5-2006

Bayesian Reference Inference on the Ratio of Poisson Rates.

Changbin Guo
East Tennessee State University

Bayesian Reference Inference on the Ratio of Two Poisson Rates

---

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

Changbin Guo

May 2006

Robert Price, Ph.D., Chair

Edith Seier, Ph.D.

Robert Gardner, Ph.D.

Yali Liu, Ph.D.

Keywords: Poisson ratio, reference analysis, posterior simulation, credible interval, HPD interval, maximum a posteriori, probability matching prior, exact coverage

ABSTRACT

Bayesian Reference Inference on the Ratio of Two Poisson Rates

by

Changbin Guo

Bayesian reference analysis is a method of determining the prior distribution under the Bayesian paradigm. It incorporates as little information as possible from the experiment. Estimation of the ratio of two independent Poisson rates is a common practical problem. In this thesis, the method of reference analysis is applied to derive the posterior distribution of the ratio of two independent Poisson rates, and then to construct point and interval estimates based on the reference posterior. In addition, the Frequentist coverage property of highest posterior density (HPD) intervals is evaluated through simulation.

DEDICATION

I dedicate this thesis to my parents, Yanping Lu and Zengqiang Guo, for their encouragement and support on my pursuit of advanced studies.

ACKNOWLEDGMENTS

I would like to thank Dr. Robert Price, my adviser and mentor, for his advice, support and patience during preparation of this thesis. He introduced me to the field of statistics and has continuously encouraged me.

Special thanks to Dr. Gardner, our graduate coordinator, who has been helping me since my entrance to the program. Many thanks to my other committee members for their time and patience. I would also like to express my gratitude to all the people in the Department of Mathematics who have taught or helped me in the past two years.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 The Problem of Estimating the Ratio of Two Poisson Rates

Many phenomena where rare events occur randomly in time or space are usually modeled by the statistician as a Poisson process, which is the most important class of stochastic processes. The following are the assumptions to ensure a Poisson process:

- The probability of one change in a short interval of length t is approximately proportional to the length of the interval and is independent of changes in other nonoverlapping intervals.

- The probability of two or more changes in a short interval of length t is essentially equal to zero.

The probability distribution that is associated with this process is called the Poisson distribution. The probability mass function has the form

$$f(y|\mu) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, \ldots$$

where $\mu > 0$.

As experiments of two-sample designs are common in decision-making applications, it may be of interest to estimate the ratio of two unknown population Poisson rates. Let Y_1, Y_2 be two independent Poisson variables and y_1, y_2 be the observed counts in each of the two samples. The amount of time or space of the sample is quantified by t_j. We have $E(Y_j) = VAR(Y_j) = \mu = t_j\lambda_j$ for $j = 1, 2$, where μ_j is an unknown mean of the Poisson distribution and λ_j is the *Poisson rate*.

Several Frequentist methods have been developed for estimating the ratio of two Poisson rates with desired accuracy using transformation techniques. See [17] for an overview and comparison of these methods. Bayesian approaches under conjugate and noninformative priors applied to each Poisson process has been studied in [17].

In this thesis, I am interested in making an inference on the ratio of two Poisson rates based on the reference posterior derived under the paradigm of Bayesian reference analysis [2]. It is interesting to observe that the resulting interval estimates have satisfactory Frequentist coverage probability, and they are in fact equivalent to the method where a noninformative prior is applied for each Poisson process in [17].

1.2 Bayesian Methodology

Unlike methods of traditional statistical inference that are primarily based on a retrospective evaluation of the distribution of possible y values conditional on the true unknown parameter θ, Bayesian methods distinguish themselves explicitly by conditioning on the observed data to quantify uncertainty in statistical data analysis.

In order to obtain such a probability statement, we should first begin with a joint probability distribution for θ and y. From probability theory, the joint probability density function can be represented as a product of two densities, namely the *prior distribution* $p(\theta)$ (sometimes we use $\pi(\theta)$) and the sampling distribution $p(y|\theta)$, that is,

$$p(\theta, y) = p(\theta)p(y|\theta).$$

Using the basic property of conditional probability known as *Bayes' formula* [10], the *posterior* density is the conditional probability of the parameter θ given the data y, i.e.

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{1}$$

where $p(y) = \sum_\theta p(\theta)p(y|\theta)$. If θ is continuous, we have $p(y) = \int p(y)p(y|\theta)d\theta$.

For the sake of computational convenience, the posterior density is usually expressed in the unnormalized form

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{2}$$

where "$\propto$" stands for "proportional to." Note that the probability density function $p(y|\theta)$ is often referred to as the *likelihood* [10] function denoted as $L(\theta)$. For an *identically and independently distributed* (i.i.d.) sample $y = (y_1, \cdots, y_n)$, the principle

of exchangeability [15] suggests writing the likelihood as

$$L(\theta) = \prod_{i=1}^{n} p(y_i|\theta) \ . \tag{3}$$

In this manner, the *Bayes' formula* can be simply interpreted in words by the statement that the posterior density is proportional to the product of the likelihood function and the prior density. This implies that Bayesian inference, like the traditional methodologies of statistics, also complies with the likelihood principle [5]. Nevertheless, a particular specification of the prior density $p(\theta)$ will make the posterior distribution not totally depend on the observed data values. It is this feature that separates Bayesian methodology from the Frequentist one.

1.3 Prior Specification

The *Bayes' formulae* (1) (2) may be viewed as a data-driven machine or probability transformation, which maps prior densities that describe prior knowledge into posterior densities. However, priors, unlike the data, are not generally known to be objective.

This situation of *insufficient reasoning* [15] makes the analysis of whether or not sensible changes in the prior would result in noticeable changes to the posterior; a fundamental question in Bayesian methodology. It disturbs some statisticians that there is often a broad range of prior distributional choices. To overcome such a difficulty, some people suggest including a scientific report to demonstrate how the posterior functionally depends on the choice of the prior [15].

Nowadays, besides the reference algorithm which we will discuss later, there are mainly two approaches to determine the prior distribution, namely, *Jeffrey's principle* and *conjugate methodology*.

1.3.1 Jeffrey's Principle

Though no general principles have been developed to specify the prior in practice, most statisticians agree that such a prior, if it exists, should incorporate as little relevant information as possible. A prior satisfying this kind of property is referred to as a *noninformative prior* [18]. The most widely used method for obtaining a noninformative prior for a one-parameter model is the *Jeffrey's principle* [10], which is motivated by the idea that Bayesian inference should not depend on how a model (or equivalently the likelihood) is parameterized.

The Fisher information matrix is defined as

$$I(\theta)_{ij} = E\Big(- \frac{\partial^2 l}{\partial\theta_i \partial\theta_j}\Big) \tag{4}$$

where l denotes the log-likelihood, $l(\theta) \triangleq \log L(\theta)$. Theoretically, since it is proportional to the expected curvature of the likelihood at the *Maximum Likelihood Estimator* (MLE), it measures sensitivity of an estimation in the neighborhood of the MLE [8].

The *Jeffrey's principle* is to take the prior to be

$$\pi_\theta(\theta) \propto det\big(I(\theta)\big)^{\frac{1}{2}} \tag{5}$$

where $det(.)$ denotes the determinant. This is applicable as long as $I(\theta)$ is well defined and positive definite [18]. It has the property of invariance as it can be easily checked that for any other parameterization γ, one has

$$\pi_\theta(\theta) = \pi_\gamma\big(\gamma(\theta)\big) \cdot \Big| det\Big(\frac{\gamma}{\theta}\Big)\Big| \ . \tag{6}$$

This means that a prior under a different parameterization will follow the change-of-variable formula. Hence, the selection of a specific parameterization is not necessary.

1.3.2 Conjugate Prior

Before powerful computational techniques such as the *Monte Carlo Markov Chain* (MCMC) method came along, a convenient choice of a prior was to take a mathematical function which simplifies the analytical calculation of the posterior density. One well-known approach is to choose a family of prior densities based on the likelihood function such that the resulting posterior belongs to the same family of functions as the prior. Prior and posterior densities determined from such a strategy are said to be *conjugate*.

Formally, if $y = (y_1, \cdots, y_n)$ are i.i.d. from an exponential family then the probability density or mass function for each observation can be expressed as

$$y_i|\theta \sim f(y_i|\theta) = A(\theta)e^{T^*(y_i)B(\theta)}\psi(y_i) \ .$$

By equation (3), the likelihood function in this case is represented as

$$l(\theta) = \prod_{i=1}^{n} f(y_i|\theta) = [A(\theta)]^n e^{T \cdot B(\theta)} H(y)$$

where $T = \sum_i T^*(y_i)$ and $H(y) = \prod_i \psi(y_i)$. Thus, a conjugate prior density is defined as

$$p(\theta) \propto \big[A(\theta)\big]^p e^{q \cdot B(\theta)}$$

and the correspondening posterior is

$$p(\theta|y) \propto \big[A(\theta)\big]^{n+p} e^{(T+q)B(\theta)} \ .$$

Interestingly, all the functions belonging to the exponential family have conjugate priors [10]. For example, the beta distribution is the conjugate prior to the binomial

model. If we combine the prior distribution $p(\theta) \propto \theta^{a-1}(1-\theta)^{a-1} \propto Beta(a,b)$ with the likelihood $L(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}$, the posterior will be $p(\theta|y) \propto \theta^{a+y-1}(1-\theta)^{n-y+b-1} \propto Beta(a+y, n-y+b)$.

Since the exponential family is broad, the conjugate methodology is quite developed. Nevertheless, the obvious limitation to apply conjugate methodology lies in that it is usually unrealistic to attempt to represent prior information in the conjugate form. There is no theoretical basis for taking certain values of parameters in the conjugate priors.

1.4 Bayesian Reference Analysis

Like the Jeffrey's principle and conjugate priors, *Bayesian reference analysis* [2] emerged as an approach for determining a prior from the likelihood function under the Bayesian paradigm. From the perspective of information theory, the available prior knowledge will affect the amount of information we expect to obtain from the designed experiment [15], or in other words, the stochastic model we choose. That is, the more prior information we have, the less information we would expect to be learned from the data. Thus, the *reference prior*, which in some sense should have a minimal effect relative to the data on the corresponding probabilistic inference, is very desirable.

Following the definition by Lindley [15], the ordered triple $\varepsilon = \{y, \Theta, p(y|\theta)\}$ is used to express an experiment, where y is the result of the experiment, Θ is the parameter space of the parameter of interest θ. After the prior density $p(\theta)$ is specified, the *expected information* from the experiment $\varepsilon = \{x, \Theta, p(x|\theta)\}$ is defined as

$$I^{\theta}\big\{\varepsilon, p(\theta)\big\} = \int p(y) \int p(\theta|y) log \frac{p(\theta|y)}{p(\theta)} d\theta dy \tag{7}$$

where the marginal density is $p(y) = \int p(y|\theta)p(\theta)d\theta$ and the posterior density is $p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$ by Bayes' formula (1).

Such a logarithmic measure of information possesses many important properties, such as invariance under different parameterization, non-negativity, concavity, etc [15]. Bernardo [2] gave a general review of such theoretical properties. Depending on whether the expected information (7) is finite or not in the limiting process, reference prior and posterior are defined as:

- When $I^{\theta}\{\varepsilon(\infty), p(\theta)\}$ is finite, it is a desirable measure for the amount of information under the prior specifications. Thus, the *reference prior* $\pi(\theta)$ is defined as the prior which maximizes the missing information $I^{\theta}\{\varepsilon, p(\theta)\}$ in the prior class C. Then the *reference posterior* density $\pi(\theta|y)$ can be obtained to be $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$ by the Bayes' formula.

- When $I^{\theta}\{\varepsilon(\infty), p(\theta)\}$ is not finite, the *reference prior* $\pi(\theta)$ is defined as the limit of a sequence of prior densities that maximizes the information we expect from an experiment.

As pointed out by Bernardo [2], there are several theoretical requirements of regularity concerning determining the reference prior by the above definition. For example, the class C of all admissible priors needs to be compact in order to guarantee the existence of a maxima. In practice, if such a requirement is not satisfied, usually an expanding convergent sequence of compact sets to C is constructed, and the limit of such a sequence is defined as the reference prior.

2 APPLYING REFERENCE ANALYSIS

2.1 Reference Algorithm for One Nuisance Parameter

Suppose the probability model is $p(y|\phi, \lambda)$, where $\theta \in \Theta, \lambda \in \Lambda(\phi)$. We are interested in the quantity of the parameter ϕ, therefore λ will be considered as the *nuisance parameter* [16], which may or may not depend on ϕ.

The common strategy [4] to eliminate the nuisance parameter is as follows: First conditioning on ϕ, determine the conditional reference prior $\pi(\lambda|\phi)$ by the Jeffrey's principle (5). Thus the model $p(y|\phi.\lambda)$ will be reduced to contain only one parameter. Next, two situations may occur:

1. If the conditional reference prior $\pi(\lambda|\phi)$ is proper, the one-parameter model can be obtained by integrating out the nuisance parameter through

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

   Then the marginal prior $\pi(\phi)$ can be determined by applying the algorithm again. Finally, the reference posterior will be obtained as $\pi(\phi|y) \propto p(y|\phi)\pi(\phi)$ by the Bayes' formula (2).

2. If the conditional reference prior $\pi(\lambda|\phi)$ is not proper, an increasing sequence of bounded approximations $\{\Lambda_i, i = 1, 2, \cdots\}$ to the nuisance parameter space $\Lambda(\phi)$ is needed as argued in the previous chapter. The corresponding reference posterior $\pi(\phi|y)$ is then obtained by taking the limit of the sequence $\{\pi_i(\phi|y), i = 1, 2, \cdots\}$.

Formally, the reference prior algorithm for one nuisance parameter is summarized by Bernardo & Ramon [4] in the following proposition.

Proposition 2.1 *: Let $p(y|\phi,\lambda), \phi \in \Phi, \lambda \in \Lambda(\phi)$ be the probability model with two real-valued parameters ϕ and λ, where ϕ is the quantity of interest, and suppose that the joint posterior distribution of (ϕ,λ) is asymptotically normal with covariance matrix $S(\tilde{\phi},\tilde{\lambda})$. Then, if $H(\phi,\lambda) = S^{-1}(\phi,\lambda)$,*

1. *the conditional reference prior of λ is $\pi(\lambda|\phi) \propto d_1^{-1}(\phi,\lambda) = h_{2,2}^{\frac{1}{2}}(\phi,\lambda)$, $\lambda \in \Lambda(\phi)$;*

2. *if $\pi(\lambda|\phi)$ is proper, the reference posterior distribution of ϕ given $\{x_1,\cdots,x_n\}$ is $\pi(\phi|x_1,\cdots,x_n) \propto \pi(\phi)\int_{\Lambda(\phi)}\{\prod p(x_l|\phi,\lambda)\}\pi(\lambda|\phi)d\lambda$, where the marginal reference prior of ϕ is $\pi(\phi) \propto exp\Big\{\int_{\Lambda(\phi)}\pi(\lambda|\phi)log\big[d_0^{-1}(\phi,\lambda)\big]d\lambda\Big\}$, $d_0(\phi,\lambda) = S_{1,1}^{\frac{1}{2}}(\phi,\lambda)$.*

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

$$\pi(\phi|x_1,\cdots,x_n) = \lim \pi_i(\phi|x_1,\cdots,x_n),$$

where $\pi(\phi|x_1,\cdots,x_n)$ is derived using $\Lambda_i(\phi)$ instead of $\Lambda(\phi)$.

Furthermore, if the nuisance parameter space $\Lambda(\phi) = \Lambda$ is independent of ϕ, and the functions d_0, d_1 can be factorized into separable forms of ϕ and λ, then this proposition can be simplified to the following corollary.

Corollary 2.2 *If the nuisance parameter space $\Lambda(\phi) = \Lambda$ is independent of ϕ, and the functions d_0, d_1 can be factorized into the forms $d_0^{-1}(\phi,\lambda) = a_0(\phi)b_0(\lambda)$, $d_1^{-1}(\phi,\lambda) =$*

$a_1(\phi)b_1(\lambda)$, *then the marginal and conditional reference priors are*

$$\begin{aligned} \pi(\phi) &\propto a_0(\phi) \\ \pi(\lambda|\phi) &\propto b_1(\lambda) \end{aligned}$$

and the reference posterior distribution of ϕ *given* $\{x_1, \cdots, x_n\}$ *is* $\pi(\phi|x_1, \cdots, x_n) \propto \pi(\phi) \int_{\Lambda(\phi)} \{ \prod p(x_l|\phi, \lambda)\} \pi(\lambda|\phi) d\lambda$.

2.2 Reference Prior and Posterior

We now derive the reference prior and posterior for the ratio of independent Poisson rates. Let y_1 and y_2 be the observed counts from each one of the two independent Poisson processes with means μ_1 and μ_2, respectively. By definition of the likelihood function (3), the probability model is expressed as

$$p(y_1, y_2|\mu_1, \mu_2) = p(y_1|\mu_1)p(y_2|\mu_2) = \frac{e^{-\mu_1}\mu_1^{y_1}}{y_1!} \cdot \frac{e^{-\mu_2}\mu_2^{y_2}}{y_2!} \tag{8}$$

where y_j, $j = 1, 2$ denotes the counts of two Poisson processes and $\mu_j = \lambda_j t_j,\ j = 1, 2$. Expressing the model in terms of the ratio of interest $\phi = \frac{\lambda_1}{\lambda_2}$, and μ_2 we have

$$p(y_1, y_2|\phi, \mu_2) = \frac{e^{\left(-\phi\mu_2\frac{t_1}{t_2}\right)}\left(\phi\mu_2\frac{t_1}{t_2}\right)^{y_1}}{y_1!} \cdot \frac{e^{-\mu_2}{\mu_2}^{y_2}}{y_2!}\ . \tag{9}$$

Applying equation (4) to the above joint density function we have the corresponding Fisher information matrix

$$F(\phi, \mu_2) = \left(\frac{t_1}{t_2}\right)\begin{bmatrix} \frac{\mu_2}{\phi} & 1 \\ 1 & \frac{\frac{t_2}{t_1}+\phi}{\mu_2} \end{bmatrix}. \tag{10}$$

It follows by the definition of $S(\phi, \mu_2)$ in Proposition 2.1

$$S(\phi, \mu_2) = F^{-1}(\phi, \mu_2) = \begin{bmatrix} \frac{\phi\left(\frac{t_2}{t_1}+\phi\right)}{\mu_2} & -\phi \\ -\phi & \mu_2 \end{bmatrix}\ . \tag{11}$$

Thus, according to the asymptotic theory [4], we conclude that the joint posterior of (ϕ, μ_2) is asymptotically normal with covariance matrix $S(\tilde{\phi}, \tilde{\mu_2})$, and

1. the marginal asymptotic posterior of ϕ is normal with standard deviation $d_0(\tilde{\phi}, \tilde{\mu_2})$,

$$d_0(\phi, \mu_2) = \frac{1}{\sqrt{n}}\left[\frac{\phi\left(\frac{t_2}{t_1}+\phi\right)}{\mu_2}\right]^{1/2};$$

2. the conditional asymptotic posterior of μ_2 given ϕ is normal with standard deviation $d_1(\phi, \tilde{\mu_2})$,

$$d_1(\phi, \mu_2) = \Big(\frac{\mu_2}{\frac{t_2}{t_1} + \phi}\Big)^{1/2} .$$

Notice the parameter space of $\phi = \frac{\lambda_1}{\lambda_2}$ is $\Lambda(\phi) = (0, +\infty)$, which is independent of $\phi = \frac{\lambda_1}{\lambda_2}$ and functions d_0^{-1} and d_1^{-1} can be factorized as

$$d_0^{-1} = \frac{1}{\sqrt{\phi\big(\frac{t_2}{t_1} + \phi\big)}} \cdot \sqrt{\mu_2} = a_0(\phi) b_0(\mu_2)$$

$$d_1^{-1} = \sqrt{\frac{t_2}{t_1} + \phi} \cdot \sqrt{\frac{1}{\mu_2}} = a_1(\phi) b_1(\mu_2).$$

Hence, according to the Corollary 2.2, the marginal reference prior for ϕ is

$$\pi(\phi) \propto a_0(\phi) = \frac{1}{\sqrt{\phi\big(\frac{t_2}{t_1} + \phi\big)}} \tag{12}$$

and the conditional reference prior of μ_2 given ϕ is

$$\pi(\mu_2|\phi) \propto b_1(\mu_2) = \sqrt{\frac{1}{\mu_2}} \, . \tag{13}$$

It is interesting to note that this conditional reference prior is the same as the marginal reference prior for μ_2 if we only apply the reference algorithm to the second Poisson process. Hence, we conclude that the two parameters ϕ and μ_2 are independent. Combining the marginal (12) and the conditional priors (13), the *joint reference prior* is then determined as

$$\pi(\phi, \mu_2) = \pi(\phi)\pi(\mu_2|\phi) \propto \phi^{-1/2}\Big(\frac{t_2}{t_1} + \phi\Big)^{-1/2} \mu_2^{-1/2} \, . \tag{14}$$

According to Proposition 2.1, by eliminating the nuisance parameter μ_2 via integration, the marginal posterior of ϕ is

$$\begin{aligned}
\pi(\phi|y_1,y_2) &\propto \pi(\phi)\int_{\Lambda(\phi)} p(y_1,y_2|\phi,\mu_2)\pi(\mu_2|\phi)d\mu_2 \\
&= \pi(\phi)\int_{\Lambda} \frac{e^{\left(-\phi\mu_2\frac{t_1}{t_2}\right)}\left(\phi\mu_2\frac{t_1}{t_2}\right)^{y_1}}{y_1!}\cdot\frac{e^{-\mu_2}\mu_2{}^{y_2}}{y_2!}\sqrt{\frac{1}{\mu_2}}d\mu_2 \\
&\propto \phi^{-1/2}\Big(\frac{t_2}{t_1}+\phi\Big)^{-1/2}\int_0^{+\infty} e^{-\left(\frac{t_1}{t_2}\phi+1\right)\mu_2}\cdot\phi^{y_1}\cdot\mu_2^{y_1+y_2-1/2}d\mu_2 \ .
\end{aligned}$$

Since ϕ and μ_2 are independent "ϕ^{y_1}" can be taken out of the integration sign. Treating ϕ as a constant during the integration, we have

$$\begin{aligned}
\pi(\phi|y_1,y_2) &\propto \Big(\frac{t_2}{t_1}+\phi\Big)^{-1/2}\phi^{y_1-1/2}\int_0^{+\infty} e^{-\left(\frac{t_1}{t_2}\phi+1\right)\mu_2}\cdot\mu_2^{y_1+y_2-1/2}d\mu_2 \\
&\propto \frac{\Gamma(y_1+y_2+1/2)}{\left(\frac{t_2}{t_1}+\phi\right)^{(y_1+y_2+1/2)}}\Big(\frac{t_2}{t_1}+\phi\Big)^{-1/2}\phi^{y_1-1/2} \\
&\propto \frac{\phi^{y_1-1/2}}{\left(\frac{t_2}{t_1}+\phi\right)^{y_1+y_2+1}} \ .
\end{aligned}$$

If we define

$$\omega = \frac{\phi}{\frac{t_2}{t_1}+\phi} = \frac{\mu_1}{\mu_1+\mu_2}$$

then $\phi = \frac{t_2}{t_1}\left(\frac{\omega}{1-\omega}\right)$ and the posterior distribution of ω can be determined using the change of variable technique. After some algebra, we obtain

$$\pi(\omega|y_1,y_2) = \pi(\phi|y_1,y_2)\left|\frac{d\phi}{d\omega}\right| \propto \omega^{y_1-1/2}(1-\omega)^{y_2-1/2} \ .$$

This implies the posterior density function is

$$\pi(\omega|y_1,y_2) = Beta(\omega|y_1+1/2, y_2+1/2)$$

or

$$\omega|y_1,y_2 \sim Beta(\alpha = y_1+1/2, \beta = y_2+1/2) \ . \tag{15}$$

Although the reference posterior distribution is a function of ω, not the parameter of interest ϕ, it won't affect upcoming inferences on ϕ. In the next chapter, inferences on ϕ such as Bayesian credible intervals, HPD intervals and Maximum a posteriori estimate will be constructed based on this posterior.

Instead of working with ω we could find the posterior of ρ where

$$\rho = \frac{t_1}{t_2} \cdot \frac{y_2 + 1/2}{y_1 + 1/2} \cdot \phi \ .$$

Solving for ϕ we find $\phi = \frac{t_2}{t_1} \cdot \frac{y_1+1/2}{y_2+1/2} \cdot \rho$ and the Jacobian to be $\left|\frac{d\phi}{d\rho}\right| = \frac{t_2}{t_1} \cdot \frac{y_1+1/2}{y_2+1/2}$. Using the change of variable technique, the posterior of ρ is

$$\begin{aligned} p(\rho|y_1, y_2) &\propto p(\phi|y_1, y_2) \cdot \left|\frac{d\phi}{d\rho}\right| \\ &\propto \frac{\left(\frac{t_2}{t_1} \cdot \frac{y_1+1/2}{y_2+1/2} \cdot \rho\right)^{y_1-1/2}}{\left(\frac{t_2}{t_1} + \frac{t_2}{t_1} \cdot \frac{y_1+1/2}{y_2+1/2} \cdot \rho\right)^{y_1+y_2+1}} \cdot \frac{t_2}{t_1} \cdot \frac{y_1 + 1/2}{y_2 + 1/2} \\ &\propto \frac{\rho^{y_1-1/2}}{\left(1 + \frac{y_1+1/2}{y_2+1/2} \cdot \rho\right)^{y_1+y_2+1}} \ . \end{aligned}$$

If

$$\nu_1 = 2(y_1 + 1/2) \ , \quad \nu_2 = 2(y_2 + 1/2)$$

then the posterior distribution of ρ is an F distribution with degrees of freedom ν_1 and ν_2,

$$\rho|y_1, y_2 \sim F\big(\nu_1 = 2(y_1 + 1/2), \nu_2 = 2(y_2 + 1/2)\big). \tag{16}$$

This result coincides with the one in [17] where the noninformative priors $\pi_j(\lambda_j) \propto \lambda_j^{-1/2}$, $j = 1, 2$ for each Poisson process were used. This means that the two posteriors (15)(16) are in fact equivalent with each other. Hence all inferences on ϕ based on the posterior of ω will be the same as the inferences based on ρ.

2.3 Posterior Simulation

Although the reference posterior has been determined in terms of ω, a function of the parameter of interest ϕ, it clearly follows that Bayesian credible interval estimates could be obtained through transformation of variables. However, the shape of the posterior is not straightforward, mainly because the *probability density function* (PDF) in general does not comply with the variable transformations. In this section, we will discuss how to obtain the posterior shape through simulations.

The joint posterior of ϕ and μ_2 can be determined by applying Bayes' formula (2) to the likelihood function and the joint prior,

$$\begin{aligned}
\pi(\phi, \mu_2|y_1, y_2) &\propto \pi(y_1, y_2|\phi, \mu_2) \cdot \pi(\phi, \mu_2) \\
&\propto \frac{e^{\left(-\phi\mu_2\frac{t_1}{t_2}\right)}\left(\phi\mu_2\frac{t_1}{t_2}\right)^{y_1}}{y_1!} \cdot \frac{e^{-\mu_2}{\mu_2}^{y_2}}{y_2!} \cdot \phi^{-1/2}\Big(\frac{t_2}{t_1}+\phi\Big)^{-1/2}{\mu_2}^{-1/2} \\
&\propto e^{-\left(\phi\frac{t_1}{t_2}+1\right)\mu_2}\phi^{y_1-1/2}\Big(\frac{t_2}{t_1}+\phi\Big)^{-1/2}{\mu_2}^{y_1+y_2-1/2}
\end{aligned}$$

which can be factored as

$$\pi(\phi, \mu_2|y_1, y_2) \propto Gamma(\mu_2|y_1+1/2, 1)\cdot Gamma\Big(\phi|y_1, \frac{1}{\mu_2\frac{t_1}{t_2}}\Big)\cdot Beta\Big(\frac{\phi}{\frac{t_2}{t_1}+\phi}|3/2, 1\Big)\ .$$

This factorization suggests that samples of ϕ and μ_2 can be drawn as a *Markov Chain* [18] from the joint posterior density as the following,

$$\mu_2|y_2 \sim Gamma(\mu_2|y_2+1/2, 1)\ ;$$

$$p(\phi|\mu_2, y_1) \propto q(\phi|\mu_2, y_1) = Gamma\Big(\phi|y_1, \frac{1}{\mu_2\frac{t_1}{t_2}}\Big)\cdot Beta\Big(\frac{\phi}{\frac{t_2}{t_1}+\phi}|3/2, 1\Big)\ .$$

Note for the second stochastic process $\phi|\mu_2, y_1$ the unnormalized functional density form $q(\phi|\mu_2, y_1)$ is sufficient for the purpose of simulation.

Instead of solving the marginal posterior $p(\phi|y_1, y_2)$ analytically via the integration $p(\phi|y_1, y_2) = \int_\Lambda p(\phi, \mu_2|y_1, y_2)d\mu_2$, we can do the following approximation,

$$\widehat{p}(\phi|y_1, y_2) = \frac{1}{m}\sum_{i=1}^{m} p\big(\phi|\mu_2^{(i)}\big) = \frac{1}{C_0}\frac{1}{m}\sum_{i=1}^{m} q\big(\phi|\mu_2^{(i)}\big)$$

where $\mu_2^{(i)}, i = 1, 2, \cdots, m$ are the i.i.d. random samples of μ_2 from the density $Gamma(\mu_2|y_2 + 1/2, 1)$; C_0 is the normalizing constant to make sure that the result is actually a probability density function, that is, $C_0 = \int_0^{+\infty} \frac{1}{m}\sum_{i=1}^{m} q\big(\phi|\mu_2^{(i)}\big)d\phi$.

There are several methods to calculate the normalizing constant C_0 numerically. By the method of numerical integration, C_0 can be approximated as

$$C_0 = \lim_{\Delta\phi\to 0}\sum_{j=0}^{+\infty}\Big[\frac{1}{m}\sum_{i=1}^{m} q\big(\phi^{(j)}|\mu_2^{(i)}\big)\Big]\Delta\phi \approx \sum_{j=0}^{M}\Big[\frac{1}{m}\sum_{i=1}^{m} q\big(\phi^{(j)}|\mu_2^{(i)}\big)\Big]\Delta\phi \qquad (17)$$

for a sufficiently large integer M and sufficiently small $\Delta\phi$, where $\phi^{(j)} = \Delta\phi \cdot j, j = 0, 1, 2, \cdots$. Empirically, we can take $M = 10,000$ and $\Delta\phi$ so that $P_{\phi|y_1,y_2}(\phi < \phi^{(M)}) = P_{\omega|y_1,y_2}\Big(\omega < \frac{\phi^{(M)}}{\frac{t_2}{t_1}+\phi^{(M)}}\Big) \geq 1 - \xi$, where $\xi = 0.0001$ to guarantee the accuracy of the approximation.

This procedure is recognized as the reduced form of the method of *substitution sampling* [9] where the conditional probability density $p(\mu_2|y_2)$ doesn't depend on the parameter ϕ. The approximated density $\widehat{p}(\phi|y_1, y_2)$ converges to $p(\phi|y_1, y_2)$ by L^1 [11]. Therefore, if sufficiently large samples are drawn, we expect this procedure to yield satisfactory accuracy.

2.4 Approximating Posterior Using Profile Likelihood Method

Another common-used method for eliminating the nuisance parameter is the *profile likelihood* (P-L) [1]. We are interested in investigating how such an approximation behaves under various situations compared to the simulation method which we believe has the best accuracy.

Consider the joint reference posterior $\pi(\phi, \mu_2|y_1, y_2)$ as the likelihood function of the parameters ϕ and μ_2 and instead of integrating out μ_2, the profile likelihood method of approximation simply substitutes the nuisance parameter with its MLE. That is,

$$\widetilde{\pi}(\phi|y_1, y_2) = \sup \pi(\phi, \mu_2) = \pi(\phi, \widehat{\mu_2}|y_1, y_2)$$

where $\widetilde{\pi}(\phi|y_1, y_2)$ is the *profile-likelihood posterior* of ϕ. Since in the Poisson process $\widehat{\mu_2} = y_2$, the approximate reference posterior is

$$\pi(\phi|y_1, y_2) = \frac{1}{C_1}\phi^{y_1-1/2}\Big(\frac{t_2}{t_1} + \phi\Big)^{-1/2} e^{-\phi\frac{t_1}{t_2}y_2}$$

where C_1 is the normalizing constant, which can be approximated by numerical integration as was presented in (17).

An R program (.1) has been developed to simulate the posterior via the method proposed in the last section as well as the computation of the profile-likelihood posterior. The results of the simulation indicate that under the same condition (without loss of generality, set $t_1 = t_2$) the approximate posterior by the profile likelihood method behaves more concentrated around its mode compared to the simulated posterior, as demonstrated in Figure 1. This can be explained by the fact that the profile likelihood simply substitutes μ_2 with its MLE $\widehat{\mu_2} = y_2$, it doesn't account for

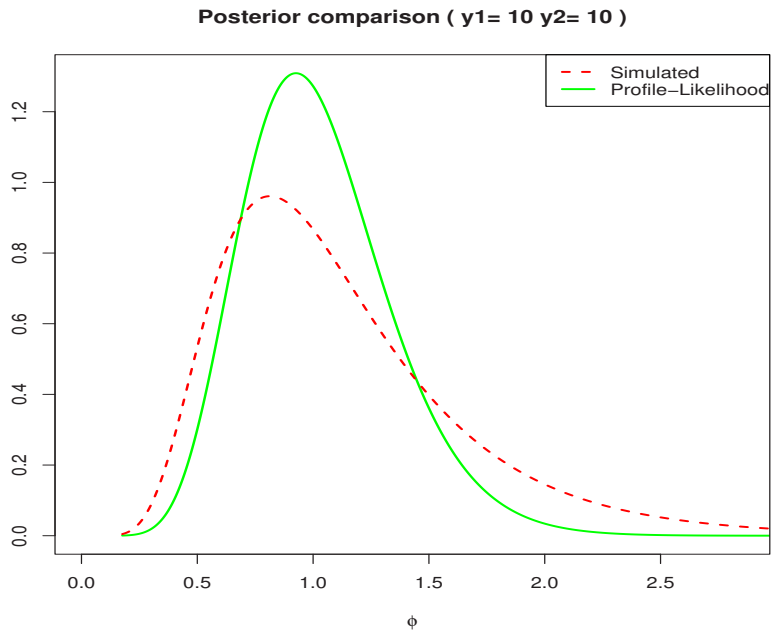

Figure 1: Comparison of Simulated and P-L Posteriors

the variability of μ_2 in the posterior distribution [1].

As y_2 increases, the two posteriors become closer to each other (exemplified in Figure 1). This is because for the Poisson process the variance of the MLE of μ_2 increases by y_2. As a result, the profile likelihood method will catch more uncertainty from μ_2. It has been argued by Berger [1] that the integrated likelihood methods should be encouraged rather than the profile likelihood method, since the profile likelihood often leads to misleading behaviors.

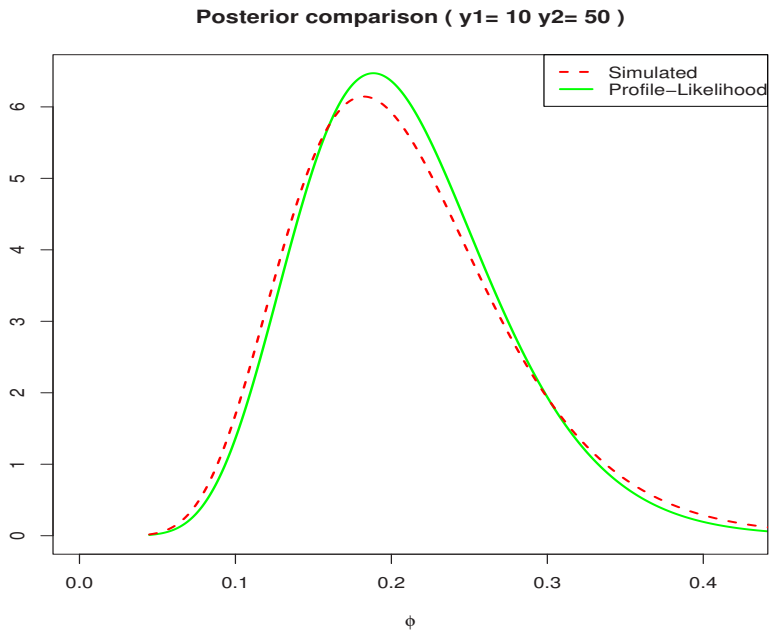

Figure 2: Comparison of Simulated and P-L Posteriors

3 POSTERIOR-BASED INFERENCES

Although a graphical presentation of the entire posterior is preferable in general, summary statistics, such as point estimates and interval estimates, which portray important features of the posterior are sometimes sufficient for the usage. In this chapter, we will first develop methods for constructing the $100(1-\alpha)\%$ Bayesian credible intervals and highest probability density intervals, then derive the maximum a posterior estimate.

3.1 Bayesian Credible Interval

Under the Frequentist paradigm, the true parameter is assumed to be fixed not random. That is, the parameter either belongs to the confidence interval or doesn't belong to it. Therefore, we need to take caution when interpreting confidence intervals.

From the Bayesian perspective, the parameter itself becomes a random variable as it is assumed to follow a particular prior distribution. It is advantageous to interpret the interval estimates in a probabilistic manner.

Definition 3.1 *Let $\pi(\theta|D)$ be the posterior distribution. A credible set is any set C such that $P_{\pi(\theta|y)}(C) = 1-\alpha$, where D denotes the data. [13]*

If the posterior distribution is continuous and unimodal, the credible set is usually constructed as an interval, which then becomes a *credible interval*. The difference between the confidence interval and the credible interval lies in the interpretation. Since the posterior is determined probabilistically by the data, the credible interval

measures the probability that the parameter θ in C is at least $1-\alpha$. On the other hand, the confidence interval means: Before observations are taken, the probability that θ belongs to C is at least $1-\alpha$.

If the cumulative posterior density $\Pi(\theta|D)$ is also available algebraically, a Bayesian credible interval can be easily constructed. First, we calculate $\theta^{(\alpha/2)}$ and $\theta^{(1-\alpha/2)}$ such that

$$\Pi\big(\theta^{(\alpha/2)}|D\big) = \alpha/2 \ , \ \Pi\big(\theta^{(1-\alpha/2)}|D\big) = 1-\alpha/2 \ .$$

Then, a $100(1-\alpha)\%$ credible interval for θ is $CI = \big(\theta^{(\alpha/2)}, \theta^{(1-\alpha/2)}\big)$.

One interesting feature of credible intervals is that they are invariant under a nonlinear transformation [5]. Thus if we assume $\eta = h(\theta)$, where h need not to be a linear function of θ, the credible interval of η can be obtained by computing $h\big(\theta^{(\alpha/2)}\big)$ and $h\big(\theta^{(1-\alpha/2)}\big)$.

Recall the definition in Section 2.2 that

$$\phi = \frac{t_2}{t_1}\Big(\frac{\omega}{1-\omega}\Big)$$

where ω follows a *Beta* distribution with $\alpha = y_1 + 1/2$ and $\beta = y_2 + 1/2$. Thus, an equal-tail $100(1-\alpha)\%$ credible interval for ϕ is obtained as

$$CI = \Big(\frac{t_2}{t_1}\cdot\frac{\omega^{(\alpha/2)}}{1-\omega^{(\alpha/2)}}, \ \frac{t_2}{t_1}\cdot\frac{\omega^{(1-\alpha/2)}}{1-\omega^{(1-\alpha/2)}}\Big)$$

where $\Pi_{\omega|y_1,y_2}\big(\omega^{(\alpha/2)}\big) = \alpha/2$ and $\Pi_{\omega|y_1,y_2}\big(\omega^{(1-\alpha/2)}\big) = 1-\alpha/2$.

3.2 Highest Probability Density Interval

Theoretically there are an infinite number of credible intervals with probability level $100(1-\alpha)\%$ by Definition 3.1. It is quite questionable to just take the equal-tailed. Unless under particular conditions they are not necessarily the shortest (Figure 3).

Definition 3.2 *Let $\pi(\theta)$ be the density function of a random variable θ. Then the $100(1-\alpha)\%$ highest probability density (HPD) interval is the subset $R(\pi_\alpha)$ of the parameter space of θ such that*

$$R(\pi_\alpha) = \{\theta : \pi(\theta) \geq \pi_\alpha\}$$

where π_α is the largest constant such that $P(\theta \in R(\pi_\alpha)) \geq 1-\alpha$. [7]

This definition guarantees that the density for every point inside the HPD interval is greater than that for every point outside the interval. Furthermore, it has been proved by Box and Tiao [5] that the HPD interval is of the shortest length for a given probability content $1-\alpha$. Thus, theoretically it is desirable to obtain the HPD interval estimates from the posterior distribution.

The central difficulty is that HPD intervals are difficult to determine analytically. For any unimodal, symmetric distribution, the HPD interval coincides with the equal-tail credible intervals [7]. However, under other conditions it is difficult to find the posterior probability of π_α, much less solve for the required probability.

Hence the HPD intervals generally need to be solved computationally. One simple algorithm is to start with a value of π_α and compute the posterior probability of the

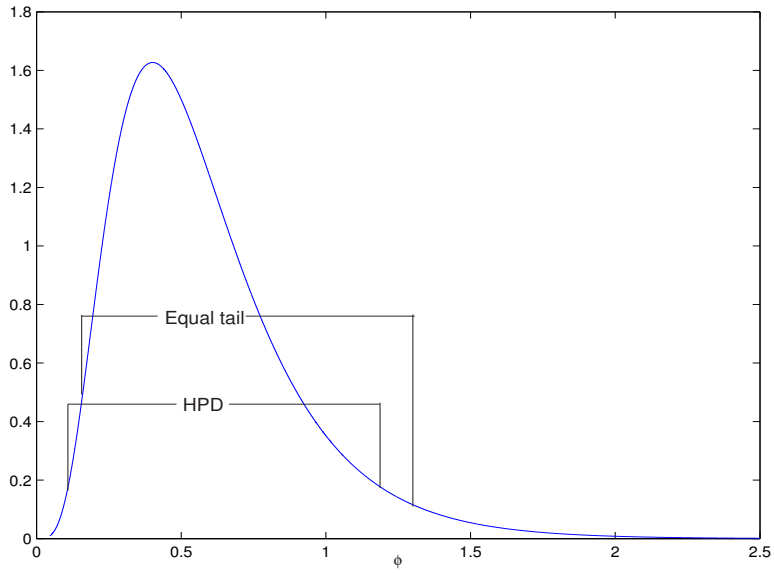

Figure 3: HPD and Equal-tailed Intervals

resulting set by numerical integration. If the probability is too small, decrease π_α. If the probability is too large, increase π_α.

3.3 Monte Carlo Estimation of the HPD Interval

Monte Carlo techniques are widely used in simulation studies nowadays. Especially in Bayesian statistics, computation is really facilitated through implementation of Monte Carlo methods. In this section, we will present a method to estimate the HPD interval using a Monte Carlo technique [7].

Let $\pi(\theta|D)$ and $\Pi(\theta|D)$ be the marginal posterior density function and the marginal posterior cumulative distribution function (CDF) of θ respectively, where D denotes data. Assume that $\pi(\theta|D)$ is unimodal for convenience and also assume that θ can be generated from $\pi(\theta|D)$ using a direct random sampling scheme. Let $\{\theta_i, i = 1, 2, \cdots, n\}$ be a Monte Carlo sample from $\pi(\theta|D)$, and let θ_j be the jth smallest of $\{\theta_i\}$. Denote

$$R_j(n) = \left(\theta_{(j)}, \theta_{(j+[(1-\alpha)n])}\right)$$

for $j = 1, 2, \cdots, n - \big[(1-\alpha)n\big]$, where [.] denotes the integer part.

Theorem 3.3 *Let $\{\theta_i, i = 1, 2, \cdots, n\}$ be an ergodic Monte Carlo sample from $\pi(\theta|D)$ and let $R_{j^*}(n) = \left(\theta_{(j^*)}, \theta_{(j^*+[(1-\alpha)n])}\right)$, where j^* is chosen so that*

$$\theta_{(j^*+[(1-\alpha)n])} - \theta_{(j^*)} = \min\left(\theta_{(j+[(1-\alpha)n])} - \theta_{(j)}\right)$$

That is, $R_{j^*}(n)$ has the smallest interval width among all $R_{j^*}(n)$'s. If $\pi(\theta|D)$ is unimodal, then we have

$$R_{j^*}(n) \to R(\pi_\alpha) \text{ almost surely as } n \to \infty,$$

where $R(\pi_\alpha)$ has defined in the previous section. Thus, to find a $100(1-\alpha)\%$ HPD interval, we look at all the $100(1-\alpha)\%$ credible intervals in the sample and then take the one with the smallest interval width.

Unlike a Bayesian credible interval, the HPD interval is not invariant under a nonlinear transformation [5]. Thus for $\eta = h(\theta)$, the HPD interval of η cannot be computed as $\left(h(\theta_{(j^*)}), h(\theta_{(j^*+[(1-\alpha)n])})\right)$ if h is not a linear function. In order to overcome such a difficulty, the above theorem can be extended to the following corollary.

Corollary 3.4 *Let* $\{\theta_i, i = 1, 2, \cdots, n\}$ *be an ergodic Monte Carlo sample from* $\pi(\theta|D)$. *Also let* $\eta_i = h(\theta_i)$ *and the* $\eta_{(i)}$ *be the ordered values of the* η_i. *Then a* $100(1-\alpha)\%$ *HPD interval of* η *can be approximated by*

$$R_{j^*}(n) = \left(\eta_{(j^*)}, \eta_{(j^*+[(1-\alpha)n])}\right)$$

where j^* *is chosen so that*

$$\eta_{(j^*+[(1-\alpha)n])} - \eta_{(j^*)} = \min\left(\eta_{(j+[(1-\alpha)n])} - \eta_{(j)}\right) .$$

According to this corollary, the HPD interval for $\phi = \frac{t_2}{t_1} \cdot \frac{\omega}{1-\omega}$ can be calculated through a random sample $\{\omega_i, i = 1, 2, \cdots, n\}$, where ω_i's are i.i.d. samples from the reference posterior distribution $Beta(\omega|y_1 + 1/2, y_2 + 1/2)$. An R program (.2) has been designed to implement this algorithm.

3.4 Maximum a Posteriori

With the reference posterior (15) available, the method of *maximum a posteriori* (MAP) can be applied to obtain a point estimate of an observed quantity based on the empirical data. Let $L(D|\theta)$ be the likelihood function and $\pi(\theta|D)$ be the posterior density function, where θ is the unknown parameter which we would like to make inference on, D denotes the data. It is known that the MLE is

$$\widehat{\theta}_{ML}(D) = \arg\max_{\theta} \big(L(D|\theta)\big)$$

Definition 3.5 *The method of maximum a posteriori estimates θ as the mode of the posterior distribution of this random variable,*

$$\widehat{\theta}_{MAP}(D) = \arg\max_{\theta} \big(\pi(\theta|D)\big) = \arg\max_{\theta} \big(L(D|\theta) \cdot \pi(\theta)\big) \ .$$

This definition implies, if the prior distribution of θ is uniform, the MAP estimate of θ coincides with the MLE estimate. The point where highest probability density occurs in Figure 4 is the MAP estimate of the posterior.

Recall in estimating the ratio of two independent Poisson processes, the transformed reference posterior (15) is

$$\omega \sim Beta(\omega|\alpha = y_1 + 1/2, \beta = y_2 + 1/2)$$

where $y_j \in \{0, 1, \cdots\}, j = 1, 2$. Theoretically [6], the mode of a $Beta$ distribution occurs at

$$\omega_M = \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{y_1 - 1/2}{y_1 + y_2 - 1}$$

with the restriction that $\alpha, \beta \geq 1$.

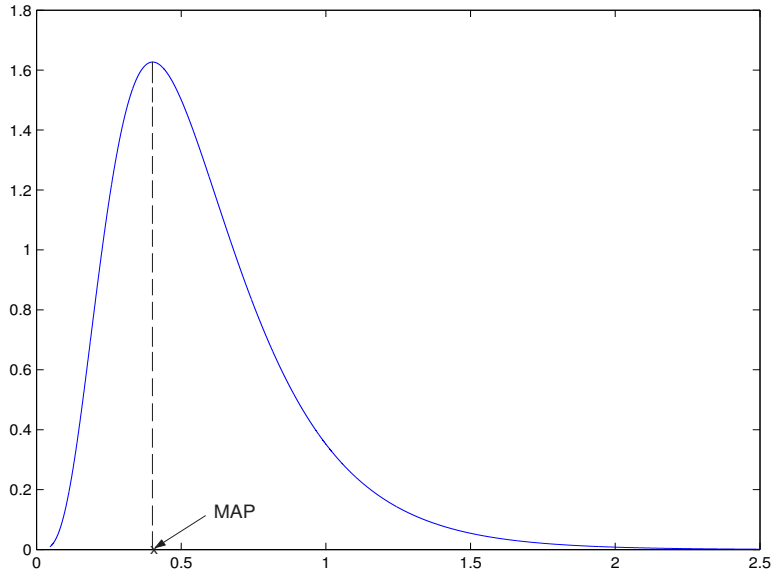

Figure 4: Maximum a Posteriori

Since the mode for a unimodal and continuous density is invariant under one-to-one continuous transformations [5], then the MAP estimate of $\phi = \frac{t_2}{t_1} \cdot \frac{\omega}{1-\omega}$ is

$$\widehat{\phi}_{MAP}(y_1, y_2) = \frac{t_2}{t_1} \cdot \frac{\omega_M}{1-\omega_M} = \frac{t_2}{t_1} \cdot \frac{y_1 - 1/2}{y_2 - 1/2}$$

where $y_j \in \mathbb{Z}^+, j = 1, 2$. Notice that the MAP estimate makes a little correction on the MLE estimate, which is $\widehat{\phi}_{MLE}(y_1, y_2) = \frac{y_1}{y_2}$, but it still fails to work when a zero count is observed.

It has been argued that although the MAP estimation uses a prior distribution, it is not generally recognized as a Bayesian method. This is because MAP estimates are point estimates, whereas Bayesian methodology is characterized by the use of distributions to summarize data and draw inferences. Bayesian methods tend to report interval estimates based on the posterior, rather than the posterior mode.

4 FREQUENTIST COVERAGE STUDY

4.1 Probability Matching Prior

Recall the major discrepancy between the Frequentists and Bayesianists lies in the interpretation of interval estimates. However, it is interesting that the Frequentist coverage probabilities of Bayesian credible intervals derived from reference posterior distributions are usually very close to their posterior probabilities [3].

Formally, if $t_\alpha = t_\alpha(D)$ denotes the $1-\alpha$ quantile which corresponds to the reference posterior $\pi(\phi|D)$, so that

$$P\big[\phi \leq |D\big] = \int_{\phi \leq t_\alpha(D)} \pi(\phi|D)d\phi = 1-\alpha \ ,$$

then the coverage probability of the $100(1-\alpha)\%$ reference posterior credible interval $(-\infty, t_\alpha)$,

$$P\big[t_\alpha \geq \phi|\phi\big] = \int_{t_\alpha(D)\geq\phi} p(D|\phi)dD$$

often satisfies

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

while, for most priors, this asymptotic approximation is only $O(n^{-\frac{1}{2}})$. This means that the reference prior is often a *probability matching* prior, that is, a prior for which the coverage probabilities of one-sided posterior credible intervals are asymptotically closer to their posterior probabilities.

Recall the joint reference prior (14) in estimating the parameter of interest ϕ is

$$\pi(\phi,\mu_2) = \pi(\phi)\pi(\mu_2|\phi) \propto \phi^{-1/2}\Big(\frac{t_2}{t_1}+\phi\Big)^{-1/2}\mu_2^{-1/2} \ .$$

It can be verified that this joint reference prior satisfies the differential equation [8] for probability matching in multiparameter models

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

where $\theta = \{\phi, \mu_2\}$, and

$$\eta(\theta) = \frac{S(\phi, \mu_2)\nabla}{\sqrt{\nabla^t S(\phi, \mu_2)\nabla}} \ ,$$

where $\nabla = \{1, 0\}^t$, and $S(\phi, \mu_2)$ is given by (11).

Hence, we conclude the reference prior for estimating the ratio of two Poisson rates is actually a probability matching prior. This means that satisfactory Frequentist coverage behavior of the interval estimates based on the reference posterior (15) is guaranteed.

4.2 Exact Coverage

Since two independent Poisson processes are involved and they are discrete distributions, then we could actually compute the exact coverage probability for the $100(1-\alpha)\%$ interval estimates under the fixed parameter values of μ_1 and μ_2 by

$$\sum_{y_1=0}^{\infty}\sum_{y_2=0}^{\infty}\frac{e^{-\mu_1}\mu_1^{y_1}}{y_1!}\frac{e^{-\mu_2}\mu_2^{y_2}}{y_2!}I(y_1,y_2)$$

where $I(y_1,y_2)$ equals 1 if the interval contains the ratio $\phi=\frac{\mu_1}{\mu_2}$ when $y_j, j=1,2$ and equals 0 if the ratio is not covered by the interval estimate.

Hartigan [12] has showed that the coverage probabilities of two-sided Bayesian posterior credible intervals have satisfactory Frequentist coverage property asymptotically by $O(n^{-1})$ for all sufficiently regular prior functions. Thus, the $(1-\alpha)$ HPD interval estimate, which is a special case of credible intervals, should have exact coverage rate of $(1-\alpha)$.

However, as in our study the HPD interval estimate always has shortest length among all interval estimates with the same probability content, we would like to actually verify its Frequentist coverage properties through simulations. An R program has been developed to implement such an idea. Table 1 summarizes the results after running the program for a wide range of values of μ_1 and μ_2. Since the average coverage rate is close to $(1-\alpha)$ under different significance levels, and also the variance of coverage rates is relatively small, we conclude that HPD interval estimates have satisfactory Frequentist coverage properties.

Table 1: Coverage Property Summary of HPD Intervals

Range of μ_1, μ_2	$1-\alpha$	Mean coverage	RMSE	Minimum coverage	Proportion below $1-\alpha$
7(3)70	.80	0.795	0.013	0.735	0.660
	.85	0.846	0.013	0.805	0.585
	.90	0.898	0.007	0.854	0.727
	.95	0.947	0.008	0.923	0.619
	.99	0.988	0.003	0.978	0.689
.1(.1)10	.80	0.794	0.014	0.772	0.668
	.85	0.851	0.009	0.829	0.587
	.90	0.903	0.018	0.835	0.541
	.95	0.952	0.012	0.903	0.562
	.99	0.991	0.009	0.964	0.539

*: 1000 samples of (y_1, y_2) are drawn.

BIBLIOGRAPHY

[1] Berger J. O., Liseo B. and Wolpert R. L. (1999), Integrated likelihood methods for eliminating nuisance parameters. *Statistical Sci.*, **14**, 1-28.

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

[3] Bernardo J. M. (2005), *Handbook of Statistics* **25** (Dipak, D. and Rao, C. R., eds.), Amsterdam: Elsevier.

[4] Bernardo J. M. and Ramon J. M. (1998), An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *J. Roy. Statist. Soc. D*, **47**, 101-135.

[5] Box G. E. P. and Tiao G. C. (1992), *Bayesian Inference in Statistical Analysis*. New York: Wiley.

[6] Casella G. and Berger R. L. (1990), *Statistical Inference, 2nd Edition*. Pacific Grove: Duxbury Press.

[7] Chen M.-H. and Shao Q.-M. (1999), Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Am. Statist. Ass.*, **8**, 69-92.

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

[9] Gelfand A. E. and Smith A. F. M. (1990), Sampling based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398-409.

[10] Gelman A., Carlin J., Stern H. and Rubin D. (2004), *Bayesian Data Analysis*, second edition. London: Chapman & Hall.

[11] Glick N. (1974), Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Mathematica*, **6**, 61-74.

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

[13] Hyndman R. J. (1996), Computing and graphing highest density regions. *The American Statistician*, **50**, 120-126.

[14] Kass R. E. and Wasserman L. (1996), The selection of prior distributions by formal rules. *J. Am. Statist. Ass.*, **91**, 1343-1370.

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

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

[17] Price R. M. and Bonett D. G. (2000), Estimating the Ratio of two Poisson rates. *Computatioanl Statistics and Data Analysis*, **34**, 345-356.

[18] Tanner M. A. (1993), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, third edition. New York: Springer-Verlag.

APPENDICES

.1 R Code of Monte Carlo Estimation of HPD Interval

```
## Monte Carlo HPD Interval Approximation


# Input posterior parameters x1,x2
hpd.mc <- function(x1,x2,t1=10,t2=10){


# Significance level alpha <- 0.05 ### specify significance level here


# Sample from transformed posterior, a beta density
n <- 1000   ### sample size
mc.phi <- t2/t1*sort(1/(1-rbeta(n,x1+.5,x2+.5))-1)
### sample, transform and sort


# Compute empirical length & highest lower rank of interval
estimates with the probability content of 1-alpha
hpd.length <- floor((1-alpha)*n)
hpd.lower.high <- n-hpd.length


# Form candidates of HPD interval
can.lower <- mc.phi[1:hpd.lower.high]
can.upper <- mc.phi[(1+hpd.length):n]
```

```
# Determine the order where the shortest interval occurs
shortest.stack <- 1     ### initial value
for (i in 2:hpd.lower.high){
  if ((can.upper-can.lower)[i] <  (can.upper-can.lower)[shortest.stack])
  shortest.stack <- i
}
hpd.rank <- shortest.stack

# Output the approximated HPD interval
hpd.appr <- c(can.lower[hpd.rank], can.upper[hpd.rank])
hpd.appr
}
```

.2 R Code for Simulating Posterior

```
### Posterior simulation from by (1) Markov Chain and (2) Profile-likelihood


## Input x1,x2 - two counts x1,x2 and t1,t2
post.sim <- function (x1=10,x2=10,t1=10,t2=10){


## Parameter setup
ipsilon <- 0.0001   ### tail-significance
N <- 1000   ### number of bins
m <- 1000   ### number of nuisance samples


## Lower & upper bounds of the ratio 'phi'
lb.phi <- t2/t1*(1/(1-qbeta(ipsilon,x1+.5,x2+.5))-1)
ub.phi <- t2/t1*(1/(1-qbeta((1-ipsilon),x1+.5,x2+.5))-1)
phi = seq(lb.phi,ub.phi,length=N)   ### vector of "phi"


#(1) Random samples of the nuisance parameter - lambda2
mc.lambda2 = rgamma(m,x2+.5,scale=1)


## Approximate posterior via averaging and normalizing
post.sum <- array(0,dim=c(1,N)) for (j in 1:m){
    post.sum <- post.sum +
    dgamma(phi,x1,scale=1/mc.lambda2[j]/(t1/t2))*dbeta(phi/(t2/t1+phi),1.5,1)
```

```
}
appr.post <- post.sum/m     ### average
# Below module computes normalizing constant by numerical integration
c <- 0 for (i in 1:(N-1)){
    c <- c+.5*(appr.post[i]+appr.post[i+1])*(ub.phi-lb.phi)/N
} appr.post <- appr.post/c   ### normalize

#(2) Profile likelihood approximation [limitation: x2 not equal 0]
pl.q.post <- dgamma(phi,x1,scale=(t2/t1)/x2)*dbeta(phi/(t2/t1+phi),1.5,1)

## Approximate posterior via averaging and normalizing
pl.q.post <- pl.q.post/sum(pl.q.post)   ### average
## Below module computes normalizing constant
c.pl <- 0 for (i in 1:(N-1)){
    c.pl <- c.pl+.5*(pl.q.post[i]+pl.q.post[i+1])*(ub.phi-lb.phi)/N
} pl.post = pl.q.post/c.pl   ### normalizing

## Plots
plot(phi,pl.post,type="l",col="green",xlim=c(0,ub.phi),main=paste("Posterior
comparison","(","y1=",x1,"y2=",x2,")"),xlab=expression(phi),ylab="",lwd=2)
lines(phi,appr.post,type="l",col="red",lty=2,lwd=2)
legend("topright",c("Simulated","Profile-Likelihood"),
lty=c(2,1),lwd=2,col=c("red","green"))}
```

.3 R Code for Studying Coverage Property of HPD Intervals

```
# Coverage property study
n1 <- 21 n2 <- 21
cover <- matrix(0,nrow=n1,ncol=n2)  ### create coverage matrix
for (i in 1:n1){
  mu1 <- 7+3*(i-1)
  for (j in 1:n2){
  mu2 <- 7+3*(j-1)
  ratio <- mu1/mu2

  I <- array(0,dim=1000)    ### true-value vector for each random sampled x1,x2
  for (k in 1:1000){
      x1.r <- rpois(1,mu1)
      x2.r <- rpois(1,mu2)
      hpd.r <- hpd.mc(x1.r,x2.r)
      I[k] <- (ratio > hpd.r[1]) & (ratio < hpd.r[2])
  } cover[i,j] <- sum(I)/1000 } }

# Coverage summary mean(cover)  ### mean coverage
sqrt(var(as.vector(cover)))     ### RMSE
min(cover)  ### minimum coverage
length(cover[cover<.95])/length(as.vector(cover))   ### proportion below 1-alpha
```

VITA

Changbin Guo

Department of Mathematics

East Tennessee State University

Johnson City, TN 37614

EDUCATION

Harbin Institute of Technology, Harbin, China

Electrical Engineering, B.E., July 2004

East Tennessee State University, Johnson City, TN

Mathematics, M.S., May 2006

EXPERIENCE

East Tennessee State University, Johnson City, TN 08/2004 - 05/2006

Graduate Assistant in Department of Mathematics