Preface

Statistical manifolds are geometric abstractions used to model information, their field of study belonging to Information Geometry, a relatively recent branch of mathematics, that uses tools of differential geometry to study statistical inference, information loss, and estimation.

This field started with the differential geometric study of the manifold of probability density functions. For instance, the set of normal distributions

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R},$$

with $(\mu, \sigma) \in \mathbb{R} \times (0, +\infty)$, can be considered as a two-dimensional surface. This can be endowed with a Riemannian metric, which measures the amount of information between the distributions. One of these possibilities is to consider the Fisher information metric. In this case, the distribution family $p(x; \mu, \sigma)$ becomes a space of constant negative curvature. Therefore, any normal distribution can be visualized as a point in the Poincaré upper-half plane.

In a similar way, we shall consider other parametric model families of probability densities that can be organized as a differentiable manifold embedded in the ambient space of all density functions. Every point on this manifold is a density function, and any curve corresponds to a one-parameter subfamily of density functions. The distance between two points (i.e., distributions), which is measured by the Fisher metric, was introduced almost simultaneously by C. R. Rao and H. Jeffreys in the mid-1940s. The role of differential geometry in statistics was first emphasized by Efron in 1975, when he introduced the concept of statistical curvature. Later Amari used the tools of differential geometry to develop this idea into an elegant representation of Fisher's theory of information loss.

A fundamental role in characterizing statistical manifolds is played by two geometric quantities, called dual connections, which describe the derivation with respect to vector fields and are interrelated in a duality relation involving the Fisher metric. The use of dual connections leads to several other dual elements, such as volume elements, Hessians, Laplacians, second fundamental forms, mean curvature vector fields, and Riemannian curvatures. The study of dual elements and the relations between them constitute the main direction of development in the study of statistical manifolds.

Even if sometimes we use computations in local coordinates, the relationships between these geometric quantities are invariant with respect to the selection of any particular coordinate system. Therefore, the study of statistical manifolds provides techniques to investigate the intrinsical properties of statistical models rather than their parametric representations. This invariance feature made statistical manifolds useful in the study of information geometry.

We shall discuss briefly the relation of this book with other previously published books on the same or closely related topic.

One of the well-known textbooks in the field is written by two of the information geometry founders, Amari and Nagaoka [8], which was published first time in Japan in 1993, and then translated into English in 2000. This book presents a concise introduction to the mathematical foundation of information geometry and contains an overview of other related areas of interest and applications. Our book intersects with Amari's book over its first three chapters, i.e. where it deals with geometric structures of statistical models and dual connections. However, the present text goes in much more differential geometric detail, studying also other new topics such as relative curvature tensors, generalized shape operators, dual mean curvature vectors, and entropy maximizing distributions. However, our textbook does not deal with any applications in the field of statistic inference, testing, or estimation. It contains only the analysis of statistical manifolds and statistical models. The question of how the new concepts introduced here apply to other fields of statistics is still under analysis.

Another book of great inspiration for us is the book of Kass and Vos [49], published in 1997. Even if this book deals mainly with the geometrical foundations of asymptotic inference and information loss, it does also contain important material regarding statistical manifolds and their geometry. This challenge is developed more geometrically in the present book than in the former.

Overview

This book is devoted to a specialized area, including Informational Geometry. This is a field that is increasingly attracting the interest of researchers from many different areas of science, including mathematics, statistics, geometry, computer science, signal processing, physics, and neuroscience. It is the authors' hope that the present book will be a valuable reference for researchers and graduate students in one of the aforementioned fields.

The book is structured into two distinct parts. The first one is an accessible introduction to the theory of statistical models, while the second part is devoted to an abstract approach of statistical manifolds.

Part I

The first part contains six chapters and relies on the understanding of the differential geometry of probability density functions viewed as surfaces.

The first two chapters present important density functions, which will offer tractable examples for later discussions in the book. The remaining four chapters devote to the geometry of entropy, which is a fundamental notion in informational geometry. The readers without a strong background in differential geometry can still follow. This part itself can be read alone as an introduction to information geometry.

Chapter 1 introduces the notion of statistical model, which is a space of density functions, and provides the exponential and mixture families as distinguished examples. The Fisher information is defined together with two dual connections of central importance to the theory. The skewness tensor is also defined and computed in certain particular cases.

Chapter 2 contains a few important examples of statistical models for which the Fisher metric and geodesics are worked out explicitly. This includes the case of normal and lognormal distributions, and also the gamma and beta distribution families.

Chapter 3 deals with an introduction to entropy on statistical manifolds and its basic properties. It contains definitions and examples, an analysis of maxima and minima of entropy, its upper and lower bounds, Boltzmann–Gibbs submanifolds, and the adiabatic flow.

Chapter 4 is dedicated to the Kullback–Leibler divergence (or relative entropy), which provides a way to measure the relative entropy between two distributions. The chapter contains explicit computations and fundamental properties regarding the first and second variations of the cross entropy, its relation with the Fisher information matrix and some variational properties involving Kullback–Leibler divergence.

Chapter 5 defines and studies the concept of informational energy on statistical models, which is a concept analogous to kinetic energy from physics. The first and second variations are studied and uncertainty relations and some thermodynamics laws are presented.

Chapter 6 discusses the significance of maximum entropy distributions in the case when the first N moments are given. A distinguished role is played by the case when $N = 1$ and $N = 2$, cases when some explicit computations can be performed. A definition and brief discussion of Maxwell–Boltzmann distributions is also made.

Part II

The second part is dedicated to a detailed study of statistical manifolds and contains seven chapters. This part is an abstractization of the results contained in Part I. Instead of statistical models, one considers here differentiable manifolds, and instead of the Fisher information metric, one takes a Riemannian metric. Thus, we are able to carry the ideas from the theory of statistical models over to Riemannian manifolds endowed with a dualistic structure defined by a pair of torsion-less dual connections.

Chapter 7 contains an introduction to the theory of differentiable manifolds, a central role being played by the Riemannian manifolds. The reader accustomed with the basics of differential geometry may skip to the next chapter. The role of this chapter is to accommodate the novice reader with the language and objects of differential geometry, which will be further developed throughout the later chapters.

A formulation of the dualistic structure is given in Chap. 8. This chapter defines and studies general properties of dual connections, relative torsion tensors and curvatures, α-connections, the skewness and difference tensors. It also contains an equivalent construction of statistical manifolds starting from a skewness tensor.

Chapter 9 describes how to associate a volume element with a given connection and discusses the properties of dual volume elements, which are associated with a pair of dual connections. The properties of α-volume elements are provided with the emphasis on the relation with the Lie derivative and vector field divergence. An explicit computation is done for the distinguished examples of exponential and mixture cases. A special section is devoted to the study of equiaffine connections, i.e. connections which admit a parallel n-volume form. The relation with the statistical manifolds of constant curvature is also emphasized.

Chapter 10 deals with a description of construction and properties of dual Laplacians, which are Laplacians defined with respect to a pair of dual connections. An α-Laplacian is also defined and studied. The relation with the dual volume elements is also emphasized. The last part of the chapter deals with trace of the metric tensor and its relation to Laplacians.

The construction of statistical manifolds starting from contrast functions is described in Chap. 11. The construction of a dualistic structure (Riemannian metric and dual connections) starting from a contrast function is due to Eguchi [38, 39, 41]. Contrast functions are also known in the literature under the name of divergences, a denomination we have tried to avoid here as much as we could.[1]

Chapter 12 presents a few classical examples of contrast functions, such as Bregman, Chernoff, Jefferey, Kagan, Kullback–Leibler, Hellinger, and f-divergence, and their values on a couple of examples of statistical models.

The study of statistical submanifolds, which are subsets of statistical manifolds with a similar induced structure, is done in Chap. 13. Many classical notions, such as second fundamental forms, shape operator, mean curvature vector, and Gauss–Codazzi equations, are presented here from the dualistic point of view. We put our emphasis on the relation between dual objects; for instance, we find a relation between the divergences of dual mean curvature vector fields and the inner product of these vector fields.

[1] A divergence in differential geometry usually refers to an operator acting on vector fields.

The present book follows the line started by Fisher and Efron and continued by Eguchi, Amari, Kaas, and Vos. The novelty of this work, besides presentation, can be found in Chaps. 5, 6, 9, and 13. The book might be of interest not only to differential geometers but also to statisticians and probabilists.

Each chapter ends with a section of proposed problems. Even if many of the problems are left as exercises from the text, there are a number of problems, aiming to deepen the reader's knowledge and skills.

It was our endeavor to make the index as complete as possible, containing all important concepts introduced in definitions. We also provide a list of usual notations of this book. It is worthy noting that for the sake of simplicity and readability, we employed the Einstain summation convention, i.e., whenever an index appears in an expression once upstairs and once downstairs, the summation is implied.

The near flowchart will help the reader navigate through the book content more easily.

Software

The book comes with a software companion, which is an Information Geometry calculator. The software is written in $C\#$ and runs on any PC computer (not a Mac) endowed with .NET Framework. It computes several useful information geometry measures for the most used probability distributions, including entropy, informational energy, cross entropy, Kullback–Leibler divergence, Hellinger distance, and Chernoff information of order α. The user instructions are included in Appendix A. Please visit http://extras.springer.com to download the software.

Bibliographical Remarks

Our presentation of differential geometry of manifolds, which forms the scene where the information geometry objects exist, is close in simplicity to the one of Millman and Parker [58]. However, a more advanced and exhaustive study of differential geometry can be found in Kobayashi and Nomizu [50], Spivak [78], Helgason [44], or Auslander and MacKenzie [9]. For the basics theory of probability distributions the reader can consult almost any textbook of probability and statistics, for instance Wackerly et al. [85].

The remaining parts of the book are based on fundamental ideas inspired from the expository books of Amari and Nagaoka [8] and Kass and Vos [49]. Another important source of information for the reader is the textbook of Murrey and Rice [60]. While the aforementioned references deal with a big deal of statistical inference, our book contains mainly a pure geometrical approach.

One of the notions playing a central role throughout the theory is the Fisher information, which forms a Riemannian metric on the space of probability distributions. This notion was first introduced almost simultaneously by Rao [70] and Jeffreys [46] in the mid-1940s, and continued to be studied by many researchers such as Akin [3], Atkinson and Mitchell [4], James [45], Oller [63], and Oller et al. [64, 65].

The role played by differential geometry in statistics was not fully acknowledged until 1975 when Efron [37] first introduced the concept of statistical curvature for one-parameter models and emphasized its importance in the theory of statistical estimation. Efron pointed out how any regular parametric family could be approximated locally by a curved exponential family and that the curvature of these models measures their departure from exponentiality. It turned out that this concept was intimately related to Fisher's theory of information loss. Efron's formal theory did not use all the bells and whistles of differential geometry. The first step to an elegant geometric theory was done by Dawid [33], who introduced a connection on the space of all positive probability distributions and showed that Efron's statistical curvature is induced by this connection.

The use of differential geometry in its elegant splendor for the elaboration of previous ideas was systematically achieved by Amari [6] and [7], who studied the informational geometric properties of a manifold with a Fisher metric on it. This is the reason why sometimes this is also called the Fisher–Efron–Amari theory.

The concept of dual connections and the theory of dually flat spaces as well as the α-connections were first introduced in 1982 by Nagaoka and Amari [61] and developed later in a monograph by Amari [5]. These concepts were proved extremely useful in the study of informational geometry, which investigates the differential geometric structure of the manifold of probability density functions. It is worthy to note the independent work of Chentsov [26] on α-connections done from a different perspective.

Entropy, from its probabilistic definition, is a measure of uncertainty of a random variable. The maximum-entropy approach was introduced by the work of Shannon [73]. However, the *entropy maximization principle* was properly introduced by Akaike [2] in the mid-1970s, as a theoretical basis for model selection. Since then it has been used in order to choose the least "biased" distribution with respect to the unknown information. The included chapter regarding maximum entropy with moments constraints is inspired by Mead [57].

The entropy of a continuous distribution is not always positive. In order to overcome this flaw one can use the relative entropy of two distributions p and q. This concept was originally introduced by S. Kullback and R. Leibler in 1951, see [51, 52]. This is also referred in the literature under the names of divergence, information divergence, information gain, relative entropy, or Kullback–Leibler divergence. The Kullback–Leibler divergence models the information between a true distribution p and a model distribution q; the reader can consult the book of Burnham and Anderson [21] for details.

In practice, the density function p of a random variable is unknown. The problem is the one of drawing inferences about the density p on the basis of N concrete realizations of the random variable. Then we can look for the density p as an element of a certain restricted class of distributions, instead of all possible distributions. One way in which this restricted class can be constructed is to consider the distributions having the same mean as the sample mean and the variance equal to the variance of the sample. Then, we need to choose the distribution that satisfies these constraints and is the most ignorant with respect to the other moments. This is realized for the distribution with the maximum entropy. The theorems regarding maximum entropy distributions subject to different constraints are inspired from Rao [71]. They treat the case of the normal distribution, as the distribution on $\mathbb{R}$ with the first two moments given, the exponential distribution, as the distribution on $[0, \infty)$ with the given mean, as well as the case of Maxwell–Boltzman distribution. The case of the maximum entropy distribution with the first n given moments is inspired from Mead and Papanicolaou [57]. The novelty brought by this chapter is the existence of maximum entropy distributions in the case when the sample space is a finite interval. The book also introduces the curves of largest entropy, whose relevance in actual physical situations is worth examining.

The second part of the book deals with statistical manifolds, which are geometrical abstractions of statistical models. Lauritzen [54] defined statistical manifolds as Riemannian manifolds endowed with a pair of torsion-free dual connections. He also introduced an equivalent way of constructing statistical manifolds, starting from a skewness tensor. A presentation of statistical structure in the language of affine connections can be found in Simon [76].

The geometry of a statistical model can be also induced by contrast functions. The dualistic structure of contrast functions was developed by Eguchi [38, 39, 41], who has shown that a contrast function induces a Riemannian metric by its second order derivatives, and a pair of dual connections by its third order derivatives. Further information on contrast geometry can be found in Pfanzagl [69]. A generalization of the geometry induced by the contrast functions is the yoke geometry, introduced by Barndorff-Nilsen [11–13] and developed by Blæsid [18, 19].

Acknowledgments

The monograph was partially supported by the NSF Grant # 0631541. The authors are indebted to Eastern Michigan University and University Politehnica of Bucharest for the excellent conditions of research. We wish to express many thanks to Professor Jun Zhang from University of Michigan for many fruitful discussions on the subject and for introducing the first author to the fascinating subject of this book. We express our gratitude to Professor Ionel Tevy from Politehnica University of Bucharest for clarifying certain mathematical problems in this book.

This work owes much of its clarity and quality to the numerous comments and observations made by several unknown reviewers whose patience and time spent on the manuscript are very much appreciated.

Finally, we would like to express our gratitude to Springer and its editors, especially Donna Chernyk for making this endeavor possible.

Ypsilanti, MI, USA — Ovidiu Calin
Bucharest, Romania — Constantin Udrişte
January, 2014

Springer

http://www.springer.com/978-3-319-07778-9

Geometric Modeling in Probability and Statistics
Calin, O.; Udriste, C.
2014, XXIII, 375 p. 22 illus., 3 illus. in color. With online files/update., Hardcover
ISBN: 978-3-319-07778-9