Edinburgh School of Economics

Discussion Paper Series

Number 23

Reference priors for the general location-scale model

Carmen Fernandez (CentER and Tilburg University)
Mark F J Steel (CentER and Tilburg University)

Date
October 1998

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

Reference Priors
for the General Location-Scale Model

Carmen Fernández and Mark F.J. Steel *
CentER for Economic Research and Department of Econometrics
Tilburg University, 5000 LE Tilburg, The Netherlands

ABSTRACT

The reference prior algorithm (Berger and Bernardo 1992) is applied to multivariate location-scale models with any regular sampling density, where we establish the irrelevance of the usual assumption of Normal sampling if our interest is in either the location or the scale. This result immediately extends to the linear regression model. On the other hand, an essentially arbitrary step in the reference prior algorithm, namely the choice of the nested sequence of sets in the parameter space is seen to play a role. Our results lend an additional motivation to the often used prior proportional to the inverse of the scale parameter, as it is found to be both the independence Jeffreys' prior and the reference prior under variation independence in the sequence of sets, for any choice of the sampling density. However, if our parameter of interest is not a one-to-one transformation of either location or scale, the choice of the sampling density is generally shown to intervene.

Keywords: Jeffreys' prior; Multivariate regression model; Posterior existence; Scale mixture of Normals.

1. INTRODUCTION

An important issue in Bayesian statistics is that of a standard "non-informative" prior distribution, often required when clear subjective prior information is lacking or for the purpose of scientific reporting. For models with only one parameter, the method introduced by Jeffreys (1961), based on invariance arguments, has gained widespread acceptance. However, in the presence of nuisance parameters, various approaches have been suggested. Jeffreys himself (1961, p. 182) considers a modification of his rule for the cases where "a previous judgement of irrelevance" seems reasonable. The result of the latter principle will be denoted as "independence Jeffreys' prior" in the sequel.

Using an information theory argument, Bernardo (1979) develops a formal methodology for multiparameter problems. He explicitly distinguishes between parameters of interest and nuisance parameters and defines the so-called "reference prior" as the prior

* Carmen Fernández is Research Fellow, CentER for Economic Research and Assistant Professor, Department of Econometrics, Tilburg University, (E-mail: carmen@kub.nl); and Mark Steel is Senior Research Fellow, CentER for Economic Research and Associate Professor, Department of Econometrics, Tilburg University, 5000 LE Tilburg, The Netherlands (E-mail: steel@kub.nl). We gratefully acknowledge comments from James O. Berger, José M. Bernardo, Jacek Osiewalski and an anonymous Referee. The first author holds a Research Training Grant (ERBFMBICT 961021) under the Training and Mobility of Researchers Programme, financed by the European Commission.

that maximizes the expected amount of information (in the Shannon sense) about the parameter of interest, as the number of independently and identically distributed (i.i.d.) replications of the experiment goes to infinity. Under certain regularity conditions, his method leads to Jeffreys' prior in the absence of nuisance parameters. An explicit and updated description of the reference prior algorithm was presented in Berger and Bernardo (1992). They introduce and discuss some technical refinements, such as a nested sequence of compact sets (hereafter denoted by $\{\Theta^l\}$), converging to the entire parameter space when the latter is non-compact. The reference prior is then derived by first considering Θ^l and afterwards taking a limit on l. In addition, in the case of more than two parameters, they consider more parameter groups than just separating the parameter of interest and the nuisance parameters. Decisions concerning the choice of the sequence of sets, as well as the choice and order of the parameter groups are found to potentially influence the form of the reference prior. On the basis of their experience with various models, Berger and Bernardo (1992) state that usually the choice of the sequence $\{\Theta^l\}$ does not matter, and they recommend defining a separate group for each parameter. For the ordering of the groups, however, they give no strict guidelines, other than ordering according to "inferential importance". Thus, the form of the reference prior can be subject to a number of essentially arbitrary decisions.

To date, the reference prior algorithm has been applied to a number of problems. Many of these results were derived under a particular distributional assumption for the sampling model, which is often Normality in location-scale contexts. The statistical analysis of real-life data, however, often requires more flexible assumptions, and novel computing techniques (such as Markov Chain Monte Carlo methods) allow us to use a much greater variety of models in empirical applications. Thus, reference priors for wide classes of sampling models are of genuine practical interest.

In the present paper, we will focus on the reference prior in the context of general location-scale models. We extend the usual assumption of Normality to any regular sampling density and explicitly consider various choices for the parameter of interest, which can involve both location and scale. We examine to what extent the distributional assumptions change the reference prior. In addition, we assess the influence of the choice of the sequence of sets $\{\Theta^l\}$.

Throughout the paper, we always assume the sufficient regularity conditions for asymptotic Normality of the likelihood function stated in DeGroot (1970, chap. 10), in which case the reference prior can be derived solely on the basis of the information matrix (Section 2.3 of Berger and Bernardo 1992, and Proposition 5.30 of Bernardo and Smith 1994).

We shall only consider the Berger and Bernardo reference prior in the present paper. Some recent publications in this area are, among others, Datta and Ghosh (1995), Clarke (1996) and Sun and Berger (1997). Bernardo and Ramón (1997) provide an account of the heuristic arguments behind the reference prior method. Alternative noninformative prior distributions are proposed in *e.g.* Tibshirani (1989), Ghosh and Mukerjee (1992), Clarke and Wasserman (1993) and Casella, DasGupta and Strawderman (1996). Kass and Wasserman (1996) give a comprehensive overview of formal prior selection rules.

In Section 2 we show that the structure of the information matrix is essentially the same for any continuous location-scale model. As a consequence, we can prove that the reference prior for the location and scale parameters is exactly the same for any sampling density, thus extending the result of Bernardo (1979) outside the Normality framework. In contrast to common belief, however, we observe that the choice of the sequence of sets in the parameter space, $\{\Theta^l\}$, can influence the form of the reference prior and we provide a general result in this respect. We also derive counterexamples indicating that the choice of sampling density may matter if we are interested in parameters that are not one-to one transformations of either location or scale. All this provides an important argument in favour of the "standard" prior which is proportional to a Uniform distribution on the location and the logarithm of the scale. Provided we are interested in location or scale, it is seen to be the reference prior in very general circumstances. Since this reference prior is improper, Section 3 discusses a sufficient condition for propriety of the resulting posterior distribution under sampling from scale mixtures of Normals. The extension to linear regression is immediate as mentioned in Section 4. Some main conclusions are summarized in Section 5.

Throughout the paper, we shall use $p(\cdot)$ to denote probability density functions on observables, whereas the notation $\pi(\cdot)$ shall be reserved for parameters. Key proofs will be sketched in the Appendix without explicit mention in the main text.

2. THE GENERAL LOCATION-SCALE MODEL

Let us consider the general continuous location-scale model, with density function

$$p(y|\mu,\sigma) = \sigma^{-k} f\{\sigma^{-1}(y-\mu)\}, \tag{2.1}$$

where $y \in \Re^k$ $(k \geq 1)$, $\mu \in \Re^k$ is a location parameter, $\sigma > 0$ is a scale parameter and $f(\cdot)$ is a probability density function (p.d.f.) in $\Re^k$. Implicitly, we shall assume throughout that the regularity conditions for asymptotic Normality mentioned in the Introduction hold. In practice, the location vector μ will often be of interest, whereas σ is often a nuisance parameter. However, other situations will also be examined.

2.1. Reference Priors for Location and Scale

In the reference prior literature, the assumption that $f(\cdot)$ is a Normal density function is usually added to (2.1). In that case, Bernardo (1979) derived the reference prior

$$\pi(\mu,\sigma) \propto \sigma^{-1}, \tag{2.2}$$

taking either μ or σ as the parameter of interest. This is also the independence Jeffreys' prior.

We shall now show that obtaining (2.2) as the reference prior does not hinge upon the Normal sampling assumption, but always holds for the model in (2.1) under "usual" choices for the sequence of sets Θ^l. Using the variable transformation from y to $z = \sigma^{-1}(y-\mu)$, we can derive that the information matrix for (2.1) is

$$I(\mu,\sigma) = \sigma^{-2} A(f), \tag{2.3}$$

where $A(f)$ is a $(k+1) \times (k+1)$-dimensional matrix with entries $a_{ij}(f)$ depending on $f(\cdot)$. Observe that the information matrix in (2.3) does not depend on μ, but only on σ and $f(\cdot)$. This fact is crucial in obtaining the following result:

Theorem 1. *For any choice of $f(\cdot)$ in* (2.1) *and with either μ or σ as the parameter of interest (two groups), we obtain* (2.2) *as the reference prior, provided the sequence of sets verifies $\Theta^l = \Theta^l_\mu \times \Theta^l_\sigma$, where $\{\Theta^l_\mu\}$ and $\{\Theta^l_\sigma\}$ are any nested sequences of compact sets for μ and σ, respectively.*

Theorem 1 extends the "usual" reference prior from Normality to a much wider context. We clearly see that the assumption of Normality is irrelevant for obtaining the prior in (2.2), since any choice of $f(\cdot)$ in (2.1) leads to this prior.

Our Theorem 1 already suggests the crucial role of the choice of the sequence $\{\Theta^l\}$. The condition on the latter mentioned in the Theorem is a sufficient condition that leads to (2.2). This absence of deterministic restrictions between μ and σ is termed "variation independence" in Basu (1977). Other choices of Θ^l may lead to different reference priors, although not necessarily. For instance, if we choose μ as the parameter of interest and

$$\Theta^l = \{(\mu, \sigma) : \|\mu\| \in [1/l, l], \sigma \in [\|\mu\|/l, l/\|\mu\|]\},$$

where $\|\cdot\|$ denotes the Euclidean norm, then we still obtain (2.2); however, the very similar sequence of

$$\Theta^l = \{(\mu, \sigma) : \|\mu\| \in [0, l], \sigma \in [(\|\mu\| + 1)/l, (l + 1)/\|\mu\|]\}$$

leads to $\pi(\mu, \sigma) \propto \sigma^{-1}\{\|\mu\|/(\|\mu\| + 1)\}^{k/2}$ as the reference prior. In general, we can deduce the following result:

Theorem 2. *With the model in* (2.1) *and using two groups, the reference prior algorithm leads to*
(i) if μ is the parameter of interest: $\pi(\mu, \sigma) \propto \sigma^{-1}t(\mu)$ for some positive function $t(\cdot)$,
(ii) if σ is the parameter of interest: $\pi(\mu, \sigma) \propto \sigma^{-1}s(\sigma)$ for some positive function $s(\cdot)$,
where the forms of $t(\cdot)$ and $s(\cdot)$ depend on the sequence $\{\Theta^l\}$.

From Theorem 2 we see that the product structure between μ and σ is always retained, irrespective of the choice of $\{\Theta^l\}$. In addition, this theorem clearly shows that the choice of $f(\cdot)$ in (2.1) **never** influences the reference prior. Thus, Normality plays no role whatsoever in obtaining the reference prior, whereas the sequence of sets chosen can affect its form. In the special case of variation independence in Θ^l, the functions $t(\cdot)$ and $s(\cdot)$ are constant and we are back in the situation of Theorem 1.

From the form of the information matrix in (2.3) we can immediately see that (2.2) also corresponds to the independence Jeffreys' prior for any $f(\cdot)$ in (2.1). The full Jeffreys' prior in this case is $\pi(\mu, \sigma) \propto \sigma^{-k-1}$.

Since the Jeffreys' prior does not depend on the parameter of interest, it will not be affected by the aims of our analysis. We may now wonder whether the fact that $f(\cdot)$ plays no role at all for the reference prior is an inherent feature of location-scale models, or is linked to the particular choice of μ or σ as parameters of interest. Proposition 5.28 of Bernardo and Smith (1994) states that, when two parameter groups are considered, the reference prior is invariant with respect to one-to-one transformations of the parameter of interest. Thus, Theorems 1 and 2 still apply in such cases. However, we now present some evidence indicating that the sampling density $f(\cdot)$ can matter when other parameters are taken to be of interest.

2.2. Reference Priors for the Norm of the Location

In a multivariate setting ($k > 1$), one may not be interested in the entire vector μ, but rather in some lower dimensional characteristic, such as its Euclidean norm. Stein (1959) pointed out serious difficulties in finding an adequate non-informative prior for this estimation problem. Bernardo (1979) presents the reference prior under Normal sampling with identity covariance matrix, whereas Berger, Philippe and Robert (1996) consider a known general covariance structure. Here we allow for an unknown scale parameter σ and any sampling density $f(\cdot)$. For reasons of simplicity, we shall derive the reference prior in the bidimensional case ($k = 2$).

Since we are interested in the Euclidean norm of μ, a polar representation of the latter seems natural. Thus, we reparameterize μ into (u, ω), with $u = \|\mu\|$, the Euclidean radius, and $\omega = \arctan(\mu_2/\mu_1)$, the polar angle. Following the general recommendation of Berger and Bernardo (1992), we consider a separate group for each of the parameters.

Theorem 3. *For the model in* (2.1) *with the order* $\{u, \omega, \sigma\}$ *(three groups), and any sequence of sets* $\Theta^l = \Theta^l_u \times \Theta^l_\omega \times \Theta^l_\sigma$, *we obtain the reference prior*

$$\pi(u, \omega, \sigma) \propto \sigma^{-1} v(\omega, f), \tag{2.4}$$

where $v(\omega, f)$ *is given in* $(A.1)$ *in the Appendix. In terms of the original parameterization, this implies*

$$\pi(\mu, \sigma) \propto \sigma^{-1} \|\mu\|^{-1} v(\arctan(\mu_2/\mu_1), f). \tag{2.5}$$

Thus, contrary to what we obtained in Theorem 2, the reference prior depends on the particular form of $f(\cdot)$ when our interest focuses on the polar radius of μ. However, in many cases the following simpler result applies:

Corollary 1. *Whenever* $f(\cdot)$ *is such that the information matrix in* (2.3) *(with* $k = 2$*) is diagonal with* $a_{11}(f) = a_{22}(f)$, $v(\omega, f)$ *becomes a constant function in* ω *and thus*

$$\pi(\mu, \sigma) \propto \sigma^{-1} \|\mu\|^{-1} \tag{2.6}$$

for any such $f(\cdot)$.

Any exchangeable and axially symmetric $f(\cdot)$ leads to an information matrix as required in Corollary 1. Both exchangeability and axial symmetry hold, for instance, for the entire spherical class (*i.e.* the class of all densities with Euclidean spheres as isodensity sets), which contains as important examples the multivariate Normal and Student-t distributions. They also hold for the l_q-spherical class, defined in Osiewalski and Steel (1993), which allows for more general shapes (spheres with respect to the l_q-norm) of the isodensity contours. In addition, the condition of Corollary 1 is fulfilled by certain classes of skewed distributions, such as the so-called multivariate Skewed Exponential Power distributions (introduced in Fernández, Osiewalski and Steel 1995).

Summarizing this Subsection, we have an example where focusing on a transformation of the location vector generally makes $f(\cdot)$ intervene. Using Proposition 5.28 of Bernardo and Smith (1994) (mentioned at the end of Subsection 2.1) in combination with our Theorem 2, it is immediate that under two groups, namely (u, ω) and σ, the reference prior does not depend on $f(\cdot)$. However, since the parameter of interest is only u, we need to place it in a separate group, thus introducing the dependence on $f(\cdot)$.

Next we investigate an example with only two groups, but where the parameter of interest involves both location and scale.

2.3. Reference Priors for Standardized Location

Consider the case where we are interested in $\phi = \mu/\sigma$. We shall take two groups with σ as the nuisance parameter (from Proposition 5.27 of Bernardo and Smith 1994, the choice of the latter does not matter). For convenience, we take the univariate case with $k = 1$:

Theorem 4. *For the model in* (2.1) *with the order* $\{\phi, \sigma\}$, *and any sequence of sets* $\Theta^l = \Theta^l_\phi \times \Theta^l_\sigma$ *or corresponding to rectangles in the parameterization* (μ, σ), *we obtain the reference prior*

$$\pi(\phi, \sigma) \propto \sigma^{-1}\{a_{11}(f)\phi^2 + 2a_{12}(f)\phi + a_{22}(f)\}^{-1/2}, \tag{2.7}$$

where the functions $a_{ij}(f)$ *correspond to the information matrix in* (2.3) *with* $k = 1$. *Therefore, in terms of the original parameterization, we obtain*

$$\pi(\mu, \sigma) \propto \sigma^{-2}\{a_{11}(f)(\mu/\sigma)^2 + 2a_{12}(f)(\mu/\sigma) + a_{22}(f)\}^{-1/2}. \tag{2.8}$$

Interestingly, we now no longer have a convenient class of distributions (as in Corollary 1) where the reference prior is the same as under Normality. Mixing scale and location into the parameter of interest really does make the reference prior depend critically upon the choice of $f(\cdot)$. As an example, under Student-t sampling the prior in (2.7) reduces to

$$\pi(\phi, \sigma) \propto \sigma^{-1}[\{(\nu + 1)/\nu\}\phi^2 + 2]^{-1/2}, \tag{2.9}$$

which depends on the degrees of freedom ν. As $\nu \to \infty$, the expression in (2.9) becomes $\pi(\phi, \sigma) \propto \sigma^{-1}(\phi^2 + 2)^{-1/2}$, which is the reference prior presented in Bernardo (1979) for the Normal case.

3. EXISTENCE OF THE POSTERIOR DISTRIBUTION

We now focus on posterior inference under the reference prior in (2.2). With a sample $Y = (y_1, \ldots, y_n)'$ of n independent replications from (2.1), the posterior distribution takes the form

$$\pi(\mu, \sigma|Y) = \frac{\pi(\mu, \sigma)\prod_{i=1}^n p(y_i|\mu, \sigma)}{p(Y)}, \text{ where } p(Y) = \int \pi(\mu, \sigma)\prod_{i=1}^n p(y_i|\mu, \sigma)d\mu d\sigma. \tag{3.1}$$

It is immediate to see that one single observation from (2.1) always leads to $p(Y) = \infty$. Thus, in order to conduct inference, we need to increase the sample size n. The following proposition provides a result under sampling from scale mixtures of Normals, which corresponds to choosing $f(\cdot)$ in (2.1) as

$$f(z) = \int_0^\infty \left(\frac{\lambda}{2\pi}\right)^{k/2} \exp\left(-\frac{\lambda}{2}z'z\right) dP_\lambda, \tag{3.2}$$

for $z \in \Re^k$ and P_λ any probability measure on $\Re_+$. This class of sampling distributions is a subset of the spherical class of considerable practical importance, containing *e.g.* the Normal distribution (when P_λ is Dirac) and the Student-t distribution with ν degrees of freedom, for which P_λ is Gamma$(\nu/2, \nu/2)$.

Proposition 1. *Consider the Bayesian model consisting of n independent observations from the sampling density in* (2.1) *with $f(\cdot)$ corresponding to a scale mixture of Normals as in* (3.2), *and the reference prior in* (2.2). *If $n \geq 2$ and the sample Y contains no repeated observations, then $p(Y) < \infty$.*

Note that the set of samples containing repeated observations has Lebesgue measure zero. This means that if $n \geq 2$, we can conduct posterior inference with almost any sample.

4. THE MULTIVARIATE LINEAR REGRESSION MODEL

Although the heuristic ideas behind reference priors are based on the case of i.i.d. replications of a basic experiment, the reference prior algorithm as such has been used in many other situations. In this section we take this wider view and apply the algorithm to a case where the observables are independent but not necessarily identically distributed. In particular, we consider the k-variate linear regression model, where we replace n i.i.d. replications from (2.1) by

$$p(y_i|\beta, \sigma) = \sigma^{-k} f\{\sigma^{-1}(y_i - \beta' x_i)\}, \quad i = 1, \ldots, n, \tag{4.1}$$

with $y_i \in \Re^k$, β a $p \times k$ matrix of regression coefficients, x_i a p-dimensional vector of exogenous explanatory variables and the entire $n \times p$ design matrix $X = (x_1, \ldots, x_n)'$ taken to be of full-column rank. Note that the pure location-scale model considered in the previous sections is a particular example of (4.1), corresponding to $p = x_i = 1$.

The structure of the information matrix for (β, σ) in the general model (4.1) can be shown to be the same as (2.3) in the way the parameters intervene, which implies that all the results presented in Section 2 for (μ, σ) also apply to (β, σ). In particular, from Theorem 1, the reference prior is

$$\pi(\beta, \sigma) \propto \sigma^{-1} \tag{4.2}$$

under variation independence of the sequence of sets, with either β or σ as the parameter of interest. For the univariate case ($k = 1$), this reference prior has also appeared in Yang and Berger (1996), albeit without proof or discussion and without mentioning the choice of sets Θ^l which we now show to be crucial to the result.

For the model in (4.1) – (4.2) with $k = 1$, Fernández and Steel (1996) deal with the issue of existence of the posterior distribution when $f(\cdot)$ is a scale mixture of Normals. Their results in this respect can be easily generalized to multivariate regression, where

$k > 1$. In addition, Fernández and Steel (1998) examine the existence of the posterior distribution when $k = 1$ and $f(\cdot)$ is a skewed scale mixture of Normals.

5. CONCLUSIONS

This paper investigates the reference prior in the context of general k-variate continuous location-scale models, with location $\mu \in \Re^k$ and scale $\sigma > 0$. Under mild regularity conditions, the usual assumption of Normality is found to be entirely irrelevant to the form of the reference prior in the location-scale model if our interest focuses on μ or σ. The reference prior only depends on the choice of the parameter of interest and the sequence of sets $\{\Theta^l\}$. If we assume variation independence in $\{\Theta^l\}$, then we always obtain $\pi(\mu, \sigma) \propto \sigma^{-1}$, no matter whether μ or σ is of interest. We can also immediately deduce that this prior coincides with the independence Jeffreys' prior. Thus, we can motivate the very frequently used prior $\pi(\mu, \sigma) \propto \sigma^{-1}$ as being both the reference prior (with a natural choice of $\{\Theta^l\}$) and the independence Jeffreys' prior in the broad setting of the general location-scale model.

However, if the parameter of interest is not a one-to-one transformation of either μ or σ, we show that the choice of the sampling density $f(\cdot)$ can influence the form of the reference prior.

Many recent studies illustrate both the feasibility of posterior analysis under non-Normal sampling distributions, using newly developed numerical techniques, and the sensitivity of posterior inference to changes in the sampling distributions. Thus, extending the reference posterior analysis to non-Normal sampling distributions is of genuine practical interest. We feel our theoretical results concerning the influence of the sampling distribution, the choice of $\{\Theta^l\}$ and the parameter of interest on the reference prior also increase our understanding of the intricate workings of this algorithm.

APPENDIX:

Proof of Theorem 1

We apply the reference prior algorithm described in Berger and Bernardo (1992), following their notation. We consider μ to be the parameter of interest (the proof for the case where σ is of interest can be done in a similar way).

From the information matrix in (2.3), we immediately obtain that $h_2(\mu, \sigma) \propto \sigma^{-2}$, whereas from the variation independence assumption we have that $\Theta^l(\mu) = \Theta^l_\sigma$. Thus

$$\pi_2^l(\sigma|\mu) \propto \sigma^{-1} I_{\Theta^l_\sigma}(\sigma).$$

In addition, we can derive that $h_1(\mu, \sigma) = \sigma^2 B(f)$ for some $k \times k$ matrix $B(f)$. Using again the variation independence, it follows that

$$\pi_1^l(\mu, \sigma) \propto \sigma^{-1} I_{\Theta^l}(\mu, \sigma),$$

which leads to the reference prior in (2.2).

Proof of Theorem 3

The information matrix takes the form

$$I(u,\omega,\sigma) = \frac{1}{\sigma^2}\begin{pmatrix} l_{11}(\omega,f) & ul_{12}(\omega,f) & l_{13}(\omega,f) \\ ul_{12}(\omega,f) & u^2 l_{22}(\omega,f) & ul_{23}(\omega,f) \\ l_{13}(\omega,f) & ul_{23}(\omega,f) & l_{33}(f) \end{pmatrix},$$

for some functions $l_{rs}(\cdot)$. It is then immediate that

$$\pi_3^l(\sigma|u,\omega) \propto \sigma^{-1} I_{\Theta_\sigma}(\sigma).$$

We can also derive that $|h_2(u,\omega,\sigma)| = \sigma^{-2}u^2v^2(\omega,f)$, where

$$v(\omega,f) = \left[a_{11}(f)\sin^2\omega + a_{22}(f)\cos^2\omega - a_{12}(f)\sin 2\omega - \frac{\{a_{13}(f)\sin\omega - a_{23}(f)\cos\omega\}^2}{a_{33}(f)}\right]^{1/2}, \tag{A.1}$$

with $a_{rs}(f)$ the (r,s) element of $A(f)$ in the information matrix in (2.3). This leads to

$$\pi_2^l(\omega,\sigma|u) \propto \sigma^{-1}v(\omega,f)I_{\Theta_\omega}(\omega)I_{\Theta_\sigma}(\sigma).$$

Finally, $h_1(u,\omega,\sigma) = \sigma^{-2}w(\omega,f)$ for some function $w(\omega,f)$, and it then follows that

$$\pi_1^l(u,\omega,\sigma) \propto \sigma^{-1}v(\omega,f)I_{\Theta}(u,\omega,\sigma),$$

which leads to the result in Theorem 3.

Proof of Proposition 1

We write $f(\cdot)$ as the integral in (3.2) replacing λ by λ_i, which is specific to observation y_i, $i = 1,\ldots,n$, and we consider the joint distribution of $(Y,\mu,\sigma,\lambda_1,\ldots,\lambda_n)$. After integrating out μ from this distribution as a multivariate Normal and σ^{-2} as a Gamma distribution, we are left with a bounded function of the λ_i's, which is therefore integrable for any probability distribution on λ_i.

REFERENCES

Basu, D. (1977), "On the Elimination of Nuisance Parameters," *Journal of the American Statistical Association,* 72, 355-366.

Berger, J.O., and Bernardo, J.M. (1992), "On the Development of Reference Priors" (with discussion), in *Bayesian Statistics 4,* eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, Oxford: Oxford University Press, pp. 35-60.

Berger, J.O., Philippe, A., and Robert, C.P. (1996), "Estimation of Quadratic Functions: Noninformative Priors for Non-centrality Parameters," Technical Report 96-10C, Purdue University, Dept. of Statistics.

Bernardo, J.M. (1979), "Reference Posterior Distributions for Bayesian Inference" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 41, 113-147.

Bernardo, J.M., and Ramón, J.M. (1997), "An Elementary Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters," Technical Report 3-97, Universidad de Valencia, Departamento de Estadística.

Bernardo, J.M., and Smith, A.F.M. (1994), *Bayesian Theory.* Chichester: John Wiley.

Casella, G., DasGupta, A., and Strawderman, W.E. (1996), “On a Theory of Proper Default Priors: With Illustrations,” Technical Report 96-1, Purdue University, Dept. of Statistics.

Clarke, B. (1996), “Implications of Reference Priors for Prior Information and for Sample Size,” *Journal of the American Statistical Association,* 91, 173-184.

Clarke, B., and Wasserman, L. (1993), “Noninformative Priors and Nuisance Parameters,” *Journal of the American Statistical Association,* 88, 1427-1432.

Datta, G.S., and Ghosh, M. (1995), “Some Remarks on Noninformative Priors,” *Journal of the American Statistical Association,* 90, 1357-1363.

DeGroot, M.H. (1970), *Optimal Statistical Decisions,* New York: McGraw-Hill.

Fernández, C., Osiewalski, J., and Steel, M.F.J. (1995), “Modeling and Inference With v-Spherical Distributions,” *Journal of the American Statistical Association,* 90, 1331-1340.

Fernández, C., and Steel, M.F.J. (1996), “On Bayesian Inference Under Sampling From Scale Mixtures of Normals,” CentER D.P. 9602, Tilburg University, CentER for Economic Research.

Fernández, C., and Steel, M.F.J. (1998), “On Bayesian Modeling of Fat Tails and Skewness”, *Journal of the American Statistical Association*, 93, forthcoming.

Ghosh, J.K., and Mukerjee, R. (1992), “Non-Informative Priors” (with discussion), in *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, Oxford: Oxford University Press, pp. 321-344.

Jeffreys, H. (1961), *Theory of Probability.* Oxford University Press, third edition.

Kass, R.E., and Wasserman, L. (1996), “The Selection of Prior Distributions by Formal Rules,” *Journal of the American Statistical Association,* 91, 1343-1370.

Osiewalski, J., and Steel, M.F.J. (1993), “Robust Bayesian Inference in l_q-Spherical Models,” *Biometrika,* 80, 456-460.

Stein, C. (1959), “An Examination of Wide Discrepancy Between Fiducial and Confidence Intervals,” *Annals of Mathematical Statistics*, 30, 877-880.

Sun, D., and Berger, J.O. (1997), “Reference Priors With Partial Information,” *Biometrika,* forthcoming.

Tibshirani, R. (1989), “Noninformative Priors for One Parameter of Many,” *Biometrika,* 76, 604-608.

Yang, R., and Berger, J.O. (1996), “A Catalog of Noninformative Priors,” mimeo, Purdue University, Dept. of Statistics.