A Catalog of Noninformative Priors *

Ruoyong Yang and James O. Berger
Department of Statistics, Purdue University

August, 1996

Abstract

[PRELIMINARY DRAFT: This draft of the catalog is incomplete, with much remaining to be filled in and/or added. We are circulating this crude draft in the hopes that readers will know of relevant information that should be added.]

A variety of methods of deriving noninformative priors have been developed, and applied to a wide variety of statistical models. In this paper we provide a catalog of many of the resulting priors, and list known properties of the priors. Emphasis is given to reference priors and the Jeffreys prior, although other approaches are also considered.

Key words and phrases. Jeffreys prior, reference prior, maximal data information prior.

1 Introduction

1.1 Motivation

The literature on noninformative priors has grown enormously over recent years. There have been several excellent books or review articles that have been concerned with discussing or comparing different approaches to developing noninformative priors (e.g., Kass and Wasserman, 1993), but there has been no systematic effort to catalog the noninformative priors that have been developed. Since use of noninformative priors is becoming routine in Bayesian practice, preparation of such a catalog seemed in order.

*This research was supported by NSF grants DMS-8923071 and DMS-9303556 at Purdue University.

Although general discussion is not the purpose of this catalog, it is useful to review the numerous reasons that noninformative priors are important to Bayesian analysis:

(i) Frequently, elicitation of subjective prior distributions is impossible, because of time or cost limitations, or resistance or lack of training of clients. Automatic or default prior distributions are then needed.

(ii) The statistical analysis is often *required* to *appear* objective. Of course, true objectivity is virtually never attainable, and the prior distribution is usually the least of the problems in terms of objectivity, but use of a subjectively elicited prior significantly reduces the *appearance* of objectivity. Noninformative priors not only preserve this appearance, but can be argued to result in analyses that are more objective than most classical analyses.

(iii) Subjective elicitation can easily result in poor prior distributions, because of systematic elicitation bias and the fact that elicitation typically yields only a few features of the prior, with the rest of the prior (e.g, its functional form) being chosen in a convenient, but possibly inappropriate, way. It is thus good practice to compare answers from a subjective analysis with answers from a noninformative prior analysis. If there are substantial differences, it is important to check that the differences are due to features of the prior that are trusted, and not due to either unelicited "convenience" features of the prior, or suspect elicitations.

(iv) In high dimensional problems, the best one can typically hope for is to develop subjective priors for the "important" parameters, with the unimportant or "nuisance" parameters being given noninformative priors.

(v) Good noninformative priors can be somewhat magical in multiparameter problems. As an example, the *Jeffreys prior* seems to almost always yield a proper posterior distribution. This is "magical," in that the common constant (or *uniform*) prior will much more frequently fail to yield a proper posterior. Even better, the *reference* prior approach has repeatedly yielded multiparameter priors that overcome limitations of the Jeffreys prior, and yield surprisingly good performance from almost any perspective. The point here is that, in multiparameter problems, inappropriate aspects of priors (even proper ones) can accumulate across dimensions in very detrimental ways; *reference* priors seem to "magically" avoid such inappropriate accumulation.

(vi) Bayesian analysis with noninformative priors is being increasingly recognized as a method for classical statisticians to obtain good classical procedures. For instance, the *frequentist-*

matching approach to developing noninformative priors is based on ensuring that one has Bayesian credible sets with good frequentist properties, and it turns out that this is probably the best way to find good frequentist confidence sets.

1.2 Approaches to Development of Noninformative Priors

We do not attempt a thorough discussion of the various approaches. See, e.g., Kass and Wasserman (1993), for such discussion. We primarily will just define the various approaches, and give relevant references.

The *Uniform* Prior: By this, we just mean the constant density, with the constant typically being chosen to be 1 (unless the constant can be chosen to yield a proper density). This choice was, of course, popularized by Laplace (1812).

The *Jeffreys* Prior: This is defined as $\pi(\theta) = \sqrt{\det(I(\theta))}$, where $I(\theta)$ is the Fisher information matrix. This was proposed in Jeffreys (1961), as a solution to the problem that the uniform prior does not yield an analysis invariant to choice of parameterization. Note that, in specific situations, Jeffreys often recommended noninformative priors that differed from the formal Jeffreys prior.

The *Reference* Prior: This approach was developed in Bernardo (1979), and modified for multiparameter problems in Berger and Bernardo (1992c). The approach cannot be simply described, but it can be roughly thought of as trying to modify the Jeffreys prior by reducing the dependence among parameters that is frequently induced by the Jeffreys prior; there are many well-known examples in which the Jeffreys prior yields poor performance (even inconsistency) because of this dependence.

The *Maximal Data Information* Prior (MDIP): This approach was developed in Zellner (1971), based on an information argument. It is given by $\pi(\theta) = \exp\{\int p(x|\theta) \log p(x|\theta) dx\}$, where $p(x|\theta)$ is the data density function.

2 Organization and Notation

The catalog is organized around statistical models, with the models being listed in alphabetical order. Each model-entry is kept as self-contained as possible. Listed for each are (i) the model density; (ii) various noninformative priors; and (iii) certain of the resulting posteriors and their properties. Category (iii) information is often very limited.

Notation is standard. This include $\pi(\theta|D)$, $|A|$ =determinant of A, $\pi(\theta)$ is a density w.r.t. $d\theta$.

Important Notation: Noninformative priors that are proper (i.e., integrate to 1) are given in bold type. Others are improper. (The distinction is important for testing problems, where proper distributions are typically needed; for estimation and prediction, improper noninformative priors are typically fine.)

3 AR(1)

The AR(1) model, in which the data $\underset{\sim}{X} = (X_1, ..., X_T)$ follow the model

$$X_t = \rho X_{t-1} + \epsilon_t \ ,$$

where the ϵ_t are i.i.d. $N(0,\ \sigma^2)$.

The expressions below are for σ^2 known. If σ^2 is unknown, multiply by $\frac{1}{\sigma}d\sigma$ (or $\frac{1}{\sigma^2}d\sigma^2$).

Prior	$\pi(\rho)$	(Marginal) Posterior
Uniform	1	all are proper
Jeffreys	$\left[\frac{T}{1-\rho^2} + \frac{1-\rho^{2T}}{1-\rho^2}\left\{\frac{E[X_0^2]}{\sigma^2} - \frac{1}{1-\rho^2}\right\}\right]^{1/2}$	
Reference[1]	$\exp\{\frac{1}{2}E[\log(\sum_{i=1}^T X_{i-1}^2)]\}$	
Reference[2]	$\begin{cases} 1/[2\pi\sqrt{1-\rho^2}] & \text{if } \mid\rho\mid < 1 \\ 1/[2\pi \mid\rho\mid \sqrt{\rho^2-1}] & \text{if } \mid\rho\mid > 1. \end{cases}$	

1. Nonasymptotic reference prior.

2. Symmetrized reference prior, recommended for typical use. See Berger and Yang (1992) for comparison of the noninformative priors.

4 Behrens-Fisher Problem

Let $x_1, \ldots, x_n$ be i.i.d. observations from $N(\xi, \sigma^2$ and $y_1, \ldots, y_n$ be i.i.d. observations from $N(\eta, \tau^2$. The parameters of interest are $\theta = \xi - \eta$ and $\lambda = \xi + \eta$.

Liseo (1993) computed the Jeffreys prior and reference prior for this problem as

Prior	$\pi(\theta, \lambda, \sigma^2, \tau^2)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$1/(\sigma\tau)^3$	proper
Reference[1]	$1/(\sigma\tau)^2$	

1. Independent of the group ordering of the parameters.

5 Beta

The $Be(\alpha,\ \beta)$, $\alpha > 0$, $\beta > 0$, density is

$$f(x|\alpha,\ \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{[0,1]}(x).$$

The Fisher information matrix is

$$I(\alpha,\ \beta) = \begin{pmatrix} PG(1,\alpha) - PG(1,\alpha+\beta) & -PG(1,\alpha+\beta) \\ -PG(1,\alpha+\beta) & PG(1,\beta) - PG(1,\alpha+\beta) \end{pmatrix},$$

where $PG(1, x) = \sum_{i=0}^{\infty}(x+i)^{-2}$ is the PolyGamma function. The Jeffreys prior is thus the square root of the Fisher information matrix.

6 Binomial

The $B(n,\ p)$, $0 \le p \le 1$, density is

$$f(x|n,\ p) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{(n-x)}.$$

Case 1: Priors for p, given n

Prior	$\pi(p\|n)$	(Marginal) Posterior
Uniform	1	$Be(p\|x+1,\ n-x+1)$
Jeffreys	$\frac{1}{\pi}p^{-1/2}(1-p)^{-1/2}$	$Be(p\|x+\frac{1}{2},\ n-x+\frac{1}{2})$
Reference		
MDIP	$1.6186p^p(1-p)^{(1-p)}$	proper
Novick and Hall's[1]	$p^{-1}(1-p)^{-1}$	$Be(p\|x,\ n-x)$

1. See Novick and Hall (1965). Note this prior is uniform in $\theta = \log p/(1-p)$.

Case 2: Priors for n

Prior	$\pi(n)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$n^{-1/2}$	
Reference[1]	1/n	
Universal[2]	$\frac{1}{2.865n\log(n)\log\log(n)\cdots\log\log\ldots\log(n)}$	

1. Discussed by Alba and Mendoza (1995).

2. Last term for which $\log\log\ldots\log(n) > 1$. See Rissanen (1983).

7 Bivariate Binomial

$$f(r,s|p,q,m) = \begin{pmatrix} m \\ r \end{pmatrix} p^r(1-p)^{m-r} \begin{pmatrix} r \\ s \end{pmatrix} q^s(1-q)^{r-s},$$

for $r = 1,\ldots,m$ and $s = 1,\ldots,r$. Polson and Wasserman (1990) compute the Fisher information matrix of this distribution as $I(p,\ q) = m\mathrm{diag}(\{\mathrm{p(1-p)}\}^{-1},\ \mathrm{p}\{\mathrm{q(1-q)}\}^{-1})$.

Prior	$\pi(p,q\|m)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\frac{1}{2\pi}(1-p)^{-/2}q^{-1/2}(1-q)^{-1/2}$	proper
Reference[1]	$\frac{1}{\pi^2}p^{-1/2}(1-p)^{-1/2}q^{-1/2}(1-q)^{-1/2}$	
Reference[2]	$\frac{1}{\pi^2}(1-p)^{-1/2}q^{-1/2}(1-q)^{-1/2}(1-pq)^{-1/2}$	
Crowder and Sweeting's[3]	$p^{-1}(1-p)^{-1}q^{-1}(1-q)^{-1}$	

1. Parameter of interest is p or q.

2. Parameter of interest is $\theta = pq$ or $\phi = p(1-q)/(1-pq)$.
3. See Crowder and Sweeting (1989).

8 Box-Cox Power Transformed Linear Model

Given observations $\{y_1, \ldots, y_n\}$, the model is

$$z^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln y_i, & \lambda = 0 \end{cases} = \mu_\lambda + x_i^t \beta_\lambda + \epsilon_i \sigma_\lambda,$$

where β_λ is a $(k \times 1)$ vector of regression coefficients, x_i is a vector of covariates, and $\epsilon_i \sim N(0,\ 1)$ truncated at $-(\mu_\lambda + x_i^t \beta_\lambda)/\sigma_\lambda$.

Jeffreys prior was obtained by Pericchi (1981),

$$\pi(\mu_\lambda,\ \beta_\lambda,\ \sigma_\lambda,\ \lambda) \propto \frac{p(\lambda)}{\sigma_\lambda^{k+1}},$$

where $p(\lambda)$ is some unspecified prior for λ.

Box and Cox (1964) proposed the prior

$$\pi(\mu_\lambda,\ \beta_\lambda,\ \sigma_\lambda,\ \lambda) \propto g^{-(k+1)(\lambda-1)} \sigma_\lambda^{-1},$$

where g is the geometric mean of the $y's$.

Based on the so called data-translated parameterization,

$$z^{(\lambda)} = \begin{cases} \frac{\theta^\lambda - 1}{\lambda} + x_i^t \beta_\lambda + \epsilon_i \sigma_\lambda, & \lambda \neq 0 \\ \ln \theta + x_i^t \beta_\lambda + \epsilon_i \sigma_\lambda, & \lambda = 0, \end{cases}$$

corresponding to $\mu_\lambda = (\theta^\lambda - 1)/\lambda$ or $\ln \theta$, Wixley (1993) proposed the following two priors,

$$\pi(\mu_\lambda,\ \beta_\lambda,\ \sigma_\lambda,\ \lambda) \propto g^{-(k+1)(\lambda-1)} \sigma_\lambda^{-1} p(\lambda),$$

where g is the geometric mean of the $y's$ and $p(\lambda)$ is some unspecified prior for λ. This prior is apart from the prior of Box and Cox (1964) only by a factor $p(\lambda)$. It is also the prior used in

Box and Tiao (1992).

$$\begin{aligned}\pi(\mu_\lambda,\ \beta_\lambda,\ \sigma_\lambda,\ \lambda) &\propto [1+\lambda\mu_\lambda]^{-(k+1)(1-1/\lambda)}\sigma_\lambda^{-1}p(\lambda)\\ &\propto \theta^{-(k+1)(\lambda-1)}\sigma_\lambda^{-1}p(\lambda),\end{aligned}$$

where g is the geometric mean of the $y's$ and $p(\lambda)$ is some unspecified prior for λ. This prior will give a very close resultant posterior distribution to that of Box and Cox (1964).

9 Cauchy

The $C(\mu,\ \sigma)$, $-\infty < \alpha < \infty$, $\sigma > 0$, density is

$$f(x|\mu,\ \sigma) = \frac{\sigma}{\pi[\sigma^2 + (x-\mu)^2]}.$$

This is a location-scale parameter problem; see that section for priors. Posterior analysis can be found in Spiegelhalter (1985) and Howlader and Weiss (1988).

10 Dirichlet

The $D(\alpha)$ density, where $\sum_{i=1}^k x_i = 1$, $0 \le x_i \le 1$, and $\alpha = (\alpha_1, \ldots, \alpha_k)^t$, $\alpha_i > 0$ for all i, is given by

$$f(x|\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{(\alpha_i - 1)},$$

where $\alpha_0 = \sum_{i=1}^k \alpha_i$,

The Fisher information matrix is

$$I(\alpha_1, \ldots, \alpha_k) = \begin{pmatrix} PG(1,\alpha_1) - PG(1,\alpha_0) & & -PG(1,\alpha_0) \\ & \ddots & \\ -PG(1,\alpha_0) & & PG(1,\alpha_k) - PG(1,\alpha_0) \end{pmatrix},$$

where $PG(1,x) = \sum_{i=0}^\infty (x+i)^{-2}$ is the PolyGamma function. The Jeffreys prior is $|I(\alpha_1, \ldots, \alpha_k)|^{1/2}$.

11 Exponential Regression Model

See Ye and Berger (1991).

$$Y_{ij} \sim N(\alpha + \beta\rho^{x+x_i a},\ \sigma^2),$$

where $\alpha,\ \beta \in R$, $0 < \rho < 1$, $x \geq 0,\ a > 0$, x and a known constants, $0 \leq i \leq k-1,\ 1 \leq j \leq m$, the x_i's are known nonnegative regressors with $x_i \neq x_j$ for $i \neq j$ and the variance $\sigma^2 > 0$ is an unknown constant. It is assumed that $x_i < x_j$ for $i < j$.

Prior	$\pi(\rho, \alpha, \beta, \sigma)$	(Marginal) Posterior
Uniform	1	
adhoc	$1/\sigma$	improper
Jeffreys	$\lvert\beta\rvert\rho^{2x-1}p_1(\rho^a)/\sigma^4$	proper; 2-dimensional numerical integration
Reference[1]	$\rho^{x-1}p(\rho^a)/\sigma$	proper; 1-dimensional numerical integration
Reference[2]	$\rho^{x-1}p(\rho^a)/\sigma^2$	
Reference[3]	$\rho^{x-1}p(\rho^a)/\sigma^3$	

where $p(\rho^a) = p_1(\rho^a)/p_2(\rho^a)$, and

$$\begin{aligned}
p_1(w) &= (\sum_{i=0}^{k-1} w^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1} w^{x_i})^2)(\sum_{i=0}^{k-1} x_i^2 w^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1} x_i w^{x_i})^2) \\
&\quad -(\sum_{i=0}^{k-1} x_i w^{2x_i} - \frac{1}{k}\sum_{i=0}^{k-1} w^{x_i} \sum_{i=0}^{k-1} x_i w^{x_i})^2, \\
p_2(w) &= \sum_{i=0}^{k-1} w^{2x_i} - \frac{1}{k}(\sum_{i=0}^{k-1} w^{x_i})^2.
\end{aligned}$$

1. Group ordering $\{\rho,\ \alpha,\ \beta,\ \sigma\}$ or $\{\rho,\ (\alpha,\ \beta),\ \sigma\}$ or with all permutations of $\alpha,\ \beta,\ \sigma$; this is recommended for typical use, and appears to be approximately frequentist matching.

2. Group ordering $\{\rho,\ \alpha,\ (\beta,\ \sigma)\}$ and with all permutations of $\alpha,\ \beta,\ \sigma$.

3. Group ordering $\{\rho,\ (\alpha,\ \beta,\ \sigma)\}$ and with all permutations of $\alpha,\ \beta,\ \sigma$.

The marginal posterior of ρ for the prior $\sigma^{-s}\rho^{x-1}p(\rho^a)$ is given by

$$\pi(\rho|y) \sim \frac{p(\rho^a)}{\rho^{1+ax_0}(1-\rho^a)h(\rho^a;\ s,\ y)},$$

where

$$h(\rho;\ s,\ y) = \{[s_{yy}^2 - md_k^2(\rho;\ y)/p_2^2(\rho)]^{km+s-3}\frac{p_2^2(\rho)}{\rho^{2x_0}(1-\rho)^2}\}^{1/2},$$

with

$$d_k(\rho;\ y) = \sum_{i=0}^{k-1}(\bar{y}_{i.} - \bar{y})\rho^{x_i}.$$

12 F Distribution

The $F(\alpha,\ \beta)$, $\alpha > 0$, $\beta > 0$, density is

$$f(x|\alpha,\ \beta) = \frac{\Gamma[(\alpha+\beta)/2]\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \cdot \frac{x^{\alpha/2-1}}{(\beta+\alpha x)^{(\alpha+\beta)/2}} I_{(0,\infty)}(x).$$

13 Gamma

The $G(\alpha,\ \beta)$, $\alpha > 0$, $\beta > 0$, density is

$$f(x|\alpha,\ \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I_{(0,\infty)}(x).$$

If α is known, β is a scale parameter. The Jeffreys, reference, and MDIP priors are all $1/\beta$, and the posterior is therefore $IG(n\alpha,\ 1/\sum_{i=1}^n x_i)$.

The Fisher information matrix is

$$I(\alpha,\beta) = \begin{pmatrix} PG(1,\alpha) & 1/\beta \\ 1/\beta & \alpha/\beta^2 \end{pmatrix},$$

where $PG(1,x) = \sum_{i=0}^\infty (x+i)^{-2}$ is the PolyGamma function.

Prior	$\pi(\alpha,\beta)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\sqrt{\alpha PG(1,\alpha)-1}/\beta$	
Reference[1]	$\sqrt{(\alpha PG(1,\alpha)-1)/\alpha}/\beta$	proper[3]
Reference[2]	$\sqrt{PG(1,\alpha)}/\beta$	

1. Group ordering $\{\alpha,\ \beta\}$.

2. Group ordering $\{\beta,\ \alpha\}$.

3. See Liseo (1993) and Sun and Ye (1994b) for marginal posterior.

14 Generalized Linear Model

Let $y_1, \ldots, y_n$ be independent observations, with the exponential density

$$f(y_i|\theta_i,\ \phi) = \exp\{a_i^{-1}(\phi)(y_i\theta_i - b(\theta_i)) + c(y_i,\phi)\},$$

where the $a_i()$, $b()$ and $c()$ are known functions, and $a_i(\phi)$ is of the form $a_i(\phi) = \phi/w_i$, where the w_i's are known weights. The θ_i's are related to regression coefficients, $\beta = (\beta_1, \ldots, \beta_p)^t$, by the link function

$$\theta_i = \theta(\eta_i), \quad i = 1, \ldots, n,$$

where $\eta_i = x_i^t\beta$, and $x_i^t = (x_{i1}, \ldots, x_{ip})$ is a $1 \times p$ vector denoting the ith row of the $n \times p$ matrix of covariates X, and θ is a monotonic differentiable function.

This model includes a large class of regression models, such as normal linear regression, logistic and probit regression, Poisson regression, gamma regression, and some proportional hazards models.

Ibrahim and Laud (1991) studied this model, using the Jeffreys prior. They focus on the case where ϕ is known. The Fisher information matrix they obtained for β is given by $I(\beta) = \phi^{-1}(X^tWV(\beta)\Delta^2(\beta)X)$, where W is an $n \times n$ diagonal matrix with ith diagonal element w_i. $V(\beta)$ and $\Delta(\beta)$ are $n \times n$ diagonal matrices with ith diagonal elements $v_i = v(x_i^t\beta) = d^2b(\theta_i)/d\theta_i^2$

and $\delta_i = \delta(x_i^t\beta) = d\theta_i/d\eta_i$, respectively. Jeffreys prior is thus given by

$$\pi(\beta) \propto |X^t W V(\beta)\Delta^2(\beta)X|^{1/2}.$$

Ibrahim and Laud (1991) show that the Jeffreys prior can lead to an improper posterior for this model, but that the posterior is proper for most GLM's. A sufficient condition for the posterior to be proper is that the integral

$$\int_S \exp\{\phi^{-1}w(yr - b(r))\}(\frac{d^2b(r)}{dr^2})^{1/2}dr$$

be finite, the likelihood function of β be bounded above, and X be of full rank. Here S denotes the range of θ.

In addition, Ibrahim and Laud (1991) give a sufficient and necessary condition for the Jeffreys prior to be proper, namely

$$\int_S (\frac{d^2b(r)}{dr^2})^{1/2}dr < \infty.$$

When ϕ is unknown, the Fisher information matrix is

$$I(\beta,\ \phi) = \begin{pmatrix} I_1(\beta,\ \phi) & 0 \\ 0 & I_2(\beta,\ \phi) \end{pmatrix},$$

where $I_1(\beta,\ \phi)\phi^{-1}(X^t W V(\beta)\Delta^2(\beta)X)$ as above, and

$$I_2(\beta,\ \phi) = -\sum_{i=1}^n \{2w_i\phi^{-3}(\dot{b}(\theta_i)\theta_i - b(\theta_i)) + E(\ddot{c}(y_i,\phi))\};$$

here $\dot{b}(\theta_i) = db(\theta_i)/d\theta_i$, and $\ddot{c} = \partial^2 c(y_i,\phi)/\partial\phi^2$. The Jeffreys prior is then the square root of the determinant of $I(\beta,\ \phi)$.

15 Inverse Gamma

The $IG(\alpha,\ \beta)$, $\alpha > 0$, $\beta > 0$, density is

$$f(x|\alpha,\ \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha x^{(\alpha+1)}}e^{-1/x\beta}I_{(0,\infty)}(x).$$

If α is known, $1/\beta$ is a scale parameter. The Jeffreys, reference, and MDIP prior is $1/\beta$, and the posterior is $IG(n\alpha,\ 1/\sum_{i=1}^n 1/x_i)$.

If α is unknown, the Fisher information matrix is the same as that of the Gamma distribution. Thus the reference prior and the Jeffreys prior are the same as for the Gamma distribution.

16 Inverse Normal or Gaussian

The $IN(\psi,\ \lambda)$, $\psi > 0$, $\lambda > 0$, density is

$$f(x|\psi,\lambda) = (\frac{\lambda}{2\pi})^{1/2} x^{-3/2} \exp\{-\frac{\lambda x}{2}(\psi - \frac{1}{x})^2\} I_{(0,\infty)}(x).$$

The Fisher information matrix is given by $I(\psi,\ \lambda) = \mathrm{diag}(\lambda/\psi,\ 1/2\lambda^2)$.

Prior	$\pi(\psi,\lambda)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$1/\sqrt{\lambda\psi}$	proper[2]
Scale	$1/\lambda$	
Reference[1]	$\psi^{-1/2}\lambda^{-1}$	

1. See Liseo (1993).

2. See Sun and Ye (1994b) for marginal posteriors.

Define $\nu = n-1$, $\xi = (u\bar{x}/\nu)^{-1/2}$, $q = (\xi\bar{x})^{-1}$, $\bar{x} = \sum x_i/n$. Banerjee and Bhattacharyya (1979) show that the marginal posterior of ψ has a left-truncated t-distribution with ν degrees of freedom, location parameter $1/\bar{x}$, and scale parameter q, the point of truncation being zero. i.e.,

$$\pi(\psi|D) \propto (nu/2)^{-n/2}\{1 + \frac{1}{\nu q^2}(\psi - \frac{1}{\bar{x}})^2\}^{-(\nu+1)/2},$$

where $u = \bar{x}_r - 1/\bar{x}$, and $\bar{x}_r = \frac{1}{n}\sum(1/x_i)$. The marginal posterior of λ is the modified gamma distribution

$$\pi(\lambda|D) \propto \frac{(nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\Phi((n\lambda/\bar{x})^{1/2})}{H_\nu(\xi)} \exp(-nu\lambda/2)\lambda^{\nu/2-1},$$

where $\Phi()$ is the standard normal cdf, and $H_\nu()$ denotes the cdf of Student's t-distribution with ν degrees of freedom. The posterior mean and variance are also available; see Banerjee and Bhattacharyya (1979) for more detail.

Another parametrization of inverse Gaussian distribution, the $IN(\mu, \sigma^2)$, $\mu > 0$, $\sigma^2 > 0$, density is

$$f(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} x^{-3/2} \exp\{-(x-\mu)^2/(2\sigma^2\mu^2 x)\} I_{(0,\infty)}(x).$$

The Fisher information matrix is given by $I(\mu, \sigma^2) = \text{diag}(\mu^{-3}\sigma{-}2, (2\sigma^4)^{-1})$.

Prior	$\pi(\mu, \sigma^2)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\mu^{-3/2}\sigma^{-3}$	
Reference[1]	$\mu^{-3/2}\sigma^{-2}$	

1. See Datta and Ghosh (1993).

17 Linear Calibration

Consider the model

$$y_i = \alpha + \beta x_i + \epsilon_{1i}, \quad i = 1, \ldots, n,$$
$$y_{0j} = \alpha + \beta x_0 + \epsilon_{2j}, \quad j = 1, \ldots, k,$$

where β, y_i, and y_{0j} are $(p \times 1)$ vectors, x_i's are known values of the precise measurements, the ϵ_{1i} and ϵ_{2j} are i.i.d. $N_p(0, \sigma^2 I_p)$. The object is to predict x_0.

Denote $\bar{x} = \sum_{i=1}^n x_i/n$, $\bar{y} = \sum_{i=1}^n y_i/n$, $\bar{y_0} = \sum_{j=1}^k y_{0j}/k$, $c_x = \sum_{i=1}^n (x_i - \bar{x})^2$, $\hat{\beta} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})/c_x$, and s_1 and s_2 be the two sums of residual squared errors based on the calibration and predication experiments, respectively. The statistics $\hat{\beta}$, $\bar{y}$, $\bar{y_0}$, and $s = s_1 + s_2$ are minimal sufficient and mutually independent. Kubokawa and Robert (1993) then consider the reduced model

$$\begin{aligned} y &\sim N_p(\beta^*, \sigma^2 I_p), \\ z &\sim N_p(x_0^* \beta^*, \sigma^2 I_p), \\ s &\sim \sigma^2 \chi^2_q, \end{aligned}$$

where $y = c_x^{1/2}\hat{\beta}$, $z = (\bar{y_0} - \bar{y})(n^{-1} + k^{-1})^{-1/2}$, $q = (n + k - 3)p$, $\beta^* = c_x^{1/2}\beta$, and $x_0^* = (x_0 - \bar{x}){c_x}^{-1/2}(n^{-1} + k^{-1})^{-1/2}$.

The Fisher information matrix for the reduced model is given by

$$I(\beta^*,\ \sigma^2,\ x_0^*) = \begin{pmatrix} & 0 & \\ \frac{1+{x_0^*}^2}{\sigma^2}I_p & \vdots & x_0^*\beta^*/\sigma^2 \\ & 0 & \\ 0 \ \cdots \ 0 & \frac{q+2p}{2\sigma^4} & 0 \\ x_0^*{\beta^*}^t/\sigma^2 & 0 & ||\beta^*||^2/\sigma^2 \end{pmatrix}.$$

Prior	$\pi(x_0^*, \beta^*, \sigma^2)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$(1 + {x_0^*}^2)^{(p-1)/2}\|\|\beta^*\|\|(\sigma^2)^{-(p+3)/2}$	
Reference[1]	$(\sigma^2)^{-(p+2)/2}/\sqrt{1 + {x_0^*}^2}$	proper

1. w.r.t. the group ordering $\{x_0^*, (\beta^*, \sigma^2)\}$. See Kubokawa and Robert (1993).

18 Location-Scale Parameter Models

Location Parameter Models:

The $LP(\beta)$, $\beta \in R^p$, density is

$$f(y|\beta) = g(y - X\beta),$$

where X is a $(n \times p)$ constant matrix and g is a n-dimensional density function.

Prior	$\pi(\beta)$	(Marginal) Posterior
Uniform[1]	1	proper if $rank(X^tX) = p$
Reference[2]	$[\beta^t(X^tX)\beta]^{-(p-2)}$	proper if $rank(X^tX) = p > 2$

1. Also Jeffreys and the usual reference prior.

2. Baranchik (1964) "shrinkage" prior; also, the reference prior for $(||\beta||, O)$, where $\beta = O\ (||\beta||, 0, \ldots, 0)^t$. This also arises from admissibility considerations; see Berger and Strawder-

man (1993).

Location-Scale Parameter Models:

The $LSP(\beta, \sigma)$, $\beta \in R^p$, $\sigma > 0$ density is

$$f(y|\beta, \sigma) = \frac{1}{\sigma^n} g(\frac{1}{\sigma}(y - X\beta)),$$

where X is a $(n \times p)$ constant matrix and g is a n-dimensional density function.

Prior	$\pi(\beta, \sigma)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$1/\sigma^{(p+1)}$	
Reference[1]	$1/\sigma$	
Reference[2]	$\frac{1}{\sigma}[\beta^t(X^tX)\beta]^{-(p-2)}$	
Reference[3]	$\pi(\phi, \sigma) \propto \sigma^{-1}(c_1\phi^2 + c_2)$	

1. Reference with respect to the group ordering $\{\beta, \sigma\}$; also the prior actually recommended by Jeffreys (1961).

2. Baranchik (1964) "shrinkage" prior; also, the reference prior with respect to the parameters $\{||\beta||, O, \sigma\}$ (see Location Parameter Models).

3. $\phi = \beta/\sigma$ is parameter of interest, selecting rectangular compacts on ϕ and nuisance parameter $c_1\beta^2 + c_2\sigma^2$. See Datta and Ghosh (1993).

Scale Parameter Models:

The $SP(\sigma_1, \ldots, \sigma_p)$, $\sigma_i > 0$ for $\forall i$, density is

$$f(y|\sigma_1, \ldots, \sigma_p) = (\prod_{i=1}^{p} \sigma_i^{-n_i}) g(\frac{1}{\sigma_1}\underset{\sim}{y}_1, \frac{1}{\sigma_2}\underset{\sim}{y}_2, \ldots, \frac{1}{\sigma_p}\underset{\sim}{y}_p),$$

where $\underset{\sim}{y}_i$ are $(n_i \times 1)$ vectors, $i = 1, \ldots, p$, g is a $n_1 + \cdots + n_p$-dimensional density function.

Prior	$\pi(\sigma_1, \ldots, \sigma_p)$	(Marginal) Posterior
Uniform	1	
Jeffreys		
Reference	$\prod_{i=1}^{p} \sigma_i^{-1}$	
MDIP		

19 Logit Model

Poirier (1992) analyzes the Jeffreys prior for the conditional logit model as follows. An experiment consists of N trials. On trial n exactly one of J_n discrete alternatives is observed. Let y_{nj} be a binary variable which equals unity iff alternative j is observed on trial n; otherwise y_{nj} equals zero. Let $z_n = [z_{n1}^t, \ldots, z_{nJ_n}^t]^t$, with each z_{nj} being a $K \times 1$ vector. γ is a $K \times 1$ vector of unknown parameters. The probability of alternative j on trial n is specified to be

$$p_{nj}(\gamma) = Prob(y_{nj} = 1 | z_n,\ \gamma) = \frac{\exp(z_{nj}^t \gamma)}{\sum_{i=1}^{J_n} \exp(z_{ni}^t \gamma)}.$$

When $J_n \equiv J$ and $z_{nj} = e_j \otimes x_n$, where e_j is a $(J-1) \times 1$ vector with all elements equal to zero except the jth which equals unity and x_n is a $M \times 1$ vector of observable characteristics of trial n, this is the multinomial logit model.

Poirier (1992) gives the Jeffreys prior for this model as

$$\pi(\gamma) \propto \left| \sum_{n=1}^{N} \sum_{j=1}^{J_n} p_{nj}(\gamma) [z_{nj} - \tilde{z}_n(\gamma)][z_{nj} - \tilde{z}_n(\gamma)]^t \right|^{1/2},$$

where $\tilde{z}_n(\gamma) = \sum_{i=1}^{J_n} p_{ni}(\gamma) z_{ni}$. The following special cases are also given therein.

Multinomial Logit Without Covariates:

Suppose $K = J - 1$ and $z_{nj} = e_j$; then the Jeffreys prior reduces to the proper density

$$\pi(\gamma) = [\frac{\Gamma(J/2)}{\pi^{J/2}}] \prod_{j=1}^{J} [p_{*j}(\gamma)]^{1/2},$$

where $p_{*j} = p_{nj}$ for any n. The density for $p_* = (p_{*1}, \ldots, p_{*J-1})$ is the Dirichlet density

$$\pi(p_*) = \frac{\Gamma(J/2)}{\pi^{J/2}} \prod_{j=1}^{J} p_{*j}^{-1/2}.$$

In the binomial case,

$$\pi(\gamma) = \pi^{-1}\{p_{*1}(\gamma)[1 - p_{*1}(\gamma)]\}^{1/2} = \frac{\exp(\gamma/2)}{\pi[1+\exp(\gamma)]},$$

and

$$\pi(p_*) = \pi^{-1}[p_{*1}(1-p_{*1})]^{-1/2}.$$

The other competing noninformative prior are: the uniform prior on γ; the uniform prior on p_{*1}; and the MDIP prior for p_{*1}. See Geisser (1984) for discussion.

Logistic Regression:

Suppose $y_i|\beta \sim$ i.i.d. Bernoulli (π_i), with

$$\pi_i = \frac{e^{x_i^t\beta}}{1+e^{x_i^t\beta}}. \quad i = 1, \ldots, n.$$

Then the likelihood function is

$$f(y|\beta) = e^{\sum_{i=1}^n y_i x_i^t \beta} / \prod_{i=1}^{n}(1+e^{x_i^t\beta}),$$

and the Fisher Information matrix is

$$I(\beta) = (x_1, \ldots, x_n) \begin{pmatrix} \pi_1(1-\pi_1) & & \\ & \ddots & \\ & & \pi_n(1-\pi_n) \end{pmatrix} \begin{pmatrix} x_1^t \\ \vdots \\ x_n^t \end{pmatrix}.$$

The Jeffreys prior is the square root of the determinant of above matrix.

For more special cases of the logit model, see Poirier (1992).

20 Mixed Model

Consider mixed model

$$y_{ijk} = B_{ijk}\alpha_b + W_{ijk}\alpha_w + T_i + C_j + \epsilon_{ijk}, \quad i = 1, ..., I, \ j = 1, ..., J, \ k = 1, ..., t_{ij},$$

assuming $\epsilon_{ijk} \sim N(0, \sigma^2)$, and independently, $C_j \sim N(0, \ \sigma_c^2)$, with α_b, α_w, T_i, σ^2, and σ_c^2 the parameters of interest.

Box and Tiao (1992) used the following prior in the balanced mixed model case, with $t_{ij} = n, \forall ij$,

$$\pi(\sigma_c^2, \sigma^2) \propto \frac{1}{\sigma^2(\sigma^2 + n\sigma_c^2)}.$$

Chaloner (1987) used three priors for this model, $(\sigma^2 + \sigma_c^2)^{-1}$, $\sigma^{-2}(\sigma^2 + \sigma_c^2)^{-1}$, and $\sigma^{-2}(\sigma^2 + \sigma_c^2)^{-3/2}$.

Yang and Pyne (1996) derived the Jeffreys prior,

$$\pi(\sigma_c^2, \sigma^2) \quad \propto \quad \left[\mathrm{Det} \left| \sum_{j=1}^{J} \begin{pmatrix} \frac{t_j^2}{2(t_j\sigma_c^2 + \sigma^2)^2} & \frac{t_j}{2(t_j\sigma_c^2 + \sigma^2)^2} \\ \frac{t_j}{2(t_j\sigma_c^2 + \sigma^2)^2} & \frac{t_j - 1}{2\sigma^4} + \frac{1}{2(t_j\sigma_c^2 + \sigma^2)^2} \end{pmatrix} \right| \right]^{1/2},$$

where $t_j = \sum_{i=1}^{I} t_{ij}$, and reference prior,

$$\pi(\sigma_c^2, \sigma^2) \propto \frac{1}{\sigma^2} \sqrt{\sum_{j=1}^{J} \frac{t_j^2}{(t_j\sigma_c^2 + \sigma^2)^2}}.$$

21 Mixture Model

For arbitrary density functions $p_1(x)$ and $p_2(x)$, consider the model

$$p(x|\lambda) = \lambda p_1(x) + (1 - \lambda)p_2(x).$$

Bernardo and Girón (1988) discussed the reference prior for this model. They found that the reference prior is always proper.

22 Multinomial

The $M(n,\ p)$ density, where $\sum_{i=1}^{k+1} x_i = n$ and each x_i is an integer between 0 and n, $p = (p_1, \ldots, p_{k+1})^t$, with $\sum_{i=1}^{k+1} p_i = 1$ and $0 \le p_i \le 1$ for all i, is given by

$$f(x|p) = \frac{n!}{\prod_{i=1}^{k+1}(x_i!)} \prod_{i=1}^{k+1} p_i^{x_i}.$$

Prior	$\pi(p)$	(Marginal) Posterior
Uniform	1	all are proper
Jeffreys	$C_k^{-1}(\prod_{i=1}^k p_i^{-1/2})(1-\delta_k)^{-1/2}$	
Reference[1]		
Reference[2]	$(\pi^{-k})\prod_{i=1}^k[p_i^{-1/2}(1-\delta_i)^{-1/2}]$	
Reference[3]	$(\prod_{i=1}^m C_{n_i}^{-1})(\prod_{i=1}^k p_i^{-1/2})(\prod_{i=1}^{m-1}(1-\delta_{N_i})^{-n_{i+1}/2})(1-\delta_{N_m})^{-1/2}$	
MDIP	$p_1^{p_1} p_2^{p_2}\cdots p_k^{p_k}(1-\sum_{i=1}^k p_i)^{1-\sum_{i=1}^k p_i}$	
Novick and Hall's[4]	$\prod_{i=1}^k p_i^{-1}$	

Here $\delta_j = \sum_{i=1}^j p_i$, $C_{2l-1} = \pi^l/(l-1)!$, and $C_{2l} = (2\pi)^l/[(2l-1)(2l-3)\cdots(1)]$, for all positive integers l. See Berger and Bernardo (1992b). See also Zellner (1993).

1. One-group reference prior.

2. k-group reference prior.

3. m-group reference prior. The posterior for the m-group reference prior, which includes the posteriors for Reference[1] and Reference[2] as special cases, is

$$\pi(p|D) \propto (\prod_{i=1}^k p_i^{x_i-\frac{1}{2}})(\prod_{i=1}^{m-1}(1-\delta_{N_i})^{-n_{i+1}/2})(1-\delta_{N_m})^{n-r-\frac{1}{2}}.$$

4. See Novick and Hall (1965).

Sono (1983) also derived a noninformative prior for this model, using the assumption of prior independence of transformed parameters and an approximate data-translated likelihood function.

23 Negative Binomial

The $NB(\alpha,\ p)$, $\alpha > 0$, $0 < p \le 1$, density is

$$f(x|\alpha,\ p) = \frac{\Gamma(\alpha+x)}{\Gamma(\alpha+1)\Gamma(\alpha)} p^{\alpha}(1-p)^{x}.$$

α is given:

Prior	$\pi(p)$	(Marginal) Posterior
Uniform	1	$Be(p\|\alpha+1,\ x-\alpha+1)$
Jeffreys	$1/[p\sqrt{1-p}]$	$Be(p\|\alpha,\ x-\alpha+1/2)$
Reference		

24 Neyman and Scott Example

This model consists of $2n$ independent observations,

$$X_{ij} \sim N(\mu_i,\ \sigma^2), \quad i = 1, \ldots, n,\ j = 1, 2.$$

Prior	$\pi(\sigma, \mu_1, \ldots, \mu_n)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\sigma^{-(n+1)}$	proper
Reference[1]	σ^{-1}	
Reference[2]	σ^{-n}	

1. Group ordering $\{\sigma,\ (\mu_1, \ldots, \mu_n)\}$ or $\{\sigma,\ \mu_1, \ldots,\ \mu_n\}$. Yields a sensible posterior. The posterior mean of σ^2 is $S^2/(n-2)$, with $S^2 = \sum_{i=1}^{n}\sum_{j=1}^{2}(X_{ij} - \bar{X}_i)^2$, $\bar{X}_i = (X_{i1} + X_{i2})/2$. See Berger and Bernardo (1992c) for discussion.

2. Group ordering $\{\mu_1,\ (\mu_2, \ldots, \mu_n, \sigma)\}$. Strong inconsistency occurs for this prior, along with the Jeffreys prior. The posterior mean of σ^2 for the Jeffreys prior is $S^2/(2n-2)$.

25 Nonlinear Regression Model

Eaves (1983) considered the following nonlinear regression model

$$y = g(\theta) + \sigma e,$$

where the $n \times 1$ vector-valued design function g of a d-dimensional vector parameter θ is no longer assumed linear as compared with linear regression, although it remains 1-1 and smooth. The noninformative prior proposed therein is

$$\pi(\theta, \sigma) \propto |I(\theta)|^{1/2}/\sigma,$$

where

$$\begin{aligned} I(\theta) &= \frac{1}{2} E(D_\theta ||y - g(\theta)||^2 | \theta) \\ &= d_\theta g(\theta)' d_\theta g(\theta) \end{aligned}$$

is the regression information matrix. Here D_θ is the $d \times d$ matrix-valued second-order partial operator: $d_\theta g(\theta)$ is the $n \times d$ Jacobian of g. Note this prior is also the reference prior provided that θ and σ are in separate group.

26 Normal

Univariate Normal:

The $N(\mu,\ \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$, density is

$$f(x|\mu,\ \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(x-\mu)^2/2\sigma^2}.$$

σ^2 Known:

Prior	$\pi(\mu)$	(Marginal) Posterior
All	1	$\pi(\mu\|D) = N(\bar{x},\ \sigma^2/n)$

Here $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

μ Known:

Prior	$\pi(\sigma^2)$	(Marginal) Posterior
Uniform	1	$\pi(\sigma^2 \mid D) \sim IG((n-2)/2,\ 2/S^2)$
Jeffreys		
Reference	$1/\sigma^2$	$\pi(\sigma^2 \mid D) \sim IG(n/2,\ 2/S^2)$
MDIP		

Here $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and $S^2 = \sum_{i=1}^n (x_i - \bar{x})^2$.

μ and σ^2 Both Unknown:

Prior	$\pi(\mu, \sigma^2)$	(Marginal) Posterior
Uniform	1	$\pi(\mu \mid D) \sim T(n-3,\ \bar{x},\ S^2/n(n-3))$ $\pi(\sigma^2 \mid D) \sim IG((n-3)/2,\ 2/S^2)$
Jeffreys	$1/\sigma^4$	$\pi(\mu \mid D) \sim T(n+1,\ \bar{x},\ S^2/n(n+1))$ $\pi(\sigma^2 \mid D) \sim IG((n+1)/2,\ 2/S^2)$
Reference[1]	$\pi(\phi,\ \sigma) \propto (2+\phi^2)^{-1/2}\sigma^{-1}$	
Reference[2]	$1/\sigma^2$	$\pi(\mu \mid D) \sim T(n-1,\ \bar{x},\ S^2/n(n-1))$
MDIP		$\pi(\sigma^2 \mid D) \sim IG((n-1)/2,\ 2/S^2)$

Here $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and $S^2 = \sum_{i=1}^n (x_i - \bar{x})^2$.

1. $\phi = \mu/\sigma$ is parameter of interest, parameter ordering $\{\phi, \sigma\}$. See Bernardo and Smith (1994).

2. If μ and σ^2 are in separate groups.

p-Variate Normal:

The $N_p(\mu,\ \Sigma)$ density, where $\mu = (\mu_1, \ldots, \mu_p) \in R^p$ and Σ is a positive definite matrix, is given by

$$f(x \mid \mu,\ \Sigma) = \frac{1}{(2\pi)^{p/2}(\det\Sigma)^{1/2}} e^{-(x-\mu)^t \Sigma^{-1}(x-\mu)/2}.$$

Σ Known:

Prior	$\pi(\mu)$	(Marginal) Posterior
Uniform		
Jeffreys	1	$\pi(\mu\|D) \sim N_p(\bar{x},\ \Sigma/n)$
Reference		
Shrinkage[1]	$(\mu^t\Sigma^{-1}\mu)^{-(p-2)}$	

Here $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

1. See Baranchik (1964) and Berger and Strawderman (1993).

μ **Known:**

Prior	$\pi(\Sigma)$	(Marginal) Posterior
Uniform	1	$\pi(\Sigma^{-1}\|D) \sim W_p(n-p-1,\ S^{-1}/n)$
Jeffreys	$1/\|\Sigma\|^{(p+1)/2}$	$\pi(\Sigma^{-1}\|D) \sim W_p(n,\ S^{-1}/n)$
Reference[1]	$1/\|\Sigma\|\prod_{i<j}(d_i - d_j)$	proper
Reference[2]	$1/\|\Sigma\|(\log d_1 - \log d_p)^{(p-2)}\prod_{i<j}(d_i - d_j)$	
Reference[3]	$1/\bar{R}(1-\bar{R}^2)^*$	$\pi(\bar{R}\|D) \sim (\bar{R}^2)^{-1/2}(1-\bar{R}^2)^{n/2-1}\ {}_2F_1(\frac{n}{2}, \frac{n}{2}; \frac{p-1}{2}; (R\bar{R})^2)/\ {}_3F_2(\frac{n}{2}, \frac{n}{2}, \frac{1}{2}; \frac{p-1}{2}; \frac{n+1}{2}; R^2)^*$
MDIP	$1/\|\Sigma\|$	

Here $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)((x_i - \mu)^t$, $d_1 < d_2 < \cdots < d_p$ are the eigenvalues of Σ, and $\bar{R}$ and R are population and sample multiple correlation coefficients, respectively. If we write

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{(1)}^t \\ \sigma_{(1)} & \Sigma_{22} \end{pmatrix}, \quad \hat{\Sigma} = S = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{(1)}^t \\ \hat{\sigma}_{(1)} & \hat{\Sigma}_{22} \end{pmatrix},$$

then, $\bar{R} = \sqrt{\sigma_{(1)}^t\Sigma_{22}^{-1}\sigma_{(1)}/\sigma_{11}}$, and $R = \sqrt{\hat{\sigma}_{(1)}^t\hat{\Sigma}_{22}^{-1}\hat{\sigma}_{(1)}/\hat{\sigma}_{11}}$.

1. Group ordering lists the ordered eigenvalues of Σ first, and is recommended for typical use. See Yang and Berger (1992).

2. Group ordering lists the eigenvalues of Σ first, with $\{d_1, d_p\}$ proceeding the other ordered eigenvalues. See Yang and Berger (1992).

3. Population multiple correlation coefficient is parameter of interest, and uses sample multiple correlation coefficient as data. Prior and posterior are w.r.t. $d\bar{R}$. ${}_2F_1$ and ${}_3F_2$ are the

hypergeometric function. See Tiwari, Chib and Jammalamadaka (1989). Also see Muirhead (1982).

μ and Σ Both Unknown:

Prior	$\pi(\mu,\sigma)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$1/\lvert\Sigma\rvert^{(p+2)/2}$	
Reference[1]	$1/\lvert\Sigma\rvert \prod_{i<j}(d_i - d_j)$	
Reference[2]	$1/\lvert\Sigma\rvert(\log d_1 - \log d_p)^{(p-2)} \prod_{i<j}(d_i - d_j)$	
MDIP	$1/\lvert\Sigma\rvert$	

Here $S = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)((x_i - \mu)^t$, and $d_1 < d_2 < \cdots < d_p$ are the eigenvalues of Σ.

1. Group ordering lists the ordered eigenvalues of Σ first. μ and Σ are in separate groups. It is recommended for typical use. See Yang and Berger (1992).

2. Group ordering lists the eigenvalues of Σ first, with $\{d_1, d_p\}$ proceeding the other ordered eigenvalues. μ and Σ are in separate groups. See Yang and Berger (1992).

Bivariate Normal:

The $N_2(\mu,\ \Sigma)$ density, where, $\mu = (\mu_1, \mu_2) \in R^2$ and $\Sigma = (\sigma_{ij})$ is a 2×2 positive definite matrix, is given by

$$f(x|\mu,\ \Sigma) = \frac{1}{(2\pi)(\det \Sigma)^{1/2}} e^{-(x-\mu)^t \Sigma^{-1}(x-\mu)/2}.$$

μ and Σ Both Unknown:

Prior	$\pi(\mu,\Sigma)$	(Marginal) Posterior
Reference[1]	$(1-\rho^2)^{-1}(\sigma_{11}\sigma_{22})^{-1/2}$	$\frac{(1-\rho^2)^{(n-3)/2}}{(1-\rho r)^{n-3/2}} F(\frac{1}{2}, \frac{1}{2}, n-\frac{1}{2}, \frac{1+\rho r}{2})$
Reference[2]	$\sigma_{11}^{1/2}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2$	
Reference[3]	$\sqrt{\sigma_{11}\sigma_{22} + \sigma_{12}^2}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^2$	
MDIP[4]	$1/\sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}$	

1. The correlation coefficient, ρ, is parameter of interest. Parameters ordered as $\{\rho, \mu_1, \mu_2, \sigma_{11}, \sigma_{22}\}$. r is the sample correlation coefficient. Prior and posterior are w.r.t. $d\rho d\mu_1 d\mu_2 d\sigma_{11} d\sigma_{22}$. F is the hypergeometric function. See Bayarri (1981).

2. σ_{11} is parameter of interest. Parameters ordered as $\{\sigma_{11}, (\sigma_{12}, \sigma_{22}, \mu_1, \mu_2)\}$. Limiting sequence of compact sets is $\{\sigma_{12}^2/\sigma_{22} \in (\sigma_{11}1^{-1}, \sigma_{11}(1-l^{-1}),\ \sigma_{22} \in (l^{-1}, l)\}$.

3. σ_{12} is parameter of interest. Parameters ordered as $\{\sigma_{12}, (\sigma_{11}, \sigma_{22}, \mu_1, \mu_2)\}$. Limiting sequence of compact sets is $\{\sigma_{11}\sigma_{22} \in (\sigma_{12}^2(1+1/l), \sigma_{12}^2 l),\ \sigma_{11} \in (l^{-1}, l)\}$.

4. Prior and posterior are w.r.t. $d\rho d\mu_1 d\mu_2 d\sigma_{11} d\sigma_{22}$.

27 Pareto

The $Pa(x_0,\ \alpha)$ density, where, $0 < x_0 < \infty$, $\alpha > 0$, is given by

$$f(x|x_0,\ \alpha) = \frac{\alpha}{x_0}(\frac{x_0}{x})^{\alpha+1} I_{(x_0,\infty)}(x).$$

If x_0 is known, this is a scale density, and the Jeffreys prior and reference prior is $1/\alpha$.

28 Poisson

The $P(\lambda)$, $\lambda > 0$, density is

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Prior	$\pi(\lambda)$	(Marginal) Posterior
Uniform	1	$G(\sum_{i=1}^n x_i + 1,\ 1/n)$
Jeffreys	$\lambda^{-1/2}$	$G(\sum_{i=1}^n x_i + 1/2,\ 1/n)$
Reference		

Possion Process:

For a Poisson process $X_1,\ X_2, \ldots$ with unknown parameter λ, Jeffreys (1961), Novick and Hall (1965), and Villegas (1977) proposed ignorance prior $\pi(\lambda) = \lambda^{-1}$, also called a logarithmic uniform prior because it implies a uniform distribution for $\log \lambda$.

29 Product of Normal Means

Consider the $N_p(\mu, \Sigma)$ density, where $\mu = (\mu_1, \ldots, \mu_p)$, and the parameter of interest is $\prod_{i=1}^p \mu_i$.

p=2, $\Sigma = I_2$:

Prior	$\pi(\mu_1, \mu_2)$	(Marginal) Posterior
Uniform	1	proper
Reference[1]	$(\mu_1^2 + \mu_2^2)^{1/2}$	

1. See Berger and Bernardo (1989).

p=n, $\Sigma = I_n$, $n > 2$:

Prior	$\pi(\mu_1, \ldots, \mu_n)$	(Marginal) Posterior
Uniform	1	proper
Reference[1]	$\prod_{i=1}^n \mu_i \sqrt{\sum_{i=1}^n \mu_i^{-2}}$	

1. See Sun and Ye (1995).

p=n, $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and $\mu_i > 0$ for $i = 1, \ldots, n$:

Prior	$\pi(\mu_1, \ldots, \mu_n)$	(Marginal) Posterior
Uniform	1	proper
Jeffreys[1]	$\prod_{i=1}^n \sigma_i^{-2}$	
Reference[2]	$\frac{(\mu_1 \cdots \mu_n)^{2/k-1} g_1(\mu_1, \ldots, \mu_n)}{\prod_{i=1}^n \sigma_i^2} \sqrt{\sum_{i=1}^n \frac{\sigma_i^2}{n_i \mu_i^2}}$	
Reference[3]	$\frac{(\mu_1 \cdots \mu_n)}{\prod_{i=1}^n \sigma_i^2} \sqrt{\sum_{i=1}^n \frac{\sigma_i^2}{n_i \mu_i^2}}$	

1. Proposed by Jeffreys (1961).

2. See Sun and Ye (1994a), where $g_1()$ is any positive real function and n_i is the number of observations from ith population.

3. Also see Sun and Ye (1994a), where n_i is the number of observations from ith population. This prior is also the Tibshirani's matching prior.

30 Random Effects Models

One-Way Model (balanced):

See Berger and Bernardo (1992a).

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, p \text{ and } j = 1, \ldots, n,$$

where the α_i are i.i.d. $N(0,\ \tau^2)$ and, independently, the ϵ_{ij} are i.i.d. $N(0,\ \sigma^2)$. The parameters $(\mu,\ \tau^2,\ \sigma^2)$ are unknown. The reference priors for this model are,

Ordered Grouping	Reference Prior	posterior
$\{(\mu, \sigma^2, \tau^2)\}$	$\sigma^{-2}(n\tau^2 + \sigma^2)^{-3/2}$	
$\{(\mu, \sigma^2), \tau^2\}$	$\sigma^{-5/2}(n\tau^2 + \sigma^2)^{-1}$	
$\{(\mu, \tau^2), \sigma^2\}$	$\tau^{-3C_n/2}\sigma^{-2}\psi(\tau^2/\sigma^2)$	
$\{\sigma^2, (\mu, \tau^2)\}$	$\sigma^{-1}(n\tau^2 + \sigma^2)^{-3/2}$	
$\{\tau^2, (\mu, \sigma^2)\}$	$\tau^{-1}\sigma^{-2}(n\tau^2 + \sigma^2)^{-1/2}\psi(\tau^2/\sigma^2)$	proper
$\{\mu, (\sigma^2, \tau^2)\}$, $\{(\sigma^2, \tau^2), \mu\}$ $\{\mu, \sigma^2, \tau^2\}$ $\{\sigma^2, \mu, \tau^2\}$, $\{\sigma^2, \tau^2, \mu\}$	$\sigma^{-2}(n\tau^2 + \sigma^2)^{-1}$	
$\{\mu, \tau^2, \sigma^2\}$, $\{\tau^2, \mu, \sigma^2\}$, $\{\tau^2, \sigma^2, \mu\}$	$\tau^{-C_n}\sigma^{-2}\psi(\tau^2/\sigma^2)$	

Here $C_n = \{1 - \sqrt{n-1}(\sqrt{n} + \sqrt{n-1})^{-3}\}$, $\psi(\tau^2/\sigma^2) = [(n-1) + (1 + n\tau^2/\sigma^2)^{-2}]^{1/2}$.

The posterior computation involves only one dimensional numerical integration. For details see Berger and Bernardo (1992a).

Suppose $\phi = n\tau^2/\sigma^2$ is the parameter of interest (see, Ye, 1991).

Prior	$\pi(\phi, \mu, \sigma^2)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\sigma^{-3}(1 + \phi)^{-3/2}$	
Reference[1]	$\sigma^{-2}(1 + \phi)^{-3/2}$	proper[4]
Reference[2]	$\sigma^{-2}(1 + \phi)^{-1}$	
Reference[3]	$\sigma^{-3}(1 + \phi)^{-1}$	

1. Group ordering $\{(\phi,\ \mu),\ \sigma^2\}$.

2. Group ordering $\{\phi,\ \mu,\ \sigma^2\}, \{\phi,\ \sigma^2,\ \mu\}, \{(\phi,\ \sigma^2),\ \mu\}$, recommended for typical use.

3. Group ordering $\{\phi,\ (\mu,\ \sigma^2)\}$.

4. The marginal posterior for ϕ, corresponding to the prior $\sigma^{-a}(1+\phi)^{-b}$ is given by

$$\pi(\phi|D) = \frac{(1+\phi)^{q-1}}{W^q B_{p,q}(W/(1+W))}(1+\frac{W}{1+\phi})^{-(p+q)},$$

where $B_{\alpha,\beta}(x) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ is the incomplete Beta function, $p = (p+2b-3)/2$, $q = [p(n-1)+a-2b]/2$, $W = S_2/S_1$, $S_1 = \sum_{i=1}^p \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2$, $\bar{Y}_i = \frac{1}{n}\sum_{j=1}^n Y_{ij}$, $S_2 = n\sum_{j=1}^n (\bar{Y}_i - \bar{Y})^2$, and $\bar{Y} = \frac{1}{pn}\sum_{i=1}^p \sum_{j=1}^n Y_{ij}$.

One-Way Model (unbalanced):

See Ye (1990).

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1,\ldots,k \text{ and } j = 1,\ldots,n_i,$$

where the α_i are i.i.d. $N(0,\ \tau^2)$ and, independently, the ϵ_{ij} are i.i.d. $N(0,\ \sigma^2)$. The parameters $(\mu,\ \tau^2,\ \sigma^2)$ are unknown.

The reference priors for this model are,

Prior	$\pi(\mu, \sigma^2, \tau^2)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\sigma^{-5}[s_{1,1}(\tau^2/\sigma^2)(ns_{2,2}(\tau^2/\sigma^2) - s_{1,1}(\tau^2/\sigma^2)^2)]^{1/2}$	proper
Reference[1]	$\sigma^{-2}\tau^{-2C_{n,\gamma}}\psi(\tau^2/\sigma^2)^{1/2}$	
Reference[2]	$\sigma^{-4}s_{2,2}(\tau^2/\sigma^2)^{1/2}$	

Here the limiting sequence of compact sets for reference prior is chosen to be $\Theta_l = [a_l, b_l] \times [c_l, d_l] \times [e_l, f_l]$, for (μ, σ^2, τ^2), and

$$\begin{aligned} s_{p,q}(x) &= \sum_{i=1}^k \frac{n_i^p}{(1+n_i x)^q}, \text{ for } p, q = 0, 1, 2, \ldots \\ \psi(x) &= n - k + \sum_{i=1}^k \frac{1}{(1+n_i x)^2}, \\ C_{n,\gamma} &= \frac{\sqrt{n-k}}{\sqrt{n}\gamma + \sqrt{n-k}} + \frac{\sqrt{n}(\sqrt{n} - \sqrt{n-k})\gamma^2}{2(\sqrt{n}\gamma + \sqrt{n-k})^2}, \end{aligned}$$

$$\gamma = \lim_{l\to\infty} \frac{\log d_l}{\log c_l^{-1}},$$

1. Group ordering $\{\mu, \tau^2, \sigma^2\}$, $\{\tau^2, \mu, \sigma^2\}$, and $\{\tau^2, \sigma^2, \mu\}$.
2. Group ordering $\{\mu, \sigma^2, \tau^2\}$, $\{\sigma^2, \mu, \tau^2\}$, $\{\sigma^2, \tau^2, \mu\}$, $\{\mu, (\sigma^2, \tau^2)\}$, and $\{(\sigma^2, \tau^2), \mu\}$.

Suppose $v = \tau^2/\sigma^2$ is the parameter of interest.

Prior	$\pi(v, \mu, \sigma^2)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\sigma^{-3}[s_{1,1}(v)(ns_{2,2}(v) - s_{1,1}(v)^2)]^{1/2}$	proper
Reference[1]	$\sigma^{-2}[ns_{2,2}(v) - s_{1,1}(v)^2]^{1/2}$	
Reference[2]	$\sigma^{-2}s_{2,2}(v)^{1/2}$	

1. Group ordering $\{\mu, v, \sigma^2\}$, $\{v, \mu, \sigma^2\}$, and $\{v, \sigma^2, \mu\}$.
2. Group ordering $\{\mu, \sigma^2, v\}$, $\{\sigma^2, \mu, v\}$, $\{\sigma^2, v, \mu\}$, $\{\mu, (\sigma^2, v)\}$, and $\{(\sigma^2, v), \mu\}$.

Random Coefficient Regression Model:

$$\underset{\sim}{y}_i = X_i\underset{\sim}{\beta}_i + \epsilon_i,$$

where $\underset{\sim}{y}_i$ is a $(t_i \times 1)$ vector of observations, X_i is a $(t_i \times p)$ constant design matrix, $\underset{\sim}{\beta}_i$ is a $(p \times 1)$ vector of random coefficients for the ith experimental subject and ϵ_i is a vector of errors for $i = 1, \ldots, n$. Furthermore, we assume that $(\underset{\sim}{\beta}_i, \epsilon_i, i = 1, \ldots, n)$ are independent, and $\underset{\sim}{\beta}_i \sim MVN(\underset{\sim}{\beta}, \Sigma)$ and $\epsilon_i \sim MVN(0, \sigma^2 I_i)$, where $\underset{\sim}{\beta}$ is the $(p \times 1)$ mean of the $\underset{\sim}{\beta}_i$ and I_i is the $(t_i \times t_i)$ identity matrix.

Defining $\Lambda = \Sigma/\sigma^2$, the Jeffreys prior is

$$\pi_J(\underset{\sim}{\beta}, \Lambda, \sigma^2) \propto \left|\sum_{i=1}^n B_i\right|^{\frac{1}{2}} \cdot \left|G[\sum_{i=1}^n (B_i \otimes B_i) - \frac{vec(\sum_{i=1}^n B_i)(vec(\sum_{i=1}^n B_i))^t}{\sum_{i=1}^n t_i}]G^t\right|^{\frac{1}{2}} /\sigma^{p+2}.$$

The reference prior w.r.t the Group Ordering $\{\underset{\sim}{\beta}, \sigma^2, \Lambda\}, \{\sigma^2, \underset{\sim}{\beta}, \Lambda\}$ or $\{\sigma^2, \Lambda, \underset{\sim}{\beta}\}$ is

$$\pi_R(\underset{\sim}{\beta}, \Lambda, \sigma^2) \propto |G\sum_{i=1}^n (B_i \otimes B_i)G^t|^{\frac{1}{2}}/\sigma^2.$$

The reference prior w.r.t the Group Ordering $\{\underset{\sim}{\beta}, \Lambda, \sigma^2\}, \{\Lambda, \underset{\sim}{\beta}, \sigma^2\}$ or $\{\Lambda, \sigma^2, \underset{\sim}{\beta}\}$ is

$$\pi_R(\underset{\sim}{\beta}, \Lambda, \sigma^2) \propto |G[\sum_{i=1}^{n}(B_i \otimes B_i) - \frac{vec(\sum_{i=1}^n B_i)(vec(\sum_{i=1}^n B_i))^t}{\sum_{i=1}^n t_i}]G^t|^{\frac{1}{2}}/\sigma^2,$$

where $B_i \overset{def}{=} X_i^t(X_i \Lambda X_i^t + I_i)^{-1} X_i$, and G denotes a $(p(p+1)/2) \times p^2$ constant matrix $\partial(vecV)/\partial(vecpV)$, where V is a $p \times p$ symmetric matrix.

The posteriors corresponding to the priors above are proper, provided we have at least $2p+1$ full rank design matrices. Computation is discussed in Yang and Chen (1993).

Random Effects Model:

$X = \{x_{ij}; j = 1, \ldots, n_i; i = 1, \ldots, k\}$ arises from k populations $\pi_1, \ldots, \pi_k$, $\pi_i \sim N_p(\mu_i,\ \Sigma)$ and $\mu_i \sim N_p(\xi,\ T)$. Fatti (1982) considers the usual diffuse prior distribution for Σ, ξ and T, and Box and Tiao's noninformative prior distribution. Specifically, Fatti (1982) consider the following type of diffuse joint prior density

$$\pi(\Sigma,\ \xi,\ T) \propto |\Sigma|^{-v_1/2}|T|^{-v_2/2}.$$

This prior has been used by Geisser (1964) and Geisser and Cornfield (1963), for $v_1 = v_2 = p+1$.

Fatti (1982) also studies the predictive density of a new observation x, under the hypothesis $x \in \pi_r$. Fatti found that v_2 must be less than 2 for the predictive density to exist. So, T cannot have the usual diffuse prior distribution with $v_2 = p+1$. If we assign the values $v_1 = p+1$ and $v_2 = 1$,

$$f(x|X,\ \pi_r) \propto |A_3^*|^{-(N-p-1)/2}\ {}_2F_1(p/2, (N-p-1)/2; (k-1)/2; {A_3^*}^{-1}A_1^*), \qquad \text{for } k > p,$$

where $A_1^* = n^* \sum_{i=1}^k (x_{i.}^* - x_{..}^*)(x_{i.}^* - x_{..}^*)^t$, $A_3^* = \sum_{i=1}^k \sum_{j=1}^{n^*}(x_{ij} - x_{..}^*)(x_{ij} - x_{..}^*)^t$, $F(a_1, a_2; b_1; \omega)$ is the hypergeometric function, and

$$n_i^* = \begin{cases} n_i & i \neq r \\ n_r + 1 & i = r, \end{cases}$$

where we assume $n_i^* = n^*,\ \forall i$, $x_{i.}^* = \frac{1}{n^*}\sum_{j=1}^{n^*} x_{ij}$, $x_{..}^* = \frac{1}{k}\sum_{i=1}^k x_{i.}^*$, and $N = \sum_{i=1}^k n_i^*$.

Box and Tiao (1992) propose the following noninformative joint prior,

$$\pi(\Sigma,\ \xi,\ T) \propto |\Sigma|^{-(p+1)/2}|\Sigma + nT|^{-(p+1)/2}.$$

This prior distribution gives the predictive density as,

$$f(x|X,\ \pi_r) \propto |A_3^*|^{-(N-1)/2}\ {}_2F_1((p+1)/2, (N-1)/2; (k+p)/2; A_3^{*\,-1}A_1^*), \qquad \text{for } N > p \text{ and } k > p.$$

31 Ratio of Exponential Means

Let $X_i \overset{ind}{\sim} Exponential(\mu_i)$, $i = 1, 2$. The parameter of interest is $\phi_1 = \mu_2/\mu_1$. With nuisance parameter $\phi_2 = \mu_1\mu_2$, Datta and Ghosh (1993) get the reference prior as $\pi(\phi_1.\phi_2) = (\phi_1\phi_2)^{-1}$.

32 Ratio of Normal Means

Suppose $X = \{X_1, \ldots, X_n\}$ and $Y = \{Y_1, \ldots, Y_m\}$ are available from two independent normal populations with unknown means $\mu,\ \eta$ and unknown common variance σ^2. The problem is to make inferences about the value of $\psi = \mu/\eta$, the ratio of the means.

The Fisher information matrix is (see Bernardo, 1977),

$$I(\psi,\ \eta,\ \sigma) = \frac{1}{\sigma^2}\begin{pmatrix} \eta^2 & \psi\eta & 0 \\ \psi\eta & 1+\psi^2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

It follows that the reference prior is

$$\pi(\psi,\ \eta,\ \sigma) \propto (1+\psi^2)^{-1/2}\sigma^{-1}$$

or, in terms of the original parameterization

$$\pi(\mu,\ \eta,\ \sigma) \propto (\mu^2+\eta^2)^{-1/2}\sigma^{-1}.$$

The reference posterior is

$$\pi(\psi|D) \propto (1+\psi^2)^{-1/2}(m+\psi^2 n)^{-1/2}\{S^2 + \frac{nm(\bar{x}-\psi\bar{y})^2}{m+\psi^2 n}\}^{-(n+m-1)/2},$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n,\ \bar{y} = \sum_{i=1}^{m} y_i/m$, and $S^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{m}(y_i-\bar{y})^2$.

Note for this problem, the usual noninformative prior $1/\sigma$ entails Fieller-Creasy problem. Kappenman, Geisser and Antle (1970) showed that $1/\sigma$ can lead to a confidence interval of ψ consisting the whole real line.

33 Ratio of Poisson Means

Let X and Y be independent Poisson random variables with means $\lambda\theta$ and λ. The parameter of interest is θ.

Liseo (1993) computed the Jeffreys prior and reference prior for this problem as

Prior	$\pi(\theta,\lambda)$	(Marginal) Posterior
Uniform	1	
Jeffreys	$\theta^{-1/2}$	proper
Reference	$1/\sqrt{\lambda\theta(1+\theta)}$	

Here the Jeffreys prior and reference prior give the same marginal posterior for θ,

$$\pi(\theta|D) \propto \frac{\theta^{x-1/2}}{(1+\theta)^{x+y+1}}.$$

34 Sequential Analysis

Suppose $X_1, X_2, \ldots,$ is an i.i.d. sample with common density function $f(x_i|\theta)$ which satisfies the regular continuous condition. Here X_i and θ are $k \times 1$ and $p \times 1$ vectors, respectively. Let N be the stopping time.

¿From Ye (1993), if $0 < E_\theta(N) < \infty$, the Jeffreys prior is

$$\pi_J^*(\theta) \propto (E_\theta[N])^{p/2}\sqrt{\det(I(\theta))},$$

where $I(\theta)$ is the Fisher information matrix of X_1.

Suppose $\theta = (\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(m)})$ is an m-ordered group. Furthermore, assume $E_\theta(N) \propto E_{\theta_{(1)}}(N)$ depends only on $\theta_{(1)}$ and $0 < E_\theta(N) < \infty$. Then the reference prior is (see Ye, 1993)

$$\pi_R^*(\theta_{(1)}, \ldots, \theta_{(m)}) \propto (E_\theta[N])^{p_1/2} \pi_R(\theta_{(1)}, \ldots, \theta_{(m)}),$$

where p_1 is the dimension of $\theta_{(1)}$ and $\pi_R(\theta_{(1)}, \ldots, \theta_{(m)})$ is the reference prior of θ for X_1, using the same group order and compact subsets.

35 Stress-Strength System

Suppose $X_1, \ldots, X_m$ are i.i.d. Weibull(η_1, β) random variables, and independently, $Y_1, \ldots, Y_m$ are i.i.d. Weibull(η_2, β) random variables. Parameter of interest is

$$\omega_1 = P(X_1 < Y_1) = \eta_2^\beta / (\eta_1^\beta + \eta_2^\beta).$$

When $\beta = 1$, this is the simple stress-strength system under exponential distribution, with parameter η as scale parameter. Thompson and Basu (1993) computed the reference prior and showed that the reference prior for (η_1, η_2), when ω_1 is the parameter of interest and $\omega_2 = \eta_1 + \eta_2$ is the nuisance parameter, coincide with the Jeffreys prior. With reparameterization of ω_1 and $\omega_2 = \eta_2^\beta / (\eta_1^\beta + \eta_2^\beta)$, Basu and Sun (1994) computed the following reference priors and gave a necessary and sufficient condition for the existence of a proper posterior,

Prior	$\pi(\omega_1, \omega_2, \beta)$	(Marginal) Posterior
Uniform	1	
Jeffreys[1]	$[\omega_1(1-\omega_1)\omega_2\beta]^{-1}$	
Reference[2]	$g_1(\omega_1, \omega_2)/[\omega_1(1-\omega_1)\omega_2\beta]$	
Reference[3]	$g_2(\omega_1, \omega_2)/[\omega_1(1-\omega_1)\omega_2\beta]$	
Reference[4]	$g_3(\omega_1, \omega_2)/[\omega_1(1-\omega_1)\omega_2\beta]$	

Where

$$\begin{aligned} g_1(\omega_1, \omega_2) &= 1/\sqrt{\gamma^* + a(1-a)\{\log[(1-\omega_1)/\omega_1]\}^2}, \\ g_2(\omega_1, \omega_2) &= \sqrt{a(1-\omega_1)^2 + (1-a)\omega_1^2}, \end{aligned}$$

$$g_3(\omega_1,\omega_2) = g_1(\omega_1,\omega_2)/\sqrt{\gamma^* + a\{\gamma+\log[(1-\omega_1)\omega_2]\}^2 + (1-a)\{\gamma+\log(\omega_1\omega_2)\}^2},$$

and

$$\begin{aligned}
\gamma &= 1 + \int_0^\infty (\log z)e^{-z}dz, \\
\gamma^* &= \int_0^\infty (\log z)^2 e^{-z}dz - \{\int_0^\infty (\log z)e^{-z}dz\}^2, \\
a &= m/(m+n).
\end{aligned}$$

1. Also the reference for the group ordering $\{(\omega_1,\omega_2,\beta)\}, \{\beta,\omega_1,\omega_2\}$, and $\{\beta,(\omega_1,\omega_2)\}$.
2. Group ordering $\{\omega_1,(\omega_2,\beta)\}$ and $\{\omega_1,\beta,\omega_2\}$.
3. Group ordering $\{(\omega_2,\beta),\omega_1\}$.
4. Group ordering $\{\omega_1,\omega_2,\beta\}$.

36 Sum of Squares of Normal Means

Consider the $N_p(\mu,\ I_p)$ density, where $\mu = (\mu_1,\ldots,\mu_p)$, and the parameter of interest is $\sum_{i=1}^p \mu_i^2$.

Prior	$\pi(\mu_1,\ldots,\mu_p$	(Marginal) Posterior
Uniform	1	proper
Reference[1]	$\|\|\mu\|\|^{-(p-1)}$	

1. See Datta and Ghosh (1993). This is also the prior obtained by Stein (1985) and Tibshirani (1989). It can be viewed as a hierachical prior with (i) $\mu_1,\ldots,\mu_p|\epsilon \overset{i.i.d.}{\sim} N(0,\epsilon^{-1})$, (ii) ϵ has the improper gamma density function $f(\epsilon) \propto^{-3/2}$.

37 T Distribution

The $T(\alpha,\ \mu,\ \sigma^2)$ density, where $\alpha > 0$, $-\infty < \mu < \infty$, and $\sigma^2 > 0$, is given by

$$f(x|\alpha,\ \ \mu,\ \sigma^2) = \frac{\Gamma[(\alpha+1)/2]}{\sigma(\alpha\pi)^{1/2}\Gamma(\alpha/2)} \cdot \left(1 + \frac{(x-\mu)^2}{\alpha\sigma^2}\right)^{-(\alpha+1)/2}.$$

This is a location-scale parameter problem; see that section for priors.

38 Weibull

The Weibull density, with shape parameter $c > 0$ and quasi-scale parameter $\alpha > 0$, is

$$f(x|\alpha,\ c) = c\alpha x^{c-1}\exp(-\alpha x^c), \quad x > 0.$$

$w = \log x$ has the extreme value density

$$f(w|\alpha,\ c) = c\alpha\exp(cw)\exp(-\alpha\exp(cw)), -\infty < w < \infty.$$

As indicated in Evans and Nigm (1980), setting $c = \sigma^{-1}$ and $\alpha = \exp(-\mu/\sigma)$, the density of w becomes

$$f(w|\mu,\ \sigma) = \sigma^{-1}\exp\{(w-\mu)/\sigma\}\exp[\exp\{(w-\mu)/\sigma\}],$$

in which μ and σ are seen to be location-scale parameters. Therefor, $\pi(\mu,\sigma) = 1/\sigma$. Prior transformed back to the original parameterization is $\pi(\alpha, c) = 1/(\alpha c^2)$. Analysis using the usual noninformative prior is discussed by Evans and Nigm (1980).

Sun (1991) computed the following reference prior

Prior	$\pi(\alpha, c)$	(Marginal) Posterior
Reference[1]	$1/(c\alpha)$	proper[3]
Reference[2]	$(c\alpha\sqrt{1+\gamma^* - 2\gamma - 2(1-\gamma)\log\alpha + (\log\alpha)^2})^{-1}$	

Here γ is Euler's constant, i.e., $\gamma = -\int_0^\infty (\log z)e^{-z}dz$ and $\gamma^* = \int_0^\infty (\log z)^2 e^{-z}dz$.

1. Group ordering $\{(\alpha, c)\}$ and $\{c, \alpha\}$ (also Jeffreys prior).

2. Group ordering $\{\alpha, c\}$.

3. Provided $n > 1$ and not all observations are equal.

This density can also be written in the form

$$f(x|\theta,\beta) = \frac{\beta x^{\beta-1}}{\theta^\beta}\exp\{-[\frac{x}{\theta}]^\beta\}.$$

Under this form, Sun (1991) computed the reference prior as

Prior	$\pi(\theta,\beta)$	(Marginal) Posterior
Reference[1]	θ^{-1}	proper[3]
Reference[2]	$(\theta\beta)^{-1}$	

1. Group ordering $\{(\theta,\beta)\}$ (also Jeffreys prior).
2. Group ordering $\{\theta,\beta\}$ and $\{\beta,\theta\}$.
3. Provided $n > 1$ and not all observations are equal.

References

Alba, E.D. and Mendoza, M. (1995). A discrete model for Bayesian forcasting with stable seasonal patterns. *Advances in Econometrics* **11**.

Banerjee, A.K. and Bhattacharyya, G.K. (1979). Bayesian results for the inverse Gaussian distribution with an application. *Technometrics* **21**, No 2, 247-251.

Baranchik, A. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Technical Report No **51**, Department of Statistics, Stanford University.

Basu, A.P. and Sun, D. (1994). Bayesian analysis for a stres-strength system via reference prior approach. Submitted.

Bayarri, M.J. (1981). Inferencia Bayesiana sobre el coefficiente de correlación de una problación normal bivariante. *Trab. Estadist.* **32**, 18-31.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag.

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. In *Bayesian Analysis in Statistics and Econometrics*, P. Goel and N.S. Iyengar(eds.). New York: Springer Verlag.

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

Berger, J.O. and Bernardo, J.M. (1992c). On the development of the reference prior method. In *Bayesian Statistics* **4**, J.M.Bernardo, J.O.Berger, D.V.Lindley, and A.F.M.Smith (eds.). London: Oxford University Press.

Berger, J.O. and Strawderman, W.E. (1994). Choice of hierarchical priors: admissibility in estimation of normal means. Technical Report **93-34C**, Department of Statistics, Purdue University.

Berger, J.O. and Yang, R. (1992). Noninformative priors and Bayesian testing for the AR(1) model. *Econometric Theory*, **10**, No 3-4, 461-482.

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

Bernardo, J.M. (1979). Reference posterior distributions for Bayes inference. *J. Roy. Statist. Soc. Ser B* **41**, 113-147 (with discussion).

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

Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory.* New York: John Wiley & Sons.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformation. *J. Roy. Statist. Soc. Ser B* **26**, 211-252. (with discussion).

Box, G.E.P. and Tiao, G.C. (1992). *Bayesian Inference in Statistical Analysis.* John Wiley & Sons, New York.

Chaloner, K. (1987). A Bayesian approach to the estimation of variance components for the unbalanced one-way random model. *Technometrics*, **29**, 323-337.

Crowder, M. and Sweeting, T. (1989). Bayesian inference for a bivariate binomial. *Biometrika*, **76**, 599-604.

Datta, G.S. and Ghosh, M. (1993). Some remarks on noninformative priors. Technical Report **93-16**, Department of Statistics, University of Georgia.

Eaves, D.M. (1983). On Bayesian nonlinear regression with an enzyme example. *Biometrika*, **70**, 2, 373-379.

Evans, I.G. and Nigm, A.M. (1980). Bayesian prediction for two-parameter Weibull lifetime models. *Commun. Statist. -Theor. Meth.*, **A9(6)**, 649-658.

Fatti, L.P. (1982). Predictive discrimination under the random effect model. *South African Statist. J.*, **16**, 55-77.

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

Geisser, S. (1964). Posterior odds for multivariate normal classifications. *J. Roy. Statist. Soc. Ser B* **26**, 69-76.

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

Howlader, H.A. and Weiss, G. (1988). Bayesian reliability estimation of a two-parameter Cauchy distribution. *Biom. J.*, **30**, 3, 329-337.

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.

Jeffreys, H. (1961). *Theory of Probability.* London: Oxford University Press.

Joshi, S. and Shah, M. (1991). Estimating the mean of an inverse Gaussian distribution with known coefficient of variation. *Commun. Statist. -Theory Meth.*, **20(9)**, 2907-2912.

Kappenman, R.F., Geisser, S. and Antle, C.F. (1970). Bayesian and fiducial solutions to the Fieller-Creasy problem. *Sankhya B*, **32**, 331-340.

Kass, R.E. and Wasserman, L. (1993). Formal rules for selecting prior distributions: a review and annotated bibliography.

Kubokawa, T. and Robert, C.P. (1993). New perspective on Linear calibration. Rapport Technique **#172**, LSTA, Université Pierre et Marie Curie. To appear in *J. Multivariate Anal.*

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

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

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

Pericchi, L.R. (1981). A Bayesian approach to transformation to normality. *Biometrika*, **68**, 35-43.

Poirier, D.J. (1992). Jeffreys prior for logit models. To appear in *Econometrics.*

Polson, N. and Wasserman, L. (1990). Prior distribution for the bivariate binomial. *Biometrika*, **77**, 4, 901-904.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Annal. Statist.*, **11**, No 2, 416-431.

Sono, S. (1983). On a noninformative prior distribution for Bayesian inference of multinomial distribution's parameters. *Ann. Inst. Statist. Math.*, **35**, Part A, 167-174.

Spiegelhalter, D.J. (1985). Exact Bayesian inference on the parameters of a Cauchy distribution with vague prior information. In *Bayesian Statistics* **2**, J.M.Bernardo, M.H. DeGroot, D.V.Lindley, and A.F.M.Smith (eds.). North-Holland: Elsevier Science Publishers B.V..

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach center publications, **16**. PWN-Polish Scientific Publishers, Warsaw.

Sun, D.C. (1991). Bayesian sequential reliability for Weibull and related distributions. *Ph.D. Dissertation.* Purdue University.

Sun, D.C. and Ye, K.Y. (1994a). Inference on a product of normal means with unknown variances. Submitted to *Biometrika.*

Sun, D.C. and Ye, K.Y. (1994b). Frequentist validity of posterior quantiles for a two-parameter exponential family. Submitted to *Biometrika.*

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

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

Tiwari, R.C., Chib, S. and Jammalamadaka, S.R. (1989). Bayes estimation of the multiple correlation coefficient. *Commun. Statist. -Theory Meth.*, **18(4)**, 1401-1413.

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

Wixley, R.A.J. (1993). Data translated likelihood and choice of prior for power transformation to normality.

Yang, R. and Berger, J.O. (1993). Estimation of a covariance matrix using the reference prior. *Annal. Statist.*, **22**, No 3, 1195-1211.

Yang, R. and Chen, M.H. (1993). Bayesian analysis for random coefficient regression models using noninformative priors. *J. Multivariate Anal.*, **55**, No 2, 283-311.

Yang, R. and Pyne, D. (1996). Bayesian analysis with mixed model in unbalanced case.

Ye, K.Y. (1990). Noninformative priors in Bayesian analysis. *Ph.D. Dissertation.* Purdue

University.

Ye, K.Y. (1991). Bayesian reference prior analysis on the ratio of one-way random effect model. Technical Report **91-22**, Department of Statistics, Virginia Polytechnic Institute and State University.

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

Ye, K.Y. and Berger, J.O. (1991). Noninformative priors for inference in exponential regression models. *Biometrika* **78**, 645-656.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley & Sons.

Zellner, A. (1984). Maximal data information prior distributions. In Zellner, A. *Basic Issues in Econometrics*, U. of Chicago Press.

Zellner, A. (1993). Models, prior information and Bayesian analysis.