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. Statistical 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.