

Objective Priors under Sequential Experimentation ¹

DONGCHU SUN, *University of Missouri-Columbia, Columbia, MO 65211, USA*

e-mail: sund@missouri.edu

and

JAMES O. BERGER, *Duke University, Durham, NC 27708, USA*

e-mail: berger@stat.duke.edu

Abstract

Objective priors for sequential experiments are considered. Common priors, such as the Jeffreys prior and the reference prior, will typically depend on the stopping rule used for the sequential experiment. New expressions for reference priors are obtained in various contexts, and computational issues involving such priors are considered.

Key words and Phrases: posterior distributions, sequential experimentation, expected stopping time, Jeffreys prior, reference prior, frequentist coverage.

1 Introduction

Bayesian analysis using objective or default priors has received considerable attention; cf. Datta and Mukerjee (2004), Bernardo (2005), Berger (2006), Ghosh, Delampady and Samanta (2006), and references therein. The latter book, in particular, contains an excellent discussion of the issues and controversies involving objective priors, reflecting the many years of leadership in the field of J.K. Ghosh (along with his many coauthors).

A common objective prior is the Jeffreys prior (Jeffreys, 1961), which is proportional to the square root of the determinant of the Fisher information matrix. The Jeffreys prior is quite useful for a single parameter model, but can be seriously deficient for multiparameter

¹This research was supported by the National Science Foundation, under grants DMS-0103265 and SES-0351523, and the National Institute of Health, under grants R01-CA100760 and R01-MH071418.

models; this has led to preference for reference priors in multiparameter situations (cf. Berger and Bernardo, 1992, and Bernardo, 2005).

Almost all results on objective priors have been for fixed sample size experiments. In practice, however, statistical experiments are often conducted sequentially, with a known stopping rule (cf. Siegmund, 1985, and Ghosh, Sen and Mukhopadhyay, 1997). Bartholomew (1965) and Geisser (1979) introduced the notion that objective priors for a sequential experiment should depend on the expected stopping time. Ye (1993) derived the reference prior for sequential experiments when the expected stopping time depends on the parameter of interest only. In this paper we generalize Ye's result in various directions, and provide some new computation tools for use with priors that depend on expected stopping times.

The paper is arranged as follows. Section 2 reviews the Fisher information matrix for sequential experiments with a known stopping rule, derives the Jeffreys/reference prior for illustrative one-parameter examples, and then provides an expression for multiparameter reference priors when the stopping rule satisfies a certain property. In Section 3, reference priors and matching priors (cf. Datta and Mukerjee, 2004) are derived for Bar-Lev and Reiser's (1982) two-parameter exponential family. Illustrations are given for normal distributions with several commonly used stopping times.

Computation of expected stopping times is often difficult, so that utilization of reference priors for sequential experiments is typically challenging. In Section 4, an approximation to the reference prior for sequential experiments is introduced which is exact under some circumstances, seems to be a reasonable approximation in general, and allows for much simpler computation.

2 Noninformative Priors with a Known Stopping Rule

2.1 Notation and the Jeffreys-rule Prior

We assume that X_1, X_2, \dots , is an iid sequence of random variables with density $f(x | \boldsymbol{\theta})$ that is *regular* (Walker, 1969). Here $\boldsymbol{\theta}$ is a $q \times 1$ vector of unknown parameters. Let N denote a proper stopping time for the sequential experiment – see Govindarajulu (1981) for definition, which also is a source for the following well-known lemma:

Lemma 1 Let $\mathbf{I}(\boldsymbol{\theta})$ be the Fisher information matrix based on X_1 . Under the proper stopping time N , the Fisher information based on (X_1, \dots, X_N) is

$$\mathbf{I}^* = \mathbb{E}_{\boldsymbol{\theta}}(N)\mathbf{I}(\boldsymbol{\theta}). \quad (1)$$

The Jeffreys-rule prior (Jeffreys, 1961) for $\boldsymbol{\theta}$ is defined as the square root of the determinant of the Fisher information matrix. In the fixed sample size case, this is $\pi_J(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}$. For the sequential experiment, it follows from the above lemma that Jeffreys prior is

$$\pi_J^*(\boldsymbol{\theta}) \propto \{\mathbb{E}_{\boldsymbol{\theta}}(N)\}^{q/2} |\mathbf{I}(\boldsymbol{\theta})|^{1/2} \propto \{\mathbb{E}_{\boldsymbol{\theta}}(N)\}^{q/2} \pi_J(\boldsymbol{\theta}). \quad (2)$$

Example 1 Let N_r be a random variable with a negative binomial distribution $NB(r, p)$, where r is a positive integer and $p \in (0, 1)$. Let X_1, X_2, \dots be a sequence of Bernoulli random variables with success probability p . N_r can be viewed as a stopping time for this Bernoulli sequence as follows:

$$N_r = \inf\{n \geq 1 : X_1 + \dots + X_n = r\}.$$

The probability of N_r is

$$P(N_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad \text{for } k = r, r+1, \dots$$

An easy computation yields $\mathbb{E}_p(N_r) = r/p$. Since the Jeffreys rule prior for a Bernoulli random variable is $\pi_J(p) \propto 1/\sqrt{p(1-p)}$, it follows from (2) that the Jeffreys rule prior for

the negative binomial distribution is $\pi_J^*(p) \propto (r/p) \pi_J(p) \propto 1/(p\sqrt{1-p})$. This, of course, is well known from a direct computation with the negative binomial distribution, as discussed in Geisser (1984) and Bernardo and Smith (1994, Example 5.14, p. 315).

We next consider an example with a continuous stopping time.

Example 2 Let $\{Z(t) : t > 0\}$ be a Brownian motion with constant drift θ and variance 1 per unit time, so $Z(t) \sim N(\theta t, t)$. Let $-\infty < a < 0 < b < \infty$, and let T_{ab} denote the random stopping time

$$T_{ab} = \inf\{t > 0 : Z(t) \leq a \text{ or } Z(t) \geq b\}. \tag{3}$$

It follows from Hall (1992) that

$$\mathbb{E}_\theta(T_{ab}) = \begin{cases} \frac{1}{\theta} \left\{ b - (b-a) \frac{e^{2b\theta} - 1}{e^{2(b-a)\theta} - 1} \right\}, & \text{if } \theta \neq 0, \\ -ab, & \text{if } \theta = 0. \end{cases}$$

Note that the constant prior is the Jeffreys prior based on stopping at a fixed time (Polson and Roberts, 1993; Sivaganesan and Lingam, 2002), from which it follows that the Jeffreys or reference prior for this situation is $\pi(\theta) = \sqrt{\mathbb{E}_\theta(T_{ab})}$.

This is of additional interest because of the study in Brown (1988), which showed that the commonly used estimate $Z(T)/T$, which is the posterior mean under a constant prior for θ , is inadmissible under estimation with squared error loss. Brown (1988) further suggested that prior distributions which behaved like $|\theta|^{-1}$ as $|\theta| \rightarrow \infty$ were optimal for this situation. The Jeffreys/reference prior has behavior $|\theta|^{-1/2}$ as $|\theta| \rightarrow \infty$, and so is not of this form, but admissibility is very dependent on the loss function used. Indeed, it can be argued that a weighted-squared error loss is appropriate for this situation, and the reference prior is likely admissible for an appropriate weight.

2.2 Reference Priors

Reference priors depend on a grouping and ordering of the parameters; see Berger and Bernardo (1992). Suppose that $\boldsymbol{\theta} = (\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(m)})$ is an m -ordered grouping, where the

dimension of component $\boldsymbol{\theta}_{(i)}$ is q_i for $i = 1, \dots, m$. Datta and Ghosh (1995) considered the special case in which the (fixed sample size) Fisher information matrix is diagonal, with the diagonal elements being products of functions of the $\boldsymbol{\theta}_{(i)}$. Our first result is a generalization of their result.

Theorem 1 Suppose that the Fisher information matrix corresponding to a single observation X_1 is of the form

$$\mathbf{I}(\boldsymbol{\theta}) = \text{diag}\left(\prod_{i=1}^m \mathbf{G}_{1i}(\boldsymbol{\theta}_{(i)}), \dots, \prod_{i=1}^m \mathbf{G}_{mi}(\boldsymbol{\theta}_{(i)})\right), \quad (4)$$

where \mathbf{G}_{li} is a $q_i \times q_i$ matrix. Assume further that the expected stopping time is of the form

$$\mathbb{E}_{\boldsymbol{\theta}}(N) = \prod_{i=1}^m g_i(\boldsymbol{\theta}_{(i)}). \quad (5)$$

Then the reference prior for $\boldsymbol{\theta}$ in the sequential experiment is

$$\pi_R^*(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(m)}) \propto \prod_{i=1}^m [g_i(\boldsymbol{\theta}_{(i)})]^{q_i/2} \pi_R(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(m)}), \quad (6)$$

where $\pi_R(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(m)})$ is the reference prior based on the single observation X_1 , given by

$$\pi_R(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(m)}) = \prod_{i=1}^m |\mathbf{G}_{ii}(\boldsymbol{\theta}_{(i)})|^{1/2}. \quad (7)$$

Proof. The proof is essentially identical to that in Datta (1996), noting that, under (5), the sequential Fisher information matrix has the product structure of Datta and Ghosh (1995). □

This theorem can also be considered to be a generalization of Ye (1993), who considered the case where $\mathbb{E}_{\boldsymbol{\theta}}(N)$ depends only on $\boldsymbol{\theta}_{(1)}$, the parameter of interest.

Berger and Bernardo (1992) suggested that one should always try to use a one-at-a-time reference prior, where each component of the grouping of parameters contains only one parameter. The following result is an immediate corollary of Theorem 1.

Corollary 1 Suppose that the conditions of Theorem 1 hold. If $q_i = 1$, for $i = 1, \dots, m = k$, then the resulting one-at-a-time reference prior for $\boldsymbol{\theta}$ in the sequential experiment is

$$\pi_R^*(\boldsymbol{\theta}) \propto \sqrt{\mathbb{E}_{\boldsymbol{\theta}}(N)} \pi_R(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k).$$

For later purposes, we also note another corollary of Theorem 1, which applies if the dimension of each component of the grouping of parameters has dimension 2.

Corollary 2 Suppose that the conditions of Theorem 1 hold. If all $q_j = 2$, then the reference prior for $\boldsymbol{\theta}$ in the sequential experiment is

$$\pi_R^*(\boldsymbol{\theta}) \propto \mathbb{E}_{\boldsymbol{\theta}}(N)\pi_R(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(m)}).$$

3 A Two-Parameter Exponential Family

3.1 The Model and Reference Priors

Bar-Lev and Reiser (1982) considered the following density function of the generic two-parameter exponential family:

$$f(x | \theta_1, \theta_2) = a(x) \exp\{\theta_1 U_1(x) - \theta_1 G_2'(\theta_2) U_2(x) - \psi(\theta_1, \theta_2)\}, \quad (8)$$

where $\theta_1 < 0$, $\theta_2 = \mathbb{E}\{U_2(X) | (\theta_1, \theta_2)\}$, $G_i(\cdot)$, ($i = 1, 2$) are infinitely differentiable functions satisfying $G_i'' > 0$, and $\psi(\theta_1, \theta_2) = -\theta_1\{\theta_2 G_2'(\theta_2) - G_2(\theta_2)\} + G_1(\theta_2)$. This is a large class of distributions, which includes, for suitable choices of G_1 , G_2 , U_1 and U_2 , many popular statistical models such as the normal, inverse normal, gamma, and inverse gamma. Table 1, reproduced from Sun (1994), indicates how each distribution arises.

Table 1. Special cases of Bar-Lev and Reiser's (1982) two parameter exponential family,

$$\text{where } h(\theta_1) = -\theta_1 + \theta_1 \log(-\theta_1) + \log(\Gamma(-\theta_1)).$$

	$G_1(\theta_1)$	$G_2(\theta_2)$	$U_1(x)$	$U_2(x)$	θ_1	θ_2
$N(\mu, \sigma^2)$	$-\frac{1}{2} \log(-2\theta_1)$	θ_2^2	x^2	x	$-1/(2\sigma^2)$	μ
Inverse Gaussian	$-\frac{1}{2} \log(-2\theta_1)$	$1/\theta_2$	$1/x$	x	$-\alpha/2$	$\sqrt{\alpha/\mu}$
Gamma	$h(\theta_1)$	$-\log \theta_2$	$-\log x$	x	$-\alpha$	μ
Inverse Gamma	$h(\theta_1)$	$-\log \theta_2$	$\log x$	$1/x$	$-\alpha$	μ

Let X_1, X_2, \dots be a sequence of random variables from (8). The Fisher information per observation is

$$\mathbf{I}(\theta_1, \theta_2) = \begin{pmatrix} G_1''(\theta_1) & 0 \\ 0 & -\theta_1 G_2''(\theta_2) \end{pmatrix}.$$

Thus the Jeffreys prior for a single observation is

$$\pi_J(\theta_1, \theta_2) \propto \sqrt{|\theta_1|} \sqrt{G_1''(\theta_1) G_2''(\theta_2)}. \quad (9)$$

When either θ_1 or θ_2 is the parameter of interest, it is shown in Sun and Ye (1996) that the one-at-a-time reference priors are

$$\pi_R(\theta_1, \theta_2) = \sqrt{G_1''(\theta_1) G_2''(\theta_2)}. \quad (10)$$

The parameter θ_2 is the expectation of $U_2(X_1)$. Bose and Boukai (1993) considered inference about θ_2 in sequential experimentation with the following stopping time:

$$N_a = \inf \left\{ n \geq m_0 : Y_n < n G_1' \left(-\frac{a^2}{n^2} \right) \right\}, \quad a \geq 0, \quad (11)$$

where $Y_n = n^{-1} \sum_{i=1}^n U_1(X_i) - G_2 \{ n^{-1} \sum_{i=1}^n U_2(X_i) \}$ and $m_0 \geq 2$ is an initial sample size.

From Theorem 2 of Bose and Boukai (1993), we have

$$\lim_{a \rightarrow \infty} \frac{N_a}{a} = \frac{1}{\sqrt{|\theta_1|}} \quad a.s. \quad (12)$$

$$\lim_{a \rightarrow \infty} \mathbf{E}_{\theta} \left(\frac{N_a}{a} \right) = \frac{1}{\sqrt{|\theta_1|}}. \quad (13)$$

Bar-Lev and Reiser (1982) showed that the distribution of Y_n does not depend on the parameter θ_2 . So condition (5) satisfies when either θ_1 or θ_2 is the parameter of interest. The following result is immediate from Theorem 1 or Corollary 1.

Fact 1 (a) The Jeffreys prior for (θ_1, θ_2) in model (8) with the stopping time (11) and when a is large is approximately

$$\pi_J^*(\theta_1, \theta_2) \propto \sqrt{G_1''(\theta_1) G_2''(\theta_2)}. \quad (14)$$

(b) The one-at-a-time reference prior for (θ_1, θ_2) in model (8), when either θ_1 or θ_2 is the parameter of interest, the stopping time (11) is used, and a is large enough, is approximately

$$\pi_R^*(\theta_1, \theta_2) \propto \frac{1}{|\theta_1|^{1/4}} \sqrt{G_1''(\theta_1)G_2''(\theta_2)}. \quad (15)$$

Example 3 Suppose X_1, X_2, \dots , are a sequence of $N(\mu, \sigma^2)$ random variables. Then $\theta_1 = -1/2\sigma^2, \theta_2 = \mu, G_1'(\theta_1) = -1/2\theta_1$, and $Y_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2$. The stopping rule (11) becomes

$$N_a = \inf \left\{ n \geq m_0 : n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 < n^2 / (2a^2) \right\}.$$

So the priors (9), (10), (14), and (15) are, respectively,

$$\pi_J(\mu, \sigma^2) \propto \frac{1}{(\sigma^2)^{3/2}}, \quad \pi_R(\mu, \sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi_J^*(\mu, \sigma^2) \propto \frac{1}{\sigma^2}, \quad \pi_R^*(\mu, \sigma^2) \propto \frac{1}{(\sigma^2)^{3/4}} \quad (16)$$

or equivalently,

$$\pi_J(\mu, \sigma) \propto \frac{1}{\sigma^2}, \quad \pi_R(\mu, \sigma) \propto \frac{1}{\sigma}, \quad \pi_J^*(\mu, \sigma) \propto \frac{1}{\sigma}, \quad \pi_R^*(\mu, \sigma) \propto \frac{1}{\sqrt{\sigma}}. \quad (17)$$

3.2 Probability Matching Priors for a sequential experiment

Asymptotic frequentist coverage is an often-used criterion to compare objective priors; see Datta and Ghosh (1995), Datta, Ghosh, and Mukerjee (2000), and Datta and Mukerjee (2004) for discussion and references. The most common approach is to find a 'matching prior,' i.e. a prior which results in posterior one-sided credible intervals that are also accurate as frequentist confidence intervals. Another type of matching prior, considered by Sun and Ye (1996), is a prior such that the confidence interval based on the signed squared root transformation of the log-likelihood ratio is also a Bayesian credible interval. Almost all of the literature considers the fixed sample case for iid observations; exceptions are Ye (1993) and Sun (1994).

For sequential experiments involving the Bar-Lev and Reiser (1982) two-parameter exponential family, let $l_n(\theta_1, \theta_2)$ be the log-likelihood function of (θ_1, θ_2) , given $\mathbf{X}_n = (X_1, \dots, X_n)$,

and let $(\hat{\theta}_{n1}, \hat{\theta}_{n2})$ be the maximum likelihood estimator of (θ_1, θ_2) . Write $Y_n = n^{-1} \sum_{i=1}^n U_1(X_i) - G_2\{n^{-1} \sum_{i=1}^n U_2(X_i)\}$. Then, on $\{Y_n \in G'_1(\Theta_1)\} \cap \{n^{-1} \sum_{i=1}^n U_2(X_i) \in \Theta_2\}$, $\hat{\theta}_{n1}$ is the solution of $Y_n = G'_1(\hat{\theta}_{n1})$, and $\hat{\theta}_{n2} = n^{-1} \sum_{i=1}^n U_2(X_i)$. Define

$$\begin{aligned} I_1(\omega_1, \theta_1) &= G_1(\theta_1) - G_1(\omega_1) - G'_1(\omega_1)(\theta_1 - \omega_1), \quad \omega_1, \theta_1 \in \Theta_1, \\ I_2(\omega_2, \mu_2) &= G_2(\omega_2) - G_2(\theta_2) - G'_2(\theta_2)(\omega_2 - \theta_2), \quad \omega_2, \theta_2 \in \Theta_2. \end{aligned}$$

From the convexity of G_1 and G_2 , these two functions are nonnegative. From Sun (1994), the log-likelihood ratio is $l_n(\hat{\theta}_{n1}, \hat{\theta}_{n2}) - l_n(\theta_1, \theta_2) = (Z_{n1}^2 + Z_{n2}^2)/2$, where

$$\begin{pmatrix} Z_{n1} \\ Z_{n2} \end{pmatrix} = \begin{pmatrix} \{2nI_1(\hat{\theta}_{n1}, \theta_1)\}^{1/2} \operatorname{sgn}(\theta_1 - \hat{\theta}_{n1}) \\ \{-2n\theta_1 I_2(\hat{\theta}_{n2}, \theta_2)\}^{1/2} \operatorname{sgn}(\theta_2 - \hat{\theta}_{n2}) \end{pmatrix}$$

is a generalized signed square root of the log-likelihood ratio.

Let $P_{(\theta_1, \theta_2)}$ denote probability over X_1, X_2, \dots , given (θ_1, θ_2) , and, for a fixed prior $\pi(\theta_1, \theta_2)$, let $P^\pi(\cdot | \mathbf{X}_n)$ denote posterior probability given \mathbf{X}_n . Suppose we are considering a stopping time, N_a , such that $N_a \rightarrow \infty$ almost surely as $a \rightarrow \infty$. An asymptotic frequentist matching prior in this sequential setting is a prior π such that

$$P^\pi(Z_{N_a,1} \leq c_1, Z_{N_a,2} \leq c_2 | \mathbf{X}_{N_a}) = P_{(\theta_1, \theta_2)}(Z_{N_a,1} \leq c_1, Z_{N_a,2} \leq c_2) + O(a^{-1}), \quad (18)$$

for all c_1 and c_2 in $P_{(\theta_1, \theta_2)}$ -probability.

Suppose now that the stopping rule satisfies

$$\frac{N_a}{a} \rightarrow \tau(\boldsymbol{\theta}), \quad \text{in } L_1. \quad (19)$$

From Sun (1994), the unique prior satisfying (18), and hence the unique asymptotic matching prior, is

$$\pi_m^*(\theta_1, \theta_2) \propto \sqrt{\tau(\boldsymbol{\theta}) G_1''(\theta_1) G_2''(\theta_2)}. \quad (20)$$

As an immediate example, for the stopping time defined in (11), property (13) establishes that (19) holds; hence the reference prior given in (15) is also the asymptotic matching prior, a very desirable situation.

Example 3 (continued). In deriving the sequential likelihood ratio test to see if $(\mu, \sigma^2) = (\mu_0, \sigma_0^2)$, Woodroffe (1982) considered the following stopping rule,

$$N_a = \min\left(b_2 a, \inf\left\{n \geq b_1 a : \sum_{i=1}^n X_i^2 - n - n \log(\hat{\sigma}_n^2) > 2a\right\}\right), \quad (21)$$

where $0 < b_1 < b_2 < \infty$ are two prespecified numbers, $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, and $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Theorem 8.3 of Woodroffe (1982) implies that

$$\frac{a}{N_a} \rightarrow \begin{cases} b_2, & \text{if } \rho^2(\boldsymbol{\theta}) < 1/b_2, \\ \rho^2(\boldsymbol{\theta}), & \text{if } 1/b_2 < \rho^2(\boldsymbol{\theta}) < 1/b_1, \\ b_1, & \text{if } \rho^2(\boldsymbol{\theta}) > 1/b_1, \end{cases}$$

in $P_{(\theta_1, \theta_2)}$ -probability, as $a \rightarrow \infty$, where $\rho^2(\boldsymbol{\theta}) = G_1(\theta_1) - G_1(-0.5) - G_1'(-0.5)(\theta_1 + 0.5) - \theta_1 \theta_2^2 = \{(\mu^2 + 1)/\sigma^2 + \log(\sigma^2) - 1\}/2$. Thus (20) gives an asymptotic matching prior for this situation. Note, however, that the expected stopping time is not of the form (5), so that we cannot assert that this prior is also a one-at-a-time reference prior.

4 Computation

If $\mathbb{E}_{\boldsymbol{\theta}}[N]$ is available in closed form, as in the examples in this paper, computation with any of the sequential priors can be done using common MCMC techniques. Hence we only consider here the case in which $\mathbb{E}_{\boldsymbol{\theta}}[N]$ can only be computed numerically.

4.1 Brute Force Computation

All the Jeffreys, reference, and matching priors that have been discussed for a sequential experiment are of the form $\Psi(\mathbb{E}_{\boldsymbol{\theta}}[N])\pi_F(\boldsymbol{\theta})$, where Ψ is some operator and π_F is the corresponding prior for the fixed sample size experiment. The posterior distribution corresponding to this prior is

$$\pi^*(\boldsymbol{\theta} \mid \mathbf{X}_N) \propto \Psi(\mathbb{E}_{\boldsymbol{\theta}}[N])\pi_F(\boldsymbol{\theta}) \prod_{i=1}^N f(X_i \mid \boldsymbol{\theta}), \quad (22)$$

where $\mathbf{X}_N = (X_1, \dots, X_N)$ is the data.

The brute force method for simulating from this posterior distribution is the following Metropolis algorithm:

Step 1. Sample a proposed $\boldsymbol{\theta}'$, from the fixed sample size posterior density of $\boldsymbol{\theta}$, which is proportional to $\pi_F(\boldsymbol{\theta}) \prod_{i=1}^N f(X_i | \boldsymbol{\theta})$.

Step 2. Numerically estimate $\mathbb{E}_{\boldsymbol{\theta}'}[N]$. For instance, one could repeatedly sample N from its distribution given $\boldsymbol{\theta}'$, by simply repeatedly simulating the sequential experiment for the given $\boldsymbol{\theta}'$, observing the N that results from each simulation, and averaging to obtain the estimate $\widehat{\mathbb{E}_{\boldsymbol{\theta}'}[N]}$.

Step 3. Perform a Metropolis step: sample $u \sim \text{uniform}(0, 1)$ and, with $\boldsymbol{\theta}$ denoting the previous value the parameter, accept $\boldsymbol{\theta}'$ if

$$u \leq \min \left\{ 1, \frac{\Psi(\widehat{\mathbb{E}_{\boldsymbol{\theta}}[N]})}{\Psi(\widehat{\mathbb{E}_{\boldsymbol{\theta}'}[N]})} \right\},$$

and set $\boldsymbol{\theta}'$ equal to the previous $\boldsymbol{\theta}$ otherwise.

If one cannot directly draw from the posterior in Step 1, one could instead using any MCMC scheme, e.g. Gibbs sampling or Metropolis-Hastings. If doing so, however, be sure to iterate Step 1 many times before moving on to Step 2. This is because Step 2 is typically extremely expensive, as it may involve thousands of simulations of the entire experiment simply to compute one Metropolis acceptance probability. In situations where one dependent step is much more expensive than others, it pays to iterate first on the others.

4.2 The Two-Dimensional Case

If using the Jeffreys prior in a two-dimensional problem or the reference prior in the situation of Corollary 2, the posterior distribution is of the form

$$\pi^*(\boldsymbol{\theta} | \mathbf{X}_N) \propto \mathbb{E}_{\boldsymbol{\theta}}[N] \pi_F(\boldsymbol{\theta}) \prod_{i=1}^N f(X_i | \boldsymbol{\theta}). \quad (23)$$

This allows a remarkable simplification in the computation, by introducing N as a latent variable.

To avoid confusion, we will label the latent variable as \tilde{N} ; it is a variable with the same distribution as N , but is independent of N . Write the density of \tilde{N} given $\boldsymbol{\theta}$ as $p(\tilde{N} | \boldsymbol{\theta})$. Then the joint density of $(\tilde{N}, \boldsymbol{\theta})$, given the data $\mathbf{X}_N = (X_1, \dots, X_N)$, is proportional to

$$p(\tilde{N} | \boldsymbol{\theta}) \tilde{N} \pi_F(\boldsymbol{\theta}) \prod_{i=1}^N f(X_i | \boldsymbol{\theta}). \quad (24)$$

Sampling $(\tilde{N}, \boldsymbol{\theta})$ from this distribution will result in $\boldsymbol{\theta}$ from (23), as can easily be seen by marginalizing over \tilde{N} in (24).

Here is a Metropolis algorithm for sampling from (24).

Step 1. Sample a proposed $\boldsymbol{\theta}'$, from the fixed sample size posterior density of $\boldsymbol{\theta}$, which is proportional to $\pi_F(\boldsymbol{\theta}) \prod_{i=1}^N f(X_i | \boldsymbol{\theta})$.

Step 2. Sample a proposed \tilde{N}' from $p(\tilde{N} | \boldsymbol{\theta}')$. This can always be done by simply simulating the sequential experiment once, given $\boldsymbol{\theta}'$.

Step 3. Perform a Metropolis step: sample $u \sim \text{uniform}(0, 1)$ and, letting $(\tilde{N}, \boldsymbol{\theta})$ denote the previous value the parameter, accept $(\tilde{N}', \boldsymbol{\theta}')$ if

$$u \leq \min \left\{ 1, \frac{\tilde{N}}{\tilde{N}'} \right\},$$

and set $(\tilde{N}', \boldsymbol{\theta}')$ equal to the previous $(\tilde{N}, \boldsymbol{\theta})$ otherwise. (Note that, if $\tilde{N}' < \tilde{N}$, one would always accept the candidate.)

The reason that this is a vastly more efficient algorithm than the brute force algorithm is that one need only simulate a single draw of \tilde{N}' in Step 2, whereas thousands of draws would be needed in Step 2 of the brute force algorithm to compute $\mathbb{E}_{\boldsymbol{\theta}'}[\widehat{N}]$. Again, of course, Step 1 could be replaced by any convenient dependent MCMC scheme. Whether one then needs to iterate Step 1 before moving on to Step 2 will be context dependent.

4.3 Modified Reference Priors

The most desirable prior is the one-at-a-time reference prior given in Corollary 1, resulting in the posterior distribution

$$\pi^*(\boldsymbol{\theta} \mid \mathbf{X}_N) \propto \sqrt{\mathbb{E}_{\boldsymbol{\theta}}[N]} \pi_R(\boldsymbol{\theta}) \prod_{i=1}^N f(X_i \mid \boldsymbol{\theta}). \quad (25)$$

Unfortunately, the latent variable trick is not available for sampling from this distribution.

Interestingly, however, it is frequently the case that

$$\sqrt{\mathbb{E}_{\boldsymbol{\theta}}[N]} \approx \mathbb{E}_{\boldsymbol{\theta}}[\sqrt{N}]. \quad (26)$$

When this is the case, the latent variable trick can be applied, and the algorithm from Section 4.2 can be utilized by simply replacing \tilde{N}/\tilde{N}' in the Metropolis step with $\sqrt{\tilde{N}/\tilde{N}'}$.

In the remainder of the section, we discuss the reason that the approximation (26) often holds. The first is that the sampling distribution of N may be rather concentrated in a region of large N , in which case the approximation is clearly good.

Example- Bar-Lev and Reiser (1982) (continued). For the stopping time N_a defined in (11), it follows from (12) and (13) that $\lim_{a \rightarrow \infty} \mathbb{E}_{\boldsymbol{\theta}}(\sqrt{N_a/a}) = 1/|\theta_1|^{1/4}$. We then have

$$\lim_{a \rightarrow \infty} \mathbb{E}_{\boldsymbol{\theta}} \sqrt{\frac{N_a}{a}} \propto \lim_{a \rightarrow \infty} \sqrt{\mathbb{E}_{\boldsymbol{\theta}} \left(\frac{N_a}{a} \right)}.$$

Example 1 (continued). Let N_r have the negative binomial distribution $NB(r, p)$. Note that $\mathbb{E}_p(N_r) = r/p$ and $Var_p(N_r) = rp/(1-p)^2$. As $r \rightarrow \infty$, we have

$$\sqrt{N_r/r} \rightarrow 1/\sqrt{p} \text{ in probability and } \mathbb{E}_p(\sqrt{N_r/r}) \rightarrow \sqrt{\mathbb{E}_p(N_r/r)} \equiv 1/\sqrt{p}.$$

To see the difference between $\mathbb{E}_p(\sqrt{N_r/r})$ and $\sqrt{\mathbb{E}_p(N_r/r)}$ for moderate r , they are plotted, as a function of p , in Figure 1 for $r = 1$ and $r = 9$. For $r = 9$, the curves are essentially indistinguishable; even for the minimal $r = 1$ they are quite close.

It is also interesting to look at the posterior distributions for this example. In Figure 2, we plot the posterior densities of p for three priors $\pi_J(p) \propto 1/\sqrt{p(1-p)}$, $\pi_R^*(p) \propto 1/\sqrt{p}$ and

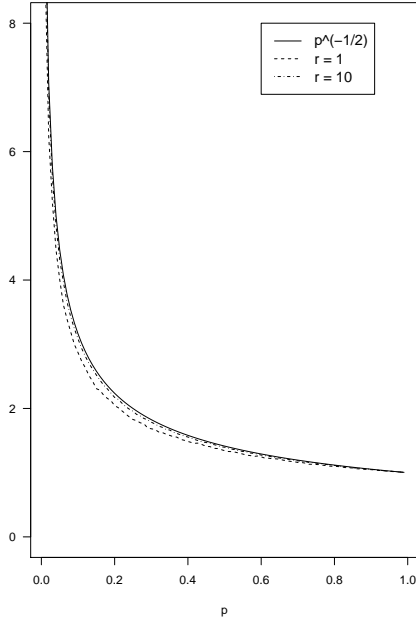


Figure 1: Negative binomial example: comparison of $\sqrt{\mathbb{E}_p(N_r/r)}$ and $\mathbb{E}_p(\sqrt{N_r/r})$ for $r = 1$ and $r = 9$.

the approximate prior $\pi_M(p) \propto \mathbb{E}_p(\sqrt{N_r^*/r})$. For even the very small $r = 2$, the posterior densities under the two priors π_R^* and π_M are quite close, yet substantially different from that under π_J . For a moderate $r = 10$, the posterior densities under π_R^* and π_M are almost identical. Note that the posterior densities of p under π_J and π_R^* are Beta $(r, N_r - r + 0.5)$ and Beta $(r = 0.5, N_r - r + 0.5)$, respectively. The posterior densities of p under π_M were computed using 5000 Metropolis samples.

As a final indication of the similarity of the true and approximate reference priors in this example, and of the value of using the sequential reference priors, we compare the frequentist coverage probabilities that result from their use in obtaining confidence intervals for p . Table 2 considers the frequentist coverage of one-sided 95% Bayesian credible regions, based on the fixed sample size Jeffreys prior π_J , the sequential Jeffreys/reference prior π_R^* and the approximate prior π_M for various combination of r and p . The fixed sample size Jeffreys prior performs worse than the other two, indicating the value of using the sequential versions, while the reference prior and the approximate prior are almost equally good.

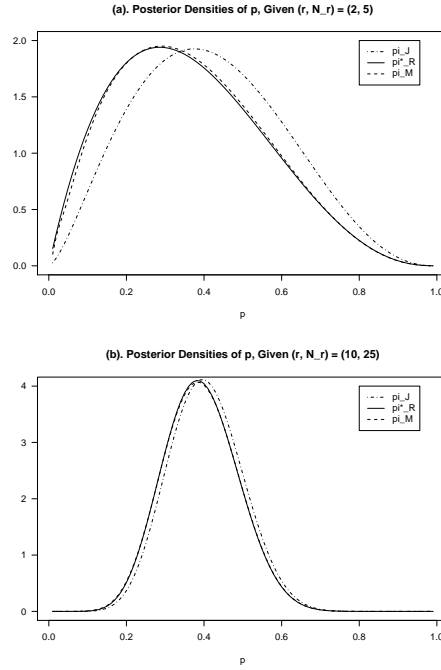


Figure 2: Posterior densities of p based on the priors $\pi_J(p) = 1/\sqrt{p(1-p)}$, $\pi_R^*(p) = 1/(p\sqrt{1-p})$, and $\pi_M(p) = \mathbb{E}_p(\sqrt{N_r^*/r})$ for $r = 1, 10$; (a) $(r, N_r) = (2, 5)$; (b) $(r, N_r) = (10, 25)$.

Table 2: Coverage Probability of one-sided 95% Bayesian credible sets for the negative binomial Example 1, under the three priors $\pi_J(p) = 1/\sqrt{p(1-p)}$, $\pi_R^*(p) = 1/(p\sqrt{1-p})$, and $\pi_M(p) = \mathbb{E}_p(\sqrt{N_r^*/r})$.

r	p	π_J	π_R^*	π_M
2	0.1	.1142(.9738)	.0516(.9511)	.0487(.9509)
2	0.5	.0002(.9652)	.0010(.9381)	.0008(.9455)
2	0.9	.0001(.9724)	.0003(.9700)	.0000(.9729)
8	0.1	.0751(.9642)	.0474(.9498)	.0465(.9534)
8	0.5	.0552(.9688)	.0522(.9536)	.0568(.9517)
8	0.9	.0000(.9307)	.0001(.9310)	.0002(.9339)
30	0.1	.0617(.9571)	.0508(.9497)	.0516(.9523)
30	0.5	.0556(.9594)	.0512(.9495)	.0525(.9503)
30	0.9	.0426(.9369)	.0438(.9410)	.0442(.9368)

References

- Bar-Lev, S.K. and Reiser, B. (1982). An exponential subfamily which admits UMPU test based on a single test statistic. *Ann. Statist.* **10** 979-989.
- Bartholomew, D. (1965). A comparison of some Bayesian and frequentist inference. *Biometrika*, **52**, 19-35.
- Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, **1**, 385–402 and 457–464.
- 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. (1992). On the development of the reference prior method (with discussion). *Bayesian Statistics*, **4**, Eds. J.M. Bernardo, J.O. Berger, A.P Dawid and A.F.M. Smith. Oxford Univ. Press, 35-60.
- Bernardo, J.M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. B* **41** 113-147.
- Bernardo, J. M. (2005). Reference analysis. In *Handbook of Statistics* **25** (D. K. Dey and C. R. Rao, eds.), 17–90. Amsterdam: Elsevier.
- Bernardo, J.M. and Smith, A.F.M. (1984). *Bayesian Theory*, Wiley, New York.
- Bose, A. and Boukai, B. (1993). Sequential estimation results for a two-parameter exponential family of distributions. *Ann. Statist.* **21** 484-502.
- Brown, L. D. (1988), "The differential inequality of a statistical estimation problem", *Statistical Decision Theory and Related Topics IV*, Volume 1, 299-324.
- Cox, D. R. and Reid, N. (1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. Roy. Statist. Soc., B*, **49** 1-39.
- Datta, G.S. (1996). On priors providing frequentist validity for Bayesian inference for multiple parametric functions. *Biometrika*, **83**, 287.
- Datta, G.S. and Ghosh, J.K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, **82**, 37-45.
- Datta, G.S. and Ghosh, M. (1995). Some remarks on noninformative priors *J. Amer. Statist. Assoc.* **90**, 1357-1363.
- Datta, G.S. and Ghosh, M. (1996). On the invariance of noninformative priors. *Ann. Statist.* **24** 141- 159.

- Datta, G.S., Ghosh, M. and Mukerjee, R. (2000). Some new results on probability matching priors. *Calcutta Statist. Assoc. Bull.*, **50**, 179–192.
- Datta, G. S. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. New York: Springer.
- Geisser, S. (1979). Comments on “Reference posterior distributions for Bayesian inference”, by J. Bernardo. *J. Roy. Statist. Soc., B*, **41**, 136-137.
- Geisser, S. (1984). On prior distributions for binary trials. *American Statisticians*, **38**, 244-251.
- Ghosh, J.K. (1994). *Higher Order Asymptotics*. Institute of Mathematical Statistics and American Statistical Association, Hayward, California, USA.
- Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer.
- Ghosh, J.K. and Mukerjee, R. (1992). Noninformative priors (with discussion). *Bayesian Statistics*, **4**, Eds. J.M. Bernardo, J.O. Berger, A.P Dawid and A.F.M. Smith. Oxford Univ. Press, 195-210.
- Ghosh, J. K. and Mukerjee, R. (1995). Frequentist validity of highest posterior density regions in the presence of nuisance parameters. *Statist. Dec.* **13**, 131–139.
- Ghosh, M., Sen, P.K., and Mukhopadhyay, N. (1997). *Sequential Estimation*. New York: Wiley.
- Govindarajulu, Z. (1981). *The sequential statistical analysis of hypothesis testing, point and interval estimation, and decision theory*, Columbus, OH: American Science Press.
- Hall, W.J. (1992). *A Course in Sequential Analysis*. Unpublished Lecture Notes, University of Rochester, Rochester, NY.
- Hu, I. (1988). Repeated significance tests for exponential families. *Ann. Statist.* **16** 1643-1666.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press.
- Peers, H.W. (1965). On confidence sets and Bayesian probability points in the case of several parameters. *J. Royal Statist. Soc., Ser. B* **27** 9-16.
- Polson, N. and Roberts, G. (1993). A utility based approach to information for stochastic differential equations. *Stochastic Processes and their Applications*, **48**, 341–356.
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer-Verlag.

- Sivaganesan, S. and Lingam, R. (2002). Bayes Factors for model selection with diffusion processes under improper priors. *The Annals of Institute of Statistical Mathematics*, **54**, 500–516.
- Sun, D. (1994). Integrable expansions for posterior distributions for a two-parameter exponential family. *The Annals of Statistics*, **22**, 1808-1830.
- Sun, D. and Ye, K. (1996). Frequentist validity of posterior quantiles for a two-parameter exponential family. *Biometrika*, **83**, 55-65.
- Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, **76** 604-608.
- Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Royal Statist. Soc., Ser. B* **31** 80-88.
- Welch, B. N. and Peers, B. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* **35** 318-329.
- Woodroffe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.
- Ye, K. (1993). Reference priors when the stopping rule depends on the parameter of interest. *J. Amer. Statist. Assoc.* **88** 360-363.