Comment

S.-I. Amari

1. STATISTICS AND GEOMETRY

It has clearly been shown by Kass that the pioneering ideas of Fisher and Jeffreys naturally lead to a geometrical theory of statistics and that differential geometrical concepts, especially curvatures, play fundamental roles in the asymptotic theory of statistics. However, one might further ask if there are any results obtained only by the geometrical method and not by ordinary analytical methods. If there are none, why do we need complicated geometrical concepts? Before answering this everlasting question, I should like to explain intuitively the reason why the geometrical method is natural and useful.

A statistical method $S = \{p(x, \theta)\}$, where $p(x, \theta)$ is a probability density of x parameterized by an n-dimensional vector parameter θ, is naturally regarded as an n-dimensional manifold imbedded in the set $\{f(x)\}$ of all the probability density functions, which is a subset of the L^1-space. Characteristics of statistical inferential procedures depend on the analytic properties of functions $p(x, \theta)$ in the model. However, we can show that relevant properties are geometrically represented by the imbedded form of S in L^1. In the first-order asymptotic case where the number of observations is large, an inferential procedure is so accurate that it suffices to take a neighborhood of the true distribution into consideration. This implies geometrically that we can approximate a curved model manifold S by a flat tangent space at the true distribution and can evaluate inferential procedures by using this linear model. The first-order theory is a linear approximation. This is the reason why we have a distribution-free first-order asymptotic theory depending only on the Fisher information matrix, because every tangent space is geometrically isomorphic (equivalent).

When we construct the second- or third-order asymptotic theory, it is natural to approximate the statistical model S by a second-order osculating manifold at the true distribution. It is then expected that we have unified distribution-free results, depending only on the Fisher metric (linear approximation) and the curvatures (which are characteristic quantities for the second-order approximation). This is true, and the curvatures play a fundamental role as Kass demonstrated.

However, the geometry of a family of probability distributions turns out to be neither Euclidean nor simply Riemannian. The geometry should represent analytical properties of $p(x, \theta)$. This requirement naturally leads us to a Riemannian manifold having a dual pair of affine connections. The dual connections introduce a new concept in differential geometry, and we are required to construct a new theory of dualistic geometry. Here is a big contribution of statistics to geometry. We have two kinds of curvatures (exponential and mixture), both of which play proper roles in statistics.

Returning to the problem we posited, it is true that we can construct an asymptotic theory without geometry. This is true in the sense that any mathematical theory can be constructed without geometry. Even the results of Euclidean geometry can be described by algebraic equations (analytical geometry). However, if instead of saying that two edges have an equal length in a triangle when two angles are equal, we write down the corresponding statement in the form of equations, we lose clear intuitive understanding. It is awkward and difficult to prove the statement without geometrical intuition. It is more natural and easier to use geometry when we study objects having geometrical structures.

I would like to emphasize that geometry can summarize necessary analytic properties of a family of probability distributions and of their inferential procedures in a unified manner. Statistial models have natural geometrical structures.

I agree that most of the higher-order asymptotics have been constructed without geometry. I would like to point out one result which was first obtained by the geometrical method (Amari, 1983, 1985; Kumon and Amari, 1983). We have known many efficient tests (the likelihood ratio test, Wald test, Rao test, etc.) of testing H_0: $t = t_0$ against H_1: $t > t_0$ (one-sided) or H_1: $t \neq t_0$ (both-sided). Their performances are equivalent in the first-order asymptotics, and their power functions are automatically equivalent up to the second-order. However, the third-order terms of the power functions are different, so that they have different characteristics represented by their third-order power-loss functions or deficiency curves. Only the geometrical method has succeeded in calculating these quantities, elucidating the higher-order

S.-I. Amari is Professor, Department of Mathematical Engineering and Information Physics. His mailing address is: Department of Mathematical Engineering and Information Physics, University of Tokyo, Tokyo 113, Japan.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science.* www.jstor.org

characteristics of these efficient tests. They could have been derived without geometry, but no one has yet succeeded in doing so.

2. RECENT RESULTS

Although its origin goes back to the 1940's, the geometry of statistical manifolds is relatively new and is now growing slowly but steadily. I would like to list here some of the recent results.

Barndorff-Nielsen has opened new directions of development such as observed geometry conditioned on ancillary statistics, string calculus and geometry of transformational models. The results are summarized in his book (Barndorff-Nielsen, 1988; see also Barndorff-Nielsen and Jupp, 1989; Mitchell, 1989; Murray, 1988). Vos (1989) studied the geometry of generalized linear models and an invariant decomposition of higher-order quantities. Differential geometry of a semiparametric model is given by Amari (1987c) and Amari and Kumon (1988) and is related to estimation in the presence of infinitely many nuisance parameters. Picard (1989) presented new characterization of invariant geometrical structures.

Mathematicians have began to pay attention to this new dual geometry. Dodson (1987) organized an international workshop to bridge mathematicians and statisticians. Syriaev and Fomenko also organized one in the USSR. It was pointed out that the dual geometry has some relation with affine differential geometry (Nomizu and Pinkall, 1986; see also Lauritzen, 1987b). Recently, Kurose (1988) developed the dual theory from the point of view of affine and projective differential geometry. Amari pointed out that a sequential inference procedure gives rise to a conformal transformation in the statistical manifold, and geometry of sequential estimation is studied by Okamoto, Amari and Takeuchi (1989).

3. INFORMATION GEOMETRY

The dual geometry is useful not only for the higher-order asymptotics of statistical inference but also for many other problems which make use of probability distributions. It is thus expected to become a unifying method of a wide range of information sciences including statistics. This may be called information geometry. The following are some examples of new directions of developments.

Time Series Analysis and Theory of Control Systems

Let $x = \{x_t\}$, $t = \cdots, 0, 1, 2, \cdots$, be a sample path of a regular stationary stochastic process. For example, we consider an ARMA (p, q) model in which x_t is determined recursively by the equation

$$x_t + \sum_{i=1}^{p} a_i x_{t-i} = \sum_{j=0}^{q} b_j \varepsilon_{t-j},$$

where $\{\varepsilon_t\}$ is a unit white Gaussian noise process. The probability distributions of the process x (which is infinite-dimensional) are specified by a $(p + q + 1)$-dimensional parameter $\theta = (a_i, b_j)$. We symbolically write this as $P(x, \theta)$. Then, the set S of all the ARMA (p, q) processes forms a $(p + q + 1)$-dimensional manifold which is equipped with the Fisher metric and the two dual affine connections (Amari, 1987b). The geometric properties represent their inner relations related to stochastic properties.

One may regard an ARMA model as a linear discrete-time stationary system driven by white noise. One may then identify a set of parameterized linear systems with the set S of the corresponding output stochastic processes. We can elucidate properties of a family of linear systems by dual geometry.

It is shown that an AR model is e-flat ($a = -1$-flat). We can solve not only inferential problems on these models but also approximation problems of stochastic processes or linear systems by one belonging to a lower-dimensional family. Geometry provides an intuitive and natural guidance to the intrinsic properties of these families.

Information Theory

An information source produces a stochastic process, so that a parameterized family of information sources has a natural geometric structure (Campbell, 1985). For example, a set of all the Markov chains on a fixed finite alphabet set forms an e-flat manifold. An encoder defines a mapping from a manifold of information sources to another one. Its characteristics can be studied geometrically by using properties of the mapping.

Geometrical structures are also useful for studying multiterminal problems. Let (X, Y) be two correlated information sources, with joint probability $p(x, y; \theta)$ parameterized by θ. Let $x_1, \cdots, x_N$, and $y_1 \cdots, y_N$ be N independent observations, where x_i and y_i are correlated. When N data $\{x_i\}$ and N data $\{y_i\}$ are summarized in the statistics $m_X(x_1, \cdots, x_N)$ and $m_Y(y_1, \cdots, y_N)$ independently, some information is lost. When their cardinalities $|\{m_X\}|$ and $|\{m_Y\}|$ or the Shannon information amounts are limited, what is the amount of loss of Fisher information or of the loss of power in the case of testing? This gives a very good example to show how geometrical structures are important for solving problems of correlated information sources (Amari and Han, 1989; Amari, 1989). I believe that differential geometry becomes a key tool for connecting information theory and statistics.

Linear Programming Problem

It is interesting that the dual geometry is useful for some other problems. When a convex function $\Psi(\theta)$ is defined, we have a Legendre transformation from θ to η with a dual convex function $\Phi(\eta)$. We can introduce a dually flat geometry when it is equipped with a pair of convex functions. In the case of statistics, Φ is the negative of the entropy function and Ψ is the cumulant generating function. We have natural convex functions derived from linear and non-linear programming problems.

It is interesting to point out that a continuous version of the Karmarkar inner method is just to proceed along an m-geodesic in the space thus equipped with the dual connections. This method can easily be generalized to a nonlinear programming problem. This shows a wide applicability and universality of dual geometry.

ADDITIONAL REFERENCES

AMARI, S.-I. (1983). Comparisons of asymptotically efficient tests in terms of geometry of statistical structures. *Bull. Internat. Statist. Inst., Proc. 44th Session, Book 2* 1190–1206.

AMARI, S.-I. (1987b). Differential geometry of a parametric family of invertible linear systems—Riemannian metric, dual affine connections and divergence. *Math. Systems Theory* **20** 53–82.

AMARI, S.-I. (1987c). Dual connections on the Hilbert bundles of statistical models. In *Geometrization of Statistical Theory* (C. T. J. Dodson, ed.) 123–152. ULDM Publications, Lancaster.

AMARI, S.-I. (1989). Fisher information under restriction of Shannon information. *Ann. Inst. Statist. Math.* To appear.

AMARI, S.-I. and HAN, T. S. (1989). Statistical inference under multiterminal rate restrictions—a differential geometrical approach. *IEEE Trans. Inform. Theory.* To appear.

AMARI, S.-I. and KUMON, M. (1988). Estimation in the presence of infinitely many nuisance parameters—geometry of estimating functions. *Ann. Statist* **16** 1044–1068.

BARNDORFF-NIELSEN, O. E. (1988). *Parametric Statistical Models and Likelihood. Lecture Notes in Statist.* **50**. Springer, New York.

BARNDORFF-NIELSEN, O. E. and JUPP, P. E. (1989). Approximating exponential models. *Ann. Inst. Statist. Math.* To appear.

CAMPBELL, L. L. (1985). The relation between information theory and the differential geometric approach to statistics. *Inform. Sci.* **35** 199–210.

DODSON, C. T. J., ed. (1987). *Geometrization of Statistical Theory.* ULDM Publications, Lancaster.

KUROSE, T. (1988). Dual connections and affine geometry. Technical Report. UTYO-MATH 88-26.

LAURITZEN, S. L. (1987b). Conjugate connections in statistical theory. In *Geometrization of Statistical Theory* (C. T. J. Dodson, ed.) 33–51. ULDM Publications, Lancaster.

MITCHELL, A. F. S. (1989). The information matrix, skewness tensors and α-connections for the general multivariate elliptic distribution. *Ann. Inst. Statist. Math.* To appear.

NOMIZU, K. and PINKALL, U. (1986). On the geometry of affine immersions. Technical Report MPI 86-28, Max Planck Inst.

OKAMOTO, I., AMARI, S.-I. and TAKEUCHI, K. (1989). Asymptotic theory of sequential estimation procedures for curved exponential families. Unpublished manuscript.

PICARD, D. B. (1989). Statistical morphisms. Unpublished manuscript.

Comment

O. E. Barndorff-Nielsen

Dr. Kass' fine account calls for little comment in itself. However, as he himself stresses, it leaves out parts of the subject, particularly of the more advanced aspects, and it may be useful here to outline briefly some of these parts so as to provide the interested reader with a fuller, though still far from comprehensive, picture of the scene. The discussion below relates mainly to work with which I have been to some degree associated, and it gives, in particular, virtually no impression of the important and extensive work of S.-I. Amari and his collaborators.

As will be indicated, the statistical problems have led to various developments and questions of a purely mathematical nature, and there are also interesting relations to theoretical physics.

INDEX NOTATION

The index notation of classical differential geometry and certain extensions thereof have turned out to be highly useful for many calculations in statistics, including some that are not of differential geometric nature (cf. McCullagh, 1987; Barndorff-Nielsen and Blæsild, 1988b; Barndorff-Nielsen and Cox, 1989, Chapter 5). The index notation makes many multivariate calculations just as easy as the corresponding

O. E. Barndorff-Nielsen is Professor in the Department of Theoretical Statistics at Aarhus University. His mailing address is: Matematisk Institut, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark.