Bayesian Estimation for Linear Functions of Poisson Rates

Lizanne Raubenheimer[a] and Abrie J. van der Merwe[b]

[a]Department of Statistics, Rhodes University, Grahamstown, South Africa
[b]Department of Mathematical Statistics and Actuarial Science, University of the Free State, Bloemfontein, South Africa

Abstract

In this paper the probability matching prior for a linear contrast of Poisson parameters is derived, this prior is extended in such a way that it is also the probability matching prior for the average of Poisson parameters. This research is an extension of the work done by Stamey & Hamilton (2006). A comparison is made between the confidence intervals obtained by Stamey & Hamilton (2006) and the intervals derived by us when using the Jeffreys' and probability matching priors. A weighted Monte Carlo method is used for the computation of the Bayesian confidence intervals, in the case of the probability matching prior. In the last section of this paper hypothesis testing about two Poisson means is considered. The power and size of the test, using Bayesian methods, is compared to tests used by Krishnamoorthy & Thomson (2004). For the Bayesian methods the Jeffreys' prior, probability matching prior and two other priors are used.

Keywords: Bayesian intervals, Poisson parameters, Power and size of test, Probability matching prior, Weighted Monte Carlo method.

1 Introduction

Research has been done on improving confidence intervals for discrete data. Barker (2002) made an attempt to find approximate confidence intervals for a single Poisson rate. Stamey & Hamilton (2006) considered three interval estimators for linear functions of Poisson rates, a Wald interval, a t interval with Satterwhaite's degrees of freedom and a Bayesian interval using non-informative priors. We are going to consider another Bayesian interval using a probability matching prior. The probability matching prior will be derived by using the method proposed by Datta & Ghosh (1995), they 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.

Krishnamoorthy & Thomson (2004) addressed the problem of hypothesis testing about two Poisson means. They compared the conditional test (C - test) to a test based on estimated p - values (E - test). Krishnamoorthy & Thomson (2004) considered the size and the power of these tests. We are going to use Bayesian methods, using the Jeffreys' prior, the probability matching prior, a prior which is proportional to $\lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, and a prior which is proportional to $\lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$. The results obtained from the Bayesian methods will be compared to the results obtained by Krishnamoorthy & Thomson (2004).

In the next section the probability matching prior for a linear contrast of Poisson parameters is derived. This prior will be extended in such a way that it can be used as the probability matching prior for the average of Poisson parameters. In Section 3 a weighted Monte Carlo simulation method is described to obtain Bayesian confidence intervals in the case of the probability matching prior, and in Section 4 an example and simulation results will be given and discussed.

2 Probability Matching Prior for a Linear Contrast of Poisson Parameters

Consider a sample from k populations. Let X_i be an observation from population i. Then $X_1, X_2, \ldots, X_k$ will be independent Poisson distributions such that $X_i \sim P(\lambda_i)$, for $i = 1, 2, \ldots, k$. Where λ_i is the expected number of events per unit sample. We assume that the interest is in a linear combination of Poisson rates. In general we can define such a linear function of Poisson parameters as $\xi = \sum_{i=1}^{k} a_i\lambda_i$, where a_i is the known coefficient value.

Theorem 1 The probability matching prior for $\xi = \sum_{i=1}^{k} a_i\lambda_i$, a linear contrast of Poisson parameters (i.e. $\sum_{i=1}^{k} a_i = 0$), is given by

$$\pi_{PM}(\underline{\lambda}) = \pi_{PM}(\lambda_1, \lambda_2, \ldots, \lambda_k) \propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}}. \tag{1}$$

Proof. The likelihood function is given by

$$L(\underline{\lambda}) = \prod_{i=1}^{k} e^{-\lambda_i} \frac{\lambda_i^{x_i}}{x_i!}.$$

The Fisher information matrix is well known , the inverse of the Fisher information matrix is then given by

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

We are interested in a probability matching prior for $t(\underline{\lambda}) = \xi = \sum_{i=1}^{k} a_i\lambda_i$, a linear contrast of Poisson parameters, where $\sum_{i=1}^{k} a_i = 0$.

Now

$$\begin{aligned} \nabla_t'(\underline{\lambda}) &= \begin{bmatrix} \frac{\partial t(\underline{\lambda})}{\partial \lambda_1} & \frac{\partial t(\underline{\lambda})}{\partial \lambda_2} & \cdots & \frac{\partial t(\underline{\lambda})}{\partial \lambda_k} \end{bmatrix} \\ &= \begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}. \end{aligned}$$

Also

$\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda}) = \left[\begin{array}{cccc} a_1\lambda_1 & a_2\lambda_2 & \cdots & a_k\lambda_k \end{array}\right]$

and

$\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda}) = \sum\limits_{i=1}^{k} a_i^2\lambda_i.$

Define

$$\begin{aligned} \eta'(\underline{\lambda}) &= \frac{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})}{\sqrt{\nabla_t'(\underline{\lambda})F^{-1}(\underline{\lambda})\nabla_t(\underline{\lambda})}} \\ &= \left[\begin{array}{cccc} \eta_1(\underline{\lambda}) & \eta_2(\underline{\lambda}) & \cdots & \eta_k(\underline{\lambda}) \end{array}\right] \end{aligned}$$

where $\eta_i(\underline{\lambda}) = \frac{a_i\lambda_i}{\sqrt{\sum\limits_{i=1}^{k} a_i^2\lambda_i}}$ $\qquad (i = 1,2,\ldots,k).$

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

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

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

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

$$\pi_{PM}(\underline{\lambda}|data) \propto \left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i} e^{-\lambda_i}. \tag{2}$$

Corollary 1.

If $\sum_{i=1}^{k} a_i \neq 0$, the following equation can be used for a probability matching prior

$$\widetilde{\pi}_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{-1} \tag{3}$$

and the posterior distribution of $\underline{\lambda}$ will be

$$\widetilde{\pi}_{PM}(\underline{\lambda}|data) \propto \left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{x_i-1} e^{-\lambda_i}. \tag{4}$$

The Jeffreys' prior, π_J, is given by

$$\pi_J(\underline{\lambda}) \propto |F(\underline{\lambda})|^{\frac{1}{2}} = \left(\prod_{i=1}^{k} \lambda_i\right)^{-\frac{1}{2}}. \tag{5}$$

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

$$\pi_J\left(\underline{\lambda}\,|data\right) \propto \left(\prod_{i=1}^{k}\lambda_i\right)^{-\frac{1}{2}}\prod_{i=1}^{k}\lambda_i^{x_i}e^{-\lambda_i} = \prod_{i=1}^{k}\lambda_i^{x_i-\frac{1}{2}}e^{-\lambda_i}. \tag{6}$$

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

In some cases the exact posterior distribution of $\xi = \sum_{i=1}^{k} a_i\lambda_i$ can be derived. For example if $a_1 = a_2 = \ldots = a_k = \frac{1}{k}$. The following theorem can now be proved.

Theorem 2 If $\pi_J\left(\underline{\lambda}\,|\text{data}\right) = \prod_{i=1}^{k}\frac{e^{-\lambda_i}\lambda_i^{x_i-\frac{1}{2}}}{\Gamma\left(x_i+\frac{1}{2}\right)}$, then the posterior distribution of $\widetilde{\xi} = \frac{1}{k}\sum_{i=1}^{k}\lambda_i$ is

$$\pi_J\left(\widetilde{\xi}\,|\text{data}\right) = \frac{k^{\sum_{i=1}^{k}x_i+\frac{k}{2}}}{\Gamma\left(\sum_{i=1}^{k}x_i+\frac{k}{2}\right)}\widetilde{z}^{\sum_{i=1}^{k}x_i+\frac{k}{2}-1}e^{-k\widetilde{z}}. \tag{7}$$

Proof. $\xi = \sum_{i=1}^{k} a_i\lambda_i$. The moment generating function of ξ is

$$\begin{aligned}
M_\xi(t) &= E\left(e^{\xi t}\right) = E\left(e^{t\sum_{i=1}^{k}a_i\lambda_i}\right)\\
&= \left\{\prod_{i=1}^{k}\frac{1}{\Gamma\left(x_i+\frac{1}{2}\right)}\right\}\int_0^\infty\cdots\int_0^\infty e^{t\sum_{i=1}^{k}a_i\lambda_i}\prod_{i=1}^{k}\left\{e^{-\lambda_i}\lambda_i^{x_i+\frac{1}{2}-1}\right\}d\lambda_1\ldots d\lambda_k\\
&= C\left(\int_0^\infty e^{ta_1\lambda_1}e^{-\lambda_1}\lambda_1^{x_1+\frac{1}{2}-1}d\lambda_1\right)\cdots\left(\int_0^\infty e^{ta_k\lambda_k}e^{-\lambda_k}\lambda_k^{x_k+\frac{1}{2}-1}d\lambda_k\right)
\end{aligned}$$

where $C = \left\{\prod_{i=1}^{k}\frac{1}{\Gamma\left(x_i+\frac{1}{2}\right)}\right\}$.

Consider

$$\begin{aligned}
I &= \int_0^\infty e^{ta_i\lambda_i}e^{-\lambda_i}\lambda_i^{x_i+\frac{1}{2}-1}d\lambda_i\\
&= \int_0^\infty e^{-\lambda_i(1-a_it)}\lambda_i^{x_i+\frac{1}{2}-1}d\lambda_i
\end{aligned}$$

Let $\lambda_i(1-a_it)=y, \therefore \lambda_i=\left(\frac{1}{1-a_it}\right)y$ and $d\lambda_i=\left(\frac{1}{1-a_it}\right)dy$.

$$
\begin{aligned}
\therefore I &= \int_0^\infty e^{-y}\left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}-1} y^{x_i+\frac{1}{2}-1}\left(\frac{1}{1-a_it}\right)dy \\
&= \left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}}\int_0^\infty e^{-y}y^{x_i+\frac{1}{2}-1}dy \\
&= \left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}}\Gamma\left(x_i+\frac{1}{2}\right)
\end{aligned}
$$

Therefore $M_\xi(t)=\prod_{i=1}^{k}\left(\frac{1}{1-a_it}\right)^{x_i+\frac{1}{2}}$.
If $a_1=a_2=\ldots=a_k=\frac{1}{k}$, then
$M_{\widetilde{\xi}}(t)=\left(\frac{k}{k-t}\right)^{\sum_{i=1}^k x_i+\frac{k}{2}}$, which is the moment generating function of a Gamma distribution.
Therefore

$$
\pi_J\left(\widetilde{\xi}\,|\text{data}\right)=\frac{k^{\sum_{i=1}^k x_i+\frac{k}{2}}}{\Gamma\left(\sum_{i=1}^k x_i+\frac{k}{2}\right)}\widetilde{\xi}^{\sum_{i=1}^k x_i+\frac{k}{2}-1}e^{-k\widetilde{\xi}} \qquad 0<\widetilde{\xi}<\infty. \blacksquare
$$

3 The Weighted Monte Carlo Method in the Case of $\xi=\sum_{i=1}^k a_i\lambda_i$

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 intervals. It does not require knowing the closed form of the marginal posterior distribution of ξ, only the kernel of the posterior distribution of $\{\lambda_1,\lambda_2,\ldots,\lambda_k\}$ is needed.

As mentioned by Chen & Shao (1999), Kim (2006), Smith & Gelfand (1992), Guttman & Menzefricke (2003), Skare *et al.* (2003) and Li (2007) the weighted Monte Carlo (sampling - importance re-sampling (SIR)) 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. 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
$\pi_{PM}(\underline{\lambda}\,|data)\propto\left\{\sum_{i=1}^k a_i^2\lambda_i\right\}^{\frac{1}{2}}\prod_{i=1}^k\lambda_i^{x_i}e^{-\lambda_i}$, if $\sum_{i=1}^k a_i=0$,
or
$\widetilde{\pi}_{PM}(\underline{\lambda}\,|data)\propto\left\{\sum_{i=1}^k a_i^2\lambda_i\right\}^{\frac{1}{2}}\prod_{i=1}^k\lambda_i^{x_i-1}e^{-\lambda_i}$, if $\sum_{i=1}^k a_i\neq 0$.

The sample probabilities are therefore proportional to

$$\frac{\pi_{PM}(\underline{\lambda}\,|data)}{q(\underline{\lambda}\,|data)} = \left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \qquad \text{if } \sum_{i=1}^{k} a_i = 0,$$

or

$$\frac{\widetilde{\pi}_{PM}(\underline{\lambda}\,|data)}{q(\underline{\lambda}\,|data)} = \left\{\sum_{i=1}^{k} a_i^2\lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k}\lambda_i^{-1} \qquad \text{if } \sum_{i=1}^{k} a_i \neq 0.$$

The normalised weights are

$$\omega_l = \frac{\left\{\sum_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}}}{\sum_{l=1}^{n}\left\{\sum_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}}} \qquad \text{if } \sum_{i=1}^{k} a_i = 0 \quad \text{for } l = 1,2,\ldots,n,$$

or

$$\omega_l = \frac{\left\{\sum_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}} \prod_{i=1}^{k}\left(\lambda_i^{-1}\right)^{(l)}}{\sum_{l=1}^{n}\left[\left\{\sum_{i=1}^{k}\left(a_i^2\lambda_i\right)^{(l)}\right\}^{\frac{1}{2}} \prod_{i=1}^{k}\left(\lambda_i^{-1}\right)^{(l)}\right]} \qquad \text{if } \sum_{i=1}^{k} a_i \neq 0 \quad \text{for } l = 1,2,\ldots,n.$$

A straightforward way of doing the weighted Monte Carlo (WMC) method was proposed by Chen & Shao (1999).

Details of the Monte Carlo method are as follows:

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 $\xi^{(l)} = \sum_{i=1}^{k}\left(a_i\lambda_i\right)^{(l)}$ for $l = 1,2,\ldots,n$.
2. Sort $\left\{\xi^{(l)},(l=1,2,\ldots,n)\right\}$ to obtain the ordered values $\xi^{[1]} \leq \xi^{[2]} \leq \cdots \leq \xi^{[n]}$.
3. Each simulated ξ value has an associated weight. Therefore compute the weighted function $\omega_{(l)}$ associated with the l^{th} ordered $\xi^{[l]}$ value.
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 $\xi^{[n_1]}$ value and denote it as $\xi_{(\alpha/2)}$. 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 $\xi^{[n_2]}$ value and denote it as $\xi_{(1-\alpha/2)}$.
5. The $100(1-\alpha)\%$ Bayesian confidence interval is:
$\left(\xi_{(\alpha/2)},\xi_{(1-\alpha/2)}\right)$.

4 Example and Simulation Studies

4.1 Example

Stamey & Hamilton (2006) considered an example where they compared four intervals. We are going to compare two Bayesian intervals, using Jeffreys' prior and a probability matching prior, with the four intervals from Stamey & Hamilton (2006). They considered the number of fatal motor vehicle accidents involving driving while intoxicated (DWI) during six major holidays for the year 2000. They obtained the data from the Crash Records Bureau of the Texas Department of Public Safety. The data are in Table 1.

Stamey & Hamilton (2006) used the following methods: a Wald interval, a t interval with Satterthwaite's degrees of freedom, and a Bayes interval using non-informative priors. In Table 2 the four methods used by Stamey & Hamilton (2006) to obtain 95% confidence intervals can be seen, as well as the two Bayesian methods that we considered. The purpose of this experiment was to estimate the number of DWI involved fatal accidents per holiday, and also to see if less such accidents occur during the summer holidays than during the winter holidays. Please note that the data are taken from the Crash Records Bureau of the Texas Department of Public Safety, the summer holidays are therefore Memorial Day, July 4 and Labor Day, and the winter holidays are Thanksgiving, Christmas and New Year's Eve. Table 2 indicates the 95% confidence intervals for the two linear functions.

Table 1: Number of DWI involved fatal motor vehicle accidents during six major holidays (2000)

Holiday	**No. of Accidents**
Memorial Day	0
July 4	5
Labor Day	2
Thanksgiving	11
Christmas	8
New Year's Eve	9

From Table 2 it can be seen that the Wald interval and the Student's t interval imply the the average number of DWI - involved fatal accidents per holiday do not exceed four, while the Bayes intervals imply that the average number exceeds four. The upper limits of the Bayesian methods exceed eight, while the Wald and Student's t upper limits are less than eight. For the contrast between the summer holidays and the winter holidays, it is noted that all the intervals are considerably greater than zero, which indicates that fatal accidents were more common in winter than in the summer. The Bayesian methods that we suggested compare well with the other intervals.

Table 2: 95% Confidence intervals for the contrasts for DWI - involved fatal motor vehicle accidents

Contrast	Wald	Student's t	Bayes	Bayes with t
Avg. No. of DWI Accidents/ Holiday $\underline{c} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$	$(3.9, 7.77)$	$(3.87, 7.80)$	$(4.31, 8.35)$	$(4.29, 8.38)$
Winter vs. Summer $\underline{c} = (-1/3, -1/3, -1/3, 1/3, 1/3, 1/3)$	$(3.13, 10.87)$	$(3.07, 10.93)$	$(2.97, 11.03)$	$(2.91, 11.09)$

Contrast	Jeffreys' prior	Probability matching prior
Avg. No. of DWI Accidents/ Holiday $\underline{c} = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$	$(4.18, 8.25)$	$(4.66, 8.63)$
Winter vs. Summer $\underline{c} = (-1/3, -1/3, -1/3, 1/3, 1/3, 1/3)$	$(2.77, 10.36)$	$(2.30, 10.25)$

4.2 Simulation Study I

In this section we are going to look into the expected widths and the coverage probabilities of six methods for constructing confidence intervals. To examine the coverage percentages the following simulation procedure was proposed by Stamey & Hamilton (2006). They first created Poisson means λ_i, $i = 1, \ldots, k$, from a Uniform distribution on the interval $(0-5)$, for a given number of theoretical populations k. They then simulated $X_i \sim P(\lambda_i)$, $i = 1, \ldots, k$, and compared the confidence intervals for each of the four methods based on the drawn observations and on the specified contrast coefficients a_i, $i = 1, \ldots, k$. To obtain the coverage probabilities the percentage of times over 100000 draws that each confidence interval contains the true value of the contrast $\xi = \sum_{i=1}^{k} a_i \lambda_i$ were calculated and to obtain the expected widths the average width of each interval were calculated.

From Table 3 it can be seen that the Wald interval is overall the poorest performer when all the Poisson rates are expected to be small, because the coverage never reached 95%. Stamey & Hamilton (2006) used the t distribution, to widen the intervals, but also did not get completely satisfactory results. When using the student's t distribution, they still got coverage results that are below nominal, but in most cases they were close to nominal, except for the case where $k = 5$. According to Stamey & Hamilton (2006) the Bayes procedure based on the Jeffreys' non-informative prior is performing the best for the small - rate cases. The coverage is above or at nominal for every case. The interval is also in many cases narrower on average than the interval based on the student's t. The interval based on student's t using the Bayesian prior estimator also has coverage above nominal in every case, but is wider than the Bayes interval using the standard normal coefficient. The procedures using the Jeffreys' and probability matching priors, the last three columns from Table 3, compare well with the other procedures. The average widths compare well with the other procedures. In Table 3, the column Bayes PMP gives the results when the prior, $\pi_{PM}(\underline{\lambda}) \propto \left\{ \sum_{i=1}^{k} a_i^2 \lambda_i \right\}^{\frac{1}{2}}$, is used and the column

Bayes PMP* gives the results when the prior, $\widetilde{\pi}_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{-1}$, is used.

Table 3: Average coverage and width for contrasts where $\lambda_i \in (0,5)$

Contrast	Wald	Student's t	Bayes	Bayes with t	Bayes Jef	Bayes PMP	Bayes PMP*
			***k* = 2**				
$(1,-1)$	91.1%	98.3%	96.9%	99.1%	96.1%	95.66%	95.32%
	8.40	10.09	9.28	11.27	9.44	10.16	9.59
$\left(\frac{1}{2},\frac{1}{2}\right)$	91.2%	97.8%	96.7%	99.2%	95.6%		96.22%
	4.12	5.08	4.64	5.63	4.53		4.46
			***k* = 3**				
$\left(1,-\frac{1}{2},-\frac{1}{2}\right)$	93.5%	96.4%	96.5%	98.2%	95.9%	96.06%	95.19%
	7.24	8.31	8.05	9.32	8.20	8.68	8.16
$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	91.6%	95.7%	96.7%	98.3%	94.8%		95.61%
	3.46	3.85	3.83	4.28	3.80		3.64
			***k* = 4**				
$\left(\frac{1}{2},\frac{1}{2},-\frac{1}{2},-\frac{1}{2}\right)$	94.8%	96.7%	97.1%	98.2%	95.5%	96.16%	95.49%
	6.05	6.51	6.68	7.20	6.81	7.05	6.59
$\left(1,-\frac{1}{3},-\frac{1}{3},-\frac{1}{3}\right)$	92.2%	94.8%	95.9%	97.5%	95.5%	96.44%	95.76%
	6.74	7.78	7.55	8.73	7.73	8.15	7.65
$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$	92.7%	95.1%	96.1%	97.4%	93.4%		95.73%
	3.02	3.25	3.38	3.60	3.33		3.14
			***k* = 5**				
$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)$	93.1%	94.8%	95.5%	96.6%	93.7%		95.52%
	2.72	2.87	3.00	3.17	2.99		2.82
$\left(1,-\frac{1}{4},-\frac{1}{4},-\frac{1}{4},-\frac{1}{4}\right)$	91.2%	93.8%	95.4%	97.1%	94.6%	96.88%	95.85%
	6.48	7.56	7.28	8.50	7.39	7.80	7.34
$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3},-\frac{1}{2},-\frac{1}{2}\right)$	94.6%	96.1%	96.9%	97.8%	96.7%	96.21%	95.56%
	5.53	5.89	6.10	6.51	6.21	6.38	5.99

Stamey & Hamilton (2006) also considered another simulation study, to see what impact larger expected counts will have on the intervals. They used exactly the same method as in the previous simulation study, the only difference is that in this case the Poisson rates were generated from a Uniform distribution on the interval $(5-10)$. In the previous simulation study the Poisson rates were generated from a Uniform distribution on the interval $(0-5)$. The results (coverage percentages and interval widths) calculated by them are given in the first four columns of Table 4. In the last three columns the results are given when using the Jeffreys' and probability matching priors. In Table 4, the column Bayes PMP gives the results when the prior, $\pi_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}}$, is used and the column Bayes PMP* gives the results when the prior, $\widetilde{\pi}_{PM}(\underline{\lambda}) \propto \left\{\sum_{i=1}^{k} a_i^2 \lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{k} \lambda_i^{-1}$, is used.

Table 4: Average coverage and width for contrasts where $\lambda_i \in (5, 10)$

Contrast	**Wald**	**Student's t**	**Bayes**	**Bayes with t**	**Bayes Jef**	**Bayes PMP**	**Bayes PMP***
***k* = 2**							
$(1, -1)$	95.1%	96.1%	95.9%	96.7%	94.8%	94.76%	93.48%
	15.01	15.68	15.53	16.22	15.69	16.24	15.66
$\left(\frac{1}{2}, \frac{1}{2}\right)$	93.4%	94.6%	95.3%	96.5%	94.2%		95.09%
	7.51	7.84	7.76	8.11	7.70		7.58
***k* = 3**							
$\left(1, -\frac{1}{2}, -\frac{1}{2}\right)$	94.5%	95.5%	95.4%	96.3%	95.6%	95.02%	93.88%
	13.00	13.55	13.44	14.03	13.43	13.87	13.39
$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$	93.9%	94.7%	95.5%	96.1%	95.9%		95.06%
	6.46	6.33	6.36	6.54	6.28		6.20
***k* = 4**							
$\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right)$	95.1%	95.6%	95.8%	96.3%	95.5%	95.00%	94.52%
	10.68	10.91	11.04	11.26	11.12	11.27	10.91
$\left(1, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right)$	94.0%	95.1%	94.7%	95.8%	94.1%	95.00%	94.66%
	12.24	12.85	12.67	13.29	12.80	13.06	12.60
$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$	94.2%	94.8%	95.4%	95.9%	94.8%		94.71%
	5.34	5.45	5.52	5.63	5.50		5.36
***k* = 5**							
$\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$	94.4%	94.8%	95.2%	95.6%	94.2%		94.74%
	4.78	4.86	4.94	5.02	4.91		4.76
$\left(1, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}\right)$	93.6%	95.0%	94.5%	95.7%	95.5%	95.46%	94.64%
	11.83	12.49	12.24	12.92	12.19	12.55	12.15
$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{2}, -\frac{1}{2}\right)$	95.0%	95.4%	95.7%	96.2%	95.3%	95.35%	94.42%
	9.75	9.94	10.08	10.28	10.1	10.27	9.93

As in the previous simulation study the Wald interval is again overall the poorest performer. Stamey & Hamilton (2006) could not clearly state whether the interval based on the Bayes or the interval based on the student's t performs the best in this case. Both have coverages that are slightly below nominal for some contrasts, but in most cases the coverage is usually quite close to nominal. From the second last column of Table 4, the probability matching prior, compares well with the other procedures in the cases where there is a linear contrast of Poisson parameters. This results in coverage at or just above nominal level in each case, the interval widths also compare well with the other procedures used.

4.3 Simulation Study II - Comparing Two Poisson Means

Krishnamoorthy & Thomson (2004) considered the problem of hypothesis testing about two Poisson means. They compared the usual conditional test (C - test) to a test based on estimated p - values (E - test). The C - test is due to Przyborowski & Wilenski (1940) and it is based on the conditional distribution of X_1 given $X_1 + X_2$, which follows a Binomial distribution whose success probability is a function of the ratio λ_1/λ_2.

Here

$$X_1 = \sum_{i=1}^{n_1} X_{1i} \sim P(n_1\lambda_1),$$

independently distributed of

$$X_2 = \sum_{i=2}^{n_2} X_{2i} \sim P(n_2\lambda_2) \tag{8}$$

where $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ are independent samples, respectively from $P(\lambda_1)$ and $P(\lambda_2)$ distributions.

The p -value for testing

$$H_0 : \lambda_1 - \lambda_2 \leq d \quad \text{vs} \quad H_a : \lambda_1 - \lambda_2 > d \tag{9}$$

is $P\left(T_{X_1,X_2} \geq T_{k_1,k_2} | H_0\right)$ which involves the unknown parameter λ_2. Here

$$T_{X_1,X_2} = \frac{X_1/n_1 - X_2/n_2 - d}{\sqrt{\hat{V}_X}}$$

is the pivot statistic for the testing problem and for given (n_1, k_1, n_2, k_2), the observed value of the pivot statistic T_{X_1,X_2} is given by

$$T_{k_1,k_2} = \frac{k_1/n_1 - k_2/n_2 - d}{\sqrt{\hat{V}_k}}$$

where

$$\hat{V}_X = \frac{X_1/n_1}{n_1} + \frac{X_2/n_2}{n_2}$$

and $\hat{V}_k$ is defined similarly with X replaced by k.

For given k_1 and k_2 an estimate of λ_2 is given by

$$\hat{\lambda}_{2k} = \frac{k_1 + k_2}{n_1 + n_2} - \frac{dn_1}{n_1 + n_2}.$$

Using this $\hat{\lambda}_{2k}$ Krishnamoorthy & Thomson (2004) estimated the p - value $P\left(T_{X_1,X_2} \geq T_{k_1,k_2} | H_0\right)$ by

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \frac{e^{-n_1\left(\hat{\lambda}_{2k}+d\right)} \left\{n_1\left(\hat{\lambda}_{2k} + d\right)\right\}^{x_1}}{x_1!} \frac{e^{-n_2\hat{\lambda}_{2k}} \left(n_2\hat{\lambda}_{2k}\right)^{x_2}}{x_2!} I\left[T_{x_1,x_2} \geq T_{k_1,k_2}\right] \tag{10}$$

where $I[\cdot]$ denotes the indicator function. For given nominal level α, the test rule is to reject H_0 in 9 whenever the estimated p - value in 9 is less than α. In view of 8, without loss of generality, we can take $n_1 = n_2 = 1$.

Krishnamoorthy & Thomson (2004) compared these two tests by looking at the size and the power of the tests at different nominal levels and also at different values for λ_1 and λ_2. They found that the E - test is almost exact and that it is more powerful than the C - test. We are going to compare three Bayesian procedures to the tests used in Krishnamoorthy & Thomson (2004). For the Bayesian procedure we will use the Jeffreys' prior, the probability matching prior, a third prior: $\pi_A(\underline{\lambda}) \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and a fourth prior: $\pi_B(\underline{\lambda}) \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$.

From Theorem 1, the probability matching prior is given by

$$\pi_{PM}(\lambda_1,\lambda_2) \propto \left\{\sum_{i=1}^{2} a_i^2\lambda_i\right\}^{\frac{1}{2}} = \sqrt{\lambda_1+\lambda_2}.$$

When using the probability matching prior, the posterior distribution of λ_1,λ_2 is given by

$$\pi_{PM}(\lambda_1,\lambda_2\,|x_1,x_2) \quad \propto \quad \left\{\sum_{i=1}^{2} a_i^2\lambda_i\right\}^{\frac{1}{2}} \prod_{i=1}^{2}\lambda_i^{x_i}e^{-\lambda_i}.$$

The Jeffreys' prior, π_J, is given by

$$\pi_J(\lambda_1,\lambda_2) \propto |F(\lambda_1,\lambda_2)|^{\frac{1}{2}} = \left(\prod_{i=1}^{2}\lambda_i\right)^{-\frac{1}{2}} = \lambda_1^{-\frac{1}{2}}\lambda_2^{-\frac{1}{2}}.$$

Where $F(\lambda_1,\lambda_2)$ is the information matrix connected with the likelihood function.

When using the Jeffreys' prior, the posterior distribution of λ_1,λ_2 is given by

$$\pi_J(\lambda_1,\lambda_2\,|x_1,x_2) \propto \left(\prod_{i=1}^{2}\lambda_i\right)^{-\frac{1}{2}} \prod_{i=1}^{2}\lambda_i^{x_i}e^{-\lambda_i} = \prod_{i=1}^{2}\lambda_i^{x_i-\frac{1}{2}}e^{-\lambda_i}.$$

The posterior distribution of λ_1,λ_2 is thus the product of 2 independently distributed $Gamma\left(x_i+\frac{1}{2},1\right)$ variates.

The third prior, π_A, is given by

$$\pi_A(\lambda_1,\lambda_2) \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}.$$

When using this prior, the posterior distribution of λ_1,λ_2 is given by

$$\pi_A(\lambda_1,\lambda_2\,|x_1,x_2) \propto \left(\prod_{i=1}^{2}\lambda_i\right)^{-\frac{1}{4}} \prod_{i=1}^{2}\lambda_i^{x_i}e^{-\lambda_i} = \prod_{i=1}^{2}\lambda_i^{x_i-\frac{1}{4}}e^{-\lambda_i}.$$

The posterior distribution of λ_1,λ_2 is thus the product of 2 independently distributed $Gamma\left(x_i+\frac{3}{4},1\right)$ variates.

The fourth prior, π_B, is given by

$$\pi_B(\lambda_1,\lambda_2) \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}.$$

When using this prior, the posterior distribution of λ_1, λ_2 is given by

$$\pi_B(\lambda_1, \lambda_2 | x_1, x_2) \propto \left(\prod_{i=1}^{2} \lambda_i\right)^{-\frac{3}{8}} \prod_{i=1}^{2} \lambda_i^{x_i} e^{-\lambda_i} = \prod_{i=1}^{2} \lambda_i^{x_i - \frac{3}{8}} e^{-\lambda_i}.$$

The posterior distribution of λ_1, λ_2 is thus the product of 2 independently distributed $Gamma\left(x_i + \frac{5}{8}, 1\right)$ variates.

Rice (1995) gives the following definition for the size of a test, which is also known as a type I error:

H_0 may be rejected when it is true. Such an error is called a type I error, and its probability is denoted by α.

In Figures 1 - 3, we compare the size of the tests using Bayesian procedures to the two tests from Krishnamoorthy & Thomson (2004). The Bayesian simulation procedure in the case of the probability matching prior is as discussed in Section 3. The size of the tests as a function of $\lambda = \lambda_1 = \lambda_2$ at the three different nominal level under the null hypothesis $H_0 : \lambda_1 - \lambda_2 = 0$ will be given in Figures 1 - 3.

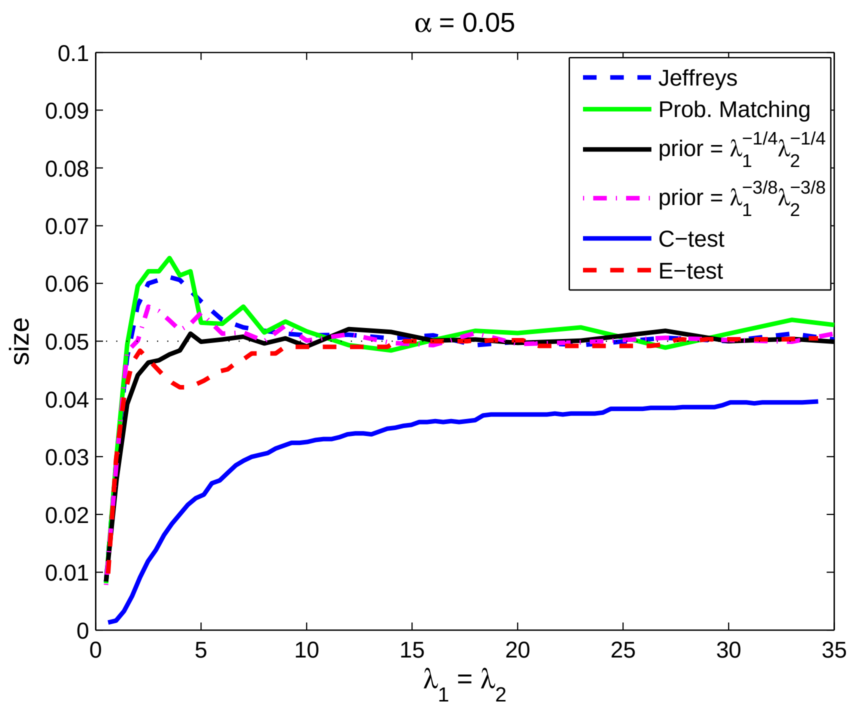

Figure 1: Size of the tests at the 5% nominal level.

From Figure 1 it can be seen that the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$, reaches the nominal level when $\lambda_1 = \lambda_2 = 5$ and from there onwards it attains this level. Where the Jeffreys' and probability matching priors reach the nominal level at $\lambda_1 = \lambda_2 = 2$, and then only at $\lambda_1 = \lambda_2 = 10$ again, from there onwards it attains this level. The prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$, is an improvement on the Jeffreys' and probability matching priors. From Krishnamoorthy & Thomson (2004) the C - test never reaches the nominal level, and the E - test reaches the nominal level only at $\lambda_1 = \lambda_2 = 10$. We must however mention that the graphs for the E - and C - tests are scanned in using a Matlab program. This means that some small technical errors may occur in the graphs. In general one can say that the Jeffreys' and probability matching priors tend o give Type 1 error rates that are somewhat larger than the chosen alpha. The error rates of the priors $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$ seem to be more accurate.

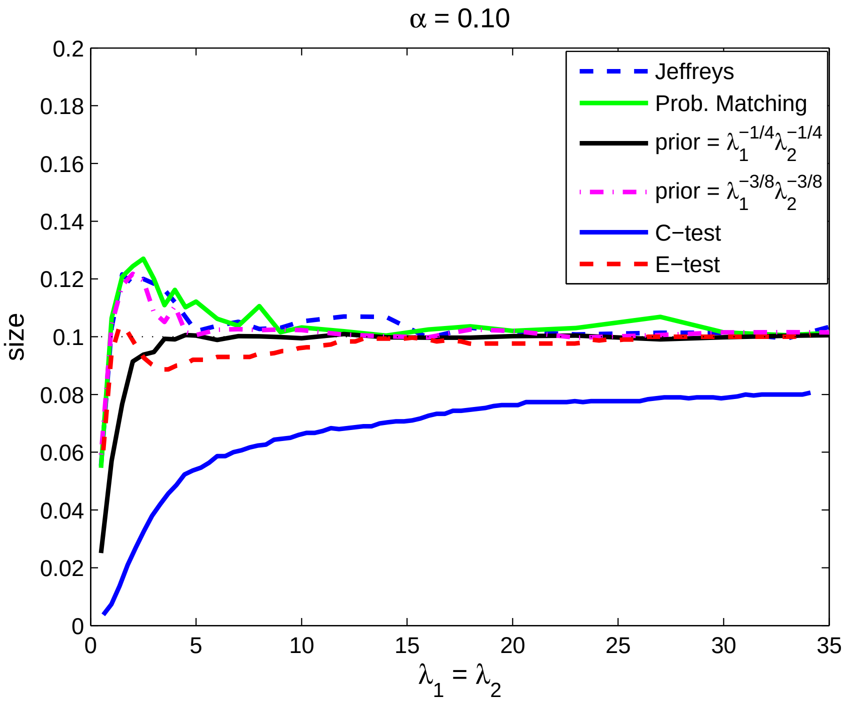

Figure 2: Size of the tests at the 10% nominal level.

From Figure 2 it can be seen that the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, reaches the nominal level when $\lambda_1 = \lambda_2 = 4$ and from here onwards it attains this level. Where the Jeffreys' prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$, reach the nominal level at $\lambda_1 = \lambda_2 = 1$, and then only at $\lambda_1 = \lambda_2 = 5$ again, from there onwards it attains this level. The probability matching prior follows a similar pattern, but it reaches the nominal level at $\lambda_1 = \lambda_2 = 1$, and then only at $\lambda_1 = \lambda_2 = 10$ again, from there onwards it attains this level. From Krishnamoorthy & Thomson (2004) the C - test never reaches the nominal level, and the E - test reaches the nominal level $\lambda_1 = \lambda_2 = 1$, and then only at $\lambda_1 = \lambda_2 = 13$ again, from there onwards it attains this level.

From Figure 3 it can be seen that the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, reaches the nominal level when $\lambda_1 = \lambda_2 = 4.5$ and from here onwards it attains this level. Where the Jeffreys' and probability matching priors never stay constant at the nominal level, it fluctuates most of the time above the nominal level. The prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$, follows a similar pattern but for a lesser extent than that of the Jeffreys' and probability matching priors. Figure 3 however enlarges the fluctuation of the observed error rate. A more direct comparison is to plot the error rate on the same scale. In terms of absolute deviations the Jeffreys' and probability matching priors are not doing worse as α decreases. The mean deviations for the Jeffreys' prior from the nominal α values 0.01, 0.005 and 0.001 are 0.009, 0.0014 and 0.000412 respectively. From Krishnamoorthy & Thomson (2004) the C - test never reaches the nominal level, and the E - test also reaches the nominal level at $\lambda_1 = \lambda_2 = 3$, and then at $\lambda_1 = \lambda_2 = 5$ again, from there onwards it attains this level.

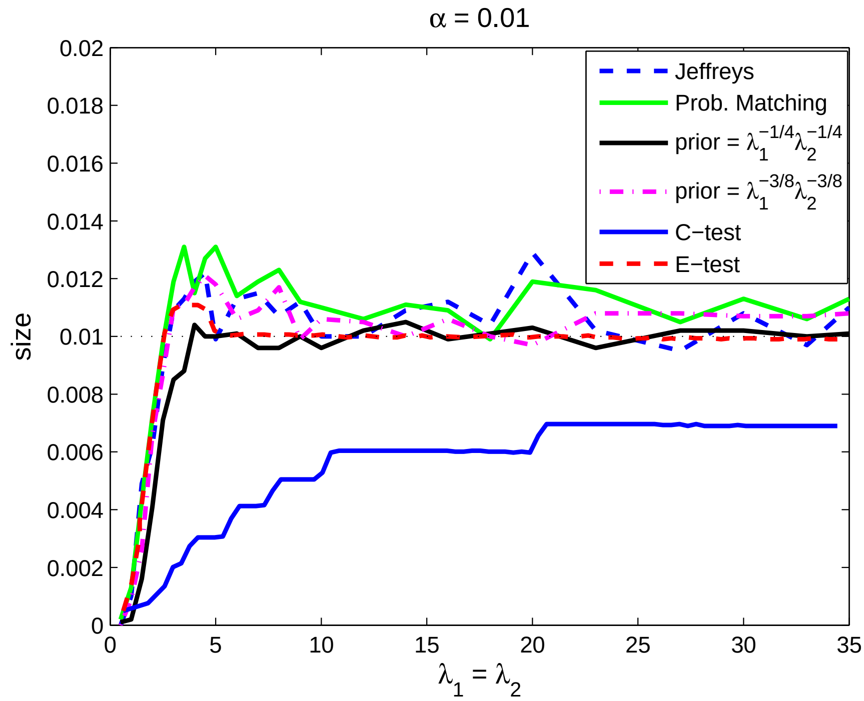

Figure 3: Size of the tests at the 1% nominal level.

From Figures 1 - 3, it can be seen that the Bayesian procedures compare relatively well with the E - test from Krishnamoorthy & Thomson (2004). From the four Bayesian procedures, the procedure when using the prior $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$ and $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$ give the best results. The C - test is the poorest performer.

In Figures 4 - 6, we compare the power of the tests using Bayesian procedures to the two tests from Krishnamoorthy & Thomson (2004). The power of the tests as a function of λ_1 at the nominal level $\alpha = 0.05$ under the alternative hypothesis $H_a : \lambda_1 - \lambda_2 > 0$ will be given in Figures 4 - 6.

Rice (1995) gives the following definition for a type II error:

H_0 may be accepted when it is false. Such an error is called a type II error, and its probability is denoted by β.

The probability that H_0 is rejected when it is false is called the power of the test, the power equals $1-\beta$.

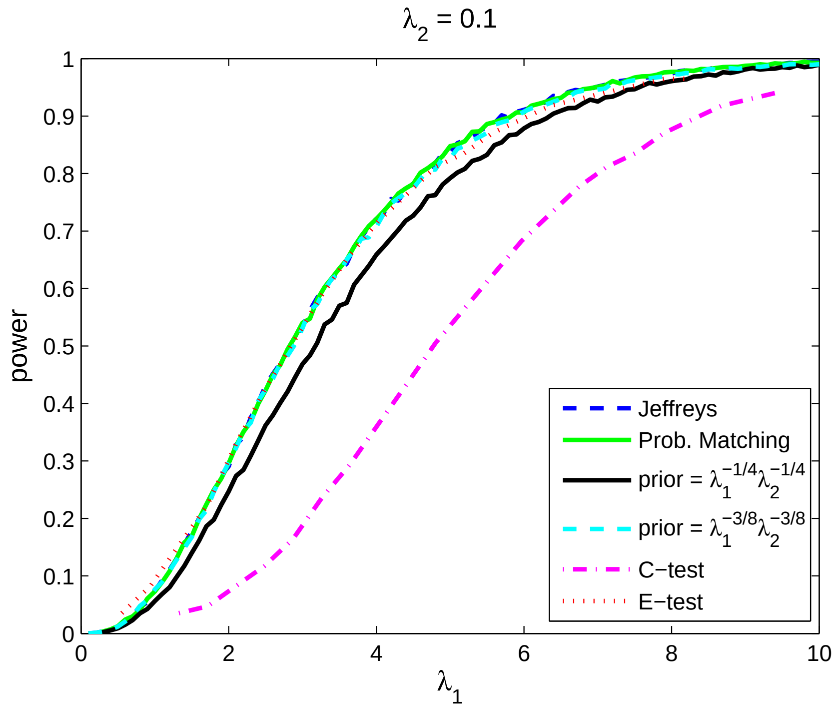

Figure 4: Power of the test as a function of λ_1 when $\lambda_2 = 0.1$.

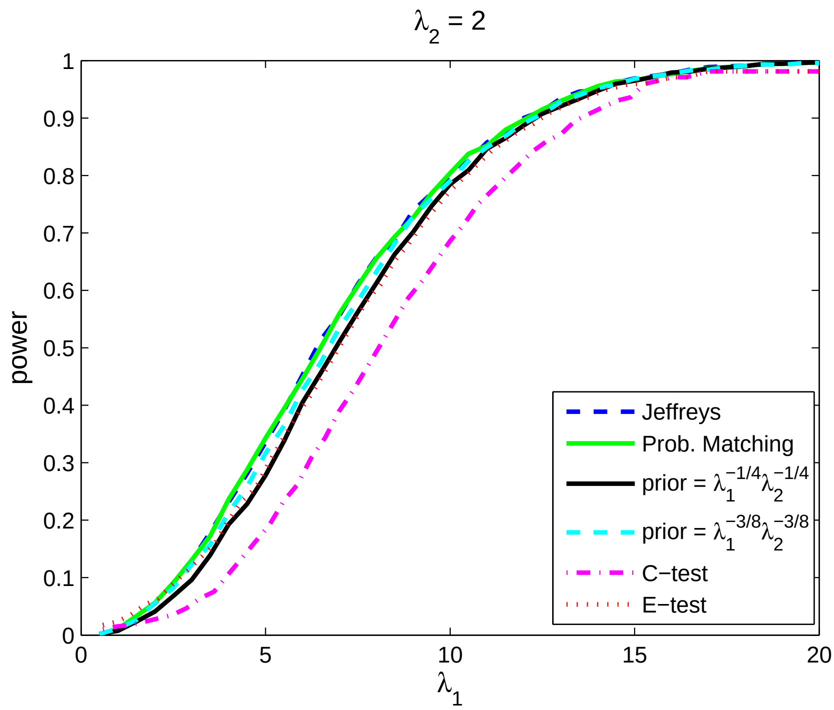

Figure 5: Power of the test as a function of λ_1 when $\lambda_2 = 2$.

From Figure 4 it can be seen that the power from the test when using the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, is smaller than the power of the tests when using the Jeffreys' prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$. The Jeffreys' prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$, and the E- test give almost exactly the same results. From Krishnamoorthy & Thomson (2004) the power of the E- test is larger than the power of the C - test.

From Figure 5 it can be seen that the power from the test when using the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, is still a bit smaller than the power of the tests when using the Jeffreys' prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$. The Jeffreys' and probability matching priors give almost exactly the same results. The prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, and the E- test give almost exactly the same results. From Krishnamoorthy & Thomson (2004) the power of the E - test is larger than the power of the C - test.

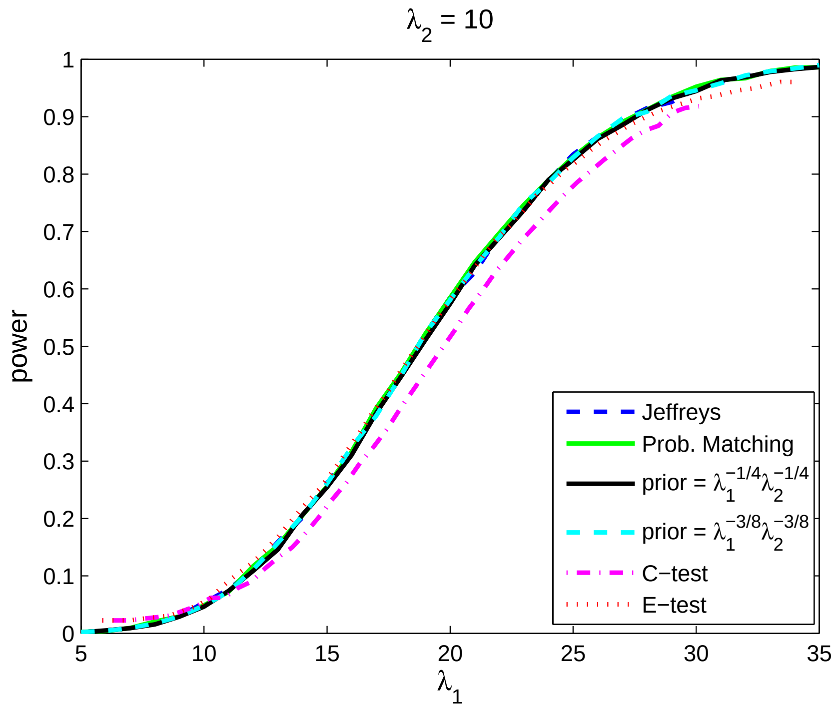

Figure 6: Power of the test as a function of λ_1 when $\lambda_2 = 10$.

From Figure 6 it can be seen that the power from the test when using the prior, $\pi_A \propto \lambda_1^{-\frac{1}{4}}\lambda_2^{-\frac{1}{4}}$, is almost the same as the power of the tests when using the Jeffreys' prior, probability matching prior and the prior, $\pi_B \propto \lambda_1^{-\frac{3}{8}}\lambda_2^{-\frac{3}{8}}$. From Krishnamoorthy & Thomson (2004) the power of the E - test is still a bit larger than the power of the C - test, but they are almost equal to each other. The four tests using Bayesian methods and the E - test all give almost exactly the same results. We can conclude from the Bayesian procedures and from the methods by Krishnamoorthy & Thomson (2004), that as the sample sizes increase, i.e. as the values of λ_1 and λ_2 increase, the powers of the tests are increasing. Also, as the values of λ_1 and λ_2 increase, the difference between the powers of the different tests are smaller. The Bayesian procedures compare well with the procedures by Krishnamoorthy & Thomson (2004). From Figures 4 -6 it is also clear that the powers of Jeffreys' and probability matching priors are larger than those of the E - test. This could be expected because the type 1 error rates of these two priors are usually somewhat larger that the chosen alpha value.

5 Conclusion

In this paper the probability matching prior for a linear contrast of Poisson parameters, $\xi = \sum_{i=1}^{k} a_i \lambda_i$, (i.e. $\sum_{i=1}^{k} a_i = 0$) was derived. We also indicated what the probability matching prior should be when $\sum_{i=1}^{k} a_i \neq 0$. We compared the four approximate confidence intervals for linear contrasts of Poisson rates proposed by Stamey & Hamilton (2006) to confidence intervals using Bayesian procedures, when using the probability matching prior. Simulation studies have shown that the Wald interval performs the poorest. The probability matching prior performs also satisfactory. We also addressed the problem of hypothesis testing about two Poisson means, by looking at the size and power of different tests. We compared three Bayesian procedures to two procedures used by Krishnamoorthy & Thomson (2004). We used the Jeffreys' prior, the probability matching prior, a third prior which is proportional to $\lambda_1^{-\frac{1}{4}} \lambda_2^{-\frac{1}{4}}$ and a fourth prior which is proportional to $\lambda_1^{-\frac{3}{8}} \lambda_2^{-\frac{3}{8}}$ and compared it to their results. The Bayesian procedures compared well with the procedures used by Krishnamoorthy & Thomson (2004). The C - test performed the worst of the five tests.

References

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.

Chen, M-H., & 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. S., & Ghosh, J. K. 1995. On Priors Providing Frequentist Validity for Bayesian Inference. *Biometrika*, **82**(1), 37–45.

Guttman, I., & Menzefricke, U. 2003. Posterior Distributions for Functions of Variance Components. *Sociedad de Estadistica e Investigacion Operativa Test*, **12**(1), 115 – 123.

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.

Krishnamoorthy, K., & Thomson, J. 2004. A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference*, **119**(1), 23 – 35.

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

Przyborowski, J., & Wilenski, H. 1940. Homogeneity of results in testing samples from Poisson series. *Biometrika*, **31**, 313 – 323.

Rice, J.A. 1995. *Mathematical Statistics and Data Analysis*. 2nd edn. Duxbury Press, California.

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

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

Stamey, J., & Hamilton, C. 2006. A Note on Confidence Intervals for a Linear Function of Poisson Rates. *Communications in Statistics: Simulation & Computation*, **35**(4), 849 – 856.