Departament d'Estadística i I.O., Universitat de València.
Facultat de Matemàtiques, 46100–Burjassot, València, Spain.
Tel. +34.96.364.3560 (direct), +34.96.386.4362 (office).
Fax +34.96.364.3560 (direct), +34.96.386.4735 (office).
Internet: jose.m.bernardo@uv.es, Web: http://www.uv.es/˜bernardo/

An Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters

JOSÉ M. BERNARDO and JOSÉ M. RAMÓN

Universitat de València, Spain

SUMMARY

This paper offers an introduction to Bayesian reference analysis, often regarded as the more successful method to produce non-subjective, model-based, posterior distributions. The ideas are illustrated with an interesting problem, the ratio of multinomial parameters, for which no model-based Bayesian analysis has been proposed. Signposts are provided to the huge related literature.

Keywords: AMOUNT OF INFORMATION; BAYESIAN ASYMPTOTICS; BAYESIAN INFERENCE; DEFAULT PRIORS; FISHER MATRIX; NON-INFORMATIVE PRIORS; REFERENCE PRIORS.

1. INTRODUCTION

From a Bayesian perspective, the outcome of *any* inference problem is the *posterior distribution of the quantity of interest*, which combines the information provided by the data with available prior information; it has been often recognised that there is a pragmatically important need for a form of prior to posterior analysis which captures, in a *well-defined sense*, the notion that the prior should have a minimal effect, relative to the data, on the posterior inference. We will generally denote by $\pi(\phi \mid \boldsymbol{x})$ a model-based, non-subjective posterior density of a quantity of interest ϕ conditional on data $\boldsymbol{x}$, for which a probability model $p(\boldsymbol{x} \mid \phi, \boldsymbol{\lambda})$ is assumed which may also depend on a vector $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_m\}$ of nuisance parameters.

In the long, fascinating history of the quest for these "baseline" posterior distributions, a number of requirements have emerged which may reasonably be regarded as necessary properties of an algorithm designed to produce such non-subjective posteriors:

(i) *Invariance with respect to one-to-one transformations*. (Jeffreys, 1946; Jaynes, 1968; Kass, 1989; Dawid, 1983; Yang, 1995; Datta and Ghosh, 1996). The posterior $\pi(\phi \mid \boldsymbol{x})$ with respect to model $p(\boldsymbol{x} \mid \phi)$ *must* be consistent with the posterior $\pi(\theta \mid \boldsymbol{x})$ with respect to $p(\boldsymbol{x} \mid \theta)$, where $\theta = \theta(\phi)$ is a one-to-one function of ϕ, so that, for all $\boldsymbol{x}$,

$$\pi(\phi \mid \boldsymbol{x}) = \pi(\theta \mid \boldsymbol{x}) \left| \frac{d\theta}{d\phi} \right|$$

José M. Bernardo is Professor of Statistics at the University of Valencia; his research is partially funded with grant PB93-1204 of the DGICYT, Madrid, Spain. José M. Ramón is a Research Student at the University of Valencia.

(ii) *No marginalization paradoxes*. (Stone and Dawid, 1972; Dawid, Stone and Zidek, 1973). If the posterior $\pi_1(\phi\,|\,\boldsymbol{x})$ for the quantity of interest ϕ conditional to data $\boldsymbol{x}$ from model $p(\boldsymbol{x}\,|\,\phi,\boldsymbol{\lambda})$ is of the form $\pi_1(\phi\,|\,\boldsymbol{x}) = \pi_1(\phi\,|\,t)$, and if the sampling distribution of t, $p(t\,|\,\phi,\boldsymbol{\lambda}) = p(t\,|\,\phi)$ only depends on ϕ, then the posterior of ϕ, $\pi_2(\phi\,|\,t)$, obtained from the simplified model $p(t\,|\,\phi)$ *must* be the same as the posterior $\pi_1(\phi\,|\,t)$ obtained from the full model $p(\boldsymbol{x}\,|\,\phi,\boldsymbol{\lambda})$.

(iii) *Consistent sampling properties*. (Neyman and Scott, 1948, Stein, 1959, 1962, 1985; Welch and Peers, 1963; Peers, 1965; Stone, 1976; Fraser, Monette and Ng, 1985; Tibshirani, 1989; Datta and Ghosh, 1995a). The properties under repeated sampling of the posterior distribution, should be consistent with the model. In particular, for any *large* sample size and for any $0 < p < 1$, the coverage probability of a credible interval with non-subjective posterior probability p should be close to p for most parameter values.

Besides those technical requirements, methods proposed to derive non-subjective posterior distributions should be *general*, *i.e.*, applicable to any properly defined inference problem, and "*admissible*" in the sense that, for each known example, no other model-based posterior could be argued to be "better" in a generally accepted, well-defined sense.

The *reference analysis*, introduced by Bernardo (1979, 1981) and further developed by Berger and Bernardo (1989, 1992a, 1992b, 1992c) is, to the best of our knowledge, the only available method to derive non-subjective posterior distributions which satisfy all those desiderata. However, reference posterior distributions have a reputation of being difficult to obtain, and the professional literature often contains formal Bayesian analysis using unjustified and often misleading, (but easily derived!), naïve "noninformative" priors. This may be partially due to the lack of an easily accessible introduction to reference analysis; in this paper, we try to offer such an introduction.

Section 2 contains an overview of reference analysis, where the definition is motivated, heuristic derivations of explicit expressions for the one parameter, two parameters, and multi-parameter cases are sequentially presented, and the behaviour of the reference posteriors under repeated sampling is discussed. In Section 3, the theory is applied to an inference problem, the ratio of multinomial parameters, for which no model-based Bayesian analysis has been previously proposed, and which has been chosen because it combines intrinsic importance and pedagogic value. Section 4 includes further discussion and provides directions for complementary reading. A number of definite integral results required in Section 3 are collected together in a final appendix.

2. AN OVERVIEW OF REFERENCE ANALYSIS

2.1. *Motivation*

The declared objective of reference Bayesian analysis is to specify a prior distribution such that, even for moderate sample sizes, the *information provided by the data should dominate the prior information* because of the "vague" nature of the prior knowledge. Reference analysis uses the concept of statistical information, in the technical sense of Shannon (1948) and Lindley (1956), to make this notion precise; see Soofi (1994) for a recent discussion of these ideas.

The amount of information to be expected from an experiment about some quantity of interest naturally depends on the available prior knowledge: the more prior information available, the less information may be expected to be learned from the data. An infinitely large experiment would eventually provide all missing information; thus, it is possible to obtain a measure of the amount of missing information as a limiting form of a functional of the prior distribution. It is natural to define “vague” prior knowledge as that with the largest missing information: the *reference prior* should then be that which *maximizes the missing information*.

Actually, due to the fact that the missing information is defined as a limit which is not necessarily finite, the reference prior is defined as some special limit of a sequence of prior distributions which maximize the information to be expected from an increasingly large experiment; we now make this formulation precise.

2.2. *One Parameter*

Given an experiment e which consists of one observation $\boldsymbol{x}$ from $p(\boldsymbol{x} \mid \phi)$, $\phi \in \Phi \subset \Re$, the amount of information $I\{e, p(\phi)\}$ which may be expected about ϕ when prior knowledge is described by $p(\phi)$ is defined by

$$I\{e, p(\phi)\} = \int_X p(\boldsymbol{x}) \int_\Phi p(\phi \mid \boldsymbol{x}) \log \frac{p(\phi \mid \boldsymbol{x})}{p(\phi)}\, d\phi d\boldsymbol{x};$$

hence, the amount of information which may be expected from k independent replications of e, $\boldsymbol{z}_k = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ is

$$I\{e(k), p(\phi)\} = \int_{X^k} p(\boldsymbol{z}_k) \int_\Phi p(\phi \mid \boldsymbol{z}_k) \log \frac{p(\phi \mid \boldsymbol{z}_k)}{p(\phi)}\, d\phi d\boldsymbol{z}_k.$$

As $k \to \infty$, $e(k)$ would provide any *missing information* about ϕ which could be obtained within this framework, and hence, as $k \to \infty$, $I\{e(k), p(\phi)\}$ will approach the missing information about ϕ when prior knowledge is described by $p(\phi)$.

It would be natural for a non-subjective prior intended to describe “lack of knowledge” about a quantity ϕ to *maximize the missing information about its value*: the reference prior would then be a special type of limit, as $k \to \infty$, of a sequence of priors $\pi_k(\phi)$ which maximize $I\{e(k), p(\phi)\}$ within the class of strictly positive priors on Φ. The amount of information $I\{e(k), p(\phi)\}$ may usefully be reexpressed as

$$I\{e(k), p(\phi)\} = \int_\Phi p(\phi) \log \frac{f_k(\phi)}{p(\phi)}\, d\phi,$$

$$f_k(\phi) = \exp\left\{ \int p(\boldsymbol{z}_k \mid \phi) \log p(\phi \mid \boldsymbol{z}_k)\, d\boldsymbol{z}_k \right\}$$

and, using a calculus of variations argument, it is easily verified that this is maximized if, and only if, the prior $p(\phi)$ is such that $p(\phi) \propto f_k(\phi)$. However, for each k, this only provides an *implicit* solution for the prior which maximizes $I\{e(k), p(\phi)\}$, since $f_k(\phi)$ depends on the prior through the posterior distribution $p(\phi \mid \boldsymbol{z}_k)$; moreover, the maximizing prior is typically discrete, even for continuous parameters (Berger, Bernardo and Mendoza, 1989).

To overcome both difficulties, consider, for large k, an asymptotic approximation to the posterior distribution, say $q(\phi \mid \boldsymbol{z}_k)$, which may certainly be chosen to be independent of $p(\phi)$. Then, under suitable regularity conditions, the sequence of *positive* functions

$$\pi_k(\phi) = \exp\left\{ \int_{X^k} p(\boldsymbol{z}_k \mid \phi) \log q(\phi \mid \boldsymbol{z}_k) d\boldsymbol{z}_k \right\}, \quad \phi \in \Phi, \quad k = 1, 2, \ldots \tag{1}$$

derived from such an asymptotic posterior may be expected to induce, by formal use of Bayes theorem, a sequence of posterior distributions

$$\pi_k(\phi \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \phi)\pi_k(\phi)}{\int_\Phi p(\boldsymbol{x} \mid \phi)\pi_k(\phi) d\phi}\,, \quad k = 1, 2, \ldots$$

with the desired reference posterior distribution $\pi(\phi \mid \boldsymbol{x})$ as its limit, so that

$$\pi(\phi \mid \boldsymbol{x}) = \lim_{k\to\infty} \pi_k(\phi \mid \boldsymbol{x}), \quad \phi \in \Phi, \quad \boldsymbol{x} \in X, \tag{2}$$

where the limit is to be understood in the *information sense, i.e.,* such that, for almost all $\boldsymbol{x}$,

$$\lim_{k\to\infty} \int_\Phi \pi_k(\phi \mid \boldsymbol{x}) \log \frac{\pi_k(\phi \mid \boldsymbol{x})}{\pi(\phi \mid \boldsymbol{x})}\, d\phi = 0.$$

For a discussion of the necessity of this type of limit, see the analysis of the *confidence paradox* of Monette, Fraser and Ng (1985), in Berger and Bernardo (1992c).

The limiting distribution (2) is *defined* to be the *reference posterior distribution* of ϕ. A *reference prior* is a function which, for any data, makes it possible to obtain the reference posterior $\pi(\phi \mid \boldsymbol{x})$ by formal use of Bayes theorem, *i.e.*, a positive function $\pi(\phi)$ such that, for all $\boldsymbol{x} \in X$,

$$\pi(\phi \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \phi)\pi(\phi)}{\int_\Phi p(\boldsymbol{x} \mid \phi)\pi(\phi) d\phi}\,.$$

Thus the reference prior $\pi(\phi)$ is the limit of the sequence $\{\pi_k(\phi),\ k = 1, 2, \ldots\}$ defined by (1) in the precise sense that the information-type limit of the corresponding sequence of posterior distributions $\{\pi_k(\phi \mid \boldsymbol{x}),\ k = 1, 2, \ldots\}$ is the posterior obtained from $\pi(\phi)$ by formal use of Bayes theorem.

Very often, the asymptotic posterior distribution $q(\phi \mid \boldsymbol{z}_k)$ only depends on the data through some asymptotically sufficient, consistent estimator $\hat\phi$. In such case, the sequence (1) may be reexpressed as

$$\begin{aligned}\pi_k(\phi) &= \exp\left\{\int_{X^k} p(\boldsymbol{z}_k \mid \phi) \log q(\phi \mid \boldsymbol{z}_k)\, d\boldsymbol{z}_k\right\}\\ &= \exp\left\{\int_{\Re} p(\hat\phi \mid \phi) \log q(\phi \mid \hat\phi)\, d\hat\phi\right\},\end{aligned}$$

which, as $k \to \infty$, converges to

$$\exp\left\{\log q(\phi \mid \hat\phi) \mid_{\hat\phi=\phi}\right\} = q(\phi|\hat\phi)\Big|_{\hat\phi=\phi.}$$

In particular, if $q(\phi|\hat\phi) = \mathrm{N}(\phi|\hat\phi, d(\hat\phi))$ so that the posterior distribution of ϕ is asymptotically normal with mean $\hat\phi$ and standard deviation $d(\hat\phi)$, then

$$q(\phi|\hat\phi)\Big|_{\hat\phi=\phi} \propto d^{-1}(\phi).$$

Summarizing, we may state:

Proposition 1. *Let $p(\boldsymbol{x} \mid \phi)$, $\boldsymbol{x} \in X$, be a probability model with one real-valued parameter $\phi \in \Phi \subset \Re$ such that there is a consistent and asymptotically sufficient estimator $\hat\phi$, and let $q(\phi \mid \hat\phi)$ be an asymptotic approximation to the posterior distribution of ϕ which only depends on the model. Then, any function of the form*

$$\pi(\phi) \propto q(\phi \mid \hat\phi)\Big|_{\hat\phi=\phi}$$

is a reference prior. In particular, if the asymptotic posterior is normal with standard deviation $d(\hat{\phi})$*, then* $\pi(\phi) \propto d(\phi)^{-1}$*. The reference posterior distribution of* ϕ *given* $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ *is*

$$\pi(\phi \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{\pi(\phi) \prod_{l=1}^n p(\boldsymbol{x}_l \,|\, \phi)}{\int_\Phi \pi(\phi) \prod_{l=1}^n p(\boldsymbol{x}_l \,|\, \phi)\, d\phi}\,.$$

It is well known that under regularity conditions, the posterior distribution is asymptotically normal with standard deviation $f(\hat{\phi})^{-1/2}$, where

$$f(\phi) = -\int_X p(\boldsymbol{x} \,|\, \phi) \frac{\partial^2}{\partial \phi^2} \log p(\boldsymbol{x} \,|\, \phi)\, d\boldsymbol{x}$$

is Fisher's function. In this case, the reference prior is

$$\pi(\phi) \propto d(\phi)^{-1} = f(\phi)^{1/2},$$

i.e., Jeffreys (1946, 1961) prior; Polson (1992) discusses in detail the necessary regularity conditions; Ghosal (1996) analyses the non-regular case. It follows that the reference prior algorithm contains Jeffreys' prior as the particular case which obtains under *normal* asymptotics in *one-parameter* continuous models.

Proposition 1 may be used to derive reference posteriors associated to models which only depend on the quantity of interest. As one should require, if the model is otherwise parametrized in terms of some one-to-one function $\theta = \theta(\phi)$ of the quantity of interest, the reference posterior of ϕ may consistently be obtained from that of θ. Indeed,

$$\pi(\phi) = q(\phi \,|\, \hat{\phi})\Big|_{\hat{\phi}=\phi} = q(\theta(\phi) \,|\, \hat{\theta})\Big|\frac{d\theta}{d\phi}\Big|\Big|_{\hat{\theta}=\theta(\phi)} = \pi(\theta(\phi))\Big|\frac{d\theta}{d\phi}\Big|.$$

We now consider models which contain nuisance parameters; it turns out that those may be handled by recursively using the one-parameter solution.

2.3. *One Nuisance Parameter*

Suppose that we are interested in the value of ϕ, given a random sample $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ from a model

$$p(\boldsymbol{x} \,|\, \phi, \lambda), \quad \phi \in \Phi \subset \Re, \quad \lambda \in \Lambda(\phi) \subset \Re,$$

which contains a nuisance parameter λ. Note that we allow for the possibility that the nuisance parameter space $\Lambda(\phi)$ *may* depend on ϕ.

Working conditionally on ϕ, this is a one-parameter problem and, hence, the one-parameter solution described above may be used to obtain a *conditional* reference prior $\pi(\lambda \,|\, \phi)$. If this is proper, then it may be used to integrate out the nuisance parameter λ and obtain a model $p(\boldsymbol{x} \,|\, \phi)$ which only depends on ϕ, thereby reducing the problem to one already solved.

If the conditional reference prior $\pi(\lambda \,|\, \phi)$ is not proper, then the procedure is performed within an increasing sequence of bounded approximations $\{\Lambda_i, i = 1, 2, \ldots\}$ to the nuisance parameter space Λ, chosen such that $\pi(\lambda \,|\, \phi)$ is integrable within each of them. The reference posterior $\pi(\phi \,|\, \boldsymbol{x})$ is then obtained as the limit of the resulting sequence $\{\pi_i(\phi \,|\, \boldsymbol{x}), i = 1, 2, \ldots\}$ of restricted reference posteriors.

We shall only consider here the regular case, where joint posterior asymptotic normality may be established. Let $F(\phi,\lambda)$ be the corresponding 2×2 Fisher's matrix in terms of ϕ and λ, and let $S(\phi,\lambda) = F^{-1}(\phi,\lambda)$, so that the posterior distribution of (ϕ,λ) is asymptotically normal with mean $(\hat\phi,\hat\lambda)$, the corresponding mle's, and covariance matrix $S(\hat\phi,\hat\lambda)$. It follows that:

(i) the *marginal* posterior distribution of ϕ is asymptotically normal with standard deviation $d_0(\hat\phi,\hat\lambda) = s_{1,1}(\hat\phi,\hat\lambda)^{1/2}$;

(ii) the *conditional* posterior distribution of λ given ϕ is asymptotically normal with standard deviation $d_1(\phi,\hat\lambda) = f_{2,2}(\phi,\hat\lambda)^{-1/2}$.

Working conditionally on ϕ, so that λ is the only relevant parameter, and using Proposition 1, we find $\pi(\lambda\,|\,\phi) \propto d_1(\phi,\lambda)^{-1}$ and, therefore

$$\pi(\lambda\,|\,\phi) = \frac{d_1^{-1}(\phi,\lambda)}{\int_{\Lambda(\phi)} d_1^{-1}(\phi,\lambda)\,d\lambda}\,,\quad \lambda\in\Lambda(\phi),$$

provided the integral exists. If it does not, an approximating sequence is

$$\pi_i(\lambda\,|\,\phi) = \frac{d_1^{-1}(\phi,\lambda)}{\int_{\Lambda_i(\phi)} d_1^{-1}(\phi,\lambda)\,d\lambda}\,,\quad \lambda\in\Lambda_i(\phi),$$

where $\{\Lambda_i(\phi), i=1,2,\ldots\}$ is an increasing sequence of compact approximations to $\Lambda(\phi)$.

The sequence of priors (1) may then be computed as

$$\begin{aligned}\pi_k(\phi) &\propto \exp\left\{\int_{X^k} p(\boldsymbol{z}_k\,|\,\phi)\log q(\phi\,|\,\boldsymbol{z}_k)\,d\boldsymbol{z}_k\right\}\\ &= \exp\left\{\int_{\Re^2} p(\hat\phi,\hat\lambda\,|\,\phi)\log q(\phi\,|\,\hat\phi,\hat\lambda)\,d\hat\phi\,d\hat\lambda\right\},\quad k=1,2,\ldots\end{aligned}$$

If $\pi(\lambda\,|\,\phi)$ is proper, we have

$$p(\hat\phi,\hat\lambda\,|\,\phi) = \int_{\Lambda(\phi)} p(\hat\phi,\hat\lambda\,|\,\phi,\lambda)\pi(\lambda\,|\,\phi)\,d\lambda$$

and therefore, substituting and changing the order of integration,

$$\pi_k(\phi) = \exp\left\{\int_{\Lambda(\phi)} \pi(\lambda\,|\,\phi)\Big(\int_{\Re^2} p(\hat\phi,\hat\lambda\,|\,\phi,\lambda)\log q(\phi\,|\,\hat\phi,\hat\lambda)\,d\hat\phi d\hat\lambda\Big)d\lambda\right\}.$$

But the inner double integral converges to

$$\log q(\phi\,|\,\hat\phi,\hat\lambda)\Big|_{(\hat\phi,\hat\lambda)=(\phi,\lambda)} = \log\left[d_0^{-1}(\phi,\lambda)\right]$$

since $q(\phi\,|\,\hat\phi,\hat\lambda)$ is normal with mean $\hat\phi$ and standard deviation $d_0^{-1}(\hat\phi,\hat\lambda)$ and, therefore,

$$\pi(\phi) \propto \exp\Big\{\int_{\Lambda(\phi)} \pi(\lambda\,|\,\phi)\log[d_0^{-1}(\phi,\lambda)]\,d\lambda\Big\}.$$

If $\pi(\lambda\,|\,\phi)$ is not proper, one would similarly obtain the approximating sequence

$$\pi_i(\phi) \propto \exp\Big\{\int_{\Lambda_i(\phi)} \pi(\lambda\,|\,\phi)\log[d_0^{-1}(\phi,\lambda)]\,d\lambda\Big\}.$$

Thus, we have:

Proposition 2. *Let $p(\boldsymbol{x}\,|\,\phi, \lambda)$, $\phi \in \Phi \subset \Re$, $\lambda \in \Lambda(\phi) \subset \Re$, be a probability model with two real-valued parameters ϕ and λ, where ϕ is the quantity of interest, and suppose that the joint posterior distribution of (ϕ, λ) is asymptotically normal with covariance matrix $S(\hat{\phi}, \hat{\lambda})$. Then, if $H(\phi, \lambda) = S^{-1}(\phi, \lambda)$,*

(i) the conditional reference prior of λ is

$$\pi(\lambda\,|\,\phi) \propto d_1^{-1}(\phi, \lambda) = h_{2,2}^{1/2}(\phi, \lambda), \quad \lambda \in \Lambda(\phi)$$

(ii) if $\pi(\lambda\,|\,\phi)$ is proper, the reference posterior distribution of ϕ given $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is

$$\pi(\phi\,|\,\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \propto \pi(\phi) \int_{\Lambda(\phi)} \Big\{\prod\nolimits_{l=1}^{n} p(\boldsymbol{x}_l\,|\,\phi, \lambda)\Big\} \pi(\lambda\,|\,\phi)\, d\lambda,$$

where the marginal reference prior of ϕ is

$$\pi(\phi) \propto \exp\Big\{\int_{\Lambda(\phi)} \pi(\lambda\,|\,\phi) \log[d_0^{-1}(\phi, \lambda)]\, d\lambda\Big\}, \quad d_0(\phi, \lambda) = s_{1,1}^{1/2}(\phi, \lambda).$$

(iii) if $\pi(\lambda\,|\,\phi)$ is not proper, a compact approximation $\{\Lambda_i(\phi), i = 1, 2, \ldots\}$ to $\Lambda(\phi)$ is required, and the reference posterior distribution of ϕ is obtained as

$$\pi(\phi\,|\,\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \lim_{i \to \infty} \pi_i(\phi\,|\,\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n),$$

where $\pi_i(\phi\,|\,\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is derived using $\Lambda_i(\phi)$ instead of $\Lambda(\phi)$.

It is often found in applications that the nuisance parameter space $\Lambda(\phi) = \Lambda$ is independent of ϕ, and that the functions d_0 and d_1 nicely factorize in the form $d_0^{-1}(\phi, \lambda) = a_0(\phi) b_0(\lambda)$, $d_1^{-1}(\phi, \lambda) = a_1(\phi) b_1(\lambda)$; if this is the case, then, for some positive constant c_i, we have

$$\pi_i(\lambda\,|\,\phi) = \frac{a_1(\phi) b_1(\lambda)}{\int_{\Lambda_i} a_1(\phi) b_1(\lambda) d\lambda} = c_i b_1(\lambda),$$

$$\pi_i(\phi) \propto \exp\left\{\int_{\Lambda_i} c_i b_1(\lambda) \log[a_0(\phi) b_1(\lambda)] d\lambda\right\} \propto a_0(\phi),$$

and hence, $\pi_i(\phi) = \pi(\phi) \propto a_0(\phi)$. Thus, we have,

Corollary. *If the nuisance parameter space $\Lambda(\phi) = \Lambda$ is independent of ϕ, and the functions d_0 and d_1 factorize in the form*

$$d_0^{-1}(\phi, \lambda) = a_0(\phi) b_0(\lambda), \quad d_1^{-1}(\phi, \lambda) = a_1(\phi) b_1(\lambda),$$

then

$$\pi(\phi) \propto a_0(\phi), \quad \pi(\lambda\,|\,\phi) \propto b_1(\lambda),$$

and there is no need for compact approximation, even if the conditional reference priors are not proper.

2.4. *The Multiparameter Case*

Proposition 2 and its corollary may easily be extended to any number of nuisance parameters. Indeed, if the model is $p(\boldsymbol{x} \mid \phi, \lambda_1, \ldots, \lambda_m)$, the quantity of interest is ϕ, the appropriate regularity conditions hold, and $F(\phi, \lambda_1, \ldots, \lambda_m)$ is the corresponding $(m+1) \times (m+1)$ Fisher's matrix, then the posterior distribution of $(\phi, \lambda_1, \ldots, \lambda_m)$ is asymptotically normal with mean $(\hat{\phi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$, the corresponding mle's, and covariance matrix $S(\hat{\phi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$, where $S = F^{-1}$.

It follows that, if S_j is the $j \times j$ upper matrix of S, $j = 1, \ldots, m+1$, $H_j = S_j^{-1}$ and $h_{j,j}(\phi, \lambda_1, \ldots, \lambda_m)$ is the (j,j) element of H_j, so that $H_{m+1} = F$ and $h_{m+1,m+1} = f_{m+1,m+1}$, then

(i) the *marginal* posterior distribution of ϕ is asymptotically normal with standard deviation

$$d_0(\hat{\phi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m) = s_{1,1}(\hat{\phi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)^{1/2} = h_{1,1}(\hat{\phi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)^{-1/2};$$

(ii) the *conditional* posterior distribution of λ_i given $\phi, \lambda_1, \ldots, \lambda_{i-1}$, is asymptotically normal with standard deviation

$$d_i(\phi, \lambda_1, \ldots, \lambda_{i-1}, \hat{\lambda}_i, \ldots, \hat{\lambda}_m) = h_{i+1,i+1}(\phi, \lambda_1, \ldots, \lambda_{i-1}, \hat{\lambda}_i, \ldots, \hat{\lambda}_m)^{-1/2},$$

and one may sequentially use the algorithm described in 2.3 to derive $\pi(\lambda_m \mid \phi, \lambda_1, \ldots, \lambda_{m-1})$, $\pi(\lambda_{m-1} \mid \phi, \lambda_1, \ldots, \lambda_{m-2})$, ..., $\pi(\lambda_1 \mid \phi)$, and $\pi(\phi)$, and produce the desired reference posterior.

Proposition 3. *Let $p(\boldsymbol{x} \mid \phi, \boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ be a probability model with $m+1$ real-valued parameters, let ϕ be the quantity of interest, and suppose that the joint distribution of $(\phi, \lambda_1, \ldots, \lambda_m)$ is asymptotically normal with covariance matrix $S(\hat{\phi}, \hat{\lambda}_1, \ldots, \hat{\lambda}_m)$. Then, if S_j is the $j \times j$ upper matrix of S, $H_j = S_j^{-1}$ and $h_{j,j}(\phi, \lambda_1, \ldots, \lambda_m)$ is the (j,j) element of H_j,*

(i) the conditional reference priors are

$$\begin{aligned}
&\pi(\lambda_m \mid \phi, \lambda_1, \ldots, \lambda_{m-1}) \propto d_m^{-1}(\phi, \lambda_1, \ldots, \lambda_m), \\
&\pi(\lambda_i \mid \phi, \lambda_1, \ldots, \lambda_{i-1}) \propto \\
&\quad \exp\Big\{ \int_{\Lambda_{i+1}} \cdots \int_{\Lambda_m} \log d_i^{-1}(\phi, \lambda_1, \ldots, \lambda_m) \Big\{ \prod_{j=i+1}^{m} \pi(\lambda_j \mid \phi, \lambda_1, \ldots, \lambda_{j-1}) \Big\} \, d\boldsymbol{\lambda}_{i+1} \Big\}
\end{aligned}$$

where $d\boldsymbol{\lambda}_j = d\lambda_j \times \cdots \times d\lambda_m$, *and*

$$d_i^{-1}(\phi, \lambda_1, \ldots, \lambda_m) = h_{i+1,i+1}(\phi, \lambda_1, \ldots, \lambda_m)^{1/2}, \quad i = 1, \ldots, m,$$

provided $\pi(\lambda_i \mid \lambda_1, \ldots, \lambda_{i-1})$, $i = 1, \ldots, m$ are all proper. If any of those conditional reference priors is not proper, then a compact approximation is required for the corresponding integrals.

(ii) The marginal reference prior of ϕ is

$$\pi(\phi) \propto \exp\Big\{ \int_{\Lambda_1} \cdots \int_{\Lambda_m} \log d_0^{-1}(\phi, \lambda_1, \ldots, \lambda_m) \Big\{ \prod_{j=1}^{m} \pi(\lambda_j \mid \phi, \lambda_1, \ldots, \lambda_{j-1}) \Big\} \, d\boldsymbol{\lambda}_1 \Big\}$$

where

$$d_0^{-1}(\phi, \lambda_1, \ldots, \lambda_m) = h_{1,1}^{1/2}(\phi, \lambda_1, \ldots, \lambda_m) = s_{1,1}^{-1/2}(\phi, \lambda_1, \ldots, \lambda_m).$$

After data $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ have been observed, the reference posterior distribution of the parameter of interest ϕ, is

$$\pi(\phi \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \propto \pi(\phi) \int_{\Lambda_1} \cdots \int_{\Lambda_m} \Big\{ \prod_{l=1}^{n} p(\boldsymbol{x}_l \,|\, \phi, \lambda_1, \ldots, \lambda_m) \Big\} \prod_{j=1}^{m} \Big\{ \pi(\lambda_j \,|\, \lambda_1, \ldots, \lambda_{j-1}) \Big\}\, d\boldsymbol{\lambda}_1 \Big\}.$$

Corollary *If the nuisance parameter spaces $\Lambda_i(\phi, \lambda_1, \ldots, \lambda_{i-1}) = \Lambda_i$ are independent of both ϕ and the λ_i's, and the functions $d_0, \ldots, d_m$, factorize in the form*

$$d_0^{-1}(\phi, \lambda_1, \ldots, \lambda_m) = h_{1,1}^{1/2}(\phi, \lambda_1, \ldots, \lambda_m) = a_0(\phi) b_0(\lambda_1, \ldots, \lambda_m)$$

$$\begin{aligned} d_i^{-1}(\phi, \lambda_1, \ldots, \lambda_m) &= h_{i+1,i+1}^{1/2}(\phi, \lambda_1, \ldots, \lambda_m) \\ &= a_i(\lambda_i) b_i(\phi, \lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_m), \quad i = 1 \ldots, m, \end{aligned}$$

then

$$\pi(\phi) \propto a_0(\phi), \quad \pi(\lambda_i \,|\, \phi, \lambda_1, \ldots, \lambda_{i-1}) \propto a_i(\lambda_i), \quad i = 1, \ldots, m$$

and there is no need for compact approximations, even if the $\pi(\lambda_i \,|\, \phi, \lambda_1, \ldots, \lambda_{i-1})$'s are not proper.

2.5. *Behaviour under repeated sampling*

The frequentist coverage probabilities of credible intervals derived from reference posterior distributions are sometimes identical, and usually very close, to their posterior probabilities; this means that even for moderate samples, an interval with reference posterior probability $1-\alpha$ may often be interpreted as an approximate confidence interval with significance level α.

More formally, if $t_\alpha = t_\alpha(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ denotes the $1-\alpha$ quantile which corresponds to the reference posterior $\pi(\phi \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, so that

$$P\Big[\phi \le t_\alpha \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\Big] = \int_{\phi \le t_\alpha(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)} \pi(\phi \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\, d\phi = 1 - \alpha,$$

then the coverage probability of the $(1-\alpha)$ reference posterior credible interval $]-\infty, t_\alpha]$,

$$P\Big[t_\alpha \ge \phi \,|\, \phi\Big] = \int_{t_\alpha(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \ge \phi} p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \,|\, \phi)\, d\boldsymbol{x}_1 \cdots d\boldsymbol{x}_n$$

often satisfies

$$P\Big[t_\alpha \ge \phi \,|\, \phi\Big] = 1 - \alpha + O(n^{-1}),$$

while, for most priors, this asymptotic approximation is only $O(n^{-1/2})$. Intuitively, this says that the reference prior is often a *probability matching* prior, *i.e.*, a prior for which the coverage probabilities of *one-sided* posterior credible intervals are asymptotically as close as possible to their posterior probabilities. Hartigan (1966) showed that the coverage probabilities of *two-sided* Bayesian posterior credible intervals satisfy this type of approximation to $O(n^{-1})$ for *all* priors.

In a pioneering paper, Welch and Peers (1963) established that in the case of the one-parameter regular continuous models Jeffreys' prior, —which in this case (Proposition 1) is also the reference prior—, is the only probability matching prior. Hartigan (1983, p. 79) showed that this result may be extended to one-parameter discrete models by using continuity corrections. Datta and Ghosh (1995a) derived a differential equation which provides a necessary and sufficient condition for a prior to be probability matching in the multiparameter continuous regular case; this has been used to verify that reference priors are typically probability matching priors. Recent work by Rousseau (1996) using continuity corrections, suggests that these results may also be extended to multiparameter discrete models. In Section 4, we summarize some additional related work.

Although the results described above only justify an *asymptotic* approximate frequentist interpretation of reference posterior probabilities, there is empirical evidence to suggest that the coverage probabilities of reference posterior credible intervals derived from relatively small samples are also typically close to their posterior probabilities; this will be illustrated in the example discussed below.

3. THE RATIO OF MULTINOMIAL PARAMETERS

3.1. *The Problem*

Consider n multinomial observations which belong to one of, say, $m+1$ categories, so that

$$p(r_1, \ldots, r_m \,|\, n, \theta_1, \ldots, \theta_m) = \frac{n!}{\prod_{i=1}^{m+1} r_i!} \prod_{i=1}^{m+1} \theta_i^{r_i}, \quad 0 \le r_i < n, \quad \sum_{i=1}^{m} r_i \le n$$

with $0 < \theta_i < 1$, $\theta_{m+1} = 1 - \sum_{i=1}^m \theta_i$, and $r_{m+1} = n - \sum_{i=1}^m r_i$.

Suppose that we are interested in the ratio of the, say, first two parameters $\phi = \theta_1/\theta_2$. For instance, in an insurance application one may be interested in assessing how many times more likely is risk 1 than risk 2 or, in a political application, one may be interested in assessing the ratio of the percentages of votes that candidates 1 and 2 may be expected to obtain.

We note that, in the absence of other information, one would expect the result to depend on r_1 and r_2, but *not* on n or the other r_i's, which intuition suggest should be irrelevant; indeed, in *the absence of information on the relationship among the θ_i's*, we cannot expect to obtain information about θ_1/θ_2 from the $n - r_1 - r_2$ observations which do not belong to either of the first two categories.

3.2. *The Two Parameters Case*

Let us first consider the two parameter case, so that there are three categories, with probabilities θ_1, θ_2 and $1-\theta_1-\theta_2$, and

$$p(r_1, r_2 \,|\, n, \theta_1, \theta_2) = \frac{n!}{r_1! r_2! (n - r_1 - r_2)!} {\theta_1}^{r_1} {\theta_2}^{r_2} (1 - \theta_1 - \theta_2)^{n - r_1 - r_2}, \quad r_1 + r_2 \le n,$$

or, in terms of $\phi = \theta_1/\theta_2$, and $\lambda = \theta_2$,

$$p(r_1, r_2 \,|\, n, \phi, \lambda) = \frac{n!}{r_1! r_2! (n - r_1 - r_2)!} \phi^{r_1} \lambda^{r_1 + r_2} \Big(1 - \lambda(1 + \phi)\Big)^{n - r_1 - r_2}.$$

The corresponding Fisher's matrix is easily found to be

$$F(\phi,\lambda) = \frac{n}{1-\lambda(1+\phi)} \begin{pmatrix} \frac{\lambda(1-\lambda)}{\phi} & 1 \\ 1 & \frac{1+\phi}{\lambda} \end{pmatrix}$$

so that

$$S(\phi,\lambda) = F^{-1}(\phi,\lambda) = \frac{1}{n} \begin{pmatrix} \frac{\phi(1+\phi)}{\lambda} & -\phi \\ -\phi & \lambda(1-\lambda) \end{pmatrix};$$

hence, the joint posterior of (ϕ,λ) is asymptotically normal with covariance matrix $S(\hat\phi,\hat\lambda)$ and, therefore,

(i) the *marginal* asymptotic posterior of ϕ is normal with standard deviation $d_0(\hat\phi,\hat\lambda)$,

$$d_0(\phi,\lambda) = \frac{1}{\sqrt{n}} \left(\frac{\phi(1+\phi)}{\lambda}\right)^{1/2};$$

(ii) the *conditional* asymptotic posterior of λ given ϕ is normal with standard deviation $d_1(\phi,\hat\lambda)$,

$$d_1(\phi,\lambda) = \frac{1}{\sqrt{n}} \left(\frac{1+\phi}{\lambda\{1-\lambda(1+\phi)\}}\right)^{-1/2}.$$

From Proposition 2 (i), $\pi(\lambda\,|\,\phi) \propto d_1^{-1}(\phi,\lambda)$; hence,

$$\pi(\lambda\,|\,\phi) = \frac{d_1^{-1}(\phi,\lambda)}{\int_{\Lambda(\phi)} d_1^{-1}(\phi,\lambda)d\lambda} = \frac{\lambda^{-1/2}\{1-\lambda(1+\phi)\}^{-1/2}}{\int_0^{(1+\phi)^{-1}} \lambda^{-1/2}\{1-\lambda(1+\phi)\}^{-1/2}d\lambda},$$

since the factor $(1+\phi)^{1/2}$ cancels out and

$$0<\theta_1+\theta_2<1 \quad \Rightarrow \quad 0<\phi\lambda+\lambda<1 \quad \Rightarrow \quad 0<\lambda<(1+\phi)^{-1};$$

thus, using Proposition A1, with $a=b=1/2$, and $c=1+\phi$, the conditional reference prior of the nuisance parameter λ given the parameter of interest ϕ is

$$\pi(\lambda\,|\,\phi) = \frac{(1+\phi)^{1/2}}{\pi}\lambda^{-1/2}\{1-\lambda(1+\phi)\}^{-1/2}, \quad 0<\lambda<(1+\phi)^{-1}, \tag{3}$$

which is a proper, Beta-like, distribution on the interval

$$\Lambda(\phi) = \left[0, \frac{1}{1+\phi}\right] = \left[0, \frac{\theta_1}{\theta_1+\theta_2}\right].$$

From Proposition 2 (ii),

$$\begin{aligned}\pi(\phi) &\propto \exp\left\{\int_{\Lambda(\phi)} \pi(\lambda\,|\,\phi) \log d_0^{-1}(\phi,\lambda)\, d\lambda\right\} \\ &\propto \exp\left\{\int_{\Lambda(\phi)} \pi(\lambda\,|\,\phi) \log \left(\frac{\phi(1+\phi)}{\lambda}\right)^{-1/2} d\lambda\right\}\end{aligned}$$

$$= \phi^{-1/2}(1+\phi)^{-1/2} \exp\left\{ \tfrac{1}{2} \int_{\Lambda(\phi)} \pi(\lambda \,|\, \phi) \log \lambda \, d\lambda \right\}$$

and, using Proposition A2 with $a = b = 1/2$ and $c = 1 + \phi$, one has

$$\pi(\phi) \propto \phi^{-1/2}(1+\phi)^{-1/2} \exp\left\{ -\tfrac{1}{2} \log[4(1+\phi)] \right\} \propto \phi^{-1/2}(1+\phi)^{-1};$$

finally , using Proposition A4 with $a = b = 1/2$,

$$\int_0^\infty \phi^{-1/2}(1+\phi)^{-1} d\phi = \pi$$

and, hence, the marginal reference prior of the parameter of interest ϕ, which is *proper* even though it is defined on the unbounded space $\Phi =]0, \infty[$, is given by

$$\pi(\phi) = \frac{1}{\pi}\phi^{-1/2}(1+\phi)^{-1}, \quad 0 < \phi < \infty. \tag{4}$$

Combining (3) and (4), the *joint* reference prior needed to obtain a reference posterior for the parameter of interest ϕ is the *proper* prior

$$\pi(\phi)\pi(\lambda \,|\, \phi) = \frac{1}{\pi^2}\phi^{-1/2}(1+\phi)^{-1/2}\lambda^{-1/2}\{1 - \lambda(1+\phi)\}^{-1/2}; \tag{5}$$

therefore, using (5) in Bayes theorem to derive the corresponding joint posterior and integrating out the nuisance parameter λ, the reference posterior for the parameter of interest is

$$\begin{aligned}
\pi(\phi \,|\, r_1, r_2, n) &\propto \pi(\phi) \int_{\Lambda(\phi)} p(r_1, r_2 \,|\, n, \phi, \lambda)\pi(\lambda \,|\, \phi) \, d\lambda \\
&\propto \phi^{-1/2}(1+\phi)^{-1} \int_0^{\frac{1}{1+\phi}} (1+\phi)^{1/2}\phi^{r_1}\lambda^{r_1+r_2-1/2}\{1-\lambda(1+\phi)\}^{n-r_1-r_2-1/2} d\lambda \\
&\propto (1+\phi)^{-1/2}\phi^{r_1-1/2} \int_0^{\frac{1}{1+\phi}} \lambda^{r_1+r_2-1/2}\{1-\lambda(1+\phi)\}^{n-r_1-r_2-1/2} d\lambda.
\end{aligned}$$

Using Proposition A1 with $a = r_1 + r_2 + 1/2$, $b = n - r_1 - r_2 + 1/2$ and $c = 1 + \phi$ to solve the last integral, we have

$$\pi(\phi \,|\, r_1, r_2, n) \propto (1+\phi)^{-1/2}\phi^{r_1-1/2}(1+\phi)^{-(r_1+r_2+1/2)} \propto \frac{\phi^{r_1-1/2}}{(1+\phi)^{r_1+r_2+1}} \, ,$$

and, using Proposition A4 with $a = r_1 + 1/2$ and $b = r_2 + 1/2$ to obtain the proportionality constant, we finally obtain the desired reference posterior distribution of the quantity of interest ϕ, as the Beta distribution of the second kind (see *e.g.*, Johnson *et al.* , 1995, p. 248)

$$\pi(\phi \,|\, r_1, r_2, n) = \pi(\phi \,|\, r_1, r_2) = \frac{\Gamma(r_1 + r_2 + 1)}{\Gamma(r_1 + 1/2)\Gamma(r_2 + 1/2)} \frac{\phi^{r_1-1/2}}{(1+\phi)^{r_1+r_2+1}} \, . \tag{6}$$

Since (6) has been derived from a proper prior, it is obviously proper for any data. Moreover, as expected, it does *not* depend on n, but only on r_1 and r_2: the $n - r_1 - r_2$ observations which belong to the third category do not directly provide any information on the value of $\phi = \theta_1/\theta_2$.

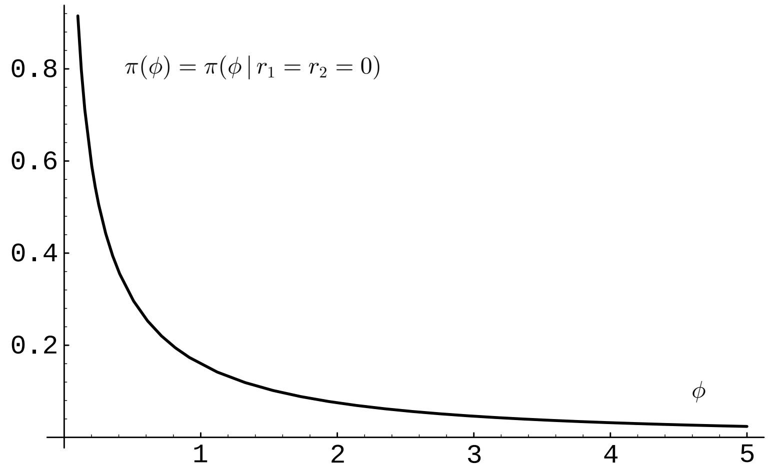

Figure 1. *Marginal reference prior of ϕ, and reference posterior for any data with $r_1 = r_2 = 0$.*

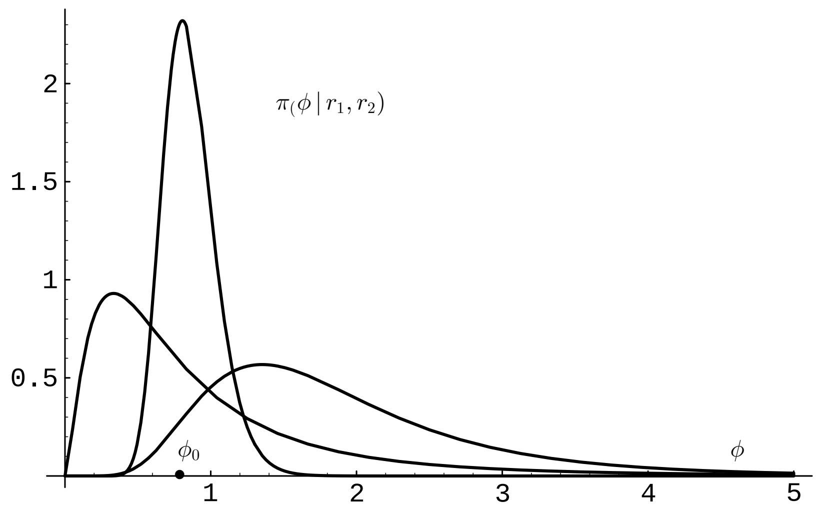

Figure 2. *Examples of reference posterior distributions from three random samples of sizes $n = 5$, $n = 20$ and $n = 100$, simulated with $\theta_1 = 0.4$ and $\theta_2 = 0.5$, so that $\phi = \phi_0 = 4/5$.*

In the particular case where $r_1 = r_2 = 0$, so that all n observations belong to the third category, we do not have any information about $\phi = \theta_1/\theta_2$ and hence, the reference posterior distribution reduces to the marginal reference prior (4), shown in Figure 1, which has no expected value, but a median equal to 1, thus making equally likely $\theta_1 > \theta_2$ than $\theta_2 > \theta_1$.

It is easily checked that, when $r_1 \geq 1$, $\pi(\phi \,|\, r_1, r_2)$ has a mode at $(r_1 - 1/2)/(r_2 + 3/2)$. Moreover, if one further defines

$$\omega = \frac{\phi}{1+\phi} = \frac{\theta_1}{\theta_1 + \theta_2}\,,$$

and hence $\phi = \omega/(1-\omega)$, one has

$$\pi(\omega \,|\, r_1, r_2) = \pi(\phi \,|\, r_1, r_2) \left| \frac{d\phi}{d\omega} \right| \propto w^{r_1 - 1/2}(1-w)^{r_2 - 1/2}$$

and, therefore,

$$\pi(w \,|\, r_1, r_2) = \mathrm{Be}(w \,|\, r_1 + 1/2, r_2 + 1/2).$$

Thus the reference posterior for ϕ is equivalent to ω, the proportion of observed elements in category 1 among those in either category 1 or category 2, having the conventional Jeffreys-like reference posterior $\text{Be}(\omega \mid r_1 + 1/2, r_2 + 1/2)$. This may be used to obtain credible regions for ϕ using the typically preprogrammed incomplete Beta routines.

Figure 2 shows the reference posterior distributions of ϕ obtained from three simulated samples of size $n = 5$, $n = 20$ and $n = 100$ from a multinomial model with $\theta_1 = 0.4$ and $\theta_2 = 0.5$, so that the true value of the quantity of interest is $\theta_1/\theta_2 = 4/5$.

3.3. *The General Case*

Let us now consider the general case, so that there are $m+1$ categories with probabilities $\theta_1, \ldots, \theta_m$ and $1 - \sum_{j=1}^m \theta_j$, and

$$p(r_1, \ldots, r_m \mid n, \theta_1, \ldots, \theta_m) = \frac{n!}{\prod_{j=1}^{m+1} r_j!} \prod_{j=1}^{m+1} \theta_j^{r_j}, \qquad \sum\nolimits_{j=1}^m r_j \le n,$$

with $r_{m+1} = n - \sum_{i=1}^m r_i$ and $\theta_{m+1} = 1 - \sum_{j=1}^m \theta_j$.

In this parametrization, the corresponding Fisher's matrix is easily found to be

$$F(\theta_1, \ldots, \theta_m) = E_{(r_1, \ldots, r_m \mid n, \theta_1, \ldots, \theta_m)} \left\{ -\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p(r_1, \ldots, r_m \mid n, \theta_1, \ldots, \theta_m) \right\}$$

$$= n \begin{pmatrix} 1 + \frac{\theta_{m+1}}{\theta_1} & 1 & \ldots & 1 \\ 1 & 1 + \frac{\theta_{m+1}}{\theta_2} & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 + \frac{\theta_{m+1}}{\theta_m} \end{pmatrix}$$

Since we are interested in $\phi = \theta_1/\theta_2$, we make the one-to-one transformation

$$\phi = \theta_1/\theta_2, \qquad \lambda = \theta_2, \qquad \theta_i = \theta_i, \quad i = 3, \ldots, m$$

The Jacobian of the inverse transformation $\theta_1 = \phi\lambda$, $\theta_2 = \lambda$, $\theta_i = \theta_i, i = 3, \ldots, m$ is then

$$J = \begin{pmatrix} \lambda & \phi & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

and, therefore, the posterior distribution of $(\phi, \lambda, \theta_3, \ldots, \theta_m)$ is asymptotically normal with covariance matrix $n^{-1} S(\hat\phi, \hat\lambda, \hat\theta_3, \ldots, \hat\theta_m)$, where $\hat\phi = r_1/r_2$, $\hat\lambda = r_2/n$, and $\hat\theta_i = r_i/n$, $i = 3, \ldots, m$ are the corresponding mle's, and

$$S(\phi, \lambda, \theta_3, \ldots, \theta_m) = H^{-1}(\phi, \lambda, \theta_3, \ldots, \theta_m),$$

with

$$H(\phi, \lambda, \theta_3, \ldots, \theta_m) = J^t F(\phi, \lambda, \theta_3, \ldots, \theta_m) J;$$

(see Mendoza, 1994, or Bernardo and Smith, 1994, p. 295). After some algebra, one finds

$$H(\phi, \lambda, \theta_3, \ldots, \theta_m) = \frac{n}{1 - \lambda(1+\phi) - \theta_m^*} \begin{pmatrix} A & B \\ B^t & C \end{pmatrix}$$

with

$$A = \begin{pmatrix} \frac{\lambda(1-\lambda-\theta_m^*)}{\phi} & 1 - \theta_m^* \\ 1 - \theta_m^* & \frac{(1+\phi)(1-\theta_m^*)}{\lambda} \end{pmatrix} \quad B = \begin{pmatrix} \lambda & \cdots & \lambda \\ 1+\phi & \cdots & 1+\phi \end{pmatrix}$$

$$C = \begin{pmatrix} \frac{1+\theta_3-\lambda(1+\phi)-\theta_m^*}{\theta_3} & 1 & \cdots & 1 \\ 1 & \ddots & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \frac{1+\theta_m-\lambda(1+\phi)-\theta_m^*}{\theta_m} \end{pmatrix} \tag{7}$$

where $\theta_m^* = \theta_3 + \cdots + \theta_m$, and

$$S(\phi, \lambda, \theta_3, \ldots, \theta_m) = H^{-1}(\phi, \lambda, \theta_3, \ldots, \theta_m)$$

$$= \frac{1}{n} \begin{pmatrix} \frac{\phi(\phi+1)}{\lambda} & -\phi & 0 & \cdots & 0 \\ -\phi & \lambda(1-\lambda) & -\lambda\theta_3 & \cdots & -\lambda\theta_m \\ 0 & -\lambda\theta_3 & \theta_3(1-\theta_3) & \cdots & -\theta_3\theta_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -\lambda\theta_m & -\theta_3\theta_m & \cdots & \theta_m(1-\theta_m) \end{pmatrix} \tag{8}$$

With the notation of Proposition 3, and using the recursive structure of (7) and (8), we find that

$$\begin{aligned} h_{i,i}(\theta_1, \ldots, \theta_m)^{1/2} &= \left(\frac{1 + \theta_i - \lambda(1+\phi) - \theta_i^*}{\theta_i\{1 - \lambda(1+\phi) - \theta_i^*\}} \right)^{1/2} \\ &= \left(\frac{1 - \lambda(1+\phi) - \delta_i}{\theta_i\{1 - \lambda(1+\phi) - \delta_i - \theta_i\}} \right)^{1/2}, \quad i = 3, \ldots, m, \end{aligned}$$

where $\theta_i^* = \theta_3 + \cdots + \theta_i$ and $\delta_i = \sum_{j=3}^{i-1} \theta_j = \theta_i^* - \theta_i$.

Using Proposition 3 (i),

$$\begin{aligned} \pi(\theta_m \,|\, \phi, \lambda, \theta_3, \ldots, \theta_{m-1}) &\propto h_{m,m}^{1/2}(\phi, \lambda, \theta_3, \ldots, \theta_{m-1}) \\ &= \left(\frac{1 - \lambda(1+\phi) - \delta_m}{\theta_m\{1 - \lambda(1+\phi) - \delta_m - \theta_m\}} \right)^{1/2}, \quad \delta_m = \sum_{i=3}^{m-1} \theta_i, \end{aligned}$$

for those θ_m values such that $0 < \theta_1 + \theta_2 + \ldots + \theta_m \leq 1$ and, therefore, such that

$$0 < \lambda(1+\phi) + \delta_m + \theta_m < 1 \quad \Rightarrow \quad 0 < \theta_m < 1 - \lambda(1+\phi) - \delta_m.$$

Hence,

$$\pi(\theta_m \,|\, \phi, \lambda, \theta_3, \ldots, \theta_{m-1}) = \frac{\theta_m^{-1/2}(c-\theta_m)^{-1/2}}{\int_0^c \theta_m^{-1/2}(c-\theta_m)^{-1/2} d\theta_m}\ ,$$

where $c = 1-\lambda(1+\phi)-\delta_m$. Using Proposition A3 with $a = b = 1/2$, and $c = 1-\lambda(1+\phi)-\delta_m$,

$$\pi(\theta_m \,|\, \phi, \lambda, \theta_3, \ldots, \theta_{m-1}) = \frac{1}{\pi}{\theta_m}^{-1/2}(1-\lambda(1+\phi)-\delta_m-\theta_m)^{-1/2},$$

for $0 \leq \theta_m \leq 1-\lambda(1+\phi)-\delta_m$.

Using again Proposition 3 (i), $\pi(\theta_i \,|\, \phi, \lambda, \ldots, \theta_{i-1})$ is proportional to

$$\exp\left\{\int_{\Theta_{i+1}} \cdots \int_{\Theta_m} \log h_{i,i}^{1/2}(\phi, \lambda, \ldots, \theta_m) \left\{\prod_{j=i+1}^{m} \pi(\theta_j \,|\, \phi, \lambda, \ldots, \theta_{j-1})\right\} d\theta_{i+1} \ldots d\theta_m\right\}$$

where

$$\begin{aligned} h_{i,i}^{1/2}(\phi, \lambda, \ldots, \theta_m) &= h_{i,i}^{1/2}(\phi, \lambda, \ldots, \theta_i) \\ &= \left(\frac{1-\lambda(1+\phi)-\delta_i}{\theta_i\{1-\lambda(1+\phi)-\delta_i-\theta_i\}}\right)^{1/2}, \quad \delta_i = \sum_{j=3}^{i-1}\theta_j \end{aligned}$$

which does *not* depend on $\theta_{i+1}, \ldots, \theta_m$. Therefore,

$$\pi(\theta_i \,|\, \phi, \lambda, \ldots, \theta_{i-1}) \propto \exp\{\log h_{i,i}^{1/2}(\phi, \lambda, \ldots, \theta_i)\} = h_{i,i}^{1/2}(\phi, \lambda, \ldots, \theta_i)$$

and, hence,

$$\pi(\theta_i \,|\, \phi, \lambda, \ldots, \theta_{i-1}) = \frac{1}{\pi}{\theta_i}^{-1/2}(1-\lambda(1+\phi)-\delta_i-\theta_i)^{-1/2}, \quad 0 < \theta_i < 1-\lambda(1+\phi)-\delta_i.$$

Moreover, S_2, the upper 2×2 submatrix of $S(\phi, \lambda, \theta_3, \ldots, \theta_m)$, equals the matrix $S(\phi, \lambda)$ obtained in two parameter case and, hence, the same results obtain, namely,

$$\pi(\lambda \,|\, \phi) = \frac{(1+\phi)^{1/2}}{\pi}\lambda^{-1/2}\{1-\lambda(1+\phi)\}^{-1/2}, \quad 0 < \lambda < (1+\phi)^{-1},$$

$$\pi(\phi) = \frac{1}{\pi}\phi^{-1/2}(1+\phi)^{-1}, \quad 0 < \phi < \infty.$$

Thus, the *joint* reference prior required to obtain the reference posterior of the quantity of interest ϕ is

$$\begin{aligned} \pi(\boldsymbol{\theta}) &= \pi(\phi)\pi(\lambda \,|\, \phi)\prod_{i=3}^{m}\pi(\theta_i \,|\, \theta_{i-1}, \ldots, \theta_3, \lambda, \phi) \\ &= \phi^{-1/2}(1+\phi)^{-1/2}\lambda^{-1/2}\{1-\lambda(1+\phi)\}^{-1/2}\prod_{i=3}^{m}\theta_i^{-1/2}\Big\{1-\lambda(1+\phi)-\sum_{j=3}^{i}\theta_j\Big\}^{-1/2}, \quad (9) \end{aligned}$$

Using Bayes theorem with this joint prior and integrating out the nuisance parameters λ, $\theta_3, \ldots, \theta_m$, the desired reference posterior may be derived as

$$\begin{aligned}
&\pi(\phi \,|\, r_1, r_2, \ldots, r_m, n) \\
&\quad \propto \pi(\phi) \int_{\Lambda(\phi)} (\phi\lambda)^{r_1} \lambda^{r_2}\, \pi(\lambda \,|\, \phi) \int_{\Theta_3} \cdots \int_{\Theta_m} \prod_{j=3}^{m} \left\{ {\theta_j}^{r_j} \right\} (1 - \lambda(1+\phi) - \theta_j^*)^{r_{m+1}} \\
&\qquad\qquad \times \left\{ \prod_{j=3}^{m} \pi(\theta_j \,|\, \phi, \lambda, \ldots, \theta_{j-1}) \right\} d\lambda\, d\theta_3 \ldots d\theta_m \\
&\quad \propto \phi^{r_1 - 1/2} (1+\phi)^{-1/2} \int_{\Lambda(\phi)} \lambda^{r_1 + r_2 - 1/2} \{1 - \lambda(1+\phi)\}^{-1/2} \int_0^{c_3} {\theta_3}^{r_3 - 1/2}\, d\theta_3 \times \ldots \\
&\qquad \times \int_0^{c_m} {\theta_m}^{r_m - 1/2} \{1 - \lambda(1+\phi) - \delta_m - \theta_m\}^{r_{m+1} - 1/2} d\theta_m,
\end{aligned}$$

where $r_{m+1} = n - \sum_{j=1}^m r_j$, and $c_j = 1 - \lambda(1+\phi) - \delta_j$.

Using Proposition A3 with $a = 1/2$, $b = r_{m+1} + 1/2$ and $c = c_m$, the last integral is proportional to $(1 - \lambda(1+\phi) - \delta_m)^{r_{m+1}}$ and, therefore,

$$\begin{aligned}
&\pi(\phi \,|\, r_1, r_2, \ldots, r_m, n) \\
&\quad \propto \phi^{r_1 - 1/2} (1+\phi)^{-1/2} \int_{\Lambda(\phi)} \lambda^{r_1 + r_2 - 1/2} \{1 - \lambda(1+\phi)\}^{-1/2} \int_0^{c_3} {\theta_3}^{r_3 - 1/2} \ldots \\
&\quad \int_0^{c_{m-1}} {\theta_{m-1}}^{r_{m-1} - 1/2} \{1 - \lambda(1+\phi) - \delta_{m-1} - \theta_{m-1}\}^{r_{m+1} - 1/2}\, d\lambda\, d\theta_3 \ldots d\theta_{m-1};
\end{aligned}$$

thus, using Proposition A3 repeatedly,

$$\begin{aligned}
&\pi(\phi \,|\, r_1, r_2, \ldots, r_m, n) \\
&\quad \propto \phi^{r_1 - 1/2} (1+\phi)^{-1/2} \int_{\Lambda(\phi)} \lambda^{r_1 + r_2 - 1/2} \{1 - \lambda(1+\phi)\}^{r_{m+1} - 1/2} d\lambda\, \cdot
\end{aligned}$$

Finally, using Proposition A1 with $a = r_1 + r_2 + 1/2$, $b = r_{m+1} + 1/2$, and $c = 1 + \phi$, we have

$$\begin{aligned}
\pi(\phi \,|\, r_1, r_2, \ldots, r_m, n) &\propto \phi^{r_1 - 1/2} (1+\phi)^{-1/2} (1+\phi)^{-(r_1 + r_2 + 1/2)} \\
&\propto \phi^{r_1 - 1/2} (1+\phi)^{-(r_1 + r_2 + 1)},
\end{aligned}$$

as in the two parameter case, so that

$$\pi(\phi \,|\, r_1, \ldots, r_m, n) = \pi(\phi \,|\, r_1, r_2) = \frac{\Gamma(r_1 + r_2 + 1)}{\Gamma(r_1 + 1/2)\Gamma(r_2 + 1/2)} \frac{\phi^{r_1 - 1/2}}{(1+\phi)^{r_1 + r_2 + 1}}\, \cdot \qquad (10)$$

Thus, as we anticipated on intuitive grounds, reference inferences about θ_1/θ_2 only depend on r_1 and r_2; the number and distribution among the other categories of the remaining $n - r_1 - r_2$ observations —and, more importantly, the essentially *arbitrary* number m of considered categories—, are all *irrelevant* for inferences solely based on the multinomial model. It is easily verified that this is *not* true if conventional "noninformative" priors, such as a uniform prior, or Jeffreys' multivariate prior, are used instead of (9) in deriving a model-based posterior for ϕ.

3.4. *Coverage Probabilities*

The joint reference prior (9) satisfies Datta and Ghosh (1995a) conditions for probability matching in continuous multiparameter models; indeed, after some algebra, it is found that

$$\sum_{j=1}^{m} \frac{\partial}{\partial \theta_i} \eta_i(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) = 0 \tag{10}$$

where $\boldsymbol{\theta} = \{\phi, \lambda, \theta_3, \ldots, \theta_m\}$,

$$\eta(\boldsymbol{\theta}) = \frac{S(\phi, \lambda, \theta_3, \ldots, \theta_m) \bigtriangledown}{\sqrt{\bigtriangledown^t S(\phi, \lambda, \theta_3, \ldots, \theta_m) \bigtriangledown}}\,,$$

$\bigtriangledown = \{1, 0, \ldots, 0\}^t$, $S(\phi, \lambda, \theta_3, \ldots, \theta_m)$ is given by (8), and $\pi(\boldsymbol{\theta})$ is given by (9).

Extending Hartigan (1983) results with the techniques developed by Rousseau (1996), this suggests that, asymptotically, the continuity corrected coverage probabilities of one-sided credible intervals for ϕ of posterior probability p are equal to p to order n^{-1}.

Table 1. *Observed coverages of one-sided reference intervals for ϕ with posterior probability p. Mean and standard deviations of five runs of 10000 simulations, for several sample sizes.*

$\phi = 1/3$			
p	$n = 10$	$n = 25$	$n = 100$
0.05	0.0512±0.0016	0.0496±0.0024	0.0542±0.0010
0.25	0.2530±0.0020	0.2528±0.0033	0.2551±0.0019
0.50	0.5028±0.0036	0.5058±0.0032	0.5037±0.0046
0.75	0.7465±0.0054	0.7504±0.0026	0.7510±0.0050
0.95	0.9479±0.0014	0.9516±0.0016	0.9489±0.0017

$\phi = 3$			
p	$n = 10$	$n = 25$	$n = 100$
0.05	0.0514±0.0008	0.0543±0.0015	0.0533±0.0018
0.25	0.2576±0.0017	0.2594±0.0034	0.2551±0.0040
0.50	0.5073±0.0030	0.5095±0.0025	0.5066±0.0024
0.75	0.7494±0.0024	0.7506±0.0027	0.7497±0.0026
0.95	0.9500±0.0017	0.9491±0.0016	0.9519±0.0016

To analyze the coverage probabilities obtained for finite samples, we simulated 10000 samples $\{r_{i1}, r_{i2}\}$ of sizes $n = 10$, $n = 25$ and $n = 100$ from a multinomial distribution with $\theta_1 = 0.1$, $\theta_2 = 0.3$ (and therefore $\phi = 1/3$), and other 10000 samples of the same sizes from a multinomial distribution with $\theta_1 = 0.6$, $\theta_2 = 0.2$ (and therefore, $\phi = 3$). In both cases, the quantiles q_p^i, for $p = 0.05, 0.1, \ldots, 0.95$, and $i = 1, \ldots, 10000$ were computed for each sample, so that

$$\int_0^{q_p^i} \pi(\phi \,|\, r_{i1}, r_{i2}) = p, \quad i = 1, \ldots, 10000,$$

and, for each sample, we verified whether or not the p-credible interval $[0, q_p^i]$ contained the true value of ϕ, and thus computed the *observed* proportion of coverages. The whole procedure was replicated *five* times.

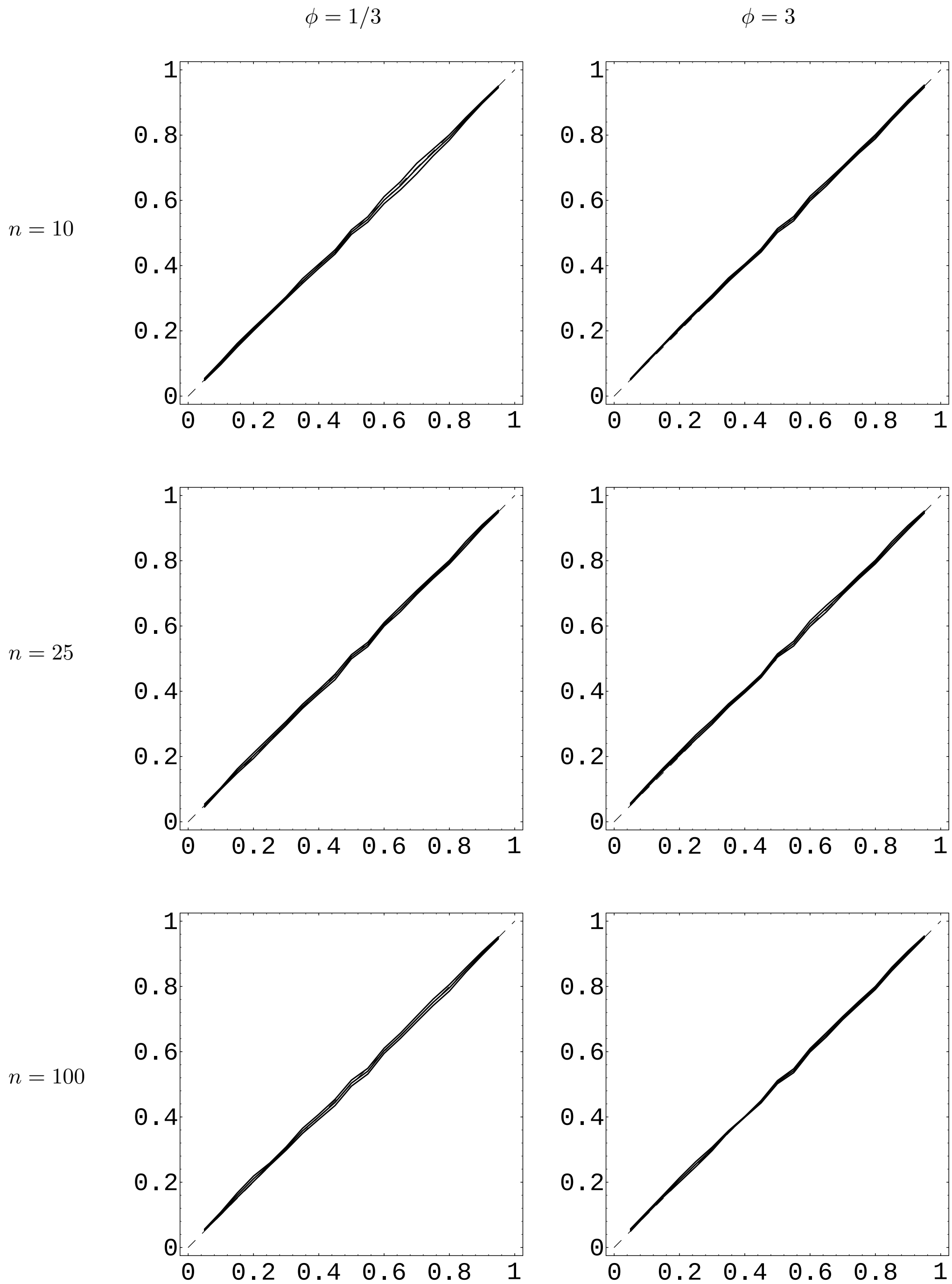

Figure 3. *Observed coverage probabilities versus reference posterior probabilities, with $\phi = 1/3$ and $\phi = 3$, for sample sizes $n = 10$, $n = 25$ and $n = 100$.*

In Table 1 we reproduce, for selected quantiles, the mean and standard deviation of the five observed coverages. In Figure 3 we offer a graphical presentation of the results, where the mean of the five observed coverages, and the bands obtained plus and minus two standard deviations, have been plotted against the corresponding reference posterior probabilities. It may be observed that even with rather moderate sample sizes, the reference posterior probabilities are not appreciably different from their observed coverages; as a matter of fact, the average coverage pattern is very much the same for the three sample sizes considered. This suggests that, even for moderate sample sizes, posterior reference intervals of ϕ are *well calibrated*, in the sense that if many samples of were to be taken from a given multinomial model, the corresponding reference intervals for ϕ with posterior probability p would contain the true value of ϕ with a relative frequency very close to p.

3.6. *Numerical Example*

In an expensive experiment designed to study possible improvements on the design of a new airbag, a random sample of 1200 airbags were destructively tested and 38 of them were found to be defective. The engineers found five different failure causes, which respectively accounted for 15, 12, 6, 3 and 2 of these failures, and judged them to be independent from each other. Moreover, it was decided that the optimal allocation of the resources available to improve the design crucially depended on the ratio of the probabilities of failure associated to the two most frequent causes of failure. Thus, with the notation above, one had $m = 5$, $n = 1200$, $r_1 = 15$, $r_2 = 12$, $r_3 = 6$, $r_4 = 3$, $r_5 = 2$, and $\phi = \theta_1/\theta_2$ is the quantity of interest.

Using (10), the reference posterior distribution of such quantity of interest is

$$\pi(\phi \,|\, \text{data}) = \pi(\phi \,|\, r_1 = 15, r_2 = 12) = \frac{\Gamma(28)}{\Gamma(15.5)\Gamma(12.5)} \frac{\phi^{14.5}}{(1+\phi)^{28}}\,, \tag{11}$$

shown in Figure 4, which does *not* depend on the total sample size 1200 or on the number, $m = 5$ of categories considered, or on the number or distribution of the failures which are not either of type 1 or type 2.

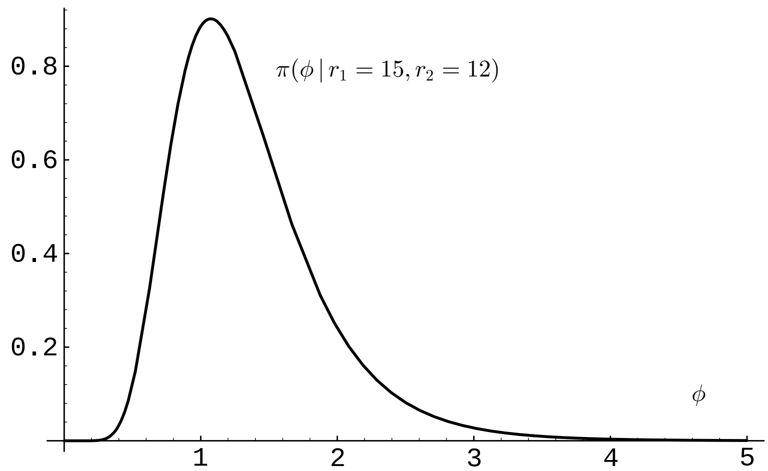

Figure 4. *Reference posterior density of the quantity of interest.*

The reference posterior has a mode at $14.5/13.5 = 1.074$; transforming to $w = \phi/(1+\phi)$ and using the incomplete Beta function, one easily finds that the median is 1.247 and that, for instance,

$$P[\phi > 1 \,|\, \text{data}] = 0.718, \quad P[\phi > 2 \,|\, \text{data}] = 0.112, \quad P[\phi > 3 \,|\, \text{data}] = 0.013.$$

Thus, given the results of the experiment, one may for instance report that type 1 is quite likely the more frequent cause of failure, and that it is probably about 1.25 times more likely than type 2, but surely less than 3 times more likely.

More formally, if a decision problem is contemplated with a utility structure of the form $u(d_i, \phi)$, then the optimal action may be found by maximizing

$$Eu[d_i \,|\, \text{data}] = \int_0^\infty u(d_i, \phi)\, \pi(\phi \,|\, \text{data})\, d\phi,$$

where $\pi(\phi \,|\, \text{data})$ is given by (11).

4. DISCUSSION AND FURTHER REFERENCES

In this paper, we have summarized the motivation, definition, and derivation of reference posterior distributions, we have illustrated the theory with an important example, and we have mentioned, —without proof—, some of the properties which may be used to substantiate the claim that they constitute the more promising available method to derive *non-subjective* prior distributions. However, the definition and possible uses of non-subjective priors, which under this and many other labels, —such as “conventional, “default”, “formal”, “neutral”, “flat” or “noninformative” —, are intended to provide Bayesian solutions which do not require to assess a subjective prior, have always been a rather polemic issue among statisticians. In this final section, we summarize some of the elements of the discussion, and provide signposts for those interested in pursuing the subject at a deeper level.

4.1. *Interpretation of Non-subjective Priors*

A major criticism to the use of non-subjective priors comes from subjectivist Bayesians, who argue that the prior should be an honest expression of the analyst’s prior knowledge and not a function of the model, specially if this involves integration over the sample space and hence may violate the likelihood principle. However, from a *foundational* viewpoint, the derivation of a reference posterior should be seen as part of a healthy *sensitivity analysis*, where it is desired to analyze the changes in the posterior of interest induced by changes in the prior: a reference posterior is just an answer to a *what if* question, namely what could be said about the quantity of interest given the data, if one’s prior knowledge were dominated by the data. If the experiment is changed the reference prior may be expected to change correspondingly; if subjective prior information is specified, the corresponding posterior could be compared with the reference posterior in order to assess the relative importance on the initial opinions in the final inference. Moreover, from a *pragmatic* point of view, it must be stressed that in the Bayesian analysis of the complex multiparameter models which are now systematically used as a consequence of the availability of numerical MCMC methods, —models typically intractable from a frequentist perspective—, there is little hope for a detailed assessment of a huge personal multivariate prior; the naïve use of some tractable “noninformative” prior may then hide important unwarranted assumptions which may easily dominate the analysis (see *e.g.*, Casella, 1996, and references therein). Careful, responsible choice of a non-subjective prior is then possibly the best available alternative.

It should also be mentioned here that some Bayesian statisticians would follow Jeffreys (1961) or Jaynes (1996) in a radical non-subjective view: they would claim that subjective priors are useless for scientific inference and so, non-subjective priors are necessary because there is nothing else to do.

4.2. *Improper Priors*

The reference priors that we have obtained in this paper have always been *proper* probability distributions; thus,

$$\int_0^\infty \pi(\phi)\, d\phi = \int_0^\infty \frac{1}{\pi}\phi^{-1/2}(1+\phi)^{-1}\, d\phi = 1,$$

even though $\Phi =]0, \infty[$ is not bounded. However, non-subjective priors associated to models with unbounded parameter spaces, —certainly including reference priors—, are typically *improper* in that, in most cases, if Φ is not compact, then $\int_\Phi \pi(\phi)\, d\phi = \infty$. This has often been criticized on the grounds that (i) foundational arguments require the use of a proper prior, and (ii) the use of improper priors may lead to unsatisfactory posteriors.

With respect to the foundational issue, we should point out that the natural axioms do *not* imply that the prior must be proper: they only lead to finite additivity, which is compatible with improper measures. However, the further natural assumption of *conglomerability* leads to σ-additivity and, hence, to proper measures; some signposts to this interesting debate are Heath and Sudderth (1978, 1989), Hartigan (1983), Cifarelli and Regazzini (1987), Seidenfeld (1987), Consonni and Veronese (1989) and Lindley (1996). It must be stressed however that, by definition, non-subjective priors are *not* intended to describe personal beliefs: they are *only* positive functions to be formally used in Bayes theorem to obtain non-subjective *posteriors*, —which indeed *should always be proper* given a minimum sample size—. Uncritical use of a "noninformative" prior may lead to an improper posterior (see *e.g.*, Berger, 1985, p. 187, for a well known example); the precise conditions for an improper prior to lead to a proper posterior are not known, but we are not aware of any example where the reference algorithm has lead to an improper posterior given a sample of minimum size. Moreover, non-subjective posteriors should be expressible as a *limit* of some sequence of posteriors derived from proper priors (Stein, 1965); this is precisely the procedure used to *define* reference distributions.

Finally, it is very important to emphasize that the use of a proper prior does certainly *not* guarantee a sensible behaviour of the resulting posterior. Indeed, if an improper prior leads to a posterior with undesirable properties, the posterior which would result from a proper approximation to that prior, —say that obtained by truncation of the parameter space—, will still have the same undesirable properties; for instance, the posterior of the sum of the squares of normal means $\phi = \sum_{j=1}^m \mu_j^2$ based on a joint uniform prior on the means $\pi(\mu_1, \ldots, \mu_m) \propto 1$ is extremely unsatisfactory as a non-subjective posterior (Stein, 1959), but so it is the posterior of ϕ based on the *proper* multinormal prior $\pi(\mu_1, \ldots, \mu_m) \propto \prod_i \mathrm{N}(\mu_i|0, \sigma)$, for large σ. Proper or improper, what must pragmatically be required from non-subjective priors is that, for any data set, they lead to sensible, data dominated, posterior distributions.

4.3. *Calibration*

Non-subjective posterior credible intervals are often numerically very close, and sometimes identical, to frequentist confidence intervals based on *sufficient* statistics (for an instructive discussion of how unsatisfactory confidence intervals may be when not based on sufficient statistics see Jaynes, 1976). Indeed, the analysis on the frequentist coverage probabilities of credible intervals derived from non-subjective posteriors, —in an attempt to verify whether or not they are "well calibrated —, has a very long history, and it does provide some bridges between frequentist and Bayesian inference. References within this topic include Lindley (1958), Welch and Peers (1963), Bartholomew (1965), Peers (1965, 1968), Welch (1965), Hartigan (1966, 1983), DeGroot (1973), Robinson (1975, 1978), Rubin (1984), Stein (1985), Chang and Villegas (1986), Tibshirani (1989), Dawid (1991), Severini (1991, 1993, 1994), Ghosh and Mukerjee (1992, 1993), Efron (1993), Mukerjee and Day (1993), Nicolau (1993), DiCiccio and Stern (1994), Samaniego and Reneau (1994), Datta and Ghosh (1995a) and Datta (1996).

This is a very active research area; indeed, the frequentist coverage probabilities of posterior credible intervals have often been an important element in arguing among competing non-subjective posteriors, as in Stein (1985), Efron (1986), Tibshirani (1989), Berger and Bernardo (1989), Ye and Berger (1991), Liseo (1993), Berger and Yang (1994), Yang and Berger (1994), Ghosh, Carlin and Srivastava (1995) and Sun and Ye (1995). Reference posteriors have consistently been found to have very attractive coverage properties, even for small samples, but no general results have been established.

4.4. *Further Signposts*

The classic books by Jeffreys (1961), Lindley (1965) and Box and Tiao (1973) are a must for anyone interested in non-subjective Bayesian inference; they prove that most “textbook” inference problems have a simple non-subjective Bayesian solution, and one which produces credible intervals which are often, *numerically*, either identical or very close to their frequentist “accepted” counterparts, but much easier to obtain. Zellner (1971) is a textbook on econometrics from a non-subjective Bayesian viewpoint; Geisser (1993) summarizes many results on non-subjective posterior *predictive* distributions.

The construction of non-subjective posterior distributions has a very interesting history, which dates back to Laplace (1812), and includes Jeffreys (1946, 1961), Perks (1947), Lindley (1961), Geisser and Cornfield (1963), Welch and Peers (1963), Hartigan (1964, 1965), Novick and Hall (1965), Jaynes (1968, 1971), Good (1969), DeGroot (1970, Ch. 10), Villegas (1971, 1977, 1981) Box and Tiao (1973, Sec. 1.3), Zellner (1977, 1986), Akaike (1978), Bernardo (1979), Geisser (1979, 1984), Rissanen (1983), Tibshirani (1989) and Berger and Bernardo (1989, 1992c) as some of the more influential contributions. The development of this long quest may conveniently be traced from Bernardo and Smith (1994, Sec. 5.6.2), Kass and Wasserman (1996), and references therein.

Some recent developments include Ghosh and Mukerjee (1992), Mukerjee and Dey (1993), Clarke and Wasserman (1993), George and McCulloch (1993), Clarke and Barron (1994), Wasserman and Clarke (1995), Datta and Ghosh (1995b, 1995c, 1996) and Zellner (1996). Yang and Berger (1996) is a partial *catalog*, alphabetically ordered by probability model, of many non-subjective priors which have been suggested in the literature. Bernardo (1997) is a non technical analysis, in a dialog format, on the *foundational* issues involved, and it is followed by a discussion.

For someone specifically interested in reference distributions, the original paper, Bernardo (1979), is easily read and it is followed by a very lively discussion; Bernardo (1981) extends the theory to general decision problems; Berger and Bernardo (1989, 1992c) contain crucial mathematical extensions. A textbook level description of reference analysis is provided in Bernardo and Smith (1994, Sec. 5.4).

Papers which contain explicit analysis of specific reference distributions include Bernardo (1977, 1978, 1979, 1980, 1982, 1985), Bayarri (1981, 1985), Ferrándiz (1982, 1985), Sendra (1982), Eaves (1983a, 1983b, 1985), Bernardo and Bayarri (1985), Chang and Villegas (1986), Hills (1987), Mendoza (1987, 1988), Bernardo and Girón (1988), Lindley (1988), Berger and Bernardo (1989, 1992a, 1992b, 1992c), Pole and West (1989), Chang and Eaves (1990), Polson and Wasserman (1990), Ye and Berger (1991), Stephens and Smith (1992), Liseo (1993), Ye

(1993, 1994, 1995), Berger and Yang, (1994) Kubokawa and Robert (1994), Yang and Berger (1994, 1996), Datta and Ghosh (1995c) Ghosh, Carlin and Srivastava (1995), Sun and Ye (1995), Ghosal (1996) and Reid (1996).

ACKNOWLEDGEMENTS

The authors are most grateful to Jim Berger, John Hartigan, Judith Rousseau and Larry Wasserman for helpful comments on an earlier version of this paper.

REFERENCES

Akaike, H. (1978). A new look at the Bayes procedure. *Biometrika* **65**, 53–59.

Barnard, G. A. (1995). Pivotal models and the fiducial argument. *Internat. Statist. Rev.* **63**, 309–323; correction 64, 137.

Bartholomew, D. J. (1965). A comparison of some Bayesian and frequentist inferences. *Biometrika* **52**, 19–35.

Bayarri, M. J. (1981). Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trab. Estadist.* **32**, 18–31.

Bayarri, M. J. (1985). Bayesian inference on the parameters of a Beta distribution. *Statistics and Decisions* **2**, 17–22.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Berlin: Springer.

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

Berger, J. O. and Bernardo, J. M. (1992a). Reference priors in a variance components problem. *Bayesian Analysis in Statistics and Econometrics* (P. K. Goel and N. S. Iyengar, eds.). Berlin: Springer, 323–340.

Berger, J. O. and Bernardo, J. M. (1992b). Ordered group reference priors with applications to a multinomial problem. *Biometrika* **79**, 25–37.

Berger, J. O. and Bernardo, J. M. (1992c). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Berger, J. O., Bernardo, J. M. and Mendoza, M. (1989). On priors that maximize expected information. *Recent Developments in Statistics and their Applications* (J. P. Klein and J. C. Lee, eds.). Seoul: Freedom Academy, 1–20.

Berger, J. O. and Strawderman, W. E. (1986). Choice of hierarchical priors: admissibility of estimation of normal means. *Ann. Statist.* **24**, 931–951.

Berger, J. O. and Yang, R. (1994). Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory* **10**, 461–482.

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

Bernardo, J. M. (1978). Unacceptable implications of the left Haar measure in a standard normal theory inference problem *Trab. Estadist.* **29**, 3–9.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference* (N. G. Polson and G. C. Tiao, eds.), Brookfield, VT: Edward Elgar, (1995), 229–263.

Bernardo, J. M. (1980). A Bayesian analysis of classical hypothesis testing. *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Valencia: University Press, 605–647 (with discussion).

Bernardo, J. M. (1981). Reference decisions. *Symposia Mathematica* **25**, 85–94.

Bernardo, J. M. (1982). Contraste de modelos probabilísticos desde una perspectiva Bayesiana. *Trab. Estadist.* **33**, 16–30.

Bernardo, J. M. (1985). On a famous problem of induction. *Trab. Estadist.* **36**, 24–30.

Bernardo, J. M. (1997). Noninformative priors do not exist: a discussion with José M. Bernardo. *J. Statist. Planning and Inference* , (to appear, with discussion).

Bernardo, J. M. and Bayarri, M. J. (1985). Bayesian model criticism. *Model Choice* (J.-P. Florens, M. Mouchart, J.-P. Raoult and L. Simar, eds.). Brussels: Pub. Fac. Univ. Saint Louis, 43–59.

Bernardo, J. M. and Girón F. J. (1988). A Bayesian analysis of simple mixture problems. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Oxford: University Press, 67–88 (with discussion).

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

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

Casella, G. (1996). Statistical inference and Monte Carlo algorithms. *Test* **5**, 249–344 (with discussion).

Clarke, B. (1996). Implications of reference priors for prior infrmation and for sample size. *J. Amer. Statist. Assoc.* **91**, 173–184.

Clarke, B. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Planning and Inference* **41**, 37–60.

Clarke, B. and Wasserman, L. (1993). Non-informative priors and nuisance parameters. *J. Amer. Statist. Assoc.* **88**, 1427–1432.

Chang, T. and Eaves, D. M. (1990). Reference priors for the orbit of a group model. *Ann. Statist.* **18**, 1595–1614.

Chang, T. and Villegas, C. (1986). On a theorem of Stein relating Bayesian and classical inferences in group models. *Canadian J. Statist.* **14**, 289–296.

Cifarelli, D. M. and Regazzini, E. (1987). Priors for exponential families which maximize the association between past and future observations. *Probability and Bayesian Statistics* (R. Viertl, ed.). London: Plenum, 83–95.

Consonni, G. and Veronese, P. (1989). Some remarks on the use of improper priors for the analysis of exponential regression problems. *Biometrika* **76**, 101–106.

Datta, G. S. (1996). On priors providing frequentist validity for Bayesian inference of multiple parametric functions. *Biometrika* **83**, 287–298.

Datta, G. S. and Ghosh, J. K. (1995a). On priors providing a frequentist validity for Bayesian inference. *Biometrika* **82**, 37–45.

Datta, G. S. and Ghosh, J. K. (1995b). Noninformative priors for maximal invariant parameter in group models. *Test* **4**, 95–114.

Datta, G. S. and Ghosh, M. (1995a). Some remarks on noninformative priors. *J. Amer. Statist. Assoc.* **90**, 1357–1363.

Datta, G. S. and Ghosh, M. (1995b). Hierarchical Bayes estimators of the error variance in one-way ANOVA models. *J. Statist. Planning and Inference* **45**, 399–411.

Datta, G. S. and Ghosh, M. (1996). On the invariance of noninformative priors. *Ann. Statist.* **24**, 141–159.

Dawid, A. P. (1983). Invariant prior distributions. *Encyclopedia of Statistical Sciences* **4** (S. Kotz, N. L. Johnson and C. B. Read, eds.). New York: Wiley, 228–236.

Dawid, A. P. (1991). Fisherian inference in likelihood and prequential frames of reference. *J. Roy. Statist. Soc. B* **53**, 79—109 (with discussion).

Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. B* **35**, 189–233 (with discussion).

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

DeGroot, M. H. (1973). Doing what comes naturally: interpreting a tail area as a posterior probability or as a likelihood ratio. *J. Amer. Statist. Assoc.* **68**, 966–969.

DiCiccio, T. J. and Stern, S. E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihood. *J. Roy. Statist. Soc. B* **56**, 397–408.

du Plessis, J. L., van der Merwe, A. J. and Groenewald, P. C. N. (1995). Reference priors for the multivariate calibration problem. *South African Statist. J.* **29**, 155-168.

Eaves, D. M. (1983a). On Bayesian nonlinear regression with an enzyme example. *Biometrika* **70**, 373-379.

Eaves, D. M. (1983b). Minimally informative prior analysis of a non-linear model. *The Statistician* **32**, 117.

Eaves, D. M. (1985). On maximizing the missing information about a hypothesis. *J. Roy. Statist. Soc. B* **47**, 263-266.

Efron, B. (1986). Why isn't everyone a Bayesian? *Amer. Statist.* **40**, 1–11 (with discussion).

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* **80**, 3–26.

Efstathiou, M. (1996). Some Aspects of Approximation and Computation for Bayesian Inference. Ph.D. Thesis, Imperial College London.

Ferrándiz, J. R. (1982). Una solución Bayesiana a la paradoja de Stein. *Trab. Estadist.* **33**, 31–46.

Ferrándiz, J. R. (1985). Bayesian inference on Mahalanobis distance: an alternative approach to Bayesian model testing. *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.), Amsterdam: North-Holland, 645–654.

Geisser, S. (1979). In discussion of Bernardo (1979). *J. Roy. Statist. Soc. B* **41**, 136–137.

Geisser, S. (1984). On prior distributions for binary trials. *J. Amer. Statist. Assoc.* **38**, 244–251 (with discussion).

Geisser, S. (1993). *Predictive Inference: an Introduction*. London: Chapman and Hall.

Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. B* **25**, 368–376.

George, E. I. and McCulloch, R. (1993). On obtaining invariant prior distributions. *J. Statist. Planning and Inference* **37**, 169–179.

Ghosal, S. (1996). Reference priors in multiparameter nonregular cases. *Test* **5**, (to appear).

Ghosh, J. K. and Mukerjee, R. (1992). Non-informative priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 195–210 (with discussion).

Ghosh, J. K. and Mukerjee, R. (1993). On priors that match posterior and frequentist distribution functions. *Canadian J. Statist.* **21**, 89–96.

Ghosh, M., Carlin, B. P. and Srivastava, M. S. (1995). Probability matching priors for linear calibration. *Test* **4**, 333–357.

Ghosh, M. and Yang, M.-Ch. (1996). Non-informative priors for the two sample normal problem. *Test* **5**, 145–157.

Good, I. J. (1969). What is the use of a distribution? *Multivariate Analysis* **2** (P. R. Krishnaiah, ed.). New York: Academic Press, 183–203.

Hartigan, J. A. (1964). Invariant prior distributions. *Ann. Math. Statist.* **35**, 836–845.

Hartigan, J. A. (1965). The asymptotically unbiased prior distribution. *Ann. Math. Statist.* **36**, 1137–1152.

Hartigan, J. A. (1966). Note on the confidence prior of Welch and Peers. *J. Roy. Statist. Soc. B* **28**, 55-56.

Hartigan, J. A. (1983). *Bayes Theory*. Berlin: Springer.

Hartigan, J. A. (1996). Locally uniform prior distributions. *Ann. Statist.* **24**, 160-173.

Heath, D. L. and Sudderth, W. D. (1978). On finitely additive priors, coherence and extended admissibility. *Ann. Statist.* **6**, 333–345.

Heath, D. L. and Sudderth, W. D. (1989). Coherent inference from improper priors and from finitely additive priors. *Ann. Statist.* **17**, 907–919.

Hills, S. E. (1987). Reference priors and identifiability problems in non-linear models. *The Statistician* **36**, 235–240.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Amer. Statist. Assoc.* **91**, 1461–1473.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys' prior. *J. Amer. Statist. Assoc.* **86**, 981–986.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Trans. Systems, Science and Cybernetics* **4**, 227–291.

Jaynes, E. T. (1971). The well posed problem. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Toronto: Holt, Rinehart and Winston, 342–356 (with discussion).

Jaynes, E. T. (1976). Confidence intervals vs. Bayesian intervals. *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science* **2** (W. L. Harper and C. A. Hooker eds.). Dordrecht: Reidel, 175–257 (with discussion).

Jaynes, E. T. (1996). *Probability Theory As Extended Logic*. Posted in the Web at <http://bayes.wustl.edu/>.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. A* **186**, 453–461.

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

Johnson, N. L., Kotz, S. and Balakrishan, N. (1995). *Continuous Univariate Distributions* **2**, (2nd. ed.) New York: Wiley.

Kadane, J. B., Schervish, M. J. and Seidenfeld, T. (1986). Statistical implications of finitely additive probability. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti* (P. K. Goel and A. Zellner, eds.). Amsterdam: North-Holland, 59–76.

Kass, R. E. (1989). The geometry of asymptotic inference. ß4, 188–234.

Kass, R. E. (1990). Data-translated likelihood and Jeffreys' rule. *Biometrika* **77**, 107–114.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.* **91**, 1343–1370.

Kubokawa, T. and Robert, C. P. (1994). New perspectives in linear calibration. *J. Multivariate Analysis* **51**, 178–200.

Lane, D. A. and Sudderth, W. D. (1983). Coherent and continuous inference. *Ann. Statist.* **11**, 114–120.

Laplace, P. S. (1812). *Théorie Analytique des Probabilités*. Paris: Courcier. Reprinted as *Oeuvres Complètes de Laplace* **7**, 1878–1912. Paris: Gauthier-Villars.

Lindley, D. V. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist.* **27**, 986–1005.

Lindley, D. V. (1958). Fiducial distribution and Bayes' theorem. *J. Roy. Statist. Soc. B* **20**, 102–107.

Lindley, D. V. (1961). The use of prior probability distributions in statistical inference and decision. *Proc. Fourth Berkeley Symp.* **1** (J. Neyman and E. L. Scott, eds.). Berkeley: Univ. California Press, 453–468.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint. Vol. 2: Statistics*. Cambridge: University Press.

Lindley, D. V. (1988). Statistical inference concerning the Hardy-Weinberg equilibrium. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Oxford: University Press, 307–326 (with discussion).

Lindley, D. V. (1997). Some comments on Bayes Factors. *J. Statist. Planning and Inference* **61**, 181–189.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80**, 295–304.

Mendoza, M. (1987). A Bayesian analysis of a generalized slope ratio bioassay. *Probability and Bayesian Statistics* (R. Viertl, ed.). London: Plenum, 357–364.

Mendoza, M. (1988). Inferences about the ratio of linear combinations of the coefficients in a multiple regression problem. *Bayesian Statistics 3* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.). Oxford: University Press, 705–711.

Mendoza, M. (1994). Asymptotic posterior normality under transformations. *Test* **3**, 173–180.

Monette, G., Fraser, D. A. S. and Ng, K. W. (1984). Marginalization, likelihood and structural models, *Multivariate Analysis* **6** (P. R. Krishnaiah, ed.). Amsterdam: North-Holland, 209–217.

Mukerjee, R. and Dey, D. K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: Higher order asymptotics. *Biometrika* **80**, 499–505.

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

Nicolau, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. Roy. Statist. Soc. B* **55**, 377–390.

Novick, M. R. and Hall, W. K. (1965). A Bayesian indifference procedure. *J. Amer. Statist. Assoc.* **60**, 1104–1117.

Peers, H. W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. Roy. Statist. Soc. B* **27**, 9–16.

Peers, H. W. (1968). Confidence properties of Bayesian interval estimates. *J. Roy. Statist. Soc. B* **30**, 535–544.

Perks, W. (1947). Some observations on inverse probability, including a new indifference rule. *J. Inst. Actuaries* **73**, 285–334 (with discussion).

Pole, A. and West, M. (1989). Reference analysis of the dynamic linear model. *J. Time Series Analysis* **10**, 13–147.

Polson, N. G. (1992). On the expected amount of information from a non-linear model. *J. Roy. Statist. Soc. B* **54**, 889–895.

Polson, N. G. and Wasserman, L. (1990). Prior distributions for the bivariate binomial. *Biometrika* **76**, 604–608.

Reid, N. (1996). Likelihood and Bayesian approximation methods. *Bayesian Statistics 5* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 349–366 (with discussion).

Rao, C. R. and Mukerjee, R. (1995). On posterior credible sets based on the score statistic. *Statistica Sinica* **5**, 781–791.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Ann. Statist.* **11**, 416–431.

Robinson, G. K. (1975). Some counter-examples to the theory of confidence intervals. *Biometrika* **62**, 155–161.

Robinson, G. K. (1978). On the necessity of Bayesian inference and the construction of measures of nearness to Bayesian form. *Biometrika* **65**, 49–52.

Rousseau, J. (1997). *Etude des Propietés Asymptotiques des Estimateurs de Bayes*. Ph.D. Thesis, Université Paris VI..

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12**, 1151–1172.

Samaniego, F. J. and Reneau, D. M. (1994). Toward a reconciliation of the Bayesian and frequentist approaches to point estimation. *J. Amer. Statist. Assoc.* **89**, 847–957.

Seidenfeld, T. (1987). Entropy and uncertainty. *Foundations of Statistical Inference*. (I. B. MacNeill and G. J. Umphrey eds.). Dordrecht: Reidel, 259–287.

Sendra, M. (1982). Distribución final de referencia para el problema de Fieller-Creasy. *Trab. Estadist.* **33**, 55–72.

Severini, T. A. (1991). On the relationship between Bayesian and non-Bayesian interval estimates. *J. Roy. Statist. Soc. B* **53**, 611–618.

Severini, T. A. (1993). Bayesian interval estimates which are also confidence intervals. *J. Roy. Statist. Soc. B* **53**, 611-618.

Severini, T. A. (1994). Approximately Bayesian inference. *J. Amer. Statist. Assoc.* **89**, 242–249.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27** 379–423 and 623–656. Reprinted in *The Mathematical Theory of Communication* (Shannon, C. E. and Weaver, W., 1949). Urbana, IL.: Univ. Illinois Press.

Soofi, E. S. (1994). Capturing the intangible concept of information. *J. Amer. Statist. Assoc.* **89**, 1243–1254.

Stephens, D. A. and Smith, A. F. M. (1992). Sampling-resampling techniques for the computation of posterior densities in normal means problems. *Test* **1**, 1–18.

Stein, C. (1959). An example of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.* **30**, 877–880.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc. B* **24**, 265–296 (with discussion).

Stein, C. (1965). Approximation of improper prior measures by proper probability measures. *Bernoulli, Bayes, Laplace Festschrift*. (J. Neyman and L. LeCam, eds.). Berlin: Springer, 217–240.

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Sequential Methods in Statistics* (R. Zielisnsi, ed.). Warsaw: Polish Scientific Pub., 485—514.

Stone, M. (1976). Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.* **71**, 114–125 (with discussion).

Stone, M. (1982). Review and analysis of some inconsistencies related to improper distributions and finite additivity. *Proc. 6th Int. Conf. Logic, Methodology and Philosophy of Science* (L. J. Cohen, J. Los, H. Pfeiffer and K. P. Podewski, eds.). Amsterdam: North-Holland, 413–426.

Stone, M. and Dawid, A. P. (1972). Un-Bayesian implications of improper Bayesian inference in routine statistical problems. *Biometrika* **59**, 369–375.

Sun, D. and Ye, K. (1995). Reference prior Bayesian analysis for normal mean products. *J. Amer. Statist. Assoc.* **90**, 589–597.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604–608.

Venegas, F. (1990). On regularity and optimality conditions for maximum entropy priors. *Rev. Bras. Probab. Estatis.* **4**, 105–136.

Villegas, C. (1971). On Haar priors. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.). Toronto: Holt, Rinehart and Winston, 409–414 (with discussion).

Villegas, C. (1977). On the representation of ignorance. *J. Amer. Statist. Assoc.* **72**, 651–654.

Villegas, C. (1981). Inner statistical inference II. *Ann. Statist.* **9**, 768–776.

Wasserman, L. and Clarke, B. (1995). Information tradeoff. *Test* **4**, 19–38.

Welch, B. L. (1965). On comparisons between confidence point procedures in the case of a single parameter. *J. Roy. Statist. Soc. B* **27**, 1–8.

Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *J. Roy. Statist. Soc. B* **25**, 318–329.

Yang, R. (1995). Invariance of the reference prior under reparametrization. *Test* **4**, 83–94.

Yang, R and Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22**, 1195–1211.

Yang, R and Berger, J. O. (1996). A catalog of noninformative priors. *Tech. Rep.*, Dept. Statistics, Purdue University, USA..

Yang, R and Chen, M.-H. (1995). Bayesian analysis of random coefficient regression models using noninformative priors. *J. Multivariate Analysis* **55**, 283–311.

Ye, K. (1993). Reference priors when the stopping rule depends on the parameter of interest. *J. Amer. Statist. Assoc.* **88**, 360–363.

Ye, K. (1994). Bayesian reference prior analysis on the ratio of variances for the balanced one-way random effect model. *J. Statist. Planning and Inference* **41**, 267–280.

Ye, K. (1995). Selection of the reference priors for a balanced random effects model. *Test* **4**, 1–17.

Ye, K. and Berger, J. O. (1991). Non-informative priors for inferences in exponential regression models. *Biometrika* **78**, 645–656.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley. Reprinted in 1987, Melbourne, FL: Krieger.

Zellner, A. (1977). Maximal data information prior distributions. *New Developments in the Applications of Bayesian Methods* (A. Aykaç and C. Brumat, eds.). Amsterdam: North-Holland, 211–232.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti* (P. K. Goel and A. Zellner, eds.). Amsterdam: North-Holland, 233–243.

Zellner, A. (1996). Models, prior information and Bayesian analysis. *J. Econometrics* **75**, 51–68.

APPENDIX

Proposition A1. *For* $a > 0$, $b > 0$, $c \geq 1$,

$$\int_0^{\frac{1}{c}} x^{a-1}(1-cx)^{b-1}dx = \frac{1}{c^a}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,.$$

Proof. The change $y = cx$ reduces this to a standard Beta integral. $\triangleleft$

Proposition A2. *For* $a > 0$, $b > 0$, $c > 1$, *if*

$$p(x\,|\,a,b,c) = c^a\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-cx)^{b-1}, \quad 0 < x < c^{-1},$$

then,

$$E[\log x] = \log\frac{1}{c} + \psi(a) - \psi(a+b),$$

where $\psi(.) = \Gamma^{'}(x)/\Gamma(x)$ *is the digamma function. In particular, if* $a = b = 1/2$, *then* $E[\log x] = -\log 4c$.

Proof. Taking logarithms in Proposition A1,

$$\log\int_0^{c^{-1}} x^{a-1}(1-cx)^{b-1}dx = -a\log c + \log\Gamma(a) + \log\Gamma(b) - \log\Gamma(a+b),$$

and taking derivatives with respect to a,

$$\frac{\int_0^{c^{-1}} \log x\; x^{a-1}(1-cx)^{b-1}dx}{\int_0^{c^{-1}} x^{a-1}(1-cx)^{b-1}dx} = -\log c + \psi(a) - \psi(a+b);$$

but the left hand side is $\int_0^{c^{-1}} \log x\; p(x\,|\,a,b,c)\,dx = E[\log x]$. The particular case follows from the fact that $\psi(1/2) - \psi(1) = -2\log 2$. $\triangleleft$

Proposition A3. *For* $a > 0$, $b > 0$, $c \geq 1$,

$$\int_0^{c} x^{a-1}(c-x)^{b-1}dx = c^{a+b-1}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,.$$

Proof. The change $y = x/c$ reduces this to a standard Beta integral. $\triangleleft$

Proposition A4. *For* $a > 0$, $b > 0$,

$$\int_0^{\infty} \frac{x^{a-1}}{(1+x)^{a+b}}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,.$$

Proof. The change $y = (1+x)^{-1}$ reduces this to a standard Beta integral. $\triangleleft$