Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg

28

Shun-ichi Amari

Differential-Geometrical Methods in Statistics

Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg

28

Shun-ichi Amari

Differential Geometrical Methods in Statistics

Springer-Verlag
Berlin Heidelberg New York Tokyo

Author

Shun-ichi Amari
University of Tokyo, Faculty of Engineering
Department of Mathematical Engineering and Instrumentation Physics
Bunkyo-ku, Tokyo 113, Japan

AMS Subject Classification: 62-03, 60E99, 62E99

ISBN 3-540-96056-2 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-96056-2 Springer-Verlag New York Heidelberg Berlin Tokyo

Library of Congress Cataloging in Publication Data. Amari, Shun'ichi. Differential-geometrical methods in statistics. (Lecture notes in statistics; 28) Bibliography: p. Includes index. 1. Mathematical statistics. 2. Geometry, Differential. I. Title. II. Series: Lecture notes in statistics (Springer-Verlag); v. 28. QA276.A56 1985 519.5 85-4646

Printed in Germany

Printing and binding: Beltz Offsetdruck, Hemsbach/Bergstr.
2147/3140-543210

CONTENTS

1. INTRODUCTION

Why Geometry?

One may ask why geometry, in particular differential geometry, is useful for statistics. The reason seems very simple and strong. A statistical model is a set of probability distributions to which we believe the true distribution belongs. It is a subset of all the possible probability distributions. In particular, a parametric model usually forms a finite-dimensional manifold imbedded in the set of all the possible probability distributions. For example a normal model consists of the probability distributions $N(\mu, \sigma^2)$ parametrized by two parameters (μ, σ). The normal model $M = \{N(\mu, \sigma^2)\}$ forms a two-dimensional manifold with coordinates μ and σ, and is imbedded in the set $S = \{p(x)\}$ of all the regular probability distributions of a random variable x. One often uses a statistical model to carry out statistical inference, assuming that the true distribution is included in the model. However, a model is merely a hypothesis. The true distribution may not be in the model but be only close to it. Therefore, in order to evaluate statistical inference procedures, it is important to know what part the statistical model occupies in the entire set of probability distrubutions and what shape the statistical model has in the entire set. This is the problem of geometry of statistical models. It is therefore expected that a fundamental role is played in statistics by the geometrical quantities such as the distance or divergence of two probability distributions, the flatness or curvature of a statistical model, etc. However, it is by no means a trivial task to define such geometrical structures in a natural and invariant manner.

Statistical inference can be carried out more and more precisely as the number of observations increases, so that one can construct a universal asymptotic theory of statistical inference in the regular case. Since the estimated probability distribution lies very close to the true distribution in this case, it is sufficient when evaluating statistical procedures to take account of only the local structure of the model in a small neighborhood of the true or estimated distribution. Hence, one can locally linearize the model at the true or estimated distribution, even if the model is curved in the entire set. Geometrically, this local linearization is an approximation to the manifold by the tangent space at a point. The tangent space has a natural inner product (Riemannian metric) given by the Fisher information matrix. From the geometrical point of view, one may say that the asymptotic theory of statistical inference has indeed been constructed by using the linear geometry of tangent spaces of a statistical model, even if it has not been explicitly stated.

Local linearization accounts only for local properties of a model. In order to elucidates larger-scale properties of a model, one needs to introduce mutual relations of two different tangent spaces at two neighboring points in the model. This can be done by defining an affine correspondence between two tangent spaces at neighboring points. This is a standard technique of differential geometry and the correspondence is called an affine connection. By an affine connection, one can study local non-linear properties, such as curvature, of a model beyond linear approximation. This suggests that a higher-order asymptotic theory can naturally be constructed in the framework of differential geometry. Moreover, one can obtain global properties of a model by connectiong tangent spaces at various points. These considerations show the usefulness and validity of the differential-geometrical approach to statistics. Although the present monograph treats mainly the higher-order asymptotic theory of statistical inference, the differential-geometrical method is useful for more general statistical anlyses. It seems rather surprising that few theories have so far been developed concerning geometrical properties of a family of probability distributions.

Historical Remark

It was Rao (1945), in his early twenties, who first noticed the importance of the differential-geometrical approach. He introduced the Riemannian metric in a statistical manifold by using the Fisher information matrix and calculated the geodesic distances between two distributions for various statistical models. This theory made an impact and not a few researchers have tried to construct a theory along this Riemannian line. Jeffreys also remarked the Riemannian distance (Jeffreys, 1948) and the invariant prior of Jeffreys (1946) was based on the Riemannian concept. The properties of the Riemannian manifold of a statistical model have further been studied by a number of researchers independently, e.g., Amari (1968), James (1973), Atkinson and Mitchell (1981), Dawid (1977), Akin (1979), Kass (1980), Skovgaard (1984), etc. Amari's unpublished results (1959) induced a number of researches in Japan; Yoshizawa (1971a, b), Takiyama (1974), Ozeki (1971), Sato et al. (1979), Ingarden et al. (1979), etc. Nevertheless, the statistical implications of the Riemannian curvature of a model did not become clear. Some additional concepts seemed necessary for proving the usefulness of the geometrical approach.

It was an isolated work by Chentsov (1972) in a Russian book (translated in English in 1982) and in some papers prior to the book that developed a new concept on statistical manifolds. He introduced a family of affine connections in a statistical manifold, whereas only the Riemannian (Levi-Civita) connection was used in the above works. He also proved that the Fisher information and these affine connections are unique in the manifold of probability distributions on a finite number of atoms. He proved this from the point of view of the categorical invariance, by considering a category whose objects are multinomial distributions and whose morphisms are Markovian mappings between them. His theory is deep and fundamental, and he elucidates the geometrical structures of the exponential family. However, he did not remark the curvature of a statistical manifold, which plays a central role in the higher-order asymptotic theory of statistical inference.

It was Efron (1975, 1978) who opened a new idea independently of Chentsov's work. He defined the statistical curvature of a statistical model, and pointed out that the statistical curvature plays a fundamental role in the higer-order asymptotic theory of statistical inference. Although he did not introduce an affine connection explicitly, a new affine connection (exponential connection) was introduced implicitly in his theory, as was elucidated by Dawid (1975). Dawid also suggested the possibility of introducing another affine connection (mixture connection). Efron's idea was generalized by Madsen (1979); see also Reads (1975).

Under the strong influence of Efron's paper and Dawid's suggestion, Amari (1980, 1982a) introduced a one-parameter family of affine connections (α-connections), which turned out to be equivalent to those Chentsov had already defined. Amari further proposed a differential-geometrical framework for constructing a higher-order asymptotic theory of statistical inference. He, defining the α-curvature of a submanifold, pointed out important roles of the exponential and mixture curvatures and their duality in statistical inference. Being stimulated by this framework, a number of papers appeared, e. g. Amari (1982b, 1983a, b), Amari and Kumon (1983), Kumon and Amari (1983, 1984, 1985), Eguchi (1983, 1984); see also Wei and Tsai (1983), Kass (1984). The theoretical background was further deepened by Nagaoka and Amari (1982), where the dualistic viewpoint was refined and some new geometrical concepts were introduced. Here statistcs contributes to differential geometry.

Professors D. R. Cox, O. E. Barndorff-Nielsen and D.V. Hinkley organized a NATO Advanced Workshop on Differential Geometry in Statistical Inference in April, 1984 in London. More than forty researchers participated, and stimulating discussions took place concerning the present achievement by and future prospects for the differential-geometrical method in statistics. New directions of developments were shown, e. g. by Amari (1984 a), Barndorff-Nielsen(1984), Lauritzen(1984), etc. I believe that the differential geometrical method will become established as one of the main and indispensable theoretical methods in statistics.

Organization of the Monograph

Part I treats fundamental geometrical properties of parametric families of probability distributions. We define in Chapter 2 the basic quantities of a statistical manifold, such as the Riemannian metric, the α-affine connection, the α-curvature of a submanifold, etc. This chapter also provides a good introduction to differential geometry, so that one can read the Monograph without any prior knowledge on differential geometry. The explanation is rather intuitive, and unnecessary rigorous treatments are avoided. The reader is asked to refer to Kobayashi and Nomizu (1963, 1969) or any other textbooks for the modern approach to differential geometry, and to Schouten (1954) for the old tensorial style of notations. Chapter 3 presents an advanced theory of differential geometry of statistical manifolds. A pair of dual connections are introduced in a differentiable manifold with a Riemannian metric. The dualistic characteristics of an α-flat manifold are especially interesting.

We can define an α-divergence measure between two probability distributions in an α-flat manifold, which fits well to the differential geometrical structures. The Kullback-Leibler information, the Chernoff distance, the f-divergence of Csiszár, the Hellinger distance etc. are all included in this class of α-divergences. This chapter is based mainly on Nagaoka and Amari (1982), which unifies the geometry of Csiszár (1967a,b; 1975) and that of Chentsov (1972) and Amari (1982a). This type of the duality theory cannot be found in any differential geometry literature.

Part II is devoted to the higher-order asymptotic theory of statistical inference in the framework of a curved exponential family. We present the fundamental method of approach in Chapter 4, by decomposing the minimal sufficient statistic into the sum of an asymptotically sufficient and asymptotically ancillary statistics in the tangent space of a model. The Edgeworth expansion of their joint probability distribution is explicitly given in geometrical terms up to the term of order 1/N, where N is the number of observations. Chapter 5 is devoted to the theory of estimation, where both the exponential and mixture curvatures play important roles. Chapter 6 treats the theory of statistical tests. We calculate the power functions of various efficient tests such as the Wald test, the Rao test (efficient score test), the likelihood ratio test, etc. up to the term of order 1/N. The characteristic of various first-order efficient tests are compared. Chapter 7 treats more basic structures concerning information such as higher-order asymptotic sufficiency and ancillarity. Conditional inference is studied from the geometrical point of view. The relation between the Fisher information and higher-order curvatures is elucidated. Chapter 8 treats statistical inference in the presence of nuisance parameters. The mixture and exponential curvatures again play important roles.

It was not possible to include in this volume the newly developing topics such as those presented and discussed at the NATO Workshop. See, e.g., Barndorff-Nielsen (1984), Lauritzen (1984) and Amari (1984 a), which together will appear as a volume of the IMS Monograph Series, and the papers by R.E. Kass, C.L.Tsai, etc. See also Kumon and Amari (1984), Amari and Kumon (1985), Amari (1984 c). The differential-geometrical method developed in statistics is also applicable to other fields of sciences such as information theory and systems theory (Amari, 1983 c, 1984 b). See Ingarden (1981) and Caianiello (1983) for applications to physics. They together will open a new field, which I would like to call <u>information geometry</u>.

Personal Remarks

It was in 1959, while I was studying for my Master's Degree at the University of Tokyo, that I became enchanted by the idea of a beautiful geometrical structure of a statistical model. I was suggested to consider the geometrical structure of the family of normal distributions, using the Fisher information as a Riemannian metric. This was Professor Rao's excellent idea proposed in 1945. I found that the family of normal distributions forms a Riemannian manifold of constant negative curvature, which is the Bolyai-Lobachevsky geometry well known in the theory of non-Euclidean geometry. My results on the geodesic, geodesic distance and curvature appeared in an unpublished report. I could not understand the statistical meaning of these results, in particular the meaning of the Riemannian curvature of a statistical manifold. Since then, I have been dreaming of constructing a theory of differential geometry for statistics, although my work has been concentrated in non-statistical areas, namely graph theory, continuum mechanics, information sciences, mathematical theory of neural nets, and other aspects of mathematical engineering. It was a paper by Professor Efron that awoke me from my dream and led me to work enthuastically on constructing a differential-geometrical theory of statistics. This Monograph is a result of several years of endeavour by myself along this line.

Finally, I list up some problems which I have now interests in and am now studying.

1. Extension of the geometric theory of statistical inference such that it is applicable to a general regular parametric model which is not necessarily a curved exponential family. This extension is possible by introducing the jet bundle which is an aggregate of local exponential families. Here, a local exponential family is attached to each point of the model such that the original model is locally (approximately) imbedded in the exponential family at that point.

2. Extension of the present theory to the function space of regular probability distributions. This enables us to construct a geometrical theory of non-parametric, semi-parametric and robust statistical inference.

3. The problem of estimating a structural parameter in the presence of as many incidental parameters as the number of observations. This classical problem can be elucidated by introducing a Hilbert bundle to the underlying statistical model.

4. Differential geometry of a statistical model which possesses an invariant transformation group. The structure of such a model is highly related to the existence of an exact ancillary statistics.

5. Geometry of statistical models of discrete random variables and categorical data analysis.

6. Geometry of multivariate statistical analysis.

7. Geometry of time-series analysis. Local and global structures of parametric time-series models are interesting.

8. Differential-geometrical theory of systems.

9. Application of differential geometry to information theory, coding theory and the theory of flow. We need to study geometrical structures of a manifold of information sources (e.g., the manifold of Markov chains and the manifold of coders, which map the manifold of all the information sources into itself.

10. Geometry of non-regular statistical models. Asymptotic properties of statistical inference in a non-regular model are related to both the Finsler geometry and the theory of stable distributions of degree α.

Acknowledgement

I would like to express my sincere gratitude to Professor Emeritus Kazuo Kondo, who organized the RAAG (Research Association of Applied Geometry) and introduced me to the world of applied geometry. The author also thanks Professor S. Moriguti for his suggestion of the geometrical approach in statistics. I especially appreciate valuable suggestions and encouragement from Professor K. Takeuchi, without which I could not complete the present work. I am greateful to many statisticians for their warm encouragement, useful comments and inspiring discussions. I would like to mention especially Professor B. Efron, Professor A.P.Dawid, Professor D.R.Cox, Professor C.R.Rao, Professor O. Barndorff-Nielsen, Professor S. Lauritzen, Professor D.V. Hinkley, Professor D.A.Pierce, Professors Ib. and L.T. Skovgaard, Professor T. Kitagawa, Professor T. Okuno, Professor T.S.Han, and Professor M. Akahira, Dr. A. Mitchell. The comments by Professor H. Kimura and Professor K. Kanatani were also useful. , Professor Th. Chang and Dr. R. Lockhart were kind enough to read the first version of the manuscript and gave me detailed and valuable suggestions both from the mathematical and editorial points of view.

My special thanks go to young Japanese researchers Dr.M. Kumon and Mr. H. Nagaoka who are actively working in this field. They collaborated with me for years when they were my students in constructing the differential-geometrical theory of statistical inference. Without their cooperation, it would have been difficult to construct the differential geometrical theory at such a speed. Mr. S. Shimada, Mr. K. Kurata and many other members of my laboratory checked the manuscript carefully. Mr. K. Shimada helped me make numerical calculations and fine illustrations. Lastly but not leastly, I would like to express my heartfelt thanks to Mrs. T. Shintani and Miss K. Enomoto for their devotion and patience in typing such a difficult manuscript.

PART I. GEOMETRICAL STRUCTURES OF A FAMILY OF PROBABILITY DISTRIBUTIONS

2. DIFFERENTIAL GEOMETRY OF STATISTICAL MODELS

The present chapter is devoted to the introduction of fundamental differential-geometrical structures of statistical models. The tangent space, the Riemannian metric and the α-connections are introduced in a statistical manifold. No differential-geometrical background is required for reading this monograph, because the present chapter provides a readable introduction to differential geometry.

2.1. Manifold of statistical model

Statisticians often treat a parametrized family of probability distributions as a statistical model. Let $S = \{p(x, \theta)\}$ be such a statistical model, where x is a random variable belonging to sample space X, and $p(x, \theta)$ is the probability density function of x, parametrized by θ, with respect to some common dominating measure P on X. Here, θ is a real n-dimensional parameter $\theta = (\theta^1, \theta^2, \ldots, \theta^n)$ belonging to some open subset Θ of the n-dimensional real space R^n. For example, the normal model is a family of probability distributions having the following density functions,

$$p(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{- \frac{(x - \mu)^2}{2\sigma 2}\}$$

where sample space X is the real R^1 with the Lebesgue measure dP = dx and the parameter θ is two-dimensional; we may put $\theta = (\theta^1, \theta^2) =$