The Annals of Statistics
1987, Vol. 15, No. 1, 346–359

CONFIDENCE REGIONS IN CURVED EXPONENTIAL FAMILIES: APPLICATION TO MATCHED CASE-CONTROL AND SURVIVAL STUDIES WITH GENERAL RELATIVE RISK FUNCTION[1]

BY SURESH H. MOOLGAVKAR AND DAVID J. VENZON

Fred Hutchinson Cancer Research Center

Differential geometric methods are used to construct approximate confidence regions for curved exponential families. The α-connection geometries discussed by Amari (1982), and another geometry introduced here, the c geometry, are exploited to construct confidence regions. Survival and case-control studies with general relative risk functions are interpreted in the context of curved exponential families, and an example illustrates the construction of confidence regions for matched case-control studies. Simulations indicate that the geometric procedures have good coverage and power properties.

1. Introduction. The goal of this paper is to set out a systematic general approach to the construction of approximate confidence regions for hypotheses in multiparameter exponential families. It is well known that when the hypothesis is nonlinear in the space of natural parameters, i.e., when, in the terminology of Efron, the hypothesis forms a curved exponential family, confidence regions based on the Wald statistic may be seriously misleading. A discussion of curvature in the case of nonlinear regression models and its effect on confidence regions can be found in papers by Beale (1960), Bates and Watts (1980, 1981), and Hamilton, Bates and Watts (1981).

Suppose that Y is a random variable belonging to the n-parameter exponential family. The density of Y can be written

$$h(y,\theta) = \exp\{\theta^t y - \psi(\theta) + \chi(y)\},$$

where $\theta \in \Omega \subset \mathscr{R}^n$ is the vector of natural parameters. Let $\beta \in \mathscr{R}^k$, $k < n$, be the parameters of interest and suppose that $\theta = f(\beta)$ and that rank f (i.e., rank of the Jacobian matrix $(\partial f/\partial\beta)$) is k. Then, in general, $f(\beta)$ is a curved exponential family, and Efron (1975) defined a curvature of embedding for this family when β is a scalar parameter, i.e., when $k = 1$.

The notion of curvature of embedding can be extended to the case $k > 1$. The appropriate concept from differential geometry is that of second fundamental form. In fact, it is possible to impose on the θ space, and therefore, on the locus $f(\beta)$, a one-parameter family of geometries, via the so called α connections (see

Received November 1984; revised April 1986.
[1]This work was supported by USPHS grant CA39949 from the National Institutes of Health.
AMS 1980 *subject classifications*. Primary 62E20; secondary 62B10.
Key words and phrases. Alpha connections, differential geometry, inference, logistic regression, partial likelihood, variance-stabilizing parametrizations.

Amari (1982); Lauritzen, (1984)). Efron's geometry corresponds to the geometry defined by $\alpha = 1$ in Amari's construction.

Likelihood-based confidence regions have the desirable property of being invariant under parameter transformations. In the single parameter case, Sprott (1973) considered specific transformations that symmetrized the likelihood, and argued that Wald intervals be constructed after such reparametrization. In the multiparameter case, suitable reparametrizations are much more difficult to find.

Our aim in this paper is to exploit differential geometry to construct approximate confidence regions in curved exponential families. Detailed simulations, to be described elsewhere, indicate that the geometric procedures have good coverage and power properties. We will be concerned with reparametrizing the model so that it looks as much like "uncurved" Euclidean space as possible. For each α geometry, such a reparametrization is accomplished by using geodesic normal coordinates. For some of the α, geodesic normal coordinates have appealing statistical interpretations as is discussed in the next section.

In addition to the α geometries, it seems natural to consider another geometry in the context of hypothesis testing. Let $\eta_\theta = E_\theta(Y) = \dot{\psi}(\theta)$ be the expectation of Y and $\Sigma_\theta = \ddot{\psi}(\theta)$ be the variance of Y. Note also that Σ_θ is the Fisher information matrix for θ. Let $\theta_0 = f(\beta_0)$ be a fixed parameter. Then, since Σ_{θ_0} is positive definite, it defines an inner product on the θ space. The embedding $f\colon \mathscr{R}^k \to \mathscr{R}^n$ then defines an inner product on $\mathscr{R}^k$ by the pull-back $G_\beta = J_\beta^t \Sigma_{\theta_0} J_\beta$, where J_β is the Jacobian of f at β. Note that G_{β_0} is the Fisher information matrix at β_0. This inner product then defines a Riemannian geometry on β space, and reflects "local" curvature near β_0. We call this the constant metric geometry or the c geometry on β space. The relationship of this geometry to the α geometries is briefly discussed in the concluding remarks.

In Section 2, we briefly review some relevant notions from differential geometry. The construction of geodesic normal coordinates requires the solution of systems of quasilinear differential equations. We provide easily computable approximations. The concepts introduced here are intimately related to the concepts of curvature described by Efron (1975, 1978) and Bates and Watts (1980). Some of these relationships have been explored in a recent paper by Kass (1984), and others are discussed briefly in the concluding remarks. The goal of this paper is the rather practical one of demonstrating that geometry may be exploited to construct adequate approximate confidence regions. In Sections 3 and 4, we discuss some of the statistical properties of the 0 geometry and the c geometry. Of particular interest is the relationship between the 0 geometry and variance stabilization. In Section 5, we show how the probability models of matched case-control and survival studies can be considered as curved exponential families. In Section 6, we consider a detailed example. In Section 7, we summarize some simulation results.

2. Geodesic coordinates and confidence regions. We give a quick introduction here to geodesics and normal coordinate neighborhoods. The reader interested in the details should consult the books by Boothby (1975), Milnor

(1963) or Hicks (1965). We begin with a discussion of the c geometry. The appropriate modifications necessary for the α geometries are described later.

The inner product $G = G_\beta$ imposes a Riemannian geometry on $\mathscr{R}^k$. We need to find the "straight" lines, or geodesics, for these geometries. Let g_{ij} be the entries of G. The Riemann–Christoffel symbols are defined by

$$\overset{c}{\Gamma}{}^{i}_{jm} = \sum_{l} \tfrac{1}{2}(\partial_j g_{ml} + \partial_m g_{jl} - \partial_l g_{jm}) g^{il}, \tag{1}$$

where, for simplicity of notation, ∂_p denotes $\partial/\partial\beta_p$, and g^{il} are the entries of G^{-1}. Here, the c on the Γ denotes that we are working in the c geometry. A geodesic in the geometry defined by G is then a curve $\beta(t)$ that satisfies the system of quasilinear differential equations

$$\frac{d^2\beta_i}{dt^2} + \sum_{j,m} \overset{c}{\Gamma}{}^{i}_{jm} \frac{d\beta_j}{dt} \cdot \frac{d\beta_m}{dt} = 0, \qquad i = 1, \dots, k. \tag{2}$$

Given a point $\beta_0 \in \mathscr{R}^k$ and a tangent vector v at β_0, the above system of differential equations has, for t small enough, a unique solution $\beta(t)$ such that $\beta(0) = \beta_0$ and $\dot\beta(0) = v$. That is, for any point β_0 and any direction and velocity, there is a unique geodesic starting at β_0 in that direction, and with prescribed velocity. Under some regularity conditions, the geodesic can be defined for all t.

In order to introduce geodesic coordinates, we now make a distinction between Euclidean β space, and β space equipped with the geometry from G. The latter space will be labelled M. At each point of M, the tangent space at that point can be identified with $\mathscr{R}^k$. For any point $\beta_0 \in M$, let $\gamma_v(t)$ be the geodesic starting at β_0 and satisfying the condition $\dot\gamma_v(0) = v$. Then, there exists an open neighborhood V of 0 in $\mathscr{R}^k$ and an open neighborhood W of β_0 in M such that the map Exp: $V \to W$ defined by $\text{Exp}(v) = \gamma_v(1)$ is a diffeomorphism. The map Exp is called the exponential map, and defines the geodesic coordinate system at β_0.

Now, if β_0 is the maximum likelihood estimate, an approximate confidence region for β may be constructed as follows. Let C be a Wald confidence region in the tangent space to M at β_0. That is, $C = \{v \in \mathscr{R}^k | v^t G_{\beta_0} v \le c\}$, where c is some positive constant. Then an approximate confidence region R for β is the image under Exp of the elliptical region C, $R = \text{Exp}(C)$. If one wanted a confidence region based on the observed information matrix rather than the Fisher information matrix, the region C may be appropriately redefined with G_{β_0} replaced by the observed information matrix. In either case, obtaining $\text{Exp}(C)$ requires the solution to a system of quasilinear differential equations, which can be quite cumbersome in practice. In many situations, the approximation given below should suffice. The confidence region $\text{Exp}(C)$ may be thought of as follows. First, β space is reparametrized in terms of the geodesic normal coordinates, the Wald confidence region is found in this system of coordinates, and then this region is reexpressed in terms of the original coordinates.

From the definition of Exp, it is clear that the Jacobian of Exp at 0 is the identity. Second derivatives may be computed from the differential equations for the geodesics. Then, an approximation to Exp is given by the first few terms of the Taylor series

$$[\mathrm{Exp}(v)]^i \approx \beta_0^i + v^i - \tfrac{1}{2}\sum_{j,m} \overset{c}{\Gamma}{}^i_{jm}(\beta_0)v^j v^m, \tag{3}$$

where u^i represents the ith component of the vector u. The first two terms of this approximation correspond to the usual Wald confidence region for β. Thus, this procedure can be thought of as refining the usual procedure when curvature is present. Additional terms can be computed if necessary.

At the risk of belaboring the obvious, we point out here that if f is linear, all the Riemann–Christoffel symbols are zero, the geodesics are the straight lines and one recovers the usual confidence region.

The entire discussion here was carried out in terms of the geometry determined by G, i.e., in terms of the c geometry. An entirely parallel development can be based on the α geometries. Geodesics are defined as solutions to the quasilinear equations (2) whether or not the geometry is metric (Riemannian). For the α geometries, the Riemann–Christoffel symbols may be defined as follows.

The Fisher information Σ_θ defines an inner product, and hence, a Riemannian geometry, on θ space. This is called the information or 0 geometry. Then for any real α, the α geometry can be defined by means of a connection that is the sum of the information connection and $-(\alpha/2)T$, where T is an appropriately defined tensor (see Amari (1982) for details). Now this allows the α connections to be easily computed on β space. Let $\tilde{G}_\beta = J_\beta^t \Sigma_{\theta_\beta} J_\beta$ be the pull-back of the Fisher information matrix to β space. Note that $\tilde{G}_\beta$ is the Fisher information matrix at β, and thus defines the information or 0 geometry on β space. The Riemann–Christoffel symbols for this geometry are defined as in (1), with the entries of $\tilde{G}$ and $\tilde{G}^{-1}$ replacing the entries of G and G^{-1} in that expression. Then, the Riemann–Christoffel symbols for the α-connection are given by

$$\overset{\alpha}{\Gamma}{}^i_{jm} = \overset{0}{\Gamma}{}^i_{jm} - \frac{\alpha}{2}\sum_l \left[\sum_{a,b,c} (\partial_l\theta_a\, \partial_j\theta_b\, \partial_m\theta_c)\frac{\partial^3\psi(\theta)}{\partial\theta_a\,\partial\theta_b\,\partial\theta_c}\right]\tilde{g}^{il}, \tag{4}$$

where as before ∂_l is $\partial/\partial\beta_l$, etc. Here, the first term on the right-hand side represents the Riemann–Christoffel symbols of the 0 connection on β space and the second term on the right-hand side represents the pull-back of the tensor T from θ space to β space. Note that $(\partial^2\psi(\theta))/\partial\theta_b\,\partial\theta_c$ are just the entries of the Fisher information matrix, Σ_θ, in θ space.

When f is linear, i.e., when β space is an affine subspace of θ space, the c geometry coincides with the 1 geometry. However, in general these geometries are different. Further, even if f is linear, the Riemann–Christoffel symbols are nonzero in all the α geometries with the exception of $\alpha = 1$. Linear hypotheses are flat submanifolds of θ space only in the 1 geometry (equivalently in the c geometry). Thus, even for linear hypotheses, it is of some interest to construct confidence regions based on the α geometries, $\alpha \neq 1$.

When $k = 1$, i.e., when the model is one-dimensional, there is a single Riemann–Christoffel symbol for any of the geometries. For example, for the 0 geometry it is given by the expression

$$\tilde{\Gamma} = \frac{1}{2}\left(\frac{\partial \tilde{g}}{\partial \beta}\right)\tilde{g}^{-1} = \frac{\partial(\ln \tilde{g}^{1/2})}{\partial \beta}. \tag{5}$$

The system of quasilinear equations (2) reduces to

$$\frac{d^2\beta}{dt^2} + \tilde{\Gamma}\left(\frac{d\beta}{dt}\right)^2 = 0. \tag{6}$$

Using expression (5) for $\tilde{\Gamma}$, equation (6) can be easily integrated once to yield the first-order equation

$$\frac{d\beta}{dt} = b\tilde{g}^{-1/2}, \tag{7}$$

where b is the constant of integration. The initial condition $d\beta/dt|_{t=0}$, of course, determines b. In order to construct the upper and lower bounds of the $1 - \varepsilon$ confidence interval the constants b are chosen, respectively, to be the upper and lower $\varepsilon/2$ points of the $\mathcal{N}(0,1)$ distribution. Thus, to obtain the 95% confidence interval, for example, $b = \pm 1.96$. Equation (7) can be integrated numerically from $t = 0$ to $t = 1$ to yield the desired confidence interval. Alternatively, approximation (3) may be used. If necessary, one more term in the Taylor series expansion can be easily found. With appropriate modifications, the entire discussion above applies to the c geometry.

In order to construct a confidence interval for a single parameter in a vector of parameters, we use the notion of profile likelihood. Suppose, without loss of generality, that a confidence interval is to be constructed for β_1. Consider the function $\tilde{l}(\beta_1) = l(\beta_1, \hat{\beta}_{-1}(\beta_1))$, where $\hat{\beta}_{-1}(\beta_1) = (\hat{\beta}_2(\beta_1)), \dots, \hat{\beta}_k(\beta_1))$ is that value of $(\beta_2, \dots, \beta_k)$ that maximizes the likelihood for a given value of β_1. The mapping $\beta_1 \mapsto (\beta_1, \hat{\beta}_{-1}(\beta_1))$ now becomes the (one-dimensional) model of interest, and a confidence interval can be constructed as before once the Riemann–Christoffel symbol is computed. Note that because of the tensorial property of the Fisher information, $\tilde{g}$ is computed as

$$\tilde{g} = J_{\beta_1}^t \tilde{G}_\beta J_{\beta_1},$$

where $J_{\beta_1} = (1, \partial[\hat{\beta}_{-1}(\beta_1)]/\partial\beta_1)^t$ is the Jacobian of the mapping $\beta_1 \mapsto (\beta_1, \hat{\beta}_{-1}(\beta_1))$. An easy calculation shows that the vector $\partial(\hat{\beta}_{-1}(\beta_1))/\partial\beta_1$ is given by $-A^{-1}w$, where A is the $(k-1) \times (k-1)$ matrix $(\partial^2 l/\partial\beta_i\, \partial\beta_j)_{i,j>1}$, and w is the $(k-1)$ vector with entries $\partial^2 l/\partial\beta_1\, \partial\beta_j$, with $j = 2, 3, \dots, k$. Similarly, for the c geometry, g is obtained by replacing $\tilde{G}$ by G in the preceding expression. The tensor T can also be pulled back to yield the α geometries, $\alpha \neq 0$. The Riemann–Christoffel symbols for the 0 geometry and the c geometry are constructed as in expression (5).

The small sample properties of these procedures need to be thoroughly investigated by Monte Carlo simulation. We did extensive simulations with the proportional hazards model for survival data, the 0 and c geometries and the profile likelihood method previously described for construction of confidence intervals. These simulations will be described in detail elsewhere (Moolgavkar and Venzon (1986)). A brief summary of the results is given in Section 7.

3. The 0 geometry and variance stabilization. Vaeth (1985) has noted that Wald confidence regions are most appropriate in variance-stabilizing parametrizations. It is well known that when $k > 1$, a variance-stabilizing parametrization does not, in general, exist. In fact, a variance-stabilizing parametrization exists if and only if there exists a 1-1 transformation $\delta \mapsto \beta$, such that the Fisher information matrix in δ space, $\tilde{G}_\delta = J_\delta^t G_{\beta(\delta)} J_\delta$, is constant. If $k = 1$, such a transformation always exists and, in fact, its existence is simply a restatement of the fact that a curve can always be parametrized by its arc length. When $k > 1$, a very stringent condition must be satisfied for a variance-stabilizing transformation to exist: A tensor constructed from $\tilde{G}_\beta$, the Riemannian curvature tensor, must be equal to zero.

EXAMPLE 1. Consider a univariate normal family with unknown mean and variance. This family has nonzero Riemannian curvature tensor. In fact, Amari (1982) has shown that this family has constant negative (sectional) curvature. A consequence is that a variance-stabilizing parametrization of this family does not exist. It is of interest to note that this family is isometric to the hyperbolic or Lobachevski plane. In this geometry, Euclid's fifth postulate is violated: Given a straight line (geodesic) and a point not on the straight line, there exists an infinite number of straight lines passing through the point and not intersecting the given line.

EXAMPLE 2. Consider a multinomial family $M(m, p_1, \dots, p_n)$, m fixed. This family is isometric to an open subset of the sphere of radius $2m^{1/2}$ in $\mathscr{R}^{n+1}$, and hence has constant positive (sectional) curvature. In fact, we show this by constructing an explicit isometry, i.e., a map f to the sphere of radius $2m^{1/2}$ such that $J^t I J = V^{-1}$, where J is the Jacobian matrix of f, I is the $(n + 1) \times (n + 1)$ identity matrix and V is the covariance matrix of the multinomial family. The map f is given explicitly by $(p_1, \dots, p_n) \mapsto m^{1/2}(-2p_0^{1/2}, 2p_1^{1/2}, \dots, 2p_n^{1/2})$, where we note that $p_0 = 1 - p_1 - \cdots - p_n$.

An immediate consequence is that there exists no submodel of dimension larger than 1 with a variance-stabilizing parametrization. For case-control and survival studies we shall be dealing with Cartesian products of multinomial distributions (see the following). The conditions under which variance-stabilizing parametrizations of submodels exist in this case can be worked out after computing the Riemann–Christoffel symbols. However, we do not discuss this further here.

Although variance-stabilizing parametrizations do not, in general, exist, geodesic normal coordinates in the 0 geometry are almost variance-stabilizing in the sense of the following geometric proposition, which was known to Riemann.

PROPOSITION. *Let $(u_1, \dots, u_k)$ define geodesic normal coordinates centered at β_0; i.e., with $\beta_0 = (0, \dots, 0)$. Let h_{ij} be the entries of the matrix G (or $\tilde{G}$) in this coordinate system. Then, $h_{ii} = c_i$, where c_i is constant along u_i and further $(\partial h_{jd}/\partial u_i)_{(0,\dots,0)} = 0, \forall i, j, d$.*

PROOF. The first part of the proposition follows immediately from the definition of geodesic normal coordinates and the fact that the tangent vectors to a geodesic have constant lengths. For the second part, note that the equations of geodesics through β_0 take the form $u_i(t) = a_i t$; $i = 1, 2, \dots, k$, a_i arbitrary constants. Thus, equations (2) for geodesics through β_0 become

$$\sum_{j,d} \Gamma^i_{jd} a_j a_d = 0,\ i = 1, \dots, k \Rightarrow \Gamma^i_{jd}(0) = 0 \quad \forall i, j, d.$$

Now, it follows from the definition of the Riemann–Christoffel symbols that

$$\frac{\partial h_{jd}}{\partial u_i} = \sum_l \Gamma^l_{ij} g_{ld} + \sum_l \Gamma^l_{id} g_{lj} \Rightarrow \left(\frac{\partial h_{jd}}{\partial u_i}\right)_{(0,\dots,0)} = 0 \quad \forall i, j, d.$$

This proves the second part of the proposition. □

This proposition makes it clear in what sense the geodesic coordinate system in the 0 geometry is almost variance-stabilizing. The diagonal entries of the Fisher information matrix are constant along coordinate curves, and the off-diagonal entries are constant to first order at β_0.

The above proposition says, in particular, that any geodesic in the 0 geometry, when considered as a one-dimensional model, is variance-stabilizing. The α geodesics, when $\alpha = -1, -\frac{1}{3}$ or $\frac{1}{3}$ can likewise be interpreted as (one-dimensional) parametrizations that reduce asymptotic bias, reduce asymptotic skewness and make the expected third derivative of the log-likelihood zero, respectively (Hougaard (1982); Kass (1984)). Thus, confidence regions based on these geometries are of interest. However, the second term in expression (4) is computationally cumbersome, and our main interest is in the 0 geometry and the c geometry.

4. The *c* geometry and Wald regions. Consider the Wald regions W_θ and W_β in θ space and β space, respectively, defined by

$$W_\theta = \left\{\theta | [\theta - \theta(\beta_1)]^t \Sigma_{\theta(\beta_1)} [\theta - \theta(\beta_1)] \le \text{constant}\right\}$$

and

$$W_\beta = \left\{\beta | (\beta - \beta_1)^t G_{\beta_1} (\beta - \beta_1) \le \text{constant}\right\},$$

where $G_{\beta_1} = J^t_{\beta_1} \Sigma_{\theta(\beta_1)} J_{\beta_1}$, for some point β_1 in β space. If the embedding of β space in θ space is linear, i.e., if $\theta = f(\beta)$ is a linear function, then $f(W_\beta) = W_\theta \cap f(\beta)$. This is not true in general. Suppose, however, that β represents the geodesic normal coordinates in the c geometry defined by $\Sigma_{\theta(\beta_1)}$. Then, by the construction of these coordinates, $(\beta_2 - \beta_1)^t G_{\beta_1} (\beta_2 - \beta_1)$ is the squared length (in θ space) of the curve $f((1-t)\beta_1 + t\beta_2)$, $0 \le t \le 1$, joining β_1 and β_2, and is thus approximately equal to

$$[\theta(\beta_2) - \theta(\beta_1)]^t \Sigma_{\theta(\beta_1)} [\theta(\beta_2) - \theta(\beta_1)],$$

if the embedding of β space into θ space is not too badly behaved. Consequently, $f(W_\beta)$ is approximately the intersection of W_θ with the submanifold $f(\beta)$.

5. Matched case-control and survival studies. The conditional likelihood of the logistic regression model for matched case-control studies is formally identical to the partial likelihood of the proportional hazards model for survival data. In either case, the appropriate likelihood function (Thomas (1981)) is

$$L = \prod_{i=1}^{n} \frac{\rho(\beta, z_{i0})}{\sum_{j=0}^{m_i} \rho(\beta, z_{ij})},$$

where n is the number of risk sets, $\rho(\beta, z)$ is a generalized relative risk function and z_{ij} is a vector of covariates for the jth individual in risk set i, with $j = 0$ corresponding to the case or the individual who failed. In this expression, m_i is the number of controls for the ith case or the number of individuals whose survival time exceeds the ith failure time. For the usual logistic regression model and the original proportional hazards model proposed by Cox (1972), $\rho(\beta, z) = \exp(\beta^t z)$.

The likelihood L can be viewed as arising from a multinominal sampling scheme as follows. Let R_i, $i = 1, \ldots, n$, be a multinominal random variable with sample size 1 and cell probabilities

$$P_{ij} = \frac{\rho(\beta, z_{ij})}{\sum_{l=0}^{m_i} \rho(\beta, z_{il})}.$$

Then L arises as the likelihood of this multinominal sampling scheme if the success for each R_i is associated with P_{i0}.

Now for each R_i, the natural parameter

$$\begin{aligned}\theta_i^t(\beta) &= \left(\ln\left(\frac{P_{i1}}{P_{i0}}\right), \ldots, \ln\left(\frac{P_{im_i}}{P_{i0}}\right)\right) \\ &= \left(\ln\left(\frac{\rho(\beta, z_{i1})}{\rho(\beta, z_{i0})}\right), \ldots, \ln\left(\frac{\rho(\beta, z_{im_i})}{\rho(\beta, z_{i0})}\right)\right).\end{aligned}$$

Let $\theta^t = (\theta_1^t, \ldots, \theta_n^t)$. Then the locus $\theta(\beta)$ defines, in general, a curved subfamily of the exponential family with natural parameter θ and covariance (Fisher information) matrix Σ_θ, where Σ_θ is block diagonal with the ith block corresponding to the covariance matrix of R_i. In matched case-control studies, the matrix $(\partial\theta/\partial\beta)^t \Sigma_{\theta_\beta} (\partial\theta/\partial\beta)$ is the Fisher information matrix in β space. This is not true for survival studies. However, this matrix has the appropriate asymptotic properties (as the number of failures, i.e., the number of risk sets, approaches infinity) for inference and may be used to construct Wald type confidence regions (Prentice and Self (1983)). It is clear that $\theta = X\beta$ (i.e., θ is a linear function of β) if and only if $\rho(\beta, z) = \exp(\beta^t z)$. In this instance, the entries of the design matrix X are the differences in covariate values between controls and the case in the risk set. Thus, when $\rho(\beta, z) = \exp(\beta^t z)$, the model is uncurved in the sense of Efron (i.e., in the $\alpha = 1$ or the c geometry).

There has been some recent interest in relative risk functions other than the exponential, e.g., in $\rho(\beta, z) = 1 + \beta^t z$. With this relative risk function it is known that inference for β based on the Wald statistic is seriously misleading (Storer, Wacholder and Breslow (1983); Lustbader, Moolgavkar and Venzon (1984)). The considerations of Section 2 may be used to construct approximate confidence regions in such situations. In the next section we illustrate the procedure with a detailed example.

6. An example. In this section, we illustrate the construction of approximate confidence regions for a matched case-control study of endometrial cancer in Los Angeles. These data are taken from Appendix III of Breslow and Day

TABLE 1

Results of the analysis of the example in Section 5. Variance stabilized intervals obtained by numerical integration of (7) using the profile likelihood (see text).

Results of analyses with $\rho(\beta,z) = 1 + \beta_1 z_1 + \beta_2 z_2$

mle = $\hat{\beta}$ = (4.263, 0.122); maximized log likelihood = −70.53

Expected covariance matrix = $\tilde{G}_{\hat{\beta}}^{-1} = \begin{pmatrix} 9.580 & 0.072 \\ 0.072 & 0.004 \end{pmatrix}$

95% confidence intervals for:

	β_1	β_2
Wald based	(−1.803, 10.329)	(0.005, 0.239)
Likelihood based	(0.502, 16.67)	(0.042, 0.374)
Variance stabilized (0 geometry)	(0.545, 16.11)	(0.047, 0.347)

Results of analyses with $\rho(\beta,z) = \exp(\beta_1 z_1 + \beta_2 z_2)$

mle = $\hat{\beta}$ = (1.06302; 0.02070); maximized log likelihood = −73.10

Expected covariance matrix = $\tilde{G}_{\hat{\beta}}^{-1} = \begin{pmatrix} 0.18696 & 0.000016 \\ 0.000016 & 0.000019 \end{pmatrix}$

95% confidence intervals for:

	β_1	β_2
Wald based	(0.216, 1.911)	(0.012, 0.029)
Likelihood based	(0.206, 1.918)	(0.012, 0.030)
Variance stabilized (0 geometry)	(0.196, 1.928)	(0.012, 0.030)

(1980), and are also available from us upon request. There are 63 cases (63 risk sets) with 4 controls per case in that data set. Two covariates, one of them discrete (presence (1) or absence (0) of gallbladder disease), and the other continuous (length of estrogen use in months, range: 0 to 96 months), were chosen for analysis. We eliminated 6 risk sets because these had a missing covariate value for the case. In addition, we eliminated 1 control from each of 8 other risk sets because a covariate value was missing on that control. Thus, we analyzed a total of 277 observations arranged in 49 risk sets consisting of a case and 4 controls each, and 8 risk sets consisting of a case and 3 controls each. Two relative risk functions were used in the analyses: $\rho(\beta, z) = 1 + \beta_1 z_1 + \beta_2 z_2$ and $\rho(\beta, z) = \exp(\beta_1 z_1 + \beta_2 z_2)$. Recall that the latter formulation defines a linear hypothesis (cf. last section) in an exponential family.

The results of the analyses are presented in Table 1. With $\rho(\beta, z) = 1 + \beta_1 z_1 + \beta_2 z_2$, the null hypothesis $\beta = 0$ is resoundingly rejected by the likelihood ratio test (38.8 on two degrees of freedom), whereas it is not rejected by the Wald test (4.56 on two degrees of freedom) at the 0.05 level of significance. Figure 1 shows that the 95%-likelihood-based joint confidence region is far from elliptical as is noted in a recent publication (Lustbader, Moolgavkar and Venzon (1984)). The correction based on the 0 geometry is shown in that figure. We note that the corrected Wald region is in excellent agreement with the likelihood-based region.

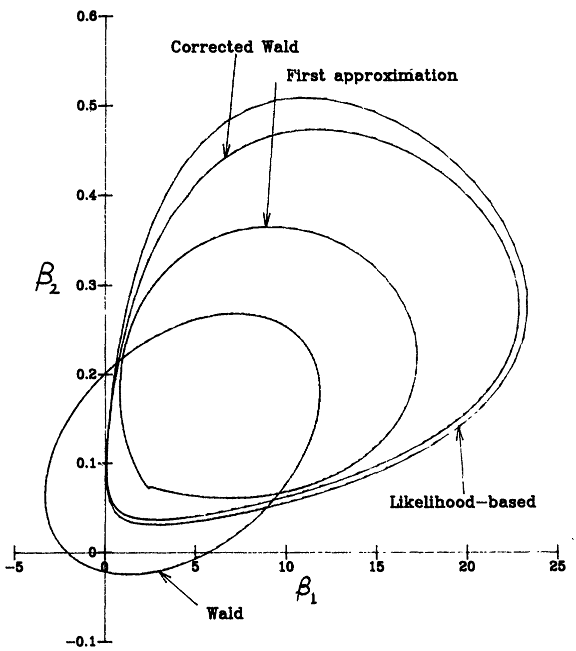

FIG. 1. *95% confidence regions for the example in Section 6. The corrected Wald region was obtained by numerical integration of the equations for geodesics in the* 0 *geometry. The first approximation was obtained by using expression* (3) *for the* 0 *geometry.*

However, the approximation based on (3) seems to be inadequate for this example. With $\rho(\beta, z) = \exp(\beta_1 z_1 + \beta_2 z_2)$, all the methods for the construction of a confidence region yield similar results.

The profile likelihood was used, as previously described, to compute confidence intervals based on the 0 geometry, i.e., to compute variance-stabilized confidence intervals. The likelihood-based confidence intervals were computed as suggested by Cox ((1970), page 88). Thus, for example, to compute the confidence interval for β_1, consider the profile likelihood $l(\beta_1, \hat{\beta}_2(\beta_1))$. A $1 - \varepsilon$ confidence interval is then defined by

$$\{\beta_1 | l(\hat{\beta}_1, \hat{\beta}_2) - l(\beta_1, \hat{\beta}_2(\beta_1)) \le \tfrac{1}{2}c\},$$

where c denotes the upper ε point of the chi-squared distribution on one degree of freedom (Table 1).

7. Some simulation results. Extensive simulations were carried out to investigate how well some of these methods did in the analysis of survival data via the proportional hazards model. The details of the simulations appear elsewhere (Moolgavkar and Venzon (1986)). Covariates were sampled once from each of three different distributions (uniform, normal, lognormal) and then used for all the replicates. With an additive form of the relative risk $\rho(\beta, z) = 1 + \beta z$,

TABLE 2

Results of simulating survival data with the proportional hazards model, no censoring and with relative risk $= 1 + \beta z$. *Covariate distribution was chosen to be uniform, lognormal, or normal, and* 50 *failures were recorded. All results are based on* 1000 *replicates. Coverage reports the proportion of computed* 95% *confidence intervals that included the true value of* β. *Power reports the proportion of computed* 95% *confidence intervals that excluded* 0.

	Coverage				Power			
β	Wald	l-based	Var. stab. (0 geom.)	c geom.	Wald	l-based	Var. stab	c geom.
			Covariate distribution: Uniform					
0	0.914	0.956	0.960	0.956	0.086	0.044	0.040	0.044
2.5	0.896	0.958	0.950	0.952	0	0.658	0.640	0.645
5.0	0.892	0.961	0.957	0.960	0	0.900	0.906	0.896
10.0	0.874	0.977	0.966	0.964	0	0.985	0.987	0.982
			Covariate distribution: Lognormal					
0	0.919	0.954	0.946	0.955	0.081	0.046	0.054	0.045
2.5	0.928	0.932	0.936	0.945	0	0.498	0.438	0.534
5.0	0.916	0.944	0.954	0.959	0	0.828	0.802	0.842
10.0	0.904	0.960	0.956	0.962	0	0.972	0.968	0.980
			Covariate distribution: Normal					
0	0.898	0.958	0.966	0.962	0.102	0.042	0.034	0.038
2.5	0.880	0.962	0.949	0.961	0.002	0.418	0.404	0.374
5.0	0.850	0.968	0.960	0.961	0.002	0.655	0.651	0.618
10.0	0.780	0.966	0.978	0.946	0	0.821	0.830	0.798

the coverage and power, defined as the probability of excluding zero, were computed for the Wald interval, the likelihood-based interval, and the intervals based on the 0 geometry and the c geometry. The Wald region behaved poorly with respect to both the coverage and power. This was especially noticeable with a normal covariate distribution. The other intervals had good coverage and power. In order to compute the geometric intervals, equation (7) was numerically integrated using the Adams–Gear algorithm. The approximation given in expression (3) was not evaluated. The results of one of the simulations are presented in Table 2. Simulations were also carried out with two covariates, one continuous and the other binary. Again, the likelihood-based and the geometric intervals had good coverage and power properties, whereas the Wald interval behaved poorly.

8. Concluding remarks. In this paper, we have discussed reparametrizations based on geometric considerations. Two of these, based on the 0 geometry and on the c geometry, are easy to implement. Geodesic normal coordinates in the 0 geometry have a natural and appealing statistical interpretation, and Wald confidence regions in these coordinates appear to have good coverage and power properties. The c geometry measures local curvature near the mle. Its statistical properties remain to be investigated. However, confidence regions based on this geometry have good properties. We note that the procedure described in this paper for the construction of confidence regions consists of transforming Wald regions from a canonical coordinate system, the geodesic coordinate system, which is unique for a given geometry. Thus, like the likelihood-based regions, these confidence regions are parametrization invariant, whereas the Wald region and the approximation based on (3) are not. This is of particular relevance to the example considered in the paper. For a binary covariate, the assignment of a 1 or 0 to one of two exposure groups is entirely arbitrary. With an additive relative risk, a change in the assignment corresponds to a nonlinear transformation of the parameters. Thus, while the likelihood-based and the geometrically constructed regions are invariant under the choice of assignment, the Wald region and the region based on (3) are not.

A few words on the relationship among the connections defined here and those defined by Amari (1982), Kass (1984) and Lauritzen (1984) are in order. The exponential or 1 connection on θ space is not usually viewed as a Riemannian connection. However, it is a Riemannian connection when θ space is endowed with a constant inner product. In fact, Efron's (1975) computations of the curvature in the single parameter case may be viewed in this way: To compute the curvature at β_0, pretend that θ space is endowed with the constant inner product Σ_{θ_0}, and then compute the geodesic curvature as usual in Euclidean geometry. An alternative point of view as in Amari (1982) and Lauritzen (1984) leads to the definition of "geodesic curvature" in a non-Riemannian manifold. This involves measuring lengths in a metric that is not compatible with the connection.

Now, there are two different ways in which connections, and therefore geometries, can be induced on β space. First, the c connection and the 0 connection may be induced via the pull-backs of the appropriate inner products as done in this paper. Convex combinations of these connections then give rise to a one-parameter family of geometries on β space. Second, a one-parameter family of connections may be constructed on β space via the log-likelihood function as in Amari (1982). This construction is equivalent to considering β space as an embedded submanifold of θ space and projecting the α connection of θ space onto β space via the Fisher information matrix as suggested by Lauritzen (1984). The 0 connections defined in these two ways are identical, although the other connections are, in general, different. In addition to the α geometries, it seems natural to consider the c geometry described in this paper. This geometry, based on G, is simply the geometry that the β space inherits from the θ space viewed as a Riemannian manifold with constant inner product Σ_{θ_0}. Since, for the maximum likelihood estimate β_0, inference usually proceeds via the Fisher information matrix at β_0, it seems natural to study the curvature in terms of this constant metric on θ space.

In normal theory regression with known covariance, the 0 geometry and the c geometry obviously are identical. Further, the tensor T involves the derivatives of the information matrix (this can be seen by examining the second term on the right-hand side of (4)), and thus is zero. It follows that all geometries coincide.

Reparametrizations of single parameter curved exponential families were systematically investigated by Hougaard (1982), who showed that parametrizations with various desirable properties could be obtained as solutions to differential equations, and that in the normal family, it was possible to find a single parametrization with all these properties. This work was extended to the multiparameter situation by Kass (1984). Our approach here is somewhat different in that we exploit the fact that both the c geometry and the 0 geometry in θ space are Riemannian with the appropriate metric. This facilitates the computation of the Riemann–Christoffel symbols and the geodesics in β space.

The 0 geometry of θ space has another interesting property. Let $I(\theta_1, \theta_2)$ be the information "distance" between θ_1 and θ_2, and consider the constant information surfaces about a fixed point θ_0, $I(\theta_0, \theta) = \text{constant}$ (see Efron (1978)). Then, in the 0 geometry, these surfaces are totally umbilical submanifolds of θ space. That is, at any fixed point of these submanifolds, the second fundamental form has only one distinct eigenvalue or, in other words, all principal curvatures are identical. Thus, this submanifold has sphere-like properties in analogy to the constant information "circles" considered by Efron (1978). An appropriate generalization of expression 4.4 of Efron (1978) can likewise be given. These issues will be discussed elsewhere.

Acknowledgment. We wish to thank a referee for many useful suggestions.

REFERENCES

AMARI, S. I. (1982). Differential geometry of curved exponential families—curvatures and information loss. *Ann. Statist.* **10** 357–385.

BATES, D. M. and WATTS, D. G. (1980). Relative curvature measures of nonlinearity. *J. Roy. Statist. Soc. Ser. B* **42** 1–25.

BATES, D. M. and WATTS, D. G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Ann. Statist.* **9** 1152–1167.

BEALE, E. M. L. (1960). Confidence regions in non-linear estimation (with discussion). *J. Roy. Statist. Soc. Ser. B* **22** 41–76.

BOOTHBY, W. M. (1975). *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic, New York.

BRESLOW, N. E. and DAY, N. E. (1980). *Statistical Methods in Cancer Research* **1**. *The Analysis of Case-Control Studies.* International Agency for Research on Cancer, Lyon, France.

COX, D. R. (1969). *Analysis of Binary Data.* Methuen, London.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

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.

HAMILTON, D. C., WATTS, D. G. and BATES, D. M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Ann. Statist.* **10** 386–393.

HICKS, N. J. (1965). *Notes on Differential Geometry.* Van Nostrand Reinhold, New York.

HOUGAARD, P. (1982). Parameterizations of non-linear models. *J. Roy. Statist. Soc. Ser. B* **44** 244–252.

KASS, R. (1984). Canonical parameterizations and zero parameter effects curvature. *J. Roy. Statist. Soc. Ser. B* **46** 86–92.

LAURITZEN, S. L. (1986). Statistical manifolds. To appear in *Differential Geometry in Statistical Inference.* IMS, Hayward, Calif.

LUSTBADER, E. D., MOOLGAVKAR, S. H. and VENZON, D. J. (1984). Tests of the null hypothesis in case-control studies. *Biometrics* **40** 1017–1024.

MILNOR, J. (1963). Morse theory. *Ann. Math. Studies* No. 51. Princeton Univ. Press.

MOOLGAVKAR, S. H. and VENZON, D. J. (1986). Confidence regions for parameters of the proportional hazards model: A simulation study. *Scand. J. Statist.* To appear.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

SPROTT, D. A. (1973). Normal likelihoods and their relation to large sample theory of estimation. *Biometrika* **60** 457–466.

STORER, B. E., WACHOLDER, S. and BRESLOW, N. E. (1983). Maximum likelihood fitting of general risk models to stratified data. *J. Roy. Statist. Soc. Ser. C* **32** 172–181.

THOMAS, D. C. (1981). General relative risk models for survival time and matched case-control analysis. *Biometrics* **37** 673–686.

VÆTH, M. (1985). On the use of Wald's test in exponential families. *Internat. Statist. Rev.* **53** 199–214.

FRED HUTCHINSON CANCER RESEARCH CENTER
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104