Bayesian Inferences and Nonlinear Functions of Poisson Rates

L. Raubenheimer

Department of Statistics, Rhodes University, Grahamstown, South Africa

e-mail: *L.Raubenheimer@ru.ac.za*

and

A.J. van der Merwe

Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa

---

Key words: Bayesian intervals, Coverage probabilities, Jeffreys' prior, Probability matching prior, Product of Poisson rates, Ratio of Poisson rates, Reference prior, Uniform prior, Weighted Monte Carlo method.

Summary: In this paper the probability matching prior for the product of different powers of k Poisson rates is derived. This is achieved by using the differential equation procedure of Datta and Ghosh (1995). The reference prior for the ratio of two Poisson rates is also obtained. Simulation studies are done to compare different methods for constructing Bayesian confidence intervals. It seems that if one is interested in making Bayesian inferences on $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$ and if $\alpha_i \geq 0$ $(i = 1, 2, \ldots, k)$ then the probability matching prior is the best. If on the other hand we want to obtain point estimates, credibility intervals or do hypotheses testing about $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates, then the uniform prior must be used.

---

1 Introduction

The Poisson distribution is often used as a probability model to describe the occurrence of rare events. For example the number of defects in items randomly selected from a production process may follow a Poisson distribution. Also the number of misprints counted on the first four pages of an early draft of a scientific paper. Events may also occur over time such as the number of radio-active decays in a fixed time interval, the number of injuries during a rugby match and the number of overseas telephone calls per hour. Research has been done improving statistical inferences on Poisson data. Methods for computing point and interval estimates of a single Poisson rate are for example discussed in Hald (1952), Guenther (1973) and Agresti and Coull (1998). Barker (2002) also made an attempt to find approximate confidence intervals for a single Poisson rate.

Our interest is to make Bayesian inferences on nonlinear functions of Poisson rates. Kim (2006) derived a non informative (probability matching) prior for $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$, the product of different powers of k Poisson rates, thereby obtaining approximate point and Bayesian confidence (credibility) intervals of the reliability of systems of k independent parallel components. In two sample situations it may be of interest to test or to construct credibility intervals for the ratio of two Poisson rates. Price and Bonett (2000) used non informative priors for small and large values of λ_j $(j = 1, 2)$ to construct credibility intervals for $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates. According to them these improper priors worked well. They also mentioned that the Jeffreys' prior has the advantage of adding 0.5 to the sample data which will avoid the problem of sampling zeros. We, however, tend to differ from Price and Bonett on the Jeffreys prior for the ratio of two Poisson rates. Our main purpose of this note is therefore to obtain probability matching priors for nonlinear functions of Poisson rates. The reference prior for the ratio of two Poisson rates will also be derived.

2 Probability Matching and Reference Prior

The Bayesian paradigm emerges as attractive in many types of statistical problems, especially in Poisson problems, but the choice of an appropriate non informative prior distribution has been controversial. As will be seen later, common non informative priors in multiparameter problems, such as Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior distribution. Datta and Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to $O\left(n^{-1}\right)$, where n is the sample size. They proved that the agreement between the posterior probability and the frequentist probability holds if and only if $\sum_{i=1}^{k} \frac{\partial}{\partial \lambda_i} \left\{\eta_i\left(\underline{\lambda}\right) \pi\left(\underline{\lambda}\right)\right\} = 0$, where $\pi\left(\underline{\lambda}\right)$ is the probability matching prior for $\underline{\lambda} = \left[\lambda_1\, \lambda_2 \ldots \lambda_k\right]'$, the vector of unknown parameters. Let $\nabla_t\left(\underline{\lambda}\right) = \left[\begin{array}{ccc} \frac{\partial}{\partial \lambda_1} t\left(\underline{\lambda}\right) & \cdots & \frac{\partial}{\partial \lambda_k} t\left(\underline{\lambda}\right) \end{array}\right]'$, where $t\left(\underline{\lambda}\right)$ is a nonlinear function of Poisson parameters, then $\eta\left(\underline{\lambda}\right) = \frac{F^{-1}\left(\underline{\lambda}\right)\nabla_t\left(\underline{\lambda}\right)}{\sqrt{\nabla_t'\left(\underline{\lambda}\right)F^{-1}\left(\underline{\lambda}\right)\nabla_t\left(\underline{\lambda}\right)}} = \left[\begin{array}{ccc} \eta_1\left(\underline{\lambda}\right) & \cdots & \eta_k\left(\underline{\lambda}\right) \end{array}\right]'$. Where $F^{-1}\left(\underline{\lambda}\right)$ is the inverse of $F\left(\underline{\lambda}\right)$, the Fisher information matrix of $\underline{\lambda}$. Reasons for using the probability matching prior is that it provides a method of constructing accurate frequentist intervals and it could also be useful for comparative purposes in Bayesian analysis. From Wolpert (2004), Berger states that frequentist reasoning will play an important role in finally obtaining good general priors for estimation and prediction. Some statisticians argue that frequency calculations are an important part of applied Bayesian statistics. (See Rubin, 1984).

The Jeffreys' and probability matching priors are but two methods to obtain useful non informative priors. As mentioned, the Jeffreys' prior is not always suitable for multiparameter problems. In recognition of this problem, Berger and Bernardo (1992) proposed the reference prior approach to the development of non informative priors, the key feature of which was a possible dependence of the reference prior on specification of parameters of interest and nuisance parameters. As mentioned by Pearn and Wu (2005) the reference prior maximises the difference in information about the parameter provided by the prior and posterior, the reference prior is derived in such a way that it provides as little information as possible about the parameter. In this section the reference prior of Berger and Bernardo (1992) will be derived for the ratio of two Poisson rates. As in the case of the Jeffreys' prior, the reference prior method is derived from the Fisher information matrix. The reference priors depend on the group ordering of the parameters. Berger and Bernardo (1992) recommended the reference prior based on having each parameter in its own group, i.e. having each conditional reference prior to be one dimensional.

The parameter $\psi = \theta/(\prod_{i=1}^{k} n_i^{\alpha_i})$ where $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$ is the product of different powers of k Poisson rates and appears in applications to system reliability. In Section 3 the probability matching prior for θ will therefore be derived as well as the reference prior for the ratio λ_1/λ_2 and in Section 4 a weighted Monte Carlo simulation method is described for obtaining credibility intervals in the case of the probability matching prior. This method is especially suitable for computing Bayesian confidence intervals, since only the kernel of the posterior distribution of the parameter is needed. In Section 5 simulation results are given for $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$ and it shows that if $\alpha_i \geq 0$ $(i = 1,2,\dots,k)$ the probability matching prior is an improvement on the Jeffreys' and uniform priors for obtaining point estimates and credibility intervals of θ. In Section 6 simulation results are given for $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson parameters. From the simulation results it becomes clear that the uniform prior is the best for making inferences about the ratio. The coverage probabilities obtained from the Jeffreys' (probability matching and reference) priors are reasonable, but the average interval lengths and the variances of the interval lengths are much too large.

3 The Ratio and Product of Poisson Rates

In this section we will firstly derive the probability matching prior for the general case where the parameter of interest is $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$. The special case $\theta_1 = \lambda_1/\lambda_2 = \nu$ $(\alpha_1 = 1, \alpha_2 = -1$ and $\alpha_3 = \alpha_4 = \dots = \alpha_k = 0)$ is also of interest. In the last part of this section the reference prior for $\nu = \lambda_1/\lambda_2$ will be derived. Kim (2006) derived the probability matching prior for the case where $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$, Kim however used a different method to derive the probability matching prior, Kim's proof is based on the method of Tibshirani (1989). Our proof is based on the procedure of Datta and Ghosh (1995).

Consider a sample from k populations. Let X_i be an observation from population i. Then $X_1, X_2, \dots, X_k$ will be independent Poisson distributions such that $X_i \sim P(\lambda_i)$, for $i = 1,2,\dots,k$. Where λ_i is the expected number of events per unit sample.

Theorem 1. *The probability matching prior for* $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$ *is given by*

$$\pi_{PM}(\underline{\lambda}) \propto \left(\sum_{i=1}^{k} \lambda_i^{-1} \alpha_i^2 \right)^{\frac{1}{2}}. \tag{1}$$

Proof. The Fisher information matrix is well known , the inverse of the Fisher information matrix is then given b

$$F^{-1}(\underline{\lambda}) = diag\begin{bmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_k \end{bmatrix}.$$

We are interested in a probability matching prior for $t(\underline{\lambda}) = \theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$.

Now

$$\begin{aligned} \nabla_t'(\underline{\lambda}) &= \begin{bmatrix} \frac{\partial t(\underline{\lambda})}{\partial \lambda_1} & \frac{\partial t(\underline{\lambda})}{\partial \lambda_2} & \cdots & \frac{\partial t(\underline{\lambda})}{\partial \lambda_k} \end{bmatrix} \\ &= \begin{bmatrix} \alpha_1 \lambda_1^{\alpha_1 - 1} \prod_{j \neq 1}^{k} \lambda_j^{\alpha_j} & \alpha_2 \lambda_2^{\alpha_2 - 1} \prod_{j \neq 2}^{k} \lambda_j^{\alpha_j} & \cdots & \alpha_k \lambda_k^{\alpha_k - 1} \prod_{j \neq k}^{k} \lambda_j^{\alpha_j} \end{bmatrix}. \end{aligned}$$

Also

$\nabla_t'(\underline{\lambda}) F^{-1}(\underline{\lambda}) = \left(\prod_{i=1}^{k} \lambda_i^{\alpha_i} \right) \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_k \end{bmatrix}$

and

$\nabla_t'(\underline{\lambda}) F^{-1}(\underline{\lambda}) \nabla_t(\underline{\lambda}) = \left(\prod_{i=1}^{k} \lambda_i^{\alpha_i} \right)^2 \sum_{i=1}^{k} \alpha_i^2 \lambda_i^{-1}$.

Define

$$\begin{aligned} \eta'(\underline{\lambda}) &= \frac{\nabla_t'(\underline{\lambda}) F^{-1}(\underline{\lambda})}{\sqrt{\nabla_t'(\underline{\lambda}) F^{-1}(\underline{\lambda}) \nabla_t(\underline{\lambda})}} \\ &= \begin{bmatrix} \eta_1(\underline{\lambda}) & \eta_2(\underline{\lambda}) & \cdots & \eta_k(\underline{\lambda}) \end{bmatrix} \end{aligned}$$

where $\eta_i(\underline{\lambda}) = \frac{\alpha_i}{\sqrt{\sum_{i=1}^{k} \alpha_i^2 \lambda_i^{-1}}}$ $(i = 1, 2, \ldots, k)$.

The prior $\pi(\underline{\lambda})$ is a probability matching prior if and only if the differential equation $\sum_{i=1}^{k} \frac{\partial}{\partial \lambda_i} \{ \eta_i(\underline{\lambda}) \pi(\underline{\lambda}) \} = 0$ is satisfied.

The differential equation will be satisfied if $\pi(\underline{\lambda})$ is

$$\pi_{PM}(\underline{\lambda}) \propto \left\{ \sum_{i=1}^{k} \alpha_i^2 \lambda_i^{-1} \right\}^{\frac{1}{2}}.$$

□

If we substitute $h = 0$ and $s = 1/2$ in equation (3.1) of Kim (2006), the two priors are identical. When using the probability matching prior, the posterior distribution of $\underline{\lambda}$ is given by

$$\pi_{PM}(\underline{\lambda} | x_1, x_2, \ldots, x_k) \propto \left\{ \sum_{i=1}^{k} \alpha_i^2 \lambda_i^{-1} \right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}. \tag{2}$$

It is easy to prove (see Kim, 2006) that equation 2 is a proper posterior distribution.

The Jefferys' prior, π_J, is given by

$$\pi_J(\underline{\lambda}) \propto |F(\underline{\lambda})|^{\frac{1}{2}} = \left(\prod_{i=1}^{k} \lambda_i \right)^{-\frac{1}{2}}. \tag{3}$$

Where $F(\underline{\lambda})$ is the information matrix connected with the likelihood function.

When using the Jeffreys' prior, the posterior distribution of $\underline{\lambda}$ is given by

$$\pi_J\left(\underline{\lambda}\,|x_1,x_2,\ldots,x_k\right) \propto \left(\prod_{i=1}^{k}\lambda_i\right)^{-\frac{1}{2}}\prod_{i=1}^{k}\lambda_i^{x_i}e^{-\lambda_i} = \prod_{i=1}^{k}\lambda_i^{x_i-\frac{1}{2}}e^{-\lambda_i}. \tag{4}$$

The posterior distribution of $\underline{\lambda}$ is thus the product of k independently distributed $Gamma\left(x_i+\frac{1}{2},1\right)$ variates.

The Uniform prior, π, is given by

$$\pi_U\left(\underline{\lambda}\right) \propto \text{constant}. \tag{5}$$

When using the Uniform prior, the posterior distribution of $\underline{\lambda}$ is given by

$$\pi_U\left(\underline{\lambda}\,|x_1,x_2,\ldots,x_k\right) \propto \prod_{i=1}^{k}\lambda_i^{x_i}e^{-\lambda_i}. \tag{6}$$

The posterior distribution of $\underline{\lambda}$ is thus the product of k independently distributed $Gamma\left(x_i+1,1\right)$ variates.

Theorem 2. *The posterior distribution for the ratio $\nu = \lambda_1/\lambda_2$ in the case of the probability matching prior is given by*

$$\pi_{PM}\left(\nu\,|x_1,x_2\right) \;\propto\; \frac{1}{B\left(x_1+\frac{1}{2},x_2+\frac{1}{2}\right)}\nu^{x_1-\frac{1}{2}}\left(\frac{1}{\nu+1}\right)^{x_1+x_2+1} \tag{7}$$

$$\nu > 0$$

a Beta distribution of the second kind.

Proof. If $\alpha_1 = 1$, $\alpha_2 = -1$ and $\alpha_3 = \alpha_4 = \ldots = \alpha_k = 0$, it easily follows from equation 1 that the probability matching prior in the case of $\nu = \lambda_1/\lambda_2$ is given by

$$\pi_{PM}\left(\lambda_1,\lambda_2\right) \;\propto\; \frac{\left(\lambda_1+\lambda_2\right)^{\tau}}{\lambda_1^{\frac{1}{2}}\lambda_2^{\frac{1}{2}}}. \tag{8}$$

where τ can take on any value. Using equation 8, the joint posterior distribution of λ_1 and λ_2 is given by

$$\pi_{PM}\left(\lambda_1,\lambda_2\,|x_1,x_2\right) \;\propto\; \frac{\left(\lambda_1+\lambda_2\right)^{\tau}}{\lambda_1^{\frac{1}{2}}\lambda_2^{\frac{1}{2}}}e^{-\left(\lambda_1+\lambda_2\right)}\lambda_1^{x_1}\lambda_2^{x_2}$$

Let $\nu = \lambda_1/\lambda_2$, thus $\lambda_1 = \nu\lambda_2$ and $d\lambda_1 = \lambda_2 d\nu$, then

$$\pi_{PM}\left(\nu,\lambda_2\,|x_1,x_2\right) \;\propto\; \nu^{x_1-\frac{1}{2}}\left(1+\nu\right)^{\tau}\lambda_2^{x_1+x_2+\tau}e^{-\lambda_2(1+\nu)}$$

and

$$\begin{aligned}\pi_{PM}(\nu\,|x_1,x_2) &= \int_0^\infty \pi_{PM}(\nu,\lambda_2\,|x_1,x_2)\,d\lambda_2 \\ &= C\nu^{x_1-\frac{1}{2}}\left(\frac{1}{\nu+1}\right)^{x_1+x_2+1} \qquad (9)\end{aligned}$$

which is a Beta distribution of the second kind and $C=\frac{1}{B\left(x_1+\frac{1}{2},x_2+\frac{1}{2}\right)}$. □

Corollary. *The Jeffreys' and probability matching priors for* $\nu=\lambda_1/\lambda_2$ *have the same posterior distribution. As will be seen from the next Theorem, equation 7 is also the posterior distribution for the reference prior of* $\nu=\lambda_1/\lambda_2$ *.*

Theorem 3. *The reference prior of* $\nu=\lambda_1/\lambda_2$ *for the group ordering* $\{\lambda_1,\lambda_2\}$ *is given by*

$$\pi_R(\lambda_1,\lambda_2) \propto \left\{\frac{1}{\lambda_1\lambda_2(\lambda_1+\lambda_2)}\right\}^{\frac{1}{2}}. \qquad (10)$$

Proof. By making a transformation we will, first derive the reference prior, $\pi_R(\nu,\lambda_2)$. The Fisher information matrix $F(\nu,\lambda_2)=A'F(\lambda_1,\lambda_2)A$ where

$$A = \frac{\partial(\lambda_1,\lambda_2)}{\partial(\nu,\lambda_2)} = \begin{bmatrix}\lambda_2 & \nu \\ 0 & 1\end{bmatrix}.$$

Therefore

$$\begin{aligned}F(\nu,\lambda_2) &= \begin{bmatrix}\lambda_2 & 0 \\ \nu & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\nu\lambda_2} & 0 \\ 0 & \frac{1}{\lambda_2}\end{bmatrix}\begin{bmatrix}\lambda_2 & \nu \\ 0 & 1\end{bmatrix} \\ &= \begin{bmatrix}\frac{\lambda_2}{\nu} & 1 \\ 1 & \frac{(\nu+1)}{\lambda_2}\end{bmatrix}.\end{aligned}$$

Now

$$h_1 = \left|\frac{\lambda_2}{\nu}-\frac{\lambda_2}{(\nu+1)}\right| = \lambda_2\left(\frac{1}{\nu(\nu+1)}\right)$$

and

$$\begin{aligned}\pi_R(\nu) &= |h_1|^{\frac{1}{2}} \propto \frac{1}{\nu^{\frac{1}{2}}(\nu+1)^{\frac{1}{2}}} \\ h_2 &= \left|\frac{1}{\lambda_2}(\nu+1)\right|.\end{aligned}$$

Therefore

$$\pi_R(\lambda_2\,|\nu) = |h_2|^{\frac{1}{2}} \propto \left(\frac{1}{\lambda_2}\right)^{\frac{1}{2}}$$

The joint prior for the group ordering $\{\nu, \lambda_2\}$ is given by

$$\pi_R(\nu, \lambda_2) = \pi_R(\nu)\,\pi_R(\lambda_2|\nu) \propto \left(\frac{1}{\lambda_2}\right)^{\frac{1}{2}} \left(\frac{1}{\nu(\nu+1)}\right)^{\frac{1}{2}}$$

and the joint reference prior for the group ordereing $\{\lambda_1, \lambda_2\}$ is given by

$$\pi_R(\lambda_1, \lambda_2) \propto \left\{\frac{1}{\lambda_1\lambda_2(\lambda_1+\lambda_2)}\right\}^{\frac{1}{2}}.$$

□

From equation 10 it follows that the reference prior is also a probability matching prior.

4 The Weighted Monte Carlo Method in the Case of $\pi_{PM}(\underline{\lambda}) \propto \left(\sum_{i=1}^{k} \lambda_i^{-1}\alpha_i^2\right)^{\frac{1}{2}}$, the Probability Matching Prior for $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$

In this section a weighted Monte Carlo method is described which will be used for simulation from the posterior distribution in the case of the probability matching prior. This method is especially suitable for computing Bayesian confidence (credibility) intervals. It does not require knowing the closed form of the marginal posterior distribution of θ, only the kernel of the posterior distribution of $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ is needed.

As mentioned by Chen and Shao (1999), Kim (2006), Smith and Gelfand (1992), Guttman and Menzefrieke (2003), Skare et al. (2003) and Li (2007) the weighted Monte Carlo (sampling - importance re-sampling) algorithm aims at drawing a random sample from a target distribution π, by first drawing a sample from a proposal distribution q, and from this a smaller sample is drawn with sample probabilities proportional to the importance ratios π/q. For the algorithm to be efficient, it is important that q is a good approximation for π. This means that q should not have too light tails when compared to π. For further details see Skare et al. (2003). In the case of credibility intervals it is not even necessary to draw the smaller sample. The weights (sample probabilities) are however important.

If a uniform prior is put on $\underline{\lambda}$, the posterior (proposal) distribution is
$q(\underline{\lambda}|data) \propto \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}$
In the case of the probability matching prior, the posterior (target) distribution is
$\pi_{PM}(\underline{\lambda}|data) \propto \left\{\sum_{i=1}^{k} \alpha_i^2 \lambda_i^{-1}\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}$
The sample probabilities are therefore proportional to

$$\frac{\pi_{PM}(\underline{\lambda}|data)}{q(\underline{\lambda}|data)} = \left\{\sum_{i=1}^{k} \alpha_i^2 \lambda_i^{-1}\right\}^{\frac{1}{2}}$$

and the normalized weights are

$$\omega_l = \frac{\left\{\sum_{i=1}^{k}\left(\alpha_i^2\lambda_i^{-1}\right)^{(l)}\right\}^{\frac{1}{2}}}{\sum_{l=1}^{n}\left\{\sum_{i=1}^{k}\left(\alpha_i^2\lambda_i^{-1}\right)^{(l)}\right\}^{\frac{1}{2}}} \qquad l = 1,2,\ldots,n.$$

A straightforward way of doing the weighted Monte Carlo (WMC) method was proposed by Chen and Shao (1999). Details of the Monte Carlo method are as follows:

Step 1 Obtain a Monte Carlo sample $\left\{\left(\lambda_1^{(l)},\lambda_2^{(l)}\ldots,\lambda_k^{(l)}\right);l = 1,2,\ldots,n\right\}$ from the proposal distribution $q(\underline{\lambda}\,|data)$ and calculate $\theta^{(l)} = \prod_{i=1}^{k}\left(\lambda_i^{\alpha_1}\right)^{(l)}$ for $l = 1,2,\ldots,n$.

Step 2 Sort $\left\{\theta^{(l)},(l = 1,2,\ldots,n)\right\}$ to obtain the ordered values $\theta^{[1]} \leq \theta^{[2]} \leq \cdots \leq \theta^{[n]}$.

Step 3 Each simulated θ value has an associated weight. Therefore compute the weighted function $\omega_{(l)}$ associated with the $l-$th ordered $\theta^{[l]}$ value.

Step 4 Add the weights up from left to right (from the first on) until one gets $\sum_{l=1}^{n_1}\omega_{(l)} = \alpha/2$. Write down the corresponding $\theta^{[n_1]}$ value and denote it as $\theta_{(\alpha/2)}$. Add the weights up from right to left (from the last back) until one gets $\sum_{l=n_2}^{n}\omega_{(l)} = \alpha/2$. Write down the corresponding $\theta^{[n_2]}$ value and denote it as $\theta_{(1-\alpha/2)}$.

Step 5 The $100(1-\alpha)\%$ Bayesian confidence interval is:

$\left(\theta_{(\alpha/2)},\theta_{(1-\alpha/2)}\right)$.

5 Simulation Studies

As mentioned, the parameter $\theta = \prod_{i=1}^{k}\lambda_i^{\alpha_i}$, the product of different powers of k Poisson parameters appears in applications to system reliability. If a system consists of k components in parallel, then the probability of system failure is $\psi = \prod_{i=1}^{k}\left(\frac{\lambda_i}{n_i}\right)^{\alpha_i}$ where $p_i = \frac{\lambda_i}{n_i}$ is the probability that the $i-$th component will fail. Also if a system requires that at least one of each of k types of components must be employed and that these components are needed in parallel, then the probability of failure of an m-component system is $\psi = \prod_{i=1}^{k}\left(\frac{\lambda_i}{n_i}\right)^{\alpha_i}$, where $k < m$, α_i is the number of components of type i and $\sum_{j=1}^{k}\alpha_j = m$.

From a Bayesian perspective a prior is needed for the parameter ψ. As mentioned common non-informative priors in multiparameter problems such as Jeffreys' priors can have features that have an unexpectedly dramatic effect on the posterior distribution. It is for this reason that the probability matching prior for $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$ was be derived in Theorem 1.

Also as mentioned a probability matching prior is a prior distribution under which the posterior probabilities match their coverage probabilities. The fact that the resulting Bayesian posterior intervals of level $1-\alpha$ are also good frequentist confidence intervals at the same level is a very desirable situation. See also Bayarri and Berger (2004) and Severini et al. (2002) for general discussion.

5.1 Simulation Study I

In Table 1 the frequentist coverage probabilities are given for $\theta = \prod_{i=1}^{k} \lambda_i$ in the case of:

1. the Jeffreys' prior, $\pi_J(\underline{\lambda}) \propto \left(\prod_{i=1}^{k} \lambda_i\right)^{-\frac{1}{2}}$;
2. the uniform prior, $\pi_U(\underline{\lambda}) \propto$ constant;
3. the probability matching prior, $\pi_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} \lambda_i^{-1}\right\}^{\frac{1}{2}}$;

50 000 samples were generated and from each sample 10 000 parameter values were simulated to obtain the Bayesian confidence intervals in the case of the Jeffreys' and uniform priors. For the probability matching prior only 20 000 samples were generated.

From the simulation results in Table 1 it is clear that the probability matching prior is better than the Jeffreys' and uniform priors in most of the situations. As mentioned by Kim (2006) if each coordinate of the parameter vector $\underline{\lambda}$ is large, the frequentist coverage percentages obtained from using the probability matching prior is close to the desired level.

The simulation results are displayed in Figures 1 and 2. The inability of the Jeffreys' and uniform priors to give good coverage probabilities is even more clear from these graphs.

Table 1. Frequentist Coverage Probabilities for 5% and 95% Posterior Quantiles of $\theta = \prod_{i=1}^{k} \lambda_i$

		Jeffreys'		Uniform		Prob. Matching	
λ	θ	5%	95%	5%	95%	5%	95%
(1 1 1)	1	0.0223	0.9512	0.1037	1.0000	0.0551	1.0000
(1 2 3)	6	0.0270	0.9133	0.0849	0.9975	0.0514	0.9819
(2 2 2)	8	0.0243	0.9142	0.0775	1.0000	0.0491	0.9674
(1 5 10)	50	0.0346	0.9576	0.0876	0.9967	0.0532	0.9862
(5 5 5)	125	0.0352	0.9069	0.0612	0.9650	0.0484	0.9417
(10 10 10)	1 000	0.0325	0.9253	0.0588	0.9625	0.0462	0.9490
(1 2 3 4 5)	120	0.0192	0.8861	0.0806	0.9922	0.0500	0.9742
(2 2 3 4 5)	240	0.0194	0.8657	0.0721	0.9847	0.0518	0.9586
(3 3 3 4 5)	540	0.0198	0.8647	0.0666	0.9753	0.0482	0.9499
(1 2 3 4 5 6 7 8)	40 325	0.0141	0.8560	0.0756	0.9877	0.0503	0.9704
(1 2 3 4 5 5 5 5)	15 000	0.0131	0.8442	0.0704	0.9870	0.0514	0.9693
(5 5 5 5 5 5 5 5)	390 625	0.0164	0.8466	0.0591	0.9614	0.0477	0.9458
(5 5 5 5 6 7 8 9)	1 890 000	0.0178	0.8617	0.0587	0.9614	0.0499	0.9455
(5 5 5 5 10 10 10 10)	6 250 000	0.0179	0.8729	0.0582	0.9627	0.0472	0.9472
(10 10 10 10 10 10 10 10)	10^8	0.0184	0.8880	0.0534	0.9546	0.0492	0.9507

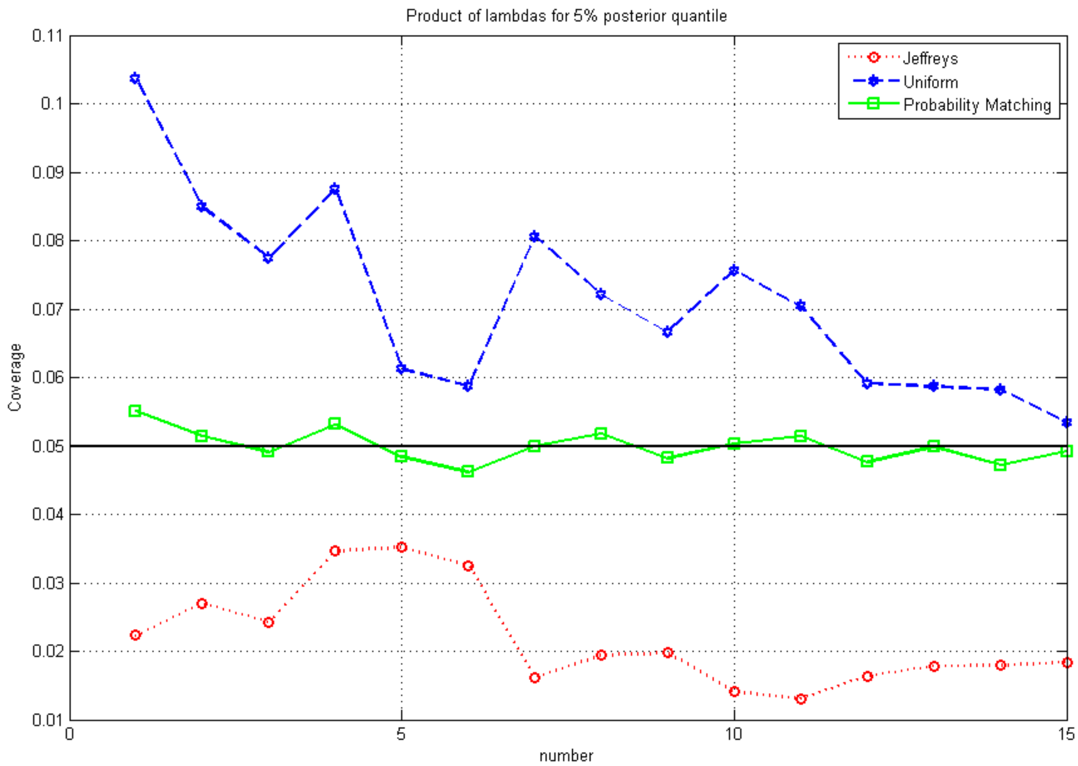

Figure 1. Illustration of the 5% Quantiles of $\theta = \prod_{i=1}^{k} \lambda_i$ in the same order as given in Table 1.

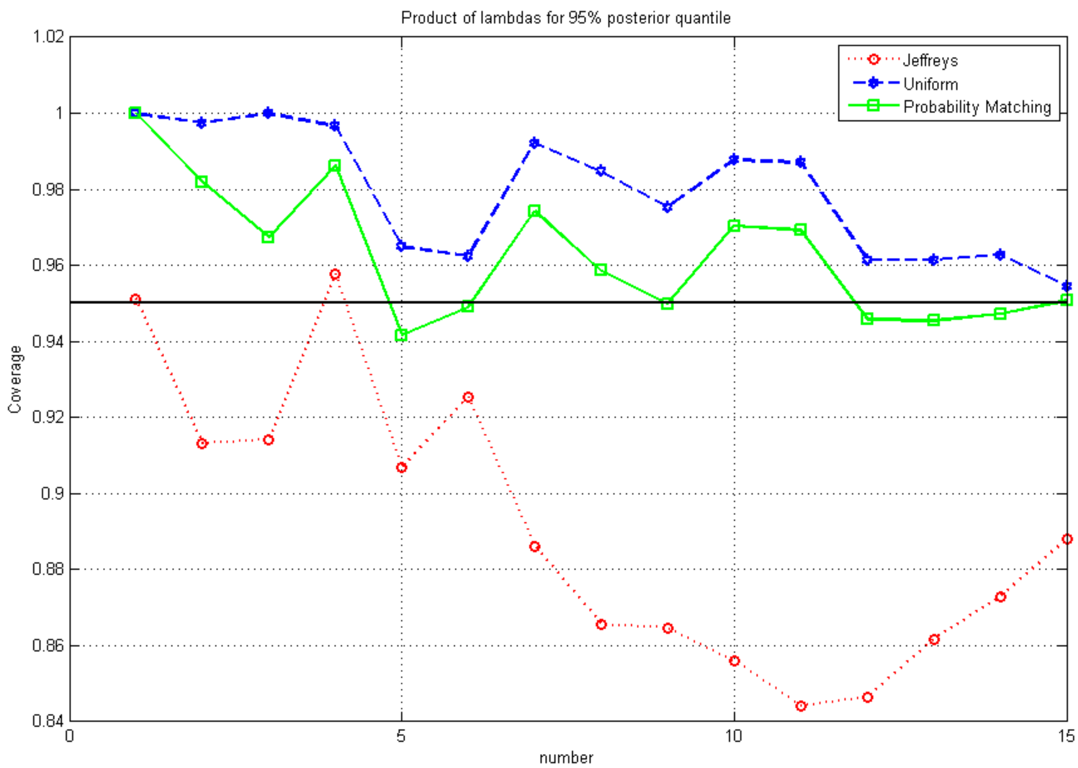

Figure 2. Illustration of the 95% Quantiles of $\theta = \prod_{i=1}^{k} \lambda_i$ in the same order as given in Table 1.

5.2 Simulation Study II - Comparing Six Priors for $\theta = \lambda_1\lambda_2$ - Reliability of Independent Parallel Components System

In this example a simulation study is done for $\theta = \lambda_1\lambda_2$, the product of two Poisson rates. The parameter values for the Poisson distributions are λ_i =2, 3, 4, 5, 6, 7, 8, 9, 10 (for $i = 1,2$).

The priors that will be compared are:

1. the uniform prior: $\pi_U(\lambda_1,\lambda_2) \propto$ constant;
2. the Jeffreys' prior: $\pi_J(\lambda_1,\lambda_2) \propto \prod_{i=1}^{2}\lambda_i^{-\frac{1}{2}}$;
3. the probability matching prior: $\pi_{PM}(\lambda_1,\lambda_2) \propto \left\{\sum_{i=1}^{2}\lambda_i^{-1}\right\}^{\frac{1}{2}}$;
4. $\pi(\lambda_1,\lambda_2) \propto \prod_{i=1}^{2}\lambda_i^{-\frac{3}{8}}$;
5. $\pi(\lambda_1,\lambda_2) \propto \prod_{i=1}^{2}\lambda_i^{-\frac{1}{4}}$;
6. $\pi(\lambda_1,\lambda_2) \propto \prod_{i=1}^{2}\lambda_i^{-\frac{1}{8}}$.

We know from experience (and this is also clear from Table 1) that the Jeffreys' prior under covers while the uniform prior tends to over cover in the case of the product of Poisson rates. Priors (4), (5) and (6) are in between priors for (1) and (2) and it is for this reason that they are included in this simulation study.

The frequentist coverage percentages of the 95% HPD (highest posterior density) intervals as well as the interval lengths are displayed in Figure 3. The graphs are averages over λ_1 for $\lambda_2 = 2$ to 10. The coverage percentage of the probability matching prior is much better than those of the Jeffreys' and uniform priors. It is also not impossible that a prior with $a = \frac{1}{5}$ or $a = \frac{1}{6}$ will give a very good coverage.

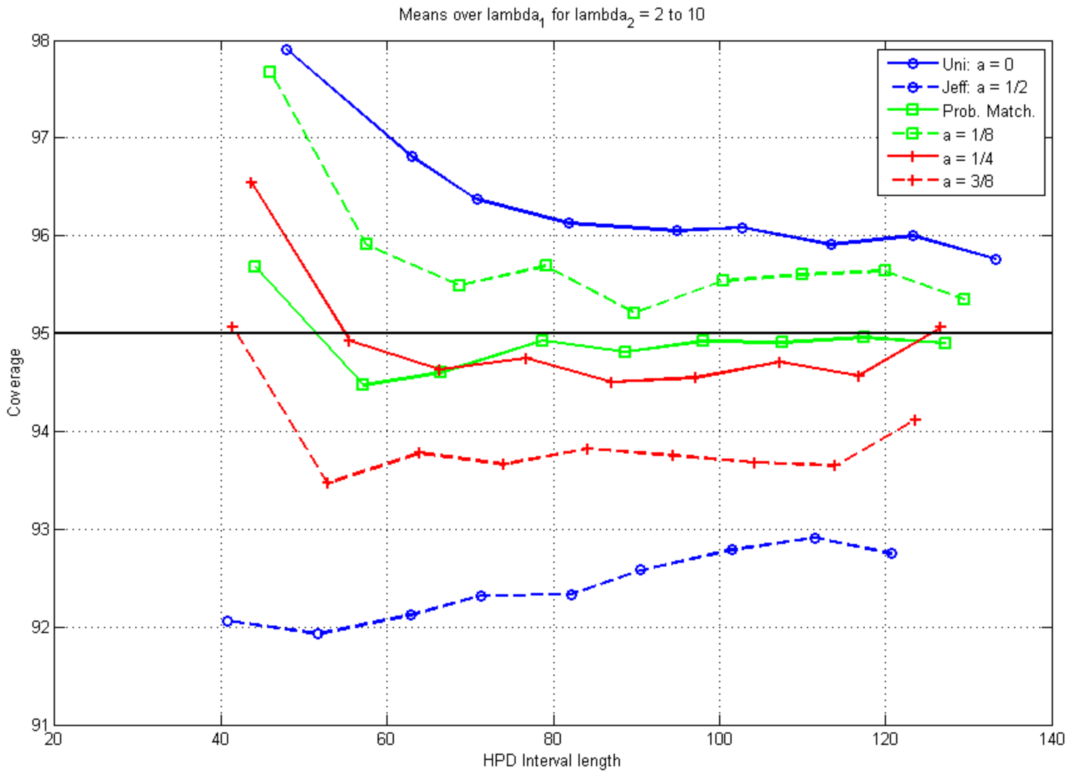

Figure 3. Illustration of the Coverage Percentages of the 95% HPD Intervals of $\theta = \lambda_1\lambda_2$.

5.3 Simulation Study III - Comparing Priors for $\theta_3 = \lambda_1^2\lambda_2$ and $\theta_4 = \lambda_1^3\lambda_2$ - Reliability of Repeated Components System

Assume a system needs three components in parallel and at least one of each of two types of components must be used. If the first component is replicated, then the probability of failure is $\psi_3 = p_1^2 p_2$. Also if four components are needed and the first component is replicated three times, then the probability of failure is $\psi_4 = p_1^3 p_2$ where $p_i = \frac{\lambda_i}{n_i}$ $(i = 1,2)$.For further details see Kim (2006). In this simulation study the coverage probabilities of these different priors for the parameters $\theta_3 = \lambda_1^2\lambda_2$ and $\theta_4 = \lambda_1^3\lambda_2$ will therefor be looked at. The parameter values for the Poisson distribution are as in Section 5.2, i.e. λ_i =2, 3, 4, 5, 6, 7, 8, 9, 10 (for $i = 1,2$) and the priors that will be compared are:

1. the uniform prior: $\pi_U(\lambda_1,\lambda_2) \propto$ constant;
2. the Jeffreys' prior: $\pi_J(\lambda_1,\lambda_2) \propto \prod_{i=1}^2 \lambda_i^{-\frac{1}{2}}$;
3. the probability matching prior: $\pi_{PM}(\lambda_1,\lambda_2) \propto \left\{\sum_{i=1}^2 \alpha_i^2 \lambda_i^{-1}\right\}^{\frac{1}{2}}$.

The frequentist coverage probabilities as well as the interval lengths of the 95% Bayesian confidence intervals for the above priors in the case of $\theta_3 = \lambda_1^2\lambda_2$ are given in Figure 4 and in Figure 5 the same graphs are given for $\theta_4 = \lambda_1^3\lambda_2$. The graphs are averages over λ_1 for $\lambda_2 = 2$ to 10.

The same patterns as in Figure 3 emerge from Figures 4 and 5, i.e. the Jeffreys' prior underestimates the coverage probabilities while the uniform prior tends to overestimate the coverage probabilities. In general the probability matching prior seems to give the best coverage probabilities.

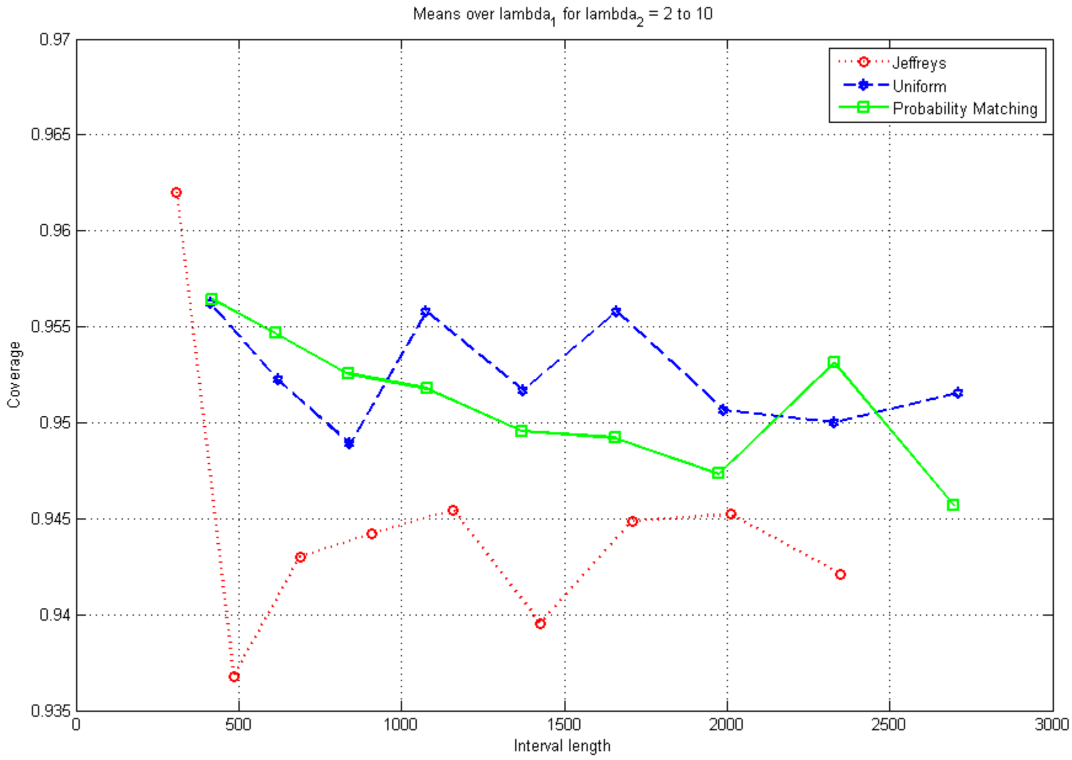

Figure 4. Illustration of the Coverage Probabilities of the 95% Bayesian Confidence Intervals for $\theta_3 = \lambda_1^2\lambda_2$.

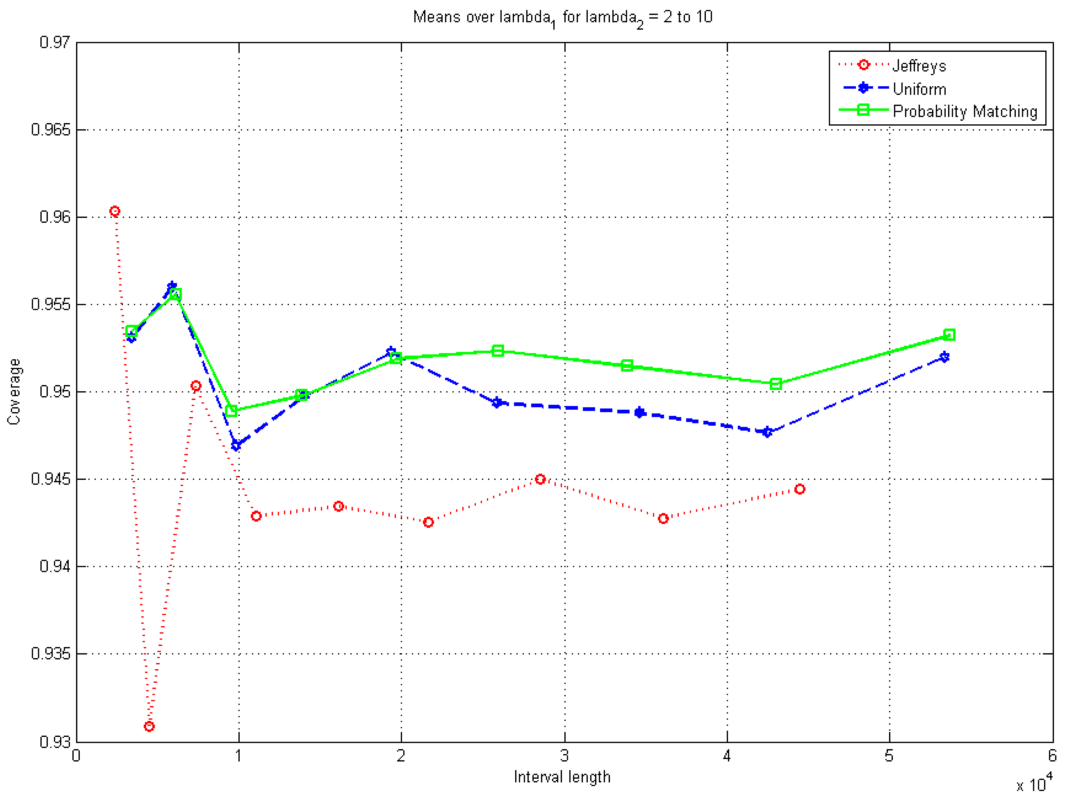

Figure 5. Illustration of the Coverage Probabilities of the 95% Bayesian Confidence Intervals for $\theta_4 = \lambda_1^3\lambda_2$.

6 Bayesian Confidence Intervals for $\nu = \lambda_1/\lambda_2$, the Ratio of Two Poisson Rates

The comparison of Poisson rates is of great interest in biological, agricultural and medical research. In two sample situations it may be of interest to test or to construct confidence intervals for the ratio of two Poisson rates. Gu et al. (2008) compared the proportions of four approaches for testing the ratio of two Poisson rates.

Price and Bonett (2004) on the other hand computed the exact coverage probabilities of the intervals of six classical methods and that of the Bayesian interval, using Jeffreys' prior, for small and large values of λ_j $(j = 1,2)$. They also looked at other plausible non informative priors for λ_j such as $\pi(\lambda_j) \propto \lambda_j^{-1}$ and $\pi(\lambda_j) \propto$ constant and mentioned that these priors work about as well as the Jeffreys' prior ($\pi(\lambda_j) \propto \lambda_j^{-\frac{1}{2}}$). According to them the Jeffreys' prior has the advantage of adding 0.5 to the sample data which would avoid the problem of sampling zeros. From their simulation studies they concluded that the non informative Bayesian intervals (using the Jeffreys' prior) is reasonable under classical evaluation. We however tend to differ from them, since we came to the conclusion that the Jeffreys' prior cannot be used for testing the ratio $\nu = \lambda_1/\lambda_2$ or obtaining confidence intervals, especially if λ_2 is small. A prior that can be used for these purposes is the uniform prior. This will become clear from the following simulation study.

In equation 9 it was shown that the posterior distribution of $\nu = \lambda_1/\lambda_2$ in the case of the Jeffreys', probability matching and reference priors is a Beta distribution of the second kind. This distribution can easily be transformed to an $F-$distribution with $2x_1 + 1$ and $2x_2 + 1$ degrees of freedom. In a similar way the posterior distribution of ν, using the uniform prior can be transformed to an $F-$distribution with $2x_1 + 2$ and $2x_2 + 2$ degrees of freedom. Bayesian confidence intervals and coverage probabilities for ν can therefore be calculated exactly.

6.1 Simulation Study IV - Exact Frequency Coverage Percentages for $\nu = \lambda_1/\lambda_2$

In this example a simulation study is done for $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates. The parameter values for the Poisson distributions are λ_1 =2, 5 and 10 and λ_2 =2, 3, 4, 5, 6, 7, 8, 9, 10.

The priors that will be compared are:

1. the uniform prior: $\pi_U(\lambda_1, \lambda_2) \propto$ constant;

2. the Jeffreys' prior: $\pi_J(\lambda_1, \lambda_2) \propto \prod_{i=1}^{2} \lambda_i^{-\frac{1}{2}}$.

In Table 2 the coverage percentages are given for the 95% Bayesian equal-tail intervals in the case of the Jeffreys' prior and in Table 3 the coverage percentages are given for the uniform prior. Ten thousand samples were generated for each parameter combination.

Table 2. Coverage Percentages of the 95% Bayesian Confidence Intervals for $\nu = \lambda_1/\lambda_2$ in the case of the Jeffreys' (Reference and Probability Matching) Priors. Coverage Percentage (a), Mean Length (b) and Standard Deviation (c).

$\downarrow\lambda_1$	$\lambda_2 \rightarrow$	2	3	4	5	6	7	8	9	10
2	(a)	95.92	94.16	94.46	94.72	95.24	95.20	95.60	95.72	95.30
	(b)	599.98	250.08	102.77	41.956	8.0429	3.7614	4.8311	1.5405	0.8661
	(c)	3.3e6	1.5e6	6.1e5	2.5e5	32752	8470	24634	1261.9	0.4405
5	(a)	95.40	94.08	94.38	94.94	94.76	94.30	94.32	95.06	94.40
	(b)	1381.8	599.82	208.86	87.282	23.004	13.445	6.6475	1.8303	1.5217
	(c)	1.5e7	7.2e6	2.5e6	1.0e6	2.4e5	1.1e5	46513	4.5629	1.5058
10	(a)	95.40	94.56	94.76	94.16	93.82	94.32	94.92	94.80	95.20
	(b)	2886.3	1054.9	3912.1	1727.3	58.639	13.765	10.434	11.167	2.4904
	(c)	5.7e7	2.3e7	8.9e6	3.5e6	1.2e6	2.3e5	2.1e5	3.2e5	3.5445

Table 3. Coverage Percentages of the 95% Bayesian Confidence Intervals for $\nu = \lambda_1/\lambda_2$ in the case of the Uniform Prior. Coverage Percentage (a), Mean Length (b) and Standard Deviation (c).

$\downarrow\lambda_1$	$\lambda_2 \rightarrow$	2	3	4	5	6	7	8	9	10
2	(a)	98.16	97.34	96.58	96.38	96.80	96.82	96.06	96.18	96.22
	(b)	21.572	10.130	5.4558	3.2781	2.1531	1.6649	1.3948	1.0562	0.9313
	(c)	1901.0	688.59	278.14	107.50	34.622	21.250	22.322	0.9839	3.4559
5	(a)	95.86	96.46	95.70	95.62	95.92	95.68	95.56	96.26	95.94
	(b)	42.687	20.474	10.341	6.1223	3.6942	2.7507	2.2019	1.7188	1.4644
	(c)	6813.1	2935.1	1075.7	415.72	83.159	38.849	38.722	2.0315	1.1217
10	(a)	96.22	95.76	96.00	95.76	94.84	95.28	95.54	95.90	95.44
	(b)	83.653	38.105	18.813	10.916	6.8133	4.9464	3.6860	3.0692	2.3601
	(c)	23576	9702.6	3713.5	1481.8	570.94	233.67	96.470	158.94	3.2491

From Table 2 it is clear that the coverage percentages of the Jeffreys' (reference and probability matching) priors are reasonably good (slight under coverage in some cases) but the mean lengths and standard deviations of the credibility (Bayesian confidence) intervals are much too large. This is especially true if λ_2 is small. The uniform prior on the other hand also give reasonably good coverage (slight over coverage) but the mean lengths and standard deviations of the credibility intervals are much smaller.

7 Conclusion

If one is interested in obtaining point estimates and Bayesian confidence intervals for $\theta = \prod_{i=1}^{k} \lambda_i^{\alpha_i}$, the product of different powers of Poisson rates and if $\alpha_i \geq 0$ $(i = 1, 2, \ldots k)$ then the probability matching prior is the best. If on the other hand we want to obtain point estimates, credibility intervals or do hypothesis testing about $\nu = \lambda_1/\lambda_2$, the ratio of two Poisson rates, then the uniform prior must be used.

Price and Bonett (2000) mentioned that the Jeffreys' prior has the advantage of adding 0.5 to the sample data which will avoid sampling zeros. From our research it seems that adding 1 to the sample data (using the uniform prior) gives better results for $\nu = \lambda_1/\lambda_2$.

References

Agresti, A. and Coull, B. (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician*, 52(2):119 – 126.

Barker, L. (2002). A Comparison of Nine Confidence Intervals for a Poisson Parameter When the Expected Number of Events is < 5. *The American Statistician*, 56(2):85–89.

Bayarri, M. J. and Berger, J. O. (2004). The Interplay of Bayesian and Frequentist Analysis. *Statistical Science*, 19(1):58–80.

Berger, J. and Bernardo, J. (1992). On the Development of Reference Priors. In *Bayesian Statistics, 4*. Oxford University Press.

Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo Estimation of Bayesian Credible and HPD Intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92.

Datta, G. and Ghosh, J. (1995). On Priors Providing Frequentist Validity for Bayesian Inference. *Biometrika*, 82(1):37–45.

Gu, K., Ng, H., Tang, M., and Schucany, W. (2008). Testing the Ratio of Two Poisson Rates. *Biometrical Journal*, 50(2):283 – 298.

Guenther, W. (1973). *Concepts of Statistical Inference*. McGraw-Hill.

Guttman, I. and Menzefrieke, U. (2003). Posterior Distributions for Functions of Variance Components. *Sociedad de Estadistica e Investigacio Operativa Test*, 12:115 – 123.

Hald, A. (1952). *Statistical Theory with Engineering Applications*. Wiley, New York.

Kim, H.-J. (2006). On Bayesian Estimation of the Product of Poisson Rates with Application to Reliability. *Communications in Statistics - Simulation and Computation*, 35(1):47 – 59.

Li, K.-H. (2007). Pool size selection for the sampling/importance resampling algorithm. *Statistica Sinica*, 17(3):895 – 907.

Pearn, W. and Wu, C. (2005). A Bayesian Approach for Assessing Process Precision Based on Multiple Samples. *European Journal of Operational Research*, 165(3):685 – 695.

Price, R. and Bonett, D. (2004). An Improved Confidence Interval for a Linear Function of Binomial Proportions. *Computational Statistics & Data Analysis*, 45(3):449 – 456.

Price, R. M. and Bonett, D. G. (2000). Estimating the Ratio of Two Poisson rates. *Computational Statistics and Data Analysis*, 34(3):345 – 356.

Rubin, D. (1984). Bayesian Justifiable and Relevant Frequency Calculation for the Applied Statistician. *The Annals of Statistics*, 12(4):1151 – 1172.

Severini, T., Mukerjee, R., and Ghosh, M. (2002). On An Exact Probability Matching Property of Right-invariant Priors. *Biometrika*, 89(4):952 – 957.

Skare, O., Bolviken, E., and Holden, L. (2003). Improved Sampling-Importance Resampling and Reduced Bias Importance Sampling. *Scandinavian Journal of Statistics*, 30(4):719–737.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian Statistics without Tears: A Sampling-Resampling Perspective. *The American Statistician*, 46(2):84–88.

Tibshirani, R. (1989). Non-informative Priors for One Parameter of Many. *Biometrika*, 76(3):604 – 608.

Wolpert, R. (2004). A Conversation with James O Berger. *Statistical Science*, 19(1):205 – 218.