Svetlozar T. Rachev · Lev B. Klebanov
Stoyan V. Stoyanov · Frank J. Fabozzi

The Methods of Distances in the Theory of Probability and Statistics

Springer

The Methods of Distances in the Theory of Probability and Statistics

Svetlozar T. Rachev • Lev B. Klebanov
Stoyan V. Stoyanov • Frank J. Fabozzi

The Methods of Distances in the Theory of Probability and Statistics

Springer

Svetlozar T. Rachev
Department of Applied
Mathematics and Statistics
College of Business
Stony Brook University
Stony Brook, NY, USA

FinAnalytica
New York, NY, USA

Stoyan V. Stoyanov
EDHEC Business School
EDHEC-Risk Institute
Singapore

Lev B. Klebanov
Department of Probability and Statistics
Charles University
Prague, Czech Republic

Frank J. Fabozzi
EDHEC Business School
EDHEC-Risk Institute
Nice, France

ISBN 978-1-4614-4868-6 ISBN 978-1-4614-4869-3 (eBook)
DOI 10.1007/978-1-4614-4869-3
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012953265

Mathematics Subject Classification (2000): 60F05, 60F17, 60K25, 62G35, 62E20

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

STR
To my grandchildren Iliana, Zoya,
and Zari

LBK
To my wife Marina

SVS
To my wife Petya

FJF
To my wife Donna
and my children Francesco, Patricia,
and Karly

Preface

The development of the theory of probability metrics – a branch of probability theory – began with the study of problems related to limit theorems in probability theory. In general, the applicability of limit theorems stems from the fact that they can be viewed as an approximation to a given stochastic model and, consequently, can be accepted as an approximate substitute. The key question that arises in adopting the approximate model is the magnitude of the error that must be accepted. Because the theory of probability metrics studies the problem of measuring distances between random quantities or stochastic processes, it can be used to address the key question of how good the approximate substitute is for the stochastic model under consideration. Moreover, it provides the fundamental principles for building probability metrics – the means of measuring such distances.

The theory of probability metrics has been applied and has become an important tool for studying a wide range of fields outside of probability theory such as statistics, queueing theory, engineering, physics, chemistry, information theory, economics, and finance. The principal reason is that because distances are not influenced by the particular stochastic model under consideration, the theory of probability metrics provides some universal principles that can be used to deal with certain kinds of large-scale stochastic models found in these fields.

The first driving force behind the development of the theory of probability metrics was Andrei N. Kolmogorov and his group. It was Kolmogorov who stated that every approximation problem has its own distance measure in which the problem can be solved in a most natural way. Kolmogorov also contended that without estimates of the rate of convergence in the central limit theorem (CLT) (and similar limit theorems such as the functional limit theorem and the max-stable limit theorem), limit theorems provide very limited information. An example worked out by Y.V. Prokhorov and his students is as follows. Regardless of how slowly a function $f(n) > 0$, $n = 1, \ldots$, decays to zero, there exists a corresponding distribution function $F(x)$ with finite variance and mean zero, for which the CLT is valid at a rate slower than $f(n)$. In other words, without estimates for convergence in the CLT, such a theorem is meaningless because the convergence to the normal law of the normalized sum of independent, identically distributed random variables with distribution function $F(x)$ can be slower than any given rate $f(n) \to 0$. The problems associated with finding the appropriate rate of convergence invoked a variety of probability distances in which the speed of convergence (i.e., convergence rate) was estimated. This included the works of Yurii V. Prokhorov, Vladimir V. Sazonov, Vladimir M. Zolotarev, Vygantas Paulauskas, Vladimir V. Senatov, and others.

The second driving force in the development of the theory of probability metrics was mass-transportation problems and duality theorems. This started with the work of Gaspard Monge in the eighteenth century and Leonid V. Kantorovich in the 1940s – for which he was awarded the Nobel Prize in Economics in 1975 – on optimal mass transportation, leading to the birth of linear programming. In mathematical terms, Kantorovich's result on mass transportation can be formulated in the following metric way. Given the marginal distributions of two probability measures P and Q on a general (separable) metric space (U, d), what is the minimal expected value – referred to as $\boldsymbol{\kappa}(P, Q)$ or the Kantorovich metric – of a distance $d(X, Y)$ over the set of all probability measures on the product space $U \times U$ with marginal distributions $P_X = P$ and $P_Y = Q$? If the measures P and Q are discrete, then this is the classic transportation problem in linear programming. If U is the real line, then $\boldsymbol{\kappa}(P, Q)$ is known as the Gini statistical index of dissimilarity formulated by Corrado Gini. The Kantorovich problem has been used in many fields of science – most notably statistical physics, information theory, statistics, and probability theory. The fundamental work in this field was done by Leonid V. Kantorovich, Johannes H. B. Kemperman, Hans G. Kellerer, Richard M. Dudley, Ludger Rüschendorf, Volker Strassen, Vladimir L. Levin, and others. Kantorovich-type duality theorems established the main relationships between metrics in the space of random variables and metrics in the space of probability laws/distributions. The unifying work on those two directions was done by V. M. Zolotarev and his students.

In this book, we concentrate on four specialized research directions in the theory of probability metrics, as well as applications to different problems of probability theory. These include:

- Description of the basic structure of probability metrics,
- Analysis of the topologies in the space of probability measures generated by different types of probability metrics,
- Characterization of the ideal metrics for a given problem, and
- Investigation of the main relationships between different types of probability metrics.

Our presentation in this book is provided 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 target audience for this book is graduate students in the areas of functional analysis, geometry, mathematical programming, probability, statistics, stochastic analysis, and measure theory. It may be partially used as a source of material for lectures for students in probability and statistics. As noted earlier in this preface, the theory of probability metrics has been applied to fields outside of probability theory such as engineering, physics, chemistry, information theory, economics, and finance. Specialists in these areas will find the book to be a useful reference to gain a greater understanding of this specialized area and its potential application.

New York, USA — Svetlozar T. Rachev
Prague, Czech Republic — Lev B. Klebanov
Singapore — Stoyan V. Stoyanov
Nice, France — Frank J. Fabozzi

Contents

1 Main Directions in the Theory of Probability Metrics 1
- 1.1 Introduction 1
- 1.2 Method of Metric Distances and Theory of Probability Metrics ... 1
- 1.3 Probability Metrics Defined 2
- 1.4 Main Directions in the Theory of Probability Metrics 3
- 1.5 Overview of the Book 4

Part I General Topics in the Theory of Probability Metrics

2 Probability Distances and Probability Metrics: Definitions 11
- 2.1 Introduction 12
- 2.2 Examples of Metrics in Probability Theory 13
- 2.3 Kolmogorov Metric: A Property and an Application 15
- 2.4 Metric and Semimetric Spaces, Distance, and Semidistance Spaces 19
- 2.5 Definitions of Probability Distance and Probability Metric 22
- 2.6 Universally Measurable Separable Metric Spaces 24
- 2.7 Equivalence of the Notions of Probability (Semi-) distance on $\mathcal{P}_2$ and on $\mathfrak{X}$ 28
- References 31

3 Primary, Simple, and Compound Probability Distances and Minimal and Maximal Distances and Norms 33
- 3.1 Introduction 34
- 3.2 Primary Distances and Primary Metrics 34
- 3.3 Simple Distances and Metrics: Cominimal Functionals and Minimal Norms 39
- 3.4 Compound Distances and Moment Functions 53
- References 66

4 A Structural Classification of Probability Distances 67
- 4.1 Introduction 68

4.2 Hausdorff Structure of Probability Semidistances 69
4.3 Λ-Structure of Probability Semidistances 85
4.4 ζ-Structure of Probability Semidistances 88
References 103

Part II Relations Between Compound, Simple, and Primary Distances

5 Monge–Kantorovich Mass Transference Problem, Minimal Distances and Minimal Norms 109
5.1 Introduction 109
5.2 Statement of Monge–Kantorovich Problem 111
5.3 Multidimensional Kantorovich Theorem 121
5.4 Dual Representation of Minimal Norms $\mathring{\mu}_c$: Generalization of Kantorovich–Rubinstein Theorem 130
5.5 Application: Explicit Representations for a Class of Minimal Norms 138
References 141

6 Quantitative Relationships Between Minimal Distances and Minimal Norms 145
6.1 Introduction 145
6.2 Equivalence Between Kantorovich Metric and Kantorovich–Rubinstein Norm 146
6.3 Inequalities Between $\widehat{\mu}_c$ and $\mathring{\mu}_c$ 148
6.4 Convergence, Compactness, and Completeness in $(\mathcal{P}(U), \widehat{\mu}_c)$ and $(\mathcal{P}(U), \mathring{\mu}_c)$ 157
6.5 $\mathring{\mu}_c$- and $\widehat{\mu}_c$-Uniformity 161
6.6 Generalized Kantorovich and Kantorovich–Rubinstein Functionals 164
References 167

7 *K*-Minimal Metrics 169
7.1 Introduction 170
7.2 Definition and General Properties 170
7.3 Two Examples of K-Minimal Metrics 175
7.4 K-Minimal Metrics of Given Probability Metrics: The Case of $U = R$ 177
7.5 The Case Where U Is a Separable Metric Space 185
7.6 Relations Between Multidimensional Kantorovich and Strassen Theorems: Convergence of Minimal Metrics and Minimal Distances 192
References 197

8 Relations Between Minimal and Maximal Distances ... 199
8.1 Introduction ... 199
8.2 Duality Theorems and Explicit Representations for $\check{\mu}_c$ and $\overset{(s)}{\mu}_c$ ... 200
8.3 Convergence of Measures with Respect to Minimal Distances and Minimal Norms ... 207
Reference ... 217

9 Moment Problems Related to the Theory of Probability Metrics: Relations Between Compound and Primary Distances ... 219
9.1 Introduction ... 219
9.2 Primary Minimal Distances: Moment Problems with One Fixed Pair of Marginal Moments ... 220
9.3 Moment Problems with Two Fixed Pairs of Marginal Moments and with Fixed Linear Combination of Moments ... 226
References ... 233

Part III Applications of Minimal Probability Distances

10 Moment Distances ... 237
10.1 Introduction ... 237
10.2 Criteria for Convergence to a Distribution with an Analytic Characteristic Function ... 238
10.3 Moment Metrics ... 255
10.3.1 Estimates of λ by Means of d_∞ ... 257
10.3.2 Estimates of λ by Means of d_α, $\alpha \in (0, \infty)$ ... 266
10.3.3 Estimates of d_α by Means of Characteristic Functions ... 267
References ... 270

11 Uniformity in Weak and Vague Convergence ... 271
11.1 Introduction ... 271
11.2 ζ-Metrics and Uniformity Classes ... 272
11.3 Metrization of the Vague Convergence ... 279
References ... 282

12 Glivenko–Cantelli Theorem and Bernstein–Kantorovich Invariance Principle ... 283
12.1 Introduction ... 283
12.2 Fortet–Mourier, Varadarajan, and Wellner Theorems ... 284
12.3 Functional Central Limit and Bernstein–Kantorovich Invariance Principle ... 291
References ... 296

13 Stability of Queueing Systems ... 297
13.1 Introduction ... 297
13.2 Stability of $G|G|1|\infty$-Systems ... 298

13.3 Stability of $GI|GI|1|\infty$-System 301
13.4 Approximation of a Random Queue by Means of Deterministic Queueing Models 305
References 315

14 Optimal Quality Usage 317
14.1 Introduction 317
14.2 Optimality of Quality Usage and the Monge–Kantorovich Problem 318
14.3 Estimates of Minimal Total Losses $\tau_\phi(\phi^*)$ 325
References 331

Part IV Ideal Metrics

15 Ideal Metrics with Respect to Summation Scheme for i.i.d. Random Variables 335
15.1 Introduction 336
15.2 Robustness of χ^2 Test of Exponentiality 336
15.3 Ideal Metrics for Sums of Independent Random Variables 343
15.4 Rates of Convergence in the CLT in Terms of Metrics with Uniform Structure 353
References 361

16 Ideal Metrics and Rate of Convergence in the CLT for Random Motions 363
16.1 Introduction 363
16.2 Ideal Metrics in the Space of Random Motions 364
16.3 Rates of Convergence in the Integral and Local CLTs for Random Motions 368
References 378

17 Applications of Ideal Metrics for Sums of i.i.d. Random Variables to the Problems of Stability and Approximation in Risk Theory 379
17.1 Introduction 379
17.2 Problem of Stability in Risk Theory 379
17.3 Problem of Continuity 381
17.4 Stability of Input Characteristics 388
References 393

18 How Close Are the Individual and Collective Models in Risk Theory? 395
18.1 Introduction 395
18.2 Stop-Loss Distances as Measures of Closeness Between Individual and Collective Models 396
18.3 Approximation by Compound Poisson Distributions 410
References 419

19 Ideal Metric with Respect to Maxima Scheme of i.i.d. Random Elements ... 421
19.1 Introduction ... 422
19.2 Rate of Convergence of Maxima of Random Vectors Via Ideal Metrics ... 422
19.3 Ideal Metrics for the Problem of Rate of Convergence to Max-Stable Processes ... 438
19.4 Double Ideal Metrics ... 463
References ... 477

20 Ideal Metrics and Stability of Characterizations of Probability Distributions ... 479
20.1 Introduction ... 479
20.2 Characterization of an Exponential Class of Distributions $\{F_p,\ 0 < p \leq \infty\}$ and Its Stability ... 482
20.3 Stability in de Finetti's Theorem ... 494
20.4 Characterization and Stability of Environmental Processes ... 500
References ... 515

Part V Euclidean-Like Distances and Their Applications

21 Positive and Negative Definite Kernels and Their Properties ... 519
21.1 Introduction ... 519
21.2 Definitions of Positive Definite Kernels ... 520
21.3 Examples of Positive Definite Kernels ... 524
21.4 Positive Definite Functions ... 527
21.5 Negative Definite Kernels ... 528
21.6 Coarse Embeddings of Metric Spaces into Hilbert Space ... 532
21.7 Strictly and Strongly Positive and Negative Definite Kernels ... 533
References ... 537

22 Negative Definite Kernels and Metrics: Recovering Measures from Potentials ... 539
22.1 Introduction ... 540
22.2 $\mathfrak{N}$-Metrics in the Set of Probability Measures ... 540
22.2.1 A Class of Positive Definite Kernels in the Set of Probabilities and $\mathfrak{N}$-Distances ... 540
22.3 m-Negative Definite Kernels and Metrics ... 544
22.3.1 m-Negative Definite Kernels and Metrics ... 544
22.4 $\mathfrak{N}$-Metrics and the Problem of Recovering Measures from Potentials ... 548
22.4.1 Recovering Measures from Potentials ... 549
22.4.2 Stability in the Problem of Recovering a Measure from its Potential ... 554
22.5 $\mathfrak{N}$-Metrics in the Study of Certain Problems of the Characterization of Distributions ... 559

22.5.1 Characterization of Gaussian and Related Distributions .. 559
22.5.2 Characterization of Distributions Symmetric to a Group of Transformations ... 567
References ... 569

23 Statistical Estimates Obtained by the Minimal Distances Method ... 571
23.1 Introduction ... 571
23.2 Estimating a Location Parameter: First Approach ... 571
23.3 Estimating a Location Parameter: Second Approach ... 575
23.4 Estimating a General Parameter ... 576
23.5 Estimating a Location Parameter: Third Approach ... 577
23.6 Semiparametric Estimation ... 578
References ... 579

24 Some Statistical Tests Based on $\mathfrak{N}$-Distances ... 581
24.1 Introduction ... 581
24.2 A Multivariate Two-Sample Test ... 581
24.3 Testing If Two Distributions Belong to the Same Additive Type ... 585
24.4 A Test for Multivariate Normality ... 587
References ... 588

25 Distances Defined by Zonoids ... 589
25.1 Introduction ... 589
25.2 Main Notions and Definitions ... 590
25.3 $\mathfrak{N}$-Distances ... 591
References ... 597

26 $\mathfrak{N}$-Distance Tests of Uniformity on the Hypersphere ... 599
26.1 Introduction ... 599
26.2 Tests of Uniformity on a Hypersphere ... 600
26.3 Asymptotic Distribution ... 602
26.3.1 Uniformity on a Circle ... 602
26.3.2 Uniformity on a Sphere ... 603
26.3.3 Alternative Approach to the Limit Distribution of T_n ... 604
26.4 Proofs ... 607
26.4.1 Proof of Proposition 26.2.1 ... 607
26.4.2 Proof of Proposition 26.3.1 ... 608
26.4.3 Proof of Theorem 26.3.4 ... 608
26.4.4 Proof of Theorem 26.3.5 ... 609
References ... 610

About the Authors ... 611

Index ... 613

Chapter 1
Main Directions in the Theory of Probability Metrics

1.1 Introduction

Increasingly, the demands of various real-world applications in the sciences, engineering, and business have resulted in the creation of new, more complicated probability models. In the construction and evaluation of these models, model builders have drawn on well-developed limit theorems in probability theory and the theory of random processes. The study of limit theorems in general spaces and a number of other questions in probability theory make it necessary to introduce functionals – defined on either classes of probability distributions or classes of random elements – and to evaluate their nearness in one or another probabilistic sense. Thus various metrics have appeared including the well-known Kolmogorov (uniform) metric, L_p metrics, the Prokhorov metric, and the metric of convergence in probability (Ky Fan metric). We discuss these measures and others in the chapters that follow.

1.2 Method of Metric Distances and Theory of Probability Metrics

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 a proposed stochastic model is required. Analysis of quantitative stability assumes the use of metrics as measures of distances or deviations. The main idea of the *method of metric distances* (MMD) – developed by Vladmir M. Zolotarev and his students to solve stability problems – is reduced to the following two problems.

S.T. Rachev et al., *The Methods of Distances in the Theory of Probability and Statistics*,
DOI 10.1007/978-1-4614-4869-3_1,

Problem 1.2.1 (Choice of ideal metrics). Find the most appropriate (i.e., ideal) metrics for the stability problem under consideration and then solve the problem in terms of these ideal metrics.

Problem 1.2.2 (Comparisons of metrics). If the solution of the stability problem must be written in terms of other metrics, then solve the problem of comparing these metrics with the chosen (i.e., ideal) metrics.

Unlike Problem 1.2.1, Problem 1.2.2 does not depend on the specific stochastic model under consideration. Thus, the independent solution of Problem 1.2.2 allows its application in any particular situation. Moreover, by addressing the two foregoing problems, a clear understanding of the specific regularities that form the stability effect emerges.

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. It is often convenient to express such stability in terms of a metric. The *theory of probability metrics* (TPM) was developed to address this. That is, TPM was developed to address Problems 1.2.1 and 1.2.2, thus providing a framework for the MMD. Figure 1.1 summarizes the problems concerning MMD and TPM.

1.3 Probability Metrics Defined

The term *probability metric*, or *p. 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. Each such metric can be viewed as a function defined on the set of probability measures on the product of two copies of a probability space. Complications connected with the question of the existence of pairs of random variables on a given space with given probability laws can be easily avoided. Fixing the marginal distributions of the probability measure on the product space, one can find the infimum of the values of our function on the class of all measures with the given marginals. Under some regularity conditions, such an infimum is a metric on the class of probability distributions, and in some concrete cases (e.g., 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].

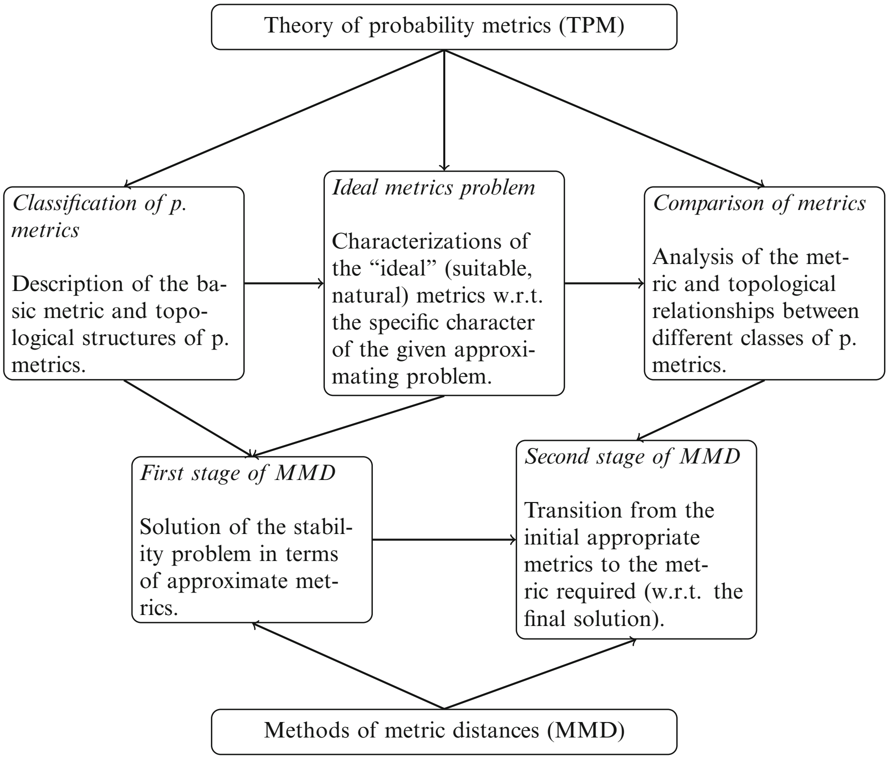

Fig. 1.1 Theory of probability metrics as a necessary tool to investigate the method of metric distances

1.4 Main Directions in the Theory of Probability Metrics

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

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

Direction 2. Analysis of topologies in space of probability measures generated by different types of p. metrics; such an 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 ideal metrics for a given problem.

Direction 4. Investigations of main relationships between different types of p. metrics.

In this book, all four directions are covered as well as applications to different problems in probability theory. Much attention is paid to the possibility of giving equivalent definitions of p. metrics (e.g., in direct and dual terms and in terms of the Hausdorff metric for sets). 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 in this book is the class of minimal metrics, an idea that goes back to the work of Kantorovich in the 1940s dealing with transportation problems in linear programming. Such metrics have been found independently by many authors in several fields of probability theory (e.g., Markov processes, statistical physics).

Another useful class of metrics studied in this book is the class of *ideal* metrics that satisfy the following properties:

1. $\mu(P_c, Q_c) \le |c|^r \mu(P, Q)$ for all $c \in [-C, C], c \ne 0$,
2. $\mu(P_1 * Q, P_2 * Q) \le \mu(P_1, P_2)$,

where $P_c(A) := P((1/c)A)$ for any Borel set A on a Banach space U and where $*$ denotes 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 we provide in this book 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.

1.5 Overview of the Book

The book is divided into five parts. In Part I, we set forth general topics in the TPM. Following the definition of a probability metric in Chap. 2, different examples of probability metrics are provided and the application of the Kolmogorov metric in mathematical statistics is discussed. Then the axiomatic construction of probability metrics is defined. There is also a discussion of an interesting property about the Kolmogorov metric, a property that is used to prove biasedness in the classic Kolmogorov test. More definitions and examples are provided in Chap. 3, where primary, simple, and compound distances and minimal and maximal distances and norms are provided and motivated. The introduction and motivation of three classifications of probability metrics according to their metric structure, as well as examples of probability metrics belonging to a particular structural group, are explained in Chap. 4. The generic properties of the structural groups and the links between them are also covered in the chapter.

In Part II, we concern ourselves with the study of the dual and explicit representations of minimal distances and norms, as well as the topologies that these metric structures induce in the space of probability measures. We do so by examining further the concepts of primary, simple, and compound distances, in particular focusing on their relationship to each other. The Kantorovich and the Kantorovich–

Rubinstein problems are introduced and illustrated in a one-dimensional and multidimensional setting in Chap. 5. These problems – more commonly referred to as the mass transportation and mass transshipment problems, respectively – are abstract formulations of optimization problems. Although the applications are important in areas such as job assignments, classification problems, and best allocation policy, our purpose for covering them in this book is due to their link to the TPM. In particular, an application leading to an explicit representation for a class of minimal norms is provided. Continuing with our coverage of Kantorovich and the Kantorovich–Rubinstein functionals in Chap. 6, we look at the conditions under which there is equality and inequalities between these two functionals. Because these two functionals generate minimal distances (Kantorovich functional) and minimal norms (Kantorovich–Rubinstein functional), the relationship between the two can be quantified, allowing us to provide criteria for convergence, compactness, and completeness of probability measures in probability spaces, as well as to analyze the problem of uniformity between these two functionals. The discussions in Chaps. 5 and 6 demonstrate that the notion of minimal distance represents the main relationship between compound and simple distances. Our focus in Chap. 7 is on the notion of K-minimal metrics, and we describe their general properties and provide representations with respect to several particular metrics such as the Lévy metric and the Kolmogorov metric. The relationship between the multidimensional Kantorovich theorem and the work by Strassen on minimal probabilistic functionals is also covered. In Chap. 8, we discuss the relationship between minimal and maximal distances, comparing them to the corresponding dual representations of the minimal metric and minimal norm, providing closed-form solutions for some special cases and studying the topographical structures of minimal distances and minimal norms. The general relations between compound and primary probability distances, which are similar to the relations between compound and simple probability distances, are the subject of Chap. 9.

The application of minimal probability distances is the subject of the five chapters in Part III. Chapter 10 contains definitions, properties, and some applications of moment distances. These distances are connected to the property of definiteness of the classic problem of moments, and one of them satisfies an inequality that is stronger than the triangle inequality. In Chap. 11, we begin with a discussion of the convergence criteria in terms of a simple metric between characteristic functions, assuming they are analytic. We then turn to providing estimates of a simple metric between characteristic functions of two distributions in terms of moment-based primary metrics and discussing the inverse problem of estimating moment-based primary metrics in terms of a simple metric between characteristic functions. In Chaps. 11 through 14, we then use our understanding of minimal distances explained in Chap. 7 to demonstrate how the minimal structure is especially useful in problems of approximations and stability of stochastic models. We explain how to apply the topological structure of the space of laws generated by minimal distance and minimal norm functionals in limit-type theorems, which provide weak convergence together with convergence of moments. We study vague convergence in Chap. 11, the Glivenko–Cantelli theorem in Chap. 12, queueing systems in Chap. 13, and optimal quality in Chap. 14.

Any concrete stochastic approximation problem requires an *appropriate* or *natural* metric (e.g., topology, convergence, uniformities) having properties that are helpful in solving the problem. If one needs to develop the solution to the approximation problem in terms of other metrics (e.g., topology), then the transition is carried out using general relationships between metrics (e.g., topologies). This two-stage approach, described in Sect. 1.2 (selection of the appropriate metric, which we labeled Problem 1.2.1, and comparison of metrics, labeled Problem 1.2.2) is the basis of the TPM. In Part IV – Chaps. 15 through 20 – we determine the structure of *appropriate* or, as we label it in this book, *ideal* probability 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 are dealing with; that is, any particular approximation problem has its own "ideal" probability distance (or distances) on which terms we can solve the problem in the most "natural" way. In the opening chapter to this part of the book, Chap. 15, we describe the notion of ideal probability metrics for summation of independent and identically distributed random variables and provide examples of ideal probability metrics. We then study the structure of such "ideal" metrics in various stochastic approximation problems such as the convergence of random motions in Chap. 16, the stability of characterizations of probability distributions in Chaps. 17 and 20, stability in risk theory in Chap. 18, and the rate of convergence for the sums and maxima of random variables in Chap. 19.

Part V is devoted to a class of distances – Euclidean-type distances. In this part of the book, we provide definitions, properties, and applications of such distances. The space of measures for these distances is isometric to a subset of a Hilbert space. We give a description of all such metrics. Some of the distances appear to be ideal with respect to additive operations on random vectors. Subclasses of the distances are very useful to obtain a characterization of distributions and especially to recover a distribution from its potential. All Euclidean-type distances are very useful for constructing nonparametric, two-sample multidimensional tests. As background material for the discussion in this part of the book, in Chap. 21 we introduce the mathematical concepts of positive and negative definite kernels, describe their properties, and provide theoretical results that characterize coarse embeddings in a Hilbert space. Because kernel functions are central to the notion of potential of probability measures, in Chap. 22 we introduce special classes of probability metrics through negative definite kernel functions and show how, for strongly negative definite kernels, a probability measure can be uniquely determined by its potential. Moreover, the distance between probability measures can be bounded by the distance between their potentials; that is, under some technical conditions, a sequence of probability measures converges to a limit if and only if the sequence of their potentials converges to the potential of the limiting probability measure. Also as explained in Chap. 22, the problem of characterizing classes of probability distributions can be reduced to the problem of recovering a measure from potential. The problem of parameter estimation by the method of minimal distances and the study of the properties of these estimators are the subject of Chap. 23. In Chap. 24, we construct multidimensional statistical tests based on the theory of distances generated by negative definite kernels in the set of probability measures described in Chap. 23. The connection between distances generated by negative definite kernels and zonoids is the subject of Chap. 25. In Chap. 26, we discuss multidimensional statistical tests of uniformity based on the theory of distances generated by negative definite kernels and calculate the asymptotic distribution of these test statistics.

Part I
General Topics in the Theory of Probability Metrics

Chapter 2
Probability Distances and Probability Metrics: Definitions

The goals of this chapter are to:

- Provide examples of metrics in probability theory;
- Introduce formally the notions of a probability metric and a probability distance;
- Consider the general setting of random variables (RVs) defined on a given probability space $(\Omega, \mathcal{A}, \Pr)$ that can take values in a separable metric space U in order to allow for a unified treatment of problems involving random elements of a general nature;
- Consider the alternative setting of probability distances on the space of probability measures $\mathcal{P}_2$ defined on the σ-algebras of Borel subsets of $U^2 = U \times U$, where U is a separable metric space;
- Examine the equivalence of the notion of a probability distance on the space of probability measures $\mathcal{P}_2$ and on the space of joint distributions $\mathcal{L}\mathfrak{X}_2$ generated by pairs of RVs (X, Y) taking values in a separable metric space U.

S.T. Rachev et al., *The Methods of Distances in the Theory of Probability and Statistics*,
DOI 10.1007/978-1-4614-4869-3_2,

Notation introduced in this chapter:

Notation	Description
EN	Engineer's metric
$\mathfrak{X}^p$	Space of real-valued random variables with $E\|X\|^p < \infty$
$\boldsymbol{\rho}$	Uniform (Kolmogorov) metric
$\mathfrak{X} = \mathfrak{X}(\mathbb{R})$	Space of real-valued random variables
L	Lévy metric
$\boldsymbol{\kappa}$	Kantorovich metric
$\boldsymbol{\theta}_p$	L_p-metric between distribution functions
K, **K***	Ky Fan metrics
$\mathcal{L}_p$	L_p-metric between random variables
$\mathbf{MOM}_p$	Metric between pth moments
(S, ρ)	Metric space with metric ρ
$\mathbb{R}^n$	n-dimensional vector space
$r(C_1, C_2)$	Hausdorff metric (semimetric between sets)
$s(F, G)$	Skorokhod metric
$\mathbb{K} = \mathbb{K}_\rho$	Parameter of a distance space
$\mathcal{H}$	Class of Orlicz's functions
ρ_H	Birnbaum–Orlicz distance
Kr	Kruglov distance
(U, d)	Separable metric space with metric d
s.m.s.	Separable metric space
U^k	k-fold Cartesian product of U
$\mathcal{B}_k = \mathcal{B}_k(U)$	Borel σ-algebra on U^k
$\mathcal{P}_k = \mathcal{P}_k(U)$	Space of probability laws on $\mathcal{B}_k$
$T_{\alpha,\beta,\ldots,\gamma} P$	Marginal of $P \in \mathcal{P}_k$ on coordinates $\alpha, \beta, \ldots, \gamma$
$\Pr_X$	Distribution of X
μ	Probability semidistance
$\mathfrak{X} := \mathfrak{X}(U)$	Set of U-valued RVs
$\mathcal{L}\mathfrak{X}_2 := \mathcal{L}\mathfrak{X}_2(U)$	Space of $\Pr_{X,Y}$, $X, Y \in \mathfrak{X}(U)$
u.m.	Universally measurable
u.m.s.m.s.	Universally measurable separable metric space

2.1 Introduction

Generally speaking, a functional that measures the distance between random quantities is called a *probability metric*.[1] In this chapter, we provide different examples of probability metrics and discuss an application of the Kolmogorov

[1] Mostafaei and Kordnourie (2011) is a more recent general publication on probability metrics and their applications.

metric in mathematical statistics. Then we proceed with the axiomatic construction of probability metrics on the space of probability measures defined on the twofold Cartesian product of a separable metric space U. This definition induces by restriction a probability metric on the space of joint distributions of random elements defined on a probability space $(\Omega, \mathcal{A}, \Pr)$ and taking values in the space U. Finally, we demonstrate that under some fairly general conditions, the two constructions are essentially the same.

2.2 Examples of Metrics in Probability Theory

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

1. *Engineer's metric*:

$$\mathbf{EN}(X,Y) := |\mathbb{E}(X) - \mathbb{E}(Y)|, \quad X, Y \in \mathfrak{X}^1, \tag{2.2.1}$$

   where $\mathfrak{X}^p$ is the space of all real-valued RVs) with $\mathbb{E}|X|^p < \infty$.
2. *Uniform (or Kolmogorov) metric*:

$$\boldsymbol{\rho}(X,Y) := \sup\{|F_X(x) - F_Y(x)| : x \in \mathbb{R}\}, \quad X, Y \in \mathfrak{X} = \mathfrak{X}(\mathbb{R}), \tag{2.2.2}$$

   where F_X is the distribution function (DF) of X, $\mathbb{R} = (-\infty, +\infty)$, and $\mathfrak{X}$ is the space of all real-valued RVs.
3. *Lévy metric*:

$$\mathbf{L}(X,Y) := \inf\{\varepsilon > 0 : F_X(x-\varepsilon) - \varepsilon \le F_Y(x) \le F_X(x+\varepsilon) + \varepsilon, \quad \forall x \in \mathbb{R}\}. \tag{2.2.3}$$

Remark 2.2.1. We see that $\boldsymbol{\rho}$ and $\mathbf{L}$ may actually be considered 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 $\boldsymbol{\rho}(X,Y) = 0 \iff \mathbf{L}(X,Y) = 0 \iff F_X = F_Y$. The Lévy metric metrizes weak convergence (convergence in distribution) in the space $\mathcal{F}$, whereas $\boldsymbol{\rho}$ is often applied in the central limit theorem (CLT).[2]

4. *Kantorovich metric*:

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

[2] See Hennequin and Tortrat (1965).

5. *L_p-metrics between distribution functions*:

$$\boldsymbol{\theta}_p(X,Y) := \left(\int_{-\infty}^{\infty} |F_X(t) - F_Y(t)|^p \mathrm{d}t\right)^{1/p}, \quad p \geq 1, \quad X, Y \in \mathfrak{X}^1. \quad (2.2.4)$$

Remark 2.2.2. 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 $1 \leq p \leq \infty$:

$$\theta_p(X,Y) = \sup_{f \in \mathcal{F}_p} |Ef(X) - Ef(Y)|, \quad X, Y \in \mathfrak{X}^1,$$

where $\mathcal{F}_p$ 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[3]

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

6. *Ky Fan metrics*:

$$\mathbf{K}(X,Y) := \inf\{\varepsilon > 0 : \Pr(|X - Y| > \varepsilon) < \varepsilon\}, \qquad X, Y \in \mathfrak{X}, \qquad (2.2.5)$$

and

$$\mathbf{K}^*(X,Y) := E\frac{|X - Y|}{1 + |X - Y|}. \qquad (2.2.6)$$

Both metrics metrize convergence in probability on $\mathfrak{X} = \mathfrak{X}(\mathbb{R})$, the space of real RVs.[4]

7. *L_p-metric*:

$$\mathcal{L}_p(X,Y) := \{E|X - Y|^p\}^{1/p}, \quad p \geq 1, \quad X, Y \in \mathfrak{X}^p. \qquad (2.2.7)$$

Remark 2.2.3. Define

$$m^p(X) := \{E|X|^p\}^{1/p}, \quad p > 1, \quad X \in \mathfrak{X}^p. \qquad (2.2.8)$$

and

$$\mathbf{MOM}_p(X,Y) := |m^p(X) - m^p(Y)|, \quad p \geq 1, \quad X, Y \in \mathfrak{X}^p. \qquad (2.2.9)$$

[3]The proof of this representation is given by (Dudley, 2002, p. 333) for the case $p = 1$.

[4]See Lukacs (1968, Chap. 3) and Dudley (1976, Theorem 3.5).

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

$$\mathcal{L}_p(X_n, X_0) \to 0 \iff \begin{cases} \mathbf{K}(X_n, X_0) \to 0, \\ \mathbf{MOM}_p(X_n, X_0) \to 0 \end{cases} \tag{2.2.10}$$

[see, e.g., Lukacs (1968, Chap. 3)].

Other probability metrics in common use include the discrepancy metric, the Hellinger distance, the relative entropy metric, the separation distance metric, the χ^2-distance, and the f-divergence metric. These probability metrics are summarized in Gibbs and Su (2002).

All of the aforementioned (semi-)metrics on subsets of $\mathfrak{X}$ 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** (2.2.1) and $\mathbf{MOM}_p$ (2.2.9). For these metrics

$$\mathbf{EN}(X, Y) = 0 \iff EX = EY,$$
$$\mathbf{MOM}_p(X, Y) = 0 \iff m^p(X) = m^p(Y). \tag{2.2.11}$$

A metric μ is *simple* if

$$\mu(X, Y) = 0 \iff F_X = F_Y. \tag{2.2.12}$$

Examples are $\boldsymbol{\rho}$ (2.2.2), **L** (2.2.3), and $\boldsymbol{\theta}_p$ (2.2.4). The third group, the *compound* (semi-)metrics, has the property

$$\mu(X, Y) = 0 \iff \Pr(X = Y) = 1. \tag{2.2.13}$$

Some examples are **K** (2.2.5), $\mathbf{K}^*$ (2.2.6), and $\mathcal{L}_p$ (2.2.7).

Later on, precise definitions of these classes will be given as well as a study of the relationships between them. Now we will begin with a common definition of probability metric that will include the types mentioned previously.

2.3 Kolmogorov Metric: A Property and an Application

In this section, we consider a paradoxical property of the Kolmogorov metric and an application in the area of mathematical statistics.

Consider the metric space $\mathfrak{F}$ of all one-dimensional distributions metrized by the Kolmogorov distance

$$\boldsymbol{\rho}(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|, \tag{2.3.1}$$

which we define now in terms of the elements of $\mathfrak{F}$ rather than in terms of RVs as in the definition in (2.2.2). Denote by $B(F, r)$ an open ball of radius $r > 0$ centered