ELIMINATION OF NUISANCE PARAMETERS WITH REFERENCE NONINFORMATIVE PRIORS

by

Brunero Liseo
University of Rome and
Purdue University

Technical Report #90-58C

Department of Statistics
Purdue University

November 1990

ELIMINATION OF NUISANCE PARAMETERS WITH REFERENCE NON INFORMATIVE PRIORS

1. INTRODUCTION .

The problem of elimination of nuisance parameters in inferential contexts based on parametric statistical models has been producing, in the last years, a wide and often confused literature. Given a statistical model $\left(\mathfrak{X},\ P_\omega,\ \Omega\right)$, where $\mathfrak{X}$ is the sample space, $\left\{P_\omega,\ \omega\epsilon\Omega\right\}$ is a class of probability distributions on $\mathfrak{X}$ and $\Omega\subseteq\mathbb{R}^k$ is the parametric space, we are very often interested only in a function $g_1{:}\Omega\rightarrow\Theta$. The standard case is when g_1 is not invertible and $\Theta\subseteq\mathbb{R}^h$, $h<k$. Actually the complementary transformation $g_2{:}\Omega\rightarrow\Lambda$, with $\Lambda\times\Theta=\Omega$ is seldom indicated and it is never a datum of the problem. In these situations we would need inferential procedures to get rid of the parameter λ and, at the same time, to provide statistical statements on ϑ.

Under special conditions, if one or more components of the sufficient statistics has a marginal or a conditional distribution depending only on ϑ, we can use it as a likelihood: in other words we consider a marginal or a conditional experiment in lieu of the entire one [Kalbfleisch and Sprott(1970); see also Basu(1977) for a critical discussion). The most general and commonly used method is the *profile likelihood* (*pl* hereafter), where the nuisance parameter λ is replaced by its maximum likelihood estimate (for each ϑ) $\hat{\lambda}_\theta$, and the analysis is based on $L(\vartheta,\hat{\lambda}_\theta)$, a likelihood function of ϑ only. However, there exist several examples of bad behaviour of *pl* which have hinted various suggestions of "modifications" to the *pl*. These changes are justified and supported on different grounds: differential geometry [Barndorff-Nielsen(1983,1988); Fraser and Reid (1988)], asymptotics [Cox and Reid (1987)], approximations [McCullogh and

Tibshirani(1989)].

From a theoretical point of view Bayesian techniques seem simpler and more intuitive. Let $\pi(\vartheta,\lambda)$ be a prior distribution over Ω; every inferential statement will be then based on the marginal posterior distribution of ϑ, after observing a sample $\underline{x}=(x_1,\cdots,x_n)$ from an unknown element of P_ω

$$\pi(\vartheta|\underline{x})\propto\int_{\Lambda}\frac{\pi(\vartheta)\pi(\lambda|\vartheta)L(\vartheta,\lambda)d\lambda}{m(\underline{x})}$$

where $L(\vartheta,\lambda)$ is the entire likelihood function and

$$m(\underline{x})=\int_{\Lambda\times\Theta}\pi(\vartheta)\pi(\lambda|\vartheta)L(\vartheta,\lambda)d\lambda d\vartheta$$

is the marginal distribution of $\underline{x}$. Problems can often rise in computional steps. Also, in many situations the parameter λ is considered a nuisance mainly because it has not a clear physical meaning and it can be hard to elicit a prior distribution on it. A possible answer to the second kind of problems can be found in using "automatic" prior distribution, depending only on the choice of the class P_ω. Provided that, in our opinion, one should use *his* prior distribution, it can be very useful to perform a reference Bayesian analysis depending only on the data and the model. Subjective information can be introduced later, with the adoption of different prior distributions, inspecting in this way the robustness of the conclusions. Among a great number of suggestions, the Jeffreys' approach (Jeffreys,1961) and the "reference prior" algorithm (Berger and Bernardo 1989a,1989b) seem the most general and satisfactory. A different way to obtain marginal posteriors, involving approximation concepts, is developped in Leonard et al.(1989) and Kass et al.(1988).

The aim of this paper is to analyze the performances of Bayesian techniques on several examples of theoretical and historical interest, such as the Fieller-Creasy paradox, which is often mentioned as a challenging example in the comparison of different inferential approach. In some examples Bayesian techniques will be compared to likelihood based methods, when available; frequentist coverage of the implied confidence procedure will be considered. In section 2 the Jeffreys' and reference methods will be briefly illustrated, in section 3 some examples will be more deeply discussed.

2. JEFFREYS' AND REFERENCE PRIORS.

The most widely used method for determining a noninformative prior is that of Jeffreys(1961). He proposed to choose

$$\pi_J(\omega)=\sqrt{|H(\omega)|} \tag{2.1}$$

where $H(\omega)$ is the *expected Fisher information*. The Jeffreys' prior is invariant under reparametrization and, therefore, does not depend on which is the parameter of interest. When ω is a vector, however, even Jeffreys is not clear on which prior we should use. The obvious generalization of his rule should be the use of

$$\pi_J(\omega)\propto\sqrt{det|H(\omega)|} \tag{2.2}$$

which is still invariant, but in some examples he seems to prefer assuming the indipendence of single real parameters and multiplying the noninformative priors obtained for each one: it can be seen as a Bayesian version of the orthogonalization ideas of Cox-Reid(1987) (but the results are generally different).

Also, the Jeffreys' priors are usually improper.

The reference prior's approach was proposed by Bernardo(1979) and it has been further developped by Berger and Bernardo(1989a,1989b). The reference prior is the one which maximizes the quantity

$$I(\pi(\omega))=\int_{\Omega}\int_{\mathfrak{X}} log\frac{\pi(\omega|\underline{x})}{\pi(\omega)}\ d\underline{x}\ d\omega. \tag{2.3}$$

i.e. the amount of expected information (in the sense of Shannon-Lindley) provided by an experiment, when $n\rightarrow\infty$. If $\Omega\subseteq\mathbb{R}$ the reference prior is equal to the Jeffreys' one; in multiparameter problems the optimal solution can depend on the ordering and grouping of the parameters according to the inferential interest. Berger and Bernardo particularly recommend the reference prior based on having one parameter per group. This algorithm (see App. A1) can be considered one of the possible generalizations of the univariate Jeffreys' approach, obtained by introducing a well defined measure of information. Even the reference priors are often improper.

Example 2.1: *Many normal means* (Neyman and Scott, 1948).
This very interesting example is known for showing a poor behaviour of the *pl*, easily corrected by any of the different types of adjusted profiles. Same problems arises with a Bayesian analysis using the Jeffreys' invariant prior but not with the reference one.

Let x_{i1}, x_{i2} be two i.i.d. observations from a random variable $X_i \sim N(\mu_i, \sigma^2)$, $i=1,\cdots,n$. We need an estimate for σ^2 and the μ_i's are nuisances. The profile likelihood is

$$pl(\sigma^2) \propto \frac{1}{\sigma^{2n}}\ exp\left\{-\frac{S^2}{2\sigma^2}\right\} \qquad (2.4)$$

with

$$S^2 = \frac{1}{2}\sum_{i=1}^{n}(x_{i1}-x_{i2})^2 \qquad \text{and } S^2 \rightarrow \frac{\sigma^2}{2}\ when\ n \rightarrow \infty.$$

This inconsistent estimate can be easily modified by using one of the several modifications to the *pl*: with different argumentations it can be shown that the *marginal likelihood* (Kalbfleisch and Sprott,1969), the *conditional profile likelihood* (*cpl*) (Cox and Reid,1987), the *modified profile likelihood* (*mpl*) (Barndorff-Nielsen,1988), the *adjusted profile likelihood* (McCullogh and Tibishirani,1989) are all equal to

$$\frac{1}{\sigma^n}\ exp\left\{-\frac{S^2}{2\sigma^2}\right\} = \sigma^n pl(\sigma^2) \qquad (2.5)$$

The Jeffreys prior is computed from the information matrix

$$H(\sigma^2,\mu_1,\cdots,\mu_n)=diag\left[\frac{n}{\sigma^4},\ \frac{2}{\sigma^2},\cdots,\frac{2}{\sigma^2}\right], \tag{2.6}$$

$$\pi_J(\sigma^2,\mu_1,\cdots,\mu_n)\propto\frac{1}{\sigma^{n+2}}. \tag{2.7}$$

Jeffreys' marginal posterior distribution will be then

$$\pi_J(\sigma^2|x_{11},x_{12},\cdots,x_{n1},x_{n2})=\frac{1}{\sigma^{2n+2}}\ exp\left\{-\frac{S^2}{2\sigma^2}\right\} \tag{2.8}$$

leading to the same problems of the profile.

The reference prior, indipendently of the ordering and grouping of the parameters, is (see App. A3)

$$\pi_R(\sigma^2,\mu_1,\cdots,\mu_n)\propto\frac{1}{\sigma^2}. \tag{2.9}$$

with marginal posterior

$$\pi_R(\sigma^2|x_{11},x_{12},\cdots,x_{n1},x_{n2})=$$

$$\frac{1}{\sigma^{n+2}}\ exp\left\{-\frac{S^2}{2\sigma^2}\right\}=\sigma^n\pi_J(\sigma^2|x_{11},x_{12},\cdots,x_{n1},x_{n2}). \tag{2.10}$$

In this example the correction of the reference posterior to the Jeffreys' one is the same as the correction to the *pl* induced by all its *modifications*. Also, note that the integrated likelihood (i.e. the likelihood obtained by integrating the entire likelihood with respect $\pi(\mu_1,\cdots,\mu_n|\sigma^2)$, the conditional prior for the nuisance parameters) would be equal to the modifications to the *pl*, either with the Jeffreys' approach or with the reference algorithm.
However, we will avoid to use integrated likelihood because, if the prior is improper, is not always clear how to factorize $\pi(\vartheta,\lambda)=\pi(\lambda|\vartheta)\pi(\vartheta)$, because of the indeterminateness of conditional probability when $\pi(\Omega)=\infty$.

Example 2.2: *Poisson Ratio.* Suppose that X and Y are indipendent Poisson random variables with means $\lambda\vartheta$ and λ. Our parameter of interest is the mean ratio ϑ. Several likelihood methods gives the same answer in this example. The likelihood function is

$$L(\vartheta,\lambda)\propto\ \lambda^{x+y}\ \vartheta^x\ exp\Big(-\lambda(\vartheta+1)\Big) \tag{2.11}$$

From $\hat{\lambda}_\vartheta = \frac{x+y}{1+\vartheta}$ we obtain the profile likelihood

$$pl(\vartheta) \propto \frac{\vartheta^x}{(1+\vartheta)^{x+y}}. \tag{2.12}$$

The modified profile likelihood is:

$$mpl(\vartheta) = pl(\vartheta)\ |j_\lambda(\vartheta,\hat{\lambda}_\vartheta)|^{-1/2} \left|\frac{\partial \hat{\lambda}}{\partial \hat{\lambda}_\vartheta}\right| \tag{2.13}$$

where $|j_\lambda(\vartheta,\lambda)|$ is the observed Fisher information for λ. In this example

$$|j_\lambda(\vartheta,\hat{\lambda}_\vartheta)| = (1+\vartheta)^2, \qquad \hat{\lambda} = Y$$

and $mpl(\vartheta) = pl(\vartheta)$. *A* conditional pseudo-likelihood is also available if we consider the sampling distribution of $X|T$ where $T = X+Y$.

$$cond.l.(\vartheta) = p(x|\,T = x+y, \lambda, \vartheta) \propto \vartheta^x\ (1+\vartheta)^{-(x+y)} = pl(\vartheta).$$

Fisher information matrix is

$$H(\vartheta,\lambda) = \begin{bmatrix} \frac{\lambda}{\vartheta} & 1 \\ 1 & \frac{1+\vartheta}{\lambda} \end{bmatrix} \tag{2.14}$$

and Jeffreys' prior is easily obtained as

$$\pi_J(\vartheta,\lambda) \propto \frac{1}{\sqrt{\vartheta}} \tag{2.15}$$

The reference prior is (see App. A2)

$$\pi_R(\vartheta,\lambda) \propto \frac{1}{\sqrt{\lambda\vartheta(1+\vartheta)}} \tag{2.16}$$

However, different prior give the same marginal posterior for ϑ,

$$\pi_J(\vartheta|x,y) = \pi_R(\vartheta|x,y) = \frac{\vartheta^{y-1/2}}{(1+\vartheta)^{x+y+1}} \tag{2.17}$$

In general, different parametrizations will give different marginal posteriors for any parameter of interest. In this example we can fully appreciate the importance of the parametrization in a noninformative Bayesian analysis.

3. EXAMPLES

In this section some examples will be showed where different approaches lead to different conclusions. Each of these situations arises interesting theoretical issues.

Example 3.1: *Fieller and Creasy's paradox.* Let x be the sample mean of n i.i.d. observations from $X \sim N(\alpha,1)$ and y the mean of n i.i.d. observations from $Y \sim N(\beta,1)$. We are interested in making inference on the ratio $\vartheta = \alpha/\beta$. As nuisance parameter, we can choose, for computional simplicity, $\lambda = \beta$ (for a different parametrization, see McCullogh and Tibshirani,1990).

This is a simpler version of a famous example proposed by Fieller(1954) and Creasy(1954). The two Authors gave two different solutions to the problem, both based on a fiducial approach, showing a counterexample against the uniqueness of the fiducial argument. In our notation the likelihood function is

$$L(\vartheta,\lambda) \propto exp\Big\{-\tfrac{n}{2}\Big[(x-\lambda\vartheta)^2 + (y-\lambda)^2\Big]\Big\}. \tag{3.1}$$

The maximum likelihood estimate for λ, given ϑ, is $\hat{\lambda}_\vartheta = \dfrac{x\vartheta + y}{1+\vartheta^2}$ and the profile likelihood is

$$pl(\vartheta) \propto exp\Big\{\frac{-n(x-\vartheta y)^2}{2(1+\vartheta^2)}\Big\}. \tag{3.2}$$

The profile likelihood takes its maximum value $\hat{\vartheta}$ at x/y and for $\vartheta \rightarrow \pm\infty$ it does not converge to 0 because

$$\lim_{\vartheta \rightarrow \pm\infty} pl(\vartheta) = exp\,\Big(\frac{-ny^2}{2}\Big).$$

The last fact, even if reasonable, is very hard to manage statistically; especially when the absolute value of the sample ratio $\frac{x}{y}$ is high. In these cases any reasonable likelihood set should be infinite because of the strong support to the tails of Θ. Similar conclusion were shown by Fieller(1954) from a fiducial argument. The *modified* profile likelihood is:

$$mpl(\vartheta)= pl(\vartheta)\ |j_\lambda(\vartheta,\hat{\lambda}_\vartheta)|^{-1/2}|\frac{\partial\hat{\lambda}}{\partial\hat{\lambda}_\vartheta}|=pl(\vartheta)\sqrt{1+\vartheta^2}. \tag{3.3}$$

where $|j_\lambda(\vartheta,\lambda)|$ is the observed Fisher information for λ. It diverges for big ϑ, so its behaviour is even worse.

The Jeffreys' prior is easily computed from the information matrix . It is

$$\pi_J \propto |\lambda|. \tag{3.4}$$

After some algebra the marginal posterior for ϑ is

$$\pi_J(\vartheta|x,y) \propto \frac{pl(\vartheta)}{1+\vartheta^2}\left[\frac{x\vartheta+y}{\sqrt{n(1+\vartheta^2)}}\left(2\Phi\left(\frac{\sqrt{n}(x\vartheta+y)}{\sqrt{1+\vartheta^2}}\right)-1\right)+\frac{2}{n}\ exp\left(-\frac{n(x\vartheta+y)^2}{2(1+\vartheta^2)}\right)\right] \tag{3.5}$$

where Φ is the cdf of a standard normal random variable. Also $\pi_J(\vartheta|x,y)$ was derived, as a fiducial distriution, by Creasy(1954).

The reference prior, already given in Bernardo(1977), with a slightly different argumentation, is

$$\pi_R(\vartheta,\lambda) \propto \frac{1}{\sqrt{1+\vartheta^2}} \tag{3.6}$$

and the marginal posterior for ϑ is

$$\pi_R(\vartheta|x,y) \propto \frac{1}{1+\vartheta^2}\ exp\left(-\frac{n(x\text{-}\vartheta y)^2}{2(1+\vartheta^2)}\right) \tag{3.7}$$

It can be easily shown that both reference and Jeffreys' priors produce proper marginal posteriors for ϑ. When $|x/y|$ is small, both the marginal posteriors show a behaviour similar to the *pl*, in a neighborhood of the maximum, but the tails are different because for each x,y

$$\lim_{\vartheta\to\pm\infty} \pi_J(\vartheta|x,y)= \lim_{\vartheta\to\pm\infty} \pi_R(\vartheta|x,y)=0$$

Jeffreys' and Reference posteriors are dramatically different from *pl* when

$|x/y|$ is bigger (see fig. 1). In fact, while posteriors are bound to be proper, pl is not. So the profile likelihood seems to give very sensible answers but they cannot be used in a probabilistic way; on the other hand Bayesian analysis is sensible *as much as it can*, compatibly with the probability laws.

Even more surprisingly, profile and reference analysis give the same results if we change parametrization, using $\tau = tg^{-1}\vartheta$. Now the parametric space is $\mathrm{T}=\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (considering only the main value of tg) and it can be easily shown that

$$pl(\tau) = \pi_{\mathrm{R}}(\tau|x,y) = exp\left\{-\frac{n}{2}\left(x\ cos\tau - y\ sen\tau\right)^2\right\} \tag{3.8}$$

Bayesian analysis is, by definition, invariant with respect of reparametrizations: this suggests strong counterevidence against the use of the profile likelihood.

It is hard to find a convincing way of comparing different techniques in this example. The profile likelihood is never integrable and cannot be treated as a posterior distribution. We compute the frequentist coverage probability $C_{.95}(\alpha,\beta)$ for the 0.95 (and 0.05) confidence sets obtained with the χ_1^2 approximation (see Kalbfleisch,1983). Following Berger and Bernardo (1989b) we calculate $Pr_{\alpha\beta}(\vartheta \leq \vartheta_\gamma)$, the frequentist probability that ϑ_γ (the posterior quantile from using Jeffreys' and reference prior) is bigger than the actual ϑ, for $\gamma=0.95$ (and 0.05) and various values of (α,β). Table 1 shows the results of simulations generating 5000 pairs (x,y) *for each* (α,β). This was done on SUN X-19 computer using Fortran-IMSL subroutines; the standard error of the entries in Table 1 is about $\sqrt{p(1-p)/\ 5000}$, where p is the entry. Comparisons between posteriors show no big differences. Much harder is to say something about the discrepancies between profile and posteriors. Frequentist properties of χ_1^2 approximation for the pl are of course better, especially when the true ratio α/β is big, but very often the obtained likelihood sets are cofinite. There are no easy conclusions from the comparisons. When β is near to 0, no one method seems to work properly. When β is big, Bayesian solutions seems better, particularly for $\gamma=0.05$. Results of comparisons does not seem to change with n.

Table 1. Frequentist Coverage Probabilities of 0.05 *and* 0.95 *Posterior Quantiles* (π_J *and* π_R) *and of pl based confidence sets.*

	$C_{.95}(\alpha,\beta)$	$P_{.95}(\alpha,\beta)$		$C_{.05}(\alpha,\beta)$	$P_{.05}(\alpha,\beta)$	
(α,β)	pl	π_J	π_R	pl	π_J	π_R
	N=3					
(0.1,0.1)	.948*	.978	.980	.006	.00	.00
(0.1,10)	.951	.947	.947	.004	.046	.046
(1,−5)	.940	.938	.938	.004	.049	.047
(2,2)	.946	.945	.948	.003	.030	.028
(−5,1)	.949	.998	.998	.006	.056	.053
(10,0.1)	.947*	.921	.921	.005	.029	.029
	N=10					
(0.1,0.1)	.952*	.967	.985	.003	.00	.00
(0.1,10)	.953	.969	.969	.00	.063	.063
(10,0.1)	.960*	.826	.825	.007	.00	.00

A "" indicates that more than* 80% *of samples gives a cofinite likelihood set*

Example 3.2. *Gamma distribution.* Let $y_1,\cdots,y_n$ be n observations from $Y\sim\Gamma(\vartheta,\lambda)$ with density function

$$p(y|\vartheta,\lambda)=\left(\frac{\lambda}{\vartheta}\right)^{-\vartheta}\frac{y^{\vartheta-1}}{\Gamma(\vartheta)}\,exp\left(-\frac{\vartheta y}{\lambda}\right). \tag{3.9}$$

This parametrization makes λ and ϑ orthogonal, to be able to use the *conditional profile likelihood* (*cpl*). The shape parameter ϑ is of interest, while $\lambda=E(Y)$ is the nuisance. The log-likelihood function is

$$\mathcal{L}(\vartheta,\lambda)\propto\ -n\ log\Gamma(\vartheta)-n\vartheta\ log\lambda+n\vartheta\ log\ \vartheta\ +(\vartheta-1)\ log\ p\ -\frac{\vartheta}{\lambda}\ t \tag{3.10}$$

where $p=\prod_{i=1}^{n}y_i$ and $t=\sum_{i=1}^{n}y_i$ are the sufficient statistics.

Conditional maximum likelihood of λ is $\hat{\lambda}_\vartheta = t/n$, which does not depend on ϑ (it means that *modified profile likelihood= conditional profile likelihood*). The *profile likelihood* is

$$pl(\vartheta) \propto \frac{1}{\left(\Gamma(\vartheta)\right)^n} \left(\frac{n\vartheta}{t}\right)^{n\vartheta} p^{\vartheta-1}\ exp(-n\vartheta) \tag{3.11}$$

The correction provided by *cpl* (or *mpl*) is $|j_\lambda(\vartheta,\hat{\lambda}_\vartheta)|^{-1/2} = 1/\sqrt{\vartheta}$, and we have

$$cpl(\vartheta) \propto \frac{1}{\left(\Gamma(\vartheta)\right)^n} \left(\frac{n}{t}\right)^{n\vartheta} \vartheta^{n\vartheta-1/2}\ p^{\vartheta-1}\ exp(-n\vartheta) \tag{3.12}$$

The information matrix is

$$H(\lambda,\vartheta) = diag\left[\xi(\vartheta) - 1/\vartheta,\ \vartheta/\lambda^2\right], \tag{3.13}$$

where $\xi(\vartheta) = \frac{\partial^2}{\partial\vartheta^2} log\ \Gamma(\vartheta)$ is the Trigamma function (see Abramowitz and Stegun,1964)

and the Jeffreys' prior is

$$\pi_J(\vartheta,\lambda) \propto \frac{1}{\lambda}\sqrt{\vartheta\xi(\vartheta)-1}. \tag{3.14}$$

The reference algorithm does not take on account the presence of ϑ in i_λ and the resulting prior is (see App. A4)

$$\pi_R(\vartheta,\lambda) \propto \frac{1}{\lambda}\sqrt{\xi(\vartheta) - \frac{1}{\vartheta}} \tag{3.15}$$

Note that $\frac{\pi_R(\vartheta,\lambda)}{\pi_J(\vartheta,\lambda)} = \frac{cpl(\vartheta)}{pl(\vartheta)} = \frac{1}{\sqrt{\vartheta}}$. This relation, far from being a general rule, is very

common in problems involving shape parameters. After some algebra, marginal posteriors are

$$\pi_J(\vartheta|y_1,\cdots,y_n) = \sqrt{\vartheta\xi(\vartheta)-1}\ \frac{\Gamma(\vartheta n)}{\left[\Gamma(\vartheta)\right]^n}\ \frac{p^{\vartheta-1}}{t^{n\vartheta}} \tag{3.16}$$

and

$$\pi_R(\vartheta|y_1,\cdots,y_n) = \sqrt{\xi(\vartheta) - \frac{1}{\vartheta}}\ \frac{\Gamma(\vartheta n)}{\left[\Gamma(\vartheta)\right]^n}\ \frac{p^{\vartheta-1}}{t^{n\vartheta}} = \frac{1}{\sqrt{\vartheta}}\ \pi_J(\vartheta|y_1,\cdots,y_n). \tag{3.17}$$

The following proposition holds:

PROPOSITION 1. For every value of (t,p) the marginal posteriors $\pi_J(\vartheta|y_1,\cdots,y_n)$ and $\pi_R(\vartheta|y_1,\cdots,y_n)$ are proper.

Proof: see App. A5.

One way of comparing the different methods is to consider *pl* and *cpl* as probability distributions over ϑ and to check frequentist coverage probability of *pl*, *cpl*, $\pi_J(\vartheta|y_1,\cdots,y_n)$ and $\pi_R(\vartheta|y_1,\cdots,y_n)$, as in the example 3.1. As stated in Berger and Bernardo (1989b): ≪ There is no guarantee that this approach will work; (.....) sensible conditional behaviour and uniform frequentist properties are often simply not compatible ≫. Table 2 shows coverage probabilities for a variety of values of (ϑ,λ). Simulations were done generating 5000 samples for each (ϑ,λ); the standard error of the entries in Table 2 can be estimated by $\sqrt{p(1-p)/\ 5000}$, where p is the entry.

Table 2 Frequentist Coverage Probabilities of .95 and .05 Posterior and Likelihood Quantiles

	$P_{.95}(\vartheta,\lambda)$				$P_{.05}(\vartheta,\lambda)$			
(ϑ,λ)	*pl*	*cpl*	π_J	π_R	*pl*	*cpl*	π_J	π_R
	N=3							
(0.5,1)	.965	.953	.937	.921	.405	.295	.157	.064
(1,2)	.996	.991	.973	.948	.414	.291	.148	.054
(2,1)	.994	.988	.970	.953	.419	.296	.161	.054
(5,3)	.997	.992	.981	.949	.430	.290	.154	.040
	N=10							
(1,2)	.902	.895	.886	.869	.181	.134	.092	.060
(2,1)	.921	.896	.879	.842	.180	.131	.092	.061

When $n=3$ evidence in favour of the reference prior is overwhelming, especially at the 0.05 quantile which is generally too high for *cpl* and *pl.* Jeffreys' posterior is better than profiles but inferior to the reference one. Further simulations, using bigger sample sizes ($n=5$ and 10), have given less clear results: the supremacy of the reference is no more so evident. The suspect is that Jeffreys and reference priors give high probability to small values of ϑ: then, if n is small, the variability of "possible likelihoods" is made smaller by the prior but, for bigger n, a contrast between peaks in the prior and in the likelihood can be possible. In these cases it can be a good device the use of *integrated likelihood* (i.e. to use the reference conditional prior $\pi_R(\lambda|\vartheta)$ only)

$$il(\vartheta)=\int_\Lambda \pi(\lambda|\vartheta)L(\vartheta,\lambda)\ d\lambda. \tag{3.18}$$

However, as stated in section 2, this must be considered only an *ad hoc* device: noninformative priors are often improper and $\pi(\lambda|\vartheta)$ can be not univocally determined. This is in my opinion, the main reason for considering the improper priors as not fully justified and treating them only as an useful starting point in a Bayesian analysis.

Example 3.3. *Inverse Gaussian distribution.* Let $x_1,\cdots,x_n$ be n i.i.d. observations from $X\sim IG(\vartheta,\lambda)$, with density function

$$p(x|\vartheta,\lambda)=\sqrt{\frac{\vartheta}{2\pi}}\ \ exp\left(\sqrt{\vartheta\lambda}\right)\ x^{-\frac{3}{2}}\ exp\left\{-\frac{1}{2}\left(\frac{\vartheta}{x}+\lambda x\right)\right\} \tag{3.19}$$

We want to estimate the shape parameter ϑ, and λ is the nuisance parameter. This problem is deeply discussed by Barndorff-Nielsen(1983,1988) as an example of the use of the modified profile likelihood. Also it is interesting because the reference algorithm provides different prior if different parametrizations of the nuisance are used. The likelihood function can be written

$$L(\vartheta,\lambda)\propto\ \vartheta^{n/2}\ exp\left\{-\frac{n}{2}\left(\vartheta h+\lambda t-2\sqrt{\vartheta\lambda}\right)\right\} \tag{3.20}$$

where $\left(h=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i},\ t=\frac{1}{n}\sum_{i=1}^{n}x_i\right)$ is the sufficient statistics. Conditional maximum likelihood estimate of λ is $\hat{\lambda}_\vartheta=\frac{\vartheta}{t^2}$ and the profile likelihood is

$$pl(\vartheta)\propto \vartheta^{\frac{n}{2}}\ exp\left\{-\frac{n\vartheta}{2}\left(h-\frac{1}{t}\right)\right\} \tag{3.21}$$

which, after normalization, can be considered a Gammma density function. The "correction" factor provided by the modified profile likelihood is $1/\sqrt{\vartheta}$ and we have

$$mpl(\vartheta)\propto \vartheta^{\frac{n-1}{2}}\ exp\left\{-\frac{n\vartheta}{2}\left(h-\frac{1}{t}\right)\right\} \tag{3.22}$$

which is also a Gamma density function.

Jeffreys' prior is computed from the information matrix

$$H(\vartheta,\lambda)=\begin{bmatrix} \frac{2+\sqrt{\lambda\vartheta}}{4\vartheta^2} & -\frac{1}{4\sqrt{\lambda\vartheta}} \\ & \\ -\frac{1}{4\sqrt{\lambda\vartheta}} & \frac{1}{4}\sqrt{\frac{\vartheta}{\lambda^3}} \end{bmatrix} \tag{3.23}$$

and it is

$$\pi_J(\vartheta,\lambda)\propto\left(\vartheta\lambda\right)^{-\frac{3}{4}} \tag{3.24}$$

The marginal posterior distribution for ϑ has not an easy analytical form:

$$\pi_J(\vartheta|x_1,\cdots,x_n)=\vartheta^{\frac{n}{2}-\frac{3}{4}}\ exp\left\{-\frac{n\vartheta h}{2}\right\}\int_0^\infty \lambda^{-\frac{3}{4}}\ exp\left\{-\frac{n}{2}\left(\lambda t-2\sqrt{\vartheta\lambda}\right)\right\}d\lambda. \tag{3.25}$$

Palmer(1973) shows that Jeffreys' posterior is proper for every (h,t).
The reference prior which comes naturally from this parametrization (i. e. by considering a sequence of increasing rectangles $\Omega_m\subseteq\Theta\times\Lambda$) is (see App. A6)

$$\pi_R(\vartheta,\lambda)\propto\vartheta^{-1}\lambda^{-\frac{3}{4}} \tag{3.26}$$

and

$$\pi_R(\vartheta|x_1,\cdots,x_n)=\vartheta^{\frac{n}{2}-1}\ exp\left\{-\frac{n\vartheta h}{2}\right\}\int_0^\infty \lambda^{-\frac{3}{4}}\ exp\left\{-\frac{n}{2}\left(\lambda t-2\sqrt{\vartheta\lambda}\right)\right\}d\lambda \tag{3.27}$$

$$=\vartheta^{-\frac{1}{4}}\pi_J(\vartheta|x_1,\cdots,x_n).$$

Let us consider now a new parametrization (ϑ,β) such that $\beta=\sqrt{\frac{\lambda}{\vartheta}}$: the new density function is

$$p(x|\vartheta,\beta)=\sqrt{\frac{\vartheta}{2\pi}}\ exp\left\{\beta\vartheta\right\}\ x^{-\frac{3}{2}}\ exp\left\{-\frac{\vartheta}{2}\left(\frac{1}{x}+\beta^2x\right)\right\} \tag{3.28}$$

The profile and the modified profile likelihoods do not change with the new parametrization; also, because of the orthogonality of β and ϑ, a conditional profile can be computed and it is equal to *mpl*. Jeffreys' prior is easily obtained via Jacobian transformation:

$$\pi^*_J(\vartheta,\beta)\propto\frac{1}{\sqrt{\vartheta\beta}} \tag{3.29}$$

With the same transformation one obtains the reference prior for (ϑ,β)

$$\pi^*_R(\vartheta,\beta)\propto\beta^{-\frac{1}{2}}\ \vartheta^{-\frac{3}{4}} \tag{3.30}$$

However this reference prior is different from the one obtained by choosing an increasing sequence of rectangles in $\Theta\times B$, that will be said π_{R1}. In fact (see App. A6)

$$\pi_{R1}(\vartheta,\beta)\propto\beta^{-\frac{1}{2}}\ \vartheta^{-1} \tag{3.31}$$

Therefore different priors are derived from different parametrizations of the nuisance and it is hard to say what is the most natural, even if the use of the euristic device of choosing the reference prior such that

$$\pi_R=\frac{1}{\sqrt{\vartheta}}\ \pi_J$$

would make chosen π_{R1}.

4. Conclusions.

We do not think there exists a "best" method of elimination of nuisance parameters. For each one it is possible to find examples of poor behaviour. Even the profile likelihood cannot be discarded in favour of other non Bayesian techniques (see Example 3.1).

Neverthless there is a great difference among Bayesian and non Bayesian methods, a difference which makes us to be Bayesians: likelihood based methods provide conclusions which does not depend on the real meaning of the parameters. The output of the analysis is a pseudolikelihood weighting different values of a Greek letter (Lindley, 1990). Bayesian analysis is much more flexible: if, in some situations, thinking of a Greek letter is not enough, a noninformative automatic prior should be avoided and one can try to improve results using a more accurate and specific prior.

REFERENCES

Abramowitz, M. & Stegun, I.A. (1964). *Handbook of Mathematical Functions.* National Bureau of Standards, Washington.

Barndorff-Nielsen, O. (1983). On a formula for the distriution of the maximum likelihood estimator. *Biometrika*,**70**, 2, 343-365.

Barndorff-Nielsen, O. (1988). *Parametric Statistical Models and Likelihood.* Lecture notes in Statistics, 50. Springer Verlag, New York.

Basu, D. (1977). On the elimination of nuisance parameters. *J. Amer. Statist. Assoc.*72, 355-366.

Berger, J.O. & Bernardo, J.M. (1989a). Ordered group reference priors with applications to multinomial and variance components. Technical report 89-31c, Purdue University, Dept of Statistics.

Berger, J.O. & Bernardo, J.M. (1989b). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* 84, 200-207.

Bernardo, J.M. (1977). Inferences about the ratio of normal means: a Bayesian approach to the Fieller-Creasy problem. In *Recent Development in Statistics* (*J. R. Barra et al., eds*), 345-349. North-Holland, Amsterdam.

Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference (with Discussion). *J.R. Statist. Soc. B* 41, 113-147.

Cox, D.R. & Reid, N. (1987). Parameter Orthogonality and Approximate Conditional Inference. *J. R. Statist.Soc. B* **49**, 1-39.

Creasy, M.A. (1954). Limits for the ratio of the means. *J. R. Statist. Soc. B* **16**, 186-194.

Fieller, E.C. (1954). Some problems in interval estimation. *J. R. Statist. Soc. B* 16, 175-185.

Fraser, D.A.S. & Reid, N. (1988). Adjustments to Profile Likelihood. Technical Report No. 10, University of Toronto, Dept. of Statist.

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

Kalbfleisch, J. (1985). *Probability and Statistical inference.* (*2nd ed.*) *Vol. II.* Springer Verlag, New York.

Kalbfleisch, J. & Sprott, D. (1970). Application of Likelihood Methods to Models Involving Large Numbers of Parameters(with Discussion). *J.R. Statist. Soc. B* 32, 175-208

Kalbfleisch, J. & Sprott, D. (1974). Marginal and Conditional Likelihood. *Sankhya, A* **35**, 311-328.

Kass, R. , Tierney, L. & Kadane, J. (1988). Asymptotics in Bayesian Computation. In *Bayesian Statistics*, 3 (*J.M. Bernardo et al. eds*), 261-278. Oxford University Press.

Leonard, T., Hsu, J.S.J. & Tsui, K. (1989). Bayesian Marginal Inference. *J. Amer. Statist. Assoc.* 84, 1051-1058.

Lindley, D.V. (1990). The Present Position in Bayesian Statistics. *Statistical science* 5, *No.*1, 44-89.

McCullogh, P. & Tibshirani, R. (1989). A simple method for the adjustment of profile likelihoods. Preliminary Report.

Neyman, J. & Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica.* 16, 1-32.

Palmer, J. (1973). Certain non classical inference procedures applied to the Inverse Gaussian distribution. *Ph.D. Dissertation.* Oklahoma State University

APPENDIX A1. *The reference algorithm.*

First we briefly sketch the algorithm from Berger and Bernardo (1989a) for computing ordered group reference priors. Let $\omega=(\omega_{(1)},\omega_{(2)},\cdots,\omega_{(r)})$ be the r groups of parameters;

each $\omega_{(i)}=(\omega_{i1},\cdots,\omega_{in_i})$ has size n_i. Let us define

$$\omega_{[i]}=(\omega_{(1)},\cdots,\omega_{(i)}) \text{ and } \omega_{[\sim i]}=(\omega_{(i+1)},\omega_{(i+2)},\cdots,\omega_{(r)})$$

Let $S(\omega)= H^{-1}(\omega)$. Corresponding to the partition $(\omega_{(1)},\omega_{(2)},\cdots,\omega_{(r)})$ we can define the decomposition of $S(\omega)$ as

$$S(\omega)=\left(S_{ij}\right)_{i,j=1,\cdots,r}$$

Define

$$S_i=\left(\mathrm{S}_{kj}\right)_{k,j=1,\cdots,i}, \quad \text{and} \quad H_i=\mathrm{S}_i^{-1}=\left(H_{kj}\right)_{k,j=1,\cdots,i}$$

using the same decomposition, and $h_i=det(H_{ii})$.

Suppose $\Omega^1\subseteq\Omega^2\subseteq\cdots$ are a sequence of compact subsets of Ω such that $\bigcup_{m=1}^{\infty}\Omega^i=\Omega$.

Define

$$\Omega^m_{\omega_{[i]}}=\left\{\omega_{(i+1)}:\ (\omega_{[i]},\omega_{(i+1)},\omega^*_{[\sim(i+1)]}\in\Omega^m \textit{ for some } \omega^*_{[\sim(i+1)]}\right\}$$

and the indicator function

$$I^m(x|\omega_{[i]})=I_{\Omega^m_{\omega_{[i]}}}(\omega)=\begin{cases}1 & x\in\Omega^m_{\omega_{[i]}}\\ 0 & x\notin\Omega^m_{\omega_{[i]}}\end{cases}.$$

Step 1. Define

$$\pi^m_r(\omega_{[-(r-1)]}|\omega_{[r-1]})=\frac{|h_r(\omega)|^{1/2}I^m(\omega_{(r)}|\omega_{[r-1]})}{\int_{\Omega^m_{\omega_{[r-1]}}}|h_r(\omega)|^{1/2}I^m(\omega_{(r)}|\omega_{[r-1]})d\omega_{(r)}}$$

Then we proceed by iterating on the groups defining, for $i=r-1,\ r-2,\ \cdots,1,\ \pi^m_i$ in terms of the previously defined $\pi^m_j,\quad j=i+1,\ \cdots,r$, as

$$\pi^m_i(\omega_{[-(i-1)]}|\omega_{[i-1]})=\ \pi^m_{i+1}(\omega_{[-i]}|\omega_{[i]})\ \cdot$$

$$\cdot\ \frac{exp\left\{\frac{1}{2}E^m_\omega\left[log|h_i(\omega)|\ |\omega_{[i]}\right]\right\}I^m(\omega_{(i)}|\omega_{[i-1]})}{\int_{\Omega^m_{\omega_{[i-1]}}}exp\left\{\frac{1}{2}E^m_\omega\left[log|h_i(\omega)|\ |\omega_{[i]}\right]\right\}I^m(\omega_{(i)}|\omega_{[i-1]})d\omega_{(i)}}$$

where

$$E^m_\omega\left[f(\omega)|\omega_{[i]}\right]=\int_{*}f(\omega)\quad\pi^m_{i+1}(\omega_{[\sim i]}|\omega_{[i]})d\omega_{[\sim i]},$$

with $*=\left\{\omega_{[\sim i]}:\left(\omega_{[i]},\omega_{[\sim i]}\right)\in\Omega^m\right\}$.

Finally $\omega_{[\sim 0]}$ is interpreted as ω and $\omega_{[0]}$ as vacuous; therefore $\pi^m(\omega)=\pi^m_1(\omega_{[\sim 0]}|\omega_{[0]})$. The reference prior is then defined as

$$\pi(\omega)=\lim_{m\to\infty}\ \frac{\pi^m(\omega)}{\pi^m(\omega_0)},$$

provided it exists for some $\omega_0\in\Omega^1$.

APPENDIX A2.

$$H^{-1}(\vartheta,\lambda)=S(\vartheta,\lambda)=\begin{bmatrix} \frac{(1+\vartheta)\vartheta}{\lambda} & -\vartheta \\ -\vartheta & \lambda \end{bmatrix}$$

$$h_1=\frac{\lambda}{(1+\vartheta)\vartheta}, \quad h_2=\frac{1+\vartheta}{\lambda}$$

We will define the ranges for ϑ and λ to be $\vartheta\in[0,a_m]$ and $\lambda\in[0,b_m]$, find the reference priors for these ranges and let a_m, $b_m \to\infty$ as $m\to\infty$. K's will indicate constants.

$$\pi^m(\lambda|\vartheta)=\frac{1}{\sqrt{\lambda}}\ K(b_m)\ I_{B_m}(\lambda)$$

where $B_m=[0,b_m]$.

$$E_\vartheta\left[log\frac{\lambda}{(1+\vartheta)\vartheta}|\vartheta\right]=K(b_m)-log\ ((1+\vartheta)\vartheta).$$

$$\pi^m(\vartheta,\lambda)=K(a_m,b_m)\frac{1}{\sqrt{\lambda(1+\vartheta)\vartheta}}\ I_{A_m\times B_m}$$

where $A_m=[0,a_m]$.

Finally

$$\pi_{\mathrm{R}}(\vartheta,\lambda)=\lim_{m\to\infty}\frac{\pi^m(\vartheta,\lambda)}{\pi^m(1,1)}=\frac{1}{\sqrt{\lambda(1+\vartheta)\vartheta}}.$$

APPENDIX A3.

$$H^{-1}(\sigma^2,\ \mu_1,\cdots,\mu_n)=diag\left[\frac{\sigma^4}{n},\ \frac{\sigma^2}{2},\cdot\ \cdot\ \cdot,\frac{\sigma^2}{2}\right]$$

For every ordering and grouping of $(\mu_1,\cdots,\mu_n)$, the reference prior will be the same. We will show calculations for $(\sigma^2,(\mu_1,\cdots,\mu_n))$.

Ranges will be $\sigma^2 \in [a_m,+\infty]$, $\mu_i \in \left[b_m^{(i)},c_m^{(i)}\right]$, $i=1,\cdots,n$, with $a_m \to 0$, $b_m^{(i)} \to -\infty$, $c_m^{(i)} \to +\infty$.

$$h_1=\frac{n}{\sigma^4}, \qquad h_2=\frac{2^n}{\sigma^{2n}}.$$

$$\pi^m(\mu_1,\cdots,\mu_n|\sigma^2)=K(b_m^{(1)},c_m^{(1)},\cdots,b_m^{(n)},c_m^{(n)})\ I_{A_m}(\mu_1,\cdots,\mu_n)$$

where $A_m=\left[[b_m^{(1)},c_m^{(1)}]\times\cdot\ \cdot\ \cdot\times[b_m^{(n)},c_m^{(n)}]\right]$

$$E_{\sigma^2}\left(\log\frac{n}{\sigma^4}|\mu_1,\cdots,\mu_n\right)=\log\frac{n}{\sigma^4}$$

and

$$\pi^m\left(\sigma^2,\mu_1,\cdots,\mu_n\right)= K(a_m,b_m^{(1)},c_m^{(1)},\cdots,b_m^{(n)},c_m^{(n)})\frac{1}{\sigma^2}\ I_{A_m}(\mu_1,\cdots,\mu_n)\ I_{[a_m,\infty]}(\sigma_2)$$

Finally

$$\pi_{\mathrm{R}}\left(\sigma^2,\mu_1,\cdots,\mu_n\right)=\lim_{m\to\infty}\frac{\pi^m\left(\sigma^2,\mu_1,\cdots,\mu_n\right)}{\pi^m\left(1,0,\cdots,0\right)}=\frac{1}{\sigma^2}.$$

APPENDIX A4.

$$H^{-1}(\vartheta,\lambda)=\ diag\left[\frac{\vartheta}{\vartheta\xi(\vartheta)-1},\ \frac{\lambda^2}{\vartheta}\right]$$

$$h_1=\xi(\vartheta)-\frac{1}{\vartheta}, \qquad h_2=\frac{\vartheta}{\lambda^2}$$

Ranges are: $\lambda\in[a_m,+\infty]$ and $\vartheta\in[b_m,c_m]$ with $a_m,b_m\to 0$ and $c_m\to+\infty$.
Since h_1 does not depend on λ,

$$\pi^m(\vartheta,\lambda)=\left(\prod_{i=1}^{2}\frac{|h_i|^{1/2}}{\int\limits_{\Omega^m(\omega_{(i)})}|h_i|^{1/2}\ d\omega_{(i)}}\right)I_{\Omega_{(i)}}(\omega)=$$

$$=\frac{|h_1|^{1/2}}{\int_{b_m}^{c_m}|h_1|^{1/2}\ d\vartheta}\ I_{[b_m,c_m]}(\vartheta)\ \frac{|h_2|^{1/2}}{\int_{a_m}^{\infty}|h_2|^{1/2}\ d\lambda}\ I_{[a_m,+\infty]}(\lambda)=$$

$$=K(a_m,b_m,c_m)\frac{1}{\lambda}\sqrt{\xi(\vartheta)-\frac{1}{\vartheta}}\; I_{[a_m,+\infty]}(\lambda)\; I_{[b_m,c_m]}(\vartheta)$$

and

$$\pi_{\mathrm{R}}(\vartheta,\lambda) = \lim_{m\to\infty} \frac{\pi^m(\vartheta,\lambda)}{\pi^m(1,1)} = \frac{1}{\lambda}\sqrt{\xi(\vartheta)-\frac{1}{\vartheta}}$$

APPENDIX A5: *Proof of Proposition* 1.

Because of the continuity of $\pi_{\mathrm{J}}(\vartheta|y_1,\cdots,y_n)$ and $\pi_{\mathrm{R}}(\vartheta|y_1,\cdots,y_n)$ and of the relation

$$\frac{\pi_{\mathrm{J}}(\vartheta|y_1,\cdots,y_n)}{\pi_{\mathrm{R}}(\vartheta|y_1,\cdots,y_n)} = \sqrt{\vartheta}$$

it will be sufficient to show that

(a) $\lim_{\vartheta\to 0} \pi_{\mathrm{R}}(\vartheta|y_1,\cdots,y_n) = 0$

(b) $\pi_{\mathrm{J}}(\vartheta|y_1,\cdots,y_n)$ has an integrable right tail.

To prove (a) we will use the following relations (Abramowitz and Stegun,1964):

$$\Gamma(n\vartheta) = (2\pi)^{\frac{1-n}{2}}\; n^{n\vartheta-\frac{1}{2}} \prod_{k=1}^{n-1} \Gamma(\vartheta+\tfrac{k}{n}) \qquad (A5.1)$$

$$\xi(\vartheta)=\sum_{k=0}^{\infty} \frac{1}{(\vartheta+k)^2} \qquad \vartheta\neq 0,-1,-2,\ldots\ldots \qquad (A5.2)$$

$$\lim_{\vartheta\to 0} \sqrt{\xi(\vartheta)-\frac{1}{\vartheta}}\; \frac{\Gamma(n\vartheta)}{\left(\Gamma(\vartheta)\right)^n}\; \frac{p^{\vartheta-1}}{t^{n\vartheta}} =$$

$$= K(t,p,n) \lim_{\vartheta\to 0} \frac{\sqrt{\sum_{k=0}^{\infty} \frac{1}{(\vartheta+k)^2} - \frac{1}{\vartheta}}}{\Gamma(\vartheta)^{n-2}} \cdot \lim_{\vartheta\to 0} \frac{n^{n\vartheta-\frac{1}{2}}}{\Gamma(\vartheta)} \prod_{k=1}^{n-1} \Gamma(\vartheta+\tfrac{k}{n}).$$

and both the limits are zero.

(b).

$$\pi_J(\vartheta|y_1,\cdots,y_n) = \sqrt{\vartheta\xi(\vartheta)-1}\, \frac{\Gamma(\vartheta \mathrm{n})}{\left[\Gamma(\vartheta)\right]^n} \frac{\mathrm{p}^{\vartheta-1}}{\mathrm{t}^{\mathrm{n}\vartheta}}$$

When $\vartheta\to+\infty$, the following relation holds (Abramowitz and Stegun, 1964)

$$\xi(\vartheta)\sim\frac{1}{\vartheta}+\frac{1}{2\vartheta^2}+\frac{1}{6\vartheta^3}-\frac{1}{30\vartheta^5}+\frac{1}{42\vartheta^7}\cdot\cdot\cdot \tag{A5.3}$$

Therefore

$$\sqrt{\vartheta\xi(\vartheta)-1} = O\Big(\frac{1}{\sqrt{\vartheta}}\Big) \tag{A5.4}$$

Also, using formulae ($A5.1$) and Stirling's approximation we obtain, for $\vartheta\to+\infty$,

$$\frac{\Gamma(\vartheta \mathrm{n})}{\left[\Gamma(\vartheta)\right]^n} \sim (2\pi)^{\frac{1-n}{2}}\, n^{n\vartheta}\, \vartheta^{\frac{n+1}{2}}$$

Finally, using a well known inequality between geometric and arithmetic means,

$$\frac{\mathrm{n}^n \mathrm{p}}{\mathrm{t}^n} \leq 1$$

and

$$\frac{\Gamma(\vartheta \mathrm{n})}{\left[\Gamma(\vartheta)\right]^n} \frac{\mathrm{p}^{\vartheta-1}}{\mathrm{t}^{\mathrm{n}\vartheta}} = O\Big(k^{-\vartheta}\Big) \qquad (k>1)$$

except for the null probability case ($y_1= y_2=\cdots=y_n= 1$), and it concludes the proof.

APPENDIX A6.

(a) Reference prior from (ϑ,λ) parametrization.

$$H^{-1}(\vartheta,\lambda)=\begin{bmatrix} 2\vartheta^2 & 2\lambda\vartheta \\ 2\lambda\vartheta & \dfrac{4\sqrt{\lambda^3}+\lambda^2\sqrt{\vartheta}}{\sqrt{\vartheta}} \end{bmatrix}$$

$$h_1=\frac{1}{\vartheta^2}, \qquad h_2=\sqrt{\frac{\vartheta}{\lambda^3}}$$

Ranges are: $\lambda\in[0,\ a_m]$ and $\vartheta\in[b_m,c_m]$ with $b_m\to 0$ and $a_m,c_m\to+\infty$.

$$\pi^m(\lambda|\vartheta)=K(a_m)\ \lambda^{-\frac{3}{4}}I_{[0,\ a_m]}$$

Since h_1 does not depend on λ,

$$\pi^m(\vartheta,\lambda)=K(a_m,b_m,c_m)\ \lambda^{-\frac{3}{4}}\ \vartheta^{-1}I_{[0,\ a_m]}\ I_{[b_m,c_m]}$$

and

$$\pi_{\mathrm{R}}(\vartheta,\lambda)=\lim_{m\to\infty}\frac{\pi^m(\vartheta,\lambda)}{\pi^m(1,1)}=\lambda^{-\frac{3}{4}}\ \vartheta^{-1}$$

(b) Reference prior for (ϑ,β) parametrization

$$H(\vartheta,\beta)=\ diag\left[\frac{1}{2\vartheta^2},\ \frac{\vartheta}{\beta}\right], \qquad S(\vartheta,\beta)=H^{-1}(\vartheta,\beta)=diag\left[2\vartheta^2,\ \ \frac{\beta}{\vartheta}\right]$$

$$h_1=\frac{1}{2\vartheta^2}, \qquad h_2=\frac{\vartheta}{\beta}$$

Ranges are $\vartheta\in[a_m,b_m]$ and $\lambda\in[0,\ c_m]$ with $a_m\to 0$ and $b_m,c_m\to+\infty$.

$$\pi^m(\beta|\vartheta)=K(c_m)\ \beta^{-\frac{1}{2}}I_{[0,\ c_m]}$$

Also in this case h_1 does not depend on λ and

$$\pi^m(\vartheta,\beta)=K(a_m,b_m,c_m)\ \beta^{-\frac{1}{2}}\ \vartheta^{-1}I_{[0,\ c_m]}\ I_{[a_m,b_m]}$$

Finally

$$\pi_{\mathrm{R}}(\vartheta,\lambda)=\lim_{m\to\infty}\frac{\pi^m(\vartheta,\lambda)}{\pi^m(1,1)}=\beta^{-\frac{1}{2}}\ \vartheta^{-1}.$$

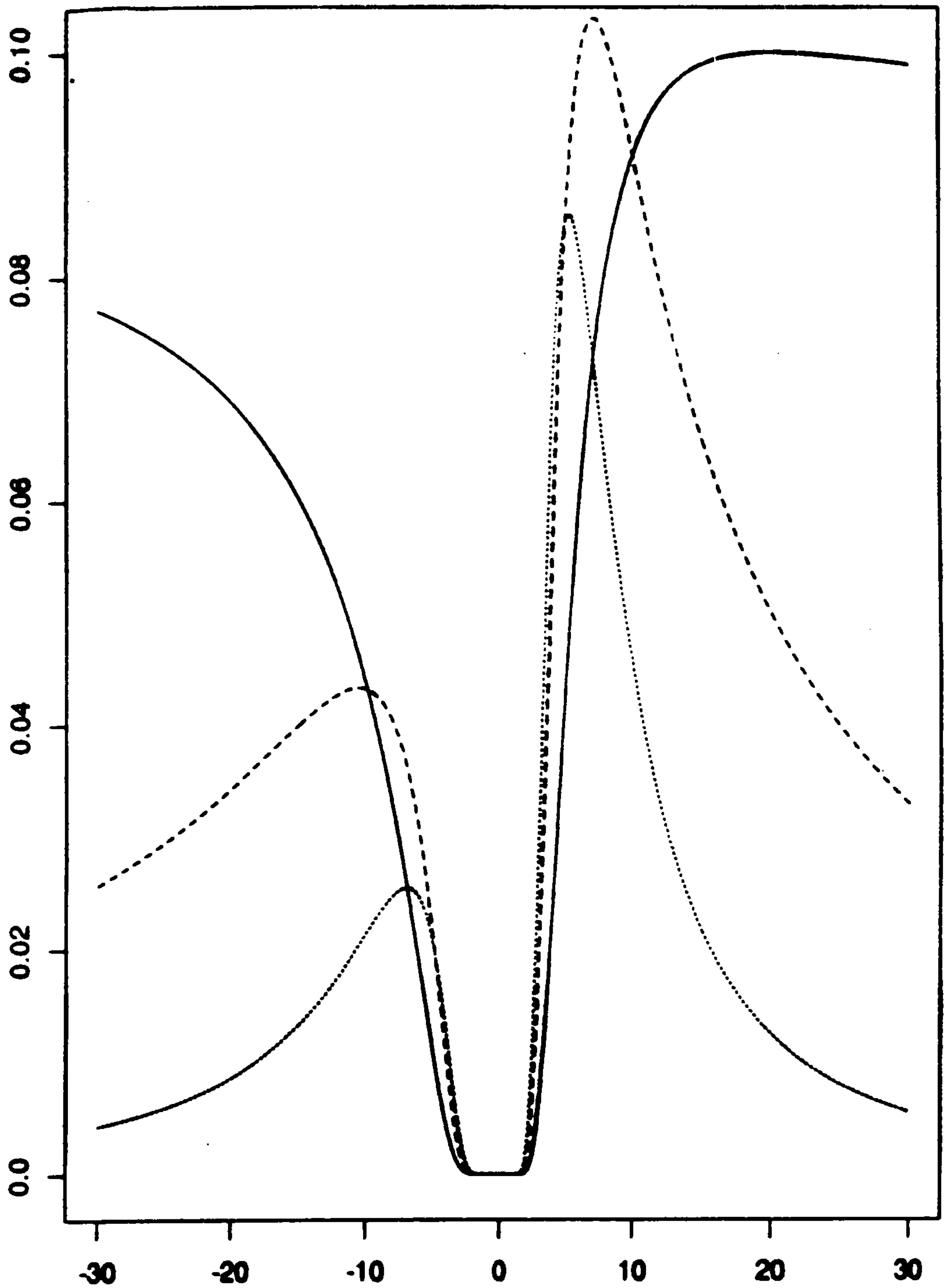

Fig. 1. Fieller-Creasy paradox: pl (ϑ) *(continuous line),*
$\pi_J(\vartheta|x,y)$ *(slash) and* $\pi_R(\vartheta|x,y)$ *(dot) when* $x=5$, $y=0.25$.