Biometrika (1990), 77, 1, *pp.* 107–14
Printed in Great Britain

Data-translated likelihood and Jeffreys's rules

BY ROBERT E. KASS
Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.

SUMMARY

According to the definition used by Box & Tiao (1973) for a likelihood to be 'data translated' it must have location form in terms of a sufficient statistic. In contrast to Jeffreys's arguments for a uniform prior, theirs does not cover cases such as the Cauchy location family, and is in this sense stronger than the group-theoretic criterion of invariance. Their definition is easily modified to cover such cases through the introduction of an ancillary statistic, and their argument in favour of a uniform prior then becomes group-theoretic. Their concept of 'approximate data-translated likelihood' may also be modified to produce a sharper local approximation.

Some key words: Ancillary statistic; Information metric; Invariant prior; Reference prior.

1. INTRODUCTION

Box & Tiao (1973, § 1·3) introduced the notion of 'data-translated likelihood' to motivate the use of uniform priors. They then introduced 'approximate data-translated likelihood' to motivate Jeffreys's general rule, which is to take the prior density to be proportional to the square-root of the determinant of the Fisher information matrix. The purpose of this note is to clarify these concepts.

Let y be a vector of observations and let $L_y(.)$ be the likelihood function on a real one-dimensional parameter space Φ. According to Box & Tiao (1973, eqn (1·3·13)), the likelihood function is data translated if it may be written in the form

$$L_y(\phi)=f\{\phi-t(y)\} \tag{1·1}$$

for some real-valued functions $f(.)$ and $t(.)$, with the definition of $f(.)$ not depending on y. When (1·1) is satisfied, Box & Tiao recommend the use of the uniform prior on Φ because two different samples y and y^* will then produce posteriors that differ only with respect to location. That is, the uniform prior does not produce posterior densities with different shapes for different samples. This feature of the uniform prior is, for Box & Tiao, what makes it 'noninformative'.

For a likelihood to be approximately data translated, Box & Tiao require it to be 'nearly independent of the data y except for its location'. Operationally, they discuss samples of size n consisting of independent and identically distributed observations and begin with the normal approximation to the likelihood

$$L_y(\theta)\simeq n(\theta;\, \hat{\theta}, \hat{\sigma}_y^2), \tag{1·2}$$

where $n(x;\, \mu, \sigma^2)$ is the normal density with argument x, mean μ and variance σ^2, and $\hat{\sigma}_y^2=\{ni(\hat{\theta})\}^{-1}$, the inverse of the expected Fisher information evaluated at the maximum likelihood estimate $\hat{\theta}$. They then take ϕ to be a variance-stabilizing parameterization, that is, $i(\phi)=c$ for some constant c, so that

$$L_y(\phi) \simeq n(\phi; \hat{\phi}, c/n). \tag{1·3}$$

The normal approximate likelihood of (1·3) has the form (1·1), so that the likelihood itself is, in a sense Box & Tiao do not make explicit, approximately data translated. Based on the analogy with (1·1), they recommend the use of a prior that is uniform on ϕ, and they note that this prior is the one determined by Jeffreys's general rule.

The first and most basic question, arising from the determination of a uniform prior based on (1·1), is whether it amounts to the selection of an invariant prior under a group action. In § 2 it is shown that (1·1) applies only in restricted cases, but a simple modification of (1·1) applies to general group transformation models. In § 3, it is shown that (1·1), or a modification of it, holds locally to order $O(n^{-1})$, even though (1·3) is only accurate to order $O(n^{-\frac{1}{2}})$. The results may be considered additional motivation for Jeffreys's general rule.

2. Exact data-translated likelihood

From (1·1) it follows that the likelihood functions based on alternative data y and y^* are translated images of one another in the sense that

$$L_y(\phi) = L_{y^*}(\phi^*) \tag{2·1}$$

for $\phi^* = \phi + \{t(y^*) - t(y)\}$. Clearly, if (2·1) holds, the translation group may be defined on Φ and on the image of $t(.)$ so that the likelihood function is invariant under its action. With the requirement that $f(.)$ is independent of y, however, $t(y)$ becomes a sufficient statistic, which is restrictive. For instance, when Φ is the real line, the only translation families having support independent of ϕ for which there exists a one-dimensional sufficient reduction of n independent and identically distributed observations are the normal and logged gamma families (Dynkin, 1951; Ferguson, 1962). Thus, for example, the Cauchy location family does not satisfy (2·1).

It is easy to modify (1·1), however, so as to obtain a version of (2·1) that holds for general location families. The modification is to allow the definition of $f(.)$ to depend on the value of an ancillary statistic $a=a(y)$. That is, the likelihood becomes data translated in an extended sense if

$$L_y(\phi) = f_a\{\phi - t(y)\}, \tag{2·2}$$

where $f_a(.)$ depends on the data only through a. A well-known argument, for example, Cox & Hinkley (1974, p. 221), shows that this extended notion covers general location families: consider the configuration statistic $a=(y_{(2)}-y_{(1)}, \ldots, y_{(n)}-y_{(n-1)})$, where $y_{(i)}$ is the ith order statistic, let $\hat{\phi}$ be the maximum likelihood estimate of the location parameter ϕ, and note that (i) a is distribution-constant, and (ii) $(\hat{\phi}, a)$ is sufficient. That is, a is ancillary in conjunction with $\hat{\phi}$. Furthermore, (iii) the conditional distribution of $\hat{\phi}$ given a is again of location type. From (ii), the likelihood function based on y is equal to the likelihood function based on $(\hat{\phi}, a)$ apart from an arbitrary multiplicative constant; i.e.

$$L_y(\phi)/L_y(\hat{\phi}) = L_{(\hat{\phi},a)}(\phi)/L_{(\hat{\phi},a)}(\hat{\phi}).$$

From (iii) the conditional density of $\hat{\phi}$ given a may be written $p(\hat{\phi}\,|\,a, \phi) = p_a(\hat{\phi}-\phi)$ and, using this together with (i), $p(\hat{\phi}, a\,|\,\phi) = p_a(\hat{\phi}-\phi)h(a)$, where $h(a)$ does not depend on ϕ. Therefore, $L_{(\hat\phi,a)}(\phi)/L_{(\hat\phi,a)}(\hat\phi)=p_a(\hat\phi-\phi)$, so that

$$L_{(\hat\phi,a)}(\phi)/L_{(\hat\phi,a)}(\hat\phi)=L_{(\hat\phi^*,a)}(\phi^*)/L_{(\hat\phi^*,a)}(\hat\phi^*) \tag{2·3}$$

for $\phi^*=\phi+(\hat\phi^*-\hat\phi)$, which generalizes (2·1).

The choice of the statistic a is irrelevant to the definition (2·2) as long as a is ancillary in conjunction with $\hat\phi$. Furthermore, (2·2) holds for all values a of the ancillary statistic, except possibly on a set of Lebesgue measure zero. In words, we may choose any statistic a satisfying (i) and (ii); then, for two alternative data vectors y and y^* such that $a(y)=a(y^*)$, the likelihood functions $L_y(\phi)$ and $L_{y^*}(\phi)$ will be identical apart from the translation $\hat\phi\to\hat\phi^*$.

The extended definition (2·2) may be recognized as essentially group-theoretic, and a version of (2·3) will hold for quite general transformation models. For simplicity, suppose that a group G acting on the sample space $\mathcal{Y}$ generates the family of densities, with G being in one-to-one correspondence with the family, and further suppose there exists a sufficient statistic of the form $(\hat g, a)$, where $\hat g$ is the maximum likelihood estimate and a is invariant; that is $a(gy)=a(y)$ for all $g\in G$ and $y\in\mathcal{Y}$. It may be shown (Fraser, 1968, Ch. 2), that the conditional density of $\hat g$ given a is again transformational. Furthermore, taking $g^*=\hat g^*\hat g^{-1}g$, we have

$$L_{\hat g}(g)/L_{\hat g}(\hat g)=L_{\hat g^*}(g^*)/L_{\hat g^*}(\hat g^*) \tag{2·4}$$

as the generalization of (2·3). See the Appendix for details.

3. Approximate data-translated likelihood

3·1. *Local approximation*

Approximation (1·3) has a multiplicative error of order $O(n^{-\frac12})$ for ϕ such that $\phi-\hat\phi=O(n^{-\frac12})$. It will now be shown that by arguing locally in a neighbourhood of some parameter value ϕ_0, the order of error may be improved to $O(n^{-1})$, thereby including a nonnormal factor that can account for some skewness in the likelihood on the variance-stabilizing parameter. For $\delta>0$, the local argument compares the likelihood at $\phi=\hat\phi+\delta n^{-\frac12}$ based on data with a maximum likelihood estimate $\hat\phi$, to the likelihood at $\phi^*=\hat\phi^*+\delta n^{-\frac12}$ based on data with a maximum likelihood estimate $\hat\phi^*$, where both $\hat\phi$ and $\hat\phi^*$ differ by $O(n^{-\frac12})$ from ϕ_0, as would maximum likelihood estimates from alternative samples under the true parameter value ϕ_0.

Consider first the exponential family case, for which observed information $I_y(\hat\theta)=-l''_y(\hat\theta)$ and expected information at the maximum likelihood estimate agree; that is $I_y(\hat\theta)=ni(\hat\theta)$. For θ such that $\theta=\hat\theta+O(n^{-\frac12})$, the log likelihood may be expanded as

$$\begin{aligned} l_y(\theta)&=l_y(\hat\theta)+\tfrac12 l''_y(\hat\theta)(\theta-\hat\theta)^2+\tfrac16(E_{\theta_0}\{l'''_Y(\theta_0)\}+[l'''_y(\hat\theta)-E_{\theta_0}\{l'''_Y(\theta_0)\}])(\theta-\hat\theta)^3+O(n^{-1})\\ &=l_y(\hat\theta)+\tfrac12 l''_y(\hat\theta)(\theta-\hat\theta)^2+\tfrac16 E_{\theta_0}\{l'''_Y(\theta_0)\}(\theta-\hat\theta)^3+O(n^{-1}),\end{aligned}$$

the latter equality holding when $\theta_0-\hat\theta=O(n^{-\frac12})$. For an exponential family, y may be replaced by the maximum likelihood estimate since it is sufficient. For a variance stabilizing parameter ϕ, with maximum likelihood estimates

$$\hat\phi=\phi_0+O(n^{-\frac12}),\quad \hat\phi^*=\phi_0+O(n^{-\frac12}),$$

we obtain

$$\frac{L_{\hat\phi}(\phi)}{L_{\hat\phi}(\hat\phi)}=\frac{L_{\hat\phi^*}(\phi^*)}{L_{\hat\phi^*}(\hat\phi^*)}\{1+O(n^{-1})\}, \tag{3·1}$$

where $\phi - \hat{\phi} = O(n^{-\frac{1}{2}})$ and $\phi^* = \phi + (\hat{\phi}^* - \hat{\phi})$. Approximation (3·1) shows that likelihoods based on independent and identically distributed samples of observations from exponential families are locally approximately data translated to order $O(n^{-1})$. Note that, if we let

$$f(\delta n^{-\frac{1}{2}}) = L_{\phi_0}(\phi_0 + \delta n^{-\frac{1}{2}})/L_{\phi_0}(\phi_0),$$

where $L_{\phi_0}(.)$ is the likelihood function that would be obtained if ϕ_0 were the maximum likelihood estimate, and we continue to let $\hat{\phi}$ be the maximum likelihood estimate based on y, then we have

$$L_y(\phi)/L_y(\hat{\phi}) = f(\phi - \hat{\phi})\{1 + O(n^{-1})\} \tag{3·2}$$

for $\phi - \hat{\phi} = O(n^{-\frac{1}{2}})$ and $\hat{\phi} - \phi_0 = O(n^{-\frac{1}{2}})$, which may be compared with (1·1).

For the nonexponential family case, the maximum likelihood estimate is no longer sufficient and an approximation analogous to (2·2) must replace (3·2). Let $A = a$ be the value of the locally ancillary statistic based on observed information,

$$a = n^{\frac{1}{2}}\{ni(\hat{\theta})\gamma(\hat{\phi})\}^{-1}\{I_y(\hat{\theta}) - ni(\hat{\theta})\}$$

(Efron & Hinkley, 1978), where $\gamma(\theta)$ is the statistical curvature at θ. The statistic $(\hat{\theta}, A)$ is approximately locally sufficient in the sense that

$$\frac{L_{(\hat{\theta},a)}(\theta)}{L_{(\hat{\theta},a)}(\hat{\theta})} = \frac{L_y(\theta)}{L_y(\hat{\theta})}\{1 + O(n^{-1})\} \tag{3·3}$$

for $\theta - \hat{\theta} = O(n^{-\frac{1}{2}})$, for example, Michel (1978). Furthermore, for the variance stabilizing parameter ϕ for which $i(\phi) = 1$ there is

$$I_y(\hat{\phi}) = \gamma(\phi_0)^{-1}n^{\frac{1}{2}}a + n + \{\gamma(\hat{\phi}) - \gamma(\phi_0)\}n^{\frac{1}{2}}a,$$

which is independent of $\hat{\phi}$ to order $O(1)$. Thus, the argument leading to (3·1) may be used with $(\hat{\phi}, a)$ and $(\hat{\phi}^*, a)$ replacing $\hat{\phi}$ and $\hat{\phi}^*$ to obtain

$$\frac{L_{(\hat{\phi},a)}(\phi)}{L_{(\hat{\phi},a)}(\hat{\phi})} = \frac{L_{(\hat{\phi},a)}(\phi^*)}{L_{(\hat{\phi}^*,a)}(\hat{\phi}^*)}\{1 + O(n^{-1})\}, \tag{3·4}$$

where

$$\phi^* = \phi + (\hat{\phi}^* - \hat{\phi}), \quad \phi - \hat{\phi} = O(n^{-\frac{1}{2}}), \quad \hat{\phi} - \phi_0 = O(n^{-\frac{1}{2}}), \quad \hat{\phi}^* - \phi_0 = O(n^{-\frac{1}{2}}).$$

To extend (3·2), for $\delta > 0$ we may define

$$f_a(\delta n^{-\frac{1}{2}}) = L_{(\phi_0,a)}(\phi_0 + \delta n^{-\frac{1}{2}})/L_{(\phi_0,a)}(\phi_0)$$

and obtain

$$L_y(\phi)/L_y(\hat{\phi}) = f_a(\phi - \hat{\phi})\{1 + O(n^{-1})\} \tag{3·5}$$

for $\phi - \hat{\phi} = O(n^{-\frac{1}{2}})$ and $\hat{\phi} - \phi_0 = O(n^{-\frac{1}{2}})$. Thus, for general independent and identically distributed samples, likelihood functions are locally approximately data translated to order $O(n^{-1})$, in the sense of (3·5), analogously to (2·2).

3·2. *Approximate posterior densities*

The approximate normal density of the maximum likelihood estimate $\hat{\theta}$ has a value of $\{ni(\theta)/(2\pi)\}^{\frac{1}{2}}$ at its peak $\hat{\theta} = \theta$. This was mentioned by Jeffreys (1961, p. 192), attributed to Diananda, as an alternative motivation for Jeffreys's general rule, which is to take the prior proportional to $\{i(\theta)\}^{\frac{1}{2}}$. Perks (1947) made a similar argument. If the prior of Jeffreys's general rule is used, then the posterior on the variance-stabilizing parameter ϕ based on the normal approximation to the distribution of the maximum likelihood estimator will have a constant value $n^{\frac{1}{2}}/(2\pi)^{\frac{1}{2}}$ for its peak density, regardless of the observed value of the maximum likelihood estimate $\hat{\phi}$.

In fact, a somewhat stronger statement is also true. By Laplace's method (Bleistein & Handelsman, 1986, Ch. 5),

$$\int L_y(\phi)\, d\phi = (2\pi)^{\frac{1}{2}} L_y(\hat{\phi}) I_y(\hat{\phi})^{-\frac{1}{2}}\{1+O(n^{-1})\},$$

so that the posterior density using Jeffreys's general rule satisfies

$$p(\phi \mid y) = (2\pi)^{-\frac{1}{2}} I_y(\hat{\phi})^{\frac{1}{2}} \frac{L_y(\phi)}{L_y(\hat{\phi})}\{1+O(n^{-1})\}. \tag{3·6}$$

For an independent and identically distributed sample from an exponential family, $I_y(\hat{\phi})=n$ and the value of the posterior density at its peak is approximately $n^{\frac{1}{2}}/(2\pi)^{\frac{1}{2}}$ with an error of order $O(n^{-1})$, regardless of the value of $\hat{\phi}$. This improves on the order $O(n^{-\frac{1}{2}})$ error provided by the normal approximation.

Furthermore, combining (3·6) with (3·1) and absorbing $(2\pi/n)^{\frac{1}{2}}$ into the definition of f we obtain

$$p(\phi \mid y) = f(\phi - \hat{\phi})\{1+O(n^{-1})\} \tag{3·7}$$

for

$$\phi = \hat{\phi} + O(n^{-\frac{1}{2}}), \quad \hat{\phi} = \phi_0 + O(n^{-\frac{1}{2}}).$$

That is, the posterior density of ϕ may be considered locally asymptotically data-translated to order $O(n^{-1})$ when the uniform prior on ϕ is used.

The nonexponential family case may be treated as in § 3·1 by noting that, for fixed values of the locally ancillary statistic A, $I_y(\hat{\phi})$ is locally independent of $\hat{\phi}$ to order $O(n^{-1})$. This provides a weakened version of (3·7) in which f is replaced by f_a as in (3·5).

3·3. *Multiparameter local approximation*

In multiparameter families, variance-stabilizing parameterizations need not exist, and the argument leading to (3·2) and (3·5) does not go through. There is, however, a generalization of equations (3·1) and (3·4). It may be obtained by suitably identifying points θ at which the likelihood based on $\hat{\theta}$ is evaluated, with points θ^* at which the likelihood based on $\hat{\theta}^*$ is evaluated. The identification is made using geodesics defined by the information metric. See Kass (1989) for a discussion of the information metric and related concepts used below.

Again, for simplicity, consider the exponential family case, and let M be an m-dimensional regular exponential family structured as a Riemannian manifold with the information metric. Let θ_0 be a point in a parameter space Θ, and let $\hat{\theta}$ and $\hat{\theta}^*$ be two maximum likelihood estimate values in an order $O(n^{-\frac{1}{2}})$ neighbourhood of θ_0. Given a point θ near $\hat{\theta}$ at which $L_{\hat{\theta}}$ will be evaluated, a corresponding point θ^* near $\hat{\theta}^*$ may be defined as follows. Letting τ_{θ_1,θ_2} be parallel transport with respect to the information metric along the geodesic from θ_1 to θ_2, take

$$\theta^* = \exp_{\hat{\theta}^*}(\tau_{\theta_0,\hat{\theta}^*}[\tau_{\hat{\theta},\theta_0}\{\exp_{\hat{\theta}}^{-1}(\theta)\}]),$$

where $\exp_\theta$ is the exponential map on the tangent space at θ. Now, letting s be arc length along the geodesic connecting θ with $\hat{\theta}$, with $s=0$ corresponding to $\hat{\theta}$, we may write $\theta=\theta(s)$ and expand $l_{\hat{\theta}}\{\theta(s)\}$ as a function of s at $s=0$.

To carry this out, let $F(s)=l_{\hat{\theta}}\{\theta(s)\}$. Since s is arc length we have $\langle d\theta/ds, d\theta/ds\rangle=1$, using the information metric, so $F''(0)=-n$. Thus,

$$l_{\hat{\theta}}(\theta)-l_{\hat{\theta}}(\hat{\theta})=-\frac{1}{2}ns^2+\frac{1}{6}\left[\frac{d^3F}{ds^3}\right]_{s=0}s^3+O(s^4),$$

and, with $F^*(s)=l_{\hat{\theta}^*}\{\theta^*(s)\}$,

$$l_{\hat{\theta}^*}(\theta)-l_{\hat{\theta}^*}(\hat{\theta})=-\frac{1}{2}ns^2+\frac{1}{6}\left[\frac{d^3F^*}{ds^3}\right]_{s=0}s^3+O(s^4).$$

Notice that in the above expansions $s=d(\theta,\hat{\theta})=d(\theta^*,\hat{\theta}^*)$, the latter equality holding due to the definition of θ^*; thus, the value of s appearing in the two expansions is the same.

To arrive at the generalization of (3·1) we need only note that

$$\left[\frac{1}{n}\frac{d^3F}{ds^3}\right]_{s=0}=\left[\frac{1}{n}\frac{d^3F^*}{ds^3}\right]_{s=0}+O(n^{-\frac{1}{2}})$$

as $\hat{\theta}$ and θ^* converge to θ_0 at the rate $O(n^{-\frac{1}{2}})$. Thus, on exponentiating the geodesic log likelihood expansions,

$$\frac{L_{\hat{\theta}}(\theta)}{L_{\hat{\theta}}(\hat{\theta})}=\frac{L_{\hat{\theta}^*}(\theta^*)}{L_{\hat{\theta}^*}(\hat{\theta}^*)}\{1+O(n^{-1})\} \tag{3·8}$$

which may be compared with (2·1). The interpretation of (3·8) is that the likelihood is locally approximately data-translated to order $O(n^{-1})$ along geodesics defined by the information metric. If used to motivate a choice of prior, it suggests choosing the prior to be uniform in the information-metric geometry, i.e. to use Jeffreys's general rule. The uniformity of Jeffreys's rule is discussed by Kass (1989).

4. CONCLUSION

One-parameter likelihoods are exactly data translated in the sense of Box & Tiao (1973) only in the normal and logged-gamma location families, among families having fixed support independent of the parameter value. In § 2 it was shown that when the concept is extended to general location families, it becomes group-theoretic; i.e. it may also be extended to general transformation families. The derivation was based on results that are well known in the theory of conditional inference, though the focus here is quite different. In § 3 it was shown that one-parameter likelihoods are locally data translated to order $O(n^{-1})$. The derivation there was based on a Taylor-series expansion in a form used by Hinkley (1980), and is closely related to results of Welch & Peers (1963), which show that one-parameter families may be considered local location families in the variance-stabilizing parameterization. An extension to multiparameter families was given using the geometry of the information metric, but it takes a weakened form, considering local data-translation separately along each geodesic, rather than simultaneously along arbitrary local paths.

The arguments given here, like those of Box & Tiao (1973) provide some heuristic motivation for using Jeffreys's rules for selecting a reference prior. To the extent that data translation of the likelihood captures an intuitive notion of what is 'noninformative' about the uniform prior on Φ of (1·1), one might wish to say that invariant priors, generally, provide 'constant information' on orbits. On the other hand, a less sanguine view of (2·4) would be that it simply shows Box & Tiao's concept to be based on symmetry.

Acknowledgement

I would like to thank Jay Kadane, Steve Stigler, Larry Wasserman and a referee for their comments.

Appendix

Derivation of (2·4)

With e being the identity of G and $J_{\hat{g}}(g)$ being the Jacobian determinant of the transformation $g_0 \to gg_0$ at $\hat{g}$,

$$p(\hat{g} \mid a, g)p(a) = p(\hat{g}, a \mid g) = p(g^{-1}\hat{g}, a \mid e)J_{\hat{g}}(g^{-1})$$
$$= p(g^{-1}\hat{g} \mid a, e)p(a)J_{\hat{g}}(g^{-1}),$$

so that $p(\hat{g} \mid a, g) = p(g^{-1}\hat{g} \mid a, e)J_{\hat{g}}(g)^{-1}$, which shows that the conditional distribution of $\hat{g}$ given a again forms a transformation family under the action of G. The likelihood function may be written

$$L_{\hat{g}}(g) = p(g^{-1}\hat{g} \mid a, e)J_{\hat{g}}(g^{-1}). \qquad (A·1)$$

Using the chain rule

$$J_{g_0}(g_2 g_1) = J_{g_1 g_0}(g_2)J_{g_0}(g_1), \qquad (A·2)$$

we get

$$J_{\hat{g}}(g^{-1}) = J_{\hat{g}}(g^{-1}\hat{g}\hat{g}^{-1}) = J_e(g^{-1}\hat{g})J_{\hat{g}}(\hat{g}^{-1}).$$

Taking $g^* = \hat{g}^*\hat{g}^{-1}g$, another application of (A·2) gives

$$J_{\hat{g}^*}(g^{-1}\hat{g}\hat{g}^{*-1}) = J_e(g^{-1}\hat{g})J_{\hat{g}^*}(\hat{g}^{*-1}),$$

and putting these Jacobian expressions in (A·1), we have (2·4).

This argument goes through under weaker conditions, as well: the sample space $\mathcal{Y}$ must be a locally compact Hausdorff space, and G must be a locally compact, σ-compact Hausdorff group acting properly on $\mathcal{Y}$, as given by Barndorff-Nielsen et al. (1982). The derivation of (2·4) in the general case would take the dominating measure for the family to be left-invariant measure, rather than Lebesgue measure; this simplifies the argument slightly, but Lebesgue measure was used here because it is more familiar.

References

Barndorff-Nielson, O., Blaesild, P., Jensen, J. L. & Jorgensen, B. (1982). Exponential transformation models. *Proc. R. Soc., Lond.* A **379**, 41–65.

Bleistein, N. & Handelsman, R. A. (1986). *Asymptotic Expansions of Integrals.* New York: Dover.

Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass: Addison Wesley.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman & Hall.

Dynkin, E. B. (1951). Necessary and sufficient statistics for a family of probability distribution. *Select. Transl. Math. Statist. Prob.* **1** (1961), 23–41.

Efron, B. & Hinkley, D. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). *Biometrika* **65**, 457–87.

Ferguson, T. S. (1962). Location and scale parameters in exponential families of distributions. *Ann. Math. Statist.* **33**, 986–1001; Correction (1963) **34**, 1603.

Fraser, D. A. S. (1968). *The Structure of Inference.* New York: Wiley.

Hinkley, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* **67**, 287–92.

Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford University Press.

KASS, R. E. (1989). The geometry of asymptotic inference (with discussion). *Statist. Sci.* **4**, 188-234.

MICHEL, R. (1978). Higher order asymptotic sufficiency. *Sankhyā* **40**, 76-84.

PERKS, W. (1947). Some observations on inverse probability including a new indifference rule. *J. Inst. Actuar.* **73**, 285-334.

WELCH, B. L. & PEERS, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Statist. Soc.* B **25**, 318-29.

[*Received October* 1988. *Revised June* 1989]