_2(U)$ satisfy the following inequality: for any $P \in \mathfrak{P}(\tilde{P}_1, \tilde{P}_2)$

$$\psi(\mu(P)) \geqslant \phi(\mu(T_{1,n+1}P), \mu_2(T_{2,n+2}P), \ldots, \mu_n(T_{n,2n}P)). \tag{7.1.4}$$

Then

$$\psi(\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2)) \geqslant \phi(\hat{\mu}(P_1^{(1)}, P_2^{(1)}), \ldots, \hat{\mu}(P_1^{(n)}, P_2^{(n)})).$$

Proof. Given $\varepsilon > 0$ there exists $P^{(\varepsilon)} \in \mathfrak{P}(\tilde{P}_1, \tilde{P}_2)$ such that $|\mathscr{D}_\varepsilon| < \varepsilon$, where $\mathscr{D}_\varepsilon = \psi(\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2)) - \psi(\mu(P^{(\varepsilon)}))$. Thus by (7.1.4),

$$\psi(\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2)) = \psi(\mu(P^{(\varepsilon)})) + \mathscr{D}_\varepsilon \geqslant \phi(\hat{\mu}_1(P_1^{(1)}, P_2^{(2)}), \ldots, \hat{\mu}_n(P_1^{(n)}, P_2^{(n)})) - \varepsilon. \quad \text{QED}$$

Theorem 7.1.2. Let $\mu_1, \ldots, \mu_k$, $\nu_1, \ldots, \nu_k$ be probability distances on $\mathscr{P}_2(U^n)$ and suppose

$$\psi(\mu_1(P_1), \ldots, \mu_k(P_k)) \geqslant \phi(\nu_1(P_1), \ldots, \nu_k(P_k)),\ P_i \in \mathscr{P}_2(U^n)$$

where ϕ is non-decreasing in each argument and ψ is an r.s.c. on $\mathbb{R}_t^n$. Then

$$\psi(\overset{n}{\mu}_1, \ldots, \overset{n}{\mu}_k) \geqslant \phi(\overset{n}{\nu}_1, \ldots, \overset{n}{\nu}_k).$$

The proof is straightforward.

In the following $P_1 \times \cdots \times P_n$ denotes the product measure generated by $P_1, \ldots, P_n$. The next theorem describes conditions providing an equality between $\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2)$ and $\mu(P_1^{(1)} \times \cdots \times P_1^{(n)}, P_2^{(1)} \times \cdots \times P_2^{(n)})$.

Theorem 7.1.3. Suppose that a p. distance μ on $\mathscr{P}_2(U^n)$ and p. distances $\mu_1, \ldots, \mu_n$ on $\mathscr{P}_2(U)$ satisfy the equality

$$\mu(P) = \phi(\mu_1(T_{1,n+1}P), \ldots, \mu_n(T_{n,2n}P)) \tag{7.1.5}$$

where ϕ is an r.s.c. function, non-decreasing in each argument. Then for any vectors of measures $\tilde{P}_1, \tilde{P}_2 \in \mathscr{P}_1(U)^n$

$$\begin{aligned}\overset{n}{\mu}(\tilde{P}_1, \tilde{P}_2) &= \hat{\mu}(P_1^{(1)} \times \cdots \times P_1^{(n)}, P_2^{(1)} \times \cdots \times P_2^{(n)}) \\ &= \phi(\hat{\mu}_1(P_1^{(1)}, P_2^{(1)}), \ldots, \hat{\mu}_n(P_1^{(n)}, P_2^{(n)})).\end{aligned} \tag{7.1.6}$$

Proof. Given $\varepsilon > 0$, choose $\delta_\varepsilon \in (0, \varepsilon)$ and $P^{(\varepsilon)} \in \mathfrak{P}(\tilde{P}_1, \tilde{P}_2)$ such that

$$\hat{\hat{\mu}}(\tilde{P}_1, \tilde{P}_2) = \mu(P^{(\varepsilon)}) - \delta_\varepsilon = \phi(\mu_1(T_{1,n+1}P^{(\varepsilon)}), \ldots, \mu_n(T_{n,2n}P^{(\varepsilon)})) - \delta_\varepsilon. \quad (7.1.7)$$

Take

$$Q^{(\varepsilon)} = T_{1,n+1}P^{(\varepsilon)} \times \cdots \times T_{n,2n}P^{(\varepsilon)}.$$

Then,

$$T_{1,\ldots,n}Q^{(\varepsilon)} = P_1^{(1)} \times \cdots \times P_1^{(n)}, T_{n+1,\ldots,2n}Q^{(\varepsilon)} = P_2^{(1)} \times \cdots \times P_2^{(n)}$$

and by (7.1.5), $\mu(P^{(\varepsilon)}) = \mu(Q^{(\varepsilon)})$, which together with (7.1.7) implies

$$\hat{\hat{\mu}}(\tilde{P}_1, \tilde{P}_2) = \mu(Q^{(\varepsilon)}) - \delta_\varepsilon \geqslant \hat{\mu}(P_1^{(1)} \times \cdots \times P_1^{(n)}, P_2^{(1)} \times \cdots \times P_2^{(n)}) - \delta_\varepsilon$$

and

$$\hat{\hat{\mu}}(\tilde{P}_1, \tilde{P}_2) \geqslant \phi(\hat{\mu}_1(P_1^{(1)}, P_2^{(1)}), \ldots, \hat{\mu}_n(P_1^{(n)}, P_2^{(n)})) - \delta_\varepsilon.$$

On the other hand, $\hat{\hat{\mu}}(\tilde{P}_1, \tilde{P}_2) \leqslant \hat{\mu}(P_1^{(1)} \times \cdots \times P_1^{(n)}, P_2^{(1)} \times \cdots \times P_2^{(n)})$ and if

$$D_\varepsilon := \phi(\hat{\mu}_1(P_1^{(1)}, P_2^{(1)}), \ldots, \hat{\mu}_n(P_1^{(n)}, P_2^{(n)})) - \phi(\mu_1(T_{1,n+1}P^{(\varepsilon)}), \ldots, \mu_n(T_{n,2n}P^{(\varepsilon)}))$$

then taking into account (7.1.5) we get

$$\phi(\hat{\mu}_1(P_1^{(1)}, P_2^{(1)}), \ldots, \hat{\mu}_n(P_1^{(n)}, P_2^{(n)})) = \mu(P^{(\varepsilon)}) + D_\varepsilon \geqslant \hat{\hat{\mu}}(\tilde{P}_1, \tilde{P}_2) + D_\varepsilon$$

where $D_\varepsilon \to 0$ as $\varepsilon \to 0$. QED

In terms of distributions of random variables the last theorem can be rewritten as follows: Let $X_i = (X_i^{(1)}, \ldots, X_i^{(n)})$ $(i = 1, 2)$ be two vectors in $\mathfrak{X}(U^n)$ with independent components and suppose that the compound metric μ in $\mathfrak{X}(U^n)$ has the following representation:

$$\mu(X, Y) = \phi(\mu_1(X^{(1)}, Y^{(1)}), \ldots, \mu_n(X^{(n)}, Y^{(n)})) \quad X, Y \in \mathfrak{X}(U^n) \quad (7.1.8)$$

where ϕ is defined as in Theorem 7.1.3. Then

$$\hat{\hat{\mu}}(X_1, X_2) = \hat{\mu}(X_1, X_2) = \phi(\hat{\mu}(X_1^{(1)}, X_2^{(1)}), \ldots, \hat{\mu}_n(X_1^{(n)}, X_2^{(n)})). \quad (7.1.9)$$

Remark 7.1.1. The implication (7.1.8) $\Rightarrow$ (7.1.9) is often used in problems of estimating the closeness between two random vectors with independent components. In many cases working with compound distances is more convenient than working with simple ones. Namely, when we are seeking inequalities, estimators and so on, then, considering all random variables on a common probability space, we deal with simple operations (as, e.g., sums and maximums) in the space of r.v.s. However, considering inequalities between simple metrics and distances, we must evaluate functionals in the space of distributions involving, e.g., convolutions or product of distribution functions. Among many specialists this simple idea is referred to as 'the method of one probability space'.

A particular case of Theorem 7.1.3 asserts that the equality

$$\mu(X_1, X_2) = \phi(\mu_1(X_1, X_2)) \quad X_1, X_2 \in \mathfrak{X}(U) \tag{7.1.10}$$

yields

$$\hat{\mu}(X_1, X_2) = \phi(\hat{\mu}_1(X_1, X_2)) \tag{7.1.11}$$

for any r.s.c. non-decreasing function ϕ on $[0, \infty)$. The next theorem is a variant of the implication (7.1.10) $\Rightarrow$ (7.1.11) and essentially says that if

$$\mu_\phi(X_1, X_2) = \mu(\phi(X_1), \phi(X_2)) \tag{7.1.12}$$

then

$$\hat{\mu}_\phi = (\hat{\mu})_\phi \tag{7.1.13}$$

for any measurable function ϕ. More precisely, let $(U, \mathscr{A})$, $(V, \mathscr{B})$ be measurable spaces and $\phi: U \to V$ be a measurable function. Let μ be a p. distance on $\mathscr{P}(V^2)$; then define

$$\mu_\phi: \mathscr{P}(V^2) \to [0, \infty] \quad \mu_\phi(Q) := \mu(Q_{(\phi,\phi)}) \quad Q \in \mathscr{P}(V^2) \tag{7.1.14}$$

where $Q_{(\phi,\phi)}$ is the image of Q under the transformation $(\phi, \phi)(x, y) = (\phi(x), \phi(y))$. Similarly, if ν is a simple distance $\nu_\phi(P_1, P_2) = \nu(P_{1\phi}, P_{2\phi})$, where $\mathscr{P}_{i,\phi}(A) = P_i(\phi^{-1}(A))$.

It is easy to see that μ_ϕ defines a p. semidistance on $P(U^2)$. In terms of random variables the above definition can also be written in the following way: $\mu_\phi(X, Y) = \mu(\phi(X), \phi(Y))$.

Definition 7.1.2. A measurable space $(U, \mathscr{A})$ is called a *Borel space*, if there exists a Borel subset $B \in \mathscr{B}_1 = \mathscr{B}(\mathbb{R}^1)$ and a Borel isomorphism $\psi: (U, \mathscr{A}) \to (B, B \cap \mathscr{B}_1)$, i.e., if U and B are Borel-isomorphic (cf. Definition 2.4.6).

Theorem 7.1.4. Let $(U, \mathscr{A})$ be a Borel-space, $(V, \mathscr{B})$ a measurable space such that $\{v\} \in \mathscr{B}$ for all $v \in V$ and $\phi: U \to V$ be a measurable mapping. Let $\hat{\mu}$, $\hat{\mu}_\phi$ denote the minimal distance corresponding to μ, μ_ϕ. Then

$$\hat{\mu}_\phi(P_1, P_2) = \hat{\mu}(P_{1\phi}, P_{2\phi}) \tag{7.1.15}$$

for all $P_1, P_2 \in \mathscr{P}_1(U)$.

Proof. We need an auxiliary result on the construction of random variables. Let $(\Omega, \mathscr{E}, \Pr)$ be a probability space and let $(S, Z): \Omega \to V \times \mathbb{R}$ be a pair of independent random variables, where S is a V-valued r.v. and Z is uniformly distributed on [0, 1]. Let P be a probability measure on $(U, \mathscr{A})$ such that $P \circ \phi^{-1}$ coincides with the law of S, $\Pr_S$.

Lemma 7.1.1. There exists a U-valued r.v. X such that

$$\Pr_X = P \quad \text{and} \quad \phi(X) = S \text{ a.e.} \tag{7.1.16}$$

Proof. We start with the special case $(U, \mathscr{A}) = (\mathbb{R}, \mathscr{B}_1)$. Let $I\colon \mathbb{R} \to \mathbb{R}$ denote the identity, $I(x) = x$, and define the set $(P_s)_{s\in V}$ of regular conditional distributions $P_s := P_{I|\phi=s}$, $s \in V$. Let F_s be the distribution function of P_s, $s \in V$. Then it is easy to check that

$$F\colon V \times \mathbb{R} \to [0, 1],\ F(s, x) := F_s(x) \tag{7.1.17}$$

is product-measurable. For $s \in V$, let $F_s^{-1}(x) := \sup\{y\colon F_s(y) < x\}$, $x \in (0, 1)$ be the generalized inverse of F_s and define the random variable $X := F_S^{-1}(Z)$. For any $A \in \mathscr{A} = \mathscr{B}_1$, we have

$$\Pr(X \in A) = \int_V \Pr_{X|S=s}(A)\Pr_S(ds).$$

For the regular conditional distributions we obtain, by the independence of S and Z,

$$\Pr_{X|S=s} = \Pr_{F_s^{-1}(Z)|S=s} = \Pr_{F_s^{-1}(Z)}.$$

Since $\Pr_{F_s^{-1}(Z)} = P_s = P_{I|\phi=s}$ then $\Pr(X \in A) = \int P_{I|\phi=s}(A)P \circ \phi^{-1}(ds) = P(A)$. Thus, the law of X is P. To show that $\phi X = S$ a.e., observe that, by $\Pr_S = P \circ \phi^{-1}$ and $\Pr_{X|S=s} = P_{I|\phi=s}$, we have

$$\begin{aligned}\Pr(\phi(X) = S) &= \int_V \Pr_{X|S=s}(x\colon \phi(x) = s)\Pr_S(ds)\\ &= \int_V P_{I|\phi=s}(x\colon \phi(s) = s)P \circ \phi^{-1}(ds) = 1.\end{aligned}$$

Let now $(U, \mathscr{A})$ be a Borel space. Let $\psi\colon (U, \mathscr{A}) \to (B\colon B_n \cap \mathscr{B}_1)$, $B \in \mathscr{B}_1$, be a measure isomorphism and define $P' := P \circ \psi^{-1}$, $\phi' := \phi \circ \psi^{-1}$. By part one of this proof there exists a random variable $X'\colon \Omega \to B$ such that $\Pr_{X'} = P'$ and $\phi' \circ X' = S$ a.e., thus, $\Pr_X = P$ and $\phi \circ X = S$ a.e. where $X = \psi^{-1} \circ X'$ as desired in (7.1.16). QED

Now let $\mathscr{P}^{(P_1,P_2)}$ be the set of all probability measures on $U \times U$ with marginals P_1, P_2. Then

$$\{Q_{(\phi,\phi)}\colon Q \in \mathscr{P}^{(P_1,P_2)}\} \subset \mathscr{P}^{(P_{1\phi},P_{2\phi})}$$

and hence

$$\hat{\mu}_\phi(P_1, P_2) = \inf\{\mu(P_{\phi,\phi})\colon P \in \mathscr{P}^{(P_1,P_2)}\} \geqslant \inf\{\mu(P)\colon P \in \mathscr{P}^{(P_{1\phi},P_{2\phi})}\} = \hat{\mu}(P_{1\phi}, P_{2\phi}).$$

On the other hand, suppose $P \in \mathscr{P}^{(P_{1\phi}, P_{2\phi})}$. Let $(\Omega, \mathscr{E}, \Pr)$ be a probability space with V-valued random variables S, S' such that $\Pr_{(S,S')} = P$ and rich enough to contain a further random variable $Z: M \to [0, 1]$ uniformly distributed on $[0, 1]$ and independent of S, S'. By Lemma 7.1.1 there exist U-valued r.v.s X and Y such that $\Pr_X = P_1$, $\Pr_Y = P_2$ and $\phi(X) = S$, $\phi(Y) = S'$ a.e. Therefore, $\mu(P) = \mu(\phi \circ X, \phi \circ Y) = \mu_\phi(X, Y)$, implying that

$$\begin{aligned}\hat{\mu}_\phi(P_1, P_2) &= \inf\{\mu_\phi(X, Y): P_X = P_1, P_Y = P_2\} \\ &\leqslant \inf\{\mu(S, S'): \Pr_S = P_{1\phi}, P_{S'} = P_{2\phi}\} = \hat{\mu}(P_{1\phi}, P_{2\phi}). \qquad \text{QED}\end{aligned}$$

Remark 7.1.2. Theorem 7.1.4 is valid under the alternative condition of U being u.m.s.m.s. and V being s.m.s.

Remark 7.1.3. Let $U = V$ be a Banach space, $d_s(x, y) = \|x\|x\|^{s-1} - y\|y\|^{s-1}\|$, $x, y \in U$, where $s \geqslant 0$ and $x|x|^{s-1} = 0$ for $x = 0$. Let $\mu_s(X, Y) = \mathbb{E}d_s(X, Y)$. Then the corresponding minimal metrics $\boldsymbol{\kappa}_s(X, Y) := \hat{\mu}_s(X, Y)$ are the *absolute pseudo-moments of order s* (see (4.3.40)–(4.3.43)). By Theorem 7.1.4 $\boldsymbol{\kappa}_s$ can be expressed in terms of the more simple metric $\boldsymbol{\kappa}_1$, $\kappa_s(P_1, P_2) = \boldsymbol{\kappa}_1(P_{1\phi}, P_{2\phi})$, where $\phi(x) = x\|x\|^{s-1}$.

7.2 TWO EXAMPLES OF K-MINIMAL METRICS

Let (U, d) be a s.m.s. with metric d and Borel σ-algebra $\mathscr{B}(U)$. Let U^n be the Cartesian product of n copies of the space U. We consider in U^n the metrics $\rho_\alpha(x, y)$, $\alpha \in [0, \infty]$, $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in U^n$ of the following form

$$\rho_\alpha(x, y) = \left(\sum_{i=1}^{n} d^\alpha(x_i, y_i)\right)^{\min(1, 1/\alpha)} \qquad \text{for } \alpha \in (0, \infty)$$

$$\rho_\infty(x, y) = \max\{d(x_i, y_i); i = 1, \ldots, n\}$$

$$\rho_0(x, y) = \sum_{i=1}^{n} I\{(x, y); x_i \neq y_i\} \tag{7.2.1}$$

where I is the indicator in U^{2n}. Let $\mathfrak{X}(U^n) = \{X = (X_1, \ldots, X_n)\}$ be the space of all n-dimensional U-valued r.v.s defined on a probability space $(\Omega, \mathscr{A}, \Pr)$ which is rich enough (see Section 2.5 and Remark 2.5.1).

Let μ be a probability semimetric in the space $\mathfrak{X}(U^n)$. For every pair of random vectors $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)$ in $\mathfrak{X}(U^n)$ we define the K-minimal metric

$$\hat{\hat{\mu}}(X, Y) = \inf \mu(X, Y)$$

where the infimum is taken over all joint distributions $\Pr_{X,Y}$ with fixed one-dimensional marginal distributions $\Pr_{X_i}$, $\Pr_{Y_i}$, $i = 1, \ldots, n$. In the case $n = 1$, $\hat{\hat{\mu}} = \hat{\mu}$ is the minimal metric with respect to μ. Following the definitions in Section 2.3, a semimetric μ in $\mathscr{X}(U^n)$ is called a simple semimetric if its values

$\mu(X, Y)$ are determined by the pair of marginal distributions $\Pr_X$, $\Pr_Y$. A semimetric $\mu(X, Y)$ in $\mathscr{X}(U^n)$ is called *componentwise simple* (or K-simple) if its values are determined by the one-dimensional marginal distributions $\Pr_{X_i}$, $\Pr_{Y_i}$, $i = 1, \dots, n$. Obviously, every K-simple semimetric is simple in $\mathfrak{X}(U^n)$.

We give two examples of K-simple semimetrics that will be used frequently in what follows:

Example 1. Suppose that in $\mathbb{R}^n$ a monotone seminorm $\|x\|$ is given; that is, (a) $\|x\| \geqslant 0$ for any $x \in \mathbb{R}^n$; (b) $\|\lambda x\| = |\lambda| \cdot \|x\|$ for $\lambda \in \mathbb{R}$, $x \in \mathbb{R}^n$; (c) $\|x + y\| \leqslant \|x\| + \|y\|$; (d) if $0 < x_i < y_i$, $i = 1, \dots, n$, then $\|x\| \leqslant \|y\|$. Examples of monotone seminorms:

(i) a monotone norm

$$\|a\|_\alpha = \left(\sum_{i=1}^n |a_i|^\alpha\right)^{1/\alpha} \quad 1 \leqslant \alpha < \infty \quad a = (a_1, \dots, a_n) \in \mathbb{R}^n, \tag{7.2.2}$$

$$\|a\|_\infty = \max\{|a_i|, i = 1, \dots, n\}; \tag{7.2.3}$$

(ii) a monotone seminorm

$$\|a\| = \left|\sum_{i=1}^n a_i\right|. \tag{7.2.4}$$

Let $\mu^{(1)}, \dots, \mu^{(n)}$ be simple metrics in $\mathfrak{X}(U)$. The semimetric $\mu(X, Y) = \|\mu^{(1)}(X_1, Y_1), \dots, \mu^{(n)}(X_n, Y_n)\|$ is K-simple in $\mathfrak{X}(U^n)$.

Example 2. Denote by E a random variable uniformly distributed on $(0, 1)$, and for every $X = (X_1, \dots, X_n) \in \mathfrak{X}(\mathbb{R}^n)$ denote by X_E the random vector $X_E = (F_{X_1}^{-1}(E), \dots, F_{X_n}^{-1}(E))$, where $F_{X_i}^{-1}(t) = \sup\{x: F_{X_i}(x) \leqslant t\}$. For any p. metric $\mu(X, Y)$ in the space $\mathfrak{X}(\mathbb{R}^n)$

$$\tilde{\mu}(X, Y) = \mu(X_E, Y_E) \tag{7.2.5}$$

is K-simple in $\mathfrak{X}(\mathbb{R}^n)$. Obviously, $\overset{n}{\mu} \leqslant \tilde{\mu}$.

In the next two sections, for some simple and compound probability metrics, we shall find the explicit form of the corresponding K-minimal metrics. We shall often use the following obvious assertion.

Theorem 7.2.1. Let $\nu = \hat{\mu}$. Then $\overset{n}{\mu} = \overset{n}{\nu}$.

7.3 *K*-MINIMAL METRICS OF GIVEN PROBABILITY METRICS; THE CASE $U = R$

In this section we shall examine the representations of the K-minimal metrics w.r.t. the following p. metrics in $\mathfrak{X}(\mathbb{R}^n)$: Lévy metric, Kolmogorov metric, and the p-average metric $\mathscr{L}_p$ (see (4.1.22), (4.1.24), (3.3.3)).

Let $0 < \alpha < \infty$ and ρ_α be defined by (7.2.1). The expression $x \leqslant y$ or $x \in (-\infty, y]$ for $x, y \in \mathbb{R}^n$ means that $x_i \leqslant y_i$, for all $i = 1, \dots, n$. As a metric d in $U = \mathbb{R}^1$ we take the uniform metric $d(x_1, y_1) = |x_1 - y_1|$ for $x_1, y_1 \in \mathbb{R}$. For every $\alpha \in (0, \infty)$ we define a Lévy metric in $\mathfrak{X}(\mathbb{R}^n)$

$$\mathbf{L}(X, Y; \alpha) = \inf\{\varepsilon > 0; \Pr(X \leqslant x) \leqslant \Pr(Y \in (-\infty, x]_\alpha^\varepsilon) + \varepsilon,$$
$$\Pr(Y \leqslant x) \leqslant \Pr(X \in (-\infty, x]_\alpha^\varepsilon) + \varepsilon, \quad \forall x \in \mathbb{R}^n\}$$

where $A_\alpha^\varepsilon = \{x : \rho_\alpha(x, A) \leqslant \varepsilon\}$ for any $A \subset \mathbb{R}^n$. As is well known, $\mathbf{L}(X, Y; \alpha)$, $\alpha \in (0, \infty]$ metrizes the weak convergence in $\mathfrak{X}(\mathbb{R}^n)$. In $\mathfrak{X}(\mathbb{R}^1)$ we define the Lévy metric $\mathbf{L}(X_1, Y_1; \alpha)$ in the above manner. Obviously, $\mathbf{L}(X_1, Y_1; \alpha) = \mathbf{L}(X_1, Y_1, 1)$ for $\alpha \in [1, \infty]$ is the usual Lévy metric (2.1.3) (see Fig. 4.1.1). We recall the uniform metric (Kolmogorov metric) $\boldsymbol{\rho}(X, Y)$ in $\mathfrak{X}(\mathbb{R}^n)$

$$\boldsymbol{\rho}(X, Y) = \sup\{|\Pr(X \leqslant x) - \Pr(Y \leqslant x)| : x \in \mathbb{R}^n\}.$$

Denote by $\mathbf{W}$ and $\boldsymbol{\delta}$ the following simple metrics in $\mathfrak{X}(\mathbb{R}^n)$

$$\mathbf{W}(X, Y; \alpha) := \inf\{\varepsilon > 0; \Pr(X \leqslant x) \leqslant \Pr(Y \in (-\infty, x]_\alpha^\varepsilon),$$
$$\Pr(Y \leqslant x) \leqslant \Pr(X \in (-\infty, x]_\alpha^\varepsilon), \quad \forall x \in \mathbb{R}^n\}$$

and $\boldsymbol{\delta}(X, Y)$ is the *discrete metric*: $\boldsymbol{\delta}(X, Y) = 0$ if $F_X = F_Y$, and $\delta(X, Y) = +\infty$ if $F_X \neq F_Y$. The following relations are valid (see Example 4.1.3)

$$\mathbf{L}\left(\frac{1}{\lambda} X, \frac{1}{\lambda} Y; \alpha\right) \to \boldsymbol{\rho}(X, Y) \quad \text{as } \lambda \to 0, \lambda > 0, \alpha \in (0, \infty] \tag{7.3.1}$$

$$\lim_{\lambda \to \infty} \lambda \mathbf{L}\left(\frac{1}{\lambda} X, \frac{1}{\lambda} Y; \alpha\right) = \mathbf{W}(X, Y; \alpha), \text{ for } \alpha \in [1, \infty] \tag{7.3.2}$$

$$\lim_{\lambda \to \infty} \lambda \mathbf{L}\left(\frac{1}{\lambda} X, \frac{1}{\lambda} Y; \alpha\right) = \boldsymbol{\delta}(X, Y), \text{ for } \alpha \in (0, 1).$$

For any $X = (X_1, \dots, X_n) \in \mathfrak{X}(\mathbb{R}^n)$ we denote by $M_X(x) = \min(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) = \Pr(F_{X_1}^{-1}(E) \leqslant x_1, \dots, F_{X_n}^{-1}(E) \leqslant x_n)$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ the maximal distribution function having fixed one-dimensional marginal distributions F_{X_i}, $i = 1, \dots, n$. For any semimetric $\mu(X, Y)$ in $\mathfrak{X}(\mathbb{R}^n)$ we denote by $\mu(X_i, Y_i)$, $i = 1, \dots, n$, the corresponding semimetric in $\mathfrak{X}(\mathbb{R})$.

Theorem 7.3.1. For any $\alpha \in (0, \infty]$ and $X, Y \in \mathfrak{X}(\mathbb{R}^n)$

$$\max_{1 \leqslant i \leqslant n} \mathbf{L}(X_i, Y_i; \alpha)^{\alpha^*} \leqslant \mathring{\mathbf{L}}(X, Y; \alpha) \leqslant \max_{1 \leqslant i \leqslant n} \mathbf{L}(n^{1/\alpha} X_i, n^{1/\alpha} Y_i; \alpha) \quad \alpha^* := \max(1, 1/\alpha). \tag{7.3.3}$$

Proof. The lower estimate for $\overset{n}{\mathbf{L}}$ follows from the inequality

$$\max\{\mathbf{L}(X_i, Y_i; \alpha); i = 1, \ldots, n\} \leqslant \mathbf{L}(X, Y; \alpha)^{\beta}, \beta := \min(1, \alpha)$$

for any $X, Y \in \mathfrak{X}(\mathbb{R}^n)$. Let $\max\{\mathbf{L}(n^{1/\alpha}X_i, n^{1/\alpha}Y_i; \alpha); i = 1, \ldots, n\} < \varepsilon$ and $x \in \mathbb{R}^n$. Then for any $i = 1, \ldots, n$, any $x_i \in \mathbb{R}$

$$\Pr(X_i \leqslant x_i) \leqslant \Pr(Y_i \leqslant x_i + n^{-1/\alpha}\varepsilon) + \varepsilon$$

and thus

$$\min_{1 \leqslant i \leqslant n} \Pr(X_i \leqslant x_i) \leqslant \min_{1 \leqslant i \leqslant n} \Pr(Y_i \leqslant x_i + n^{-1/\alpha} \cdot \varepsilon^{\max(1, 1/\alpha)}) + \varepsilon. \tag{7.3.4}$$

Given $X, Y \in \mathfrak{X}(\mathbb{R}^n)$ denote $\tilde{X} = X_E$, $\tilde{Y} = Y_E$ (see Example 2 in Section 7.2). Then X and Y have d.f.s M_X and M_Y, respectively. Now, (7.3.4) implies that $M_X(x) = \Pr(\tilde{X} \leqslant x) \leqslant \Pr(\tilde{Y} \in (-\infty, x]^{\varepsilon}_{\alpha}) + \varepsilon$. Therefore, $\mathbf{L}(X, Y; \alpha) < \varepsilon$ and thus the upper bound in (7.3.3) is established. QED

Letting $\alpha = \infty$ in (7.3.3) we obtain the following corollary immediately.

Corollary 7.3.1. For any X and $Y \in \mathfrak{X}(\mathbb{R}^n)$,

$$\overset{n}{\mathbf{L}}(X, Y; \infty) = \mathbf{L}(X_E, Y_E; \infty) = \max\{\mathbf{L}(X_i, Y_i, \infty); i = 1, \ldots, n\}. \tag{7.3.5}$$

Corollary 7.3.2. For any X and $Y \in \mathfrak{X}(\mathbb{R}^n)$

$$\overset{n}{\boldsymbol{\rho}}(X, Y) = \boldsymbol{\rho}(X_E, Y_E) = \max\{\boldsymbol{\rho}(X_i, Y_i); i = 1, \ldots, n\}. \tag{7.3.6}$$

Proof. One can prove (7.3.6) using the same arguments as in the proof of Theorem 7.3.1. Another way is to use (7.3.5) and (7.3.1). QED

Corollary 7.3.3. For every $\alpha \in (0, \infty]$ and $X, Y \in \mathfrak{X}(\mathbb{R}^n)$,

$$\max_{1 \leqslant i \leqslant n} \mathbf{W}(X_i, Y_i; \alpha)^{\alpha^*} \leqslant \overset{n}{\mathbf{W}}(X, Y; \alpha) \leqslant \max_{1 \leqslant i \leqslant n} \mathbf{W}(n^{1/\alpha}X_i, n^{1/\alpha}Y_i; \alpha)$$

$$\overset{n}{\mathbf{W}}(X, Y; \infty) = \sup\{|F_{X_i}^{-1}(t) - F_{Y_i}^{-1}(t)|; t \in [0, 1], i = 1, \ldots, n\}. \tag{7.3.7}$$

Proof. The first estimates follow from (7.3.3) and (7.3.2). The representation for $\overset{n}{\mathbf{W}}(X, Y, \infty)$ is a consequence of the above estimates. QED

Corollary 7.3.4. For any $X, Y \in \mathfrak{X}(\mathbb{R}^n)$

$$\overset{n}{\boldsymbol{\delta}}(X, Y) = \boldsymbol{\delta}(X_E, Y_E) = \max\{\boldsymbol{\delta}(X_i, Y_i); i = 1, \ldots, n\}. \tag{7.3.8}$$

The equalities (7.3.5)–(7.3.8) describe the sharp lower bounds of the simple metrics $\mathbf{L}(X, Y)$, $\boldsymbol{\rho}(X, Y)$, $\mathbf{W}(X, Y)$ and $\boldsymbol{\delta}(X, Y)$ in $\mathfrak{X}(\mathbb{R}^n)$ in case of fixed one-dimensional distributions, F_{X_i}, F_{Y_i} $(i = 1, \ldots, n)$.

We shall next consider the K-minimal metric with respect to the average compound distance

$$\mathscr{L}_H(X, Y) = \mathbb{E}H(d(X, Y)),\ X, Y \in \mathfrak{X}(\mathbb{R}^n) \tag{7.3.9}$$

(see Example 3.3.1 and (3.3.3)) where $d(x, y) = \rho_\alpha(x, y)$ ((7.2.2)–(7.2.4)) $(\alpha \geqslant 1)$ and H is a convex function on $[0, \infty)$, $H(0) = 0$. We shall examine minimal functionals that are more general than $\overset{n}{\hat{\mathbf{L}}}_H$.

Definition 7.3.1. (Cambanis *et al.*, 1976). A function ϕ: $E \subset \mathbb{R}^2 \to \mathbb{R}$ is said to be *quasiantitone* if

$$\phi(x, y) + \phi(x', y') \leqslant \phi(x', y) + \phi(x, y') \tag{7.3.10}$$

for all $x' > x$, $y' > y$, $x, x', y, y' \in E$. We call ϕ: $E \subset \mathbb{R}^N \to \mathbb{R}$ quasiantitone if it is an quasiantitone function of any two coordinates considered separately.

Some examples of quasiantitone functions are the following: $f(x - y)$ where f is a non-negative convex function on $\mathbb{R}$, $|x - y|^p$, for $p \geqslant 1$, $\max(x, y)$, $x, y \in \mathbb{R}$, any concave function on $\mathbb{R}^N$ and any distribution function of a non-positive measure in $\mathbb{R}^n$ are quasiantitone.

At first we shall find an explicit solution of the multi-dimensional Kantorovich problem (see Section 5.1, VI, and (5.1.36)) in the case of $U = \mathbb{R}$, $d = \rho\alpha$, and a cost function c being quasiantitone. Namely, let $\tilde{F} = \{F_i, i = 1, \ldots, N\}$ be the vector of N distribution functions $F_1, \ldots, F_N$ on $\mathbb{R}$ and $\mathfrak{P}(\tilde{F})$ be the set of all distribution functions F on $\mathbb{R}^N$ with fixed one-dimensional marginals $F_1, \ldots, F_N$. The pointwise upper bound of the distributions F in $\mathfrak{P}(\tilde{F})$ is obtained at the Hoeffding distribution

$$M(x) := \min(F_1(x_1), \ldots, F_N(x_N)) \quad x = (x_1, \ldots, x_N). \tag{7.3.11}$$

The next theorem shows that the minimal total cost in the multi-dimensional Kantorovich transportation problem

$$\mathscr{A}_c(\tilde{F}) = \inf\left\{\int_{\mathbb{R}^N} c\, \mathrm{d}F : F \in \mathfrak{P}(\tilde{F})\right\} \tag{7.3.12}$$

coincides with the total cost of $\int_{\mathbb{R}^N} c\, \mathrm{d}M$, i.e., M describes the optimal plan of transportation.

Lemma 7.3.1. (Lorentz, 1953) For a p-tuple $(x_1, \ldots, x_p)$ let $(\bar{x}_1, \ldots, \bar{x}_p)$ denote its rearrangement in increasing order. Then, given N p-tuples $(x_1^{(1)}, \ldots, x_p^{(1)})$, $\ldots$, $(x_1^{(N)}, \ldots, x_p^{(N)})$, for any quasiantitone function ϕ, the minimum of

$\sum_{i=1}^{p} \phi(x_i^{(1)}, \ldots, x_i^{(N)})$ over all the rearrangements of the p-tuples is attained at $(\bar{x}_1^{(1)}, \ldots, \bar{x}_p^{(1)}), \ldots, (\bar{x}_1^{(N)}, \ldots, \bar{x}_p^{(N)})$.

Proof. Let $X^{(k)} = (x_1^{(k)}, \ldots, x_p^{(k)})$. Further, in inequalities containing values of the function ϕ at different points, we shall omit those of the arrangements which take the same but arbitrary values. For a group I of indices i, $1 \leqslant i \leqslant N$, we denote $U_I := \{u_i\}_{i \in I}$, $U'_I = \{u'_i\}_{i \in I}$, and $U_I + U'_I = \{u_i + u'_i\}_{i \in I}$.

Claim. For any two disjoint groups of indices I, J and h_i, $h_i \geqslant 0$,

$$\phi(U_I + H_I, U_J + H_J) - \phi(U_I + H_I, U_J) - \phi(U_I, U_J + H_J) + \phi(U_I, U_J) \leqslant 0. \tag{7.3.13}$$

Proof of the claim. Let I' be the group consisting of I and the index k, which belongs neither to I nor to J. Then

$$\begin{aligned}
&\phi(U_{I'} + H_{I'}, U_J + H_J) - \phi(U_{I'} + H_{I'}, U_J) - \phi(U_{I'}, U_J + H_J) + \phi(U_{I'}, U_J) \\
&\quad = \{\phi(U_I + H_I, u_k + h_k, U_J + H_J) - \phi(U_I + H_I, u_k + h_k, U_J) \\
&\qquad - \phi(U_I, u_k + h_k, U_J + H_J) + \phi(U_I, u_k + h_k, U_J)\} \\
&\qquad + \{\phi(U_I, u_k + h_k, U_J + H_J) - \phi(U_I, u_k + h_k, U_J) \\
&\qquad - \phi(U_I, u_k, U_J + H_J) + \phi(U_I, u_k, U_J)\}.
\end{aligned}$$

Starting the inductive arguments with the inequality

$$\phi(x', y') - \phi(x', y) - \phi(x, y') + \phi(x, y) \leqslant 0 \quad x' \geqslant y, \quad y' \geqslant y$$

we prove the claim by induction with respect to the number of elements of I and J.

Further, for any $1 \leqslant s < p$ we consider the following operation which gives a new set of p-tuples $\tilde{X}^{(k)}$. We put $\tilde{x}_i^{(k)} = x_i^{(k)}$ for $i \neq s$, $i \neq s+1$ and $\tilde{x}_s^{(k)} = \min(x_s^{(k)}, x_{s+1}^{(k)})$, $\tilde{x}_{s+1}^{(k)} = \max(x_s^{(k)}, x_{s+1}^{(k)})$. If I consists of indices k, for which $x_s^{(k)} \leqslant x_{s+1}^{(k)}$, J of indices k for which $x_s^{(k)} \geqslant x_{s+1}^{(k)}$, u_k is the smaller, $u_k + h_k$ is the larger of the two values, then

$$\sum_{i=1}^{p} \phi(x_i^{(1)}, \ldots, x_i^{(N)}) \geqslant \sum_{i=1}^{p} \phi(\tilde{x}_i^{(1)}, \ldots, \tilde{x}_i^{(N)}) \tag{7.3.14}$$

is exactly the inequality (7.3.13). Continuing in the same manner we prove the theorem after a finite number of steps. QED

Theorem 7.3.2. (Tchen, 1980). Let $\tilde{F} = (F_1, \ldots, F_N)$ be a set of N distribution functions on $\mathbb{R}$ and M be defined by (7.3.11). Given a quasiantitone function ϕ: $\mathbb{R}^N \to \mathbb{R}$, suppose that the family $\{\phi(X), X$ distributed as $F \in \mathfrak{B}(\tilde{F})\}$ is uniformly integrable. Then

$$\mathscr{A}_\phi(\tilde{F}) = \int \phi \, \mathrm{d}M. \tag{7.3.15}$$

Remark 7.3.1. For $N = 2$ this theorem is known as the Cambanis *et al.* (1976) Theorem (see Kalashnikov and Rachev 1988, Theorem 7.1.1).

Proof. Suppose first that the F_is have compact support. Let $X = (X_1, \ldots, X_N)$ be distributed as $F \in \mathfrak{P}(\tilde{F})$ and defined on $[0, 1]$ with Lebesgue measure. By Lemma 7.3.1, if the distribution F is concentrated on p atoms $(x_i^{(1)}, \ldots, x_i^{(N)})$ $(i = 1, \ldots, p)$ of mass $1/p$, $\mathbb{E}\phi(X) \geqslant \mathbb{E}\phi(X_E)$ where $X_E = (F_{X_1}^{-1}(E), \ldots, F_{X_N}^{-1}(E))$, $E(\omega) = \omega$, $\omega \in [0, 1]$ (see Section 7.2, Example 2). In the general case, let

$$x_{i,k}^m = 2^m \mathbb{E}\{X_i I[k2^{-m} \leqslant X_i \leqslant (k+1)2^{-m}]\}$$

and

$$X_i^m(\omega) = \sum_{k=0}^{2^m - 1} x_{i,k}^m \cdot I[k2^{-m} \leqslant \omega \leqslant (k+1)2^{-m}] \quad i = 1, \ldots, N, \omega \in [0, 1].$$

$X_1^m, X_2^m, \ldots, X_N^m$ are step functions and bounded martingales converging a.s. to $X_1, \ldots, X_N$ respectively, cf. Breiman (1968), p. 94. Call $\overline{X_i^m}$, $i = 1, \ldots, N$ the reorderings of X_i^m. $\overline{X_i^m}$ and X_i^m have the same distribution and $\overline{X_i^m} = F_{X_i^m}^{-1}(E)$; hence $\overline{X_i^m} \to F_{X_i}^{-1}(E)$ a.s., so that in the bounded case the theorem follows by bounded convergence.

Consider the general case. Let $\mathbb{B}_N = (-B, B)^N$ and let F_B be the distribution which is F outside $\mathbb{B}_N$ and $F_B\{A\} = F\{A \cap \mathbb{B}_N^c\} + \bar{F}_B\{A \cap \mathbb{B}_N\}$ for all Borel sets on $\mathbb{R}^N$, where $\bar{F}_B$ is the maximal sub-probability with sub-distribution function

$$M_B(x) = \min_{1 \leqslant i \leqslant N} F\{(-B, B]^{i-1} \times (-B, x_i] \times (-B, B]^{N-i}\}$$

for

$$x = (x_1, \ldots, x_N) \in \mathbb{B}_N.$$

Clearly, $F_B \in \mathfrak{P}(\tilde{F})$ and F_B converges weakly to M as $B \to \infty$, which completes the proof of the theorem. QED

As a consequence of the explicit solution of the N-dimensional Kantorovich problem, we shall find an explicit representation of the following minimal functional

$$\mathscr{L}_{p,q}(\tilde{F}) := \inf\{\mathbb{E}D_{p,q}(X): X = (X_1, \ldots, X_N) \in \mathfrak{X}(\mathbb{R}^N), F_{X_i} = F_i, i = 1, \ldots, N\} \tag{7.3.16}$$

where $D_{p,q}(x) = [\sum_{1 \leqslant i < j \leqslant N} |x_i - x_j|^p]^q$, $p \geqslant 1$, $q \leqslant 1$, and $\tilde{F} = (F_1, \ldots, F_N)$ is a vector of one-dimensional d.f.s.

Corollary 7.3.5. For any $p \geqslant 1$ and $q \leqslant 1$

$$\mathscr{L}(\tilde{F}) = \int_0^\infty D_{p,q}(F_1^{-1}(t), \ldots, F_N^{-1}(t))\, dt. \tag{7.3.17}$$

As a special case of Theorem 7.3.2 ($N = 2$, $\phi(x, y) = H(|x - y|)$, H convex on $[0, \infty)$, $H \in \mathscr{H}$, cf. Example 2.2.1), we obtain the following corollary.

Corollary 7.3.6. Let H be a convex function from $\mathscr{H}$ and

$$\mathscr{L}_H(X, Y) = \mathbb{E}H(|X - Y|)$$

be the H-average distance on $\mathfrak{X}(\mathbb{R})$ (cf. Example 3.3.1). Then

$$\hat{\mathscr{L}}_H(X, Y) = \tilde{\mathscr{L}}_H(X, Y) = \int_0^1 H(|F_X^{-1}(t) - F_Y^{-1}(t)|)\,dt. \tag{7.3.18}$$

Further, we shall consider other examples of explicit formulae for K-minimal and minimal distances and metrics. Denote by $\mathbf{m}(X, Y)$ the following probability metric:

$$\mathbf{m}(X, Y) = \mathbb{E}\left[2\max(X_1, \ldots, X_n, Y_1, \ldots, Y_n) - \frac{1}{n}\sum_{i=1}^{n}(X_i + Y_i)\right]. \tag{7.3.19}$$

Theorem 7.3.3. Suppose that the set of random vectors X and Y with fixed one-dimensional marginals is uniformly integrable. Then

$$\hat{\mathbf{m}}(X, Y) = \tilde{\mathbf{m}}(X, Y) = \int_{-\infty}^{\infty} \frac{1}{n}\sum_{i=1}^{n}[F_{X_i}(u) + F_{Y_i}(u)]$$
$$- 2\min[F_{X_1}(u), \ldots, F_{X_n}(u), F_{Y_1}(u), \ldots, F_{Y_n}(u)]\,du. \tag{7.3.20}$$

Proof. Suppose $\mathbb{E}|X_i| + \mathbb{E}|Y_i| < \infty$, $i = 1, \ldots, n$. Then from the representation

$$\mathbf{m}(X, Y) = \int_{-\infty}^{\infty} \frac{1}{n}\sum_{i=1}^{n}[F_{X_i}(u) + F_{Y_i}(u)] - 2\Pr(\max(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \leqslant u)\,du$$

and the Hoeffding inequality,

$$\Pr(\max(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \leqslant u) \leqslant \min(F_{X_1}(u), \ldots, F_{X_n}(u), F_{Y_1}(u), \ldots, F_{Y_n}(u)),$$

we obtain (7.3.20). The weaker regularity condition is obtained as in the previous theorem. QED

Consider the special case $n = 1$. We shall prove the equality

$$\hat{\mu}(X, Y) = \tilde{\mu}(X, Y) := \mu(F_X^{-1}(E), F_Y^{-1}(E)) \tag{7.3.21}$$

(E is uniformly distributed on $(0, 1)$) for various compound distances in $\mathfrak{X}(\mathbb{R})$.

In Example 3.3.3 we introduced the Birnbaum–Orlicz compound distances

$$\Theta_H(X_1, X_2) = \int_{-\infty}^{\infty} H\Big(\Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1)\Big)\,\mathrm{d}t \quad H \in \mathscr{H}$$

$$\mathbf{R}_H(X_1, X_2) = \sup_{t\in\mathbb{R}} H(\Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1)) \tag{7.3.22}$$

and compound metrics

$$\Theta_p(X_1, X_2) = \left\{\int_{-\infty}^{\infty} [\Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1)]^p\,\mathrm{d}t\right\}^{p'} \quad p' = \min(1, 1/p)$$

$$\Theta_\infty(X_1, X_2) = \sup_{t\in\mathbb{R}} [\Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1)].$$

Note that $\Theta_1(X_1, X_2) = \mathbb{E}|X_1 - X_2|$ for $H(t) = t$. In Example 3.2.4 we consider the corresponding simple Birnbaum–Orlicz distances

$$\theta_H(F_1, F_2) = \int_{-\infty}^{\infty} H(|F_1(x) - F_2(x)|)\,\mathrm{d}x \quad H \in \mathscr{H} \tag{7.3.23}$$

$$\rho_H(F_1, F_2) = \sup_{x\in\mathbb{R}} H(|F_1(x) - F_2(x)|)$$

and simple metrics

$$\theta_p(F_1, F_2) = \left(\int_{-\infty}^{\infty} |F_1(x) - F_2(x)|^p\,\mathrm{d}x\right)^{p'}$$

$$\theta_\infty(F_1, F_2) = \rho(F_1, F_2) = \sup_{x\in\mathbb{R}} |F_1(x) - F_2(x)|.$$

Theorem 7.3.4

$$\theta_H = \tilde{\Theta}_H = \hat{\Theta}_H \quad \rho_H = \tilde{\mathbf{R}}_H = \hat{\mathbf{R}} \quad \theta_p = \tilde{\Theta}_p = \hat{\Theta}_p \quad 0 < p \leqslant \infty. \tag{7.3.24}$$

Proof. To prove the first equality in (7.3.24), consider the set of all random pairs (X_1, X_2) with marginal d.f.s F_1 and F_2. For any such pair

$$\begin{aligned}\Theta_H(X_1, X_2) &= \int_{-\infty}^{\infty} H(F_1(t) + F_2(t) - 2\Pr(X_1 \vee X_2 \leqslant t))\,\mathrm{d}t \\ &\geqslant \tilde{\Theta}_H(X_1, X_2) = \int_{-\infty}^{\infty} H(F_1(t) + F_2(t) - 2\min(F_1(t), F_2(t))\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} H(|F_1(t) - F_2(t)|)\,\mathrm{d}t = \theta_H(F_1, F_2).\end{aligned}$$

Thus $\hat{\Theta}_H - \tilde{\Theta}_H = \theta_H$. In a similar way one proves the other equalities in (7.3.24). QED

Remark 7.3.2. Theorem 7.3.2 for $N = 2$ shows that the infimum of $\mathbb{E}\phi(X_1, X_2)$ (ϕ is a quasiantitone function (7.3.10) over $\mathfrak{P}(F_1, F_2)$ (the set of all possible joint d.f. $H = F_{X_1, X_2}$ with fixed marginals $F_{X_i} = F_i$) is attained at the upper Hoeffding–Fréchet bound $\bar{H}(x, y) = \min(F_1(x), F_2(y))$. Similarly (see Cambanis *et al.*, 1976, Tchen 1980)),

$$\sup\{\mathbb{E}\phi(X_1, X_2)\colon H \in \mathfrak{P}(F_1, F_2) = \int_0^1 \phi(F_1(t), F_2(1-t))\,dt \qquad (7.3.25)$$

i.e., the supremum of $\mathbb{E}\phi(X_1, X_2)$ is attained at the lower Hoeffding–Fréchet bound $\underline{H}(x, y) = \max(0, F_1(x) + F_2(y) - 1)$. The multi-dimensional analogs of (7.3.25) are not known. Notice that the multivariate lower Hoeffding–Fréchet bound $\underline{H}(x_1, \ldots, x_N) = \max(0, F_1(x_1) + \cdots + F_N(x_n) - N + 1)$ is not a d.f. on $\mathbb{R}^N$ in contrast to the upper bound $\bar{H}(x_1, \ldots, x_N) = \min(F_1(x_1), \ldots, F_N(x_N))$ which is a d.f. on $\mathbb{R}^N$. That is why we do not have an analog of Theorem 7.3.2 when the supremum of $\mathbb{E}\phi(X_1, \ldots, X_N)$ over the set of N-dimensional d.f.s with fixed one-dimensional marginals is considered.

Remark 7.3.3. Kolmogorov in 1981 stated the following problem to Makarov: find the infimum and supremum of $\Pr(X + Y < z)$ over $\mathfrak{P}(F_1, F_2)$ for any fixed z. The problem was solved independently by Makarov (1981) and Rüschendorf (1982). Rüschendorf (1982) considered also the multivariate extension. Another solution was given by Frank *et al.* (1987). Their solution was based on the notion of 'copula' linking the multi-dimensional d.f.s to their one-dimensional marginals (see Sklar 1959, Schweizer and Sklar 1983, Sect. 6.5, Wolff and Schweizer 1981, Genest and MacKay 1986).

7.4 THE CASE: U IS A SEPARABLE METRIC SPACE

We begin with a multivariate extension of the Strassen theorem, $\pi = \hat{\mathbf{K}}$; π being the Prokhorov metric (see Example 3.2.3 and (3.2.18)).

The following theorem was proved by Schay (1979) in the case where (U, d) is a complete separable space. We shall use the method of Dudley (1976, Theorem 18.1) to extend this result in the case of a separable space.

Denote by $\mathscr{P}(U)$ the space of all Borel probability measures (laws) on a s.m.s. (U, d). Let $N \geqslant 2$ be an integer, let $\|x\|$, $x \in \mathbb{R}^m$, be a monotone norm (if $0 < x < y$ then $\|x\| \leqslant \|y\|$) in $\mathbb{R}^m$, where $m = \binom{N}{2}$, and let

$$\mathscr{D}(x_1, \ldots, x_N) = \|d(x_1, x_2), \ldots, d(x_1, x_N), d(x_2, x_3), \ldots, d(x_{N-1}, x_N)\|. \quad (7.4.1)$$

Theorem 7.4.1. For any $P_1, \ldots, P_N$ in $\mathscr{P}(U)$, $\alpha \geqslant 0$, $\beta \geqslant 0$ the following two assertions are equivalent:

(I) For any $a > \alpha$ there exists a $\mu \in \mathscr{P}(U^N)$ with marginal distributions $P_1, \ldots, P_N$ such that

$$\mu\{\mathscr{D}(x_1, \ldots, x_N) > a\} \leqslant \beta. \tag{7.4.2}$$

(II) For any Borel sets $B_1, \ldots, B_{N-1} \in \mathscr{B}(U)$

$$P_1(B_1) + \cdots + P_{N-1}(B_{N-1}) \leqslant P_N B^{(\alpha)} + \beta + N - 2 \tag{7.4.3}$$

where $B^{(\alpha)} = \{x_N \in U : \mathscr{D}(x_1, \ldots, x_N) \leqslant \alpha$, for some $x_1 \in B_1, \ldots, x_{N-1} \in B_{N-1}\}$. If $P_1, \ldots, P_N$ are tight measures, then $a = \alpha$.

Proof. Assertion (I) implies (II), since

$$\begin{aligned} P_1(B_1) &\leqslant \mu(\mathscr{D}(x_1, \ldots, x_N) > a) + \mu\left(\bigcap_{i=1}^{N-1} \{x_i \in B_i\}, \mathscr{D}(x_1, \ldots, x_N) \leqslant a\right) \\ &\quad + \mu\left(x_1 \in B_1, \bigcup_{i=2}^{N-1} \{x_i \notin B_i\}, \mathscr{D}(x_1, \ldots, x_N) \leqslant a\right) \\ &\leqslant \beta + \mu(B^{(a)}) + \sum_{i=2}^{N-1} (1 - P_i(B_i)). \end{aligned}$$

As $a \to \alpha$ we obtain (II).

To prove that (II) $\Rightarrow$ (I) suppose first that $P_1, \ldots, P_N$ are tight measures. Let $\{x_i : i = 1, 2, \ldots\}$ be a dense sequence in U, and $P_{i,n}(i = 1, \ldots, N)$ be probability measures on the set $U_n := \{x_1, \ldots, x_n\}$. We first fix n, and prove (II) $\to$ (I) for $a = \alpha$, U_n and $P_{1,n}, \ldots, P_{N,n}$ in place of U and $P_1, \ldots, P_N$ and then let $n \to \infty$.

For any $I = (i_1, \ldots, i_N) \in \{1, \ldots, n\}^N$ and $X_I = (x_{i_1}, \ldots, x_{i_N})$ define the indicator: $\operatorname{Ind}(X_I) = 1$ if $\mathscr{D}(X_I) \leqslant \alpha$ and $\operatorname{Ind}(X_I) = 0$ otherwise. To obtain the μ of the theorem we consider μ_n on U_n^N. We denote

$$\xi_I = \mu_n(\{X_I\}) \quad P_{i_k, j} = P_{j,n}(\{x_{i_k}\}) \quad i_k = 1, \ldots, n, \quad k, j = 1, \ldots, N.$$

Since we want μ_n to have $P_{1,n}, \ldots, P_{N,n}$ as one-dimensional projections, we require the constraints

$$\sum_{i_\ell} \xi_I \leqslant P_{i_k, j} \quad j = 1, \ldots, N \quad i_k = 1, \ldots, n, \tag{7.4.4}$$

$$\xi_I \geqslant 0$$

where in (7.4.4) i_ℓ runs from 1 to n for all $\ell \in \{1, \ldots, k-1, k+1, \ldots, N\}$.

If we denote by μ_n^* the 'optimal' μ_n that assigns as much probability as possible to the 'diagonal cylinder' C_α in U_n^N given by $\mathscr{D}(X_I) \leqslant \alpha$, then we shall determine $\mu_n^*(C_\alpha)$ by looking at the following linear programming problem of canonical form

$$\text{maximize } Z = \sum_{I \in \{1,,\ldots,n\}^N} \operatorname{Ind}(X_I)\xi_I \quad \text{subject to (7.4.4)} \tag{7.4.5}$$

The dual of the above problem is easily seen to be

$$\text{minimize } W = \sum_{i_k=1}^{n} \sum_{j=1}^{N} P_{i_k,j} u_{i_k,j}$$

subject to $u_{i_k,j} \geqslant 0$

$$\sum_{j=1}^{N} u_{i_k,j} \geqslant \operatorname{Ind}(X_I) \quad \forall i_k = 1,\ldots,n,\ k = 1,\ldots,N,\ j = 1,\ldots,N \tag{7.4.6}$$

and by the duality theorem (see for example, Section 5.2, Berge and Ghouila-Houri 1965) the maximum of Z equals the minimum of W. Let us write $\bar{u}_{i_k,j} = 1 - u_{i_k,j}$. Then (7.4.6) becomes

$$\text{minimize } W = N - 1 - \sum_{i_k=1}^{n} \sum_{j=1}^{N-1} P_{i_k,j}\bar{u}_{i_k,j} + \sum_{i_k=1}^{n} p_{i_k,N} u_{i_k,N}$$

$$\text{subject to } \bar{u}_{i_k,j} \leqslant 1,\ j = 1,\ldots,N-1,\ u_{i_k,N} \geqslant 0,$$

$$\text{and } u_{i_k,N} \geqslant (\operatorname{Ind}(X_I) - N - 1) + \sum_{j=1}^{N-1} \bar{u}_{i_k,j}$$

$$\forall i_k = 1,\ldots,n \quad k = 1,\ldots,N. \tag{7.4.7}$$

We may also assume

$$\bar{u}_{i_k,j} \geqslant 0 \quad j = 1,\ldots,N-1 \quad u_{i_k,N} \leqslant 1 \tag{7.4.8}$$

since these additional constraints cannot affect the minimum of W. Now the set of 'feasible' solutions $u_{i_k,j}, j = 1,\ldots,N-1, u_{i_k,N}, i_k = 1,\ldots,n, k = 1,\ldots,N$ for the dual problem (7.4.7), (7.4.8) is a convex polyhedron contained in the unit cube $[0,1]^{Nn}$, the extreme points of which are the vertices of the cube. Since the minimum of W is attained at one of these extreme points, there exists $\bar{u}_{i_k,j}, u_{i_k,N}$ equal to 0 or 1 which minimize W under the constraints in (7.4.7) and (7.4.8). Thus without loss of generality we may assume that $\bar{u}_{i_k,j}, u_{i_k,N}$ are 0s and 1s.

Define the sets $F_j \subset U^n, j = 1,\ldots,N-1$ such that $\bar{u}_{i_k,j} = 1$ for all j such that $x_{i_k} \in F_j$ and $u_{i_k,j} = 0$ otherwise. Then, by (7.4.7), $u_{i_k,N} = 1$ for all k such that $\operatorname{Ind}(X_I) = 1$ when $\bar{u}_{i_k,j} = 1, j = 1,\ldots,N-1$, that is, whenever x_{i_N} satisfies $\mathscr{D}(X_I) \leqslant \alpha$ with $x_{i_j} \in F_j, j = 1,\ldots,N-1$. Hence

$$\min W = N - 1 - \max[P_{1,n}(F_1) + \cdots + P_{N-1,n}(F_{N-1}) - P_{N,n}(F_n^{(\alpha)})]$$

where

$$F_n^{(\alpha)} := \{x_{i_N}: D(X_I) \leqslant \alpha \text{ for some } x_{i_j} \in F_j, j = 1,\ldots,N-1\}.$$

Thus, by the duality theorem in linear programming, maximum Z = minimum W, and then

$$\mu_n^*(\mathscr{D}(X_I) > \alpha) = 1 - \mu_n^*(C_\alpha)$$
$$= 2 - N + \max\{[P_{1,n}(F_1) + \cdots + P_{N-1,n}(F_{N-1}) - P_{N,n}(F_n^{(\alpha)})]: F_1, \ldots, F_{N-1} \subset U_n\}.$$

The latter inequality is true for any $\alpha > 0$ and therefore,

$$\inf\{\alpha: \mu_n^*(\mathscr{D}(X_I) \geqslant \alpha) \leqslant \alpha\}$$
$$= \inf\left\{\alpha: \max_{F_1, \ldots, F_{N-1} \subset U_n} [P_{1,n}(F_1) + \cdots + P_{N-1,n}(F_{N-1}) - P_{N,n}(F_n^{(\alpha)})] + 2 - N \leqslant \alpha\right\}.$$

Given P_j ($j = 1, \ldots, N$) one can take $P_{j,n}$ concentrated in finitely many atoms, say in U_n, such that the Prokhorov distance $\pi(P_{j,n}, P_j) \leqslant \varepsilon$. The latter follows, for example, by the Glivenko–Cantelli–Varadarajan Theorem (see Dudley 1989, Theorem 11.4.1). As P_j is tight then $P_{j,n}$ is uniformly tight and thus there is a weakly convergent subsequence $P_{j,n(k)} \to P_j$. The corresponding sequence of optimal measures $\mu_{n(k)}^*$ with marginals $P_{j,n(k)}$ ($j = 1, \ldots, N$) is also uniformly tight. Now the same 'tightness' argument implies the existence of a measure μ for which (7.4.2) holds.

Remark 7.4.1. It is easy to see that (II) is equivalent to (7.4.3) for all closed sets B_j ($j = 1, \ldots, N - 1$) and/or $B^{(\alpha)}$ given by $\{x_N \in U: \mathscr{D}(x_1, \ldots, x_N) < \alpha\}$.

Now, suppose that $P_1, \ldots, P_N$ are not tight. Let $\bar{U}$ be a completion of the space U. For a given $a > \alpha$ let $\varepsilon \in (0, (a - \alpha)/2\|e\|)$, where $e = (1, \ldots, 1)$ and A is a maximal subset of U such that $d(x, y) \geqslant \varepsilon/2$ for $x \neq y$ in A. Then A is countable; $A = \{x_k\}_{k=1}^{\infty}$. Let $\bar{A}_k = \{x \in \bar{U};\ d(x, x_k) < \varepsilon/2 \leqslant d(x, x_j),\ j = 1, \ldots, k - 1\}$ and $A = \bar{A}_k \cap U$. The measures $P_1, \ldots, P_N$ on U determine probability measures $\bar{P}_1, \ldots, \bar{P}_N$ on $\bar{U}$. Then $\bar{P}_1, \ldots, \bar{P}_N$ are tight and, consequently, there exists $\bar{\mu} \in \mathscr{P}(\bar{U}^N)$ with marginal distributions $\bar{P}_1, \ldots, \bar{P}_N$ for which (I) holds for $a = \alpha$. Let $P_{k,m}(B) = P_k(B \cap A_m)$, $k = 1, \ldots, N$, for any $B \in \mathscr{B}(U)$. We define the measures

$$\mu_{m_1, \ldots, m_N} = c_{m_1, \ldots, m_N} P_{m_1} \times \cdots \times P_{m_N}$$

where the number $c_{m_1, \ldots, m_N}$ is chosen so that

$$\mu_{m_1, \ldots, m_N}(A_{m_1} \times \cdots \times A_{m_N}) = \bar{\mu}(\bar{A}_{m_1} \times \cdots \times \bar{A}_{m_N}).$$

We set

$$\mu_\varepsilon = \sum_{m_1, \ldots, m_N} \mu_{m_1, \ldots, m_N}.$$

Then μ_ε has marginal distributions $P_1, \ldots, P_N$ (see the proof of Case 3, Theorem 5.2.1) and

$$\begin{aligned} \mu_\varepsilon(\mathscr{D}(y_1, \ldots, y_N) > a) &\leqslant \sum_{m_1, \ldots, m_N} \mu_{m_1, \ldots, m_N}(\mathscr{D}(y_1, \ldots, y_N) > \alpha + 2\varepsilon\|e\|) \\ &\leqslant \sum_{m_1, \ldots, m_N} \bar{\mu}\{(\bar{A}_{m_1} \times \cdots \times \bar{A}_{m_N}): \mathscr{D}(x_1, \ldots, x_N) > \alpha + \varepsilon\|e\|\} \\ &\leqslant \bar{\mu}(\mathscr{D}(y_1, \ldots, y_N) > \alpha) \leqslant \beta. \end{aligned}$$

Thus (II) $\rightarrow$ (I) as desired. QED

Let us apply Theorem 7.4.1 to the set $\mathfrak{X}(U)$ of random variables defined on a rich enough probability space (see Remark 2.5.1), taking values in the s.m.s. (U, d).

Given $\alpha > 0$ and a vector of laws $\tilde{P} = (P_1, \ldots, P_N) \in (\mathscr{P}(U))^N$, define

$$S_1(\tilde{P}; \alpha) = \inf\{\Pr(\mathscr{D}(X) > \alpha): X = (X_1, \ldots, X_N) \in \mathfrak{X}(U^N),\ \Pr_{X_i} = P_i,\ i = 1, \ldots, N\} \tag{7.4.9}$$

and

$$S_2(\tilde{P}; \alpha) = \sup\{P_1(B_1) + \cdots + P_{N-1}(B_{N-1}) - P_N(B_N^{(\alpha)}) - N + 2: \\ B_1, B_2, \ldots, B_{N-1} \in \mathscr{B}(U)\} \tag{7.4.10}$$

where $\mathscr{D}(x_1, \ldots, x_N) = \|d(x_1, x_2), \ldots, d(x_1, x_N), \ldots, d(x_{N-1}, x_N)\|$, $\|\cdot\|$ is a monotone seminorm and $B_N^{(\alpha)}$ is defined as in Theorem 7.4.1. Then the following duality theorem holds.

Corollary 7.4.1. For any $\alpha > 0$

$$S_1(\tilde{P}; \alpha) = S_2(\tilde{P}; \alpha). \tag{7.4.11}$$

If P_is are tight measures then the infimum in (7.4.9) is attained.

In the case $N = 1$ we obtain the Strassen–Dudley theorem.

Corollary 7.4.2. Let $\mathbf{K}_\lambda (\lambda > 0)$ be the Ky Fan metric (see (3.3.10)), and $\boldsymbol{\pi}_\lambda$ the Prokhorov metric (see (3.2.22)). Then $\boldsymbol{\pi}_\lambda$ is the minimal metric relative to $\mathbf{K}_\lambda$, i.e.

$$\hat{\mathbf{K}}_\lambda = \boldsymbol{\pi}_\lambda. \tag{7.4.12}$$

In particular, by the limit relations: $\boldsymbol{\pi}_\lambda \rightarrow_{\lambda \rightarrow 0} \ell_0 = \boldsymbol{\sigma}$ (see Lemma 3.2.1) and

$\mathbf{K}_\lambda \to_{\lambda\to 0} \mathscr{L}_0$ (see (3.3.11), (3.3.6)) we have that the minimal metric relative to the indicator metric $\mathscr{L}_0(X, Y) = \mathbb{E}I\{X \neq Y\}$ equals the total variation metric

$$\boldsymbol{\sigma}(X, Y) = \sup_{A \in \mathscr{B}(U)} |\Pr(X \in A) - \Pr(Y \in A)|$$

i.e., (Dobrushin, 1970) $\hat{\mathscr{L}}_0 = \boldsymbol{\sigma}$.

By the duality theorem 7.4.1, for any $\lambda > 0$ and $\tilde{P} = (P_1, \dots, P_N) \in \mathscr{P}(U)^N$,

$$\inf_{\substack{X \in \mathfrak{X}(U^N)\\ \Pr_{X_i} = P_i,\, i=1,\dots,N}} \mathscr{K}\mathscr{F}_\lambda(X) = \Pi_\lambda(\tilde{P}) \tag{7.4.13}$$

where $\mathscr{K}\mathscr{F}_\lambda$ is the *Ky Fan functional* in $\mathfrak{X}(U^N)$

$$\mathscr{K}\mathscr{F}_\lambda(X) := \inf\{\varepsilon > 0\colon \Pr(\mathscr{D}(X) > \lambda\varepsilon) \leqslant \varepsilon\}$$

and $\Pi_\lambda(\tilde{P})$ is the *Prokhorov functional in* $(\mathscr{P}(U))^N$ *with parameter* $\lambda > 0$

$$\Pi_\lambda(\tilde{P}) = \inf\{\varepsilon > 0\colon S_2(\tilde{P}, \lambda\varepsilon) \leqslant \varepsilon\}.$$

Letting $\lambda \to 0$ in (7.4.13) we obtain the following multivariate version of the Dobrushin (1970) duality theorem

$$\begin{aligned}
&\inf_{\substack{X \in \mathfrak{X}(U^N)\\ \Pr_{X_i} = P_i,\, i=1,\dots,N}} \Pr(X_i \neq X_j \quad \forall 1 \leqslant i < j \leqslant N)\\
&= \sup_{B_1,\dots,B_{N-1} \in \mathscr{B}(U)} \left[P_1(B_1) + \cdots + P_{N-1}(B_{N-1}) - P_N\left(\bigcap_{i=1}^{N-1} B_i\right) - N + 2\right]\\
&= \sup_{B_1,\dots,B_{N-1} \in \mathscr{B}(U)} \left[P_N\left(\bigcup_{i=1}^{N-1} B_i\right) - P_1(B_1) - \cdots - P_{N-1}(B_{N-1})\right].
\end{aligned} \tag{7.4.14}$$

Note that the above quantities are symmetric with respect to any rearrangement of the vector $\tilde{P}$.

Multiplying both sides of (7.4.13) by λ and then letting $\lambda \to \infty$ (or simply using (7.4.11)), we obtain

$$\inf_{\substack{X \in \mathfrak{X}(U^N)\\ \Pr_{X_i} = P_i,\, i=1,\dots,N}} \operatorname{ess\,sup} \mathscr{D}(X) = \inf\{\varepsilon > 0\colon S_2(\tilde{P}; \varepsilon) = 0\}.$$

Using the above equality for $N = 2$, we obtain that the minimal metric relative to $\mathscr{L}_\infty(X, Y) = \operatorname{ess\,sup} d(X, Y)$ (see (3.3.5), (3.3.7), (3.3.11)) is equal to ℓ_∞ (see (3.2.14) and Lemma 3.2.1), i.e.,

$$\hat{\mathscr{L}}_\infty = \ell_\infty. \tag{7.4.15}$$

Suppose that $d_1, \dots, d_n$ are metrics in U and that U is a separable metric space with respect to each d_i, $i = 1, \dots, n$. We introduce in U^n the metric

$$d_\Sigma(x, y) = \sum_{i=1}^{n} d_i(x_i, y_i),\ x = (x_1, \dots, x_n),\ y = (y_1, \dots, y_n) \in U^n. \tag{7.4.16}$$

We consider in $\mathfrak{X}(U^n)$ the metric $\tau_\Sigma(X, Y) := \mathbb{E} d_\Sigma(X, Y)$. Denote by $\kappa(X_i, Y_i; d_i)$ the Kantorovich metric in the space $\mathfrak{X}(U, d_i)$

$$\kappa(X_i, Y_i; d_i) = \sup\left\{ |\mathbb{E}[f(X_i) - f(Y_i)]| : \|f\|_L^{(i)} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_i(x, y)} \leqslant 1 \right\} \tag{7.4.17}$$

(see Example 3.2.2).

Theorem 7.4.2. Suppose that for $X = (X_1, \dots, X_n)$, $Y = (Y_1, \dots, Y_n) \in \mathfrak{X}(U^n)$, $\tau_\Sigma(X, a) + \tau_\Sigma(Y, a) < +\infty$ for some $a \in U^n$. Then

$$\hat{\tau}_\Sigma(X, Y) = \sum_{i=1}^{n} \kappa(X_i, Y_i; d_i). \tag{7.4.18}$$

Proof. By the Kantorovich theorem (Corollary 6.1.1) the minimal metric relative to the metric $\tau(X_i, Y_i; d_i) = \mathbb{E} d_i(X_i, Y_i)$ in $\mathfrak{X}(U)$ is $\kappa(X_i, Y_i; d_i)$. Hence

$$\hat{\tau}_\Sigma(X, Y) \geqslant \sum_{i=1}^{n} \hat{\tau}(X_i, Y_i; d_i) = \sum_{i=1}^{n} \kappa(X_i, Y_i; d_i). \tag{7.4.19}$$

Conversely, let $\mathfrak{L}_m^{(i)}$, $i = 1, 2, \dots$, be a sequence of joint distributions of random variables X_i, Y_i such that $\kappa(X_i, Y_i; d_i) = \lim_{m \to \infty} \tau(X_i, Y_i; d_i, \mathfrak{L}_m^{(i)})$, where $\tau(X_i, Y_i; d_i; \mathfrak{L}_m^{(i)})$ is the value of the metric τ for the joint distribution $\mathfrak{L}_m^{(i)}$. Then $\hat{\tau}_\Sigma(X, Y) \leqslant \sum_{i=1}^{n} \tau(X_i, Y_i; d_i, \mathfrak{L}_m^{(i)})$ and as $m \to +\infty$ we get the inequality

$$\hat{\tau}_\Sigma(X, Y) \leqslant \sum_{i=1}^{n} \kappa(X_i, Y_i; d_i). \tag{7.4.20}$$

Inequalities (7.4.19) and (7.4.20) imply Equality (7.4.18). QED

Corollary 7.4.3. For any $\alpha \in [0, 1]$

$$\hat{\tau}(X, Y; d^\alpha) = \sum_{i=1}^{n} \kappa(X_i, Y_i; d^\alpha) \quad \text{for } 0 < \alpha \leqslant 1 \tag{7.4.21}$$

$$\hat{\tau}(X, Y; d^0) = \sum_{i=1}^{n} \sigma(X_i, Y_i). \tag{7.4.22}$$

The proof of (7.4.21) follows from (7.4.18) if we set $d_i = d^\alpha$. Equality (7.4.22) follows from Equality (7.4.21) as $\alpha \to 0$.

7.5 RELATIONS BETWEEN MULTI-DIMENSIONAL KANTOROVICH AND STRASSEN THEOREMS; CONVERGENCE OF MINIMAL METRICS AND MINIMAL DISTANCES

Recall the multi-dimensional Kantorovich theorem (see (5.2.1), (5.2.2) and (5.2.4))

$$A_D(\tilde{P}) = \mathbb{K}(\tilde{P}) := \mathbb{K}(\tilde{P}, \mathfrak{G}(U)) \tag{7.5.1}$$

where $\tilde{P} = (\tilde{P}_1, \dots, \tilde{P}_N) \in (\mathscr{P}(U))^N$

$$A_D(\tilde{P}) = \inf\left\{\int_{U^N} D \, \mathrm{d}P,\ P \in \mathfrak{P}(\tilde{P})\right\}, \quad D = H(\mathscr{D}). \tag{7.5.2}$$

In the above minimal functional $\mathscr{D}(x)$ is given by

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

$\|\cdot\|$ is a monotone seminorm on $\mathbb{R}^m$, $m = \binom{N}{2}$, and $\mathfrak{P}(\tilde{P})$ is the space of all Borel probability measures P on U^N with fixed one-dimensional marginals $P_1, \dots, P_N$ (see Section 5.2).

Next we turn our attention to the relationship between (7.5.1) and the multi-dimensional Strassen theorem (cf. (7.4.13)).

Theorem 7.5.1. Suppose that (U, d) is a s.m.s.

$$\mathscr{K}\mathscr{F}(P) = \inf\{\alpha > 0: P(\mathscr{D}(x) > \alpha) < \alpha\} \tag{7.5.3}$$

is the *Ky Fan functional* in $\mathscr{P}(U^N)$ and

$$\Pi(\tilde{P}) = \inf\{\alpha > 0: P_1(B_1) + \dots + P_{N-1}(B_{N-1}) \leqslant P_N(B^{(\alpha)}) + \alpha + N - 2 \text{ for all } B_1, \dots, B_{N-1}, \text{ Borel subsets of } U\} \tag{7.5.4}$$

is the *Prokhorov functional* in $(\mathscr{P}(U))^N$, where $B^{(\alpha)} = \{x_N \in U: \mathscr{D}(x_1, \dots, x_N) \leqslant \alpha$ for some $x_1 \in B_1, \dots, x_{N-1} \in B_{N-1}\}$. Then

$$\inf\{\mathscr{K}\mathscr{F}(P): P \in \mathfrak{P}(\tilde{P})\} = \Pi(\tilde{P}) \tag{7.5.5}$$

and if $\tilde{P}$ is a set of tight measures, then the infimum is attained in (7.5.3).

The next inequality represents the main relationship between the Kantorovich functional $A_D(\tilde{P})$ and the Prokhorov functional $\Pi(\tilde{P})$.

Theorem 7.5.2. For any $H \in \mathscr{H}^*$ (i.e., $H \in \mathscr{H}$, Example 2.2.1, and H is convex), $M > 0$ and $a \in U$

$$\Pi(\tilde{P})H(\Pi(\tilde{P})) \leqslant \mathbb{K}(\tilde{P}) \leqslant H(\Pi(\tilde{P})) + c_1 H(M)\Pi(\tilde{P}) + c_2 \sum_{i=1}^{N} \int_U H(d(x, a)) I(d(x, a) > M) P_i(\mathrm{d}x) \tag{7.5.6}$$

where $c_2 := K_H^\ell$ (see (2.2.3)), $\ell := [\log_2(A_m N^2)] + 1$, $c_1 = Nc_2$, $[x]$ is the integer part of x and

$$A_m := \max_{1 \leqslant j \leqslant m} \{\|(i_1, \dots, i_m)\| : i_k = 0, k \neq j, i_j = 1\} \quad m = \binom{N}{2}.$$

Proof. For any probability measure P on U^N and $\varepsilon > 0$ the inequality $\int_{U^N} H(\mathscr{D}(x))P(\mathrm{d}x) < \delta = \varepsilon H(\varepsilon)$ follows from $P(\mathscr{D}(x) > \varepsilon) < \varepsilon$; hence,

$$\mathscr{K}\mathscr{F}(P) \cdot H(\mathscr{K}\mathscr{F}(P)) \leqslant \int_{U^N} H(\mathscr{D}(x))P(\mathrm{d}x).$$

From (7.5.1), (7.5.2) and (7.5.5), it follows that $\Pi(\tilde{P})H(\Pi(\tilde{P})) \leqslant A_D(\tilde{P})$. We shall now prove the right-hand side inequality in (7.5.6). Given $\mathscr{K}\mathscr{F}(P) < \delta$ and $a \in U$ we have

$$\begin{aligned}
\int H(\mathscr{D}(x))P(\mathrm{d}x) &= \left(\int_{\mathscr{D}(x) \leqslant \delta} + \int_{\mathscr{D}(x) > \delta}\right) H(\mathscr{D}(x))P(\mathrm{d}x) \\
&\leqslant H(\delta) + \int_{\mathscr{D}(x) > \delta} H\left(A_m \sum_{i<j} d(x_i, x_j)\right) P(\mathrm{d}x) \\
&\leqslant H(\delta) + \int_{\mathscr{D}(x) > \delta} H\left(A_m N^2 \max_{1 \leqslant i \leqslant N} d(x_i, a)\right) P(\mathrm{d}x)
\end{aligned}$$

(by (2.2.3), $H(2^k t) \leqslant K_H^k H(t)$)

$$\leqslant H(\delta) + K_H^\ell \sum_{i=1}^N I_i$$

where

$$\begin{aligned}
I_i &:= \int_{\mathscr{D}(x) > \delta} H(d(x_i, a))P(\mathrm{d}x) \\
&= \left(\int_{\mathscr{D}(x) > \delta, d(x_i, a) > M} + \int_{\mathscr{D}(x) > \delta, d(x_i, a) \leqslant M}\right) H(d(x_i, a))P(\mathrm{d}x) \\
&\leqslant \int_{d(x_i, a) \geqslant M} H(d(x_i, a))P(\mathrm{d}x) + H(M)\delta.
\end{aligned}$$

Hence,

$$\begin{aligned}
\int H(\mathscr{D}(x))P(\mathrm{d}x) &\leqslant H(\mathscr{K}\mathscr{F}(P)) + c_2 N H(M) \mathscr{K}\mathscr{F}(P) \\
&\quad + c_2 \sum_{i=1}^N \int_{d(x_i, a) > M} H(d(x_i, a))P_i(\mathrm{d}x)).
\end{aligned}$$

Together with (7.5.1) and (7.5.5) the latter inequality yields the required estimate (7.5.6). QED

The inequality (7.5.6) provides a 'merging' criterion for a sequence of vectors $\tilde{P}^{(n)} = (P_1^{(n)}, \ldots, P_N^{(n)})$.

As in Diaconis and Freedman (1984b), D'Aristotile *et al.* (1988), Dudley (1989) (Section 11.7), we call two sequences $\{P^{(n)}\}_{n\geqslant 1}$, $\{Q^{(n)}\}_{n\geqslant 1} \in \mathscr{P}(U)$ *μ-merging* where μ is a simple p. metric if $\mu(P^{(n)}, Q^{(n)}) \to 0$ as $n \to \infty$. More generally, we say *the sequence $\{\tilde{P}^{(n)}\}_{n\geqslant 1} \subset (\mathscr{P}(U))^N$ is μ-merging* if

$$\mu(P_i^{(n)}, P_j^{(n)}) \to 0 \quad \text{as } n \to \infty$$

for any $i, j = 1, \ldots, N$.

The next corollary gives criteria for merging with respect to the Prokhorov metric $\boldsymbol{\pi}$ and the minimal distance ℓ_H (3.2.10).

Corollary 7.5.1. Let $\{\tilde{P}^{(n)}\}_{n\geqslant 1} \subset (\mathscr{P}(U))^N$. Then the following hold:

(i) $\{\tilde{P}^{(n)}\}_{n\geqslant 1}$ is $\boldsymbol{\pi}$-merging if and only if

$$\Pi(\tilde{P}^{(n)}) \to 0 \quad \textit{as } n \to \infty. \tag{7.5.7}$$

(ii) if $H \in \mathscr{H}^*$ and $\int H(d(x, a))P_i(\mathrm{d}x) < \infty$, $i = 1, \ldots, N$ then $\{\tilde{P}^{(n)}\}_{n\geqslant 1}$ is ℓ_H-merging if and only if

$$\mathbb{K}(\tilde{P}^{(n)}) \to 0 \quad \text{as } n \to \infty.$$

Proof.

(i) There exist constants C_1 and C_2 depending on the seminorm $\|\cdot\|$ such that

$$C_1 \sum_{1\leqslant i\leqslant j\leqslant N} \mathbf{K}(T_{ij}P) \leqslant \mathscr{K}\mathscr{F}(P) \leqslant C_2 \sum_{1\leqslant i\leqslant j<N} \mathbf{K}(T_{ij}P)$$

where $\mathbf{K}$ is the Ky Fan distance in $\mathscr{P}_i(U)$ (cf. Example 3.3.2). Now, Theorem 7.5.1 can be used to yield the assertion.

(ii) The same argument is applied. Here we make use of the multi-dimensional Kantorovich theorem (7.5.1). QED

Theorem 7.5.2 and the last corollary show that ℓ_H-merging implies π-merging. On the other hand, if

$$\lim_{M\to\infty} \max_{n\geqslant 1, 1\leqslant i\leqslant N} \int H(d(x, a))I\{d(x, a) > M\}P_i^{(n)}(\mathrm{d}x) = 0$$

then ℓ_H-merging and $\boldsymbol{\pi}$-merging of $\{\tilde{P}^{(n)}\}_{n\geqslant 1}$ are equivalent.

Regarding the K-minimal metric $\hat{\mathfrak{r}}_\Sigma$ (see (7.4.16) and (7.4.18)), we have the following criterion for the $\hat{\mathfrak{r}}_\Sigma$-convergence.

Corollary 7.5.2. Given $X^{(k)} = (X_1^{(k)}, \dots, X_n^{(k)}) \in \mathfrak{X}(U^n)$ such that $\mathbb{E}d_j(X_j^{(k)}, a) < \infty$, $j = 1, \dots, n$, $k = 0, 1, \dots,$ the convergence $\hat{\mathfrak{t}}_\Sigma(X^{(k)}, X^{(0)}) \to 0$ as $k \to \infty$ is equivalent to convergence in distributions, $X_j^{(k)} \xrightarrow{w} X_j^{(0)}$, and the moment convergence $\mathbb{E}d_j(X_j^{(k)}, a) \to \mathbb{E}d_j(X_j^{(0)}, a)$ $\forall j = 1, \dots, n$.

The latter corollary is a consequence of Theorem 7.4.2 and Theorem 6.3.1 (for $\lambda(x) = d(x, a)$, $c(x, y) = d(x, y)$).

To conclude, we turn our attention to the inequalities between minimal distances $\hat{\mathscr{L}}_H$, Kantorovich distance ℓ_H (see (3.2.10), (3.2.15), (5.2.17)) and the Prokhorov metric π (see (3.2.18)).

Corollary 7.5.3. (i) For any $H \in \mathscr{H}$, $M > 0$, $a \in U$ and $P_1, P_2 \in \mathscr{P}(U)$ such that

$$\int H(d(x, a))(P_1 + P_2)(\mathrm{d}x) < \infty \tag{7.5.8}$$

the following inequality holds:

$$\begin{aligned} H(\pi(P_1, P_2))\pi(P_1, P_2) &\leqslant \hat{\mathscr{L}}_H(P_1, P_2) \\ &\leqslant H(\pi(P_1, P_2)) + K_H\left[2\pi(P_1, P_2)H(M) + \int_{d(x,a)>M} H(d(x, a))(P_1 + P_2)(\mathrm{d}x)\right]. \end{aligned} \tag{7.5.9}$$

If $H \in \mathscr{H}$ is a convex function, then one can replace $\hat{\mathscr{L}}_H$ with ℓ_H in (7.5.9).

(ii) Given a sequence $P_0, P_1, \dots \in \mathscr{P}(U)$ with $\int H(d(x, a))P_j(\mathrm{d}x) < \infty$, ($j = 0, 1, \dots$) the following assertions are equivalent as $n \to \infty$:

(a) $\hat{\mathscr{L}}_H(P_n, P_0) \to 0$,

(b) P_n converges weakly to $P(P_n \xrightarrow{w} P)$ and $\displaystyle\int H(d(x, a))(P_n - P)(\mathrm{d}x) \to 0$,

(c) $P_n \xrightarrow{w} P$ and $\displaystyle\lim_{N\to\infty} \limsup_n \int H(d(x, a))I\{d(x, a) > N\}P_n(\mathrm{d}x) = 0$.

This theorem is a particular case of more general theorems (see further Theorems 8.2.1 and 10.1.1).

CHAPTER 8

Relations between Minimal and Maximal Distances

The metric structure of the functionals $\hat{\mu}_c$, $\mathring{\mu}$, $\check{\mu}_c$, $\overset{(s)}{\mu}$ has been discussed in Chapter 3 (see Fig. 3.3.1). Dual and explicit representations for the minimal distance $\hat{\mu}_c$ and minimal norm $\mathring{\mu}_c$ have been found in Chapter 6 for some special choice of the function c. Here we shall deal mainly with the following two questions:

(1) What are the dual representations and the explicit forms of $\check{\mu}_c$, $\overset{(s)}{\mu}_c$?

(2) What are the necessary and sufficient conditions for $\lim_{n\to\infty} \hat{\mu}_c(P_n, P) = 0$, resp. $\lim_{n\to\infty} \mathring{\mu}_c(P_n, P) = 0$?

8.1 DUALITY THEOREMS AND EXPLICIT REPRESENTATIONS FOR $\check{\mu}_c$ AND $\overset{(s)}{\mu}_c$

First of all let us consider the dual form for the maximal distance $\check{\mu}_c$ and $\overset{(s)}{\mu}_c$ and let us compare them with the corresponding dual representations for the minimal metric $\hat{\mu}_c$ and minimal norm $\mathring{\mu}_c$ (see Definitions 3.2.2, 3.2.5, 3.3.4 and 3.3.5). Recall that

$$\mathring{\mu}_c(P_1, P_2) \leqslant \hat{\mu}_c(P_1, P_2) \leqslant \check{\mu}_c(P_1, P_2) \leqslant \overset{(s)}{\mu}_c(P_1, P_2). \tag{8.1.1}$$

In the following, we shall use the following notation

$$L_\alpha = \{f\colon U \to \mathbb{R}^1; |f(x) - f(y)| \leqslant \alpha d(x, y),\ x, y \in U\}$$

$$\mathrm{Lip} := \bigcup_{\alpha > 0} L_\alpha$$

$$\mathrm{Lip}^b := \{f \in \mathrm{Lip}\colon \sup\{|f(x)|\colon x \in U\} < \infty\}$$

$$c(x, y) := H(d(x, y)),\ x, y \in U,\ H \in \mathscr{H} \quad \text{(see Example 2.2.1)}$$

$$\mathscr{P}_H := \left\{P \in \mathscr{P}(U)\colon \int c(x, a)P(\mathrm{d}x) < \infty\right\}$$

$$\underline{\mathscr{G}}_H := \{(f, g)\colon f, g \in \mathrm{Lip}^b,\ f(x) + g(y) \leqslant c(x, y),\ x, y \in U\} \tag{8.1.2}$$

$$\overline{\mathscr{G}}_H := \{(f, g): f, g \in \mathrm{Lip}^b, f(x) \geqslant 0, g(y) \geqslant 0, f(x) + g(y) \geqslant c(x, y)\ x, y \in U\} \tag{8.1.3}$$

$$h(x, y) := d(x, y)h_0(d(x, a) \vee d(y, a)) \quad x, y \in U, \ \vee := \max \tag{8.1.4}$$

where a is a fixed point of U and h_0 is a non-negative, non-decreasing, continuous function on $[0, \infty)$

$$\mathrm{Lip}_h := \{f: U \to \mathbb{R}^1: |f(x) - f(y)| \leqslant h(x, y), x, y \in U\}$$

$$\mathscr{H}^* := \{\text{convex } H \in \mathscr{H}\}$$

$$\mathscr{F} := \{f \in \mathrm{Lip}^b: f(x) + f(y) \geqslant c(x, y), x, y \in U\} \tag{8.1.5}$$

and

$$\mathbb{T}(P_1, P_2; \mathscr{F}) := \inf\left\{\int f\, \mathrm{d}(P_1 + P_2): f \in \mathscr{F}\right\}. \tag{8.1.6}$$

Theorem 8.1.1. Let (U, d) be a s.m.s.

(i) If $H \in \mathscr{H}^*$ and $P_1, P_2 \in \mathscr{P}_H$ then the minimal distance,

$$\hat{\mu}_c(P_1, P_2) := \inf\{\mu_c(P): P \in \mathscr{P}(U \times U), T_i P = P_i, i = 1, 2\} \tag{8.1.7}$$

relative to the compound distance,

$$\mu_c(P) = \int_{U \times U} c(x, y)P(\mathrm{d}x, \mathrm{d}y) \tag{8.1.8}$$

admits the dual representation

$$\hat{\mu}_c(P_1, P_2) = \sup\left\{\int f\, \mathrm{d}P_1 + \int g\, \mathrm{d}P_2: (f, g) \in \mathscr{G}_H\right\}. \tag{8.1.9}$$

If P_1 and P_2 are tight measures then the infimum in (8.1.7) is attained.

(ii) If $\int h(x, a)(P_1 + P_2)(\mathrm{d}x) < \infty$ then the minimal norm

$$\mathring{\mu}_h(P_1, P_2) := \inf\{\mu_h(m): m\text{-bounded non-negative measures with fixed } T_1 m - T_2 m = P_1 - P_2\} \tag{8.1.10}$$

has a dual form

$$\mathring{\mu}_h(P_1, P_2) = \sup\left\{\left|\int f\, \mathrm{d}(P_1 - P_2)\right|: f \in \mathrm{Lip}_h\right\} \tag{8.1.11}$$

and the supremum in (8.1.11) is attained.

(iii) If $H \in \mathscr{H}^*$ and $P_1, P_2 \in \mathscr{P}_H$ then the maximal distance

$$\check{\mu}(P_1, P_2) := \sup\{\mu_c(P_1, P_2): P \in \mathscr{P}(U \times U), T_i P = P_i, i = 1, 2\} \tag{8.1.12}$$

has the dual representation

$$\check{\mu}_c(P_1, P_2) = \inf\left\{\int f\,\mathrm{d}P_1 + \int g\,\mathrm{d}P_2 \colon (f, g) \in \overline{\mathscr{G}}_H\right\}. \tag{8.1.13}$$

If P_1 and P_2 are tight measures then the supremum in (8.1.12) is attained.

(iv) If $H \in \mathscr{H}^*$ and $P_1, P_2 \in \mathscr{P}_H$ then

$$\overset{(s)}{\mu}_c(P_1, P_2) := \overset{(s)}{\mu}_c(P_1 + P_2)$$
$$:= \sup\{\mu(P) \colon P \in \mathscr{P}(U \times U),\ T_1P + T_2P = P_1 + P_2\} \tag{8.1.14}$$

has the dual representation

$$\overset{(s)}{\mu}_c(P_1, P_2) = \mathbb{T}(P_1, P_2; \mathscr{F}). \tag{8.1.15}$$

If P_1 and P_2 are tight measures then the supremum in (8.1.14) is attained.

Proof. (i) This is Corollary 5.2.2.

(ii) This is a special case of Theorems 5.3.2 and 5.3.3 with $c(x, y) = h(x, y)$ given by (8.1.9).

(iii) The proof here is quite similar to that of Corollary 5.2.2 and Theorem 5.2.1 and thus it is omitted.

(iv) For any probability measures P_1 and P_2 on U, any $P \in \mathscr{P}(U \times U)$ with fixed sum of marginals, $T_1P + T_2P = P_1 + P_2$, and any $f \in \mathscr{F}$ (see (8.1.5)) we have

$$\int f\,\mathrm{d}(P_1 + P_2) = \int f\,\mathrm{d}(T_1P + T_2P) = \int f(x) + f(y)P(\mathrm{d}x, \mathrm{d}y) \geqslant \int c(x, y)P(\mathrm{d}x, \mathrm{d}y)$$

hence

$$\overset{(s)}{\mu}_c(P_1 + P_2) \leqslant \mathbb{T}(P_1, P_2; \mathscr{F}). \tag{8.1.16}$$

Our next step is to prove the inequality

$$\overset{(s)}{\mu}_c(P_1 + P_2) \geqslant \mathbb{T}(P_1, P_2; \mathscr{F}) \tag{8.1.17}$$

and here we shall use the main idea of the proof of Theorem 5.2.1. To prove (8.1.17) we first treat

Case A. (U, d) is a bounded s.m.s. For any subset $U_1 \subset U$ define

$$\bar{\mathscr{F}}(U_1) = \{f \colon U \to \mathbb{R}^1,\ f(x) + f(y) \geqslant c(x, y) \text{ for all } x, y \in U_1\}$$
$$\mathscr{F}(U_1) = \bar{\mathscr{F}}(U_1) \cap \mathrm{Lip}_\tau(U_1)$$

where $\mathrm{Lip}_\tau(U_1) := \{f \colon U \to \mathbb{R}^1 \colon |f(x) - f(y)| \leqslant \tau(x, y),$ for all $x, y \in U_1\}$ and $\tau(x, y) := \sup\{|c(x, z) - c(y, z)| \colon z \in U\}$, $x, y \in U$. We need the following equality: if $P_1(U_1) = P_2(U_1) = 1$ then

$$\mathbb{T}(P_1, P_2; \bar{\mathscr{F}}(U_1)) = \mathbb{T}(P_1, P_2; \mathscr{F}(U)). \tag{8.1.18}$$

Let $f \in \mathscr{F}(U_1)$. We extend f to a function on the whole U letting $f(x) = \infty$ for $x \notin U_1$ and hence

$$f(x) \geqslant f^*(x) := \sup\{c(x, y) - f(y) : y \in U\} \quad \forall x \in U. \tag{8.1.19}$$

Since for any $x, y \in U$

$$\begin{aligned} f^*(x) - f^*(y) &= \sup_{z \in U}\{c(x, z) - f(z)\} - \sup_{w \in U}\{c(y, w) - f(w)\} \\ &\leqslant \sup_{z \in U}\{c(x, z) - c(y, z)\} \leqslant \tau(x, y) \end{aligned}$$

then $f^* \in \mathscr{F}(U)$. Moreover, if $P_1(U_1) = P_2(U_1) = 1$ then by (8.1.19)

$$\mathbb{T}(P_1, P_2; \mathscr{F}(U_1)) \geqslant \mathbb{T}(P_1, P_2; \mathscr{F}(U)) \tag{8.1.20}$$

which yields (8.1.18).

Case A1. Let (U, d) be a finite set, say, $U = \{u_1, \ldots, u_n\}$. By (8.1.18) and the duality theorem in the linear programming we obtain

$$\overset{(s)}{\mathring{\mu}}_c(P_1 + P_2) = \mathbb{T}(P_1, P_2; \mathscr{F}(U)) = \mathbb{T}(P_1, P_2; \mathscr{F}(U)) \tag{8.1.21}$$

as desired.

The remaining cases A2 ((U, d) is a compact space), A3 ((U, d) is a bounded s.m.s.) and $B((U, d)$ is a s.m.s.) are treated in a way quite similar to that in Theorem 5.2.1. QED

In the special case $c = d$ one can get more refined duality representations for $\overset{(s)}{\mathring{\mu}}_c$. This is the following corollary.

Corollary 8.1.1. If (U, d) is a s.m.s. and $P_1, P_2 \in \mathscr{P}(U)$, $\int d(x\, a)(P_1 + P_2)(\mathrm{d}x) < \infty$, then

$$\overset{(s)}{\mathring{\mu}}_d(P_1, P_2) = \inf\left\{\int f\,\mathrm{d}(P_1 + P_2) : f \in L_1,\ f(x) + f(y) \geqslant \mathrm{d}(x, y) \quad \forall x, y \in U\right\}. \tag{8.1.22}$$

Here the proof is a copy of the proof of (iv) in Theorem 8.1.1 with some simplifications due to the fact that $c = d$.

Open problem 8.1.1. Let us compare the dual forms of $\hat{\mu}_d$, $\mathring{\mu}_d$, $\check{\mu}_d$ and $\overset{(s)}{\mathring{\mu}}_d$. The Kantorovich metric $\hat{\mu}_d$ in the space $\mathscr{P}^1$ of all measures P with finite moment

$\int d(x, a)P(\mathrm{d}x) < \infty$ has two dual representations

$$\hat{\mu}_d(P_1, P_2) = \sup\left\{\int f\,\mathrm{d}P_1 + \int g\,\mathrm{d}P_2 : f, g \in L,\ f(x) + g(y) \leqslant d(x, y),\ x, y \in U\right\}$$

$$= \sup\left\{\int f\,\mathrm{d}(P_1 - P_2) : f \in L_1\right\} = \mathring{\mu}_d(P_1, P_2) \tag{8.1.23}$$

(see Section 6.1, (5.3.15) and (8.1.9)). On the other hand, by (8.1.13) and (8.1.16), a dual form of $\check{\mu}_d$ is

$$\check{\mu}_d(P_1, P_2) = \inf\left\{\int f\,\mathrm{d}P_1 + \int g\,\mathrm{d}P_2 : f, g \in L_1, f(x) + g(y) \geqslant d(x, y) \quad \forall x, y \in U\right\} \tag{8.1.24}$$

which corresponds to the first expression for $\hat{\mu}_d$ in (8.1.23) so an open problem is to check whether the equality

$$\check{\mu}_d(P_1, P_2) = \overset{(s)}{\mu}_d(P_1, P_2) \tag{8.1.25}$$

holds (here, $\overset{(s)}{\mu}_d$ is given by (8.1.22)). In the special case $(U, d) = (\mathbb{R}, |\cdot|)$ the equality (8.1.25) is true (see further Remark 8.1.1).

Next we shall concern ourselves with the explicit representations for $\hat{\mu}_c$, $\mathring{\mu}_c$, $\check{\mu}_c$ and $\overset{(s)}{\mu}_c$ in the case $U = \mathbb{R}$, $d(x, y) = |x - y|$.

Suppose ϕ: $\mathbb{R}^2 \to \mathbb{R}$ is a quasiantitone upper-semicontinuous function (cf. Section 7.3). Then $\hat{\mu}_\phi$, $\check{\mu}_\phi$ and $\overset{(s)}{\mu}_\phi$ have the following representations:

Lemma 8.1.1. Given P_1 and $P_2 \in \mathscr{P}(\mathbb{R})$ with finite moments $\int \phi(x, a)\,\mathrm{d}P_i(x) < \infty$, $i = 1, 2$, we have:

(i) (Cambanis–Simons–Stout)

$$\hat{\mu}_\phi(P_1, P_2) = \int_0^1 \phi(F_1^{-1}(t), F_2^{-1}(t))\,\mathrm{d}t \tag{8.1.26}$$

where F_i is the d.f. of P_i and

$$\check{\mu}_\phi(P_1, P_2) = \int_0^1 \phi(F_1^{-1}(t), F_2^{-1}(1 - t))\,\mathrm{d}t. \tag{8.1.27}$$

(ii) Assuming that $\phi(x, y)$ is symmetric,

$$\overset{(s)}{\mu}_\phi(P_1 + P_2) = \int_0^1 \phi(A(t), A(1 - t))\,\mathrm{d}t \tag{8.1.28}$$

where $A(t) = \frac{1}{2}(F_1(t) + F_2(t))$.

Proof. (i) The Equality (8.1.26) follows from Theorem 7.3.2 (with $N = 2$). Analogously, one can prove (8.1.27). Namely, let $\mathscr{F}(F_1, F_2)$ be the set of all d.f.s F on $\mathbb{R}^2$ with marginals F_1 and F_2. By the well known Hoeffding–Fréchet inequality, $\mathscr{F}(F_1, F_2)$ has a lower bound

$$F_-(x_1, x_2) := \max(0, F_1(x_1) + F_2(x_2) - 1),\ F_- \in \mathscr{F}(F_1, F_2) \tag{8.1.29}$$

and a lower bound

$$F_+(x_1, x_2) = \min(F_1(x_1), F_2(x_2)),\ F_+ \in \mathscr{F}(F_1, F_2). \tag{8.1.30}$$

Consider the space $\mathfrak{X}(\mathbb{R})$ of all random variables on a non-atomic probability space (see Remark 2.5.2). Then

$$\hat{\mu}_\phi(P_1, P_2) = \inf\{\mathbb{E}\phi(X_1, X_2) \colon X_i \in \mathfrak{X}(\mathbb{R}),\ F_{X_i} = F_i,\ i = 1, 2\} \tag{8.1.31}$$

$$\check{\mu}_\phi(P_1, P_2) = \sup\{\mathbb{E}\phi(X_1, X_2) \colon X_i \in \mathfrak{X}(\mathbb{R}),\ F_{X_i} = F_i,\ i = 1, 2\}. \tag{8.1.32}$$

If E is a $(0, 1)$-uniformly distributed r.v. then $F_-(x_1, x_2) = P(X_1^- \leqslant x_1, X_2^- \leqslant x_2)$ where $X_1^- := F_1^{-1}(E)$, $X_2^- := F_2^{-1}(1 - E)$ and $F_i^{-1}(u) := \inf\{t \colon F_i(t) \geqslant u\}$ is the generalized inverse function to F_i. Similarly, $F_+(x_1, x_2) = P(X_1^+ \leqslant x_1, X_2^+ \leqslant x_2)$ where $X_i^+ = F_i^{-1}(E)$, $i = 1, 2$. Thus

$$\check{\mu}_\phi(P_1, P_2) \geqslant \mathbb{E}\phi(X_1^-, X_2^-) = \int_0^1 \phi(F_1^{-1}(t), F_2^{-1}(1 - t))\, dt \tag{8.1.33}$$

and

$$\hat{\mu}_\phi(P_1, P_2) \leqslant \mathbb{E}\phi(X_1^+, X_2^+) = \int_0^1 \phi(F_1^{-1}(t), F_2^{-1}(t))\, dt. \tag{8.1.34}$$

In Theorem 7.3.2 (in the special case $N = 2$), we showed that (8.1.34) is true with equality sign. Using the same method one can check that $\check{\mu}_\phi(P_1, P_2) = \mathbb{E}\phi(X_1^-, X_2^-)$, see Kalashnikov and Rachev (1988), Theorem 7.1.1.

(ii) From the definition of $\overset{(s)}{\check{\mu}}(P_1, P_2)$ (see (8.1.14)) it follows that

$$\begin{aligned}\overset{(s)}{\check{\mu}}_\phi(P_1 + P_2) &= \overset{(s)}{\check{\mu}}_\phi(F_1 + F_2)\\ &:= \sup\{\mathbb{E}\phi(X_1, X_2) \colon X_1, X_2 \in \mathfrak{X}(\mathbb{R}),\ F_{X_1} + F_{X_2} = F_1 + F_2 =: 2A\}\end{aligned}$$

or in other words,

$$\overset{(s)}{\check{\mu}}_\phi(F_1 + F_2) = \sup\left\{\int_{\mathbb{R}^2} \phi(x, y)\, dF(x, y) \colon F \in \mathscr{F}(F_1, F_2),\ \tfrac{1}{2}(F_1 + F_2) = A\right\}.$$

For any $F \in \mathscr{F}(F_1, F_2)$ denote $\tilde{F}(x, y) = \frac{1}{2}[F(x, y) + F(y, x)]$; then, by the symmetry of $\phi(x, y)$,

$$\overset{(s)}{\mu}_\phi(F_1, F_2) = \sup\left\{\int_{\mathbb{R}^2} \phi(x, y)\,\mathrm{d}\tilde{F}(x, y) \colon \tilde{F} \in \mathscr{F}(A, A)\right\}$$

$$= \check{\mu}_\phi(A, A) = \int_0^1 \phi(A^{-1}(t), A^{-1}(1-t))\,\mathrm{d}t. \qquad \text{QED}$$

Remark 8.1.1. It is easy to see that for any symmetric cost function c

$$\overset{(s)}{\mu}_c(P_1 + P_2) = \check{\mu}_c(\tfrac{1}{2}(P_1 + P_2), \tfrac{1}{2}(P_1 + P_2)), \quad P_i \in \mathscr{P}(U). \tag{8.1.35}$$

On the other hand, in the case $U = \mathbb{R}$, $c(x, y) = |x - y|$, by Lemma 8.1.1.

$$\check{\mu}_c(P_1, P_2) = \int_0^1 |F_1^{-1}(t) - F_2^{-1}(1-t)|\,\mathrm{d}t$$

$$= \int_{-\infty}^{\infty} |x - a|\,\mathrm{d}(F_1(x) + F_2(x)) \tag{8.1.36}$$

where a is the point of 'intersection' of the graphs of F_1 and $1 - F_2$, i.e. $F_1(a - 0) \leqslant 1 - F_2(a - 0)$ but $F_1(a + 0) \geqslant 1 - F_2(a + 0)$. Hence, by (8.1.1) and (8.1.37)

$$\begin{aligned}\overset{(s)}{\mu}_c(P_1, P_2) &\geqslant \hat{\mu}_c(P_1, P_2)\\ &= \sup\{\mathbb{E}|X_1 - a| + \mathbb{E}|X_2 - a| \colon X_1, X_2 \in \mathfrak{X}(\mathbb{R}), F_{X_1} + F_{X_2} = F_1 + F_2\}\\ &\geqslant \overset{(s)}{\mu}_c(P_1, P_2)\end{aligned}$$

i.e., $\overset{(s)}{\mu}_c = \hat{\mu}_c$.

By virtue of Lemma 8.1.1 (with $\phi(x, y) = c(x, y) := H(|x - y|)$, H convex on $[0, \infty)$), we obtain the following explicit expressions for $\hat{\mu}_c$, $\check{\mu}_c$, $\overset{(s)}{\mu}_c$ and $\mathring{\mu}_h$.

Theorem 8.1.2. (i) Suppose $P_1, P_2 \in \mathscr{P}(\mathbb{R})$ have finite H-absolute moments, $\int H(|x|)(P_1 + P_2)(\mathrm{d}x) < \infty$, where $H \in \mathscr{H}^*$. Then

$$\hat{\mu}_c(P_1, P_2) = \int_0^1 c(F_1^{-1}(t), F_2^{-1}(t))\,\mathrm{d}t \tag{8.1.37}$$

$$\check{\mu}_c(P_1, P_2) = \int_0^1 c(F_1^{-1}(t), F_2^{-1}(1-t))\,\mathrm{d}t \tag{8.1.38}$$

and

$$\overset{(s)}{\check{\mu}}_c(P_1, P_2) = \int_0^1 c(A^{-1}(t), A^{-1}(1-t))\,\mathrm{d}t \tag{8.1.39}$$

where F_i is the d.f. of P_i, F_i^{-1} is the inverse of F_i and $A = \frac{1}{2}(F_1 + F_2)$.

(ii) Suppose $h\colon \mathbb{R}^2 \to \mathbb{R}$ is given by (8.1.4), where $d(x, y) = |x - y|$ and $h(t) > 0$ for $t > 0$. Then

$$\mathring{\mu}_h(P_1, P_2) = \int_{-\infty}^{\infty} h(|x - a|)|F_1(x) - F_2(x)|\,\mathrm{d}x. \tag{8.1.40}$$

8.2 CONVERGENCE OF MEASURES WITH RESPECT TO MINIMAL DISTANCES AND MINIMAL NORMS

In this section we investigate the topological structure of minimal distances $(\hat{\mu}_c)$ and minimal norms $\mathring{\mu}_h$ defined as in Section 8.1.

First, note that the definition of a simple distance v (say $v = \hat{\mu}_c$ or $v = \mathring{\mu}_h$) does not exclude infinite values of v. Hence, the space $\mathscr{P}_1 = \mathscr{P}(U)$ of all laws P on a s.m.s. (U, d) is divided into the classes $\mathscr{D}(v, P_0) := \{P \in \mathscr{P}_1 : v(P, P_0) < \infty\}$, $P_0 \in \mathscr{P}_1$ with respect to the equivalence relation $P_1 \sim P_2 \Leftrightarrow v(P_1, P_2) < \infty$. In the previous Sections 6.2, 6.3 and 7.5 the 'topological' structure of the Kantorovich distance $\hat{\mathscr{L}}_H = \hat{\mu}_c$, where $c(x, y) = H(d(x, y))$, $H \in \mathscr{H}$ (see Example 3.2.2 and (5.2.17)) was analyzed *only* in the set $\mathscr{D}(\hat{\mu}_c, \delta_\alpha)$, $\alpha \in U$, where $\delta_\alpha(\{\alpha\}) = 1$. Here we shall consider the $\hat{\mu}_c$ convergence in the following sets: $\mathscr{D}(\hat{\mu}_c, P_0)$, $\tilde{\mathscr{D}}_c(P_0) := \{P \in \mathscr{P}_1 : \mu_c(P \times P_0) := \int_{U \times U} c(x, y)P(\mathrm{d}x)P_0(\mathrm{d}y) < \infty\}$ and $\mathscr{D}(\check{\mu}_c, P_0) := \{P \in \mathscr{P}_1 : \check{\mu}_c(P, P_0) \leqslant \infty\}$, where $\hat{\mu}$ is the maximal distance relative to μ_c (see (8.1.12)) and *P_0 is an arbitrary law in $\mathscr{P}_1$*. Obviously, $\mathscr{D}(\check{\mu}_c, P_0) \subset \tilde{\mathscr{D}}_c(P_0) \subset \mathscr{D}(\hat{\mu}_c, P_0)$ for any $P_0 \in \mathscr{P}_1$ and $\mathscr{D}(\check{\mu}_c, \delta_\alpha) \equiv \tilde{\mathscr{D}}_c(\delta_\alpha) \equiv \mathscr{D}(\hat{\mu}_c; \delta_\alpha)$, $\alpha \in U$.

Let $H_N(t) = H(t)I\{t > N\}$ for $H \in \mathscr{H}$, $t \geqslant 0$, $N > 0$ and define $c_N(x, y) := H_N(d(x, y))$, μ_{c_N}, $\hat{\mu}_{c_N}$, $\check{\mu}_{c_N}$ by (8.1.8), (8.1.7) and (8.1.12), respectively. Therefore

$$\mathscr{D}(\hat{\mu}_c, P_0) = \left\{P \in \mathscr{P}_1 : \lim_{N \to \infty} \hat{\mu}_{c_N}(P, P_0) = 0\right\} \tag{8.2.1}$$

$$\tilde{\mathscr{D}}_c(P_0) = \left\{P \in \mathscr{P}_1 : \lim_{N \to \infty} \hat{\mu}_{c_N}(P \times P_0) = 0\right\} \tag{8.2.2}$$

$$\mathscr{D}(\check{\mu}_c, P_0) \supset \tilde{\mathscr{D}}(\check{\mu}_c, P_0) := \left\{P \in \mathscr{P}_1 : \lim_{N \to \infty} \check{\mu}_{c_N}(P, P_0) = 0\right\}. \tag{8.2.3}$$

As usual, we denote the weak convergence of laws $\{P_n\}_{n=1}^{\infty}$ to the law P by $P_n \xrightarrow{w} P$.

Theorem 8.2.1. Let (U, d) be a u.m. s.m.s. (see Section 2.4), $H \in \mathscr{H}$ ($H(t) > 0$ for $t > 0$) and P_0 be a law in $\mathscr{P}_1$.

(i) If $\{P_1, P_2, \ldots\} \subset \mathscr{D}(\hat{\mu}_c, P_0)$ and $Q \in \tilde{\mathscr{D}}(\check{\mu}_c, P_0)$ then

$$\lim_{N \to \infty} \hat{\mu}_c(P_n, Q) = 0 \tag{8.2.4}$$

if and only if the following two conditions are satisfied:

(1*) $P_n \xrightarrow{w} Q$

(2*) $\lim_{N \to \infty} \sup_n \hat{\mu}_{c_N}(P_n, P_0) = 0.$

(ii) If $\{Q, P_1, P_2, \ldots,\} \subset \tilde{\mathscr{D}}_c(P_0)$ then (8.2.4) holds if and only if the conditions (1*) and

(3*) $\lim_{N \to \infty} \sup_n \mu_c(P_n \times P_0) = 0$

are fulfilled.

(iii) If $\{P_1, P_2, \ldots,\} \subset \tilde{\mathscr{D}}(\check{\mu}_c, P_0)$ and $Q \in \mathscr{D}(\hat{\mu}_c, P_0)$, then (8.2.4) holds if and only if the conditions (1*) and

(4*) $\lim_{N \to \infty} \sup_n \check{\mu}_c(P_n, P_0) = 0$

are fulfilled.

Theorem 8.2.1 is an immediate corollary of the following lemma. Further we use the same notation as in (8.1.1), (8.1.2) and (8.1.3).

Lemma 8.2.1. Let U be a u.m. s.m.s., π is the Prokhorov metric in $\mathscr{P}$, and $H \in \mathscr{H}$. For any $P_0, P_1, P_2 \in \mathscr{P}_1$ and $N > 0$, the following inequalities are satisfied

$$\check{\mu}_c(P_1, P_2) \leqslant H(\pi(P_1, P_2)) + K_H\{2\pi(P_1, P_2)H(N) + \hat{\mu}_{c_N}(P_1, P_0) + \check{\mu}_{c_N}(P_2, P_0)\} \tag{8.2.5}$$

$$\hat{\mu}_c(P_1, P_2) \leqslant H(\pi(P_1, P_2)) + K_H\{2\pi(P_1, P_2)H(N) + \mu_{c_N}(P_1 \times P_0) + \mu_{c_N}(P_2 \times P_0)\} \tag{8.2.6}$$

$$\pi(P_1, P_2)H(\pi(P_1, P_2)) \leqslant \hat{\mu}_c(P_1, P_2) \tag{8.2.7}$$

$$\hat{\mu}_{c_N}(P_1, P_0) \leqslant K(\hat{\mu}_c(P_1, P_2) + \hat{\mu}_{c_{N/2}}(P_2, P_0)) \tag{8.2.8}$$

$$\mu_{c_N}(P_1 \times P_0) \leqslant K(\hat{\mu}_c(P_1, P_2) + \mu_{c_{N/2}}(P_2 \times P_0)) \tag{8.2.9}$$

$$\check{\mu}_{c_N}(P_1, P_0) \leqslant K(\hat{\mu}_c(P_1, P_2) + \check{\mu}_{c_{N/2}}(P_2, P_0)) \tag{8.2.10}$$

where K_H is given by (2.2.3) and $K = K_H + K_H^2$.

Remark 8.2.1. Relationships (8.2.5) to (8.2.10) give us necessary and sufficient conditions for $\hat{\mu}_c$-convergence as well as quantitative representations of these conditions. Clearly, such treatment of the $\hat{\mu}_c$-convergence is preferable because it gives not only a qualitative answer when $\hat{\mu}_c(P_n, Q) \to 0$ but also establishes a quantitative estimate of the convergence $\hat{\mu}_c(P_n, Q) \to 0$.

Proof of Lemma 8.2.1. In order to get (8.2.5) we require the following relation between the H-average compound $\mu_c = \mathscr{L}_H$ and the Ky Fan metric $\mathbf{K}$ (see Examples 3.3.1 and 3.3.2):

$$\mathscr{L}_H(P) \leqslant H(\mathbf{K}(P)) + K_H\{2\mathbf{K}(P)H(N) + \mathscr{L}_{H_N}(P') + \mathscr{L}_{H_N}(P'')\} \quad \mathscr{L}_{H_N} := \mu_{c_N} \tag{8.2.11}$$

for $N > 0$ and any triplet of laws $(P, P', P'') \in \mathscr{P}_2$ such that there exists a law $Q \in \mathscr{P}_3$ with marginals

$$T_{12}Q = P \quad T_{13}Q = P' \quad T_{23}Q = P''. \tag{8.2.12}$$

If $\mathbf{K}(P) > \delta$ then, by (2.2.3)

$$\int H(d(x, y))P(dx, dy)$$

$$\leqslant K_H \int [H(d(x, x_0)) + H(d(y, x_0))]I\{d(x, y) > \delta\}Q(\mathrm{d}x, \mathrm{d}y, \mathrm{d}x_0) + H(\delta)$$

$$\leqslant H(\delta) + K_H\{2H(N)\delta + \mathscr{L}_{H_N}(P') + \mathscr{L}_{H_N}(P'')\}.$$

Letting $\delta \to \mathbf{K}(P)$ completes the proof of (8.2.11).

For any $\varepsilon > 0$ we choose $P \in \mathscr{P}_2$ with marginals P_1, P_2 and $P' \in \mathscr{P}_2$ with marginals P_1 and P_0 such that

$$\hat{\mathbf{K}}(P_1, P_2) > \mathbf{K}(P) - \varepsilon \quad \hat{\mathscr{L}}_{H_N}(P_1, P_0) > \mathscr{L}_{H_N}(P') - \varepsilon. \tag{8.2.13}$$

Choosing Q with property (8.2.12) (cf. (3.2.5)) we obtain

$$\begin{aligned}\hat{\mathscr{L}}_H(P_1, P_2) &\leqslant \mathscr{L}(P)\\ &\leqslant H(\hat{\mathbf{K}}(P_1, P_2) + \varepsilon) + \mathbb{K}_H\{2(\hat{\mathbf{K}}(P_1, P_2) + \varepsilon)H(N) + \hat{\mathscr{L}}_{H_N}(P_1, P_0)\\ &\qquad + \varepsilon + \mathscr{L}_{H_N}(T_{23}Q)\}\end{aligned}$$

by (8.2.11) and (8.2.13). The last inequality together with the Strassen theorem (see Corollary 7.4.2) proves (8.2.5).

If $P_1 \times P_0$ and $P_2 \times P_0$ stand for P' and P'', respectively, then (8.2.11) implies (8.2.6). To prove that (8.2.7) to (8.2.10) hold we use the following two inequalities: for any $P \in \mathscr{P}_2$ with marginals P_1 and P_2

$$\mathbf{K}(P)H(\mathbf{K}(P)) \leqslant \mathscr{L}_H(P) \tag{8.2.14}$$

and

$$\mathscr{L}_{H_N}(P') \leqslant K[\mathscr{L}_H(P) + \mathscr{L}_{H_{N/2}}(P'')] \tag{8.2.15}$$

where (P, P', P'') are subject to the conditions (8.2.12) and $N > 0$. Using the same arguments as in the proof of (8.2.5), we get (8.2.7) to (8.2.10) by means of (8.2.14) and (8.2.15). QED

Given an u.m. s.m.s. (U, d) and a s.m.s. (V, g) let $\phi\colon U \to V$ be a measurable function. For any p. distance μ on $\mathscr{P}(V^2)$ define the p. distance μ_ϕ on $\mathscr{P}(U^2)$ by (7.1.14). Theorem 7.1.4 states that (cf. Remark 7.1.2)

$$\hat{\mu}_\phi(P_1, P_2) = \hat{\mu}(P_{1,\phi}, P_{2,\phi}) \tag{8.2.16}$$

or in terms of U-valued random variables

$$\hat{\mu}_\phi(X_1, X_2) = \hat{\mu}(\phi(X_1), \phi(X_2)),\ X_1, X_2 \in \mathfrak{X}(U). \tag{8.2.17}$$

Next we shall generalize Theorem 8.2.1, considering criteria for $\hat{\mu}_{c,\phi}$-convergence. We start with the special but important case of $\mu_c = \mathscr{L}_p^p (p \geqslant 1)$. Define the $\mathscr{L}_p$-metric in $\mathscr{P}(V^2)$

$$\mathscr{L}_p(Q) := \left(\int_{V\times V} g^p(x, y) Q(\mathrm{d}x, \mathrm{d}y)\right)^{1/p} \quad p \geqslant 1 \quad Q \in \mathscr{P}(V^2).$$

Then, by (7.1.14), $\mathscr{L}_{p,\phi}$ is a probability metric in $\mathscr{P}(U^2)$ and $\hat{\mathscr{L}}_{p,\phi}$ is the corresponding minimal metric. In the next corollary we apply Theorems 7.1.4 and 8.1.1 in order to get a criterion for $\hat{\mathscr{L}}_{p,\phi}$-convergence.

Let $Q, P_1, P_2, \ldots,$ be probability measures on $\mathscr{P}(U)$. Denote $\pi_{n,\phi} = \pi(P_{n,\phi}, Q_\phi)$, π being the Prokhorov metric in $\mathscr{P}(V)$

$$D_{n,\phi} := \mathbf{D}(P_{n,\phi}, Q_\phi) := \left|\left(\int_V g^p(x, a) P_{n,\phi}(\mathrm{d}x)\right)^{1/p} - \left(\int_V g^p(x, a) Q_\phi(\mathrm{d}x)\right)^{1/p}\right|$$

(a is a fixed point in V)

$$\mathscr{A}(Q_\phi) := \left(p \int_V (g(x, a) + 1)^{p-1} Q_\phi(\mathrm{d}x)\right)^{1/p}$$

$$M(Q_\phi, N) := \left(\int_V g^p(x, a) I\{g(x, a) > N\} Q_\phi(\mathrm{d}x)\right)^{1/p}$$

$$M(Q_\phi) := \left(\int_V g^p(x, a) Q_\phi(\mathrm{d}x)\right)^{1/p}.$$

Corollary 8.2.1. For all $n = 1, 2, \ldots$ let

$$M(P_{n,\phi}) + M(Q_\phi) < \infty. \tag{8.2.18}$$

Then $\hat{\mathscr{L}}_{p,\phi}(P_n, Q) \to 0$ as $n \to \infty$ if and only if $P_{n,\phi}$ weakly tends to Q_ϕ and

$D_{n,\phi} \to 0$ as $n \to \infty$. Moreover, the following quantitative estimates are valid

$$\hat{\mathscr{L}}_{p,\phi}(P_n, Q) \geqslant \max(D_{n,\phi}, (\pi_{n,\phi})^{1+1/p}) \tag{8.2.19}$$

$$\hat{\mathscr{L}}_{p,\phi}(P_n, Q) \leqslant (1+2N)\pi_{n,\phi} + 5M(Q_\phi, N) + (\pi_{n,\phi})^{1/p}(3\mathscr{A}(Q_\phi) + 2^{2+1/p}N) + D_{n,\phi} \tag{8.2.20}$$

for each positive N.

Proof. The first part of Corollary 8.2.1 follows immediately from (8.2.19), (8.2.20) (for the 'if' part put, for instance, $N = (\pi_{n,\phi})^{-1/2p}$). Relations (8.2.19) and (8.2.20) establish additionally a quantitative estimate of the convergence of $\hat{\mathscr{L}}_{p,\phi}(P_n, Q)$ to zero. To prove the latter relations we use (8.2.16) and the following inequalities

$$\hat{\mathscr{L}}_p(Q_1, Q_2) \geqslant \max(\pi(Q_1, Q_2)^{1+1/p}, \mathbf{D}(Q_1, Q_2)) \tag{8.2.21}$$

$$\hat{\mathscr{L}}_p(Q_1, Q_2) \leqslant (1+2N)\pi(Q_1, Q_2) + M(Q_1, N) + M(Q_2, N) \tag{8.2.22}$$

and

$$M(Q_1, 2N) \leqslant \mathbf{D}(Q_1, Q_2) + 4M(Q_2, N) + \pi(Q_1, Q_2)^{1/p}(3\mathscr{A}(Q_2) + 2^{2+1/p}N) \tag{8.2.23}$$

for each positive N and $Q_1, Q_2 \in \mathscr{P}(V)$, where $\mathbf{D}$ is the primary metric given by

$$\mathbf{D}(Q_1, Q_2) = \left| \left(\int_V g^p(x, a) Q_1(\mathrm{d}x) \right)^{1/p} - \left(\int_V g^p(x, a) Q_2(\mathrm{d}x) \right)^{1/p} \right|. \tag{8.2.24}$$

Claim 1 (8.2.21) holds.
For any V-valued random variables X_1 and X_2 with distributions Q_1 and Q_2 respectively,

$$\mathscr{L}_p(X_1, X_2) = [\mathbb{E}g^p(X_1, X_2)]^{1/p} > \mathbf{D}(Q_1, Q_2)$$

by the Minkovski inequality. Thus $\hat{\mathscr{L}}_p(Q_1, Q_2) \geqslant \mathbf{D}(Q_1, Q_2)$. Using (8.2.7) with $H(t) = t^p$ we have also that $\hat{\mathscr{L}}_p \geqslant \pi^{1+1/p}$.

Claim 2 (8.2.22) holds.
We start with the Chebyshev's inequality: for any X_i with laws Q_i

$$\mathscr{L}_p(X_1, X_2) \leqslant (1+2N)\mathbf{K}(X_1, X_2) + M(Q_1, N) + M(Q_2, N)$$

where $\mathbf{K}$ is the Ky Fan metric in $\mathfrak{X}(V)$ and $M(Q_i) = (\int_V g^p(x, a) Q_i(\mathrm{d}x))^{1/p}$. The proof is analogous to that of (8.2.11). By virtue of the Strassen theorem $\hat{\mathbf{K}} = \pi$ the above inequality yields (8.2.22).

Claim 3 (8.2.23) holds.

Observe that

$$M(Q_1, 2N) := \left(\int_V g^p(x, a) I\{g(x, a) > 2N\} Q(\mathrm{d}x)\right)^{1/p}$$
$$\leqslant \mathbf{D}(Q_1, Q_2) + \left|\int_V g^p(x, a) I\{g(x, a) \leqslant 2N\}(Q_1 - Q_2)(\mathrm{d}x)\right|^{1/p}$$
$$+ M(Q_2, 2N).$$

Denote $f(x) := \min\{g^p(x, a), (2N)^p\}$, $h(x) := \min\{2^p g^p(x, O(a, N)), (2N)^p\}$ where $O(a, N) := \{x \in V: g(x, a) \leqslant N\}$. Then

$$I := \left|\int_V g^p(x, a) I\{g(x, a) \leqslant 2N\}(Q_1 - Q_2)(\mathrm{d}x)\right|^{1/p}$$
$$\leqslant \left|\int_V f(x)(Q_1 - Q_2)(\mathrm{d}x)\right|^{1/p} + 2N\left|\int_V I\{g(x, a) > 2N\}(Q_1 - Q_2)(\mathrm{d}x)\right|^{1/p}$$
$$=: I_1 + I_2.$$

Using the inequality

$$\begin{aligned}|f(x) - f(y)| &\leqslant |g^p(x, a) - g^p(y, a)| \\ &\leqslant p \max(g^{p-1}(x, a), g^{p-1}(y, a))|g(x, a) - g(y, a)| \\ &\leqslant p \max(g^{p-1}(x, a), g^{p-1}(y, a)) g(x, y) \quad x, y \in V\end{aligned}$$

we get for any pair (X_1, X_2) of V-valued random variables with marginal distributions Q_1 and Q_2

$$\begin{aligned}I_1^p &:= |\mathbb{E}(f(X_1) - f(X_2))| \\ &\leqslant \mathbb{E}|f(X_1) - f(X_2)| I\{g(X_1, X_2) \leqslant \gamma\} \\ &\quad + \mathbb{E}[|f(X_1)| + |f(X_2)|] I\{g(X_1, X_2) \geqslant \gamma\} \\ &\leqslant \gamma p \mathbb{E}(g(X_2, a) + \gamma)^{p-1} + 2(2N)^p \Pr(g(X_1, X_2) \geqslant \gamma) \quad \text{for any } \gamma \in [0, 1].\end{aligned}$$

Let $K = \mathbf{K}(X_1, X_2)$ be the Ky Fan metric in $\mathfrak{X}(V)$. Then from the above bound

$$I_1 \leqslant K^{1/p}[\mathscr{A}(Q_2)^p + 2(2N)^p]^{1/p} \leqslant K^{1/p}[\mathscr{A}(Q_2) + 2^{1+1/p} N].$$

Now let us estimate the second term in the upper bound for I

$$I_2 := \left|\int_V (2N)^p I\{g(x, a) > 2N\}(Q_1 - Q_2)(\mathrm{d}x)\right|^{1/p}$$
$$\leqslant \left(\int_V (2N)^p I\{g(x, a) > 2N\} Q_1(\mathrm{d}x)\right)^{1/p} + M(Q_2, 2N).$$

If $g(x, c) > 2N$, then $g(x, O(c, N)) \geqslant N$ and, therefore

$$\left[\int_V (2N)^p I\{g(x, a) > 2N\} Q_1(\mathrm{d}x)\right]^{1/p} \leqslant [\mathbb{E}h(X_1)]^{1/p}$$
$$\leqslant |\mathbb{E}h(X_1) - \mathbb{E}h(X_2)|^{1/p} + [\mathbb{E}h(X_2)]^{1/p} =: I_1' + I_2'.$$

The inequality

$$|h(x) - h(y)| \leqslant 2^p |g^p(x, O(a, N)) - g^p(y, O(a, N))|$$
$$\leqslant 2^p p \max[g^{p-1}(x, O(a, N)), g^{p-1}(y, O(a, N))] g(x, y)$$

implies

$$I_1' \leqslant [\mathbb{E}|h(X_1) - h(X_2)| I\{g(X_1, X_2) \leqslant \gamma\}]^{1/p}$$
$$+ [\mathbb{E}(h(X_1) + h(X_2)) I\{g(X_1, X_2) > \gamma\}]^{1/p}$$
$$\leqslant 2\{\gamma \mathbb{E} p[g(X_2, O(a, N)) + 1]^{p-1}\}^{1/p} + \{2(2N)^p \Pr(g(X_1, X_2) > \gamma)^{1/p}, \text{for } K < \gamma.$$

On the other hand, by the definition of h,

$$I_2' := [\mathbb{E}h(X_2)]^{1/p} \leqslant [\mathbb{E}(2N)^p I\{g(X_2, a) > N\}]^{1/p} \leqslant 2M(Q_2, N).$$

Combining the above estimates we get

$$I_2 \leqslant 3M(Q_2, N) + 2K^{1/p}\mathscr{A}(Q_2) + 2^{1+1/p} N K^{1/p}.$$

Making use of the estimates for I_1 and I_2 and the Strassen theorem, we get

$$I \leqslant I_1 + I_2 \leqslant 3M(Q_2, N) + \pi(Q_1, Q_2)^{1/p}(3\mathscr{A}(Q_2) + 2^{2+1/p}N).$$

This completes the proof of (8.2.23). QED

We can extend Corollary 8.2.1 considering the H-average compound distance

$$\mu_c(Q) := \mathscr{L}_H(Q) := \int_{V^2} c(x, y) Q(\mathrm{d}x, \mathrm{d}y) \quad Q \in \mathscr{P}(V^2) \tag{8.2.25}$$

where $c(x, y) = H(g(x, y))$ and $H(t)$ is a non-decreasing continuous function on $[0, \infty)$ vanishing at zero (and only there) and satisfying the Orlicz condition

$$K_H := \sup\{H(2t)/H(t);\ t > 0\} < \infty \tag{8.2.26}$$

see (3.3.1) and Example 2.2.1.

Corollary 8.2.2. Assume that $\int_V c(x, a)(P_{n,\phi} + Q_\phi)(\mathrm{d}x) < \infty$. Then the convergence $\hat{\mu}_{c,\phi}(P_n, Q) \to 0$ as $n \to \infty$ is equivalent to the following relations: $P_{n,\phi}$ tends weakly to Q_ϕ as $n \to \infty$, and for some $a \in U$,

$$\lim_{N \to \infty} \overline{\lim_{n}} \int_V c(x, a) I\{g(x, a) > N\} P_{n,\phi}(\mathrm{d}x) = 0.$$

Proof: See Theorems 8.2.1 and 7.1.4. QED

Note that the Orlicz condition (8.2.26) implies a power growth of the function H. In order to extend the $\hat{\mu}_{c,\phi}$ convergence criterion in the last corollary we consider functions H in (8.2.25) with exponential growth. Let RB be the class of all bounded from above real-valued r.v.s. Then

$$\xi \in RB \Leftrightarrow \tag{8.2.27}$$

$$\tau(\xi) := \inf\{a > 0: \mathbb{E}\exp \lambda\xi \leqslant \exp \lambda a \ \forall \lambda > 0\} = \sup_{\lambda > 0} \frac{1}{\lambda} \ln \mathbb{E}\exp(\lambda\xi) < \infty.$$

In fact, clearly, if $\xi \in RB$ then $\tau(\xi) < \infty$. On the other hand, if $F_\xi(x) < 1$ for $x \in \mathbb{R}$ then for any $a > 0$, $\mathbb{E}\exp[\lambda(\xi - a)] \geqslant \exp(\lambda a)\Pr(\xi > 2a) \to \infty$ as $\lambda \to \infty$. By the Hölder inequality one gets

$$\tau(\xi + \eta) \leqslant \tau(\xi) + \tau(\eta) \tag{8.2.28}$$

and hence if $Q \in \mathscr{P}(V^2)$ and (Y_1, Y_2) is a pair of V-valued r.v.s with joint distribution Q, then

$$\boldsymbol{\tau}(Q) := \tau(g(Y_1, Y_2)) \tag{8.2.29}$$

determines a compound metric on $\mathscr{P}(V^2)$ (see Section 2.3). The next theorem gives us a criterion for $\hat{\boldsymbol{\tau}}_\phi$-convergence, where $\hat{\boldsymbol{\tau}}_\phi$ is defined by (8.2.16) and (8.2.17).

Theorem 8.2.2. Let X_n, $n = 1, 2, \ldots$, and Y be U-valued r.v.s with distributions P_n and Q respectively and let $\boldsymbol{\tau}(g(\phi(X_n), a)) + \boldsymbol{\tau}(g(\phi(Y), a)) < \infty$. Then the convergence $\hat{\boldsymbol{\tau}}_\phi(P_n, Q) \to 0$ as $n \to \infty$ is equivalent to the following relations:

(a) $P_{n,\phi}$ tends weakly to Q_ϕ, and
(b) $\lim_{N\to\infty} \overline{\lim}_n \, \boldsymbol{\tau}(g(\phi(X_n), a)I\{g(\phi(X_n), a) > N) = 0$.

Proof. As in Corollary 8.2.1 the assertion of the theorem is a consequence of (8.2.16) and the following three claims. Let V-valued random variables Y_1 and Y_2 have distributions Q_1 and Q_2, respectively.

Claim 1

$$\pi^2(Q_1, Q_2) \leqslant \hat{\boldsymbol{\tau}}(Q_1, Q_2). \tag{8.2.30}$$

By the Strassen theorem, $\hat{\mathbf{K}} = \pi$, see Corollary 7.4.2, it is enough to prove that $\tau(g(Y_1, Y_2)) \geqslant \mathbf{K}^2(Y_1, Y_2)$. Let $\xi = g(Y_1, Y_2)$ and $\tau(\xi) < \varepsilon^2 \leqslant 1$; then

$$\Pr(\xi > \varepsilon) \leqslant \frac{\mathbb{E}\,e^\xi - 1}{e^\varepsilon - 1} \leqslant \frac{e^{\tau(\xi)} - 1}{e^\varepsilon - 1} \leqslant \frac{e^{\varepsilon^2} - 1}{e^\varepsilon - 1} \leqslant \varepsilon.$$

Letting $\varepsilon^2 \to \tau(\xi)$ we obtain (8.2.30).

Claim 2

$$\tau(g(Y_1, a)I\{g(Y_1, a) > N\}) \leqslant 2\hat{\tau}(Q_1, Q_2) + 2\tau(g(Y_2, c)I\{g(Y_2, a) > N/2\}). \quad (8.2.31)$$

Note that the inequality $\xi \leqslant \eta$ with probability one implies $\tau(\xi) \leqslant \tau(\eta)$. Hence

$$\begin{aligned}
&\tau(g(Y_1, a)I\{g(Y_1, a) > N\}) \\
&\quad \leqslant \tau[(g(Y_1, Y_2) + g(Y_2, a))I\{g(Y_2, a) + g(Y_1, Y_2) > N\}] \\
&\quad \leqslant \tau[(g(Y_1, Y_2) + g(Y_2, a))\max(I\{g(Y_2, a) > N/2\}, I\{g(Y_1, Y_2) > N/2\})] \\
&\quad \leqslant \tau(g(Y_1, Y_2)I\{g(Y_1, Y_2) > N/2\}) + \tau(g(Y_1, Y_2)I\{g(Y_1, Y_2) \leqslant N/2\} \\
&\qquad \times I\{g(Y_2, a) > N/2\}) + \tau(g(Y_2, a)I\{g(Y_2, a) > N/2\}) + \tau(g(Y_2, a) \\
&\qquad \times I\{g(Y_2, a) \leqslant N/2\}I\{g(Y_1, Y_2) > N/2\}) \\
&\quad \leqslant 2\tau(g(Y_1, Y_2)I\{g(Y_1, Y_2) > N/2\}) + 2\tau(g(Y_2, a)I\{g(Y_2, a) \geqslant N/2\}) \\
&\quad \leqslant 2\tau(g(Y_1, Y_2)) + 2\tau(g(Y_2, a)I\{g(Y_2, a) > N/2\}).
\end{aligned}$$

Passing to the minimal metric $\hat{\tau}$ we get (8.2.31).

Claim 3

$$\begin{aligned}
\hat{\tau}(Q_1, Q_2) \leqslant \pi(Q_1, Q_2)(1 + 2N) + \tau(g(Y_1, a)I\{g(Y_1, a) > N\}) \quad &(8.2.32) \\
+ \tau(g(Y_2, a)I\{g(Y_2, a) > N\}) \quad \forall N > 0, a \in V.&
\end{aligned}$$

For each δ the following holds: $\tau(g(Y_1, Y_2)) \leqslant \tau(g(Y_1, Y_2)I\{g(Y_1, Y_2) \leqslant \delta\}) + \tau(g(Y_1, Y_2)I\{g(Y_1, Y_2) > \delta\}) =: I_1 + I_2$. For I_1 we obtain the estimate

$$I_1 = \sup_{\lambda > 0} 1/\lambda \ln \mathbb{E} \exp(\lambda g(Y_1, Y_2)I\{g(Y_1, Y_2) \leqslant \delta\}) \leqslant \sup_{\lambda > 0} 1/\lambda \ln \mathbb{E} \exp \lambda\delta = \delta.$$

For I_2 we have:

$$\begin{aligned}
I_2 &\leqslant \tau(g(Y_1, a) + g(Y_2, a))I\{g(Y_1, Y_2) > \delta\}) \\
&\leqslant \tau(g(Y_1, a)I\{g(Y_1, Y_2) \geqslant \delta\}) \\
&\quad + \tau(g(Y_2, a)I\{g(Y_1, Y_2) \geqslant \delta\}) =: A_1 + A_2.
\end{aligned}$$

Furthermore

$$\begin{aligned}
A_1 &\leqslant \tau(gY_1, a)I\{g(Y_1, Y_2) > \delta\}I\{g(Y_1, a) \leqslant N\}) \\
&\quad + \tau(g(Y_1, a)I\{g(Y_1, Y_2) > \delta\}I\{g(Y_1, a) > N\}) \\
&\leqslant \tau(NI\{g(Y_1, Y_2) > \delta\}) + \tau(g(Y_1, a)I\{g(Y_1, a) \geqslant N\}).
\end{aligned}$$

Hence, if $\mathbf{K}(Y_1, Y_2) < \delta$, then

$$\tau(g(Y_1, Y_2)) \leqslant (1 + 2N)\delta + \tau(g(Y_1, c)I\{g(Y_1, c) > N\}) + \tau(g(Y_2, a)I\{g(Y_2, a) > N\}).$$

Letting $\delta \to \mathbf{K}(Y_1, Y_2)$ and passing to the minimal metrics we obtain (8.2.32).

QED

In the rest of this section we look at the topological structure of the minimal norms $\mathring{\mu}_h(P_1, P_2)$, $P_1, P_2 \in \mathscr{P}_1$ (see (8.1.10)) where the function $h(x, y) = d(x, y) h_0(d(x, a) \vee d(y, a))$, $x, y \in U$ is defined as in (8.1.4).

Theorem 8.2.3. Let (U, d) be a s.m.s.

(a) If $g := d/(1 + d)$ and $a_n := \sup_{t>0} h_0(2t)/h_0(t) < \infty$, then

$$\mathring{\mu}_h(P_1, P_2) \leqslant (1 + N)\mathring{\mu}_g(P_1, P_2) + (2a_h + 4) \int h(x, a)I\{d(x, a) > N\}(P_1 + P_2)(\mathrm{d}x) \text{ for } N \geqslant 1.$$

(b) If $b_h := \sup_{0<s<t} [(1 + t - s)h_0(t)]^{-1} < \infty$ then $\mathring{\mu}_g(P_1, P_2) \leqslant \mathring{\mu}_h(P_1, P_2)$.

(c) If

$$c_h := \sup_{0<s<t} [th_0(t) - sh_0(s)]/[(t - s)h_0(t)] < \infty$$

then

$$\left| \int h(x, a)(P_1 - P_2)(\mathrm{d}x) \right| \leqslant c_h \mathring{\mu}_h(P_1, P_2).$$

(d) If $a_h + b_h + c_h < \infty$ and $\int h(x, a)(P_n + P)(\mathrm{d}x) < \infty$, $n = 1, 2, \ldots$ then

$$\lim_{n\to\infty} \mathring{\mu}_h(P_n, P) = 0$$

if and only if $P_n \xrightarrow{w} P$ and

$$\lim_{n\to\infty} \left| \int h(x, a)(P_n - P)(\mathrm{d}x) \right| = 0.$$

The proof of the theorem is similar to that of Theorem 6.3.1 and can therefore be omitted. Note that, in contrast to Theorems 6.2.2 and 6.2.3, the above bounds are based only on the relationships between minimal norms.

Open problem 8.2.1. A question of great interest concerning the topological structure of minimal distances is the *necessary and sufficient conditions for the convergence $\hat{\mu}_c(P_n, P) \to 0$ where $\{P, P_n, n = 1, 2, \ldots\} \subset \mathscr{D}(\hat{\mu}_c, P_0)$ and P_0 is an arbitrary law of $\mathscr{P}_1$.* Note that in the case $(U, d) = (\mathbb{R}^1, |\cdot|)$, if $\{P, P_n, n = 1, 2, \ldots\} \subset \mathscr{D}(\hat{\mu}_d, P_0)$ then $\hat{\mu}_d(P_n, P) = \int_{-\infty}^{\infty} |F_n(x) - F(x)| \,\mathrm{d}x = \mathscr{L}_1(P_n, P) \to 0$ if and only if $P_n \xrightarrow{w} P$ and

$$\lim_{N\to\infty} \sup_n \int_{|x|>N} |F_n(x) - F_0(x)| \,\mathrm{d}x = 0$$

where F_n is a d.f. of P_n, $n = 0, 1, \ldots$, and F is the d.f. of P.

CHAPTER 9

Moment Problems Related to the Theory of Probability Metrics. Relations between Compound and Primary Distances

In Chapters 5 to 8, we investigated the relationships between compound and simple distances. The main method we used was based on the dual and explicit solutions of the following.

(A) *Marginal problem.* For fixed probability measures (laws) P_1 and P_2 on a s.m.s. (U, d) and a continuous function c on the product space $U^2 = U \times U$

$$\text{minimize (maximize)} \int_{U^2} c(x, y)P(\mathrm{d}x, \mathrm{d}y)$$

where the laws P on U^2 have marginals P_1 and P_2, i.e. $T_iP = P_i$, $i = 1, 2$.

Similarly, we shall study the connection between compound and primary distances (see Section 3.1) solving the following:

(B) *Moment problem.* For fixed real numbers a_{ij} and real-valued continuous functions f_{ij} $(i = 1, 2, j = 1, \dots, n)$

$$\text{minimize (maximize)} \int_{U^2} c(x, y)P(\mathrm{d}x, \mathrm{d}y)$$

where the law P on U^2 satisfies the marginal moment conditions

$$\int_U f_{ij}\,\mathrm{d}P_i = a_{ij} \quad i = 1, 2, j = 1, \dots, n.$$

9.1 PRIMARY MINIMAL DISTANCES; MOMENT PROBLEMS WITH ONE FIXED PAIR OF MARGINAL MOMENTS

Let U be a separable norm space with norm $\|\cdot\|$, $\mathfrak{X} = \mathfrak{X}(U)$ the space of all U-valued r.v.s, μ be a compound metric in $\mathfrak{X}(U)$ and $\mathscr{M}$ be the class of all strictly increasing continuous functions f: $[0, \infty] \to [0, \infty]$, $f(0) = 0$, $f(\infty) = \infty$. Following the definition of primary distances (see Section 3.1) let us define the spaces $h(\mathfrak{X}) = \{\mathbb{E}h(\|X\|): X \in \mathfrak{X}\}$ (cf. (3.1.3)), for a fixed $h \in \mathcal{M}$ and a primary minimal distance $\tilde{\mu}_h$ (in $h(\mathfrak{X})$)

$$\tilde{\mu}_h(a, b) := \inf\{\mu(X, Y): X, Y \in \mathfrak{X}, \mathbb{E}h(\|X\|) = a, \mathbb{E}h(\|Y\|) = b\}. \quad (9.1.1)$$

Given the H-average compound distance

$$\mu(X, Y) = \mathscr{L}_H(X, Y) = \mathbb{E}H(\|X - Y\|) \quad H \in \mathcal{M} \cap \mathcal{H} \quad (9.1.2)$$

(see Example 3.3.1) we shall treat the explicit representations of the following extremal functional

$$I(H, h, a, b) := \tilde{\mu}_h(a, b). \quad (9.1.3)$$

Moreover, we shall consider the following upper bound for $\mu(X, Y) := \mathscr{L}_H(X, Y)$

$$S(H, h; a, b) := \sup\{\mu(X, Y): X, Y \in \mathfrak{X}, \mathbb{E}h(\|X\|) = a, \mathbb{E}h(\|Y\| = b\} \quad (9.1.4)$$

which explicit form will lead to the expression for the moment functions, discussed in Section 3.3 (see Definition 3.3.6). Denote for all $p \geqslant 0$, $q \geqslant 0$ the values

$$I(p, q; a, b) := I(H, h; a, b)(H(t) = t^p, h(t) = t^q) \quad (9.1.5)$$

$$S(p, q; a, b) := S(H, h; a, b)(H(t) = t^p, h(t) = t^q) \quad (9.1.6)$$

where here and in the sequel 0^0 means 0 and thus $\mathbb{E}\|X - Y\|^0$ means $\Pr(X \neq Y)$. Clearly, $I(p, q; a, b)^{\min(1, 1/p)}$ represents the primary h-minimal metric ($\tilde{\mathscr{L}}_{p,h}$) with respect to the $\mathscr{L}_p$-metric

$$\mathscr{L}_p(X, Y) := \{\mathbb{E}\|X - Y\|^p\}^{\min(1, 1/p)}, \ \mathscr{L}_0(X, Y) = \operatorname{ess\,sup}\|X - Y\|$$

where $hX := \mathbb{E}\|X\|^q$, $q \geqslant 0$, cf. Definition 3.1.2, i.e.,

$$I(p, q; a, b)^{\min(1, 1/p)} = \tilde{\mathscr{L}}_{p,h}(a, b) := \inf(\mathscr{L}_p(X, Y): hX = a, hY = b\}.$$

Further (Corollary 9.1.1), we shall find explicit expressions for $\tilde{\mathscr{L}}_{p,h}$ for any $p \geqslant 0$ and any $q \geqslant 0$. The scheme of the proofs of all statements here is as follows: first we prove the necessary inequalities that give us the required bounds and then we construct pairs of random variables which achieve the bounds or approximate them with arbitrary precision.

Let $f, f_1, f_2 \in \mathcal{M}$ and consider the following conditions (in the following f^{-1} is the inverse function of $f \in \mathcal{M}$)

A (f_1, f_2): $f_1 \circ f_2^{-1}(t)$ $(t \geqslant 0)$ is convex
B (f): $f^{-1}(\mathbb{E}f(\|X + Y)) \leqslant f^{-1}(\mathbb{E}f(\|X\|)) + f^{-1}(\mathbb{E}f(\|Y\|))$ for any $X, Y \in \mathfrak{X}$
C (f): $\mathbb{E}f(\|X + Y\|) \leqslant \mathbb{E}f(\|X\|) + \mathbb{E}f(Y))$ for any $X, Y \in \mathfrak{X}$
D (f_1, f_2): $\lim_{t\to\infty} f_1(t)/f_2(t) = 0$
E (f_1, f_2): $f_1 \circ f_2(t)$ $(t \geqslant 0)$ is concave
F (f_1, f_2): f_1 is concave and f_2 is convex
G (f_1, f_2): $\lim_{t\to\infty} f_1(t)/f_2(t) = \infty$.

Obviously, if $H(t) = t^p$, $h(t) = t^q (p > 0, q > 0)$, then $A(H, h) \Leftrightarrow p \geqslant q$, $B(h) \Leftrightarrow q \geqslant 1$, $C(h) \Leftrightarrow q \leqslant 1$, $D(H, h) \Leftrightarrow q > p$, $E(H, h) \Leftrightarrow q \geqslant p$, $F(H, h) \Leftrightarrow p \leqslant 1 \leqslant q$, $G(H, h) \Leftrightarrow p > q$ and hence the conditions A to G cover all possible values of the pairs (p, q).

Theorem 9.1.1. For any $a \geqslant 0$ and $b \geqslant 0$, $a + b > 0$ the following equalities hold

(i)
$$I(H, h; a, b) = \begin{cases} H(|h^{-1}(a) - h^{-1}(b)|) & \text{if } A(H,h) \text{ and } B(h) \text{ hold} \\ H \circ h^{-1}(|a - b|) & \text{if } A(H,b) \text{ and } C(h) \text{ hold} \\ 0 & \text{if } D(H,h) \text{ holds.} \end{cases} \tag{9.1.7}$$

(ii) For any $H \in \mathscr{M}$ and $h \in \mathscr{M}$

$$\inf\{\Pr\{X \neq Y\} \colon \mathbb{E}h(\|X\|) = a, \mathbb{E}h(\|Y\|) = b\} = 0 \tag{9.1.8}$$

$$\inf\{\mathbb{E}H(\|X - Y\|) \colon \Pr\{X \neq u\} = a, \Pr\{Y \neq u\} = 0 \, (u \in U, a, b \in [0, 1]). \tag{9.1.9}$$

(iii)

$$S(H, h; a, b) = \begin{cases} H(h^{-1}(a) + h^{-1}(b)) & \text{if } F(H, h) \text{ holds or if } B(h) \text{ and } E(H, h) \text{ hold,} \\ H \circ h^{-1}(\alpha + \beta) & \text{if } C(h) \text{ and } E(H, h) \text{ hold,} \\ \infty & \text{if } G(H, h) \text{ holds.} \end{cases} \tag{9.1.10}$$

(iv) For any $u \in U$, $H \in \mathscr{M}$, $h \in \mathscr{M}$,

$$\sup\{\Pr\{X \neq Y\} \colon \mathbb{E}h(\|X\|) = a, \mathbb{E}h(\|Y\|) = b\} = 1 \tag{9.1.11}$$

$$\sup\{\Pr\{X \neq Y\} \colon \Pr\{X \neq u\} = a, \Pr\{Y \neq u\} = b\} = \min(a + b, 1)(a, b \in [0, 1]) \tag{9.1.12}$$

$$\sup\{\mathbb{E}H(\|X - Y\|) \colon \Pr(X \neq u) = a, \Pr\{Y \neq u\} = b\} = \infty. \tag{9.1.13}$$

Proof. (i) *Case 1.* Let $A(H, h)$ and $B(h)$ be fulfilled. Denote $\phi(a, b) := H(|h^{-1}(a) - h^{-1}(b)|)$, $a \geqslant 0$, $b \geqslant 0$.

Claim 1. $I(H, h, a, b) \geqslant \phi(a, b)$.

By Jensen's inequality and $A(H, h)$

$$H \circ h^{-1}(\mathbb{E}Z) \leqslant \mathbb{E}H \circ h^{-1}(Z). \tag{9.1.14}$$

Taking $Z = h(\|X - Y\|)$ and using $B(h)$ we obtain $H^{-1}(\mathbb{E}H(\|X - Y\|)) = H^{-1}(\mathbb{E}H \circ h^{-1}(Z)) \geqslant h^{-1}(\mathbb{E}h(\|X - Y\|)) \geqslant |h^{-1}(\mathbb{E}h(\|X\|)) - h^{-1}(\mathbb{E}h(\|Y\|))|$ for any $X, Y \in \mathfrak{X}$, which proves the claim.

Claim 2. There exists an 'optimal' pair (X^*, Y^*) of r.v.s such that $\mathbb{E}h(\|X^*\|) = a$, $\mathbb{E}h(\|Y^*\|) = b$, $\mathbb{E}H(\|X^* - Y^*\|) = \phi(a, b)$.

Let $\bar{e}$ here and in the following be a fixed point of U with $\|\bar{e}\| = 1$. Then the required pair (X^*, Y^*) is given by

$$X^* = h^{-1}(a)\bar{e} \quad Y^* = h^{-1}(b)\bar{e} \tag{9.1.15}$$

which proves the claim.

Case 2. Let $A(H, h)$ and $C(h)$ be fulfilled. Denote $\phi_1(t) := H \circ h^{-1}(t)$, $t \geqslant 0$. As in Claim 1 we get $I(H, h; a, b) \geqslant \phi_1(|a-b|)$. Suppose that $a > b$ and for each $\varepsilon > 0$ define a pair $(X_\varepsilon, Y_\varepsilon)$ of r.v.s as follows: $\Pr\{X_\varepsilon = c_\varepsilon \bar{e},\ Y_\varepsilon = \bar{0}\} = p_\varepsilon$, $\Pr\{X_\varepsilon = d_\varepsilon \bar{e},\ Y_\varepsilon = d_\varepsilon \bar{e}\} = 1 - p_\varepsilon$, where

$$\bar{0} := 0\bar{e} \quad p_\varepsilon := \frac{a-b}{a-b+\varepsilon} \quad c_\varepsilon := h^{-1}(a-b+\varepsilon) \quad d_\varepsilon := h^{-1}\left(\frac{b}{1-p_\varepsilon}\right). \tag{9.1.16}$$

Then $(X_\varepsilon, Y_\varepsilon)$ enjoys the side conditions in (9.1.1) and $\mathbb{E}H(\|X_\varepsilon - Y_\varepsilon\|) = \phi_1(a-b+\varepsilon)(a-b)/(a-b+\varepsilon)$. Letting $\varepsilon \to 0$ we claim (9.1.7).

Case 3. Let $D(H, h)$ be fulfilled. In order to obtain (9.1.7) it is sufficient to define a sequence (X_n, Y_n) $(n \geqslant N)$ such that $\lim_{n\to\infty} \mathbb{E}H(\|X_n - Y_n\|) = 0$, $\mathbb{E}h(\|X_n\|) = a$, $\mathbb{E}h(\|Y_n\|) = b$. An example of such sequence is the following one: $\Pr\{X_n = \bar{0},\ Y_n = \bar{0}\} = 1 - c_n - d_n$, $\Pr\{X_n = na\bar{e},\ Y_n = \bar{0}\} = c_n$, $\Pr\{X_n = \bar{0},\ Y_n = nb\bar{e}\} = d_n$, where $c_n = a/h(na)$, $d_n = b/h(nb)$ and N satisfies $c_N + d_N < 1$.

(ii) Define the sequence (X_n, Y_n) $(n = 2, 3, \ldots)$ such that $\Pr\{X_n = h^{-1}(na)\bar{e},\ Y_n = h^{-1}(nb)\} = 1/n$, $\Pr(X_n = \bar{0},\ Y_n = \bar{0}\} = (n-1)/n$. Hence, $\mathbb{E}h(\|X_n\|) = a$, $\mathbb{E}h(\|Y_n\|) = b$ and $\Pr(X_n \neq Y_n) = 1/n$ which shows (9.1.8).

Further, suppose $a \geqslant b$. Without loss of generality we may assume that $u = \bar{0}$. Then consider the random pair $(\tilde{X}_n, \tilde{Y}_n)$ with the following joint distribution: $\Pr\{\tilde{X}_n = \bar{0},\ \tilde{Y}_n = \bar{0}\} = 1 - a$, $\Pr\{\tilde{X}_n = (1/n)\bar{e},\ \tilde{Y}_n = \bar{0}\} = a - b$, $\Pr\{\tilde{X}_n = (1/n)\bar{e},\ \tilde{Y}_n = (1/n)\bar{e}\} = b$. Obviously $(\tilde{X}_n, \tilde{Y}_n)$ satisfies the constraints $\Pr(\tilde{X}_n \neq \bar{0}) = a$, $\Pr(\tilde{Y}_n \neq \bar{0}) = b$ and $\lim_{n\to\infty} \mathbb{E}H(\|X_n - Y_n\|) = 0$, which proves (9.1.9).

The proofs of (iii) and (iv) are quite analogous to those of (i) and (ii), respectively. QED

Remark 9.1.1 If $A(H, h)$ and $B(h)$ hold we have constructed an *optimal* pair (X^*, Y^*) (see (9.1.15)), i.e. (X^*, Y^*) realizes the infimum in $I(H, h; a, b)$. However, if $D(H, h)$ holds and $a \neq b$ then optimal pairs *do not exist*, because $\mathbb{E}H(\|X - Y\|) = 0$ implies $a = b$. Note that the latter was not the case when we studied the minimal or maximal distances on a u.m.s.m.s. (U, d) since, by Theorem 8.1.1, there do exist (X^*, Y^*) and (X^{**}, Y^{**}) such that

$$\hat{\mu}_c(X, Y) := \inf\{\mu_c(\tilde{X}, \tilde{Y}) \colon \tilde{X}, \tilde{Y} \in \mathfrak{X}(U, d), \Pr_{\tilde{X}} = \Pr_X, \Pr_{\tilde{Y}} = \Pr_Y\} = \mu_c(X^*, Y^*)$$

and

$$\check{\mu}_c(X, Y) := \sup\{\mu_c(\tilde{X}, \tilde{Y}) \colon \tilde{X}, \tilde{Y} \in \mathfrak{X}(U, d), \Pr_{\tilde{X}} = \Pr_X, \Pr_{\tilde{Y}} = \Pr_Y\} = \mu_c(X^{**}, Y^{**}).$$

Corollary 9.1.1 For any $a \geqslant 0$, $b \geqslant 0$, $a + b > 0$, $p \geqslant 0$, $q \geqslant 0$,

$$I(p, q; a, b) = \begin{cases} |a^{1/q} - b^{1/q}|^p & \text{if } p \geqslant q \geqslant 1, \\ |a - b|^{p/q} & \text{if } p \geqslant q \quad 0 < q < 1, \\ 0 & \text{if } 0 \leqslant p < q \quad \text{or} \quad q = 0, p > 0, \\ |a - b| & \text{if } p = q = 0, \end{cases} \tag{9.1.17}$$

and in particular, the primary h-minimal metric, $\tilde{\mathscr{L}}_{p,h}$ ($hX = \mathbb{E}\|X\|^q$), admits the following representation

$$\tilde{\mathscr{L}}_{p,h}(a, b) = \begin{cases} |a^{1/q} - b^{1/q}| & \text{if } p \geqslant q \geqslant 1, \\ |a - b|^{1/q} & \text{if } p \geqslant 1, 0 < q < 1, \\ |a - b|^{p/q} & \text{if } 1 \geqslant p \geqslant q > 0, \\ 0 & \text{if } 0 \leqslant p < q \text{ or } q = 0, p > 0, \\ |a - b| & \text{if } p = q = 0. \end{cases} \tag{9.1.18}$$

One can easily check that if μ is a compound or simple probability distance with parameter $\mathbb{K}_H$ then

$$M(P_1, P_2) := \sup\left\{\mu(X_1, X_2): X_1, X_2 \in \mathfrak{X}, \mathbb{E}h(\|X_i\|) = \int_U h(\|x\|)P_i(\mathrm{d}x), i = 1, 2\right\} \tag{9.1.19}$$

is a moment function with the same parameter $\mathbb{K}_M = \mathbb{K}_\mu$ (see Definition 3.3.2). In particular, in (9.1.4) $S(H, h: a, b)$ ($H \in \mathscr{H} \cap \mathscr{M}$) and in (9.1.6) $M_p(P_1, P_2) = S(p, q; a, b)^{\min(1,1/p)}$, $a = \int \|x\|^q P_1(\mathrm{d}x)$, $b = \int \|x\|^q P_2(\mathrm{d}x)$ may be viewed as moment functions with parameters $\mathbb{K}_M = K_H$ (cf. (2.2.3)) and $\mathbb{K}_M = 1$, respectively.

Corollary 9.1.2. For any $a \geqslant 0$, $b \geqslant 0$, $a + b > 0$, $p \geqslant 0$, $q \geqslant 0$,

$$S(p, q; a, b) = \begin{cases} (a^{1/q} + b^{1/q})^q & \text{if } 0 \leqslant p \leqslant q, q \geqslant 1, \\ (a + b)^{p/q} & \text{if } 0 \leqslant p \leqslant q < 1, q \neq 0, \\ \infty & \text{if } p > q \geqslant 0, \\ \min(a + b, 1) & \text{if } p = q = 0. \end{cases} \tag{9.1.20}$$

Obviously, if $q = 0$ in (9.1.17), (9.1.18) or (9.1.20) the values of I and S make sense for $a, b \in [0, 1]$.

The following theorem is an extension for $p = q = 1$ of Corollaries 1 and 2 to a non-normed space U such as the Skorokhod space $D[0, 1]$ (see, for example, Billingsley 1968).

Theorem 9.1.2. Let (U, d) be a separable metric space, $\mathfrak{X} = \mathfrak{X}(U)$, the space of all U-valued r.v.s, and let $u \in U$, $a \geqslant 0$, $b \geqslant 0$. Assume that there exists $z \in U$ such that $d(z, u) \geqslant \max(a, b)$. Then

$$\min\{\mathbb{E}d(X, Y): X, Y \in \mathfrak{X}, \mathbb{E}d(X, u) = a, \mathbb{E}d(Y, u) = b\} = |a - b| \tag{9.1.21}$$

and

$$\max\{\mathbb{E}d(X, Y): X, Y \in \mathfrak{X}, \mathbb{E}d(X, u) = a, \mathbb{E}d(Y, u) = b\} = a + b. \tag{9.1.22}$$

Proof. Let $a \leqslant b$, $\gamma = d(z, u)$. By the triangle inequality the minimum in (9.1.21) is greater than $b - a$. On the other hand, if $\Pr(X = u, Y = u) = 1 - b/\gamma$,

$\Pr(X = u, Y = z) = (b - a)/\gamma$, $\Pr(X = z, Y = u) = 0$, $\Pr(X = z, Y = z) = a/\gamma$, then $\mathbb{E}d(X, u) = a$, $\mathbb{E}d(Y, u) = b$, $\mathbb{E}d(X, Y) = b - a$, which proves (9.1.21). Analogously, one proves (9.1.22). QED

From (9.1.21) it follows that the primary h-minimal metric, $\tilde{\mathscr{L}}_{1,h}(hX, hY)$, with respect to the average metric $\mathscr{L}_1(X, Y) = \mathbb{E}d(X, Y)$ with $hX = \mathbb{E}d(X, u)$ is equal to $|hX - hY|$. The above theorem provides the exact values of the bounds (3.3.48) and (3.3.52).

Open problem 9.1.1. Find the explicit solutions of moment problems with one fixed pair of marginal moments for r.v.s with values in a separable metric space U. In particular, find the primary h-minimal metric $\tilde{\mathscr{L}}_{p,h}(hX, hY)$, with respect to $\mathscr{L}_p(X, Y) = \{\mathbb{E}d^p(X, Y)\}^{1/p}$, $p > 1$ with $hX = \mathbb{E}d^q(X, u)$, $q \geqslant 0$.

Suppose that $U = \mathbb{R}^n$ and $d(x, y) = \|x - y\|_1$, where $\|(x_1, \ldots, x_n)\|_1 = |x_1| + \cdots + |x_n|$. Consider the H-average distance $\mathscr{L}_H(X, Y) := \mathbb{E}H(\|X - Y\|_1)$, with convex $H \in \mathscr{H}$, and the L_p-metric $\mathscr{L}_p(X, Y) = \{\mathbb{E}\|X - Y\|_1^p\}^{1/p}$. Define the engineer distance $\mathbf{EN}(X, Y; H) = H(\|\mathbb{E}X - \mathbb{E}Y\|_1)$, where $\mathbb{E}X = (\mathbb{E}X_1, \ldots, \mathbb{E}X_n)$ in the space of $\tilde{\mathfrak{X}}(\mathbb{R}^n)$ of all n-dimensional random vectors with integrable components (see Example 3.1.5). Similarly, define '$\mathscr{L}_p$-engineer metric', $\mathbf{EN}(X, Y, p) = (\sum_{i=1}^n |\mathbb{E}X_i - \mathbb{E}Y_i|^p)^{1/p}$, $p \geqslant 1$ (cf. (3.1.14)). Let $hX = \mathbb{E}X$ for any $X \in \tilde{\mathfrak{X}}(\mathbb{R}^n)$. Then the following relations between the compound distances $\mathscr{L}_H$, $\mathscr{L}_p$ and the primary distances $\mathbf{EN}(\cdot, \cdot; H)$, $\mathbf{EN}(\cdot, \cdot; p)$ hold.

Corollary 9.1.3. (i) If H is convex then

$$\tilde{\mathscr{L}}_{H,h}(hX, hY) := \min\{\mathscr{L}_H(\tilde{X}, \tilde{Y}): h\tilde{X} = hX, h\tilde{Y} = hY\} = \mathbf{EN}(X, Y; H). \tag{9.1.23}$$

(ii) For any $p \geqslant 1$

$$\tilde{\mathscr{L}}_{p,h}(hX, hY) = \mathbf{EN}(X, Y; p). \tag{9.1.24}$$

Proof. Use Jensen's inequality to obtain the necessary lower bounds. The 'optimal pair' is $\tilde{X} = \mathbb{E}X$, $\tilde{Y} = \mathbb{E}Y$. QED

Combining Theorems 8.1.1, 8.1.2 and 9.1.1, we obtain the following sharp bounds of the extremal functionals $\hat{\mathscr{L}}_H(P, Q)$ $(P, Q) \in \mathscr{P}(U))$ and $\check{\mathscr{L}}_H(P, Q)$ (see Theorem 8.1.1) in terms of the moments

$$a = \int_U h(x)P(\mathrm{d}x) \quad b = \int_U h(x)Q(\mathrm{d}x). \tag{9.1.25}$$

Corollary 9.1.4. Let $(U, \|\cdot\|)$ be a separable normed space and $H \in \mathscr{H}$.

(i) If $A(H, h)$ and $B(h)$ hold, then $\hat{\mathscr{L}}_H(P, Q) \geqslant H(|h^{-1}(a) - h^{-1}(b)|)$.
(ii) If $B(h)$ and $E(H, h)$ hold, then $\check{\mathscr{L}}_H(P, Q) \leqslant H(h^{-1}(a) + h^{-1}(b))$.

Moreover, there exist P_i, $Q_i \in \mathscr{P}(U)$, $i = 1, 2$ with $a = \int_U h(x)P_i(dx)$, $b = \int_U h(x)Q_i(dx)$ such that $\mathscr{L}_H(P_1, Q_1) = H(|h^{-1}(a) - h^{-1}(b)|)$ and $\breve{\mathscr{L}}_H(P_2, Q_2) = H(h^{-1}(a) + h^{-1}(b))$.

9.2 MOMENT PROBLEMS WITH TWO FIXED PAIRS OF MARGINAL MOMENTS; MOMENT PROBLEMS WITH FIXED LINEAR COMBINATION OF MOMENTS

In this section we shall consider the explicit representation of the following bounds

$$I(H, h_1, h_2; a_1, b_1, a_2, b_2) := \inf \mathbb{E}H(\|X - Y\|) \tag{9.2.1}$$

$$S(H, h_1, h_2; a_1, b_1, a_2, b_2) := \sup \mathbb{E}H(\|X - Y\|) \tag{9.2.2}$$

where $H, h_1, h_2 \in \mathscr{M}$, and the infimum in (9.2.1) and the supremum in (9.2.2) are taken over the set of all pairs of r.v.s $X, Y \in \mathfrak{X}(U)$, satisfying the moment conditions

$$\mathbb{E}h_i(\|X\|) = a_i \quad \mathbb{E}h_i(\|Y\|) = b_i \quad i = 1, 2 \tag{9.2.3}$$

and U is a separable normed space with norm $\|\cdot\|$. In particular, if $H(t) = t^p$, $h_i(t) = t^{q_i}$, $i = 1, 2$ ($p \geqslant 0$, $q_2 > q_1 \geqslant 0$), we write

$$I(p, q_1, q_2; a_1, b_1, a_2, b_2) := I(H, h_1, h_2; a_1, b_1, a_2, b_2) \tag{9.2.4}$$

$$S(p, q_1, q_2; a_1, b_1, a_2, b_2) := S(H, h_1, h_2; a_1, b_1, a_2, b_2). \tag{9.2.5}$$

If $H \in \mathscr{H}$, the functional I represents a primary h-minimal distance with respect to $\mathscr{L}_H(X, Y) = \mathbb{E}H(\|X - Y\|)$ with $hX = (\mathbb{E}h_1(\|X\|), \mathbb{E}h_2(\|X\|))$ (cf. Section 3.1). In particular, $I(p, q_1, q_2; a_1, b_1, a_2, b_2)^{\min(1, 1/p)}$ is a primary h-minimal metric with respect to $\mathscr{L}_p(X, Y) = \{\mathbb{E}\|X - Y\|^p\}^{\min(1, 1/p)}$. The functionals (9.2.2) and (9.2.5) may be viewed as moment functions with parameters K_H and $2^{\min(1, p)}$, respectively, see Definition 3.3.2. The moment problem with two pairs of marginal conditions is considerably more complicated and in the present section, our results are not as complete as in the previous one. Further, the conditions A to G are defined as in the previous section.

Theorem 9.2.1. Let the conditions $A(h_2, h_1)$ and $G(h_2, h_1)$ hold. Let $a_i \geqslant 0$, $b_i \geqslant 0$, $i = 1, 2$, $a_1 + a_2 > 0$, $b_1 + b_2 > 0$ and

$$h_1^{-1}(a_1) \leqslant h_2^{-1}(a_2) \quad h_1^{-1}(b_1) \leqslant h_2^{-1}(b_2). \tag{9.2.6}$$

(i) If $A(H, h_1)$, $B(h_1)$ and $D(H, h_2)$ are fulfilled, then

$$I(H, h_1, h_2; a_1, b_1, a_2, b_2) = I(H, h_1; a_1, b_1) = H(|h_1^{-1}(a_1) - h_1^{-1}(b_1)|). \tag{9.2.7}$$

(ii) Let $D(H, h_2)$ be fulfilled. If $F(H, h_1)$ holds or if $B(h_1)$ and $E(H, h_1)$ hold, then

$$S(H, h_1, h_2; a_1, b_1, a_2, b_2) = S(H, h_1; a_1, b_1) = H(h_1^{-1}(a_1) + h_1^{-1}(b_1)). \quad (9.2.8)$$

(iii) If $G(H, h_2)$ is fulfilled and $h_1^{-1}(a_1) \neq h_2^{-1}(a_2)$ or $h_1^{-1}(b_1) \neq h_2^{-1}(b_2)$, then

$$S(H, h_1, h_2; a_1, b_1, a_2, b_2) = S(H, h_1; a_1, b_1) = \infty. \quad (9.2.9)$$

Proof. By Theorem 9.1.1 (i) we have

$$\begin{aligned} I(H, h_1, h_2; a_1, b_1, a_2, b_2) &\geqslant I(H, h_1; a_1, b_1) \\ &= \phi(a_1, b_1) := H(|h_1^{-1}(a_1) - h_1^{-1}(b_1)|). \end{aligned} \quad (9.2.10)$$

Further, we shall define an appropriate sequence of r.v.s (X_t, Y_t) that satisfy the side conditions (9.2.3) and $\lim_{t\to\infty} \mathbb{E}H(\|X_t - Y_t\|) = \phi(a_1, b_1)$. Let $f(x) = h_2 \circ h_1^{-1}(x)$. Then, by Jensen's inequality and $A(h_2, h_1)$

$$f(a_1) = f(\mathbb{E}h_1(\|X\|)) \leqslant \mathbb{E}f \circ h_1(\|X\|) = a_2 \quad (9.2.11)$$

as well as $f(b_1) \leqslant b_2$. Moreover, $\lim_{t\to\infty} f(t)/t = \infty$ by $G(h_1, h_2)$.

Case 1. Suppose that $f(a_1) < a_2, f(b_1) < b_2$.

Claim. If the convex function $f \in \mathcal{M}$ and the real numbers c_1, c_2 possess the properties

$$f(c_1) < c_2 \quad \lim_{t\to\infty} f(t)/t = \infty \quad (9.2.12)$$

then there exist a positive t_0 and a function $k(t)$ $(t \geqslant t_0)$ such that the following relations hold for any $t \geqslant t_0$

$$0 < k(t) < c_1 \quad (9.2.13)$$

$$tf(c_1 - k(t)) + k(t)f(c_1 + t) = c_2(k(t) + t) \quad (9.2.14)$$

$$\frac{k(t)}{k(t) + t} \leqslant \frac{c_2}{f(c_1 + t)} \quad (9.2.15)$$

and

$$\lim_{t\to\infty} k(t) = 0. \quad (9.2.16)$$

Proof of the claim: Let us take such t_0 that $f(c_1 + t)/(c_1 + t) > c_2/c_1, t \geqslant t_0$ and consider the equation

$$F(t, x) = c_2$$

where $F(t, x) := (f(c_1 - x)t + f(c_1 + t)x)/(x + t)$. For each $t \geqslant t_0$ we have $F(t, c_1) > c_2$, $F(t, 0) = f(c_1) < c_2$. Hence, for each $t \geqslant t_0$ there exists such $x =$

$k(t)$ that $k(t) \in (0, c_1)$ and $F(t, k(t)) = c_2$, which provides (9.2.13) and (9.2.14). Further, (9.2.14) implies (9.2.15), and (9.2.13), (9.2.15) imply (9.1.16). The claim is established.

From the claim, we see that there exist $t_0 > 0$ and functions $\ell(t)$ and $m(t)$ $(t \geqslant t_0)$ such that for all $t > t_0$ we have

$$0 < \ell(t) < a_1 \quad 0 < m(t) < b_1 \tag{9.2.17}$$

$$tf(a_1 - \ell(t)) + \ell(t)f(a_1 + t) = a_2(\ell(t) + t) \tag{9.2.18}$$

$$tf(b_1 - m(t)) + m(t)f(b_1 + t) = b_2(m(t) + t) \tag{9.2.19}$$

$$\lim_{t\to\infty} \ell(t) = 0 \quad \lim_{t\to\infty} m(t) = 0. \tag{9.2.20}$$

Using (9.2.17) to (9.2.20) and the conditions $A(H, h_1)$, $D(H, h_2)$ and $G(h_2, h_1)$, one can readily obtain that the r.v.s $(X_t, Y_t)(t > t_0)$ determined by the equalities

$$\Pr\{X_t = x_i(t),\ Y_t = y_j(t)\} = p_{ij}(t),\ i, j = 1, 2,$$

where

$$x_1(t) := h_1^{-1}(a_1 - \ell(t))\bar{e}, \qquad x_2(t) := h_1^{-1}(a_1 + t)\bar{e}$$

$$y_1(t) := h_1^{-1}(b_1 - m(t))\bar{e}, \qquad y_2(t) := h_1^{-1}(b_1 + t)\bar{e}$$

$$p_{11}(t) := \min\{t/(\ell(t) + t),\ t/(m(t) + t)\}, \qquad p_{12}(t) := t/(\ell(t) + t) - p_{11}(t)$$

$$p_{21}(t) := t/(m(t) + t) - p_{11}(t), \quad p_{22}(t) := \min\{\ell(t)/(\ell(t) + t),\ m(t)/m(t) + t)\}$$

possess all the desired 'optimal' properties.

Case 2. Suppose $f(a_1) = a_2$ (i.e., $h_1^{-1}(a_1) = h_2^{-1}(a_2)$), $f(b_1) < b_2$. Then we can determine (X_t, Y_t) by the equalities $\Pr\{X_t = h_1^{-1}(a_1),\ Y_t = y_1(t)\} = t/(m(t) + t)$, $\Pr\{X_t = h_1^{-1}(a_1),\ Y_t = y_2(t)\} = m(t)/(m(t) + t)$.

Case 3. The cases $(f(a_1) < a_2, f(b_1) = b_2)$, $(f(a_1) = a_2, f(b_1) = b_2)$ are considered in the same way as in Case 2.

(ii) and (iii) are proved by analogous arguments. QED

Corollary 9.2.1. Let $a_i \geqslant 0$, $b_i \geqslant 0$, $a_1 + a_2 > 0$, $b_1 + b_2 > 0$, $a_1^{1/q_1} \leqslant a_2^{1/q_2}$, $b_1^{1/q_1} \leqslant b_2^{1/q_2}$.

(i) If $1 \leqslant q_1 \leqslant p < q_2$, then

$$I(p, q_1, q_2; a_1, b_1, a_2, b_2) = I(p, q_1; a_1, b_1) = (a_1^{1/q_1} - b_1^{1/q_1})^p. \tag{9.2.21}$$

(ii) If $0 < p \leqslant q_1$, $1 \leqslant q_1 < q_2$ then

$$S(p, q_1, q_2; a_1, b_1, a_2, b_2) = S(p, q_1; a_1, b_1) = (a_1^{1/q_1} + b_1^{1/q_1})^p. \tag{9.2.22}$$

(iii) If $0 < q_1 < q_2 < p$ and $a_1^{1/q_1} = a_2^{1/q_2}$ or $b_1^{1/q_1} = b_2^{1/q_2}$ then

$$S(p, q_1, q_2; a_1, b_1, a_2, b_2) = S(p, q_1; a_1, b_1) = \infty.$$

Corollary 9.2.1 describes situations in which the 'additional moment information' $a_2 = \mathbb{E}\|X\|^{q_2}$, $b_2 = \mathbb{E}\|Y\|^{q_2}$ does not affect the bounds

$$I(p, q_1, q_2; a_1, b_1, a_2, b_2) = I(p, q_1; a_1, a_2)$$

$$S(q_1, q_2; a_1, b_1, a_2, b_2) = S(p, q_1; a_1, a_2)$$

(and likewise Theorem 9.2.1).

Open problem 9.2.1. Find the explicit expression of $I(p, q_1, q_2; a_1, b_1, a_2, b_2)$ and $S(p, q_1, q_2; a_1, b_1, a_2, b_2)$ for all $p \geqslant 0$, $q_2 > 0$, $q_1 \geqslant 0$, (see 9.2.4), Corollary 9.2.1 and Theorem 9.2.1). One could start with the following one-dimensional version of the problem. Let h_i: $[0, \infty) \to \mathbb{R}(i = 1, 2)$ and $H: \mathbb{R} \to \mathbb{R}$ be given continuous functions with H symmetric and strictly increasing on $[0, \infty)$. Further, let X and Y be nonnegative random variables having fixed moments $a_i = \mathbb{E}h_i(X)$, $b_i = \mathbb{E}h_i(Y)$, $i = 1, 2$. The problem is to evaluate

$$I = \inf \mathbb{E}H(X - Y) \quad S = \sup \mathbb{E}H(X + Y). \tag{9.2.23}$$

If desired one could think of $X = X(t)$ and $Y = Y(t)$ as functions on the unit interval (with Lebesgue measure), see Karlin and Studden (1966), Chapter 3 and Rogosinski (1958). The five moments a_1, a_2, b_1, b_2 and $\mathbb{E}H(X \pm Y)$ depend only on the joint distribution of the pair (X, Y) and the extremal values in (9.2.23) are realized by a probability measure supported by six points. (See Rogosinski 1958, Theorem 1, Karlin and Studden 1966, Chapter 3, Kemperman 1983.) Thus the problem can also be formulated as the nonlinear programming problem to find

$$I = \inf \sum_{j=1}^{6} p_j H(u_j - v_j) \quad S = \sup \sum_{j=1}^{6} p_j H(u_j + v_j)$$

subject to

$$p_j \geqslant 0 \quad \sum_{j=1}^{6} p_j = 1 \quad u_j \geqslant 0 \quad v_j \geqslant 0, j = 1, \ldots, 6$$

$$\sum_{j=1}^{6} p_j h_i(u_j) = a_i \quad \sum_{j=1}^{6} p_j h_i(v_j) = b_i \quad i = 1, 2.$$

Such a problem becomes simpler when the functions h_i and H on $\mathbb{R}_+$ are convex (see, for example, Karlin and Studden 1966, Chapter 14). Note that in the case when U is a normed space, the moment problem was easily reduced to the one-dimensional moment problem ($U = \mathbb{R}$). This is no longer possible for general (non-normed) spaces U rendering the problem quite different from that considered in Sections 9.1 and 9.2.

Open problem 9.2.2. Let $\mu(X, Y)$ be a given compound p. metric in $(U, \|\cdot\|)$, I an arbitrary index set, α_i, β_i $(i \in I)$ be positive constants and $h_i \in \mathscr{M}$, $i \in I$. Find

$$I\{\mu; \alpha_i, \beta_i, i \in I\} = \inf\{\mu(X, Y): X, Y \in \mathfrak{X}(U), \\ \mathbb{E}h_i(\|X\|) = \alpha_i \ \mathbb{E}h_i(\|Y\|) = \beta_i, i \in I\} \quad (9.2.24)$$

and define $S\{\mu; \alpha_i, \beta_i, i \in I\}$ by changing inf to sup in (9.2.24). One very special case of the problem is

$$\mu(X, Y) = \delta(X, Y) = \begin{cases} 0 & \text{if } \Pr(X = Y) = 1 \\ 1 & \text{if } \Pr(X \neq Y) = 1 \end{cases}$$

see Example 3.1.4. Then one can easily see that

$$I\{\delta; \alpha_i, \beta_i, i \in I\} = \begin{cases} 0 & \text{if } \alpha_i = \beta_i \quad \forall i \in I \\ 1 & \text{otherwise} \end{cases} \quad (9.2.25)$$

and

$$S(\delta; \alpha_i, \beta_i, i \in I) = 1. \quad (9.2.26)$$

In Section 3.3 we introduced the μ-upper bound with fixed sum of marginal qth moments

$$\bar{\mu}(c; m, q) := \sup\{\mu(X, Y): X, Y \in \mathfrak{X}(U), m_q(X) + m_q(Y) = c\} \quad (9.2.27)$$

where μ is a compound p. distance in $\mathfrak{X}(U)$ and $m_p(X)$ is the 'qth moment'

$$m_q(X) := \mathbb{E}d(X, a)^q \quad q > 0$$

$$m_0(X) := \mathbb{E}I\{d(X, a) \neq 0\} = \Pr(X \neq a).$$

Similarly, we defined the μ-lower bound with fixed difference of marginal qth moments,

$$\underline{\mu}(c; m, q) := \inf\{\mu(X, Y): X, Y \in \mathfrak{X}(U), m_q(X) - m_q(Y) = c\}. \quad (9.2.28)$$

The next theorem gives us explicit expressions for $\underline{\mu}(c; m, q)$ and $\bar{\mu}(c; m, q)$ when μ is the p-average metric (cf. Example 3.3.1), $\mu(X, Y) = \mathscr{L}_p(X, Y) = \{\mathbb{E}\|X - Y\|^p\}^{p'}$, $p' = \min(1, 1/p)(p > 0)$ or μ is the indicator metric, $\mu(X, Y) = \mathscr{L}_0(X, Y) = \mathbb{E}\|X - Y\|^0 = \mathbb{E}I\{X \neq Y\}$. (We assume as before that (U, d), $d(x, y) := \|x - y\|$, is a separable normed space.)

We shall generalize the functionals $\underline{\mu}$ and $\bar{\mu}$ given by (9.2.27) and (9.2.28) in the following way. For any $p \geqslant 0$, $q \geqslant 0$, $\alpha, \beta, c \in \mathbb{R}$, consider

$$I(p, q, c, \alpha, \beta) := \inf\{\mathbb{E}\|X - Y\|^p: \alpha m_q + \beta m_q = c\} \quad (9.2.29)$$

and

$$S(p, q, c, \alpha, \beta) := \sup\{\mathbb{E}\|X - Y\|^p: \alpha m_q + \beta m_q = c\}. \quad (9.2.30)$$

Theorem 9.2.2. For any $\alpha > 0$, $\beta > 0$, $c > 0$, $p \geqslant 0$, $q \geqslant 0$, the following hold

$$I(p, q, c, \alpha, \beta) = \begin{cases} 0 & \text{if } q \neq 0, \text{or if } q = 0, c \leqslant \alpha + \beta \\ +\infty & \text{if } q = 0, c > \alpha + \beta \end{cases} \tag{9.2.31}$$

(the value $+\infty$ means that the infimum in (9.2.29) is taken over an empty set)

$$\begin{aligned} I(p, q, c, \alpha, -\beta) &= 0 \text{ if } \beta < \alpha,\ p^2 + q^2 \neq 0, \text{ or } 0 \leqslant p < q, \\ &\qquad\qquad \text{or } q = 0,\ p > 0 \text{ and } c \leqslant \alpha \\ &= [c(\beta^{1/(q-1)} - \alpha^{1/(q-1)})^{q-1}/(\alpha\beta)]^{p/q} \quad \text{if } \alpha \leqslant \beta,\ p \geqslant \beta > 1 \\ &= (c/\alpha)^{p/q} \quad \text{if } \alpha \leqslant \beta,\ p \geqslant q,\ 0 < q \leqslant 1 \\ &= \max\left(\frac{c - \alpha + \beta}{\alpha}, 0\right) \quad \text{if } p = q = 0,\ \beta < \alpha,\ c \leqslant \alpha \\ &= +\infty \qquad\qquad \text{if } q = 0,\ c > \alpha. \end{aligned} \tag{9.2.32}$$

Proof. Clearly, if $c > \alpha + \beta$ then there is no (X, Y) such that $\alpha m_0(X) + \beta m_0(Y) = c$. Suppose $q > 0$. Define the 'optimal' pair (X^*, Y^*) by $X^* = Y^* = (c/(\alpha + \beta))^{1/q}\bar{e}$, where $\|\bar{e}\| = 1$. Then $\mathscr{L}_p(X^*, Y^*) = 0$ for all $0 \leqslant p < \infty$ and clearly $\alpha m_q + \beta m_q = c$, i.e., (9.2.31) holds.

To prove (9.2.32) we will make use of Corollary 9.1.1 (cf. (9.1.17)). By the definition of the extremal functional $I(p, q; a, b)$ (9.1.5)

$$I(p, q, c, \alpha, -\beta) = \inf\{I(p, q; d, f): d \geqslant 0, f \geqslant 0, \alpha d - \beta f = c\} \tag{9.2.33}$$

where $I(p, q; a, b)$ admits the explicit representation (9.1.17). Solving the minimization problem (9.2.33) yields to (9.2.32). QED

Similarly we have the following explicit formulae for $S(p, q, c, \alpha, \beta)$ (9.2.30).

Theorem 9.2.3. For any $\alpha > 0$, $\beta > 0$, $c > 0$, $p \geqslant 0$, $q \geqslant 0$

$$S(p, q, c, \alpha, -\beta) = \begin{cases} +\infty & \text{if } p > 0, q > 0, \text{or } p > 0, q = 0, c \leqslant \alpha \\ 1 & \text{if } p = q = 0, c \leqslant \alpha, \text{or } p = 0, q > 0 \\ -\infty & \text{if } q = 0, c > \alpha, \end{cases}$$

(the value $-\infty$ means that the supremum in (9.2.30) is taken over an empty set)

$$\begin{aligned} S(p, q, c, \alpha, \beta) &= [c(\alpha^{1/(q-1)} + \beta^{1/(q-1)})^{q-1}/(\alpha\beta)]^{p/q} && \text{if } 0 \leqslant p \leqslant q,\ q > 1 \\ &= \left(\frac{1}{c}\min(\alpha, \beta)\right)^{p/q} && \text{if } 0 \leqslant p \leqslant q \leqslant 1,\ q > 0 \\ &= +\infty && \text{if } p > q > 0 \text{ or } p > q = 0, \\ & && \qquad c \leqslant \alpha + \beta \\ &= \min[1, c/\min(\alpha, \beta)] && \text{if } p = q = 0,\ c \leqslant \alpha + \beta \\ &= -\infty && \text{if } q = 0,\ c > \alpha + \beta. \end{aligned}$$

Using Corollary 9.2.1 one can study similar but more general moment problems:

$$\begin{array}{l}\text{minimize} \\ \text{(maximize)}\end{array} \{\mathscr{L}_p(X, Y)\colon F(m_{q_1}(X), m_{q_2}(X), m_{q_1}(Y), m_{q_2}(Y)) = 0\}.$$

PART III

Applications of Minimal p. Distances

As was clarified in the preceding chapter, the minimal distance studies natural metrical structures. The Kantorovich functional $\mathscr{A}_c$ (5.1.2), minimal distance $\hat{\mu}_c$ (5.1.16) and minimal norm $\mathring{\mu}_c$ (5.1.17) have nice metrical and topological properties. In this chapter, we shall show that the minimal structure of $\hat{\mu}_c$ and $\mathring{\mu}_c$ is especially useful in problems of approximations and stability of stochastic models. It is natural to use the topological structure of the space of laws with the metric functionals $\hat{\mu}_c$ and $\mathring{\mu}_c$ in the limit-type theorems providing weak convergence plus convergence of moments.

In this part we shall study the vague convergence, the Glivenko–Cantelli theorem, a functional limit theorem, and the stability of queueing systems in terms of minimal distances and metrics.

CHAPTER 10

Uniformity in Weak and Vague Convergence

10.1 ζ-METRICS AND UNIFORMITY CLASSES

Let (U, d) be a separable metric space (s.m.s) with Borel σ-algebra $\mathfrak{B}$. Let $\mathfrak{M}$ denote the set of all bounded non-negative measures on $\mathfrak{B}$ and $\mathscr{P}_1 = \mathscr{P}(U)$ the subset of probability measures. Let $\mathfrak{M}' \subset \mathfrak{M}$. For each class $\mathscr{F}$ of μ-integrable functions f on U ($\mu \in \mathfrak{M}'$) define on $\mathfrak{M}'$ the semimetric

$$\zeta_{\mathscr{F}}(\mu', \mu'') = \sup\left\{\left|\int f \,\mathrm{d}(\mu' - \mu'')\right| : f \in \mathscr{F}\right\} \tag{10.1.1}$$

with ζ-structure (cf. Definition 4.3.1). There is a special interest in finding, for a given semimetric ρ on $\mathfrak{M}'$, a semimetric $\zeta_{\mathscr{F}}$ which is topologically equivalent to ρ. Note that this is not always possible, see Lemma 4.3.4.

Definition 10.1.1. The class $\mathscr{F}$ is said to be *ρ-uniform* if $\zeta_{\mathscr{F}}(\mu_n, \mu) \to 0$ as $n \to \infty$ for any sequence $\{\mu_1, \mu_2, \ldots\} \subset \mathfrak{M}'$ ρ-convergent to $\mu \in \mathfrak{M}'$.

Such ρ-uniform classes were studied in Section 4.3. Here we shall investigate ρ-uniform classes in a more general setting. We generalize the notion of ρ-uniform class as follows. Let $\mathbb{K}$ be the class of pairs (f, g) of real measurable functions on U which are μ-integrable for any $\mu \in \mathfrak{M}' \subset \mathfrak{M}$. Consider the functional

$$\eta_{\mathbb{K}}(\mu', \mu'') = \sup\left\{\int f \,\mathrm{d}\mu' + \int g \,\mathrm{d}\mu'' : (f, g) \in \mathbb{K}\right\} \quad \mu', \mu'' \in \mathbb{M}. \tag{10.1.2}$$

The functional $\mu_{\mathbb{K}}$ may provide dual and explicit expressions for minimal distances. For example, define for any measures μ', μ'' with $\mu'(U) = \mu''(U)$ the class $\mathfrak{A}(\mu', \mu'')$ of all Borel measures $\tilde{\mu}$ on the direct product $U \times U$ with fixed marginals $\mu'(A) = \tilde{\mu}(A \times U)$, $\mu''(A) = \tilde{\mu}(U \times A)$, $A \in \mathfrak{B}$. Then (see Corollary 5.2.2), for $1 \leqslant p < \infty$, if $\int d^p(x, a)(\mu' + \mu'')(\mathrm{d}x) < \infty$, we have that (10.1.2) gives the dual form of the p-average metric, i.e.,

$$\eta_{\mathbb{K}(p)}(\mu', \mu'') = \inf\left\{\int d^p(x, y)\tilde{\mu}(dx \times dy) \colon \tilde{\mu} \in \mathfrak{A}(\mu', \mu'')\right\} \tag{10.1.3}$$

where $\mathbb{K}(p)$ is the set of all pairs (f, g) for which $f(x) + g(y) \leqslant d^p(x, y)$, $x, y \in U$.

Definition 10.1.2. We call the class $\mathbb{K}$ a *ρ-uniform class* (*in a broad sense*) if for any sequence $\{\mu_1, \mu_2, \ldots\} \subset \mathfrak{M}' \subset \mathfrak{M}$, the ρ-convergence to $\mu \in \mathfrak{M}'$ implies $\lim_{n\to\infty} \eta_{\mathbb{K}}(\mu_n, \mu) = 0$.

The notation $\mu_n \xrightarrow{w} \mu$ denotes as usual the weak convergence of the sequence $\{\mu_1, \mu_2, \ldots\} \subset \mathfrak{M}$ to $\mu \in \mathfrak{M}$.

Theorem 10.1.1. Let $\mu, \mu_1, \mu_2, \ldots$ be a sequence of measures in $\mathfrak{M}$ and $\mu(U) = \mu_n(U)$, $n = 1, 2, \ldots$. Let $B(t)$, $t \geqslant 0$, be a convex non-negative function, $B(0) = 0$, satisfying the Orlicz condition: $\sup\{B(2t)/B(t) \colon t > 0\} < \infty$. If

$$\int B(d(x, a))(\mu_n + \mu)(dx) < \infty$$

then the joint convergence

$$\mu_n \xrightarrow{w} \mu \qquad \int B(d(x, a))(\mu_n - \mu)(dx) \to 0 \tag{10.1.4}$$

is equivalent to the convergence $\eta_{\mathbb{B}}(\mu_n, \mu) \to 0$, where $\mathbb{B}$ is the class of pairs (f, g) such that $f(x) + g(y) \leqslant B(d(x, y))$, $x, y \in U$.

Proof. Let π be the Prokhorov metric in $\mathfrak{M}$, i.e.,

$$\pi(\mu', \mu'') = \inf\{\varepsilon > 0 \colon \mu'(A) \leqslant \mu''(A^\varepsilon) + \varepsilon,\ \mu''(A) \leqslant \mu'(A^\varepsilon) + \varepsilon \text{ for any closed set } A \subset U\} \tag{10.1.5}$$

(see, for example, Hennequin and Tortrat 1965).

Then, as in Lemma 8.2.1 (see (8.2.5), (8.2.7)) we conclude that

$$\begin{aligned} B(\pi(\mu', \mu''))\pi(\mu', \mu'') &\leqslant \eta_{\mathbb{B}}(\mu', \mu'') \\ &\leqslant B(\pi(\mu', \mu'')) + K_B\left[2\pi(\mu', \mu'')B(M) + \int_{d(x,a)>M} B(d(x, a))(\mu' + \mu'')(dx)\right] \end{aligned} \tag{10.1.6}$$

for any $\mu', \mu'' \in \mathfrak{M}$, $M > 0$, $a \in U$ and $K_B := \sup\{B(2t)/B(t); t > 0\}$. Hence, (10.1.4) provides $\eta_{\mathbb{B}}(\mu_n, \mu) \to 0$.

To prove (10.1.4) provided that $\eta_{\mathbb{B}}(\mu_n, \mu) \to 0$, we use the following inequality: for any $\mu', \mu'' \in \mathfrak{M}$ with $\mu'(U) = \mu''(U)$ and $\int B(d(x, a))(\mu' + \mu'')(dx) < \infty$, and for any $M > 0$ and $a \in U$ we have

$$\int B(d(x, a))I\{d(x, a) > M\}\mu'(dx)$$
$$\leqslant (K_B + K_B^2)(\hat{\mathscr{L}}_B(\mu', \mu'') + \int B(d(x, a))I\{d(x, a) > M/2\}\mu''(\mathrm{d}x). \tag{10.1.7}$$

In the above inequality, $\hat{\mathscr{L}}_B(\mu', \mu'') := \inf\{\mathscr{L}_B(\tilde{\mu}): \tilde{\mu} \in \mathfrak{A}(\mu', \mu'')\}$ is the minimal distance relative to $\mathscr{L}_B(\tilde{\mu}) := \int B(d(x, y))\tilde{\mu}(\mathrm{d}x, \mathrm{d}y)$. To prove (10.1.7) observe that for any $\mu \in \mathfrak{A}(\mu', \mu'')$ we have

$$\int B(d(x, a))I\{d(x, a) > M\}\mu'(\mathrm{d}x)$$
$$\leqslant K_B \int B(d(y, a))I\{d(x, a) > M\}\tilde{\mu}(\mathrm{d}x, \mathrm{d}y) + K_B\mathscr{L}_B(\tilde{\mu})$$

where

$$\int B(d(y, a))I\{d(x, a) > M\}\tilde{\mu}(\mathrm{d}x, \mathrm{d}y)$$
$$\leqslant B(M)\mu'(d(x, a) > M) + \int B(d(y, a))I\{d(y, a) > M\}\mu''(\mathrm{d}y)$$

and

$$\mu'(d(x, a) > M) \leqslant \frac{1}{B(M/2)}\left(\mathscr{L}_B(\tilde{\mu}) + \int B(d(y, a))I\{d(y, a) > M/2\}\mu''(\mathrm{d}y)\right).$$

Combining the last three inequalities we obtain

$$\int B(d(x, a))I\{d(x, a) > M\}\mu'(\mathrm{d}x)$$
$$\leqslant K_B\mathscr{L}_B(\tilde{\mu}) + K_B^2\mathscr{L}_B(\tilde{\mu}) + K_B \int B(d(y, a))I\{d(y, a) > M\}\mu''(\mathrm{d}y)$$
$$+ K_B^2 \int B(d(y, a))I\{d(y, a) > M/2\}\mu''(\mathrm{d}y).$$

Passing to the minimal distances $\hat{\mathscr{L}}_B$ in the last estimate yields the required (10.1.7). Then $\eta_{\mathbb{B}}(\mu_n, \mu) \to 0$ together with (10.1.6), and (10.1.7) implies

$$\mu_n \to \mu, \quad \lim_{M \to \infty} \sup_n \int B(d(x, a))I\{d(x, a) > M\}\mu_n(\mathrm{d}x) = 0.$$

The above limit relations complete the proof of (10.1.4), cf. Billingsley (1968), Theorem 5.4. QED

Recall that if $G(x)$ is a non-negative continuous function on U and $\{\mu_0, \mu_1, \dots\} \subset \mathfrak{M}$, $\int G\,\mathrm{d}\mu_n < \infty$, $n = 0, 1, \dots$, then the joint convergence

$$\mu_n \xrightarrow{w} \mu_0 \quad \int G\,\mathrm{d}(\mu_n - \mu_0) \to 0 \quad n \to \infty \tag{10.1.8}$$

is called *G-weak convergence*, cf. Definition 4.2.2.

Theorem 10.1.2. The G-weak convergence (10.1.8) in

$$\mathfrak{M}_G := \left\{\mu \in \mathfrak{M} : \int G\,\mathrm{d}\mu < \infty\right\}$$

is equivalent to the weak convergence $\lambda_n \xrightarrow{w} \lambda_0$, where

$$\lambda_n(A) = \int_A (1 + G(x))\mu_n(\mathrm{d}x) \quad n = 0, 1, \dots, \quad A \in \mathfrak{B}. \tag{10.1.9}$$

Proof. Suppose (10.1.8) holds, then define the measures $\nu_i(B) := \int_B G\,\mathrm{d}\mu_i$ $(i = 0, 1, \dots)$ on $A_G \cap \mathfrak{B}$ where $A_G := \{x : G(x) > 0\}$. For any continuous and bounded function f,

$$\left|\int f\,\mathrm{d}(\nu_n - \nu_0)\right|$$
$$\leqslant \left|\int f(1 + G)I\{G \leqslant N\}\,\mathrm{d}(\mu_n - \mu)\right| + \int \|f\|(1 + G)I\{G > N\}\,\mathrm{d}(\mu_n + \mu) \tag{10.1.10}$$

where $\|f\| := \sup\{|f(x)| : x \in U\}$ and $N > 0$. For any N with $\mu_0(G(x) = N) = 0$, by the weak convergence $\mu_n \xrightarrow{w} \mu_0$, we have that the first integral on the right-hand side of (10.1.10) converges to zero and hence (10.1.10) and (10.1.8) imply $\lambda_n \xrightarrow{w} \lambda_0$.

Conversely, if $\lambda_n \xrightarrow{w} \lambda_0$ then for any continuous and bounded function f and $g = f/(1 + G)$ we have $\int g\,\mathrm{d}\lambda_n \to \int g\,\mathrm{d}\lambda_0$ since g is also continuous and bounded i.e., $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu_0$. Finally, by $\int_U \mathrm{d}\lambda_n \to \int_U \mathrm{d}\lambda_0$, we have $\mu_n(U) + \int G\,\mathrm{d}\mu_n \to \mu_0(U) + \int G\,\mathrm{d}\mu_0$ and thus (10.1.8) holds. QED

Recall the *G-weighted Prokhorov metric* (cf. (4.2.5)),

$$\pi_{\lambda, G}(\mu_1, \mu_2) = \inf\{\varepsilon > 0 : \lambda_1(A) \leqslant \lambda_1(A^{\lambda\varepsilon}) + \varepsilon$$
$$\lambda_2(A) \leqslant \lambda_1(A^{\lambda\varepsilon}) + \varepsilon \quad \forall A \in \mathfrak{B}\} \tag{10.1.11}$$

where λ_i is defined by (10.1.9) and $\lambda > 0$.

Corollary 10.1.1. $\pi_{\lambda,G}$ metrizes the G-weak convergence in $\mathfrak{M}_G$.

Proof. For any $\mu_0, \mu_n \in \mathfrak{M}_G$, $\pi_{\lambda,G}(\mu_n, \mu_0) \to 0$ if and only if $\pi_{1,G}(\mu_n, \mu_0) \to 0$ which by the Prokhorov (1956) theorem is equivalent to $\lambda_n \xrightarrow{w} \lambda_0$. An appeal to Theorem 10.1.2 proves the corollary. QED

In the next theorem, 10.1.3, we shall omit the basic restriction in Theorem 10.1.1, $\mu_n(U) = \mu(U)$, $n = 1, 2, \ldots$. Define the class $\mathcal{OR}$ of continuous non-negative functions $B(t)$, $t \geqslant 0$, $\overline{\lim}_{t\to 0} \sup_{0 \leqslant s \leqslant t} B(s) = 0$, satisfying the following condition: there exist a point $t_0 \geqslant 0$ and a non-decreasing continuous function $B_0(t)$, $t \geqslant 0$, $K_{B_0} := \sup\{B_0(2t)/B_0(t): t > 0\} < \infty$, $B_0(0) = 0$, such that $B(t) = B_0(t)$ for $t \geqslant t_0$.

Lemma 10.1.1. Let $B \in \mathcal{OR}$ and $\mu, \mu_1, \mu_2, \ldots$ be a sequence of measures in $\mathfrak{M}$ satisfying (10.1.4), $\mu(U) = \mu_n(U)$, $\int B(d(x, a))(\mu_n + \mu)(\mathrm{d}x) < \infty$, $n = 1, 2, \ldots$. Then $\eta_{\mathbb{B}}(\mu_n, \mu) \to 0$ as $n \to \infty$, where $\mathbb{B}$ is defined as in Theorem 10.1.1.

Proof. One can easily see that the joint convergence (10.1.4) is equivalent to

$$\mu_n \to \mu, \quad \lim_{M\to\infty} \sup_n \int B_0(d(x, a))I\{d(x, a) > M\}\mu_n(\mathrm{d}x) = 0 \qquad (10.1.12)$$

(see Billingsley 1968, Theorem 5.4 and the proof of Theorem 6.3.1). Then, as in the proof of (8.2.6) we conclude that for any $M \geqslant t_0$

$$\begin{aligned}
\sup&\left\{\int f\,\mathrm{d}\mu_n + \int g\,\mathrm{d}\mu : f(x) + g(y) \leqslant B(d(x, y)) \quad \forall x, y \in U\right\} \\
&\leqslant \inf\left\{\int B(d(x, y))\tilde{\mu}(\mathrm{d}x, \mathrm{d}y) : \tilde{\mu} \in \mathfrak{A}(\mu, \mu_n)\right\} \\
&\leqslant \tilde{B}(\pi(\mu_n, \mu)) + K_{B_0}\left[2\pi(\mu_n, \mu)B_0(M)\right. \\
&\qquad \left. + \int B_0(d(x, a))I\{d(x, a) > M\}(\mu_n + \mu)(\mathrm{d}x)\right]
\end{aligned}$$

where $\tilde{B}(t) = \sup\{B(s): 0 \leqslant s \leqslant t\}$. The last inequality implies $\eta_{\mathbb{B}}(\mu_n, \mu) \to 0$, see (10.1.2). QED

Theorem 10.1.3. Let $B \in \mathcal{OR}$ and $\mu, \mu_1, \mu_2, \ldots$ be a sequence of measures in $\mathfrak{M}$ satisfying (10.1.4) and $\int B(d(x, a))(\mu_n + \mu)(\mathrm{d}x) < \infty$. Then

$$\lim_{n\to\infty} \eta_{\mathbb{K}_1}(\mu_n, \mu) = 0 \qquad (10.1.13)$$

where $\mathbb{K}_1 = \{(f, g): f(x) + g(y) \leqslant B(d(x, y)), |g(x)| \leqslant B(d(x, b)), x, y \in U\}$, and b is an arbitrary point in U.

Proof. As $B \in \mathcal{OR}$ it is enough to prove (10.1.13) for $b = a$. Setting $c_n = \mu(U)/\mu_n(U)$ we have $\lim_{n\to\infty} \eta_{\mathbb{B}}(c_n\mu_n, \mu) = 0$ by Lemma 10.1.1. Hence, as $n \to \infty$, $0 \leqslant \eta_{\mathbb{K}_1}(\mu_n, \mu) \leqslant 1/c_n\eta_{\mathbb{B}}(c_n\mu_n, \mu) + |1/c_n - 1| \int B(d(x, b))\mu(dx) \to 0$. QED

In the next theorem we omit the condition $B \in \mathcal{OR}$ but shall assume that the class $\mathcal{G} = \{g: (f, g) \in \mathbb{K}_1\}$ is *equicontinuous*, i.e.,

$$\lim_{y\to x} \sup\{|g(x) - g(y)|: g \in \mathcal{G}\} = 0 \quad x \in U. \tag{10.1.14}$$

Theorem 10.1.4 (cf. Ranga-Rao, 1962). Let G be a non-negative continuous function on U and h a non-negative function on $U \times U$. Let $\mathbb{K}$ be the class of pairs (f, g) of measurable functions on U such that $(0, 0) \in \mathbb{K}$ and $f(x) + g(y) \leqslant h(x, y)$, $x, y \in U$. Then $\mathbb{K}$ is a π_G-uniform class (cf. Definition 10.1.2 with $\mathfrak{M}' = \mathfrak{M}_G$) if at least one of the following conditions holds:

(a) $\lim_{y\to x} h(x, y) = h(x, x) = 0$ for all $x \in U$, the class $\mathcal{F} = \{f: (f, g) \in \mathbb{K}\}$ is equicontinuous, and $|f(x)| \leqslant G(x)$ for all $x \in U$, $f \in \mathcal{F}$;

(b) $\lim_{y\to x} h(y, x) = h(x, x) = 0$ for all $x \in U$ and the class $\mathcal{G} = \{g: (f, g) \in \mathbb{K}\}$ is equicontinuous, $|g(x)| \leqslant G(x)$ for all $x \in U$, $g \in \mathcal{G}$.

Proof. Suppose that $G \equiv 1$. Let $\varepsilon > 0$ and (a) holds. For any $z \in U$ there is $\delta = \delta(z) > 0$ such that if $B(z) := \{x: d(x, z) < \delta\}$ then

$$\sup_{x\in B(z)} h(z, x) \leqslant \varepsilon/2 \quad \sup_{f\in\mathcal{F}} \sup_{x\in B(z)} |f(x) - f(z)| < \varepsilon/2. \tag{10.1.15}$$

Without loss of generality we assume that $\mu(\mathring{B}(z)) = 0$ ($\mathring{B}$ is the boundary of B). As U is a s.m.s, there exists $z_1, z_2, \ldots$ such that $\bigcup_{j=1}^{\infty} B(z_j)$. Setting $A_1 := B(z_1)$, $A_j := B(z_j)\setminus\bigcup_{k=1}^{j-1} B(z_k)$, $j = 2, 3, \ldots,$ we have $f(x)+g(y) = f(x)-f(z_j)+f(z_j)+g(y) \leqslant \varepsilon/2 + h(z_j, y) \leqslant \varepsilon$ for any $x, y \in A_j$. Let $x_j \in A_j$, $j = 1, 2, \ldots$. Then, by

$$f(x) + g(y) \leqslant \varepsilon \quad \forall x, y \in A_j, j = 1, 2, \ldots \tag{10.1.16}$$

it follows that

$$\sum_{j=1}^{\infty} f(x_j)\mu(A_j) + \int g\,d\mu = \sum_{j=1}^{\infty} \int_{A_j} f(x_j) + g(x)\mu(dx) \leqslant \varepsilon\mu(U) \text{ for any } (f, g) \in \mathbb{K}. \tag{10.1.17}$$

Also,

$$\left|\int f(x)\mu_n(dx) - \sum_{j=1}^{\infty} f(x_j)\mu_n(A_j)\right| \leqslant \varepsilon\mu_n(U) \tag{10.1.18}$$

and

$$\sum_{j=1}^{\infty} |f(x_j)(\mu_n(A_j) - \mu(A_j))| \leqslant \sum_{j=1}^{\infty} |\mu_n(A_j) - \mu(A_j)| \to 0 \quad \text{as } n \to \infty \tag{10.1.19}$$

by (10.1.15) and $\mu_n(A_j) \to \mu(A_j)$, respectively. Combining relations (10.1.17) to (10.1.19) and taking into account that $\mu_n(U) \to \mu(U)$ we have that $0 \leqslant \eta_{\mathbb{K}}(\mu_n, \mu) \to 0$.

In the general case, let $A_G := \{x: G(x) > 0\}$. Define measures ν_n and ν on $\mathfrak{B}$ by $\nu_n(B) = \int_B G\,\mathrm{d}\mu_n$ and $\nu(B) = \int_B G\,\mathrm{d}\mu$ respectively. Convergence $\pi_G(\mu_n, \mu) \to 0$ implies $\nu_n \xrightarrow{\mathrm{w}} \nu$ as $n \to \infty$. To reduce the general case to the case $G \equiv 1$, denote $f_1(x) := f(x)/G(x)$, $g_1(x) := g(x)/G(x)$ for $x \in A_G$, $\mathbb{K}_1 := \{(f_1, g_1): (f, g) \in \mathbb{K}\}$, $\mathscr{F}_1 := \{f_1: f \in \mathscr{F}\}$ and

$$h_1(x, y) := \frac{h(x, y)}{G(y)} + \left|1 - \frac{G(x)}{G(y)}\right|.$$

Then

$$f_1(x) + g_1(y) = \frac{f(x) + g(y)}{G(y)} + \frac{f(x)}{G(x)} - \frac{f(x)}{G(y)} \leqslant h_1(x, y)$$

and thus $\eta_{\mathbb{K}}(\mu_n, \mu) = \eta_{\mathbb{K}_1}(\nu_n, \nu) \to 0$ as $n \to \infty$. By the symmetry (*b*) also implies $\eta_{\mathbb{K}_1}(\nu_n, \nu) \to 0$. QED

For any continuous non-negative function $b(t)$, $t \geqslant 0$, $b(0) = 0$, we define the class $A_b = A_b(c)$, $c \in U$, of all real functions f on U with $f(c) = 0$ and norm

$$\mathrm{Lip}_b(f) = \sup\{|f(x) - f(y)|/D(x, y): x \neq y,\ x, y \in U\} \leqslant 1$$

where $D(x, y) = d(x, y)\{1 + b(d(x, c)) + b(d(y, c))\}$.

Let $C(t) = t(1 + b(t))$, $t \geqslant 0$, and $p(x, y)$ be a non-negative function on $U \times U$ continuous in each argument, $p(x, x) = 0$, $x \in U$, and let $\mathfrak{C}$ be the set of pairs $(f, g) \in A_b \times A_b$ for which $f(x) + g(y) \leqslant p(x, y)$, $x, y \in U$.

Corollary to 10.1.2 (cf. Fortet and Mourier (1953)). Let

$$\int C(d(x, c))(\mu_n + \mu)(\mathrm{d}x) < \infty, \ n = 1, 2, \ldots.$$

Then

(a) If

$$\mu_n \xrightarrow{\mathrm{w}} \mu \qquad \int C(d(x, c))(\mu_n - \mu)(\mathrm{d}x) \to 0 \tag{10.1.20}$$

then

$$\eta_{\mathfrak{C}}(\mu_n, \mu) \to 0. \tag{10.1.21}$$

(b) If $p(x, y) \geqslant D(x, y)$, $x, y \in U$ and

$$K := \sup\{|C(s) - C(t)|/[(s - t)(1 + b(s) + b(t))]: s > t \geqslant 0\} < \infty \tag{10.1.22}$$

then (10.1.21) implies (10.1.20).

Proof. (a) For any $x \in U$ and $f \in A_b$, $|f(x)| \leqslant C(d(x, c))$. The class A_b is clearly equicontinuous and thus (10.1.21) follows from Theorem 10.1.4.

(b) As $p(x, y) \geqslant D(x, y)$, $x, y \in U$, it follows that

$$\eta_{\mathfrak{C}}(\mu', \mu'') \geqslant \zeta_{A_b}(\mu', \mu'') \quad \mu', \mu'' \in \mathfrak{M}. \tag{10.1.23}$$

Applying Theorem 10.1.4 with $g = -f$ and $h = D$ we see that ζ_{A_b}-convergence yields $\mu_n \xrightarrow{w} \mu$. As $K < \infty$ in (10.1.22), the function $(1/K)C(d(x, c))$, $x \in U$ belongs to the class A_b and hence (10.1.21) implies $\int C(d(x, c))(\mu_n - \mu)(\mathrm{d}x) \to 0$. QED

10.2 METRIZATION OF THE VAGUE CONVERGENCE

In this section we shall study ρ-uniform classes in the space $\mathfrak{N}$ of all Borel measures $\nu\colon \mathfrak{B} \to [0, \infty]$ finite on the ring $\mathfrak{B}_0$ of all bounded Borel subsets of (U, d). In particular, this will give two types of metrics metrizing the vague convergence in $\mathfrak{N}$.

Definition 10.2.1. The sequence of measures $\{\nu_1, \nu_2, \dots\} \subset \mathfrak{N}$ *vaguely converges* to $\nu \in \mathfrak{N}(\nu_n \xrightarrow{v} \nu)$ if

$$\int f \,\mathrm{d}\nu_n \to \int f \,\mathrm{d}\nu \quad \text{for } f \in \bigcup_{m=1}^{\infty} \mathscr{F}_m \tag{10.2.1}$$

where $\mathscr{F}_m$, $m = 1, 2, \dots$, is the set of all bounded continuous functions on U equal to zero on $S_m = \{x\colon d(x, a) < m\}$ (see Kallenberg 1975, Kesten *et al.* 1978).

Theorem 10.2.1. Let h be a non-negative function on $U \times U$, $\lim_{y \to x} h(x, y) = h(x, x) = 0$. Let $\mathbb{K}_m$ be the class of pairs (f, g) of measurable functions such that $(0, 0) \in \mathbb{K}_m$, $f(x) + g(y) \leqslant h(x, y)$, $x, y \in U$, $f(x) = g(x) = 0$, $x \notin S_m$, and let the class $\Phi_m = \{f\colon (f, g) \in \mathbb{K}_m\}$ be equicontinuous and uniformly bounded. Then, if for the sequence $\{\nu_0, \nu_1, \dots\}$, $\nu_n \xrightarrow{v} \nu_0$, then $\lim_{n\to\infty} \eta_{\mathbb{K}_m}(\nu_n, \nu_0) = 0$, where $\eta_{\mathbb{K}_m}$ is given by (10.1.2), with $\mathfrak{M}'$ replaced by $\mathfrak{N}$.

Proof. Let $\theta(x) := \max(0, 1 - d(x, S_m))$, $x \in U$ and $\mu_n(A) := \int_A \theta \,\mathrm{d}\nu_n$, $A \in \mathfrak{B}$, $n = 0, 1, \dots$. Then by $\nu_n \xrightarrow{v} \nu_0$, we have $\mu_n \xrightarrow{w} \mu_0$. By virtue of Theorem 10.1.4 we obtain $\eta_{\mathbb{K}_m}(\mu_n, \mu_0) = \eta_{\mathbb{K}_m}(\nu_n, \nu_0) \to 0$ as $n \to \infty$. QED

Now we shall look into the question of metrization of vague convergence. Known methods of metrization (see Kallenberg 1975, Szacz 1975, Kersten *et al.* 1978) are too complicated from the viewpoint of the structure of the introduced metrics or use additional restrictions on the space $\mathfrak{N}$.

Let $\mathscr{F}\mathscr{L}_m = \{f\colon U \to \mathbb{R},\ |f(x) - f(y)| \leqslant d(x, y),\ x, y \in U,\ f(x) = 0,\ x \notin S_m\}$, $m = 1, 2, \ldots$. Set $\mathbf{K}_m$ to be the following ζ-metric (cf. (10.1.1)

$$\mathbf{K}_m(v', v'') = \zeta_{\mathscr{F}\mathscr{L}_m}(v', v''),\ v', v'' \in \mathfrak{N} \quad m = 1, 2, \ldots \tag{10.2.2}$$

and define the metric

$$\mathbf{K}(v', v'') = \sum_{m=1}^{\infty} 2^{-m}\mathbf{K}_m(v', v'')/[1 + \mathbf{K}_m(v', v'')] \quad v', v'' \in \mathfrak{N}. \tag{10.2.3}$$

Clearly, in the subspace $\mathfrak{M}_0$ of all Borel non-negative measures with common bounded support the metric $\mathbf{K}$ is topologically equivalent to the Kantorovich metric

$$\ell_1(v', v'') := \sup\left\{\left|\int f\,\mathrm{d}(v' - v'')\right| : f\colon U \to \mathbb{R},\ \text{bounded},\right.$$
$$\left. |f(x) - f(y)| \leqslant d(x, y),\ x, y \in U\right\} \tag{10.2.4}$$

see Example 3.2.2.

Corollary 10.2.1. For any separable metric space (U, d) the metric $\mathbf{K}$ metrizes the vague convergence in $\mathfrak{N}$.

Proof. For any metric space (U, d), a necessary and sufficient condition for $v_n \xrightarrow{\mathrm{v}} v$ is

$$\int f\,\mathrm{d}v_n \to \int f\,\mathrm{d}v \quad \text{for any } f \in \mathscr{F}\mathscr{L} := \bigcup_m \mathscr{F}\mathscr{L}_m. \tag{10.2.5}$$

Actually, if (10.2.5) holds, then for any $\varepsilon > 0$, $B \in \mathfrak{B}_0$ (i.e. B is a bounded Borel set) we have

$$\int f_{\varepsilon, B}\,\mathrm{d}v_n \to \int f_{\varepsilon, B}\,\mathrm{d}v \tag{10.2.6}$$

where $f_{\varepsilon, B}(x) := \max(0, 1 - d(x, B)/\varepsilon)$. For any $\varepsilon > 0$, $B \in \mathfrak{B}$, define the sets $B^\varepsilon := \{x\colon d(x, B) < \varepsilon\}$, $B_{-\varepsilon} := \{x\colon d(x, U\backslash B) \geqslant \varepsilon\}$, $\mathbb{B}^\varepsilon B := B^\varepsilon \backslash B_{-\varepsilon}$. For any μ', $\mu'' \in \mathfrak{N}$ and $B \in \mathfrak{B}_0$,

$$\mu'(B) \leqslant \int f_{\varepsilon, B}\,\mathrm{d}\mu' \leqslant \int f_{\varepsilon, B}\,\mathrm{d}(\mu' - \mu'') + \mu''(B^\varepsilon)$$

and hence,

$$\mu'(B) \leqslant \mu'(B_{-\varepsilon}) + \mu'(\mathbb{B}^\varepsilon B) \leqslant \int f_{\varepsilon, B_{-\varepsilon}}\,\mathrm{d}(\mu' - \mu'') + \mu''(B) + \mu'(\mathbb{B}^\varepsilon B)$$

and

$$\mu'(B) \leqslant \int f_{\varepsilon,B}\,\mathrm{d}(\mu' - \mu'') + \mu''(B^\varepsilon) \leqslant \int f_{\varepsilon,B}\,\mathrm{d}(\mu' - \mu'') + \mu''(B) + \mu''(\mathbb{B}^\varepsilon B).$$

By the symmetry,

$$|\mu'(B) - \mu''(B)| \leqslant \left|\int f_{\varepsilon,B_{-\varepsilon}}\,\mathrm{d}(\mu' - \mu'')\right| + \left|\int f_{\varepsilon,B}\,\mathrm{d}(\mu' - \mu'')\right| + \min(\mu'(\mathbb{B}^\varepsilon B), \mu''(\mathbb{B}^\varepsilon B)).$$

Hence, $\limsup_{n\to\infty} |\mu_n(B) - \mu(B)| \leqslant \mu(\mathbb{B}^\varepsilon B)$ and thus $\nu_n \xrightarrow{\mathrm{v}} \nu$.

In particular, from (10.2.5) it follows that the convergence $\mathbf{K}(\nu_n, \nu) \to 0$ implies $\nu_n \xrightarrow{\mathrm{v}} \nu$.

Conversely, suppose $\nu_n \xrightarrow{\mathrm{v}} \nu$. By virtue of Theorem 10.2.1, if Θ_m is a class of equicontinuous and uniformly bounded functions $f(x)$, $x \in U$ such that $f(x) = 0$ for $x \notin S_m$, then $\sup\{|\int f\,\mathrm{d}(\nu_n - \nu)|: f \in \Theta_m\} \to 0$ as $n \to \infty$. Setting $\Phi_m = \mathscr{F}\mathscr{L}_m$, $m = 1, 2, \ldots$ we get $\mathbf{K}(\nu_n, \nu) \to 0$. QED

For all $m = 1, 2, \ldots l$ define

$$\mathbf{\Pi}_m(\nu', \nu'') := \inf\{\varepsilon > 0: \nu'(B) \leqslant \nu''(B^\varepsilon) + \varepsilon,\ \nu''(B) \leqslant \nu'(B^\varepsilon) + \varepsilon,\ \forall B \in \mathfrak{B},\ B \subset S_m\} \quad \nu', \nu'' \in \mathfrak{N}$$

and the *Prokhorov metric* in $\mathfrak{N}$

$$\mathbf{\Pi}(\nu', \nu'') = \sum_{m=1}^{\infty} 2^{-m}\mathbf{\Pi}_m(\nu', \nu'')/[1 + \mathbf{\Pi}_m(\nu', \nu'')]. \tag{10.2.7}$$

Obviously, the metric $\mathbf{\Pi}$ does not change if we replace $\mathfrak{B}$ by the set of all closed subsets of U as well as if we replace $B^\varepsilon = \{x: d(x, B) < \varepsilon\}$ by its closure. In $\mathfrak{M}_0$ (the space of Borel non-negative measures with common bounded support) the metric $\mathbf{\Pi}$ is equivalent to $\boldsymbol{\pi}$. We find from Corollary 10.2.1 that $\mathbf{\Pi}$ metrizes the vague convergence in $\mathfrak{N}$. If (U, d) is a complete separable metric space, then $(\mathfrak{N}, \mathbf{K})$ and $(\mathfrak{N}, \mathbf{\Pi})$ are also complete separable metric spaces. Here we refer to Hennequin and Tortrat (1965) for the similar problem (the Prokhorov completeness theorem) concerning the metric space $\mathfrak{M} = \mathfrak{M}(U)$ of all bounded non-negative measures with Prokhorov metric

$$\boldsymbol{\pi}(\mu, \nu) = \sup\{\varepsilon > 0: \mu(F) \leqslant \nu(F^\varepsilon) + \varepsilon,\ \nu(F) \leqslant \mu(F^\varepsilon) + \varepsilon \quad \forall \text{ closed } F \subset A\}. \tag{10.2.8}$$

CHAPTER 11

Glivenko–Cantelli Theorem and Bernstein–Kantorovich Invariance Principle

11.1 FORTET-MOURIER, VARADARAJAN AND WELLNER THEOREMS

Let (U, d) be a s.m.s. and let $\mathscr{P}(U)$ be the set of all probability measures on U. Let $X_1, X_2, \ldots$ be a sequence of r.v.s with values in U and corresponding distributions $P_1, P_2, \ldots \in \mathscr{P}(U)$. For any $n \geqslant 1$, define the 'empirical measure'

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

and the 'average' measure

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

Let $\mathscr{A}_c$ be the Kantorovich functional (5.1.2),

$$\mathscr{A}_c(P_1, P_2) = \inf\left\{\int_{U \times U} c(x, y)P(\mathrm{d}x, \mathrm{d}y)\colon P \in \mathscr{P}^{(P_1, P_2)}\right\} \tag{11.1.1}$$

where $c \in \mathfrak{C}$. Recall that $\mathscr{P}^{(P_1, P_2)}$ is the set of all laws on $U \times U$ with fixed marginals P_1 and P_2 and $\mathfrak{C}$ is the class of all functions $c(x, y) = H(d(x, y))$, $x, y \in U$, where the function H belongs to the class $\mathscr{H}$ of all non-decreasing functions on $[0, \infty)$ for which $H(0) = 0$ and which satisfy the Orlicz condition

$$K_H = \sup\{H(2t)/H(t)\colon t > 0\} < \infty \tag{11.1.2}$$

see Example 2.2.1.

We now state the well known theorems of Fortet and Mourier (1953), Varadarajan (1958a), and Wellner (1981) in terms of $\mathscr{A}_c$ relying on the following criterion for μ_c-convergence of measures (cf. Theorem 10.1.1).

Theorem 11.1.1. Let $c \in \mathfrak{C}$ and $\int_U c(x, a)P_n(\mathrm{d}x) < \infty$, $n = 0, 1, \ldots$. Then

$$\lim_{n\to\infty} \mathscr{A}_c(P_n, P_0) = 0 \text{ if and only if } P_n \xrightarrow{w} P_0, \quad \lim_{n\to\infty} \int_U c(x, b)(P_n - P_0)(\mathrm{d}x) = 0 \tag{11.1.3}$$

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

Theorem 11.1.2. (Fortet–Mourier 1953). If $P_1 = P_2 = \cdots = \mu$ and $c_0(x, y) = d(x, y)/(1 + d(x, y))$, then $\mathscr{A}_{c_0}(\mu_n, \mu) \to 0$ almost surely (a.s.) as $n \to \infty$.

Theorem 11.1.3 (Varadarajan 1958a). If $P_1 = P_2 = \cdots = \mu$ and $c(c \in \mathfrak{C})$ is a bounded function, then $\mathscr{A}_c(\mu_n, \mu) \to 0$ a.s. as $n \to \infty$.

Theorem 11.1.4 (Wellner 1981). If $\bar{P}_1, \bar{P}_2, \ldots$ is a tight sequence, then $\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.

Proof. We follow the proof of the original Wellner's Theorem (see Wellner 1981, Dudley 1969, Theorem 8.3). By the SLLN

$$\int_U f \, \mathrm{d}(\mu_n - \bar{P}_n) \to 0 \quad \text{a.s., as } n \to \infty \tag{11.1.4}$$

for any bounded continuous function on U. Since $\{\bar{P}_n\}_{n\geq 1}$ is a tight sequence; then for any $\varepsilon > 0$ there exists a compact K_ε, such that $\bar{P}_n(K_\varepsilon) \geq 1 - \varepsilon$ for all $n = 1, 2, \ldots$. Denote

$$\mathrm{Lip}_{c_0}(U) = \{f: U \to \mathbb{R}: |f(x) - f(y)| \leq c_0(x, y) \quad \forall x, y \in U\}. \tag{11.1.5}$$

Thus, for some finite m there are $f_1, f_2, \ldots, f_m \in \mathrm{Lip}_{c_0}(U)$ such that

$$\sup_{f \in \mathrm{Lip}_{c_0}(U)} \inf_{1 \leq k \leq m} \sup_{x \in K_\varepsilon} |f(x) - f_k(x)| < \varepsilon$$

consequently

$$\sup_{f \in \mathrm{Lip}_{c_0}(U)} \inf_{1 \leq k \leq m} \sup_{x \in K_\varepsilon^\varepsilon} |f(x) - f_k(x)| < 3\varepsilon \tag{11.1.6}$$

where $K_\varepsilon^\varepsilon$ means the ε-neighborhood of K_ε with respect to the metric c_0. Let $g(x) := \max(0, 1 - d(x, K_\varepsilon)/\varepsilon)$. Then, by (11.1.4) and $\bar{P}_n(K^\varepsilon) \geq \int g \, \mathrm{d}\bar{P}_n \geq \bar{P}_n(K) \geq 1 - \varepsilon$, we have

$$\mu_n(K^\varepsilon) \geq \int g \, \mathrm{d}\mu_n \geq \int g \, \mathrm{d}(\mu_n - \bar{P}_n) + 1 - \varepsilon \geq 1 - 2\varepsilon \text{ a.s.} \tag{11.1.7}$$

for n large enough. The inequalities (11.1.6) and (11.1.7) imply that

$$\sup\left\{\left|\int_U f\,\mathrm{d}(\bar{\mu}_n - \bar{P}_n)\right| : f \in \mathrm{Lip}_{c_0}(U)\right\} \leq 10\varepsilon \text{ a.s.} \tag{11.1.8}$$

for n large enough. Note that the left-hand side of (11.1.8) is equal to the minimal norm $\mathring{\mu}_{c_0}(\bar{\mu}_n, \bar{P}_n)$ and thus coincides with $\hat{\mu}_{c_0}(\bar{\mu}_n, \bar{P}_n)$ (see Theorem 6.1.1). QED

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

Theorem 11.1.5 (generalized Wellner theorem). Suppose that $s_1, s_2, \ldots$ is a sequence of operators in U and denote

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

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

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

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

$$\bar{\Theta}_n = (\Theta_1 + \cdots + Q_n)/n \to 0 \quad \text{a.s.} \quad n \to \infty \tag{11.1.9}$$

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

$$\lim_{M\to\infty} \sup_n \int_U c(x, a) I\{d(x, a) > M\}(\mu_n + \bar{P}_n)(\mathrm{d}x) = 0 \quad \text{a.s.} \tag{11.1.10}$$

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

Proof. From Wellner's theorem it follows that $\lim_n \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) = 0$ a.s.. We next estimate $\mathscr{A}_{c_0}(\mu_n, \bar{P}_n)$ obtaining

$$\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \leqslant \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) + (B_1 + \cdots + B_n)/n \tag{11.1.11}$$

where

$$B_i = \sup\left\{\left|\int_U [f(s_i x) - f(x)](\delta_{X_i} - P_i)(\mathrm{d}x)\right| : f \in \mathrm{Lip}_{c_0}(U)\right\}.$$

In fact, by the duality representation of $\mathscr{A}_{c_0}$ (see Corollary 6.1.1)

$$\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) = \sup\left\{\left|\frac{1}{n}\sum_{i=1}^n \int_U f(x)(\delta_{X_i} - P_i)(\mathrm{d}x)\right| : f \in \mathrm{Lip}_{c_0}(U)\right\}$$

and thus

$$
\begin{aligned}
&\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \\
&\quad \leqslant \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) + \sup_{f \in \mathrm{Lip}_{c_0}(U)} \left| \frac{1}{n} \int f(x)(\delta_{Y_i} - Q_i)(\mathrm{d}x) - \frac{1}{n} \int f(x)(\delta_{X_i} - P_i)(\mathrm{d}x) \right| \\
&\quad = \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) + \sup_{f \in \mathrm{Lip}_{c_0}(U)} \left| \frac{1}{n} \sum_{i=1}^{n} (f(s_i X_i) - \mathbb{E} f(s_i X_i) - f(X_i) + \mathbb{E} f(X_i)) \right| \\
&\quad = \mathscr{A}_{c_0}(\nu_n, \bar{Q}_n) + (B_1 + \cdots + B_n)/n.
\end{aligned}
$$

We estimate B_i as follows

$$
\begin{aligned}
B_i &\leqslant \sup_{f \in \mathrm{Lip}_{c_0}(U)} \int |f(s_i X_i) - f(x)| (\delta_{X_i} + P_i)(\mathrm{d}x) \\
&\leqslant \sup_{x \in U} \frac{d(s_i x, x)}{1 + d(s_i x, x)} \int (\delta_{X_i} + P_i)(\mathrm{d}x) \leqslant 2 \min(D_i, 1)
\end{aligned}
$$

and moreover, since for $g(x) := f(s_k x) - f(x)$, $f \in \mathrm{Lip}_{c_0}(U)$ we have

$$
\begin{aligned}
|g(x) - g(y)| &\leqslant d(s_i x, s_i y) + d(x, y) \leqslant (L_i + 1) d(x, y) \\
|g(x) - g(y)| &\leqslant 2 \frac{d(x, y)}{1 + d(x, y)} (L_i + 1) \text{ if } d(x, y) \leqslant 1 \\
\|g\|_\infty &:= \sup\{|g(x)| : x \in U\} \leqslant 2, \\
\tfrac{1}{4} |g(x) - g(y)| &\leqslant \tfrac{1}{4} \{|g(x)| + |g(y)|\} \\
&\leqslant 2 \frac{d(x, y)}{1 + d(x, y)} \quad \text{if } d(x, y) > 1
\end{aligned}
$$

and thus

$$
\begin{aligned}
B_i &\leqslant \sup\left\{ \left| \int_U g(x)(\delta_{X_i} - P_i)(\mathrm{d}x) \right| : g : U \to \mathbb{R}, |g(x) - g(y)| \leqslant 8(L_i + 1) c_0(x, y) \right\} \\
&\leqslant 8(L_i + 1) \mathscr{A}_{c_0}(\delta_{X_i}, P_i).
\end{aligned}
$$

Using the above estimates for B_i and the assumption (11.1.9), we obtain that $(B_1 + \cdots + B_n)/n \to 0$. According to (11.1.11)

$$
\mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \to 0 \quad \text{a.s. as } n \to \infty. \tag{11.1.12}
$$

If $\mathbf{K}$ is the Ky Fan metric (see Example 3.3.2) and μ_{c_0} is the probability metric

$$
\mu_{c_0}(P) := \int_{U \times U} c_0(x, y) P(\mathrm{d}x, \mathrm{d}y), \; P \in \mathscr{P}_2(U)
$$

then by Chebyshev's inequality we have

$$\frac{\mathbf{K}^2}{1+\mathbf{K}} \leqslant \mu_{c_0} \leqslant \mathbf{K} + \frac{\mathbf{K}}{1+\mathbf{K}}.$$

Passing to the minimal metrics in the last inequality and using the Strasen–Dudley theorem (see Corollary 7.4.2) we get

$$\frac{\pi^2}{1+\pi} \leqslant \mathscr{A}_{c_0} \leqslant \pi + \frac{\pi}{1+\pi} \tag{11.1.13}$$

where π is the Prokhorov metric in $\mathscr{P}(U)$. Applying (11.1.13) and (7.5.9) (see also Lemma 8.2.1) we have, for any positive M

$$\frac{\pi^2(\mu_n, \bar{P}_n)}{1+\pi(\mu_n, \bar{P}_n)} \leqslant \mathscr{A}_{c_0}(\mu_n, \bar{P}_n) \leqslant \pi(\mu_n, \bar{P}_n) + \frac{\pi(\mu_n, \bar{P}_n)}{1+\pi(\mu_n, \bar{P}_n)} \tag{11.1.14}$$

and

$$\mathscr{A}_c(\mu_n, \bar{P}_n) \leqslant H(\pi(\mu_n, \bar{P}_n)) + 2K_H\pi(\mu_n, \bar{P}_n)H(M) + K_H \int_U c(x, a)I\{d(x, a) > M\}(\mu_n + P_n)(\mathrm{d}x). \tag{11.1.15}$$

From (11.1.12), (11.1.14), (11.1.15) and (11.1.10), it follows that $\mathscr{A}_c(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$. QED

Corollary 11.1.1 If $c (c \in \mathfrak{C})$ is a bounded function and $\Theta_n \to 0$ a.s., then $\mathscr{A}_c(\mu_n, \bar{P}_n) \to 0$ a.s. as $n \to \infty$.

Corollary 11.1.1 is a consequence of the above theorem when $s_i(x) = x$, $x \in U$, and clearly is a generalization of the Varadarajan theorem 11.1.3. It is also clear that Theorem 11.1.3 implies Theorem 11.1.2. The following example shows that the conditions imposed in Corollary 11.1.1 are actually weaker as compared to the conditions of Wellner's theorem 11.1.4.

Example 11.1.1 Let $(U, \|\cdot\|)$ be a separable normed space. Let x_k be $k\bar{e}$ where $\bar{e}$ is the unit vector in U and let $X_k = x_k$ a.s.. Set $s_k(x) := x - x_k$, then $\bar{Q}_n = \delta_0$ and $\Theta_n = 0$ a.s.. Clearly, $\mathscr{A}_c(\mu_n, \bar{P}_n) = 0$ a.s. but $\bar{P}_n$ is not a tight sequence.

In the following we shall assume that $P_1 = P_2 = \cdots = = \mu$. In this case, the Glivenko–Cantelli theorem can be stated as follows in terms of $\mathscr{A}_c$ and the minimal metric $\ell_p = \mathscr{L}_p$ $(0 < p < \infty)$ (see the definitions (3.2.11), (3.2.12); the representations (3.3.18), (5.3.16); and Theorem 8.1.1).

Corollary 11.1.2 (Generalized Glivenko–Cantelli–Varadarajan theorem). Let

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

$$\int_U d^p(x, a)\mu(\mathrm{d}x) < \infty \quad 0 < p < \infty \tag{11.1.16}$$

then $\ell_p(\mu_n, \mu) \to 0$ a.s. as $n \to \infty$.

According to Theorem 6.2.1 and Corollary 7.5.3, the minimal norm $\mathring{\mu}_{c_p}$ with

$$c_p(x, y) = d(x, y)\max(1, d^{p-1}(x, a), d^{p-1}(y, a)) \quad p \geqslant 1$$

and the minimal metric ℓ_p metrizes one and the same convergence in the space $\mathscr{P}_p(U) = \{P \in \mathscr{P}(U): \int_U d^p(x, a)P(dx) < \infty\}$, namely

$$\ell_p(P_n, P) \to 0 \Leftrightarrow \mathring{\mu}_{c_p}(P_n, P) \to 0 \Leftrightarrow \begin{cases} P_n \xrightarrow{w} P \text{ and} \\ \displaystyle\int_U d^p(x, a)(P_n - P)(\mathrm{d}x) \to 0 \end{cases} \tag{11.1.17}$$

Thus Corollary 11.1.2 implies the following theorem stated by Fortet and Mourier (1953).

Corollary 11.1.3. If (11.1.16) holds, then

$$\mathring{\mu}_{c_p}(\mu_n, \mu) \to 0 \quad \text{a.s. as } n \to \infty.$$

Remark 11.1.1. One could generalize Corollaries 11.1.2 and 11.1.3 by means of Theorem 6.3.1, see also Ranga Rao (1962) for extensions of the original Fortet–Mourier result. We write $\mathscr{H}^*$ for the subset of all convex functions in $\mathscr{H}$ and $\mathfrak{C}^*$ for the set $\{H \circ d: H \in \mathscr{H}^*\}$. Theorem 8.1.2 gives an explicit representation of the functionals $\mathscr{A}_c$, $c \in \mathfrak{C}^*$, when $U = \mathbb{R}^1$ (see (8.1.38)). Corollary 11.1.2 may be formulated in this case as follows.

Corollary 11.1.4. Let $c \in \mathfrak{C}^*$, $U = \mathbb{R}^1$, and $d(x, y) = |x - y|$. Let $F_n(x)$ be the empirical distribution function corresponding to the distribution function $F(x)$ with $\int c(x, 0)\, \mathrm{d}F(x)$ finite. Then

$$\int_0^1 c(F_n^{-1}(x), F^{-1}(x))\, \mathrm{d}x \to 0 \quad \text{a.s.} \tag{11.1.18}$$

In particular, if

$$\int |x|^p\, \mathrm{d}F(x) < \infty \quad p \geqslant 1 \tag{11.1.19}$$

then

$$\ell_p^p(F_n, F) = \int_0^1 |F_n^{-1}(x) - F^{-1}(x)|^p\, \mathrm{d}x \to 0 \quad \text{a.s.} \tag{11.1.20}$$

and

$$\mathring{\mu}_{c_p}(F_n, F) = \int_{-\infty}^{\infty} \max(1, |x|^{p-1})|F_n(x) - F(x)|\,dx \to 0 \quad \text{a.s.} \tag{11.1.21}$$

Remark 11.1.2. Corollary 11.1.4 was proved for the case $p = 1$ by Fortet and Mourier (1953), and $p = 2$, $F(x)$ a continuous strictly increasing function by Samuel and Bachi (1964).

We study next the estimation of the convergence speed in the Glivenko–Cantelli theorem in terms of $\mathscr{A}_c$. Estimates of this sort are useful if one has to estimate not only the speed of convergence of the distribution μ_n to μ in weak metrics but also the speed of convergence of their moments. Thus, for example, if $\mathbb{E}\ell_p(\mu_n, \mu) = O(\phi(n))$, $n \to \infty$, for some $p \in (0, \infty)$, then Lemma 8.2.1 implies that $(\mathbb{E}(\pi(\mu_n, \mu))^{(p+1)/p'} = O(\phi(n))$, $n \to \infty$, where $p' = \max(1, p)$ (cf. (8.2.7)) and by Minkowski's inequality it follows that

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

for any point $a \in U$.

We shall estimate $\mathbb{E}\mathscr{A}_c(\mu_n, \mu)$ in terms of the ε-entropy of the measure μ as was suggested by Dudley in 1969. Let $N(\mu, \varepsilon, \delta)$ be the smallest number of sets of diameter at most 2ε whose union covers U except for a set A_0 with $\mu(A_0) \leqslant \delta$. Using Kolmogorov's definition of the ε-entropy of a set U, we call $\log N(\mu, \varepsilon, \varepsilon)$ the *ε-entropy of the measure* μ. The next theorem was proved by Dudley (1969) for $c = c_0$.

Theorem 11.1.6. (Dudley 1969). Let $c = H \circ d \in \mathfrak{C}$ and $H(t) = t^\alpha h(t)$, where $0 < \alpha < 1$ and $h(t)$ is a non-decreasing function on $[0, \infty)$. Let $\beta_r = \int_U c^r(x, a)\,\mu(dx) < \infty$ for some $r > 1$ and $a \in U$.

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

$$N(\mu, \varepsilon^{1/\alpha}, \varepsilon^{k/(k-2)}) \leqslant K\varepsilon^{-k} \tag{11.1.22}$$

then

$$\mathbb{E}\mathscr{A}_c(\mu_n, \mu) \leqslant Cn^{-(1-1/r)/k}$$

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

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

$$N(\mu, \varepsilon^{1/\alpha}, 1/2) \geqslant c_1\varepsilon^{-k} \tag{11.1.23}$$

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

$$\mathbb{E}\mathscr{A}_c(\mu_n, \mu) \geqslant c_2 n^{-1/k}. \tag{11.1.24}$$

The proof of Theorem 11.1.6 is based on Dudley (1969) and the inequality

$$\mathscr{A}_c(\mu, \nu) \leqslant 2H(N)\mathscr{A}_{c_\alpha}(\mu, \nu) + 2c_H \int c(x, a)I\{d(x, a) > N/2\}(\mu + \nu)(\mathrm{d}x) \quad (11.1.25)$$

where $c_\alpha = d^\alpha/(1 + d^\alpha)$, $N > 0$ and μ and ν are arbitrary measures on $\mathscr{P}(U)$.

The detailed proof is given in Kalashnikov and Rachev (1988) (Theorem 9.7, pp. 147–150), where the constant C is bounded from above by $\frac{4}{3}(\sqrt{K}3^{2k+1})$.

If $(U, d) = (\mathbb{R}^d, \|\cdot\|)$, $m_\gamma = \int \|x\|^\gamma \mu(\mathrm{d}x) < \infty$, where $\gamma = k\alpha d/[(k\alpha - d)(k - 2)]$, $k\alpha > d$, $k > 2$, then requirement (11.1.22) is satisfied. If $(U, d) = (\mathbb{R}^{k\alpha}, \|\cdot\|)$, where $k\alpha$ is an integer and μ is an absolutely continuous distribution, then condition (11.1.23) is satisfied. The estimate $\mathbb{E}\mathscr{A}_c(\mu_n, \mu) \leqslant cn^{-1/k}$ has exact exponent $(1/k)$ when $k\alpha$ is an integer, $U = \mathbb{R}^{k\alpha}$ and μ is an absolutely continuous distribution having uniformly bounded moments β_r, $r > 1$, and m_γ, $\gamma > 1$.

Open problem 11.1.1. What is the exact order of n as $\mathscr{A}_c(\mu_n, \mu) \to 0$ a.s.? For the case μ being uniform in $[0, 1]$ and

$$c(x, y) = c_0(x, y) = \frac{|x - y|}{1 + |x - y|}$$

it follows immediately from a result of Yukich (1989) that there exist constants c and C such that

$$\lim_{n\to\infty} \Pr\left\{c \leqslant \left(\frac{n}{\log n}\right)^{1/2} \mathscr{A}_{c_0}(\mu_n, \mu) \leqslant C\right\} = 1. \quad (11.1.26)$$

11.2 FUNCTIONAL CENTRAL LIMIT THEOREM AND BERNSTEIN–KANTOROVICH INVARIANCE PRINCIPLE

Let $\xi_{n1}, \xi_{n2}, \ldots, \xi_{nk_n}$, $n = 1, 2, \ldots$, be an array of independent r.v.s with d.f. F_{nk}, $k = 1, \ldots, k_n$, obeying the condition of limiting negligibility

$$\lim_n \max_{1 \leqslant k \leqslant k_n} \Pr(|\xi_{nk}| > \varepsilon) = 0 \quad (11.2.1)$$

and the conditions

$$\mathbb{E}\xi_{nk} = 0 \quad \mathbb{E}\xi_{nk}^2 = \sigma_{nk}^2 > 0 \quad \sum_{k=1}^{k_n} \sigma_{nk}^2 = 1. \quad (11.2.2)$$

Let $\zeta_{n0} = 0$ and $\zeta_{nk} = \xi_{n1} + \cdots + \xi_{nk}$, $1 \leqslant k \leqslant k_n$, and form a random polygonal line $\zeta_n(t)$ with vertices $(\mathbb{E}\zeta_{nk}^2, \zeta_{nk})$ (see Prokhorov 1956). Let P_n, from the space of laws on $\mathbb{C}[0, 1]$ with the supremum norm $\|x\| = \sup\{|x(t)|: t \in [0, 1]\}$), be the distribution of $\zeta_n(t)$ and let W be a Wiener measure in $\mathbb{C}[0, 1]$. On the basis of Theorem 8.2.1 we have the following $\mathscr{A}_c$-convergence criterion

$$\mathscr{A}_c(P_n, W) \to 0 \Leftrightarrow \begin{cases} P_n \xrightarrow{w} W \\ \displaystyle\int_{\mathbb{C}[0,1]} c(x, 0)(P_n - W)(\mathrm{d}x) \to 0 \end{cases} \tag{11.2.3}$$

for any $c \in \mathfrak{C} = \{c(x, y) = H(\|x - y\|),\ H \in \mathscr{H}, \text{ see (11.1.2)}\}$.

The limit relation (11.2.3) implies the following version of the classical Donsker–Prokhorov theorem (see, for example, Billingsley 1968, Theorem 10.1).

Theorem 11.2.1 (Bernstein–Kantorovich functional limit theorem). Suppose that conditions (11.2.1) and (11.2.2) hold and that $\mathbb{E}H(|\zeta_{nk}|) < \infty$, $k = 1, 2, \ldots, k_n$, $n = 1, 2, \ldots$, $H \in \mathscr{H}$. Then the convergence $\mathscr{A}_c(P_n, W) \to 0$, $n \to \infty$, is equivalent to the fulfilment of the Lindeberg condition

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x|>\varepsilon} x^2 \,\mathrm{d}F_{nk}(x) = 0 \quad \varepsilon > 0 \tag{11.2.4}$$

and the Bernstein condition

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sum_{k=1}^{k_n} \int_{|x|>N} H(|x|)\,\mathrm{d}F_{nk}(x) = 0. \tag{11.2.5}$$

Proof. By the well known Prokhorov (1956) theorem, the necessity of (11.2.4) is a straightforward consequence of $P_n \xrightarrow{w} W$. Let us prove the necessity of (11.2.5). Define the functional b: $\mathbb{C}[0, 1] \to \mathbb{R}$ by $b(x) = x(1)$. Since, for any $N > 2\sqrt{2}$

$$\int_N^\infty \Pr(\|\zeta_n\| > t)\,\mathrm{d}H(t) \leqslant 2\int_N^\infty \Pr(|\zeta_{n,k_n}| \geqslant t - \sqrt{2})\,\mathrm{d}H(t)$$

$$\leqslant 2\int_{N/2}^\infty \Pr(|\zeta_{n,k_n}| > t)\,\mathrm{d}H(2t) \leqslant 2\,K_H \int_{M(N)} \Pr(|\zeta_{n,k_n}| \geqslant t)\,\mathrm{d}H(t)$$

where $M(N)$ increases to infinity with $N \uparrow \infty$ (see, for example, Billingsley 1968, p. 69). From the last inequality, it follows that $\mathbb{E}H(\|\zeta_n\|) < \infty$ for all $n = 1, 2, \ldots$. By Theorem 11.1.1 and $\mathscr{A}_c(P_n, W) \to 0$ the relations $P_n \xrightarrow{w} W$ and

$$\int H(\|x\|)(P_n - W)(\mathrm{d}x) \to 0$$

hold as $n \to \infty$ and since, for any N

$$\mathbb{E}H(|b(\zeta_n)|)I\{|b(\zeta_n)| > N\} \leqslant \mathbb{E}H(\|\zeta_n\|)I\{\|\zeta_n\| > N\}$$
$$\leqslant 2\int_{M_1(N)}^{\infty} \Pr(\|\zeta_n\| > t)\,\mathrm{d}H(t)$$

where $M_1(N) \uparrow \infty$ together with $N \uparrow \infty$, we have (i) $P_n \circ b^{-1} \xrightarrow{w} W \circ b^{-1}$ and (ii) $\int h(\|x\|)(P_n \circ b^{-1} - W \circ b^{-1})(\mathrm{d}x) \to 0$ as $n \to \infty$.

By virtue of Kruglov's moment limit theorem (given (i), then (ii) is equivalent to (11.2.5), cf. Kruglov 1973, Theorem 1), the necessity of condition (11.2.5), has been proved. The sufficiency of (11.2.4) and (11.2.5) is proved in a similar way.

QED

Next we state a functional limit theorem which is based on the Bernstein CLT (cf. Bernstein 1964, p. 358). We formulate the result in terms of the minimal metric ℓ_p (3.2.11), (3.3.18) and (11.1.17).

Corollary 11.2.1. Let $\xi_1, \xi_2, \dots$ be a sequence of independent r.v.s such that $\mathbb{E}\xi_i^2 = b_i$ and $\mathbb{E}|\xi_i|^p < \infty$, $i = 1, 2, \dots, p > 2$. Let $B_n = b_1 + \cdots + b_n$, $\zeta_n = \xi_1 + \cdots + \xi_n$ and let the sequence $B_n^{-1/2}\xi_j$, $j = 1, 2, \dots$, satisfy the limiting negligibility condition. Let $X_n(t)$ be the random polygonal line with vertices $(B_k/B_n, B_n^{-1/2}\zeta_k)$ and let P_n be its distribution. Then the convergence

$$\ell_p(P_n, W) \to 0 \quad n \to \infty \tag{11.2.6}$$

is equivalent to the fulfilment of the condition

$$\lim_{n\to\infty} B_n^{-p/2} \sum_{i=1}^{n} \mathbb{E}|\xi|^p = 0. \tag{11.2.7}$$

Proof. The proof is analogous to that of Theorem 11.2.1. Here, conditions (11.2.4) and (11.2.5) are equivalent to (11.2.7) (cf. Bernstein 1969, Kruglov 1973, Acosta and Gine 1979). QED

Corollary 11.2.2 (Bernstein–Kantorovich invariance principle). Suppose that c, $c' \in \mathfrak{C}$, the array $\{\xi_{nk}\}$ satisfies the conditions of Theorem 11.2.1, and conditions (11.2.4) and (11.2.5) hold. Then $\mathscr{A}_{c'}(P_n \circ b^{-1}, W \circ b^{-1}) \to 0$ as $n \to \infty$ for any functional on $\mathbb{C}[0, 1]$ for which $N(b; c, c') = \sup\{c'(b(x), b(y))/c(x, y)\colon x \neq y, x, y \in \mathbb{C}[0, 1]\} < \infty$.

Proof. Observe that $\mathscr{A}_c(P_n, W) \to 0$ implies $\mathscr{A}_{c'}(P_n \circ b^{-1}, W \circ b^{-1}) \to 0$ as $n \to \infty$, provided $N(b; c, c') < \infty$. Now apply Theorem 11.2.1. QED

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

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

$$\|b\|_L = \sup\{|b(x) - b(y)|/\|x - y\| : x \neq y,\ x, y \in \mathbb{C}[0, 1]\} < \infty \quad (11.2.8)$$

then $N(b; c, c') < \infty$. Functionals such as these are $b_1(x) = x(a)$, $a \in [0, 1]$; $b_2(x) = \max\{x(t)\colon\ t \in [0, 1]\}$; $b_2(x) = \|x\|$ and $b_4(x) = \int_0^1 \phi(x(t))\,dt$, where $\|\phi\|_L := \sup\{|\phi(x) - \phi(y)|/|x - y| : x, y \in [0, 1]\} \leqslant 1$.

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

$$|\phi(x) - \phi(y)| \leqslant |x - y|^{p/p'} \quad x, y \in [0, 1]. \quad (11.2.9)$$

Further, as an example of Corollary 11.2.2 we shall consider the functional b_4 and the following moment limit theorem.

Corollary 11.2.3. Suppose $\xi_1, \xi_2, \ldots$ are independent random variables with $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = \sigma^2 > 0$ and

$$\lim_{n\to\infty} n^{-p/2} \sum_{j=1}^{n} \mathbb{E}|\xi_j|^p = 0 \quad \text{for some } p > 2. \quad (11.2.10)$$

Suppose also that $\phi\colon [0, 1] \to \mathbb{R}$ has a finite Lipschitz seminorm $\|\phi\|_L$. Then

$$\ell_p\left(\frac{1}{n}\sum_{k=1}^{n} \phi\left(\frac{\xi_1 + \cdots + \xi_k}{\sigma\sqrt{n}}\right), \int_0^1 \phi(w(t))\,dt\right) \to 0 \quad \text{as } n \to \infty \quad (11.2.11)$$

where the law of w is W.

Proof. Let $X_n(\cdot)$ be the random polygon line with vertices $(k/n, S_k/\sigma\sqrt{n})$ where $S_0 = 0$, $S_k = \xi_1 + \cdots + \xi_k$. From Corollaries 11.2.1 and 11.2.2, it follows that

$$\lim_{n\to\infty} \ell_p\left(\int_0^1 \phi(X_n(t))\,dt, \int_0^1 \phi(w(t))\,dt\right) = 0. \quad (11.2.12)$$

Readily (cf. Gikhman and Skorokhod 1971, p. 491, or p. 416 of the English edition), we have

$$\mathbf{K}\left(\left|\int_0^1 \phi(X_n(t))\,dt - \frac{1}{n}\sum_{k=1}^{n}\phi\left(\frac{S_k}{\sigma\sqrt{n}}\right)\right|, 0\right) \to 0 \quad (11.2.13)$$

$\mathbf{K}$ being the Ky Fan metric. By virtue of the maximal inequality (Billingsley 1968, p. 69),

$$\int_N^\infty \Pr\left\{\left|\int_0^1 \phi(X_n(t))\,dt - \frac{1}{n}\sum_{i=1}^{n}\phi\left(\frac{S_k}{\sigma\sqrt{n}}\right)\right| > u\right\} du^p$$

$$\leqslant \int_N^\infty \Pr\left\{\frac{|S_n|}{\sigma\sqrt{n}} > \frac{t}{2\|\phi\|_L} - \sqrt{2}\right\} dt^p. \quad (11.2.14)$$

Corollary 11.2.1 and (11.2.10) imply that the right-hand side of (11.2.14) goes to zero uniformly on n as $N \to \infty$. From (11.2.13) and (11.2.14), it follows that

$$\mathbb{E}\left|\int_0^1 \phi(X_n(t))\,dt - \frac{1}{n}\sum_{k=1}^n \phi\left(\frac{S_k}{\sigma\sqrt{n}}\right)\right|^p \to 0. \tag{11.2.15}$$

Finally, (11.2.15) and (11.2.13) imply

$$\ell_p\left(\int_0^1 \phi(X_n(t))\,dt, \frac{1}{n}\sum_{k=1}^n \phi\left(\frac{S_k}{\sigma\sqrt{n}}\right)\right) \to 0 \quad \text{as } n \to \infty$$

which, together with (11.2.12), completes the proof of (11.2.11). QED

We state one further consequence of Theorem 11.2.1. Let the series scheme $\{\zeta_{nk}\}$ satisfy the conditions of Theorem 11.2.1 and let $\eta_n(t) = \zeta_{nk}$ for $t \in (t_{n(k-1)}, t_{nk})$, $t_{nk} := \mathbb{E}\zeta_{nk}^2$, $k = 1, \ldots, k_n$, $\eta_n(0) = 0$. Let $\hat{P}_n$ be the distribution of η_n. The distribution $\hat{P}_n$ belongs to the space of probability measures defined on the Skorokhod space $D[0, 1]$ (Billingsley 1968, Chap. 3).

Corollary 11.2.4. The convergence $\mathscr{A}_c(\hat{P}_n, W) \to 0$ as $n \to \infty$ is equivalent to the fulfilment of (11.2.4) and (11.2.5).

CHAPTER 12

Stability of Queueing Systems

12.1 STABILITY OF $G|G|1|\infty$-SYSTEM

As a model example of the applicability of Kantorovich's theorem in the stability problem for queueing systems, we consider the stability of the system $G|G|1|\infty$. (A detailed discussion of this problem is presented in Kalashnikov and Rachev (1988).) The notation $G|G|1|\infty$ means that we consider single-server queue with 'input flow' $\{e_n\}_{n=0}^{\infty}$ and 'service flow' $\{s_n\}_{n=0}^{\infty}$ consisting of dependent nonidentically distributed components. Here, $\{e_n\}_{n=0}^{\infty}$ and $\{s_n\}_{n=0}^{\infty}$ are treated as sequences of the lengths of the time intervals between the nth and $(n+1)$th arrivals and the service times of the nth arrival, respectively. Define the recursive sequence

$$w_0 = 0 \quad w_{n+1} = \max(w_n + s_n - e_n, 0) \quad n = 1, 2, \dots. \tag{12.1.1}$$

The quantity w_n may be viewed as the waiting time for the nth arrival to begin to be serviced. Introduce the notation: $\mathbf{e}_{j,k} = (e_j, \dots, e_k)$, $\mathbf{s}_{j,k} = (s_j, \dots, s_k)$, $k > j$, $\mathbf{e} = (e_0, e_1, \dots)$, and $\mathbf{s} = (s_0, s_1, \dots)$. Along with the model defined by relations (12.1.1), we consider a sequence of analogous models by indexing it with the letter $r(r \geqslant 1)$. Namely, all quantities pertaining to the rth model will be designated in the same way as the model (12.1.1) but will have a superscript r: $e_n^{(r)}$, $s_n^{(r)}$, $w_n^{(r)}$, and so on. It is convenient to regard the value $r = \infty$ (which can be omitted) as corresponding to the original model. All of the random variables are assumed to be defined on the same probability space. For brevity, functionals Φ depending just on the distributions of the r.v.s X and Y will be denoted by $\Phi(X, Y)$.

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

$$\begin{aligned}&\phi_k(\xi_1, \dots, \xi_k, \eta_1, \dots, \eta_k)\\ &\quad := \max[0, \eta_k - \xi_k, (\eta_k - \xi_k) + (\eta_{k-1} - \xi_{k-1}), \dots, (\eta_k - \xi_k) + \cdots + (\eta_1 - \xi_1)].\end{aligned}$$

It is not hard to see that $\phi_k(\mathbf{e}_{n-k,n-1}, \mathbf{s}_{n-k,n-1})$ is the waiting time for the nth arrival under the condition that $w_{n-k} = 0$.

Let $c \in \mathfrak{C} = \{c(x, y) = H(d(x, y)), H \in \mathscr{H}$, see (11.1.2)$\}$. *The system $G|G|1|\infty$ is uniformly stable with respect to the functional $\mathscr{A}_c$ on finite time intervals if,* *for every positive T, the following limit relation holds: as $r \to \infty$*

$$\delta_{(r)}(T; \mathscr{A}_c) := \sup_{n \geqslant 0} \max_{1 \leqslant k \leqslant T} \mathscr{A}_c(\phi_k(\mathbf{e}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}^{(r)}_{n,n+k-1})) \to 0 \tag{12.1.2}$$

where $\mathscr{A}_c$ is the Kantorovich functional on $\mathfrak{X}(\mathbb{R}^k)$

$$\mathscr{A}_c(X, Y) = \inf\{\mathbb{E}c(\tilde{X}, \tilde{Y}): \tilde{X} \stackrel{d}{=} X, \tilde{Y} \stackrel{d}{=} Y\} \tag{12.1.3}$$

(cf. (11.1.1)).

Similarly, we define $\delta^{(r)}(T; \ell_p)$, where $\ell_p = \hat{\mathscr{L}}_p (0 < p < \infty)$ is the minimal metric w.r.t. $\mathscr{L}_p$-distance (see (3.2.11), (3.2.12), (3.3.18), (5.3.16), Theorem 6.1.1).

The relation (12.1.3) means that the largest deviation between the variables w_{n+k} and $w^{(r)}_{n+k}$, $k = 1, \ldots, T$, converges to zero as $r \to \infty$ if at time n both compared systems are free of 'customers', and for any positive T this convergence is uniform in n.

Theorem 12.1.1. If for each $r = 1, 2, \ldots, \infty$ the sequences $\mathbf{e}^{(r)}$ and $\mathbf{s}^{(r)}$ are independent, then

$$\delta^{(r)}_c(T; \mathscr{A}_c) \leqslant K_H \sup_{n \geqslant 0} \mathscr{A}_c(\mathbf{e}_{n,n+T-1}, \mathbf{e}^{(r)}_{n,n+T-1}) + K_H \sup_{n \geqslant 0} \mathscr{A}_c(\mathbf{s}_{n,n+T-1}, \mathbf{s}^{(r)}_{n,n+T-1}) \tag{12.1.4}$$

where K_H is given by (11.1.2). In particular

$$\delta^{(r)}_c(T; \ell_p) \leqslant \sup_{n \geqslant 0} \ell_p(\mathbf{e}_{n,n+T-1}, \mathbf{e}^{(r)}_{n,n+T-1}) + \sup_{n \geqslant 0} \ell_p(\mathbf{s}_{n,n+T-1}, \mathbf{s}^{(r)}_{n,n+T-1}). \tag{12.1.5}$$

Proof. We shall prove (12.1.4) only. The proof of (12.1.3) is carried out in a similar way. For any $1 \leqslant k \leqslant T$, we have the triangle inequality

$$\begin{aligned}
&\mathscr{L}_p(\phi_k(\mathbf{e}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}^{(r)}_{n,n+k-1}))\\
&\quad \leqslant \mathscr{L}_p(\phi_k(\mathbf{e}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}))\\
&\qquad + \mathscr{L}_p(\phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}^{(r)}_{n,n+k-1}))\\
&\quad \leqslant \mathscr{L}_p(\phi_k(\mathbf{e}_{n,n+T-1}, \mathbf{s}_{n,n+T-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+T-1}, \mathbf{s}_{n,n+T-1}))\\
&\qquad + \mathscr{L}_p(\phi_k(\mathbf{e}^{(r)}_{n,n+T-1}, \mathbf{s}_{n,n+T-1}, \phi_k(\mathbf{e}^{(r)}_{n,n+T-1}, \mathbf{s}^{(r)}_{n,n+T-1})).
\end{aligned}$$

Changing over to minimal metric ℓ_p and using the assumption that $\mathbf{e}^{(r)}$ and $\mathbf{s}^{(r)}$ are independent $(r = 1, \ldots, \infty)$

$$\begin{aligned}
&\inf\{\mathscr{L}_p(\phi_k(\mathbf{e}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}^{(r)}_{n,n+k-1})\}\\
&\quad \leqslant \ell_p(\mathbf{e}_{n,n+T-1}, \mathbf{e}^{(r)}_{n,n+T-1}) + \ell_p(\mathbf{s}_{n,n+T-1}, \mathbf{s}^{(r)}_{n,n+T-1}).
\end{aligned} \tag{12.1.6}$$

The infimum in the last inequality is taken over all joint distributions

$$F(x, y, \xi, \eta) = \Pr(\mathbf{e}_{n,n+k-1} < x, \mathbf{e}^{(r)}_{n,n+k-1} < y)\Pr(\mathbf{s}_{n,n+k-1} < \xi, \mathbf{s}^{(r)}_{n,n+k-1} < \eta) \quad x, y, \xi, \eta \in \mathbb{R}^k$$

with fixed marginal distributions

$$F_1(x, \xi) = \Pr(\mathbf{e}_{n,n+k-1} < x, \mathbf{s}_{n,n+k-1} < \xi)$$

$$F_2(y, \eta) = \Pr(\mathbf{e}^{(r)}_{n,n+k-1} < y, \mathbf{s}^{(r)}_{n,n+k-1} < \eta)$$

and thus the left-hand side of (12.1.5) is not greater than

$$\ell_p(\phi_k(\mathbf{e}_{n,n+k-1}, \mathbf{s}_{n,n+k-1}), \phi_k(\mathbf{e}^{(r)}_{n,n+k-1}, \mathbf{s}^{(r)}_{n,n+k-1}))$$

which proves (12.1.4). QED

From (12.1.3) and (12.1.4) it is possible to derive an estimate of the stability of the system $G|G|1|\infty$ in the sense of (12.1.2). It can be expressed in terms of the deviations of the vectors $\mathbf{e}^{(r)}_{n,n+T-1}$ and $\mathbf{s}^{(r)}_{n,n,+T-1}$ from $\mathbf{e}_{n,n+T-1}$ and $\mathbf{s}_{n,n,+T-1}$, respectively. Such deviations are easy to estimate if we impose additional restrictions on $\mathbf{e}^{(r)}$ and $\mathbf{s}^{(r)}$, $r = 1, 2, \ldots$. For example, when the terms of the sequences are independent, the following estimates hold

$$\mathscr{A}_c(\mathbf{e}_{n,n+T-1}, \mathbf{e}^{(r)}_{n,n+T-1}) \leqslant K_H^q \sum_{j=n}^{n+T-1} \mathscr{A}_c(e_j, e_j^{(r)}) \quad q = [\log_2 T] + 1 \tag{12.1.7}$$

$$\ell_p(\mathbf{e}_{n,n+T-1}, \mathbf{e}^{(r)}_{n,n+T-1}) \leqslant \sum_{j=n}^{n+T-1} \ell_p(e_j, e_j^{(r)}) \quad \text{for } 0 \leqslant p \leqslant \infty. \tag{12.1.8}$$

Let us check (12.1.8). (Similarly, one gets (12.1.7).) By the minimality of ℓ_p, for any vectors $X = (X_1, \ldots, X_T)$, $Y = (Y_1, \ldots, Y_T) \in \mathfrak{X}(\mathbb{R}^T)$ with independent components, we have that the Minkovski inequality

$$\mathscr{L}_p(X, Y) = [\mathbb{E}\|X - Y\|^p]^{1/p'} \leqslant \sum_{i=1}^{T} \mathscr{L}_p(X_i, Y_i) \quad p' = \max(1, p) \tag{12.1.9}$$

implies

$$\ell_p(X, Y) \leqslant \sum_{i=1}^{T} \ell_p(X_i, Y_i) \tag{12.1.10}$$

i.e., (12.1.8) holds.

The estimates (12.1.7) and (12.1.8) can be even further simplified when the terms of these sequences are identically distributed. On the basis of (12.1.3) and (12.1.4), it is possible to construct stability estimates for the system that are uniform over the entire time axis (see Kalashnikov and Rachev 1989, Chap. 5).

12.2 STABILITY OF $GI|GI|1|\infty$-SYSTEM

The system $GI|GI|1|\infty$ is a special case of $G|G|1|\infty$. For this model the random variables $\zeta_n = s_n - e_n$ are i.i.d. and we assume that $\mathbb{E}\zeta_1 < 0$. Then the one-dimensional stationary distribution of the waiting time coincides with the distribution of the following maximum

$$w = \sup_{k \geq 0} Y_k \quad Y_k = \sum_{j=-k}^{-1} \zeta_j \quad Y_0 = 0 \quad \zeta_{-j} \overset{d}{=} \zeta_j. \tag{12.2.1}$$

The perturbed model (i.e., $e_k^{(r)}, s_k^{(r)}, Y_k^{(r)}$) is assumed to be also of type $GI|GI|1|\infty$. For reference to these problems we refer to Gnedenko (1970), Kennedy (1972), Iglehart (1973), Whitt (1974 a,b), Borovkov (1984), Chap. IV, and Kalashnikov and Rachev (1988). Borovkov (1984), p. 239, noticed that one of the aims of the stability theorems is to estimate the closeness of $\mathbb{E}f^{(r)}(W^{(r)})$ and $\mathbb{E}f(W)$ for various kind of functions $f, f^{(r)}$. On pp. 239, 240 he proposed to consider the case

$$f^{(r)}(x) - f(y) \leqslant A|x-y| \quad \forall x, y \in \mathbb{R}. \tag{12.2.2}$$

He proved on p. 270 that

$$D = \sup\{|\mathbb{E}f(w^{(r)}) - \mathbb{E}f(w)| : |f(x) - f(y)| \leqslant A|x-y|, x, y \in \mathbb{R}\} \leqslant c\varepsilon \tag{12.2.3}$$

assuming that $|\zeta_1^{(r)} - \zeta_1| \leqslant \varepsilon$ a.s. Here and in the following c stands for an absolute constant which may be different in different places.

By (3.2.12), (3.3.18), (5.3.16) and Theorem 6.1.1, we have for the minimal metric $\ell_1 = \hat{\mathscr{L}}_1$

$$A\ell_1(w^{(r)}, w) = \sup\{\mathbb{E}f^{(r)}(w^{(r)}) - \mathbb{E}f(w); (f^{(r)}, f) \text{ satisfy } (12.2.2)\} = D \tag{12.2.4}$$

provided that $\mathbb{E}|w^{(r)}| + \mathbb{E}|w| < \infty$. So the estimate in (12.2.3) essentially says that

$$\ell_1(w^{(r)}, w) \leqslant c\ell_\infty(\zeta_1^{(r)}, \zeta_1) \tag{12.2.5}$$

where for any $X, Y \in \mathfrak{X}(\mathbb{R})$,

$$\ell_\infty(X, Y) = \hat{\mathscr{L}}_\infty(X, Y) = \sup_{0 \leqslant t \leqslant 1} |F_X^{-1}(t) - F_Y^{-1}(t)| \tag{12.2.6}$$

see (2.3.4), (3.2.14), (3.3.18), (7.4.15), Corollary 7.3.2. Actually, using (7.3.18) with $H(t) = t^p$, we have

$$\ell_p(X, Y) = \hat{\mathscr{L}}_p(X, Y) = \left(\int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)^p \, dt\right)^{1/p} \tag{12.2.7}$$

where F_X^{-1} is the generalized inverse of the d.f. F_X

$$F_X^{-1}(t) := \sup\{x : F_X(x) \leqslant t\}. \tag{12.2.8}$$

Letting $p \to \infty$ we obtain (12.2.6).

The estimate in (12.2.5) needs strong assumptions on the disturbances in order to conclude stability. We shall refine the bound (12.2.5) considering bounds of

$$A\ell_p^p(w^{(r)}, w) = \sup\{\mathbb{E}f^{(r)}(w^{(r)}) - \mathbb{E}f(w); f^{(r)}(x) - f(y) \leqslant A|x-y|^p \quad \forall x, y \in \mathbb{R}^1\} \quad 0 < p < \infty \tag{12.2.9}$$

assuming that $\mathbb{E}|w^{(r)}|^p + \mathbb{E}|w|^p < \infty$. The next lemma considers the closeness of the prestationary distributions of $w_n = \max(0, w_{n-1} + \zeta_{n-1})$, $w_0 = 0$ and of $w_n^{(r)}$, see (12.1.1).

Lemma 12.2.1. For any $0 < p < \infty$ and $\mathbb{E}\zeta_1 = \mathbb{E}\zeta_1^{(r)}$ the following holds

$$\ell_p(w_n^{(r)}, w_n) \leqslant A_p \tag{12.2.10}$$

$$A_p := \min\left(\frac{n(n+1)}{2}\varepsilon_p, c \min_{1/p - 1 < \delta < 2/p - 1} n^{1/(1+\delta)}\varepsilon_{p(1+\delta)}^p\right) \quad \text{for } p \in (0, 1]$$

where

$$A_p := cn^{1/p}\varepsilon_p \quad \text{for } 1 < p \leqslant 2$$

$$A_p := cn^{1/2}\varepsilon_p \quad \text{for } p > 2$$

and

$$\varepsilon_p := \ell_p(\zeta_1, \zeta_1^{(r)}).$$

Remark 12.2.1. The condition $\mathbb{E}\zeta_1 = \mathbb{E}\zeta_1^{(r)}$ means that we know the mean of $\zeta_1^{(r)}$ for the perturbed 'real' model and we chose an 'ideal' model with ζ_1 having the same mean.

Proof. The distributions of the waiting times w_n and w_n^* can be determined as follows

$$w_n = \max(0, \zeta_{n-1}, \zeta_{n-1} + \zeta_{n-2}, \ldots, \zeta_{n-1} + \cdots + \zeta_1) \stackrel{d}{=} \max_{0 \leqslant j \leqslant n} Y_j,$$

$$w_n^{(r)} = \max(0, \zeta_{n-1}^{(r)}, \zeta_{n-1}^{(r)} + \zeta_{n-2}^{(r)}, \ldots, \zeta_{n-1}^{(r)} + \cdots + \zeta_1^{(r)}) \stackrel{d}{=} \max_{0 \leqslant j \leqslant n} Y_j^{(r)}.$$

Further (see (18.3.41), Theorem 18.3.6), we shall prove the following estimates of the closeness between $w_n^{(r)}$ and w_n

$$\ell_p(w_n^{(r)}, w_n) \leqslant \frac{n(n+1)}{2}\ell_p(\zeta_1, \zeta_1^{(r)}) \quad \text{if } 0 < p \leqslant 1 \tag{12.2.11}$$

and

$$\ell_p(w_n^{(r)}, w_n) \leqslant \frac{p}{p-1}B_p n^{1/p}\varepsilon_p \quad \text{if } 1 < p \leqslant 2 \tag{12.2.12}$$

where $B_1 = 1$, $B_p = 18p^{3/2}/(p-1)^{1/2}$ for $1 < p \leqslant 2$. From (12.2.12) and $\ell_p \leqslant \ell_{p(1+\delta)}$ for any $0 < p < 1$, and $(1/p) - 1 < \delta \leqslant 2/p - 1$ we have $1 \leqslant p(1+\delta) \leqslant 2$ and

$$\ell_p(w_n^{(r)}, w_n) \leqslant cn^{1/(1+\delta)}\varepsilon_{p(1+\delta)}^p. \tag{12.2.13}$$

For $p \geqslant 2$ we have

$$\mathscr{L}_p^p(w_n, w_n^{(r)}) = \mathbb{E}\left|\bigvee_{k=1}^n Y_k - \bigvee_{k=1}^n Y_k^{(r)}\right|^p$$

$$\leqslant \frac{p}{p-1}\mathbb{E}|Y_n - Y_n^{(r)}|^p \leqslant cn^{p/2}\mathscr{L}_p(\zeta_1, \zeta_1^{(r)})^p. \tag{12.2.14}$$

This last inequality is a consequence of the Marcinkiewicz–Zygmund inequality (cf. Chow and Teicher 1978, p. 357). Passing to the minimal metrics $\ell_p = \hat{\mathscr{L}}_p$ in (12.2.14) we get (12.2.10). QED

Remark 12.2.2. (a) The estimates in (12.2.10) are of the right order as can be seen by examples. If, for example, $p \geqslant 2$, consider $\zeta_i \overset{d}{=} N(0, 1)$ and $\zeta_i^{(r)} = 0$, then $\ell_p(w_n^{(r)}, w_n) = cn^{1/2}$.

(b) If $p = \infty$, then $\ell_\infty(w_n^{(r)}, w_n) \leqslant n\varepsilon_\infty$.

Define the stopping times

$$\theta = \inf\left\{k\colon w_k = \max_{0 \leqslant j \leqslant k} Y_j = w = \sup_{j \geqslant 0} Y_j\right\} \tag{12.2.15}$$

$$\theta^{(r)} = \inf\{k\colon w_k^{(r)} = w^{(r)}\}.$$

From Lemma 12.2.1 we now obtain estimates for $\ell_p(w^{(r)}, w)$ in terms of the distributions of $\theta, \theta^{(r)}$. Define $G(n) := \Pr(\max(\theta^{(r)}, \theta) = n) \leqslant \Pr(\theta^{(r)} = n) + \Pr(\theta = n)$.

Theorem 12.2.1. If $1 < p \leqslant 2$, λ, $\mu \geqslant 1$ with $(1/\lambda) + (1/\mu) = 1$ and $\mathbb{E}\zeta_1 = \mathbb{E}\zeta_1^{(r)} < 0$, then

$$\ell_p^p(w^{(r)}, w) \leqslant c\varepsilon_{p\lambda}\sum_{n=1}^\infty n^{1/\lambda}G(n)^{1/\mu}. \tag{12.2.16}$$

Proof.

$$\mathscr{L}_p^p(w^{(r)}, w) = \mathbb{E}|w^{(r)} - w|^p = \sum_{n=0}^\infty \mathbb{E}|w^{(r)} - w|^p I\{\max(\theta^{(r)}, \theta) = n\}$$

$$= \sum_{n=0}^\infty \mathbb{E}|w_n - w_n^{(r)}|^p I\{\max(\theta^{(r)}, \theta) = n\} \leqslant \sum_{n=0}^\infty (\mathbb{E}|w_n - w_n^{(r)}|^{p\lambda})^{1/\lambda}G(n)^{1/\mu}$$

and thus by (12.2.10)

$$\ell_p^p(w^{(r)}, w) \leqslant \sum_{n=0}^{\infty} A_{p\lambda}^p G(n)^{1/\mu} = \sum_{n=0}^{\infty} cn^{1/\lambda}\varepsilon_{p\lambda} G(n)^{1/\mu}. \qquad \text{QED}$$

Remark 12.2.3. (a) if

$$G(n) \leqslant cn^{-\mu(1/\lambda+1+\varepsilon)} \tag{12.2.17}$$

for some $\varepsilon > 0$, then $\sum_{n=1}^{\infty} n^{1/\lambda} G(n)^{1/\mu} \leqslant c\sum_{n=1}^{\infty} n^{-1/(1+\varepsilon)} \leqslant \infty$. For conditions on ζ_1, ζ_1^* ensuring (12.2.17), compare Borovkov (1984), pp. 229, 230, 240.

(b) For $0 < p \leqslant 1$ and $p > 2$, in the same way we get from Lemma 12.2.1 corresponding estimates for $\ell_p(w^{(r)}, w)$.

(c) Note that $\ell_1(w^{(r)}, w) \leqslant \ell_p(w^{(r)}, w)$, i.e., ℓ_p-metric represents more functions w.r.t. the deviation (see the side conditions in (12.2.9)) than ℓ_1. Moreover, $\varepsilon_{p\lambda} = \ell_{p\lambda}(\zeta_1^{(r)}, \zeta_1) \leqslant \ell_\infty(\zeta_1^{(r)}, \zeta_1)$. Therefore, Theorem 12.2.1 is a refinement of the estimates given by Borovkov (1984), p. 270.

12.3 APPROXIMATION OF A RANDOM QUEUE BY MEANS OF DETERMINISTIC QUEUEING MODELS

The conceptually simplest class of queueing models are those of the deterministic type. Such models are usually explored under the assumption that the underlying (real) queueing system is close (in some sense) to a deterministic system. It is common practice to change the random variables governing the queueing model with constants in the neighborhood of their mean values. In this section we evaluate the possible error involved by approximating the random queueing model with the deterministic one. In order to get precise estimates we explore relationships between distances in the space of random sequences, precise moment inequalities and the Kemperman (1968, 1987) geometric approach to a certain trigonometric moment problem.

More precisely, as in Section 12.1, we consider a single-channel queueing system $G|G|1|\infty$ with sequences $\mathbf{e} = (e_0, e_1, \dots)$ and $\mathbf{s} = (s_0, s_1, \dots)$ of inter-arrival times and service times respectively, assuming that $\{e_j\}_{j\geqslant 1}$ and $\{s_j\}_{j\geqslant 1}$ are *dependent and nonidentically distributed r.v.s.* We denote by $\boldsymbol{\zeta} = (f_0, \zeta_1, \dots)$ the difference $\mathbf{s} - \mathbf{e}$ and let $\mathbf{w} = (w_0, w_1, \dots)$ be the sequence of waiting times, determined by (12.1.1).

Along with the queueing model $G|G|1|\infty$ defined by the input random characteristics $\mathbf{e}$, $\mathbf{s}$, ζ and the output characteristic $\mathbf{w}$, we consider an approximating model with corresponding inputs $\mathbf{e}^*$, $\mathbf{s}^*$, $\boldsymbol{\zeta}^*$ and output $\mathbf{w}^*$,

$$w_0^* = 0 \quad w_{n+1}^* = (w_n^* + s_n^* - e_n^*)_+ \quad n = 1, 2, \dots \tag{12.3.1}$$

where $(\cdot)_+ = \max(0, \cdot)$. The latter model has a simpler structure; namely, we assume that $\mathbf{e}^*$ and/or $\mathbf{s}^*$ are deterministic. We also assume that estimates of the deviations between certain moments of e_j and e_j^* (resp. s_j and s_j^* or ζ_j and ζ_j^*) are given.

We shall consider two types of approximating models

(a) $D|G|1|\infty$ (i.e., e_j^* are constants and in general, $e_j^* \neq e_i^*$ for $i \neq j$) and
(b) $D|D|1|\infty$ (i.e., e_j^* and s_j^* are constants).

The next theorem provides a bound for the deviation between the sequences $\mathbf{w} = (w_0, w_1, \ldots)$ and $\mathbf{w}^* = (w_0^*, w_1^*, \ldots)$ in terms of the Prokhorov metric π (see Example 3.2.3 and (3.2.18)). We denote by $U = \mathbb{R}^\infty$ the space of all sequences with metric

$$d(\bar{x}, \bar{y}) = \sum_{i=0}^{\infty} 2^{-i}|x_i - y_i| \quad (\bar{x} := (x_0, x_1, \ldots), \bar{y} := (y_0, y_1, \ldots))$$

which may take infinite values. Let $\mathfrak{X}^\infty = \mathfrak{X}(\mathbb{R}^\infty)$ be the space of all random sequences defined on a 'rich enough' probability space $(\Omega, \mathscr{A}, \Pr)$, see Remark 2.5.2. Then the Prokhorov metric in $\mathfrak{X}^\infty$ is given by

$$\pi(X, Y) := \inf\{\varepsilon > 0 \colon \Pr(X \in A) \leqslant \Pr(Y \in A^\varepsilon) + \varepsilon, \quad \forall \text{ Borel sets } A \subset \mathbb{R}^\infty\} \tag{12.3.2}$$

where A^ε is the open ε-neighborhood of A. Recall the Strassen–Dudley theorem (see Corollary 7.4.2)

$$\pi(X, Y) = \hat{\mathbf{K}}(X, Y) := \inf\{\mathbf{K}(\bar{X}, \bar{Y}) \colon \bar{X}, \bar{Y} \in \mathfrak{X}^\infty, \bar{X} \stackrel{d}{=} X, \bar{Y} \stackrel{d}{=} Y\} \tag{12.3.3}$$

where $\mathbf{K}$ is the Ky Fan metric

$$\mathbf{K}(X, Y) := \inf\{\varepsilon > 0 \colon \Pr(d(X, Y) > \varepsilon) < \varepsilon\} \quad X, Y \in \mathfrak{X}^\infty \tag{12.3.4}$$

see Example 3.3.2.

In stability problems for characterizations of ε-independence the following concept is frequently used (see Kalashnikov and Rachev 1988, Chapter 4). Let $\varepsilon > 0$ and $X = (X_0, X_1, \ldots) \in \mathfrak{X}^\infty$; the components of X are said to be ε-*independent* if

$$\mathrm{IND}(X) = \pi(X, \underline{X}) \leqslant \varepsilon$$

where the components $\underline{X}_i$ of $\underline{X}$ are independent and $\underline{X}_i \stackrel{d}{=} X_i (i \geqslant 0)$. The Strassen–Dudley theorem gives upper bounds for $\mathrm{IND}(X)$ in terms of the Ky Fan metric $\mathbf{K}(X, \underline{X})$.

Lemma 12.3.1. Let the approximating model be of the type $D|G|1|\infty$. Assume that the sequences $\mathbf{e}$ and $\mathbf{s}$ of the queueing model $G|G|1|\infty$ are independent. Then

$$\pi(\mathbf{w}, \mathbf{w}^*) \leqslant \mathrm{IND}(\mathbf{s}) + \mathrm{IND}(\mathbf{s}^*) + \sum_{j=1}^{\infty} (\pi(e_j, e_j^*) + \pi(s_j, s_j^*)). \tag{12.3.5}$$

Proof. By (12.1.1) and (12.3.1),

$$
\begin{aligned}
d(\mathbf{w}, \mathbf{w}^*) &= \sum_{n=1}^{\infty} 2^{-n} |w_n - w_n^*| \\
&= \sum_{n=1}^{\infty} 2^{-n} |\max(0, s_{n-1} - e_{n-1}, \ldots, (s_{n-1} - e_{n-1}) + \cdots + (s_0 - e_0)) \\
&\quad -\max(0, s_{n-1}^* - e_{n-1}^*, \ldots, (s_{n-1}^* - e_{n-1}^*) + \cdots + (s_0^* - e_0^*))| \\
&\leqslant \sum_{n=1}^{\infty} 2^{-n} |\max(0, s_{n-1} - e_{n-1}, \ldots, (s_{n-1} - e_{n-1}) + \cdots + (s_0 - e_0)) \\
&\quad -\max(0, s_{n-1} - e_{n-1}^*, \ldots, (s_{n-1} - e_{n-1}^*) + \cdots + (s_0 - e_0^*))| \\
&\quad + \sum_{n=1}^{\infty} 2^{-n} |\max(0, s_{n-1} - e_{n-1}^*, \ldots, (s_{n-1} - e_{n-1}^*) + \cdots + (s_0 - e_0^*) \\
&\quad -\max(0, s_{n-1}^* - e_{n-1}^*, \ldots, (s_{n-1}^* - e_{n-1}^*) + \cdots + (s_0^* - e_0^*))| \\
&\leqslant \sum_{n=1}^{\infty} 2^{-n} \max(|e_{n-1} - e_{n-1}^*|, \ldots, |e_{n-1} - e_{n-1}^*| + \cdots + |e_0 - e_0^*|) \\
&\quad + \sum_{n=1}^{\infty} 2^{-n} \max(|s_{n-1} - s_{n-1}^*|, \ldots, |s_{n-1} - s_{n-1}^*| + \cdots + |s_0 - s_0^*|) \\
&\leqslant \sum_{n=1}^{\infty} 2^{-n} \sum_{j=0}^{n-1} (|e_j - e_j^*| + |s_j - s_j^*|) \\
&\leqslant d(\mathbf{e}, \mathbf{e}^*) + d(\mathbf{s}, \mathbf{s}^*).
\end{aligned}
$$

Hence, by the definition of the Ky Fan metric (12.3.4), we obtain $\mathbf{K}(\mathbf{w}, \mathbf{w}^*) \leqslant \mathbf{K}(\mathbf{e}, \mathbf{e}^*) + \mathbf{K}(\mathbf{s}, \mathbf{s}^*)$. Next, using the representation (12.3.3) let us choose independent pairs $(\mathbf{e}_\varepsilon, \mathbf{e}_\varepsilon^*)$, $(\mathbf{s}_\varepsilon, \mathbf{s}_\varepsilon^*)$ $(\varepsilon > 0)$ such that $\pi(\mathbf{e}, \mathbf{e}^*) > \mathbf{K}(\mathbf{e}_\varepsilon, \mathbf{e}_\varepsilon^*) - \varepsilon$, $\pi(\mathbf{s}, \mathbf{s}^*) > \mathbf{K}(\mathbf{s}_\varepsilon, \mathbf{s}_\varepsilon^*) - \varepsilon$, and $\mathbf{e} \overset{d}{=} \mathbf{e}_\varepsilon$, $\mathbf{e}^* \overset{d}{=} \mathbf{e}_\varepsilon^*$, $\mathbf{s} \overset{d}{=} \mathbf{s}_\varepsilon$, $\mathbf{s}^* \overset{d}{=} \mathbf{s}_\varepsilon^*$. Then by the independence of $\mathbf{e}$ and $\mathbf{s}$ (resp. $\mathbf{e}^*$ and $\mathbf{s}^*$) we have

$$
\begin{aligned}
\pi(\mathbf{w}, \mathbf{w}^*) &= \inf\{\mathbf{K}(\mathbf{w}_0, \mathbf{w}_0^*) : \mathbf{w}_0 \overset{d}{=} \mathbf{w}, \mathbf{w}_0^* \overset{d}{=} \mathbf{w}^*\} \\
&\leqslant \inf\{\mathbf{K}(\mathbf{e}_0, \mathbf{e}_0^*) + \mathbf{K}(\mathbf{s}_0, \mathbf{s}_0^*) : (\mathbf{e}_0, \mathbf{s}_0) \overset{d}{=} (\mathbf{e}, \mathbf{s}), (\mathbf{e}_0^*, \mathbf{s}_0^*) \overset{d}{=} (\mathbf{e}^*, \mathbf{s}^*)\} \\
&\leqslant \inf\{\mathbf{K}(\mathbf{e}_0, \mathbf{e}_0^*) + \mathbf{K}(\mathbf{s}, \mathbf{s}_0^*) : \mathbf{e}_0 \overset{d}{=} \mathbf{e}, \mathbf{s}_0 \overset{d}{=} \mathbf{s}, \mathbf{e}_0^* \overset{d}{=} \mathbf{e}^*, \mathbf{s}_0^* \overset{d}{=} \mathbf{s}^*, \\
&\qquad \mathbf{e}_0 \text{ is independent of } \mathbf{s}_0, \mathbf{e} \text{ is independent of } \mathbf{s}, \\
&\qquad \mathbf{e}_0^* \text{ is independent of } \mathbf{s}_0^*, \mathbf{e}^* \text{ is independent of } \mathbf{s}^*\} \\
&\leqslant \mathbf{K}(\mathbf{e}_\varepsilon, \mathbf{e}_\varepsilon^*) + \mathbf{K}(\mathbf{s}_\varepsilon, \mathbf{s}_\varepsilon^*) \leqslant \pi(\mathbf{e}, \mathbf{e}^*) + \pi(\mathbf{s}, \mathbf{s}^*) + 2\varepsilon
\end{aligned}
$$

which proves that

$$\pi(\mathbf{w}, \mathbf{w}^*) \leqslant \pi(\mathbf{e}, \mathbf{e}^*) + \pi(\mathbf{s}, \mathbf{s}^*). \tag{12.3.6}$$

Next let us estimate $\pi(\mathbf{e}, \mathbf{e}^*))$ in the above inequality. Observe that

$$\mathbf{K}(X, Y) \leqslant \sum_{i=0}^{\infty} \mathbf{K}(X_i, Y_i) \tag{12.3.7}$$

for any $X, Y \in \mathfrak{X}^\infty$. In fact, if $\mathbf{K}(X_i, Y_i) \leqslant \varepsilon_i$ and $1 > \varepsilon = \sum_{i=0}^{\infty} \varepsilon_i$ then

$$\varepsilon > \sum_{i=0}^{\infty} \Pr(|X_i - Y_i| > \varepsilon_i) \geqslant \sum_{i=0}^{\infty} \Pr(2^{-i}|X_i - Y_i| > \varepsilon_i)$$
$$\geq \Pr\left(\sum_{i=0}^{\infty} 2^{-i}|X_i - Y_i| > \varepsilon\right).$$

Letting $\varepsilon_i \to \mathbf{K}(X_i, Y_i)$ we obtain (12.3.7). By (12.3.7) and $\pi(\mathbf{e}, \mathbf{e}^*) = \mathbf{K}(\mathbf{e}, \mathbf{e}^*)$ we have

$$\pi(\mathbf{e}, \mathbf{e}^*) \leqslant \sum_{i=0}^{\infty} \mathbf{K}(e_i, e_i^*) = \sum_{i=0}^{\infty} \pi(e_i, e_i^*). \tag{12.3.8}$$

Next we shall estimate $\pi(\mathbf{s}, \mathbf{s}^*)$ on the right-hand side of (12.3.6). By the triangle inequality for the metric π we have

$$\pi(\mathbf{s}, \mathbf{s}^*) \leqslant \mathrm{IND}(\mathbf{s}) + \mathrm{IND}(\mathbf{s}^*) + \pi(\underline{\mathbf{s}}, \underline{\mathbf{s}}^*), \tag{12.3.9}$$

where the sequence $\underline{\mathbf{s}}$ (resp. $\underline{\mathbf{s}}^*$) in the last inequality consists of independent components such that $\underline{s}_j \stackrel{\mathrm{d}}{=} s_j$ (resp. $\underline{s}_j^* \stackrel{\mathrm{d}}{=} s_j^*$). We now need the 'regularity' property of the Prokhorov metric

$$\pi(X + Z, Y + Z) \leqslant \pi(X, Y) \tag{12.3.10}$$

for any Z independent of X and Y in $\mathfrak{X}^\infty$. In fact, (12.3.10) follows from the Strassen–Dudley theorem (12.3.3) and the corresponding inequality for the Ky Fan metric

$$\mathbf{K}(X + Z, Y + Z) \leqslant \mathbf{K}(X, Y) \tag{12.3.11}$$

for all X, Y and Z in $\mathfrak{X}^\infty$. By the triangle inequality and (12.3.10) we have

$$\pi\left(\sum_{i=0}^{\infty} X_i, \sum_{i=0}^{\infty} Y_i\right) \leqslant \sum_{i=0}^{\infty} \pi(X_i, Y_i) \tag{12.3.12}$$

for all $X, Y \in \mathfrak{X}^\infty$, $X = (X_0, X_1, \ldots)$ and $Y = (Y_0, Y_1, \ldots)$ with independent components. Thus $\pi(\underline{\mathbf{s}}, \underline{\mathbf{s}}^*) \leqslant \sum_{j=0}^{\infty} \pi(s_j, s_j^*)$, which together with (12.3.6), (12.3.8) and (12.3.9) completes the proof of (12.3.5). QED

In the next theorem we shall omit the restriction that $\mathbf{e}$ and $\mathbf{s}$ are independent, but we shall assume that the approximation model is of a completely deterministic type $D|D|1|\infty$. (Note that for this approximation model e_j^* as well as s_j^* can be different constants for different j.)

Lemma 12.3.2. Under the above assumptions we have the following estimates

$$\pi(\mathbf{w}, \mathbf{w}^*) = \mathbf{K}(\mathbf{w}, \mathbf{w}^*) \leqslant \pi(\boldsymbol{\zeta}, \boldsymbol{\zeta}^*) \leqslant \sum_{j=0}^{\infty} \pi(\zeta_j, \zeta_j^*) = \sum_{j=0}^{\infty} \mathbf{K}(\zeta_j, \zeta_j^*) \quad (12.3.13)$$

$$\pi(\mathbf{w}, \mathbf{w}^*) \leqslant \sum_{j=1}^{\infty} (\pi(e_j, e_j^*) + \pi(s_j, s_j^*)) = \sum_{j=1}^{\infty} (\mathbf{K}(e_j, e_j^*) + \mathbf{K}(s_j, s_j^*)). \quad (12.3.14)$$

The proof is similar to that of the previous theorem.

Lemmas 12.3.1 and 12.3.2 transfer our original problem of estimating the deviation between $\mathbf{w}$ and $\mathbf{w}^*$ to the problem of obtaining sharp or nearly sharp upper bounds for $\mathbf{K}(e_j, e_j^*) = \pi(e_j, e_j^*)$ (resp. $\mathbf{K}(\zeta_j, \zeta_j^*)$), assuming that certain moment characteristics of e_j (resp. ζ_j) are given. The problem of estimating $\pi(s_j, s_j^*)$ in (12.3.5) was considered in Kalashnikov and Rachev (1988), Chapter 4, under different assumptions such as s_j^* being exponentially distributed and s_j possessing certain 'aging' or 'lack of memory' properties; therefore, the problem is to estimate the terms IND($\mathbf{s}$), IND($\mathbf{s}^*$) and $\pi(e_j, e_j^*)$ in (12.3.5). IND($\mathbf{s}$) and IND($\mathbf{s}^*$) can be estimated using the Strassen–Dudley theorem and the Chebychev inequalities. The estimates for $\pi(e_j, e_j^*)$, $\pi(\zeta_j, \zeta_j^*)$, e_j^*, ζ_j^* being constants, are given in the next Lemmas 12.3.3 to 12.3.8.

Lemma 12.3.3. Let $\alpha > 0$, $\delta \in [0, 1]$ and ϕ be a nondecreasing continuous function on $[0, \infty)$. Then the *Ky Fan radius* (*with fixed moment* ϕ)

$$R = R(\alpha, \delta, \phi) := \max\{\mathbf{K}(X, \alpha) \colon \mathbb{E}\phi(|X - \alpha|) \leqslant \delta\} \quad (12.3.15)$$

is equal to $\min(1, \psi(\delta))$, where ψ is the inverse function of $t\phi(t)$, $t \geqslant 0$.

Proof. By Chebychev's inequality $\mathbf{K}(X, \alpha) \leqslant \psi(\delta)$ if $\mathbb{E}\phi(|X - \alpha|) \leqslant \delta$, and thus $R \leqslant \min(1, \psi(\delta))$. Moreover, if $\psi(\delta) < 1$ (otherwise, we have trivially that $R = 1$), then by letting $X = X_0 + \alpha$, where X_0 takes the values $-\varepsilon$, 0, $\varepsilon := \psi(\delta)$ with probabilities $\varepsilon/2$, $1 - \varepsilon$, $\varepsilon/2$ respectively, we obtain $\mathbf{K}(X, \alpha) = \psi(\delta)$ as is required.

QED

Using Lemma 12.3.3 we obtain a sharp estimate of $\mathbf{K}(\zeta_j, \zeta_j^*)$($\zeta_j^*$ constant) if it is known that $\mathbb{E}\phi(|\zeta_j - \zeta_j^*|) \leqslant \delta$. However, the problem becomes more difficult if one assumes that

$$\zeta_j \in S_{\zeta_j^*}(\varepsilon_{1j}, \varepsilon_{2j}, f_j, g_j) \quad (12.3.16)$$

where for fixed constants $\alpha \in \mathbb{R}$, $\varepsilon_1 \geqslant 0$ and $\varepsilon_2 > 0$

$$S_\alpha(\varepsilon_1, \varepsilon_2, f, g) := \{X \in \mathfrak{X} \colon |\mathbb{E}f(X) - f(\alpha)| \leqslant \varepsilon_1, |\mathbb{E}g(X) - g(\alpha)| \leqslant \varepsilon_2\} \quad (12.3.17)$$

and $\mathfrak{X}$ is the set of real-valued r.v.s for which $\mathbb{E}f(X)$ and $\mathbb{E}g(X)$ exist.

Suppose that the only information we have on hand concerns estimates of the deviations $|\mathbb{E}f(\zeta_j) - f(\zeta_j^*)|$ and $|\mathbb{E}g(\zeta_j) - g(\zeta_j^*)|$. Here, the main problem is the evaluation of the *Ky Fan radius*

$$D = D_\alpha(\varepsilon_1, \varepsilon_2, f, g) = \sup_{X \in S_\alpha(\varepsilon_1, \varepsilon_2, f, g)} \mathbf{K}(X, \alpha) = \sup_{X \in S_\alpha(\varepsilon_1, \varepsilon_2, f, g)} \pi(X, \alpha). \quad (12.3.18)$$

The next theorem deals with an estimate of $D_\alpha(\varepsilon_1, \varepsilon_2, f, g)$ for the 'classical' case

$$f(x) = x \quad g(x) = x^2. \quad (12.3.19)$$

Lemma 12.3.4. If $f(x) = x$, $g(x) = x^2$ then

$$\varepsilon_2^{1/3} \leqslant D_\alpha(\varepsilon_1, \varepsilon_2, f, g) \leqslant \min(1, \gamma) \quad (12.3.20)$$

where $\gamma = (\varepsilon_2 + 2|\alpha|\varepsilon_1)^{1/3}$.

Proof. By Chebychev's inequality for any $X \in S_\alpha(\varepsilon_1, \varepsilon_2, f, g)$, we have $\mathbf{K}^3(X, \alpha) \leqslant \mathbb{E}X^2 - 2\alpha\mathbb{E}X + \alpha^2 := I$. We consider two cases

If $\alpha > 0$ then $I \leqslant \alpha^2 + \varepsilon_2 - 2\alpha(\alpha - \varepsilon_1) + \alpha^2 = \gamma^3$.
If $\alpha \leqslant 0$ then $I \leqslant 2\alpha^2 + \varepsilon_2 - 2\alpha(\alpha + \varepsilon_1) = \gamma^3$.

Hence the upper bound of D (12.3.20) is established.

Consider the r.v. X which takes the values $\alpha - \varepsilon$, α, $\alpha + \varepsilon$ with probabilities p, q, $p(2p + q = 1)$ respectively. Then $\mathbb{E}X = \alpha$, so that $|\mathbb{E}X - \alpha| = 0 \leqslant \varepsilon_1$. Further $\mathbb{E}X^2 = \alpha^2 + 2\varepsilon^2 p = \varepsilon_2 + \alpha^2$, if we choose $\varepsilon = \varepsilon_2^{1/3}$, $p = \varepsilon_2^{1/3}/2$. Then $F_X(\alpha + \varepsilon - 0) - F_X(\alpha - \varepsilon) = q = 1 - \varepsilon_2^{1/3}$ and thus $\mathbf{K}(X, \alpha) \geqslant \varepsilon_2^{1/3}$, which proves the lower bound of D in (12.3.20). QED

Using Lemma 12.3.4, we can easily obtain estimates for $D_\alpha(\varepsilon_1, \varepsilon_2, f, g)$ where

$$f(x) := \lambda + \mu x + \zeta x^2 \quad x, \lambda, \mu, \zeta \in \mathbb{R}$$

and

$$g(x) := a + bx + cx^2 \quad x, a, b, c \in \mathbb{R}$$

are polynomials of degree two. Namely, assuming $c \neq 0$, we may represent f as follows: $f(x) = A + Bx + Cg(x)$, where $A = \lambda - \zeta a/c$, $B = \mu - \zeta b/c$, $C = \zeta/c$. QED

Lemma 12.3.5. Let f and g be defined as above. Assume $c \neq 0$, and $B \neq 0$. Then

$$D_\alpha(\varepsilon_1, \varepsilon_2, f, g) \leqslant D_\alpha(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \tilde{f}, \tilde{g})$$

where.

$$\tilde{\varepsilon}_1 := \frac{1}{|B|}(|C|\varepsilon_2 + \varepsilon_1) \quad \tilde{\varepsilon}_2 := \frac{1}{|c|}\left[\left|\frac{b}{B}\right|(|C|\varepsilon_2 + \varepsilon_1) + \varepsilon_2\right] \quad \tilde{f}(x) = x,\ \tilde{g}(x) = x^2.$$

In particular, $D_\alpha(\varepsilon_1, \varepsilon_2, f, g) \leqslant (\tilde{\varepsilon}_2 + 2|\alpha|\tilde{\varepsilon}_1)^{1/3} = (c_1\varepsilon_2 + c_2\varepsilon_1)^{1/3}$, where

$$c_1 = \frac{1}{|c||\mu - \zeta b|}(|b\zeta| + |\mu - \zeta b| + 2|\alpha||\zeta c|)$$

and

$$c_2 = \left|\frac{b}{\mu - \zeta b}\right| + 2|\alpha|.$$

Proof. First we consider the special case $f(x) = x$ and $g(x) = a + bx + cx^2$, $x \in \mathbb{R}$, where a, b, $c \neq 0$ are real constants. We prove first that

$$D_\alpha(\varepsilon_1, \varepsilon_2, f, g) \leqslant D_\alpha(\varepsilon_1, \tilde{\varepsilon}_2, f, \tilde{g}) \tag{12.3.21}$$

where $\tilde{\varepsilon}_2 := (1/|c|)(|b|\varepsilon_1 + \varepsilon_2)$ and $\tilde{g}(x) = x^2$. Thus, by (12.3.20) we get

$$D_\alpha(\varepsilon_1, \varepsilon_2, f, g) \leqslant (\tilde{\varepsilon}_2 + 2|\alpha|\varepsilon_1)^{1/3}. \tag{12.3.22}$$

Since $|\mathbb{E}f(X) - f(\alpha)| = |\mathbb{E}X - \alpha| \leqslant \varepsilon_1$ and $|\mathbb{E}g(X) - g(\alpha)| = |b(\mathbb{E}X - \alpha) + c(\mathbb{E}X^2 - \alpha^2)| \leqslant \varepsilon_2$, we have that $|c||\mathbb{E}X^2 - \alpha^2| \leqslant |b||\mathbb{E}X - \alpha| + \varepsilon_2 \leqslant |b|\varepsilon_1 + \varepsilon_2$. That is, $|\mathbb{E}X^2 - \alpha^2| \leqslant \tilde{\varepsilon}_2$, which establishes the required estimate (12.3.21).

Now we consider the general case of $f(x) = \lambda + \mu x + \zeta x^2$. From $f(x) = A + Bx + Cg(x)$ and the assumptions that $|\mathbb{E}f(X) - f(\alpha)| \leqslant \varepsilon_1$, $|\mathbb{E}g(X) - g(\alpha)| \leqslant \varepsilon_2$, we have $|B||\mathbb{E}X - \alpha| \leqslant |\mathbb{E}f(X) - f(\alpha)| + |C||\mathbb{E}g(X) - g(\alpha)| \leqslant \varepsilon_1 + |C|\varepsilon_2$, that is, $|\mathbb{E}X - \alpha| \leqslant \tilde{\varepsilon}_1$. Therefore, $D_\alpha(\varepsilon_1, \varepsilon_2, f, g) \leqslant D_\alpha(\tilde{\varepsilon}_1, \varepsilon_2, \tilde{f}, g)$, where $\tilde{f}(x) = x$. Using (12.3.22), we have that $D_\alpha(\tilde{\varepsilon}_1, \varepsilon_2, \tilde{f}, g) \leqslant D_\alpha(\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \tilde{f}, \tilde{g})$, where

$$\tilde{\varepsilon}_2 = \frac{1}{|c|}(|b|\tilde{\varepsilon}_1 + \varepsilon_2)$$

which by means of Lemma 12.3.4 completes the proof of Lemma 12.3.5. QED

The main assumption in Lemmas 12.3.3 to 12.3.5 was the monotonicity of ϕ, f and g, which allows us to use the Chebychev inequality. More difficult is the problem of finding $D_\alpha(\varepsilon_1, \varepsilon_2, f, g)$ when f and g are not polynomials of degree two.

We are going to meet a quite difficult case, namely, when

$$f(x) = \cos x \quad g(x) = \sin x \quad x \in [0, 2\pi].$$

Remark 12.3.1. If fact, we shall investigate the rate of the convergence of $\mathbf{K}(X_n, \alpha) \to 0$ $(0 \leqslant X_n \leqslant 2\pi)$ as $n \to \infty$, provided that $\mathbb{E}\cos X_n \to \cos\alpha$ and

$\mathbb{E} \sin X_n \to \sin \alpha$. In the next lemma we show Berry–Essen type bounds for the implication

$$\mathbb{E}\exp(iX_n) \to \exp(i\alpha) \Rightarrow \mathbf{K}(X_n, \alpha) = \pi(X_n, \alpha) \to 0.$$

In the following we consider probability measures μ on $[0, 2\pi]$ and let

$$M(\varepsilon) = \left\{\mu: \left|\int \cos t \, d\mu - \cos \alpha\right| \leqslant \varepsilon, \left|\int \sin t \, d\mu - \sin \alpha\right| \leqslant \varepsilon\right\}. \quad (12.3.23)$$

We would like to evaluate the *trigonometric Ky Fan (or Prokhorov) radius for* $M(\varepsilon)$ defined by

$$D = \sup\{\pi(\mu, \delta_\alpha): \mu \in M(\varepsilon)\} \quad (12.3.24)$$

where δ_α is the point mass at α and $\pi(\mu, \delta_\alpha)$ is the Ky Fan (or Prokhorov) metric

$$\pi(\mu, \delta_\alpha) = \inf\{r > 0: \mu([\alpha - r, \alpha + r]) \geqslant 1 - r\} \quad (12.3.25)$$

Our main result is as follows.

Lemma 12.3.6. Let fixed $\alpha \in [1, 2\pi - 1]$ and $\varepsilon \in (0, (1/\sqrt{2})(1 - \cos 1))$. We get D as the unique solution of

$$D - D\cos D = \varepsilon(|\cos \alpha| + |\sin \alpha|). \quad (12.3.26)$$

Here we have that $D \in (0, 1)$.

Remark 12.3.2. By (12.3.24) one obtains

$$D \leqslant [2\varepsilon(|\cos \alpha| + |\sin \alpha|)]^{1/3}. \quad (12.3.27)$$

From (12.3.26), (12.3.27) (and see also further (12.3.28), we have that $D \to 0$ as $\varepsilon \to 0$. The last implies that $\pi(\mu, \delta_\alpha) \to 0$ which in turn gives that $\mu \xrightarrow{w} \delta_\alpha$, where δ_α is the point mass at α. In fact, D converges to zero quantitatively through (12.3.24), (12.3.27), i.e., the knowledge of D gives the rate of weak convergence of μ to δ_α, see also further Lemma 12.3.7.

The proofs of Lemmas 12.3.6 and 12.3.7, while based on the solution of certain moment problems (see Section 9), need more facts on the Kemperman geometric approach for the solution of the general moment problem (see Kemperman 1968, 1987), and will be omitted. For the necessary proofs we refer to Anastassiou and Rachev (1990).

Lemma 12.3.7. Let $f(x) = \cos x$, $g(x) = \sin x$; $\alpha \in [0, 1)$ or $\alpha \in (2\pi - 1, 2\pi)$. Define

$$D = D_\alpha(\varepsilon, f, g) = \sup\{\mathbf{K}(X, \alpha): |\mathbb{E}\cos X - \cos \alpha| \leqslant \varepsilon, |\mathbb{E}\sin X - \sin \alpha| \leqslant \varepsilon\}.$$

Let $\beta = \alpha + 1$ if $\alpha \in [0, 1)$, and $\beta = \alpha - 1$ if $\alpha \in (2\pi - 1, 2\pi)$. Then

$$D_\alpha(\varepsilon, f, g) \leqslant D_\beta(\varepsilon(\cos 1 + \sin 1), f, g).$$

In particular, by (12.3.27),

$$D_\alpha(\varepsilon, f, g) \leqslant [2\varepsilon(\cos 1 + \sin 1)(|\cos \alpha| + |\sin \alpha|)^{1/3} \tag{12.3.28}$$

for any $0 \leqslant \alpha < 2\pi$ and $\varepsilon \in (0, (1/\sqrt{2})(1 - \cos 1))$.

Further, we are going to use (12.3.28) in order to obtain estimates for $D_\alpha(\varepsilon, f, g)$ where $f(x) = \lambda + \mu \cos x + \zeta \sin x$, $x \in [0, 2\pi]$, $\lambda, \mu, \zeta \in \mathbb{R}$, and $g(x) = a + b \cos x + c \sin x$, $x \in [0, 2\pi]$, $a, b, c \in \mathbb{R}$. Assuming $c \neq 0$ we have $f(x) = A + B \cos x + Cg(x)$, where $A = \lambda - \zeta a/c$, $B = \mu - \zeta b/c$, $C = \zeta/c$.

Lemma 12.3.8. Let the trigonometric polynomials f and g be defined as above. Assume $c \neq 0$ and $B \neq 0$. Then $D_\alpha(\varepsilon, f, g) \leqslant D_\alpha(\varepsilon\tau\eta, \tilde{f}, \tilde{g})$, for any $0 \leqslant \alpha < 2\pi$, where

$$\tau = \max\left(1, \frac{1}{|c|}(|b| + 1)\right)$$

and

$$\eta = \max\left(1, \frac{1}{|B|}(|C| + 1)\right)$$

$\tilde{f}(x) = \cos x$, $\tilde{g}(x) = \sin x$. If

$$0 \leqslant \varepsilon \leqslant \frac{1}{\tau\eta\sqrt{2}}(1 - \cos 1)$$

then we obtain

$$D_\alpha(\varepsilon, f, g) \leqslant [2\varepsilon\tau\eta(\cos 1 + \sin 1)(|\cos \alpha| + |\sin \alpha|)]^{1/3} \tag{12.3.29}$$

for any $0 \leqslant \alpha < 2\pi$.

The proof is similar to that of Lemma 12.3.5.

Now we can state the main result determining the deviation between the waiting times of a deterministic and random queueing models.

Theorem 12.3.1. (i) Let the approximating queueing model be of type $D|G|1|\infty$. Assume that the sequences $\mathbf{e}$ and $\mathbf{s}$ of the 'real' queue of type $G|G|1|\infty$ are independent. Then the Prokhorov metric between the sequences of waiting times of $D|G|1|\infty$ queue and $G|G|1|\infty$ queue is estimated as follows

$$\pi(\mathbf{w}, \mathbf{w}^*) \leqslant \mathrm{IND}(\mathbf{s}) + \mathrm{IND}(\mathbf{s}^*) + \sum_{j=1}^{\infty} (\pi(e_j, e_j^*) + \pi(s_j, s_j^*)). \tag{12.3.30}$$

(ii) Assume that the approximating model is of type $D|D|1|\infty$ and the 'real' queue is of type $G|G|1|\infty$. Then

$$\pi(\mathbf{w}, \mathbf{w}^*) \leqslant 2 \sum_{j=1}^{\infty} \pi(\zeta_j, \zeta_j^*) \tag{12.3.31}$$

and

$$\pi(\mathbf{w}, \mathbf{w}^*) \leqslant 2 \sum_{j=1}^{\infty} (\pi(e_j, e_j^*) + \pi(s_j, s_j^*)). \tag{12.3.32}$$

(iii) The right-hand sides of (12.3.30), (12.3.31) and (12.3.32) can be estimated as follows: let $\pi(X, X^*)$ denote $\pi(e_j, e_j^*)$ in (12.3.30) or $\pi(\zeta_j, \zeta_j^*)$ in (12.3.31) or $\pi(e_j, e_j^*)(\pi(s_j, s_j^*))$ in (12.3.32) (note that X^* is a constant). Then

(a) If for a function ϕ which is non-decreasing on $[0, \infty)$ and continuous on $[0, 1]$ holds

$$\mathbb{E}\phi(|X - X^*|) \leqslant \delta \leqslant 1 \tag{12.3.33}$$

then

$$\pi(X, X^*) \leqslant \min(1, \psi(\delta)) \tag{12.3.34}$$

where ψ is the inverse function of $t\phi(t)$

(b) If $|\mathbb{E}f(X) - f(X^*)| \leqslant \varepsilon_1$, $|\mathbb{E}g(X) - g(X^*)| \leqslant \varepsilon_2$, where

$$f(x) = \lambda + \mu x + \zeta x^2 \quad x, \lambda, \mu, \zeta \in \mathbb{R}$$

$$g(x) = \alpha + bx + cx^2 \quad x, \alpha, b, c \in \mathbb{R}$$

$c \neq 0$, $\mu \neq \zeta b/c$, then for any $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$,

$$\pi(X, X^*) \leqslant (\tilde{\varepsilon}_2 + 2|X^*|\tilde{\varepsilon}_1)^{1/3}$$

where $\tilde{\varepsilon}_1$ and $\tilde{\varepsilon}_2$ are linear combinations of ε_1 and ε_2 defined as in Lemma 12.3.5.

(c) If $X \in [0, 2\pi]$ a.e. and $|\mathbb{E}f(X) - f(X^*)| \leqslant \varepsilon$, $|\mathbb{E}g(X) - g(X^*)| \leqslant \varepsilon$, where $f(x) = \lambda + \mu \cos x + \zeta \sin x$, and $g(x) = a + b \cos x + c \sin x$ for $x \in [0, 2\pi]$, a, b, c, λ, μ, $\zeta \in \mathbb{R}$, $c \neq 0$, $\mu \neq \zeta b/c$, then

$$\mathbf{K}(X, X^*) \leqslant [2\varepsilon\tau\eta(\cos 1 + \sin 1)(|\cos X^*| + |\sin X^*|)]^{1/3}$$

where the constants τ and η are defined as in Lemma 12.3.8.

Open problem 12.3.1. First, one can easily combine the results of this section with those of Kalashnikov and Rachev (1988, Chapter 5), in order to obtain estimates between the outputs of general multi-channel and multi-stage models and approximating queueing models of types $G|D|1|\infty$ and $D|G|1|\infty$. However, much more interesting and difficult is to obtain *sharp* estimates for $\pi(\mathbf{e}, \mathbf{e}^*)$, assuming that $\mathbf{e}$ and $\mathbf{e}^*$ are random sequences satisfying

$$|\mathbb{E}(e_j - e_j^*)| \leqslant \varepsilon_{1j} \quad |\mathbb{E}f_j(|e_j|) - \mathbb{E}f_j(|e_j^*|) \leqslant \varepsilon_{2j}.$$

Here even the case $f_j(x) = x^2$ is open (see Section 9).

Open problem 12.3.2. It is interesting to get estimates for $\ell_p(\mathbf{w}, \mathbf{w}^*)$, $(0 < p \leqslant \infty)$, where $\ell_p = \hat{\mathscr{L}}_p$ (see Sections 12.1 and 12.2).

CHAPTER 13

Optimal Quality Usage

13.1 OPTIMALITY OF QUALITY USAGE AND THE MONGE–KANTOROVICH PROBLEM

The quality of an item of a product is usually described by a collection of its characteristics $x = (x_1, \ldots, x_m)$ where m is a required number of quality characteristics and x_i is the real value of the ith characteristic.

The quality of all produced items of a given type is described by a probability measure $\mu(A)$, $A \in \mathfrak{B}^m$, where, as before, $\mathfrak{B}^m$ is the Borel σ-algebra sets in $\mathbb{R}^m$. The measure $\mu(A)$ represents the proportion of items with quality x satisfying $x \in A$. On the other hand the usage (consumption) of all produced items can be represented by another probability measure $\nu(B)$, $B \in \mathfrak{B}^m$, where $\nu(B)$ describes the necessary consumption product for which the quality characteristics satisfy $x \in B$. We call $\mu(A)$ the *production quality measure* and $\nu(B)$ the *consumption quality measure* $A, B \in \mathfrak{B}^m$ and assume that $\mu(\mathbb{R}^m) = \nu(\mathbb{R}^m) = 1$. Clearly, it happens often that $\mu(A) \neq \nu(A)$ at least for some $A \in \mathfrak{B}^m$.

Following the formulation of the Monge-Kantorovich problem (see Section 5.1) we introduce the loss function $\phi(x, y)$ defined for all $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^m$ and taking positive values whenever an item with quality x is used in place of an item with required quality y. Finally, we propose the notion of a distribution plan for production quality (with given measure $\mu(A)$) to satisfy the demand for consumption (with given measure $\nu(B)$). We define for any distribution plan (for short 'plan') a non-negative Borel measure $\theta(A, B)$ on the direct product $\mathbb{R}^m \times \mathbb{R}^m = \mathbb{R}^{2m}$. The measure $\theta(A, B)$ indicates the part of produced items with quality $x \in A$, which is planned to satisfy a required consumption of items with quality $y \in B$. The plan $\theta(A, B)$ is called admissible if it satisfies the balance equation

$$\theta(A, \mathbb{R}^m) = \mu(A); \; \theta(\mathbb{R}^m, B) = \nu(B) \quad \forall A, B \in \mathfrak{B}^m. \tag{13.1.1}$$

In reality the balance equations express the fact that any produced item will be consumed and any demand for an item will be satisfied.

Denote $\Theta(\mu, \nu)$ the collection of all admissible plans. For a given plan $\theta \in \Theta(\mu, \nu)$ the total loss of consumption quality is defined by the following integral

$$\tau(\theta) := \tau_\phi(\theta) := \int_{\mathbb{R}^{2m}} \phi(x, y)\theta(\mathrm{d}x, \mathrm{d}y). \tag{13.1.2}$$

θ^* is said to be the *optimal plan for consumption quality* if it satisfies the relationship

$$\tau_\phi(\theta^*) = \hat{\tau}_\phi(\mu, \nu) := \inf_{\theta \in \Theta(\mu, \nu)} \tau(\theta). \tag{13.1.3}$$

Relations (13.1.1) express the balances between the production quality measure $\mu(A)$, the consumption quality measure $\nu(B)$ and the distribution plan $\theta(A, B)$. It assumes that full information on the marginals μ and ν is available when constructing the plan. In most practical cases the information about production and consumption quality concerns only the set of distributions of x_is ($i = 1, \ldots, m$). In this case it is assumed that the balance equations can be expressed in terms of the corresponding one-dimensional marginal measures. This leads to formulation of the multi-dimensional Kantorovich problem (see Section 5.1, IV and (5.1.36)). If we denote the ith marginal measure of production quality by $\mu_i(A_i)$ and the jth marginal measure of the consumption quality by $\nu_j(B_j)$, then the following hold

$$\mu_i(A_i) = \mu(\mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{m-i}) \quad A_i \in \mathfrak{B}^1$$

$$\nu_j(B_j) = \nu(\mathbb{R}^{j-1} \times B_j \times \mathbb{R}^{m-j}) \quad B_j \in \mathfrak{B}^1.$$

We say a distribution plan $\theta(A, B)$ is *weakly admissible* when it satisfies the conditions

$$\theta(\mathbb{R}^{i-1} \times A_i \times \mathbb{R}^{m-i}, \mathbb{R}^m) = \mu_i(A_i) \quad i = 1, \ldots, m \tag{13.1.4}$$

$$\theta(\mathbb{R}^m, \mathbb{R}^{j-1} \times B_j \times \mathbb{R}^{m-j}) = \nu_j(B_j) \quad j = 1, \ldots, m. \tag{13.1.5}$$

Denote by $\bar{\Theta}(\mu_1, \ldots, \mu_m; \nu_1, \ldots, \nu_m)$ the collection of all weakly admissible plans. Obviously

$$\Theta(\mu, \nu) \subseteq \bar{\Theta}(\mu_1, \ldots, \mu_m; \nu_1, \ldots, \nu_m). \tag{13.1.6}$$

A distribution plan θ^o is called *weakly optimal* if it satisfies the relation

$$\tau(\theta^0) = \inf_{\theta \in \bar{\Theta}} \tau(\theta) \tag{13.1.7}$$

where $\tau(\theta)$ is defined by (13.1.2) for a given loss function $\phi(x, y)$. (13.1.6) mean we have $\tau(\theta^o) \leqslant \tau(\theta^*)$, where θ^o and θ^* are determined by (13.1.3) and (13.1.7). Therefore $\tau(\theta^o)$ is an essential lower bound on the minimal total losses.

First we shall evaluate $\tau(\theta^*)$ and determine θ^*. We consider two types of loss functions $\phi(x, y)$, when the item with quality x is used instead of an item with required quality y.

The first type has the following form

$$\phi(x, y) = a(x) + b(x, y) \tag{13.1.8}$$

where $a(x)$ is the production cost of an item with quality x and $b(x, y) = b_o(x, y) + b_o(y, x)$, $b_o(x, y)$ is the consumer's expenses resulting from replacing the required item with quality y by a product with quality x. We can assume that $b(x, y) = 0$ for all $x = y$ and $b(x, y) \geqslant 0$, $a(x) \geqslant 0$ $\forall x \in \mathbb{R}^m$.

From (13.1.3) and (13.1.8)

$$\begin{aligned}\tau_\phi(\theta^*) &:= \inf_{\theta \in \Theta(\mu, \nu)} \left\{ \int_{\mathbb{R}^{2m}} [a(x) + b(x, y)]\theta(\mathrm{d}x, \mathrm{d}y) \right\} \\ &= \int_{\mathbb{R}^m} a(x)\mu(\mathrm{d}x) + \inf_{\theta \in \Theta(\mu, \nu)} \int_{\mathbb{R}^{2m}} b(x, y)\theta(\mathrm{d}x, \mathrm{d}y) =: I_1 + I_2.\end{aligned} \tag{13.1.9}$$

Here

$$I_1 := \int_{\mathbb{R}^m} a(x)\mu(\mathrm{d}x) \tag{13.1.10}$$

represents the expected (complete) production price of items with quality measure μ, whereas

$$I_2 := \inf_{\theta \in \Theta(\mu, \nu)} \int_{\mathbb{R}^{2m}} b(x, y)\theta(\mathrm{d}x, \mathrm{d}y) \tag{13.1.11}$$

represents the minimal (expected) means of consumer's expenses from exchanging the required product for consumption with quality ν by the produced item with quality μ, under its optimal distribution among consumers, according to plan θ^*. Since I_1 in (13.1.10) is completely determined by the measure μ, the only problem is the evaluation of I_2.

The second type of loss function which is of interest has the form

$$\phi(x, y) = H(d(x, y)) \tag{13.1.12}$$

where $H(t)$ is a non-decreasing function and d is a metric in $\mathbb{R}^m$, characterizing the deviation between the production quality x and the required consumption quality y. The function $H(t)$ is defined for all $t \geqslant 0$, and represents the user's expenses as a function of the deviation $d(x, y)$. Notice that the function $b(x, y)$ in (13.1.11) may also be written in the form (13.1.12), so without loss of generality we may assume that ϕ has the form (13.1.12).

The dual representation for $\hat{\tau}_\phi$ (13.1.3) is given by Corollary 5.2.1, i.e., if the loss function $\phi(x, y)$ is given by (13.1.12) where H is convex and $K_H := \sup_{t<\infty} [H(2t)/H(t)] < \infty$ (see 2.2.3), then $\hat{\tau}_\phi$ is a minimal distance with dual representation

$$\hat{\tau}_\phi(\mu, \nu) = \sup\left\{\int_{\mathbb{R}^m} f\,\mathrm{d}\mu + \int_{\mathbb{R}^m} g\,\mathrm{d}\nu : f, g \in \operatorname{Lip}\mathbb{R}^m,\right.$$

$$\left. f(x) + g(y) \leqslant H(d(x, y));\ x, y \in \mathbb{R}^m\right\} \tag{13.1.13}$$

where

$$\operatorname{Lip}\mathbb{R}^m := \left\{f\colon \mathbb{R}^m \to \mathbb{R}^1\colon \|f\|_\infty = \sup_{x\in\mathbb{R}^m} |f(x)| < \infty,\right.$$

$$\left.\sup_{x,y\in\mathbb{R}^m} |f(x) - f(y)|/d(x, y) < \infty\right\}.$$

By the Cambanis–Simons–Stout formula, see (8.1.26) in the case of $m = 1$ and $d(x, y) = |x - y|$, the minimal total losses can be expressed by

$$\hat{\tau}_\phi(\mu, \nu) = \tau_\phi(\theta^*) = \int_0^1 H(|F^{-1}(x) - G^{-1}(x)|)\,\mathrm{d}x \tag{13.1.14}$$

where $F(x) = \mu((-\infty, x])$ and $G(x) = \nu((-\infty, x])$ are the distribution functions of the production quality and the required quality characteristics for usage, respectively. The functions $F^{-1}(x)$ and $G^{-1}(x)$ are their generalized inverses defined by $F^{-1}(x) := \sup\{t\colon F(t) \leqslant x\}$. Furthermore, the optimal distribution plan is given by

$$\theta^*((-\infty, x] \times (-\infty, y]) = \min(F(x), G(y)). \tag{13.1.15}$$

The equality (13.1.15) essentially means that if $F(x)$ is a continuous d.f., then the optimal correspondence between the item of quality x and the item with required quality y is given by

$$y = G^{-1}(F(x)). \tag{13.1.16}$$

The last formula follows immediately from (13.1.14), (13.1.15), since the minimal distance

$$\tau_\phi(\theta^*) = \inf\{\mathbb{E}H(|X - Y|)\colon F_X = F, F_Y = G\} \tag{13.1.17}$$

is equal to $\mathbb{E}H(|X^* - Y^*|)$, where $Y^* = G^{-1}(F(X^*))$ and the joint distribution of X^*, Y^* is given by θ^*. Thus, the case of $m = 1$ is solved for any ϕ given by (13.1.2). However, (13.1.16) holds in a more general situation when ϕ is a quasiantitone function (see Definition 7.3.1, Theorem 7.3.2 and Remark 7.3.1).

The next theorem deals with the special case where $\phi(x, y)$ is $\|x - y\|^2$ and $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^m$. Let μ and ν be two probability measures on $\mathfrak{B}^m$ such that

$$\int_{\mathbb{R}^m} \|x\|^2(\mu + \nu)(\mathrm{d}x) < \infty.$$

Recall that the pair of m-dimensional vectors (X^*, Y^*) with joint distribution θ^* and marginal distributions μ and ν is *optimal* if

$$\tau_\phi(\theta^*) = \mathbb{E}\|X^* - Y^*\|^2 = \inf\{\mathbb{E}\|X - Y\|^2 \colon \Pr_X = \mu, \Pr_Y = \nu\}. \quad (13.1.18)$$

In the next theorem we describe the necessary and sufficient condition for a pair (X^*, Y^*) to be optimal. To this end we recall the definition of a subdifferential (cf. Rockafellar, 1970). For a lower semicontinuous convex (LSC) function f on $\mathbb{R}^m$ let f^* denote the conjugate function

$$f^*(y) := \sup_{x \in \mathbb{R}^m} \{\langle x, y\rangle - f(x)\}$$

$$\left(\langle x, y\rangle := \sum_{i=1}^m x_i y_i,\ x = (x_1, \ldots, x_m),\ y = (y_1, \ldots, y_m)\right) \quad (13.1.19)$$

and denote the *subdifferential* of f in x by

$$\partial f(x) = \{y \in \mathbb{R}^m \colon f(z) - f(x) \geqslant \langle z - x, y\rangle,\ z \in \mathbb{R}^m\}. \quad (13.1.20)$$

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

$$f(x) + f^*(y) \geqslant \langle x, y\rangle \quad (13.1.21)$$

with equality if and only if $y \in \partial f(x)$.

Theorem 13.1.1. (X^*, Y^*) is optimal if for some LSC function f

$$Y^* \in \partial f(X^*)(\Pr \text{ a.s.}). \quad (13.1.22)$$

Remark 13.1.1. Note that we can always reduce to the case where the means $m_\mu := \{\int_{\mathbb{R}^m} x_i \mu(\mathrm{d}x), i = 1, \ldots, m\}$ and m_ν are zero vectors. Simply note that if X and Y are $\mathbb{R}^m$-valued random variables with distributions $\Pr_X = \mu$, $\Pr_Y = \nu$ and $\xi = X - m_\mu$, $\eta = Y - m_\nu$, then

$$\mathbb{E}\|X - Y\|^2 = \mathbb{E}\|\xi - \eta\|^2 + \|m_\mu - m_\nu\|^2. \quad (13.1.23)$$

Proof. We begin with

$$\mathbb{E}\|X - Y\|^2 = \mathbb{E}\|X\|^2 + \mathbb{E}\|Y\|^2 - 2\mathbb{E}\langle X, Y\rangle. \quad (13.1.24)$$

Therefore, the problem (13.1.18) is equivalent to

Find (X^*, Y^*) such that

$$\mathbb{E}\langle X^*, Y^*\rangle = \sup\{\mathbb{E}\langle X, Y\rangle \colon \Pr_X = \mu, \Pr_Y = \nu\}. \quad (13.1.25)$$

By the duality theorem (cf. Theorem 8.1.1, (13.1.13)) and (13.1.24), it follows that

$$\begin{aligned}
&\sup\{\mathbb{E}\langle X, Y\rangle : \Pr_X = \mu, \Pr_Y = \nu\} \\
&\quad = \int \|x\|^2(\mu + \nu)(\mathrm{d}x) - \inf\{\mathbb{E}\|X - Y\| : \Pr_X = \mu, \Pr_Y = \nu\} \\
&\quad = \int \|x\|^2(\mu + \nu)(\mathrm{d}x) - \sup\left\{\int g\,\mathrm{d}\mu + \int h\,\mathrm{d}\nu : g, h \in \mathrm{Lip}\,\mathbb{R}^m,\right. \\
&\qquad\qquad\qquad \left.\text{and} \quad \forall x, y \in \mathbb{R}^m, g(x) + h(y) \leqslant \|x - y\|^2\right\} \\
&\quad \geqslant \inf\left\{\int \tilde{g}\,\mathrm{d}\mu + \int \tilde{h}\,\mathrm{d}\nu : \int |\tilde{g}|\,\mathrm{d}\mu < \infty, \int |\tilde{h}|\,\mathrm{d}\nu < \infty, \tilde{g}(x) + \tilde{h}(y) \geqslant \langle x, y\rangle\right\} \\
&\quad \geqslant \sup\{\mathbb{E}\langle X, Y\rangle : \Pr_X = \mu, \Pr_Y = \nu\}.
\end{aligned} \tag{13.1.26}$$

Here, the last inequality follows from the 'trivial' part of the duality theorem and therefore the last two inequalities are valid with equality signs.

Now, let $\Pr_{X^*} = \mu$, $\Pr_{Y^*} = \nu$ and assume that $Y^* \in \partial f(X^*)$ (Pr-a.s.) for a LSS function f. Then for any other random variables $\tilde{X}$ and $\tilde{Y}$ with distributions μ and ν we have

$$\mathbb{E}\langle \tilde{X}, \tilde{Y}\rangle \leqslant \mathbb{E}(f(\tilde{X}) + f^*(\tilde{Y})) = \mathbb{E}(f(X^*) + f^*(X^*)) = \mathbb{E}\langle X^*, Y^*\rangle.$$

Therefore, (13.1.25) holds. QED

Remark 13.1.2. Condition (13.1.18) is also necessary.

Sketch of the proof: Let, conversely, $\langle X^*, Y^*\rangle$ be a solution of (13.1.25). Then, by (13.1.26),

$$\begin{aligned}
&\sup\{\mathbb{E}\langle X, Y\rangle : \Pr_X = \mu, \Pr_Y = \nu\} \\
&\quad = \inf\left\{\int g\,\mathrm{d}\mu + \int h\,\mathrm{d}\nu : g(x) + h(y) \geqslant \langle x, y\rangle, \int |g|\,\mathrm{d}\mu + \int |h|\,\mathrm{d}\nu < \infty\right\}.
\end{aligned} \tag{13.1.27}$$

Note that the supremum in (13.1.27) is attained, see Corollary 5.2.1 and (13.1.26). Moreover, one could see that the infimum in (13.1.26) is also attained, see the proof of Theorem 5.2.1, Kellerer (1984a) (Theorem 2.21), and Knott and Smith (1984) (Theorem 3.2). Suppose $f(x)$ and $g(y)$ are 'optimal', i.e., $\mathbb{E}\langle X^*, Y^*\rangle = \int g\,\mathrm{d}\mu + \int h\,\mathrm{d}\nu$ and $g(x) + h(y) \geqslant \langle x, y\rangle$. Then $g^*(y) = \sup_x \{\langle x, y\rangle - g(x)\} \leqslant h(y)$, and thus (g, g^*) is also optimal. In the same way, defining $f = g^{**}$ we see that $\langle x, y\rangle \leqslant f(x) + f^*(y)$ and also f is a LSC function. This implies that $\langle X^*, Y^*\rangle = f(X^*) + f^*(Y^*)$ (Pr-a.s.) and therefore by (13.1.21) that $Y^* \in \partial f(X^*)$(Pr-a.s.).

Remark 13.1.3. If $m = 1$ and F, G are d.f.s of μ and ν, then, as we have seen by (13.1.14) (with $H(t) = t^2$), the optimal pair X^*, Y^* is given by $X^* = F^{-1}(V)$, $Y^* = G^{-1}(V)$, where V is uniform on $(0, 1)$. Defining $\theta(x) := G^{-1} \circ F(x)$ and $f(x) = \int_0^x \theta(y)\,dy$, f is convex and $Y = G^{-1}(V) \in \partial f(F^{-1}(V))$. So, (13.1.14) is a consequence of Theorem 13.1.1.

Remark 13.1.4. For a symmetric positive semidefinite $(m \times m)$ matrix T define $f(x) = \frac{1}{2}\langle x, Tx\rangle$ and $g(y) = \frac{1}{2}\langle y, T^{-1}y\rangle$. Then $f(x) + g(Tx) = \langle x, Tx\rangle$. Therefore, if $\nu = \mu \circ T^{-1}$ (T^{-1} denotes the More–Penrose inverse), then the pair (X^*, TX^*) is optimal. This leads to the explicit expression for $\tau_\phi(\theta^*)$ (13.1.14) when μ and ν are Gaussian measures on $\mathbb{R}^m$ with means m_μ, m_ν and non-singular covariance matrices Σ_μ and Σ_ν.

Corollary 13.1.1 (Olkin and Pukelheim 1982). In the Gaussian case, μ and ν being normal laws with means m_μ and m_ν and covariance matrices $\sum_\mu$ and $\sum_\nu$,

$$\tau_\phi(\theta^*) = \|m_\mu - m_\nu\|^2 + \operatorname{tr}(\textstyle\sum_\mu) + \operatorname{tr}(\sum_\nu) - 2\operatorname{tr}(\sum_\mu^{1/2} \sum_\nu \sum_\mu^{1/2})^{1/2}. \quad (13.1.28)$$

Proof. We can assume that $m_\mu = m_\nu = 0$, see (13.1.22). Applying Remark 13.1.4 we have that the pair (X^*, TX^*) with

$$T = \textstyle\sum_\nu^{1/2} (\sum_\nu^{1/2} \sum_\mu \sum_\nu^{1/2})^{-1/2} \sum_\nu^{1/2} \quad (13.1.29)$$

is optimal. Hence, by (13.1.29) $\mathbb{E}\langle X^*, TX^*\rangle = \operatorname{tr}(\sum_\mu^{1/2} \sum_\nu \sum_\mu^{1/2})^{1/2}$ is the maximal possible value for $\mathbb{E}\langle X, Y\rangle$ with $\Pr_X = \mu$ and $\Pr_Y = \nu$. QED

Thus, if both production quality measure μ and the consumption quality measure ν are Gaussian, then the optimal plan for consumption quality θ^* is determined by the joint distribution of (X^*, TX^*), where T is given by (13.1.29).

In order to determine θ^* we need to have complete information on the measures μ and ν. It is much more likely that we can have only the one-dimensional distributions μ_i and ν_j, see (13.1.4), (13.1.5), i.e. we deal with the set of weakly admissible plants $\bar{\theta}(\mu_1, \ldots, \mu_m; \nu_1, \ldots, \nu_m)$ and we would like to determine the weakly optimal plan θ^o and evaluate $\tau(\theta^o)$, see (13.1.7).

We make use of the multi-dimensional Kantorovich Theorem (see Section 5.2) in order to obtain a dual representation for $\tau(\theta^o)$. As in (13.1.13), suppose the cost function ϕ is given by (13.1.12), where H is convex and $K_H < \infty$. Then, by Theorem 5.2.1, there exists a weakly optimal plan θ^o for which the minimal value of the total loss function is

$$\tau_\phi(\theta^o) = \sup_{f_i, g_j \in C_\phi} \left(\sum_{i=1}^{m} \int_{\mathbb{R}} f_i(x)\mu_i(dx) + \sum_{j=1}^{m} \int_{\mathbb{R}} g_j(y)\nu_j(dy) \right) \quad (13.1.30)$$

where C_ϕ denotes the collection of all functions $f_i(x_i)$, $g_j(y_j)$ on $\mathbb{R}$ satisfying the constraints

$$\mathrm{Lip}(f_i) := \sup_{x \neq y} |f_i(x) - f_i(y)|/|x - y| < \infty \quad \mathrm{Lip}(g_j) < \infty \qquad (13.1.31)$$

and

$$\sum_{i,j=1}^{m} [f_i(x_i) + g_j(y_j)] < \phi(x, y), \quad \forall x, y \in \mathbb{R}^m. \qquad (13.1.32)$$

Moreover, by Theorem 7.3.2 we can get explicit representations for θ^o and $\tau_\phi(\theta^o)$, for any cost function ϕ that is quasiantitone, see Definition 7.3.1. Denote F_i the d.f.s of μ_i and G_i the d.f.s of ν_i. Define the random variables $\mathring{X}_i = F_i^{-1}(V)$; $\mathring{Y}_j = G_j^{-1}(V)$, $i, j = 1, \ldots, m$ and the random vectors $\mathring{X} = (\mathring{X}_1, \ldots, \mathring{X}_m); \mathring{Y} = (\mathring{Y}_1, \ldots, \mathring{Y}_m)$ where $F_i^{-1}(x)$, $G_j^{-1}(x)$ are the inverses of the distribution functions $F_i(x)$, $G_j(x)$ respectively, and V is uniform on $[0, 1]$.

Theorem 13.1.2. For any cost function ϕ: $\mathbb{R}^{2m} \to \mathbb{R}$ which is *quasiantitone*, the weakly distribution plan θ^o with distribution function F_o given by

$$F_o(x_1, \ldots, x_m; y_1, \ldots, y_m) = \min(F_1(x_1), \ldots, F_m(x_m), G_1(y_1), \ldots, G_m(y_m)) \qquad (13.1.33)$$

is optimal. Moreover, in this case the minimal total cost is given by

$$\tau_\phi(\theta^o) = \mathbb{E}\phi(\mathring{X}, \mathring{Y}) = \int_0^1 \phi(F^{-1}(t)), \ldots, F_m^{-1}(t), G_1^{-1}(t), \ldots, G_m^{-1}(t))\,\mathrm{d}t. \qquad (13.1.34)$$

For example, let ϕ be the following metric in $\mathbb{R}^m$, for $x = (x_1, \ldots, x_m)$, $y = (y_1, \ldots, y_m)$

$$\phi(x, y) = 2\max(x_1, \ldots, x_m, y_1, \ldots, y_m) - \frac{1}{m}\sum_{i=1}^{m}(x_i + y_i)$$

(see also (7.3.19)). Then, by Theorem 13.1.2 and Theorem 13.1.3, θ^o with d.f. F_o is an optimal plan and

$$\tau_\phi(\theta^o) = \int_{-\infty}^{\infty} \frac{1}{n}\sum_{i=1}^{n}(F_i(u) + G_i(u)) - 2\min[F_1(u), \ldots, F_n(u), G_1(u), \ldots, G_n(u)]\,\mathrm{d}u.$$

13.2 ESTIMATES OF THE MINIMAL TOTAL LOSSES $\tau_\phi(\phi^*)$

Consider the multi-dimensional case, when the quality vector $x = (x_1, \ldots, x_m)$ has $m > 1$ one-dimensional characteristics. We derive an upper bound for $\tau_\phi(\theta^*)$

(see (13.1.3)) in the special case when the loss function has the form

$$\phi(x, y) = K \sum_{i=1}^{m} |x_i - y_i|, \ x, y \in \mathbb{R}^m \quad x := (x_1, \ldots, x_m) \quad y := (y_1, \ldots, y_m). \tag{13.2.1}$$

Remark 13.2.1. In this particular case

$$\tau_\phi(\theta^*) = K\ell_1(\mu, \nu) := K \sup\left\{\left|\int_{\mathbb{R}^m} u \, \mathrm{d}(\mu - \nu)\right| : u \in \mathrm{Lip}^b_{1.1}(\mathbb{R}^m)\right\} \tag{13.2.2}$$

where ℓ_1 is the minimal metric with respect to the $\mathscr{L}_1$-distance

$$\mathscr{L}_1(X, Y) = \mathbb{E}\|X - Y\|_1 \quad X, Y \in \mathfrak{X}(\mathbb{R}^m) \quad \|x - y\|_1 := \sum_{i=1}^{m} |x_i - y_i| \quad x, y \in \mathbb{R}^m \tag{13.2.3}$$

see (3.2.2), (3.3.3) and (5.2.18).

Remark 13.2.2. Dobrushin (1970) called ℓ_1 the Vasershtein (Wasserstein) distance. In our terminology ℓ_1 is the Kantorovich metric, see Example 3.2.2. The problem of estimating ℓ_1 from above also arises in connection with the sufficient conditions implying the uniqueness of the Gibbs random fields, see Dobrushin (1970, Sections 4 and 5).

By (13.2.2) we need to find precise estimates for ℓ_1 in the space $\mathscr{P}(\mathbb{R}^m)$ of all laws on $(\mathbb{R}^m, \|\cdot\|_1)$. The next two theorems provide such estimates and in certain cases even explicit representations of ℓ_1.

We suppose that $P_1, P_2 \in \mathscr{P}(\mathbb{R}^m)$ have densities p_1 and p_2, respectively.

Theorem 13.2.1. (i) The following inequality holds

$$\ell_1(P_1, P_2) \leqslant \alpha_1(P_1, P_2) \tag{13.2.4}$$

with

$$\alpha_1(P_1, P_2) := \int_{\mathbb{R}^m} \|x\|_1 \left| \int_0^1 t^{-m-1}(p_1 - p_2)(x/t) \, \mathrm{d}t \right| \mathrm{d}x.$$

(ii) If

$$\int_{\mathbb{R}_m} \|x\|_1 d(P_1 + P_2) < \infty \tag{13.2.5}$$

and if a continuous function $g: \mathbb{R}^m \to \mathbb{R}^1$ exists with derivatives $\partial g/\partial x_i$, $i = 1, \ldots, m$, defined almost everywhere (a.e.) and satisfying

$$\frac{\partial g}{\partial x_i}(x) = \operatorname{sgn}\left[x_i \int_0^1 t^{-m-1}(p_1 - p_2)(x/t)\,\mathrm{d}t\right] \text{a.e.} \quad i = 1, \ldots, m \tag{13.2.6}$$

then (13.2.4) holds with the equality sign.

Proof. (i) It is easy to see that the constraints set for

$$\ell_1(P_1, P_2) = \sup\left\{\left|\int_{\mathbb{R}^m} u\,\mathrm{d}(P_1 - P_2)\right| : u: \mathbb{R}^m \to \mathbb{R}, \text{ bounded} \right.$$
$$\left. |u(x) - u(y)| \leqslant \|x - y\|,\ x, y \in \mathbb{R}^m\right\} \tag{13.2.7}$$

coincides with the class of continuous bounded functions u ($u \in C_b(U)$) which have partial derivatives u_i' defined a.e. and satisfying the inequalities $|u_i'(x)| \leqslant 1$ a.e., $i = 1, \ldots, m$. Now, using the identity

$$u(x) = u(0) + \sum_{i=1}^m x_i \int_0^1 u_i'(tx)\,\mathrm{d}t$$

passing on from the coordinates t, x to the coordinates $t' = t$, $x' = tx$, and denoting these new coordinates again by t, x, one obtains

$$\ell_1(P_1, P_2) = \sup\left\{\left|\int_{\mathbb{R}^m} \sum_{i=1}^m u_i'(x)x_i\left(\int_0^1 t^{-m-1}(p_1 - p_2)(x/t)\,\mathrm{d}t\right)\mathrm{d}x\right| : u \in C_b(\mathbb{R}^m), \right.$$
$$\left. |u_i'| \leqslant 1, \ldots, |u_m'| \leqslant 1 \text{ a.e.}\right\}. \tag{13.2.8}$$

The estimate (13.2.4) follows obviously from here.

(ii) If the moment condition (13.2.5) holds, then, by Corollary 6.1.1,

$$\ell_1(P_1, P_2) = \sup\left\{\left|\int_{\mathbb{R}^m} u\,\mathrm{d}(P_1 - P_2)\right| : u \in C(\mathbb{R}^m), \right.$$
$$\left. |u(x) - u(y)| \leqslant \|x - y\| \quad \forall x, y \in \mathbb{R}^m\right\}$$
$$= \sup\left\{\left|\int_{\mathbb{R}^m} u\,\mathrm{d}(P_1 - P_2)\right| : u \in C(\mathbb{R}^m),\ |u_i'| \leqslant 1 \text{ a.e.}, i = 1, \ldots, m\right\} \tag{13.2.9}$$

where $C(\mathbb{R}^m)$ is the space of all continuous functions on $\mathbb{R}^m$.

Then, in (13.2.8), $C_b(\mathbb{R}^m)$ may also be replaced by $C(\mathbb{R}^m)$. It follows from (13.2.6) and (13.2.8) that the supremum in (13.2.8) is attained by $u = g$, and hence

$$\ell_1(P_1, P_2) = \boldsymbol{\alpha}_1(P_1, P_2). \tag{13.2.10}$$

The proof is complete. QED

Next we give simple sufficient conditions assuring the equality (13.2.10). Denote

$$J(P_1, P_2; x) := \int_0^1 t^{-m-1}(p_1 - p_2)(x/t)\, dt.$$

Corollary 13.2.1. If the moment condition (13.2.5) holds and $J(P_1, P_2; x) \geqslant 0$ a.e. or $J(P_1, P_2; x) \leqslant 0$ a.e., then the equality (13.2.10) takes place.

Proof. Indeed, one can take $g(x) := \|x\|_1$ in Theorem 3.2.1 (ii) if $J(P_1, P_2 \colon x) \geqslant 0$ a.e. and $g(x) = -\|x\|_1$ if $J(P_1, P_2; x) \leqslant 0$ a.e. QED

Remark 13.2.3. The inequality $J(P_1, P_2; x) \geqslant 0$ a.e. holds, for example, in the following cases:

(a) $0 < \underline{\lambda} \leqslant \bar{\lambda}$: $p_1(x) = \text{Veib}_{\underline{\lambda}}(x) := \prod_{i=1}^{m} \alpha_i \underline{\lambda}(\underline{\lambda} x_i)^{\alpha_i - 1} \exp(-(\underline{\lambda} x_i)^{\alpha_i}\ \alpha_i > 0$, and $p_2(x) = \text{Weib}_{\bar{\lambda}}(x)$;

(b) $0 < \underline{\lambda} \leqslant \bar{\lambda}$: $p_1(x) = \text{Gam}_{\underline{\lambda}}(x) := \prod_{i=1}^{m} \underline{\lambda}^{\alpha_i} x_i^{\alpha_i - 1}(\Gamma(\alpha_i))^{-1} \exp(-\underline{\lambda} x_i)$, $\alpha_i > 0$, and $p_2(x) = \text{Gam}_{\bar{\lambda}}(x)$;

(c) $\bar{\lambda} \geqslant \underline{\lambda} > 0$: $p_1(x) = \text{Norm}_{\bar{\lambda}}(x) := \prod_{i=1}^{m} (1/\bar{\lambda}\sqrt{2\pi}) \exp[-(x_i^2/2\bar{\lambda}^2)]$ and $p_2(x) = \text{Norm}_{\underline{\lambda}}(x)$.

Theorem 13.2.2. (i) The inequality

$$\ell_1(P_1, P_2) \leqslant \alpha_2(P_1, P_2) \tag{13.2.11}$$

holds with

$$\begin{aligned}\alpha_2(P_1, P_2) := &\int_{-\infty}^{\infty} \left| \int_{-\infty}^{t} q_1(x_{(1)})\, dx_1 \right| dt \\ &+ \sum_{i=2}^{m} \int_{\mathbb{R}^{i-1}} \left(\int_{-\infty}^{0} \left| \int_{-\infty}^{t} q_i(x_{(i)})\, dx_i \right| dt \right. \\ &\left. + \int_0^{\infty} \left| \int_t^{\infty} q_i(x_{(i)})\, dx_i \right| dt \right) dx_1 \cdots dx_{i-1}\end{aligned} \tag{13.2.12}$$

where

$$x_{(i)} := (x_1, \ldots, x_i),$$

$$q_i(x_{(i)}) := \int_{\mathbb{R}^{m-i}} (p_1 - p_2)(x_1, \ldots, x_m)\, \mathrm{d}x_{i+1}, \ldots, \mathrm{d}x_m \quad i = 1, \ldots, m-1,$$

$$q_m(x_{(m)}) := (p_1 - p_2)(x_1, \ldots, x_m).$$

(ii) If (13.2.5) holds and if a continuous function $h: \mathbb{R}^m \to \mathbb{R}^1$ exists with derivatives h_i', $i = 1, \ldots, m$, defined a.e. and satisfying the conditions

$$h_1'(t, 0, \ldots, 0) = \operatorname{sgn}[F_{11}(t) - F_{21}(t)]$$

$$h_2'(x_1, t, 0, \ldots, 0) = \begin{cases} \operatorname{sgn} \displaystyle\int_{-\infty}^{t} q_2(x_{(2)})\, \mathrm{d}x_2 & \text{if } t \in (-\infty, 0], x_1 \in \mathbb{R}^1 \\ -\operatorname{sgn} \displaystyle\int_{t}^{\infty} q_2(x_{(2)})\, \mathrm{d}x_2 & \text{if } t \in (0, +\infty), x_1 \in \mathbb{R}^1 \end{cases}$$

$$h_m'(x_1, \ldots, x_{m-1}, t) = \begin{cases} \operatorname{sgn} \displaystyle\int_{-\infty}^{t} q_m(x_{(m)})\, \mathrm{d}x_m & \text{if } t \in (-\infty, 0], x_1, \ldots, x_{m-1} \in \mathbb{R}^1 \\ -\operatorname{sgn} \displaystyle\int_{t}^{\infty} q_m(x_{(m)})\, \mathrm{d}x_m & \text{if } t \in (0, +\infty), x_1, \ldots, x_{m-1} \in \mathbb{R}^1 \end{cases}$$

then (13.2.11) holds with the equality sign. Here F_{ji} stands for the d.f. of the projection $(T_i P_j)$ of P_j over the ith coordinate.

Proof. (i) Using the obvious formulae

$$q_i(x_{(i)}) = \int_{-\infty}^{\infty} q_{i+1}(x_{(i+1)})\, \mathrm{d}x_{i+1} \quad i = 1, \ldots, m-1$$

$$\int_{-\infty}^{\infty} q_1(x_{(1)})\, \mathrm{d}x_1 = \int_{\mathbb{R}^m} (p_1 - p_2)(x)\, \mathrm{d}x = 0$$

and applying repeatedly the identity

$$\int_{-\infty}^{\infty} a(t)b(t)\, \mathrm{d}t = \int_{-\infty}^{\infty} a(0)b(t)\, \mathrm{d}t - \int_{-\infty}^{0} a'(t)\left(\int_{-\infty}^{t} b(s)\, \mathrm{d}s\right) \mathrm{d}t + \int_{0}^{\infty} a'(t)\left(\int_{t}^{\infty} b(s)\, \mathrm{d}s\right) \mathrm{d}t$$

for $a(t) = u(x_1, \ldots, x_{i-1}, t, \ldots, 0)$, $b(t) = q_i(x_1, \ldots, x_{i-1}, t)$, $i = 1, \ldots, m$, one obtains

$$\ell_1(P_1, P_2) = \sup\left\{\left| -\int_{-\infty}^{\infty} u_1'(t, 0, \ldots, 0) \int_{-\infty}^{t} q_1(x_{(1)})\, \mathrm{d}x_1\, \mathrm{d}t \right.\right.$$
$$+ \sum_{i=2}^{m} \int_{\mathbb{R}^{i-1}} \left(-\int_{-\infty}^{0} u_i'(x_1, \ldots, x_{i-1}, t, \ldots, 0) \int_{-\infty}^{t} q_i(x_{(i)})\, \mathrm{d}x_i\, \mathrm{d}t \right.$$
$$\left.\left. + \int_{0}^{\infty} u_i'(x_1, \ldots, x_{i-1}, t, \ldots, 0) \int_{t}^{\infty} q_i(x_{(i)})\, \mathrm{d}x_i\, \mathrm{d}t \right) \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \right| :$$
$$\left. u \in C_b(\mathbb{R}^m), |u_1'| \leqslant 1, \ldots, |u_m'| \leqslant 1 \text{ a.e.} \right\} \quad (13.2.13)$$

which obviously implies (13.2.11).

(ii) In view of (13.2.5), $C_b(\mathbb{R}^m)$ in (13.2.13) may be replaced by $C(\mathbb{R}^m)$. Then the function $u = h$ yields the supremum in the right-hand side of (13.2.13), and hence

$$\ell_1(P_1, P_2) = \alpha_2(P_1, P_2). \quad (13.2.14)$$

QED

Remark 13.2.4. The bounds (13.2.4) and (13.2.14) are of interest by themselves. They give two improvements of the following bound (Zolotarev, 1986, Section 1.5):

$$\ell_1(P_1, P_2) \leqslant \nu(P_1, P_2)$$

where

$$\nu(P_1, P_2) := \int_{\mathbb{R}^m} \|x\|_1 |p_1(x) - p_2(x)|\, \mathrm{d}x$$

is the first absolute pseudomoment. Indeed, one can easily check that

$$\alpha_i(P_1, P_2) \leqslant \nu(P_1, P_2) \quad i = 1, 2.$$

Remark 13.2.5. Consider the sth-difference pseudomoment (see Section 4.3, Case D).

$$\boldsymbol{\kappa}_s(P_1, P_2) = \sup\left\{ \left| \int_{\mathbb{R}^m} u\, \mathrm{d}(P_1 - P_2) \right| : u: \mathbb{R}^m \to \mathbb{R}^1, \right.$$
$$\left. |u(x) - u(y)| \leqslant d_s(x, y) \right\} \quad s > 0 \quad (13.2.15)$$

where

$$d_s(x, y) := \|\mathscr{Q}_s(x) - \mathscr{Q}_s(y)\| \quad \mathscr{Q}_s: \mathbb{R}^m \to \mathbb{R}^m \quad (13.2.16)$$
$$\mathscr{Q}_s(t) := t\|t\|^{s-1}.$$

Since

$$\boldsymbol{\kappa}_s(P_1, P_2) = \ell_1(P_1 \circ \mathcal{Q}_s^{-1}, P_2 \circ \mathcal{Q}_s^{-1}) \tag{13.2.17}$$

then by (13.2.4) and (13.2.11) we obtain the bounds

$$\boldsymbol{\kappa}_s(P_1, P_2) \leqslant \alpha_i(P_1 \circ \mathcal{Q}_s^{-1}, P_2 \circ \mathcal{Q}_s^{-1}) \quad i = 1, 2 \tag{13.2.18}$$

which are better than the following one (see Zolotarev 1986, Section 1.5, (1.5.37))

$$\boldsymbol{\kappa}_s(P_1, P_2) \leqslant \nu(P_1 \circ \mathcal{Q}_s^{-1}, P_2 \circ \mathcal{Q}_s^{-1}). \tag{13.2.19}$$

PART IV

Ideal Metrics

Any concrete stochastic approximation problem requires 'appropriate' or 'natural' metric (topology, convergence, uniformities, etc.) having properties which are helpful in solving the problem. If one needs to develop the solution of the approximation problem in terms of other metrics (topology, etc.) the transition is carried out by using general relationships between metrics (topologies, etc.). This two-stage approach (selection of the appropriate metric and comparison of metrics) is the basis of the theory of probability metrics, see Fig. 1.1.1.

In this part we shall determine the structure of 'appropriate' ('ideal') p. distances for various probabilistic problems. The fact that a certain metric is (or is not) appropriate depends on the concrete approximation (or stability) problem we deal with, i.e. any particular approximation problem has its own 'ideal' p. distance (or distances) on which terms we can solve the problem in the most 'natural' way. We shall study the structure of such 'ideal' metrics in various stochastic approximation problems, like stability of characterizations of probability distributions, rate of convergence for the sums and maxima of random variables, convergence of random motions and stability in risk theory.

CHAPTER 14

Ideal Metrics with Respect to Summation Scheme for i.i.d. Random Variables

Since there is not a satisfactory definition of an 'ideal' ('natural') metric, we shall use different examples to explain this approach. The first illustrative example deals with approximation of exponential families of distributions.

14.1 ROBUSTNESS OF χ^2-TEST OF EXPONENTIALITY

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

(A) *The case when X's belong to an aging class distribution.* A non-negative random variable X with distribution function F is said to be HNBUE (*harmonic new better than used in expectation*) if $\int_x^\infty \bar{F}(u)\,du \leqslant a\exp(-x/a)$ for all $x \geqslant 0$ where $a = \mathbb{E}(X)$ and $\bar{F} = 1 - F$. It is easily seen that if X is HNBUE, moments of all orders exist. Similarly, X is said to be HNWUE (*harmonic new worse than used in expectation*) if $\int_x^\infty \bar{F}(u)\,du \geqslant a\exp(-x/a)$ for all $x \geqslant 0$ assuming that a is finite. The class of HNBUE (HNWUE) distributions include all the standard 'aging' ('anti-aging') classes—IFR, IFRA, NBU and NBUE (DFR, DFRA, NWU and NWUE); see Barlow and Proschan (1975), Chapter 4, and Kalashnikov and Rachev (1988), Chapter 4, for the necessary definitions.

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

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

(b) *μ satisfies the inequalities*

$$\phi_1(|a-\sigma|) \leqslant \mu(X, E(a)) \leqslant \phi_2(|a-\sigma|) \quad (X \in \text{HNBUE}, \mathbb{E}X = a, \sigma^2 = \operatorname{var} X)$$

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

Clearly, the most appropriate metric μ should satisfy (a) and (b) with $\phi_1 \equiv \phi_2$. Such a metric is the so-called Zolotarev ζ_2-metric

$$\zeta_2(X, Y) := \zeta_2(F_X, F_Y) = \sup_{f \in \mathbb{F}_2} |\mathbb{E}(f(X) - f(Y))| \quad \mathbb{E}X^2 < \infty, \mathbb{E}Y^2 < \infty \tag{14.1.1}$$

where $\mathbb{F}_2$ is the class of all functions f having almost everywhere second derivative f'' and $|f''| \leqslant 1$ a.e. To check (a) and (b) for $\mu = \zeta_2$ first notice that the finiteness of ζ_2 implies $\infty > \zeta_2(X, Y) \geqslant \sup_{a>0} |\mathbb{E}(aX) - \mathbb{E}(aY)|$, i.e. $\mathbb{E}X = \mathbb{E}Y$. Secondly, if $\mathbb{E}X = \mathbb{E}Y$ then $\zeta_2(X, Y)$ admits second representation

$$\zeta_2(X, Y) = \int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} (F_X(t) - F_Y(t))\, dt \right| dx. \tag{14.1.2}$$

In fact, by Taylor's theorem, or integrating by parts

$$\begin{aligned}\zeta_2(X, Y) &= \sup_{f \in \mathbb{F}_2} \left| \int_{-\infty}^{\infty} f(t)\, d(F_X(t) - F_Y(t)) \right| \\ &= \sup_{f \in \mathbb{F}_2} \left| \int_{-\infty}^{\infty} f''(x) \int_{-\infty}^{x} (F_X(t) - F_Y(t))\, dt \right| dx.\end{aligned}$$

Now use the isometric isomorphism between L_1^*- and L_∞-spaces (see, for example, Dunford and Schwartz (1988) Theorem IV.8.3.5) in order to obtain the equality (14.1.2).

In the next lemma, using both representations for ζ_2, we show that $\mu = \zeta_2$ satisfies (a).

Lemma 14.1.1. (i) In the space $\mathfrak{X}^2(\mathbb{R})$ of all square integrable random variables

$$\tfrac{1}{2} |\mathbb{E}X^2 - \mathbb{E}Y^2| \leqslant \zeta_2(X, Y) \tag{14.1.3}$$

and

$$\mathrm{L}(X, Y) \leqslant [4\zeta_2(X, Y)]^{1/3} \tag{14.1.4}$$

where $\mathbf{L}$ is the Lévy metric (2.1.3). In particular, if $X_n, X \in \mathfrak{X}^2(\mathbb{R})$ then

$$\zeta_2(X_n, X) \to 0 \Rightarrow \begin{cases} X_n \to X \text{ in distribution} \\ \mathbb{E}X_n^2 \to \mathbb{E}X^2. \end{cases} \tag{14.1.5}$$

(ii) Given $X_0 \in \mathfrak{X}^2(\mathbb{R})$ let $\mathfrak{X}^2(\mathbb{R}, X_0)$ be the space of all $X \in \mathfrak{X}^2(\mathbb{R})$ with $\mathbb{E}X = \mathbb{E}X_0$. Then for any $X, Y \in \mathfrak{X}^2(\mathbb{R}, X_0)$

$$2\zeta_2(X, Y) \leqslant \boldsymbol{\kappa}_2(X, Y) \tag{14.1.6}$$

where $\boldsymbol{\kappa}_2$ is the second pseudomoment

$$\boldsymbol{\kappa}_2(X, Y) = 2\int_{-\infty}^{\infty} |x||F_X(x) - F_Y(x)|\,\mathrm{d}x. \tag{14.1.7}$$

In particular, for $X_n, X \in \mathfrak{X}^2(\mathbb{R}, X_0)$

$$\begin{cases} X_n \to X \text{ in distribution} \\ \mathbb{E}X_n^2 \to \mathbb{E}X^2 \end{cases} \Rightarrow \zeta_2(X_n, X) \to 0. \tag{14.1.8}$$

Proof. (i) Clearly, the representation (14.1.1) implies (14.1.3). To prove (14.1.4), let $\mathbf{L}(X, Y) > \varepsilon > 0$; then there exists $z \in \mathbb{R}$ such that either

$$F_X(z) - F_Y(z + \varepsilon) > \varepsilon \tag{14.1.9}$$

or $F_Y(z) - F_X(z + \varepsilon) > \varepsilon$. Suppose (14.1.9) holds; then put

$$f_0\left(z + \frac{\varepsilon}{2} + h\right) := \left\{\left[\left(1 - \frac{2|h|}{\varepsilon}\right)_+\right]^2 - 1\right\}\operatorname{sgn} h \tag{14.1.10}$$

where $(\cdot)_+ = \max(0, \cdot)$. Then $f_0(x) = 1$ for $x \leqslant z$, $= -1$ for $x > z + \varepsilon$ and $|f_0| \leqslant 1$. Since $\|f_0''\|_\infty := \operatorname{ess\,sup}|f''(x)| = 8\varepsilon^{-2}$, we have

$$\begin{aligned}
\zeta_2(X, Y) &\geqslant \|f_0''\|_\infty^{-1}\left|\int (f_0(x) + 1)\,\mathrm{d}(F_X(x) - F_Y(x))\right| \\
&\geq (\varepsilon^2/8)\left(\int_{-\infty}^{z} (f_0(x) + 1)\,\mathrm{d}F_X(x) - \int_{z+\varepsilon}^{\infty} (f_0(x) + 1)\,\mathrm{d}F_Y(x)\right) \geqslant \varepsilon^2/4.
\end{aligned}$$

Letting $\varepsilon \to \mathbf{L}(X, Y)$ implies (14.1.4).

(ii) Using the representation (14.1.2) and $\mathbb{E}X = \mathbb{E}Y$, one obtains (14.1.6). Clearly

$$\boldsymbol{\kappa}_2(X, Y) = \ell_1(X|X|, Y|Y|) \tag{14.1.11}$$

where ℓ_1 is the Kantorovich metric $\ell_1(X, Y) = \int_{-\infty}^{\infty} |F_X(x) - F_Y(x)|\,\mathrm{d}x$, see also (4.3.39), (13.2.17). For any X_n and X with $\mathbb{E}|X_n| + \mathbb{E}|X| < \infty$ we have by

Theorem 6.2.1, that

$$\ell_1(X_n, X) \to 0 \Leftrightarrow \begin{cases} X_n \to X \text{ in distribution} \\ \mathbb{E}|X_n| \to \mathbb{E}|X| \end{cases} \tag{14.1.12}$$

which together with (14.1.11) completes the proof of (ii). QED

Thus ζ_2-convergence preserves the convergence in distribution plus convergence of the second moments and so the requirement (a) holds. Concerning the property (b), we use the second representation of ζ_2, (14.1.2), to get

$$\begin{aligned} \zeta_2(X, Y) &= \int_0^\infty \left| \int_x^\infty \bar{F}_X(t)\,\mathrm{d}t - a\ \exp(-x/a) \right| \mathrm{d}x \\ &= \int_0^\infty \left(a \exp(-x/a) - \int_x^\infty \bar{F}_X(t)\,\mathrm{d}t \right) \mathrm{d}x \\ &= \frac{1}{2}(a^2 - \sigma^2) \quad \text{for } X \text{ being HNBUE, } Y := E(a). \end{aligned} \tag{14.1.13}$$

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

$$\boldsymbol{\rho}(X, Y) := \sup_{x \in \mathbb{R}} \| F_X(x) - F_Y(x)| \tag{14.1.14}$$

then one simply compares ζ_2 with $\boldsymbol{\rho}$. Namely, by the well known inequality between the Lévy distance $\mathbf{L}$ and the Kolmogorov distance $\boldsymbol{\rho}$, we have

$$\boldsymbol{\rho}(X, Y) \leqslant \left[1 + \sup_x f_X(x) \right] \mathbf{L}(X, Y) \tag{14.1.15}$$

if $f_X = F_X'$ exists. Thus, by (14.1.4) and (14.1.15)

$$\begin{aligned} \boldsymbol{\rho}(X, Y) &\leqslant \left[1 + \sup_t f_{cX}(t) \right] \left[4\zeta_2(cX, cY) \right]^{1/3} \\ &= (c^{2/3} + M_X c^{-1/3})[4\zeta_2(X, Y)]^{1/3} \quad \text{for any } c > 0 \end{aligned}$$

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

$$\boldsymbol{\rho}(X, Y) \leqslant 3M_X^{2/3}(\zeta_2(X, Y))^{1/3} \tag{14.1.16}$$

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

$$\boldsymbol{\rho}(X, E(a)) \leqslant 3(\alpha/2)^{1/3} \quad \alpha = 1 - \sigma^2/a^2. \tag{14.1.17}$$

Remark 14.1.1. Note that the order 1/3 of α is precise, see Daley (1988), for an appropriate example.

Next, using the 'natural' metric ζ_2, we derive a bound on the uniform distance between the χ^2_{2n} distribution and the distribution of $2\sum_{i=1}^n X_i/a$, assuming that X is HNBUE. Define $\bar{X}_i = (X_i - a)/a$ and $\bar{Y}_i = (Y_i - a)/a$ $(i = 1, 2, \ldots, n)$ and write $W_n = 2\sum_{i=1}^n X_i/a$, $\bar{W}_n = \sum_{i=1}^n \bar{X}_i/\sqrt{n}$, $Z_n = 2\sum_{i=1}^n Y_i/a$ and $\bar{Z}_n = \sum_{i=1}^n \bar{Y}_i/\sqrt{n}$. Let $f_{\bar{Z}_n}$ denote the p.d.f. of $\bar{Z}_n$ and let $M_n = \sup_x f_{\bar{Z}_n}(x)$. Then by (14.1.16)

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

Now we use the fact that ζ_2 is *the ideal metric of order* 2 (see further Section 14.2), i.e., *for any vectors* $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ *with independent components and constants* $c_1, \ldots, c_n$

$$\zeta_2\left(\sum_{i=1}^n c_i X_i, \sum_{i=1}^n c_i Y_i\right) \leqslant \sum_{i=1}^n |c_i|^2 \zeta_2(X_i, Y_i) \tag{14.1.18}$$

Remark 14.1.2. Since ζ_2 is a simple metric, without loss of generality, we may assume that $\{X_i\}$ and $\{Y_i\}$ are independent. Then (14.1.18) follows from single induction arguments, triangle inequality and the following two properties: for any independent X, Y and Z, and any $c \in \mathbb{R}$

$$\mu(X + Z, Y + Z) \leqslant \mu(X, Y) \quad \text{(regularity)} \tag{14.1.19}$$

and

$$\mu(cX, cY) = c^2\mu(X, Y) \quad \text{(homogeneity of order 2)} \tag{14.1.20}$$

cf. further Definition 14.2.1. Thus, by (14.1.18), $\zeta_2(\bar{W}_n, \bar{Z}_n) \leqslant \zeta_2(X, Y)/a^2$, and finally the required estimate is

$$\rho(W_n, Z_n) \leqslant \frac{3}{2^{1/3}} M_n^{2/3}[1 - (\sigma/a)^2]^{1/3} \tag{14.1.21}$$

and it is a straightforward matter to show that

$$M_n = \frac{\sqrt{n}(n-1)^{n-1}\exp[-(n-1)]}{(n-1)!}. \tag{14.1.22}$$

Expression (14.1.22) may be simplified by using the Robbins–Stirling inequality

$$n^n \mathrm{e}^{-n}(2\pi n)^{1/2} \exp[1/(12n + 1)] < n! < n^n \mathrm{e}^{-n}(2\pi n)^{1/2} \exp(1/12n)$$

(e.g., Erdös and Spencer 1974, page 17) to give the following simple bound

$$M_n < \left(\frac{n}{2\pi(n-1)}\right)^{1/2} \exp[-1/(12n - 11)]. \tag{14.1.23}$$

Further, if X is HNWUE, a similar calculation shows that $\zeta_2(X, Y) = \frac{1}{2}(\sigma^2 - a^2)$, assuming that σ^2 is finite. In summary, we have shown that if X is

HNBUE or HNWUE, then

$$\rho(W_n, Z_n) \leqslant \frac{3}{2^{1/3}} M_n^{2/3} |1 - (\sigma/a)^2|^{1/3} \tag{14.1.24}$$

where M_n can be estimated by (14.1.23).

It follows from (14.1.22) that if X is HNBUE or HNWUE, and if the coefficient of variation of X is close to unity, then the distribution of $2\sum_{i=1}^{n} X_i/a$ is uniformly close to the χ^2_{2n} distribution.

(A) *The case where X is arbitrary. Contamination by mixture.* In practice a 'perturbation' of an exponential random variable does not necessarily yield an HNBUE or HNWUE variable, in which case the bound (14.1.24) will not hold. If we make no assumptions concerning X, it is necessary to make an assumption concerning the 'mechanism' by which the exponential distribution is 'perturbed'. Further, we shall deduce bounds for the three most common possible 'mechanisms': contamination by mixture, contamination by an additive error and right-censoring.

Suppose that an exponential random variable is contaminated by an arbitrary non-negative random variable with distribution function H, i.e. $\bar{F}_X(t) = (1-\varepsilon)\exp(-t/\lambda) + \varepsilon\bar{H}(t)$. Then $a = (1-\varepsilon)\lambda + \varepsilon h$ where $h = \int_0^\infty t\,\mathrm{d}H(t)$. It is assumed that $\varepsilon > 0$ is small. Now since $Y = E(a)$

$$\begin{aligned}
\zeta_2(X, Y) &= \int_0^\infty \left| \int_x^\infty [\bar{F}_X(t) - \exp(-t/a)]\,\mathrm{d}t \right| \mathrm{d}x \leqslant \int_0^\infty t|\bar{F}_X(t) - \exp(-t/a)|\,\mathrm{d}t \\
&\leqslant (1-\varepsilon)\int_0^\infty t|\exp(-t/\lambda) - \exp(-t/a)|\,\mathrm{d}t \\
&\quad + \varepsilon \int_0^\infty t|\bar{H}(t) - \exp(-t/a)|\,\mathrm{d}t \\
&\leqslant (1-\varepsilon)|\lambda_2 - a^2| + \varepsilon(b/2 + a^2) \quad \left(\text{where } b = \int_0^\infty t^2\,\mathrm{d}H(t)\right) \\
&= \varepsilon[|h-\lambda|(\lambda + a) + b/2 + a^2].
\end{aligned}$$

Thus $\zeta_2(X, Y) = \mathrm{O}(\varepsilon)$ and so, from (14.1.16), it follows that $\rho(W_n, Z_n) = \mathrm{O}(\varepsilon^{1/3})$.

(B) *Contamination by additive error.* Suppose now that an exponential random variable is contaminated by an arbitrary additive error, i.e. $X \overset{\mathrm{d}}{=} Y_\lambda + V$, V is an arbitrary random variable and Y_λ is an exponential random variable independent of V with mean $\lambda = a - \mathbb{E}(V)$. Consider the metric $\boldsymbol{\kappa}_2$ (14.1.7). For any $N > 0$ simply estimate $\boldsymbol{\kappa}_2$, by the Kantorovich metric ℓ_1,

$$\begin{aligned}
\tfrac{1}{2}\boldsymbol{\kappa}_2(X, Y) &= \int |t||F_X(t) - F_Y(t)|\,\mathrm{d}t \\
&\leqslant N\ell_1(X, Y) + N^{-\delta}[\mathbb{E}(|X|^{2+\delta}) + \mathbb{E}(|Y|^{2+\delta})]
\end{aligned}$$

and hence the least upper bound of $\boldsymbol{\kappa}_2(X, Y)$ obtained by varying N is

$$\boldsymbol{\kappa}_2(X, Y) \leqslant 2(1 + 1/\delta)[\ell_1(X, Y)]^{\delta/(1+\delta)}(\delta\beta)^{1/(1+\delta)} \tag{14.1.25}$$

where $\beta = \mathbb{E}(|X|^{2+\delta}) + \mathbb{E}(|Y|^{2+\delta})$. By the triangle inequality

$$\ell_1(X, Y) = \ell_1(Y_\lambda + V, Y_a) \leqslant \ell_1(Y_\lambda + V, Y_\lambda) + \ell_1(Y_\lambda, Y_a)$$
$$\leqslant \ell_1(V, 0) + \int_0^\infty |\exp(-x/\lambda) - \exp(-x/a)|\,\mathrm{d}x = \mathbb{E}|V| + |\mathbb{E}V| \leqslant 2\mathbb{E}|V|. \tag{14.1.26}$$

It follows from (14.1.25) and (14.1.26) that

$$\boldsymbol{\kappa}_2(X, Y) \leqslant 2(1 + 1/\delta)[2\mathbb{E}(|V|)]^{\delta/(1+\delta)}(\delta\beta)^{1/(1+\delta)}. \tag{14.1.27}$$

Clearly from (14.1.27), we see that if $\mathbb{E}|V|$ is close to zero, then $\boldsymbol{\kappa}_2(X, Y)$ is small. But $\boldsymbol{\kappa}_2(X, Y) \geqslant 2\zeta_2(X, Y)$ (cf. (14.1.6)) and so, from (14.1.16) it follows that if $\mathbb{E}|V|$ is small, the uniform distance between the distribution of $2\sum_{i=1}^n X_i/a$ and the χ^2_{2n} distribution is small.

(C) *Right-censoring*. Suppose, lastly, that $X = Y_\lambda \wedge N$ where N is a non-negative random variable independent of $Y_\lambda \stackrel{\mathrm{d}}{=} \mathbf{E}(\lambda)$ so that $a = \mathbb{E}(Y_\lambda \wedge N)$. Now, for $\eta > 0$

$$\zeta_2(X, Y) \leqslant \tfrac{1}{2}\boldsymbol{\kappa}_2(X, Y) = \tfrac{1}{2}\boldsymbol{\kappa}_2(Y_\lambda \wedge N, Y_a)$$
$$= \int_0^\eta t|\exp(-t/a) - \exp(-t/\lambda)\bar{F}_N(t)|\,\mathrm{d}t$$
$$+ \int_\eta^\infty t|\exp(-t/a) - \exp(-t/\lambda)\bar{F}_N(t)|\,\mathrm{d}t.$$

It can easily be shown that

$$\int_0^\eta t|\exp(-t/a) - \exp(-t/\lambda)\bar{F}_N(t)|\,\mathrm{d}t \leqslant |\lambda^2 - a^2| + \lambda^2 F_N(\eta)$$

and that

$$\int_\eta^\infty t|\exp(-t/a) - \exp(-t/\lambda)\bar{F}_N(t)|\,\mathrm{d}t \leqslant a(\eta + a)\exp(-\eta/a) + \lambda(\eta + \lambda)\exp(-\eta/\lambda).$$

Hence, for any $\eta > 0$,

$$2\zeta_2(X, Y) \leqslant |\lambda^2 - a^2| + \lambda^2 F_N(\eta) + 2\gamma(\eta + \gamma)\exp(-\eta/\gamma),\ \gamma := \max(a, \lambda). \tag{14.1.28}$$

For fixed η the value of $F_N(\eta)$ is small if N is big enough. Thus (14.1.28) together with (14.1.16) gives an estimate of $\zeta_2(X, Y)$ as $N \xrightarrow{\mathrm{d}} \infty$. Finally, by $\zeta_2(\bar{W}_n, \bar{Z}_n) \leqslant \zeta_2(X, Y)/a^2$ and $\boldsymbol{\rho}(W_n, Z_n) = \boldsymbol{\rho}(\bar{W}_n, \bar{Z}_n) \leqslant 3M_n^{2/3}[\zeta_2(\bar{W}_n, \bar{Z}_n)]^{1/3}$, it follows that the distribution of $2\sum_{i=1}^{n} X_i/a$ is uniformly close to the χ^2_{2n} distribution.

The derivation of the estimates for $\boldsymbol{\rho}(W_n, Z_n)$ is just an illustrative example of how one can use the theory of probability metrics. Clearly, in this simple case one can get similar results by traditional methods. However, in order to study stability of characterization of multivariate distributions, rapidity in the multivariate CLT and other approximation type stochastic problems one should use the general relationships between probability distances which will considerably simplify the task.

14.2 IDEAL METRICS FOR SUMS OF INDEPENDENT RANDOM VARIABLES

Let $(U, \|\cdot\|)$ be a complete separable Banach space equipped with the usual algebra of Borel sets $\mathscr{B}(U)$ and let $\mathfrak{X} := \mathfrak{X}(U)$ be the vector space of all random variables defined on a probability space $(\Omega, \mathscr{A}, \Pr)$ and taking values in U. We will choose to work with simple probability metrics on the space $\mathfrak{X}$ instead of the space $\mathscr{P}(U)$, cf. Section 2.3 and 3.2. We will show that certain 'convolution' metrics on $\mathfrak{X}$ may be used to provide exact rates of convergence of normalized sums to a stable limit law. They will play the role of 'ideal metrics' for the approximation problems under consideration. 'Traditional' metrics for the rate of convergence in the CLT are uniform type metrics. Having exact estimates in terms of the 'ideal' metrics we shall pass to the uniform estimates by using the Bergström convolution method. The rates of convergence, which hold uniformly in n will be expressed in terms of a variety of uniform metrics on $\mathfrak{X}$.

Definition 14.2.1 (Zolotarev). A p. semimetric $\mu: \mathfrak{X} \times \mathfrak{X} \to [0, \infty]$ is called an *ideal (probability) metric* of order $r \in \mathbb{R}$ if for any random variables X_1, X_2, $Z \in \mathfrak{X}$ and any non-zero constant c the following two properties are satisfied

(i) *Regularity:* $\mu(X_1 + Z, X_2 + Z) \leqslant \mu(X_1 X_2)$, and
(ii) *Homogeneity of order r:* $\mu(cX_1, cX_2) = |c|^r \mu(X_1, X_2)$.

When μ is a simple metric (see Section 3.2), i.e., its values are determined by the marginal distributions of the random variables being compared, then it is assumed in addition that the random variable Z is independent of X_1 and X_2 in condition (i). All metrics μ in this section are simple.

Remark 14.2.1. Zolotarev (1976 b,d) (cf. Zolotarev 1986, Chapter 1) showed the existence of an ideal metric of a given order $r \geqslant 0$ and he defined the ideal metric

$$\zeta_r(X_1, X_2) := \sup\{|\mathbb{E}(f(X_1) - f(X_2))| : |f^{(m)}(x) - f^{(m)}(y)| \leqslant \|x - y\|^\beta\} \quad (14.2.1)$$

where $m = 0, 1, \ldots$ and $\beta \in (0, 1]$ satisfy $m + \beta = r$ and $f^{(m)}$ denotes the mth

Fréchet derivative of f for $m \geqslant 0$ and $f^{(0)}(x) = f(x)$. He also obtained an upper bound for ζ_k (k integer) in terms of the *difference pseudomoment* κ_k where for $r > 0$

$$\kappa_r(X_1, X_2) := \sup\{|\mathbb{E}(f(X_1) - f(X_2))| : |f(x) - f(y)| \leqslant \|x\|x\|^{r-1} - y\|y\|^{r-1}\|\}$$

(see (4.3.40) and (4.3.42)). If $U = \mathbb{R}$, $\|x\| = |x|$, then (see (4.3.43))

$$\kappa_r(X_1, X_2) := r \int |x|^{r-1} |F_{X_1}(x) - F_{X_2}(x)| \, dx \quad r > 0 \tag{14.2.2}$$

where F_X denotes the distribution function for X.

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

Remark 14.2.2. Further in this and the next section for each $X_1, X_2 \in \mathfrak{X}$ we write $X_1 + X_2$ to mean the sum of independent random variables with laws $\Pr_{X_1}$ and $\Pr_{X_2}$ respectively. For any $X \in \mathfrak{X}$, p_X denotes the density of X if it exists. We reserve the letter Y_α (or Y) to denote a *symmetric stable random variable* with parameter $\alpha \in (0, 2]$, i.e. $Y_\alpha \overset{d}{=} -Y_\alpha$ and for any $n = 1, 2, \ldots, X'_1 + \cdots + X'_n \overset{d}{=} n^{1/\alpha} Y_\alpha$, where $X'_1, X'_2, \ldots, X'_n$ are i.i.d. random variables with the same distribution as Y_α. If $Y_\alpha \in \mathfrak{X}(\mathbb{R})$ we assume that Y_α has characteristic function

$$\phi_Y(t) = \exp\{-|t|^\alpha\} \quad t \in \mathbb{R}.$$

For any $f: U \to \mathbb{R}$,

$$\|f\|_L := \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|}$$

denotes the Lipschitz norm of f, $\|f\|_\infty$ the essential supremum of f, and when $U = \mathbb{R}^k$, $\|f\|_p$ denotes the L^p norm,

$$\|f\|_p^p := \int_{\mathbb{R}^k} |f(x)|^p \, dx, \quad p \geqslant 1.$$

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

$$\frac{X_1 + \cdots + X_n}{n^{1/\alpha}} \xrightarrow{w} Y_\alpha \tag{14.2.3}$$

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

Total variation metrics

$$\begin{aligned}\boldsymbol{\sigma}(X_1, X_2) := \sup_{A \in \mathscr{B}(U)} &|\Pr\{X_1 \in A\} - \Pr\{X_2 \in A\}|,\\
:= \sup\{&|\mathbb{E}f(X_1) - \mathbb{E}f(X_2)| : f\colon U \to \mathbb{R} \text{ is measurable and}\\
&\text{for any } x, y \in B, |f(x) - f(y)| \leqslant \mathbb{I}(x, y) \text{ where}\\
&\mathbb{I}(x, y) = 1 \text{ if } x \neq y \text{ and } 0 \text{ otherwise}\}, X_1, X_2 \in \mathfrak{X}(U)\end{aligned} \qquad (14.2.4)$$

(see Lemma 3.2.1, (3.3.18) and (3.2.13)) and

$$\begin{aligned}\mathbf{Var}(X_1, X_2) &:= \sup\{|\mathbb{E}f(X_1) - \mathbb{E}f(X_2)| : f\colon U \to \mathbb{R} \text{ is measurable and } \|f\|_\infty \leqslant 1\}\\
&= 2\boldsymbol{\sigma}(X_1, X_2),\ X_1, X_2 \in \mathfrak{X}(U).\end{aligned} \qquad (14.2.5)$$

In $\mathfrak{X}(\mathbb{R}^n)$ we have $\mathbf{Var}(X_1, X_2) := \int |\mathrm{d}(F_{X_1} - F_{X_2})|$.

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

$$\ell(X_1, X_2) := \operatorname*{ess\,sup}_x |p_{X_1}(x) - p_{X_2}(x)|. \qquad (14.2.6)$$

Uniform metric between characteristic functions

$$\boldsymbol{\chi}(X_1, X_2) := \sup_{t \in \mathbb{R}} |\phi_{X_1}(t) - \phi_{X_2}(t)| \quad X_1, X_2 \in \mathfrak{X}(\mathbb{R}) \qquad (14.2.7)$$

where ϕ_X denotes the characteristic function of X. The metric $\boldsymbol{\chi}$ is topologically weaker than **Var**, which is itself topologically weaker than ℓ by Scheffe''s theorem, see Billingsley (1968, p. 224).

We will use the following simple metrics on $\mathfrak{X}(\mathbb{R})$.

Kolmogorov metric

$$\boldsymbol{\rho}(X_1, X_2) := \sup_{x \in \mathbb{R}} |F_{X_1}(x) - F_{X_2}(x)|. \qquad (14.2.8)$$

Weighted χ-metric

$$\boldsymbol{\chi}_r(X_1, X_2) := \sup_{t \in \mathbb{R}} |t|^{-r} |\phi_{X_1}(t) - \phi_{X_2}(t)|. \qquad (14.2.9)$$

L^p-version of ζ_m

$$\begin{aligned}\zeta_{m,p}(X_1, X_2) &:= \sup\{|\mathbb{E}f(X_1) - f(X_2)| : \|f^{(m+1)}\|_q \leqslant 1\}\\
&\qquad 1/p + 1/q = 1 \quad m = 0, 1, 2, \ldots.\end{aligned} \qquad (14.2.10)$$

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

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

see Kalashnikov and Rachev (1988), Chapter 3, Section 8.2 and further Lemma 17.1.1.

Kantorovich ℓ_p-metric:

$$\ell_p^p(X_1, X_2) := \sup\left\{ \int f \mathrm{d}F_{X_1} + \int g \, \mathrm{d}F_{X_2} : \|f\|_\infty + \|f\|_L \leqslant \infty \right.$$
$$\left. \|g\|_\infty + \|g\|_L < \infty, f(x) + g(y) \leqslant \|x - y\|^p \quad \forall x, y \in \mathbb{R} \right\}, p \geqslant 1 \tag{14.2.11}$$

(see (3.2.11) and (3.3.18)).

Now we define the ideal metrics of order $r - 1$ and r, respectively.

Let $\theta \in \mathfrak{X}(\mathbb{R}^k)$, $\theta \stackrel{\mathrm{d}}{=} -\theta$ and define for every $r > 0$ the *convolution* (probability) metric

$$\boldsymbol{\mu}_{\theta,r}(X_1, X_2) := \sup_{h \in \mathbb{R}} |h|^r \ell(X_1 + h\theta, X_2 + h\theta) \quad X_1, X_2 \in \mathfrak{X}(\mathbb{R}^k). \tag{14.2.12}$$

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

$$\boldsymbol{\nu}_{\theta,r}(X_1, X_2) := \sup_{h \in \mathbb{R}} |h|^r \, \mathbf{Var}(X_1 + h\theta, X_2 + h\theta) \quad X_1, X_2 \in \mathfrak{X}(U). \tag{14.2.13}$$

Lemmas 14.2.1 and 14.2.2 below show that $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$ are ideal of order $r - 1$ and r, respectively. In general, $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$ are actually only semimetrics, but this distinction is not of importance in what follows and so we omit it, see Sections 2.2 and 2.3.

When θ is a symmetric α-stable random variable we will write $\boldsymbol{\mu}_{\alpha,r}$ and $\boldsymbol{\nu}_{\alpha,r}$ or simply $\boldsymbol{\mu}_r$ and $\boldsymbol{\nu}_r$ when it is understood, in place of $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$.

The remainder of this section describes the special properties of the ideal convolution (or smoothing) metrics $\boldsymbol{\mu}_{\theta,r}$ and $\boldsymbol{\nu}_{\theta,r}$. We first verify ideality.

Lemma 14.2.1. For all $\theta \in \mathfrak{X}$ and $r > 0$, $\boldsymbol{\mu}_{\theta,r}$ is an ideal metric of order $r - 1$.

Proof. If Z does not depend upon X_1 and X_2 then $\ell(X_1 + Z, X_2 + Z) \leqslant \ell(X_1, X_2)$, and hence $\boldsymbol{\mu}_{\theta,r}(X_1 + Z, X_2 + Z) \leqslant \boldsymbol{\mu}_{\theta,r}(X_1, X_2)$. Additionally, for any $c \neq 0$

$$\begin{aligned}\boldsymbol{\mu}_{\theta,r}(cX_1, cX_2) &= \sup_{h\in\mathbb{R}} |h|^r \ell(cX_1 + h\theta,\, cX_2 + h\theta) \\ &= \sup_{h\in\mathbb{R}} |ch|^r \ell(cX_1 + ch\theta,\, cX_2 + ch\theta) = |c|^{r-1}\boldsymbol{\mu}_{\theta,r}(X_1, X_2).\end{aligned}$$

QED

The proof of the next lemma is analogous to the one above.

Lemma 14.2.2. For all $\theta \in \mathfrak{X}$ and $r > 0$ $\boldsymbol{\nu}_{\theta,r}$ is an ideal metric of order r.

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

Lemma 14.2.3. Let $k \in \mathbb{N}^+ := \{0, 1, 2, \dots\}$ and suppose that $X, Y \in \mathfrak{X}(\mathbb{R})$ satisfy $\mathbb{E}X^j = \mathbb{E}Y^j, j = 1, \dots, k-2$. Then for every $\theta \in \mathfrak{X}(\mathbb{R})$ with a density g which is $k-1$ times differentiable

$$\boldsymbol{\mu}_{\theta,k}(X_1, X_2) \leqslant \frac{\|g^{(k-1)}\|_\infty}{(k-1)!} \boldsymbol{\kappa}_{k-1}(X_1, X_2). \tag{14.2.14}$$

Proof. In view of the inequality (Zolotarev 1986, Chapter 3, Kalashnikov and Rachev 1988, Theorem 10.1.1),

$$\zeta_{k-1}(X_1, X_2) \leqslant \frac{1}{(k-1)!} \boldsymbol{\kappa}_{k-1}(X_1, X_2) \tag{14.2.15}$$

it suffices to show that

$$\boldsymbol{\mu}_{\theta,k}(X_1, X_2) \leqslant \zeta_{k-1}(X_1, X_2) \tag{14.2.16}$$

but with $H(t) = F_{X_1}(t) - F_{X_2}(t)$ we have

$$\begin{aligned}\boldsymbol{\mu}_{\theta,k}(X_1, X_2) &= \sup_{h\in\mathbb{R}} |h|^k \sup_{x\in\mathbb{R}} \frac{1}{|h|} \left| \int g\left(\frac{x-y}{h}\right) \mathrm{d}H(y) \right| \\ &= \sup_{h\in\mathbb{R}} |h|^{k-1} \sup_{x\in\mathbb{R}} \left| \int H(y) g^{(1)}\left(\frac{x-y}{h}\right) \frac{1}{h} \mathrm{d}y \right| \\ &= \sup_{h\in\mathbb{R}} |h|^{k-2} \sup_{x\in\mathbb{R}} \left| \int g^{(1)}\left(\frac{x-y}{h}\right) \mathrm{d}H^{(-1)}(y) \right| \\ &\;\;\vdots \\ &= \sup_{h\in\mathbb{R}} |h| \sup_{x\in\mathbb{R}} \left| \int g^{(k-1)}\left(\frac{x-y}{h}\right) \frac{1}{h} H^{(-k+2)}(y)\, \mathrm{d}y \right|\end{aligned}$$

where

$$F_X^{-k}(x) := \int_{-\infty}^{x} \frac{(x-t)^k}{k!} \, \mathrm{d}F_X(t). \tag{14.2.18}$$

Therefore, by (14.2.10) and $\zeta_{k-1} = \zeta_{k-2,1}$, we have

$$\mu_{\theta,k}(X_1, X_2) \leqslant \|g^{(k-1)}\|_\infty \int |H^{(2-k)}(y)| \, \mathrm{d}y = \|g^{(k-1)}\|_\infty \zeta_{k-1}(X_1, X_2).$$

QED

Similar to Lemma 14.2.3 one can prove a slightly better estimate.

Lemma 14.2.4. For every $\theta \in \mathfrak{X}(\mathbb{R})$ with a density g which is m times differentiable and for all $X_1, X_2 \in \mathfrak{X}(\mathbb{R})$,

$$\mu_{\theta,r}(X_1, X_2) \leqslant C(m, p, g)\zeta_{m-1,p}(X_1, X_2) \tag{14.2.19}$$

where $r = m + 1/p$, $m \in \mathbb{N}^+$, and

$$C(m, p, g) := \|g^{(m)}\|_q \quad 1/p + 1/q = 1. \tag{14.2.20}$$

Proof. For any $r > 0$ and X_1, X_2, $H(t) = F_{X_1}(t) - F_{X_2}(t)$, we have, using integration by parts (cf. (14.2.17)) and Hölder's inequality

$$\begin{aligned}
\mu_{\theta,r}(X_1, X_2) &= \sup_{h>0} h^r \sup_{x\in\mathbb{R}} |p_{X_1+h\theta}(x) - p_{X_2+h\theta}(x)| \\
&= \sup_{h>0} h^{r-m-1} \sup_{x\in\mathbb{R}} \left| \int g^{(m)}\left(\frac{x-y}{h}\right) H^{(1-m)}(y) \, \mathrm{d}y \right| \\
&\leqslant \sup_{h>0} h^{r-m-1} \sup_{x\in\mathbb{R}} \left[\int \left| g^{(m)}\left(\frac{x-y}{h}\right) \right|^q \mathrm{d}y \right]^{1/q} \|H^{(1-m)}\|_p \\
&= C(m, p, g) \|H^{(1-m)}\|_p.
\end{aligned}$$

By Theorem 10.2.1 of Kalashnikov and Rachev (1988), $\zeta_{m-1,p}(X_1, X_2) < \infty$ implies $\zeta_{m-1,p}(X_1, X_2) = \|H^{(1-m)}\|_p$, completing the proof of the lemma. QED

Lemma 14.2.5. Under the hypotheses of Lemma 14.2.4 we have

$$\nu_{\theta,r}(X_1, X_2) \leqslant C(r, g)\zeta_r(X_1, X_2) \tag{14.2.21}$$

where $C(r, g)$ is a finite constant, $r \in \mathbb{N}^+$.

The proof is similar to the proof of Lemma 14.2.4 and left to the reader.

Lemma 14.2.6. (See Section 3, Theorem 10.1 of Kalashnikov and Rachev

(1988).) Let $m \in \mathbb{N}^+$ and suppose $\mathbb{E}(X_1^j - X_2^j) = 0$, $j = 0, 1, \ldots, m$. Then for $p \in [1, \infty)$

$$\zeta_{m,p}(X_1, X_2) \leqslant \begin{cases} \kappa_1^{1/p}(X_1, X_2) & \text{if } m = 0 \\ \dfrac{\Gamma(1 + 1/p)}{\Gamma(r)} \kappa_r(X_1, X_2) & \text{if } m = 1, 2, \ldots, r = m + 1/p. \end{cases} \tag{14.2.22}$$

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

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

Lemmas 14.2.4, 14.2.5 and 14.2.6 describe the conditions under which $\zeta_{\theta,r}$ (resp. $\nu_{\theta,r}$) is finite. Thus, by (14.2.19) and (14.2.22) we have that for $r > 1$

$$\begin{cases} \mathbb{E}(X_1^j - X_2^j) = 0, j = 0, 1, \ldots, m-1, \\ r := m + 1/p \\ \kappa_{r-1}(X_1, X_2) < \infty \end{cases} \Rightarrow \mu_{\theta,r}(X_1, X_2) < \infty \tag{14.2.23}$$

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

$$\begin{cases} \displaystyle\int x^j \, d(F_{X_1} - F_{X_2})(x) = 0, j = 0, 1, \ldots, m-1, \\ r := m + 1/p \\ \kappa_{r-1}(X_1, X_2) < \infty \end{cases} \Rightarrow \mu_{\alpha,r}(X_1, X_2) < \infty. \tag{14.2.24}$$

Similarly,

$$\begin{cases} \displaystyle\int x^j \, d(F_{X_1} - F_{X_2})(x) = 0, j = 0, \ldots, r-1, \\ r \in \mathbb{N}^+ \\ \kappa_r(X_1, X_2) < \infty \end{cases} \Rightarrow \nu_{\alpha,r}(X_1, X_2) < \infty. \tag{14.2.25}$$

We conclude our discussion of the ideal metrics $\mu_{\alpha,r}$ and $\nu_{\alpha,r}$ by showing that they satisfy the same weak convergence properties as do the Kantorovich distance ℓ_p and the pseudomoments κ_r.

Theorem 14.2.1. Let $k \in \mathbb{N}^+$, $0 < \alpha \leqslant 2$, and X_n, $U \in \mathfrak{X}(\mathbb{R})$ with $\mathbb{E}X_n^j = \mathbb{E}U^j$, $j = 1, \ldots, k-2$ and $\mathbb{E}|X_n|^{k-1} + \mathbb{E}|U|^{k-1} < \infty$. If k is odd then the following are equivalent as $n \to \infty$:

(i) $\mu_{\alpha,k}(X_n, U) \to 0$

(ii) (a) $X_n \xrightarrow{w} U$ and (b) $\mathbb{E}|X_n|^{k-1} \to \mathbb{E}|U|^{k-1}$

(iii) $\ell_{k-1}(X_n, U) \to 0$

(iv) $\boldsymbol{\kappa}_{k-1}(X_n, U) \to 0$
(v) $\boldsymbol{\nu}_{\alpha, k-1}(X_n, U) \to 0$.

Proof. We note that (ii)$\Leftrightarrow$(iii) follows immediately from Theorem 8.2.1 with $c(x, y) = |x - y|^{k-1}$ or from (8.2.21) to (8.2.24) and $\ell_{k-1} = \hat{\mathscr{L}}_{k-1}$. Also, (ii)$\Leftrightarrow$(iv) follows from the following three relations

$$\ell_1(X, Y) = \boldsymbol{\kappa}_1(X, Y) = \int_{\mathbb{R}} |F_X(x) - F_Y(x)|\, \mathrm{d}x$$

(see Corollary 5.4.1 and Theorem 6.1.1),

$$\boldsymbol{\kappa}_r(X, Y) = \boldsymbol{\kappa}_1(X^{\uparrow r}, Y^{\uparrow r})$$

for any $r > 0$ and $X^{\uparrow r} = |X|^r \operatorname{sgn} X$, and

$$\ell_1(X_n^{\uparrow r}, U^{\uparrow r}) \to 0 \Leftrightarrow X_n^{\uparrow r} \xrightarrow{w} U^{\uparrow r} \text{ and } \mathbb{E}|X_n^{\uparrow r}| \to \mathbb{E}|U^{\uparrow r}|$$

(see Theorem 6.3.1) (or Theorem 8.2.1) with $c(x, y) = |x - y|$). Finally, (iv)$\Rightarrow$(i) by (14.2.24) and (iv)$\Rightarrow$(v) by (14.2.25).

Thus the only new result here are the implications (i)$\Rightarrow$(ii) and (v)$\Rightarrow$(ii).

Now (i)$\Rightarrow$(ii) (a) follows easily from Fourier transform arguments since the Fourier transform of g never vanishes. Similarly, if (v) holds, then $X_n + Y_\alpha \xrightarrow{w} U + Y_\alpha$ and thus (ii) (a) follows. To prove (i)$\Rightarrow$(ii) (b) we need the following estimate for $\mu_{\alpha, k}(X, U)$.

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

$$\boldsymbol{\mu}_k(X, U) \geqslant \beta \left| \int_{\mathbb{R}} F_X^{(2-k)}(z) - F_U^{(2-k)}(z)\, \mathrm{d}z \right|. \tag{14.2.26}$$

Here $F^{(2-k)}$ is as in (14.2.18).

Proof of the claim. Integration by parts yields

$$\begin{aligned}
\boldsymbol{\mu}_k(X, U) &= \sup_{h \in \mathbb{R}} |h|^k \sup_{x \in \mathbb{R}} |p_{X+hY}(x) - p_{U+hY}(x)| \quad (Y := Y_\alpha) \\
&= \sup_{h \in \mathbb{R}} |h|^k \sup_{x \in \mathbb{R}} \left| \int p_{hY}(z)\, \mathrm{d}H(x - z) \right| \quad (H := F_X - F_U) \\
&= \sup_{h \in \mathbb{R}} |h|^k \sup_{x \in \mathbb{R}} \left| \int H^{(2-k)}(x - z) p_{hY}^{(k-1)}(z)\, \mathrm{d}z \right|.
\end{aligned} \tag{14.2.27}$$

Now $2\pi\ p_{hY}(z) = \int \exp(-itz)\exp(-|ht|^{\alpha})\,\mathrm{d}t$ and differentiating $p_{hY}\,k-1$ times gives (setting $\tilde{t} = th$)

$$2\pi|h^k p_{hY}^{(k-1)}(z)| = \left|h^k \int (\mathrm{i}t)^{k-1}\exp(-\mathrm{i}tz - |ht|^{\alpha})\,\mathrm{d}t\right|$$
$$= \left|h^k \int \left(\mathrm{i}\frac{\tilde{t}}{h}\right)^{k-1} \exp(-\mathrm{i}\tilde{t}z/h - |\tilde{t}|^{\alpha})\,\mathrm{d}\left(\frac{\tilde{t}}{h}\right)\right|$$
$$= \left|\int (\mathrm{i}\tilde{t})^{k-1} \exp(-\mathrm{i}\tilde{t}z/h - |\tilde{t}|^{\alpha})\,\mathrm{d}\tilde{t}\right|.$$

Since

$$\beta := \beta(\alpha, k) := \frac{1}{2\pi}\int |t|^{k-1} \exp(-|t|^{\alpha})\,\mathrm{d}t < \infty$$

we obtain

$$\lim_{h\to\infty} |h^k p_{hY}^{(k-1)}(z)| = \frac{1}{2\pi}\left|\int \lim_{h\to\infty} (\mathrm{i}t)^{k-1} \exp(\mathrm{i}tz/h - |t|^{\alpha})\,\mathrm{d}t\right| = \beta.$$

Now multiply both sides of (14.2.27) by β^{-1}. Since β and $\zeta_{k-1}(X, U)$ are both finite, $\beta^{-1}\boldsymbol{\mu}_k(X, U)$ is

$$\geqslant \beta^{-1} \sup_{x\in\mathbb{R}}\left|\int H^{(2-k)}(x-z) \lim_{h\to\infty} h^k p_{hY}^{(k-1)}(z)\,\mathrm{d}z\right| = \sup_{x\in\mathbb{R}}\left|\int H^{(2-k)}(x-z)\,\mathrm{d}z\right|$$
$$= \left|\int H^{(2-k)}(z)\,\mathrm{d}z\right|$$

which proves the claim.

Now, using equality of the first $k-2$ moments and applying (14.2.26) to X_n and U yields

$$\beta^{-1}\mu_k(X_n, U) \geqslant \left|\int_{-\infty}^{\infty} \frac{(z-t)^{k-2}}{(k-2)!} H_n(\mathrm{d}t)\,\mathrm{d}z\right| \quad (H_n := F_{X_n} - F_U)$$
$$= \left|\int_{-\infty}^{0} (\bullet)\,\mathrm{d}t + \int_{0}^{\infty} (\bullet)\,\mathrm{d}t\right| := |I_1 + I_2|. \tag{14.2.28}$$

To estimate I_1 and I_2 we first note that since

$$\int_{\mathbb{R}} (z-t)^{k-2} H_n(\mathrm{d}t) = \mathbb{E}(z - X_n)^{k-2} - \mathbb{E}(z-U)^{k-2} = 0$$

we obtain

$$\int_{-\infty}^{z} \frac{(z-t)^{k-2}}{(k-2)!} H_n(\mathrm{d}t) = -\int_{z}^{\infty} \frac{(z-t)^{k-2}}{(k-2)!} H_n(\mathrm{d}t)$$
$$= (-1)^{k-1} \int_{z}^{\infty} \frac{(t-z)^{k-2}}{(k-2)!} H_n(\mathrm{d}t). \qquad (14.2.29)$$

Thus by (14.2.29) and Fubini's theorem we obtain

$$I_2 = (-1)^{k-1} \int_0^{\infty} \int_z^{\infty} \frac{(t-z)^{k-2}}{(k-2)!} H_n(\mathrm{d}t)\,\mathrm{d}z$$
$$= (-1)^{k-1} \int_0^{\infty} \int_0^{t} \frac{(t-z)^{k-2}}{(k-2)!} \mathrm{d}z\, H_n(\mathrm{d}t) = \int_0^{\infty} \frac{(-t)^{k-1}}{(k-1)!} H_n(\mathrm{d}t). \qquad (14.2.30)$$

Another application of Fubini's theorem gives

$$I_1 = \int_{-\infty}^{0} \int_t^{0} \frac{(z-t)^{k-2}}{(k-2)!} \mathrm{d}z\, H_n(\mathrm{d}t) = \int_{-\infty}^{0} \frac{(-t)^{k-1}}{(k-1)!} H_n(\mathrm{d}t). \qquad (14.2.31)$$

Combining (14.2.29), (14.2.30) and (14.2.31) gives

$$\beta^{-1}\boldsymbol{\mu}_k(X_n, U) \geqslant \left| \int \frac{(-t)^{k-1}}{(k-1)!} H_n(\mathrm{d}t) \right| = \frac{1}{(k-1)!} |\mathbb{E}(X_n^{k-1} - U^{k-1})|$$

which gives the desired implication (i) $\Rightarrow$ (ii) (b).

To prove (v) $\Rightarrow$ (ii) (b) we integrate by parts to obtain

$$\boldsymbol{\nu}_k(X_n, U) \geqslant \int_{\mathbb{R}} |p_{X_n+Y}(x) - p_{U+Y}(x)|\,\mathrm{d}x$$
$$= \int_{\mathbb{R}} \left| \int p_Y^{(k)}(x-z) \int_{-\infty}^{z} \frac{(z-t)^{k-1}}{(k-1)!} \mathrm{d}H_n(t)\,\mathrm{d}z \right| \mathrm{d}x$$
$$\geqslant \left| \iint p_Y^{(k)}(x-z)\mathrm{d}x \int_{-\infty}^{z} \frac{(z-t)^{k-1}}{(k-1)!} \mathrm{d}H_n(t)\,\mathrm{d}z \right|$$
$$= \left| \iint_{-\infty}^{z} \frac{(z-t)^{k-1}}{(k-1)!} \mathrm{d}H_n(t)\,\mathrm{d}z \right| \left| \int p_Y^{(k)}(x)\,\mathrm{d}x \right|.$$

By (14.2.28) to (14.2.31) we obtain

$$\boldsymbol{\nu}_k(X_n, U) \geqslant \left| \int p_Y^{(k)}(x)\,\mathrm{d}x \right| |\mathbb{E}(X_n^k - U^k)|$$

showing (v) $\Rightarrow$ (ii) (b) and completing Theorem 14.2.1. QED

14.3 RATES OF CONVERGENCE IN THE CLT IN TERMS OF METRICS WITH UNIFORM STRUCTURE

First we develop rates of convergence with respect to the **Var**-metric (14.2.5). We suppose that $X, X_1, X_2, \ldots$ denotes a sequence of i.i.d. random variables in $\mathfrak{X}(U)$, where U is a separable Banach space. $Y \in \mathfrak{X}(U)$ denotes a symmetric α-stable random variable. The ideal convolution metric $\mathbf{v}_r := \mathbf{v}_{\alpha,r}$ (see (14.2.13) with $\theta = Y$) will play a central role.

Our main theorem is the following.

Theorem 14.3.1. Let Y be an α-stable random variable. Let $r = s + 1/p > \alpha$ for some integer s and $p \in [1, \infty)$, $a = 1/2^{r/\alpha}A$, and $A := 2(2^{(r/\alpha)-1} + 3^{r/\alpha})$. If $X \in \mathfrak{X}(U)$ satisfies

$$\tau_0 := \tau_0(X, Y) := \max(\mathbf{Var}(X, Y), \mathbf{v}_{\alpha,r}(X, Y)) \leqslant a \tag{14.3.1}$$

then for any $n \geqslant 1$

$$\mathbf{Var}\left(\frac{X_1 + \cdots + X_n}{n^{1/\alpha}}, Y\right) \leqslant A(a)\tau_0 n^{1-r/\alpha} \leqslant 2^{-r/\alpha} n^{1-r/\alpha}. \tag{14.3.2}$$

Remark 14.3.1. A result of this type was proved by Senatov (1980) for the case $U = \mathbb{R}^k$, $s = 3$, and $\alpha = 2$ via the ζ_r metric (14.2.1). We will follow Senatov's method with some refinements.

Before proving Theorem 14.3.1 we need a few auxiliary results.

Lemma 14.3.1. For any $X_1, X_2 \in \mathfrak{X}(U)$ and $\sigma > 0$

$$\mathbf{Var}(X_1 + \sigma Y, X_2 + \sigma Y) \leqslant \sigma^{-r}\mathbf{v}_r(X_1, X_2). \tag{14.3.3}$$

Proof. Since Y and $(-Y)$ have the same distribution

$$\mathbf{v}_r(X_1, X_2) = \sup_{h>0} h^r\, \mathbf{Var}(X_1 + hY, X_2 + hY)$$

and thus

$$\begin{aligned}\mathbf{Var}(X_1 + hY, X_2 + hY) &\leqslant h^{-r} \sup_{h>0} h^r\, \mathbf{Var}(X_1 + hY, X_2 + hY)\\ &= h^{-r}\mathbf{v}_r(X_1, X_2).\end{aligned}$$ QED

Lemma 14.3.2. For any $X_1, X_2, U, V \in \mathfrak{X}(U)$ the following inequality holds

$$\mathbf{Var}(X_1 + U, X_2 + U) \leqslant \mathbf{Var}(X_1, X_2)\mathbf{Var}(U, V) + \mathbf{Var}(X_1 + V, X_2 + V).$$

Proof. By the definition (14.2.5) and the triangle inequality

$$\begin{aligned}\mathbf{Var}(X_1 + U, X_2 + U) &= \sup\{|\mathbb{E}f(X_1 + U) - \mathbb{E}f(X_2 + U)|: \|f\|_\infty \leqslant 1\} \\ &= \sup\left\{\left|\int f(u)(\Pr_{X_1+U} - \Pr_{X_2+U})(du)\right|: \|f\|_\infty \leqslant 1\right\} \\ &\leqslant \sup\left\{\left|\int \bar{f}(x)(\Pr_{X_1} - \Pr_{X_2})(dx)\right|: \|f\|_\infty \leqslant 1\right\} \\ &\quad + \mathrm{Var}(X_1 + V, X_2 + V)\end{aligned}$$

where

$$\bar{f}(x) := \int f(u)(\Pr_U - \Pr_V)(du - x) = \int f(u + x)(\Pr_U - \Pr_V)(du)$$

and where $\Pr_X$ denotes the law of the U-valued random variable X. Since $\|f\|_\infty \leqslant 1$ then

$$\begin{aligned}\|\bar{f}\|_\infty &= \sup_{x\in U}\left|\int f(u + x)(\Pr_U - \Pr_V)(du)\right| \\ &\leqslant \mathbf{Var}(U, V), \quad \text{by (14.2.5)}\end{aligned}$$

and thus

$$\sup\left\{\left|\int_U \bar{f}(x)(\Pr_{X_1} - \Pr_{X_2})(dx)\right|: \|f\|_\infty \leqslant 1\right\}$$

is bounded by

$$\begin{aligned}&\leqslant \sup\left\{\left|\int_U g(x)(\Pr_{X_1} - \Pr_{X_2})(dx)\right|: \|g\|_\infty \leqslant \mathbf{Var}(U, V)\right\} \\ &= \mathbf{Var}(X_1, X_2)\mathbf{Var}(U, V).\end{aligned}$$

QED

We now prove Theorem 14.3.1; throughout, $Y_1, Y_2, \ldots$ denote i.i.d copies of Y.

Proof. We proceed by induction; for $n = 1$ the assertion of the theorem is trivial. For $n = 2$ the assertion follows from the inequality

$$\begin{aligned}\mathbf{Var}\left(\frac{X_1 + X_2}{2^{1/\alpha}}, Y\right) &= \mathbf{Var}\left(\frac{X_1 + X_2}{2^{1/\alpha}}, \frac{Y_1 + Y_2}{2^{1/\alpha}}\right) = \mathbf{Var}(X_1 + X_2, Y_1 + Y_2) \\ &\leqslant 2\,\mathbf{Var}(X_1, Y_2) \leqslant A(a)\tau_0 2^{1-r/\alpha}\end{aligned}$$

since $A(a) \geqslant 2^{r/\alpha}$. A similar calculation holds for $n = 3$. Suppose now that the estimate

$$\mathbf{Var}\left(\frac{X_1 + \cdots + X_j}{j^{1/\alpha}}, Y\right) \leqslant A(a)\tau_0 j^{1-r/\alpha} \tag{14.3.4}$$

holds for all $j < n$. To complete the induction we only need to show that (14.3.4) holds for $j = n$.

Thus assuming (14.3.4) we have by (14.3.1)

$$\mathbf{Var}\left(\frac{X_1 + \cdots + X_j}{j^{1/\alpha}}, Y\right) \leqslant A(a)a = 2^{-r/\alpha}. \tag{14.3.5}$$

For any integer $n \geqslant 4$ and $m = [n/2]$, where $[\cdot]$ denotes integer part, the triangle inequality gives

$$\begin{aligned}
V := \mathbf{Var}\left(\frac{X_1 + \cdots + X_n}{n^{1/\alpha}}, Y\right) &= \mathbf{Var}\left(\frac{X_1 + \cdots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \cdots + Y_n}{n^{1/\alpha}}\right) \\
&\leqslant \mathbf{Var}\left(\frac{X_1 + \cdots + X_m}{n^{1/\alpha}} + \frac{X_{m+1} + \cdots + X_n}{n^{1/\alpha}}, \right. \\
&\qquad\qquad \left. \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \frac{X_{m+1} + \cdots + X_n}{n^{1/\alpha}}\right) \\
&\quad + \mathbf{Var}\left(\frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \frac{X_{m+1} + \cdots + X_n}{n^{1/\alpha}}, \right. \\
&\qquad\qquad \left. \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \frac{Y_{m+1} + \cdots + Y_n}{n^{1/\alpha}}\right).
\end{aligned}$$

Hence, by Lemma 14.3.2

$$V \leqslant I_1 + I_2 + I_3 \tag{14.3.6}$$

where

$$\begin{aligned}
I_1 &:= \mathbf{Var}\left(\frac{X_1 + \cdots + X_m}{n^{1/\alpha}}, \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}}\right) \\
&\qquad\qquad \mathbf{Var}\left(\frac{X_{m+1} + \cdots + X_n}{n^{1/\alpha}}, \frac{Y_{m+1} + \cdots + Y_n}{n^{1/\alpha}}\right) \\
I_2 &:= \mathbf{Var}\left(\frac{X_1 + \cdots + X_m}{n^{1/\alpha}} + \frac{Y_{m+1} + \cdots + Y_n}{n^{1/\alpha}}, \right. \\
&\qquad\qquad \left. \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \frac{Y_{m+1} + \cdots + Y_n}{n^{1/\alpha}}\right)
\end{aligned}$$

and

$$I_3 := \mathbf{Var}\left(\frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \frac{X_{m+1} + \cdots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \frac{Y_{m+1} + \cdots + Y_n}{n^{1/\alpha}}\right).$$

We first estimate I_1. By (14.3.5)

$$I_1 \leqslant 2^{-r/\alpha} A(a)\tau_0 (n-m)^{1-r/\alpha} \leqslant \tfrac{1}{2} A(a)\tau_0 n^{1-r/\alpha}. \tag{14.3.7}$$

In order to estimate I_2 and I_3 we will use Lemma 14.3.1 and the relation

$$\frac{Y_1 + \cdots + Y_n}{n^{1/\alpha}} \overset{\mathrm{d}}{=} Y_1. \tag{14.3.8}$$

Thus, by (14.3.8), Lemma 14.3.1, and the fact that $\mathbf{v}_r$ is ideal of order r, we deduce

$$\begin{aligned} I_2 &= \mathbf{Var}\left(\frac{X_1 + \cdots + X_n}{n^{1/\alpha}} + \left(\frac{n-m}{n}\right)^{1/\alpha} Y, \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}} + \left(\frac{n-m}{n}\right)^{1/\alpha} Y\right) \\ &\leqslant \left(\frac{n-m}{n}\right)^{-r/\alpha} \mathbf{v}_r\left(\frac{X_1 + \cdots + X_m}{n^{1/\alpha}}, \frac{Y_1 + \cdots + Y_m}{n^{1/\alpha}}\right) \\ &\leqslant 2^{r/\alpha} m \mathbf{v}_r\left(\frac{X_1}{n^{1/\alpha}}, \frac{Y_1}{n^{1/\alpha}}\right) \leqslant 2^{(r/\alpha)-1} n^{1-r/\alpha} \mathbf{v}_r(X_1, Y_1). \end{aligned} \tag{14.3.9}$$

Analogously, we estimate I_3 by

$$\begin{aligned} I_3 &= \mathbf{Var}\left(\frac{X_1 + \cdots + X_{n-m}}{n^{1/\alpha}} + \left(\frac{m}{n}\right)^{1/\alpha} Y, \frac{Y_1 + \cdots + Y_{n-m}}{n^{1/\alpha}} + \left(\frac{m}{n}\right)^{1/\alpha} Y\right) \\ &\leqslant \left(\frac{m}{n}\right)^{-r/\alpha} \mathbf{v}_r\left(\frac{X_1 + \cdots + X_{n-m}}{n^{1/\alpha}}, \frac{Y_1 + \cdots + Y_{n-m}}{n^{1/\alpha}}\right) \\ &\leqslant 3^{r/\alpha} n^{1-r/\alpha} \mathbf{v}_r(X_1, Y_1). \end{aligned} \tag{14.3.10}$$

Taking (14.3.6), (14.3.7), (14.3.9) and (14.3.10) into account we obtain

$$V \leqslant (\tfrac{1}{2}A(a) + 2^{r/\alpha - 1} + 3^{r/\alpha})\tau_0 n^{1-r/\alpha} \leqslant A(a)\tau_0 n^{1-r/\alpha}$$

since $A(a)/2 = 2^{(r/\alpha)-1} + 3^{r/\alpha}$. QED

Further, we develop rates of convergence in (14.2.3) with respect to the χ metric (14.2.7); our purpose here is to show that the methods of proof for Theorem 14.3.1 can be easily extended to deduce analogous results with respect to $\boldsymbol{\chi}$. The metric $\boldsymbol{\chi}_r$ (14.2.9) will play a role analogous to that played by $\mathbf{v}_r$ in Theorem 14.3.1.

Theorem 14.3.2. Let Y be an α-stable random variable in $\mathfrak{X}(\mathbb{R})$. Let $r > \alpha$, $b := 1/2^{r/\alpha}B$, and $B := \max(3^{r/\alpha}, 2C_r(2^{(r/\alpha)-1} + 3^{r/\alpha}))$ where $C_r := (r/\alpha e)^{r/\alpha}$. If $X \in \mathfrak{X}(\mathbb{R})$ satisfies

$$\tau_r := \tau_r(X, Y) := \max\{\chi(X, Y), \chi_r(X, Y)\} \leqslant b \tag{14.3.11}$$

then for all $n \geqslant 1$

$$\chi\left(\frac{X_1 + \cdots + X_n}{n^{1/\alpha}}, Y\right) \leqslant B\tau_r n^{1-r/\alpha} \leqslant 2^{-r/\alpha} n^{1-r/\alpha}. \tag{14.3.12}$$

Remark 14.3.2. (i) In comparing conditions (14.3.1) and (14.3.11) it is useful to note that the metric χ is topologically weaker than **Var**, i.e., $\mathbf{Var}(X_n, Y) \to 0$ implies $\chi(X_n, Y) \to 0$ but not conversely. Also, if $r = m + \beta$, $m = 0, 1, \ldots,$ $\beta \in (0, 1]$ then (cf. (14.2.1) and (14.2.9)),

$$\chi_r \leqslant C_\beta \zeta_r \tag{14.3.13}$$

where $C_\beta := \sup_t |t|^{-\beta}|1 - e^{it}|$.

Proof of inequality (14.3.13). By the definitions of χ_r and ζ_r we have

$$\chi_r(X, Y) := \sup_{t \in \mathbb{R}} |\mathbb{E}(f_t(X) - f_t(Y))|$$

where $f_t(x) := t^{-r}\exp(itx)$ and

$$\zeta_r(X, Y) := \sup\{|\mathbb{E}(f(X) - f(Y)|: f\colon \mathbb{R} \to \mathbb{C}, \text{ and } |f^{(m)}(x) - f^{(m)}(y)| \leqslant |x - y|^\beta\}$$

where $r = m + \beta$, $m = 0, 1, \ldots$ and $\beta \in (0, 1]$. For any $t \in \mathbb{R}$

$$f_t^{(m)}(x) = t^{-\beta} i^m \exp(itx)$$

and thus

$$\frac{|f_t^{(m)}(x) - f_t^{(m)}(y)|}{|s|^\beta} = \frac{|t|^{-\beta}|\exp(itx) - \exp(ity)|}{|s|^\beta} = \frac{|t|^{-\beta}|1 - \exp(its)|}{|s|^\beta}$$

where $s := x - y$. We observe that for any $D_r > 0$

$$D_r\zeta_r(X, Y) = \sup\{|\mathbb{E}(f(X) - f(Y))|: |f^{(m)}(x) - f^{(m)}(y)| \leqslant D_r|x - y|^\beta\}$$

and

$$\sup_{x, y \in \mathbb{R}} \frac{|f_t^{(m)}(x) - f_t^{(m)}(y)|}{|x - y|^\beta} \leqslant \sup_{s \in \mathbb{R}} |st|^{-\beta}|1 - \exp(its)| := C_\beta.$$

A simple calculation shows that $C_\beta < \infty$ and this completes the proof of inequality (14.3.13). QED

Finally, we note that since $\zeta_m(X, Y) := \sup\{|\mathbb{E}(f(X) - f(Y))| : |f^{(m+1)}(x)| \leqslant 1$ a.e.$\}$ and since $|f_t^{(m+1)}(x)| = |i^{m+1}\exp(itx)| = 1$, we obtain $\chi_m \leqslant \zeta_m$.

Remark 14.3.3. One may show that for $r \in \mathbb{N}^+$ the metric χ_r has a convolution type structure. In fact, with a slight abuse of notation,

$$\chi_r(F_{X_1}, F_{X_2}) = \chi(F_{X_1} * p_r, F_{X_2} * p_r)$$

where $p_r(t) = (t^r/r!)I_{\{t>0\}}$ is the density of an unbounded positive measure on the half-line $[0, \infty)$.

The proof of Theorem 14.3.2 is very similar to that of Theorem 14.3.1 and uses the following auxiliary results, which are completely analogous to Lemma 14.3.1 and 14.3.2. We leave the details to the reader to complete the proof of Theorem 14.3.2.

Lemma 14.3.3. For any $X_1, X_2 \in \mathfrak{X}(\mathbb{R})$, $\sigma > 0$, and $r > \alpha$

$$\chi(X_1 + \sigma Y, X_2 + \sigma Y) \leqslant C_r \sigma^{-r} \chi_r(X_1, X_2)$$

where $C_r := (r/\alpha e)^{r/\alpha}$.

Proof. We have

$$\begin{aligned}
\chi(X_1 + \sigma Y, X_2 + \sigma Y) &:= \sup_{t \in \mathbb{R}} |\phi_{X_1}(t) - \phi_{X_2}(t)| \phi_{\sigma Y}(t) \\
&= \sup_{t \in \mathbb{R}} |\phi_{X_1}(t) - \phi_{X_2}(t)| \exp\{-|\sigma t|^\alpha\} \\
&\leqslant \sup_{t \in \mathbb{R}} |\sigma t|^{-r} |\phi_{X_1}(t) - \phi_{X_2}(t)| \sup_{u > 0} u^r \exp(-u^\alpha) \\
&= C_r \sigma^{-r} \chi_r(X, Y)
\end{aligned}$$

since $C_r = \sup_{u>0} u^r \exp(-u^\alpha)$ by a simple computation. QED

Lemma 14.3.4. For any $X_1, X_2, Z, W \in \mathfrak{X}(\mathbb{R})$ the following inequality holds:

$$\chi(X_1 + Z, X_2 + Z) \leqslant \chi(X_1, X_2)\chi(Z, W) + \chi(X_1 + W, X_2 + W).$$

Proof. From the inequality

$$\begin{aligned}
|\phi_{X_1+Z}(t) - \phi_{X_2+Z}(t)| \leq |\phi_{X_1}(t) - \phi_{X_2}(t)| |\phi_Z(t) - \phi_W(t)| \\
+ |\phi_{X_1}(t) - \phi_{X_2}(t)| |\phi_W(t)|
\end{aligned}$$

we obtain the desired result. QED

Finally, we develop convergence rates with respect to the ℓ-metric (14.2.6) and thus we naturally restrict attention to the subset $\mathfrak{X}^*$ of $\mathfrak{X}(\mathbb{R}^k)$ of random variables with densities. Let $X, X_1, X_2, \ldots,$ denote a sequence of i.i.d. random variables in $\mathfrak{X}^*$ and $Y = Y_\alpha$ denote a symmetric α-stable random vector. The ideal convolution metric $\mu_r := \mu_{\alpha,r}$ and $\nu_r := \nu_{\alpha,r}$ (i.e., $\theta = Y$) will play a central role.

Theorem 14.3.3. Let Y be a symmetric α-stable random variable in $\mathfrak{X}(\mathbb{R}^k)$. Let $r = m + 1/p > \alpha$ for some integer m and $p \in [1, \infty)$, $a := 1/2^{r/\alpha}A$, $A := 2(2^{r/\alpha - 1} + 3^{(r+1)/\alpha})$ and $D := 3^{1/\alpha}\, 2^{r/\alpha}$. If $X \in \mathfrak{X}^*$ satisfies

(i) $$\tau(X, Y) := \max(\ell(X, Y), \boldsymbol{\mu}_{\alpha,r}(X, Y)) \leqslant a$$

(ii) $$\tau_0(X, Y) := \max(\mathbf{Var}(X, Y), \boldsymbol{\nu}_{\alpha,r}(X, Y)) \leqslant \frac{1}{A(a)D} \tag{14.3.14}$$

then

$$\ell\left(\frac{X_1 + \cdots + X_n}{n^{1/\alpha}}, Y\right) \leqslant A(a)\tau(X, Y)n^{1 - r/\alpha}. \tag{14.3.15}$$

Remark 14.3.4. (i) Conditions (i) and (ii) guarantee ℓ-closeness (of order $n^{1-r/\alpha}$) between Y and the normalized sums $n^{-1/\alpha}(X_1 + \cdots + X_n)$.

(ii) From Lemmas 14.2.3, 14.2.5 and 14.2.6 we know that $\boldsymbol{\mu}_{\alpha,r+1}(X, Y)$ and $\boldsymbol{\nu}_{\alpha,r}(X, Y)$, $r = m - 1 + 1/p$, $m = 1, 2, \ldots$ can be approximated from above by the rth difference pseudomoment κ_r whenever X and Y share the same first $(m-1)$ moments (see (14.2.23) to (14.2.25)). Thus conditions (i) and (ii) could be expressed in terms of difference pseudomoments, which of course amounts to conditions on the tails of X.

To prove Theorem 14.3.3 we need a few auxiliary results similar in spirit to Lemma 14.3.1 and 14.3.2.

Lemma 14.3.5. Let $X_1, X_2 \in \mathfrak{X}(\mathbb{R}^k)$. Then

$$\ell(X_1 + \sigma Y, X_2 + \sigma Y) \leqslant \sigma^{-r}\boldsymbol{\mu}_r(X_1, X_2).$$

Proof. $\ell(X_1 + \sigma Y, X_2 + \sigma Y) \leqslant \sigma^{-r}\sigma^r\ell(X_1 + \sigma Y, X_2 + \sigma Y) \leqslant \sigma^{-r}\boldsymbol{\mu}_r(X_1, X_2)$. QED

Lemma 14.3.6. For any (independent) $X, Y, U, V \in \mathfrak{X}^*(\mathbb{R}^k)$ the following inequality holds:

$$\ell(X + U, Y + U) \leqslant \ell(X, Y)\,\mathbf{Var}(U, V) + \ell(X + V, Y + V).$$

Proof. Using the triangle inequality we obtain

$$\begin{aligned}\ell(X+U, Y+U) &= \sup_{x\in\mathbb{R}^k}\left|\int (p_X(x-y)-p_Y(x-y))\Pr(U\in \mathrm{d}y)\right| \\ &\leqslant \sup_{x\in\mathbb{R}^k}\left|\int (p_X(x-y)-p_Y(x-y))(\Pr\{U\in \mathrm{d}y\}-\Pr\{V\in \mathrm{d}y\}\right| \\ &\quad + \sup_{x\in\mathbb{R}^k}\left|\int (p_X(x-y)-p_Y(x-y))\Pr\{V\in \mathrm{d}y\}\right| \\ &\leqslant \ell(X, Y)\,\mathbf{Var}(U, V)+\ell(X+V, Y+V).\end{aligned}$$

QED

To prove Theorem 14.3.3 one only needs to use the method of proof for Theorem 14.3.1 combined with the above two auxiliary results. The complete details are left to the reader. A more general theorem will be proved in the next section, see Theorem 15.2.2.

The results above show that the 'ideal' structure of the convolution metrics $\boldsymbol{\mu}_r$ and $\mathbf{v}_r$ may be used to determine optimal rates of convergence in the general central limit theorem problem. The rates are expressed in terms of the uniform metrics **Var**, $\boldsymbol{\chi}$ and ℓ and hold uniformly in n under the sufficient conditions (14.3.1), (14.3.11), and (14.3.14), respectively. We have not explored the possible weakening of these conditions or even their possible necessity.

The ideal convolution metrics $\boldsymbol{\mu}_r$ and $\mathbf{v}_r$ are not limited to the context of Theorems 14.3.1, 14.3.2 and 14.3.3, but they can also be successfully employed to study other questions of interest. For example, we only mention here that v_r can be used to prove a Berry–Esseen type of estimate for the Kolmogorov metric $\boldsymbol{\rho}$ (14.2.8).

More precisely, if $X, X_1, X_2, \ldots$ denotes a sequence of i.i.d. random variables in $\mathfrak{X}(\mathbb{R})$ and $Y\in\mathfrak{X}(\mathbb{R})$ a symmetric α-stable random variable, then for all $r>\alpha$ and $n\geqslant 1$

$$\begin{aligned}&\boldsymbol{\rho}\left(\frac{X_1+\cdots+X_n}{n^{1/\alpha}}, Y\right) \\ &\qquad\leqslant C\mathbf{v}_{\alpha,r}(X, Y)n^{1-r/\alpha}+C\max\{\boldsymbol{\rho}(X, Y), \mathbf{v}_{\alpha,1}(X, Y), \mathbf{v}_{\alpha,r}^{1/(r-\alpha)}(X, Y))\}n^{-1/\alpha}\end{aligned} \tag{14.3.16}$$

where C is an absolute constant. Whenever $\mathbf{v}_{\alpha,1}(X, Y)<\infty$ and $\mathbf{v}_{\alpha,r}(X, Y)<\infty$ we obtain the right order estimate in the Berry–Esseen theorem in terms of the metric $v_{\alpha,r}$.

Thus, metrics of the convolution type, especially those with the 'ideal' structure, are appropriate when investigating sums of independent random variables converging to a stable limit law. We can only conjecture that there are other ideal convolution metrics, other than those explored in this section, which may furnish additional results in related limit theorem problems.

CHAPTER 15

Ideal Metrics and Rate of Convergence in the CLT for Random Motions

This chapter considers the rate of the convergence problem for α-stable distributions on the non-commutative group of motions in $\mathbb{R}^d$. The techniques involve the careful use of convolution ideal metrics on the space of random motions.

15.1 IDEAL METRICS IN THE SPACE OF RANDOM MOTIONS

Let $\mathbb{M}(d)$ be the group of rigid motions on $\mathbb{R}^d$, i.e. the group of one-to-one transformations of $\mathbb{R}^d$ to $\mathbb{R}^d$ preserving the orientation of the space and the inner product. $\mathbb{M}(d)$ is known as the Euclidean group of motions of d-dimensional Euclidean space. Letting $\mathrm{SO}(d)$ denote the special orthogonal group in $\mathbb{R}^d$, any element $g \in \mathbb{M}(d)$ can be written in the form $g = (y, u)$, where $y \in \mathbb{R}^d$ represents the translation parameter and $u \in \mathrm{SO}(d)$ is a rotation about the origin. Note that for all $x \in \mathbb{R}^d$, $g(x) = y + ux$. If $g_i = (y_i, u_i)$ $1 \leqslant i \leqslant n$, then the product $g(n) = g_1 \circ g_2 \circ \cdots \circ g_n$ has the form $g(n) = (y(n), u(n))$, where $u(n) = u_1, \ldots, u_n$ and $y(n) = y_1 + u_1 y_2 + \cdots + u_1 \cdots u_{n-1} y_n$. For any $c \in \mathbb{R}$ and $g = (y, u) \in \mathbb{M}(d)$, define $cg = (cy, u)$.

Next, let $(\Omega, \mathscr{F}, \Pr)$ be a probability space on which is defined a sequence of i.i.d. random variables G_i, $i \geqslant 1$, with values in $\mathbb{M}(d)$. A natural problem involves finding the limiting distribution (i.e. CLT) of the product $G_1 \circ \cdots \circ G_n$, which leads to the notion of α-stable random motion. The definition of an α-stable random motion resembles that for a spherically symmetric α-stable random vector, that is, H_α is an α-stable random motion if for any sequence of i.i.d. random motions G_i, with $G_1 \overset{\mathrm{d}}{=} H_\alpha$

$$H_\alpha \overset{\mathrm{d}}{=} n^{-1/\alpha}(G_1 \circ \cdots \circ G_n) \quad \text{for any } n \geqslant 1, \text{ and } H_\alpha \overset{\mathrm{d}}{=} uH_\alpha \quad \text{for any } u \in \mathrm{SO}(d). \tag{15.1.1}$$

Baldi (1979) proved that $H_\alpha = (Y_\alpha, U_\alpha)$ is an α-stable random motion if and only if Y_α has a spherically symmetric α-stable distribution on $\mathbb{R}^d$ and U_α is uniformly distributed on $\mathrm{SO}(d)$. Henceforth, we write $H_\alpha = (Y_\alpha, U_\alpha)$ to denote an α-stable random motion. In this section, we will be interested in the rate of convergence of i.i.d. random motions to a stable random motion. First we shall define and examine the properties of ideal metrics related to this particular approximation problem.

Let $\mathfrak{X}(\mathbb{M}(d))$ be the space of all random motions $G = (Y, U)$ on $(\Omega, \mathscr{F}, \Pr)$, $Y \in \mathfrak{X}(\mathbb{R}^d)$ the space of all d-dimensional random vectors, and $U \in \mathfrak{X}(\mathrm{SO}(d))$ the space of all random 'rotations' in $\mathbb{R}^d$. $\mathfrak{X}^*(\mathbb{R}^d)$ denotes the subspace of $\mathfrak{X}(\mathbb{R}^d)$ of all random variables with densities: $\mathfrak{X}^*(\mathbb{M}(d))$ is defined by $\mathfrak{X}^*(\mathbb{R}^d) \times \mathfrak{X}(\mathrm{SO}(d))$.

Define the total variation distance between elements G and G^* of $\mathfrak{X}(\mathbb{M}(d))$ by

$$\mathbf{Var}(G, G^*) := \sup_{x \in \mathbb{R}^d} \mathbf{Var}(G(x), G^*(x)) \tag{15.1.2}$$

where for X and Y in $\mathfrak{X}(\mathbb{R}^d)$

$$\mathbf{Var}(X, Y) := 2 \sup\{|\Pr\{X \in A\} - \Pr\{Y \in A\}|, \quad A \in \mathscr{B}(\mathbb{R}^d)\}$$

$\mathscr{B}(\mathbb{R}^d)$ denoting the Borel sets in $\mathbb{R}^d$, see (14.2.5).

Let $\theta \in \mathfrak{X}(\mathbb{R}^d)$ have a spherically symmetric α-stable distribution on $\mathbb{R}^d$. As in Section 14.2, define smoothing metrics associated with the **Var** and ℓ distances

$$\mathbf{v}_r(X, Y) := \sup_{h \in \mathbb{R}} |h|^r \, \mathbf{Var}(X + h\theta, Y + h\theta) \quad X, Y \in \mathfrak{X}(\mathbb{R}^d) \tag{15.1.3}$$

and

$$\boldsymbol{\mu}_r(X, Y) := \sup_{h \in \mathbb{R}} |h|^r \ell(X + h\theta, Y + h\theta) \quad X, Y \in \mathfrak{X}^*(\mathbb{R}^d) \tag{15.1.4}$$

where $\ell(X, Y)$, $X, Y \in \mathfrak{X}^*(\mathbb{R}^d)$, is the ess sup norm distance between the densities p_X and p_Y of X and Y, respectively, that is,

$$\ell(X, Y) := \operatorname*{ess\,sup}_{y \in \mathbb{R}^d} |p_X(y) - p_Y(y)| \tag{15.1.5}$$

see (14.2.6), (14.2.12), (14.2.13).

Next, extend the definitions of $\mathbf{v}_r$ and $\boldsymbol{\mu}_r$ to $\mathfrak{X}(\mathbb{M}(d))$ and $\mathfrak{X}^*(\mathbb{M}(d))$, respectively

$$\mathbf{v}_r(G_1, G_2) := \sup_{x \in \mathbb{R}^d} \mathbf{v}_r(G_1(x), G_2(x)) \quad G_1, G_2 \in \mathfrak{X}(\mathbb{M}(d)) \tag{15.1.6}$$

and

$$\boldsymbol{\mu}_r(G_1, G_2) := \sup_{x \in \mathbb{R}^d} \boldsymbol{\mu}_r(G_1(x), G_2(x)) \quad G_1, G_2 \in \mathfrak{X}^*(\mathbb{M}(d)). \tag{15.1.7}$$

As in Chapter 14, $\mathbf{v}_r$ and $\boldsymbol{\mu}_r$ will play important roles in establishing rates of convergence in the integral and local limit CLT theorems. Zolotarev's $\boldsymbol{\zeta}_r$ metric defined by (14.2.1) on $\mathfrak{X}(\mathbb{R}^d)$ is similarly extended in $\mathfrak{X}(\mathbb{M}(d))$

$$\zeta_r(G_1, G_2) := \sup_{x \in \mathbb{R}^d} \zeta_r(G_1(x), G_2(x)). \tag{15.1.8}$$

The following two theorems record some special properties of $\mathbf{v}_r$ and $\boldsymbol{\mu}_r$ and is proved by exploiting their ideality on $\mathfrak{X}(\mathbb{R}^d)$ (cf. Lemmas 14.2.1 and 14.2.2).

Theorem 15.1.1. $\boldsymbol{\mu}_r$ is an ideal metric on $\mathfrak{X}^*(\mathbb{M}(d))$ of order $r - 1$, i.e., $\boldsymbol{\mu}_r$ satisfies the following two conditions:

(i) *Regularity*: $\boldsymbol{\mu}_r(G_1 \circ G, G_2 \circ G) \leqslant \boldsymbol{\mu}_r(G_1, G_2)$ and

$$\boldsymbol{\mu}_r(G \circ G_1, G \circ G_2) \leqslant \boldsymbol{\mu}_r(G_1, G_2)$$

for any G_1 and G_2 which are independent of G;

(ii) *Homogeneity*: $\boldsymbol{\mu}_r(cG_1, cG_2) \leqslant |c|^{r-1}\boldsymbol{\mu}_r(G_1, G_2)$ for any $c \in \mathbb{R}$.

Proof. The proof rests upon two auxiliary lemmas.

Lemma 15.1.1. For any independent $G \in \mathfrak{X}^*(\mathbb{M}(d))$, Y_1, $Y_2 \in \mathfrak{X}^*(\mathbb{R}^d)$

$$\boldsymbol{\mu}_r(G(Y_1), G(Y_2)) \leqslant \boldsymbol{\mu}_r(Y_1, Y_2). \tag{15.1.9}$$

Proof of Lemma 15.1.1. By definition of $\boldsymbol{\mu}_r$ and the regularity of ℓ, we have, for $G := (Y, U)$

$$\begin{aligned}\boldsymbol{\mu}_r(G(Y_1), G(Y_2)) &= \sup_{x \in \mathbb{R}} |h|^r \ell(G(Y_1) + h\theta, G(Y_2) + h\theta)\\ &= \sup_{x \in \mathbb{R}} |h|^r \ell(Y + UY_1 + h\theta, Y + UY_2 + h\theta)\\ &\leqslant \sup_{x \in \mathbb{R}} |h|^r \ell(UY_1 + h\theta, UY_2 + h\theta). \end{aligned}\tag{15.1.10}$$

Next, we show $\ell(UY_1, UY_2) \leqslant \ell(Y_1, Y_2)$ for any independent $U \in \mathfrak{X}(\mathrm{SO}(d))$. To see this, notice that

$$\ell(UY_1, UY_2) \leqslant \sup_{x \in \mathbb{R}^d} \sup_{u \in \mathrm{SO}(d)} |p_{uY_1}(x) - p_{uY_2}(x)|$$

$$= \sup_{u \in \mathrm{SO}(d)} \sup_{z = u \circ x \in \mathbb{R}^d} |(p_{Y_1} - p_{Y_2})(x_1(z_1, \ldots, z_d), \ldots, x_d(z_1, \ldots, z_d))| \left| \frac{\partial x_1 \cdots \partial x_d}{\partial z_1 \cdots \partial z_d} \right|.$$

Since the determinant of the Jacobian equals 1

$$\ell(UY_1, UY_2) \leqslant \ell(Y_1, Y_2). \tag{15.1.11}$$

Combining (15.1.10) and (15.1.11) and using $U^{-1}\theta \stackrel{\mathrm{d}}{=} \theta$, where $UU^{-1} = I$, we have

$$\mu_r(G(Y_1), G(Y_2)) \leqslant \sup_{h\in\mathbb{R}} |h|^r \ell(U(Y_1 + hU^{-1}\theta), U(Y_2 + hU^{-1}\theta))$$

$$\leqslant \sup_{h\in\mathbb{R}} |h|^r \ell(Y_1 + hU^{-1}\theta, Y_2 + hU^{-1}\theta) = \boldsymbol{\mu}_r(Y_1, Y_2).$$

QED

Lemma 15.1.2. If G_1, G_2, and Y are independent, then

$$\boldsymbol{\mu}_r(G_1(Y), G_2(Y)) \leqslant \sup_{x\in\mathbb{R}^d} \boldsymbol{\mu}_r(G_1(x), G_2(x)). \tag{15.1.12}$$

Proof of Lemma 15.1.2. We have, for $G_i = (Y_i, U_i)$

$$\begin{aligned}
\boldsymbol{\mu}_r(G_1(Y), G_2(Y)) &= \sup_{h\in\mathbb{R}} |h|^r \ell(Y_1 + U_1 Y + h\theta, Y_2 + U_2 Y + h\theta) \\
&= \sup_{h\in\mathbb{R}} |h|^r \sup_{x\in\mathbb{R}^d} |p_{Y_1+U_1Y+h\theta}(x) - p_{Y_2+U_2Y+h\theta}(x)) \\
&= \sup_{h\in\mathbb{R}} |h|^r \sup_{x\in\mathbb{R}^d} \bigg| \int_{\mathbb{R}^d} \bigg\{ \int p_{Y_1+h\theta}(x - u_1 y)\Pr(U_1 \in \mathrm{d}u_1) \\
&\qquad - \int p_{Y_2+h\theta}(x - u_2 y)\Pr(U_2 \in \mathrm{d}u_2) \bigg\} \Pr(Y \in \mathrm{d}y) \bigg| \\
&\leqslant \sup_{h\in\mathbb{R}} |h|^r \sup_{y\in\mathbb{R}^d} \sup_{x\in\mathbb{R}^d} \bigg| \int p_{Y_1+h\theta}(x - u_1 y)\Pr(U_1 \in \mathrm{d}u_1) \\
&\qquad - \int p_{Y_2+h\theta}(x - u_2 y)\Pr(U_2 \in \mathrm{d}u_2) \bigg| \\
&= \sup_{y\in\mathbb{R}^d} \sup_{h\in\mathbb{R}} |h|^r \sup_{x\in\mathbb{R}^d} |p_{Y_1+h\theta+U_1y}(x) - p_{Y_2+h\theta+U_2y}(x)| \\
&= \sup_{y\in\mathbb{R}^d} \boldsymbol{\mu}_r(G_1(y), G_2(y)).
\end{aligned}$$

QED

Now we can prove property (i) of the theorem. By (15.1.9),

$$\boldsymbol{\mu}_r(G \circ G_1, G \circ G_2) = \sup_{x\in\mathbb{R}^d} \mu_r(G \circ G_1(x), G \circ G_2(x))$$

$$\leqslant \sup_{x\in\mathbb{R}^d} \boldsymbol{\mu}_r(G_1(x), G_2(x)) = \boldsymbol{\mu}_r(G_1, G_2).$$

Similarly, by (15.1.12)

$$\begin{aligned}\boldsymbol{\mu}_r(G_1 \circ G, G_2 \circ G) &= \sup_{x \in \mathbb{R}^d} \boldsymbol{\mu}_r(G_1 \circ G(x), G_2 \circ G(x)) \\ &\leqslant \sup_{x \in \mathbb{R}^d} \boldsymbol{\mu}_r(G_1(x), G_2(x)) = \boldsymbol{\mu}_r(G_1, G_2)\end{aligned}$$

which completes the proof of the 'regularity' property. To prove the 'homogeneity', observe that by the ideality of $\boldsymbol{\mu}_r$ on $\mathfrak{X}(\mathbb{R}^d)$

$$\begin{aligned}\boldsymbol{\mu}_r(cG_1, cG_2) &= \sup_{x \in \mathbb{R}^d} \boldsymbol{\mu}_r(cY_1 + U_1 x, cY_2 + U_2 x) \\ &= \sup_{x \in \mathbb{R}^d} \boldsymbol{\mu}_r\left(c\left(Y_1 + U_1\left(\frac{1}{c}x\right)\right), c\left(Y_2 + U_2\left(\frac{1}{c}x\right)\right)\right) \\ &= |c|^{r-1}\boldsymbol{\mu}_r(G_1, G_2).\end{aligned}$$

QED

Theorem 15.1.2. $\mathbf{v}_r$ is an ideal metric on $\mathfrak{X}(\mathbb{M}(d))$ of order r.

The proof is similar to that of the previous theorem.

The usefulness of ideality may be illustrated in the following way. If μ is ideal of order r on $\mathfrak{X}^*(\mathbb{M}(d))$, then for any sequence of i.i.d. random motions G_1, $G_2, \ldots$ it easily follows that

$$\mu(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n); H_\alpha) \leqslant n^{1-(r/\alpha)}\mu(G_1, H_\alpha) \tag{15.1.13}$$

is a 'right order' estimate for the rate of convergence in the CLT. Estimates such as these will play a crucial role in all that follows.

The next result clarifies the relation between the ideal metrics $\boldsymbol{\mu}_r$, $\mathbf{v}_r$ and $\boldsymbol{\zeta}_r$. It shows that upper bounds for the rate of the convergence problem, when expressed in terms of $\boldsymbol{\zeta}_r$ are necessarily weaker than bounds expressed in terms of either $\boldsymbol{\mu}_r$ or $\mathbf{v}_r$ (as in Theorems 15.2.1 and 15.2.2 below).

Theorem 15.1.3. For any G_1 and $G_2 \in \mathfrak{X}(\mathbb{M}(d))$

$$\boldsymbol{\mu}_r(G_1, G_2) \leqslant C_1(r)\boldsymbol{\zeta}_{r-1}(G_1, G_2) \quad r \geqslant 1 \tag{15.1.14}$$

and

$$\mathbf{v}_r(G_1, G_2) \leqslant C_2(r)\boldsymbol{\zeta}_r(G_1, G_2) \quad r \geqslant 0, r\text{-integer} \tag{15.1.15}$$

where $C_i(r)$ is a constant depending only on r.

The proof follows from the similar inequalities between $\boldsymbol{\mu}_r$, $\mathbf{v}_r$ and $\boldsymbol{\zeta}_r$ in the space $\mathfrak{X}(\mathbb{R}^d)$ (see Section 14.2, Lemmas 14.2.4, 14.2.5, 14.2.6). As far as the finiteness of ζ_r (G_1, G_2) is concerned we have that the condition

$$\left| \sum_{\substack{0 \leqslant i_1, \ldots, i_d \leqslant d \\ i_1 + \cdots + i_d = j}} \int_{\mathbb{R}^d} y_1^{i_1} \circ \cdots \circ y_d^{i_d} (\Pr(G_1(x) \in \mathrm{d}y) - \Pr(G_2(x) \in \mathrm{d}y)) \right| = 0 \quad (15.1.16)$$

for all $x \in \mathbb{R}^d$, $j = 0, 1, \ldots, m$, $m + \beta = r$, $\beta \in (0, 1]$, m-integer, implies

$$\zeta_r(G_1, G_2) \leqslant \frac{1}{\Gamma(1 + r)} \mathbf{Var}_r(G_1, G_2), \quad (15.1.17)$$

where the metric $\mathbf{Var}_r$ is the rth absolute pseudomoment in $\mathfrak{X}(\mathbb{M}(d))$, that is

$$\mathbf{Var}_r(G_1, G_2) := \sup_{x \in \mathbb{R}^d} \int \|y\|^r |\Pr_{G_1(x)} - \Pr_{G_2(x)}|(\mathrm{d}y). \quad (15.1.18)$$

15.2 RATES OF CONVERGENCE IN THE INTEGRAL AND LOCAL CLTS FOR RANDOM MOTIONS

Let $G_1, G_2 \ldots$ be a sequence of i.i.d random motions and H_α an α-stable random motion. We seek precise order estimates for the rate of convergence

$$n^{-1/\alpha}(G_1 \circ \cdots \circ G_n) \to H_\alpha \quad (15.2.1)$$

in terms of Kolmogorov's metric $\boldsymbol{\rho}$, $\mathbf{Var}$ and ℓ distances on $\mathfrak{X}(\mathbb{M}(d))$. Here, the uniform (Kolmogorov's) metric between random motions G and G^* is defined by

$$\boldsymbol{\rho}(G, G^*) := \sup_{x \in \mathbb{R}^d} \boldsymbol{\rho}(G(x), G^*(x)) \quad (15.2.2)$$

where $\boldsymbol{\rho}(X, Y)$ is the usual Kolmogorov distance between the d-dimensional random vectors X and Y in $\mathfrak{X}(\mathbb{R}^d)$, that is

$$\boldsymbol{\rho}(X, Y) := \sup_{A \in \mathbb{C}} |\Pr\{X \in A\} - \Pr\{Y \in A\}| \quad (15.2.3)$$

$\mathbb{C}$ being the convex Borel sets in $\mathbb{R}^d$. Recall that the total variation metric $\mathbf{Var}$ in $\mathfrak{X}(\mathbb{M}(d))$ is defined by (15.1.2) and ℓ in $\mathfrak{X}^*(\mathbb{M}(d))$ is given by

$$\ell(G, G^*) := \sup_{x \in \mathbb{R}^d} \ell(G(x), G^*(x)) \quad (15.2.4)$$

(cf. (15.1.5)).

The first result obtains rates with respect to $\boldsymbol{\rho}$. Here and henceforth C denotes an absolute constant whose value may change from line to line.

The next theorem establishes the estimates of the uniform rate of convergence in the *integral* CLT for random motions.

Theorem 15.2.1. Let $r > \alpha$ and set $\rho := \boldsymbol{\rho}(G_1, H_\alpha)$ and $\tau_r := \tau_r(G_1, H_\alpha) := \max\{\rho, v_r, v_\alpha^{1/(r-\alpha)}\}$. Then

$$\boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \leqslant C(v_r n^{1-r/\alpha} + \tau_r n^{-1/\alpha}). \tag{15.2.5}$$

Proof. As in Section 14.3, it is helpful to establish first three smoothing inequalities for $\boldsymbol{\rho}$ and $\mathbf{Var}$. Throughout, recall that H_α has components $Y_\alpha (\stackrel{d}{=} \theta)$ and U_α and let $\bar{H}_\alpha$ denote the projection of H_α on $\mathbb{R}^d$. The purpose of the next lemma is to transfer the problem of estimating the $\boldsymbol{\rho}$-distance between two random motions to the same problem involving smoothed random motions. Here and in the following, $G \circ \tilde{G}$ means that $G \circ \tilde{G}$ is a random motion whose distribution is a convolution of the distributions of G and $\tilde{G}$. QED

Lemma 15.2.1. For any G and G^* in $\mathfrak{X}(\mathbb{M}(d))$ and $\delta > 0$

$$\boldsymbol{\rho}(G, G^*) \leqslant C\boldsymbol{\rho}(\delta\bar{H}_\alpha \circ G, \delta\bar{H}_\alpha \circ G^*) + C\delta \tag{15.2.6}$$

where C is an absolute constant.

Proof of Lemma 15.2.1. The required inequality is a slight extension of the 'smoothing inequality' in $\mathfrak{X}(\mathbb{R}^d)$ (see Paulauskas, 1974, 1976); Zolotarev, 1986 (Lemma 5.4.2) and Bhattacharya and Ranga Rao, 1976 (Lemma 12.1)):

$$\boldsymbol{\rho}(X, Y) \leqslant C\boldsymbol{\rho}(X + \delta\theta, Y + \delta\theta) + C\delta \quad X, Y \in \mathfrak{X}(\mathbb{R}^d) \tag{15.2.7}$$

where θ is a spherically symmetric α-stable random vector independent of X and Y and C is a constant depending upon α and d only. By (15.2.7) we have

$$\begin{aligned}
\boldsymbol{\rho}(G, G^*) &= \sup_{x \in \mathbb{R}^d} \boldsymbol{\rho}(Y + Ux, Y^* + U^*x) \\
&\leqslant C \sup_{x \in \mathbb{R}^d} \boldsymbol{\rho}(\delta\theta + Y + Ux, \delta\theta + Y^* + U^*x) + C\delta \\
&= C\boldsymbol{\rho}(\delta\bar{H}_\alpha \circ G, \delta\bar{H}_\alpha \circ G^*) + C\delta.
\end{aligned}$$

QED

The next estimate is the analog of Lemma 14.3.1 and will be used several times in the proof.

Lemma 15.2.2. Let $G, \tilde{G} \in \mathfrak{X}(\mathbb{M}(d))$, $\lambda_i \geqslant 0$, $i = 1, 2$; $\lambda^\alpha := \lambda_1^\alpha + \lambda_2^\alpha$; $\tilde{H}_\alpha \stackrel{d}{=} H_\alpha$. For all $r > 0$

$$\mathbf{Var}(\lambda_1 H_\alpha \circ G \circ \lambda_2 \tilde{H}_\alpha, \lambda_1 H_\alpha \circ \tilde{G} \circ \lambda_2 \tilde{H}_\alpha) \leqslant \lambda^{-r} \mathbf{v}_r(G, \tilde{G}). \tag{15.2.8}$$

Proof of Lemma 15.2.2. Let $\tilde{H}_\alpha := (\tilde{Y}_\alpha, \tilde{U}_\alpha)$, $G := (Y, U)$, and $\tilde{G} := (\tilde{Y}, \tilde{U})$. Then by definition of the **Var** metric

$$
\begin{aligned}
&\mathbf{Var}(\lambda_1 H_\alpha \circ G \circ \lambda_2 \tilde{H}_\alpha, \lambda_1 H_\alpha \circ \tilde{G} \circ \lambda_2 \tilde{H}_\alpha)\\
&\qquad = \sup_x \mathbf{Var}(\lambda_1 H_\alpha \circ G \circ (\lambda_2 \tilde{Y}_\alpha + \tilde{U}_\alpha x), \lambda_1 H_\alpha \circ \tilde{G} \circ (\lambda_2 \tilde{Y}_\alpha + \tilde{U}_\alpha x))\\
&\qquad = \sup_x \mathbf{Var}(\lambda_1 H_\alpha (Y + U\lambda_2 \tilde{Y}_\alpha + U\tilde{U}_\alpha x), \lambda_1 H_\alpha(\tilde{Y} + \tilde{U}\lambda_2 \tilde{Y}_\alpha + U\tilde{U}_\alpha x))\\
&\qquad = \sup_x \mathbf{Var}(\lambda_1 Y_\alpha + U_\alpha(Y + U\lambda_2 \tilde{Y}_\alpha) + U_\alpha U\tilde{U}_\alpha x, \lambda_1 Y_\alpha\\
&\qquad\qquad + U_\alpha(\tilde{Y} + \tilde{U}\lambda_2 \tilde{Y}_\alpha) + U_\alpha \tilde{U}\tilde{U}_\alpha x).\\
&\qquad = \sup_x \mathbf{Var}(\lambda_1 Y_\alpha + U_\alpha Y + \lambda_2 \tilde{Y}_\alpha + U_\alpha U\tilde{U}_\alpha x, \lambda_1 Y_\alpha\\
&\qquad\qquad + U_\alpha \tilde{Y} + \lambda_2 \tilde{Y} + U_\alpha U\tilde{U}_\alpha x).
\end{aligned}
$$

Using $\lambda_1 Y_\alpha + \lambda_2 \tilde{Y}_\alpha \stackrel{\text{d}}{=} \lambda Y_\alpha$, the right-hand side equals

$$
\begin{aligned}
&\sup_x \mathbf{Var}(\lambda Y_\alpha + U_\alpha(Y + U\tilde{U}_\alpha x), \lambda Y_\alpha + U_\alpha(\tilde{Y} + \tilde{U}\tilde{U}_\alpha x))\\
&\qquad \leqslant \lambda^{-r} \sup_x \sup_{h \in \mathbb{R}} |\lambda h|^r \mathbf{Var}(h\lambda Y_\alpha + U_\alpha(Y + U\tilde{U}_\alpha x), h\lambda Y_\alpha + U_\alpha(\tilde{Y} + \tilde{U}\tilde{U}_\alpha x))\\
&\qquad = \lambda^{-r} \sup_x \mathbf{v}_r(U_\alpha(Y + U\tilde{U}_\alpha x), U_\alpha(\tilde{Y} + \tilde{U}\tilde{U}_\alpha x))\\
&\qquad = \lambda^{-r} \sup_x \mathbf{v}_r(Y + U\tilde{U}_\alpha x, \tilde{Y} + \tilde{U}U_\alpha x))\\
&\qquad = \lambda^{-r} \mathbf{v}_r(G, \tilde{G})
\end{aligned}
$$

by definition of $\mathbf{v}_r$, and since **Var** (and hence $\mathbf{v}_r$) is invariant with respect to rotations. QED

The third and final lemma may be considered as the analog of Lemma 14.3.2.

Lemma 15.2.3. For any G_1^*, G_2^*, $\tilde{G}_1$, $\tilde{G}_2$ in $\mathfrak{X}(\mathbb{M}(d))$ and $\lambda \geqslant 0$

$$
\begin{aligned}
&\boldsymbol{\rho}(\lambda \bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda \bar{H}_\alpha \circ G_1^* \circ \tilde{G}_2)\\
&\qquad \leqslant \boldsymbol{\rho}(G_1^*, G_2^*)\, \mathbf{Var}(\lambda \bar{H}_\alpha \circ \tilde{G}_1, \lambda \bar{H}_\alpha \circ \tilde{G}_2) + \boldsymbol{\rho}(\lambda \bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1, \lambda \bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2).
\end{aligned} \tag{15.2.9}
$$

Also,

$$\rho(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1) \leqslant \rho(G_1^*, G_2^*)\ \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) + \rho(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_2, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2)$$

and

$$\mathbf{Var}(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1) \leqslant \mathbf{Var}(G_1^*, G_2^*)\ \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) + \mathbf{Var}(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_2, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2). \tag{15.2.10}$$

Proof. We shall prove only (15.2.9). The proof of the other two inequalities is similar. We have

$$\begin{aligned}
\rho(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_2) &= \rho(G_1^* \circ \lambda\bar{H}_\alpha \circ G_1, G_1^* \circ \lambda\bar{H}_\alpha \circ G_2) \\
&= \sup_{x \in \mathbb{R}^d} \sup_{A \in \mathbb{C}} \left| \int_{\mathbb{M}(d)} \Pr\{G_1^* \circ g(x) \in A\}(\lambda\bar{H}_\alpha \circ \tilde{G}_1 - \lambda\bar{H}_\alpha \circ \tilde{G}_2)\, \mathrm{d}g \right| \\
&\leqslant \sup_{x \in \mathbb{R}^d} \sup_{A \in \mathbb{C}} \left| \int (\Pr\{G_1^* \circ g(x) \in A\} \right. \\
&\qquad \left. - \Pr\{G_2^* \circ g(x) \in A\})(\lambda\bar{H}_\alpha \circ \tilde{G}_1 - \lambda\bar{H}_\alpha \circ \tilde{G}_2)\, \mathrm{d}g \right| \\
&\qquad + \sup_{x \in \mathbb{R}^d} \sup_{A \in \mathbb{C}} \left| \int \Pr\{G_2^* \circ g(x) \in A\}(\lambda\bar{H}_\alpha \circ \tilde{G}_1 - \lambda\bar{H}_\alpha \circ \tilde{G}_2)\, \mathrm{d}g \right| \\
&\leqslant \rho(G_1^*, G_2^*)\ \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) + \rho(G_2^* \circ \lambda\bar{H}_\alpha \circ \tilde{G}_1, G_2^* \circ \lambda\bar{H}_\alpha \circ \tilde{G}_2) \\
&= \rho(G_1^*, G_2^*)\ \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) + \rho(\lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2).
\end{aligned}$$

QED

From these three lemmas, Theorem 15.2.1 may now be proved. The proof uses induction on n. First, note that for a fixed n_0 and $n \leqslant n_0$, the estimate (15.2.5) is an obvious consequence of the hypotheses. Thus, let $n \geqslant n_0$ and assume that for any $j < n$

$$\rho(j^{-1/\alpha}(G_1 \circ \cdots \circ G_j), H_\alpha) \leqslant B(v_r j^{1-r/\alpha} + \tau_r j^{-1/\alpha}) \tag{15.2.11}$$

where B is an absolute constant.

Remark 15.2.1. We shall use the main idea of Theorem 2 of Senatov (1980), where the case $\alpha = 2$ is considered and rates of convergence for CLT of random vectors in terms of ζ_r are investigated.

Set $m = [n/2]$ and

$$\delta := A \max(\mathbf{v}_1(G_1, H_\alpha), \mathbf{v}_r^{1/(r-\alpha)}(G_1, H_\alpha))n^{-1/\alpha} \tag{15.2.12}$$

where A is a constant to be determined later. Note that $\delta \leqslant A\tau_r n^{-1/\alpha}$, which will be used in the sequel.

Let $G_1', G_2', \ldots$ be a sequence of i.i.d. random motions with $G_i' \stackrel{d}{=} H_\alpha$. By the definition of symmetric α-stable and Lemma 15.2.1, it follows that

$$\begin{aligned}
&\boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \\
&\quad = \boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), n^{-1/\alpha}(G_1' \circ \cdots \circ G_n')), \\
&\quad \leqslant C\boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), \delta\bar{H}_\alpha \circ n^{-1/\alpha}(G_1' \circ \cdots \circ G_n')) + C\delta.
\end{aligned} \tag{15.2.13}$$

Applying the triangle inequality m times, the first term in (15.2.13) is bounded by

$$\begin{aligned}
&\boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1 \circ \cdots \circ n^{-1/\alpha}G_n, \delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1 \circ \cdots \circ n^{-1/\alpha}G_{n-1} \circ n^{-1/\alpha}G_n') \\
&\quad + \sum_{j=1}^{m} \boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1 \circ \cdots \circ n^{-1/\alpha}G_{n-j} \circ n^{-1/\alpha}G_{n-j+1}' \circ \cdots \circ n^{-1/\alpha}G_n', \\
&\qquad\qquad \delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1 \circ \cdots \circ n^{-1/\alpha}G_{n-j-1} \circ n^{-1/\alpha}G_{n-j}' \circ \cdots \circ n^{-1/\alpha}G_n') \\
&\quad + \boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1 \circ \cdots \circ n^{-1/\alpha}G_{n-m-1} \circ n^{-1/\alpha}G_{n-m}' \circ \cdots \circ n^{-1/\alpha}G_n', \\
&\qquad\qquad \delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1' \circ \cdots \circ n^{-1/\alpha}G_n') \\
&\quad =: A_1 + A_2 + A_3.
\end{aligned} \tag{15.2.14}$$

Next, using Lemma 15.2.3, A_1 and A_2 may be bounded as follows

$$\begin{aligned}
A_1 &\leqslant \boldsymbol{\rho}(n^{-1/\alpha}G_1 \circ \cdots \circ n^{-1/\alpha}G_{n-1}, n^{-1/\alpha}G_1' \circ \cdots \circ n^{-1/\alpha}G_{n-1}') \\
&\quad \times \mathbf{Var}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_n, \delta\bar{H}_\alpha \circ n^{-1/\alpha}G_n') \\
&\quad + \boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1' \circ \cdots \circ n^{-1/\alpha}G_{n-1}' \circ n^{-1/\alpha}G_n, \delta\bar{H}_\alpha \circ n^{-1/\alpha}G_1' \circ \cdots \circ n^{-1/\alpha}G_n') \\
&=: I_1 + I_3'.
\end{aligned} \tag{15.2.15}$$

Similarly,

$$\begin{aligned}
A_2 &\leqslant \sum_{j=1}^{m} \boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_{n-j-1}), n^{-1/\alpha}(G_1' \circ \cdots \circ G_{n-j-1}')) \\
&\quad \times \mathbf{Var}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_{n-j} \circ n^{-1/\alpha}G_{n-j+1}' \circ \cdots \circ n^{-1/\alpha}G_n', \\
&\qquad\qquad \delta\bar{H}_\alpha \circ n^{-1/\alpha}G_{n-j}' \circ n^{-1/\alpha}G_{n-j+1}' \circ \cdots \circ n^{-1/\alpha}G_n') \\
&\quad + \sum_{j=1}^{m} \boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{1/\alpha}G_1' \circ \cdots \circ n^{-1/\alpha}G_{n-j-1}' \circ n^{-1/\alpha}G_{n-j} \circ \cdots \circ n^{-1/\alpha}G_n', \\
&\qquad\qquad \delta\bar{H}_\alpha \circ n^{-1/\alpha}(G_1' \circ \cdots \circ G_n')) := I_2 + I_3''.
\end{aligned} \tag{15.2.16}$$

Combining (15.2.13)–(15.2.16) and letting $I_3 := I_3' + I_3''$, $I_4 := A_3$ yields

$$\boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \leqslant C(I_1 + I_2 + I_3 + I_4) + C\delta. \tag{15.2.17}$$

Next, Lemma 15.2.2 will be used to successively estimate each of the quantities I_1, I_2, I_3 and I_4.

By the induction hypothesis, Lemma 15.2.2 (with $\lambda_1 = \lambda = \delta$ and $\lambda_2 = 0$ there), and the ideality of $\mathbf{v}_r$ it follows that

$$I_1 \leqslant B(v_r(n-1)^{1-r/\alpha} + \tau_r(n-1)^{1-r/\alpha})v_1(n^{-1/\alpha}G_1, n^{-1/\alpha}\tilde{G}_1)/\delta$$
$$\leqslant C(B/A)(v_r n^{1-r/\alpha} + \tau_r n^{1-r/\alpha}) \tag{15.2.18}$$

by definition of δ.

To estimate I_2, apply the induction hypothesis again, Lemma 15.2.2 (with $\lambda_1 = \delta$, $\lambda_2 = (j/n)^{1/\alpha}$), ideality of $\mathbf{v}_r$ and the definition of δ to obtain

$$I_2 = \sum_{j=1}^{m} \boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_{n-j-1}), n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_{n-j-1}))$$
$$\mathbf{Var}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}G_{n-j} \circ (j/n)^{1/\alpha}H_\alpha, \delta\bar{H}_\alpha \circ n^{-1/\alpha}G'_{n-j} \circ (j/n)^{1/\alpha}H_\alpha)$$
$$\leqslant B(v_r(n-m)^{1-r/\alpha} + \tau_r(n-m)^{-1/\alpha}) \sum_{j=1}^{m} \frac{1}{(\delta^\alpha + j/n)^{r/\alpha}} v_r(n^{-1/\alpha}G_1, n^{-1/\alpha}G'_1)$$
$$\leqslant B(v_r n^{1-r/\alpha} + \tau_r n^{-1/\alpha}) \sum_{j=1}^{\infty} v_r/(A^\alpha v_r^{\alpha/(r-\alpha)} + j)^{r/\alpha}$$
$$\leqslant B(v_n n^{1-r/\alpha} + \tau_r n^{-1/\alpha})C(A^\alpha v_r^{\alpha/(r-\alpha)})^{1-(r/\alpha)} v_r$$
$$\leqslant CB(v_r n^{1-r/\alpha} + \tau_r n^{-1/\alpha})/A^{\alpha-r} \tag{15.2.19}$$

where C again denotes some absolute constant.

To estimate

$$I_3 = \sum_{j=0}^{m} \boldsymbol{\rho}(\delta\bar{H}_\alpha \circ n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_{n-j-1} \circ G_{n-j} \circ G'_{n-j+1} \circ \cdots \circ G'_n),$$
$$\delta\bar{H}_\alpha \circ n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_n))$$

use $2\boldsymbol{\rho} \leqslant \mathbf{Var}$, Lemma 15.2.2 (with $\lambda_1 = ((n-j-1)/(n-1))^{1/\alpha}$, $\lambda_2 = (j/(n-1))^{1/\alpha}$ and $\lambda = 1$) and the ideality of $\mathbf{v}_r$ to obtain

$$I_3 \leqslant \sum_{j=0}^{m} \mathbf{v}_r\left(\left(\frac{n-j-1}{n-1}\right)^{1/\alpha} H_\alpha \circ (n-1)^{-1/\alpha}G_{n-j} \circ \left(\frac{j}{n-1}\right)^{1/\alpha} H_\alpha,\right.$$
$$\left.\left(\frac{n-j-1}{n-1}\right)^{1/\alpha} H_\alpha \circ (n-1)^{1/\alpha}G'_{n-j} \circ \left(\frac{j}{n-1}\right)^{1/\alpha} H_\alpha\right)$$
$$\leqslant \sum_{j=0}^{m} \mathbf{v}_r((n-1)^{-1/\alpha}G_1, (n-1)^{-1/\alpha}G'_1)$$
$$\leqslant n^{1-r/\alpha}\mathbf{v}_r \tag{15.2.20}$$

where it is assumed that n_0 is chosen so that $(n/(n-1))^{r/\alpha} \leqslant 2$.

Similarly, using Lemma 15.2.2 with $\lambda_1 = 0$ and $\lambda_2 = 1$, we may bound I_4

$$\begin{aligned}
I_4 &\leqslant \boldsymbol{\rho}(m^{-1/\alpha}(G_1 \circ \cdots \circ G_{n-m-1}) \circ m^{-1/\alpha}(G'_{n-m} \circ \cdots \circ G'_n), \\
&\qquad\qquad m^{-1/\alpha}(G'_1 \circ \cdots \circ G'_{n-m-1}) \circ m^{-1/\alpha}(G'_{n-m} \circ \cdots \circ G'_n)) \\
&= \boldsymbol{\rho}(\delta \bar{H}_\alpha \circ m^{-1/\alpha}(G_1 \circ \cdots \circ G_{n-m-1}) \circ H_\alpha, \delta \bar{H}_\alpha \circ m^{-1/\alpha}(G'_1 \circ \cdots \circ G'_{n-m-1}) \circ H_\alpha) \\
&\leqslant \mathbf{Var}(m^{-1/\alpha}(G_1 \circ \cdots \circ G_{n-m-1}) \circ H_\alpha, m^{-1/\alpha}(G'_1 \circ \cdots \circ G'_{n-m-1}) \circ H_\alpha) \\
&\leqslant \mathbf{v}_r(m^{-1/\alpha}(G_1 \circ \cdots \circ G_{n-m-1}), m^{-1/\alpha}(G'_1 \circ \cdots \circ G'_{n-m-1})) \\
&\leqslant m^{-r/\alpha}(n - m - 1)\mathbf{v}_r(G_1, G'_1) \leqslant 2^{r/\alpha} n^{1-r/\alpha} v_r \qquad (15.2.21)
\end{aligned}$$

since we may assume that $((n - m - 1)/n)(n/m)^{r/\alpha}$ is bounded by $2^{r/\alpha}$ for $n \geqslant n_0$.

Finally, combining the estimates (15.2.17)–(15.2.21) and the definition of δ yields

$$\begin{aligned}
\boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) &\leqslant C(I_1 + I_2 + I_3 + I_4) + C\delta \\
&\leqslant C(A^{-1} + A^{\alpha-r})B(v_r n^{1-r/\alpha} + \tau_r n^{-1/\alpha}) + Cv_r n^{1-r/\alpha} + CA\tau_r n^{-1/\alpha}.
\end{aligned}$$

Choosing the absolute constant A such that $C(A^{-1} + A^{\alpha-r}) \leqslant \frac{1}{2}$ shows, for sufficiently large B

$$\boldsymbol{\rho}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \leqslant B(v_r n^{1-r/\alpha} + \tau_r n^{-1/\alpha})$$

completing the proof of Theorem 15.2.1. QED

The main theorem in the second part of this section deals with uniform rates of convergence in the *local* limit theorem on $\mathbb{M}(d)$ (see further Theorem 15.2.2). Again, ideal smoothing metrics play a considerable role. More precisely, if $\{G_i\} = \{Y_i, U_i\}_{i \geqslant 1}$ are i.i.d. random motions and $G_1(x)$ has a density $p_{G_1(x)}$ for any $X \in \mathbb{R}^d$, then ideal metrics are used to determine the rate of convergence in the limit relationship

$$\ell(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \to 0 \qquad (15.2.22)$$

where ℓ is determined by (15.1.5) and (15.2.4).

The result considers rates in (15.2.22) under hypotheses on $v_r := \mathbf{v}_r(G_1, H_\alpha)$, $\ell := \ell(G_1, H_\alpha)$ and, $\mu_r := \boldsymbol{\mu}_r(G_1, H_\alpha)$, where

$$\begin{aligned}
\boldsymbol{\mu}_r(G_1, H_\alpha) &:= \sup_{x \in \mathbb{R}^d} \sup_{h \in \mathbb{R}} |h|^r \ell(G_1(x) + h\theta, H_\alpha(x) + h\theta) \\
&= \sup_{x \in \mathbb{R}} |h|^r \ell((h\bar{H}_\alpha) \circ G_1, (h\bar{H}_\alpha) \circ H_\alpha) \qquad (15.2.23)
\end{aligned}$$

and $\bar{H}_\alpha := (Y_\alpha, I)$ denotes, as before, the projection of H_α on $\mathbb{R}^d$.

The proof of the next theorem depends heavily upon the ideality of $\mathbf{v}_r$ and $\boldsymbol{\mu}_r$. As in the proof of Theorem 15.2.1 ideality is first used to establish some critical smoothing inequalities. The first smoothing inequality provides a rate of convergence in (15.2.1) with respect to the **Var**-metric and could actually be considered as a companion lemma to the main result. The proof of the next lemma is similar to that of Theorem 14.3.1 and is thus omitted.

Lemma 15.2.4. Let $r > \alpha$ and

$$K_r := K_r(G_1, H_\alpha) := \max\{\mathbf{Var}(G_1, H_\alpha), \mathbf{v}_r(G_1, H_\alpha)\} \leqslant a$$

where $a^{-1} := 2^{1+r/\alpha}(2^{(r/\alpha)-1} + 3^{r/\alpha})$. If $A := 2(2^{(r/\alpha)-1} + 3^{r/\alpha})$ then

$$\mathbf{Var}(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \leqslant AK_r n^{1-r/\alpha}.$$

The next estimate, the companion to Lemma 15.2.2 is the analog of Lemma 14.3.5.The proof is similar to that of Lemma 15.2.2 and will be omitted.

Lemma 15.2.5. Let $G_1, G_2 \in \mathfrak{X}(\mathbb{M}(d))$, $\lambda_i > 0$, $i = 1, 2$; $\lambda^\alpha = \lambda_1^\alpha + \lambda_2^\alpha$, $\tilde{H}_\alpha \stackrel{\mathrm{d}}{=} H_\alpha$. For all $r > 0$

$$\ell(\lambda_1 H_\alpha \circ G_1 \circ \lambda_2 \tilde{H}_\alpha, \lambda_1 H_\alpha \circ G_2 \circ \lambda_2 \tilde{H}_\alpha) \leqslant \lambda^{r-1} \boldsymbol{\mu}_r(G_1, G_2) \tag{15.2.24}$$

and

$$\ell(\lambda_1 \bar{H}_\alpha \circ G_1 \lambda_2 \tilde{H}_\alpha, \lambda_1 \bar{H}_\alpha \circ G_2 \circ \lambda_2 \tilde{H}_\alpha) \leqslant \lambda^{r-1} \boldsymbol{\mu}_r(G_1, G_2). \tag{15.2.25}$$

The following smoothing inequality may be considered as the analog of Lemma 14.3.7. Only (15.2.27) is used in the sequel.

Lemma 15.2.6. Let $G_1^*, G_2^*, \tilde{G}_1, \tilde{G}_2 \in \mathfrak{X}(\mathbb{M}(d))$ and $\lambda \geqslant 0$. Then

$$\begin{aligned} \ell(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_2) &\leqslant \ell(G_1^*, G_2^*)\, \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) \\ &\quad + \ell(\lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2) \end{aligned} \tag{15.2.26}$$

and

$$\begin{aligned} \ell(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1) &\leqslant \ell(G_1^*, G_2^*)\, \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) \\ &\quad + \ell(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_2, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2). \end{aligned} \tag{15.2.27}$$

Proof. Since $\bar{H}_\alpha \circ G = G \circ \bar{H}_\alpha$ we see that $\ell(\lambda\bar{H}_\alpha \circ G_1^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2)$ equals

$$\begin{aligned} &\ell(G_1^* \circ \lambda\bar{H}_\alpha \circ \tilde{G}_1, G_1^* \circ \lambda\bar{H}_\alpha \circ \tilde{G}_2) \\ &= \sup_{x \in \mathbb{R}^d} \operatorname*{ess\,sup}_{z \in \mathbb{R}^d} |p_{G_1^* \circ \lambda\bar{H}_\alpha \circ \tilde{G}_1(x)}(z) - p_{G_1^* \circ \lambda\bar{H}_\alpha \circ \tilde{G}_2(x)}(z)| \\ &= \sup_x \operatorname*{ess\,sup}_z \left| \int_{\mathbb{M}(d)} p_{G_1^* \circ g(x)}(z)[\Pr(\lambda\bar{H}_\alpha \circ \tilde{G}_1 \in \mathrm{d}g) - \Pr(\lambda\bar{H}_\alpha \circ \tilde{G}_2 \in \mathrm{d}g] \right| \\ &\leqslant \sup_x \operatorname*{ess\,sup}_z |p_{G_1^* \circ g(x)}(z) - p_{G_2^* \circ g(x)}(z)| \int_{\mathbb{M}(d)} |\Pr(\lambda\bar{H}_\alpha \circ \tilde{G}_1 \in \mathrm{d}g) \\ &\quad - \Pr(\lambda\bar{H}_\alpha \circ \tilde{G}_2 \circ \mathrm{d}g)| \\ &\quad + \ell(\lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2) \\ &\leqslant \ell(G_1^*, G_2^*)\, \mathbf{Var}(\lambda\bar{H}_\alpha \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ \tilde{G}_2) + \ell(\lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_1, \lambda\bar{H}_\alpha \circ G_2^* \circ \tilde{G}_2). \end{aligned}$$

This proves (15.2.26); (15.2.27) is proved similarly. QED

With these three smoothing inequalities the main result may now be proved.

Theorem 15.2.2. Let the following two conditions hold

$$\lambda_r(G_1, H_\alpha) := \max\{\ell(G_1, H_\alpha), \boldsymbol{\mu}_{r+1}(G_1, H_\alpha)\} < \infty \tag{15.2.28}$$

and

$$K_r := K_r(G_1, H_\alpha) := \max\{\mathbf{Var}(G_1, H_\alpha), \mathbf{v}_r(G_1, H_\alpha)\} \leqslant 1/DA \tag{15.2.29}$$

where $r > \alpha$, $A := 2(2^{(r/\alpha)-1} + 3^{r/\alpha})$ and $D := 2(3^{-1+(r+1)/\alpha})$. Then

$$\ell(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \leqslant A\lambda_r(G_1, H_\alpha)n^{1-r/\alpha}. \tag{15.2.30}$$

Proof. Let $G'_1, G'_2, \ldots$ be a sequence of i.i.d. random motions with $G'_i \stackrel{d}{=} H_\alpha$. Now (15.2.30) holds for $n = 1$, 2 and 3. Let $n > 3$.

Suppose that for all $j < n$

$$\ell(j^{-1/\alpha}(G_1 \circ \cdots \circ G_j), H_\alpha) \leqslant A\lambda_r j^{1-r/\alpha}. \tag{15.2.31}$$

To complete the induction proof, it only remains to show that (15.2.31) holds for $j = n$. By (15.2.27) with $\lambda = 0$ and $m = [n/2]$

$$\ell(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_n))$$

is bounded by

$$\begin{aligned}
&\leqslant \ell((n^{-1/\alpha}(G_1 \circ \cdots \circ G_m) \circ n^{-1/\alpha}(G_{m+1} \circ \cdots \circ G_n),\\
&\qquad\qquad n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m) \circ n^{-1/\alpha}(G_{m+1} \circ \cdots \circ G_n))\\
&\quad + \ell(n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m) \circ n^{-1/\alpha}(G_{m+1} \circ \cdots \circ G_n),\\
&\qquad\qquad n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m) \circ n^{-1/\alpha}(G'_{m+1} \circ \cdots \circ G'_n))\\
&\leqslant \ell(n^{-1/\alpha}(G_1 \circ \cdots \circ G_m), n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m))\, \mathbf{Var}(n^{-1/\alpha}(G_{m+1} \circ \cdots \circ G_n),\\
&\qquad\qquad n^{-1/\alpha}(G'_{m+1} \circ \cdots \circ G'_n))\\
&\quad + \ell(n^{-1/\alpha}(G_1 \circ \cdots \circ G_m) \circ n^{-1/\alpha}(G'_{m+1} \circ \cdots \circ G'_n),\\
&\qquad\qquad n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m) \circ n^{-1/\alpha}(G'_{m+1} \circ \cdots \circ G'_n))\\
&\quad + \ell(n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m) \circ n^{-1/\alpha}(G_{m+1} \circ \cdots \circ G_n),\\
&\qquad\qquad n^{-1/\alpha}(G'_1 \circ \cdots \circ G'_m) \circ n^{-1/\alpha}(G'_{m+1} \circ \cdots \circ G'_n))\\
&:= I_1 + I_2 + I_3.
\end{aligned}$$

As in the proof of Lemma 15.2.4, it may be shown via Lemma 15.2.5 that

$$I_2 + I_3 \leqslant (2^{(r/\alpha)-1} + 3^{r/\alpha})\mu_{r+1} n^{1-r/\alpha} \leqslant \tfrac{1}{2}A\lambda_r n^{1-r/\alpha}.$$

It remains to estimate I_1. By the homogeneity property $\ell(X, Y) = c\ell(cX, cY)$

and the induction hypothesis, the first factor in I_1 is bounded by

$$(n/m)^{1/\alpha}\ell(m^{-1/\alpha}(G_1 \circ \cdots \circ G_m), H_\alpha) \leqslant (n/m)^{1/\alpha} A\lambda_r m^{1-r/\alpha} \leqslant 3^{(r+1)/\alpha - 1} A\lambda_r n^{1-r/\alpha}.$$

By Lemma 15.2.4 the second factor in I_1 is bounded by

$$\begin{aligned}&\mathbf{Var}(n^{-1/\alpha}(G_{m+1} \circ \cdots \circ G_n), n^{-1/\alpha}(G'_{m+1} \circ \cdots \circ G'_n))\\ &\qquad \leqslant AK_r(n-m)^{1-r/\alpha} \leqslant AK_r \leqslant D^{-1}.\end{aligned}$$

Hence, $I_1 \leqslant \frac{1}{2}A\lambda_r n^{1-r/\alpha}$. Combining this with the displayed bound on $I_1 + I_2$ shows that

$$\ell(n^{-1/\alpha}(G_1 \circ \cdots \circ G_n), H_\alpha) \leqslant A\lambda_r n^{1-r/\alpha}$$

as desired. QED

The conditions (15.2.28), (15.2.29) in Theorem 15.2.2 as well as conditions $v_r = \mathbf{v}_r(G_1, H_\alpha) < \infty$, $\tau_r = \boldsymbol{\tau}_r(G_1, H_\alpha) < \infty$ in Theorem 15.2.1 can be examined via Theorem 15.1.3 and the estimate (15.1.16).

CHAPTER 16

Applications of Ideal Metrics for Sums of i.i.d. Random Variables to the Problems of Stability and Approximation in Risk Theory

16.1 THE PROBLEM OF STABILITY IN RISK THEORY

When using a stochastic model in insurance risk theory one has to consider this model as an approximation of the real insurance activities. The stochastic elements derived from these models represent an idealization of the real insurance phenomena under consideration. Hence the problem arises of establishing the limits in which one can use our 'ideal' model. The practitioner has to know the accuracy of our recommendations, resulting from our investigations based on the ideal model.

Mostly one deals with the real insurance phenomena including the following main elements: input data (epochs of claims, size of claims, ...) and, resulting from these, the output data (number of claims up to time t, total claim amount ...)

In this section we apply the method of metric distances (cf. Fig. 1.1.1) to investigate the 'horizon' of widely used stochastic models in insurance mathematics. The main stochastic elements of these models are the following.

(a) *model input elements*:
the *epochs of the claims*, denoted by $T_0 = 0, T_1, T_2, \ldots,$ where $\{W_i = T_i - T_{i-1}; i = 1, 2, \ldots\}$ is a sequence of positive r.v.s;
the sequences of *claim sizes* $X_0 = 0, X_1, X_2, \ldots,$ where X_n is the claim occurring at time T_n.

(b) *model output elements:*
the *number of claims* up to time t

$$N(t) = \sup\{n: T_n \leqslant t\} \tag{16.1.1}$$

the *total claim amount* at time t

$$X(t) = \sum_{i=0}^{N(t)} X_i. \tag{16.1.2}$$

In particular let us consider the problem of calculating the distribution of $X(t)$. As in Teugels (1985) it is written:

'The evaluation of the compound distribution $G_t(x)$ of $X(t)$

$$G_t(x) = \sum_{n=1}^{\infty} \Pr\left\{\sum_{i=1}^{n} X_i \leqslant x \mid N(t) = n\right\}\Pr\{N(t) = n\} + \Pr\{N(t) = 0\} \quad x \geqslant 0 \tag{16.1.3}$$

is generally extremely complicated and one is forced to rely on approximations, even in the case when the sequences $\{X_i\}$ and $\{W_i\}$ are independent and consist of i.i.d. r.v.s.'

Using approximations means here that we investigate 'ideal' models which are rather simple, but nevertheless close in some sense to the real (disturbed) model.

For example, as an ideal model we can consider $\tilde{W}_i = \tilde{T}_i - \tilde{T}_{i-1}$; $i = 1, 2, \ldots$, to be independent with a common simple distribution (e.g. an exponential). Moreover, one often supposes that the claim sizes $\tilde{X}_i$ in the ideal model are i.i.d. and independent of $\tilde{W}_i$.

We consider $\tilde{W}_i$, $\tilde{X}_i$ as input elements for our ideal model. Correspondingly, we define

$$\tilde{N}(t) = \sup\{n: \tilde{T}_n \leqslant t\} \tag{16.1.4}$$

$$\tilde{X}(t) = \sum_{i=0}^{\tilde{N}(t)} \tilde{X}_i \tag{16.1.5}$$

as the output elements of our ideal model, related to the output elements $N(t)$, $X(t)$ of the real model.

More concretely, our approximation problem can be stated in the following way: if the input elements of the ideal and real model are 'close' to each other, then can we estimate the deviation between the corresponding outputs? Translating the concept of closeness in a mathematical way one uses some measures of comparisons between the characteristics of the random elements involved.

In this section we confine ourselves to investigating the sketched problems when the sequences $\{X_i\}$ and $\{W_i\}$ have i.i.d. components and are mutually independent. Then we can state our mathematical problem in the following way.

PR I. Let μ, ν, τ be simple probability metrics on $\mathfrak{X}(\mathbb{R})$, i.e. metrics (as before, we shall write $\mu(X, Y)$, $\nu(X, Y)$, $\tau(X, Y)$ instead of $\mu(F_X, F_Y)$, $\nu(F_X, F_Y)$, $\tau(F_X, F_Y)$) in the distribution function space. Find a function ψ: $[0, \infty) \times [0, \infty) \to [0, \infty)$ non-decreasing in both arguments, vanishing and continuous at the origin such that for every ε, $\delta > 0$

$$\left.\begin{array}{l}\mu(W_1, \tilde{W}_1) < \varepsilon \\ \nu(X_1, \tilde{X}_1) < \delta\end{array}\right\} \Rightarrow \tau(X(t), \tilde{X}(t)) \leqslant \psi(\varepsilon, \delta). \tag{16.1.6}$$

The choice of τ is dictated by the user who also wants to be able to check the left-hand side of (16.1.6).

For this reason one has to solve the next stability problem.

PR II. Find a qualitative description of the ε-(resp. δ-) neighborhood of the set of ideal model distributions $F_{\tilde{W}_1}$ (resp. $F_{\tilde{X}_1}$).

16.2 THE PROBLEM OF CONTINUITY

In this section we consider PR I, see (16.1.6). Usually, in practice, the metric τ is chosen to be the Kolmogorov (uniform) metric

$$\boldsymbol{\rho}(X, Y) = \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|. \tag{16.2.1}$$

Moreover, we will choose $\mu = \nu = \boldsymbol{\kappa}_r$, where

$$\boldsymbol{\kappa}_r(X, Y) = r \int_{\mathbb{R}} |x|^{r-1} |F_X(x) - F_Y(x)| \, \mathrm{d}x \quad r > 0 \tag{16.2.2}$$

is the difference pseudomoment, see Section 4.3, Case D. The usefulness of $\boldsymbol{\kappa}_r$ will follow from the considerations in the next section, where PR II is treated. The metric $\boldsymbol{\kappa}_r$ metrizes the weak convergence, plus the convergence of the rth absolute moments in the space of r.v.s X with $\mathbb{E}|X|^r < \infty$, i.e.,

$$\boldsymbol{\kappa}_r(X_n, X) \to 0 \Leftrightarrow \begin{cases} X_n \xrightarrow{w} F_X \\ \mathbb{E}|X_n|^r \to \mathbb{E}|X|^r \end{cases} \quad \text{as } n \to \infty$$

see Theorems 5.4.1 and 6.3.1. Note that

$$\boldsymbol{\kappa}_r(X, Y) = \boldsymbol{\kappa}_1(X|X|^{r-1}, Y|Y|^{r-1}). \tag{16.2.3}$$

First, let us simplify the right-hand side of (16.1.6). Using the triangle inequality we get

$$\begin{aligned} \boldsymbol{\rho}(X(t), \tilde{X}(t)) &= \boldsymbol{\rho}\left(\sum_{i=0}^{N(t)} X_i, \sum_{i=0}^{\tilde{N}(t)} \tilde{X}_i\right) \\ &\leqslant \boldsymbol{\rho}\left(\sum_{i=0}^{N(t)} X_i, \sum_{i=0}^{N(t)} \tilde{X}_i\right) + \boldsymbol{\rho}\left(\sum_{i=0}^{N(t)} \tilde{X}_i, \sum_{i=0}^{\tilde{N}(t)} \tilde{X}_i\right) =: I_1 + I_2. \end{aligned} \tag{16.2.4}$$

Assuming $H(t) = \mathbb{E}N(t)$ to be finite we have

$$I_1 = \boldsymbol{\rho}\left(\sum_{i=0}^{N(t)} \tilde{X}_i, \sum_{i=0}^{N(t)} X_i\right) = \boldsymbol{\rho}\left(\frac{1}{H(t)} \sum_{i=0}^{N(t)} X_i, \frac{1}{H(t)} \sum_{i=0}^{N(t)} \tilde{X}_i\right). \tag{16.2.5}$$

From this expression we are going to estimate I_1 from above, by $\boldsymbol{\kappa}_r(X_1, \tilde{X}_1)$. This will be achieved in two steps:

(i) estimation of the closeness between the r.v.s

$$Z(t) = \frac{1}{H(t)} \sum_{i=0}^{N(t)} X_i \quad \tilde{Z}(t) = \frac{1}{H(t)} \sum_{i=0}^{N(t)} \tilde{X}_i \tag{16.2.6}$$

in terms of an appropriate ('ideal' for this purpose) metric;

(ii) passing from the ideal metric to $\boldsymbol{\rho}$ and $\boldsymbol{\kappa}_r$ respectively, via inequalities of the type

$$\phi_1(\boldsymbol{\rho}) \leqslant \text{ideal metric} \leqslant \phi_2(\boldsymbol{\kappa}_r) \tag{16.2.7}$$

for some non-negative, continuous functions ϕ_i: $[0, \infty) \to [0, \infty)$ with $\phi_i(0) = 0$, $\phi_i(t) > 0$ if $t > 0$, $i = 1, 2$.

Considering the first step, we choose $\zeta_{m,p}$ ($m = 0, 1, \ldots, p \geqslant 1$) as our ideal metric where $\zeta_{m,p}(X, Y)$ is given by (14.2.10). The metric $\zeta_{m,p}$ is ideal of order $r = m + 1/p$, see Definition 14.2.1, i.e. for each X, Y, Z with Z independent of X and Y and every $c \in \mathbb{R}$

$$\zeta_{m,p}(cX + Z, cY + Z) \leqslant |c|^r \zeta_{m,p}(X, Y). \tag{16.2.8}$$

These and other properties for $\zeta_{m,p}$ will be considered in the next section, cf. Lemma 17.1.2.

Lemma 16.2.1 Let $\{X_i\}$ and $\{\tilde{X}_i\}$ be two sequences of i.i.d. r.v.s and let $N(t)$ be independent of the sequences $\{X_i\}$, $\{\tilde{X}_i\}$ and have finite moment $H(t) = \mathbb{E}N(t) < \infty$. Then

$$\zeta_{m,p}(Z(t), \tilde{Z}(t)) \leqslant H(t)^{1-r} \zeta_{m,p}(X_1, \tilde{X}_1) \tag{16.2.9}$$

where $r = m + 1/p$.

Proof. The following chain of inequalities proves the required estimate:

$$\begin{aligned}
&\zeta_{m,p}(Z(t), \tilde{Z}(t)) \\
\text{(a)}\quad &\leqslant H(t)^{-r} \zeta_{m,p}\left(\sum_{i=0}^{N(t)} X_i, \sum_{i=0}^{N(t)} \tilde{X}_i \right) \\
\text{(b)}\quad &\leqslant H(t)^{-r} \sum_{k=1}^{\infty} \Pr(N(t) = k) \zeta_{m,p}\left(\sum_{i=1}^{k} X_i, \sum_{i=1}^{k} \tilde{X}_i \right) \\
\text{(c)}\quad &\leqslant H(t)^{-r} \sum_{k=1}^{\infty} \Pr(N(t) = k) \sum_{i=1}^{k} \zeta_{m,p}(X_i, \tilde{X}_i) \\
&= H(t)^{-r} \sum_{k=1}^{\infty} k \Pr(N(t) = k) \zeta_{m,p}(X_1, \tilde{X}_1) \\
&= H(t)^{1-r} \zeta_{m,p}(X_1, \tilde{X}_1).
\end{aligned}$$

Here (a) follows from (16.2.8) with $Z = 0$ and $c = H(t)^{-1}$. Inequality (b) results from the independence of $N(t)$ with respect to $\{X_i\}$, $\{\tilde{X}_i\}$. Finally (c) can be proved by induction using the triangle inequality and (16.2.8) with $c = 1$.

QED

The obtained estimate (16.2.9) is meaningful if $\zeta_{m,p}(X_1, \tilde{X}_1) \leqslant \infty$. This implies, however, that

$$\int_{\mathbb{R}} x^j \, \mathrm{d}(F_{X_1}(x) - F_{\tilde{X}_1}(x)) = 0 \quad \text{for } j = 0, 1, \ldots, m. \tag{16.2.10}$$

(Indeed, if (16.2.10) fails for some $j = 0, 1, \ldots, m$, then $\zeta_{m,p}(X_1, \tilde{X}_1) \geqslant \sup_{c>0} |\mathbb{E}(cX_1^j - c\tilde{X}_1^j)| = +\infty$.)

Let us now find a lower bound for $\zeta_{m,p}(Z(t), \tilde{Z}(t))$ in terms of $\boldsymbol{\rho}$. (An upper bound for $\zeta_{m,p}(X_1, \tilde{X}_1)$ in terms of $\boldsymbol{\kappa}_r$ ($r = m + 1/p$) is given by Lemma 14.2.6.)

Lemma 16.2.2. If Y has a bounded density p_Y, then

$$\boldsymbol{\rho}(X, Y) \leqslant \left(1 + \sup_{x \in \mathbb{R}} p_Y(x)\right)(c_{m,p}\zeta_{m,p}(X, Y))^{1/(r+1)} \tag{16.2.11}$$

where

$$c_{m,p} = \frac{(2m+2)!(2m+3)^{1/2}}{(m+1)!(3-2/p)^{1/2}}.$$

Proof. To prove (16.2.11) we use similar estimates between the Lévy metric $\mathbf{L} = \mathbf{L}_1$ (see (4.1.3)) and $\zeta_{m,p}$; for any r.v.s X and Y,

$$\mathbf{L}(X, Y)^{r+1} \leqslant c_{m,p}\zeta_{m,p}(X, Y) \tag{16.2.12}$$

see Theorem 3.10.2 (Kalashnikov and Rachev, 1988). Next, since the density of Y exists and is bounded we have

$$\boldsymbol{\rho}(X, Y) \leqslant (1 + \sup_{x \in \mathbb{R}} p_Y(x))\mathbf{L}(X, Y) \tag{16.2.13}$$

which implies (16.2.11). QED

In addition, let us remark that $\zeta_{0,\infty} = \boldsymbol{\rho}$ and $\zeta_{0,1} = \boldsymbol{\kappa}_1$. So, combining Lemmas 14.2.6, 16.2.1, and 16.2.2, we find immediately the following lemma.

Lemma 16.2.3. Let $\{X_i\}$, $\{\tilde{X}_i\}$ be two sequences of i.i.d. r.v.s and let $N(t)$ be independent of $\{X_i\}$, $\{\tilde{X}_i\}$ with $H(t) = \mathbb{E}N(t) < \infty$. Suppose that

$$\boldsymbol{\kappa}_r(X_1, \tilde{X}_1) < \infty$$

and

$$\int x^j \, \mathrm{d}(F_{X_1}(x) - F_{\tilde{X}_1}(x)) = 0 \quad j = 0, 1, \ldots, m \tag{16.2.14}$$

for some $r = m + 1/p \geqslant 1$ ($m = 1, 2, \ldots$; $1 \leqslant p < \infty$). Moreover, let $\tilde{Z}(t)$ (see (16.2.6)) have a bounded density $p_{\tilde{Z}(t)}$. Then

$$I_1 = \boldsymbol{\rho}\left(\sum_{i=0}^{N(t)} X_i, \sum_{i=0}^{N(t)} \tilde{X}_i\right) \leqslant \psi_1(\kappa_r(X_1, \tilde{X}_1))$$
$$:= (1 + \sup p_{\tilde{Z}(t)}(x))(c_{m,p}\phi_2(\kappa_r(X_1, \tilde{X}_1)))^{1/(1+r)} H(t)^{(1-r)/(1+r)} \tag{16.2.15}$$

where

$$\phi_2(\boldsymbol{\kappa}_r) = \begin{cases} \boldsymbol{\kappa}_1^{1/p} & m = 0, 1 \leqslant p < \infty \\ \dfrac{\Gamma(1 + p^{-1})}{\Gamma(r)} \boldsymbol{\kappa}_r & m > 0, 1 \leqslant p < \infty. \end{cases} \tag{16.2.16}$$

Now, going back to (16.2.4), we need also to estimate

$$I_2 = \boldsymbol{\rho}\left(\sum_{i=0}^{N(t)} \tilde{X}_i, \sum_{i=0}^{\tilde{N}(t)} \tilde{X}_i\right)$$

from above by some function, ψ_2, say, of $\boldsymbol{\kappa}_r(W_1, \tilde{W}_1)$.

Lemma 16.2.4. Let $\{W_i\}$, $\{\tilde{W}_i\}$ be two sequences of i.i.d. positive r.v.s, both independent of $\{\tilde{X}_i\}$. Suppose that $H(t) = \mathbb{E}N(t) < \infty$, $\tilde{H}(t) = \mathbb{E}\tilde{N}(t) < \infty$

$$\theta(\tilde{W}_1) = \sup_k \sup_x p_{k^{-1/2}\sum_{i=1}^k \tilde{W}_i}(x) < \infty \quad \kappa_r(W_1, \tilde{W}_1) < \infty \tag{16.2.17}$$

and

$$\int_0^\infty x^j \, \mathrm{d}(F_{W_1}(x) - F_{\tilde{W}_1}(x)) = 0 \quad j = 0, 1, \ldots, m \tag{16.2.18}$$

for some $r = m + 1/p (\geqslant 2)$ ($m = 1, 2, \ldots$; $1 \leqslant p < \infty$). Finally let $F_{\tilde{X}_1}(a) < 1$ $\forall a > 0$, and $\mathbb{E}\tilde{X}_1 < \infty$. Then

$$I_2 = \boldsymbol{\rho}\left(\sum_{i=1}^{N(t)} \tilde{X}_i, \sum_{i=1}^{\tilde{N}(t)} \tilde{X}_i\right) \leqslant \psi_2(\boldsymbol{\kappa}_r(W_1, \tilde{W}_1))$$
$$:= (1 + \theta(\tilde{W}_1))\boldsymbol{\kappa}_r(W_1, \tilde{W}_1)^{1/(r+1)}$$
$$+ \inf_{a>0} \{2(c_{m,p}(1 + \theta(\tilde{W}_1))\phi_2(\kappa_r(W_1, \tilde{W}_1)))^{1/(1+r)} \chi_{\tilde{X}_1, r}(a)$$
$$+ a^{-1}\mathbb{E}\tilde{X}_1 \max(H(t), \tilde{H}(t))\} \tag{16.2.19}$$

where ϕ_2 is given by (16.2.16) and

$$\chi_{\tilde{X}_1,r}(a) := \sum_{k=1}^{\infty} k^{(1-r/2)/(1+r)} F^k_{\tilde{X}_1}(a).$$

Remark 16.2.1. The normalization $k^{-1/2}$ of the sum $\sum_{i=1}^{k} \tilde{W}_i$ in (16.2.17) comes from the quite natural assumption that the $\tilde{W}_i$s, the claim's inter-arrival times for the ideal model, are in the domain of attraction of the normal law. Actually, this is the case which we are going to consider in the next section. However, for example, if we need to approximate W_is with $\tilde{W}_i$s, where $\tilde{W}_i$s are in the normal domain of attraction of symmetric α-stable distribution with $0 < 2$, we should use the normalization $k^{-1/\alpha}$ in (16.2.17).

Remark 16.2.2. Note that if $\boldsymbol{\kappa}_r(W_1, \tilde{W}_1)$ tends to zero, then the right-hand side of (16.2.19) also tends to zero since, for each $a > 0$, $\chi_{\tilde{X}_1,r}(a) < \infty$.

Proof of Lemma 16.2.4. By the independence of $N(t)$ and $\tilde{N}(t)$ with respect to $\{\tilde{X}_i\}$, we find that, for every $a > 0$

$$\begin{aligned} I_2 = {} & \sup_{0 \leqslant x \leqslant a} \left| \sum_{k=1}^{\infty} [\Pr(N(t) = k) - \Pr(\tilde{N}(t) = k)] \Pr\left(\sum_{i=1}^{k} \tilde{X}_i \leqslant x \right) \right| \\ & + \sup_{x > a} \left| \Pr\left(\sum_{i=1}^{N(t)} \tilde{X}_i > x \right) - \Pr\left(\sum_{i=1}^{\tilde{N}(t)} \tilde{X}_i > x \right) \right| \\ & + |\Pr(N(t) = 0) - \Pr(\tilde{N}(t) = 0)| =: J_{1,a} + J_{2,a} + J_3. \end{aligned}$$

Estimating $J_{1,a}$ we get

(a)

$$\begin{aligned} J_{1,a} \leqslant {} & \sum_{k=1}^{\infty} \left(\left| \Pr\left(\sum_{i=1}^{k} W_i \leqslant T \right) - \Pr\left(\sum_{i=1}^{k} \tilde{W}_i \leqslant t \right) \right| \right. \\ & \left. + \left| \Pr\left(\sum_{i=1}^{k+1} W_i \leqslant t \right) - \Pr\left(\sum_{i=1}^{k+1} \tilde{W}_i \leqslant t \right) \right| \right) \Pr\left(\max_{i=1,\dots,k} \tilde{X}_i \leqslant x \right) \\ \leqslant {} & \sum_{k=1}^{\infty} \left\{ \boldsymbol{\rho}\left(\sum_{i=1}^{k} W_i, \sum_{i=1}^{k} \tilde{W}_i \right) + \boldsymbol{\rho}\left(\sum_{i=1}^{k+1} W_i, \sum_{i=1}^{k+1} \tilde{W}_i \right) \right\} F^k_{\tilde{X}_1}(a) \end{aligned}$$

(b)

$$\begin{aligned} & \leqslant \sum_{k=1}^{\infty} (c_{m,p} \phi_2(\boldsymbol{\kappa}_r(W_1, \tilde{W}_1)))^{1/(1+r)} \\ & \qquad \times \{k^{(1-r/2)/(1+r)} + (k+1)^{(1-r/2)/(1+r)}\}(1 + \theta(\tilde{W}_1)) F^k_{\tilde{X}_1}(a) \\ & \leqslant 2(1 + \theta(\tilde{W}_1))(c_{m,p} \phi_2(\kappa_r(W_1, \tilde{W}_1)))^{1/(1+r)} \chi_{\tilde{X}_1,r}(a). \end{aligned}$$

Inequality (a) follows from

$$\Pr(N(t) = k) = \Pr\left(\sum_{i=1}^{k} W_i \leqslant t\right) - \Pr\left(\sum_{i=1}^{k+1} W_i \leqslant t\right).$$

We derived (b) from Lemmas 16.2.2 and 14.2.6, see also (16.2.16) and (16.2.17). Furthermore, one finds with Chebychev's inequality that

$$J_{2,a} \leqslant \max\left(\Pr\left(\sum_{i=1}^{N(t)} \tilde{X}_i > a\right), \Pr\left(\sum_{i=1}^{\tilde{N}(t)} \tilde{X}_i > a\right)\right)$$
$$\leqslant a^{-1}(\mathbb{E}X_1)\max(H(t), \tilde{H}(t)).$$

The inequality (14.2.22) can be extended in the case $m = 1$, $p = \infty$ (so $\zeta_{0,\infty} = \boldsymbol{\rho}$) to

$$\boldsymbol{\rho}(W_1, \tilde{W}_1) \leqslant \left(1 + \sup_x p_{\tilde{W}_1}(x)\right)\boldsymbol{\kappa}_r(W_1, \tilde{W}_1)^{1/(r+1)}. \tag{16.2.20}$$

By virtue of (16.2.13) we see that to prove (16.2.20) it is enough to show the following estimate.

Claim. For any non-negative r.v.s X and Y

$$\mathbf{L}(X, Y) \leqslant \boldsymbol{\kappa}_r(X, Y)^{1/(1+r)}. \tag{16.2.21}$$

Indeed, if the Lévy metric $\mathbf{L}(X, Y)$ is greater than $\varepsilon \in (0, 1)$, then there is $x_0 \geqslant 0$ such that $|F_X(x) - F_Y(x)| \geqslant \varepsilon\ \forall x \in [x_0, x_0 + \varepsilon]$. Thus

$$\kappa_r(X, Y) \geqslant r \int_{x_0}^{x_0+\varepsilon} x^{r-1}|F_X(x) - F_Y(x)|\,dx \geqslant \varepsilon^{r+1}.$$

Letting $\varepsilon \to \mathbf{L}(X, Y)$ proves the claim. Finally, since $J_3 \leqslant \boldsymbol{\rho}(W_1, \tilde{W}_1)$, the lemma follows. QED

We can conclude by the following theorem which follows immediately by combining (16.2.4) and Lemmas 16.2.3 and 16.2.4.

Theorem 16.2.1 Under the conditions of Lemma 16.2.3 and 16.2.4,

$$\boldsymbol{\rho}(X(t), \tilde{X}(t)) \leqslant \psi(\boldsymbol{\kappa}_r(W_1, \tilde{W}_1), \boldsymbol{\kappa}_r(X_1, \tilde{X}_1))$$
$$=: \psi_1(\boldsymbol{\kappa}_r(X_1, \tilde{X}_1)) + \psi_2(\boldsymbol{\kappa}_r(W_1, \tilde{W}_1))$$

where ψ_1(resp. ψ_2) is given by (16.2.15) (resp. (16.2.19)).

The preceding theorem gives us a solution to *PR I* (see (16.1.6)) with $\mu = \nu = \boldsymbol{\kappa}_r$, and $\tau = \rho$ under some moment conditions (see Lemmas 16.2.3 and 16.2.4).

16.3 STABILITY OF THE INPUT CHARACTERISTICS

In order to solve *PR II* (see Section 16.1.1) we will investigate the conditions on the real input characteristics that imply $\mu(W_1, \tilde{W}_1) < \varepsilon$ and $\nu(X_1, \tilde{X}_1) < \delta$ for $\mu = \nu = \boldsymbol{\kappa}_r$ (see (16.1.6)). We consider only $r = 2$ and qualitative conditions on the distribution of W_1 implying $\boldsymbol{\kappa}_2(W_1, \tilde{W}_1) < \varepsilon$. One can follow the same idea to check $\boldsymbol{\kappa}_r(W_1, \tilde{W}_1) < \varepsilon$, $r \neq 2$, and $\boldsymbol{\kappa}_r(X_1, \tilde{X}_1) < \delta$. We will characterize the input 'ideal' distribution F_{W_i} supposing that the 'real' F_{W_1} belongs to one of the so-called 'aging' classes of distribution

$$\text{IFR} \subset \text{IFRA} \subset \text{NBU} \subset \text{NBUE} \subset \text{HNBUE} \tag{16.3.1}$$

(cf. Section 14.1). These kinds of quantitative conditions on F_{W_1} (see Barlow and Proschan, 1975; Kalashnikov and Rachev, 1988) are quite natural in an insurance risk setting. For example, $F_{W_1} \in \text{IFR}$ if and only if the residual lifelength distribution $\Pr(W_1 \leqslant x + t \mid W_1 > t)$ is non-decreasing in t for all positive x.

The above assumption leads in a natural way to the choice of an exponential ideal distribution in view of the characterizations of the exponential law given in the next Lemma 16.3.1. Moreover, we emphasize here the use of the NBUE and HNBUE classes as we want to impose the weakest possible conditions on the 'real' (unknown) F_{W_1}. Let us recall the definitions of these classes.

Definition 16.3.1. Let W be a positive r.v. with $\mathbb{E}W < \infty$, and denote $\bar{F} = 1 - F$. Then $F_W \in \text{NBUE}$ if

$$\int_t^\infty \bar{F}_W(u)\,du \leqslant (\mathbb{E}W)\bar{F}_W(t) \quad \forall t > 0 \tag{16.3.2}$$

and $F_W \in \text{HNBUE}$ if

$$\int_t^\infty \bar{F}_W(u)\,du \leqslant (\mathbb{E}W)\exp(-t/\mathbb{E}W) \quad \forall t > 0. \tag{16.3.3}$$

Lemma 16.3.1. (i) If $F_W \in \text{NBUE}$ and $m_i = \mathbb{E}W^i < \infty$, $i = 1, 2, 3$, then

$$\bar{F}_W(t) = \exp(-t/m_1) \quad \text{iff } \alpha := m_1^2 + \frac{m_2}{2} - \frac{m_3}{3m_1} = 0. \tag{16.3.4}$$

(ii) If $F_W \in \text{HNBUE}$ and $m_i = \mathbb{E}W^i < \infty$, $i = 1, 2$, then

$$\bar{F}_W(t) = \exp(-t/m_1) \quad \text{iff } \beta := 2 - \frac{m_2}{m_1^2} = 0. \tag{16.3.5}$$

The 'only if' parts of Lemma 16.3.1 are obvious. The 'iff' parts result from the following estimates of the stability of exponential law characterizations

(i) and (ii) in Lemma 16.3.1. Further, denote $E(\lambda)$, an exponentially distributed r.v. with parameter $\lambda > 0$.

Lemma 16.3.2. (i) If $F_W \in \text{NBUE}$ and $m_i = \mathbb{E}W^i < \infty$, $i = 1, 2, 3$, then

$$\boldsymbol{\kappa}_2(W, E(\lambda)) \leqslant 2\alpha + 2|\lambda^{-2} - m_1^2|. \tag{16.3.6}$$

(ii) If $F_W \in \text{HNBUE}$ and $m_i = \mathbb{E}W^i < \infty$, $i = 1, 2$, then

$$\boldsymbol{\kappa}_2(W, E(\lambda)) \leqslant C(m_1, m_2)\beta^{1/8} + 2|\lambda^{-2} - m_1^2| \tag{16.3.7}$$

where

$$C(m_1, m_2) = 8\sqrt{6}\, m_1(\sqrt{m_2} + m_1\sqrt{2}). \tag{16.3.8}$$

Proof. (i) The proof of the first part relies on the following claim concerning the stability of the exponential law characterizations in the class NBU. Let us recall that if F_W has a density then $F_W \in \text{NBU}$ if the 'hazard rate function' $h_W(t) = F'_W(t)/\bar{F}_W(t)$ satisfies

$$h_W(t) \geqslant h = h_W(0) \quad \forall t \geqslant 0. \tag{16.3.9}$$

Claim. Let $F_W \in \text{NBU}$ and $\mu_i = \mu_i(W) = \mathbb{E}W^i < \infty$, $i = 1, 2$. Then

$$\int_0^\infty t|F'_W(t) - h\exp(ht)|\,dt \leqslant \mu_1 - h\mu_2 + h^{-1}. \tag{16.3.10}$$

Proof of the claim. By (16.3.9) it follows that $H(t) = h\bar{F}_W(t) - F'_W(t)$ is a non-positive function on $[0, \infty)$. Clearly,

$$\bar{F}_W(t) = \exp(-ht)\left(1 + \int_0^t H(u)\exp(hu)\,du\right).$$

Hence

$$\begin{aligned}\int_0^\infty t|F'_W(t) - h\exp(-ht)|\,dt &= \int_0^\infty t\left|h\exp(-ht)\int_0^t H(u)\exp(hu)\,du - H(t)\right|dt\\ &\leqslant \int_0^\infty ht\exp(-ht)\int_0^t |H(u)|\exp(hu)\,du\,dt + \int_0^\infty t|H(t)|\,dt\\ &= -\int_0^\infty\left(\int_0^\infty ht\exp(ht)\,dt\right)H(u)\exp(hu)\,du - \int_0^\infty tH(t)\,dt.\end{aligned}$$

Integrating by parts in the first integral and replacing $H(t)$ by $h\bar{F}_W(t) - F'_W(t)$ we obtain the required inequality (16.3.10).

Now, continuing the proof of Lemma 16.3.2 (i), note that $F_W \in \text{NBUE}$

implies $F_{W^*} \in \text{NBU}$ where $F_{W^*}(t) = m_1^{-1} \int_0^t \bar{F}_W(u)\,du$, $t \geqslant 0$. Also

$$\kappa_2(W, E(m_1^{-1})) = 2m_1 \int_0^\infty t\,|F'_{W^*}(t) - h_{W^*}(0)\exp(-t/h_{W^*}(0))|\,dt \quad (16.3.11)$$

where

$$h_{W^*}(0) = m_1^{-1} \quad \mathbb{E}W^* = m_2/2m_1 \quad (16.3.12)$$

and

$$\mathbb{E}(W^*)^2 = m_3/3m_1. \quad (16.3.13)$$

Using the claim (16.3.11)–(16.3.13), we get

$$\tfrac{1}{2}\kappa_2(W, E(m_1^{-1})) \leqslant \frac{m_2}{2} - \frac{m_3}{3m_1} + m_1^2. \quad (16.3.14)$$

On the other hand, for each $\lambda > 0$ one easily shows that

$$\kappa_2(E(\lambda), E(m_1^{-1})) = 2|m_1^2 - \lambda^{-2}|. \quad (16.3.15)$$

From (16.3.14), (16.3.15), using the triangle inequality, (16.3.6) follows.

(ii) To derive (16.3.7) we use the representation of κ_2 as a minimal metric: for any two non-negative r.v.s X and Y with finite second moment

$$\kappa_2(X, Y) = \inf\{\mathbb{E}|\tilde{X}^2 - \tilde{Y}^2| : \tilde{X} \overset{d}{=} X, \tilde{Y} \overset{d}{=} Y\}. \quad (16.3.16)$$

(Use Theorem 8.1.2 with $c(x, y) = |x - y|$ and the representation (16.2.3), see also Remark 7.1.3.) Similarly, by Theorem 8.1.2 with $c(x, y) = |x - y|^2$

$$\ell_2(X, Y) = \left(\int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)|^2\,dt\right)^{1/2}$$

$$= \inf\{(\mathbb{E}(\tilde{X} - \tilde{Y})^2)^{1/2} : \tilde{X} \overset{d}{=} X, \tilde{Y} \overset{d}{=} Y\}. \quad (16.3.17)$$

By Hölder's inequality we obtain that

$$\mathbb{E}|\tilde{X}^2 - \tilde{Y}^2| \leqslant (\mathbb{E}(\tilde{X} - \tilde{Y})^2)^{1/2}((\mathbb{E}\tilde{X}^2)^{1/2} + (\mathbb{E}\tilde{Y}^2)^{1/2}).$$

Hence, by (16.3.16) and (16.3.17)

$$\kappa_r(X, Y) \leqslant \ell_2(X, Y)((\mathbb{E}X^2)^{1/2} + (\mathbb{E}Y^2)^{1/2}). \quad (16.3.18)$$

In Kalashnikov and Rachev (1988) (Lemma 4.2.1) it is shown that for $W \in \text{NBUE}$

$$\ell_2(W, E(m_1^{-1})) \leqslant 8\sqrt{6}\,m_1\beta^{1/8}. \quad (16.3.19)$$

By (16.3.18) and (16.3.19) we now get that

$$\kappa_2(W, E(m_1^{-1})) \leqslant C(m_1, m_2)\beta^{1/8}. \quad (16.3.20)$$

The result in (ii) is a consequence of (16.3.15) and (16.3.20). QED

Remark 16.3.1. Note that the term $|\lambda^{-2} - m_1^2|$ in (16.3.6) and (16.3.7) is zero if we choose the parameter λ in our 'ideal' exponential distribution F_W to be m_1^{-1}, and hence the 'if' parts of Lemma 16.3.1 follow.

Reformulating Lemma 16.3.2 towards our original problem *PR II* we can state the following theorem.

Theorem 16.3.1. Let $\tilde{W} \stackrel{\mathrm{d}}{=} E(\lambda)$. Then

$$\boldsymbol{\kappa}_2(W, \tilde{W}) \leqslant \varepsilon \tag{6.3.21}$$

where $\varepsilon = 2\alpha + 2|\lambda^{-2} - m_1^2|$ if $F_W \in \text{NBUE}$, and

$$\varepsilon = C(m_1, m_2)\beta^{1/8} + 2|\lambda^{-2} - m_1^2|$$

if $F_W \in \text{HNBUE}$.

Remark 16.3.2. In the case where F_W belongs to IFR, IFRA, or NBU the preceding estimate (16.3.21) can be improved by using more refined estimates than (16.3.19) (see Kalashnikov and Rachev (1988), Lemma 4.2.1).

The above results concerning *PR I* and *PR II* lead to the following recommendations.

(i) One checks if F_{W_1} belongs to some of the classes in (16.3.1). (Basu and Ebrahimi, 1985) proposed some statistical procedures for checking that $F_{W_1} \in \text{HNBUE}$. See the references in the above paper for testing whether F_{W_1} belongs to the ageing classes.)

(ii) If, for example, $F_{W_1} \in \text{HNBUE}$, one computes $m_1 = \mathbb{E}W_1$, $m_2 = \mathbb{E}W_2$ and $\beta = 2 - m_2/m_1^2$. If β is close to zero, we can choose the 'ideal' distribution $F_{\tilde{W}}(x) = 1 - \exp(x/m_1)$. Then the possible deviation between F_{W_1} and $F_{\tilde{W}_1}$ in $\boldsymbol{\kappa}_2$-metric is given by Theorem 16.3.1

$$\boldsymbol{\kappa}_2(W_1, \tilde{W}_1) \leqslant C(m_1, m_2)\beta^{1/8} = \varepsilon. \tag{16.3.22}$$

(iii) In a similar way choose $F_{\tilde{X}_1}$ and estimate the deviation

$$\boldsymbol{\kappa}_2(X_1, \tilde{X}_1) \leqslant \delta. \tag{16.3.23}$$

(iv) Compute the approximating compound Poisson distribution $F_{\sum_{i=1}^{\tilde{N}(t)} \tilde{X}_i}$ (see Teugels, 1985). Then the possible deviation between the 'real' compound distribution $F_{\sum_{i=1}^{N(t)} X_i}$ the 'ideal' $F_{\sum_{i=1}^{\tilde{N}(t)} X_i}$ in terms of the uniform metric is

$$\boldsymbol{\rho}\left(\sum_{i=1}^{N(t)} X_i, \sum_{i=1}^{\tilde{N}(t)} \tilde{X}_i\right) \leqslant \psi(\varepsilon, \delta) \tag{16.3.24}$$

(see Theorem 16.2.1).

If F_W does not belong to any of the classes in (16.3.1), then one can compute the empirical distribution function $\hat{F}^{(N)}_{W_1}(\cdot, \omega)$ based on N observations $W_1, W_2, \ldots, W_N$. Choosing $\lambda > 0$ (or $F_{\tilde{W}_1}(x) = 1 - \exp(-\lambda x)$) such that $\mathbb{E}\boldsymbol{\kappa}_2(\hat{F}^{(N)}_{W_1}, F_{\tilde{W}_1}) < \varepsilon$, we get that

$$\boldsymbol{\kappa}_2(F_{W_1}, F_{\tilde{W}_1}) < \varepsilon + \mathbb{E}\boldsymbol{\kappa}_2(\hat{F}^{(N)}_{W_1}, F_{W_1}). \tag{16.3.25}$$

Dudley's theorem (see Theorems 4.9.7 and 4.9.8 of Kalashnikov and Rachev (1988)) implies that the second term in the right-hand side of (16.3.25) can be estimated by some function $\phi(N)$, tending to zero as $N \to \infty$.

CHAPTER 17

How Close are the Individual and Collective Models in the Risk Theory?

17.1 STOP-LOSS DISTANCES AS MEASURES OF CLOSENESS BETWEEN INDIVIDUAL AND COLLECTIVE MODELS

In Chapters 15 and 16 we defined and used an ideal metric of order $r = m + 1/p > 0$,

$$\zeta_{m,p}(X, Y) = \sup\{|\mathbb{E}f(X) - \mathbb{E}f(Y)|: \|f^{(m+1)}\|_q \leqslant 1\} \tag{17.1.1}$$

$m = 0, 1, 2, \ldots, p \in [1, \infty]$, $1/p + 1/q = 1$. The 'dual' representation of

$$\zeta_{1,\infty}(X, Y)$$

gives for any X and Y with equal means

$$\zeta_{1,\infty}(X, Y) = \sup_{x \in \mathbb{R}} \left| \int_x^\infty (x - t)\, \mathrm{d}(F_X(t) - F_Y(t)) \right| \tag{17.1.2}$$

where F_X stands for the d.f. of X.

The latter distance is well known in risk theory as *stop-loss metric* (cf. Gerber 1981, p. 97) and d is used to measure the distance between the so-called *individual and collective models*. More precisely, let $X_1, \ldots, X_n$ be independent real valued variables with d.f.s F_i, $1 \leqslant i \leqslant n$, of the form

$$F_i = (1 - p_i)E_0 + p_i V_i \quad 0 \leqslant p_i \leqslant 1. \tag{17.1.3}$$

Here E_0 is the one point mass d.f. concentrated at zero and V_i is any d.f. on $\mathbb{R}$. We can, therefore, write $X_i = C_i D_i$, where C_i has d.f. V_i, D_i is Bernoulli distributed with success probability p_i and C_i, D_i are independent. Then

$$S^{\mathrm{ind}} := \sum_{i=1}^n X_i = \sum_{i=1}^n C_i D_i \tag{17.1.4}$$

has the d.f. $F = F_1 * \cdots * F_n$ where $*$ denotes the convolution of d.f.s.

The notation S^{ind} comes from risk theory (see Gerber 1981, Chapter 4), where S^{ind} is the so-called *aggregate claim in the individual model*. Each of n policies leads with (small) probability p_i to a 'claim amount' C_i with d.f. V_i.

Consider approximations of S^{ind} by compound Poisson distributed r.v.s

$$S^{\text{coll}} := \sum_{i=1}^{N} Z_i \tag{17.1.5}$$

where $\{Z_i\}$ are i.i.d., $Z_i \stackrel{d}{=} V$ (i.e. Z_i has d.f. V), N is Poisson distributed with parameter μ and $\{Z_i\}$, N are independent. The empty sum in (17.1.5) is defined to be zero. S^{coll} is called *collective model* in the risk theory setting. The usual choice of V and μ in the collective model is

$$\mu = \tilde{\mu} := \sum_{i=1}^{n} p_i \quad V = \tilde{V} := \sum_{i=1}^{n} \frac{p_i}{\mu} V_i = \sum_{i=1}^{n} \frac{p_i}{\mu} F_{C_i} \tag{17.1.6}$$

see Gerber (1981), Section 1, Chapter 4. This choice leads to the following representation of S^{coll}

$$S^{\text{coll}} = \sum_{i=1}^{n} S_i^{\text{coll}}. \tag{17.1.7}$$

Here, $S_i^{\text{coll}} = \sum_{j=1}^{N_i} Z_{ij}$, $N_i \stackrel{d}{=} \mathscr{P}(p_i)$ (i.e. Poisson distribution with parameter p_i), $Z_{ij} \stackrel{d}{=} V_i$, and N_i, Z_{ij} are independent, i.e. one approximates each summand X_i by a compound Poisson distributed r.v. S_i^{coll}.

Our further objective is to replace the usual choice (17.1.6) in the compound Poisson model by a *scaled model*, i.e. we choose $Z_{ij} \stackrel{d}{=} u_i C_i$, $\mu = \sum_{i=1}^{n} \mu_i$ with scale factors u_i and with μ_i such that *the first two moments of* S^{ind} *and* S^{coll} *coincide.*

Remark 17.1.1. In the usual collective model (17.1.6)

$$\mathbb{E}S^{\text{ind}} = \sum_{i=1}^{n} p_i \mathbb{E}C_i = \mathbb{E}S^{\text{coll}} \tag{17.1.8}$$

and if $q_i = 1 - p_i$,

$$\begin{aligned} \operatorname{Var}(S^{\text{ind}}) &= \sum_{i=1}^{n} p_i \operatorname{Var}(C_i) + \sum_{i=1}^{n} p_i q_i (\mathbb{E}C_i)^2 \\ &< \operatorname{Var}(S^{\text{coll}}) = \sum_{i=1}^{n} p_i \operatorname{Var}(C_i)) + \sum_{i=1}^{n} p_i (\mathbb{E}C_i)^2 \end{aligned} \tag{17.1.9}$$

(cf. Gerber 1981, p. 50).

To compare the scaled and individual model we shall use several distances well known in risk theory. Among them is the *stop-loss metric of order m*

$$\mathbf{d}_m(X, Y) := \sup_t \left| \int_t^\infty \frac{(x-t)^m}{m!} \, \mathrm{d}(F_X(x) - F_Y(x)) \right|$$
$$= \sup_t (1/m!) | \mathbb{E}(X-t)_+^m - \mathbb{E}(Y-t)_+^m |, \; m \in \mathbb{N} := \{1, 2, \ldots\}, \; (\cdot)_+$$
$$:= \max(\cdot, 0). \tag{17.1.10}$$

This choice is motivated by risk theory and allows us to estimate the difference of two stop-loss premiums (cf. Gerber 1981, p. 97 for $s = 1$).

We shall also consider the L_p-version of $\mathbf{d}_s$, namely

$$\mathbf{d}_{m,p}(X, Y) := \left(\int |D_m(t)|^p \, \mathrm{d}t \right)^{1/p} \quad 1 \leqslant p < \infty$$
$$\mathbf{d}_{m,\infty}(X, Y) := \mathbf{d}_m(X, Y) \tag{17.1.11}$$

where

$$D_m(t) := D_{m,X,Y}(t) := (1/m!)(\mathbb{E}(X-t)_+^m - \mathbb{E}(Y-t)_+^m). \tag{17.1.12}$$

The rest of this section is devoted to the study of the stop-loss metrics $\mathbf{d}_m$ and $\mathbf{d}_{m,p}$.

Lemma 17.1.1. If $\mathbb{E}(X^j - Y^j) = 0$, $1 \leqslant j \leqslant m$, then

$$\mathbf{d}_m(X, Y) = \zeta_{m,\infty}(X, Y), \; |\mathbb{E}(X-Y)| \leqslant \mathbf{d}_1(X, Y) \leqslant \int |F_X(x) - F_Y(x)| \, \mathrm{d}x \tag{17.1.13}$$

and

$$\mathbf{d}_{m,p}(X, Y) = \zeta_{m,p}(X, Y). \tag{17.1.14}$$

Proof. We shall prove (17.1.13) only. The proof of (17.1.14) is similar.

Here and in the following we use the notation

$$H_0(t) := H(t) := F_X(t) - F_Y(t) \tag{17.1.15}$$

and

$$H_1(t) := \int_t^\infty H(u) \, \mathrm{d}u \quad H_k(t) := \int_t^\infty H_{k-1}(u) \, \mathrm{d}u \text{ for } k \geqslant 2. \tag{17.1.16}$$

Claim 1. (a) If $xH(x) \to 0$ for $x \to \infty$, then for $k = 1, \ldots, m$,

$$
\begin{aligned}
D_m(t) &= -\frac{1}{(m-1)!}\int_t^\infty (x-t)^{m-1}H(x)\,\mathrm{d}x \\
&= -\frac{1}{(m-k)!}\int_t^\infty (x-t)^{m-k}H_{k-1}(x)\,\mathrm{d}x \\
&= -H_m(t).
\end{aligned}
\tag{17.1.17}
$$

(b) $|\mathbb{E}X - \mathbb{E}Y| \leqslant d_1(X, Y) \leqslant \int_{-\infty}^{\infty} |H(x)|\,\mathrm{d}x$.

The proof of (a) follows from repeated partial integration. (b) follows from (a).

Claim 2. If f is $(m+1)$-times differentiable, $\mathbb{E}(X^j - Y^j)$ exists, $1 \leqslant j \leqslant m$, and $f(X)$, $f(Y)$ are integrable, then

$$
\begin{aligned}
\mathbb{E}(f(X) - f(Y)) = {} & \sum_{j=0}^{m} \frac{f^{(j)}(0)}{j!}\,\mathbb{E}(X^j - Y^j) \\
& + (-1)^{m+1}\int_{-\infty}^{0} \bar{D}_m(t) f^{(m+1)}(t)\,\mathrm{d}t + \int_0^\infty D_m(t) f^{(m+1)}(t)\,\mathrm{d}t
\end{aligned}
\tag{17.1.18}
$$

and

$$
\mathbb{E}(f(X) - f(Y)) = \int_{\mathbb{R}} D_m(t) f^{(m+1)}(t)\,\mathrm{d}t = (-1)^{m+1}\int_{\mathbb{R}} \bar{D}_m(t) f^{(m+1)}(t)\,\mathrm{d}t
\tag{17.1.19}
$$

where

$$
\bar{D}_m(t) := \bar{D}_{m,X,Y}(t) := (1/m!)(\mathbb{E}(t-X)_+^m - \mathbb{E}(t-Y)_+^m) \quad s \geqslant 1. \tag{17.1.20}
$$

The proof of (17.1.18) follows from the Taylor series expansion,

$$
\begin{aligned}
\mathbb{E}(f(X) - f(Y)) &= \int_{\mathbb{R}} f(x)\,\mathrm{d}H(x) \\
&= \int_{\mathbb{R}} \left[f(0) + \cdots + \frac{x^m}{m!} f^{(m)}(0) + \int_0^x \frac{(x-t)^m}{m!} f^{(m+1)}(t)\,\mathrm{d}t \right] \mathrm{d}H(x) \\
&= \sum_{j=0}^{m} \frac{f^{(j)}(0)}{j!}\,\mathbb{E}(X^j - Y^j) + \int_{-\infty}^{0} (-1)^{m+1} \bar{D}_m(t) f^{(m+1)}(t)\,\mathrm{d}t \\
&\quad + \int_0^\infty D_m(t) f^{(m+1)}(t)\,\mathrm{d}t.
\end{aligned}
$$

To prove (17.1.19) observe that if $\mathbb{E}(X^j - Y^j)$ is finite, $1 \leqslant j \leqslant m$, then

$$D_m(t) = (1/m!) \sum_{j=0}^{m} \binom{m}{j} \mathbb{E}(X^j - Y^j)(-t)^{m-j} + (-1)^{m+1} \bar{D}_m(t). \quad (17.1.21)$$

Now (17.1.19) follows from (17.1.18) and (17.1.21) and thus the proof of Claim 2 is completed.

It is known that for a function h on $\mathbb{R}$ with

$$\|h\|_\infty = \operatorname*{ess\,sup}_{x \in \mathbb{R}} |h(x)|$$

the following dual representation holds:

$$\|h\|_\infty = \sup\left\{\int h(t)g(t)\,\mathrm{d}t \colon \|g\|_1 \leqslant 1\right\} \quad (17.1.22)$$

see, for example, Dunford and Schwartz (1988), Section IV.8 and Neveu (1965). Recall that

$$\zeta_{m,\infty}(X, Y) := \sup\{|\mathbb{E}(f(X) - f(Y))|; f \in \mathscr{F}_m\} \quad (17.1.23)$$

where $\mathscr{F}_m := \{f\colon \mathbb{R}^1 \to \mathbb{R}^1; f^{(m+1)}$ exists and $\|f^{(m+1)}\|_1 \leqslant 1\}$.

Thus (17.1.19), (17.1.22) and (17.1.23) imply

$$\zeta_{m,\infty}(X, Y) = \sup_{f \in \mathscr{F}_m} \left| \int D_m(t) f^{(m+1)}(t)\,\mathrm{d}t \right|$$
$$= \|D_m\|_\infty = \|\bar{D}_m\|_\infty = \mathbf{d}_m(X, Y). \qquad \text{QED}$$

The next lemma shows that the moment condition in Lemma 17.1.1. is necessary for the finiteness of $\zeta_{m,\infty}$ (cf. condition (16.2.10) for $\zeta_{m,p}$ in the previous section).

Lemma 17.1.2. (a) $\zeta_{m,\infty}(X, Y) < \infty$ implies that

$$\mathbb{E}(X^j - Y^j) = 0 \quad 1 \leqslant j \leqslant m. \quad (17.1.24)$$

(b) $\zeta_{m,\infty}$ is an ideal metric of order m, i.e., $\zeta_{m,\infty}$ is a simple probability metric such that

$$\zeta_{m,\infty}(X + Z, Y + Z) \leqslant \zeta_{m,\infty}(X, Y)$$

for Z independent of X, Y and

$$\zeta_{m,\infty}(cX, cY) = |c|^m \zeta_{m,\infty}(X, Y) \quad \text{for } c \in \mathbb{R} \quad (17.1.25)$$

(cf. Definition 14.2.1).

(c) For independent $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ and for $c_i \in \mathbb{R}$ the following inequality holds.

$$\zeta_{m,\infty}\left(\sum_{i=1}^{n} c_i X_i, \sum_{i=1}^{n} c_i Y_i\right) \leqslant \sum_{i=1}^{n} |c_i|^m \zeta_{m,\infty}(X_i, Y_i). \quad (17.1.26)$$

Proof. (a) For any $a > 0$ and $1 \leqslant j \leqslant m$, $f_a(x) := ax^j \in \mathscr{F}_m$ and, therefore,

$$\zeta_{m,\infty}(X, Y) \geqslant \sup_{a>0} a|\mathbb{E}(X^j - Y^j)|$$

i.e., $\mathbb{E}(X^j - Y^j) = 0$.

(b) Since for $z \in \mathbb{R}$ and $f \in \mathscr{F}_m$, $f_z(x) := f(x + z) \in \mathscr{F}_m$, the first part follows from conditioning on $Z = z$. For the second part note that for $c \in \mathbb{R}^1$: $f \in \mathscr{F}_m$ if and only if $|c|^{-m} f_c \in \mathscr{F}_m$ with $f_c(x) = f(cx)$.

Finally, (c) follows from (b) and the triangle inequality for $\zeta_{m,\infty}$. QED

The proof of the next lemma is similar.

Lemma 17.1.3. (a) $\mathbf{d}_m$ is an ideal metric of order m.

(b) For $X_1, \ldots, X_n$ independent, $Y_1, \ldots, Y_n$ independent and $c_i > 0$

$$\mathbf{d}_m\left(\sum_{i=1}^n c_i X_i, \sum_{i=1}^n c_i Y_i\right) \leqslant \sum_{i=1}^n c_i^m \mathbf{d}_m(X_i, Y_i). \tag{17.1.27}$$

(c) $\mathbf{d}_m(X + a, Y + a) = \mathbf{d}_m(X, Y)$, for all $a \in \mathbb{R}$.

(d) If $\mathbb{E}X = \mathbb{E}Y = \mu$, $\sigma^2 = \mathrm{Var}(X) = \mathrm{Var}(Y)$, then with $\tilde{X} = (X - \mu)/\sigma$, $\tilde{Y} = (Y - \mu)/\sigma$

$$\mathbf{d}_m(\tilde{X}, \tilde{Y}) = \sigma^{-m} \mathbf{d}_m(X, Y). \tag{17.1.28}$$

Recall the definition of the difference pseudomoment of order m

$$\boldsymbol{\kappa}_m(X, Y) := m \int_{-\infty}^{\infty} |x|^{m-1} |H(x)| \, dx. \tag{17.1.29}$$

In the next lemma we prove that the finiteness of $\mathbf{d}_{m+1}$ implies the moment condition (17.1.24).

Lemma 17.1.4. (a) If $X, Y \geqslant 0$ a.s., $\mathbb{E}(X^j - Y^j)$ exists and is finite, $1 \leqslant j \leqslant m$, and $\mathbf{d}_m(X, Y) < \infty$, then $\mathbb{E}(X^j - Y^j) = 0$, $1 \leqslant j \leqslant m - 1$.

(b) If $\mathbf{d}_m(X, Y) < \infty$ and $\boldsymbol{\kappa}_m(X, Y) < \infty$, then $\mathbb{E}(X^j - Y^j) = 0$, $1 \leqslant j \leqslant m - 1$.

Proof. (a) From (17.1.16) we obtain for $t \leqslant 0$

$$\begin{aligned}
&(m-1)! D_m(t) \\
&= \int_t^{\infty} (x - t)^{m-1} H(x) \, dx \\
&= \int_0^{\infty} (x - t)^{m-1} H(x) \, dx = \sum_{j=0}^{m-1} (-t)^{m-1-j} \left(\int_0^{\infty} x^j H(x) \, dx \right) \binom{m-1}{j} \\
&= \sum_{j=0}^{m-1} \binom{m-1}{j} (-t)^{m-1-j} \frac{\mathbb{E}(Y^{j+1} - X^{j+1})}{j+1}.
\end{aligned}$$

Since $\mathbf{d}_m(X, Y) = \sup_t D_m(t) < \infty$, all coefficients of the above polynomial for $j = 0, \ldots, m-2$ have to be zero.

(b) By the assumptions

$$m!\, d_m(X, Y) = \sup_{x \in \mathbb{R}} \left| \int_x^\infty \frac{(t-x)^{m-1}}{(m-1)!} H(t)\, dt \right| < \infty$$

and thus

$$\limsup_{x \to -\infty} \left| \sum_{j=0}^{m-1} \binom{m-1}{j} (-x)^{m-1-j} \int_x^\infty t^j H(t)\, dt \right| < \infty. \tag{17.1.30}$$

Further, by $\boldsymbol{\kappa}_m(X, Y) < \infty$ (see (17.1.29))

$$\limsup_{x \to -\infty} \left| \int_x^\infty t^{m-1} H(t)\, dt \right| \leqslant (1/m)\boldsymbol{\kappa}_m(X, Y) < \infty.$$

Thus,

$$\limsup_{x \to -\infty} \left| \sum_{j=0}^{m-2} \binom{m-j}{j} (-x)^{m-2-j} \int_x^\infty t^j H(t)\, dt \right| = 0. \tag{17.1.31}$$

Since

$$\limsup_{x \to -\infty} \left| \int_x^\infty t^{m-2} H(t)\, dt \right| \leqslant \frac{1}{m-1} \boldsymbol{\kappa}_{m-1}(X, Y) \leqslant 2 + (1/m)\boldsymbol{\kappa}_m(X, Y) < \infty$$

by (17.1.31), we have

$$\limsup_{x \to -\infty} \left| \sum_{j=0}^{m-3} \binom{m-1}{j} (-x)^{m-3-j} \int_x^\infty t^j H(t)\, dt \right| = 0. \tag{17.1.32}$$

Similar to (17.1.31) and (17.1.32) we obtain

$$\limsup_{x \to -\infty} \left| \sum_{j=0}^{m-k} \binom{m-1}{j} (-x)^{m-k-j} \int_x^\infty t^j H(t)\, dt \right| = 0 \tag{17.1.33}$$

for all $k = 2, \ldots, m$. In the case where $k = m$, we have

$$0 = \limsup_{x \to -\infty} \left| \int_x^\infty H(t)\, dt \right| = \limsup_{x \to -\infty} \left| \int_x^\infty t\, dH(t) \right| \tag{17.1.34}$$

and thus $\mathbb{E}(X - Y) = 0$. Put $k = m - 1$ in (17.1.33), then

$$0 = \limsup_{x \to -\infty} \left| (-x) \int_x^\infty H(t)\, dt + (m-1) \int_x^\infty tH(t)\, dt \right|. \tag{17.1.35}$$

By (17.1.34) and $\boldsymbol{\kappa}_m(X, Y) < \infty$

$$\limsup_{x\to-\infty}\left|x\int_x^{\infty} H(t)\,dt\right| = \limsup_{x\to-\infty}\left|x\int_{-\infty}^{x} H(t)\,dt\right|$$
$$\leqslant \limsup_{x\to-\infty}\int_{-\infty}^{x} |t||H(t)|\,dt = 0. \qquad (17.1.36)$$

Combining (17.1.35) and (17.1.36) implies

$$\limsup_{x\to-\infty}\left|\int_x^{\infty} t\,H(t)\,dt\right| = 0$$

and hence $\mathbb{E}(X^2 - Y^2) = 0$. In the same way we get $\mathbb{E}(X^j - Y^j) = 0$ for all $j = 1, \ldots, m-1$. QED

We next establish some relations between the different metrics considered so far. (Further, $\zeta_m := \zeta_{m,1}$ stands for the Zolotarev metric, see (14.2.1).)

Lemma 17.1.5 (a) If $X, Y \geqslant 0$ a.s., $\mathbb{E}(X^j - Y^j)$ is finite, $1 \leqslant j \leqslant m$ and $\mathbf{d}_m(X, Y) < \infty$, then

$$\mathbf{d}_m(X, Y) \leqslant (1/m!)\max\{|\mathbb{E}(X^m - Y^m)|, \boldsymbol{\kappa}_m(X, Y)\}. \qquad (17.1.37)$$

(b) $$\mathbf{d}_m(X, Y) \leqslant \mathbf{d}_{m-1,1}(X, Y) \quad \text{if } x^s H(x) \underset{x\to\infty}{\to} 0$$

$$\zeta_{m,\infty}(X, Y) \leqslant \zeta_m(X, Y) \quad \text{if } \zeta_{m,\infty}(X, Y) < \infty \qquad (17.1.38)$$

$$\mathbf{d}_{m,p}(X, Y) = \zeta_{m,p}(X, Y) \leqslant \zeta_{m+1/p}(X, Y) \quad \text{if } 1 \leqslant p < \infty \text{ and } \zeta_{m,p}(X, Y) < \infty.$$

(c) If $\mathbb{E}(X^j - Y^j) = 0$, $1 < j \leqslant m$, then

$$\mathbf{d}_m(X, Y) = \zeta_{m,\infty}(X, Y) \leqslant (1/m!)\boldsymbol{\kappa}_m(X, Y). \qquad (17.1.39)$$

(d) $$\boldsymbol{\kappa}_m(X, Y) \leqslant \mathbb{E}|X|X|^{m-1} - Y|Y|^{m-1}| \leqslant \mathbb{E}|X|^m + \mathbb{E}|Y|^m.$$

Proof. (a) For $t \geqslant 0$ it follows from (17.1.16) that

$$(m-1)!|D_m(t)| = \left|\int_t^{\infty} (x-t)^{m-1} H(x)\,dx\right| \leqslant \int_t^{\infty} (x-t)^{m-1}|H(x)|\,dx$$
$$\leqslant \int_0^{\infty} x^{m-1}|H(x)|\,dx = (1/m)\boldsymbol{\kappa}_m(X, Y).$$

For $t \leqslant 0$ it follows from Lemma 17.1.4(a) that

$$(m-1)!D_m(t) = \int_t^\infty (x-t)^{m-1}H(x)\,dx = \int_0^\infty (x-t)^{m-1}H(x)\,dx$$
$$= (1/m)\mathbb{E}(Y^m - X^m).$$

(b) From (17.1.16) it follows that if $x^m H(x) \to 0$, then

$$\mathbf{d}_m(X, Y) = \sup_t |D_m(t)| = \sup_t |H_m(t)| = \sup_t \left| \int_t^\infty H_{m-1}(u)\,du \right|$$
$$\leqslant \sup_t \int_t^\infty |D_{m-1}(u)|\,du = \int_{-\infty}^\infty |D_{m-1}(u)|\,du = \mathbf{d}_{m-1,1}(X, Y).$$

If $\mathbb{E}(X^j - Y^j) = 0$, $1 \leqslant j \leqslant m$, then $\zeta_{m,\infty}(X, Y) = \mathbf{d}_m(X, Y) \leqslant \mathbf{d}_{m-1,1}(X, Y) = \zeta_m(X, Y)$. The relation $\zeta_{m,p}(X, Y) \leqslant \zeta_{m+1/p}(X, Y)$ follows from the inequality

$$|f^m(x) - f^m(y)| \leqslant \|f^{(m+1)}\|_q |x-y|^{1/p} \leqslant |x-y|^{1/p}$$

for any function f with $\|f^{(m+1)}\|_q \leqslant 1$ and $1/p + 1/q = 1$.

(c) By (b) and Lemma 17.1.1,

$$\mathbf{d}_m(X, Y) = \zeta_{m,\infty}(X, Y) \leqslant \zeta_m(X, Y). \tag{17.1.40}$$

Further, by (17.1.14) with $p = 1$,

$$\zeta_m(X, Y) = \int_{-\infty}^\infty \left| \int_x^\infty \frac{(t-x)^m}{m!}\,dH(t) \right| dx. \tag{17.1.41}$$

By the assumption, $\mathbb{E}(X^j - Y^j) = 0, j = 1, \ldots, m$

$$\zeta_m(X, Y) = \int_{-\infty}^\infty \left| \int_x^\infty \frac{(t-x)^{m-1}}{(m-1)!} H(t)\,dt \right| dx$$
$$= \int_0^\infty \left| \int_x^\infty \frac{(t-x)^{m-1}}{(m-1)!} H(t)\,dt \right| dx + \int_{-\infty}^0 \left| \int_{-\infty}^x \frac{(x-t)^{m-1}}{(m-1)!} H(t)\,dt \right| dx$$
$$\leq \int_0^\infty \int_x^\infty \frac{(t-x)^{m-1}}{(m-1)!} |H(t)|\,dt\,dx + \int_{-\infty}^0 \int_{-\infty}^x \frac{(x-t)^{m-1}}{(m-1)!} |H(t)|\,dt\,dx$$
$$= (1/m!)\boldsymbol{\kappa}_m(X, Y).$$

(d) Clearly, for any X and Y,

$$\boldsymbol{\kappa}_1(X, Y) = \int_{-\infty}^\infty |F_X(x) - F_Y(x)|\,dx \leqslant E|X - Y|$$

see, for example, (6.4.11). Now use the representation

$$\kappa_m(X, Y) = \kappa_1(X|X|^{m-1}, Y|Y|^{m-1})$$

to complete the proof of (d). QED

The next relations concern the *uniform metric*

$$\boldsymbol{\rho}(X, Y) := \sup_{x \in \mathbb{R}} |F_X(x) - F_y(x)| \tag{17.1.42}$$

the stop-loss distance $\mathbf{d}_m$ (17.1.10) and the pseudomoment $\boldsymbol{\kappa}_m(X, Y)$ (17.1.29).

Lemma 17.1.6. (a) If X has a bounded Lebesgue density f_X, $|f_X(t)| \leqslant M$, then

$$\mathbf{d}_m(X, Y) \geqslant K(m)(1 + M)^{-m-1}\boldsymbol{\rho}(X, Y)^{m+1} \tag{17.1.43}$$

where $K(m) = \dfrac{(m+1)\sqrt{3}}{(2m+2)!\sqrt{2m+3}}$.

(b) If for some $\delta > 0$, $\tilde{m}_\delta := \mathbb{E}(|X|^{m+\delta} + |Y|^{m+\delta}|) < \infty$, then

$$\boldsymbol{\kappa}_m(X, Y) \leqslant 2\left(\frac{\delta \tilde{m}_\delta}{2m}\right)^{m/(m+d)} (\boldsymbol{\rho}(X, Y))^{d/(m+d)} \frac{m+\delta}{\delta}. \tag{17.1.44}$$

Proof. (a) We first apply Lemma 17.1.1, $\mathbf{d}_m = \zeta_{m,\infty}$. Then Lemma 16.2.2 completes the proof of (17.1.43).

(b) For $\alpha > 0$ and $\boldsymbol{\rho} = \boldsymbol{\rho}(X, Y)$

$$\begin{aligned}\boldsymbol{\kappa}_m(X, Y) &= m\int_{-\infty}^{\infty} |x|^{m-1}|H(x)|\, dx \\ &\leqslant m\int_{-\alpha}^{\alpha} |x|^{m-1}|H(x)|\, dx + \mathbb{E}|X|^m\{|X| > \alpha\} + \mathbb{E}|Y|^m\{|Y| > \alpha\} \\ &\leqslant 2\rho\, \alpha^m + \frac{\tilde{m}_\delta}{\alpha^\delta} =: f(\alpha).\end{aligned}$$

Minimizing $f(\alpha)$ we obtain (17.1.44) QED

Remark 17.1.2. The estimate (17.1.44) combined with (17.1.39) shows that

$$\mathbf{d}_m(X, Y) \leqslant (2/m!)\left(\frac{\delta \tilde{m}_\delta}{2m}\right)^{m/(m+\delta)} (\boldsymbol{\rho}(X, Y))^{\delta/(m+\delta)} \frac{m+\delta}{m} \quad \text{if } \zeta_{m,\infty}(X, Y) < \infty. \tag{17.1.45}$$

Under the assumption of a finite moment generating function this can be improved to $\boldsymbol{\rho}(X, Y)\{\log(\zeta(X, Y))\}^\alpha$ for some $\alpha > 0$.

An important step in the proof of the precise rate of convergence in the CLT, the so-called *Berry–Esseen type theorems*, is the smoothing inequality (cf. Lemma 15.2.1, (15.2.7) and Lemma 15.2.3). For the stop-loss metrics there are some similar inequalities leading also to the Berry–Esseen type theorems.

Lemma 17.1.7 (Smoothing inequality). (a) Let Z be independent of X and Y, $\zeta_{m,\infty}(X, Y) < \infty$, then for any $\varepsilon > 0$ holds

$$\mathbf{d}_m(X, Y) \leqslant \mathbf{d}_m(X + \varepsilon Z, Y + \varepsilon Z) + 2\frac{\varepsilon^m}{m!}\mathbb{E}|Z|^m. \tag{17.1.46}$$

(b) If X, Y, Z, W are independent, $x^m H(x) \to 0$, $x \to \infty$, then

$$\mathbf{d}_m(X + Z, Y + Z) \leqslant 2\mathbf{d}_m(Z, W)\boldsymbol{\sigma}(X, Y) + \mathbf{d}_m(X + W, Y + W) \tag{17.1.47}$$

and

$$\mathbf{d}_m(X + Z, Y + Z) \leqslant 2\mathbf{d}_m(X, Y)\boldsymbol{\sigma}W, Z) + \mathbf{d}_m(X + W, Z + W) \tag{17.1.48}$$

where $\boldsymbol{\sigma}$ is the total variation metric, see (14.2.4).

Proof. (a) From Lemmas 17.1.3 and 17.1.5

$$\begin{aligned}\mathbf{d}_m(X, Y) &\leqslant \mathbf{d}_m(X, X + \varepsilon Z) + \mathbf{d}_m(X + \varepsilon Z, Y + \varepsilon Z) + \mathbf{d}_m(Y + \varepsilon Z, Y)\\ &\leqslant \mathbf{d}_m(X + \varepsilon Z, Y + \varepsilon Z) + 2\mathbf{d}_m(0, \varepsilon Z)\\ &\leqslant \mathbf{d}_m(X + \varepsilon Z, Y + \varepsilon Z) + 2\frac{\varepsilon^m}{m!}\boldsymbol{\kappa}_m(0, Z)\\ &= \mathbf{d}_m(X + \varepsilon Z, Y + \varepsilon Z) + 2\frac{\varepsilon^m}{m!}\mathbb{E}|Z|^m.\end{aligned}$$

(b) $\mathbf{d}_m(X + Z, Y + Z)$

$$\begin{aligned}&= \frac{1}{(m-1)!}\sup_x\left|\int_x^\infty (t-x)^{m-1}(F_{X+Z}(t) - F_{Y+Z}(t))\,\mathrm{d}t\right|\\ &= \frac{1}{(m-1)!}\sup_x\left|\int_x^\infty\left[\int_{-\infty}^\infty (t-x)^{m-1}F_Z(t-u)\,\mathrm{d}(F_X(u) - F_Y(u))\right]\mathrm{d}t\right|\\ &\leqslant \frac{1}{(m-1)!}\sup_x\left|\int_x^\infty\left[\int_{-\infty}^\infty (t-x)^{m-1}\{F_Z(t-u) - F_W(t-u)\}\,\mathrm{d}H(u)\right]\mathrm{d}t\right|\\ &\quad + \frac{1}{(m-1)!}\sup_x\left|\int_x^\infty\left[\int_{-\infty}^\infty (t-x)^{m-1}F_W(t-u)\,\mathrm{d}H(u)\right]\mathrm{d}t\right|\\ &\leqslant \int_{-\infty}^\infty \mathbf{d}_m(Z, W)|\mathrm{d}H(u)| + \mathbf{d}_m(X + W, Y + W)\\ &= 2\mathbf{d}_m(Z, W)\sigma(X, Y) + \mathbf{d}_m(X + W, Y + W).\end{aligned}$$

The inequality (17.1.48) is derived similarly. QED

From the smoothing inequality we obtain the following relation between $\mathbf{d}_1$, $\mathbf{d}_m$.

Lemma 17.1.8. If $\mathbb{E}(X^j - Y^j) = 0$, $1 \leqslant j \leqslant m$, then

$$\mathbf{d}_1(X, Y) \leqslant \lambda_m(\mathbf{d}_m(X, Y))^{1/m} \tag{17.1.49}$$

where

$$\lambda_m := \mathbb{K}_m^{1/m}\left(\frac{2\mathbb{K}_2}{m-1}\right)^{(m-1)/m} m \quad \mathbb{K}_m := \int |\mathbb{H}_{m-1}(x)| \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\, dx$$

$\mathbb{H}_m$ is the Hermite polynomial of order m; $\mathbb{K}_1 = 1$, $\mathbb{K}_2 = (2/\pi)^{1/2}$. In particular,

$$\mathbf{d}_1(X, Y) \leqslant (4/\sqrt{\pi})(\mathbf{d}_2(X, Y))^{1/2}. \tag{17.1.50}$$

Proof. Let Z be a $N(0, 1)$-distributed r.v. independent of X, Y. Then for $\varepsilon > 0$ from (17.1.46)

$$\mathbf{d}_1(X, Y) \leqslant \mathbf{d}_1(X + \varepsilon Z, Y + \varepsilon Z) + 2\varepsilon(2/\pi)^{1/2}. \tag{17.1.51}$$

With $W := \varepsilon Z$ it follows from Lemma 17.1.1 (see (17.1.13)) that

$$\begin{aligned}\mathbf{d}_1(X + W, Y + W) &= \sup\{|\mathbb{E}(f(X + W) - f(Y + W))|;\ \|f^{(2)}\|_1 \leqslant 1\}\\ &= \sup\{|\mathbb{E}(g_f(X) - g_f(Y)|;\ \|f^{(2)}\|_1 \leqslant 1\}\end{aligned}$$

where

$$g_f(t) := \int_{-\infty}^{\infty} f(x) f_W(x - t)\, dx = f * f_W(t) \quad f_W := F'_W.$$

The derivatives of g_f have the following representation:

$$\begin{aligned}g_f^{(m)}(t) &= (-1)^m \int_{-\infty}^{\infty} f(x) f_W^{(m)}(x - t)\, dx = (-1)^m \int_{-\infty}^{\infty} f(x + t) f_W^{(m)}(x)\, dx\\ &= (-1)^{m-1} \int_{-\infty}^{\infty} f^{(1)}(x + t) f_W^{(m-1)}(x)\, dx\end{aligned}$$

and

$$g_f^{(m+1)}(t) = (-1)^{m-1} \int_{-\infty}^{\infty} f^{(2)}(x + t) f_W^{(m-1)}(x)\, dx = (-1)^{m-1} f^{(2)} * f_W^{(m-1)}(t).$$

For the L^1-norm we therefore obtain

$$\begin{aligned}\|g_f^{(m-1)}\|_1 &= \int |g_f^{(m+1)}(t)|\, dt = \|f^{(2)} * f_W^{(m-1)}\|_1\\ &\leqslant \|f^{(2)}\|_1 \|f_W^{(m-1)}\|_1 \leqslant \frac{1}{\varepsilon^{m-1}} \|f_Z^{(m-1)}\|_1 = \frac{1}{\varepsilon^{m-1}} \mathbb{K}_m.\end{aligned}$$

Therefore, from Lemma 17.1.1

$$\mathbf{d}_1(X + \varepsilon Z, Y + \varepsilon Z) = \zeta_{1,\infty}(X + \varepsilon Z, Y + \varepsilon Z) \leqslant \frac{1}{\varepsilon^{m-1}} \mathbb{K}_m \mathbf{d}_m(X, Y). \quad (17.1.52)$$

From the smoothing inequality (17.1.46) we obtain

$$\mathbf{d}_1(X, Y) \leqslant \frac{1}{\varepsilon^{m-1}} \mathbb{K}_m \mathbf{d}_m(X, Y) + 2\mathbb{K}_2 \varepsilon.$$

Minimizing the right-hand side with respect to ε, we obtain (17.1.49). QED

Lemma 17.1.9. Let Z be independent from X, Y with Lebesgue density f_Z.

(a) If $C_{3,Z} := \|f_Z^{(3)}\|_1 < \infty$, then

$$\sigma(X + Z, Y + Z) \leqslant \tfrac{1}{2} C_{3,Z} \mathbf{d}_{2,1}(X, Y). \quad (17.1.53)$$

(b) If $C_{s,Z} := \|f_Z^{(s)}\|_1 < \infty$, and if $\zeta_{m,\infty}(X, Y) < \infty$, then for $m \geqslant 1$

$$\mathbf{d}_m(X + Z, Y + Z) \leqslant C_{s,Z} \zeta_{m+s}(X, Y). \quad (17.1.54)$$

Proof. (a) With $H(t) = F_X(t) - F_Y(t)$

$$\begin{aligned}
2\sigma(X + Z, Y + Z)& \\
&= \int |f_{X+Z}(x) - f_{Y+Z}(x)|\, dx = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f_Z(x - t)\, dH(t) \right| dx \\
&= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} H(t) f_Z'(x - t)\, dt \right| dx = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f_Z'(x - t)\, d\left(\int_x^\infty H(u)\, du \right) \right| dx \\
&= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} \left(\int_x^\infty H(u)\, du \right) f_Z''(x - t)\, dt \right| dx \\
&\leqslant \int_{\mathbb{R}} \int_{\mathbb{R}} \left| \int_x^\infty \frac{(u - x)}{1!} H(u)\, du \right| |f_Z^{(3)}(x - t)|\, dt\, dx \\
&= \tfrac{1}{2} C_{3,Z} \int_{\mathbb{R}} |E(X - t)_+^2 - E(Y - t)_+^2|\, dt = C_{3,Z} \mathbf{d}_{2,1}(X, Y).
\end{aligned}$$

(b) If $C_{s,Z} = \|f_Z^{(s)}\|_1 < \infty$ and $\zeta_{m,\infty}(X, Y) < \infty$, then by (17.1.38), similar to (a), we get

$$\mathbf{d}_m(X + Z, Y + Z) \leqslant \zeta_m(X + Z, Y + Z) \leqslant C_{s,Z} \zeta_{m+s}(X, Y)$$

(cf. Zolotarev (1986), Theorem 1.4.5, and Kalashnikov and Rachev (1988), Chapter 3, p. 10, Theorem 10).

Theorem 17.1.1. Let $\{X_n\}$ be i.i.d. $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$, $S_n = n^{-1/2}\sum_{i=1}^n X_i$, and Y be a standard normal r.v. Then, for $m = 1, 2$,

$$\mathbf{d}_m(S_n, Y) \leqslant C(m)\max\{\mathbf{d}_{m,1}(X_1, Y), \mathbf{d}_{3,1}(X, Y)\}n^{-1/2} \qquad (17.1.55)$$

where $C(m)$ is an absolute constant, and

$$\mathbf{d}_3(S_n, Y) \leqslant d_3(X_1, Y)n^{-1/2}. \qquad (17.1.56)$$

Proof. The inequality (17.1.56) is a direct consequence of Lemma 17.1.3(b). The proof of (17.1.55) is based on Lemmas 17.1.3, 17.1.5, 17.1.7 and 17.1.9, and follows the proof of Theorem 15.2.1 step by step. QED

Remark 17.1.3. In terms of ζ_s metrics, a similar inequality is given by Zolotarev (1986), Theorem 5.4.7,

$$\zeta_1(S_n, Y) \leqslant 11.5\max(\zeta_1(X_1, Y)\zeta_3(X_1, Y))n^{-1/2}. \qquad (17.1.57)$$

Open problem 17.1.1. Regarding the right-hand side of (17.1.55) one could expect that the better bound should be $C(m)\max\{\mathbf{d}_m(X_1, Y), \mathbf{d}_{3,1}(X_1, Y)\}n^{-1/2}$. Is it true that for $m = 1, 2, p \in (1, \infty]$

$$\mathbf{d}_{m,p}(S_n, Y) \leqslant C(m, p)\max\{\mathbf{d}_{m,p}(X_1, Y), \mathbf{d}_{3,1}(X_1, Y)\}n^{-1/2}? \qquad (17.1.58)$$

What is a good bound for $C(m, p)$ in (17.1.58)?

17.2 APPROXIMATION BY COMPOUND POISSON DISTRIBUTIONS

We now consider the problem of approximation of the individual model $S^{\text{ind}} = \sum_{i=1}^n X_i = \sum_{i=1}^n C_i D_i$ (see (17.1.3), (17.1.4)) by a compound model (cf. (17.1.5)), i.e., by a compound Poisson distributed r.v.

$$S^{\text{coll}} = \sum_{i=1}^{N} Z_i \stackrel{\text{d}}{=} \sum_{i=1}^{n} S_i^{\text{coll}}, \; S_i^{\text{coll}} = \sum_{j=1}^{N_i} Z_{ij}.$$

Choose $Z_{ij} \stackrel{\text{d}}{=} u_i C_i$ and N_i to be Poisson (μ_i)-distributed. Then N is Poisson (μ) $(N \stackrel{\text{d}}{=} \mathscr{P}(\mu))$, $\mu = \sum_{i=1}^n \mu_i$ and

$$F_{Z_j} = \sum_{i=1}^{n} \frac{\mu_i}{\mu} F_{u_i C_i}. \qquad (17.2.1)$$

We choose μ_i, u_i in a way such that the first two moments of S_i^{coll} coincide with the corresponding moments of X_i.

Lemma 17.2.1. Let $a_i := \mathbb{E}C_i$, $b_i := \mathbb{E}C_i^2$, and define

$$\mu_i := \frac{p_i b_i}{b_i - p_i a_i^2} \quad u_i := \frac{p_i}{\mu_i} = \frac{b_i - p_i a_i^2}{b_i}. \tag{17.2.2}$$

Then

$$\mathbb{E}S_i^{\mathrm{coll}} = \mathbb{E}X_i = p_i a_i \quad \text{and} \quad \mathrm{Var}(S_i^{\mathrm{coll}}) = \mathrm{Var}(X_i) = p_i b_i - (p_i a_i)^2. \tag{17.2.3}$$

Proof. Since $N_i \stackrel{\mathrm{d}}{=} \mathscr{P}(\mu_i)$, $Z_{ij} \stackrel{\mathrm{d}}{=} u_i C_i$ we obtain $\mathbb{E}Z_{ij} = u_i a_i$, $\mathbb{E}Z_{ij}^2 = u_i^2 b_i$

$$\mathbb{E}S_i^{\mathrm{coll}} = \mathbb{E}\sum_{j=1}^{N_i} Z_{ij} = \mu_i u_i a_i = p_i a_i = \mathbb{E}X_i$$

and

$$\begin{aligned}\mathrm{Var}(X_i) &= p_i b_i - (p_i a_i)^2 = p_i(b_i - p_i a_i^2) = \frac{p_i^2 b_i}{\mu_i} = \mu_i u_i^2 b_i = (\mathbb{E}N_i)\mathbb{E}Z_{ij}^2 \\ &= \mathrm{Var}(S_i^{\mathrm{coll}}).\end{aligned}$$

QED

So in contrast to the 'usual' choice (17.1.6) of $\mu = \tilde{\mu}$, $v = \tilde{v}$ we use a scaling factor u_i and μ_i such that the first two moments agree. We see that

$$\mu_i > p_i \quad \text{for } p_i > 0 \tag{17.2.4}$$

and $u_i < 1$.

Theorem 17.2.1. Let μ_i, u_i be as defined in (17.2.2), then

$$\mathbf{d}_1(S^{\mathrm{ind}}, S^{\mathrm{coll}}) \leqslant \frac{4}{\sqrt{\pi}}\left(\sum_{i=1}^{n} p_i b_i\right)^{1/2} \tag{17.2.5}$$

$$\mathbf{d}_2(S^{\mathrm{ind}}, S^{\mathrm{coll}}) \leqslant \sum_{i=1}^{n} p_i b_i.$$

Proof. By (17.1.50) we have $\mathbf{d}_1(S^{\mathrm{ind}}, S^{\mathrm{coll}}) \leqslant (4/\sqrt{\pi})\,(\mathbf{d}_2(S^{\mathrm{ind}}, S^{\mathrm{coll}})^{1/2}$. Now

$$\mathbf{d}_2(S^{\mathrm{ind}}, S^{\mathrm{coll}}) = \mathbf{d}_2\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} S_i^{\mathrm{coll}}\right) \leqslant \sum_{i=1}^{n} \mathbf{d}_2(X_i, S_i^{\mathrm{coll}})$$

by Lemma 17.1.3. By Lemma 17.1.5(b), (d), it follows that

$$\mathbf{d}_2(X_i, S_i^{\mathrm{coll}}) \leqslant \tfrac{1}{2}(\mathbb{E}X_i^2 + \mathbb{E}(S_i^{\mathrm{coll}})^2) = \mathbb{E}X_i^2 = p_i b_i.$$

QED

Remark 17.2.1. Note that in our model $\mathbf{d}_2(S^{\mathrm{ind}}, S^{\mathrm{coll}}) = \frac{1}{2}\sup_t |\mathbb{E}(S^{\mathrm{ind}} - t)_+^2 - \mathbb{E}(S^{\mathrm{coll}} - t)_+^2|$ is finite. In view of Lemma 17.1.4 this is not necessarily true for the usual model. By Lemma 17.1.2 the $\zeta_{2,\infty}$-distance between S^{ind} and S^{coll} is infinite in the usual model while $\zeta_{2,\infty}(S^{\mathrm{ind}}, S^{\mathrm{coll}}) = \mathbf{d}_2(S^{\mathrm{ind}}, S^{\mathrm{coll}})$ is finite in our 'scaled' model determined by (17.2.1)–(17.2.3). Moreover, the $\mathbf{d}_3$-metric for the usual model is infinite as follows from Lemma 17.1.4. This indicates more stability in our new scaled approximation.

We shall next consider the special case where $\{X_i\}_{i\geqslant 1}$ are i.i.d. For this purpose we shall use a Berry–Esseen type estimate for $\mathbf{d}_1$, see Theorem 17.1.1 and Remark 17.1.2. In the next theorem we use the following moment characteristic

$$\tau_3(X, Y) := \max((\mathbb{E}|\tilde{X}| + \mathbb{E}|Y|), \tfrac{1}{3}(\mathbb{E}|\tilde{X}|^3 + \mathbb{E}|Y|^3)) \quad \tilde{X} := \frac{X - \mathbb{E}X}{\mathrm{Var}(X)}.$$

Theorem 17.2.2. If $\{X_i\}$ are i.i.d. with $a = \mathbb{E}C_1$, $\sigma^2 = \mathrm{Var}(C_1)$, $p = \Pr(D_i = 1)$, then

$$\mathbf{d}_1(S^{\mathrm{ind}}, S^{\mathrm{coll}}) \leqslant 11.5[p\sigma^2 + p(1-p)a^2]^{1/2}\left(\tau_3(X_1, Y) + \tau_3\left(\sum_{i=1}^{N_1} C_i, Y\right)\right) \tag{17.2.6}$$

where Y has a standard normal distribution and $N_1 \overset{\mathrm{d}}{=} \mathscr{P}(\mu_1)$.

Proof. By the ideality of $\mathbf{d}_1$ (see Lemma 17.1.3)

$$\begin{aligned}\mathbf{d}_1(S^{\mathrm{ind}}, S^{\mathrm{coll}}) &= \mathbf{d}_1\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} S_i^{\mathrm{coll}}\right)\\ &= (n\ \mathrm{Var}(X_1))^{1/2}\mathbf{d}_1\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_i\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right)\end{aligned}$$

where $Y_i := (\tilde{S}_i^{\mathrm{coll}})$. By the triangle inequality, (17.1.57) and Lemma 17.1.5(c),

$$\begin{aligned}\mathbf{d}_1\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_i, \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right) &\leqslant \boldsymbol{\kappa}_1\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_i, Y\right) + \boldsymbol{\kappa}_1\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i, Y\right)\\ &\leqslant 11.5\{\max(\boldsymbol{\kappa}_1(\tilde{X}_1, Y), \mathbf{d}_{2,1}(\tilde{X}_1, Y))\\ &\quad + \max(\boldsymbol{\kappa}_1(Y_1, Y), \mathbf{d}_{2,1}(Y_1, Y))\}n^{-1/2}\end{aligned}$$

where $\boldsymbol{\kappa}_1 = \zeta_1 = \zeta_{1,1}$ is the first difference pseudomoment, see (17.1.29). With

$$\boldsymbol{\kappa}_1(\tilde{X}_1, Y) \leqslant \mathbb{E}|\tilde{X}_1| + \mathbb{E}|Y| \quad d_{2,1}(\tilde{X}_1, Y) \leqslant \tfrac{1}{2}(\mathbb{E}|\tilde{X}_1|^3 + \mathbb{E}|Y|^3)$$

and similarly for the second term we get

$$\max(\boldsymbol{\kappa}_1(Y_1, Y), \mathbf{d}_{2,1}(Y_1, Y)) \leqslant \tau_3(Y_1, Y) \leqslant \tau_3\left(\sum_{i=1}^{N_1} C_i, Y\right). \qquad \text{QED}$$

The next theorem gives a better estimate for $\mathbf{d}_1(S^{\text{ind}}, S^{\text{coll}})$ when p_i are relatively small.

Theorem 17.2.3. Let μ_i, u_i be as in (17.2.2) and let $C_i \geqslant 0$ a.s., then for any $\Delta_i > 1$

$$\mathbf{d}_1(S^{\text{coll}}, S^{\text{ind}}) \leqslant \sum_{i=1}^{n} p_i^2 \tau_i \tag{17.2.7}$$

where $\tau_i := a_i + \Delta_i v_i + \max(\Delta_i a_i v_i, 2a_i \tilde{v}_i + (1 + \Delta_i a_i v_i p_i) u_i)$, $v_i := a_i^2/b_i \leqslant 1/p_i$, $p_i v_i \leqslant 1 - \Delta_i^{-1}$, and $\tilde{v}_i := a_i^2/(b_i - p_i a_i^2)$.

Proof. Since $\mathbf{d}_1$ is an ideal metric, for the proof it is enough to establish

$$\mathbf{d}_1(S_i^{\text{coll}}, X_i) \leqslant p_i^2 \tau_i \tag{17.2.8}$$

see (17.1.4), (17.1.7) and (17.1.26). We shall omit the index i in the following. Since the first moments of S^{ind}, S^{coll} are the same, $|D_{1,S^{\text{coll}},X}(t)|$ determined by (17.1.12) admits the form

$$|D_{1,S^{\text{coll}},X}(t)| = \left| \int_{-\infty}^{t} (t - x)(\mathrm{d}F_{S^{\text{coll}}}(x) - \mathrm{d}F_X(x)) \right|.$$

Further, we shall consider only the case where $t > 0$ since the case where $t < 0$ can be handled in the same manner using the above equality. For $t > 0$,

$$\begin{aligned}
|D_{1,S^{\text{coll}},X}(t)| &:= \left| \int_t^{\infty} (x - t)(\mathrm{d}F_{S^{\text{coll}}}(x) - \mathrm{d}F_X(x)) \right| \\
&= \left| \sum_{k=1}^{\infty} \frac{\mu^k}{k!} \exp(-\mu) \int_t^{\infty} (x - t)\, \mathrm{d}F_{uC}^{*k}(x) - p \int_t^{\infty} (x - t)\, \mathrm{d}F_C(x) \right| \\
&\leqslant I_1 + I_2,
\end{aligned}$$

where

$$I_1 := \left| \mu \exp(-\mu) \int_t^{\infty} (x - t)\, \mathrm{d}F_{uC}(x) - p \int_t^{\infty} (x - t)\, \mathrm{d}F_C(x) \right| \tag{17.2.9}$$

and

$$I_2 := \sum_{n=2}^{\infty} \frac{\mu^k}{k!} \exp(-\mu) \int_t^{\infty} (x - t)\, \mathrm{d}F_{uC}^{*k}(x).$$

Since $u = p/\mu$, it follows that

$$I_2 \leqslant \sum_{k=2}^{\infty} \frac{\mu^k}{k!} \exp(-\mu) kua = \mu \exp(-\mu) ua(\exp(\mu) - 1) \leqslant ua\mu^2 = pa\mu. \tag{17.2.10}$$

Using $\mu = pb/(b - pa^2) = p/(1 - pa^2/b)$ we obtain

$$p \leqslant \mu \leqslant p\left(1 + \Delta \frac{a^2 p}{b}\right) \tag{17.2.11}$$

therefore,

$$I_2 \leqslant pa\mu \leqslant pap\left(1 + \Delta \frac{a^2 p}{b}\right) = p^2 a + \Delta \frac{a^2 p^3}{b} \tag{17.2.12}$$

For the estimate of I_1 we use $\bar{F}_C := 1 - F_C$ to obtain

$$I_1 = \left| \mu \exp(-\mu) \int_t^\infty \bar{F}_{uC}(x)\, \mathrm{d}x - p \int_t^\infty \bar{F}_C(x)\, \mathrm{d}x \right|.$$

Since by (17.2.11) $u = p/\mu \leqslant 1$ and, therefore, $\bar{F}_{uC}(x) \leqslant \bar{F}_C(x)$, we obtain

$$\begin{aligned} &\mu \exp(-\mu) \int_t^\infty \bar{F}_{uC}(x)\, \mathrm{d}x - p \int_t^\infty \bar{F}_C(x)\, \mathrm{d}x \\ &\leqslant p\left(1 + \Delta \frac{a^2 p}{b}\right) \exp(-p) \int_t^\infty \bar{F}_C(x)\, \mathrm{d}x - p \int_t^\infty \bar{F}_C(x)\, \mathrm{d}x \\ &\leqslant \Delta\left(\frac{a^2}{b}\right) p^2(\mathbb{E}C) = \Delta \frac{a^3 p^2}{b} \end{aligned} \tag{17.2.13}$$

On the other hand, by (17.2.11), $\exp(-\mu) \geqslant 1 - \mu \geqslant 1 - p(1 + \Delta a^2 p/b)$ implying

$$\begin{aligned} A &:= p \int_t^\infty \bar{F}_C(x)\, \mathrm{d}x - \mu \exp(-\mu) \int_t^\infty \bar{F}_{uC}(x)\, \mathrm{d}x \\ &\leqslant p \int_t^\infty \bar{F}_C(x)\, \mathrm{d}x - p\left(1 - p - \Delta \frac{a^2 p^2}{b}\right) \int_t^\infty \bar{F}_{uC}(x)\, \mathrm{d}x \\ &\leqslant p\left(\int_t^\infty \bar{F}_C(x)\, \mathrm{d}x - \int_t^\infty \bar{F}_{uC}(x)\, \mathrm{d}x\right) + p^2\left(1 + \Delta \frac{a^2 p}{b}\right) ua. \end{aligned} \tag{17.2.14}$$

Now since

$$u = p/\mu = \frac{(b - pa^2)}{b} = 1 - \frac{pa^2}{b}$$

then

$$p\left(\int_t^\infty \bar{F}_C(x)\, \mathrm{d}x - \int_{t/u}^\infty \bar{F}_C(y)\, \mathrm{d}y\right) + p^2\left(\frac{a^2}{b}\right) \int_{t/u}^\infty \bar{F}_C(x)\, \mathrm{d}x$$

$$\leqslant p \int_t^{t/u} \bar{F}_C(x)\,\mathrm{d}x + \frac{p^2a^3}{b} \leqslant p\bar{F}_C(t)t\left(\frac{1}{u} - 1\right) + \frac{p^2a^3}{b}$$

$$\leqslant pa\left(\frac{1}{u} - 1\right) + \frac{p^2a^3}{b} \leqslant 2p^2 \frac{a^3}{b - pa^2}. \tag{17.2.15}$$

Thus by (17.2.14), (17.2.15)

$$A \leqslant 2p^2 \frac{a^3}{b - pa^2} + p^2\left(1 + \Delta \frac{a^2p}{b}\right)au. \tag{17.2.16}$$

The estimates (17.2.16) and (17.2.14) imply

$$I_1 \leqslant \max\left(\Delta \frac{a^3p^2}{b}, 2p^2 \frac{a^3}{b - pa^2} + ap^2\left(1 + \Delta \frac{a^2p}{b}\right)u\right)$$

$$= ap^2 \max\left(\Delta \frac{a^2}{b}, 2 \frac{a^2}{b - pa^2} + u + \Delta \frac{a^2u}{b} p\right). \tag{17.2.17}$$

Thus the required bound (17.2.7) follows from (17.2.12) and (17.2.17). QED

Remark 17.2.2. From the regularity of $\mathbf{d}_1$ it follows that

$$\mathbf{d}_1(S^{\text{coll}}, S^{\text{ind}}) \leqslant \sum_{i=1}^{n} \mathbf{d}_1(S_i^{\text{coll}}, X_i)$$

$$\leqslant \sum_{i=1}^{n} (\mathbb{E}S_i^{\text{coll}} + \mathbb{E}X_i) = 2\sum_{i=1}^{n} a_ip_i. \tag{17.2.18}$$

Clearly, for p_i small the estimate (17.2.7) is a refinement of the above bound.

We next give a direct estimate for $\mathbf{d}_2$ and use the relation between $\mathbf{d}_2$ and $\mathbf{d}_1$ in order to obtain an improved estimate for $\mathbf{d}_1$ for p_i not too small.

Theorem 17.2.4. Let $C_i \geqslant 0$ a.s. and let μ_i, u_i be as in (17.2.2); then

$$\mathbf{d}_2(S^{\text{ind}}, S^{\text{coll}}) \leqslant \tfrac{1}{2} \sum_{i=1}^{n} p_i^2\tau_i^* \tag{17.2.19}$$

where

$$\tau_i^* := b_i + 3a_i^2 + \Delta_ia_i^2 + 2\tilde{v}_ib_i^2 + b_iu_i^2 + \Delta_ia_ip_i \tag{17.2.20}$$

and Δ_i, $\tilde{v}_i$ are defined as in Theorem 17.2.3. Moreover,

$$\mathbf{d}_1(S^{\text{ind}}, S^{\text{coll}}) \leqslant (4/\sqrt{\pi})\left(\sum_{i=1}^{n} p_i^2\tau_i^*\right)^{1/2}. \tag{17.2.21}$$

Proof. Again it is enough to consider $\mathbf{d}_1(S_i^{\text{coll}}, X_i)$ and we shall omit the index i. Then for $t > 0$

$$\left|\int_t^\infty (x-t)^2\, \mathrm{d}(F_{S^{\text{coll}}}(x) - F_X(x))\right|$$

$$= \left|\sum_{k=1}^\infty \frac{\mu^k}{k!}\exp(-\mu)\int_t^\infty (x-t)^2\, \mathrm{d}F_{uC}^{*k}(x) - p\int_t^\infty (x-t)^2\, \mathrm{d}F_C(x)\right| \leqslant I_1 + I_2 \tag{17.2.22}$$

where

$$I_1 := \left|\mu\exp(-\mu)\int_t^\infty (x-t)^2\, \mathrm{d}F_{uC}(x) - p\int_t^\infty (x-t)^2\, \mathrm{d}F_C(x)\right|$$

and

$$I_2 := \sum_{k=2}^\infty \frac{\mu^k}{k!}\exp(-\mu)\int_t^\infty (x-t)^2\, \mathrm{d}F_{uC}^{*k}(x).$$

Since $u = p/\mu$, we obtain the following

$$\begin{aligned} I_2 &\leqslant \sum_{k=2}^\infty \frac{\mu^k}{k!}\exp(-\mu)\mathbb{E}\left(\sum_{i=1}^k uC_i\right)^2 = u^2\sum_{k=2}^\infty \frac{\mu^k}{k!}\exp(-\mu)(kb + k(k-1)a^2) \\ &= u^2 b\mu\exp(-\mu)(\exp(\mu)-1) + u^2a^2\mu^2\exp(-\mu)\exp(\mu) \leqslant u^2(b+a^2)\mu^2 \\ &= p^2(b+a^2). \end{aligned} \tag{17.2.23}$$

Furthermore, from the fact that $uC \leqslant C$ and (17.2.11)

$$\tfrac{1}{2}I_1 = \left|\mu\exp(-\mu)\int_t^\infty (x-t)\bar{F}_{uC}(x)\,\mathrm{d}x - p\int_t^\infty (x-t)\bar{F}_C(x)\,\mathrm{d}x\right| =: |A| \tag{17.2.24}$$

and

$$A \leqslant p\left(1 + \Delta\frac{a^2p}{b}\right)\int_t^\infty (x-t)\bar{F}_C(x)\,\mathrm{d}x - p\int_t^\infty (x-t)\bar{F}_C(x)\,\mathrm{d}x \leqslant \tfrac{1}{2}\Delta a^2p^2. \tag{17.2.25}$$

On the other hand, by $\mu \geqslant p$, $\exp(-\mu) \geqslant 1 - \mu \geqslant 1 - p(1 + \Delta(a^2/b)p)$ and

$$\begin{aligned} -A &\leqslant p\int_t^\infty (x-t)\,\bar{F}_C(x)\,\mathrm{d}x - p\left(1 - p - \Delta\frac{a^2p^2}{b}\right)\int_t^\infty (x-t)\bar{F}_{uC}(x)\,\mathrm{d}x \\ &\leqslant p\left(\int_t^\infty (x-t)\bar{F}_C(x)\,\mathrm{d}x - \int_t^\infty (x-t)\bar{F}_{uC}(x)\,\mathrm{d}x\right) + p^2\left(1 + \Delta\frac{a^2p}{b}\right)u^2b/2 \end{aligned}$$

and

$$\int_t^\infty (x-t)\bar{F}_C(x)\,dx - \int_t^\infty (x-t)\bar{F}_{uC}(x)\,dx$$

$$= \int_t^\infty x\bar{F}_C(x)\,dx - \int_t^\infty x\bar{F}_{uC}(x)\,dx + \int_t^\infty t(\bar{F}_{uC}(x) - \bar{F}_C(x))\,dx$$

$$\leqslant \int_t^\infty x\bar{F}_c(x)\,dx - \int_{t/u}^\infty y\bar{F}_C(y)u^2\,dy$$

$$\leqslant \int_t^\infty x\bar{F}_C(x)\,dx - \int_{t/u}^\infty y\bar{F}_C(y)\left(1 - \frac{pa^2}{b}\right)^2 dy$$

$$= \int_t^{t/u} x\bar{F}_C(x)\,dx + \left(\frac{2pa^2}{b}\right)\int_t^\infty x\bar{F}_C(x)\,dx$$

$$\leqslant \frac{t}{u}\,F_C(t)\left(\frac{t}{u} - t\right) + \left(\frac{2pa^2}{b}\right)\frac{b}{2}$$

$$\leqslant \frac{b(1-u)}{u^2} + (pa^2) = \frac{b^2pa^2}{(b-pa^2)^2} + pa^2.$$

Thus,

$$-A \leqslant \frac{b^2p^2a^2}{(b-pa^2)^2} + p^2a^2 + p^2\left(1 + \frac{\Delta a^2 p}{b}\right)\frac{bu^2}{2}.$$

So we obtain

$$I_1 \leqslant \max\left(\Delta a^2p^2, \frac{2b^2p^2a^2}{(b-pa^2)^2} + 2p^2a^2 + p^2\left(1 + \frac{\Delta a^2 p}{b}\right)bu^2\right),$$

which together with (17.2.23) implies (17.2.19).
(17.2.21) is a consequence of (17.2.19) and (17.1.50). QED

As corollary we obtain an estimate for

$$\mathbf{V}_2(S^{\text{ind}}, S^{\text{coll}}) := \sup_t |\mathbf{Var}((S^{\text{ind}} - t)_+) - \mathbf{Var}((S^{\text{coll}} - t)_+)|.$$

Corollary 17.2.1.

$$\mathbf{V}_2(S^{\text{ind}}, S^{\text{coll}}) \leqslant 2\sum_{i=1}^n p_i^2\tau_i^* + \left(\sum_{i=1}^n p_i^2\tau_i\right)2\sum_{i=1}^n p_i a_i$$

where τ_i^* is defined by (17.2.20) and τ_i is the same as in (17.2.7).

Proof.

$$\begin{aligned}
&\mathbf{V}_2(S^{\text{ind}}, S^{\text{coll}}) \\
&\quad \leqslant \sup_t |\mathbb{E}(S^{\text{coll}} - t)_+^2 - \mathbb{E}(S^{\text{ind}} - t)_+^2| + \sup_t |(\mathbb{E}(S^{\text{coll}} - t)_+)^2 - (\mathbb{E}(S^{\text{ind}} - t)_+)^2| \\
&\quad \leqslant 2\mathbf{d}_2(S^{\text{coll}}, S^{\text{ind}}) + \mathbf{d}_1(S^{\text{ind}}, S^{\text{coll}}) \sup_t (\mathbb{E}(S^{\text{coll}} - t)_+ + \mathbb{E}(S^{\text{ind}} - t)_+) \\
&\quad \leqslant 2\mathbf{d}_2(S^{\text{coll}}, S^{\text{ind}}) + \mathbf{d}_1(S^{\text{ind}}, S^{\text{coll}})(\mathbb{E}S^{\text{coll}} + \mathbb{E}S^{\text{ind}}) \\
&\quad \leqslant 2\sum_{i=1}^n p_i^2\tau_i^* + \left(\sum_{i=1}^n p_i^2\tau_i\right)2\sum_{i=1}^n p_i a_i
\end{aligned}$$

the last inequality follows from (17.2.19) and (17.2.7). QED

Remark 17.2.3. One could try to find r.v. $\{Z_{ij}\}_{j\geqslant 1}$ (not necessarily scaled versions of X_i) such that the first k moments of S_i^{coll} coincide with those of X_i. For this purpose (omitting the index i) let $\phi_X(s) = \mathbb{E}s^X$ denote the generating function of X. Then for $Y = \sum_{j=1}^N Z_j$, a compound Poisson distributed r.v. with N, Poisson $\mathscr{P}(\mu)$, we have $\phi_Y(s) = \phi_N(\phi_Z(s))$, where $\phi_Z := \phi_{Z_1}$.

Now for the Poisson r.v. N we obtain the factorial moments $\phi_N^{(k)}(1) = \mathbb{E}N(N-1)\cdots(N-k+1) = \mu^k$. Denote the factorial moments $b_k := \mathbb{E}X(X-1)\cdots(X-k+1)$, $a_k := \mathbb{E}Z(Z-1)\cdots(Z-k+1)$. This implies $\mathbb{E}Y = \phi_Y^{(1)}(1) = \mu a_1$, $\mathbb{E}Y(Y-1) = \phi_Y^{(2)}(1) = \mu^2 a_1^2 + \mu a_2$, $\mathbb{E}Y(Y-1)(Y-2) = \phi_Y^{(3)}(1) = \mu^3 a_1 a_2 + \mu a_3$.

So we obtain the equations

$$\begin{aligned}
\phi_Y^{(1)}(1) &= \mu a_1 = b_1 \\
\phi_Y^{(2)}(1) &= \mu^2 a_1^2 + \mu a_2 = b_2 \\
\phi_Y^{(3)}(1) &= \mu^3 a_1^3 + 3\mu^2 a_1 a_2 + \mu a_3 = b_3
\end{aligned}$$

etc., i.e.,

$$\mu a_1 = b_1 \quad \mu a_2 = b_2 - b_1^2 \quad \mu a_3 = b_3 - b_1^3 - 3b_1(b_2 - b_1^2) = b_3 - 3b_1 b_2 + 2b_1^3$$

etc. In contrast to the scaled model where we have two free parameters μ, u, we here have more 'nearly' free parameters. These equations can easily be solved but one has to find solutions $\mu > 0$ such that $\{a_i\}$ are factorial moments of a distribution. In our case where $X = CD$ this is seen to be possible for p small. With $\lambda = p/\mu$ we obtain for the first three moments A_i of Z: $A_1 = \lambda c_1$, $A_2 = \lambda(c_2 + 2c_1 - pc_1)$, $A_3 = \lambda(c_3 - O(p))$, where c_i are the corresponding moments of C. For p small A_1, A_2, A_3 is a moment sequence. For an example concerning the approximation of a binomial r.v. by compound Poisson distributed r.v.s with three coinciding moments and further three moments close to each other see Arak and Zaitsev (1988), p. 80. Arak and Zaitsev used the closeness in this case to derive the optimal bounds for the variation distance.

By Lemmas 17.1.8 and 17.1.5(c),(d), it follows that if one can match the first s moments of X_i and S_i^{coll}, then

$$\mathbf{d}_1(S^{\text{ind}}, S^{\text{coll}}) \leqslant \lambda_s(\mathbf{d}_s(S^{\text{ind}}, S^{\text{coll}}))^{1/s} \leqslant \lambda_s\left[\frac{1}{s!}\sum_{i=1}^{n}(\mathbb{E}|X_i|^s + \mathbb{E}|S_i^{\text{coll}}|^s)\right]^{1/s}. \quad (17.2.26)$$

This implies that in the case of $\mathbb{E}|X_i|^s + \mathbb{E}|S_i^{\text{coll}}|^s \leqslant C$ we have the order $n^{1/s}$ as $n \to \infty$ and, in particular, finiteness of the $\mathbf{d}_s$ distance.

CHAPTER 18

Ideal Metric with Respect to Maxima Scheme of i.i.d. Random Elements

18.1 RATE OF CONVERGENCE OF MAXIMA OF RANDOM VECTORS VIA IDEAL METRICS

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. random vectors (r.v.s) in $\mathbb{R}^m$ with d.f. F. Define the sample maxima as $M_n = (M_n^{(1)}, \dots, M_n^{(m)})$ where $M_n^{(i)} = \max_{1 \leqslant j \leqslant n} X_j^{(i)}$. For many d.f. F there exist normalizing constants $a_n^{(i)} > 0$, $b_n^{(i)} \in \mathbb{R}$ $(n \geqslant 1, 1 < i \leqslant m)$ such that

$$\left(\frac{M_n^{(1)} - b_n^{(1)}}{a_n^{(1)}}, \dots, \frac{M_n^{(m)} - b_n^{(m)}}{a_n^{(m)}}\right) \stackrel{\mathrm{d}}{\Rightarrow} Y \tag{18.1.1}$$

where Y is a r.v. with non-degenerate marginals. The d.f. H of Y is said to be a *max-extreme value d.f.* The marginals H_i of H must be one of the three extreme value types $\phi_\alpha(x) = \exp(-x^{-\alpha})$, $(x \geqslant 0, \alpha > 0)$, $\psi_\alpha(x) = \psi_\alpha(-x^{-1})$ or $\Lambda(x) = \phi_1(e^x)$. Moreover, necessary and sufficient conditions on F for convergence in (18.1.1) are known (see, e.g., Resnick 1987a and the references therein).

Throughout we will assume that the limit d.f. H of Y in (18.1.1) is *simple max-stable*, i.e., each marginal $Y^{(i)}$ has d.f. $H_i(x) = \phi_1(x) = \exp(-x^{-1})(x \geqslant 0)$. Note that if $Y_1, Y_2, \dots$ are i.i.d. copies of Y, then

$$\frac{1}{n}\left(\max_{1 \leqslant j \leqslant n} Y_j^{(i)}, \dots, \max_{1 \leqslant j \leqslant n} Y_j^{(m)}\right) \stackrel{\mathrm{d}}{=} Y.$$

In this section we are interested in the rate of convergence in (18.1.1) with respect to different 'max-ideal' metrics. In the next section we shall investigate similar rate of convergence problems, but with respect to compound metrics and their corresponding minimal metrics.

Definition 18.1.1. A probability metric μ on space $\mathfrak{X} := \mathfrak{X}(\mathbb{R}^m)$ of r.v.s is called a *max-ideal metric of order* $r > 0$ if for any r.v.s X, Y, $Z \in \mathfrak{X}$ and positive constant c the following two properties are satisfied:

(i) *Max-regularity*: $\mu(X_1 \vee Z, X_2 \vee Z) \leqslant \mu(X_1, X_2)$, where $x \vee y := (x^{(1)} \vee y^{(1)}, \dots, x^{(m)} \vee y^{(m)})$ for $x, y \in \mathbb{R}^m$, $\vee := \max$;

(ii) *Homogeneity of order r*: $\mu(cX_1, cX_2) = c^r\mu(X_1, Y_2)$.

If μ is a simple p. metric, i.e., $\mu(X_1, X_2) = \mu(\Pr_{X_1}, \Pr_{X_2})$, it is assumed that Z is independent of X and Y in (i).

The above definition is similar to Definition 14.2.1 of an ideal metric of order r w.r.t. the summation scheme. Taking into account the metric structure of the convolution metrics $\mu_{\theta,r}$ and $\nu_{\theta,r}$ (cf. (14.2.12) and (14.2.13)) we can construct in a similar way a *max-smoothing metric* $(\tilde{\mathbf{v}}_r)$ *of order r* as follows: for any r.v.s X' and X'' in $\mathfrak{X}$, and Y being a simple max-stable r.v., define.

$$\begin{aligned}\tilde{\mathbf{v}}_r(X', X'') &= \sup_{h>0} h^r \boldsymbol{\rho}(X' \vee hY, X'' \vee hY)\\ &= \sup_{h>0} h^r \sup_{x\in\mathbb{R}^m} |F_{X'}(x) - F_{X''}(x)| F_Y(x/h) \qquad (18.1.2)\end{aligned}$$

where $\boldsymbol{\rho}$ is the Kolmogorov metric in $\mathfrak{X}$

$$\boldsymbol{\rho}(X', X'') = \sup_{x\in\mathbb{R}^m} |F_{X'}(x) - F_{X''}(x)|. \qquad (18.1.3)$$

Here and in the following in this section, $X' \vee X''$ means a r.v. with d.f. $F_{X'}(x)F_{X''}(x)$.

Lemma 18.1.1. The max-smoothing metric $\mathbf{v}_r$ is max-ideal of order $r > 0$.

The proof is similar to that of Lemma 14.2.1 and thus omitted.

Another example of max-ideal metric is given by the *weighted Kolmogorov metric*

$$\boldsymbol{\rho}_r(X', X'') := \sup_{x\in\mathbb{R}^m} M^r(x)|F_{X'}(x) - F_{X''}(x)| \qquad (18.1.4)$$

where $M(x) := \min_{1\leqslant i\leqslant m} |x^{(i)}|$ for $x := (x^{(1)}, \ldots, x^{(m)})$.

Lemma 18.1.2. $\boldsymbol{\rho}_r$ is a max-ideal metric of order $r > 0$.

Proof. The max-regularity property follows easily from $|F_{X'\vee Z}(x) - F_{X''\vee Z}(x)| \leqslant |F_{X'}(x) - F_{X''}(x)|$ for any Z independent of X' and X''. The homogeneity property is also obvious. QED

Next we consider the rate of convergence in (18.1.1) with $a_n^{(i)} = 1/n$ and $b_n^{(i)} = 0$ by means of a max-ideal metric μ. In the sequence, for any X we write $\tilde{X} := n^{-1}X$.

Lemma 18.1.3. Suppose $X_1, X_2, \ldots$ are i.i.d. r.v.s, $M_n := \bigvee_{i=1}^n X_i$, Y is simple max-stable and μ_r is a max-ideal simple p. metric of order $r > 1$. Then

$$\mu_r(\tilde{M}_n, Y) \leqslant n^{1-r}\mu_r(X_1, Y). \qquad (18.1.5)$$

Proof. Take $Y_1, Y_2, \ldots$ to be i.i.d. copies of Y, $N_n := Y_1 \vee \cdots \vee Y_n$. Then

$$\begin{aligned}
&\mu_r(\tilde{M}_n, Y)\\
&\quad = \mu_r(\tilde{M}_n, \tilde{N}_n) \quad \text{(by the max-stability of } Y)\\
&\quad = n^{-r}\mu_r(M_n, N_n) \quad \text{(by the homogeneity property)}\\
&\quad \leqslant n^{-r}\sum_{i=1}^{n} \mu_r(X_i, Y_i) \quad \text{(by the triangle inequality and max-regularity of } \mu_r)\\
&\quad = n^{1-r}\mu_r(X_1, Y).
\end{aligned}$$

QED

By virtue of Lemmas 18.1.1 to 18.1.3 we have that for $r > 1$ and $n \to \infty$

$$\tilde{\mathbf{v}}_r(X_1, Y) < \infty \Rightarrow \tilde{\mathbf{v}}_r(\tilde{M}_n, Y) \leqslant n^{1-r}\tilde{\mathbf{v}}_r(X_1, Y) \to 0 \tag{18.1.6}$$

as well as

$$\boldsymbol{\rho}_r(X_1, Y) < \infty \Rightarrow \boldsymbol{\rho}_r(\tilde{M}_n, Y) \leqslant n^{1-r}\boldsymbol{\rho}_r(X_1, Y) \to 0. \tag{18.1.7}$$

The last two implications indicate that the right order of the rate of the uniform convergence $\boldsymbol{\rho}(\tilde{M}_n, Y) \to 0$ should be $O(n^{1-r})$ provided that $\mathbf{v}_r(X_1, Y) < \infty$ or $\boldsymbol{\rho}_r(X_1, Y) < \infty$. The next theorem gives the proof of this statement for $1 < r \leqslant 2$.

Theorem 18.1.1. Let $r > 1$.

(a) If

$$\tilde{\mathbf{v}}_r(X_1, Y) < \infty \tag{18.1.8}$$

then

$$\boldsymbol{\rho}(\tilde{M}_n, Y) \leqslant A(r)[\tilde{v}_r n^{1-r} + \kappa n^{-1}] \to 0 \quad \text{as } n \to \infty. \tag{18.1.9}$$

In (18.1.9) the absolute constant $A(r)$ is given by

$$A(r) := 2[c_1(4^r + 2^r) \vee c_1c_2 4(3/2)^{\tilde{r}} \vee c_2(4c_1 4^{\tilde{r}}/(r-1))^{1/(r-1)}] \tag{18.1.10}$$

where $c_1 := 1 + 4e^{-2}m$, $c_2 := mc_1$, $\tilde{r} := 1 \vee (r-1)$, and $\tilde{v}_r$, κ are the following measures of deviation of F_{X_1} from F_Y

$$\kappa := \max(\rho, \tilde{v}_1, \tilde{v}_r^{r/(r-1)}) \quad \rho := \boldsymbol{\rho}(X_1, Y) \quad \tilde{v}_1 := \tilde{\mathbf{v}}_1(X_1, Y) \quad \tilde{v}_r := \tilde{\mathbf{v}}_r(X_1, Y). \tag{18.1.11}$$

(b) If $\boldsymbol{\rho}_r(X_1, Y) < \infty$ then

$$\boldsymbol{\rho}(\tilde{M}_n, Y) \leqslant B(r)[\rho_r n^{1-r} + \tau n^{-1}] \to 0 \quad \text{as } n \to \infty \tag{18.1.12}$$

where

$$B(r) := (1 \vee K_1 \vee K_r \vee K_r^{1/(r-1)})A(r) \quad K_r := (r/e)^r \tag{18.1.13}$$

and

$$\tau := \max(\rho, \rho_1, \rho_r^{1/(r-1)}) \quad \rho := \boldsymbol{\rho}(X_1, Y) \quad \rho_1 := \boldsymbol{\rho}_1(X_1, Y) \quad \rho_r := \boldsymbol{\rho}_r(X_1, Y). \tag{18.1.14}$$

Remark 18.1.1. Since the simple max-stable r.v. Y is concentrated on $\mathbb{R}_+^m$ then

$$\boldsymbol{\rho}(\tilde{M}_n, Y) = \boldsymbol{\rho}\left(\left(\bigvee_{i=1}^n \tilde{X}_i\right)_+, Y\right) = \boldsymbol{\rho}\left(n^{-1}\bigvee_{i=1}^n (X_i)_+, Y\right) \tag{18.1.15}$$

where $(x)_+ := ((x^{(1)})_+, \ldots, (x^{(m)})_+)$, $(x^{(i)})_+ := 0 \vee x^{(i)}$. Therefore, without loss of generality we may consider X_is being non-negative r.v.s. Thus further, we assume that all r.v.s X under consideration are non-negative.

Similar to the proof of Theorem 15.2.1 the proof of the above theorem is based on relationships between the max-ideal metrics $\mathbf{v}_r$ and $\boldsymbol{\rho}_r$ and the uniform metric $\boldsymbol{\rho}$. These relationships have the form of max-smoothing type inequalities, see further Lemmas 18.1.4 to 18.1.7. Recall that in our notations $X' \vee X''$ means maximum of independent copies of X' and X''. The first lemma is an analog of Lemma 15.2.1 concerning the smoothing property of stable random motion.

Lemma 18.1.4 (Max-smoothing inequality). For any $\delta > 0$

$$\boldsymbol{\rho}(X, Y) \leqslant c_1\boldsymbol{\rho}(X \vee \delta Y, Y \vee \delta Y) + c_2\delta \tag{18.1.16}$$

where

$$c_1 = 1 + 4e^{-2}m \quad c_2 = mc_1. \tag{18.1.17}$$

Proof. Let $\mathbf{L}(X'X'')$ be the Lévy metric

$$\mathbf{L}(X', X'') = \inf\{\varepsilon > 0: F_{X'}(x - \varepsilon\mathbf{e}) - \varepsilon \leqslant F_{X''}(x) \leqslant F_{X'}(x + \varepsilon\mathbf{e}) + \varepsilon\} \tag{18.1.18}$$

in $\mathfrak{X}^m = \mathfrak{X}(\mathbb{R}_+^m)$, $\mathbf{e} = (1, 1, \ldots, 1) \in \mathbb{R}^m$, see Example 4.1.3.

Claim 1. If c_1 is given by (18.1.17), then

$$\boldsymbol{\rho}(X, Y) \leqslant (1 + c_1)\mathbf{L}(X, Y). \tag{18.1.19}$$

Since $F_{Y^{(i)}}(t) = \exp(-1/t)$, $t > 0$, it is easy to see that

$$\boldsymbol{\rho}(X, Y) \leqslant \left(1 + \sum_{j=1}^m \sup_{t>0}\left(\frac{\mathrm{d}}{\mathrm{d}t} F_{Y^{(j)}}(t)\right)\right)\mathbf{L}(X, Y) = (1 + 4e^{-2}m)\mathbf{L}(X, Y)$$

which proves (18.1.19).

Claim 2. For any $X \in \mathfrak{X}^m$ and a simple max-stable r.v. Y,

$$\mathbf{L}(X, Y) \leqslant \boldsymbol{\rho}(X \vee \delta Y, Y \vee \delta Y) + \delta m \quad \delta > 0. \tag{18.1.20}$$

Proof of claim 2. Let $\mathbf{L}(X, Y) > \gamma$. Then there exists $x_0 \in \mathbb{R}_+^m$ (i.e., $x_0 \geqslant \bar{0}$, i.e., $x_0^{(i)} \geqslant 0$, $i = 1, \ldots, m$) such that

$$|F_X(x) - F_Y(x)| > \gamma \quad \text{for any } x_0 \leqslant x \leqslant x_0 + \gamma\mathbf{e}. \tag{18.1.21}$$

By (18.1.21) and the Hoeffding–Fréchet inequality

$$F_Y(x) \geqslant \max(0, \textstyle\sum_{j=1}^m F_{Y^{(j)}}(x) - m + 1)$$

we have that

$$\begin{aligned}
&|F_X(x_0+\gamma\mathbf{e})-F_Y(x_0+\gamma\mathbf{e})|F_{\delta Y}(x_0+\gamma\mathbf{e})\\
&\quad\geq \gamma F_{\delta Y}(\gamma\mathbf{e})=\gamma F_Y\left(\frac{\gamma}{\delta}\,\mathbf{e}\right)\geq \gamma\left(\sum_{j=1}^{m}F_{Y^{(j)}}(\gamma/\delta)-m+1\right)\\
&\quad=\gamma\left(\sum_{j=1}^{m}\exp(-\delta/\gamma)-m+1\right)\geq \gamma(m(1-\delta/\gamma)-m+1)=\gamma-m\delta.
\end{aligned}$$

Therefore, $\rho(X\vee\delta Y, Y\vee\delta Y)\geq\gamma-\delta m$. Letting $\gamma\to\mathbf{L}(X,Y)$ we obtain (18.1.20).

Now (18.1.16) is a consequence of Claims 1 and 2. QED

The next lemma is an analog of Lemmas 15.2.2 and 14.3.1.

Lemma 18.1.5. For any $X', X''\in\mathfrak{X}^m$

$$\boldsymbol{\rho}(X'\vee\delta Y, X''\vee\delta Y)\leq\delta^{-r}\tilde{\mathbf{v}}_r(X',X'') \tag{18.1.22}$$

and

$$\boldsymbol{\rho}(X'\vee\delta Y, X''\vee\delta Y)\leq K_r\delta^{-r}\boldsymbol{\rho}_r(X',X'') \tag{18.1.23}$$

where

$$K_r := (r/e)^r. \tag{18.1.24}$$

Proof of Lemma 18.1.5. The inequality (18.1.22) follows immediately from the definition of $\tilde{\mathbf{v}}_r$ (see (18.1.2)). Using the Hoeffding–Fréchet inequality

$$F_Y(x)\leq\min_{1\leq i\leq m}F_{Y^{(i)}}(x^{(i)})=\min_{1\leq i\leq m}\exp\{-1/x^{(i)}\} \tag{18.1.25}$$

we have

$$\begin{aligned}
\rho(X'\vee\delta Y, X''\vee\delta Y)&=\sup_{x\in\mathbb{R}^m}F_{\delta Y}(x)|F_{X'}(x)-F_{X''}(x)|\\
&\leq\sup_{x\in\mathbb{R}^m}\min_{1\leq i\leq m}\exp(-\delta/x^{(i)})|F_{X'}(x)-F_{X''}(x)|\\
&=\sup_{x\in\mathbb{R}^m}\min_{1\leq i\leq m}\left[\left(\frac{\delta}{x^{(i)}}\right)^r\exp\left(-\frac{\delta}{x^{(i)}}\right)\right]\\
&\qquad\times\left(\frac{\delta}{x^{(i)}}\right)^{-r}|F_{X'}(x)-F_{X''}(x)|\\
&\leq K_r\sup_{x\in\mathbb{R}^m}\min_{1\leq i\leq m}\left(\frac{\delta}{x^{(i)}}\right)^{-r}|F_{X'}(x)-F_{X''}(x)|\\
&=K_r\delta^{-r}\boldsymbol{\rho}_r(X',X'')
\end{aligned}$$

which proves (18.1.23). QED

Lemma 18.1.6. For any X' and X''

$$\tilde{\mathbf{v}}_r(X', X'') \leqslant K_r \boldsymbol{\rho}_r(X', X'') \tag{18.1.26}$$

Proof. Apply (18.1.23) and (18.1.2) to get the above inequality. QED

Lemma 18.1.7. For any independent r.v.s $X', X'', Z, W \in \mathfrak{X}^m$

$$\boldsymbol{\rho}(X' \vee Z, X'' \vee Z) \leqslant \boldsymbol{\rho}(Z, W)\rho(X', X'') + \boldsymbol{\rho}(X' \vee W, X'' \vee W). \tag{18.1.27}$$

Proof. For any $x \in \mathbb{R}^m$

$$\begin{aligned} &F_Z(x)|F_{X'}(x) - F_{X''}(x)| \\ &\quad \leqslant |F_Z(x) - F_W(x)||F_{X'}(x) - F_{X''}(x)| + F_W(x)|F_{X'}(x) - F_{X''}(x)| \end{aligned}$$

which proves (18.1.27). QED

The last lemma resembles Lemmas 14.3.2 and 14.3.4 dealing with 'smoothing for sums of i.i.d.' Now we are ready for the proof of the theorem.

Proof of Theorem 18.1.1. (a) Let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. copies of Y, $N_n := \bigvee_{i=1}^n Y_i$. Hence

$$\boldsymbol{\rho}(\tilde{M}_n, Y) = \boldsymbol{\rho}(\tilde{M}_n, \tilde{N}_n). \tag{18.1.28}$$

By the smoothing inequality (18.1.16),

$$\boldsymbol{\rho}(\tilde{M}_n, \tilde{N}_n) \leqslant c_1 \rho(\tilde{M}_n \vee \delta Y, \tilde{N}_n \vee \delta Y) + c_2 \delta. \tag{18.1.29}$$

Consider the right-hand side of (18.1.29) and obtain for $n \geqslant 2$

$$\begin{aligned} &\boldsymbol{\rho}(\tilde{M}_n \vee \delta Y, \tilde{N}_n \vee \delta Y) \\ &\quad \leqslant \sum_{j=1}^{m+1} \boldsymbol{\rho}\left(\bigvee_{i=1}^{j-1} \tilde{Y}_i \vee \bigvee_{i=j}^{n} \tilde{X}_i \vee \delta Y, \bigvee_{i=1}^{j} \tilde{Y}_i \vee \bigvee_{i=j+1}^{n} \tilde{X}_i \vee \delta Y\right) \\ &\quad\quad + \boldsymbol{\rho}\left(\bigvee_{j=1}^{m+1} \tilde{Y}_j \vee \bigvee_{j=m+2}^{n} \tilde{X}_j \vee \delta Y, \bigvee_{j=1}^{m+1} \tilde{Y}_j \vee \bigvee_{j=m+2}^{n} \tilde{Y}_j \vee \delta Y\right) \end{aligned} \tag{18.1.30}$$

where m is the integer part of $n/2$ and $\bigvee_{j=1}^{0} := 0$. By Lemma 18.1.7 we can estimate each term in the right-hand side of (18.1.30) as follows

$$\begin{aligned} &\boldsymbol{\rho}\left(\bigvee_{i=1}^{j-1} \tilde{Y}_i \vee \bigvee_{i=j}^{n} \tilde{X}_i \vee \delta Y, \bigvee_{i=1}^{j} \tilde{Y}_i \vee \bigvee_{i=j+1}^{n} \tilde{X}_i \vee \delta Y\right) \\ &\quad \leqslant \boldsymbol{\rho}\left(\bigvee_{i=j+1}^{n} \tilde{X}_i, \bigvee_{i=j+1}^{n} \tilde{Y}_i\right) \boldsymbol{\rho}\left(\bigvee_{i=1}^{j-1} \tilde{Y}_i \vee \tilde{X}_j \vee \delta Y, \bigvee_{i=1}^{j} \tilde{Y}_i \vee \delta Y\right) \\ &\quad\quad + \boldsymbol{\rho}\left(\bigvee_{i=1}^{j-1} \tilde{Y}_i \vee \tilde{X}_j \vee \delta Y \vee \bigvee_{i=j+1}^{n} \tilde{Y}_i, \bigvee_{i=1}^{j} \tilde{Y}_i \vee \delta Y \vee \bigvee_{i=j+1}^{n} \tilde{Y}_i\right). \end{aligned} \tag{18.1.31}$$

Combining (18.1.28) to (18.1.31) and using Lemma 18.1.7 again we have

$$\rho\left(\bigvee_{j=1}^{n} \tilde{X}_j, Y\right) \leqslant c_1(I_1 + I_2 + I_3 + I_n) + c_2\delta \tag{18.1.32}$$

where

$$I_1 := \rho\left(\bigvee_{i=2}^{n} \tilde{X}_i, \bigvee_{i=2}^{n} \tilde{Y}_i\right)\rho(\tilde{X}_1 \vee \delta Y, \tilde{Y}_1 \vee \delta Y)$$

$$I_2 := \sum_{j=2}^{m+1} \rho\left(\bigvee_{i=j+1}^{n} \tilde{X}_i, \bigvee_{i=j+1}^{n} \tilde{Y}_i\right)\rho\left(\bigvee_{i=1}^{j-1} \tilde{Y}_i \vee \tilde{X}_j \vee \delta Y, \bigvee_{i=1}^{j} \tilde{Y}_i \vee \delta Y\right)$$

$$I_3 := \sum_{j=1}^{m+1} \rho\left(\tilde{X}_j \vee \bigvee_{i=m+2}^{n} \tilde{Y}_i, \tilde{Y}_j \vee \bigvee_{i=m+2}^{n} \tilde{Y}_i\right)$$

and

$$I_4 := \rho\left(\bigvee_{j=1}^{m+1} \tilde{Y}_j \vee \bigvee_{j=m+2}^{n} \tilde{X}_j, \bigvee_{j=1}^{m+1} \tilde{Y}_j \vee \bigvee_{j=m+2}^{n} \tilde{Y}_j\right).$$

Take $n \geqslant 3$. We estimate I_3 by making use of Lemmas 18.1.1 and 18.1.5,

$$\begin{aligned} I_3 &\leqslant \sum_{j=1}^{m+1} \tilde{\mathbf{v}}_r(\tilde{X}_j, \tilde{Y}_j)\left(\frac{n-m-1}{n}\right)^{-r} \leqslant \sum_{j=1}^{m+1} \tilde{\mathbf{v}}_r(\tilde{X}_j, \tilde{Y}_j)4^r \\ &\leqslant (m+1)n^{-r}\tilde{\mathbf{v}}_r(X_1, Y_1)4^r \leqslant 4^r n^{1-r} v_r. \end{aligned} \tag{18.1.33}$$

In the same way we estimate I_4

$$\begin{aligned} I_4 &\leqslant \tilde{\mathbf{v}}_r\left(\bigvee_{j=m+2}^{n} \tilde{X}_j, \bigvee_{j=m+2}^{n} \tilde{Y}_j\right)\left(\frac{m+1}{n}\right)^{-r} \\ &\leqslant 2^r(n-m)n^{-r}\tilde{\mathbf{v}}_r(X_1, Y_1) \leqslant 2^r n^{1-r}\tilde{\mathbf{v}}_r. \end{aligned} \tag{18.1.34}$$

Set

$$\delta := A \max(\tilde{v}_r, \tilde{v}_r^{1/(r-1)})n^{-1} \tag{18.1.35}$$

where $A > 0$ will be chosen later. Suppose that for all $k < n$

$$\rho\left(k^{-1}\bigvee_{j=1}^{k} X_j, k^{-1}\bigvee_{j=1}^{k} Y_j\right) \leqslant A(r)(\tilde{v}_r k^{1-r} + \kappa k^{-1}) \tag{18.1.36}$$

where $\tilde{v}_r = \tilde{\mathbf{v}}_r(X_1, Y)$, $\kappa = \kappa(X_1, Y) = \max(\rho, \tilde{v}_1, \tilde{v}_r^{1/(r-1)})$ (cf. (18.1.11)). Here $A(r)$ is an absolute constant to be determined later. For $k = 1$, (18.1.36) holds with $A(r) \geqslant 1$. For $k = 2$, $\rho(2^{-1}\bigvee_{j=1}^{2} X_j, 2^{-1}\bigvee_{j=1}^{2} Y_j) \leqslant 2\rho(X_1, Y_2)$, i.e., (18.1.36) is valid with $A(r) \geqslant 4 \vee 2^r$.

Let us estimate I_1 in (18.1.32). By (18.1.22), (18.1.35) and (18.1.36),

$$I_1 \leqslant A(r)(\tilde{\nu}_r(n-1)^{1-r} + \kappa(n-1)^{-1})\tilde{\mathbf{v}}_1(n^{-1}X_1, n^{-1}Y_1)\frac{1}{A\tilde{\nu}_1 n^{-1}}$$

$$\leqslant (\tfrac{3}{2})^{(r-1)\vee 1}\frac{A(r)}{A}(\tilde{\nu}_r n^{1-r} + \kappa n^{-1}).$$

Similarly, we estimate I_2

$$I_2 = \sum_{j=2}^{m+1}\rho\left((n-j)^{-1}\bigvee_{j=1}^{n-j}X_i, (n-j)^{-1}\bigvee_{i=1}^{n-j}Y_i\right)\rho\left(\left(\frac{j-1}{n}+\delta\right)Y \vee \tilde{X}_j, \left(\frac{j-1}{n}+\delta\right)Y \vee \tilde{Y}_j\right)$$

$$\leqslant \sum_{j=2}^{m+1}A(r)(\tilde{\nu}_r(n-j)^{1-r} + \kappa(n-j)^{-1})\frac{\tilde{\mathbf{v}}_r(\tilde{X}_j, \tilde{Y}_j)}{\left(\dfrac{j-1}{n}+\delta\right)^r}$$

$$\leqslant \sum_{j=2}^{m+1}A(r)(\tilde{\mathbf{v}}_r(n-m-1)^{1-r} + \kappa(n-m-1)^{-1})\frac{n^{-r}\tilde{\mathbf{v}}_r(X_1, Y)}{n^{-r}(j-1+\delta n)^r}$$

$$\leqslant A(r)(4^{r-1}\tilde{\nu}_r n^{1-r} + 4\kappa n^{-1})\sum_{j=2}^{\infty}\frac{\nu_r}{(j-1+A\nu_r^{1/(r-1)})^r}$$

$$\leqslant 4^{(r-1)\vee 1}\frac{1}{r-1}\frac{A(r)}{A^{r-1}}(\tilde{\nu}_r n^{1-r} + \kappa n^{-1}).$$

Now we can use the above estimates for I_1, I_2, combine them with (18.1.33) to (18.1.35) and (18.1.32) to get

$$\rho\left(\bigvee_{j=1}^{n}\tilde{X}_j, Y\right) \leqslant \left(c_1(\tfrac{3}{2})^{\tilde{r}}(1/A) + c_1 4^{\tilde{r}}\frac{1}{r-1}\frac{1}{A^{r-1}}\right)A(r)(\tilde{\nu}_r n^{1-r} + \kappa n^{-2})$$
$$+ c_2(4^r + 2^r)\tilde{\nu}_r n^{1-r} + c_2 A\kappa n^{-1} \quad \tilde{r} := \max(1, r-1).$$

Now choose $A = \max\left(4c_1(3/2)^{\tilde{r}}, \left(4c_1 4^{\tilde{r}}\dfrac{1}{r-1}\right)^{1/(r-1)}\right)$. Then

$$\rho\left(\bigvee_{j=1}^{n}\tilde{X}_j, Y\right) \leqslant \tfrac{1}{2}A(r)(\tilde{\nu}_r n^{1-r} + \kappa n^{-1})$$
$$+ (c_1(4^r + 2^r) \vee c_2 A)(\tilde{\nu}_r n^{1-r} + \kappa n^{-2}).$$

Finally, letting $\frac{1}{2}A(r) := c_1(4^r + 2^r) \vee c_2 A$ completes the proof. QED

(b) By (a) and Lemma 18.1.6,

$$\boldsymbol{\rho}\left(\bigvee_{j=1}^{n}\tilde{X}_j, Y\right) \leqslant A(r)[K_r\rho_r n^{1-r} + \max(\rho, K_1\rho_r, K_r^{1/(r-1)}\rho_r)n^{-1}]$$

$$\leqslant (1 \vee K_1 \vee K_r \vee K_r^{(1/(r-1))})A(r)[\rho_r n^{1-r} + \tau n^{-1}]. \quad \text{QED}$$

Further, we shall prove that the order $O(n^{1-r})$ of the rate of convergence in (18.1.9) and (18.1.12) is precise for any $r > 1$ under the conditions $\tilde{\nu}_r < \infty$ or $\boldsymbol{\rho}_r < \infty$. Moreover, we shall investigate more general 'tail' conditions than $\rho_r = \boldsymbol{\rho}_r(X_1, Y) < \infty$.

Let $\psi: [0, \infty) \to [0, \infty)$ denote a continuous function, increasing to ∞ as $x \to \infty$. Let us consider the metrics $\boldsymbol{\rho}_\psi$ and $\boldsymbol{\mu}_\psi$ defined by

$$\boldsymbol{\rho}_\psi(X', X'') := \sup_{x \in \mathbb{R}_+^m} \psi(M(x))|F_{X'}(x) - F_{X''}(x)| \quad X', X'' \in \mathfrak{X}^m$$

(cf. the definition of $\boldsymbol{\rho}_r$, see (18.1.4)) and

$$\boldsymbol{\mu}_\psi(X', X'') := \sup_{x \in \mathbb{R}_+^m} \psi(M(x))|\log F_{X'}(x) - \log F_{X''}(x)|$$

and recall that $M(x) := \min\{x^{(i)}: i = 1, \ldots, m\}$, $x \in \mathbb{R}_+^m$.

We shall investigate the rate of convergence in $\tilde{M}_n \overset{\mathrm{d}}{\Rightarrow} Y$, assuming that either $\boldsymbol{\rho}_\psi(X_1, Y) < \infty$ or $\boldsymbol{\mu}_\psi(X_1, Y) < \infty$. Obviously, $\boldsymbol{\rho}_\psi(X_1, Y) < \infty$ implies that for each i, $\boldsymbol{\rho}_\psi(X_1^{(i)}, Y^{(i)}) := \sup\{\psi(x)|F_i(x) - \phi_1(x)|: x \in \mathbb{R}_+\} \leqslant \boldsymbol{\rho}_\psi(X_1, Y) < \infty$, where F_i is the d.f. of $X_1^{(i)}$. We also define

$$\tilde{\rho}_\psi := \max\{\boldsymbol{\rho}_\psi(X_1^{(i)}, Y^{(i)}): i = 1, \ldots, m\}$$

and whenever $\boldsymbol{\mu}_\psi(X_1^{(i)}, Y^{(i)}) := \sup\{\psi(x)|\log F_i(x) - \log \phi_1(x)|: x \in \mathbb{R}_+\} < \infty$ we also define

$$\tilde{\mu}_\psi := \max\{\boldsymbol{\mu}_\psi(X_1^{(i)}, Y^{(i)}): i = 1, \ldots, k\}.$$

In the proofs of the results below we shall often use the following inequalities. Since $H(x) := \Pr(Y \leqslant x) \leqslant H_i(x^{(i)}) := \Pr(Y^{(i)} \leqslant x^{(i)}) = \phi_1(x^{(i)})$ for each i, we have

$$H(x) \leqslant \phi_1(M(x)). \tag{18.1.37}$$

For $a, b > 0$ we have

$$n|a - b| \min(a^{n-1}, b^{n-1}) \leqslant |a^n - b^n| \leqslant n|a - b| \max(a^{n-1}, b^{n-1}) \tag{18.1.38}$$

and

$$\min(a, b)\left|\log \frac{a}{b}\right| \leqslant |a - b| \leqslant \max(a, b)\left|\log \frac{a}{b}\right|. \tag{18.1.39}$$

Theorem 18.1.2. Assume that

$$g(a) := \sup_{x \geqslant 0} \frac{\phi_1(xa)}{\psi(x)}$$

is finite for all $a \geqslant 0$. For $n \geqslant 2$ define $R(n) := ng(1/(n-1))$.

(i) If $\rho_\psi := \boldsymbol{\rho}_\psi(X_1, Y) < \infty$ and $\tilde{\mu}_\psi < \infty$, then for all $n \geqslant 2$, $\boldsymbol{\rho}(\tilde{M}_n, Y) \leqslant R(n)\tau$, where $\tau := \max(\rho_\psi \exp \tilde{\mu}_\psi, 1 + \exp \tilde{\mu}_\psi)$.

(ii) If $\rho_\psi < \infty$, then

$$\limsup_{n \to \infty} \frac{1}{R(n)} \boldsymbol{\rho}(\tilde{M}_n, Y) \leqslant \bar{\tau}$$

where $\bar{\tau} := \max(\rho_\psi \exp \tilde{\rho}_\psi, 1 + \exp \tilde{\rho}_\psi)$.

(iii) If $\rho_\psi < \infty$ and if there exists a sequence δ_n of positive numbers such that

$$\lim_{n \to \infty} \frac{n}{\psi(\delta_n)} = \lim_{n \to \infty} \frac{1}{R(n)} \phi_1\left(\frac{\delta_n}{n-1}\right) = 0$$

then in (ii) $\bar{\tau}$ may be replaced by ρ_ψ.

Remark 18.1.2. (a) In Theorem 18.1.2 we normalize the partial maxima M_n by n; in Theorem 18.1.4 below we prove a result in which other normalizations are allowed.

(b) If Y is not simple max-stable but has marginals $H_i = \phi_{\alpha_i}(x)$ $(\alpha_i > 0)$, then using simple monotone transformations Theorem 18.1.2 can be used to estimate

$$\boldsymbol{\rho}((M_n^{(1)} n^{-1/\alpha_1}, \ldots, M_n^{(k)} n^{-1/\alpha_k}), Y) \quad \text{where } M_n^{(i)} := \bigvee_{j=1}^{n} X_j^{(i)}.$$

Proof of Theorem 18.1.2. Using (18.1.38) and $H(x) = H^n(nx)$ we have

$$I := |F^n(nx) - H(x)| \leqslant n|F(nx) - H(nx)| \max(F^{n-1}(nx), H^{n-1}(nx))$$

where F is the d.f. of X_i. Let us consider $I_1 := n|F(nx) - H(nx)|H^{n-1}(nx)$. Using (18.1.37) we have

$$H^{n-1}(nx) \leqslant \phi_1\left(\frac{nM(x)}{n-1}\right).$$

Hence $I_1 \leqslant ng(1/(n-1))\psi(nM(x))|F(nx) - H(nx)|$ and we obtain

$$I_1 \leqslant R(n)\rho_\psi. \tag{18.1.40}$$

Next consider $I_2 := n|F(nx) - H(nx)|F^{n-1}(nx)$ and let δ_n denote a sequence of positive numbers to be determined later. Observe that for each i and $u \geqslant \delta_n$

we have

$$|\log F_i(u) - \log \phi_1(u)| \leqslant \frac{1}{\psi(\delta_n)} \sup_{u \geqslant \delta_n} \psi(u) \left| \log \frac{F_i(u)}{\phi_1(u)} \right| =: \frac{1}{\psi(\delta_n)} \mu_n^{(i)}$$

so that

$$F_i(u) \leqslant \phi_1(u) \exp \frac{1}{\psi(\delta_n)} \mu_n^{(i)}.$$

If $nx_i \geqslant \delta_n$ for each i we obtain

$$F^{n-1}(nx) \leqslant F_i^{n-1}(nx_i) \leqslant \phi_1^{n-1}(nx_i) \exp \frac{n-1}{\psi(\delta_n)} \mu_n^{(i)}. \tag{18.1.41}$$

This implies that

$$I_2 \leqslant R(n)\psi(nx_i)|F(nx) - H(nx)| \exp \frac{n-1}{\psi(\delta_n)} \mu_n^{(i)}.$$

Choosing i such that $x_i = M(x)$ it follows that

$$I_2 \leqslant R(n)\rho_\psi \exp \frac{n-1}{\psi(\delta_n)} \mu_n^{(i)}. \tag{18.1.42}$$

On the other hand, if $nx_i \leqslant \delta_n$ for some index i we have $I \leqslant F_i^n(\delta_n) + \phi_1^n(\delta_n)$. Using (18.1.41) with $nx_i = \delta_n$ it follows that

$$I \leqslant \phi_1^{n-1}(\delta_n)\left(1 + \exp \frac{n-1}{\psi(\delta_n)} \mu_n^{(i)}\right). \tag{18.1.43}$$

Using $\phi_1^{n-1}(\delta_n) = \phi_1\left(\dfrac{\delta_n}{n-1}\right) \leqslant \psi(\delta_n)g(1/(n-1))$ we obtain

$$I \leqslant \psi(\delta_n)g(1/(n-1))\left(1 + \exp \frac{n-1}{\psi(\delta_n)} \mu_n^{(i)}\right). \tag{18.1.44}$$

Proof of (i). Choose δ_n such that $n - 1 \leqslant \psi(\delta_n) \leqslant n$; since $\mu_n^{(i)} \leqslant \tilde{\mu}_\psi$, the inequalities (18.1.40), (18.1.42), and (18.1.44) yield

$$I \leqslant \begin{cases} R(n)\rho_\psi \exp \tilde{\mu}_\psi & \text{if } nM(x) \geqslant \delta_n \\ R(n)(1 + \exp \tilde{\mu}_\psi) & \text{if } nM(x) < \delta_n. \end{cases}$$

This proves (i).

Proof of (ii). Again choose δ_n such that $n - 1 \leqslant \psi(\delta_n) \leqslant n$. Using (18.1.39) we obtain

$$\limsup_{n \to \infty} \mu_n^{(i)} \leqslant \rho_\psi(X_1^{(i)}, Y^{(i)}) \leqslant \tilde{\rho}_\psi. \tag{18.1.45}$$

Combining (18.1.40), (18.1.42), (18.1.44) and (18.1.45) we obtain the proof of (ii).

Proof of (iii). Using the sequence δ_n satisfying the assumption of the theorem, it follows from (18.1.40), (18.1.42), (18.1.43) and (18.1.45) that

$$\limsup_{n\to\infty} \frac{1}{R(n)}\, \rho(\tilde{M}_n, Y) \leqslant \rho_\psi$$

which completes the proof. QED

Suppose now that ψ is *regularly varying* with index $r \geqslant 1$ (see Resnick, 1987a), i.e., $\psi(x) \sim x^r L(x)$ as $x \to \infty$ and $L(x)$ varying slowly, $\psi \in RV_r$. We may assume that ψ' is positive and $\psi' \in RV_{r-1}$. In this case $g(a) = \phi_1(\bar{x}a)/\psi(\bar{x})$ where $\bar{x}$ is a solution of the equation $x^2\psi'(x)/\psi(x) = 1/a$. It follows that $\bar{x}a \to 1/r$ as $a \to 0$ and hence that $g(a) \sim K(r)\, 1/\psi(1/a)$ $(a \to 0)$, where $K(r) = (r/e)^r$. In particular, if $\psi(t) = t^r$, then $\boldsymbol{\rho}_\psi = \boldsymbol{\rho}_r$ (see (18.1.4)) and both Theorem 18.2.1 (for $1 < r \leqslant 2$) and Theorem 18.2.2 (for any $r > 1$) state that $\boldsymbol{\rho}_r(X_1, Y) < \infty$ implies $\boldsymbol{\rho}(\tilde{M}_n, Y) = O(n^{1-r})$. Moreover, in Theorem 18.2.1 we obtain an estimate for $\boldsymbol{\rho}(\tilde{M}_n, Y)$ (cf. (18.1.12)), which is *uniform on* $n = 1, 2, \ldots$. The next theorem shows that the condition $\boldsymbol{\rho}_r(X_1, Y) < \infty$ is necessary for having rate $O(n^{1-r})$ in the uniform convergence $\boldsymbol{\rho}(\tilde{M}_n, Y) \to 0$ as $n \to \infty$.

Theorem 18.1.3. Assume that $\psi \in RV_r$, $r \geqslant 1$ and that $\lim_{x\to\infty} \psi(x)/x = \infty$. Let Y denote a r.v. with a simple max-stable d.f. H, and $X_1, X_2, \ldots$ be i.i.d. with common d.f. F. Then

(i) $\boldsymbol{\rho}_\psi(X_1, Y) < \infty$ holds if and only if $\limsup_{n\to\infty} (\psi(n)/n)\boldsymbol{\rho}(\tilde{M}_n, Y) < \infty$ and

(ii) if $r > 1$, then $\limsup_{M(x)\to\infty} \psi(M(x))|F(x) - H(x)| = 0$ if and only if

$$\lim_{n\to\infty} \frac{\psi(n)}{n}\, \boldsymbol{\rho}(\tilde{M}_n, Y) = 0.$$

Proof. (i) If $\rho_\psi < \infty$ the result is a consequence of Theorem 18.1.2. To prove the 'if' part, use the inequality (18.1.38) to obtain

$$n|F(x) - H(x)| \min(F^{n-1}(x), H^{n-1}(x)) \leqslant \boldsymbol{\rho}(\tilde{M}_n, Y).$$

Now if $M(x) \to \infty$, choose n such that $n \leqslant M(x) < n + 1$; then $F^{n-1}(x) \geqslant F^{n-1}(n\mathbf{e})$ and $H^{n-1}(x) \geqslant H^{n-1}(n\mathbf{e})$ and it follows that

$$\psi(M(x))|F(x) - H(x)| \leqslant \frac{\psi(n+1)}{\psi(n)} \frac{(\psi(n)/n)\rho(\tilde{M}_n, Y)}{\min(F^{n-1}(n\mathbf{e}), H^{n-1}(n\mathbf{e}))}.$$

Since $\psi(n+1) \sim \psi(n)$ $(n \to \infty)$ it follows that

$$\limsup_{M(x)\to\infty} \psi(M(x))|F(x) - H(x)| < \infty$$

and consequently that $\boldsymbol{\rho}_\psi(X_1, Y) < \infty$.

(ii) If

$$\lim_{n\to\infty} \frac{\psi(n)}{n} \boldsymbol{\rho}(\tilde{M}_n, Y) = 0$$

it follows as in the proof of (i) that $\limsup_{M(x)\to\infty} \psi(M(x))|F(x) - H(x)| = 0$. To prove the 'only if' part choose A such that $\psi(M(x))|F(x) - H(x)| \leqslant \varepsilon$, $M(x) \geqslant A$.

Now we proceed as in the proof of Theorem 18.1.2: if $M(nx) \geqslant \delta_n > A$ we have

$$I_1 \leqslant \varepsilon \boldsymbol{\rho}_\psi R(n) \quad I_2 \leqslant \varepsilon \boldsymbol{\rho}_\psi R(n) \exp \frac{n-1}{\psi(\delta_n)} \mu_n^{(i)}.$$

If $M(nx) \leqslant \delta_n$, (18.1.43) remains valid. If we choose δ_n such that $\psi(\delta_n) = n^s$ with $1 < s < r$, it follows that

$$\lim_{n\to\infty} \frac{\psi(n)}{n} \phi_1\left(\frac{\delta_n}{n-1}\right) = \lim_{n\to\infty} \frac{n}{\psi(\delta_n)} = 0$$

and hence that

$$\limsup_{n\to\infty} \frac{1}{R(n)} \boldsymbol{\rho}(\tilde{M}_n, Y) \leqslant \varepsilon \boldsymbol{\rho}_\psi.$$

Now let $\varepsilon \downarrow 0$ to obtain the proof of (ii). QED

Remark 18.1.3. (a) In a similar way one can prove that $\rho_\psi < \infty$ holds if and only if for each marginal,

$$\limsup_{n\to\infty} \frac{\psi(n)}{n} \boldsymbol{\rho}\left(\frac{M_n^{(i)}}{n}, Y^{(i)}\right) < \infty.$$

(b) If $\psi(0) = 0$, if ψ is *0-regularly varying* ($\psi \in ORV$, i.e., for any $x > 0$, $\limsup_{t\to\infty} \psi(xt)/\psi(t) < \infty$) and if $\limsup_{x\to\infty} (\psi(x)/x) = \infty$, Theorem 18.1.3 (i) remains valid. To prove this assertion we only prove that $\limsup_{a\to\infty} \psi(a)g(1/a) < \infty$. Indeed, since ψ is increasing we have $\psi(a) \leqslant \psi(x)$ if $a \leqslant x$, and since $\psi \in ORV$ we have $\psi(a) \leqslant A(a/x)^\alpha \psi(x)$ if $x \geqslant a \geqslant x_0$ for some positive numbers x_0, A and α. Using $\sup_{p\geqslant 0} p^\alpha \phi_1(1/p) < \infty$ we obtain

$$\limsup_{a\to\infty} \psi(a)g(1/a) = \limsup_{a\to\infty} \sup_{x\geqslant 0} \frac{\psi(a)\phi_1(x/a)}{\psi(x)} < \infty.$$

Remark 18.1.4. Up to now we have normalized all partial maxima by n^{-1} and we have always assumed that the limit d.f. H of Y in (18.1.1) is simple max-stable. We can remove these restrictions as follows. For simplicity we analyze the situation in $\mathbb{R}$. Assume that $H(x)$ is a simple max-stable and assume there exists an increasing and continuous function r: $[0, \infty) \to [0, \infty)$ with an inverse s such that for the d.f. F of X_1 we have

$$F(r(x)) = H(x) \tag{18.1.46}$$

or equivalently

$$F(x) = H(s(x)). \tag{18.1.47}$$

For a sequence a_n of positive numbers to be determined later, it follows from (18.1.47) that

$$\Pr(M_n \leqslant a_n x) = F^n(a_n x) = H^n(s(a_n x)) = H\left(\frac{s(a_n x)}{n}\right).$$

For $a > 0$ we obtain

$$|F^n(a_n x) - H(x^\alpha)| = \left|\phi_1\left(\frac{s(a_n x)}{n}\right) - \phi_1(x^\alpha)\right|. \tag{18.1.48}$$

If $s \in RV_\alpha$ (or equivalently $r \in RV_{1/\alpha}$) and if we choose $a_n = r(n)$ it follows that (18.1.1) holds, i.e.,

$$\lim_{n \to \infty} F^n(a_n x) = H(x^\alpha). \tag{18.1.49}$$

If s is regularly varying, we expect to obtain a rate of convergence which results in (18.1.49). We quote the following result from the theory of regular variation functions.

Lemma 18.1.8 (Omey and Rachev, 1990). Suppose $h \in RV_\eta$, $(\eta > 0)$ and that h is bounded on bounded intervals of $[0, \infty)$. Suppose $0 \leqslant p \in ORV$ and such that

$$A_1(x/y)^\zeta \leqslant \frac{p(x)}{p(y)} \leqslant A_2(x/y)^\xi \quad \text{for each } x \geqslant y \geqslant x_0$$

for some constants $A_i > 0$, $x_0 \in \mathbb{R}$, $\xi < \eta$ and $\zeta \in \mathbb{R}$. If for each $x > 0$,

$$\limsup_{t \to \infty} \frac{h(t)}{t^\eta p(t)} \left|\frac{h(tx)}{h(t)} - x^\eta\right| < \infty \tag{18.1.50}$$

then

$$\limsup_{t \to \infty} \frac{h(t)}{t^\eta p(t)} \sup_{x \geqslant 0} \left|\phi_1\left(\frac{h(tx)}{h(t)}\right) - \phi_1(x^\eta)\right| < \infty. \tag{18.1.51}$$

If s satisfies the hypothesis of Lemma 18.1.8 (with an auxiliary function p and $\eta = \alpha$) then take $h(t) = s(t) = n$, $t = a_n$, in (19.1.51) to obtain

$$\limsup_{n \to \infty} \frac{n}{a_n^\alpha p(a_n)} \sup_{x \geqslant \infty} \left|\phi_1\left(\frac{s(a_n x)}{n}\right) - \phi_1(x^\alpha)\right| < \infty.$$

Combining these results with (18.1.48) we obtain the following theorem.

Theorem 18.1.4. Suppose $H(x) = \phi_1(x)$ and assume there exists an increasing and continuous function r: $[0, \infty) \to [0, \infty)$ with an inverse s such that $F(r(x)) = H(x)$.

(a) If $s \in RV_\alpha(\alpha > 0)$, then $\lim_{n\to\infty} \Pr\{M_n \leqslant a_n x\} = H(x^\alpha)$ where $a_n = r(n)$.
(b) If $s \in RV_\alpha$ with remainder term as in (18.1.50), then

$$\limsup_{n\to\infty} \frac{n}{a_n^\alpha p(a_n)} \rho(M_n/a_n, Y_\alpha) < \infty$$

where Y_α has d.f. $H(x^\alpha)$. QED

18.2 IDEAL METRICS FOR THE PROBLEM OF RATE OF CONVERGENCE TO MAX-STABLE PROCESSES

In this section we extend the results on the rate of convergence for maxima of random vectors (Section 18.1) by investigating maxima of random processes. In the new setup we need another class of ideal metrics simply because the weighted Kolmogorov metrics $\boldsymbol{\rho}_r$ and $\boldsymbol{\rho}_\psi$ (see (18.14) and Theorems 18.1.1, 18.1.2) cannot be extended to measure the distance between processes, cf. Open problem 4.3.2.

Let $\mathbf{B} = (\mathbf{L}_r[T], \|\cdot\|_r)$, $1 \leqslant r \leqslant \infty$, be the separable Banach space of all measurable functions x: $T \to \mathbb{R}$ (T is a Borel subset of $\mathbb{R}$) with finite norm $\|x\|_r$, where

$$\|x\|_r = \left\{\int_T |x(t)|^r \,\mathrm{d}t\right\}^{1/r} \quad 1 \leqslant r < \infty \tag{18.2.1}$$

and if $r = \infty$, $\mathbf{L}_\infty(T]$ is assumed to be the space of all continuous functions on a compact subset T with the norm

$$\|x\|_\infty = \sup_{t\in T} |x(t)|. \tag{18.2.2}$$

Suppose $\mathbf{X} = \{X_n, n \geqslant 1\}$ is a sequence of (dependent) random variables taking values in $\mathbf{B}$. Let $\mathscr{C}$ be the class of all sequences $\mathbf{C} = \{c_j(n); j, n = 1, 2, \dots\}$ satisfying the conditions

$$c_1(n) > 0 \quad c_j(n) \geqslant 0 \quad j = 1, 2, \dots, \quad \sum_{j=1}^\infty c_j(n) = 1. \tag{18.2.3}$$

For any $\mathbf{X}$ and $\mathbf{C}$ define the normalized maxima $\bar{X}_n := \bigvee_{j=1}^\infty c_j(n) X_j$, where $\bigvee := \max$ and $\bar{X}_n(t) := \bigvee_{j=1}^\infty c_j(n) X_j(t)$, $t \in T$.

In the previous section we have considered the special case of the sequence

$c_j(n)$, namely $c_j(n) = 1/n$ for $j \leqslant n$ and $c_j(n) = 0$ for $j > n$, and that X_n were i.i.d. random vectors. Here we are interested in the limit behavior of $\bar{X}_n$ in the general setting determined above. To this end we explore an approximation (denoted by $\bar{Y}_n$) of $\bar{X}_n$ with a known limit behavior. More precisely, let $\mathbf{Y} = \{Y_n, n \geqslant 1\}$ be a sequence of i.i.d. r.v.s and define $\bar{Y}_n = \bigvee_{j=1}^{\infty} c_j(n) Y_j$. Assuming that

$$\bar{Y}_n \stackrel{\mathrm{d}}{=} Y_1 \quad \text{for any } \mathbf{C} \in \mathscr{C} \tag{18.2.4}$$

we are interested in estimates of the deviation between $\bar{X}_n$ and $\bar{Y}_n$. The r.v. Y_1 satisfying (18.2.4) is called a *simple max-stable process.*

Example 18.2.1 (de Haan 1984). Consider a Poisson point process on $\mathbb{R}_+ \times [0, 1]$ with intensity measure $(\mathrm{d}x/x^2)\,\mathrm{d}y$. With probability 1 there are denumerably many points in the point process. Let $\{\xi_k, \eta_k\}$, $k = 1, 2, \ldots$, be an enumeration of the points in the process. Consider a family of non-negative functions $\{f_t(\cdot),\ t \in T\}$ defined on $[0, 1]$. Suppose for fixed $t \in T$ the function $f_t(\cdot)$ is measurable and $\int_0^1 f_t(v)\,\mathrm{d}v < \infty$. We claim that the family of random variables $Y(t) := \sup_{k \geqslant 1} f_t(\eta_k)\,\xi_k$ form a simple max-stable process. Clearly, it is sufficient to show that for any $\mathbf{C} \in \mathscr{C}$ and any $0 < t_1 < \cdots < t_k \in T$ the joint distribution of $(Y(t_1), \ldots, Y(t_k))$ satisfies the equality

$$\prod_{j=1}^{\infty} \Pr\{c_j Y(t_1) \leqslant y_1, \ldots, c_j Y(t_k) \leqslant y_k\}$$
$$= \Pr\{Y(t_1) \leqslant y_1, \ldots, Y(t_k) < y_k\} \quad \text{where } c_j = c_j(n).$$

Now

$$\prod_{j=1}^{\infty} \Pr\{c_j Y(t_1) \leqslant y_1, \ldots, c_j Y(t_k) \leqslant y_k\}$$
$$= \prod_{j=1}^{\infty} \Pr\{f_{t_i}(\eta_m)\xi_m \leqslant y_i/c_j, \quad i = 1, \ldots, k; \quad m = 1, 2, \ldots\}$$
$$= \prod_{j=1}^{\infty} \Pr\{\text{there are no points of the point process above the graph of the function } g(v) = (1/c_j) \min_{i<k} y_i/f_{t_i}(v),\ v \in [0, 1]\}$$
$$= \prod_{j=1}^{\infty} \exp\left(- \int_0^1 \left[\int_{\{x > g(v)\}} x^{-2}\,\mathrm{d}x \right] \mathrm{d}v \right) = \prod_{j=1}^{\infty} \exp\left(- \int_0^1 \left(c_j \max_{i \leqslant k} f_{t_i}(v)/y_i \right) \mathrm{d}v \right)$$
$$= \exp \sum_{j=1}^{\infty} c_j \left(- \int_0^1 \max_{i \leqslant k} f_{t_i}(v)/y_i \right) \mathrm{d}v = \exp\left(- \int_0^1 \left(\max_{i \leqslant k} f_{t_i}(v)/y_i \right) \mathrm{d}v \right)$$
$$= \Pr\{Y(t_1) \leqslant y_1, \ldots, Y(t_k) \leqslant y_k\}.$$

In this section we seek the weakest conditions providing an estimate of the deviation $\mu(\bar{X}_n, \bar{Y}_n)$ with respect to a given compound or simple p. metric μ. Such a metric will be defined on the space $\mathfrak{X}(\mathbf{B})$ of all r.v.s $X: (\Omega, \mathscr{A}, \Pr) \to (\mathbf{B}, \mathfrak{B}(\mathbf{B}))$, where the probability space $(\Omega, \mathscr{A}, \Pr)$ is assumed to be nonatomic, see Section 2.5 and Remark 2.5.2.

Our method is based on exploring *compound max-ideal metrics of order* $r > 0$, i.e. compound p. metrics μ_r satisfying

$$\mu_r(c(X_1 \vee Y), c(X_2 \vee Y)) \leqslant c^r \mu_r(X_1, X_2) \quad X_1, Y_2, Y \in \mathfrak{X}(\mathbf{B}) \quad c > 0 \tag{18.2.5}$$

(cf. Definition 18.1.1).

In particular, if the sequence $\mathbf{X}$ consists of i.i.d. r.v.s we will derive estimates of the rate of convergence of $\bar{X}_n$ to Y_1 in terms of the *minimal metric* $\hat{\mu}_r$ defined by

$$\hat{\mu}_r(X, Y) := \hat{\mu}_r(\Pr_X, \Pr_Y) := \inf\{\mu_r(X', Y') : X', Y' \in \mathfrak{X}(\mathbf{B}), X' \stackrel{\mathrm{d}}{=} X, Y' \stackrel{\mathrm{d}}{=} Y\} \tag{18.2.6}$$

see Definition 3.2.2. By virtue of μ_r-ideality, $\hat{\mu}_r$ is a *simple max-ideal metric of order* $r > 0$, i.e. (18.2.5) holds for Y independent of X_is.

We start with estimates of the deviation between $\bar{X}_n$ and $\bar{Y}_n$ in terms of the $\mathscr{L}_p$*-probability compound metric*, cf. Example 3.3.1 with $d(x, y) = \|x - y\|_r$. For any $r \in [1, \infty]$ define

$$\mathscr{L}_{p,r}(X, Y) := [\mathbb{E}\|X - Y\|_r^p]^{1/p} \quad p \geqslant 1 \tag{18.2.7}$$

$$\mathscr{L}_{\infty,r}(X, Y) := \operatorname{ess\,sup}\|X - Y\|_r. \tag{18.2.8}$$

Let

$$\ell_{p,r} := \hat{\mathscr{L}}_{p,r}. \tag{18.2.9}$$

Let us recall some of the metric and topological properties of $\ell_{p,r}$ and $(\mathscr{P}(\mathbf{B}), \ell_{p,r})$. The duality theorem for the minimal metric w.r.t. $\mathscr{L}_{p,r}$ implies

$$\begin{aligned}
\ell_{p,r}^p(X, Y) = \sup\{&\mathbb{E}f(X) + \mathbb{E}g(Y) : f: \mathbf{B} \to \mathbb{R}, g: \mathbf{B} \to \mathbb{R},\\
&\|f\|_\infty := \sup\{|f(x)| : x \in \mathbf{B}\} < \infty, \quad \|g\|_\infty < \infty,\\
&\operatorname{Lip}(f) := \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|_r} < \infty, \quad \operatorname{Lip}(g) < \infty,\\
&f(x) + g(y) \leqslant \|x - y\|_r \text{ for any } x, y \in \mathbf{B}\}, \quad \text{for any } p \in [1, \infty)
\end{aligned} \tag{18.2.10}$$

see Corollary 5.2.2 and (3.2.12). Moreover, by Corollary 6.1.1, the representation (18.2.10) can be refined in the special case of $p = 1$:

$$\ell_{1,r}(X, Y) = \sup\{|\mathbb{E}f(X) - \mathbb{E}f(Y)| : f: \mathbf{B} \to \mathbb{R}, \|f\|_\infty \leqslant \infty, \operatorname{Lip}(f) \leqslant 1\}.$$

Corollary 7.4.2 and (7.4.15) give the dual form for $\ell_{\infty,r}$

$$\ell_{\infty,r}(X, Y) = \inf\{\varepsilon > 0: \Pi_\varepsilon(X, Y) = 0\} \tag{18.2.11}$$

where $\Pi_\varepsilon(X, Y) := \sup\{\Pr\{X \in A\} - \Pr\{Y \in A^\varepsilon\}: A \in \mathfrak{B}(\mathbf{B})\}$ and A^ε is the ε-neighborhood of A w.r.t. the norm $\|\cdot\|_r$.

If $\mathbf{B} = \mathbb{R}$, then $\ell_p = \ell_{p,r}$ has the explicit representation

$$\ell_p(X, Y) = \left[\int_0^1 |F_X^{-1}(x) - F_Y^{-1}(x)|^p \, dx\right]^{1/p} \quad 1 \leqslant p < \infty \tag{18.2.12}$$

$$\ell_\infty(X, Y) = \sup\{|F_X^{-1}(x) - F_Y^{-1}(x)|: x \in [0, 1]\} \tag{18.2.13}$$

where F_X^{-1} is the generalized inverse of the d.f. F_X of X, see Corollary 7.3.2 and (7.4.15).

As far as the ℓ_p-convergence in $\mathscr{P}(\mathbf{B})$ is concerned, if π is the Prokhorov metric

$$\pi(X, Y) := \inf\{\varepsilon > 0: \Pi_\varepsilon(X, Y) < \varepsilon\} \tag{18.2.14}$$

and $\omega_X(N) := \{\mathbb{E}\|X\|_r^p I\{\|X\|_r > N\}\}^{1/p}$, $N > 0$, $X \in \mathfrak{X}(\mathbf{B})$, then for any $N > 0$, $X, Y \in \mathfrak{X}(\mathbf{B})$

$$\ell_{p,r}(X, Y) \leqslant \pi(X, Y) + 2N\pi^{1/p}(X, Y) + \omega_X(N) + \omega_Y(N), \tag{18.2.15}$$

$$\ell_{p,r}(X, Y) \geqslant \pi(X, Y)^{(p+1)/p} \quad \ell_{\infty,r}(X, Y) \geqslant \pi(X, Y) \tag{18.2.16}$$

and

$$\omega_X(N) \leqslant 3(\ell_{p,r}(X, Y) + \omega_Y(N)). \tag{18.2.17}$$

In particular, if $\mathbb{E}\|X_n\|^p + \mathbb{E}\|X\|^p < \infty$, $n = 1, 2, \ldots$, then

$$\ell_{p,r}(X_n, X) \to 0 \Leftrightarrow \underline{\pi}(X_n, X) \to 0 \text{ and } \lim_{N\to\infty} \sup_n \omega_{X_n}(N) = 0 \tag{18.2.18}$$

see Lemma 8.2.1 and Corollary 8.2.1.

Define the sample maxima with normalizing constants $c_j(n)$ by

$$\bar{X}_n = \bigvee_{j=1}^{\infty} c_j(n) X_j \quad \bar{Y}_n = \bigvee_{j=1}^{\infty} c_j(n) Y_j. \tag{18.2.19}$$

In the next theorem we obtain estimates of the closeness between $\bar{X}_n$ and $\bar{Y}_n$ in terms of the metric $\mathscr{L}_{p,r}$. In particular, if $\mathbf{X}$ and $\mathbf{Y}$ have i.i.d. components and Y_1 is a simple max-stable process (cf. (18.2.4)) we obtain the rate of convergence of $\bar{X}_n$ to Y_1 in terms of the minimal metric $\ell_{p,r}$. With this aim in mind we need some conditions on the sequences $\mathbf{X}, \mathbf{Y}$ and $\mathbf{C}$, see (18.2.3).

Condition 1. Let

$$a_p(n) := \left[\sum_{j=1}^{\infty} c_j^p(n)\right]^{\bar{p}} \text{ for } p \in (0, \infty),\ \bar{p} := \min(1, 1/p) \tag{18.2.20}$$

and

$$a_\infty(n) := \sup_{j \geqslant 1} c_j(n). \tag{18.2.21}$$

Assume that

$$a_\alpha(n) < \infty \text{ for some fixed } \alpha \in (0, 1) \text{ and all } n \geqslant 1 \tag{18.2.22}$$

$$a_1(n) = 1 \quad \forall n \geqslant 1,\ a_p(n) \to 0 \text{ as } n \to \infty \quad \forall p > 1.$$

The main examples of **C** satisfying Condition 1 are the Cesàro and Abel summation schemes.

Cesàro sum:

$$c_j(n) = \begin{cases} 1/n & j = 1, 2, \ldots, n \\ 0 & j = n+1, n+2, \ldots \end{cases} \tag{18.2.23}$$

$$a_p(n) = \begin{cases} n^{1-p} & \text{for } p \in (0, 1] \\ n^{-1+1/p} & \text{for } p \in [1, \infty]. \end{cases} \tag{18.2.24}$$

Abel sum:

$$c_j(n) = (\exp(1/n) - 1)\exp(-j/n) \quad j = 1, 2, \ldots, n = 1, 2, \ldots \tag{18.2.25}$$

$$a_p(n) = (1 - \exp(-1/n))^p/(1 - \exp(-p/n)) \sim (1/p)n^{1-p}$$
$$\text{as } n \to \infty \text{ for any } p \in (0, 1) \tag{18.2.26}$$

$$a_p(n) = (1 - \exp(-1/n))(1 - \exp(-p/n))^{-1/p} \sim p^{-1/p} n^{-1+1/p}$$
$$\text{as } n \to \infty \text{ for any } p \in [1, \infty)$$

$$a_p(n) = 1 - \exp(-1/n) \sim 1/n \quad \text{as } n \to \infty \text{ for } p = \infty.$$

The following condition concerns the sequences **X** and **Y**.

Condition 2. Let $\alpha \in (0, 1)$ be such that $a_\alpha(n) < \infty$ (see (18.2.22)) and assume that

$$\sup_{j \geqslant 1} \mathbb{E}|X_j(t)|^\alpha < \infty \quad \text{for any } t \in T \tag{18.2.27}$$

$$\sup_{j \geqslant 1} \mathbb{E}|Y_j(t)|^\alpha < \infty \quad \text{for any } t \in T. \tag{18.2.28}$$

Condition 2 is quite natural. For example, if Y_j, $j \geqslant 1$, are independent copies of a max-stable process (see, e.g., Resnick, 1987a), then all one-dimensional marginal d.f.s are of the form $\exp(-\beta(t)/x)$, $x \geqslant 0$ (for some $\beta(t) \geqslant 0$) and hence (18.2.28) holds. In the simplest m-dimensional case $T = \{t_k\}_{k=1}^m$,

$X_j = \{X_j(t_k)\}_{k=1}^m, j \geqslant 1$ are i.i.d. random vectors (r.v.s) as well as $Y_j = \{Y_j(t_k)\}_{k=1}^m$, $j \geqslant 1$ are i.i.d. r.v.s with simple max-stable distribution, see Section 18.1. One can check that the condition (18.2.27) is necessary to have a rate $O(n^{1-r})$ $(r > 1)$ of the uniform convergence of the d.f. of $(1/n)\bigvee_{j=1}^n X_j$ to the simple max-stable distribution F_{Y_1}, see Theorem 18.1.3.

Theorem 18.2.1. (a) Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{C}$ satisfy Condition 1 and Condition 2. Let $1 < p \leqslant r \leqslant \infty$ and

$$\mathscr{L}_{p,r}(X_j, Y_j) \leqslant \mathscr{L}_{p,r}(X_1, Y_1) < \infty \quad \forall j = 1, 2, \ldots \tag{18.2.29}$$

Then

$$\mathscr{L}_{p,r}(\bar{X}_n, \bar{Y}_n) \leqslant a_p(n)\mathscr{L}_{p,r}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty. \tag{18.2.30}$$

(b) If $\mathbf{X}$ and $\mathbf{Y}$ have i.i.d. components, $1 < p \leqslant t \leqslant \infty$ and $\ell_{p,r}(X_1, Y_1) < \infty$, then

$$\ell_{p,r}(\bar{X}_n, \bar{Y}_n) \leqslant a_p(n)\ell_{p,r}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty \tag{18.2.31}$$

where $\ell_{p,r}$ is determined by (18.2.10).

In particular, if $\mathbf{Y}$ satisfies the 'max-stable property'

$$\bar{Y}_n \stackrel{d}{=} Y_1 \tag{18.2.32}$$

then

$$\ell_{p,r}(\bar{X}_n, Y_1) \leqslant a_p(n)\ell_{p,r}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty. \tag{18.2.33}$$

Proof. (a) Let $1 < p < r < \infty$. By Condition 1, Condition 2 and Chebychev's inequality, we have

$$\Pr\{\bar{X}_n(t) > \lambda\} \leqslant \lambda^{-\alpha} a_\alpha(n) \sup_{j \geqslant 1} \mathbb{E} X_j(t)^\alpha \to 0 \quad \text{as } \lambda \to \infty$$

and hence

$$\Pr\{\bar{X}_n(t) + \bar{Y}_n(t) < \infty\} = 1 \text{ for any } t \in T.$$

For any $\omega \in \Omega$ such that $\bar{X}_n(t)(\omega) + \bar{Y}_n(t)(\omega) < \infty$, we have

$$\bar{X}_n(t)(\omega) = \bigvee_{j=1}^m c_j(n) X_j(t)(\omega) + \varepsilon_\omega(m) \quad \lim_{m \to \infty} \varepsilon_\omega(m) = 0$$

$$\bar{Y}_n(t)(\omega) = \bigvee_{j=1}^m c_j(n) Y_j(t)(\omega) + \delta_\omega(m) \quad \lim_{m \to \infty} \delta_\omega(m) = 0$$

and hence

$$|\bar{X}_n(t)(\omega) - \bar{Y}_n(t)(\omega)| \leqslant \bigvee_{j=1}^m |c_j(n) X_j(t)(\omega) - c_j(n) Y_j(t)(\omega)| + |\varepsilon_\omega(m)| + |\delta_\omega(m)|.$$

So, with probability 1,

$$|\bar{X}_n(t) - \bar{Y}_n(t)| \leqslant \bigvee_{j=1}^{\infty} c_j(n)|X_j(t) - Y_j(t)|. \tag{18.2.34}$$

Using the Minkowski inequality and the fact that $p/r \leqslant 1$, we obtain

$$\begin{aligned}\mathscr{L}_{p,r}(\bar{X}_n, \bar{Y}_n) &= \left\{\mathbb{E}\left[\int_T |\bar{X}_n(t) - \bar{Y}_n(t)|^r \,\mathrm{d}t\right]^{p/r}\right\}^{1/p} \\ &\leqslant \left\{\mathbb{E}\left[\int_T \left[\bigvee_{j=1}^{\infty} c_j(n)|X_j(t) - Y_j(t)|\right]^r \mathrm{d}t\right]^{p/r}\right\}^{1/p} \\ &\leqslant \left\{\mathbb{E}\left[\int_T \sum_{j=1}^{\infty} c_j^r(n)|X_j(t) - Y_j(t)|^r \,\mathrm{d}t\right]^{p/r}\right\}^{1/p} \\ &\leqslant \left\{\mathbb{E} \sum_{j=1}^{\infty} c_j^p(n)\left[\int_T |X_j(t) - Y_j(t)|^r \,\mathrm{d}t\right]^{p/r}\right\}^{1/p} \\ &\leqslant a_p(n)\mathscr{L}_{p,r}(X_1, Y_1).\end{aligned}$$

If $p < r = \infty$ then

$$\begin{aligned}\mathscr{L}_{p,\infty}(\bar{X}_n, \bar{Y}_n) &\leqslant \left\{\mathbb{E}\left[\sup_{t\in T} \bigvee_{j=1}^{\infty} c_j(n)|X_j(t) - Y_j(t)|\right]^p\right\}^{1/p} \\ &\leqslant \left\{\mathbb{E} \sum_{j=1}^{\infty} c_j^p(n) \sup_{t\in T} |X_j(t) - Y_j(t)|\right\}^{1/p} \\ &\leqslant \sum_{j=1}^{\infty} c_j(n)\mathscr{L}_{p,\infty}(X_1, Y_1).\end{aligned}$$

The statement for $p = r = \infty$ can be proved in an analogous way.

(b) By the definition of the minimal metric (cf. (18.2.6) and Section 7.1) we have

$$\begin{aligned}\hat{\mathscr{L}}_{p,r}(\bar{X}_n, \bar{Y}_n) &= \inf\{\mathscr{L}_{p,r}(\tilde{X}, \tilde{Y}) : \tilde{X} \stackrel{\mathrm{d}}{=} \bar{X}_n,\ \tilde{Y} \stackrel{\mathrm{d}}{=} \bar{Y}_n\} \\ &\leqslant \inf\left\{\left[\sum_{j=1}^{\infty} c_j^p(n)\mathscr{L}_{p,r}^p(\tilde{X}_j, \tilde{Y}_j)\right]^{1/p} : \{\tilde{X}_j, j \geqslant 1\} \text{ are i.i.d.,}\right. \\ &\qquad \left.\{\tilde{Y}_j, j \geqslant 1\} \text{ are i.i.d., } (\tilde{X}_j, \tilde{Y}_j) \stackrel{\mathrm{d}}{=} (\tilde{X}_1, \tilde{Y}_1),\ \tilde{X}_1 \stackrel{\mathrm{d}}{=} X_1,\ \tilde{Y}_1 \stackrel{\mathrm{d}}{=} Y_1\right\} \\ &\leqslant \inf\left\{\left[\sum_{j=1}^{\infty} c_j^p(n)\mathscr{L}_{p,r}^p(\tilde{X}_1, \tilde{Y}_1)\right]^{1/p} : \tilde{X}_1 \stackrel{\mathrm{d}}{=} X_1,\ \tilde{Y}_1 \stackrel{\mathrm{d}}{=} Y_1\right\} \\ &= a_p(n)\hat{\mathscr{L}}_{p,r}(X_1, Y_1).\end{aligned}$$

By (18.2.9) and (18.2.10) we obtain (18.2.31).

Finally, (18.2.33) follows immediately from (18.2.31) and (18.2.32). QED

Corollary 18.2.1. Let $\{X_j, j \geqslant 1\}$ and $\{Y_j, j \geqslant 1\}$ be random sequences with i.i.d. real-valued components and $F_{Y_1}(x) = \exp\{-1/x\}$, $x \geqslant 0$. Then

$$\ell_p\left(\bigvee_{j=1}^{\infty} c_j(n)X_j,\ Y_1\right) \leqslant a_p(n)\ell_p(X_1, Y_1) \quad p \in [1, \infty] \tag{18.2.35}$$

where the metric ℓ_p is given by (18.2.12) and (18.2.13). In particular, if for some $1 < p \leqslant \infty$, $\ell_p(X_1, Y_1) < \infty$ then $\ell_p(\bigvee_{j=1}^{\infty} c_j(n)\, X_j, Y_1) \to 0$ as $n \to \infty$.

Note that $\ell_p(X_1, Y_1) < \infty$ for $1 < p < \infty$ may be viewed as a 'tail' condition similar to the condition $\boldsymbol{\rho}_r(X_1, Y_1) < \infty$ $(r > 1)$ in Theorem 18.1.1 (b).

Open problem 18.2.1. It is not difficult to check that if $\mathbb{E}|X_n|^p + \mathbb{E}|X|^p < \infty$ then, as $n \to \infty$

$$\ell_p(X_n, X) = \left\{\int_0^1 |F_{X_n}^{-1}(t) - F_X^{-1}(t)|^p\, dt\right\}^{1/p} \to 0 \tag{18.2.36}$$

provided that for some $r > p$

$$\boldsymbol{\rho}_r(X_n, X) := \sup_{x \in \mathbb{R}} |x|^r |F_{X_n}(x) - F_X(x)| \to 0. \tag{18.2.37}$$

Since on the right-hand side of (18.2.35) the conditions $\ell_p(X_1, Y_1) < \infty$ and $\mathbb{E}|Y_1|^p = \infty$ imply $\mathbb{E}|X_1|^p = \infty$ it is a matter of interest to find necessary and sufficient conditions for $\boldsymbol{\rho}_r(X_n, X) \to 0$ $(r > p)$, respective $\ell_p(X_n, X) \to 0$, in the case of X_n and X having infinite pth absolute moments; for example under the assumption: $\ell_p(X_n, Y) + \ell_p(X, Y) < \infty$, $p > 1$, where Y is a simple max-stable r.v.

Let π be the Prokhorov metric (18.2.14) in the space $\mathfrak{X}(\mathbf{B}, \|\cdot\|_r)$. Using the relationship between π and $\ell_{p,r}$ (18.2.16), we get the following rate of convergence of $\bar{X}(n)$ to Y_1 under the assumptions of Theorem 18.2.1(b).

Corollary 18.2.2. Let the assumptions of Theorem 18.2.1(b) be valid and (18.2.32) hold.

Then

$$\pi(\bar{X}(n), Y_1) \leqslant a_p(n)^{p/(1+p)} \ell_{p,r}(X_1, Y_1)^{p/(1+p)}. \tag{18.2.38}$$

The next theorem is devoted to a similar estimate of the closeness between $\bar{X}_n$ and $\bar{Y}_n$ but now in terms of the compound Q-difference pseudomoment

$$\tau_{p,r}(X, Y) = \mathbb{E}\|Q_p X - Q_p Y\|_r \quad p > 0 \tag{18.2.39}$$

where the homeomorphism Q_p on $\mathbf{B}$ is defined by

$$(Q_p x)(t) = |x(t)|^p \operatorname{sgn} x(t) \tag{18.2.40}$$

see Example 4.3.3 and (4.3.41). Recall that the minimal metric $\kappa_{p,r} = \tilde{\tau}_{p,r}$ admits the following form of Q_p*-difference pseudomoment*, see (4.3.42), (4.3.43) and Remark 7.1.3

$$\begin{aligned}\kappa_{p,r}(X, Y) = \sup\{|\mathbb{E}f(X) - \mathbb{E}f(Y)|: f: \mathbf{B} \to \mathbb{R},\ \|f\|_\infty < \infty,\\ |f(x) - f(y)| \leqslant \|Q_p x - Q_p y\|_r \quad \forall x, y \in \mathbf{B}\}\end{aligned} \tag{18.2.41}$$

and if $\mathbf{B} = \mathbb{R}$, $\kappa_{p,r} =: \kappa_p$ is the pth *difference pseudomoment*

$$\begin{aligned}\kappa_p(X, Y) &= p \int_{-\infty}^{\infty} |x|^{p-1} |F_X(x) - F_Y(x)|\, dx\\ &= \int_{-\infty}^{\infty} |F_{Q_pX}(x) - F_{Q_pY}(x)|\, dx.\end{aligned} \tag{18.2.42}$$

Recall also (see Remark 7.1.3) that

$$\kappa_{p,r}(X, Y) = \ell_{1,p}(Q_p X, Q_p Y) = \hat{\tau}_{p,r}(X, Y) \quad \forall X, Y \in \mathfrak{X}(\mathbf{B})$$

and thus, by (18.2.18), if $\mathbb{E}\|X_n\|_r^p + \mathbb{E}\|X\|_r^p < \infty$, $n = 1, 2, \ldots$, then

$$\kappa_{p,r}(X_n, X) \to 0 \Leftrightarrow \pi(X_n, X) \to 0 \quad \text{and } \mathbb{E}\|X_n\|_r^p \to \mathbb{E}\|X\|_r^p.$$

In the next theorem we relax the restriction $1 < p \leqslant r \leqslant \infty$ imposed in Theorem 18.2.1.

Theorem 18.2.2. (a) Let Condition 1 and Condition 2 hold, $p > 0$ and $1/p < r \leqslant \infty$. Assume that

$$\tau_{p,r}(X_j, Y_j) \leqslant \tau_{p,r}(X_1, Y_1) < \infty \quad j = 1, 2, \ldots \tag{18.2.43}$$

Then

$$\tau_{p,r}(\bar{X}_n, \bar{Y}_n) \leqslant \alpha_{\bar{p}}(n)\tau_{p,r}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty \tag{18.2.44}$$

where $\alpha_{\bar{p}}(n) := \sum_{j=1}^{\infty} c_j^{\bar{p}}(n)$, $\bar{p} := p \min(1, r)$.

(b) If $\mathbf{X}$ and $\mathbf{Y}$ consist of i.i.d. r.v.s, then $\kappa_{p,r}(X_1, Y_1) < \infty$ implies

$$\kappa_{p,r}(\bar{X}_n, \bar{Y}_n) \leqslant \alpha_{\bar{p}}(n)\kappa_{p,r}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty. \tag{18.2.45}$$

Moreover, assuming that (18.2.32) holds, we have

$$\kappa_{p,r}(\bar{X}_n, Y_1) \leqslant \alpha_{\bar{p}}(n)\kappa_{p,r}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty. \tag{18.2.46}$$

Proof. (a) By Condition 1 and Condition 2,

$$\Pr\left(\bigvee_{j=1}^{\infty} c_j^p(n)(Q_p X_j)(t) + \bigvee_{j=1}^{\infty} c_j^p(n)(Q_p Y_j)(t) < \infty\right) = 1.$$

Hence, as in Theorem 18.2.1 we have

$$\begin{aligned}
&\left|Q_p\left(\bigvee_{j=1}^{\infty} c_j(n)X_j\right)(t) - Q_p\left(\bigvee_{j=1}^{\infty} c_j(n)Y_j\right)(t)\right| \\
&\qquad = \left|\bigvee_{j=1}^{\infty} c_j^p(n)(Q_pX_j)(t) - \bigvee_{j=1}^{\infty} c_j^p(n)(Q_pY_j)(t)\right| \\
&\qquad \leqslant \bigvee_{j=1}^{\infty} c_j^p(n)|(Q_pX_j)(t) - (Q_pY_j)(t)|.
\end{aligned}$$

Next, denote $\tilde{r} = \min(1, 1/r)$ and then

$$\begin{aligned}
\tau_{p,r}(\tilde{X}_n, \tilde{Y}_n) &= \mathbb{E}\left[\int_T \left|Q_p\left(\bigvee_{j=1}^{\infty} c_j(n)X_j\right)(t) - Q_p\left(\bigvee_{j=1}^{\infty} c_j(n)Y_j\right)(t)\right|^r \mathrm{d}t\right]^{\tilde{r}} \\
&\leqslant \mathbb{E}\left[\sum_{j=1}^{\infty} \int_T c_j^{pr}(n)|(Q_pX_j)(t) - (Q_pY_j)(t)|^r \,\mathrm{d}t\right]^{\tilde{r}} \\
&\leqslant \sum_{j=1}^{\infty} c_j^{pr\tilde{r}}(n)\tau_{p,r}(X_j, Y_j) \leqslant \alpha_{\tilde{p}}(n)\tau_{p,r}(X_1, Y_1).
\end{aligned}$$

(b) Passing to the minimal metrics, as in Theorem 18.2.1(b), we obtain (18.2.45) and (18.2.46). QED

The next corollary can also be proved directly by using Lemma 18.1.3, noting that $\boldsymbol{\kappa}_p$ is a max-ideal metric of order p.

Corollary 18.2.3. Let $\mathbf{X}$ and $\mathbf{Y}$ consist of i.i.d. real-valued r.v.s and $F_{Y_1}(x) = \exp\{-1/x\}$, $x \geqslant 0$. Then

$$\boldsymbol{\kappa}_p(X_n, Y_1) \leqslant \alpha_p(n)\boldsymbol{\kappa}_p(X_1, Y_1) \quad p > 1 \tag{18.2.47}$$

where $\alpha_p(n) = \sum_{j=1}^{\infty} c_j^p(n)$ and $\boldsymbol{\kappa}_p$ is given by (18.2.42).

The main assumption in Corollary (18.2.38) is $\ell_{p,r}(X_1, Y_1) < \infty$. In order to relax it we shall consider a more refined estimate than (18.2.38). For this purpose we introduce the following metric

$$\chi_{p,r}(X, Y) := \left[\sup_{t>0} t^p \Pr\{\|X - Y\|_r > t\}\right]^{1/(1+p)} \quad p > 0,\ r \in [1, \infty]. \tag{18.2.48}$$

Lemma 18.2.1. For any $p > 0$, $\chi_{p,r}$ is a compound probability metric in $\mathfrak{X}(\mathbf{B})$.

Proof. Let us check the triangle inequality. For any $\alpha \in [0, 1]$ and any $t > 0$,

$$\Pr\{\|X - Y\|_r > t\} \leqslant \Pr\{\|X - Z\|_r > \alpha t\} + \Pr\{\|Z - Y\|_r > (1 - \alpha)t\}$$

and hence $\chi_{p,r}^{p+1}(X, Y) \leqslant \alpha^{-p}\chi_{p,r}^{p+1}(X, Z) + (1-\alpha)^{-p}\chi_{p,r}^{p+1}(Z, Y)$. Minimizing the right-hand side of the last inequality over all $\alpha \in (0, 1)$, we obtain $\chi_{p,r}(X, Y) \leqslant \chi_{p,r}(X, Z) + \chi_{p,r}(Z, Y)$. QED

We shall also use the *minimal metric w.r.t.* $\chi_{p,r}$

$$\xi_{p,r}(X, Y) := \hat{\chi}_{p,r}(X, Y) \quad p > 0. \tag{18.2.49}$$

The fact that $\xi_{p,r}$ is a metric follows from Theorem 3.2.1.

Lemma 18.2.2. (a) Let

$$\tilde{\omega}_X(N) := \left[\sup_{t>N} t^p \Pr\{\|X\|_r > t\}\right]^{1/(p+1)} \quad N > 0,\ X \in \mathfrak{X}(\mathbf{B}) \tag{18.2.50}$$

and

$$\boldsymbol{\eta}_{p,r}(X, Y) := \left[\sup_{t>0} t^p \Pi_t(X, Y)\right]^{1/(1+p)} \tag{18.2.51}$$

where Π_t is defined as in (18.2.11). Then for any $N > 0$ and $p > 0$

$$\pi \leqslant \boldsymbol{\eta}_{p,r} \leqslant \xi_{p,r} \leqslant \begin{cases} \ell_{p,r}^{p/(1+p)} & \text{if } p \geqslant 1 \\ \ell_{p,r}^{1/(1+p)} & \text{if } p \leqslant 1 \end{cases} \tag{18.2.52}$$

where $\ell_{p,r}$, $p < 1$, is determined by (3.2.12), (3.3.18) with $d(x, y) = \|x - y\|_r^p$:

$$\mathscr{L}_p(X, Y) = \ell_{p,r}(X, Y) := \sup\{|\mathbb{E}f(X) - \mathbb{E}f(Y)| : f: \mathbb{B} \to \mathbb{R} \text{ bounded},$$
$$|f(x) - f(y)| \leqslant \|x - y\|_r^p \quad \forall x, y \in \mathbb{B}\}.$$

Moreover,

$$\tilde{\omega}_X(N) \leqslant 2^{p/(1+p)}[\eta_{p,r}(X, Y) + \tilde{\omega}_Y(N/2)] \tag{18.2.53}$$

and

$$\xi_{p,r}^{p+1}(X, Y) \leqslant \max[\pi^p(X, Y), (2N)^p\pi(X, Y), 2^p(\tilde{\omega}_X^p(N) + \tilde{\omega}_Y^p(N))]. \tag{18.2.54}$$

(b) In particular, if $\lim_{N\to\infty} (\tilde{\omega}_{X_n}(N) + \tilde{\omega}_X(N)) = 0$, $n \geqslant 1$, then the following statements are equivalent:

$$\xi_{p,r}(X_n, X) \to 0 \tag{18.2.55}$$

$$\boldsymbol{\eta}_{p,r}(X_n, X) \to 0 \tag{18.2.56}$$

$$\pi(X_n, X) \to 0 \text{ and } \lim_{N\to\infty} \sup_{n\geqslant 1} \tilde{\omega}_{X_n}(N) = 0. \tag{18.2.57}$$

Proof. Suppose $\pi(X, Y) > \varepsilon > 0$. Then $\Pi_\varepsilon(X, Y) > \varepsilon$ (cf. (18.2.14)) and thus $\boldsymbol{\eta}_{p,r}(X, Y) \geqslant \varepsilon$ which gives $\boldsymbol{\eta}_{p,r} \geqslant \pi$. By using $\boldsymbol{\eta}_{p,r} \leqslant \chi_{p,r}$ and passing to the minimal metric $\xi_{p,r} = \hat{\chi}_{p,r}$ we get $\eta_{p,r} \leqslant \xi_{p,r}$. For $p \geqslant 1$, by Chebychev's inequality, $\chi_{p,r} \leqslant \mathscr{L}_{p,r}^{p/(1+p)}$ which implies $\xi_{p,r} \leqslant \ell_{p,r}^{p/(1+p)}$. The case of $p \in (0, 1)$ is handled in the same way which completes the proof of (18.2.52).

The proof of (18.2.54) and (b) is similar to that of Lemma 8.2.1 and Theorem 8.2.1. For details see Lemmas 2.4.1, 2.4.2 and Theorem 2.4.1 in Kakosjan *et al.* (1988). QED

Open problem 18.2.2. The equality $\eta_{p,r} = \xi_{p,r}$ may fail in general. The problem of getting dual representation for $\xi_{p,r}$, similar to that of $\mathscr{L}_{p,r}$ (see (18.2.9), (18.2.10)), is open.

The main purpose of the next theorem is to refine the estimate (18.2.38) in the case of $r = \infty$. By Lemma 18.2.2 (b) and (18.2.18) we know that $\ell_{p,\infty}$ is topologically stronger than $\xi_{p,\infty} = \hat{\chi}_{p,\infty}$. So, in the next theorem we shall show that it is possible to replace $\ell_{p,\infty}$ with $\xi_{p,\infty}$ in the right-hand side of the inequality (18.2.38) ($r = \infty$).

Theorem 18.2.3. (a) Let Condition 1 and Condition 2 hold and $\mathbf{X}$ and $\mathbf{Y}$ be sequences of r.v.s taking values in $\mathfrak{X}(\mathbf{L}_\infty)$ such that

$$\chi_{p,\infty}(X_j, Y_j) \leqslant \chi_{p,\infty}(X_1, Y_1) < \infty \quad \forall j \geqslant 1. \tag{18.2.58}$$

Then

$$\chi_{p,\infty}(\bar{X}_n, \bar{Y}_n) \leqslant \alpha_p^{1/(1+p)}(n)\chi_{p,\infty}(X_1, Y_1) \to 0 \quad \text{as } n \to \infty \tag{18.2.59}$$

where $\alpha_p := a_p^p$, $p > 1$.

(b) If $\mathbf{X}$ and $\mathbf{Y}$ have i.i.d. components and $\bar{Y}_n \stackrel{\mathrm{d}}{=} Y_1$, then

$$\xi_{p,\infty}(\bar{X}_n, Y_1) \leqslant \alpha_p^{1/(1+p)}(n)\xi_{p,\infty}(X_1, Y_1). \tag{18.2.60}$$

In particular,

$$\begin{aligned}\pi(\bar{X}_n, Y_1) &\leqslant \alpha_p^{1/(1+p)}(n)\xi_{p,\infty}(X_1, Y_1)\\ &\leqslant \alpha_p^{1/(1+p)}(n)\ell_{p,\infty}(X_1, Y_1)^{p/(1+p)}.\end{aligned} \tag{18.2.61}$$

Proof. (a) By (18.2.2) and (18.2.34)

$$\begin{aligned}\chi_{p,\infty}^{1+p}(\bar{X}_n, \bar{Y}_n) &\leqslant \sup_{u>0} u^p \Pr\left\{\sup_{t\in T} \bigvee_{j=1}^{\infty} |c_j(n)X_j(t) - c_j(n)Y_j(t)| > u\right\}\\ &\leqslant \sum_{j=1}^{\infty} \sup_{u>0} u^p \Pr\left\{\sup_{t\in T} |X_j(t) - Y_j(t)| > u/c_j(n)\right\}\\ &= \sum_{j=1}^{\infty} c_j^p(n)\chi_{p,\infty}^{1+p}(X_j, Y_j) \leqslant \alpha_p(n)\chi_{p,\infty}^{1+p}(X_1, Y_1).\end{aligned}$$

(b) Passing to the minimal metrics in (18.2.59), similar to part (b) of

Theorem 18.2.1, we get (18.2.60). Finally, using the inequality (18.2.52), we obtain (18.2.61). QED

Further, we shall investigate the *uniform rate of convergence of the distributions of maxima of random sequences.* Here we assume that $\mathbf{X} := \{X, X_j, j \geqslant 1\}$, $\mathbf{Y} := \{Y, Y_j, j \geqslant 1\}$ are sequences of i.i.d. r.v.s taking on values in $\mathbb{R}_+^\infty$ and

$$\bar{X}_n := \bigvee_{j=1}^{\infty} c_j(n) X_j \quad \bar{Y}_n := \bigvee_{j=1}^{\infty} c_j(n) Y_j \tag{18.2.62}$$

where the components $Y^{(i)}$, $i \geqslant 1$, of $\mathbf{Y}$ have extreme-value distribution $F_{Y^{(i)}}(x) = \phi_1(x) = \exp(-1/x)$, $x \geqslant 0$.

Further, we shall consider $\mathbf{C} \in \mathscr{C}$ (see (18.2.3)) subject to the condition

$$\alpha_p(n) := \sum_{j=1}^{\infty} c_j^p(n) \to 0 \quad \text{as } n \to \infty \text{ for any } p > 1. \tag{18.2.63}$$

Denote $a \circ x := (a^{(1)}x^{(1)}, a^{(2)}x^{(2)}, \ldots)$, $bx := (bx^{(1)}, bx^{(2)}, \ldots)$ for any $a = (a^{(1)}, a^{(2)}, \ldots) \in \mathbb{R}^\infty$, $x = (x^1, x^{(2)}, \ldots) \in \mathbb{R}^\infty$, $b \in \mathbb{R}$.

We shall examine the uniform rate of convergence $\boldsymbol{\rho}(\bar{X}_n, Y) \to 0$ (as $n \to \infty$) where $\boldsymbol{\rho}$ is the *Kolmogorov* (*uniform*) *metric*

$$\boldsymbol{\rho}(X, Y) := \sup\{|F_X(x) - F_Y(x)| : x \in \mathbb{R}^\infty\}. \tag{18.2.64}$$

Here, $F_X(x) := \Pr\{\bigcap_{i=1}^{\infty} [X^{(i)} \leqslant x^{(i)}]\}$, $x = (x^{(1)}, x^{(2)}, \ldots)$ is the d.f. of X. Our aim is to prove an infinite-dimensional analog of Theorem 18.1.1 concerning the uniform rate of convergence for maxima of m-dimensional random vectors (see 18.1.12)).

First, note that the assumption, that the components $X_j^{(k)}$ of X_j are non-negative, is not a restriction since $\boldsymbol{\rho}(\bar{X}_n, Y) = \boldsymbol{\rho}(\bigvee_{j=1}^{\infty} c_j(n)\tilde{X}_j, Y)$, where $\tilde{X}_j^{(k)} := \max(X_j^{(k)}, 0)$ $k \geqslant 1$ (see Remark 18.1.1)). As in (18.1.4) we define the *weighted Kolmogorov probability metric*

$$\boldsymbol{\rho}_p(X, Y) := \sup\{M^p(x)|F_X(x) - F_Y(x)| : x \in \mathbb{R}^\infty\} \quad p > 0 \tag{18.2.65}$$

where $M(x) := \inf_{i \geqslant 1} |x^{(i)}|$, $x \in \mathbb{R}^\infty$.

First, we shall obtain an estimate of the rate of convergence of $\bar{X}_n$ to Y in terms of $\boldsymbol{\rho}_p$, $p > 1$ (see Lemmas 18.1.2 and 18.1.3 for similar results).

Lemma 18.2.3. Let $p > 1$. Then

$$\boldsymbol{\rho}_p(\bar{X}_n, Y) \leqslant \alpha_p(n)\boldsymbol{\rho}_p(X, Y). \tag{18.2.66}$$

Proof. For any $x \in \mathbb{R}^\infty$

$$\begin{aligned} M^p(x)|F_{\bar{X}_n}(x) - F_Y(x)| &= M^p(x)|F_{\bar{X}_n}(x) - F_{\bar{Y}_n}(x)| \\ &\leqslant \sum_{j=1}^{\infty} M^p(x)|F_{X_j}(x/c_j(n)) - F_{Y_j}(x/c_j(n))| \leqslant \alpha_p(n)\boldsymbol{\rho}_p(X, Y). \end{aligned}$$

QED

The problem now is how to pass from the estimate (18.2.63) to a similar estimate for $\boldsymbol{\rho}(\bar{X}_n, Y)$. We were able to solve this problem for the case of finite-dimensional random vectors (see Theorem 18.1.1). A close look at the proof of Theorem 18.1.1 shows that in the infinite-dimensional case the max-smoothing inequality (Lemma 18.1.4) is not valid. (The same is true for the summation scheme, see (15.2.7).) Further, we shall use relationships between $\boldsymbol{\rho}$, $\boldsymbol{\rho}_p$ and other metric structures which will provide estimates for $\rho(\bar{X}_n, Y)$ 'close' to that in the right-hand side of (18.2.63).

The next lemma deals with inequalities between $\boldsymbol{\rho}$, $\boldsymbol{\rho}_p$ and the Lévy metric in the space $\mathfrak{X}^\infty = \mathfrak{X}(\mathbb{R}^\infty)$ of random sequences. We define the *Lévy metric* as follows

$$\mathbf{L}(X, Y) := \inf\{\varepsilon > 0 \colon F_X(x - \varepsilon\mathbf{e}) - \varepsilon \leqslant F_Y(x) \leqslant F_X(x + \varepsilon\mathbf{e}) + \varepsilon\} \quad (18.2.67)$$

for all $x \in \mathbb{R}^\infty$, where $\mathbf{e} := (1, 1, \dots)$.

Open problem 18.2.3. What are the convergence criteria for $\mathbf{L}$, $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_p$ in $\mathfrak{X}^\infty$? Since $\mathbf{L}$, $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_p$ are simple metrics the answer to this question depends on the choice of the norm

$$\|x\|_p = \left[\sum_{i=1}^{\infty} |x^{(i)}|^p\right]^{1/p}, \quad \|x\|_\infty = \sup_{1 \leqslant i < \infty} |x^{(i)}|$$

in the space of probability laws $\mathscr{P}((\mathbb{R}^\infty, \|\cdot\|_p))$.

Lemma 18.2.4. (a) For any $\beta > 0$, $X, Y \in \mathfrak{X}^\infty$

$$\mathbf{L}^{\beta+1}(X, Y) \leqslant \mathbb{E}\|X - Y\|_\infty^\beta \quad (18.2.68)$$

where $\|x\|_\infty := \sup_{i \geqslant 1} |x^{(i)}|$

$$\mathbf{L}(X, Y) \leqslant \boldsymbol{\rho}(X, Y) \quad (18.2.69)$$

and

$$\mathbf{L}^{p+1}(X, Y) \leqslant 2^p \boldsymbol{\rho}_p(X, Y). \quad (18.2.70)$$

(b) If $Y = (Y^{(1)}, Y^{(2)}, \dots)$ has bounded marginal densities $p_{Y^{(i)}}$, $i = 1, 2, \dots$, with $A_i := \sup_{x \in \mathbb{R}} p_{Y^{(i)}}(x) < \infty$ and $A := \sum_{i=1}^\infty A_i$, then

$$\boldsymbol{\rho}(X, Y) \leqslant (1 + A)\mathbf{L}(X, Y). \quad (18.2.71)$$

Moreover, if $X, Y \in \mathfrak{X}_+^\infty = \mathfrak{X}(\mathbb{R}_+^\infty)$ (i.e. X, Y have non-negative components), then

$$\mathbf{L}^{p+1}(X, Y) \leqslant \boldsymbol{\rho}_p(X, Y) \quad (18.2.72)$$

and

$$\boldsymbol{\rho}(X, Y) \leqslant \Lambda(p) A^{p/(1+p)} \boldsymbol{\rho}_p^{1/(1+p)}(X, Y) \quad p > 0 \quad (18.2.73)$$

where

$$\Lambda(p) := (1 + p)p^{-p/(1+p)}. \tag{18.2.74}$$

Proof. (a) The inequalities (18.2.68) and (18.2.69) are obvious. The first follows from Chebychev's inequality, the second from the definitions of $\mathbf{L}$ and $\boldsymbol{\rho}$. One can obtain (18.2.70) in the same manner as (18.2.71) which we are going to prove completely.

(b) Let $\mathbf{L}(X, Y) < \varepsilon$. Further, for each $x \in \mathbb{R}^\infty$ and $n = 1, 2, \ldots$, let $x_n := (x^{(1)}, \ldots, x^{(n)}, \infty, \infty, \ldots)$. Then $F_X(x_n) - F_Y(x_n) \leqslant \varepsilon + F_Y(x_n + \varepsilon\mathbf{e}) - F_Y(x_n) \leqslant \varepsilon + [A_1 + \cdots + A_n]\varepsilon$. Analogously,

$$F_Y(x_n) - F_X(x_n) \leqslant F_Y(x_n) - F_Y(x_n - \varepsilon\mathbf{e}) + \varepsilon \leqslant \varepsilon + [A_1 + \cdots + A_n]\varepsilon.$$

Letting $n \to \infty$, we obtain $\boldsymbol{\rho}(X, Y) \leqslant (1 + A)\varepsilon$, which proves (18.2.71).

Further, let $\mathbf{L}(X, Y) > \varepsilon > 0$. Then there exists $x_0 \in \mathbb{R}^\infty_+$ such that $|F_X(x) - F_Y(x)| > \varepsilon$ for all $x \in [x_0, x_0 + \varepsilon\mathbf{e}]$ (i.e., $x^{(i)} \in [x_0^{(i)}, x_0^{(i)} + \varepsilon]$ for all $i \geqslant 1$). Hence

$$\begin{aligned}\boldsymbol{\rho}_p(X, Y) &\geqslant \sup\{M^p(x)\varepsilon : x \in [x_0, x_0 + \varepsilon\mathbf{e}]\} \\ &\geqslant \varepsilon \inf_{z \in \mathbb{R}^\infty_+} \sup_{x \in [z, z+\varepsilon\mathbf{e}]} M^p(x) = \varepsilon^{1+p}.\end{aligned}$$

Letting $\varepsilon \to \mathbf{L}(X, Y)$, we obtain (18.2.72).

By (18.2.71) and (18.2.72) we obtain

$$\boldsymbol{\rho}(X, Y) \leqslant (1 + A)\boldsymbol{\rho}_p^{1/(1+p)}(X, Y). \tag{18.2.75}$$

Next we shall use the homogeneity of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_p$ in order to improve (18.2.75). Namely, using the equality

$$\boldsymbol{\rho}(cX, cY) = \boldsymbol{\rho}(X, Y) \quad \boldsymbol{\rho}_p(cX, cY) = c^p\boldsymbol{\rho}_p(X, Y) \quad c > 0 \tag{18.2.76}$$

we have, by (18.2.75)

$$\begin{aligned}\boldsymbol{\rho}(X, Y) &\leqslant \left(1 + \frac{1}{c}A\right)\boldsymbol{\rho}_p^{1/(1+p)}(cX, cY) \\ &= (c^{p/(1+p)} + c^{1/(1+p)}A)\boldsymbol{\rho}_p^{1/(1+p)}(X, Y).\end{aligned} \tag{18.2.77}$$

Minimizing the right-hand side of (18.2.77) w.r.t. $c > 0$, we obtain (18.2.73).

QED

Theorem 18.2.4. Let $\gamma > 0$ and $a = (a^{(1)}, a^{(2)}, \ldots) \in \mathbb{R}^\infty_+$ be such that $A(a, \gamma) := \sum_{k=1}^\infty (a^{(k)})^{1/\gamma} < \infty$. Then for any $p > 1$ there exists a constant $c = c(a, p, \gamma)$ such that

$$\boldsymbol{\rho}(\bar{X}_n, Y) \leqslant c\alpha_p(n)^{1/(1+p\gamma)}\boldsymbol{\rho}_p(a \circ X, a \circ Y)^{1/(1+p\gamma)}. \tag{18.2.78}$$

Remark 18.2.1. In the estimate (18.2.78) the 'convergence index' $\alpha_p(n)^{1/(1+p\gamma)}$ tends to the right one $\alpha_p(n)$ as $\gamma \to 0$, see Lemma 18.2.3. The constant c has the form

$$c := (1+\tilde{p})\tilde{p}^{\tilde{p}/(1+\tilde{p})}[A(a,\gamma)\lambda(\gamma)]^{\tilde{p}/(1+\tilde{p})} \tag{18.2.79}$$

where $\tilde{p} := p\gamma$ and

$$\lambda(\gamma) := \gamma \exp[(1+1/\gamma)(\ln(1+1/\gamma)-1)]. \tag{18.2.80}$$

Choosing $a = a(\gamma) \in \mathbb{R}^\infty$ such that $(a^{(k)})^{-1/\gamma}\,\lambda(\gamma) = k^{-\theta}$ for any $k \geqslant 1$ and some $\theta > 1$, one can obtain that $c = c(a(\gamma), p, \gamma) \to 1$ as $\gamma \to 0$. However, in this case, $a^{(k)} = a^{(k)}(\gamma) \to \infty$ as $\gamma \to 0$ for any $k \geqslant 1$ and hence $\boldsymbol{\rho}_p(a \circ X, a \circ Y) \to \infty$ as $\gamma \to 0$.

Proof of Theorem 18.2.4. Denote

$$\tilde{X}_j := a \circ X_j \quad \tilde{Y}_j =: a \circ Y_j,\ p_k(\gamma) := \sup_{x \geqslant 0} p_{(\tilde{Y}^{(k)})^{1/\gamma}}(x) \tag{18.2.81}$$

where $p_X(\cdot)$ means the density of a real-valued r.v. X. Using the inequality (18.2.73), we have that for any $\tilde{p} > \gamma$, i.e. $p > 1$,

$$\begin{aligned}\boldsymbol{\rho}(\tilde{X}_n, Y) &= \boldsymbol{\rho}\left(\bigvee_{j=1}^{\infty} c_j(n)^{1/\gamma}\tilde{X}_j^{1/\gamma},\ \tilde{Y}^{1/\gamma}\right)\\ &\leqslant \Lambda(\tilde{p})\left(\sum_{k=1}^{\infty} p_k(\gamma)\right)^{\tilde{p}/(1+\tilde{p})} \boldsymbol{\rho}_{\tilde{p}}^{1/(1+\tilde{p})}\left(\bigvee_{j=1}^{\infty} c_j(n)^{1/\gamma}\tilde{X}_j^{1/\gamma},\ \tilde{Y}^{1/\gamma}\right)\end{aligned} \tag{18.2.82}$$

where $\Lambda(\tilde{p})$ is given by (18.2.74). Next we exploit Lemma 18.2.3 and obtain

$$\boldsymbol{\rho}_{\tilde{p}}\left(\bigvee_{j=1}^{\infty} c_j(n)^{1/\gamma}\tilde{X}^{1/\gamma},\ \tilde{Y}^{1/\gamma}\right) \leqslant \alpha_{\tilde{p}/\gamma}(n)\boldsymbol{\rho}_{\tilde{p}/\gamma}(\tilde{X}_j, \tilde{Y}_j). \tag{18.2.83}$$

Now we can choose $\tilde{p} := p\gamma$. Then, by (18.2.82) and (18.2.83)

$$\boldsymbol{\rho}(X_n, Y) \leqslant \Lambda(\tilde{p})\left(\sum_{k=1}^{\infty} p_k(\gamma)\right)^{\tilde{p}/(1+\tilde{p})} \alpha_p(n)^{1/(1+\tilde{p})}\boldsymbol{\rho}_p(\tilde{X}_j, \tilde{Y}_j)^{1/(1+\tilde{p})}. \tag{18.2.84}$$

Finally, note that since the components of Y have common d.f. ϕ_1, then $p_k(\gamma) = (a^{(k)})^{-1/\gamma}\lambda(\gamma)$, where $\lambda(\gamma)$ is given by (18.2.80). QED

In Theorem 18.2.4 we have no restrictions on the sequence of **C** of normalizing constants $c_j(n)$ (see (18.2.3) and (18.2.63)). However, the rate of convergence $\alpha_p(n)^{1/(1+p\gamma)}$ is close but not equal to the exact rate of convergence, namely $\alpha_p(n)$.

In the next theorem we impose the following conditions on **C** which allow us to reach the exact rate of convergence.

(A.1) There exist absolute constants $K_1 > 0$ and a sequence of integers $m(n)$,

$n = 2, 3, \ldots$, such that

$$\sum_{j=1}^{m(n)} c_j(n) \geqslant K_1 \leqslant \sum_{j=m(n)+1}^{\infty} c_j(n) \tag{18.2.85}$$

and $m(n) < n$.

(A.2) There exist constants $\beta \in (0, 1)$, $\theta \geqslant 0$, $\varepsilon_m(n)$ and $\delta_{im}(n)$, $i = 1, 2, \ldots$, $n = 2, 3, \ldots$, such that

$$c_{i+m}(n) = \varepsilon_m(n) c_i(n-m) + \delta_{im}(n) \tag{18.2.86}$$

and

$$\left\{\sum_{i=1}^{\infty} |\delta_{im}(n)|^{\beta}\right\}^{1/(1+\beta)} \leqslant \theta \alpha_p(n) \tag{18.2.87}$$

for all $i = 1, 2, \ldots, n = 2, 3, \ldots$ and $m = m(n)$ defined by (A.1).

(A.3). There exists a constant K_2 such that

$$\alpha_p(n - m(n)) \leqslant K_2 \alpha_p(n). \tag{18.2.88}$$

We shall check now that the Cesàro sum (for any $p > 1$) satisfy (A.1) to (A.3).

Example 18.2.2 Cesàro sum (see (18.2.23)). For any $p \geqslant 1$ we have $\alpha_p(n) = n^{1-p}$.

(A.1) Take $m(n) = [n/2]$, where $[a]$ means the integer part of $a \geqslant 0$. Then (18.2.85) holds with $K_1 \leqslant \frac{1}{2}$ and, obviously, $m(n) < n$.

(A.2) The equality (18.2.86) is valid with $\varepsilon_m(n) = (n-m)/n$ and $\delta_{im} = 0$. Hence, $\theta = 0$ in (18.2.87).

(A.3) $K_2 := 2^{p-1}$.

Theorem 18.2.5. Let Y be max-stable sequences (see (18.2.4)) and $\mathbf{C}$ satisfy (A.1) to (A.3). Let $a \in \mathbb{R}_+^{\infty}$ be such that

$$\mathscr{A}(a) := \sum_{i=1}^{\infty} 1/a^{(i)} < \infty. \tag{18.2.89}$$

Let $p > 1$, $\tilde{X} = a \circ X$, $\tilde{Y} = a \circ Y$ and

$$\lambda_p := \lambda_p(\tilde{X}, \tilde{Y}) := \max(\boldsymbol{\rho}_p^{1/(p+1)}(\tilde{X}, \tilde{Y}), \boldsymbol{\rho}_p(\tilde{X}, \tilde{Y}), \Gamma_\beta)$$

where

$$\Gamma_\beta := \theta\{[\mathbb{E}\|\tilde{X}\|_\infty^\beta]^{1/(1+\beta)} + [\mathbb{E}\|\tilde{Y}\|_\infty^\beta]^{1/(1+\beta)}\}$$

and β, θ are given by (A.2). Then there exist absolute constants A and B such that

$$\lambda_p \leqslant A \Rightarrow \boldsymbol{\rho}(\bar{X}_n, Y) \leqslant B \lambda_p \alpha_p(n). \tag{18.2.90}$$

Remark 18.2.2. As appropriate pairs (A, B) satisfying (18.2.90) one can take any A and B such that $A \leqslant C_8(p, a)$, $B \geqslant C_9(p, a)$, where the constants C_8 and C_9 are defined in the following way. Denote

$$C_1(a) := 1 + (2/e)^2 \mathscr{A}(a)/K_1 \quad C_2(a) := c_1(a)(1 + K_2) \tag{18.2.91}$$

$$C_3(a) := (2/e)^2 \mathscr{A}(a) \quad C_4(p, a) := (p/e)^p \mathscr{B}(a)^{-p} \tag{18.2.92}$$

where $\mathscr{B}(a) := \min_{i \geqslant 1} a^{(i)} > 0$, cf. (18.2.89),

$$C_5(p, a) := 4C_4(p, a)K_1^{-p} \quad C_6(p, a) := \Lambda(p)\left(\frac{C_3(a)}{K_1}\right)^{p/(1+p)} \tag{18.2.93}$$

where $\Lambda(p)$ is given by (18.2.74),

$$C_7(p, a) := \Lambda(p) C_3(a)^{p/(1+p)}$$

$$C_8(p, a) := (2C_6(p, a)C_2(a))^{-1-p} \tag{18.2.94}$$

and

$$C_9(p, a) := \max\{1, C_5(p, a), C_7(p, a)(1 \vee \alpha_p(2))^{-p/(1+p)}\}.$$

The proof of Theorem 18.2.5 is essentially based on the next lemma. (In the following $X' \vee X''$, for $X', X'' \in \mathfrak{X}(\mathbb{R}_+^\infty)$, always means a random sequence with d.f. $F_{X'}(x)F_{X''}(x)$, $x \in \mathbb{R}_+^\infty$, as well as $\tilde{X}$ means $a \circ X$ where $a \in \mathbb{R}_+^\infty$ satisfies (18.2.89).)

Lemma 18.2.5. (a) (ρ_p *is a max-ideal metric of order* $p > 0$) For any $X', X'', Z \in \mathfrak{X}(\mathbb{R}_+^\infty)$ and $c > 0$, $\rho_p(cX', cX'') = c^p \rho_p(X', X'')$, $p > 0$, and

$$\rho_p(X' \vee Z, X'' \vee Z) \leqslant \rho_p(X', X'').$$

(b) (*Max-smoothing inequality*) If Y is a simple max-stable sequence, then for any $X', X'' \in \mathfrak{X}(\mathbb{R}_+^\infty)$ and $\delta > 0$,

$$\rho(X' \vee \delta\tilde{Y}, X'' \vee \delta\tilde{Y}) \leqslant C_4(p, a)\delta^{-p}\rho_p(X', X'') \tag{18.2.95}$$

$$\rho(X', \tilde{Y}) \leqslant C_7(p, a)\rho_p^{1/(1+p)}(X', \tilde{Y}) \tag{18.2.96}$$

where C_4, C_7 are given by (4.19) and (4.21) respectively.

(c) For any $X', X'', U, V \in \mathfrak{X}(\mathbb{R}_+^\infty)$,

$$\rho(X' \vee U, X'' \vee U) \leqslant \rho(X', X'')\rho(U, V) + \rho(X' \vee V, X'' \vee V). \tag{18.2.97}$$

Remark 18.2.3. Lemma 18.2.5 is the analogue of Lemmas 14.2.2, 14.3.1, and 14.3.2 concerning the summation scheme of i.i.d. r.v.s.

Proof. (a) and (c) are obvious, see Lemmas 18.1.2 and 18.1.7.

(b) Let $G(x) := \exp(-1/x)$, $x \geqslant 0$, and

$$C(p) := (p/e)^p = \sup_{x>0} x^{-p}G(x). \tag{18.2.98}$$

Then

$$F_{\tilde{Y}}(x/\delta) \leqslant \min_{i \geqslant 1} F_{a^{(i)}Y^{(i)}}(x^{(i)}/\delta) = \min_{i \geqslant 1} G(x^{(i)}/a^{(i)}\delta) \leqslant C(p)\mathscr{B}(a)^{-p}M(x)^p\delta^{-p}.$$

Hence, by (18.2.92) and (18.2.98), $\boldsymbol{\rho}(X' \vee \delta\tilde{Y}, X'' \vee \delta\tilde{Y}) \leqslant C_4(p, a)\delta^{-p}\boldsymbol{\rho}_p(X', X'')$, which proves (18.2.95). Further, by Lemma 18.2.4 (see (18.2.73)) we have

$$\begin{aligned}\boldsymbol{\rho}(X', \tilde{Y}) &\leqslant \Lambda(p)\left(C(2)\sum_{i=1}^{\infty} 1/a^{(i)}\right)^{p/(1+p)} \boldsymbol{\rho}_p(X', \tilde{Y})^{1/(1+p)} \\ &= C_7(p, a)\boldsymbol{\rho}_p(X', \tilde{Y})^{1/(1+p)}.\end{aligned}$$

QED

Proof of Theorem 18.2.5. The main idea of the proof is to follow the 'max-Bergström' method as in Theorem 18.1.1 but avoiding the use of max-smoothing inequality (18.1.16). If $n = 1, 2$, then by (18.2.96) and Lemma 18.2.4 we have

$$\begin{aligned}\boldsymbol{\rho}(\bar{X}_n, Y) &\leqslant C_7(p, a)\boldsymbol{\rho}_p^{1/(1+p)}\left(\bigvee_{i=1}^{\infty} c_i(n)\tilde{X}_i, \tilde{Y}_i\right) \\ &\leqslant C_7(p, a)\alpha_p(n)^{1/(1+p)}\boldsymbol{\rho}_p^{1/(1+p)}(\tilde{X}, \tilde{Y}).\end{aligned}$$

Since $\lambda_p \geqslant \boldsymbol{\rho}_p^{1/(1+p)}(\tilde{X}, \tilde{Y})$ and $C_7(p, a)\alpha_p(n)^{1/(1+p)} \leqslant \mathscr{B}\alpha_p(n)$ for $n = 1, 2$, we have proved (18.2.90) for any A and $n = 1, 2$.

We now proceed by induction. Suppose that

$$\boldsymbol{\rho}\left(\bigvee_{j=1}^{\infty} c_j(k)\tilde{X}_j, Y\right) \leqslant \mathscr{B}\lambda_p\alpha_p(k) \quad \forall k = 1, \ldots, n-1. \tag{18.2.99}$$

Let $m = m(n)$, $n \geqslant 3$, be given by (A.1). Then using the triangle inequality we obtain

$$\boldsymbol{\rho}\left(\bigvee_{j=1}^{\infty} c_j(n)\tilde{X}_j, \tilde{Y}\right) \leqslant J_1 + J_2 \tag{18.2.100}$$

where

$$J_1 := \boldsymbol{\rho}\left(\bigvee_{j=1}^{m} c_j(n)\tilde{X}_j \vee \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j, \bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j \vee \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j\right)$$

and

$$J_2 := \boldsymbol{\rho}\left(\bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j \vee \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j, \tilde{Y}\right).$$

Now we will use the inequality (18.2.97) in order to estimate J_1

$$J_1 \leqslant J_1' + J_1'' \tag{18.2.101}$$

where

$$J_1' := \boldsymbol{\rho}\left(\bigvee_{j=1}^{m} c_j(n)\tilde{X}_j, \bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j\right)\boldsymbol{\rho}\left(\bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j, \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j\right)$$

and

$$J_1'' := \boldsymbol{\rho}\left(\bigvee_{j=1}^{m} c_j(n)\tilde{X}_j \vee \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j, \bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j \vee \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j\right).$$

Let us estimate J_1'. Since Y is a simple max-stable sequence

$$\bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j \stackrel{\mathrm{d}}{=} a \circ \left(\sum_{j=m+1}^{\infty} c_j(n)\right) Y \tag{18.2.102}$$

(see (18.2.3) and (18.2.4)). Hence, by (18.2.102), (18.2.71), (A.1) and (A.2), we have

$$\begin{aligned}
&\boldsymbol{\rho}\left(\bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j, \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j\right)\\
&\quad \leqslant \left(1 + \left(\frac{2}{e}\right)^2 \sum_{i=1}^{\infty}\left(a^{(i)} \sum_{j=m+1}^{\infty} c_j(n)\right)^{-1}\right)\mathbf{L}\left(\bigvee_{j=1}^{\infty} c_{j+m}(n)\tilde{X}_j, \bigvee_{j=1}^{\infty} c_{j+m}(n)\tilde{Y}_j\right)\\
&\quad \leqslant C_1(a)\mathbf{L}\left(\bigvee_{j=1}^{\infty} (\varepsilon_m(n)c_j(n-m) + \delta_{jm}(n))\tilde{X}_j, \bigvee_{j=1}^{\infty} (\varepsilon_m(n)c_j(n-m) + \delta_{jm}(n))\tilde{Y}_j\right)\\
&\quad \leqslant C_1(a)\left[\mathbf{L}\left(\bigvee_{j=1}^{\infty} (\varepsilon_m(n)c_j(n-m) + \delta_{jm}(n))\tilde{X}_j, \bigvee_{j=1}^{\infty} \varepsilon_m(n)c_j(n-m)\tilde{X}_j\right)\right.\\
&\qquad + \mathbf{L}\left(\bigvee_{j=1}^{\infty} \varepsilon_m(n)c_j(n-m)\tilde{X}_j, \bigvee_{j=1}^{\infty} \varepsilon_m(n)c_j(n-m)\tilde{Y}_j\right)\\
&\qquad \left. + \mathbf{L}\left(\bigvee_{j=1}^{\infty} \varepsilon_m(n)c_j(n-m)\tilde{Y}_j, \bigvee_{j=1}^{\infty} (\varepsilon_m(n)c_j(n-m) + \delta_{jm}(n))\tilde{Y}_j\right)\right]\\
&\quad =: C_1(a)(I_1 + I_2 + I_3)
\end{aligned} \tag{18.2.103}$$

where $C_1(a)$ is given by (18.2.91). Let us estimate I_1 by using (A.2) and the inequality (18.2.68):

$$\begin{aligned}
I_1 &\leqslant \left\{\mathbb{E} \bigvee_{j=1}^{\infty} \|(\varepsilon_m(n)c_j(n-m) + \delta_{jm}(n))\tilde{X}_j - \varepsilon_m(n)c_j(n-m)\tilde{X}_j\|_\infty^\beta\right\}^{1/(1+\beta)}\\
&\leqslant \left\{\mathbb{E} \sum_{j=1}^{\infty} |\delta_{jm}|^\beta \|\tilde{X}_j\|_\infty^\beta\right\}^{1/(1+\beta)} \leqslant \theta\alpha_p(n)\{\mathbb{E}\|\tilde{X}\|_\infty^\beta\}^{1/1(1+\beta)}.
\end{aligned} \tag{18.2.104}$$

Analogously,

$$I_3 \leqslant \theta\alpha_p(n)\{\mathbb{E}\|\tilde{Y}\|_\infty^\beta\}^{1/(1+\beta)}. \tag{18.2.105}$$

In order to estimate I_2 we use the inductive assumption (18.2.99) condition (A.3) and (18.2.69)

$$\begin{aligned} I_2 &\leqslant \boldsymbol{\rho}\left(\bigvee_{j=1}^{\infty} \varepsilon_m(n)c_j(n-m)\tilde{X}_j, \bigvee_{j=1}^{\infty} \varepsilon_m(n)c_j(n-m)\tilde{Y}_j\right) \\ &\leqslant \mathscr{B}\lambda_p\alpha_p(n-m) \leqslant K_2\mathscr{B}\lambda_p\alpha_p(n). \end{aligned} \tag{18.2.106}$$

Hence, by (18.2.103)–(18.2.106) and (18.2.91) we have

$$\begin{aligned} \boldsymbol{\rho}\left(\bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j, \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j\right) &\leqslant C_1(a)[\Gamma_\beta + K_2\mathscr{B}\lambda_p]\alpha_p(n) \\ &\leqslant C_2(a)\mathscr{B}\lambda_p\alpha_p(n). \end{aligned} \tag{8.2.107}$$

Next, let us estimate $\boldsymbol{\rho}(\bigvee_{j=1}^m c_j(n)\tilde{X}_j, \bigvee_{j=1}^m c_j(n)\tilde{Y}_j)$ in J'_1. Since Y is a simple max-stable sequence (cf. (18.2.3) and (18.2.4)) we have

$$\bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j \overset{\mathrm{d}}{=} \sum_{j=1}^{m} c_j(n)\tilde{Y}. \tag{18.2.108}$$

Thus, by (18.2.73), (18.2.93) and (A.1),

$$\begin{aligned} &\boldsymbol{\rho}\left(\bigvee_{j=1}^{m} c_j(n)\tilde{X}_j, \bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j\right) \\ &\leqslant \Lambda(p)\left[(2/e)^2 \sum_{i=1}^{\infty}\left(a^{(i)}\sum_{j=1}^{m} c_j(n)\right)^{-1}\right]^{p/(1+p)} \boldsymbol{\rho}_p(\tilde{X}, Y)^{1/(1+p)} \\ &\leqslant C_6(p,a)\lambda_p^{1/(1+p)} \leqslant C_6(p,a)A^{1/(1+p)}. \end{aligned} \tag{8.2.109}$$

Using the estimates in (18.2.107) and (18.2.109), we obtain the following bound for J'_1

$$J'_1 \leqslant C_6(p,a)A^{1/(1+p)}C_2(a)\mathscr{B}\lambda_p\alpha_p(n) \leqslant \tfrac{1}{2}\mathscr{B}\lambda_p\alpha_p(n). \tag{18.2.110}$$

Now let us estimate J''_1. By (18.2.95), (18.2.102), (A.1) and (18.2.66), we have

$$\begin{aligned} J''_1 &\leqslant C_4(p,a)\boldsymbol{\rho}_p\left(\bigvee_{j=1}^{m} c_j(n)\tilde{X}_j, \bigvee_{j=1}^{m} c_j(n)\tilde{Y}_j\right)\left(\sum_{j=m+1}^{\infty} c_j(n)\right)^{-p} \\ &\leqslant C_4(p,a)K_1^{-p}\lambda_p\alpha_p(n). \end{aligned} \tag{18.2.111}$$

Analogously, we estimate J_2 (see (18.2.100))

$$\begin{aligned} J_2 &\leqslant C_4(p,a)\rho_p\left(\bigvee_{j=m+1}^{\infty} c_j(n)\tilde{X}_j, \bigvee_{j=m+1}^{\infty} c_j(n)\tilde{Y}_j\right)\left(\sum_{j=1}^{m} c_j(n)\right)^{-p} \\ &\leqslant C_4(p,a)K_1^{-p}\lambda_p\alpha_p(n). \end{aligned} \tag{18.2.112}$$

Since $2C_4(p, a)K_1^{-p} \leqslant \mathscr{B}/2$ (see Remark 18.2.2),

$$J_1'' + J_2 \leqslant \tfrac{1}{2}\mathscr{B}\lambda_p\alpha_p(n) \tag{18.2.113}$$

by (18.2.111) and (18.2.112). Finally, using (18.2.100), (18.2.101), (18.2.110) and (18.2.113), we obtain (18.2.99) for $k = n$. QED

In the case of the Cesàro sum (18.2.23) one can refine Theorem 18.2.5 following the proof of the theorem and using some simplifications (see Example (18.2.1)). Namely, the following assertion holds.

Corollary 18.2.4. Let $X, X_1, X_2, \dots$ be a sequence of i.i.d. r.v.s taking values in $\mathbb{R}_+^\infty$. Let $Y = (Y^{(1)}, Y^{(2)}, \dots)$ be a *max-stable sequence* (cf. (18.2.4)) with $F_{Y^{(i)}}(x) = \exp(-1/x), x > 0$. Let $a \in \mathbb{R}^\infty$ satisfy (18.2.89). Denote $\bar{\lambda}_p := \bar{\lambda}_p(\tilde{X}, \tilde{Y}) := \max\{\rho(\tilde{X}, \tilde{Y}), \rho_p(\tilde{X}, \tilde{Y})\}$, $\tilde{X} := a \circ X$, $\tilde{Y} := a \circ \tilde{Y}$. Then there exist constants C and D such that

$$\bar{\lambda}_p \leqslant C \Rightarrow \rho\left((1/n) \bigvee_{k=1}^{n} X_k,\ Y\right) \leqslant D\bar{\lambda}_p n^{1-p}. \tag{18.2.114}$$

Remark 18.2.4. As an example of a pair (C, D) that fulfils (18.2.114) one can choose any (C, D) satisfying the inequalities

$$CD(\tfrac{2}{3})^{p-1} \leqslant \tfrac{1}{2} \quad D \geqslant \max(2^p, 4C_4(p, a)(2^{p-1} + 6^p))$$

where $C_4(p, a)$ is defined by (18.2.92).

Remark 18.2.5. Let $Z_1, Z_2, \dots$ be a sequence of i.i.d. r.v.s taking values in the Hilbert space $H = (\mathbb{R}^\infty, \|\cdot\|_2)$ with $\mathbb{E}Z_1 = 0$ and covariance operator $\mathbf{V}$. The central limit theorem in H states that the distribution of the normalized sums $\tilde{Z}_n = n^{-1/2}\sum_{i=1}^n Z_i$ weakly tends to the normal distribution of a r.v. $Z \in \mathfrak{X}(H)$ with mean 0 and covariance operator V; however, the uniform convergence

$$\rho(F_{\tilde{Z}_n}, F_Z) := \sup_{x \in \mathbb{R}^\infty} |F_{\tilde{Z}_n}(x) - F_Z(x)| \to 0 \quad n \to \infty$$

may fail (see, for example, Sazonov 1981, pp. 69–70). In contrast to the summation scheme, Theorem 18.2.5 shows that under some tail conditions the distribution function of the normalized maxima $\bar{X}_n$ of i.i.d. r.v.s $X_i \in \mathfrak{X}(\mathbb{R}^\infty)$ converge uniformly to the d.f. of a simple max-stable sequence Y. Moreover, the rate of uniform convergence is nearly the same as in the finite-dimensional case (see Theorems 18.1.1 and 18.1.3). Furthermore, in our investigations we did not assume that $\mathbb{R}^\infty$ has the structure of Hilbert or even normed space.

Open problem 18.2.4. Smith (1982), Cohen (1982), Resnick (1987b) and Balkema and de Haan (1988) consider the univariate case $(X, X_1, X_2, \dots \in \mathfrak{X}(\mathbb{R}))$ of general normalized maxima

$$\boldsymbol{\rho}\left(a_n \bigvee_{i=1}^{n} X_i - b_n, Y\right) \leqslant c(X_1, Y)\phi_{X_1}(n),\ n = 1, 2, \ldots$$

(see also Theorem 18.2.4). In order to extend results of this type to the multivariate case ($X, X_1, X_2, \ldots \in \mathfrak{X}(\mathbf{B})$) using the method developed here, one needs to generalize the notions of compound and simple max-stable metrics (see (18.2.5) and Lemma 18.2.5(a)) by determining a metric μ_ϕ in $\mathfrak{X}(\mathbf{B})$ such that for any $X_1, X_2, Y \in \mathfrak{X}(\mathbf{B})$ and $c > 0$

$$\mu_\phi(c(X_1 \vee Y), c(X_2 \vee Y)) \leqslant \phi(c)\mu_\phi(X_1, X_2)$$

where ϕ: $[0, \infty) \to [0, \infty)$ is suitably chosen regular-varying with non-negative index, strictly increasing continuous function, $\phi(0) = 0$.

18.3 DOUBLY IDEAL METRICS

The minimal $\hat{\mathscr{L}}_p$-metrics are ideal w.r.t. summation and maxima of order $r_p = \min(p, 1)$. Indeed, by Definitions 14.2.1 and 18.1.1 the *p-average probability metrics*

$$\mathscr{L}_p(X, Y) = (\mathbb{E}\|X - Y\|^p)^{\min(1, 1/p)} \quad 0 < p < \infty$$
$$\mathscr{L}_\infty(X, Y) = \operatorname{ess\,sup} \|X - Y\| \quad X, Y \in \mathfrak{X}^d := \mathfrak{X}(\mathbb{R}^d) \tag{18.3.1}$$

(see Example 3.3.1) are compound ideal metrics w.r.t. sum and maxima of random vectors, i.e., for any X, Y, $Z \in \mathfrak{X}^d$

$$\mathscr{L}_p(cX + Z, cY + Z) \leqslant |c|^{r_p}\mathscr{L}_p(X, Y) \quad c \in \mathbb{R} \tag{18.3.2}$$

and

$$\mathscr{L}_p(cX \vee Z, cY \vee Z) \leqslant c^{r_p}\mathscr{L}_p(X, Y) \quad c \geqslant 0 \tag{18.3.3}$$

where $x \vee y := (x^{(1)} \vee y^{(1)}, \ldots, x^{(d)} \vee y^{(d)})$. Denote as before by $\hat{\mathscr{L}}_p$, the corresponding *minimal* metric, i.e,

$$\hat{\mathscr{L}}_p(X, Y) = \inf\{\mathscr{L}_p(\tilde{X}, \tilde{Y}); \tilde{X} \stackrel{\mathrm{d}}{=} X,\ \tilde{Y} \stackrel{\mathrm{d}}{=} Y\} \quad 0 < p \leqslant \infty. \tag{18.3.4}$$

Then, by (18.3.2) and (18.3.4), the 'ideality' properties hold:

$$\hat{\mathscr{L}}_p(cX + Z, cY + Z) \leqslant |c|^{r_p}\hat{\mathscr{L}}_p(X, Y) \quad c \in \mathbb{R} \tag{18.3.5}$$

and

$$\mathscr{L}_p(cX \vee Z, cY \vee Z) \leqslant c^{r_p}\hat{\mathscr{L}}_p(X, Y) \quad c > 0 \tag{18.3.6}$$

for any X, $Y \in \mathfrak{X}^d$ and Z independent of X and Y, see, for example, Theorem 7.1.2. In particular, if $X_1, X_2, \ldots$ are i.i.d. r.v.s and Y_α has a symmetric stable distribution with parameter $\alpha \in (0, 1)$, and $p \in (\alpha, 1]$, then one gets from (18.3.2)

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)}\right) \leqslant n^{1-p/\alpha} \hat{\mathscr{L}}_p(X_1, Y_{(\alpha)}) \tag{18.3.7}$$

which gives a precise estimate in the CLT under the only assumption that $\hat{\mathscr{L}}_p(X_1, Y_{(\alpha)}) < \infty$. Note that $\hat{\mathscr{L}}_p(X, Y) < \infty$ $(0 < p \leqslant 1)$ does not imply the finiteness of pth moments of $\|X\|$ and $\|Y\|$; for example, in the one-dimensional case, $d = 1$,

$$\hat{\mathscr{L}}_1(X, Y) = \int_{\mathbb{R}} |F_X(x) - F_Y(x)| \, \mathrm{d}x \quad X, Y \in \mathfrak{X}^1 \tag{18.3.8}$$

see Corollary 7.3.2, and therefore, $\hat{\mathscr{L}}_1(X_1, Y_{(\alpha)}) < \infty$ is a 'tail' condition on the d.f. F_X implying $\mathbb{E}|X_1| = +\infty$. Similarly, by (18.3.6), if $Z_{(\alpha)}$ is α-max-stable distributed r.v. on $\mathbb{R}^1$ (i.e., $F_{Z_{(\alpha)}}(x) := \exp(-x^{-\alpha})$, $x \geqslant 0$), then for $0 < \alpha < p \leqslant 1$

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}\right) \leqslant n^{1-p/\alpha} \hat{\mathscr{L}}_p(X_1, Z_{(\alpha)}) \tag{18.3.9}$$

for any i.i.d. r.v.s X_i.

In this section we shall investigate the following problems posed by Zolotarev (1983b, p. 300): 'It is known that there are ideal metrics or order $s \leqslant 1$ both in relation to the operation of ordinary addition of random variables and in the relation to the operation $\max(X, Y)$. Such a metric of first order is the Kantorovich metric (cf. (18.3.8)). "Doubly ideal metrics" may be useful in analyzing schemes in which both operations are present (schemes of this kind are actually encountered in certain queueing systems). However, not a single "doubly ideal" metric of order $s > 1$ is known. The study of the properties of these doubly ideal metrics and those of general type is an important and interesting problem.'

We shall prove that the problem of the existence of doubly ideal metrics of order $r > 1$ has an essential negative answer. In spite of this, the minimal $\hat{\mathscr{L}}_p$-metrics behave like ideal metrics of order $r > 1$ with respect to maxima and sums.

First we shall show that $\hat{\mathscr{L}}_p$, in spite of being only a *simple* $(r_p, +)$*-ideal metric*, i.e. ideal metric of order r_p w.r.t. summation scheme, (cf. Definition 14.2.1) $r_p = \min(1, p)$, it acts as an ideal $(r, +)$ metric of order $r = 1 + \alpha - \alpha/p$ for $0 < \alpha \leqslant p \leqslant 2$. We formulate this result for Banach spaces U of type p. Let $\{Y_i\}_{i \geqslant 1}$ be a sequence of independent random 'signs',

$$P(Y_i = 1) = P(Y_i = -1) = 1/2.$$

Definition 18.3.1 (cf. Hoffman-Jörgensen 1977). For any $p \in [1, 2]$, a separable *Banach space* $(U, \|\cdot\|)$ *is said to be of type* p, if there exists a constant C such that for all $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in U$

$$\mathbb{E}\left\|\sum_{i=1}^{n} Y_i x_i\right\|^p \leqslant C \sum_{i=1}^{n} \|x_i\|^p. \tag{18.3.10}$$

The above definition implies the following condition (cf. Hoffman-Jörgensen and Pisier 1976): there exists $A > 0$ such that for all $n \in \mathbb{N} := \{1, 2, \ldots\}$ and independent $X_1, \ldots, X_n \in \mathfrak{X}(U)$ with $\mathbb{E}X_i = 0$ and finite $\mathbb{E}\|X_i\|^p$ the following holds

$$\mathbb{E}\left\|\sum_{i=1}^{n} X_i\right\|^p \leqslant A \sum_{i=1}^{n} \mathbb{E}\|X_i\|^p. \tag{18.3.11}$$

Remark 18.3.1. (a) Every separable Banach space is of type 1.

(b) Every finite-dimensional Banach space and every separable Hilbert space is of type 2.

(c) $\mathscr{L}^q := \{X \in \mathfrak{X}^1 : \mathbb{E}|X|^q < \infty\}$ is of the type $p = \min(2, q)$ $\forall q \geqslant 1$.

(d) $\ell_q := \left\{X \in \mathbb{R}^\infty, \|x\|_q^q := \sum_{j=1}^{\infty} |x^{(j)}|^q < \infty\right\}$ is of type $p = \min(2, q)$, $q \geqslant 1$.

Theorem 18.3.1. If U is of type p, $1 \leqslant p \leqslant 2$ and $0 < \alpha < p \leqslant 2$, then for any i.i.d. r.v.s $X_1, \ldots, X_n \in \mathfrak{X}(U)$ with $\mathbb{E}X_i = 0$ and for a symmetric stable r.v. $Y_{(\alpha)}$ the following bound holds

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)}\right) \leqslant B_p n^{1/p - 1/\alpha} \hat{\mathscr{L}}_p(X_1, Y_{(\alpha)}) \tag{18.3.12}$$

where B_p is an absolute constant.

Proof. We use the following result of Woyczynski (1980): if U is of type p then for some constant B_p and any independent $Z_1, \ldots, Z_n \in \mathfrak{X}(U)$ with $\mathbb{E}Z_i = 0$

$$\mathbb{E}\left\|\sum_{i=1}^{n} Z_i\right\|^q \leqslant B_p^p \mathbb{E}\left(\sum_{i=1}^{n} \|Z_i\|^p\right)^{q/p} \quad q \geqslant 1. \tag{18.3.13}$$

Let $Y_1, \ldots, Y_n \in \mathfrak{X}(U)$ be independent, $Y_i \stackrel{\mathrm{d}}{=} Y_{(\alpha)}$; then $Z_i = X_i - Y_i$, $1 \leqslant i \leqslant n$ are also independent. Take $Y_{(\alpha)} = n^{-1/\alpha} \sum_{i=1}^{n} Y_i$. Then from (18.3.13) with $q = p$ follows

$$\mathscr{L}_p^p\left(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)}\right) \leqslant B_p^p n^{1 - p/\alpha} \mathscr{L}_p^p(X_1, Y_1). \tag{18.3.14}$$

By passing to the minimal metrics in the last inequality we establish (18.3.12).

QED

From the well known inequality between the Prokhorov metric π and $\hat{\mathscr{L}}_p$ (see (8.2.7) and (8.2.21)):

$$\pi^{p+1} \leqslant (\hat{\mathscr{L}}_p)^p \quad p \geqslant 1 \tag{18.3.15}$$

we immediately obtain the following corollary.

Corollary 18.3.1. Under the assumptions of Theorem 18.3.1,

$$\pi\left(n^{-1/\alpha} \sum_{i=1}^{n} X_i, Y_{(\alpha)}\right) \leqslant B_p^{p/(p+1)} n^{(1-p/\alpha)/(p+1)} \hat{\mathscr{L}}_p(X_1, Y_{(\alpha)})^{p/(p+1)} \tag{18.3.16}$$

for any $p \in [1, 2]$ and $p > \alpha$.

Remark 18.3.2. For $r = p \in [1, 2]$ the rate in (18.3.16) and in Zolotarev's estimate

$$\pi\left(n^{-1/\alpha} \left\| \sum_{i=1}^{n} X_i \right\|, \|Y_{(\alpha)}\|\right) \leqslant C n^{(1-r/\alpha)/(r+1)} \zeta_r^{1/(r+1)}(X_1, Y_{(\alpha)}) \tag{18.3.17}$$

($0 \leqslant \alpha \leqslant r < \infty$, C is an absolute constant) are the same. On the right-hand side ζ_r is Zolotarev's metric (14.2.1). A problem with the application of ζ_r for $r > 1$ in the infinite-dimensional case was pointed out by Bentkus and Rachkauskas (1985). In Banach spaces the convergence w.r.t. ζ_r, $r > 1$, does *not* imply weak convergence. Gine and Leon (1970) showed that in Hilbert spaces ζ_r does imply the weak convergence while by results of Senatov (1981) there is no inequality of the type $\zeta_r \geqslant c\pi^a$, $a > 0$, where c is an absolute constant. Under some smoothness conditions on the Banach space, Zolotarev (1976b), Theorem 5, obtained the estimate

$$\pi^{1+r}(\|X\|, \|Y\|) \leqslant C\zeta_r(X, Y) \tag{18.3.18}$$

where $C = C(r)$. Therefore, under these conditions (18.3.17) follows from the ideality of ζ_r: $\zeta_r(n^{-1/\alpha}\sum_{i=1}^n X_i, Y_{(\alpha)}) \leqslant n^{1-r/\alpha}\zeta_r(X_1, Y_{(\alpha)})$. It was proved by Senatov (1981) that the order in (18.3.17) is the right one for $r = 3$, $\alpha = 2$, namely $n^{-1/8}$. The only known upper estimate for ζ_r applicable in the stable case is (cf. Zolotarev 1978, Theorem 4)

$$\zeta_r \leqslant \frac{\Gamma(1+\alpha)}{\Gamma(1+r)} \mathrm{v}_r \quad r = m + \alpha \quad 0 < \alpha \leqslant 1, m \in \mathbb{N} \tag{18.3.19}$$

where

$$\mathrm{v}_r(X, Y) = \int \|\mathrm{x}\|^r |\Pr_X - \Pr_Y|(\mathrm{dx}) \tag{18.3.20}$$

is the rth *absolute pseudomoment*. So $\mathrm{v}_r(X_1, Y)_{(\alpha)}) < \infty$ ensures the validity of (18.3.17).

In contrast to the bound (18.3.17), which concerns only the distance between the norms of X and Y, the estimate (18.3.16) concerns the Prokhorov distance $\pi(X, Y)$ itself which is topologically strictly stronger than $\pi(\|X\|, \|Y\|)$ in the Banach space setting and it is more informative. Furthermore, from Zolotarev (1978), p. 272, it follows that

$$\hat{\mathscr{L}}_p^p(X, Y) \leqslant 2^p \boldsymbol{\kappa}_p(X, Y) \leqslant 2^p \mathbf{v}_p(X, Y) \tag{18.3.21}$$

where $\boldsymbol{\kappa}_r$, $r > 0$, is the rth *difference pseudomoment*

$$\begin{aligned} \boldsymbol{\kappa}_r(X, Y) &= \inf\{\mathbb{E} d_r(\tilde{X}, \tilde{Y}); \tilde{X} \stackrel{\mathrm{d}}{=} X, \tilde{Y} \stackrel{\mathrm{d}}{=} Y\} \\ &= \sup\{|\mathbb{E} f(X) - \mathbb{E} f(Y)| : f\colon U \to \mathbb{R} \text{ bounded} \\ &\qquad |f(x) - f(y)| \leqslant d_r(x, y),\ x, y \in U\} \end{aligned} \tag{18.3.22}$$

and $d_r(x, y) = \|x\|x\|^{r-1} - y\|y\|^{r-1}\|$ (see Remark 7.1.3). Since *the problem, whether* $\boldsymbol{\kappa}_r(X, Y) < \infty$, $\mathbb{E}(X - Y) = 0$ *implies* $\zeta_r(X, Y) < \infty$ *is still open for* $1 < r < 2$, the right-hand side of (18.3.16) seems to contain weaker conditions than the right-hand side of (18.3.17).

Remark 18.3.3. If $U = \mathscr{L}^p$ (cf. Remark 18.3.1 (c)) then with an appeal to the Burkholder inequality one can choose the constants B_p in (18.3.16) as follows

$$B_1 = 1 \quad B_p = 18p^{3/2}/(p-1)^{1/2} \quad \text{for } 1 < p \leqslant 2 \tag{18.3.23}$$

see Chow and Teicher (1978), p. 396.

Remark 18.3.4. Let $1 \leqslant p \leqslant 2$, let $(E, \mathbb{E}, \mu)$ be a measurable space and define

$$\ell_{p,\mu} := \{X : (E, \mathscr{E}) \times (\Omega, \mathscr{A}) \to (\mathbb{R}^1, \mathscr{B}^1) : \|X\|_{p,\mu} < \infty\} \tag{18.3.24}$$

where $\|X\|_{p,\mu} := \mathbb{E}(\int |X(t)|^p \,\mathrm{d}\mu(t))^{1/p}$; $(\ell_{p,\mu}, \| \ \|_{p,\mu})$ is a Banach space of stochastic processes. (It is identical to $\mathscr{L}^p$ for one point measures μ). Let $X_1, \ldots, X_n \in \mathfrak{X}(\ell_{p,\mu})$ with $\mathbb{E} X_i = 0$. Recall the *Marcinkiewicz–Zygmund inequality*: if $\{\xi_n, n \geqslant 1\}$ are independent integrable r.v.s with $\mathbb{E}\xi_n = 0$, then for every $p > 1$ there exist positive constants A_p and B_p such that

$$A_p \left\| \left(\sum_{j=1}^n \xi_j^2 \right)^{1/2} \right\|_p \leqslant \left\| \sum_{j=1}^n \xi_j \right\|_p \leqslant B_p \left\| \left(\sum_{j=1}^n \xi_j^2 \right)^{1/2} \right\|_p, \tag{18.3.25}$$

see Shiryaev (1984, p. 469) and Chow and Teicher (1988, p. 367).

By the Marcinkiewicz–Zygmund inequality

$$\mathbb{E} \left\| \sum_{i=1}^n X_i \right\|_{p,\mu}^p = \mathbb{E} \int \left| \sum_{i=1}^n X_i(t) \right|^p \mathrm{d}\mu(t) \leqslant \int B_p^p \mathbb{E} \left(\sum_{i=1}^n X_i^2(t) \right)^{p/2} \mathrm{d}\mu(t).$$

Since $p \leqslant 2$ we obtain from the Minkowski inequality

$$\mathbb{E}\left\|\sum_{i=1}^{n} X_i\right\|_{p,\mu}^{p} \leqslant B_p \sum_{i=1}^{n} \mathbb{E}\int |X_i(t)|^p \,\mathrm{d}\mu(t) = B_p \sum_{i=1}^{n} \|X_i\|_{p,\mu}^{p}$$

i.e., $\ell_{p,\mu}$ is of type p and, therefore, one can apply Theorem 18.3.1 and Corollary 18.3.1 to stochastic processes in $\ell_{p,\mu}$.

For $0 < \alpha < 2p \leqslant 1$ we have the following analog of Theorem 18.3.1. Using the same metric as in Section 18.2, see Lemma 18.2.2 and Theorem 18.2.3. Again let $(U, \|\cdot\|)$ be of type p and let ξ_p be the minimal metric w.r.t. the compound metric

$$\chi_p(X, Y) := \left[\sup_{t>0} t^p \Pr\{\|X - Y\| > t\}\right]^{1/(1+p)} \qquad p > 0.$$

Then the following bound for the $\mathscr{L}_p$-distance between the normalized sums of i.i.d. random elements in $\mathfrak{X} = \mathfrak{X}(U)$ holds.

Theorem 18.3.2. Let $X_1, \ldots, X_n \in \mathfrak{X}$ be i.i.d., let $Y_1, \ldots, Y_n \in \mathfrak{X}$ be i.i.d. and let $0 < \alpha < 2p < 1$. Then

$$\mathscr{L}_p\left(n^{-1/\alpha}\sum_{i=1}^{n} X_i, n^{-1/\alpha}\sum_{i=1}^{n} Y_i\right) \leqslant B_p n^{1/2p-1/\alpha}(\xi_{2p}(X_1, Y_1))^{p+1/2} \tag{18.3.26}$$

where B_p is an absolute constant.

Proof. We have

$$\begin{aligned}
\mathbb{E}\left\|n^{-1/\alpha}\sum_{i=1}^{n} X_i - n^{-1/\alpha}\sum_{i=1}^{n} Y_i\right\|^p &= n^{-p/\alpha}\mathbb{E}\left\|\sum_{i=1}^{n}(X_i - Y_i)\right\|^p \\
&\leqslant n^{-p/\alpha}\mathbb{E}\left(\sum_{i=1}^{n}\|X_i - Y_i\|\right)^p \\
&\leqslant B_p n^{-p/\alpha}\sqrt{n}\left(\sup_{c>0} c^2 \Pr(\|X_1 - Y_1\|^p > c)\right)^{1/2} \\
&= B_p n^{-p/\alpha}\sqrt{n}(\chi_{2p}(X, Y))^{p+1/2}
\end{aligned}$$

the last inequality follows from Lemma 5.3 of Pisier and Zinn (1977). Passing to the minimal metrics (18.3.26) follows. QED

Remark 18.3.5. From the ideality of order p of $\hat{\mathscr{L}}_p$ (cf. (18.3.5)) for $0 < p \leqslant 1$, one obtains for $0 < \alpha < 2p \leqslant 1$ the bound

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha}\sum_{i=1}^{n} X_i, n^{-1/\alpha}\sum_{i=1}^{n} Y_i\right) \leqslant n^{1-p/\alpha}\hat{\mathscr{L}}_p(X_1, Y_1) \tag{18.3.27}$$

and by the Hölder inequality

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha}\sum_{i=1}^{n} X_i, n^{-1/\alpha}\sum_{i=1}^{n} Y_i\right) \leqslant \hat{\mathscr{L}}_{2p}\left(n^{-1/\alpha}\sum_{i=1}^{n} X_i, n^{-1/\alpha}\sum_{i=1}^{n} Y_i\right)$$
$$\leqslant n^{1-2p/\alpha}\hat{\mathscr{L}}_{2p}(X_1, Y_1). \tag{18.3.28}$$

Since $(\xi_{2p}(X_1, Y_1))^{1+2p} \leqslant \hat{\mathscr{L}}_{2p}(X_1, Y_1)$ for $p < 1/2$, see (18.2.52), the condition $\xi_{2p}(X_1, Y_1) < \infty$ is weaker than the condition $\hat{\mathscr{L}}_{2p}(X_1, Y_1) < \infty$. Comparing the estimates (18.3.27) and (18.3.26) it is clear that (18.3.26) has the better order $(1 - p/\alpha > 1 - 2p/\alpha > (1/2p) - (1/\alpha))$. However,

$$\hat{\mathscr{L}}_p(X_1, Y_1) \leqslant 2\xi_{2p}(X_1, Y_1)^{(p+1)/2} \tag{18.3.29}$$

and thus the 'tail condition' in (18.3.27) is weaker than that in (18.3.26). To prove (18.3.29), it is enough to show that

$$\hat{\mathscr{L}}_p(X_1, Y_1) \leqslant 2\chi_{2p}(X_1, Y_1)^{(p+1)/2}. \tag{18.3.30}$$

The last inequality follows from the bound

$$\mathbb{E}d^p(X_1, Y_1) \leqslant T^p + \int_T^{\infty} \Pr(d(X_1, Y_1) > t)pt^{p-1}\,\mathrm{d}t$$
$$\leqslant T^p + (\chi_{2p}(X_1, Y_1))^{p+1}T^{-p} \quad T > 0$$

after a minimization with respect to T.

Up to now we have investigated the ideal properties of $\hat{\mathscr{L}}_p$ w.r.t. the sums of i.i.d. r.v.s. Next we shall look at the max-ideality of $\hat{\mathscr{L}}_p$ and this will lead us to the 'doubly ideal' properties of $\hat{\mathscr{L}}_p$.

First let us point out that *there is no compound ideal metric of order $r > 1$ for the summation scheme while compound max-ideal metrics of order $r > 1$ exist.*

Remark 18.3.6. It is easy to see that there is *no* non-trivial compound ideal metric μ w.r.t. the summation scheme when $r > 1$. Since the ideality (see Definition 14.2.1) would imply

$$\mu(X, Y) = \mu\left(\frac{X + \cdots + X}{n}, \frac{Y + \cdots + Y}{n}\right) \leqslant n^{1-r}\mu(X, Y), \quad \forall n \in \mathbb{N}$$

i.e., $\mu(X, Y) \in \{0, \infty\}$ $\forall X, Y \in \mathfrak{X}(U)$.

On the other hand the following metrics are examples of compound max-ideal metrics of any order. For $U = \mathbb{R}^1$ and any $0 < p \leqslant \infty$ define for $X, Y \in \mathfrak{X}(\mathbb{R}^1)$

$$\Delta_{r,p}(X, Y) = \left(\int_{-\infty}^{\infty} \phi_{X,Y}^p(x)|x|^{rp-1}\,\mathrm{d}x\right)^q \tag{18.3.31}$$

and

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

where $q = \min(1, 1/p)$ and $\phi_{X,Y}(x) = \Pr(X \leqslant x < Y) + \Pr(Y \leqslant x < X)$. It is easy to see that $\Delta_{r,p}$ is a compound probability metric. Obviously, for any $c > 0$ the following holds

$$\Delta_{r,p}(cX, cY) = \left(\int_{-\infty}^{\infty} \phi_{X,Y}^p(x/c)|x|^{rp-1}\,\mathrm{d}x\right)^q = c^{rpq}\Delta_{r,p}(X, Y)$$

and $\Delta_{r,\infty}(cX, cY) = c^r\Delta_{r,\infty}(X, Y)$. Furthermore, from $\{X \vee Z \leqslant x < Y \vee Z\} \subset \{X \leqslant x < Y\}$, which can be established for any r.v.s X, Y, Z by considering the different possible order relations between X, Y, Z, it follows that $\Delta_{r,p}$ is a *compound max-ideal metric or order* $r(1 \wedge p)$ *for* $0 < p \leqslant \infty$, $0 < r < \infty$.

Note that $\Delta_{r,p}$ is an extension of the metric $\boldsymbol{\Theta}_p$ defined in Example 3.3.3; $\boldsymbol{\Theta}_p = \Delta_{1,p}$. Following step by step the proof of Theorem 7.3.4 one can see that the minimal metric $\hat{\Delta}_{r,p}$ has the form of the difference pseudomoment

$$\hat{\Delta}_{r,p}(X, Y) = \left(\int_{-\infty}^{\infty} |F_X(x) - F_Y(x)|^p |x|^{rp-1}\,\mathrm{d}x\right)^q \tag{18.3.32}$$

for $p \in (0, \infty)$ and $\hat{\Delta}_{r,\infty}(X, Y) = \sup_{x \in \mathbb{R}^1} |x|^r |F_X(x) - F_Y(x)|$ is the weighted Kolmogorov metric $\boldsymbol{\rho}_r$, see (18.1.4). Thus if $Z_{(\alpha)}$ is a α-max-stable distributed r.v., then as in (18.1.5), (18.3.9) we obtain

$$\hat{\Delta}_{r,p}\left(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}\right) \leqslant n^{1-r^*/\alpha}\hat{\Delta}_{r,p}(X_1, Z_{(\alpha)})$$

where $r^* := r(1 \wedge p)$.

We next want to investigate the properties of the $\mathscr{L}_p$-metrics (cf. (18.3.1) and Example 3.3.1) w.r.t. maxima. Following the notations in Remark 18.3.4 we consider for $0 < \lambda \leqslant \infty$ the Banach space $U = \ell_{\lambda,\mu} = \{X: (E, \mathscr{E}) \times (\Omega, \mathscr{A}) \to (\mathbb{R}^1, \mathscr{B}^1);\ \|X\|_{\lambda,\mu} < \infty\}$ where

$$\|X\|_{\lambda,\mu} := \mathbb{E}\left(\int |X(t)|^\lambda\,\mathrm{d}\mu(t)\right)^{1/\lambda^*} \quad \text{for } 0 < \lambda < \infty,\ \lambda^* = 1 \vee \lambda$$

and define for $X, Y \in U$, $X \vee Y$ to be the pointwise maximum, $(X \vee Y)(t) = X(t) \vee Y(t), t \in E$. Following the definition of a simple max-stable process (see (18.2.41)) we call $Z_{(\alpha)}$ *α-max-stable process* if

$$Z_{(\alpha)} \stackrel{\mathrm{d}}{=} n^{-1/\alpha} \bigvee_{i=1} Y_i \tag{18.3.33}$$

for any $n \in \mathbb{N}$ and Y_is are i.i.d. copies of $Z_{(\alpha)}$.

The proof of the next lemma and theorem are similar to that in Theorem 17.8.1 and thus left to the reader.

Lemma 18.3.1. (a) For $0 < \lambda \leqslant \infty$ and $0 < p \leqslant \infty$ $\mathscr{L}_p$ is a compound ideal metric of order $r = 1 \wedge p$, with respect to maxima scheme, i.e., (18.3.6) holds.

(b) If $X_1, \ldots, X_n \in \mathfrak{X}(\ell_{\lambda,\mu})$ are i.i.d. and if $Z_{(\alpha)}$ is a α-max stable process, then for $r = 1 \wedge p$

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}\right) \leqslant n^{1-r/\alpha} \hat{\mathscr{L}}_p(X_1, Z_{(\alpha)}). \tag{18.3.34}$$

The estimate (18.3.34) is interesting for $r \leqslant \alpha$ only; for $1 < p \leqslant \lambda < \infty$ one can improve it as follows (cf. Theorem 18.2.1):

Theorem 18.3.3. Let $1 \leqslant p \leqslant \lambda < \infty$, then for $X_1, \ldots, X_n \in \mathfrak{X}(\ell_{\lambda,\mu})$ i.i.d. the following holds

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha} \bigvee_{i=1}^{n} X_i, Z_{(\alpha)}\right) \leqslant n^{1/p-1/\alpha} \hat{\mathscr{L}}_p(X_1, Z_{(\alpha)}). \tag{18.3.35}$$

Remark 18.3.7. (a) Comparing (18.3.35) with (18.3.34) we see that actually $\hat{\mathscr{L}}_p$ 'acts' in this important case as a simple max-ideal metric of order $\alpha + 1 - \alpha/p$. For $1 < p$ it holds that $1/p - 1/\alpha < 1 - 1/\alpha$, i.e. (18.3.35) is an improvement over (18.3.34).

(b) An analog of Theorem 18.3.3 holds also for the sequence space $\ell_\lambda \subset \mathbb{R}^\infty$, see Remark 18.3.1 (d).

Now we are ready to investigate the question of the existence and construction of doubly ideal metrics. Let U be a Banach space with maximum operation $\vee$.

Definition 18.3.2. (double ideal metrics). A probability metric μ on $\mathfrak{X}(U)$ is called

(a) (r, I)-ideal, if μ is *compound* $(r, +)$-*ideal* and *compound* $(r, \vee)$-*ideal*, i.e. for any X_1, X_2, Y and $Z \in \mathfrak{X}(U)$ and $c > 0$

$$\mu(X_1 + Y, X_2 + Y) \leqslant \mu(X_1, X_2) \tag{18.3.36}$$

$$\mu(X_1 \vee Z, X_2 \vee Z) \leqslant \mu(X_1, X_2) \tag{18.3.37}$$

and

$$\mu(cX_1, cX_2) = c^r \mu(X_1, X_2). \tag{18.3.38}$$

(b) (r, II)-ideal, if μ is compound $(r, \vee)$-ideal and simple $(r, +)$-ideal, i.e. (18.3.36)–(18.3.38) hold with Y independent of X_is,

(c) (r, III)-ideal, if μ is simple $(r, \vee)$-ideal and simple $(r, +)$-ideal, i.e. (18.3.36)–(18.3.38) hold with Y and Z independent of X_is.

Remark 18.3.8. In the above definition (c) the metric μ can be compound or simple. An example of a compound $(1/p, \text{III})$-ideal metric is Θ_p-metric ($p \geqslant 1$) (see (3.3.12) and (18.3.31))

$$\Theta_p(X, Y) := \left\{\int_{-\infty}^{\infty} (\Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1))^p \, dt\right\}^{1/p} \qquad 1 \leqslant p < \infty$$

$$\Theta_\infty(X, Y) := \sup_{t \in \mathbb{R}^1} (\Pr(X_1 \leqslant t < X_2) + \Pr(X_2 \leqslant t < X_1)).$$

Remark 18.3.9. Note that if μ is a (r, II)-ideal metric, then one obtains for $\{X_i\}$ i.i.d., $\{X_i^*\}$ i.i.d.

$$S_k := \sum_{i=1}^{k} X_i, \; S_k^* := \sum_{i=1}^{k} X_i^* \quad Z_n := n^{-1/\alpha} \bigvee_{k=1}^{n} S_k \quad Z_n^* := n^{-1/\alpha} \bigvee_{k=1}^{n} S_k^* \tag{18.3.39}$$

the estimate

$$\begin{aligned} \mu(Z_n, Z_n^*) &\leqslant n^{-r/\alpha} \mu\left(\bigvee_{k=1}^{n} S_k, \bigvee_{k=1}^{n} S_k^*\right) \\ &\leqslant n^{-r/\alpha} \sum_{k=1}^{n} \mu(S_k, S_k^*) \leqslant n^{-r/\alpha} \sum_{k=1}^{n} \sum_{j=1}^{k} \mu(X_j, X_j^*) \end{aligned} \tag{18.3.40}$$

and hence for the minimal metric $\hat{\mu}$ we get

$$\hat{\mu}(Z_n, Z_n^*) \leqslant \frac{n(n+1)}{2} n^{-r/\alpha} \hat{\mu}(X_1, X_1^*) < n^{2-r/\alpha} \hat{\mu}(X_1, X_1^*) \tag{18.3.41}$$

which gives us a rate of convergence if $0 < \alpha < r/2$. Therefore, from the known ideal metrics of order $r \leqslant 1$ one gets a rate of convergence for $\alpha \in (0, \frac{1}{2})$. It is therefore of interest to study Zolotarev's question for the construction of doubly ideal metrics of order $r > 1$.

Remark 18.3.10. $\mathscr{L}_p$, $0 < p < \infty$, is an example of a $(1 \wedge p, \text{I})$-ideal metric. We have seen in Remark 18.3.6 that there is no (r, I)-ideal metric for $r > 1$. $\hat{\mathscr{L}}_p$ is (r, III)-ideal metric of order $r = \min(1, p)$.

We now show that Zolotarev's question on the existence of a (r, II) or a (r, III)-ideal metric has essentially a negative answer.

Theorem 18.3.4. Let $r > 1$ and let the simple probability metric μ be (r, III)-ideal on $\mathfrak{X}(\mathbb{R})$ and assume that μ satisfies the following regularity conditions.

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

$$\overline{\lim_{n \to \infty}} \, \mu(X_n, Y_n) \geqslant \mu(a, b) \tag{18.3.42}$$

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

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

Proof. If μ is a simple $(r, +)$-ideal metric, then for integrable $X, Y \in \mathfrak{X}(\mathbb{R}^1)$ the following holds: $\mu((1/n)\sum_{i=1}^n X_i, (1/n)\sum_{i=1}^n Y_i) \leqslant n^{1-r}\mu(X, Y)$, where (X_1, Y_i) are i.i.d. copies of (X, Y). By the weak law of large numbers and *Condition 1* we have

$$\mu(\mathbb{E}X, \mathbb{E}Y) \leqslant \overline{\lim}\ \mu\left(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n Y_i\right).$$

Assuming that $\mu(X, Y) < \infty$, we have $\mu(\mathbb{E}X, \mathbb{E}Y) = 0$, i.e., $\mathbb{E}X = \mathbb{E}Y$ by *Condition 2*. Therefore, $\mu(X, Y) < \infty$ implies that $\mathbb{E}X = \mathbb{E}Y$. Therefore, by $\mu(X \vee a, Y \vee a) \leqslant \mu(X, Y)$ we have that $\mathbb{E}(X \vee a) = \mathbb{E}(Y \vee a)$, for all $a \in \mathbb{R}^1$, i.e. $\int_{-\infty}^a \Pr(X < x) - \Pr(Y < x)\,\mathrm{d}x = 0$ for all $a \in \mathbb{R}^1$. Thus $X \stackrel{\mathrm{d}}{=} Y$ and therefore $\mu(X, Y) = 0$. QED

Remark 18.3.11. Condition 1 seems to be quite natural. For example, let $\mathscr{F}$ be a class of non-negative lower semicontinuous functions (l.s.c.) on $\mathbb{R}^2$ and $\phi\colon [0, \infty) \to [0, \infty)$ continuous, nondecreasing. Suppose μ has the form of a 'minimal' functional,

$$\mu(X, Y) = \inf\left\{\phi\left(\sup_{f\in\mathscr{F}} \mathbb{E}f(\tilde{X}, \tilde{Y})\right): \tilde{X} \stackrel{\mathrm{d}}{=} X,\ \tilde{Y} \stackrel{\mathrm{d}}{=} Y\right\} \tag{18.3.43}$$

with respect to a compound metric $\mathbb{E}f(\tilde{X}, \tilde{Y})$ with ζ-structure, see (4.3.64). Then μ is l.s.c. on $\mathfrak{X}(\mathbb{R}^2)$, i.e., $(X_n, Y_n) \xrightarrow{\mathrm{w}} (X, Y)$, implies

$$\liminf_{n\to\infty} \mu(X_n, Y_n) \geqslant \mu(X, Y) \tag{18.3.44}$$

so *Condition 1* is fulfilled. Actually, suppose $\liminf_{n\to\infty} \mu(X_n, Y_n) < \mu(X, Y)$. Then for some subsequence $\{m\} \subset \mathbb{N}$ $\mu(\tilde{X}_m, \tilde{Y}_m)$ converges for some $a < \mu(X, Y)$. For $f \in \mathscr{F}$ the mapping $h_f\colon \mathfrak{X}(\mathbb{R}^2) \to \mathbb{R}$, $h_f(X, Y) := \mathbb{E}f(X, Y)$ is l.s.c. Therefore, also $\phi(\sup_{f\in\mathscr{F}} h_f)$ is l.s.c. and there exists a sequence $(\tilde{X}_m, \tilde{Y}_m)$ with $\tilde{X}_m \stackrel{\mathrm{d}}{=} X_m$, $\tilde{Y}_m \stackrel{\mathrm{d}}{=} Y_m$ such that $\mu(X_m, Y_m) = \phi(\sup_{f\in\mathscr{F}} h_f(\tilde{X}_m, \tilde{Y}_m))$. The sequence $\{\lambda_m := \Pr_{\tilde{X}_m, \tilde{Y}_m}\}_{m\geqslant 1}$ is tight. For any weakly convergent subsequence λ_{m_k} with limit λ, obviously λ has marginals $\Pr_X$ and $\Pr_Y$. Then for $(\tilde{X}, \tilde{Y})$ with distribution λ

$$\begin{aligned} a = \liminf_k \mu(X_{m_k}, Y_{m_k}) &= \liminf_k \mathbb{E}\phi\left(\sup_{f\in\mathscr{F}} h_f(\tilde{X}_{m_k}, \tilde{Y}_{m_k})\right) \\ &\geqslant \mathbb{E}\phi\left(\sup_{f\in\mathscr{F}} h_f(\tilde{X}, \tilde{Y})\right) \geqslant \mu(X, Y) \end{aligned}$$

which is in contradiction with our assumption. Therefore (18.3.44) holds.

Despite the fact that (r, III)-ideal and thus (r, II)-ideal metrics do not exist, we shall show next that for $0 < \alpha \leqslant 2$ the metrics $\mathscr{L}_p$ for $1 < p \leqslant 2$ 'act' as (r, II)-ideal metrics in the rate of convergence problem $\mathscr{L}_p(Z_n, Z_n^*) \to 0$ $(n \to \infty)$, where Z_n and Z_n^* are given by (18.3.39). The order of (r, II)-ideality is $r = 2\alpha + 1 - \alpha/p > 2\alpha$ and, therefore, we obtain a rate of convergence $n^{2-r/\alpha}$ (cf. below (18.3.48)).

We consider at first the case that $\{X_i\}$, $\{X_i^*\}$ in (18.3.39) are i.i.d. r.v.s in $(U, \|\cdot\|) = (\ell_p, \|\cdot\|_p)$, where for $x = \{x^{(j)}\} \in \ell_p$, $\|x\|_p := (\sum_{j=1}^{\infty} |x^{(j)}|^p)^{1/p}$, see Remark 18.3.1 (d). For $x, y \in \ell_p$ we define $x \vee y = \{x^{(j)} \vee y^{(j)}\}$.

Theorem 18.3.5. Let $0 \leqslant \alpha < p \leqslant 2$, $1 \leqslant p \leqslant 2$ and $\mathbb{E}(X_1 - X_1^*) = 0$; then for Z_n and Z_n^* *given by* (18.3.39)

$$\hat{\mathscr{L}}_p(Z_n, Z_n^*) \leqslant (p/(p-1)) B_p n^{1/p - 1/\alpha} \hat{\mathscr{L}}_p(X_1, X_1^*) \tag{18.3.45}$$

where the constant B_p is the same as in the Marcinkiewicz–Zygmund inequality (18.3.25). In the Hilbert space $(\ell_2, \|\cdot\|_2)$ the following holds

$$\hat{\mathscr{L}}_2(Z_n, Z_n^*) \leqslant \sqrt{2} n^{1/2 - 1/\alpha} \hat{\mathscr{L}}_2(X_1, X_1^*). \tag{18.3.46}$$

In particular, for the Prokhorov metric π we have

$$\pi(Z_n, Z_n^*) \leqslant (p/(p-1))^{p/(p+1)} B_p^{p/(p+1)} n^{(1-p/\alpha)/(p+1)} \hat{\mathscr{L}}_p^{p/(p+1)}(X_1, X_1^*). \tag{18.3.47}$$

Proof. Let $(\tilde{X}_i, \tilde{X}_i^*)$ be independent pairs of random variables in $\mathfrak{X}(\ell_p)$. Then for $\tilde{S}_k = \sum_{i=1}^k \tilde{X}_i$, $\tilde{S}_k^* = \sum_{i=1}^k \tilde{X}_i^*$ we have

$$\begin{aligned}
\mathscr{L}_p^p\left(n^{-1/\alpha} \bigvee_{k=1}^n \tilde{S}_k, n^{-1/\alpha} \bigvee_{i=1}^n \tilde{S}_k^*\right) &= n^{-p/\alpha} \mathscr{L}_p^p\left(\bigvee_{k=1}^n \tilde{S}_k, \bigvee_{k=1}^n \tilde{S}_k^*\right) \\
&= n^{-p/\alpha} \mathbb{E}\left[\sum_{j=1}^{\infty} \left| \bigvee_{k=1}^n \tilde{S}_k^{(j)} - \bigvee_{i=1}^n \tilde{S}_k^{*(j)} \right|^p\right] \\
&\leqslant n^{-p/\alpha} \mathbb{E} \sum_{j=1}^{\infty} \bigvee_{k=1}^n |\tilde{S}_k^{(j)} - \tilde{S}_k^{*(j)}|^p \\
&= n^{-p/\alpha} \sum_{j=1}^{\infty} \mathbb{E} \bigvee_{k=1}^n |\tilde{S}_k^{(j)} - \tilde{S}_k^{*(j)}|^p \\
&\leq n^{-p/\alpha} \sum_{j=1}^{\infty} (p/(p-1))^p \mathbb{E} |\tilde{S}_n^{(j)} - \tilde{S}_n^{*(j)}|^p
\end{aligned}$$

the last inequality following from Doob's inequality, see Chow and Teicher (1988) p. 247. Therefore, we can continue applying the Marcinkiewicz–Zygmund inequality (18.3.25) with

$$\leq n^{-p/\alpha}\sum_{j=1}^{\infty}(p/(p-1))^p B_p^p \mathbb{E}\left[\sum_{i=1}^{n}(\tilde{X}_i^{(j)}-\tilde{X}_i^{*(j)})^2\right]^{p/2}$$

$$\leqslant (p/(p-1))^p B_p^p n^{-p/\alpha}\sum_{j=1}^{\infty}\sum_{i=1}^{n}\mathbb{E}|\tilde{X}_i^{(j)}-\tilde{X}_i^{*(j)}|^p$$

$$= (p/(p-1))^p B_p^p n^{1-p/\alpha}\mathscr{L}_p^p(\tilde{X}_1, \tilde{X}_1^*) \qquad (18.3.48)$$

the last inequality following from the assumption that $p/2 \leqslant 1$. Passing to the minimal metrics we obtain (18.3.45) and (18.3.46). Finally, by means of $\pi^{p+1} \leqslant \hat{\mathscr{L}}_p^p$ we have (18.3.47). QED

The same proof also applies to the Banach space $\ell_{p,\mu}$ (cf. (18.3.24) and Theorem 18.3.3).

Theorem 18.3.6. If $0 \leqslant \alpha < p \leqslant 2$, $1 \leqslant p \leqslant 2$ and $X_1, \ldots, X_n \in \mathfrak{X}(\ell_{p,\mu})$ are i.i.d. and $X_1^*, \ldots, X_n^* \in \mathfrak{X}(\ell_{p,\mu})$ are i.i.d. such that $\mathbb{E}(X_1 - X_1^*) = 0$, then

$$\hat{\mathscr{L}}_p\left(n^{-1/\alpha}\bigvee_{k=1}^{n} S_k, n^{-1/\alpha}\bigvee_{k=1}^{n} S^*k\right) \leqslant (p/(p-1))B_p n^{1/p-1/\alpha}\hat{\mathscr{L}}_p(X_1, X_1^*) \quad (18.3.49)$$

and

$$\pi(Z_n, Z_n^*) \leqslant (p/(p-1))^{p/(1+p)}B_p^{p/(1+p)}n^{(1-p/\alpha)/(p+1)}\hat{\mathscr{L}}_p^{p/(p+1)}(X_1, X_1^*). \quad (18.3.50)$$

For an application of Theorem 18.3.6 to the problem of stability for queuing models, we refer to Section 12.2.

CHAPTER 19

Ideal Metrics and Stability of Characterizations of Probability Distributions

No probability distribution is a true representation of the probabilistic law of a given random phenomenon: assumptions such as normality, exponentiality, etc., are seldom if ever satisfied in practice. This is not necessarily a cause for concern, as many stochastic models characterizing certain probability distributions are relatively insensitive to 'small' violations of the assumptions. On the other hand, there are models where even a slight perturbation of the assumptions that determine the choice of a distribution will cause a substantial change in the properties of the model. It is therefore of interest to investigate the invariance or *stability* of the set of assumptions characterizing certain distributions by examining the effects of perturbations of the assumptions.

There are several approaches to this problem. One is based on the concept of statistical robustness (e.g., Hampel 1971; Huber 1977; Papantoni-Kazakos 1977; Roussas 1972); another makes use of information measures (e.g. Akaike 1981; Csiszar 1967; Kullback 1959; Ljung 1978; Wasserstein 1969); a third utilizes different measures of distance (see Zolotarev 1977, 1983a; Kalashnikov and Rachev 1985, 1986a,b, 1988; Hernández-Lerma and Marcus 1984; Rachev 1989a).

It is this third approach which we adopt in this book; it allows us to derive not merely qualitative results but also bounds on the distance between a particular attribute of the *ideal* distribution, the theoretical representation of the law of the physical random phenomenon under consideration, and a *perturbed* distribution, obtained from the ideal distribution by an appropriate weakening of the assumptions.

This *stability analysis* is formalized as follows: given a specific ideal model, we denote by $\mathscr{U}$ the class of all possible 'input' distributions and by $\mathscr{V}$ the class of all possible 'output' distributions of interest. Let $\mathscr{F}:\mathscr{U}\to\mathscr{V}$ be the transformation that maps $\mathscr{U}$ on $\mathscr{V}$. For example, in the next section, 19.1, $\mathscr{U}$ is the class of all distribution functions F on $(0,\infty)$ satisfying the moment normalizing condition: $\int x^p \, \mathrm{d}F(x) = 1$ for some positive p. For a given $F \in \mathscr{U}$ the 'output'

$\mathscr{F}(F) \in \mathscr{V}$ is the set of distributions of random variables

$$X_{k,n,p} := \sum_{j=1}^{k} \zeta_j^p \Big/ \sum_{j=1}^{n} \zeta_j^p \quad 1 \leqslant k \leqslant n,\ n \in \mathbb{N} := \{1, 2, \ldots,\}$$

where $\zeta_1, \zeta_2, \ldots$ is a sequence of i.i.d r.v.s with d.f. F.

The characterization problem we are interested in is the following: *Does there exist a (unique?) d.f. $F = F_p$ such that $X_{k,n,p}$ has a beta $B(k/p, (n-k)/p)$-distribution for any $k \leqslant n$, $n \in \mathbb{N}$?*

It is well known that F_1 is the standard exponential distribution and F_2 is the absolute value of a standard normal r.v. (see, for example, Cramer 1946, Section 18, and Diaconis and Freedman 1987a). Having a positive answer to the problem our next task is to investigate the *stability of the characterization of the input distribution F_p*. The stability analysis may be described as follows: given $\varepsilon > 0$, we seek conditions under which there exist strictly increasing functions f_1 and f_2, both continuous strictly increasing and vanishing at the origin such that the following two implications hold:

(a) Given a simple probability metric μ_1 on $\mathfrak{X}(\mathbb{R})$
(a1) $\mu_1(\tilde{F}_p, F_p) = \mu_1(\tilde{\zeta}_1, \zeta_1) < \varepsilon$ implies
(a2) $\sup_{k,n} \mu_1(\tilde{X}_{k,n,p}, X_{k,n,p}) < f_1(\varepsilon)$.

In (a2) $X_{k,n,p}$ is determined as above where ζ_is are F_p-distributed (and thus $X_{k,n,p}$ has a $B(k/p, (n-k)/p)$-distribution). Further, in (a2) the r.v. $\tilde{X}_{k,n,p} := \sum_{i=1}^{k} \tilde{\zeta}_i^p / \sum_{i=1}^{n} \tilde{\zeta}_i^p$ is determined by a 'disturbed' sequence $\tilde{\zeta}_1, \tilde{\zeta}_2, \ldots$ of i.i.d. non-negative r.v.s with common d.f. $\tilde{F}_p$ close to F_p in the sense that (a1) holds for some 'small' $\varepsilon > 0$.

Together with (a) we shall prove the continuity of the inverse mapping $\mathscr{F}^{-1}$

(b) Given a simple p. metric μ_2 on $\mathfrak{X}(\mathbb{R})$ the following implication holds

$$\sup_{k,n} \mu_2(\tilde{X}_{k,n,p}, X_{k,n,p}) < \varepsilon \Rightarrow \mu_2(\tilde{F}_p, F_p) < f_2(\varepsilon).$$

If a small value of $\varepsilon > 0$ yields a small value of $f_2(\varepsilon) > 0$, $i = 1, 2, \ldots$ then the *characterization of the input distribution* $U \in \mathscr{U}$ (in our case $U = F_p$) can be regarded as being relatively insensitive to small perturbations of the assumptions, or *stable*. In practice, the principal difficulty in performing such a stability analysis is in determining the appropriate metrics μ_i such that (a) and (b) hold. The procedure we use is first to determine *ideal metrics* μ_1 and μ_2. These are the metrics most appropriate to the characterization problem under consideration. What is meant by 'most appropriate' will vary from characterization to characterization, but ideal metrics have so far been identified for a large class of problems, see Chapters 14 to 17. The detailed discussion of the above problem of stability will be given in Sections 19.1 and 19.2.

In Section 19.3 we shall consider the *stability of the input distributions.* Here the characterization problem arises from the soil erosion model of Todorovich and Gani (1987) and its generalization (Rachev and Todorovich 1989, Rachev and Samorodnitski 1990). The outline of the generalized erosion model is as follows: Let $Y, Y_1, Y_2, \ldots$ be i.i.d. sequence of random variables; Y_i represents the yield of a given crop in the ith year. Let $Z, Z_1, Z_2, \ldots$ be independent of Ys sequence of i.i.d. r.v.s; Z_i represents the proportion of crop yield maintained in the year i, $Z_1 < 1$ corresponds to a 'bad' year due to erosion, $Z_i > 1$ corresponds to a 'good' year in which rain comes at the right time. Further, let τ be a geometric random variable independent of Ys and Zs representing a disastrous event such as a drought. The total crop yield until the disastrous year is

$$G = \sum_{k=1}^{\tau} Y_k \prod_{i=1}^{k} Z_i.$$

Now the input distributions are $U := (F_Y, F_Z)$ and the output distribution $V = F(U)$ is the law of G. In general, the description of the class of compound distributions $V(x) = \Pr(G \leqslant x)$ is a complicated problem (cf. the problem of stability in risk theory, Section 16.1). Consider the simple example of V being $E(\lambda)$, i.e., exponential with parameter $\lambda > 0$ (for the general case, see Section 19.3). Here, the 'input' $U = (F_Y, F_Z)$ consists of a constant $Z = z \in (0, 1)$ and the mixture $F_Y(x) = F_{\bar{Y}}(x) := zE(\lambda/p) + (1 - z)(E(\lambda/p) * E(\lambda z))$, where $p := (1 + E\tau)^{-1}$ and $*$ stands for the convolution. Again we can put the problem of *stability of* the exponential distribution $E(\lambda)$ as *an output of the characterization problem*

$$U = (F_{\bar{Y}}, F_{\bar{z}}) \overset{\mathscr{F}}{\to} \bar{V} = E(\lambda).$$

As in the previous example, the problem is to choose an 'ideal metric' providing the implication

$$\left.\begin{aligned} \nu(Y^*, \bar{Y}) &\leqslant \varepsilon \\ \nu(Z^*, z) &\leqslant \delta \end{aligned}\right\} \Rightarrow \nu(F_{V^*}, E(\lambda)) \leqslant \phi(\varepsilon, \delta)$$

where $V^* = \mathscr{F}(F_{Y^*}, F_{Z^*})$ and ϕ is a continuous strictly increasing function in both arguments on $\mathbb{R}^2_+$ and vanishing at the origin.

19.1 CHARACTERIZATION OF AN EXPONENTIAL CLASS OF DISTRIBUTIONS $\{F_p, 0 < p \leqslant \infty\}$ AND ITS STABILITY

Let $\zeta_1, \zeta_2, \ldots$ be a sequence of i.i.d. r.v.s with d.f. F satisfying the normalization $\mathbb{E}\zeta_1^p = 1$, $\infty > p > 0$ and define

$$X_{k,n,p} := \sum_{j=1}^{k} \zeta_j^p \Big/ \sum_{j=1}^{n} \zeta_j^p \quad 1 \leqslant k \leqslant n \quad n \in \mathbb{N} := \{1, 2, \ldots,\}. \tag{19.1.1}$$

Theorem 19.1.1. For any $0 < p < \infty$ there exists exactly one distribution $F = F_p$, such that for all $k \leqslant n$, $n \in \mathbb{N}$, $X_{k,n,p}$ has a beta distribution $B(k/p, (n-k)/p)$†. F_p has the density

$$f_p(x) = \frac{p^{1-1/p}}{\Gamma(1/p)} \exp\left(-\frac{x^p}{p}\right) \quad x \geqslant 0. \tag{19.1.2}$$

Proof. Let the r.v.s $\{\zeta_i\}_{i \in \mathbb{N}}$ have the common density f_p. Then

$$f_{\zeta_i^p}(x) = \frac{1}{p^{1/p}\Gamma(1/p)} x^{-1+1/p} \exp(-x/p) \quad x \geqslant 0$$

is the $\Gamma(1/p, 1/p)$-density. Recall that $\Gamma(\alpha, \nu)$-density is given by

$$\frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} \exp(-\alpha x) \quad x \geqslant 0, \nu > 0, \alpha > 0.$$

The family of gamma densities is closed under convolutions, $\Gamma(\alpha, \mu) * \Gamma(\alpha, \nu) = \Gamma(\alpha, \mu + \nu)$ (Feller 1971) and hence $\sum_{i=1}^k \zeta_i^p$ is $\Gamma(1/p, k/p)$-distributed.

Usual calculations show that

$$f_\kappa(x) = B\left(\frac{k}{p}, \frac{n-k}{p}\right) \frac{x^{-1+k/p}}{(x+1)^{n/p}} \quad x > 0$$

where

$$\kappa := \frac{\sum_{i=1}^{n} \zeta_i^p}{\sum_{i=k+1}^{n} \zeta_i^p}$$

and

$$B\left(\frac{k}{p}, \frac{n-k}{p}\right) := \frac{\Gamma(n/p)}{\Gamma(k/p)\Gamma((n-k)/p)}.$$

This leads to the $B(k/p, (n-k)/p)$-distribution of $X_{k,n,p}$, cf. Cramer (1946), Section 18 for the case $p = 2$.

On the other hand, assuming that $X_{1,n,p}$ has a $B(1/p, (n-1)/p)$-distribution for all $n \in \mathbb{N}$, by the SLLN, $nX_{1,n,p} \to \zeta_1^p$ a.s. Further, the density of $(nX_{1,n,p})^{1/p}$

† The density of beta distribution with parameters α and β is given by

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1.$$

given by

$$\frac{\Gamma\left(\frac{n}{p}\right)}{\Gamma\left(\frac{1}{p}\right)\Gamma\left(\frac{n-1}{p}\right)} \times \left(\frac{x^p}{n}\right)^{-1+1/p}\left(1-\frac{x^p}{n}\right)^{(n-1)/p-1}\frac{p}{n}x^{p-1}$$

converges pointwise to $f_p(x)$, since

$$\left(\frac{p}{n}\right)^{1/p}\frac{\Gamma\left(\frac{n}{p}\right)}{\Gamma\left(\frac{n-1}{p}\right)} \to 1$$

as $n \to \infty$ (see e.g., Abramowitz and Stegun 1970, page 257). Thus $f_{\zeta_1} = f_p$ as required. QED

Remark 19.1.1. One simple extension is to the case that $\tilde{\zeta}_1, \tilde{\zeta}_2, \ldots$ be i.i.d. r.v.s on the whole real line satisfying the conditions $\mathbb{E}\tilde{\zeta}_1 = 0$, $\mathbb{E}|\tilde{\zeta}_1|^p = 1$. Then $\tilde{X}_{k,n,p} = \sum_{j=1}^{k} |\zeta_j|^p / \sum_{j=1}^{n} |\tilde{\zeta}_1|^p$ is $B(k/p, (n-k)/p)$-distributed if and only if the density $\tilde{f}_p$ of $\tilde{\zeta}_1$ satisfies $\tilde{f}_p(x) + \tilde{f}_p(-x) = f_p(|x|)$. In this way one gets for $p = 2$ the normal distribution (cf. Cramer 1986, Section 18), for $p = 1$ one gets the Laplace distribution. Uniqueness can be obtained by the additional assumption of symmetry of F.

Remark 19.1.2. To obtain a meaningful result for $p = \infty$, we have to normalize $X_{k,n,p}$ in (19.1.1) by looking at the limit distribution of

$$X_{k,n,p}^{1/p} = \frac{\left(\sum_{j=1}^{k} \zeta_j^p\right)^{1/p}}{\left(\sum_{j=1}^{n} \zeta_j^p\right)^{1/p}}$$

as $p \to \infty$.

Let β be a $B(k/p,\ (n-k)/p)$-distributed r.v. and define $\gamma_{k,n,p} = \beta^{1/p}$, then $\gamma_{k,n,p}$ has a density given by

$$f\gamma_{k,n,p}(x) = B\left(\frac{k}{p}, \frac{n-k}{p}\right) p x^{k-1}(1-x^p)^{(n-1)/p} \quad 0 \leqslant x \leqslant 1.$$

By Theorem 19.1.1 ζ_j are F_p-distributed if and only if $X_{k,n,p}^{1/p} \stackrel{\mathrm{d}}{=} \gamma_{k,n,p}$.

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

$$\Pr(\gamma_{k,n,\infty} \leqslant x) = \begin{cases} \dfrac{n-k}{n} x^k & \text{if } 0 \leqslant x < 1, \\ 1 & \text{if } x \geqslant 1. \end{cases} \tag{19.1.3}$$

Thus the above d.f. plays the role of the normalized $B(k/p, (n-k)/p)$-distribution as $p \to \infty$. Clearly, $X_{k,n,p}^{1/p}$ converges to

$$X_{k,n,\infty} := \bigvee_{i=1}^{k} \zeta_i \Big/ \bigvee_{i=1}^{n} \zeta_i \quad \left(\bigvee \zeta_i := \max \zeta_i\right) \tag{19.1.4}$$

as $p \to \infty$. Now, similarly to the case $p \in (0, \infty)$, we pose the following question: Does there exist a (unique?) d.f. F_∞ of ζ_1 such that $X_{k,n,\infty} \stackrel{\mathrm{d}}{=} \gamma_{k,n,\infty}$ for any $k \leqslant n$, $n \in \mathbb{N}$?

Theorem 19.1.2. Let $\zeta_1, \zeta_2, \ldots$ be a sequences of positive i.i.d. r.v.s and let F_∞ stand for the uniform distribution on $[0, 1]$. Then $X_{k,n,\infty}$ and $\gamma_{k,n,\infty}$ are equally distributed for any $k \leqslant n$, $n \in \mathbb{N}$ if and only if ζ_1 is F_∞-distributed.

Proof. Assuming that ζ_1 is F_∞-distributed, the d.f. of $X_{k,n,\infty}$ has the form $\Pr(X \leqslant x(X \vee Y))$ where X, Y are independent with d.f.s $F_X(t) = t^k$, $F_Y(t) = t^{n-k}$, $0 \leqslant t \leqslant 1$. Therefore, for $0 \leqslant x \leqslant 1$

$$\begin{aligned} F_{X_{k,n,\infty}}(x) &= \int_0^x \Pr(t \leqslant x(t \vee Y))\, \mathrm{d}t^k \\ &= \int_0^x \Pr(t \leqslant xY, Y > t)\, \mathrm{d}t^k + \int_0^x \Pr(t \leqslant xt, Y \leqslant t)\, \mathrm{d}t^k \\ &=: I_1(x) + I_2(x). \end{aligned}$$

Now $I_1(x) = [(n-k)/n]x^k$ for $x \in [0, 1]$ and $I_2(x) = 0$ for $0 < x < 1$, $I_2(1) = k/n$. This implies that $X_{k,n,\infty}$ has a distribution given by (19.1.3).

On the other hand, if $X_{1,n,\infty} := \zeta_1 / \bigvee_{i=1}^n \zeta_i$ has the same distribution as $\gamma_{1,n,\infty}$, then letting $n \to \infty$ the distribution of $\bigvee_{i=1}^n \zeta_i$ converges weakly to 1 and therefore the limit of $F_{X_{1,n,\infty}} = F_{\gamma_{1,n,\infty}}$ is $F_{\zeta_1} = F_\infty$. QED

Theorems 19.1.1 and 19.1.2 show that the basic probability distributions—exponential, normal, and uniform—correspond to F_1, F_2 and F_∞ in our characterization problem.

Next we shall examine the stability of the exponential class F_p, $0 < p \leqslant \infty$.

We now consider a 'disturbed' sequence $\tilde{\zeta}_1, \tilde{\zeta}_2, \ldots$ of i.i.d. non-negative r.v.s with common d.f. $\tilde{F}_p$ close to F_p in the sense that the uniform metric

$$\rho := \boldsymbol{\rho}(\tilde{\zeta}_1, \zeta_1) = \boldsymbol{\rho}(\tilde{F}_p, F_p) \tag{19.1.5}$$

is close to zero. (Here as before $\boldsymbol{\rho}(X, Y) := \sup_x |F_X(x) - F_Y(x)|$.) The next theorem says that the distribution of $\tilde{X}_{k,n,p} = \sum_{i=1}^{k} \tilde{\zeta}_i^p / \sum_{i=1}^{n} \tilde{\zeta}_i^p$ is close to the beta $B(k/p, (n-k)/p)$-distribution w.r.t. the uniform metric. In the following c denotes absolute constants which may be different in different places, and $c(\dots)$ denotes quantities depending only on the arguments in the parenthesis.

Remark 19.1.3. In view of the comments at the beginning of the section, the choice of the metric ρ as a 'suitable' metric for the problem of stability is dictated by the following observation. In the stability analysis of the characterization of the 'input' distribution F_p, we require the existence of simple metrics μ_1 and μ_2 such that

$$\mu_1(\tilde{F}_p, F_p) \leqslant \varepsilon \Rightarrow \sup_{k,n} \mu_1(\tilde{X}_{k,n,p}, X_{k,n,p}) \leqslant f_1(\varepsilon) \tag{19.1.6}$$

and

$$\sup_{k,n} \mu_2(\tilde{X}_{k,n,p}, X_{k,n,p}) \leqslant \varepsilon \Rightarrow \mu_2(\tilde{F}_p, F_p) \leqslant f_2(\varepsilon) \tag{19.1.7}$$

(cf. (a1), (a2) and (b)). Clearly we would like to select metrics μ_1 and μ_2 in such a way that as $n \to \infty$

$$\mu_1(X_n, Y_n) \to 0 \Leftrightarrow \mu_2(X_n, Y_n) \to 0$$

i.e., μ_is generate one and the same uniformities (see Dudley 1989, Section 11.7) and, in particular, μ_is metrize one and the same topology in the space of laws. The 'ideal' choice will be to find a metric such that both (19.1.6) and (19.1.7) are valid with $\mu = \mu_1 = \mu_2$. The next two theorems show that this choice is possible with $\mu = \boldsymbol{\rho}$.

Theorem 19.1.3. For any $0 < p < \infty$ and $\{\tilde{\zeta}_i\}$ i.i.d. with $\mathbb{E}\tilde{\zeta}_1^p = 1$ and $\tilde{m}_\delta := \mathbb{E}\tilde{\zeta}_1^{(2+\delta)p} < \infty$ ($\delta > 0$), we have

$$\Delta := \sup_{k,n} \boldsymbol{\rho}(X_{k,n,p}, \tilde{X}_{k,n,p}) \leqslant c(\delta, \tilde{m}_\delta, p)\rho^{\delta/(3(2+\delta))}. \tag{19.1.8}$$

Proof. The proof follows the two stage approach of the MMD, see Fig. 1.1.1.

(a) First stage: solution of the problem in terms of 'ideal' metric (Claim 2).

(b) Transition from the 'ideal' metric to the 'traditional' metric (see Claim 1, 2, 4).

We start with the first claim:

Claim 1 The 'traditional' metric ρ *is a regular metric* (cf. Definition 14.2.1 (i)). In particular

$$\rho(X_{k,n,p}, \tilde{X}_{k,n,p}) \leqslant \rho\left(\sum_{i=1}^{k} \zeta_i^p, \sum_{i=1}^{k} \tilde{\zeta}_i^p\right) + \rho\left(\sum_{i=k+1}^{n} \zeta_i^p, \sum_{i=k+1}^{n} \tilde{\zeta}_i^p\right) \leqslant n\rho(\zeta_1, \tilde{\zeta}_1). \tag{19.1.9}$$

To prove (19.1.9) observe that

$$X_{k,n,p} = \frac{X_1}{X_1 + X_2} \quad \tilde{X}_{k,n,p} = \frac{\tilde{X}_1}{\tilde{X}_1 + \tilde{X}_2}$$

where $X_1 = \sum_{i=1}^k \zeta_i^p$, $X_2 = \sum_{i=k+1}^n \zeta_i^p$, $\tilde{X}_1 = \sum_{i=1}^k \tilde{\zeta}_i^p$, $\tilde{X}_2 = \sum_{i=k+1}^n \tilde{\zeta}_i^p$. Since $\phi(t) = t/(1+t)$ is strictly monotone and $X_{k,n,p} = \phi(X_1/X_2)$, we have that

$$\rho(X_{k,n,p}, \tilde{X}_{k,n,p}) = \rho\left(\frac{X_1}{X_2}, \frac{\tilde{X}_1}{\tilde{X}_2}\right).$$

Choosing $X_1^* \stackrel{\mathrm{d}}{=} X_1$, X_1^* independent of $\tilde{X}_2$, we obtain

$$\begin{aligned}
\rho\left(\frac{X_1}{X_2}, \frac{\tilde{X}_1}{\tilde{X}_2}\right) &\leqslant \rho\left(\frac{X_1}{X_2}, \frac{X_1^*}{\tilde{X}_2}\right) + \rho\left(\frac{X_1^*}{\tilde{X}_2}, \frac{\tilde{X}_1}{\tilde{X}_2}\right)\\
&= \sup_{x \geqslant 0}\left|\int_0^\infty \left[\Pr\left(\frac{y}{X_2} \leqslant x\right) - \Pr\left(\frac{y}{\tilde{X}_2} \leqslant x\right)\right] \mathrm{d}F_{X_1}(y)\right|\\
&\quad + \sup_{x \geqslant 0}\left|\int_0^\infty \left[\Pr\left(\frac{X_1}{y} \leqslant x\right) - \Pr\left(\frac{\tilde{X}_1}{y} \leqslant x\right)\right] \mathrm{d}F_{\tilde{X}_2}(y)\right|\\
&\leqslant \int_0^\infty \sup_{x \geqslant 0}\left|P\left(X_2 \geqslant \frac{y}{x}\right) - P\left(\tilde{X}_2 \geqslant \frac{y}{x}\right)\right| \mathrm{d}F_{X_1}(y)\\
&\quad + \int_0^\infty \sup_{x \geqslant 0} |P(X_1 \leqslant xy) - P(\tilde{X}_1 \leqslant xy)| \, \mathrm{d}F_{\tilde{X}_2}(y)\\
&= \rho(X_1, \tilde{X}_1) + \rho(X_2, \tilde{X}_2).
\end{aligned}$$

The second part of (19.1.9) follows from the regularity of ρ, i.e.

$$\rho(X + Z, Y + Z) \leqslant \rho(X, Y)$$

for Z independent of X, Y.

Claim 2 (Bound from above of the 'traditional' metric ρ *by the 'ideal' metric* ζ_2*).* Let $n > p$, $\mathbb{E}\zeta_1^p = \mathbb{E}\tilde{\zeta}_1^p = 1$, $\sigma_p^2 := \operatorname{Var}(\zeta_1^p)$, $\tilde{\sigma}_p^2 := \operatorname{Var}(\tilde{\zeta}_1^p) < \infty$. Then

$$\rho\left(\sum_{i=1}^n \zeta_i^p, \sum_{i=1}^n \tilde{\zeta}_i^p\right) \leqslant 3\sigma_p^{2/3}\left(2\pi\left(1 - \frac{p}{n}\right)\right)^{-1/3} \zeta_2^{1/3}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i, \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{Z}_i\right) \tag{19.1.10}$$

where

$$Z_i := \frac{\zeta_i^p - 1}{\sigma_p} \quad \tilde{Z}_i := \frac{\tilde{\zeta}_i^p - 1}{\sigma_p}$$

and

$$\zeta_2(X, Y) := \int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} (F_X(t) - F_Y(t))\, \mathrm{d}t \right| \mathrm{d}x$$

is the Zolotarev ζ_2-metric (14.1.1) and (14.1.2).

Proof. For any $n = 1, 2, \ldots$ the following holds:

$$\rho\left(\sum_{i=1}^{n} \zeta_i^p, \sum_{i=1}^{n} \tilde{\zeta}_i^p\right) = \rho\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{Z}_i\right).$$

From (14.1.16) we have

$$\rho(X, Y) \leqslant 3M^{2/3}(\zeta_2(X, Y))^{1/3} \tag{19.1.11}$$

where $M = \sup_{x \in \mathbb{R}} f_X(x)$ and the density of X is assumed to exist. We have

$$f_{1/\sqrt{n}\sum_{i=1}^n Z_i}(x) = \sigma_p \sqrt{n} f_{\sum_{i=i}^n \zeta_i^p}(\sqrt{n}\sigma_p x + 1)$$

and

$$f'_{\sum_{i=1}^n \zeta_i^p}(x) = \frac{1}{p^{n/p}\Gamma\left(\frac{n}{p}\right)} \left[\left(\frac{n}{p} - 1\right) x^{-2+n/p} \exp(-x/p) - \frac{1}{p} x^{-1+n/p} \exp(-x/p)\right] = 0$$

if and only if $(n/p) - 1 = (1/p)x$.

The sum $\sum_{i=1}^n \zeta_i^p$ is $\Gamma(1/p, n/p)$-distributed and hence for $n > p$ the following holds

$$f_{\sum_{i=1}^n \zeta_i^p}(x) \leqslant \frac{p^{(n-p)/p}(-1 + n/p)^{(n-p)/p} \exp(1 - n/p)}{p^{n/p}\left(\frac{n}{p} - 1\right)\left[\left(\frac{n}{p} - 1\right)^{n/p - 3/2} \exp\left(-\frac{n}{p} + 1\right)(2\pi)^{1/2}\right]}$$

using $\Gamma(z) \geqslant z^{z-1/2} \mathrm{e}^{-z}(2\pi)^{1/2}$.

This implies that

$$\begin{aligned} \sigma_p \sqrt{n} f_{\sum_{i=1}^n \zeta_i^p}(x) &\leqslant \sigma_p \frac{\sqrt{n} p^{n/p - 1}}{p^{n/p}\left(\frac{n}{p} - 1\right)^{1/2} (2\pi)^{1/2}} \\ &= \sigma_p \left(2\pi\left(1 - \frac{p}{n}\right)\right)^{-1/2} \end{aligned} \tag{19.1.12}$$

and thus (19.1.11) and (19.1.12) together imply (19.1.10).

Since the metric ζ_2 is an ideal metric of order 2 (see (14.1.18)) we get

Claim 3 (Solution of the estimation problem in terms of ideal metric ζ_2).

$$\zeta_2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i, \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \tilde{Z}_i\right) \leqslant \zeta_2(Z_1, \tilde{Z}_1). \tag{19.1.13}$$

Claim 4 (Bound from above of the ideal metric ζ_2 by the 'traditional' metric ρ). If $m_\delta < \infty$, *then*

$$\zeta_2(Z_1, \tilde{Z}_1) \leqslant c(\delta, \tilde{m}_\delta, p)\rho^{\delta/(2+\delta)}. \tag{19.1.14}$$

Proof. For r.v.s X, Y with $\mathbb{E}(X - Y) = 0$ the following holds

$$\begin{aligned}\zeta_2(X, Y) &\leqslant \int_{-\infty}^{\infty} |x||F_X(x) - F_Y(x)|\,dx \\ &\leqslant N^2\rho(X, Y) + \tfrac{1}{2}\mathbb{E}X^2 I\{|X| > N\} + \tfrac{1}{2}\mathbb{E}Y^2 I\{|Y| > N\} \\ &\leqslant N^2\rho(X, Y) + \tfrac{1}{2}N^{-\delta}(\mathbb{E}|X|^{2+\delta} + \mathbb{E}|Y|^{2+\delta}).\end{aligned}$$

Minimizing the right-hand side over $N > 0$ we get (19.1.14).

Combining Claims 2, 3 and 4 we get $\rho(\sum_{i=1}^{n} \zeta_i^p, \sum_{i=1}^{n} \tilde{\zeta}_i^p) \leqslant c(\delta, \tilde{m}_\delta, p)^{\delta/3(2+\delta)}$ if $p/n < 1$. From Claim 1 we then obtain

$$\rho(X_{k,n,p}, \tilde{X}_{k,n,p}) \leqslant \begin{cases} 2p\rho & \text{if } p \geqslant \dfrac{n}{2} \\ p\rho + c\rho^{\delta/3(2+\delta)} & \text{if } p \geqslant k, p < \dfrac{n}{2} \\ c\rho^{\delta/3(2+\delta)} & \text{if } p < k \end{cases} \tag{19.1.15}$$

which proves (19.1.8). QED

Remark 19.1.4. Claim 1 of the proof of Theorem 19.1.3 remains also true for the total variation metric σ (see (3.2.13)). But ρ seems to be the appropriate metric for this problem since ρ is related to the ideal metric ζ_2 of order 2, see (19.1.11), while the total variation metric is too 'strong' to be estimated from above by ζ_2 or any other ideal metric of order 2.

Open problem 19.1.1. (*Topological structure of the metric space* ($\mathscr{F}(\mathbb{R})$, μ) *of d.f.s where μ is an ideal metric of order $r > 1$*). Consider the space $\mathfrak{X}_r(X_0)$, $r > 1$, of all random variables X such that $\mathbb{E}X^j = \mathbb{E}X_0^j$, $j = 0, 1, \ldots, [r]$ and $\mathbb{E}|X|^r < \infty$. Let μ be an ideal metric of order $r > 1$ in $\mathfrak{X}(X_0)$, i.e., μ is a simple metric, and for any X, Y and $Z \in \mathfrak{X}_r(X_0)$ (Z is independent of X and Y) and any $c \in \mathbb{R}$,

$$\mu(cX + Z, cY + Z) \leqslant |c|^r\mu(X, Y)$$

(see Remark 18.3.6). What is the topological structure of the space of laws of $X \in \mathfrak{X}_r(X_0)$ endowed with the metric μ?

Theorem 19.1.3 implies the following result on qualitative stability (cf. (19.1.6) with $\mu_1 = \rho$)

$$\zeta_1 \xrightarrow{w} \zeta_1 \quad m_\delta < \infty \Rightarrow \tilde{X}_{k,n} \xrightarrow{w} X_{k,n}.$$

For the stability in the opposite direction we prove the following result (cf. (19.1.7) with $\mu_2 = \rho$).

Theorem 19.1.4. For any $0 < p < \infty$ and any i.i.d. sequences $\{\zeta_i\}$, $\{\tilde{\zeta}_i\}$ with $\mathbb{E}\zeta_1^p = \mathbb{E}\tilde{\zeta}_1^p = 1$ and $\zeta_1, \tilde{\zeta}_1$ having continuous distribution functions, the following holds

$$\rho(\zeta_1, \tilde{\zeta}_1) \leqslant \sup_{k,n} \rho(X_{k,n,p}, \tilde{X}_{k,n,p}). \tag{19.1.16}$$

Proof. Denote $X_i = \zeta_i^p$, $\tilde{X}_i = \tilde{\zeta}_i^p$. Then

$$\begin{aligned}
\sup_{k,n} \rho\left(\sum_{i=1}^k \zeta_i^p \Big/ \sum_{i=1}^n \zeta_i^p, \sum_{i=1}^k \tilde{\zeta}_i^p \Big/ \sum_{i=1}^n \tilde{\zeta}_i^p\right) &\geqslant \sup_n \rho\left(\frac{X_1}{\sum_{i=1}^n X_i}, \frac{\tilde{X}_1}{\sum_{i=1}^n \tilde{X}_i}\right) \\
&= \sup_n \rho\left(\frac{X_1}{\frac{1}{n}\sum_{i=2}^{n+1} X_i}, \frac{\tilde{X}_1}{\frac{1}{n}\sum_{i=2}^{n+1} \tilde{X}_i}\right) \\
&\geqslant \rho(X_1, \tilde{X}_1) - \overline{\lim_{n\to\infty}}\, \rho\left(\frac{X_1}{\frac{1}{n}\sum_{i=2}^{n+1} X_i}, X_1\right) \\
&\quad - \overline{\lim_{n\to\infty}}\, \rho\left(\frac{\tilde{X}_1}{\frac{1}{n}\sum_{i=2}^{n+1} \tilde{X}_1}, \tilde{X}_1\right).
\end{aligned}$$

By the SLLN and the assumption $\mathbb{E}X_1 = \mathbb{E}\tilde{X} = 1$

$$\frac{X_1}{\frac{1}{n}\sum_{i=2}^{n+1} X_i} \to X_1 \text{ a.s. and } \frac{\tilde{X}_1}{\frac{1}{n}\sum_{i=2}^{n+1} \tilde{X}_i} \to \tilde{X}_1 \text{ a.s.} \tag{19.1.17}$$

Since X_1 and $\tilde{X}_1$ have continuous d.f.s the convergence in (19.1.17) is valid w.r.t. the uniform metric ρ. Hence $\sup_{k,n} \rho(X_{k,n,p}, \tilde{X}_{k,n,p}) \geqslant \rho(X_1, \tilde{X}_1) = \rho(\zeta_1, \tilde{\zeta}_1)$ as required. QED

Next we would like to prove similar results for the case $p = \infty$ and

$\tilde{X}_{k,n,\infty} = \bigvee_{i=1}^{n} \tilde{\zeta}_i / \bigvee_{i=1}^{n} \tilde{\zeta}_i$. In this case the structure of the ideal metric is totally different. Instead of ζ_2 (an ideal metric for the summation scheme) we shall explore the weighted Kolmogorov metrics ρ_r, $r > 0$ (cf. (18.1.4)) which are ideal for the maxima scheme.

We shall use the following condition.

Condition 1. There exists a non-decreasing continuous function $\phi(t) = \phi_{\tilde{\zeta}_1}(t)$: $[0, 1] \to [0, \infty)$, $\phi(0) = 0$ and such that

$$\phi(t) \geqslant \sup_{1-t \leqslant x \leqslant 1} (-\log x)^{-1} |F_{\tilde{\zeta}_1}(x) - x|.$$

Obviously *Condition 1* is satisfied for $\tilde{\zeta}_1 \stackrel{d}{=} \zeta_1$, uniformly distributed on $[0, 1]$. Let $\psi(t) = -\log(1-t)\phi(t)$ and let ψ^{-1} be the inverse of ψ.

Theorem 19.1.5. (i) If *Condition 1* holds and if $F_{\tilde{\zeta}_1}(1) = 1$, then

$$\Delta := \sup_{k,n} \rho(X_{k,n,\infty}, \tilde{X}_{k,n,\infty}) \leqslant c(\phi \circ \psi^{-1}(\rho))^{1/2} \quad \text{where } \rho := \rho(\zeta_1, \tilde{\zeta}_1).$$

(ii) If $\tilde{\zeta}_1$ has a continuous d.f., then $\Delta \geqslant \rho$.

Proof. (i) *Claim 1.* For any $1 \leqslant k \leqslant n$ the following holds:

$$\rho(X_{k,n,\infty}, \tilde{X}_{k,n,\infty}) \leqslant \rho\left(\bigvee_{i=1}^{k} \zeta_i, \bigvee_{i=1}^{k} \tilde{\zeta}_i\right) + \rho\left(\bigvee_{i=k+1}^{n} \zeta_i, \bigvee_{i=k+1}^{n} \tilde{\zeta}_i\right). \tag{19.1.18}$$

Proof. We use the representation

$$X_{k,n,\infty} = \frac{X_1}{X_1 \vee X_2} \quad \tilde{X}_{k,n,\infty} = \frac{\tilde{X}_1}{\tilde{X}_1 + \tilde{X}_2}$$

where

$$X_1 = \bigvee_{i=1}^{k} \zeta_i \quad X_2 = \bigvee_{i=k+1}^{n} \zeta_i \quad \tilde{X}_1 = \bigvee_{i=1}^{k} \tilde{\zeta}_i \quad \tilde{X}_2 = \bigvee_{i=k+1}^{n} \tilde{\zeta}_i.$$

Following the proof of (19.1.9), since ρ is a simple metric, we may assume (X_1, X_2) is independent of $(\tilde{X}_1, \tilde{X}_2)$. Thus by the regularity of the uniform metric and its invariance w.r.t monotone transformations we get

$$\begin{aligned}\rho(X_{k,n,\infty}, \tilde{X}_{k,n,\infty}) &= \rho\left(\frac{X_1}{X_1 \vee X_2}, \frac{\tilde{X}_1}{\tilde{X}_1 \vee \tilde{X}_2}\right) = \rho\left(1 \vee \frac{X_2}{X_1}, 1 \vee \frac{\tilde{X}_2}{\tilde{X}_1}\right) \\ &\leqslant \rho\left(\frac{X_2}{X_1}, \frac{\tilde{X}_2}{\tilde{X}_1}\right) \leqslant \rho\left(\frac{X_2}{X_1}, \frac{\tilde{X}_2}{X_1}\right) + \rho\left(\frac{\tilde{X}_2}{X_1}, \frac{\tilde{X}_2}{\tilde{X}_1}\right) \\ &\leqslant \rho(X_2, \tilde{X}_2) + \rho(X_1, \tilde{X}_1)\end{aligned}$$

the last inequality following by taking conditional expectations.

Claim 2. Let

$$\rho_* = \rho_*(\zeta_1, \tilde{\zeta}_1) := \sup_{0 \leqslant x \leqslant 1} (-\log x)^{-1} |F_{\zeta_1}(x) - F_{\tilde{\zeta}_1}(x)|.$$

(ρ_* plays the role of 'ideal metric' for our problem.) Then

$$\rho\left(\bigvee_{i=1}^{n} \zeta_i, \bigvee_{i=1}^{n} \tilde{\zeta}\right) \leqslant c\sqrt{\rho_*}. \tag{19.1.19}$$

Proof. Consider the transformation $f(t) = (-\log t)^{-1/\alpha}$ $(0 < t < 1)$. Then

$$\rho\left(\bigvee_{i=1}^{n} \zeta_i, \bigvee_{i=1}^{n} \tilde{\zeta}_i\right) = \rho\left(f\left(\bigvee_{i=1}^{n} \zeta_i\right), f\left(\bigvee_{i=1}^{n} \tilde{\zeta}\right)\right) = \rho\left(\bigvee_{i=1}^{n} X_i, \bigvee_{i=1}^{n} \tilde{X}_i\right) \tag{19.1.20}$$

where $X_i = f(\zeta_i)$, $\tilde{X}_i = f(\tilde{\zeta}_i)$. Since X_1 has extreme value distribution with parameter α, so does $Z_n := n^{-1/\alpha} \bigvee_{i=1}^{n} X_i$. The density of Z_n is given by

$$F_{Z_n}(x) = \frac{\mathrm{d}}{\mathrm{d}x} \exp(-x^{-\alpha}) = \alpha x^{-\alpha-1} \exp(-x^{-\alpha})$$

and thus

$$C_n := \sup_{x>0} f_{Z_n}(x) = \alpha\left(\frac{\alpha+1}{\alpha}\right)^{\alpha+1/\alpha} \exp\left(-\frac{\alpha+1}{\alpha}\right). \tag{19.1.21}$$

Let ρ_α be the *weighted Kolmogorov metric*

$$\rho_\alpha(X, Y) = \sup_{x>0} x^\alpha |F_X(x) - F_Y(x)| \tag{19.1.22}$$

see Lemma 18.1.2. Then by (18.2.73), Lemma 18.2.4

$$\rho(X, Y) \leqslant \Lambda_\alpha A^{\alpha/(1+\alpha)} \rho_\alpha^{1/(1+\alpha)}(X, Y) \tag{19.1.23}$$

where $\Lambda_\alpha := (1+\alpha)\alpha^{-\alpha(1+\alpha)}$ and $A := \sup_{x>0} F'_Y(x)$ (the existence of the density being assumed). Hence, by (19.1.20), (19.1.21) and (19.1.23),

$$\rho\left(\bigvee_{i=1}^{n} \zeta_i, \bigvee_{i=1}^{n} \tilde{\zeta}_i\right) = \rho(Z_n, \tilde{Z}_n) \leqslant \Lambda_\alpha C_n^{\alpha/(1+\alpha)} \rho_\alpha^{1/(1+\alpha)}(Z_n, \tilde{Z}_n) \tag{19.1.24}$$

where $\tilde{Z}_n = n^{-1/\alpha} \bigvee_{i=1}^{n} \tilde{X}_i$. The metric ρ_α is an ideal metric of order α w.r.t. the maxima scheme for i.i.d. r.v.s (see Lemma 18.1.2) and in particular

$$\rho_\alpha(Z_n, \tilde{Z}_n) \leqslant \rho_\alpha(X_1, \tilde{X}_1) = \rho_*(\zeta_1, \tilde{\zeta}_1). \tag{19.1.25}$$

From *Condition 1* we now obtain

Claim 3. $\rho_* \leqslant \phi \circ \psi^{-1}(\rho)$.

Proof. For any $0 \leqslant t \leqslant 1$ the following holds:

$$\rho_* = \max\left\{\sup_{0\leqslant x\leqslant 1-\varepsilon}(-\log x)^{-1}|F_{\zeta_1}(x) - x|,\ \sup_{1-\varepsilon\leqslant x\leqslant 1}(-\log x)^{-1}|F_{\zeta_1}(x) - x|\right\}$$

$$\leqslant \max((-\log(1-\varepsilon))^{-1}\rho,\ \phi(\varepsilon)). \tag{19.1.26}$$

Choosing ε by $\phi(\varepsilon) = (-\log(1-\varepsilon))^{-1}\rho$, i.e., $\rho = \psi(\varepsilon)$, one obtains the claim.

From Claims 1, 2 and 3 we obtain

$$\boldsymbol{\rho}(X_{k,n,p}, \tilde{X}_{k,n,p}) \leqslant \min(n\rho,\ c(\phi\circ\psi^{-1}(\rho))^{1/2}) \tag{19.1.27}$$

which proves (i).

(ii) For the proof of (ii) observe that $F_{\bigvee_{i=1}^n \tilde{\zeta}_i}(x) = F_{\tilde{\zeta}_i}^n \to 1$ for any x with $F_{\tilde{\zeta}_1}(x) > 0$. As in the proof of Theorem 19.1.4 we then obtain

$$\sup_{k,n}\boldsymbol{\rho}\left(\frac{\bigvee_{i=1}^k \zeta_i}{\bigvee_{i=1}^n \zeta_i}, \frac{\bigvee_{i=1}^k \tilde{\zeta}_i}{\bigvee_{i=1}^n \tilde{\zeta}_i}\right) \geqslant \limsup_n \boldsymbol{\rho}\left(\frac{\zeta_1}{\bigvee_{i=1}^n \zeta_i}, \frac{\tilde{\zeta}_1}{\bigvee_{i=1}^n \tilde{\zeta}_i}\right)$$

$$\geqslant \boldsymbol{\rho}(\zeta_1, \tilde{\zeta}_1) - \overline{\lim_n}\,\boldsymbol{\rho}\left(\frac{\zeta_1}{\bigvee_{i=1}^n \zeta_i}, \zeta_1\right) - \overline{\lim_n}\,\boldsymbol{\rho}\left(\frac{\tilde{\zeta}_1}{\bigvee_{i=1}^n \tilde{\zeta}_i}, \tilde{\zeta}_1\right) = \boldsymbol{\rho}(\zeta_1, \tilde{\zeta}_1)$$

since ζ_1, $\tilde{\zeta}_1$ have continuous d.f.s. QED

Remark 19.1.5. In Theorem 19.1.5(i) the constant c depends on $\alpha > 0$ (see (19.1.23)). Thus, one can optimize c by choosing α appropriately in (19.1.22).

19.2 STABILITY IN DE FINETTI'S THEOREM

In this section we apply the characterization of distributions F_p $(0 < p \leqslant \infty)$ (cf. (19.1.2), Theorems 19.1.1 and 19.1.2) to show that the uniform distribution on the 'positive p-sphere' $S_{p,n}$,

$$\begin{aligned} S_{p,n} &:= \left\{x = (x_1, \dots, x_n) \in \mathbb{R}_+^n : \sum_{i=1}^n x_i^p = n\right\}, \\ S_{\infty,n} &:= \left\{x \in \mathbb{R}_+^n : \bigvee_{i=1}^n x_i = n\right\} \end{aligned} \tag{19.2.1}$$

has approximately independent F_p-distributed components. This will lead us to the stability of the following de Finetti's type theorem: Let $\zeta = (\zeta_1, \dots, \zeta_n)$

be non-negative r.v.s and $C_{n,p}$ be the class of ζ-laws with the property that given $\sum_{i=1}^n \zeta_i^p = s$ (for $p = \infty$ given $\bigvee_{i=1}^n \zeta_i = s$), the conditional distribution of ζ is uniform on $S_{p,n}$. Then the joint distribution of i.i.d. ζ_i with common F_p-distribution is in the class $C_{n,p}$. Moreover, if $P \in \mathscr{P}(\mathbb{R}_+^\infty)$ and for any $n \geq 1$ the projection $T_{1,2,\ldots,n}P$ on the first n-coordinates belongs to C_n then p is a mixture of i.i.d F_p-distributed random variables (*de Finetti's theorem*).

The de Finetti theorem will follow from the following stability theorem: if n non-negative r.v.s ζ_i are conditionally uniform on $S_{p,n}$ given $\sum_{i=1}^n \zeta_i^p = s$ (resp. $\bigvee_{i=1}^n \zeta_i = s$ for $p = \infty$) then the total variation metric $\boldsymbol{\sigma}$ between the law of $(\zeta_1, \ldots, \zeta_k)$ (k fix, n large enough) and a mixture of i.i.d. F_p-distributed r.v. $(\tilde{\zeta}_1 \ldots, \tilde{\zeta}_k)$ is less than const $\cdot\, k/n$.

Remark 19.2.1. An excellent survey on de Finetti's theorem is given by Diaconis and Freedman (1987a), where the cases $p = 1$ and $p = 2$ have been considered in detail.

We start with another characterization of the exponential class of distributions F_p (cf. Theorem 19.1.1).

Let

$$S_{p,s,n} := \left\{x \in \mathbb{R}_+^n : \sum_{i=1}^n x_i^p = s\right\}$$

denote the p-sphere of radius s in $\mathbb{R}_+^n$, $0 < p < \infty$. (The next two lemmas are simple applications of the well known formulae for conditional distributions.)

Lemma 19.2.1. Let $\zeta_1, \ldots, \zeta_n$ be i.i.d. r.v.s with common d.f. F_p where $0 < p < \infty$. Then the conditional distribution of $(\zeta_1, \ldots, \zeta_n)$ given $\sum_{i=1}^n \zeta_i^p = s$, denoted by

$$P_{s,p} := P_{(\zeta_1, \ldots, \zeta_n) | \sum_{i=1}^n \zeta_i^p = s}$$

is uniform on $S_{p,s,n}$.

Similarly, we examine the case $p = \infty$; let $\zeta_1, \ldots, \zeta_n$ be i.i.d. F_∞-distributed (recall that F_∞ is the (0, 1)-uniform distribution). Denote the conditional distribution of $(\zeta_1, \ldots, \zeta_n)$ given $\bigvee_{i=1}^n \zeta_i = s$ by $P_{s,\infty} := \Pr_{(\zeta_1, \ldots, \zeta_n) | \bigvee_{i=1}^n \zeta_i = s}$.

Lemma 19.2.2. $P_{s,\infty}$ is uniform on $S_{\infty,s,n} := \{x \in \mathbb{R}_+^n : \bigvee_{i=1}^n x_i = s\}$ for almost all $s \in [0, 1]$.

Now, using the above lemma, we can prove a stability theorem related to de Finetti's theorem for $p = \infty$.

Let $P_\sigma^{n,\infty}$ for $\sigma > 0$ be the law of $(\sigma\zeta_1, \ldots, \sigma\zeta_n)$ and let $Q_{n,s,k}^{(\infty)}$ be the law of $(\eta_1, \ldots, \eta_k)$ where $\eta = (\eta_1, \ldots, \eta_n)$ ($n > k$) is uniform on $S_{\infty,s,n}$. In the next theorem we evaluate the deviation between $Q_{n,s,k}^{(\infty)}$ and $P_s^{k,\infty}$ in terms of the *total variation metric*

$$\boldsymbol{\sigma}(Q_{nsk}^{(\infty)}, P_s^{k,\infty}) := \sup_{A \in \mathfrak{B}^k} |Q_{nsk}^{(\infty)}(A) - P_s^{k,\infty}(A)|$$

$\mathfrak{B}^k$ being the σ-algebra of Borel sets in $\mathbb{R}^k$.

Theorem 19.2.1. For any $s > 0$ and $0 < k \leqslant n$

$$\boldsymbol{\sigma}(Q_{nsk}^{(\infty)}, P_s^{k,\infty}) = k/n. \tag{19.2.2}$$

Proof. We need the following invariant property of the total variation metric $\boldsymbol{\sigma}$.

Claim 1 (Sufficiency theorem). If T: $\mathbb{R}^n \to \mathbb{R}$ is a sufficient statistic for P, $Q \in \mathscr{P}(\mathbb{R}^n)$ then

$$\boldsymbol{\sigma}(P, Q) = \boldsymbol{\sigma}(P \circ T^{-1}, Q \circ T^{-1}). \tag{19.2.3}$$

Proof. Take $\mu = \frac{1}{2}(P + Q)$ and let $f := \mathrm{d}P/\mathrm{d}\mu$, $g := \mathrm{d}Q/\mathrm{d}\mu$. Since T is sufficient then $f = h_1 \circ T$, $g = h_2 \circ T$ and

$$h_1 = \frac{\mathrm{d}P \circ T^{-1}}{\mathrm{d}\mu \circ T^{-1}} \quad h_2 := \frac{\mathrm{d}Q \circ T^{-1}}{\mathrm{d}\mu \circ T^{-1}}.$$

Clearly, $\boldsymbol{\sigma}(P \circ T^{-1}, Q \circ T^{-1}) \leqslant \boldsymbol{\sigma}(P, Q)$. On the other hand,

$$\begin{aligned}
\boldsymbol{\sigma}(P, Q) &= \sup_{A \in \mathfrak{B}^k} \left| \int_A (h_1 \circ T - h_2 \circ T)\, \mathrm{d}\mu \right| \\
&\leqslant \sup_{A \in \mathfrak{B}^k} \left| \int_{T \circ A} (h_1 - h_2)\, \mathrm{d}\mu \circ T^{-1} \right| \\
&= \sup_{A \in \mathfrak{B}^k} \left| \int_{T \circ A} \left(\frac{\mathrm{d}P \circ T^{-1}}{\mathrm{d}\mu \circ T^{-1}} - \frac{\mathrm{d}Q \circ T^{-1}}{\mathrm{d}\mu \circ T^{-1}} \right) \mathrm{d}\mu \circ T^{-1} \right| \\
&\leqslant \boldsymbol{\sigma}(P \circ T^{-1}, Q \circ T^{-1})
\end{aligned}$$

which proves the claim.

Further, without loss of generality we may assume $s = 1$ since

$$\boldsymbol{\sigma}(Q_{nsk}^{(\infty)}, P_s^{k,\infty}) = \boldsymbol{\sigma}(\mathrm{Pr}_{(\eta_1, \ldots, \eta_k)/\bigvee_{i=1}^n \eta_i = s}, \mathrm{Pr}_{(s\zeta_1, \ldots, s\zeta_k)}) = \boldsymbol{\sigma}(Q_{n1k}^{\infty}, P_1^{k,\infty})$$

by the zero-order ideality of $\boldsymbol{\sigma}$ (cf. Definition 14.2.1). Let $\tilde{Q}$ be the law of $\eta_1 \vee \cdots \vee \eta_k$ determined by $Q_{n1k}^{(\infty)}$, the distribution of $(\eta_1, \ldots, \eta_k)$ where the vector $\eta = (\eta_1, \ldots, \eta_k, \eta_{k+1}, \ldots, \eta_n)$ is uniformly distributed on the simplex $S_{\infty,1,n} = \{x \in \mathbb{R}_+^n : \bigvee_{i=1}^{\infty} x_i = 1\}$. Let $\tilde{P}$ be the law of $\zeta_1 \vee \cdots \vee \zeta_k$ where ζ_is are i.i.d. uniforms. Then with $\gamma_{k,n,\infty} = \bigvee_{i=1}^k \zeta_i / \bigvee_{i=1}^n \zeta_i$, $\tilde{Q} = \mathrm{Pr}_{\gamma_{k,n,\infty}}$ and $\tilde{Q}$ has a d.f.

given by (19.1.3). On the other hand $\tilde{P}((-\infty, x]) = x^k$, $0 \leqslant x \leqslant 1$. Hence

$$\tilde{Q} = \frac{n-k}{n}\tilde{P} + \frac{k}{n}\delta_1$$

is the mixture of $\tilde{P}$ and δ_1, the point measure at 1. Consider the total variation distance

$$\sigma(Q^{\infty}_{n,1,k}, P^{k,\infty}_1) = \sup_{A \in \mathfrak{B}^k}\left|\Pr\left((\eta_1, \ldots, \eta_k) \in A \,\middle|\, \bigvee_{i=1}^n \eta_i = 1\right) - \Pr((\zeta_1, \ldots, \zeta_k) \in A)\right|.$$

We realize $Q_{n,1,k}$ as the law of $\zeta_1/M, \ldots, \zeta_k/M$ where $M = \bigvee_{i=1}^n \zeta_i$, so $\tilde{Q}$ is the law of $\max(\zeta_1/M, \ldots, \zeta_k/M)$. By Claim 1,

$$\sigma(Q^{\infty}_{n,1,k}, P^{k,\infty}_1) = \sigma(\tilde{Q}, \tilde{P}) = \sup_{A \in \mathfrak{B}^k}\left|\frac{n-k}{n}\tilde{P}(A) + \frac{k}{n}\delta_1(A) - \tilde{P}(A)\right|$$

$$= \frac{k}{n}\sup_{A \in \mathfrak{B}^k}|\delta_1(A) - \tilde{P}(A)| = \frac{k}{n}$$

as required. QED

Let C_n be the class of distributions of $X = (X_1, \ldots, X_n)$ on $\mathbb{R}^n_+$ which share with the i.i.d. uniforms (see Lemma 19.2.1) the property that given $M := \bigvee_{i=1}^n X_i = s$, the conditional joint distribution of X is uniform on $S_{\infty,s,n}$. Clearly $P^{n,\infty}_s \in C_n$. As a consequence of Theorem 19.2.1 we get the *stability form of de Finetti's theorem:*

Corollary 19.2.1. If $P \in C_n$, then there is a μ such that for all $k < n$

$$\|P_k - P_{\mu k}\| \leqslant k/n \tag{19.2.4}$$

where P_k is the P-law of the first k-coordinates $(X_1, \ldots, X_k)$ and $P_{\mu k} := \int P^{k,\infty}_\sigma \, \mu(d\sigma)$.

Proof. Define $\mu = \Pr_{\bigvee_{i=1}^n X_i}$, then $P_k = \int Q^{(\infty)}_{n,s,k}\,\mu(ds)$, $P_{\mu k} = \int P^{k,\infty}_s\,\mu(ds)$ and, therefore, $\sigma(P_k, P_{\mu k}) \leqslant \int \sigma(Q^{(\infty)}_{nsk}, P^{k,\infty}_s)\,\mu(ds) = k/n$. QED

In particular, one gets the de Finetti's type characterization of scale mixtures of i.i.d. uniform variables.

Corollary 19.2.2. Let P be a probability on $\mathbb{R}^\infty_+$ with P_n being the P-law of the first n coordinates. Then P is a uniform scale mixture of i.i.d. uniform distributed r.v.s, if and only if $P_n \in C_n$ for every n.

Following the same method we shall consider the case $p \in (0, \infty)$. Let $\zeta_1, \zeta_2, \ldots$ be i.i.d r.v.s with d.f. F_p given by Theorem 19.1.1. Then by Lemma

19.2.1 the conditional distribution of $(\zeta_1, \ldots, \zeta_n)$ given $\sum_{i=1}^n \zeta_i^p = s$ is $Q_{n,s,k}^{(p)}$ where $Q_{n,s,k}^{(p)}$ is the distribution of the first k coordinates of a random vector $(\eta_1, \ldots, \eta_n)$ uniformly distributed on the p-sphere of radius s, denoted by $S_{p,s,n}$. Let $P_\sigma^{n,p}$ be the law of the vector $(\sigma\zeta_1, \ldots, \sigma\zeta_n)$. The next result shows that $Q_{n,s,k}^{(p)}$ is close to $P_{(s/n)^{1/p}}^{k,p}$ w.r.t. the total variation metric.

Theorem 19.2.2. Let $0 < p < \infty$; then for $k < n - p$ and k, n big enough,

$$\sigma(Q_{n,s,k}^{(p)}, P_{(s/n)^{1/p}}^{k,p}) \leqslant \text{const} \cdot k/n. \tag{19.2.5}$$

Proof. By the zero-ideality of σ,

$$\begin{aligned}
\sigma(Q_{n,s,k}^{(p)}, P_{(s/n)^{1/p}}^{k,p}) &= \sup_{A \in \mathfrak{B}^k} |\Pr(\eta_1, \ldots, \eta_k) \in (A/\eta_1^p + \cdots + \eta_n^p = s) \\
&\quad - \Pr(((s/n)^{1/p}\zeta_1, \ldots, (s/n)^{1/p}\zeta_k) \in A)| \\
&= \sup_{A \in \mathfrak{B}^k} \left| \Pr(((n/s)^{1/p}\eta_1, \ldots, (n/s)^{1/p}\eta_k) \in A) \Big/ \sum_{i=1}^n ((n/s)^{1/p}\eta_i)^p = n) \right. \\
&\quad \left. - \Pr((\zeta_1, \ldots, \zeta_k) \in A) \right| = \sigma(Q_{n,n,k}^{(p)}, P_1^{k,p}).
\end{aligned}$$

Thus it suffices to take $s = n$. Let $\tilde{Q}_k$ be the $Q_{n,n,k}^{(p)}$-law of $\eta_1^p + \cdots + \eta_k^p$ and $\tilde{P}_k$ be the $P_1^{k,p}$-law of $\zeta_1^p + \cdots + \zeta_k^p$. Then $\sigma(Q_{n,n,k}^{(p)}, P_1^{k,p}) = \sigma(\tilde{Q}_k, \tilde{P}_k)$ as in the proof of Theorem 19.1.1. By Lemma 19.1.1 we may consider $Q_{n,n,k}$ as the law of $(\zeta_1/R, \ldots, \zeta_k/R)$, where $R^p := (1/n)\sum_{i=1}^n \zeta_i^p$. So $\tilde{Q}_k$ is the law of $\sum_{i=1}^k (\zeta_i/R)^p = n\sum_{i=1}^k \zeta_i^p/(\sum_{i=1}^n \zeta_i^p)$. Thus, as in the proof of Theorem 19.1.1 $\tilde{Q}_k$ has a density

$$f(x) = \frac{1}{n} B\left(\frac{k}{p}, \frac{n-k}{p}\right)\left(\frac{x}{n}\right)^{(k/p)-1}\left(1 - \frac{x}{n}\right)^{-1+(n-k)/p}$$

$$B\left(\frac{k}{p}, \frac{n-k}{p}\right) := \frac{\Gamma\left(\frac{n}{p}\right)}{\Gamma\left(\frac{k}{p}\right)\Gamma\left(\frac{n-k}{p}\right)} \tag{19.2.6}$$

for $0 \leqslant x \leqslant n$ and $f(x) = 0$ for $x > n$. On the other hand, $\tilde{P}_k$ has a gamma $(1/p, k/p)$-density

$$g(x) := \frac{1}{p^{k/p}\Gamma(k/p)} \exp(-x/p)x^{-1+k/p} \quad \text{for } 0 \leqslant x \leqslant \infty. \tag{19.2.7}$$

By Scheffé's theorem (see Billingsley, 1968)

$$\sigma(\tilde{Q}_k, \tilde{P}_k) = \int_0^\infty |f(x) - g(x)|\, dx$$

$$= 2\int_0^\infty \max(0, f(x) - g(x))\, dx = \int_0^\infty \max\left(0, \frac{f(x)}{g(x)} - 1\right) g(x)\, dx. \tag{19.2.8}$$

By (19.2.6) and (19.2.7) $f/g = Ah$, where

$$A = \left(\frac{p}{n}\right)^{k/p} \Gamma\left(\frac{n}{p}\right) \Big/ \Gamma\left(\frac{n-k}{p}\right)$$

and

$$h(x) = \exp\left(\frac{x}{p}\right)\left(1 - \frac{x}{n}\right)^{-1+(n-k)/p}$$

for $x \in [0, n]$ and $h(x) = 0$ for $x > n$. We have

$$\log h(x) = \frac{x}{p} + (-1 + (n-k)/p)\log\left(1 - \frac{x}{n}\right)$$

and

$$\frac{\partial}{\partial x} \log h(x) \geqslant 0$$

if and only if $x \leqslant k + p$. Hence, if $k + p \leqslant n$

$$\log h(x) \leqslant \frac{k+p}{p} + \left(\frac{n-k}{p} - 1\right)\log\left(1 - \frac{k+p}{n}\right). \tag{19.2.9}$$

We use the following consequence of the Stirling expansion of the gamma function (cf. Abramowitz and Stegun 1970, p. 257).

$$\Gamma(x) = \exp(-x)x^{x-1/2}(2\pi)^{1/2}\exp(\theta/12x) \quad 0 \leqslant \theta < 1. \tag{19.2.10}$$

This imples that

$$A = \left(\frac{n}{n-k}\right)^{(n-k)/p+1/2} \exp\left(-\frac{k}{p}\right)\tilde{\theta}$$

with

$$\tilde{\theta} = \exp\left[\frac{p}{12}\left(\frac{\theta_1}{n} - \frac{\theta_2}{n-k}\right)\right] \leqslant \exp\left(\frac{p}{12n}\right)$$

and $0 \leqslant \theta_i < 1$. Hence,

$$\begin{aligned}
Ah &\leqslant e\left(\frac{n}{n-k}\right)^{(n-k)/p+1/2}\left(\frac{n-k-p}{n}\right)^{(n-k)/p-1}\tilde{\theta} \\
&= e\left(\frac{n-k-p}{n-k}\right)^{(n-k)/p}\frac{n}{n-k-p}\left(\frac{n}{n-k}\right)^{1/2}\tilde{\theta} \\
&= e\left(1-\frac{p}{n-k}\right)^{(n-k)/p}\frac{n}{n-k-p}\left(\frac{1}{1-k/n}\right)^{1/2}\tilde{\theta}.
\end{aligned}$$

We use the following estimate

$$\sup_{0\leqslant x<a}\left|\exp(-x)-\left(1-\frac{x}{a}\right)^a\right| \leqslant c/a \quad \text{with } c := \sup_{0\leqslant x<a} x\exp(-x) = 1/e \quad a>1 \tag{19.2.11}$$

implying that

$$\left|e\left(1-\frac{p}{n-k}\right)^{(n-k)/p}-1\right| \leqslant \frac{p}{n-k}.$$

Furthermore, we use the estimates

$$\left(1-\frac{k}{n}\right)^{-1/2} \leqslant 1+\frac{k}{2n} \quad \text{and} \quad \tilde{\theta} \leqslant \exp\left(\frac{p}{12n}\right) \leqslant 1+\frac{p}{12n}\exp(1/12)$$

to obtain

$$Ah \leqslant \left(1+\frac{p}{n-k}\right)\frac{n}{n-k-p}\left(1+\frac{k}{n}\right)\left(1+\frac{p\exp(1/12)}{12n}\right)$$

implying that $Ah - 1$ is bounded by the right-hand side of (19.2.5). QED

Analogous to Corollary 19.2.1 and 19.2.2 we can state de Finetti's theorem (and its stable version) for the class $C_{n,p}$ of distributions of $X_1, \ldots, X_n$ which share with i.i.d. F_p-distributed r.v.s $(\zeta_1, \ldots, \zeta_n)$ the property that given $\sum_{i=1}^n X_i^p = s$, the conditional joint distribution of X is uniform on the positive pth sphere $S_{p,s,n}$.

19.3 CHARACTERIZATION AND STABILITY OF ENVIRONMENTAL PROCESSES

The objective of this section is the study of four stochastic models which take into account the effect of erosion on the annual crop production. More precisely, we are concerned with the limit behavior of four recursive equations modelling environmental processes:

$$S_0 = 0, \quad S_n \stackrel{\mathrm{d}}{=} (Y + S_{n-1})Z \tag{19.3.1}$$

$$M_0 = 0 \quad M_n \stackrel{\mathrm{d}}{=} (Y \vee M_{n-1})Z \tag{19.3.2}$$

$$G \stackrel{\mathrm{d}}{=} (Y + \delta G)Z \tag{19.3.3}$$

and

$$H \stackrel{\mathrm{d}}{=} (Y \vee \delta H)Z \tag{19.3.4}$$

where the random variables in the right-hand sides of (19.3.1)–(19.3.4) are assumed to be independent. $Y, Z, S_{(\cdot)}, M_{(\cdot)}, G$ and H are random variables taking on values in the Banach space $\mathbb{B} = C(T)$ of continuous functions x on the compact set T with the usual supremum norm $\|x\|$. For any $x, y \in \mathbb{B}$ define the pointwise maximum and multiplication: $(x \vee y)(t) = x(t) \vee y(t)$ and $(x \cdot y)(t) = x(t) \cdot y(t)$. Z in (19.3.2) and (19.3.4) is assumed to be non-negative, i.e. $Z(t) \geqslant 0$ for all $t \in T$. Finally, $\delta = \delta(d)$ is a Bernoulli random variable independent of Y, G, H, Z with success probability d.

Equation (19.3.1) arises in modelling the total crop yield over n years. Namely, consider a set of crop-producing areas $A_t (t \in T)$ and denote by $\{Y_n(t)\}_{n \geqslant 1}$ the sequence of annual yields. For fixed n, the real-valued r.v.s $Y_n(t)$, $t \in T$ are dependent. Let $Z_n(t)$ be the proportion of crop yield maintained in year n after the environmental effect from the previous year: $Z_n(t) < 1$ corresponds to a 'bad' year, probably due to erosion, while $Z_n(t) \geqslant 1$ corresponds to a good year. The random variables Z_n are assumed to be i.i.d. and independent of $\{Y_n\}$. Assuming that the crop-growing area A_t is subject to environmental effects, the resulting sequence of annual yields is

$$X_n(t) = Y_n(t) \prod_{i=1}^{n} Z_i(t). \tag{19.3.5}$$

Let us denote by

$$S_n(t) = \sum_{k=1}^{n} X_k(t) \quad n \in \mathbb{N} \tag{19.3.6}$$

the total crop yield over n years. Then clearly the process S_n satisfies the recursive equation (19.3.1) where here and in the following Y and Z are generic independent random variables with $Y \stackrel{\mathrm{d}}{=} Y_1$ and $Z \stackrel{\mathrm{d}}{=} Z_1$, and independent of Y_is and Z_is.

Analogously, the maximal crop yield over n years

$$M_n = \bigvee_{k=1}^{n} X_k \tag{19.3.7}$$

has distribution determined by (19.3.2).

Next we consider the situation when each year a disastrous event may occur with probability $1 - d \in (0, 1)$. The year of the disaster is a geometric random variable $\tau = \tau(d)$, $\Pr(\tau(d) = k) = (1 - d)d^{k-1}$, $k \in \mathbb{N}$. Thus, the total crop yield until the disastrous year can be modelled by

$$\begin{aligned} G := S_\tau &= \sum_{k=1}^{\tau} X_k \overset{\mathrm{d}}{=} \sum_{k=1}^{1+\delta\tau} X_k \overset{\mathrm{d}}{=} X_1 + \delta \sum_{k=2}^{1+\tau} X_k \\ &\overset{\mathrm{d}}{=} Y_1 Z_1 + \delta \sum_{k=2}^{1+\tau} Y_k \prod_{i=1}^{k} Z_i \overset{\mathrm{d}}{=} YZ + \delta Z \sum_{k=2}^{1+\tau} Y_{k-1} \prod_{i=2}^{k} Z_i \overset{\mathrm{d}}{=} (Y + \delta G)Z \end{aligned} \tag{19.3.8}$$

i.e., G satisfies the recurrence (19.3.3). Analogously, the maximal crop yield until the year of the disaster

$$H := M_\tau = \bigvee_{k=1}^{\tau} X_k \tag{19.3.9}$$

satisfies (19.3.4).

Further, our goal is to prove that S_n has a limit S (a.s.) and S satisfies

$$S \overset{\mathrm{d}}{=} (Y + S)Z. \tag{19.3.10}$$

Similarly, the limit M of M_n in (19.3.2) satisfies

$$M \overset{\mathrm{d}}{=} (Y \vee M)Z. \tag{19.3.11}$$

The problem is to characterize the set of solutions of the equations (19.3.10), (19.3.11), (19.3.3) and (19.3.4). Since the general solution seems to be difficult to obtain we shall use appropriate approximations and will evaluate the error involved in these approximations.

Following the main idea of this book that each approximation problem has 'natural' ('suitable', 'ideal') metrics in terms of which the problem can be solved easily and completely, we choose $\mathscr{L}_p$-metric and its minimal ℓ_p for our approximation problem. Recall that $\mathfrak{X}(\mathbb{B})$ is the set of all random elements on a non-atomic probability space $\{\Omega, \mathscr{A}, \Pr\}$ with values in $\mathbb{B}$ and

$$\mathscr{L}_p(X, Y) := \begin{cases} (\mathbb{E}\|X - Y\|^p)^{p'} & \text{if } 0 < \infty, p' = \min(1, p^{-1}) \\ \Pr\{X \neq Y\} & \text{if } p = 0 \\ \operatorname{ess\,sup}\|X - Y\| & \text{if } p = \infty, X, Y \in \mathfrak{X}(\mathbb{B}). \end{cases} \tag{19.3.12}$$

The corresponding minimal (simple) metric $\ell_p(X, Y) = \ell_p(\Pr_X, \Pr_Y)$ is given by

$$\mathscr{L}_p(X, Y) = \inf\{\mathscr{L}_p(\tilde{X}, \tilde{Y});\ \tilde{X}, \tilde{Y} \in \mathfrak{X}(B), \tilde{X} \overset{\mathrm{d}}{=} X, \tilde{Y} \overset{\mathrm{d}}{=} Y\}. \tag{19.3.13}$$

The basic properties of ℓ_p-metrics were summarized in Section 18, see (18.2.9)–(18.2.18).

In what follows we shall need some analogs to the $\mathscr{L}_p$-metric in the space $\mathfrak{X}(B^\infty)$. The space B^∞ is a Banach space with usual supremum norm defined by $\|\bar{X}\| = \sup\{\|X_i\|: i \geqslant 1\}$ where $\bar{X} = (X_1, X_2, \ldots)$. Now, on $\mathfrak{X}(B^\infty)$ we consider the following metrics

$$\mathbf{K}(\bar{X}, \bar{Y}) = \inf\{\varepsilon > 0: \Pr(\|\bar{X} - \bar{Y}\| > \varepsilon) < \varepsilon\} \quad \text{(Ky Fan)} \tag{19.3.14}$$

$$\mathscr{L}_p(\bar{X}, \bar{Y}) = (\mathbb{E}\|\bar{X} - \bar{Y}\|^p)^{1 \wedge p^{-1}} \quad \text{for } 0 < p < \infty \tag{19.3.15}$$

$$\mathscr{L}_0(\bar{X}, \bar{Y}) = \Pr\{\bar{X} \neq \bar{Y}\} \text{ and } \mathscr{L}_\infty(\bar{X}, \bar{Y}) = \operatorname{ess\,sup}\|\bar{X} - \bar{Y}\|.$$

Clearly, if X_n and X are random elements in $\mathfrak{X}(B)$ then $X_n \to X(\Pr - \text{a.s.})$ if and only if $\mathbf{K}(X_n^*, X^*) \to 0$, where $X^* := (X_n, X_{n+1}, \ldots)$ and $X^* = (X, X, \ldots)$. Similar to the proof of Lemma 8.2.1, we have that if

$$\mathbb{E}\{\sup_{n \geqslant 1} \|X_n\|^p\} + \mathbb{E}\|\mathrm{X}\|^p < \infty$$

for some $p \in [1, \infty)$ then as $n \to \infty$

$$\mathscr{L}_p(X_n^*, X^*) \to 0 \text{ if and only if } X_n \to X \text{ (Pr-a.s.) and } \mathbb{E} \sup_{m \geqslant n} \|X_m\|^p \to \mathbb{E}\|X\|^p. \tag{19.3.16}$$

In both 'limit' cases $p = 0$, $p = \infty$,

$$\mathscr{L}_p(X_n^*, X^*) \to 0 \Rightarrow X_n \to X \text{ (Pr-a.s.)}. \tag{19.3.17}$$

Theorem 19.3.1. (a) *(Existence of the limit S)* Suppose that $\{Y_n\}_{n \geqslant \mathbb{N}} \subset \mathfrak{X}(B)$ is an i.i.d. sequence with $N_p(Y) < \infty$, where $0 \leqslant p \leqslant \infty$ and

$$N_p(Y) := \mathscr{L}_p(Y, 0) = \ell_p(Y, 0) = \begin{cases} [\mathbb{E}\|Y\|^p]^{\min(1, 1/p)} & 0 < p < \infty \\ \operatorname{ess\,sup}\|Y\| & p = \infty \\ \Pr(Y \neq 0), & p = 0. \end{cases} \tag{19.3.18}$$

Assume also that $\{Z_n\}_{n \in \mathbb{N}} \subset \mathfrak{X}(B)$ is an i.i.d. sequence independent of $\{Y_n\}_{n \in \mathbb{N}}$, such that $N_p(Z) < 1$. Given S_n by (19.3.6) there exists S such that $S_n \to S$ (Pr-a.s.). Moreover, S satisfies the equation (19.3.10) with Y, Z and S mutually independent.

(b) (*Rate of convergence of S_n to S*) Let $p \in [0, \infty]$, $N_p(Y) < \infty$ and $N_p(Z) < 1$. Assume that the laws of S_n and S are specified by the equations (19.3.1) and (19.3.10), respectively. Then

$$\ell_p(S_n, S) \leqslant N_p^n(Z) \frac{N_p(Y)}{1 - N_p(Z)}. \tag{19.3.19}$$

Proof. (a) For any $k, n \in \mathbb{N}$, $1 \leqslant p < \infty$,

$$\begin{aligned}\mathscr{L}_p(S_n^*, S_{n+k}^*) &= \mathscr{L}_p((S_n, S_{n+1}, \dots), (S_{n+k}, S_{n+k+1}, \dots)) \\ &= \left(\mathbb{E} \max_{m \geqslant n} \left\| \sum_{i=m}^{m+k} Y_i \prod_{j=1}^{i} Z_j \right\|^p\right)^{1/p} \\ &\leqslant \sum_{i \geqslant n} \left(\mathbb{E}\|Y_i\|^p \prod_{j=1}^{i} \|Z_j\|^p\right)^{1/p} \\ &= N_p(Y)(N_p(Z))^n/(1 - N_p(Z)).\end{aligned}$$

On the other hand, the space of all sequences $\bar{X}$ with $E\|\bar{X}\|^p < \infty$ is complete with respect to $\mathscr{L}_p$ and thus, $S^* = (S, S, \dots)$ exists. Finally, notice that $\mathscr{L}_p(S_n^*, S^*) \leqslant (N_p(Z))^n \ N_p(Y)/(1 - N_p(Z))$ holds. This proves the assertion for $1 \leqslant p < \infty$. The cases $0 \leqslant p < 1$ and $p = \infty$ are treated analogously. The equation (19.3.10) follows from $S_n \to S$ (Pr-a.s.) and (19.3.1)

(b) From (19.3.1), (19.3.10) we have for $0 < p < \infty$

$$\begin{aligned}\ell_p(S_n, S) &\leqslant \mathscr{L}_p((Y + S_{n-1})Z, (Y + S)Z) \leqslant \mathscr{L}_p(S_{n-1}Z, SZ) \\ &\leqslant \{\mathbb{E}\|S_{n-1} - S\|^p\|Z\|^p\}^{1 \wedge p^{-1}} = \mathscr{L}_p(S_{n-1}, S)N_p(Z) \quad (19.3.20)\end{aligned}$$

the last inequality follows from the independence of (S_{n-1}, S) and Z. Taking the minimum of the right-hand side of (19.3.20) over all the joint distributions of S_{n-1} and S we obtain

$$\ell_p(S_n, S) \leqslant \ell(S_{n-1}, S)N_p(Z). \quad (19.3.21)$$

Hence,

$$\ell_p(S_n, S) \leqslant \ell_p(0, S)N_p^n(Z) = N_p(S)N_p^n(Z). \quad (19.3.22)$$

From the Minkovski inequality

$$N_p(S) \leqslant N_p(Z)N_p(Y + S) \leqslant N_p(Z)\{N_p(Y) + N_p(S)\}$$

which implies that $N_p(S) \leqslant N_p(Y) \ \{1 - N_p(Z)\}^{-1}$. This and (19.3.22) prove (19.3.10). The cases $p = 0$ and $p = \infty$ can be handled similarly. QED

The problem of characterization of distribution of S as a solution of $S \stackrel{d}{=} (S + Y)Z$ is still open. Here we consider two examples.

Example 19.3.1. Let the distribution of $S \in \mathfrak{X}(B)$ by symmetric α-stable. In other words, the characteristic function of $S_{\bar{t}} = (S(t_1), \dots, S(t_n))$, $\bar{t} = (t_1, \dots, t_n)$, $0 < t_1 < \cdots < t_n \leqslant 1$, is

$$E \exp\{i(\theta, S_{\bar{t}})\} = \exp\left\{-\int_{\mathbb{R}^n} |(\theta, \bar{s})|^\alpha \Gamma_{S_{\bar{t}}}(d\bar{s})\right\}$$

where $\Gamma_{S_{\bar{t}}}(\cdot)$ is the spectral, finite symmetric measure of a symmetric α-stable random vector $S_{\bar{t}}$ (see Kuelbs 1973; Samorodnitsky and Taqqu 1989). For any $z \in (0, 1)$ let us choose an α-stable $Y_{\bar{t}} = (Y(t_1), \ldots, Y(t_n))$ with spectral measure

$$\Gamma_{Y_{\bar{t}}}(\mathrm{d}\bar{s}) = \frac{1 - z^\alpha}{z^\alpha} \Gamma_{S_{\bar{t}}}(\mathrm{d}\bar{s})$$

then S satisfies (19.3.8) with $Z = z$ and Y having marginals $Y_{\bar{t}}$. For other similar examples, see the class L, Feller (1971), Section 8, Chapter XVII.

Example 19.3.2 (Rachev and Samorodnitsky, 1990). Let $\mathbb{B} = \mathbb{R}$, Z be uniformly (0, 1)-distributed r.v. Consider the equation (19.3.10) with non-negative Y and S. If ϕ_X stands for the Laplace transform of a non-negative random variable X, then by (19.3.10), $\theta\phi_S(\theta) = \int_0^\theta \phi_S(x)\theta_Y(x)\,\mathrm{d}x$ for all $\theta > 0$. By differentiating we obtain

$$\theta_Y(\theta) = 1 + \theta\phi_S'(\theta)/\phi_S(\theta)$$

and thus

$$\phi_S(\theta) = \exp\left(-\int_0^\theta \frac{1 - \theta_Y(x)}{x}\,\mathrm{d}x\right). \tag{19.3.23}$$

It follows, from (19.3.23), that

$$\infty > \int_0^\theta \frac{(1 - \phi_Y(x))}{x}\,\mathrm{d}x = \int_0^\theta \left(\int_0^\infty \exp(-xy)(1 - F_Y(y))\,\mathrm{d}y\right)\mathrm{d}x$$
$$= \int_0^\infty (1 - F_Y(y))\,\frac{(1 - \exp(-y\theta)}{y}\,\mathrm{d}y.$$

Thus,

$$\int_1^\infty \frac{(1 - F_Y(y))}{y}\,\mathrm{d}y < \infty \quad \text{or} \quad \int_1^\infty (\ln y)F_Y(\mathrm{d}y) < \infty.$$

Thus, in the equation

$$S \stackrel{\mathrm{d}}{=} (Y + S)Z \quad Z \text{ is uniformly distributed} \tag{19.3.24}$$

Y must satisfy

$$\mathbb{E}\ln(1 + Y) < \infty. \tag{19.3.25}$$

With an appeal to Feller (1971), Theorem XIII 4.2, we conclude the following.

(a) *Any r.v. Y satisfying (19.3.25) gives a unique solution F_S of (19.3.24) for which the Laplace transform is given by $\phi_S(\theta)$ in (19.3.23).* More detailed analysis of Equation (19.3.24) shows:

(b) *Any distribution* F_S *determined by (19.3.24) is infinitely divisible. More precisely, let* Y *correspond to* S *in (19.3.24), and let* $0 < \beta < 1$. *Then there is a distribution* F_{S_β} *with Laplace transform* $\phi_S(\cdot)^\beta$; F_{S_β} *solves (19.3.24) and the corresponding* F_{Y_β} *is the mixture* $F_{Y_\beta}(x) = (1-\beta)F_0(x) + \beta F_Y(x)$.

(c) *The class* $\mathbb{S}$ *of random variables* S *solving (19.3.24) consists of infinitely divisible random variables, whose Lévy measure* λ (cf. Shiryaev 1984, p. 337) *is of the following form*

$$\lambda \ll \text{Leb } \textit{and } \lambda(\mathrm{d}x) = H(x)\,\mathrm{d}x \tag{19.3.26}$$

where $H(0) \in [0, 1]$, H *is non-increasing and* $H(x) \downarrow 0$ *as* $x \to \infty$. *The corresponding* Y *has* $1 - H$ *as its distribution function.*

Finally note that if S is the solution of (19.3.24) with given Y and uniformly distributed Z, then for any $\alpha > 0$,

$$S \stackrel{\mathrm{d}}{=} (S + Y_\alpha)Z_\alpha \quad S, Y_\alpha, Z_\alpha \text{ independent} \tag{19.3.27}$$

where F_{Y_α} is the mixture

$$\frac{\alpha}{1+\alpha} F_0 + \frac{1}{1+\alpha} F_Y$$

and Z_α has density $f_{Z_\alpha}(z) = (1 + \alpha)z^\alpha$, $0 \leqslant z \leqslant 1$.

As we have seen, in general the problem of evaluating the distribution of S is a difficult one, and in most cases we have to resort to approximations. Here we start with the analysis of stability of the set of solutions $\Pr_S$ of

$$S \stackrel{\mathrm{d}}{=} (Y + S)Z \quad Y, S, Z \text{ are independent r.v.s in } \mathfrak{X}(\mathbb{B}) \tag{19.3.28}$$

with some Y and Z for which we only know that they are 'close' to some given Y^* and Z^*.

Suppose we want to approximate the distribution of S in (19.3.28) by the distribution of S^* defined by

$$S^* \stackrel{\mathrm{d}}{=} (Y^* + S^*)Z^* \quad Y^*, S^*, Z^* \text{ are independent} \tag{19.3.29}$$

and such that we can evaluate the law of S^* given the laws of Z^* and Y^*, respectively. Assume also that the distributions of Z and Z^* (respectively, Y and Y^*) are close w.r.t. the minimal metric ℓ_p; i.e., for some small $\varepsilon > 0$ and $\delta > 0$

$$\ell_p(Z, Z^*) < \varepsilon \quad \text{and} \quad \ell_p(Y, Y^*) < \delta. \tag{19.3.30}$$

Theorem 19.3.2. Assume that (19.3.30) holds,

$$N_p(Z^*) < 1 - \varepsilon,$$

and

$$N_p(Y^*) + N_p(S^*) < \infty.$$

Then

$$\ell_p(S, S^*) \leqslant \frac{(\varepsilon + N_p(Z^*))\delta + \{N_p(Y^*) + N_p(S^*)\}\varepsilon}{1 - \varepsilon - N_p(Z^*)}. \tag{19.3.31}$$

Proof. From the definition of S and S^*

$$\begin{aligned}\ell_p(S, S^*) &= \ell_p(Z(Y + S), Z^*(Y^* + S^*))\\ &\leqslant \ell_p(Z(Y + S), Z(Y^* + S^*)) + \ell_p(Z(Y^* + S^*), Z^*(Y^* + S^*))\\ &\leqslant N_p(Z)\ell_p(Y + S, Y^* + S^*) + N_p(Y^* + S^*)\ell_p(Z, Z^*)\\ &\leqslant N_p(Z)[\ell_p(Y, Y^*) + \ell_p(S, S^*)] + \ell_p(Z, Z^*)[N_p(Y^*) + N_p(S^*)]\end{aligned}$$

and thus

$$\ell_p(S, S^*) \leqslant \frac{N_p(Z)\delta + \{N_p(Y^*) + N_p(S^*)\}\varepsilon}{1 - N_p(Z)}.$$

Finally, by the triangle inequality and (19.3.18), $N_p(Z) = \ell_p(Z, 0) \leqslant \varepsilon + N_pZ^*)$, which proves the assertion. QED

In a similar fashion one may evaluate the rate of convergence of $M_n \to M$, where $M_n = \sup_{1 \leqslant k \leqslant n} X_k$, $M_n \stackrel{\mathrm{d}}{=} (Y \vee M_{n-1})Z$ (here $Z \geqslant 0$ and the product and maximum are pointwise). Similar to Theorem 19.3.1, by letting $n \to \infty$, one obtains (19.3.11). Further, since Z and Y are independent,

$$\begin{aligned}\ell_p(M_n, M) &\leqslant \mathscr{L}_p((Y \vee M_{n-1})Z, (Y \vee M)Z)\\ &\leqslant \mathscr{L}_p(M_{n-1}Z, M \vee Z) \leqslant \mathscr{L}_p(M_{n-1}, M)N_p(Z).\end{aligned}$$

From this, as in Theorem 19.3.1 (b), we get

$$\ell_p(M_n, M) \leqslant N_p^n(Z)\frac{N_p(Y)}{1 - N_p(Z)}.$$

Suppose now that the assumption of Theorem 19.3.1 (b) holds, then $M_n \to M$(a.s.), and

$$\mathscr{L}_p(M_n^*, M^*) \leqslant N_p^n(Z)\frac{N_p(Y)}{1 - N_p(Z)}$$

where $M_n^* = (M_n, M_{n+1}, \dots)$, $M^* = (M, M, \dots)$.

Example 19.3.3. All simple max-stable processes are solutions of $M \stackrel{\mathrm{d}}{=} (Y \vee M)Z$. Given an α-max-stable process M, i.e. one whose marginal distributions are specified by

$$\Pr\{M(t_1) \leqslant x_1, \dots, M(t_n) \leqslant x_n\} = \exp\left\{-\int_\Omega\left(\max_{1 \leqslant i \leqslant n}(\lambda_i x_i^{-\alpha})U_t(\mathrm{d}\lambda_1, \dots, \mathrm{d}\lambda_n\right\}\right.$$

where $\alpha > 0$, $\bar{t} = (t_1, \ldots, t_n)$ and $U_{(\bullet)}$ is a finite measure on

$$\Omega = \left\{(\lambda_1, \ldots, \lambda_n);\ \lambda_i > 0,\ i = 1, \ldots, n,\ \sum_1^n \lambda_i = 1\right\}$$

(see Resnick 1987b). For any $z \in (0, 1)$ define the max-stable

$$Y_{\bar{t}} = (Y(t_1), \ldots, Y(t_n))$$

with max-stable measure

$$U_{Y_{\bar{t}}}(\mathrm{d}\lambda) = \frac{1 - z^\alpha}{z^\alpha} U_{M_{\bar{t}}}(\mathrm{d}\lambda)$$

then M satisfies (19.3.11) with $Z = z$ and Y being α-max-stable with marginals having spectral measure $M_{Y_{\bar{t}}}$. A more general example for M as a solution of the above equation is given by Balkema *et al.* (1990), where the class L for maxima scheme is studied.

Example 19.3.4. Suppose $\mathbb{B} = \mathbb{R}$ and Z is $(0, 1)$-uniformly distributed. Then $M \stackrel{\mathrm{d}}{=} (Y \vee M)Z$ implies that

$$F_M(x) = \exp\left(-\int_x^\infty \frac{1}{t} \bar{F}_Y(t)\, \mathrm{d}t\right) \quad \bar{F}_Y := 1 - F_Y. \tag{19.3.22}$$

For example, if Y has *Pareto distribution*, $\bar{F}_Y(t) = \min(1, t^{-\beta})$, $\beta > 1$, then M has a truncated extreme value distribution

$$F_M(x) = \begin{cases} \exp\left(-1 + x - \dfrac{x^{1-\beta}}{\beta - 1}\right) & \text{for } 0 \leqslant x \leqslant 1 \\ \exp\left(-\dfrac{x^{1-\beta}}{\beta - 1}\right) & \text{for } x \geqslant 1. \end{cases} \tag{19.3.33}$$

From (19.3.11) it also follows that

$$F_M(x) > 0 \quad \forall x > 0 \Leftrightarrow \mathbb{E} \ln(1 + Y) < \infty. \tag{19.3.34}$$

Note that if M has an atom at the origin, then

$$0 < \Pr(M \leqslant 0) = \Pr(M \leqslant 0)\Pr(Y \leqslant 0)$$

i.e., $Y \equiv 0$, the degenerate case. Moreover, *the condition* $\mathbb{E} \ln(1 + Y) < \infty$ *is necessary and sufficient for the existence of the non-degenerate solution of* $M \stackrel{\mathrm{d}}{=} (Y \vee M)Z$ *given by* (19.3.32) (*Rachev and Samorodnitsky*, 1990). Clearly, the latter assertion can be extended for any Z such that Z^α is $(0, 1)$-uniformly distributed for some $\alpha > 0$, since $M \stackrel{\mathrm{d}}{=} (M \vee Y)Z \Rightarrow M^\alpha \stackrel{\mathrm{d}}{=} (M^\alpha \vee Y^\alpha)Z^\alpha$.

As far as the approximation of M is concerned we have the following theorem,

Theorem 19.3.3. Suppose the distribution of M is determined by (19.3.11) and

$$M^* \stackrel{d}{=} Z^*(Y^* \vee M^*) \quad \ell_p(Z, Z^*) < \varepsilon \quad \ell_p(Y, Y^*) < \delta. \tag{19.3.35}$$

Assume also that $N_p(Z^*) \leqslant 1 - \varepsilon$ and $N_p(Y) + N_p(M^*) < \infty$, then as in Theorem 19.3.2 we arrive at

$$\ell_p(M, M^*) \leqslant \frac{(\varepsilon + N_p(Z^*))\delta + [N_p(Y^*) + N_p(M^*)]\varepsilon}{1 - N_p(Z^*) - \varepsilon}. \tag{19.3.36}$$

Proof. Recall that ℓ_p-metric is regular with respect to the sum and maxima of independent r.v.s, i.e., $\ell_p(X + Z, Y + Z) \leqslant \ell_p(X, Y)$ and $\ell_p(X \vee Z, Y \vee Z) \leqslant \ell_p(X, Y)$ for any $X, Y, Z \in \mathfrak{X}(\mathbb{B})$ and Z-independent of X and Y, see (18.3.2)–(18.3.6). Thus, one can repeat step by step the proof of Theorem 19.3.2 by replacing the equation $S \stackrel{d}{=} (S + Y)Z$ with $M \stackrel{d}{=} (M \vee Y)Z$. QED

Remark that in both Theorem 19.3.2 and 19.3.3, the ℓ_p-metric was chosen as a suitable metric for the stability problems under consideration. The reason is the 'double ideality' of ℓ_p (cf. Section 18.3), i.e., ℓ_p plays the role of ideal metric for both summation and maxima schemes.

Next we consider the relation

$$G \stackrel{d}{=} Z(Y + \delta G) \tag{19.3.37}$$

where as before, Z, Y and G are independent elements of $B = C(T)$, and δ is a Bernoulli r.v., independent of them, with $\Pr\{\delta = 1\} = d$. If $Z \equiv 1$, G could be chosen to have a '*geometric infinitely divisible distribution*', i.e., the law of G admits the representation

$$G \stackrel{d}{=} \sum_{i=1}^{\tau(d)} Y_i \tag{19.3.38}$$

where Y_is are i.i.d. and $\tau(d)$ is independent of Y_is geometric random variable with mean $1/(1 - d)$, see (19.3.8).

Lemma 19.3.1. In the finite-dimensional case $B = \mathbb{R}^m$, a necessary and sufficient condition for G to be geometric infinitely divisible is that its characteristic function is of the form

$$f_G(t) = (1 - \log \phi(t))^{-1} \tag{19.3.39}$$

where $\phi(\bullet)$ is an infinitely divisible characteristic function.

The proof is similar (but slightly more complicated) to the proof of Lemma 19.3.2 which we shall write in detail.

Example 19.3.5. Suppose Z has a density

$$p_Z(z) = (1 + \alpha)z^\alpha \quad \text{for } z \in (0, 1). \tag{19.3.40}$$

Then from (19.3.37) we have

$$\theta^{\alpha+1} f_G(\theta) = (\alpha + 1) \int_0^\theta u^\alpha f_Y(u)\{(1 - d) + d f_G(u)\}\, \mathrm{d}u$$

where $f_{(\bullet)}$ stands for the characteristic function of the r.v. ($\bullet$). By differentiating we obtain the equation

$$f'_G(\theta) + \frac{\alpha + 1}{\theta}[1 - d f_Y(\theta)] f_G(\theta) = \frac{\alpha + 1}{\theta}(1 - d) f_Y(\theta)$$

which solution clearly describes the distribution of G for given Z and Y. Next let us consider the approximation problem assuming that $Z = z$ is a constant 'close to 1'. Suppose further that the distribution of Y belongs to the class of 'aging' distributions HNBUE (see Definition 16.3.1). Then our problem is to approximate the distribution of G defined by

$$G \stackrel{\mathrm{d}}{=} Z(Y + \delta G) \quad Z = z(\text{const}) \quad F_Y \in \text{HNBUE},\ Y, G \tag{19.3.41}$$

independent by means of G^* specified by

$$G^* \stackrel{\mathrm{d}}{=} Y^* + \delta G \quad F_Y^*(t) = 1 - \exp(-t/\mu) \quad t \geqslant 0,\ Y^*, G^* \text{ independent.} \tag{19.3.42}$$

Given that $\mathbb{E}Y = \mu$ and $\operatorname{Var} Y = v^2$, we obtain the following estimate of the deviation between the distributions of Y and Y^* in terms of the metric ℓ_p (see Kalashnikov and Rachev (1988), Chapter 4, Section 2, Lemma 10)

$$\ell_1(Y, Y^*) \leqslant 2(\mu^2 - v^2)^{1/2} \tag{19.3.43}$$

and for $p > 1$

$$\ell_p(Y, Y^*) \leqslant (\mu^2 - v^2)^{1/4p} 8\mu\Gamma(2p)^{1/p}. \tag{19.3.44}$$

The following proposition gives an estimate of the distance between G and G^*.

Theorem 19.3.4. Suppose that G satisfies (19.3.37) where Z, Y and G are independent elements of $B = C(T)$ and δ is a Bernoulli r.v. independent of them, and consider

$$G^* \stackrel{\mathrm{d}}{=} (Y^* + \delta G^*)Z^* \quad G^*, Z^*, Y^* \in \mathfrak{X}(B)$$

where G^*, Z^*, Y^* and δ are again independent. Assume also that

$$\ell_p(Z, Z^*) \leqslant \varepsilon,\ \ell_p(Y, Y^*) \leqslant \delta, \quad N_p(Z^*)d < 1 - \varepsilon d.$$

Then

$$\ell_p(G, G^*) \leqslant \frac{(\varepsilon + N_p(Z^*))\delta + [N_p(Y^*) + dN_p(G^*)]\varepsilon}{1 - dN_p(Z^*) - d\varepsilon}.$$

Proof. Analogous to Theorem 19.3.2,

$$\begin{aligned}\ell_p(G, G^*) &\leqslant \ell_p(Z(Y + \delta G), Z(Y^* + \delta G^*)) + \ell_p(Z(Y^* + \delta G^*), Z^*(Y^* + \delta G^*))\\ &\leqslant N_p(Z)\ell_p(Y + \delta G, Y^* + \delta G^*) + N_p(Y^* + \delta G^*)\ell_p(Z, Z^*)\\ &\leqslant N_p(Z)[\ell_p(Y, Y^*) + d\ell_p(G, G^*)] + [N_p(Y^*) + dN_p(G^*)]\ell_p(Z, Z^*).\end{aligned} \tag{19.3.45}$$

From this and $N_p(Z) \leqslant \varepsilon + N_p(Z^*)$ the assertion follows. QED

In the special case given in (19.3.41) and (19.3.42) the inequality

$$\ell_p(G, G^*) \leqslant \frac{N_p(Z)\delta + [N_p(Y^*) + dN_p(G^*)]\varepsilon}{1 - dN_p(Z)} \tag{19.3.46}$$

holds and moreover $N_p(Y^*) + dN_p(G^*) \leqslant (1 + d)/(1 - d)\mu(\Gamma(p + 1))^{1/p}$. Finally, since $\varepsilon = \ell_p(Z, Z^*) = 1 - z$, and δ can be defined to be the right-hand side of (19.3.43) or (19.3.44), we have the following theorem.

Theorem 19.3.5. If G and G^* are given by (19.3.41) and (19.3.42) respectively,

$$\ell_p(G, G^*) \leqslant \frac{1 - z}{1 - zd}\mu(\Gamma(p + 1))^{1/p}\frac{1 + d}{1 - d} + \frac{z\delta_p}{1 - zd} \tag{19.3.47}$$

where

$$\delta_p := \begin{cases} 2(\mu^2 - v^2)^{1/2} & \text{if } p = 1\\ \Gamma(2p)^{1/p}(\mu^2 - v^2)^{1/4p}8\mu & \text{if } p > 1\end{cases}$$

and $N_p(Y) = (\mathbb{E}Y^p)^{1/p}$.

For $p = 1$ we obtain from (19.3.47)

$$\int_{-\infty}^{\infty} |F_G(x) - F_{G^*}(x)|\,dx \leqslant \frac{2}{1 - zd}\left[\left(\frac{1 - z}{1 - d}\right)\mu + z(\mu^2 - v^2)^{1/2}\right]$$

Finally, consider the geometric maxima H defined by

$$H \overset{d}{=} Z(Y \vee \delta H) \text{ or equivalently } H \overset{d}{=} \bigvee_{k=1}^{\tau(d)} Y_k \prod_{j=1}^{k} Z_j \tag{19.3.48}$$

where Z, Y, δ, H, $\tau(d)$, Y_k and Z_j are assumed to be independent $Y_k \overset{d}{=} Y$, $Z_k \overset{d}{=} Z$, $Z > 0$, $Y > 0$, $H > 0$.

Example 19.3.6. If $Z = 1$, then H has a *geometric maxima infinitely divisible (GMID) distribution*, i.e. for any $d \in (0, 1)$

$$H \stackrel{d}{=} \bigvee_{1}^{\tau(d)} Y_k \tag{19.3.49}$$

where $Y_k = Y_k^{(d)}$, $k \in \mathbb{N}$ are i.i.d non-negative r.v. and $\tau(d)$ is independent of Y_ks geometric random variable

$$\Pr(\tau(d) = k) = (1 - d)d^{k-1} \quad k \geqslant 1. \tag{19.3.50}$$

Let $\mathbb{B} = \mathbb{R}^m$. Let $\Pr(H \leqslant x) = G(x)$, $x \in \mathbb{R}_+^m$, $\Pr(Y_1^{(d)} \leqslant x) = G_d(x)$ and then (19.3.49) is the same as

$$G(x) = \sum_{j=1}^{\infty} (1 - d)d^{k-1} G_d(x) = \frac{(1 - d)G_d(x)}{1 - dG_d(x)}. \tag{19.3.51}$$

If we solve for G_d in (19.3.51) we get

$$G_d(x) = G(x)/(1 - d + dG(x)) \tag{19.3.52}$$

which is clearly equivalent to

$$H \stackrel{d}{=} Y \vee \delta H \tag{19.3.53}$$

where $Y \stackrel{d}{=} Y_1^{(d)}$ (cf. (19.3.49)).

We now characterize the class GMID. We shall consider the slightly more general case of H and Y_k being not necessarily non-negative. The characterizations are in terms of max-infinitely divisible (MID) distributions, exponent measures and multivariate extremal processes. For background on these concepts see Resnick (1987a), Chapter 5. A MID-distribution F with exponent measure μ has the property that the support $[x\colon F(x) > 0]$ is a rectangle. Let $\ell \in \mathbb{R}^m$ be the 'bottom' of this rectangle (cf. Resnick 1987a, p. 260). Clearly, in the one-dimensional case $m = 1$, any d.f. F is MID-distribution.

Lemma 19.3.2. For a distribution G on $\mathbb{R}^m$ the following are equivalent:

(i) $G \in \mathrm{GMID}$.

(ii) $\exp(1 - 1/G)$ is a MID-distribution.

(iii) There exists $\ell \in [-\infty, \infty)^m$ and an exponent measure μ concentrating on the rectangle $\{x \in \mathbb{R}^m, x \in \ell\}$ such that for any $x \geqslant \ell$

$$G(x) = \frac{1}{1 + \mu(\mathbb{R}^m \backslash [\ell, x])}.$$

(iv) There exists an extremal process $\{Y(t), t > 0\}$ with values in $\mathbb{R}^m$ and an independent exponential random variable E with mean 1 such that $G(x) = \Pr(Y(E) \leqslant x)$.

Proof. (i) $\Rightarrow$ (ii) We have the following identity

$$G = \lim_{\alpha \downarrow 0} 1 \bigg/ \left[1 + \frac{1}{\alpha} \left(1 - \frac{G}{\alpha + (1 - \alpha)G} \right) \right]. \tag{19.3.54}$$

Therefore,

$$\exp(1 - 1/G) = \lim_{\alpha \downarrow 0} \exp\left[-\frac{1}{\alpha} \left(1 - \frac{G}{\alpha + (1 - \alpha)G} \right) \right].$$

If $G \in GMID$, then $G/(\alpha + (1 - \alpha)G)$ is a d.f. for any $\alpha \in (0, 1)$ which implies (from Resnick (1987a), pp. 257–258) that

$$\exp\left[-\frac{1}{\alpha} - \frac{G}{\alpha + (1 - \alpha)G)} \right) \right]$$

is MID-distribution. Since the class of MID-distributions is closed under weak convergence, it follows that $\exp(1 - 1/G)$ is MID-distribution.

(ii) $\Rightarrow$ (iii) If $\exp(1 - 1/G)$ is MID-distribution then, by the characterization of MID-distribution, there exists $\ell \in [-\infty, \infty)^m$ and an exponent measure μ concentrating on $\{x \colon x \geqslant \ell\}$ such that for

$$x \geqslant \ell, \exp\left\{ 1 - \frac{1}{G(x)} \right\} = \exp\{-\mu(\mathbb{R}^m \backslash [\ell, x])\}$$

and equating exponents yields (iii).

(iii) $\Rightarrow$ (iv) Suppose μ is the exponent measure assumed to exist by (iii) and let $\{Y(t), t > 0\}$ be an extremal process with

$$\Pr(Y(t) \leqslant x) = \exp\{-t\mu(\mathbb{R}^m \backslash [\ell, x])\}. \tag{19.3.55}$$

Then

$$\begin{aligned}\Pr(Y(E) \leqslant x) &= \int_0^\infty e^{-t} \Pr(Y(t) \leqslant x)\, dt = \int_0^\infty e^{-t} \exp\{-t\mu(\mathbb{R}^m \backslash [\ell, x])\}\, dt \\ &= 1/(1 + \mu(\mathbb{R}^m \backslash [\ell, x]))\end{aligned}$$

as required.

(iv) $\Rightarrow$ (i) Suppose $G(x) = \Pr(Y(E) \leqslant x)$. If (19.3.55) holds then

$$G(x) = 1/(1 + \mu(\mathbb{R}^m \backslash [\ell, x])).$$

To show $G \in GMID$ we need to show that

$$\frac{G(x)}{1 - d + dG(x)} = \frac{1}{1 + (1 - d)\mu(\mathbb{R}^m \backslash [\ell, x])}$$

is a distribution and this follows readily by observing

$$\Pr(Y((1 - d)E) \leqslant x) = \frac{1}{1 + (1 - d)\mu(\mathbb{R}^m \backslash [\ell, x])}. \qquad \text{QED}$$

In particular, Lemma 19.3.2 implies that the real-valued r.v. H has a GMID-distribution if and only if its d.f. F_H can be represented as $F_H(t) = (1 - \log \Phi(t))^{-1}$, where $\Phi(t)$ is an arbitrary d.f. For instance, if

$$\Phi(x) = \exp(-x^{-\alpha})\ (x > 0)$$

then

$$F_H(x) = \frac{x^\alpha}{1 + x^\alpha} \quad (x \geqslant 0)$$

is the 'log logistic' distribution with parameter $\alpha > 0$. If Φ is the Gumbel distribution, i.e., $\Phi(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$, then clearly F_H is the exponential distribution with parameter 1.

Example 19.3.7. Consider the Equation (19.3.53) for real-valued r.v.s Z, Y and H.

Assume Z has the density (19.3.40); then

$$F_H(x) = \int_0^1 F_Y\left(\frac{x}{z}\right)\left[1 + dF_H\left(\frac{x}{z}\right)\right](\alpha + 1)z^\alpha \,\mathrm{d}z$$

or

$$x^{-\alpha-1}F_H(x) = (\alpha + 1)\int_x^\infty F_Y(y)[1 + dF_H(y)]y^{-\alpha-2}\,\mathrm{d}y.$$

This equation is easily solved to obtain:

$$F_H(x) = \left(\exp - (\alpha + 1)\int_x^\infty \frac{1}{u}[1 - dF_Y(u)]\,\mathrm{d}u\right)$$
$$\times (\alpha + 1)\int_x^\infty \left(\exp(\alpha + 1)\int_y^\infty \frac{1}{u}[1 - dF_Y(u)]\,\mathrm{d}u\right)\frac{1}{y}F_Y(y)\,\mathrm{d}y. \tag{19.3.56}$$

The stability analysis is handled in a similar way to Theorem 19.3.4. Consider the equations

$$H \stackrel{\mathrm{d}}{=} (Y \vee H)Z \quad \text{and} \quad H^* \stackrel{\mathrm{d}}{=} (Y^* \vee H^*)Z^* \tag{19.3.57}$$

where Y, H, Z (respectively Y^*, H^*, Z^*) are independent non-negative elements of $\mathfrak{X}(\mathbb{B})$. Following the model in the beginning of Chapter 19, suppose that the 'input' distributions $(\Pr_Y, \Pr_Z)$ and $(\Pr_{Y^*}, \Pr_{Z^*})$ are close in the sense that

$$\ell_p(Z, Z^*) \leqslant \varepsilon \quad \ell_p(Y, Y^*) \leqslant \delta. \tag{19.3.58}$$

Then the 'output' distributions $\Pr_H$, $\Pr_{H^*}$ are also close as the following theorem asserts.

Theorem 19.3.6. Suppose H and H^* satisfy (19.3.57) and (19.3.58) holds. Suppose also that $N_p(Z^*) < 1 - \varepsilon d$. Then

$$\ell_p(H, H^*) \leqslant \frac{(\varepsilon + N_p(Z^*))\delta + [N_p(Y^*) + dN_p(H^*)]\varepsilon}{1 - dN_p(Z^*) - d\varepsilon}$$

The proof is similar to that of Theorem 19.3.4.

Bibliographical Notes

PART I

Metrics in the space of probability measures and in the space of random elements have been studied by a great number of researchers; among them are Lévy (1925, 1950), Fréchet (1937), Fortét and Mourier (1953), Kolmogorov (1953), Prokhorov (1956), Kantorovich and Rubinstein (1958), Strassen (1965), Dudley (1976, 1989) and Zolotarev (1976 a–d, 1986).

Below are a few monographs which treat the probability metrics as a separate subject:

(1) Lévy (1925) (pp. 199–200; definition of Lévy metric);
(2) Fréchet (1937) (pp. 193–195; definition of a metric between random variables);
(3) Lukacs (1968) (Chap. 3; definition and properties of the Ky Fan metric and $\mathscr{L}_p$-metrics between random variables);
(4) Hennequin and Tortrat (1965) (Prokhorov metric, total variation metric);
(5) Dudley (1976, 1989) (duality representation of Prokhorov and Kantorovich–Rubinstein metrics in the space of probability measures, properties of metrics that metrize the weak convergence);
(6) Shorack and Wellner (1986) (Section 3.6; Kantorovich–Wasserstein metric);
(7) Le Cam (1986) (Section 16.4; Hellinger metric);
(8) Zolotarev (1986) (Chap. 1; main concepts and applications of the theory of probability metrics (TPM));
(9) Kalashnikov and Rachev (1988) (Chap. 3; review on the results in the TPM until 1984);
(10) Kakosyan *et al.* (1988) (Chap. 2; compactness criteria in the space of measures endowed with different metrics);
(11) Reiss (1989) (Section 5.2; Hellinger metric).

Chapters 2 and 3

The results of Chapters 2 and 3 are given in Rachev and Shortt (1989). The dual representation for $\boldsymbol{\theta}_p$ (see (2.1.4) and Remark 2.1.2) is probably well known;

for some extensions see Maejima and Rachev (1987). More examples of metric and semimetric (pseudometric) spaces can be found in Dunford and Schwartz (1988), Kelley (1955), Billingsley (1968). The notion of probability metric on the space of distributions $\Pr_{X,Y}$ of pairs or random variables X, Y taking on values in a complete separable metric space (U, d) was introduced in the basic paper of Zolotarev (1976b). The extension of p. metrics on universally measurable separable metric space is given by Rachev (1985a) and Rachev and Shortt (1989). For additional facts on universally measurable spaces and analytic sets, we refer to Dudley (1989). Lemma 2.5.1 is due to Berkes and Phillip (1979). The notions of simple and compound metrics were introduced by Zolotarev (1976b, 1983b). The measure Q in (3.2.5) is defined as in Berkes and Phillip (1979), p. 53. Regarding the open problem 3.2.1, A. V. Skorokhod posed the problem of characterization of the set of simple metrics generating the weak topology.

Chapter 4

The necessary facts on Hausdorff metric can be found in Section 29, Hausdorff (1949) and Sections 21, 31, Kuratovski (1966, 1969). The Hausdorff metric between functions $f\colon [0, 1] \to \mathbb{R}$ (cf. (4.1.9)) was introduced by Sendov and Penkov (1962) as an appropriate metric in the space $D[0, 1]$. The problem of characterization of the class of 'suitable' metrics in $D[0, 1]$ was posed by Kolmogorov (see Kolmogorov 1956, Skorokhod 1956, Section 2; Prokhorov 1956, Appendix 1; Billingsley 1968, Sendov, 1969).

The Hausdorff distance has been used in approximation theory very extensively (see, e.g., the survey Sendov (1969) and the monograph Sendov (1979)). Lemma 4.11, Properties 4.1.1 and 4.1.2 are due to Sendov (1969). For more details concerning Remark 4.1.4 see Rachev (1980). The Hausdorff structure of probability measures was introduced and studied by Rachev (1979a,b, 1980, 1985b).

The λ-structure was introduced by Zolotarev (1976b). The 'duality' between the λ-representation and the Hausdorff representation of a p. semidistance (Theorem 4.2.1) was given in Rachev (1983a).

ζ-structure for simple p. metrics was introduced and studied by Zolotarev (1976b, 1983b). Section 4.3 is based on the results of Fortét and Mourier (1953), Bartoszynski (1961), Rao, (1962), Dudley (1966a, 1976), Bhattacharya and Ranga Rao (1976), Zolotarev (1977a,b, 1978), Neveu and Dudley (1980), A. Szulga (private communication), Rachev (1984b, 1985c).

PART II

The Monge–Kantorovich problem (MKP) has been the center of attention of many specialists in various areas of mathematics for a long time:

(1) differential geometry—see Germaine (1886), Appell (1884, 1928), Vygodskii (1936), Sudakov (1976);
(2) functional analysis—Kantorovich and Akilov (1984), Levin and Milyutin (1979);
(3) linear programming—Hoffman (1961), Berge and Ghouila-Houri (1965, Chap. 9), Vershik (1970), Rubinstein (1970), Kemperman (1983), Knott and Smith (1984), Smith and Knott (1987), Barnes and Hoffman (1985), Anderson and Nash (1987), Kellerer (1988);
(4) probability theory—Kantorovich and Rubinstein (1958), Fernique (1981), Dudley (1976, 1989), Zolotarev (1983b), Kellerer (1984a,b), Rüschendorf (1985b);
(5) mathematical statistics—Huber (1981), Bickel and Freedman (1981);
(6) information theory and cybernetics—Wasserstein (1969), Gray *et al.* (1975), Gray and Ornstein (1979), Gray *et al.* (1980);
(7) statistical physics—Dobrushin (1970), Cassando *et al.* (1978);
(8) the theory of dynamical systems—Ornstein (1971), (1985), Gray (1988);
(9) matrix theory—Olkin and Pukelheim (1982).

Currently, it is now appropriate to talk about the MKP as being a whole range of problems with applications to many mathematical theories that seem different at first glance. Entire schools have been formed developing different offshoots of the MKP by making use of diversified mathematical language.

Gini introduced his 'L'indice di dissomiiglianza' in 1914 for a goodness-of-fits test. There exists a vast literature in Italy on statistical aspects of the Gini's index of dissimilarity (cf. Bertino 1971, 1972, Forcina and Galmacei 1974, Gini 1951, 1965, Herzel 1963, Landenna 1956, 1961, Leti 1962, 1963, Lucia 1965, Pompilei 1956, Salvemini 1939, 1943, 1949, 1953, 1957a,b, Samuel and Bachi 1964).

The fundamental work of Hoeffding (1940) and Fréchet (1951, 1957) on the marginal problem was extended in three areas of mathematics:

(a) *Linear programming*—Hoffman (1961), Barnes and Hoffman (1985), Derigs, Goecke and Schrader (1986), Goldfarb (1985), Kleinschmidt, Lee and Schannath (1987).

Note that the Hoffman's North–West corner rule for solutions of transportation problems in linear programming gives a 'greedy algorithm' for computing the Fréchet bounds in the class of bivariate distributions with fixed marginals. Balinski and Rachev (1989) proposed to exploit the connection between discrete transportation and apportionment problems (Balinski and Gonzalez 1988, Balinski and Rachev 1989, Balinski and Young 1982), and the 'continuous' Monge–Kantorovich problem (MKP) (Rachev 1984d). For example, the explicit solutions of MKP for laws on $\mathbb{R}$ give rise to new 'greedy algorithms' for transportation linear programming problems. Vice versa, greedy algorithms for marginal problems on $\mathbb{R}^k$ ($k \geqslant 2$) could suggest explicit solutions of the continuous MKP.

(b) *Statistics*—concordance, associativeness and correlation of random quantities—Kruskal (1958), Lehman (1966), Robbins (1975), Lai and Robbins (1976), Mallows (1972), Rinott (1973), Bartfai (1970), Whitt (1976), Cambanis *et al.* (1976), Cambanis and Simons (1982), Wolff and Schweizer (1981), Rüschendorf (1980, 1981, 1985a,b, 1989), Gaffke and Rüschendorf (1981, 1984), Cuesta and Matran (1989).

(c) *Marginal problems and doubly stochastic matrices*—Beasley and Gibson (1975), Berkes and Phillip (1979), Brown (1965), Brown and Shiflett (1970), Brunaldi and Czima (1975), Deland and Shiflett (1980), Isbell (1955, 1962), Douglas (1964, 1966), Dowson and Landau (1982), Kellerer (1961–88), Hoffmann-Jorgensen (1977), Major (1978), Mukerjee (1985), Shortt (1983–87), Vorob'ev (1962, 1963), Sudakov (1972–76), Levin (1977–86), Levin and Milyutin (1979), Haneveld (1985).

Ornstein's $\bar{d}$ distance (cf. Ornstein 1971, 1975) as well as the $\bar{\rho}$-distance (Gray 1988) were applied to communication theory (Gray *et al.* 1975).

The term 'Wasserstein metric' appeared first in Dobrushin's 1970 paper to denote the Kantorovich metric (in our notations). The work of Wasserstein (1969) was very influential among specialists in statistical physics, dynamical systems and probability (cf. Dobrushin 1970, Gray *et al.* 1975, Dudley 1976, Vallander 1973, De Acosta 1982, Rüschendorf 1985b, Tchen 1980, Valkeila 1986, Gelbrich 1988, Gray 1988, Cuesta and Matran 1989).

Chapters 5 and 6

Problem (5.1.2) was first formulated and studied by Kantorovich (1940, 1942, 1948) for a compact metric space (U, d) and cost function $c = d$. The MKP with a continuous cost function on a compact space U was studied by Levin (1974–78) and Levin and Milyutin (1979) (see also Vershik 1970, Rubinstein 1970). Duality in MKP for u.m. s.m.s. U and continuous cost function c was studied by Kellerer (1984a, see also the reference there), Kemperman (1983) and Levin (1984). In our representation (Chapter 5) we consider cost function c such that the corresponding Kantorovich functional (5.1.2) is a probability semi-distance. In this case we obtain duality theorems for s.m.s. U which are more refined than the general duality theorems for the MKP.

Kantorovich's formulation differs from the Monge problem in that the class of all laws on U^2 with fixed marginals is broader than the class of one-to-one transference plans in Monge's sense (Vershik 1970). Sudakov (1976) showed that if the marginals are given on a bounded subset of a finite-dimensional Banach space and are absolutely continuous with respect to Lebesgue measure, then there exists an optimal one-to-one transference plan.

In Chapters 5 and 6 we follow the results of Rachev (1984d), Rachev and

Taksar (1989), Rachev and Shortt (1990). In the proof of Theorem 5.2, we borrow some ideas from Dudley (1968), Szulga (1978, 1982), Fernique (1981). De Acosta (1982) referred to the Kantorovich metric as the Wasserstein metric and called the Kantorovich–Rubinstein functional the Wasserstein norm. The concept of uniformity (merging) in the set of laws was investigated by Senatov (1977), Dudley (1989), D'Aristotile, Diaconis and Freedman (1988), Rachev, Rüschendorf and Schief (1988).

Chapter 7

Relationships between minimal metrics similar to these in Theorems 7.1.1 to 7.1.3 were studied by Zolotarev (1976b, 1977a; 1986, Theorem 1.2.4). Theorem 7.1.4 was proved by Rachev and Rüschendorf (1990a) (see also Ignatov and Rachev (1983) for some particular cases). Its extension (see Remark 7.1.2) was pointed out by Shortt.

Kellerer (1984a,b, 1988) investigated the dual representation of the generalized Monge–Kantorovich problem. K-minimal metrics were introduced by Rachev (1981b). Chapter 7 includes results of Lorentz (1953) Lemma 7.3.1; Tchen (1980), see Theorem 7.3.2; Cambanis *et al.* (1976), see Corollary 7.3.2; Dudley (1968), Schay (1974, 1979), see Theorem 7.4.1; Strassen (1965), Dudley (1968), Garsia-Palomares and Gine (1977), see Corollary 7.4.2; Dobrushin (1970), see Corollary 7.4.2; Rachev (1981b), see Theorems 7.3.1, 7.3.3; Rachev (1984d), see Theorems 7.5.1, 7.5.2; Rachev (1982a,b, 1984c), see Corollary 7.5.3. Special cases of Theorem 7.3.2 were considered by Salvemini (1943), Bass (1955), Bertino (1966), Kantorovich and Rubinstein (1958), Bartfai (1970), Dall'Aglio (1956), Day (1972), Hardy *et al.*, (1952, p. 278), Rinott (1973), Robbins (1975), Lai and Robbins (1976), Rachev (1981b), Rüschendorf (1980). The linear programming transportation problems related to the discrete version of Theorem 7.3.2 were considered by Hoffman (1961) and Balinski and Rachev (1989). Hoffman (1961) proposed a 'greedy algorithm' (the famous North–West corner rule) for determining the optimal distribution in the Lorentz Lemma 7.3.1 (for $N = 2$). The general case ($N \geqslant 2$) was considered by Balinski and Rachev (1989).

The definitions 'Monge sequence' (Hoffman 1961), quasimonotone function (Cambanis *et al.* 1976), and superadditive function (Tchen 1980) are used to explain equivalent notions. Quasimonotone is a function f for which $(-f)$ is quasiantitone (cf. Definition 7.3.1).

Chapter 8

Dall'Aglio (1956) found the explicit representation (8.1.27) for $\phi(x, y) = |x - y|^p$, $p \geqslant 1$. Cambanis *et al.* extended his result for any antitone function ϕ. Theorem 8.2.1 was proved in Rachev (1985a). Dual representations for $\check{\mu}_c$ (8.1.12) were also given by Kellerer (1988). Special cases of Corollary 8.2.1 (U-complete separable metric space) was proved by Bickel and Freedman (1981), de Acosta (1982), Givens and Shortt (1984), Dotto (1989). The case $p = 1$, U separable metric space was considered by Dudley ((1966a, 1968) (U, d) bounded) and Dobrushin (1970). Zolotarev (1975a) proved inequalities similar to (8.2.21) and (8.2.22). The results in Chapter 8 are due to Rachev (1982a,b, 1984a,b,c, 1985a) and Rachev and Rüschendorf (1990a).

Chapter 9

Owing to the number of its important applications (see Shohat and Tamarkin 1943, Ahiezer and Krein 1962, Karlin and Studden 1966, Hoeffding 1955, Hoeffding and Shrikhande 1955, Basu and Simons 1983, Kemperman 1972, 1983, Haneveld 1985), the moment problem can also be treated as an approximation of the marginal problem. The significance of the moment problem for the theory of probability metrics was also stressed by Sholpo (1983), Rachev (1985a, 1989b), Anastassiou and Rachev (1989, 1990).

General dual representations of moment problems on a compact space U are given in Kemperman (1972, 1983), Haneveld (1985). In a more general case of a completely regular topological space U, dual expressions are given in Kemperman (1983) under some 'tightness' conditions. The results of Chapter 9 are due to Kuznezova-Sholpo and Rachev (1989).

PART III

Chapter 10

A systematic study of the structural and topological properties of $\zeta_{\mathscr{F}}$-metrics was made in the articles of Fortét and Mourier (1953), Rao (1962), Dudley (1966a) and Zolotarev (1976–83). ρ-uniform classes were studied by Prokhorov (1956), Bartoszynski (161), Rao (1962), Dudley (1966a), Bhattacharya and Ranga Rao (1976). The results of this section were stated in Rachev (1985c). Rao (1962) proved Theorem 10.1.3 for $g = -f$.

Chapter 11

In 1953, Fortét and Mourier introduced the metrics $\zeta(\cdot, \cdot; \mathscr{G}^p)$ (see Section 4.3, Example B, Lemma 4.3.1). Their goal was the Glivenko–Cantelli theorem for measures on a separable metric space. Among other results they showed that $\zeta(\mu_n, \mu; \mathscr{G}^1) \to 0$ a.s. where μ_n is the empirical measure based on n observations of μ. By virtue of Theorem 11.1.1 and 6.1.1, this result is equivalent to $\mathscr{A}_c(\mu_n, \mu) \to 0$ a.s. (Theorem 11.1.2). Surprisingly, they did not check that their metric $(\zeta(\cdot, \cdot; \mathscr{G}^1))$ metrizes the weak convergence in $\mathscr{P}(U)$. Three years later Prokhorov introduced his famous metric π in order to metrize the space of non-negative bounded Borel measures on a complete s.m.s. Independently of Fortét and Mourier's work, Kantorovich and Rubinstein (1957, 1958) considered $\zeta(\cdot,\cdot;\mathscr{G}^1)$ on the space of laws on a compact space and noticed that in that case $\zeta(\cdot,\cdot;\mathscr{G}^1)$ metrizes the weak convergence. Dudley (1966a) proved the equivalence of $\zeta(\cdot,\cdot\mid\mathscr{G}^1)$ and π. The results of Section 11.1 are based on the work of Wellner (1981), Dudley (1976) and Rachev (1984d). For some related results, see Shorack and Wellner (1986, pp. 62–65), and Yukich (1989).

The Bernstein CLT deals with the joint weak and moment convergence of the normalized sums of i.i.d. r.v.s (cf. Kruglov 1973, 1976, Acosta and Gine 1979). The functional limit theorems 11.2.1 (Rachev 1984d) is an application of the classical Donsker–Prokhorov theorem (see Prokhorov 1956, Billingsley 1968, Gikhman and Skorokhod 1971).

Chapter 12

In the investigation of queuing systems it is a common technique to replace the complicated stochastic elements that govern the queuing systems by simpler ones which are in some sense 'close' to the real stochastic elements. The constructed queuing model represents the idealization of the real queuing systems and hence the 'stability' problem arises in establishing the limits in which one can use the 'ideal' queuing model. In other words, the stability problem in queuing theory is concerned with the 'domain' within which the ideal queuing model may be applied as a good approximation to the real queuing system under consideration; see Gnedenko (1970), Kennedy (1972), Iglehart (1973), Whitt (1974a,b), Zolotarev (1977a,b), Borovkov (1984, Chap. 2), Stoyan (1983, Chap. 8), Kalashnikov and Rachev (1988), Rachev (1989b). The results in this section are based on Rachev (1984b), Kalashnikov and Rachev (1984) (cf. Section 12.1), Rachev and Rüschendorf (1989b) (cf. Section 12.2), and Anastassiou and Rachev (1990) (cf. Section 12.3).

Chapter 13

Corollary 13.1.1 was proved first by Olkin and Pukelheim (1982) (see also Dowson and Landau 1982, Givens and Shortt 1984). The L^2-Wasserstein metric in Hilbert spaces was investigated by Gelbricht (1988), Cuesta and Matran (1989). Theorem 13.1.1 was proved by Rüschendorf and Rachev (1990); see also Knott and Smith (1984), Smith and Knott (1987).

Explicit representations for $\ell_p(X, Y)$ $(X, Y \in \mathfrak{X}(\mathbb{R}^m), m > 1)$ are not known. The special case considered in Theorem 13.2.1 (Rachev (1984d), Levin and Rachev (1990)) deals only with $p = 1$ and $\|\cdot\| = \|\cdot\|_1$. The concept of optimality of quality usage was essentially introduced by Kantorovich in the late thirties (for details, see Rachev *et al.* 1990).

PART IV

Chapter 14

The approach adopted in Chapter 14 is based on the notion of 'ideal' metric (Zolotarev 1976–1983, 1986, Rachev 1987). 'Ideality' of ζ_2 for stability of characterizations of the exponential law via 'aging' property was noticed in Obretenev and Rachev (1985b). The main results in this section are stated in Baxter and Rachev (1990a); see Tibshirani and Wasserman (1988) for an alternative approach to the use of probability metrics to analyze robustness.

14.2. The structure of ideal metrics is currently an important topic of study in the theory of probability metrics. Zolotarev (1976–1983), Senatov (1980), Kalashnikov and Rachev (1988), Kakosyan *et al.* (1988). Ideal metrics provide a useful apparatus in approximation problems which arise in probability theory and its applications (Zolotarev 1986, Kalashnikov and Rachev 1988). Zolotarev (1976b) introduced ideal metrics with ζ-structure in the space $\mathfrak{X} = \mathfrak{X}(\mathbb{B})$ of Banach space-valued random variables

$$\zeta(X, Y, \mathscr{F}) = \sup\{|\mathbb{E}[f(X) - f(Y)]| : f \in \mathscr{F}\},$$

where $\mathscr{F}$ is a class of integrable functions (cf. (14.2.1)). Other examples of ideal metrics on $\mathfrak{X}(\mathbb{R})$ are given by Rachev and Ignatov (1984) and Maejima and Rachev (1987); see the definition of $\zeta_{m,p}$ (14.2.10). In applications we often use ideal metrics which have several representations. Thus, for instance (cf. Zolotarev (1979a), $p = 1$, Maejima and Rachev (1987), $p > 1$), $\zeta_{m,p}(X, Y) = \|\mathscr{T}_m(F - G)\|_p$, where $\mathscr{T}_m$ is the m-fold integration operator

$$(\mathscr{T}_m f)(x) = \int_{-\infty}^{x} \frac{(x-t)^m}{m!}\, \mathrm{d}f(t) \quad s \geqslant 1$$

$(\mathscr{T}_0 f)(x) = f^{(1)}(x)$ and $F(x)$, $G(x)$ are the distribution functions of the r.v.s X and Y, respectively. The Plancherel theorem ensures yet another representation for the metric $\zeta_{m,2}$,

$$\zeta_{m,2}(X, Y) = \frac{1}{(2\pi)^{1/2}} \left(\int t^{-2s} |f_X(t) - f_Y(t)|^2 \, \mathrm{d}t \right)^{1/2}$$

where $f_X(t)$ is the characteristic function of X.

The use of ideal metrics goes back to Kolmogorov (1931, 1933). Further we shall sketch out the *Kolmogorov method for the CLT by use of ideal type metrics.* (For details, see Borovkov (1983), Rachev and Ignatov (1984).) Let μ be a (K, r)*r-weakly perfect metric* in $\mathfrak{X}(\mathbb{B})$, i.e., μ is a simple probability metric and μ satisfies the following properties: for any $X, Y \in \mathfrak{X}(\mathbb{B})$, (a) $\mu(X + Z, Y + Z) \leqslant \mu(X, Y)$ for any Z independent of X and Y, and (b) $\mu(cX, cY) \leqslant |c|^r \mu(X, Y)$ for any $c \neq 0$, $|c| \leqslant K$ (see Zolotarev 1976b). Clearly a (∞, r)-weakly perfect metric is ideal of order r (cf. Definition 14.2.1). For some $k > 0$ define a non-negative function $g_k(h)$, $h \in \mathbb{R}$ with the following property: if $|c| \leqslant K$, then $g_k(ch) \leqslant |c|^k g_k(h)$, $h \in \mathbb{R}$. Let $1 \leqslant p \leqslant \infty$. Define the 'smoothing' metric $\mu_{k,p}(X, Y) := \|\mu^{(\cdot)}(X, Y)\|_p$, $X, Y \in \mathfrak{X}(\mathbb{B})$, where $\mu^{(h)}(X, Y) := g_k(h)\mu(X + h\Theta, Y + h\Theta)$ and the r.v. Θ is independent of the pair (X, Y). It is easily seen that $\mu_{k,p}$ is a $(K, r + k + 1/p)$-weakly perfect metric. To ensure nontriviality of the metric $\mu_{k,p}$ (see Theorem 2, Rachev and Ignatov 1984), we assume that Θ depends on k, e.g., $\Theta = \sum_{i=1}^n \Theta_i$, where $n \geqslant k$ and $\Theta_1, \ldots, \Theta_n$ are independent identically distributed r.v.s having a density, $\mathbb{B} = \mathbb{R}^k$. The metrics $\mu_{k,p}$ are essentially used in Kolmogorov (1931, 1933). As an example of the use of $\mu_{k,p}$, let us consider the proof of the central limit theorem assuming the Lindeberg condition for absolute pseudomoments (cf. Zolotarev 1967ab, 1970). Let $\xi_{1,n}, \xi_{2,n}, \ldots$, $n = 1, 2, \ldots$ be a sequence of independent r.v.s in the series scheme $S_{k,n} = \sum_{j=1}^n \xi_{j,n}$, $\mathbb{E}\xi_{j,n} = 0$, $B_{j,n}^2 = \operatorname{Var} \xi_{j,n}$, $B_n^2 = \sum_{j=1}^n b_{j,n}^2$, $F_{k,n}(x) = \Pr(\xi_{k,n} \leqslant xB_n)$, $\zeta_n = S_{n,n}/B_n$. By $N(t)$ we denote a $(0, t)$-normally distributed random variable. Let $N_{k,n}$ be the distribution function $N(b_{k,n}^2/B_n^2)$. Denote

$$M_{k,n}(\tau) = \int_{|u|>\tau} u^2 |\,\mathrm{d}(F_{k,n}(u) - N_{k,n}(u))| \quad M_n(\tau) = \sum_{k=1}^n M_{k,n}(\tau).$$

Let τ_n be the solution of the equation $M_n(\tau) = \tau$. By the Lindeberg condition, $\tau_n \to 0$ for $n \to \infty$. Consider the $(1, 2)$-perfect metric

$$\mathbf{W}(X, Y) = \sup_{h>0} (h^2 \wedge h^3))\boldsymbol{\rho}(X + hU^{(3)}, Y + hU^{(3)})$$

where $\wedge$ = min and $U^{(3)}$ is a sum of three i.i.d. uniforms independent of X and Y. It can be shown (cf. Kolmogorov 1931, 1933, Borovkov 1983, Rachev and Ignatov 1984) that $\mathbf{W}(\zeta_n, N(1)) \leqslant 2\tau_n$. Also note that from the inequality $\boldsymbol{\rho}(X, Y) \leqslant 4M^{3/4}\mathbf{W}^{1/4}(X, Y)$, for

$$\mathbf{W}(X, Y) \leqslant M = \sup_{x \in \mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}x} \Pr(Y < x)$$

we obtain $\boldsymbol{\rho}(\zeta_n, N(1)) < 4\tau_n^{1/4}$. The metric approach used to construct the above bounds is simple and clear and provides a construction to prove limit theorems for different schemes, summation, maxima, generalized sums (Urbanik convolutions), convolution of random motion, etc.

14.3. The rate of convergence to α-stable distribution was investigated by Zolotarev (1962), Cramer (1963), Paulauskas (1974), Mitalauskas and Statulevicius; (1976), Christoph (1979a,b, 1980), Hall (1981), Dubinskaite (1983), Mijnheer (1975, 1983).

The approach we accepted (see Rachev and Yukich 1989, 1990) is based on the combination of the method of metric distances (the use of ideal metrics) and the classic method of Bergström, which provides the rate of convergence in the CLT with respect to the uniform metric (cf. Bergström 1945, Senatov 1980, Sazonov 1972, Sazonov and Ul'yanov 1979).

The topological structure of the space of measures supplied with a convolution metric was studied by Ignatov and Rachev (1983), Yukich (1985), Rachev and Ignatov (1984), Zolotarev (1986), and Rachev and Yukich (1990).

Chapter 15

The rate of convergence problem in the central limit theorem (CLT) for α-stable random variables with respect to summation and maxima of i.i.d. random elements arises in various areas of probability theory and its applications, including

(a) limit theorems for sums of random elements (see Ibragimov and Linnik 1971, Paulauskas 1976, Hall 1981, Banys 1976, Sazonov 1981, Zolotarev 1986, Maejima and Rachev 1987, Rachev and Yukich 1989, Rachev and Rüschendorf 1989b and references there),
(b) limit theorems for maxima of random elements (see Smith 1982, Cohen 1982, Zolotarev and Rachev 1985, Resnick 1987b, Omey and Rachev 1988, de Haan and Rachev 1989, Balkema and de Haan 1988),
(c) stochastic models of economic phenomena (see Du Mouchel 1971, Mittnik and Rachev 1989, Rachev and Todorovich 1989), and
(d) queuing models (see Borovkov 1984, Kalashnikov and Vsekhsvyatskii 1985, Rachev 1989a,b).

The main similarity in (a) and (b) involves the fact that summation and maxima are commutative operations. Hohlov (1986) considers the rate of convergence problem for α-stable distributions on the noncommutative group of motions in $\mathbb{R}^d$. His results are based on the CLT for random motions studied by Tutubalin (1967), Roynette (1974), Hohlov (1982, 1986), Grincevicius (1985). The integral and local limit theorems 15.2.1 and 15.2.2 are proved by Rachev and Yukich (1990).

Chapter 16

A general scheme of stability of stochastic models is given by Zolotarev (1977a, 1983a,b). Stability of queuing models was analyzed by many authors (see the references in Kalashnikov and Rachev (1988)). The problem of stability in insurance mathematics was analyzed by Beirlant and Rachev (1987). Similar results can be obtained regarding the maximal and minimal claim amounts.

Chapter 17

Gerber (1981) and Goovaerts *et al.* (1984) are excellent surveys on mathematical risk theory. For the usual compound Poisson model S^{coll} (17.1.5), several bounds have been proved via different choices of V_i and μ (cf. (17.1.3)–(17.1.6)): for the total variation metric

$$\sigma(S^{\text{ind}}, S^{\text{coll}}) \leqslant \sum_{j=1}^{n} p_j^2 \quad \text{(cf. Le Cam 1986 p. 424),}$$

and for the uniform metric,

$$\rho(S^{\text{ind}}, S^{\text{coll}}) \leqslant c \max_{1 \leqslant j \leqslant n} p_j \quad \text{(cf. Zaitsev 1983).}$$

For $V_i = E_i$—the one point distribution at 1—i.e., for the problem of the Poisson approximation of sums of binomials (cf. (17.1.6)), some different choices of μ turned out to be better for 'small' p_i, i.e., $\sum_{j=1}^{n} p_j \to a$, $0 \leqslant a \leqslant \sqrt{2}$ (cf. Serfling 1975, Deheuvels and Pfeifer 1986, Presman 1983). For a comprehensive discussion of these results we refer to the recent book by Arak and Zaitsev (1988). The results of this section are due to Rachev and Rüschendorf (1990b), where a numerical example comparing the distributions of S^{ind}, S^{coll} (traditional model (17.1.5), (17.1.6)) and S^{coll} (scaled model (79.2.1), 17.2.2)) is also given.

Chapter 18

The extreme value theory is the main topic of five well-known monographs (de Haan 1970, Galambos 1978, Leadbetter *et al.* 1980, Resnick 1987a, Reiss 1989).

The results of Section 18.1 are based on Omey and Rachev (1988, 1990). The rate of convergence in univariate extreme value theory has been studied extensively. Smith (1982) relates uniform rates of convergence to slow variation with remainder. In his paper he considers domain of attraction of $\phi_\alpha(x)$ and $\psi_\alpha(x)$. Balkema and de Haan (1988) obtain uniform rates of convergence to the limit d.f. $\Lambda(x)$. Omey and Rachev (1988) consider rates of convergence in terms of probability metrics. We also refer to Leadbetter *et al.*, (1980), Resnick (1987a,b), Omey (1988), Zolotarev (1983), de Haan and Rachev (1989), Reiss (1989) for the discussion related to this topic. For some generalizations of the results in Section 18.1, including asymptotic expansions, see Omey and Rachev (1988, 1990). Zolotarev (1985) suggested $\boldsymbol{\rho}_r$ as an ideal metric for maxima of i.i.d. random variables. Rachev (1986), de Haan and Rachev (1989) studied the max-ideality of ℓ_p-minimal metrics in the space of Banach-valued random variables. Zolotarev (1976b) introduced the difference pseudomoments; their 'ideality' properties for maxima scheme were studied by de Haan and Rachev (1989), Levin and Rachev (1990). The results of Section 18.2 are due to de Haan and Rachev (1989). One could easily reformulate the results of Section 18.2 considering the rate of convergence for the CLT for summability methods (see Maejima (1988) and the references there). The role of an ideal metric in this case will be played by $\zeta_{m,p}$ (14.2.10).

Zolotarev (1976) conjectured that, w.r.t. the summation scheme, compound ideal metrics of order greater than 1 do not exist (see Ignatov and Rachev (1983) for a detailed discussion). Rachev (1986) and Rachev and Rüschendorf (1989b) showed that max-ideal compound metrics for any order do exist. For a comprehensive discussion on limit theorems for Banach-valued random variables, we refer to Hoffman-Jorgensen and Pisier (1976), Hoffman-Jorgensen (1977), Pisier and Zinn (1977), Woyczynski (1980). The properties of the ideal metrics ζ_s were studied by Zolotarev (1976a,b, 1977a,b, 1979a,b, 1981, 1983b), Senatov (1980, 1981), Bentkus and Rachkauskas (1985). The main results in Section 18.3 are due to Rachev and Rüschendorf (1989b) where an application to the stability of queuing models is also given.

Chapter 19

Stability problems for characterization of distributions were discussed in the monographs of Kagan *et al.* (1972), Kakosyan *et al.* (1984), Azlarov and Volodin (1986) (see also Galambos and Kotz 1978, Kakosyan *et al.* 1988). A general scheme for stability of characterizations of distributions was proposed by Zolotarev (1976a, 1977a, 1979a). Sections 19.1 and 19.2 are motivated by the paper of Diaconis and Freedman (1987a). The results in these two sections are due to Rachev and Rüschendorf (1989a), where some additional results concerning stability of F_p-class can be found. For additional relevant references we refer to Berk (1977), Csiszar (1967), Diaconis and Freedman (1984a,b, 1987a,b), Kullback (1959), Ressel (1985), Zabell (1980). Diaconis and Freedman (1987a) investigated the stability of the F_p-distributions (for $p = 1, 2$) with respect to the total variation metric. The proof of Theorem 19.2.2 is based on their method. The Kolmogorov metric $\boldsymbol{\rho}$ seems to be suitable for the stability problem since two-side estimates for F_p-stability can be found (Section 19.1). To carry out an analysis of F_p-stability ($p < \infty$) w.r.t. to the total variation metric $\boldsymbol{\sigma}$, one needs refine Berry–Esseen estimates in the CLT w.r.t. $\boldsymbol{\sigma}$ (see Chapter 14, Senatov 1980, Sazonov 1981). Similarly in the case of $p = \infty$ one needs estimates of the rate of convergence in the extreme value limit theorems w.r.t. $\boldsymbol{\sigma}$ (cf. Omey 1988). The characterization of the 'discrete' version of the class F_p is related to the famous Waring problem in number theory (see Vinogradov 1975, Ellison 1971 and Rachev and Rüschendorf 1989a for details).

19.3. Section 19.3 is inspired by the paper of Todorovich and Gani (1987) who consider the stochastic model (19.3.1) in the special case of $Z \in (0, 1)$; see also Todorovich *et al.* (1987). Puri (1987) proved the following: if $\{Y_n\}_1^\infty$ and $\{Z_n\}_1^\infty$ are independent, $\{Y_n\}_1^\infty$ strictly stationary, $\{Z_n\}_1^\infty$ is an i.i.d. sequence with $P\{0 < Z \leqslant 1\}$ and $S_n = \sum_{k=1}^n Y_k \prod_{j=1}^k Z_j$, then $S_n \to S$ a.s. provided that $\mathbb{E}(\ln|Y|) < \infty$. The equation (19.3.2) regarding maximal crop yield for n years was studied by Todorovich (1987) where the weak convergence of M_n to M (cf. (19.3.11)) was proved. Equation (19.3.3) with $Z = 1$ determines the class of geometric infinitely divisible laws (Klebanov *et al.* 1984). Similarly, the distribution of H satisfying (19.3.4) is called geometric max-infinitely divisible distribution (Rachev and Resnick 1989). Theorem 19.3.1–19.3.6 are due to Rachev and Todorovich (1989). Lemma 19.3.2 is proved by Rachev and Resnick (1989). For additional references and possible generalizations see Rachev and Samorodnitski (1990).

Additional examples of the use of the theory of probability metrics in the problems of stability of stochastic models can be found in the following proceedings.

(1) (1980) *Stability Problems for Stochastic Models*, *Proceedings*, VNIISI, Moscow (in Russian); (Engl. transl. (1986) *J. Sov. Math.*, **32**)
(2) (1981) *Stability problems for Stochastic Models*, *Proceedings*, VNIISI, Moscow (in Russian); (Engl. transl., (1986) *J. Sov. Math.*, **34**)
(3) (1982) *Stability Problems for Stochastic Models*, *Proceedings*, VNIISI, Moscow (in Russian); (Engl. trans. (1986) *J. Sov. Math.*, **35**)
(4) (1983) *Stability Problems for Stochastic Models*, *Proceedings*, VNIISI, Moscow (in Russian); (Engl. transl., (1986) *J. Sov. Math.*, **32**)
(5) (1983) *Stability Problems for Stochastic Models*, (*Lect. Notes Math.* **982**), Springer, Berlin.
(6) (1984) *Stability Problems for Stochastic Models*, *Proceedings*, VNIISI, Moscow, (in Russian); (Engl. transl., (1986) *J. Sov. Math.*, **35**)
(7) (1985) *Stability Problems for Stochastic Models*, *Proceedings*, VNIISI, Moscow, (in Russian); (Engl. transl. (1988) *J. Sov. Math.*, **40**).
(8) (1985) *Stability Problems for Stochastic Models*, (*Lect. Notes Math.* **1152**) Springer, Berlin.
(9) (1986) *Stability Problems for Stochastic Models*, (*Lect. Notes Math.*, **1233**) Springer, Berlin.
(10) (1989) *Stability Problems for Stochastic Models*, (*Lect. Notes Math.*, **1412**) Springer, Berlin.

Ideal metrics and their applications to the rate of convergence to self-similar processes were studied by Maejima and Rachev (1987). Extensive study of ideal metrics related to the CLT are given by Zolotarev (1986). Ideal metrics for approximation of queuing models are described by Kalashnikov and Rachev (1988).

References

* Additional references related to the context of the book

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions.* 9th Printing, Dover, New York.

Ahiezer, N. I. and Krein, M. (1962). *Some Questions in the Theory of Moments.* American Mathematical Society, Providence, RI.

*Akaike, H. (1981). Modern development of statistical methods. In: *Trends and Progress in System Identification.* P. Eykhoff (ed.), Pergamon, Oxford, pp. 169–184.

Anastassiou, G. A. and Rachev, S. T. (1989). Approximation of a random queue by means of deterministic queueing models. In: *Approximation Theory VI*, Vol. 1, C. K. Chui, L. L. Shumaker, and J. D. Ward (eds.), Academic Press, New York, pp. 9–11.

Anastassiou, G. A. and Rachev, S. T. (1990). Moment problems and their applications to the stability of queueing models. *Comput. Math. Appl.* (to appear).

Anderson, E. J. and Nash, P. (1987). *Linear Programming in Infinite-Dimensional Spaces.* Wiley, New York.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

Appell, P. (1884). Mémoires sur les děblais et remblais des systèmes continués et discontinués. *Mémoires des sav. etr., 2 ser., 29, 3 mémoire.*

Appell, P. (1928). Le probléme géomětrique des děblais et remblais. *Mem. des Sci. Math., Paris.: Fasc.* **27**, 1–34.

*Arak, T. V. (1981). On the convergence rate in Komogorov's uniform limit theorem I and II. *Theory Prob. Appl.*, **26**, 219–239, 437–451.

Arak, T. V. and Zaitsev, A. Yu. (1988). *Uniform Limit Theorems for Sums of Independent Random Variables.* Proc. Steklov Institute of Math. 174, AMS.

*Arjas, E. (1981). A stochastic process approach to multivariate reliability systems: notions based on conditional stochastic order. *Math. Oper. Res.*, 263–276.

*Arnold, B. C. (1975). A characterization of the exponential distribution by multivariate geometric compounding. *Sankhya*, **37A**, 164–173.

Azlarov, T. A. and Volodin, N. A. (1986). *Characterization Problems Associated With the Exponential Distribution.* Springer, Berlin.

*Baldessari, B. (1974). Metodi di quantificazione basati sugli indici di dissomiglianza di inequaglianza, *Metron*, **XXXII**, 197–219.

Baldi, P. (1979). Lois stables sur les deplacements de R^n. In: *Probability Measures on Groups.* (*Lect. Notes Math.*, **706**,) Springer, Berlin, pp. 1–9.

Balinski, M. L. and Gonzalez, J. (1988). *Maximum matchings in bipartite graphs via strong spanning trees.* Technical Report, January.

*Balinski, M. L. and Rispoli, F. (1988). On signature methods and the Hirsch conjecture for transportation polytopes. In preparation.

Balinski, M. and Rachev, S. T. (1989). *On Monge–Kantorovich problems.* Preprint, SUNY at Stony Brook, Dept. of Applied Mathematics and Statistics.

*Balinski, M. L. and Russakoff, A. (1984). Faces of dual transportation polyhedra. *Math. Progr. Study*, **22**, 1–8.

*Balinski, M. L. and Young, H. P. (1982). *Fair Representation: Meeting the Ideal of One Man, One Vote.* Yale University Press, New Haven.

Balkema, A. A. and De Haan, L. (1988). *A convergence rate in extreme-value theory.* Preprint, Erasmus Universiteit, Rotterdam.

Balkema, A. A., de Haan, L., and Karandikar, R. L. (1990). *The maximum of independent stochastic processes.* Preprint, Erasmus University, Rotterdam.

Banys, J. (1976). On the convergence rate in the multidimensional local limit theorem with limiting stable law. *Lithuanian Math. J.* **16**, 320–325.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart, and Winston, New York.

Barnes, R. and Hoffman, J. (1985). On transportation problems with upper bounds on leading rectangles. *SIAM J. Alge. Discr. Meth.*, **6**, 487–496.

*Barsilovich, I. Ju., Beljaev, Ju. K., Solovyev, A. D. and Ushakov, I. A. (1983). *Problems of Mathematical Reliability Theory.* B. U. Gnedenko (ed.) Radio and Svjaz, Moscow. (in Russian).

Bartfai, P. (1970). Uber die Entfernung der Irrahrttswegge. *Studia Sci. Math. Hungar.*, **V**, 41–49.

Bartoszynski, R. (1961). A characterization of the weak convergence of measures. *Ann. Math. Stat.*, **32**, pp. 561–576.

Bass, J. (1955). Sur la compatibilité des fonctions de repartition. *C.R. Acad. Sci.* **240A**, 839–841.

*Basu, A. P. and Block, H. W. (1975). On characterizing univariate and multivariate exponential distributions with applications. *Statistical Distributions in Scientific Work*, Vol. 3. G. P. Patil, S. Kotz, and J. Ord (eds.) Reidel, Dordrecht, pp. 399–491.

Basu, A. P. and Ebrahimi, N. (1985). Testing whether survival function is harmonic new better than used in expectation. *Ann. Inst. Stat. Math.*, **37**, 347–359.

Basu, A. P. and Ebrahimi, N. (1986). HNBUE and HNWUE distributions. In: *A survey. Reliability and Quality Control*, A. P. Baum, (ed.), North-Holland, Amsterdam.

*Basu, A. P., Ebrahimi, H. and Klefsijo, B. (1983a). Multivariate HNBUE distributions. *Scand. J. Statist.*, **10**, 19–25.

*Basu, A. P., Ebrahimi, N., and Klefsijo, B. (1983b). Multivariate harmonic new better than used in expectation. *Ann. Inst. Statist. Math.*, **37A**, 347–359.

*Basu, S. (1976). On a local limit theorem concerning variables in the domain of normal attraction of a stable law of index α, $1 < \alpha < 2$. *Ann. Prob.*, **4**, 486–489.

Basu, S. K. and Simons, G. (1983). Moment spaces for IFR distributions, applications, and related material. *Contributions in Statistics: Essays in Honour of Norman L. Johnon*, P. K. Sen (ed.), North-Holland, Amsterdam, pp. 27–46.

*Bauer, H. (1972). *Probability Theory and Elements of Measure Theory*. Holt, Rinehart, and Winston, New York.

Baxter, L. A. and Rachev, S. T. (1990a). A note on the stability of the estimation of the exponential distribution. *Stat. Prob. Lett.*, **10**, 37–41.

Baxter, L. A. and Rachev, S. T. (1990b). The stability of a characterization of the bivariate Marshall-Olkin distribution. *J. Math. Anal. Appl.* To appear.

Bazaraa, M. S. and Jarvis, J. J. (1977). *Linear Programming and Network Flows*. Wiley, New York.

*Beasley, E. T. and Gibson, P. M. (1975). A generalization of doubly stochastic matrices over a field. *Lin. Algebra. Appl.*, **10**, 241–257.

*Behnen, K. and Neuhaus, G. (1975). A central limit theorem under contiguous alternatives. *Ann. Statist.* **3**, 1349–1353.

Beirlant, J. and Rachev, S. T. (1987). The problems of stability in insurance mathematics, *Insurance: Math. Econ.* **6**, 179–188.

*Benson, R. (1966). *Euclidean Geometry and Convexity*. McGraw-Hill, New York.

Bentkus, V. Yu. and Rackauskas, A. (1985). Estimates of the distance between sums of independent random elements in Banach spaces. *Theory Prob. Appl.* **29**, 50–65.

Berge, C. and Chouila-Houri, A. (1965). *Programming, Games, and Transportation Networks*. Wiley, New York.

Bergstrom, H. (1945). On the central limit theorem in the space R^k, $k > 1$. *Skand. Aktuatietdskr.*, **28**, 106–127.

Berk, R. H. (1977). Characterizations via conditional distributions. *J. Appl. Prob.*, **14**, 806–816.

Berkes, I. and Phillip, W. (1979). Approximation theorems for independent and weakly independent random vectors. *Ann. Prob.* **7**, 29–54.

*Berkes, I., Dabrowski, A., Dehling, H. and Phillip, W. (1986). A strong approximation theorem for sums of random vectors in the domain of attraction to a stable law. *Acta. Math. Hungar.*, **48**, 161–172.

Bernshtein, S. N. (1964). *Collected Works*. Vol. 4. Nauka, Moscow. (In Russian).

Bertino, S. (1966). Una generalizzazione della dissomiglianza di C. Gini tra variabili casuali semplici. *Giorn. Ist. Italiano Attuari*, **29**, 155–178.

Bertino, S. (1971). Gli indici di dissomiglianza e la stima dei parametri. In: *Studi di* *probabilità, Statistica e ricerca operativa in onore di Guiseppe Pomlilej*, *Oderisi*, Gubbio, Rome.

Bertino, S. (1972). Sulla media e la varianza dell'indice semplice di dissomiglianza caso di campioni provenienti da una stessa variable casuale assolutamente continua. *Metron*, **XXX**, 256–281.

Bhattacharya, R. M. and Ranga Rao, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

*Bickel, P. J. and Breiman, L. (1983). Sum of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Ann. Prob.*, **11**, 185–219.

*Bickel, P. J. and Doksum, K. (1960). Tests for monotone failure rate based on normalized spacings. *Ann. Math. Stat.*, **40**, 1216–1235.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Stat.*, **9**, 1196–1217.

*Bildikar, S. and Patil, G. P. (1968). Multivariate exponential type distributions. *Ann. Math. Stat.*, **39**, 1316–1326.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

Birnbaum, Z. W. and Orliz, W. (1931). Über die verallgemeinerung des begriffes der zueinander Konjugierten Potenzen. *Studia Math.*, **3**, 1–67.

*Blackwell, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Stat.*, **24**, 265–272.

*Blackwell, D. (1956). *On a class of probability spaces. Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 2, University of California Press, Berkeley, pp. 1–6.

*Block, H. W. and Basu, A. P. (1974). A continuous bivariate exponential distribution. *J. Amer. Statist. Assoc.*, **69**, 1031–1037.

*Block, H. W. and Savits, T. H. (1980). Multivariate IFRA distributions. *Ann. Prob.*, **8**, 793–801.

*Block, H. W. and Savits, T. H. (1981). Multivariate classes in reliability theory. *Math. Oper. Res.*, **6**, 453–461.

*Block, H. W., Savits, T. H., and Shaked, M. (1982). Some concepts of negative dependence. *Ann. Prob.*, **10**, 765–772.

*Bochner, S. (1956). Positive zonal functions on spheres. *Proc. Nat. Acad. Sci. USA*, **40**, 1141–1167.

*Borovkov, A. A. (1983). Limit problems, the invariance principle, large deviations. *Usp. Math. Nauk.*, **28**, 227–254.

Borovkov, A. A. (1984). *Asymptotic Methods in Queueing Theory*. Wiley, New York.

Breiman, L. (1968). *Probability*. Addison-Wesley, Reading.

*Brown, J. R. (1965). Doubly stochastic measures and Markov operators. *Michigan Math. J.*, **2**, 367–375.

*Brown, J. R. and Shiflett, R. C. (1970). On extreme doubly stochastic measures. *Michigan Math. J.*, **2**, 249–254.

*Brown, M. (1980). Bounds, inequalities, and monoicity properties for some specialized renewal processes. *Ann. Prob.*, **8**, 227–240.

*Brown, M. (1983). Approximating IMRL distributions by exponential distributions, with applications to first passage times. *Ann. Prob.*, **11**, 419–427.

*Brown, M. (1987). Inequalities for distributions with increasing failure rate. In: *Contributions to the theory and Application of Statistics*. A. E. Gelfand (ed.) Academic, New York, pp. 3–17.

*Brown, M. and Ge, G. (1984). Exponential approximations for two classes of aging distributions. *Ann. Prob.*, **12**, 869–875.

*Brualdi, R. A. (1968). Convex sets of non-negative matrices. *Can. J. Math.*, **20**, 144–157.

*Brualdi, R. A. and Czima, J. (1975). Extremal plane stochastic matrices of dimensional three. *Linear Algebra Appl.*, **11**, 105–133.
*Brualdi, R. A., Parter, S. V., and Schneider, H. (1966). The diagonal equivalence of a non-negative matrix to a stochastic matrix. *J. Math. Anal. Appl.*, **16**, 31–50.
Cambanis, S. and Simons, G. (1982). Probability and expectation inequalities. *Z. Wahrsch. Verw. Geb.*, **59**, 1–25.
Cambanis, S., Simons, G., and Stout, W. (1976). Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrsch. Verw. Geb.*, **36**, 285–294.
Cassando, E., Olivieri, E., Pellegrinotti, A., and Presutti, E. (1978). Existence and uniqueness of DLR measures for unbounded spin systems. *Z. Wahrsch. Verw. Geb.*, **41**, 313–334.
Chow, Y. S. and Teicher, H. (1978). *Probability Theory: Independance, interchangeability, martingales.* Springer, New York.
*Christensen, J. P. R. (1974). *Topology and Borel Structure.* North-Holland–Elsevier, Amsterdam.
Christoph, G. (1979a). Convergence rate in integral limit theorem with stable limit law. *Lithuanian Math. Trans.*, **XIX**, 91–101.
Christoph, G. (1979b). *Konvergenzaussagen und asymptotische Entwicklungen für die Verteilungsfunktion einer Summe unabhangiger identisch verteilter Zufallsgroben im Falle einer stabilen grenzverteilungs funktion.* Akademie der Wissenschaften der DDR, Berlin.
Christoph, G. (1980). Über notwendige und hinreichende Bedingen für Konvergenzgeschwindigkeitsaussagen im Falle einer stabilen Grenzverteilung. *Z. Wahrsch. Verw. Geb.*, **54B**, 29–40.
*Christoph, G. (1983). *Asymptotic expansions for sums of random variables and of reciprocals of these with a non-normal stable limit law.* Preprint. Technische Universität, Dresden.
*Clayton, D. G. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.
Cohen, J. P. (1982). Convergence rate for the ultimate and penultimate approximation in extreme-value theory. *Adv. Appl. Prob.*, **14**, 833–854.
Cohn, D. L. (1980). *Measure Theory.* Birkhauser, Boston.
*Cosares, S. (1988). *Some properties of primal and dual transhipment polyhedra.* Preprint, Dept. of Industrial Engineering and Operations Research, University of California, Berkeley.
*Cox, G. V. (1980). Lusin properties in the product space S. *Fund. Math.*, **108**, 199–120.
*Cox, L. H. and Ernst, L. R. (1982). Controlled rounding. *INFOR*, **20**, 423–432.
Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, New York.
Cramer, H. (1963). On asymptotic expansions for sums of independent random variables with a limiting stable distribution. *Sankhya* **a25**, 13–24.
Csiszar, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, **2**, 299–318.
*Csiszar, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Ann. Prob.* **3**, 146–158.
Cuesta, J. A. and Matran, C. (1989). Notes on the Wasserstein metric in Hilbert spaces. *Ann. Prob.*, **17**, 1264–1278.
*Daley, D. J. (1988). Tight bounds on the exponential approximation of some aging distributions. *Ann. Prob.*, **16**, 414–423.
Dall'Aglio, G. (1956). Sugli estremi dei momenti delle funzioni di ripartizione doppie. *Ann. Scuola Normale Superiore Di Pisa, Cl. Sci.*, **3**, 33–74.

*Dall'Aglio, G. (1960). Les fonctions extrèmes de la classe de Fréchet à 3 dimensions, *Publ. Inst. Stat. Univ. Paris*, **IX**, 175–188.

*Dall'Aglio, G. (1961). Sulle distribuzioni doppie con margini assegnati soggette a delle limitazioni. *Giorn. Ist. Ital. Attuari*, **1**, 94.

*Dall'Aglio, G. (1972). Fréchet classes and compatibility of distribution functions, *Symp. Math.*, **9**, 131–150.

*D'Aristotile, A., Diaconis, P., and Freedman, D. (1988). *On a merging of probabilities.* Technical Report, vol. 301, Dept of Statistics, Stanford University.

*Darst, R. B. (1971). On universal measurability and perfect probability. *Ann. Math. Stat.*, **42**, 352–354.

*Davies, R. O. (1971). A non-Prokhorov space. *Bull London Math. Soc.*, **3**, 341–342.

*Day, P. W. (1972). Rearrangement inequalities. *Can. J. Math.* **XXIV**, 930–943.

de Acosta, A. (1982). Invariance principles in probability for triangle arrays of B-valued random vectors and some applications. *Ann. Prob.*, **10**, 346–373.

*de Acosta, A. and Gine, E. (1979). Convergence of moments and related functionals in the general central limit theorem in Banach spaces. *Z. Wahrsch. Verw. Geb.*, **48**, 213–231.

de Haan, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tract 32, Mathematics Centre, Amsterdam, Holland.

de Haan, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.*, **12**, 1197–1204.

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

*de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate samples extremes. *Z. Wahrsch. Verw. Geb.*, **40**, 317–333.

*Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation. *Ann. Prob.* **14**, 663–678.

*Deland, P. N. and Shiflett, R. C. (1980). Covariance linear functionals on doubly stochastic measures. *Appl. Anal.*, **10**, 213–220.

*Dellacherie, C. and Meyer, P. A. (1983). *Probabilités et potentiel.* Herrmann, Paris. Ch. IX–XI.

Derigs, U., Goecke, O., and Schrader, R. (1986). Monge sequences and a simple assignment algorithm. *Discr. Appl. Math.*, **15**, 241–248.

*Deshpande, J. V. (1983). A class of tests against IFRA alternatives. *Biometrika*, **70**, 531–535.

*Deshpande, J. V. and Kochar, S. C. (1983). A linear combination of two U-statistics for testing new better than used. *Comm. Statist.*, **12**, 153–160.

*Deshpande, J. V. and Kochar, S. C. (1985). A new class of tests for testing exponentiality against positive aging. *J. Ind. Statist. Assoc.*, **23**, 89–96.

*Deshpande, J. V. and Kusum, K. (1987). Two sample distribution free tests for shape. *J. Ind. Stat. Assoc.*, **25**, 13–20.

*Deshpande, J. V., Kalpana, K. and Kochav, S. C. (1986). Testing the star equivalence of two probability distributions. *J. Ind. Statist. Assoc.*, **24**, 21–30.

*Deshpande, J. V., Bagai, J. and Kochar, S. C. (1989a). A distribution free test for the equality of failure rates due to two competing risks. *Comm. Statist. (Theor. Meth.)*, **18**, 107–120.

*Deshpande, J. V., Bagai, J. and Kochar, S. C. (1989b). Distribution free tests for stochastic ordering among two independent competing risks. *Biometrika.* Submitted.

*Diaconis, P. and Freedman, D. (1979). On rounding percentages. *J. Amer. Stat. Assoc.*, **74**, 359–364.

*Diaconis, P. and Freedman, D. (1984a). Parital exchangeability and sufficiency. In: *Sufficiency in Statistics, Applications and new Directions.* J. K. Ghosh and J. Roy, (eds.) *Proc. Indian Stat. Inst., Calcutta*, pp. 205–236.

Diaconis, P. and Freedman, D. (1984b). *A note on weak star uniformities.* Technical report, 39, Dept. of Statistics, University of California, Berkeley.

Diaconis, P. and Freedman, D. (1987a). A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincare*, **23**, 397–623.

Diaconis, P. and Freedman, D. (1987b). *A finite version of de Finetti's theorem for exponential families with uniform asymptotic estimates.* Technical Report, University of California.

*Diego, H. and Germani, H. (1972). Extreme measures with prescribed marginals. *J. Combin. Theor.*, **13**, 353–366.

*Dimitrov, B., Klebanov, L., and Rachev, S. T. (1982). Stability of the characterization of the exponential law. In: *Stability Problems for Stochastic Models. Proceedings*, Moscow, VNIISI, pp. 39–76. (In Russian). (Engl. transl. (1986), *J. Sov. Math.*, **35**, 2479–2485)

*Dobrushin, R. L. (1968a). Gibbsian random fields for lattice systems with pairwise interaction. *Funct. Anal. Appl.*, **2**, 292–301.

*Dobrushin, R. L. (1968b). Uniqueness problem for a Gibbsian random field and the problem of phase transitions. *Funct. Anal. Appl.*, **2**, 302–312.

Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theor. Prob. Appl.*, **15**, 458–486.

*Dobrushin, R. L. (1979). Gaussian and their subordinated generalized fields. *Ann. Prob.*, **7**, 1–28.

Dotto, O. J. (1989). A note on p-metrics on $M_p(S)$. *J. Math. Anal. Appl.*, **142**, 403–412.

*Douglas, R. G. (1964). On extremal measures and subspace density. *Michigan Math. J.*, **11**, 243–246.

*Douglas, R. G. (1966). Extremal measures and subspace density, II. *Proc. Am. Math. Soc.*, **17**, 1363–1365.

Dowson, C. D. and Landau, B. U. (1982). The Fréchet distance between multivariate normal distributions, *J. Multivar. Anal.*, **12**, 450–455.

Dubinskaite, J. (1983). On the accuracy of stable approximation of the distribution functions of sums independent random variables. *Lithuanian Math. Trans.*, **XXIII**, 74–91.

Dudley, R. M. (1966a). Convergence of Baire measures, *Studia Math.*, **XXVII**, 251–268.

*Dudley, R. M. (1966b). Weak convergence of probability measures on non separable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.*, **10**, 109–126.

*Dudley, R. M. (1967). Measures on non separable metric spaces. *Illinois J. Math.*, **11**, 449–453.

Dudley, R. M. (1968). Distances of probability measures and random variables. *Ann. Math. Statist.*, **39**, 1563–1572.

Dudley, R. M. (1969). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Statist.*, **40**, 40–50.

*Dudley, R. M. (1972). Speeds of metric probability convergence. *Z. Wahrsch. Verw. Geb.*, **22**, 323–332.

Dudley, R. M. (1976). *Probabilities and Metrics: Convergence of Laws on Metric Spaces, With a View to Statistical Testing.* Aarhus University Mathematics Institute Lecture Notes Series no. 45, Aarhus.

Dudley, R. M. (1989). *Real Analysis and Probability.* Wadsworth & Brooks–Cole, Pacific Grove, California.

*Duffin, R. J. (1956). Infinite programmes, linear inequalities and related systems. *Ann. Meth. Stud.*, **38**, 157–170.

*Duffin, R. J. and Karlowitz, L. A. (1965). An infinite linear program with a duality gap. *Management Sci.*, **12**, 122–134.

*Du Mouchel, W. (1971). *Stable Distributions in Statistical Inference.* PhD, Yale University.

Dunford, N. and Schwartz, J. (1988). *Linear operators. Vol. 1.* Wiley, New York.

*Ebrahimi, N. and Ghosh, M. (1981). Multivariate negative dependence. *Comm. Statist.*, **A10**, 307–337.

*Edwards, D. A. (1978). On the existence of probability measures with given marginals. *Ann. Inst. Fourier*, **28**, 53–78.

Ellison, W. J. (1971). Waring's problem. *Amer. Math. Monthly*, **78**, 10–36.

Erdos, P. and Spencer, J. (1974). *Probabilistic Methods in Combinatorics.* Academic, New York.

*Esary, J. D. and Marshall, A. W. (1974). Multivariate distributions with exponential minimum. *Ann. Stat.*, **2**, 84–93.

*Esary, J. D. and Marshall, A. W. (1979). Multivariate distributions with increasing hazard rate average. *Ann. Prob.*, **7**, 354–370.

*Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *Ann. Math. Stat.*, **38**, 1466–1474.

*Feller, W. (1946). A limit theorem for random variables with infinite moments. *Am. J. Math.*, **68**, 257–262.

Feller, W. (1970). *An Introduction to Probability Theory and its Applications.* Vol. 1, 3rd edn., Wiley, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* Vol. II. 2nd ed. Wiley, New York.

Fernique, X. (1981). *Sur le théoreme de Kantorovich–Rubinstein dans les espaces polonais.* (*Lect. Notes Math.*, **850**) Springer, Berlin, pp. 6–10.

*Fishburn, P. C. (1980). Stochastic dominance and moments of distributions. *Math. Oper. Res.*, **5**, 94–100.

*Fomenko, A. T. and Rachev, S. T. (1990). Volume functions of historical texts and the amplitude correlation principle. *Computers and Humanities.* To appear.

Forcina, A. and Galmacei, G. (1974). Sulla distribuzione dell'indice semplice di dissomiglianza. *Metron*, **XXII**, 361–378.

Fortet, R. and Mourier, B. (1953). Convergence de la répartition empirique vers la répétition théorétique. *Ann. Sci. Ecole Norm.*, **70**, 267–285.

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

Fréchet, M. (1937). Sur la distance de deux lois de probabilité. *Publ. Inst. Stat. Univ. Paris*, **6**, pp. 185–198.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon*, **14**, 53–77.

Fréchet, M. (1957). Les tableaux de corrélation dont les marges sone donñes. *Ann. Univ. Lyon S. A* **20**, 13–31.

*Friendrich, V. (1972). *Stetige Transportoptimierung ihre Beziehungen zur Theorie der holderstetigen. Funktionen und einige ihrer Anwendungen.* VEB, Berlin.

*Gaffke, N. and Rüschendorf, L. (1981). On a class of extremal problems in statistics. *Math. Operations Forsch. Stat., Optimization*, **12**, 123–136.

Gaffke, N. and Rüschendorf, L. (1984). On the existence of probability measures with given marginals. *Stat. Decisions*, **2**, 163–174.

Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics.* Wiley, New York.
Galambos, J. and Kotz, S. (1978). *Characterizations of Probability Distributions*, Springer, Berlin.
*Gale, D. (1960) *Theory of linear economic models.* McGraw-Hill, New York.
*Ganssler, P. (1970). A note on a result of Dudley on the speed of Glivenko–Cantelli convergence. *Ann. Math. Stat.*, **41**, 1339–1343.
*Ganssler, P. (1971). Compactness and sequential compactness in spaces of measures. *Z. Wahrsch. Verw. Geb.*, **17**, 124–146.
*Garel, B. B. (1981). Calcul approchéde la distance de Prokhorov. *C.R: Acad. Sci.*, **293**, 697–690.
Garsia-Palomares, U. and Gine, E. M. (1977). On the linear programming approach to the optimality property of Prokhorov's distance, *J. Math. Anal. Appl.*, **60**, 596–600.
*Geffroy, M. J. and Quidel, M. P. (1974). *Convergence stochastiques des répartitions ponctuelles aleatoires.* Revista da Faculdad de Ciencias da Universidade de Coimbra, vol. XLIX.
*Geffroy, M. J. and Zeboulon, M. (1975). Sur certaines convergences stochastiques des mesures aleatoires et des processus ponctuels. *C.R.A.S.P.*, **280A**, 291–293.
Gelbrich, M. (1988). *On a formula for the L^p Wasserstein metric between measures on Euclidean and Hilbert spaces.* Preprint, Sektion Mathematic der Humboldt—Universität zu Berlin, no. 179, Berlin.
Genest, C. and MacKay, J. (1986). The joy of copulas, bivariate distributions with uniform marginals. *Amer. Statist.*, **40**, 280–283.
Gerber, H. (1981). *An Introduction to Mathematical Risk Theory.* Huebner Foundation Monograph.
*Gerber, H. (1984). Error bounds for the compound Poisson approximation. *Insurance: Math. Econ.*, **3**, 191–194.
Germaine, A. (1886). *Etudes sur le probléme des déblais et réblais.* Extrait des Mémoires de l'Academie Nationale de Caen.
*Getoor, K. R. (1961). Infinitely divisible probabilities on the hyperbole plane. *Pacific J. Math.*, **11**, 81–115.
Ghosh, M. and Ebrahimi, N. (1981). Multivariate NBU and NBUE distributions. *Egyptian Statist. J.*, **25**, 36–55.
Gikhman, I. I. and Skorokhod, A. V. (1971). *The Theory of Stochastic Processes.* (In Russian). Nauka, Moscow (Engl. Transl. (1976) Springer, Berlin).
Gine, E. and Leon, J. R. (1980). On the central limit theorem in Hilbert space. *Stochastica*, **4**, 43–71.
Gini, C. (1914). Di una misura delle relazioni tra le graduatorie di due caratteri, In: *L'Elezioni Generali Politiche del 1913 nel comune di Roma*, A. Hancini (ed), Ludovico Cecehini, Rome.
*Gini, C. (1914–1915). Di una misura della dissomiglianza tra due gruppi di quatıta e della sue applicazioni allo studio delle relazioni statitiche. Atti del Reale Instituo Veneto di Sci., *Lettera ed Arti*, **LXXIV**, 185–213.
*Gini, C. (1915–1916). Sul criterio di concordanza tra due caratteri. Atti del Reale Instituto Vento di Sci., *Lettera ed Arti*, **LXXV**, 309–331.
Gini, C. (1951). *Corso di statistica.* Veschi, Rome.
Gini, C. (1965). La dissomiglianza. *Metron*, **XXIV**, 309–331.
Givens, C. R. and Shortt, R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.*, **31**, 231–240.
*Gnedenko, B. V. (1970). *On some unsolved problems in queueing theory.* Sixth International Telegraphic Conference, Munich. (In Russian).

*Gnedenko, B. V. (1983). *On some stability theorem. (Lect. Notes in Math.*, **982**) Springer, Berlin.

*Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA.

*Gnedenko, C. R. (1943). Sur la distribution limité du terme maximum d'une série aléatorie. *Ann. Math.*, **44**, 423–453.

Goldfarb, D. (1985). Efficient dual simplex algorithms for the assignment problem. *Math. Prog.*, **33**, 187–203.

*Gol'shtein, E. G. (1978). Dual problems of convex and fractional convex programming in function spaces, in the collection: *Studies in Mathematical Programming*, Nauka, Moscow. (In Russian).

*Goovaerts, M. J., de Vylder, F., and Haezandonck, J. (1984). *Insurance Premiums*. North-Holland, Amsterdam.

*Gorostiza, L. G. (1973). The central limit theorem for random motions of d-dimensional Euclidean space. *Ann. Prob.*, **1**, 603–612.

Gray, R. M. (1988). *Probability, Random Processes, and Ergodic properties*. Springer, New York.

Gray, R. M. and Ornstein, D. S. (1979). Block coding for discrete stationary $\bar{d}$-continuous channels. *IEEE Trans. Inform. Theory*, **IT25**, 292–306.

Gray, R. M., Neuhoff, D. L., and Shields, P. C. (1975). A generalization to Ornstein's $\bar{d}$-distance with applications to information theory. *Ann. Prob.*, **3**, 315–328.

Gray, R. M., Ornstein, D. S., and Dobrushin, R. L. (1980). Block synchronization, sliding-block coding, invulnerable sources and zero error codes for discrete noisy channels. *Ann. Prob.*, **8**, 639–674.

*Grigorevski, N. N. and Shiganov, I. S. (1976). On certain modifications of Dudley's metric. *Zap. Naucn. Sem. Leningrad. Otdel. Math. Inst. Steklov (LOMI)*, **41**, 17–27.

Grincevicius, A. K. (1985). *Some limit theorems for weakly dependent random variables*. Fourth International Vilnius conference on probability theory and mathematical statistics, Abstracts of communications, IV, pp. 106–108.

*Gross, L. (1979). Decay of correlation in classical lattice models at high temperature. *Comm. Math. Phys.*, **68**, 9–27.

*Gross, L. (1981). Absence of second order phase transition in the Dobrushin uniqueness region. *J. Stat. Phys.*, **25**, 57–72.

*Grzegorek, E. and Ryll-Nardzewski, C. (1980). A remark on absolutely measurable sets. *Bull. Acad. Polon. Sci.*, **28**, 229–232.

*Gumbel, E. J. (1960). Bivariate exponential distributions. *J. Am. Stat. Assoc.*, **55**, 698–707.

Hall, P. (1981). Two-sided bounds on the rate of convergence to a stable law. *Z. Wahrsch. Verw. Geb.*, **57**, 349–364.

*Halmos, P. R. (1950). *Measure Theory*. Van Nostrand, Princeton.

*Hammersely, J. M. (1952). An extension of the Slutzky-Fréchet theory. *Acta Math.*, **87**, 243–257.

Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.*, **42**, 1887–1896.

Haneveld, W. K. K. (1985). *Duality in Stochastic Linear and Dynamic Programming*. Centrum voor Wiskunde en Informatica, Amsterdam.

*Hansel, G. and Troallic, J. P. (1978). Mesures marginales et theoremelde Ford–Fulkerson. *Z. Wahrsch. Verw. Geb.*, **43**, 245–251.

*Hansel, G. and Troallic, J. R. (1986). Sur le probleme des marges. *Prob. Theo. Rel. Fields*, **71**, 357–366.

*Hardy, G. H. (1969). *Divergent Series*. Oxford University Press, London.

Hardy, G. H., Littlewood, J. E. and Polya, G. (1952). *Inequalities.* Cambridge University Press, London.
*Harris, R. (1970). A multivariate definition for increasing hazard rate distribution function. *Ann. Math. Stat.*, **41**, 713–717.
Hausdorff, F. (1949). *Set Theory.* Dover, New York.
Hennequin, P. L. and Tortrat, A. (1965). *Théorie des probabilités et quelques applications*, Masson, Paris.
*Hettich, R. (Editor). (1979). *Semi-infinite programming. (Lect. Notes Control Inform. Sci.*, **15**) Springer, Berlin.
Hernandez-Lerma, O. and Marcus, S. I. (1984). Identification and Approximation of Queueing Systems, *IEEE Trans. Automatic Control.* **AC29**, 472–474.
Herzel, A. (1963). *Il valore medio e la varianza dell'indice semplice di dissomiglianza negli universi dei campioni* (Biblioteka del "Metron", serie C: Note e commentivol.), Roma.
Hewitt, E. and Stromberg, K. (1965). *Real and Abstract Analysis.* Springer, New York.
*Hipp, C. (1985). Approximation of aggregate chains distributions by compound Poisson distributions. *Insurance: Math. Econ.*, **4**, 227–232, (Correction: (1987), **6**, 165).
*Hipp, C. (1987). Improved approximations for the aggregate chains distribution in the individual model. *Astin Bull.*, **16**, 89–100.
Hoeffding, W. (1940). Masstabvariate korrelations theorie, *Sehr. Math. Inst. Univ. Berlin*, **5**, 181–233.
Hoeffding, W. (1955). The extrema of the expected value of a function of independent random variables. *Ann. Math. Stat.*, **26**, 268–275.
Hoeffding, W. and Shrikhande, S. S. (1955). Bounds for the distribution function of a sum of independent random variables. *Ann. Math. Stat.* **26**, 439–449.
*Hoeffding, W. and Wolfowitz, J. (1958). Distinguishability of sets of distributions. *Ann. Math. Statist.*, **29**, 700–718.
Hoffman, A. J. (1961). *On simple linear programming problems. Convexity, Proceedings of Symposia in Pure Mathematics*, Vol. 7, *American Mathematical Society*, Providence, RI, pp. 317–327.
*Hoffman-Jorgensen, J. (1970). *The Theory of Analytic Spaces.* Publications Series, no. 10, Aarhus.
Hoffman-Jorgensen, J. (1977). *Probability in Banach spaces. Ecole D'Ete de Probabilités de Saint-Flour VI. (Lect. Math.*, **598**), Springer, Berlin, pp. 1–186.
Hoffman-Jorgensen, J. and Pisier, G. (1976). The law of large numbers and the central limit theorem in Banach spaces. *Ann. Prob.*, **4**, 587–599.
Hohlov, Yu. S. (1982). On the convergence of distribution of the translation parameter of the composition of Euclidean random motions to multidimensional stable distribution. *Theory Prob. Appl.*, **26**, 342–344. (In Russian).
Hohlov, Yu. S. (1986). *The estimation of the rate of convergence in the integral limit theorem in the Euclidean motion group. (Lect. Notes Math.*, **1233**) Springer, Berlin, pp. 1–10.
*Hougaard, P. (1987). Modelling multivariate survival. *Scand. J. Stat.*, **14**, 291–304.
Huber, P. J. (1977). *Robust Statistical Procedures.* CBMS Regional Conference Series in Applied Mathematics, 27, Society for Industrial and Applied Mathematics, Philadelphia.
Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.
*Ibragimov, I. A. (1966). On the accuracy of Gaussian approximation to the distribution functions of sums of independent variables. *Theory Prob. Appl.*, **XI**, 559–580.
Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and stationary sequences of random variables.* Wolters-Noorhoff, Groningen.
*Ibragimov, I. A. and Presman, E. L. (1973). On the rate of approach of the distributions

of sums of independent random variables to accompanying distributions. *Theo. Prob. Appl.*, **18**, 713–727.

Iglehart, D. L. (1973). Weak convergence in queueing theory. *Adv. Appl. Prob.*, **5**, 570–594.

Ignatov, Ts. G. and Rachev, S. T. (1983). Minimal ideal probability metrics. In: *Stability of Stochastic Models*, VNIISI, Moscow, pp. 36–48. (In Russian).

*Ioffe, A. D. and Tikhomirov, V. M. (1968). Duality of convex functions and extremal problems. *Russian Math. Surveys*, **23**, 53–124.

*Ireland, C. T. and Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, **55**, 179–188.

*Isbell, F. R. (1955). Birkhoff's problem III. *Proc. Am. Math. Soc.*, **6**, 217–218.

*Isbell, F. R. (1962). Infinite doubly stochastic matrices. *Can. Math. Bull.*, **5**, 1–4.

*Isii, K. (1960). The extrema of probability determined by generalized moments (I): Bounded random variables. *Ann. Inst. Math. Statist.*, **12**, 119–133.

*Jacob, P. (1979a). Convergence uniforme a distance finie des mesures positives. *C. R. Acad. Sci.*, **288A**, 295–297.

*Jacob, P. (1979b). Convergence uniforme a distance finie des mesures signees. *Ann. Inst. Poincare*, **XV**, 355–373.

*Joag-Dev, K. (1984). *Measures of dependence. Handbook of Statistics*, Vol. 4, (P. R. Krishnaiah and P. K. Sen (eds), North-Holland, Amsterdam, pp. 79–85.

*Johnson, J. (1985). An elementary characterization of weak convergence of measures. *Am. Math. Monthly*, **92**, 136–137.

Kagan, A. M., Linnik, Yu. V., and Rao, S. R. (1972). *Characterization Problems of Mathematical Statistics.* Nauka, Moscow (in Russian), (Engl. Transl. (1973) Wiley, New York).

Kakosyan, A. B., Klebanov, L., and Melamed, J. A. (1984). *Characterization of Distributions by the Method of Intensively Monotone Operators. (Lect. Notes Math.*, **1088**) Springer, Berlin.

Kakosyan, A. B., Klebanov, L., and Rachev, S.T. (1988). *Quantitative Criteria for Convergence of Probability Measures.* Ayastan Press, Erevan. (In Russian). (Engl. Transl. Springer, Berlin, to appear).

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

*Kalashnikov, V. V. (1983). Estimates of the rate of convergence in Renyi theorem, In: *Proc. Stability problems for stochastic models.* Moscow: Institute for System Studies. pp. 48–57. (In Russian).

*Kalashnikov, V. V. and Anichkin, S. A. (1981). Continuity of random sequences and approximation of Markov chains. *Adv. Appl. Prob.*, **13**, 402–414.

Kalashnikov, V. V. and Rachev, S. T. (1984). Characterization problems in queueing and their stability. In: *Stability Problems for Stochastic Models, Proceedings*, VNIISI, Moscow, pp. 49–86. (In Russian); (Engl. Transl. (1986) *J. Sov. Math.*, **35**, 2336–2360).

*Kalashnikov, V. V. and Rachev, S. T. (1985). Characterization problems in queueing and their stability. *Adv. Appl. Prob.*, **17**, 868–886.

*Kalashnikov, V. V. and Rachev, S. T. (1986a). Characterization of inverse problems in queueing and their stability. *J. Appl. Prob.*, **23**, 459–473.

*Kalashnikov, V. V. and Rachev, S. T. (1986b). *Characterization of queueing models and their stability in Probability theory and mathematical Statistics.* Yu. K. Prohorov *et al.* (eds) VNU Science Press, **2**, 37–53.

*Kalashnikov, V. V. and Rachev, S. T. (1986c). *Average stability of characterization of queueing models in Stability Problems for Stochastic Models, Proceedings*, VNIISI, Moscow, pp. 67–75. (In Russian). (Engl. transl.: *J. Sov. Math.*, **40**, 502–509.)

Kalashnikov, V. V. and Rachev, S. T. (1988). *Mathematical Methods for Construction of Stochastic Queueing Models.* Nauka, Moscow. (In Russian). (Engl. transl., (1990) Wadsworth, Brooks–Cole, Pacific Grove, California).

Kalashnikov, V. V. and Vsekhvyatskii, S. Yu. (1985). Metric estimates of the first occurrence time in regenerative processes, In: *Proc. Stability problems for stochastic models, (Lect. Notes Math.*, **1155**.) Springer, Berlin, pp. 102–130.

*Kalashnikov, V. V. and Zolotarev, V. M. (1984). *Stability Problems for Stochastic Models. Proceedings, (Lect. Notes Math.*, **1155**) Springer, Berlin, pp. 102–130.

*Kalashnikov, V. V., Rachev, S. T., and Fomenko, A. T. (1986). New methods for comparison of volume functions of historical texts, In: *Stability Problems for Stochastic Models, Proceedings*, VNIISI, Moscow. (In Russian), pp. 33–45.

Kallenberg, O. (1975). *Random Measures.* Akademie, Berlin.

*Kamae, T. and Krengel, U. (1978). Stochastic partial ordering. *Ann. Prob.*, **6**, 1044–1049.

*Kamae, T., Krenkel, U., and O'Brien, G. I. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Prob.*, **5**, 899–912.

*Kandelaki, N. P. (1976). On a metric in the space of distributions. *Soobshch. Akad. Nauk. Gruzin. SSR*, **82**, 549–552. (In Russian).

Kantorovich, L. V. (1940). On one effective method of solving certain classes of extremal problems. *Dokl. Akad. Nauk, USSR*, **28**, 212–215.

Kantorovich, L. V. (1942). On the transfer of masses. *Dokl. Akad. Nauk. USSR*, **37**, 227–229.

Kantorovich, L. V. (1948). On a problem of Monge. *Usp. Mat. Nauk*, **3**, 225–226. (In Russian).

Kantorovich, L. V. and Akilov, G. P. (1984). *Functional Analysis*, Nauka, Moscow. (In Russian).

*Kantorovich, L. V. and Gavurin, M. K. (1949). Application of mathematical methods to questions of the analysis to freight traffic. Collection 'Problems of improving the efficiency of transport,' *Akad. Nauk SSSR.* (In Russian).

Kantorovich, L. V. and Rubinshtein, G. Sh. (1957). On a function space and some extremum problems. *Dokl. Akad. Nauk. SSSR*, **115**, 1058–1061. (In Russian).

Kantorovich, L. V. and Rubinshtein, G. Sh. (1958). On the space of completely additive functions. *Vestni LGU. Ser. Mat., Mekh. i Astron.*, **7**, 52–59.

Karatzas, I. (1984). Gittins indices in the dynamic allocation problem for diffusion processes. *Ann. Prob.* **12**, 173–192.

*Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. *J. Mult. Anal.* **10**, 467–498.

Karlin, S. and Studden, W. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics.* Wiley, New York.

Kaufman, R. (1984). Fourier transforms and descriptive set theory. *Mathematika* **31**, 336–339.

*Keisler, H. J. and Tarski, A. (1964). From accessible to inaccessible cardinals. *Fund. Math.*, **53**, 225–308.

Kellerer, H. G. (1961). Funktionen auf Produktraumen mit vorgegebenen Marginal funktionen. *Math. Ann.*, **144**, 323–344.

Kellerer, H. G. (1962). Die Schnittma-Funktionen Me barer Teilmengen eines Produktraumes. *Math. Ann.*, **146**, 123–128.

Kellerer, H. G. (1963a). Distribution functions with given marginal distributions. *Ann. Math. Stat.*, **32**, 1120.

Kellerer, H. G. (1963b). Ein Zerlegungsproblem fur Lebesguesche Mengen. *Math. Z.*, **80**, 306–313.

Kellerer, H. G. (1964a). Bemerkung zu einem Satz von H. Richter. *Archiv Math.*, **15**, 204–207.

Kellerer, H. G. (1964b). Schittmab-Funktionen in mebrfachen Produkträumen. *Math. Ann.*, **155**, 369–391.

Kellerer, H. G. (1964c). Maßtheoretische Marginal probleme. *Math. Ann.*, **153**, 168–198.

Kellerer, H. G. (1964d). Marginal probleme für Funktionen. *Math Ann.*, **154**, 147-156.

Kellerer, H. G. (1964e). Verteilungsfunktionen mit gegebenen Marginalverteilungen. *Z. Wahrsche.*, **3**, 247–270.

Kellerer, H. G. (1984a). Duality theorems for marginal problems. *Z. Wahrsch. Verw. Geb.*, **67**, 399–432.

Kellerer, H. G. (1984b). *Duality theorems and probability metrics.* Proc. 7th Brasov Conf., Bucuresti, 211–220.

Kellerer, H. G. (1988). Measure-theoretic versions of linear programming. *Math. Z.*, **198**, 367–400.

*Kelley, J. L. (1955). *General Topology.* Van Nostrand, Princeton.

*Kelley, J. L. and Namioka, I. (1963). *Linear Topological Spaces.* Van Nostrand, Princeton.

Kemperman, J. H. B. (1968). The general moment problem, a geometric approach. *Ann. Math. Stat.*, **39**, pp. 93–122.

Kemperman, J. M. B. (1972). *On a class of moment problems.* Proceedings Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 101–126.

*Kemperman, J. H. B. (1977). On the FKG-inequality for measures on a partially ordered space. *Proc. Nederl. Akad. Wet.*, **80**, 313–331.

Kemperman, J. H. B. (1983). *On the role of duality in the theory of moments. Semi-infinite programming and applications (Lect. Notes Economic Math. Syst.*, **215**) Springer, Berlin, pp. 63–92.

Kemperman, J. H. B. (1987). Geometry of the moment problem, a geometric approach. *Proc. SIAM*, **5**, 16–53.

*Kendall, D. G. (1960). On infinite doubly-stochastic matrices and Birkhoff's problem III. *J. London Math. Soc.*, **35**, 81–84.

*Kendall, D. G. (1968). Delphic semigroups, infinitely divisible regenerate phenomena and arithmetic of p-functions. *Z. Wahrsch. Verw. Geb.*, **9**, 163–195.

Kennedy, D. (1972). The continuity of the single server queue. *J. Appl. Prob.*, **9**, 370–381.

Kerstan, J., Matthes, K., and Mecke, J. (1978). *Infinitely Divisible Point Processes.* Wiley, New York.

*Kesten, M. and Spitzer, F. (1979). A limit theorem related to a new class of self-similar processes. *Z. Wahrsch. Verw. Geb.*, **50**, 5–25.

*Kimeldorff, G. and Sampson, A. R. (1975). Uniform representations of bivariate distributions. *Commun. Stat.*, **4(7)**, pp. 617–627.

*Kimeldorff, G. and Sampson, A. R. (1978). Monotone dependence. *Ann. Stat.*, **6**, 895–903.

*Kingman, T. F. (1963). Random walks with spherical symmetry. *Acta Math.*, **109**, 11–53.

*Klebanov, L. B. and Rachev, S. T. (1985). *Stability of lack of memory property of multivariate exponential distributions in finite number of points. (Lect. Notes Math.*, **1155**), Springer, Berlin, pp. 131–143.

Klebanov, L. B., Maniya, G. M., and Melamed, I. A. (1984). A problem of Zolotarev and analogs of infinitely divisible and stable distributions in a scheme for summing a random number of random variables. *Theor. Prob. Appl.*, **29**, 791–794.

*Klefsjo, B. (1982). The HNBUE and HNWUE classes of life distributions. *Naval Res. Logistics Q.*, **29**, 331–344.

*Klein, D. (1982). Dobrushin uniqueness techniques and decay of correlations in continuum statistical mechanics. *Comm. Math. Phys.*, **86**, 149–160.

*Kleinschmidt, P., Lee, C. W., and Schannath, H. (1987). Transportation problems which can be solved by the use of Hirsch paths for the dual problems. *Math. Prog.*, **37**, 153–168.

Knott, M. and Smith, C. S. (1984). On the optimal mapping of distributions. *J. Opt. Theor. Appl.*, **43**, 39–49.

Kolmogoroff, A. N. (1931). Eine Verallgemeinerung des Laplace–Liapounoffschen Satsen. *Izs. Akad. Nauk. SSSR, Math. Nat. Sci.*, **7**, 959–962.

Kolmogoroff, A. N. (1933). Uber die Grenzwersatze der Wahrscheinlichkeitsrechnung. *Izv. Akad. Nauk SSSR, Math. Nat. Sci.*, **3**, 363–373.

Kolmogorov, A. N. (1953). Some latest works on limit theorems in probability theory. *Vestnik MGU*, **10**, 28–39. (In Russian).

Kolmogorov, A. N. (1956). On the convergence of A. V. Skorokhod. *Theory Prob. Appl.*, **1**, 237–247.

*Krein, M. G. (1951). The ideas of P. L. Cebysev and A. A. Markov in the theory of limiting values of integrals and their further development. *Usp. Mat. Nauk.*, **6**, 3–130. (In Russian); (Engl. Transl. (1959), *Am. Math. Soc. Transl.*, Series 2, **12**, 1–121).

*Krein, M. G. and Nudel'man, A. A. (1977). *The Markov moment problem and extremal problems.* American Mathematical Society, Providence, RI.

*Kretschmer, K. S. (1961). Programmes in paired spaces. *Can. J. Math.*, **13**, 221–229.

Kruglov, V. M. (1973). Convergence of numerical characteristics of independent random variables with values in a Hilbert space. *Theory Prob. Appl.*, **18**, 694–712.

Kruglov, V. M. (1976). Global limit theorems. *Zap. Nauchn. Sem. LOMI*, **61**, 84–101. (In Russian).

Kuelbs, T. (1973). A representation theorem for symmetric stable processes and stable measures. *Z. Wahrsch. Verw. Geb.*, **26**, 259–271.

*Kullback, S. (1959). *Information Theory and Statistics.* Wiley, New York.

Kruskal, W. M. (1958). Ordinal measures of association. *J. Am. Stat. Assoc.*, **53**, 814–861.

*Kumar, A. and Pathak, P. K. (1977). Sufficiency and tests of goodness of fit. *Scand. J. Stat.*, **4**, 39–43.

*Kunsch, M. (1982). Decay of correlations under Dobrushin's uniqueness condition and its applications. *Comm. Math. Phys.*, **84**, 207–222.

Kuratowski, K. (1966). *Topology*, Vol. I. Academic, New York.

Kuratowski, K. (1969). *Topology*, Vol. II. Academic, New York.

*Kuznezova-Sholpo, I. A. (1985). *Compound metrics with fixed minimal ones.* (Lect. Notes Math., **1155**) Springer, Berlin.

Kuznezova-Sholpo and Rachev, S. T. (1989). Explicit solutions of moment problems. *Probability and Mathematical Statistics*, **X**, Fasc. 2, pp. 297–312.

*Ky Fan (1959). *Convex sets and their applications.* Argonne National Laboratory, Argonne.

Lai, T. L. and Robbins, M. (1976). Maximally dependent random variables. *Proc. Nat. Acad. Sci. USA*, **73**, 286–288.

*Lamperti, A. (1964). On extreme order statistics. *Ann. Math. Stat.*, **35**, 1726–1737.

Landenna, G. (1956). La dissomiglianza. *Statistica*, **XVI**, 21–57.

Landenna, G. (1961). L'indice semplice di dissomiglianza tra campioni ottenuti con estrazione in blocco. *Statistica*, **XXI** H., Luglio settembre.

*Landford, III, O. E. (1973). *Entropy and Equilibrium States in Classical Statistical Mechanics.* (*Lect. Notes Phys.*, **20**), Springer, Berlin.

*Lange, K. (1973). Borel sets of probability measures. *Pacific J. Math.*, **48**, 141–161.

Leadbetter, M. R., Lindren, G., and Rootzen, H. (1980). *Extremes and Related Properties of Random Sequences and Processes.* Springer, New York.

Lebesgue, H. (1905). Sur les fonctions representables analytiquement. *J. Math. Pures Appl.*, **V**, 139–216.

*Le Cam, L. (1957). Convergence in distribution of stochastic processes. *Univ. California Publ. Stat.*, **2**, 207–236.
Le Cam, L. (1986). *Asymptotic Methods on Statistical Decision Theory*. Springer, New York.
*Lee, M. L. T. (1985). Dependence by total positivity. *Ann. Prob.*, **13**, 572–582.
*Lee, M. L. T., Rachev, S. T. and Samorodnitsky, G. (1990). Association of stable random variables, *Ann. Prob.* (to be published).
*Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
*Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Stat.*, **37**, 1137–1153.
Leti, G. (1962). Il termine generico delle tabelle di congraduazione e di contro-graduazione. *Biblioteca del Metron*, 1–27.
Leti, G. (1963). *Una misura della mutabilita dell'universo dei campioni.* Atti della XXII Riunione Scientifica della Societa Italiana di Statistica, Roma 27–28. Ottobre, Roma.
Levin, S. L. (1980). Applications of Dobrushin's uniqueness condition and its applications. *Comm. Math. Phys.*, **84**, 65–74.
Levin, V. L. (1974). Duality and approximation in the problem of mass transfer. In: *Mathematical Economics and Functional Analysis*, Nauka, Moscow, pp. 94–108. (In Russian).
Levin, V. L. (1975). On the problem of mass transfer. *Sov. Math. Dokl.*, **16**, 1349–1353.
Levin, V. L. (1977). On duality theorems in the Monge–Kantorovich problem. *Usp. Mat. Nauk.*, **32**, 171–172. (In Russian).
Levin, V. L. (1978a). The Monge–Kantorovich problem on mass transfer. In: *Methods of Functional Analysis in Mathematical Economics*, Nauka, Moscow, pp. 23–55. (In Russian).
Levin, V. L. (1978b). *The mass transfer problem, strong stochastic domination, and probability measures on the product of two compact spaces with given projections.* Preprint, TsEMI, Moscow. (In Russian).
Levin, V. L. (1984). The mass transfer problem in topological space and probability measures on the product of two spaces with given marginal measures. *Dokl. Akad. Nauk USSR*, **276**, 1059–1064. (Engl. Transl. *Sov. Math. Dokl.*, 638–643.)
*Levin, V. L. (1985). *Convex Analysis in Spaces of Measurable Functions and Its Applications in Mathematics and Economics.* Nauka, Moscow. (In Russian).
Levin, V. L. (1986). Measurable selectors of multivalued mappings and the mass transfer problem. *Dokl. Acad. Sci. USSR.* (In Russian).
Levin, V. L. and Milyutin, A. A. (1979). The mass transfer problem with discontinuous cost function and a mass setting for the problem of duality of convex extremum problems. *Usp. Mat. Nauk*, **34**, 3–68. (In Russian). (Engl. Transl. (1979) *Russian Math. Surveys* **34**, 1–78.)
Levin, V. L. and Rachev, S. T. (1990). *New duality theorems for marginal problems with some applications in stochastics.* (*Lect. Notes Math.*, **1412**) Springer, Berlin. pp. 137–171.
Lévy, P. (1925). *Calcul des Probabilités.* Gauthier-Villars, Paris.
Lévy, P. (1950). Distance de deux variables aléatoires et distance de deux probabilités. In: *Géněralites sur les probabilités.* M. Fréchet (ed.) Gauthiers-Villars, Paris.
*Lindenstrauss, J. (1965). A remark on extreme doubly-stochastic measures. *Am. Math. Monthly*, **72**, 379–382.
Ljung, L. (1978). Convergence analysis of parametric identification models. *IEEE Trans. Automat. Contr.*, **AC-23**, 770–783.
Loeve, M. (1963). *Probability Theory*, 3rd Edn. Van Nostrand, Princeton.
*Loomis, H. (1975). Dilations and extremal measures. *Adv. Math.* **17**, 1–13.
Lorentz, G. G. (1953). An inequality for rearrangements. *Am. Math. Monthly*, **60**, 176–179.

Lucia, L. (1965). Variabilita superficiale e dissomiglianza tra distribuzioni semplici. *Metron*, **XXIV**, 225–331.

Lukacs, E. (1968). *Stochastic Convergence*. D. C. Heath, Lexington, Mass.

*Lusin, N. (1914). Sur un problème de M. Baire. *C.R. Acad. Sci.* **158**, 1258–1260.

Lusin, N. (1930). *Lecons Sur les Ensembles Analytiques*. Gauthier-Villars, Paris.

*Maharam, D. (1942). On homogeneous measure algebras. *Proc. Nat. Acad. Sci. USA* **28**, 108–111.

*Maharam, D. (1971). *Consistent extensions of linear functionals and of probability measures*. Proc. 6th Berkeley Sympos. Math. Statist. Probab. Univ. Calif. **2**, pp. 127–147.

Maejima, M. (1988). *Some limit theorems for stability methods of r.r.d. random variables. (Lect. Notes Math.* **1233**) Springer, Berlin, pp. 57–68.

Maejima, M. and Rachev, S. T. (1987). An ideal metric and the rate of convergence to a self-similar process. *Ann. Prob.*, **15**, 702–727.

Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.*, **8**, 487–517.

*Maitra, A. and Ryll-Nardzewski, C. (1970). On the existence of two analytic non-Borel sets which are not isomorphic. *Bull. Acad. Polon Sci.*, **18**, 177–178.

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

Mallows, C. L. (1972). A note on asymptotic joint normality. *Ann. Math. Stat.*, **43**, 508–515.

*Marshall, A. W. (1975). Multivariate distributions with monotone hazard rate. In: *Reliability and Fault Tree Analysis*. R. E. Barlow *et al.* (eds.) SIAM, Philadelphia, pp. 259–284.

Marshall, A. W. and Olkin, I. (1967). A generalized bivariate exponential distribution. *J. Appl. Prob.*, **4**, 291–302.

*Marshall, A. W. and Olkin, I. (1983). Domains of attraction of multivariate extreme value distributions. *Ann. Prob.*, **11**, 168–177.

*Marshall, A. W. and Shaked, M. (1982). A class of multivariate new better than used distributions. *Ann. Prob.*, **10**, 259–264.

*Mazukiewicz, S. (1936). Uber die Menge der differenzierbaren Funktionen. *Fund Math.*, **27**, 247–248.

*Meyer, P. A. (1966). *Probability and Potentials*. Blaisdell, Waltham, MA.

Mijnheer, J. L. (1975). *Sample path properties of stable processes*. Math. Centre Tracts 59. Amsterdam, Math. Centrum.

Mijnheer, J. L. (1983). On the rate of convergence to a stable limit law. Colloquia mathematica societatis Janos Bolyai. *Limit theorems in probability and statistics*. Hungary, Veszprem.

*Mirksy, L. (1963). Results and problems in the theory of doubly stochastic matrices. *Z. Wahrsch. Verw. Geb.*, **1**, 319–334.

Mitalauskas, A. and Statulevicius, V. (1976). An asymptotic expansion in the case of a stable approximation law. *Lith. Math. J.* **16**, 574–587.

*Mittnik, S. and Rachev, S. T. (1989). Stable distributions for asset returns. *Appl. Math. Lett.*, **2/3**, 301–304.

*Moeschberger, M. L., (1974). Life test under competing causes of failure. *Technometrics*, **16**, pp. 39–47.

*Monge, G. (1781). Mémoire sur la théorie des dèblais et des remblais. *Histoire de l'Académie des sciences de Paris*, avec les Mémoires de mathématique et de physique pour la même année, pp. 666–704, avec 2 pl.

*Montague, J. S. and Plemmons, R. J. (1973). Doubly stochastic matrix equations. *Israel J. Math.*, **15**, 216–229.

Mukerjee, H. G. (1985). Supports of extremal measures with given marginals. *Illinois J. Math.*, **29**, 248–260.

*Musial, K. (1972). Existence of proper conditional probabilities. *Z. Wahrsch. Verw. Geb.*, **22**, 8–12.

*Nachbin, L. (1951). Linear continuous functionals positive on the increasing continuous functions. *Summa Brazil. Math.*, **2**, 135–150.

*Nachbin, L. (1965). *Topology and Order.* New York, Van Nostrand.

Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

Neveu, J. and Dudley, R. M. (1980). *On Kantorovich–Rubinstein theorems* (transcript).

*Novak, J. (1965). On convergence spaces and their sequential envelopes. *Czech. Math. J.* **15**, 74–100.

*Obretenov, A. and Rachev, S. T. (1983a). *Characterization of the bivariate exponential distribution and Marshall–Olkin distribution and stability. (Lect. Notes Math.*, **982**), Springer, Berlin, pp. 136–150.

*Obretenov, A. and Rachev, S. T. (1983b). Stability of some characterization properties of the exponential distribution, In: *Stability Problems for Stochastic Models. Proceedings*, Moscow, VNIISI, 1983, pp. 79–87. (In Russian). (Engl. Transl., (1986) *J. Sov. Math.*, **32**, 643–651.)

*Obretenov, A. and Rachev, S. T. (1987). *Estimates of the deviation between the exponential and new classes of bivariate distribution.* (*Lect. Notes Math.*, **1233**), Springer, Berlin. 93–102.

*Obretenov, A., Dimitrov, B., and Rachev, S. T. (1983). Stability of the service process in a system of type M/M/I, In: *Stability Problems for Stochastic Models, Proceedings*, Moscow, VNIISI, pp. 71–79. (In Russian). (Engl. transl. (1986), *J. Sov. Math.*, **32**, 634–643).

Olkin, I. and Pukelheim, F. (1982). The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl.*, **48**, 257–263.

Omey, E. (1988) Rates of convergence for densities in extreme value theory. *Ann. Prob.*, **16**, 479–486.

Omey, E. and Rachev, S. T. (1988). On the rate of convergence in extreme value theory. *Theory Prob. Appl.*, **33**, 601–607.

Omey, E. and Rachev, S. T. (1990). *Rates of convergence in multivariate extreme value theory.* Preprint, EHSAL, Brussels.

Ornstein, D. S. (1971). Entropy is enough to classify Bernoulli shifts but not K-automorphisms. *In*: *The Collection Congress International Mathematicians*, Vol. 2, Nauka, Moscow, pp. 571–575.

Ornstein, D. S. (1975). *Ergodic Theory, Randomness and Dynamical Systems.* Yale Math. Monographs No. 5, Yale University Press, New Haven, CT.

*Oxtoby, J. (1971). *Measure and Category.* Springer, New York.

Papantoni-Kazakos, P. (1977). Robustness in Parameter Estimation. *IEEE Trans. Inform. Theory*, **IT-23**, 223–231.

*Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic, New York.

*Partasarathy, K. R., Ranga, R. R., and Varadhan, S. R. S. (1962). On the category of idecomposable distributions on topological groups. *Trans. Am. Math Soc.* **102**, 200–217.

*Partasarathy, K. R., Ranga, R. R., and Varadhan, S. R. S. (1963). Probability distributions on locally compact abelian groups. *Illinois J. Math.*, **7**, 337–369.

*Paulauskas, V. I, (1973). Some nonuniform estimates in limit theorems of probability theory. *Sov. Math. Dokl.*, **14**, 1125–1127.

Paulauskas, V. I. (1974). Estimates of the remainder term in limit theorems in the case of stable limit law. *Lith. Math. Trans.*, **XIV**, 127–147.

Paulauskas, V. I. (1976). On the rate of convergence in the central limit theorem in certain Banach spaces. *Theory Prob. Appl.*, **21**, 754–769.

*Peck, J. E. L. (1959). Doubly stochastic measures. *Michigan Math. J.*, **6**, 217–220.

*Peterson, A. V. (1976). Bounds for a joint distribution function with fixed subdistribution functions, applications to competing risks. *Proc. Nat. Acad. Sci. USA*, **73**, 11–13.

Petrov, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

*Phellps, R. R. (1966). *Lectures on Choquet's Theorem.* Princeton University Press, Princeton, NJ.

*Phillip, W. (1980). Weak and L^p-invariance principles for sums of B-valued random variables. *Ann. Prob.*, **8**, 68–82.

*Phillip, W. and Stout, S. (1975). Almost sure invariance principles of partial sums of weakly dependent random variables. *Mem. Am. Math. Soc.*, 161.

*Pitt, L. D. (1982). Positively correlated normal variables are associated. *Ann. Prob.* **10**, 496–499.

Pisier, G. and Zinn, J. (1977). On the limit theorems for random variables with values in the spaces L^p. *Z. Wahrsch. Verw. Geb.*, **41**, 289–305.

Pompilij, G. (1956). *La mutabilita dell'universo dei campioni.* Scritti matematici in onore di Filippo Sibirani. Bologna.

*Potts, R. B. and Oliver, R. M. (eds.) (1972). *Editors, Social Accounting Matrices: A Basis for Planning.* World Bank, Washington, DC.

*Preiss, D. (1973). Metric spaces in which Prohorov's theorem is not valid. *Z. Wahrsch. Verw. Geb.*, **27**, 109–116.

*Presman, E. L. (1983). Approximation of binomial distributions by infinitely divisible ones. *Theory Prob. Appl.*, **28**, 393–403.

Prokhorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Prob. Appl.*, **1**, 157–214.

Puri, P. S. (1987). On almost sure convergence of an erosion process due to Todorovic and Gani. *J. Appl. Prob.* **24**, 1001–1005.

*Puri, P. S. and Rubin, H. (1974). On a characterization of the family of distributions with constant multivariate failure rates. *Ann. Prob.* **2**, 738–740.

*Rachev, S. T. (1978). Hausdorff metric structures of the space of probability measures. *Zap. Nauchn. Sem. Leningrad. Otdel Math. Inst. Steklov. (LOMI)*, **87**, 87–103. (In Russian). (Engl. transl., (1981) *J. Sov. Math.*, **17**, 2218–2232).

*Rachev, S. T. (1979a). *Metric structure in the space of random variables and distributions.* Ph.D. Thesis, Lomonosov State University, Moscow. (In Russian).

*Rachev, S. T. (1979b). On a Hausdorff metric construction in the space of probability measures. *Zap. Nauchn. Sem. LOMI*, **87**, 87–103. (In Russian). (Engl. transl. 1981, *J. Sov. Math.*, **17**, 2218–2232).

Rachev, S. T. (1980). Levy–Prokhorov distance in a space of semi-continuous set functions, In: *Stability Problems for Stochastic Models. Proceedings*, Moscow, VNIISI, 1980, pp. 76–88. (In Russian). (Engl. transl. (1986) *J. Sov. Math.*, **32**, 64–74.)

*Rachev, S. T. (1981a). On minimal metrics in the space of real-valued random variables, *Sov. Dokl. Math.*, **23**, 425–432.

Rachev, S. T. (1981b). Minimal metrics in a space of random vectors with fixed one-dimensional marginal distributions, In: *Stability Problems for Stochastic Models. Proceedings*, Moscow, VNIISI, pp. 112–128. (In Russian). (Engl. transl. (1986) *J. Sov. Math.*, **34**, 1542–1555).

Rachev, S. T. (1982a). Minimal metrics in the minimal variables space. *Publ. Inst. Stat. Univ. Paris.*, **XXVI**, 27–47.

Rachev, S. T. (1982b). Minimal metrics in the random variables space. In: *Probability and Statistical Inference Proceedings of the 2nd Pannonian Symp.*, W. Grossmann et al., (eds.) Reidel, Dordrecht, pp. 319–327.

*Rachev, S. T. (1982c). Minimal metrics in the random variable space. *Pub. Inst. Stat., Univ. Paris,* **XXVII**, 27–47.

Rachev, S. T. (1983a). *Minimal metrics in the real valued random variables space. (Lect. Notes in Math.*, **982**), Springer, Berlin, pp. 172–190.

*Rachev, S. T. (1983b). Compactness in the probability measures space. In: *Proceedings of the third European Young Statisticians Meeting*, Galyare, M. et al. (eds.), Leuven, Katholieke University, Leuven, pp. 148–152.

Rachev, S. T. (1984a). Hausdorff metric construction in the probability measure space. *Pliska Studia Mathem. Bulgarica*, 7, 152–162.

Rachev, S. T. (1984b). On the ζ-structure of the average and uniform distances. *Dokl. Akad. Nauk.*, **278**, 282–285. (In Russian). (Engl. transl., (1984) *J. Sov. Math. Dokl.*, **30**, 369–372).

Rachev, S. T. (1984c). On a class of minimal functionals on a space of probability measures. *Theory Prob. Appl.*, **29**, 41–49.

Rachev, S. T. (1984d). The Monge–Kantorovich mass transfer problem and its stochastic applications. *Theory Prob. Appl.*, **29**, 647–676.

*Rachev, S. T. (1984e). On a problem of Dudley. *Sov. Math. Dokl.*, **29**, 162–164.

Rachev, S. T. (1985a). *Extreme functionals in the space of probability measures. Proc. Stability problems for stochastic models. (Lect. Notes in Math.*, **1155**,) Springer, New York, pp. 320–348.

*Rachev, S. T. (1985b). *Probability metrics and their applications to the stability problems for stochastic models.* Author's review of the Doctor of Sciences Thesis, Steklov Mathematical Institute, USSR Academy of Sciences, Moscow. (In Russian).

Rachev, S. T. (1985c). Uniformity in weak and vague convergences. *Theory Prob. Appl.*, **30**, 573–576.

Rachev, S. T. (1986). Extreme functionals in the space of probability measures, *Prob. Theory and Math. Statist.*, Yu. V. Prokhorov *et al.* (Eds.), *VNU Science Press*, 471–476.

Rachev, S. T. (1987). *Probability metrics and their applications to problems of stability of stochastic models.* Proceedings of the Sixteenth Spring Conference of the Union of Bulgarian Mathematicians, Sunny Beach, April 1987, pp. 53–60.

Rachev, S. T. (1989a). The stability of stochastic models. *Applied Prob. Newsletter, ORSA*, **12**, 3–4.

Rachev, S. T. (1989b). The problem of stability in queueing theory. *Queueing Systems Theory Appl.*, **4**, 287–318.

*Rachev, S. T. and Chobanov, G. S. (1983). Existence and uniqueness on a boundary Gibbsian distribution. Dobrushin–Sinai approach. In: *Lectures on Stochastic Problems in Contemporary Physics*, Izd-vo Sofia University, Sofia, pp. 42–60. (In Bulgarian).

Rachev, S. T., Dimitrov, B., and Khalil, Z. (1990). A probabilistic approach to optimal quality usage. *Comput. Math. Appl.* To appear.

*Rachev, S. T. and Ignatov, Zv. (1984). Ideal quadratic metrics, In: *Stability problems for stochastic models. Proceedings*, Moscow, VNIISI, 1984, pp. 119–128. (In Russian). (Engl. transl. (1989) *J. Sov. Math.*, **35**, 2386–2394.)

*Rachev, S. T., Klebanov, L. B., and Yakovlev, A. I. (1988). An estimate of the rate of convergence to the limit distribution for the minima scheme for random number of identically distributed random variables: *Stability Problems of Stochastic Models Proceedings*, Moscow, VNIISI, pp. 120–124. (In Russian).

Rachev, S. T. and Resnick, S. (1989). Geometric max-infinitely divisible and geometric max-stable distributions. *Stochastic Models* (to appear).

Rachev, S. T. and Rüschendorf, L. (1989a). *Approximation independence of distributions on spheres and their stability properties.* Preprint, Angewandte Mathematik und Informatik, Universitat Munster, *Ann. Probability* to appear.

Rachev, S. T. and Rüschendorf, L. (1989b). Rate of convergence for sums and maxima and doubly ideal metrics. *Theory Prob. Appl.* (to appear).

Rachev, S. T. and Rüschendorf, L. (1990a). A transformation property of minimal metrics. *Theory Prob. Appl.*, **96**, 131–137.

Rachev, S. T. and Rüschendorf, L. (1990b). Approximation of sums by compound Poisson distributions with respect to stop-loss distances. *Adv. Appl. Prob.*, **22**, 350–374.

*Rachev, S. T. and Rüschendorf, L. (1990c). Recent results in the theory of probability metrics. *Statistics and Decisions* (to appear).

*Rachev, S. T. and Rüschendorf, L. (1990d). A counterexample to a.s. constructions. *Stat. Prob. Lett.*, **9**, 307–309.

*Rachev, S. T., Rüschendorf, L., and Schief, A. (1988). *On the construction of almost surely convergent random variables.* Preprint, Angewandte Mathematik und Informatik, Universitat Münster.

*Rachev, S. T. and Samorodnitsky, G. (1990) *Distributions arising in modelling environmental processes.* Preprint. Operations Research and Industrial Engineering, Cornell University.

Rachev, S. T. and Shortt, R. M. (1989). Classification problem for probability metrics. Contemporary Mathematics, "Measure and Measurable Dynamics," *J. Am. Math. Soc.*, **96**, 221–262.

*Rachev, S. T. and Shortt, R. M. (1990) *Duality Theorems for Kantorovich–Rubinstein and Wasserstein Functionals.* Dissertationes Mathematicae. Vol. 299.

Rachev, S. T. and Taksar, M. (1989). *Kantorovich's functionals in space of measures.* Technical Report no. 103, Nov. 1989, Dept. of Statistics and Applied Probability, University of California, Santa Barbara.

Rachev, S. T. and Todorovic, P. (1990). On the rate of convergence of some functionals of a stochastic process. *J. Appl. Prob.*, **27** (to be published).

*Rachev, S. T. and Yakovlev, A. Yu. (1988a). Bounds for the probabilistic characteristics of latent failure times within competing risks framework. *Serdica*, **14**, 325–332.

*Rachev, S. T. and Yakovlev, A. Yu. (1988b). Theoretical bounds for the tumor treatment efficacy. *Syst. Anal. Model Simul.*, **5**, 37–42.

Rachev, S. T. and Yukich (1989a). Rates for the CLT via new ideal metrics. *Ann. Prob.*, **17**, 775–778.

Rachev, S.T. and Yukich, J. E. (1989b). Smoothing metrics for measures on groups. *Ann. Inst. H. Poincare*, **25**, 429–441.

Rachev, S. T. and Yukich, J. E. (1990). Rates of convergence of α-stable random motions. *J. Theor. Prob.* (to appear).

*Rachev, S. T., Yakovlev, A. J., Kadyrova, N. O., and Myasnikova, E. M. (1988). On the statistical inference from survival experiments with two types of failure. *Biometrical J.* **30**, 835–842.

Ranga, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Stat.*, **33**, 659–680.

*Rao, B. V. (1969). On discrete Borel sets and projective spaces. *Bull. Am. Math. Soc.*, **75**, 614–617.

*Rao, B. V. (1971). Lattice of Borel structures. *Colloq. Math.*, **23**, 213–216.

*Rao, B. V. and Bhaskara Rao, K. P. S. (1981). *Borel Spaces.* Dissertationes Mathematicae CXC.

*Rao, C. R. and Shanbhag, D. N. (1986). Recent results on characterization of probability distributions: a unified approach through extensions of Denny's theorem. *Adv. Appl. Prob.*, **18**, p. 660–1986.
*Rao, M. (1971). Measurable selections of representing measures. *Q. J. Math.*, **22**, 571–573.
*Rattray, B. A. and Peck, J. E. L. (1955). Infinite stochastic matrices. *Trans. Roy. Soc. Canada, Sec. III*, **49**, 55–57.
*Reiss, R. D. (1981). Approximation of Product Measures With an Application to Order Statistics. *Ann. Prob.* **9**, 335–341.
*Reiss, R. D. (1989). *Approximate Distributions of Order Statistics.* Springer, New York.
Resnick, S. I. (1987a). *Extreme values, Regular Variation and Point Processes.* Springer, New York.
Resnick, S. I. (1987b). Uniform rates of convergence to extreme-value distributions. In: *Probability and Statistics: Essays in honor of Franklin A. Graybill* (J. Srivastava, (ed.). North-Holland, Amsterdam.
*Ressel, P. (1985). De Finetti-type theorems: an analytic approach. *Ann. Prob.* **13**, 898–922.
*Revesz, P. (1962). A probabilistic solution of problem III of G. Birkhoff. *Acta Math. Acad. Sci. Hungar.*, **13**, 187–198.
*Richardson, G. D. (1982). A convergence structure on a class of probability measures. *Rocky Mountain J. Math.*, **12**, 30–41.
*Richter, H. (1957). Parameterfreie Abschatzung und Realisierung von Erwartungswerten. *Blatter Deutscher Gesellschaft Versicherungsmathematik*, 3, 147–161.
Rinott, Y. (1973). Multivariate majorization and rearrangement inequalities with some applications to probability and statistics. *Israel J. Math.*, **45**, 60–77.
Robbins, H. (1975). The maximum of identically distributed random variables. *I.M.S. Bull.*, March (abstract), pp. 75–76.
Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press.
Rogosinky, W. W. (1958). Moments of non-negative mass. *Proc. R. Soc.* **245A**, 1–27.
*Rolski, T. (1975). Mean residual life. *Bull. ISI*, **46**, 266–270.
*Rolski, T. (1976). *Order relations in the set of probability distributions and their applications in queueing theory.* Dissertationes Mathematicae, Warszawa.
*Rolski, T. (1983). Mean residual life. *Z. Wahrsch. Verw. Geb.*, **33**, 714–718.
*Rosenblatt, M. (1979). Some limit theorems for partial sums of quadratic forms of stationary Gaussian variables. *Z. Wahrsch. Verw. Geb.*, **49**, 125–132.
*Roussas, G. G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge University Press, Cambridge.
*Rota, G. C. (1960). On the representation of averaging operators. *Rend. Sem. Math. Univ. Padova*, **30**, 52–64.
*Rota, G. C. (1962). An "alternierende verfahren" for general positive-operators. *Bull. Amer. Math. Soc.*, **68**, 95–102.
*Royden, H. L. (1968). *Real Analysis.* McGraw Hill, New York.
*Rubinstein, G. Sh. (1965). *Studies of dual external problems*, DSc. Dissertation, Novosibirsk.
Roynette, B. (1974). Theoreme central-limite pour le group de deplacements de R^d *Ann. Inst. H. Poincare*, **10B**, 391.
Rubinstein, G. Sh. (1970). Duality in mathematical programming and some questions in convex analysis. *Usp. Mat. Nauk.*, **25**, 171–201. (In Russian).
*Rudin, M. E. (1977). McFartin's Axiom. In: *The Handbook of Mathematical Logic*, North-Holland, Amsterdam, pp. 491–501.

Rüschendorf, L. (1980). Inequalities for the expectation of Δ-monotone functions. *Z. Wahrsch., Verw. Geb.*, **54**, 341–349.
Rüschendorf, L. (1981). Sharpness of Fréchet-Bounds. *Z. Wahrsch. Verw. Geb.*, **57**, 293–302.
Rüschendorf, L. (1982). Random variables with maximum sums. *Adv. Appl. Prob.* **14**, 623–632.
Rüschendorf, L. (1983a). Solution of a statistical optimization problem by rearranging methods. *Metrika* **30**, 55–62.
Rüschendorf, L. (1983b). On the multidimensional assignment problem. *Meth. OR* **47**, 107–113.
*Rüschendorf, L. (1984). On the minimum discrimination information theorem. *Stat. Decisions, Suppl.* **1**, 263–283.
*Rüschendorf, L. (1985a). Construction of multivariate distributions with given marginals. *Ann. Inst. Stat. Math.*, **37**, 225–233.
Rüschendorf, L. (1985b). The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Geb.*, **70**, 117–129.
Rüschendorf, L. (1989). *Conditional stochastic order and partial sufficiency.* Preprint 'Angewandte Mathematik und Informatik', Universitat Munster.
*Rüschendorf, L. and Rachev, S. T. (1990). A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, **32**, 48–54.
*Ryff, J. V. (1963). On the representation of doubly stochastic operators. *Pacific J. Math.*, **13**, 1379–1386.
*Saks, S. (1932). On the functions of Besicovitch. *Fund. Math.*, **19**, 211–219.
Salvemini, T. (1939). *Sugli indici di omofilia.* Atti 1 Riunione della Soc. Ital. di Statistica, Pisa.
Salvemini, T. (1943). *Sul calcolo degli indici di concordanza tra due caratteri quatitativi.* Atti della VI Riunione della Soc. Ital. di Statistica, Roma.
Salvemini, T. (1949). Nuovi procedimenti per il calcolo degli indici di dissomiglianza e di connessione, *Statistica*, **IX**, pp. 3–36.
Salvemini, T. (1953). *L'indice quadratico di dissolianza fra distruzioni gaussiane.* Atti dell XIII e XIV Riunione Scientifica, Societa Italiana di Statistica, pp. 297–301.
Salvemini, T. (1957a). L'indice di dissomiglianza fra distributioni continue. *Metron*, **16**, 75–100.
Salvemini, T. (1957b). *L'indice quadratico di dissomiglianza fra distribuzioni guassiane.* Atti della XIII e XIV Riunione Scientifica, Società Italiana di Statistica, 1953–1954, pp. 297–301.
Samorodnitski, G. and Taqqu, M. S. (1989). *Stable Random Processes.* In preparation.
Samuel, E. and Bachi, R. (1964). Measures of the distance of distribution functions and some applications. *Metron*, **XXIII**, 83–122.
*Sarbadhikari, H., Bhaskara Rao, K. P. S., and Grzegorek, E. (1974). Complementation in the lattice of Borel structures. *Colloq. Math.*, **31**, 29–32.
*Sazonov, V. V. (1954). On perfect measures, *Am. Math. Soc. Trans. Series 2*, **(48)**, 229–254.
*Sazonov, V. V. (1972). *On a bound for the rate of convergence in the multi-dimensional central limit theorem.* Proc. Sixth Berkeley Symp. Math. Statist. Probab. **2**, pp. 563–581. University Calif. Press.
Sazonov, V. V. (1981). *Normal Approximation—Some Recent Advances. (Lect. Notes Math.*, **879**) Springer, New York.
Sazonov, V. V. and Ul'Yanov, V. V. (1979). On the speed of convergence in the central limit theorem. *Adv. Appl. Prob.*, **11**, 269–270.
*Schaefer, M. (1976). Note on the k-dimensional Jensen inequality. *Ann. Prob.*, **2**, 502–504.

Schay, G. (1979). Nearest random variables with given distributions. *Ann. Prob.*, **2**, 163–166.

Schay, G. (1979). Optimal joint distributions of several random variables with given marginals. *Stud. Appl. Math.*, **LXI**, 179–183.

*Schilling, M. F. (1983). Goodness of fit testing based on the weighted empirical distribution of certain nearest neighbor distances. *Ann. Stat.*, **11**, 1–12.

Schweizer, B. and Sklar, A. (1983). *Probabilistic metric spaces.* Elsevier–North-Holland, New York.

Senatov, V. V. (1977). On some properties of metrics in the set of distribution functions. *Math, USSR Sbornik*, **31**, 379–387.

Senatov, V. V. (1980). Uniform estimates of the rate of convergence in the multidimensional central limit theorem. *Theory Prob. Appl.*, **25**, 745–759.

Senatov, V. V. (1981). Some lower estimates for the rate of convergence in the central limit theorem in Hilbert space. *Sov. Math. Dokl.*, **23**, 188–192.

Sendov, B. (1969). Some problems of the theory of approximations of functions and sets in the Hausdorff metrics. *Usp. Mat. Nauk*, **24/5**, 141–173.

Sendov, B. (1979). *Hausdorff Approximations.* Bulgarian Academic Scientific Press, Sofia.

Sendov, B. and Penkov, B. (1962). ε-entropy and ε-capacity on the space of continuous functions. *Izv. Math. Inst. Bulgarian Akad. Nauk*, **6**, 27–50. (In Bulgarian).

*Serfling, R. J. (1975). A general Poisson approximation theorem. *Ann. Prob.*, **3**, 726–731.

*Shaked, M. (1977b). A concept of positive dependence for exchangeable random variables. *Ann. Stat.*, **5**, 505–515.

*Sherman, S. (1952). On a conjecture concerning doubly stochastic matrices. *Proc. Am. Math. Soc.*, **3**, 511–513.

*Shiganov, I. S. (1970). A refinement of the upper bound of the constant in the remainder term of the central limit theorem. In: *Stability problems for Stochastic Models*, VNIISI, Moscow, pp. 109–115. (In Russian).

Shiryayev, A. N. (1984). *Probability.* Springer, New York.

*Shnirman, M. G. (1980). On the question of ergodicity of a Markov chain, *Problemy Kib.*, **20**, 125–222. (In Russian).

Sholpo, I. A. (1983). ε-minimal metrics. *Theory Prob. Appl.*, **28**, 854–855.

Shohat, J. A. and Tamarkin, J. D. (1943). *The Problem of Moments.* American Mathematical Society, Providence, RI.

*Shor, P. W. and Yukich, J. E. (1990). Minimal grid matching and empirical measures. *Ann. Prob.* To appear.

Shorak, G. R. and Wellner, J . A. (1986). *Empirical Processes With Applications to Statistics.* Wiley, New York.

*Shortt, R. M. (1983a). Universally measurable spaces: an invariance theorem and diverse characterizations. *Fund. Math.*, **121** 35–42.

*Shortt, R. M. (1983b). Strassen's marginal problems in two or more dimensions. *Z. Wahrsch. Verw. Geb.*, **64**, 313–325.

*Shortt, R. M. (1983c). Combinatorial methods in the study of marginal problems over separable spaces. *J. Math. Anal. Appl.*, **97**, 462–479.

*Shortt, R. M. (1983d). Empirical measures. *Am. Math. Monthly*, **91**, 358–360.

*Shortt, R. M. (1984a). Products of Blackwell spaces and regular conditional probabilities. *Real Analysis Exchange*, **10**, 31–41.

*Shortt, R. M. (1984b). Minimal completion and maximal conjugation for partitions, with an application to Blackwell sets. *Fund. Math.*, **CXXIII**, 211–223.

*Shortt, R. M. (1984c). Borel-density, the marginal problem and isomorphism types of analytic sets. *Pacific J. Math.*, **113**, 183–200.

*Shortt, R. M. (1985a). Uniqueness and extremality for a class of multiply-stochastic measures. *Prob. Math. Stat.*, **5**, 225–233.

*Shortt, R. M. (1985b). Reticulated sets and the isomorphism of analytic powers. *Pacific J. Math.*, **119**, 215–226.
*Shortt, R. M. (1986a). Product sigma-ideals. *Topology Appl.*, **32**, 279–290.
*Shortt, R. M. (1986b). The singularity of extremal measures. *Real Analysis*, **12**, 205–215.
*Sierpinski, W. (1930). Sur l'extension des fonctions de Baire definies sur les ensembles lineaires quelconques, *Fund. Math.*, **16**, 31.
*Sierpinski, W. (1931). Sur deux complementaires analytiques non separables. *B.*, *Fund Math.*, **17**, 296.
*Sierpinski, W. (1934). Sur une properté des ensembles lineaires quelconques. *Fund Math.*, **23**, 125–134.
*Simon, B. A. (1979). A remark on Dobrushin's uniqueness theorem. *Comm. Math. Phys.*, **68**, 183–185.
*Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *Am. Math. Monthly*, **74**, 402–405.
*Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific J. Math.*, **21**, 343–348.
Sklar, M. (1959). *Fonctions de répartition à dimensions et leurs marges.* Institute of Statistics, University of Paris 8, Paris, pp. 229–231.
Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theory Prob. Appl.*, **1**, 261–290.
Smith, R. L. (1982). Uniform rates of convergence in extreme-value theory. *Adv. Appl. Prob.*, **14**, 600–622.
Smith, C. S. and Knott, M. (1987). Note on the optimal transportation of distributions. *J. Optim. Theory Appl.*, **52**, 323–329.
Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Stat.*, **36**, 423–439.
*Steerneman, T. (1983). On the total variation and Hellinger distance between signed measure; an application to product measures. *Proc. Am. Math. Soc.*, **88**, 684–688.
*Stout, W. F. (1974). *Almost sure convergence.* Academic Press, New York.
Stoyan, P. (1983). *Comparison Methods for Queues and Other Stochastic Models.* Wiley, Chichester.
*Sudakov, V. N. (1972a). Measures on subsets of direct products. *Zap. Nauchn. Sem. LOMI*, **29**, 113–128. (In Russian).
*Sudakov, V. N. (1972b). On the existence of bistochastic densities. *Zap. Nauchn. Sem. LOMI*, **29**, 129–146.
*Sudakov, V. N. (1975). Measures on subsets of direct products, *J. Sov. Math.*, **3**, 825–839.
Sudakov, V. N. (1976). *Geometric problems in the theory of infinite-dimensional probability distributions*, Tr. Mat. Inst. V. A. Steklov, Akad. Nauk, SSSR, **141.** (In Russian). (Engl. transl. (1979) *Proc. Steklov Inst. Math.* (1979). no 2, 141).
Szasz, D. (1975). On a metrization of the vague convergence. *Studia Sci. Math. Hungarica*, **9**, 219–222.
Szulga, A. (1978) On the Wasserstein metric. In: *Transactions of the 8th Prague Conference on Information theory, Statistical Decision Functions, and Random Processes, Prague, vol. B.* Akademia, Prague, pp. 267–273.
Szulga, A. (1982). On minimal metrics in the space of random variables. *Theory Prob. Appl.*, **27**, 424–430.
*Szulga, A. (1983). On the metrics of the type ζ. *(Lect. Notes Math.*, **982**) Springer, Berlin, pp. 238–243.
*Szwarc, W. and Posner, M. (1984). The tridiagonal transportation problem. *Op. Res. Lett.*, **3**, 25–30.

*Tanaka, H. (1978). Probabilistic treatment of Boltzmann equation of Maxwellian molecules. *Z. Wahrsch. Verw. Geb.*, **46**, 67–105.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Ann. Prob.*, **8**, 814–827.

*Temel't, V. (1969) Duality of infinite-dimensional problems of linear programming and some related questions. *Kibernetika* **4**, 80–87. (In Russian).

Teugels, J. L. (1985). *Selected Topics in Insurance Mathematics.* Katholieke Universiteit Leuven, Leuven.

*Thompson, W. A. and Brindley, E. C. Jr. (1972). Dependence and aging aspects of multivariate survival. *J. Am. Stat. Assoc.*, **67**, pp. 822–830.

*Tibshirani, R. and Wasserman, L. A. (1988). Sensitive parameters. *Can. J. Stat.*, **16**, 185–192.

*Timan, A. F. (1960). *Theory of Approximation of Real Functions.* Fimatgiz, Moscow. (In Russian).

Todorovic, P. (1987). An extremal problem arising in soil erosion modeling. In: *Applied Probability, Stochastic Processes and Sampling Theory*, I. B. MacNeil and G. J. Umphrey (eds), Reidel, Dordrecht, pp. 65–73.

Todorovic, P. and Gani, J. (1987). Modeling of the effect of erosion on crop production. *J. Appl. Prob.*, **24**, 787–797.

Todorovic, P. , Woolhiser, D. A., and Renard, K. G. (1987). Mathematical model for evaluation of the effect of soil erosion on crop productivity. *Hydrolog. Process.*, **1**, 181–198.

*Toom, A. L. (1968). A family of networks of formal neutrons. *Sov. Math. Dokl.*, **9**, 1338–1344.

*Topsoe, F. (1970). *Topology and Measure.* (*Lect. Notes Math.* **133**) Springer, Berlin.

Tutubalin, V. N. (1967). The central limit theorem for random motions of Euclidean space. *Vestnik MGU, ser. I, Math. and Mech.*, **22**, 100–108 (In Russian).

*Ulam, S. (1930). Zur Masstheorie in der allgemeinen Mengenlehre. *Fund. Math. Dokl.*, **9**, 1338–1344.

*Urbanik, K. (1964). Generalized convolutions. *Studia Math.*, **23**, 217–245.

Valkeila, E. (1986). *Studies on distributional properties of counting processes.* PhD Thesis, University of Helsinki.

Vallander, S. S. (1973). Calculation of the Wasserstein distance between probability distributions on the line. *Theory Prob. Appl.*, **18**, 784–786.

Varadarajan, V. S. (1958a). Weak convergence of measures on separable metric space, *Sankhya*, **19**, 15–22.

*Varadarajan, V. S. (1958b). Measures on topological spaces. *Am. Math. Soc. Trans. Series 2*, **48**, 15–22.

*Varadarajan, V. S. (1963). Groups of automorphism of Borel spaces. *Trans. Am. Math. Soc.*, **109**, 191–220.

*Varadarajan, V. S. (1965). Measures on topological spaces. *Am. Math. Soc. Transl.*, **11**, 161–228.

Vershik, A. M. (1970). Some remarks on infinite-dimensional linear programming problems. *Usp. Mat. Nauk*, **25**, 117–124. (In Russian).

*Vershik, A. M. and Temel't, V. (1968). Some questions of approximation of the optimal value of infinite-dimensional linear programming problems. *Siberian Math. J.*, **9**, 591–601.

Vinogradov, I. M. (1975). *Trigonometrical sums in number theory.* Statistical Publishing Society, Calcutta.

*Vitale, R. E. and Pirkin, A. C. (1976). Conditions on the regression function when both variables are uniformly distributed. *Ann. Prob.*, **4**, 869–873.

*Vlach, M. (1983). *On the three planar sums transportation polytope.* Report n. 83312-OR, Institut fur Okonmetrie und Operations Research, Bonn, FRG.

*Von Alexits, G. (1930). Uber die Erweiterung einer Baireschen Funktion. *Fund. Math.*, **15**, 51–56.

*Von Bahr, B. and Esseen, C. G. (1965). Inequalities for the rth absolute moment of a sum of random variables, 1 r 2. *Ann. Math. Stat.*, **36**, 299–303.

*Vorob'ev, N. N. (1962). Consistent families of measures and their extensions. *Theory Prob. Appl.*, **7**, 147–163.

*Vorob'ev, N. N. (1963). Markov measures and Markov extension. *Theory Prob. Appl.*, **8**, 420–429.

Vygodskii, M. Ya. (1936). The origins of differential geometry. In: *G. Monge, Applications of Analysis to Geometry*, Main Editorial Staff of General Technical Disciplines, Moscow. (In Russian).

*Vyndev, S., Ignatov, T., and Rachev, S. T. (1982). Metrics that are invariant relative to monotone transformations. In: *Stability problems for stochastic models. Proceedings*, VNIISI, Moscow, pp. 25–36. (In Russian). (Engl. transl (1986), *J. Sov. Math.*, **35**, 2466–2476).

Wasserstein, L. N. (1969). Markov processes over denumerable products of spaces describing large systems of automata. *Prob. Inf. Transmission*, **5**, 47–52.

Wellner, Jon A. (1981). A Glivenko–Cantelli theorem for empirical measures of independent but nonidentically distributed random variables, *Stochastic Process. Appl.*, **11**, 309–312.

*Whitmore, G. A. (1970). Third degree stochastic dominance. *Am. Economic Rev.*, **60**, 457–459.

Whitt, W. (1974). *Heavy traffic limit theorems for queues: A survey. (Lect. Notes Economics Math. Syst.*, **98**) Springer, Berlin.

Whitt, W. (1974). The continuity of queues. *Adv. Appl. Prob.*, **6**, 175–183.

*Whitt, W. (1976). Bivariate distributions with given marginals. *Ann. Stat.*, **4**, 1280–1289.

*Whittle, P. (1982). *Optimization Over Time: Dynamic Programming and Stochastic Control.* Wiley, New York.

Wolff, E. F. and Schweizer, B. (1981). On nonparametric measures of dependence for random variables. *Ann. Stat.*, **9**, 879–885.

Woyczynski, W. A. (1980). On Marcinkiewicz–Zygmund laws of large numbers in Banach spaces and related rates of convergence. *Prob. Math. Stat.*, **1**, 117–131.

*Yakovlev, A. Ju. and Rachev, S. T. (1988). Bounds for crude survival probabilities within competing risks framework and their statistical application. *Stat. Prob. Lett.*, **6**, 389–394.

*Yamukov, G. I. (1977). Estimates of generalized Dudley metrics in spaces of finite-dimensional distributions. *Theory Prob. Appl.*, **22**, 579–584.

*Yanagimoto, T. (1972). Families of positively dependent random variables. *Ann. Inst. Stat. Math.*, **24**, 559–573.

*Yosida, K. (1978). *Functional Analysis.* Springer, Berlin.

*Young, H. P. (1987). On dividing an amount according to individual claims or liabilities. *Math. Oper. Res.*, **12**, 398–414.

Yukich, J. E. (1985). The convolution metric d_g. *Math. Proc. Cam. Phil. Soc.*, **98**, 533–540.

Yukich, J. E. (1989). Optimal matching and empirical measures. *Proc. Am. Math. Soc.*, **107**, 1051–1059.

Zabell, S. L. (1980). Rates of convergence for conditional expectations. *Ann. Prob.*, **8**, 928–941.

Zaitsev, A. Yu. (1983). On the accuracy of approximation of distributions of sums of independent random variables that are nonzero with small probability by accompanying laws. *Theory Prob. Appl.*, **28**, 657–669.

Zolotarev, V. M. (1962). *Counterpart of the Edgeworth–Cramer asymptotic expansion for the case of approximation with stable distribution law.* Proc. Sixth All-Union Conf. Vilnius, pp. 49–50.

Zolotarev, V. M. (1967a). Sur une généralization du théoreme de Lindeberg–Feller et des questions qui s'y rattachent. *C.R. Acad. Sci.*, **265**, 260–267.

Zolotarev, V. M. (1967b) Théoremes limités pourles sommes des variables aléatoires indépendentes qui ne sont pas infinitesimales. *C.R. Acad. Sci.*, **264**, 799–800.

Zolotarev, V. M. (1970). Théoremes limités généraux pour les sommes des variables aléatoires indépendentes. *C.R. Acad. Sci.*, **270**, 899–902.

Zolotarev, V. M. (1975a). On the continuity of stochastic sequences generated by recurrence procedures. *Theory Prob. Appl.*, **20**, 819–832.

*Zolotarev, V. M. (1975b). Qualitative estimates in problems of continuity of queueing systems. *Theory Prob. Appl.*, **20**, 211–213.

Zolotarev, V. M. (1976a). The effect of stability for characterization of probability distributions. *Zap. Nauchn. Sem. LOMI*, **61**, 38–55. (In Russian).

Zolotarev, V. M. (1976b). Metric distances in spaces of random variables and their distributions. *Math. USSR sb.*, **30**, 373–401.

Zolotarev, V. M. (1976c). On stochastic continuity of queueing systems of type G\G\1. *Theory Prob. Appl.*, **21**, 250–269.

Zolotarev, V. M. (1976d). Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory Prob. Appl.*, **21**, 721–737.

Zolotarev, V. M. (1977a). *General problems of mathematical models.* Proceedings of 41st session of ISI, New Delhi, pp. 382–401.

Zolotarev, V. M. (1977b). Ideal metrics in the problem of approximating distributions of sums of independent random variables. *Theory Prob. Appl.*, **22**, 433–449.

Zolotarev, V. M. (1978). On pseudomoments. *Theory Prob. Appl.*, **23**, 269–278.

Zolotarev, V. M. (1979a). Ideal metrics in the problem in topological space and probability theory. *Austral. J. Stat.*, **21**, 193–208.

Zolotarev, V. M. (1979b). On properties and relationships between certain types of metrics. *Zap. Nauchn. Sem. LOMI*, **87**, 18–35. (In Russian).

Zolotarev, V. M. (1981). On the structure of ideal metrics, In: *Stability Problems for stochastic models.* VNISI, Moscow. pp. 30–38. (In Russian).

Zolotarev, V. M. (1983a). *Foreword. (Lec. Notes Math.*, **982**) Springer, New York. pp. VII–XVII.

Zolotarev, V. M. (1983b). Probability metrics. *Theory Prob. Appl.*, **28**, 278–302.

Zolotarev, V. M. (1986). *Contemporary Theory of Summation of Independent Random Variables.* Nauka, Moscow. (In Russian).

Zolotarev, V. M. and Rachev, S. T. (1985). *Rate of convergence in limit theorems for the Max-scheme. (Lect. Notes in Math.* **1155**) Springer, Berlin, pp. 415–442.

*Zubkov, A. M. (1985). *On compound probabilistic metrics with given minimal metrics. (Lect. Notes Math.*, **1155**) Springer, Berlin, pp. 443–447.

Table of Probability Semidistances

Probability semidistance	Definition	Type	Properties	Applications and relationships
AS_p (Szulga's metric)	(4.3.53)	Simple metric, ideal metric of order 1	Convergence, see Lemma 4.3.3	Dual representation for $\mathscr{L}_p$, see pp. 81, 82
$\mathbf{D}$ (the metric in the space of pth moments)	(8.2.24)	Primary metric	Inequalities, see (8.2.21), (8.2.22), (8.2.23)	Quantitative criteria for ℓ_p-convergence
$\mathbf{D}_{n,\alpha}$ (Ornstein type metric)	(5.1.34)	Simple metric, ideal metric of order 1	See ℓ_1-metric	Dynamical systems and information theory, see pp. 97, 98
$\mathbf{d}_m$ (stop-loss metric of order m)	(17.1.10)	Simple metric, ideal metric of order m	Relation with other metrics, see (17.1.14), (17.1.37), (17.1.39), (17.1.40), (17.1.43), (17.1.45), (17.1.49), (17.1.54); smoothing in equality, see Lemma 17.1.7	Risk theory, see Theorems 17.2.1, 17.2.2, 17.2.3, 17.2.4; CLT, see Theorem 17.1.1
$\mathbf{d}_{m,p}$(L_p-version of $\mathbf{d}_m$)	(17.1.11)	Simple metric, ideal of order $r = m + \frac{1}{p}$	Dual representation, $\mathbf{d}_{m,p} = \zeta_{m,p}$, see Lemmas 17.1.1, 17.1.5	See the applications of $\zeta_{m,p}$
EN (engineer's metric in the space of random variables)	(2.1.1)	Primary metric, ideal metric of order 1	Relation with $\mathscr{L}_1$, see Corollary 9.1.3	Primary h-minimal metric w.r.t. $\mathscr{L}_1$
EN $(\cdot, \cdot; p)$ (L_p-engineer's metric)	(3.1.15)	Primary metric, ideal metric of order 1	Relation with $\mathscr{L}_p$ see (9.1.24)	Primary h-minimal metric w.r.t. $\mathscr{L}_p$

$\mathbf{EN}(\cdot, \cdot; H)$ (engineer's distance between random vectors)	(3.1.14)	Primary distance	Relation with $\mathscr{L}_H$, see (9.1.23)	Primary h-minimal metric w.r.t. $\mathscr{L}_H$
$\mathbf{H}_\lambda$ (Hausdorff metric with parameter λ)	(4.1.32)	Simple metric, (1, 0)-weakly perfect metric, see p. 430	Limiting properties, see (4.1.33) p. 65; convergence, completeness, see pp. 65–67	H_λ metrizes the same topology as **SB**-metric and has Hausdorff representation
$\mathbf{K}$ (Ky Fan metric)	(2.1.5), (2.3.2), (3.38), p. 40	Compound metric	Minimality: $\hat{\mathbf{K}} = \pi$, Corollary 7.4.2	**K**-metrizes convergence in probability. The minimal metric $\pi = \hat{\mathbf{K}}$ has a Hausdorff representation
$\mathbf{K}^*$ (Ky Fan metric)	(2.1.6)	Compound metric	Version of $\mathbf{K}(X, Y)$, see (2.1.5); minimal metric: $\hat{\mathbf{K}}^* = \mathscr{A}_c = \hat{\mu}_c = \mathring{\mu}_c$, see Sections 6.1.1, 11.1.1 with $c(x, y) = \dfrac{d(x, y)}{1 + d(x, y)}$	$\mathbf{K}^* \overset{\text{top}}{\sim} \mathbf{K}$. The minimal metric $\hat{\mathbf{K}}^*$ has a ζ-representation
$\mathbf{K}_\lambda$, $0 < \lambda < \infty$ (parametric family of Ky Fan metrics)	(3.3.10), pp. 40, 41	Compound metric	Minimality $\hat{K}_\lambda = \pi_\lambda$, see (7.4.12); limit relations, see (3.3.11); Λ-structure, see Example 4.2.3	$\mathbf{K}_\lambda \overset{\text{top}}{\sim} \mathbf{K}$, $\lim_{\lambda \to 0} \mathbf{K}_\lambda = \mathscr{L}_0$, $\lim_{\lambda \to \infty} \lambda \mathbf{K}_\lambda = \mathscr{L}_\infty$
$\mathbf{KF}_H$ (Ky Fan distance)	(3.3.9)	Compound distance	Λ-structure, see Example 4.2.3	For any strictly increasing $H \in \mathscr{H}$, $\mathbf{KF}_H \overset{\text{top}}{\sim} \mathbf{K}$

Continued

Probability semidistance	Definition	Type	Properties	Applications and relationships
Kr (Kruglov's distance in the space of d.f.'s on $\mathbb{R}$)	(2.2.5)	Simple distance	Convergence: if $\mathbf{Kr}(F_n, G) < \infty$ for $n \geqslant 0$, then $\mathbf{Kr}(F_n, F_0) \to 0$ iff $F_n \xrightarrow{d} F_0$ and $\mathbf{Kr}(F_n, G) \to \mathbf{Kr}(F_0, G)$ (see p. 183)	**Kr** is a suitable metric for the problem of rate of convergence in the moment (global) limit theorems, see p. 220, Kruglov (1973)
L (Lévy metric between the d.f.'s)	(2.1.3), p. 148, (18.2.67)	Simple metric; weakly (1, 0)-perfect metric	K-minimality: see (7.3.5); estimates from above: see (16.2.12), (16.2.21); estimates from below: see (18.1.19)	L metrizes the weak convergence in the space $\mathscr{F}(\mathbb{R}^n)$ of multivariate d.f.'s
$\mathbf{L}_\lambda$, $0 < \lambda < \infty$ (parametric version of the Lévy metric)	(4.1.3), (4.1.22)	Simple metric; weakly (1, 0)-perfect metric	Limit relations: see (4.1.4), (4.1.5), (4.1.24), (4.1.25); Hausdorff representations: see (4.1.13), (4.1.17), (4.1.23); Λ-structure, see Example 4.2.1	$\mathbf{L}_\lambda \overset{\text{top}}{\sim} \mathbf{L}$, $\lim_{\lambda \to 0} \mathbf{L}_\lambda = \rho$, $\lim_{\lambda \to \infty} \lambda \mathbf{L}_\lambda = \mathbf{W}$
$\mathbf{L}_{\lambda, H}$ (Lévy distance in the space of n-dimensional d.f.'s)	(4.1.27)	Simple distance	Limit relations: see (4.1.29), (4.1.30); Hausdorff representation: see (4.1.28); Λ-structure: see p. 70	For any strictly increasing $H \in \mathscr{H}$, $\mathbf{L}_{\lambda, H} \overset{\text{top}}{\sim} L$, $\lim_{\lambda \to 0} \mathbf{L}_{\lambda H} = \rho_H$, $\lim_{\lambda \to \infty} \lambda \mathbf{L}_{\lambda, H} = \mathbf{W}$
$\mathscr{L}_p$ (L_p-metric in the space of U-valued	(2.3.2)–(2.3.5), (3.3.3)–(3.3.7), (2.1.7)	Compound metric; $(1 \wedge p, I)$-ideal metric, see pp. 381,	Dual representations: (4.3.63); ℓ_p is the minimal metric w.r.t. $\mathscr{L}_p$, see	Rate of convergence for maxima of sums of i.i.d. r.v.'s, see Theorems 18.3.5, 18.3.6

r.v.'s, or in the space $\mathscr{P}_2$)		382	Corollary 5.2.2.(ii); the primary h-minimal metric has explicit representation (9.1.18)	
$\mathscr{L}_{p,r}$ ($\mathscr{L}_p$-probability metric in the space of $L_r(T)$-valued r.v.'s	(18.2.7)	Compound metric $(1 \wedge p, I)$-ideal metric, see pp. 381, 382	The minimal metric, $\hat{\mathscr{L}}_{p,r}$, equals $\ell_{p,r}$, see (18.2.10)	Rate of convergence for maxima of dependent r.v.'s, see Theorem 18.2.1
$\mathscr{L}_H$ (H-average compound distance in the space of U-valued random variables, or in the space $\mathscr{P}_2 = \mathscr{P}(U \times U)$)	(3.3.1), p. 12	Compound distance	Minimality property, see Corollary 5.2.2 (i)	$\mathscr{L}_H(P)$ is the total cost of transportation of mass in MKP with cost function $c(x, y) = H(d(x, y))$, see (5.1.1), (13.1.2)
ℓ (uniform metric between densities)	On $\mathbb{R}^n$, see (14.2.6); on the space of random motions, see (15.2.4)	Simple metric, ideal metric of order (-1)	Convergence, see p. 266	CLT, see Theorems 14.3.3, 15.2.2
ℓ_p (the minimal distance w.r.t. $\mathscr{L}_p$)	(3.2.12)–(3.2.14)	Simple metric	Dual representation: $\ell_p = \hat{\mathscr{L}}_p$, (5.2.18). Explicit representation: Corollary 7.3.6, p. 153	Solution of an infinite dimensional transportation problem with cost function $c(x, y) = d^p(x, y)$, Section 5.2; metrization of the vague convergence, Section 10.2; Glivenko–Cantelli theorem, Corollary 11.1.4, Theorem 11.1.6; functional CLT,

Continued

PROBABILITY SEMIDISTANCE	DEFINITION	TYPE	PROPERTIES	APPLICATIONS AND RELATIONSHIPS
				Corollaries 11.2.1, 11.2.3; stability of queueing systems, see Theorems 12.1.1, 12.2.1; convergence of processes, (19.3.19), (19.3.31), Theorem 19.3.4
$\ell_{p,r}$ (the minimal metric w.r.t $\mathscr{L}_{p,r}$)	(18.2.9)	Simple metric, ideal of order 1, max-ideal of order 1	Duality representation, (18.2.10); relations with other metrics, see (18.2.15), (18.2.16); convergence, see (18.2.18)	Rate of convergence of maxima of i.i.d. r.v.'s, see Theorem 18.2.1
ℓ_H (the minimal distance w.r.t. $\mathscr{L}_H$)	(3.2.10)	Simple distance	Dual representation: $\ell_H = \hat{\mathscr{L}}_H$, (5.2.17); explict representation, (7.3.18)	Solution of an infinite dimensional transportation problem with cost function $c(x, y) = H(d(x, y))$, see Section 5.2; Bernstein–Kantorovich functional limit theorem, see Theorem 11.2.1, Corollary 11.2.7; quality usage, see Section 13.1
$\mathscr{M}$ (moment metric)	p. 39	Primary metric	Primary h-minimal metric $\mathscr{L}_1$, see Theorem 9.1.2	Solution of a moment problem on a s.m.s. (U, d), Theorem 9.1.2
$\mathscr{M}(g)_{H,p}$, $p > 0$	(3.18)	Primary distance	Primary h-minimal metric, see Theorems 9.1.1, 9.2.1 if $d(x, y) = \|x - y\|$	Solution of moment problems (9.1.1), (9.2.1)

$\mathbf{MOM}_p$, $p \geqslant 0$ (metric in the space of pth absolute moments)	(2.1.9)	Primary metric	Primary h-minimal metric, see (9.1.18), (9.2.21)	Solution of moment problems (9.1.5), (9.2.4)
$\mathbf{m}$ (an n-dimensional extension of $E\|X - Y\|$)	(7.3.19)	Compound metric, ideal metric of order 1	K-minimal metric w.r.t. $\mathbf{m}$, see (7.3.20)	Optimal quality usage, p. 248
$\mathbf{R}_H$ (Birnbaum–Orlicz compound uniform distance)	(3.3.16)	Simple distance	Minimality: $\hat{\mathbf{R}}_H = \rho_H$, see (7.3.24)	$\mathbf{R}_H = \lim\limits_{\lambda \to 0} L_{\lambda, H}$
s (Skorokhod metric in the space of d.f.'s on $\mathbb{R}$)	p. 8	Simple metric	Convergence, see p. 8	s metrizes topology that is 'between' the weak-star topology and that of the Kolmogorov metric
SB (Skorokhod–Billingsley metric in the space of d.f.'s)	p. 66	Simple metric	Compactness, completeness, see pp. 66, 67	$\mathbf{SB} \overset{\text{top}}{\sim} s \overset{\text{top}}{\sim} H$, and $(\mathbf{SB}, \mathscr{F})$ is a complete metric space
Var (total variation metric)	(14.2.5)	Ideal metric of order 0	Smoothing inequalities: Lemmas 14.3.1, 14.3.2	Rate of convergence in the CLT, see Theorem 14.3.1, Lemma 15.2.4
W (uniform metric between the inverse functions of the d.f.'s)	(4.1.7) multivariate analog, see p. 148	Simple metric, ideal metric of order 1, max-ideal metric of order 1	Limit relations, see p. 53, (4.1.25); K-minimality, see Corollary 7.3.3	Rate of convergence for maxima of i.i.d. r.v.'s, see (18.2.35) with $\ell_\infty = W$

Continued

Probability semidistance	Definition	Type	Properties	Applications and relationships
$\tilde{\mathbf{W}}$	(4.1.34)	Simple metric, ideal of order 1	$\tilde{W} = \lim_{\lambda\to\infty} \lambda \mathbf{H}_\lambda$, see pp. 64, 65	$\tilde{W} \equiv W$ in the space of strictly increasing d.f.'s
β (Dudley metric)	p. 79	Simple metric	Convergence, see Corollary 4.3.6.	β metrizes the weak topology and has ζ-structure, i.e. it can be viewed as a seminorm in the space of signed measures with total mass zero
δ	p.148	Simple metric, ideal metric of order $r = +\infty$	Limiting relation with $\mathbf{L}_\lambda$, p. 148	Any ideal simple metric of order ∞ coincides with δ on the space of distributions that are completely determined by their moments
$\Delta_{r,p}$	(18.3.31)	Compound max-ideal metric of order $r(1 \wedge p)$, $r > 0$, $0 < p \leqslant \infty$	Explicit form for the minimal metric $\hat{\Delta}_{r,p}$, see (18.3.32)	Rate of convergence for maxima of dependent random variables, see p. 380
$\zeta_{\mathfrak{F}}$ (Zolotarev's semimetric $\zeta_{\mathfrak{F}}$)	p. 36, (4.3.5)	Simple semi-metric	Special cases of $\zeta_{\mathfrak{F}}$ are **EN**, ρ, κ, $l = l_0$, see p. 73	Most simple metrics μ have ζ-representation, or at least $\mu \overset{\text{top}}{\sim} \zeta_{\mathfrak{F}}$ for an appropriate choice of $\mathfrak{F}$. One counter-example is the H-metric, see Lemma 4.3.4
$\zeta(\cdot, \cdot; \mathscr{G}^p)$ (Fortet–Mourier metric)	(4.3.34)	Simple metric	Other representations, see (4.3.34), p. 93, convergence, Corollary 4.3.5(i)	Describe optimal value of a network flow problem with special cost function, see p. 93

$\zeta(\cdot, \cdot; \mathscr{G}^p)$ (Fortet–Mourier bounded metric)	(4.3.35)	Simple metric	Other representations, see (4.3.35); convergence, see Corollary 4.3.5(ii)	$\zeta(\cdot, \cdot, \mathscr{G}^p)$ metrizes the weak convergence. The structure of $\zeta(\cdot, \cdot, \mathscr{G}^p)$ is similar to that of β
ζ_r (Zolotarev's ideal metric of order r)	(14.2.1); on the space of random motions ζ_r is defined by (15.1.8)	Simple metric, ideal metric of order $r > 0$	Convergence, see (14.1.5); bounds from above (14.2.15), (15.1.17), (18.3.20); bounds from below (14.2.21), p. 270, (18.3.19)	Characterization of aging distributions, see Section 14.2; limit theorems, see Section 14.2
$\zeta_{m,p}$ (L^p-version of ζ_m)	(14.2.10)	Simple metric, ideal metric of order $r = m + 1/p$, see (16.2.8)	Relations with other ideal metrics, see (14.2.19), p. 270; bounds from below, see (16.2.11), (16.2.12)	Rate of convergence to self-similar processes, p. 435; stability in risk theory, see pp. 302, 303, 313
$\eta_{p,r}$ (weighted Prokhorov metric)	(18.2.51)	Simple metric, ideal of order $\dfrac{p}{1+p}$	Bounds, see (18.2.52); convergence, see Lemma 18.2.2(b)	$\eta_{p,r}$ metrizes convergence that is 'between' the weak convergence and $l_{p,r}$-convergence. It can be used in rate of convergence problem when the limit law is stable with index $\alpha < 2$, see Kakosyan *et al.* (1988)
Θ_p	(3.3.12), p. 382	Compound metric $(1/p, \text{III})$ ideal metric	$\hat{\Theta}_p \equiv \theta_p$, see (7.3.24); for possible generalizations, see Θ_H and $\Delta_{r,p}$	See generalized Kantorovich and Kantorovich–Rubinstein functionals, pp. 137, 138
Θ_H (Birnbaum–Orlicz compound distance)	(3.3.12), (3.3.15)	Compound distance	Minimality property: $\hat{\Theta}_H = \theta_H$, see (7.3.24)	Θ_H is the only known 'protominimal' distance for θ_H, see Rachev (1981a), Zolotarev (1986)

Continued

Probability semidistance	Definition	Type	Properties	Applications and relationships
θ_H (Birnbaum–Orlicz distance between distribution functions)	(3.2.26)	Simple distance	Representation as a minimal metric, see Theorem 7.3.4	Global limit theorems, see Kruglov (1973)
θ_p (L_p-metric between the d.f.'s of X and Y)	(2.1.4), (3.2.28), p. 32	Simple metric, ideal metric of order $\min(1, 1, p)$, max-ideal metric of order $\min(1, 1/p)$	Dual representation, p. 6; representation as a minimal metric, see (7.3.24)	Rate of convergence in the CLT in terms of θ_p, see Ibragimov and Linnik (1971)
$\theta_p(\cdot, \cdot; g)$ (weighted L_p-metric between d.f.'s on $\mathbb{R}^n$)	(4.3.12)	Simple metric	ζ-representation, see Theorem 4.3.1	Possible multivariate analog of θ_p
$\boldsymbol{\kappa}$ (Kantorovich metric in the space of d.f.'s)	p. 5	Simple metric, ideal metric of order 1, max-ideal metric of order 1	Dual representation, pp. 6, 28	Gini's index of dissimilarity, marginal problem, see p. 97
$\boldsymbol{\kappa}_r$ (difference pseudomoment of order $r > 0$)	(13.2.15), pp. 265, 359	Simple metric, max-ideal metric of order $r > 0$	Relations with ideal metrics, (14.2.14), (14.2.15); convergence $\kappa_r \overset{\text{top}}{\sim} \ell_r$, see p. 271; bounds, (17.1.44), (18.3.21)	The problem of stability in risk theory, see Sections 16.2, 16.3; rate of convergence for maxima of i.i.d. random vectors, see Theorem 18.2.2
$\boldsymbol{\kappa}_Q$ (Q-difference pseudomoment)	(4.3.40)	Simple metric	Explicit representation, (4.3.43)	CLT, see Hall (1981), Zolotarev (1976b,d)

$\mu(P)$, $P \in \mathscr{P}_2$ (probability (p) semidistance on $\mathscr{P}_2 = \mathscr{P}(U \times U)$)	Definition 2.3.1, p. 10	p. semidistance	General definition	Basic notion in the TPM
$\mu(X, Y)$ (probability (p.) semidistance on the space of U-valued r.v.'s)	Definition 2.3.2	p. semidistance	Given a 'rich enough' probability space (Remark 2.5.1) any μ on $\mathscr{P}_2$ is induced by $\mu(X, Y)$, see Section 2.5	Basic notion in the TPM
$\hat{\mu}(P_1, P_2)$, P_1, $P_2 \in \mathscr{P}(U)$ (minimal distance w.r.t. a compound distance μ)	Definition 3.2.2, p. 26	Simple distance	Example 3.2.1, pp. 26, 27; dual representations, Theorem 5.2.1 ($N = 1$); relationships with $\mathring{\mu}$, see (6.1.1), Theorem 6.2.1; convergence see Theorems 6.3.1, 8.2.1; uniformity, Section 6.4; explicit representations, see (8.1.26), (8.1.37)	Optimal translocation of masses, marginal problems, pp. 92, 93; $\hat{\mu}_c(P_1, P_2)$ is a solution of a transportation problem if P_i's are discrete measures
$\mathring{\mu}_c(P_1, P_2)$, P_1, $P_2 \in \mathscr{P}(U)$ (minimal norm)	(3.2.39)	Simple semimetric	Example 3.2.6, pp. 35, 36, 37; dual representation, see (5.4.2); explicit representation, see Theorem 5.4.1, (8.1.40); convergence, see Theorem 6.3.1; compactness, see Theorem 6.3.2; completeness, see Theorem 6.3.3; uniformity, see Theorem 6.4.1	Optimal translocation of masses in case of transits permitted, see pp. 92, 93; $\mathring{\mu}_c(P_1, P_2)$ is a solution of a network flow problem if P_i's are discrete measures; Glivenko–Cantelli theorem, see Corollaries 11.1.3, 11.1.4

Continued

Probability semidistance	Definition	Type	Properties	Applications and relationships
$\mu_h(P)$, $P \in \mathscr{P}_2$ (primary distance determined by a p. semidistance μ and a partition of $\mathscr{P}_2$)	Definition 3.1.1, p. 21	Primary distance	Section 3.1	Most of the primary distances in the TPM arise as solutions moment problems, see Sections 9.1, 9.2
$\mu\nu(P_1, P_2)$ (co-minimal distance w.r.t. compound distances μ and ν).	(3.2.36)	Simple metric	Example 3.2.5, pp. 32–35	If P_1 and P_2 are discete measures, $\mu\nu(P_1, P_2, \alpha)$ and $\mu\nu(P_1, P_2)$ are solutions of transportation problems with additional capacity constraints, see Barnes and Hoffman (1985)
$\overline{\mu\nu}(P_1, P_2)$	p. 34	Simple semidistance	Example 3.2.5, pp. 32–35	$\overline{\mu\nu}(P_1, P_2)$ is a limiting case of an infinite dimensional mass transportation problem with additional constraints
$\mu_{\theta,r}$ (convolution metric, a smoothing version of ℓ). If θ is α-stable, $\mu_{0,r} = \mu_r$, see (15.1.4), (15.1.7)	(14.2.12)	Simple metric, ideal metric of order $r-1$; in case of random motions, see Theorem 15.1.1	Bounds from above, see (14.2.14), (14.2.19); bounds from below, see (14.2.26); smoothing inequality, see Lemma 14.3.5	Rate of convergence in the local CLT, see Theorems 14.3.3, 15.2.2
ν (the first absolute pseudomoment) (ν_r	p. 253, (18.3.20)	Simple metric	Bounds, pp. 253, 254, (18.3.19), (18.3.21)	Rate of convergence for sums of i.i.d. random motions, see pp. 288, 376

the rth absolute pseudomoment)				
$\nu_{\theta,r}$ (convolution metric, a smoothing version of **Var**). If θ is α-stable $\nu_{\theta,r} = \nu_r$, see (15.1.3), (15.1.6)	(14.2.13)	Simple metric, ideal metric of order $r > 0$; r-ideal metric for random motions, see Theorem 15.1.2	Bounds from above, see (14.2.21), (14.2.15); bounds from below, see p. 273; convergence, see Theorem 14.2.1	Rate of convergence in the CLT w.r.t. the total variation metric, see Theorem 14.3.1, Lemma 15.2.4
$\tilde{\nu}_r$ (max-smoothing metric of order $r > 0$)	(18.1.1)	Simple metric, max-ideal metric of order $r > 0$	Max-smoothing inequality, (18.1.22); bounds from above, (18.1.26)	Rate of convergence for maxima of i.i.d. random vectors, see Theorem 18.1.1
$\xi_{p,r}$ (the minimal metric w.r.t. $\chi_{p,r}$)	(18.2.49)	Simple metric, ideal metric of order $\dfrac{p}{1+p}$, max-ideal metric of order $\dfrac{p}{1+p}$	Bounds, see (18.2.52), (18.2.54); convergence, see Lemma 18.2.2(b)	Rate of convergence for maxima of i.i.d. random variable, Theorem 18.2.3(b); rate of convergence for sums of i.i.d. r.v.'s, see Theorem 18.3.2
π (Prokhorov metric in the space of laws on a s.m.s.)	(3.2.18), (3.2.20)	Simple metric; if (U, d) is a Banach space, it is $(1, 0)$-weakly perfect metric, see p. 430	Hausdorff representation, see (4.1.18) with $\lambda = 1$; bounds, see (18.3.15), (18.3.18)	Metrization of the weak, G-weak and vague convergences, see Corollary 10.1.1, p. 210; rate of convergence for maxima of random processes, see (18.2.61); stability of queueing models, see Theorem 12.3.1; rate of convergence for maxima of sums of i.i.d. r.v.'s, see (18.3.47), (18.3.50)

Continued

PROBABILITY SEMIDISTANCE	DEFINITION	TYPE	PROPERTIES	APPLICATIONS AND RELATIONSHIPS
π_λ (parametric version of the Prokhorov metric)	(3.2.22)	Simple metric; if (U, d) is a Banach space, π_λ is $(1, 0)$-weakly perfect metric	Limit relations, see Lemma 3.2.1, p. 31; Hausdorff representation, see (4.1.18); Λ-structure, see Example 4.2.2, p. 70; minimality w.r.t. $\mathbf{K}_\lambda$, see Corollary 7.4.2	$\pi_\lambda \overset{\text{top}}{\sim} \pi$, $\lim\limits_{\lambda\to 0} \pi_\lambda = \sigma$, $\lim\limits_{\lambda\to\infty} \lambda\pi_\lambda = \ell_\infty$
π_H (Prokhorov distance in the space of laws on a s.m.s.)	(3.2.24), (3.2.25)	Simple metric	$\pi_H = \hat{\mathbf{K}}\mathbf{F}_H$, see Theorem 7.4.1, with $N = 2$	For any strictly increasing $H \in \mathscr{H}$, $\pi_H \overset{\text{top}}{\sim} \pi$
πH_λ (parametric version of πH_λ)	(4.1.38)	Simple metric	Limit relations, see (4.1.39), (4.1.40); topological structure, see Lemma 4.3.4	πH_λ is a Hausdorff metric that is 'between' π and σ and it is an important example in the TPM, see Lemma 4.3.4
ρ (Kolmogorov (uniform) metric)	On $\mathbb{R}$, see (2.12); on $\mathbb{R}^n$ see (4.1.24); on $\mathbb{R}^\infty$, see (18.2.65); on the space of random motions, see (15.2.2)	Simple metric, ideal metric of order 0, max-ideal metric of order	Dual representation, see p. 6; limiting relation, see (4.1.4); multivariate analogs, (4.1.24), (4.3.24), (4.3.25); smoothing inequality, (15.2.7); max-smoothing inequality, see (18.1.16), (18.1.27)	Berry–Esseen estimates in the CLT, see (14.3.16), (15.2.5); rate of convergence for maxima of random vectors and sequences, see Theorems 18.1.1, 18.1.2, 18.1.3, 18.2.5; characterization of distributions, see Theorems 19.1.3, 19.1.5
ρ_r (weighted Kolmogorov metric)	On $\mathbb{R}^n$, see (18.1.14); on $\mathbb{R}^\infty$, see (18.2.65)	Simple metric, max-ideal metric of order $r > 0$	max-smoothing inequality, see (18.1.23), (18.2.95); bound from below, see (18.1.26), (18.2.96)	Rate of convergence for maxima of i.i.d. random vectors, see Theorems 18.1.1, 18.2.5

ρ_H (Birnbaum–Orlicz uniform distance)	(3.2.27)	Simple distance	Representation as a minimal distance, see (7.3.24)	ρ_H is a limiting case for the Lévy distance $L_{\lambda,H}$, see (4.1.29)
σ (total variation metric)	(3.2.13), (14.2.4)	Simple metric, ideal metric of order 0	Dual representation $\sigma \equiv \ell_0$, see (3.2.13); limiting relations, see (3.2.23), (4.1.39), p. 80; minimality w.r.t. $\mathscr{L}_0$, see p. 160; sufficiency theorem, see (19.2.3)	Berry–Esseen estimates, see Theorem 14.3.1 with $\mathbf{Var} = 2\sigma$ (14.2.5); stability of characterizations of distribution, see Corollary 19.2.1, Theorem 19.2.2
τ_Q (compound Q-difference pseudomoment)	(4.3.41)	Compound metric	$\boldsymbol{\kappa}_Q = \hat{\tau}_Q$, see p. 79	Rate of convergence for maxima of dependent r. v.'s, see Theorem 18.2.2(a)
χ (uniform metric between characteristic functions)	(14.2.7)	Simple metric, ideal metric of order 0	Convergences, see p. 266	CLT, see Theorem 14.3.2
χ_r (weighted χ-metric)	(14.2.9)	Simple metric, ideal metric of order r	Smoothing inequalities, Lemmas 14.3.3, 14.3.4; bounds, see (14.3.13)	Rate of convergence in the CLT w.r.t. χ, see Theorem 14.3.2
$\chi_{p,r}$ (weighted Ky Fan metric)	(18.2.48)	Compound	Bounds: $\eta_{p,r} \leqslant \chi_{p,r} \leqslant p/(1+p)$; $\mathscr{L}_{p,r}$, see pp. 361, 362	Rate of convergence for maxima of dependent r.v.'s, see Theorem 18.2.3
$\boldsymbol{\Omega}$ (discrete metric in the space of moments of random variables)	(3.1.13)	Primary metric	A p. metric with simplest structure	Moment problems, see p. 195

Index of Symbols and Abbreviations

SYMBOL	DESCRIPTION	PAGE
$\Lambda_{\lambda,\nu}$	Λ-structure of a p.semidistance	68
$C^b(U)$	The set of bounded continuous functions of U	72, 84
$\zeta(\cdot,\cdot;\mathfrak{G}^p)$, $\zeta(\cdot,\cdot;\tilde{\mathfrak{G}}^p)$	Fortet–Mourier metrics	77, 78, 93
$\boldsymbol{\beta}$	The Dudley metric	79
$\boldsymbol{\kappa}_Q$	Q-difference pseudomoment	79
τ_Q	Compound Q-difference pseudomoment	79
$\mathbf{AS}_p$	The A. Szulga metric	81, 82
$\mathscr{A}_c$	The Kantorovich functional	90
$\mathscr{P}^{(P_1,P_2)}$	The space of all laws on $U \times U$ with marginals P_1 and P_2, or in other words, the space of all translocations of masses without transits permitted	90, 92
P^*	The optimal transference plan	92, 95
$c(x, y)$	The cost of transferring the mass from x to y	92
$\mathscr{Q}^{(P_1,P_2)}$	The space of all translocations of masses with transits permitted	92
$D_{n,\alpha}$	Ornstein type metric	97
$\tilde{P}$	A vector of probability measures $P_1, \ldots, P_N$	98
$\mathfrak{P}(\tilde{P})$	The space of laws on U^N with fixed one dimensional marginals	98
$A_c(\tilde{P})$	The multidimensional version of the Kantorovich functional $\mathscr{A}_c(P_1, P_2)$	98, 162
$\mathscr{H}^*$	All convex functions in $\mathscr{H}$	99, 106
$\mathscr{P} = \mathscr{P}_U = \mathscr{P}(U)$	The space of laws on U	99, 130, 155
$\mathscr{P}^H$	The space of laws on (U, d) with finite $H(d(\cdot, a))$-moment	99
$\mathscr{D}(x)$	$\\|(d(x_1, x_2), d(x_1, x_3), \ldots, d(x_{N-1}, x_N))\\|$	99, 155, 162
$D(x)$	$H(\mathscr{D}(x))$	99, 162
$\mathbb{K}(\tilde{P}, \mathfrak{U})$	The dual form of $A_c(\tilde{P})$	99, 100, 162
$\mathfrak{K}_H$	A multivariate analog of the Kantorovich distance ℓ_H	105
$\mathring{\mu}_c$	Kantorovich–Rubinstein functional (minimal norm w.r.t. μ_c)	107, 115
$m = m^+ + m^-$	The Jordan decomposition of a signed measure m	109
$\\|\cdot\\|_w$	Kantorovich–Rubinstein or Wasserstein norm	109
$m_1 \times m_2$	Product measure	109
$\mathscr{P}_\lambda = \mathscr{P}_\lambda(U)$	The space of laws with finite λ-moment	123
$\hat{\Lambda}$	The generalized Kantorovich functional	137

Symbol	Description	Page
$\mathring{\Lambda}$	The generalized Kantorovich–Rubinstein functional	137
$\overset{n}{\hat{\mu}}$	The K-minimal distance w.r.t. a p. distance μ on $\mathscr{P}_2(U^n)$	141, 146
ρ_α	A metric on the Cartesian product U^n	146
E	(0, 1)-distributed r.v.	147
$X_E = (X_1, \dots, X_n)_E$	Random vector with components $F_{X_i}(E)$, $i = 1, \dots, n$	147
$\tilde{\mu}$	For a given compound distance μ, $\tilde{\mu}(X, Y) := \mu(X_E, Y_E)$	147
$\mathbf{L}(X, Y; \alpha)$	The Lévy distance in the space of random vectors on $(\mathbb{R}^2, \rho_\alpha)$	148
$\mathbf{W}(X, Y; \alpha)$	The limit $\lambda L\left(\frac{1}{\lambda} X, \frac{1}{\lambda} Y; \alpha\right)$ as $\lambda \to \infty$	148
$\boldsymbol{\delta}$	The discrete metric in the space of d.f.'s on $\mathbb{R}^n$	148
$\mathscr{K}\mathscr{F}_\alpha$	The Ky Fan functional in $\mathfrak{X}(U^N)$	160
$\Pi_\alpha(\tilde{P})$	The Prokhorov functional in $\mathscr{P}(U)^N$	160
$[x]$	The integer part of x	163, 296
$\mathscr{F}(F_1, F_2)$	The set of bivariate d.f.'s with fixed marginals F_1 and F_2	172
$F_-(x_1, x_2)$	The Hoeffding–Frechet lower bound in $\mathscr{F}(F_1, F_2)$	172
$F_+(x_1, x_2)$	The Hoeffding–Frechet upper bound in $\mathscr{F}(F_1, F_2)$	172
$\mathbf{D}$	The metric between the pth moments	178
$\mathfrak{M}$	The space of bounded non-negative measures	201
$\eta_{\mathbb{K}}$	A generalization of ℓ_H on $\mathfrak{M}$	201
$\pi_{\lambda, G}$	G-weighted Prokhorov metric	204
$\mathfrak{N}$	The space of non-negative measures finite on any bounded set	208
$\xrightarrow{v}$	Vague convergence	209
$\mathbf{K}(\nu', \nu'')$	Kantorovich type metric in $\mathfrak{N}$	209
$\mathbf{\Pi}(\nu', \nu'')$	The Prokhorov metric in $\mathfrak{N}$	210
$\mathbb{C}[0, 1]$	The space of continuous functions on [0, 1]	218
W	The Wiener measure	219
$D[0, 1]$	The Skorokhod space	222
$G\|G\|1\|_\infty$	Single-server queue with general flows of interarrival times and service times, and infinitely large 'waiting room'.	223
$\mathbf{e} = (e_0, e_1, \dots)$	The 'input' flow of interarrival times	223
$\mathbf{s} = (s_0, s_1, \dots)$	The flow of service times	223

Index

Greek letters are alphabetized by their Roman equivalent (m for μ, and so on). Names in the bibliography are not indexed separately.