# CHAPTER 1. INTRODUCTION

Robert E. Kass*

Geometrical analyses of parametric inference problems have developed from two appealing ideas: that a local measure of distance between members of a family of distributions could be based on Fisher information, and that the special place of exponential families in statistical theory could be understood as being intimately connected with their loglinear structure. The first led Jeffreys (1946) and Rao (1945) to introduce a Riemannian metric defined by Fisher information, while the second led Efron (1975) to quantify departures from exponentiality by defining the curvature of a statistical model. The papers collected in this volume summarize subsequent research carried out by Professors Amari, Barndorff-Nielsen, Lauritzen, and Rao together with their coworkers, and by other authors as well, which has substantially extended both the applicability of differential geometry and our understanding of the role it plays in statistical theory.**

The most basic success of the geometrical method remains its concise summary of information loss, Fisher's fundamental quantification of departure from sufficiency, and information recovery, his justification for conditioning. Fisher claimed, but never showed, that the MLE minimized the loss of information among efficient estimators, and that successive portions of the loss could be

---

*Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA

**These papers were presented at the NATO Advanced Workshop on Differential Geometry in Statistical Inference at Imperial College, April, 1934.

recovered by conditioning on the second and higher derivatives of the log-likelihood function, evaluated at the MLE. Concerning information loss, recall that according to the Koopman-Darmois theorem, under regularity conditions, the families of continuous distributions with fixed support that admit finite-dimensional sufficient reductions of i.i.d. sequences are precisely the exponential families. It is thus intuitive that (for such regular families) departures from sufficiency, that is, information loss, should correspond to deviations from exponentiality. The remarkable reality is that the correspondence takes a beautifully simple form. The most transparent case, especially for the untrained eye, occurs for a one-parameter subfamily of a two-dimensional exponential family. There, the relative information loss, in Fisher's sense, from using a statistic T in place of the whole sample is

$$\lim i(\theta)^{-1}[ni(\theta)-i^{T}(\theta)] = \gamma^2 + \frac{1}{2}\beta^2 \tag{1}$$

where $ni(\theta)$ is the Fisher information in the whole sample, $i^T(\theta)$ is the Fisher information calculated from the distribution of T, $\gamma$ is the statistical curvature of the family and $\beta$ is the mixture curvature of the "ancillary family" associated with the estimator T. When the estimator T is the MLE, $\beta$ vanishes; this substantiates Fisher's first claim.

In his 1975 paper, Efron derived the two-term expression for information loss (in his equation (10.25)), discussed the geometrical interpretation of the first term, and noted that the second term is zero for the MLE. He defined $\gamma$ to be the curvature of the curve in the natural parameter space that describes the subfamily, with the inner product defined by Fisher information replacing the usual Euclidean inner product. The definition of $\beta$ is exactly analogous to that of $\gamma$, with the mean value parameter space used instead of the natural parameter space, but Efron did not recognize this and so did not identify the mixture curvature. He did stress the role of the ancillary family associated with the estimator T (see his Remark 3 of Section 9 and his reply to discussants, p. 1240), and he also noticed a special case of (1) (in his reply, p. 1241). The final simplicity of the complete geometrical version of (1) appeared in Amari's 1982 Annals paper. There it was derived in the multiparameter case; see equation (4.8) of Amari's paper in this volume.

Prior to Efron's paper, Rao (1961) had introduced definitions of efficiency and second-order efficiency that were intended to classify estimators just as Fisher's definitions did, but using more tractable expressions. This led to the same measure of minimum information loss used by Fisher (corresponding to $\gamma^2$ in equation (1)). Rao (1962) computed the information loss in the case of the multinomial distribution for several different methods of estimation. Rao (1963) then went on to provide a decision-theoretic definition of second-order efficiency of an estimator T, measuring it according to the magnitude of the second-order term in the asymptotic expansion of the bias-corrected version of T. Efron's analysis clarified the relationship between Fisher's definition and Rao's first definition. Efron then provided a decomposition of the second-order variance term in which the right-hand side of (1) appeared, together with a parameterization-dependent third term. The extension to the multiparameter case was derived by Madsen (1979) following the outline of Reeds (1975). It appears here in Amari's paper as Theorem 3.4.

An analytically and conceptually important first step of Efron's analysis was to begin by considering smooth subfamilies of regular exponential families, which he called curved exponential families. Analytically, this made possible rigorous derivations of results, and for this reason such families were analyzed concurrently by Ghosh and Subramaniam (1974). Conceptually, it allowed specification of the ancillary families associated with an estimator: the ancillary family associated with T at t is the set of points y in the sample space of the full exponential family - equivalently, the mean value parameter space for the family - for which $T(y) = t$. The terminology and subsequent detailed analysis is due to Amari but, as noted above, the importance of the ancillary family, at once emphasized and obscured by Fisher, was apparent from Efron's presentation.

The ancillary family is also important in understanding information recovery, which is the reason Amari has chosen to use the modifier "ancillary." In the discussion of Efron's paper, Pierce (1975) noted another interpretation of statistical curvature: it furnishes the asymptotic standard deviation of observed information. More precisely, it is the asymptotic standard deviation of the asymptotically ancillary statistic $n^{-1/2}i(\hat{\theta})^{-1}[I(\hat{\theta}) - ni(\hat{\theta})]$, where $ni(\theta)$ is expected information and $I(\hat{\theta})$ is observed information; the one-parameter statement appears in Efron and Hinkley, (1978), and the multiparameter version is in Skovgaard (1985). When fitting a curved exponential family by the method of maximum likelihood, this statistic becomes a normalized component of the residual (in the direction normal to the model within the plane spanned by the first two derivatives of the natural parameter for the full exponential family). Furthermore, conditioning on this statistic recovers (in Fisher's sense) the information lost by the MLE, at least approximately. When this conditional distribution is used, the asymptotic variance of the MLE may be estimated by the inverse of observed rather than expected information; in some problems observed information is clearly superior.

This argument, sketched by Pierce and presented in more detail by Efron and Hinkley, represented an attempt to make sense of some of Fisher's remarks on conditioning. In Section 4 of his paper in this volume, Amari presents a comprehensive approach to information recovery as measured by Fisher information. He begins by defining a statistic T to be asymptotically sufficient of order q when

$$ni(\theta) - i^T(\theta) = O(n^{-q+1})$$

and asymptotically ancillary of order q when

$$i^T(\theta) = O(n^{-q}) .$$

These definitions differ from some used by other authors, such as Cox (1980), McCullagh (1984a), and Skovgaard (1985). They are, however, clearly in the spirit of Fisher's apparent feeling that $i^T(\theta)$ is an appropriate measure of information. To analyze Fisher's suggestion that higher derivatives of the loglikelihood function could be used to create successive higher-order approximate ancillary statistics, Amari defines an explicit sequence of combinations of the derivatives: he takes successive components of the residual in spaces spanned by the first p derivatives - of the natural parameter for the ambient exponential family - but perpendicular to the space spanned by the first p-1, then normalizes by higher-order curvatures. In Theorems 4.1 and 4.2 Amari achieves a complete decomposition of the information. He thereby makes specific, justifies, and provides a geometrical interpretation for Fisher's second claim. In Amari's decomposition the p-th term is attributable to the p-th statistic in his sequence and has magnitude equal to $n^{-p+2}$ times the square of the p-th order curvature. (Actually, Amari's treatment is more general than the rough description here would imply since he allows for the use of efficient estimators other than the MLE.)

As far as the basic issue of observed versus expected information is concerned, Amari (1982b) used an Edgeworth expansion involving geometrically interpretable terms (as in Amari and Kumon, 1983) to provide a general motivation for using the inverse of observed information as the estimate of the conditional variance of the MLE. See Section 4.4 of the paper here. (In truth, the result is not as strong as it may appear. When we have an approximation $v_n$ to a variance v satisfying $v(\theta) = v_n(\theta)\{1 + O(n^{-1})\}$, and we use it to estimate $v(\theta)$, we substitute $v_n(\hat{\theta})$, where $\hat{\theta}$ is some estimator of $\theta$, and then all we usually get is $v(\theta) = v_n(\hat{\theta})\{1 + O_p(n^{-1/2})\}$. For essentially this reason, observed information does not in general provide an approximation to the conditional variance of the MLE based on the underlying true value $\theta$, having relative error $O_p(n^{-1})$ - although it does do so whenever expected information is constant, as it is for a location parameter. Similarly, as Skovgaard, 1985, points out in his careful consideration of the role of observed information in inference, when estimated cumulants are used in an Edgeworth expansion it loses its higher-order approximation to the underlying density at the true value. This practical limitation of asymptotics does not affect Bayesian inference, in which observed information furnishes a better approximation to the posterior variance than does expected information for all regular families.)

For curved exponential families, then, the results summarized in the first few sections of Amari's paper provide a thorough geometrical interpretation of the Fisherian concepts of information loss and recovery and also Rao's concept of second-order efficiency. In addition, in section 3.4 Amari discusses the geometry of testing, as had Efron, providing comparisons of several commonly-used test procedures with the locally most powerful test. Curved exponential families were introduced, however, for their mathemetical and conceptual simplicity rather than their applicability. To extend his one-parameter results, Efron, in his 1975 paper, did two things: he noted that any smooth family could be locally approximated by a curved exponential family, and he provided an explicit formula for statistical curvature in the general case. In Section 5 of his paper, Amari shows how results established for curved exponential families may be extended by constructing an appropriate Hilbert bundle, about which I will say a bit more below. With the Hilbert bundle, Amari provides a geometrical foundation, and generalization, for Efron's suggestion. From it, necessary formulas can be derived.

One reason that the role of the mixture curvature in (1) and in the variance decomposition went unnoticed in Efron's paper was that he had not made the underlying geometrical structure explicit: to calculate statistical curvature at a given value $\theta_0$ of a single parameter $\theta$ in a curved exponential family, Efron used the natural parameter space with the inner product defined by Fisher information at the natural parameter point corresponding to $\theta_0$. In order to calculate the curvature at a new point $\theta_1$, another copy of the natural parameter space with a different inner product (namely, that defined by Fisher information at the natural parameter point corresponding to $\theta_1$) would have to be used. The appropriate gluing together of these spaces into a single structure involves three basic elements: a manifold, a Riemannian metric, and an affine connection. Riemannian geometry involves the study of geometry determined by the metric and its uniquely associated Riemannian connection. In his discussion to Efron's paper, Dawid (1975) pointed out that Efron had used the Riemannian metric defined by Fisher information, but that he had effectively used a non-Riemannian affine connection, now called the exponential connection, in calculating statistical curvature. Although Dawid did not identify the role of the mixture curvature in (1), he did draw attention to the mixture connection as an alternative to the exponential connection. (Geodesics with respect to the exponential connection form exponential families, while geodesics with respect to the mixture connection form families of mixtures; thus, the terminology.) Amari, who had much earlier researched the Riemannian geometry of Fisher information, picked up on Dawid's observation, specified the framework, and provided the results outlined above.

The manifold with the associated linear spaces is structured in what is usually called a tangent bundle, the elements of the linear spaces being tangent vectors. For curved exponential families, the linear spaces are finite-dimensional, but to analyze general families this does not suffice so Amari uses Hilbert spaces. When these are appropriately glued together, the result is a Hilbert bundle. The idea stems from Dawid's remark that the tangent vectors can be identified with score functions, and these in turn are functions having zero expectation. As his Hilbert space at a distribution P, Amari takes the subspace of the usual $L_2(P)$ Hilbert space consisting of functions that have zero expectation with respect to P. This clearly furnishes the extension of the information metric, and has been used by other authors as well, e.g., Beran (1977). Amari then defines the exponential and mixture connections and notes that these make the Hilbert bundle flat, and that the inherited connections on the usual tangent bundles agree with those already defined there. He then decomposes each Hilbert space into tangential and normal components, which is exactly what is needed to define statistical curvature in the general setting. Amari goes on to construct an "exponential bundle" by associating with each distribution a finite-dimensional linear space containing vectors defined by higher derivatives of the loglikelihood function, and using structure inherited from the Hilbert bundle. With this he obtains a satisfactory version of the local approximation by a curved exponential family that Efron had suggested.

This pretty construction allows results derived for curved exponential families to be extended to more general regular families, yet it is not quite the all-encompassing structure one might hope for: the underlying manifold is still a particular parametric family of densities rather than the collection of all possible densities on the given sample space. Constructions for the latter have so far proved too difficult.

In his Annals paper, Amari also noted an interesting relationship between the exponential and mixture connections: they are, in a sense he defined, mutually dual. Furthermore, a one-parameter family of connections, which Amari called the $\alpha$-connections, may be defined in such a way that for each $\alpha$ the $\alpha$-connection and the $-\alpha$-connection are mutually dual, while $\alpha=1$ and $-1$ correspond to the exponential and mixture connections. See Amari's Theorem 2.1. This family coincides with that introduced by Centsov (1971) for multinomial distributions. When the family of densities on which these connections are defined is an exponential family, the space is flat with respect to the exponential and mixture connections, and the natural parametrization and mean-value parameterization play special roles: they become affine coordinate systems for the two respective connections and are related by a Legendre transformation. The duality in this case can incorporate the convex duality theory of exponential families (see Barndorff-Nielsen, 1978, and also Section 2 of his paper in this volume). In Theorem 2.2 Amari points out that such a pair of coordinate systems exists whenever a space is flat with respect to an $\alpha$-connection (with $\alpha \neq 0$). For such spaces, Amari defines $\alpha$-divergence, a quasi-distance between two members of the family based on the relationship provided by the Legendre transformation. In Theorem 2.4 he shows that the element of a curved exponential family that minimizes the $\alpha$-divergence from a point in the exponential family parameter space may be found by following the $\alpha$-geodesic that contains the given point and is perpendicular to the curved family. This generates a new class of minimum $\alpha$-divergence estimators, the MLE being the minimum -1-divergence estimator, an interpretation also discussed by Efron (1978).

As applications of his general methods based on $\alpha$-connections on Hilbert bundles, Amari treats the problems of combining independent samples (at the end of section 5), making inferences when the number of nuisance parameters increases with the sample size (in section 6), and performing spectral estimation in Gaussian time series (in section 7).

As soon as the $\alpha$-connections are constructed a mathematical question arises. On one hand, the $\alpha$-connections may be considered objects of differential geometry without special reference to their statistical origin. On the other hand, they are not at all arbitrary. They are the simplest one-parameter family of connections based on the first three moments of the score function. What is it about their special form that leads to the many special properties of $\alpha$-connections (outlined by Amari in Section 2)?

Lauritzen has posed this question and has provided a substantial part of the answer. Given any Riemannian manifold M with metric g there is a unique Riemannian connection $\bar{\nabla}$. Given a covariant 3-tensor D that is symmetric in its first two arguments and a nonzero number c, a new (symmetric) connection is defined by

$$\nabla = \bar{\nabla} + c \cdot D \tag{2}$$

which means that given vector fields X and Y,

$$\nabla_X Y = \bar{\nabla}_X Y + c \cdot \tilde{D}(X,Y)$$

where

$$g(\tilde{D}(X,Y),Z) \equiv D(X,Y,Z)$$

for all vector fields Z. Now, when M is a family of densities and g and D are defined, in terms of an arbitrary parameterization, as

$$g(\partial_i, \partial_j) = E(\partial_i \ell \partial_j \ell)$$

$$D(\partial_i, \partial_j, \partial_k) = E(\partial_i \ell \partial_j \ell \partial_k \ell)$$

where $\ell$ is the loglikelihood function, and if $c = -\alpha/2$, then (2) defines the $\alpha$-connection.

In this statistical case, D is not only symmetric in its first two arguments, as it must be in (2), it is symmetric in all three. Lauritzen therefore defines an abstract statistical manifold to be a triple (M,g,D) in which M is a smooth m-dimensional manifold, g is a Riemannian metric, and D is a completely symmetric covariant 3-tensor. With this additional symmetry constraint alone, he then proceeds to establish a large number of basic properties, especially those relating to the duality structure Amari described. The treatment is "fully geometrical" or "coordinate-free." This is aesthetically appealing, especially to those who learned linear models in the coordinate-free setting. Lauritzen's primary purpose is to show that the appropriate mathematical object of study is one that is not part of the standard differential geometry, but does have many special features arising from an apparently simple structure. He not only presents the abstract generalities about $\alpha$-connections on statistical manifolds, he also examines five examples in full detail. The first is the univariate Gaussian model, the second is the inverse Gaussian model, the third is the two-parameter gamma model, and the last two are specially constructed models that display interesting possibilities of the non-standard geometries of $\alpha$-connections. In particular, the latter two statistical manifolds are not what Lauritzen calls "conjugate symmetric" and so the sectional curvatures do not determine the Riemann tensor (as they do in Riemannian geometry). He also discusses the construction of geodesic foliations, which are decompositions of the manifold and are important because they generate potentially interesting decompositions of the sample space. At the end of his paper, Lauritzen calls attention to several outstanding problems.

Amari's $\alpha$-connections, based on the first three moments of the score function, do not furnish the only examples of statistical manifolds. In his contribution to this volume, Barndorff-Nielsen presents another class of examples based instead on certain "observed" rather than expected derivatives of the loglikelihood.

Although the idea of using observed derivatives might occur to any casual listener on being told of Amari's use of expectations, it is not obvious how to implement it. First of all, in order to define an observed information Riemannian metric, one needs a definition of observed information at each point of the parameter space. Apparently one would want to treat each $\theta$ as if it were an MLE and then use $I(\theta)$. However, $I(\hat{\theta})$ depends on the whole sample $y$ rather than on $\hat{\theta}$ alone, so this scheme does not yet provide an explicit definition. Barndorff-Nielsen's solution is natural in the context of his research on conditionality: he replaces the sample $y$ with a sufficient pair $(\hat{\theta},a)$ where $a$ is the observed value of an asymptotically ancillary statistic $A$. This is always possible for curved exponential families, and in more general models $A$ could at least be taken so that $(\hat{\theta},A)$ is asymptotically sufficient. With this replacement, the second component may be held fixed at $A=a$ while $\hat{\theta}$ varies. Writing $I(\hat{\theta}) = I_{(\hat{\theta},a)}(\hat{\theta})$ thus allows the definition $I(\theta) \equiv I_{(\theta,a)}(\theta)$ to be made at each point $\theta$ in the parameter space. Using this definition of the Riemannian metric, Barndorff-Nielsen derives the coefficients that determine the Riemannian connection. From the transformation properties of tensors, he then succeeds in finding an analogue of the exponential connection based on a certain mixed third derivative of the loglikelihood function (two derivatives being taken with respect to $\theta$ as parameter, one with respect to $\hat{\theta}$ as MLE). In so doing, he defines the tensor $D$ in the statistical manifold and thus arrives at his "observed conditional geometry."

Barndorff-Nielsen's interest in this geometry lies not with analogues of statistical curvature and other expected-geometry constructs, but rather with an alternative derivation, interpretation, and extension of an approximation to the conditional density of the MLE, which had been obtained earlier (in Barndorff-Nielsen and Cox, 1979). In several papers, Barndorff-Nielsen (1980, 1983) has discussed generalizations and approximate versions of Fisher's fundamental density-likelihood formula for location models

$$p(\hat{\theta}\ |a,\theta) = c \cdot L(\theta)/L(\hat{\theta}) \qquad (3)$$

where $\hat{\theta}$ is the MLE, a is an ancillary statistic, p is the conditional density of the MLE, and L is the likelihood function. (This is discussed in Efron and Hinkley, 1978; Fisher actually treated the location-scale case.) The formula is of great importance both practically, since it provides a way of computing the conditional density, and philosophically, since it entails the formal agreement of conditional inference and Bayesian inference using an invariant prior. Inspection of the derivation indicates that the formula results from the transformational nature of the location problem, and Barndorff-Nielsen has shown that a version of it (with an additional factor for the volume element) holds for very general transformation models. He has also shown that for non-transformation models, a version of the right-hand side of (3) while not exactly equal to the left-hand side, remains a good asymptotic approximation for it. (See also Hinkley, 1980, and McCullagh, 1984a.) In his paper in this volume, Barndorff-Nielsen reviews these results, shows how the various observed conditional geometrical quantities are calculated, and then derives his desired expansion (of a version of the right-hand side of (3)) in terms of the geometrical quantities that correspond to those used by Amari in his expected geometry expansions. Barndorff-Nielsen devotes substantial attention to transformation models, which may be treated within his framework of observed conditional geometry. In this context, the models become Lie Groups, for which there is a rich mathematical theory.

In the fourth paper in this volume, Professor Rao returns to the characterization of the information metric that originally led him (and also Jeffreys) to introduce it: it is an infinitesimal measure of divergence based on what is now called Shannon entropy. Rao considers here a more general class of divergence measures, which he has found useful in the study of genetic diversity, leading to a wide variety of metrics. He derives the quadratic and cubic terms in Taylor series expansions of these measures and shows how, in the case of Shannon entropy, the cubic term is related to the $\alpha$-connections.

The papers here collectively show that geometrical structures of statistical models can provide both conceptual simplifications and new methods of analysis for problems of statistical inference. There is interesting mathematics involved, but does the interesting mathematics lead to interesting statistics?

The question arises because geometry has provided new techniques, and its formalism produces convenient summaries for complicated multivariate expressions in asymptotic expansions (as in Amari and Kumon, 1983, and McCullagh, 1984b), but it has not yet created new methodology with clearly important practical applications. Thus, it is already apparent from (1) that there exists a wide class of estimators that minimize information loss (and are second-order efficient): it consists of those having zero mixture curvature for their associated ancillary families. It is interesting that the MLE is only one member of this class, and it is nice to have Eguchi's (1983) derivation that certain minimum contrast estimators are other members, but it seems unlikely - though admittedly possible - that any competitor will replace maximum likelihood estimation as the primary method of choice in practice. Similarly, there is not yet any reason to think that alternative minimum $\alpha$-divergence estimators or their observed conditional geometry counterparts will be considered superior to the MLE.

On the other hand, as I indicated at the outset, geometry does give a definitive description of information loss and recovery. Since Fisher remains our wisest yet most enigmatic sage, it is worth our while to try to understand his pronouncements.** Together with the triumvirate of consistency,

---

**Since Rao's work on second order efficiency arose in an attempt to understand Fisher's computation of information loss in estimation, it might appear that Efron's investigation also began as an attempt to understand Fisher. He has informed me, however, that he set out to define the curvature of a statistical model and came later to its use in information loss and second-order efficiency.

sufficiency, and efficiency, information loss and recovery form the core of Fisher's theory of estimation. On the basis of the geometrical results, it is fair to say that we now know what Fisher was talking about, and that what he said was true. Here, as in other problems (such as inference with nuisance parameters, discussed in Amari's section 5, or in nonlinear regression, e.g., Bates and Watts, 1980, Cook and Tsai, 1985, Kass, 1984, McCullagh and Cox, 1986), the geometrical formulation tends to shift the burden of derivation of results away from proofs, toward definitions. Thus, once the statement of a proposition is understood, its truth is easier to see and in this there is great simplification. One could make this argument about much abstract mathematical development, but it is particularly appropriate here.

Furthermore, there are reasons to think that future work in this area could lead to useful results that would otherwise be difficult to obtain. One important problem that structural research might solve is that of determining useful conditions under which a particular root of the likelihood equation will actually maximize the likelihood. Global results on foliations might be very helpful, as might be formulas relating computable characteristics of statistical manifolds to the behavior of geodesics. The results in these papers could turn out to play a central role in the solution of this or some other practical problem of statistical theory. We will have to wait and see. Until then, readers may enjoy the papers as informative excursions into an intriguing realm of mathematical statistics.

Acknowledgements

I thank O. E. Barndorff-Nielsen, D. R. Cox, and C. R. Rao for their comments on an earlier draft. This paper was prepared with support from the National Science Foundation under Grant No. NSF/DMS - 8503019.

## REFERENCES


Amari, S. (1982a). Differential geometry of curved exponential families - curvatures and information loss. Ann. Statist. 10, 357-387.

Amari, S. (1982b). Geometrical theory of asymptotic ancillarity and conditional inference. Biometrika 69, 1-17.

Amari, S. and Kumon, M. (1983). Differential geometry of Edgeworth expansions in curved exponential family. Ann. Inst. Statist. Math. 35A, 1-24.

Barndorff-Nielsen, O. E. (1978). Information and Exponential Families, New York: Wiley.

Barndorff-Nielsen, O. E. (1980). Conditionality resolutions. Biometrika 67, 293-310.

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. Biometrika 70, 343-305.

Barndorff-Nielsen, O. E. and Cox, D. R. (1979). Edgeworth and Saddlepoint approximations with statistical applications, (with Discussion). J. R. Statist. Soc. B41, 279-312.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of non-linearity. J. R. Statist. Soc. B42, 1-25.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. Ann. Statist. 5, 445-463.

Centsov, N. N. (1971). Statistical Decision Rules and Optimal Inference (in Russian). Translated into English (1982), AMS, Rhode Island.

Cook, R. D. and Tsai, C.-L. (1985). Residuals in nonlinear regression. Biometrika 72, 23-29.

Cox, D. R. (1980). Local ancillarity. Biometrika 62, 269-276.

Dawid, A. P. (1975). Discussion to Efron's paper. Ann. Statist. 3, 1231-1234.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second-order efficiency), (with Discussion). Ann. Statist. 3, 1189-1242.

Efron, B. (1978). The geometry of exponential families. Ann. Statist. 6, 362-376.

Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, (with discussion). Biometrika 65, 457-487.

Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. Ann. Statist. 11, 793-803.

Fisher, R. A. (1925). Theory of statistical estimation. Proc. Camb. Phil. Soc. 22, 700-725.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. Proc. R. Soc. A144, 285-307.

Ghosh, J. K. and Subramaniam, K. (1974). Second order efficiency of maximum likelihood estimators. Sankya 36A, 325-358.

Hinkley, D. V. (1980). Likelihood as approximate pivotal distribution. Biometrika 67, 287-292.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. Proc. Roy. Soc. A186, 453-461.

Kass, R. E. (1984). Canonical parameterizations and zero parameter-effects curvature. J. Roy. Statist. Soc. B46, 1, 86-92.

Madsen, L. T. (1979). The geometry of statistical model - a generalization of curvature. Res. Report 79-1. Statist. Res. Unit, Danish Medical Res. Council.

McCullagh, P. (1984a). On local sufficiency. Biometrika 71, 233-244.

McCullagh, P. (1984b). Tensor notation and cumulants of polynomials. Biometrika 71, 461-476.

McCullagh, P. and Cox, D. R. (1986). Invariants and likelihood ratio statistics. Ann. Statist. 14, 1419-1430.

Pierce, D. A. (1975). Discussion to Efron's paper. Ann. Statist. 3, 1219-1221.

Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc. 37, 81-89.

Rao, C. R. (1961). Asymptotic efficiency and limiting information. Proc. Fourth Berkeley Symp. Math. Statist. Prob., Edited by J. Neyman, 1, 531-545.

Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion). J. Roy. Statist. Soc. B24, 46-72.

Rao, C. R. (1963). Criteria of estimation in large samples. Sankya 25, 189-206.

Reeds, J. (1975). Discussion to Efron's paper. Ann. Statist. 3, 1234-1238.

Skovgaard, I. (1985). A second-order investigation of asymptotic ancillarity. Ann. Statist. 13, 534-551.