University of Cincinnati

Date: 12/15/2014

I, Yixuan Guo , hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Mathematical Sciences.

It is entitled:

Bayesian model selection for Poisson and related models

Student's name: **Yixuan Guo**

This work and its defense approved by:

Committee chair: Siva Sivaganesan, Ph.D.

Committee member: Emily Kang, Ph.D.

Committee member: Seongho Song, Ph.D.

16227

Bayesian model selection for Poisson and related models

A dissertation submitted to the

Graduate School of the University of Cincinnati

in partial fulfillment of the requirements

for the degree of

Ph.D. in Mathematical Sciences

in the Department of Mathematical Sciences

of the McMicken College of Arts and Sciences

December, 2014

by

Yixuan Guo

M.S. University of Cincinnati, Cincinnati, OH, 2009

Committee Chair: Siva Sivaganesan, Ph.D.

Abstract

Poisson model is one of the fundamental discrete models used to model count data in various fields. It assumes that the mean and variance of data are approximately equal. In practice, the observed data often violates this assumption because the variance can be larger than the mean commonly referred to as over-dispersion. Several models have been developed based on the Poisson model to address the issue of dispersion occurred in data; generalized Poisson model, zero-inflated Poisson model, and zero-inflated generalized Poisson model are examples of such distributions. In this thesis, I will focus on developing a method within the Bayesian framework to compare these four models. This method is generic and can be readily generalized to the comparison of any number of models. We will use non-informative prior and importance sampling to calculate the posterior probability for each model. We also use the same method to compare regression models, namely Poisson, generalized Poisson and negative binomial regression models.

Acknowledgements

I would like to express my gratitude to my dissertation advisor, Dr. Siva Sivaganesan, for the understanding, patience, wisdom and encouragement he provided throughout my dissertation.

I would also like to thank my committee members, Dr. Seongho Song and Dr. Emily Kang, for being on my committee and giving valuable comments.

I am deeply indebted to Dr. James Deddens who encouraged me to pursue research in statistics and provided persistent help and advice during my study.

A very special thank goes out to Dr. Zhoushen Huang for his comments on my research, and for his patience and encouragement.

Most of all, I would like to thank my parents for their support and love which help me survive the stress and keep me going, and this dissertation would not have been possible without them.

Contents

List of Figures

Figure

List of Tables

Table

Chapter 1

Introduction

1.1 Motivation

When modeling count data in practice, one often finds the frequency of zero is higher than the one predicted by of the model being fitted. This problem is very common, and it occurs in various fields, such as engineering, manufacturing, economics, public health, epidemiology, psychology, sociology, political science, agriculture, road safety, species abundance, etc.

Example (Fetal movement data): This data set was collected in a study of breathing and body movements in fetal lambs designed to examine the possible changes in the amount of pattern of fetal activity during the last two thirds of the gestation period. The numbers of movements by a fetal lamb observed through ultrasound were recorded.

Count N	0	1	2	3	4	7
Frequency	182	41	12	2	2	1

Poisson distribution is a standard distribution which is used for fitting count data. A unique property of the Poisson distribution is the equality of its mean and variance. That implies the spread of the data is strictly limited for Poisson model. When this assumption fails and the data displays overdispersion, i.e. the ratio of variance to mean is larger than one, and if this phenomenon is not taken into account, Poisson distribution underestimates the variability and makes incorrect explanation of data. For this case, several distributions have been developed based on Poisson distribution to solve this overdispersion problem. These are generalized Poisson, zero-inflated Poisson, and zero-inflated generalized Poisson distributions. How to choose the appropriate distribution among these distributions is a important question widely concerned in the past.

Suppose we are analyzing one count data and have belief that the data arise from the Poisson Family. The problem of which model is appropriate for the data turns out to be the model selection question in statistics. For a given data, we want to know which model gives a better fit. In the Bayesian perspective, posterior probability of a model is the Bayesian method used for model comparison. Bayes factor is also equivalent measure for the model comparison on the basis of the evidence.

The frequentist estimation of GP distribution based on maximum likelihood estiamtion (MLE) often encounters the convergence problem on the expectation-maximization (EM) and the Newton-Raphson (NR) algorithms. Bayesian Inference does have theoretic guarantee that MCMC algorithm will

surely converge after the process run long enough. Hence Bayesian Inference allows more complicated models and is computationally efficient.

1.2 Literature review

Poisson distribution is the fundamental and widely used discrete probability distribution. It is commonly used for the count of events in a fixed interval of time or space. However, in some applications, the observed data can not be explained by the Poisson distribution due to the assumption of Poisson distribution, that is, the equality of the mean and variance of the data.

This kind of violation of the assumption leads to inadequate model fitting by the Poisson distribution which has only one parameter. In practice, the count data is often observed overdispersion with greater variability in the data. The underdispersion that the variance of a Poisson distribution is smaller than its mean is rarely present in practice, thus we mainly focus on the overdispersion situations. Zero-inflation is one special case of overdispersion such that the number of zeros is more than expected.

The overdispersion was originally analyzed by a finite mixture of Poisson models, but the number of parameters normally becomes quite large and the analysis is complex. For that, several distributions are developed based on Poisson distribution to allow the flexibility of variance, these are generalized Poisson distribution, zero-inflated Poisson distribution, and zero-inflated generalized Poisson distribution.

Bayarri et al. (2008) uses the Bayes Factor based on objective priors to

compare the Poisson versus zero-inflated Poisson model. *Bhattacharya et al.* (2008) presented a Bayesian method to test the zero-inflation of zero-inflated Poisson model by computing a certain posterior probability of the alternative hypothesis.

Gupta et al. (2005) tests the zero-inflation of zero-inflated generalized Poisson model using score test, the author developed the testing procedure both with covariates and without covariates. In the paper of *Xie et al.* (2009), score tests are presented for testing the zero-inflated Poisson mixed regression model against the zero-inflated generalized Poisson mixed regression model, and for testing the significance of regression coefficients in zero-inflation and generalized Poisson portion.

The overdispersion of generalized Poisson regression model has been tested using Score test by *Yang et al.* (2007), and the authors also compared the generalized Poisson regression model versus Negative Binomial regression model by Vuong test. *Joe and Zhu* (2005) compared probability mass functions and skewnesses of the generalized Poisson and negative binomial distributions with the first two moments fixed.

1.3 Poisson distribution

Poisson distribution is an essential probability distribution in statistics. It is commonly used for fitting count data which means a number of events occurring in a fixed interval of time or space given a constant occurrance

probability. The probability mass function for count $Y = y$ is

$$f(y|\lambda) = \Pr(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad y = 0, 1, 2, ...$$

where $0 < \lambda$

It is a discrete probability distribution of the number of observed events of interest. The event is assumed independent of time since the last event.

$$\mathrm{E}(Y) = \mathrm{Var}(Y) = \lambda$$

The parameter λ represents the expected value and variance of random variable Y. Poisson distribution always hold the mean and variance equal. This assumption makes the application of Poisson distribution restricted to specific data, but real data normally does not satisfy this condition. Other distributions are developed to include the control of variance by adding more parameters.

1.4 Generalized Poisson distribution

The generalized Poisson (GP) distribution is of interest for modeling count data because it control the over-dispersion and under-dispersion by one additional parameter. Generalized Poisson distribution is introduced by *Consul and Jain* (1973). The properties of the distribution has been well studied in papers *Consul* (1989); *Lerner et al.* (1997); *Tuenter* (2000). The generalized Poisson regression model is further discussed by *Consul and Famoye* (1992).

Similar to the Negative Binomial distribution, *Joe and Zhu* (2005) has proved that the GP distribution is a mixture of Poisson distributions.

$$f(y|\alpha,\lambda) = \frac{(1+\alpha y)^{y-1}(\lambda e^{-\alpha\lambda})^y}{y!e^{\lambda}} \qquad y = 0,1,2,...$$

where $0 \leq \alpha < 1/\lambda$, $0 < \lambda$. α is the parameter indicating the degree of overdispersion. The larger the value α is, the thicker the tail will be.

The mean and variance are

$$\mathrm{E}(Y) = \frac{\lambda}{1-\alpha\lambda}$$
$$\mathrm{Var}(Y) = \frac{\lambda}{(1-\alpha\lambda)^3}$$

Joe and Zhu (2005) studied the difference between GP and NB distribution, and for a simple initial analysis, they did not use covariate information at that time. In the paper, the author displayed the difference between GP and NB distribution according to the dispersion index D (the ratio of variance to mean) and the mean of distribution. They say that the difference is not significant when the the mean of distribution is smaller than 5 even though the dispersion index D is large. Otherwise, the difference between two distributions becomes obvious when dispersion index D goes away from 1.

1.5 Zero-inflated Poisson distribution

A special type of count data is the zero-inflated data in which there is a relatively large proportion of zero counts. Zero-inflated Poisson Model (ZIP)

is especially useful for the analysis of count data with a large amount of zeros (inflated). Zero-inflated Poisson regression model is derived by *Lambert* (1992). To model such data, a binary mixture is employed to the zero-inflated modification of Poisson model.

$$f(y|\phi,\lambda) = \phi \times \mathrm{I}_{[y=0]} + (1-\phi) \times \frac{\lambda^y e^{-\lambda}}{y!} \qquad y = 0, 1, 2, ...$$

where $(1-e^{\lambda})^{-1} < \phi < 1$, $0 < \lambda$.

The mean and variance are

$$\mathrm{E}(Y) = (1-\phi)\lambda$$
$$\mathrm{Var}(Y) = (1-\phi)(1+\lambda\phi)$$

It allows the excess zeros present in the data. It does not provide good estimates of the nonzero counts as generalized Poisson distribution, when the variability of data is large.

1.6 Zero-inflated generalized Poisson distribution

Zero-inflated Generalized Poisson Model (ZIGP) improve the data fitting and obtain more accurate results for both extreme zeros and overdispersion. Zero-inflated generalized Poisson distribution combines the features of generalized Poisson distribution and zero-inflated Poisson distribution. It can involve the over-dispersion and the zero-inflation effects both.

zero-inflated generalized Poisson regression model

$$f(y|\phi,\alpha,\lambda) = \phi \times \mathrm{I}_{[y=0]} + (1-\phi) \times \frac{(1+\alpha y)^{y-1}(\lambda e^{-\alpha\lambda})^y}{y!e^{\lambda}} \qquad y = 0,1,2,...$$

where $(1-e^{\lambda})^{-1} < \phi \leq 1$, $0 \leq \alpha < 1/\lambda$, $0 < \lambda$.

The mean and variance are

$$\mathrm{E}(Y) = (1-\phi)\frac{\lambda}{1-\alpha\lambda}$$
$$\mathrm{Var}(Y) = (1-\phi)\frac{\lambda}{1-\alpha\lambda}\left[\left(1+\frac{\alpha\lambda}{1-\alpha\lambda}\right)^2 + \phi\frac{\lambda}{1-\alpha\lambda}\right]$$

The maximum likelihood estimation is studied by *Gupta et al.* (1996), they applied iterative solution to the estimators. The Posterior distribution of ZIGP is studied with non-informative priors by *Angers and Biswas* (2003), within which the integration is evaluated through importance sampling method.

For the ZIGP to be a valid probability distribution, all that required is that ϕ satisfies the inequality given in the definition. Negative values of ϕ are therefore allowed, in addition to the (conventional) values between 0 and 1. Positive values indicate zero-inflation and negative values indicate zero deflation. In this dissertation we assume phi is in the range 0 to 1.

1.7 Negative binomial distribution

The first application of this characterization of the negative binomial distribution was presented by Greenwood and Yule (1920) to model accident statistics. It has subsequently been used to model phenomena as diverse as

the purchasing of consumer packaged goods(Ehrenberg 1959), salesperson productivity (Carroll, Lee, and Rao 1986), and library circulation (Burrell 1990). Negative binomial distribution is similar to generalized Poisson distribution that one additional parameter control the variance and make it flexible to be different from the mean value, and it is also nested with Poisson distribution.

The probability mass function of the negative binomial (NB) distribution is

$$f(y|\tau,\mu) = \frac{\Gamma(y+1/\tau)}{\Gamma(y+1)\Gamma(1/\tau)}\frac{(\tau\mu)^y}{(1+\tau\mu)^{y+1/\tau}} \qquad y = 0,1,2,...$$

$\tau > 0$ controls the degree of overdispersion. $\mu > 0$ is the parameter equal to the mean. This NB distribution reduces to Poisson distribution as τ goes to zero.

The mean and variance are

$$\mathrm{E}(Y) = \mu$$
$$\mathrm{Var}(Y) = (1+\tau\mu)\mu$$

The covariates are included through the log link function that $\mu = exp(x\beta)$. $x = (x_1, x_2, ..., x_p)$ is a covariate vector and β is a vector of unknown regression parameters to be estimated. Since the parameter μ is the mean of both Poisson and NB regression models, the meaning of μ are same in both models. Thus, so do the covariate links for both models.

Therefore, the comparison of Poisson regression model against NB regression model becomes

$$H_0 : \tau = 0 \quad vs. \quad H_1 : \tau > 0$$

To test the hypothesis, Bayes Factor is selected as the bayesian comparison method. The priors for parameters are required for all compared models. The prior for the regression parameters β is same as Poisson regression model.

Joe and Zhu (2005) studied the fitting of data by the generalized Poisson and negative binomial distribution and they show both of them are equally well on explanation of count data in many cases, and their fits are quite similar to each other. Generalized Poisson distribution includes the heavy tail property, whereas negative binomial distribution is able to handle excess zeroes in the data. When the mean of data and the ratio of variance to mean (dispersion ratio) are both small, they are quite close to each other. Also, author says that the zero-inflated GP distribution could fit better than the zero-inflated NB distribution when there is a large zero fraction and a heavy tail. Because the zero-inflated GP distribution contain the additional parameters control both excess zeros and overdispersion of data, while zero-inflated NB distribution does not include the overdispersion feature in the distribution.

Chapter 2

Bayesian Inference and Bayes Factor

2.1 Bayesian inference

Bayesian Inference is the method of drawing conclusions from data on the foundation of Bayes' rule which is used to measure the probability of evidence for a hypothesis. Bayes' rule expresses the posterior probability through the likelihood function and prior probability.

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

H is the hypothesis which is normally compared with others through the observed data, E is the evidence from data.
$P(H|E)$ is the posterior probability, the probability of H given evidence E.
$P(H)$ is the prior probability of hypothesis before data is observed.
$P(E|H)$ is the likelihood which is the probability of observing data given the hypothesis H.

$P(E)$ is called marginal likelihood which is the same for all hypotheses being considered.

In frequentist inference, model parameters are fixed constants. The parameters are estimated from the distribution of estimators. In Bayesian inference, model parameters are regarded as random variables, and are estimated by its posterior distribution updated from the data, that is, the distribution of the parameters given data.

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

$P(\theta|y)$ is the posterior distribution of the parameters θ given the data y
$P(y|\theta)$ is the likelihood conditional on parameters θ
$P(\theta)$ is the prior distribution of parameters before observing data
$P(y)$ is the marginal likelihood which is the distribution of the data marginalized over the parameters.

The posterior distribution contains all information of the parameters, and it is straightforward to deliver credible intervals for parameters, while the frequentist confidence intervals is difficult to understand. The posterior distribution yields more realistic information including the uncertainty of parameters which is not included using frequentist inference. The variance provided in frequentist estimation is only the sampling uncertainty. Since frequentist approach is established on the asymptotic normality assumption, the variance of the MLE is not accurate when sample size is small. The mean and variance of parameters obtained from posterior distributions are trustworthy and not

affected by the size of data. For large data size, they both give similar results. One big advantage of posterior distribution is that it can handle a large number of random variables in the model. Due to the numerical approximation development, the computation of complex integration can be implemented iteratively by high-speed computer, i.e. Markov chain Monte Carlo method. Thus, bayesian inference In contrast, frequentist inference

2.2 Prior probability

The prior probability distribution, normally called prior, expresses the information of an uncertain quantity θ before the data is taken into account. The unknown quantity could be a parameter or latent variable. The prior probability is totally subjective due to the selection of priors made by experts before any knowledge from data. The prior distribution should be assigned on a range covering all possilbe values for the unknown quantity θ. It is convenient of posterior distribution calculation to choose conjugate prior when it is available, however the conjugate prior is not always easy to find. There are two classes of priors, one is informative priors (or "subjective priors") and the other is called non-informative priors (also known as "non-informative" or "objective" priors). The later term "non-informative" is sort of misleading, as every prior includes some information actually, and it should be reasonable to denote as "weakly informative prior" instead. The informative priors normally contain specific, definite information or belief on the quantity θ before observing any data. The information might come from knowledge gained from

previous experiments or experts. In contrast, the use of non-informative priors assigns equal probabilities to all possibilities. It has minimal impact on the posterior distribution. The structure of priors is set up based on two key issues:

- The information of the quantity θ wants to be included in the prior distribution
- The properties of the posterior distribution

The information of the prior comes from either Expert opinions or the previous experiments. The most widely used method for finding the non-informative prior is the Jeffreys's rule. Jeffreys priors are sometimes improper priors. The improper prior is named for that the integral over the parameter space diverges. For example, $\pi(\theta) = 1$ over 0 to ∞ is improper, since the area under the prior density is infinite. In most instances, there are no big problem in use of improper priors in bayesian analyses. Sometimes, the use of improper priors can lead to improper posteriors. In model selection or hypothesis testing, improper priors bring in the uncertainty to the results.

Flat priors are the densities with wide spread, such as a normal distribution with large variance for infinite range, i.e. $\mathrm{N}(\mu, 100^2)$, or a uniform distribution over a bounded range, i.e. $\mathrm{Uniform}(0, 1)$. Flat piror is one of the most common priors and easy for approach. It is non-informative prior, since no specific preference of the quantity is indicated from the flat prior distribution. Often in bayesian analysis, three common choices of non-informative priors are the uniform prior, Jeffreys Prior.

2.2.1 Jeffreys Prior

The Jeffreys Prior is named after Harold Jeffreys. It is a non-informative prior which is proportional to the square root of the determinant of the Fisher information. The advantage of Jeffreys priors is that it is invariant under reparameterization of parameter vector $\vec{\theta}$.

One-Parameter case First, let's consider the simple situation with single parameter θ. Jeffreys Prior is

$$\pi(\theta) \propto \sqrt{\det I(\theta)}$$

where $I(\theta)$ is the Fisher information defined as

$$I(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial \theta^2} lnL(\theta|y) \middle| \theta \right]$$

$L(\theta|y) = \mathrm{P}(y|\theta) = \prod_i f(y_i|\theta)$ is likelihood probability of parameter θ.

For a alternative parameterization ϕ, the Jeffreys prior can be derived from the change of variables theorem

$$\pi(\phi) = \pi(\theta) \left| \frac{d\theta}{d\phi} \right| \propto \sqrt{\det I(\phi)}$$

that is, Jeffreys priors are independent to reparameterization.

Multiple-Parameter case When there are more than one parameter in the model, i.e. $\vec{\theta} = (\theta_1, \theta_2, ..., \theta_p)^T$, the Fisher information will be a $p \times p$ matrix

$$I(\vec{\theta}) = [I(\theta)_{i,j}]_{p \times p}$$

the elements of the Fisher information matrix are

$$I(\theta)_{i,j} = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} lnL(\theta|y) \middle| \theta \right]$$

the Jeffreys prior is

$$\pi(\vec{\theta}) \propto \sqrt{\det I(\vec{\theta})}$$

Also, the reparameterization is easy to conduct on Jeffreys priors. For the alternative parameterization $\vec{\phi}$.

$$\pi(\vec{\phi}) = \pi(\vec{\theta}) \left| \det \frac{d\theta_i}{d\phi_j} \right| \propto \sqrt{\det I(\vec{\phi})}$$

Jeffreys prior is sometimes a improper prior which cause the arbitrary constant of the normalizing constant, and this arise the hypothesis testing problem by the arbitrary constant. One

2.3 Bayes factors

Bayes Factor is a broadly used bayesian alternative to frequentist hypothesis testing. In frequentist statistics, the hypothesis testing is focus on the probability of test statistic under null hypothesis. In bayesian framework,

the probability of testing the hypothesis is to compute the probability of the model given the true data. Bayes Factors relies on the posterior probability odds which gives the alternative of hypothesis testing. Bayes Factors follows from Bayes Theorem.

$$P(M_i|y) = \frac{P(y|M_i)P(M_i)}{P(y|M_1)P(M_1) + P(y|M_2)P(M_2)} \qquad for\ i = 1, 2$$

where y is the given data, M_i is the ith model of the proposed models to be compared. $P(M_i)$ is the prior belief in model M_i. so that,

$$\frac{P(M_1|y)}{P(M_2|y)} = \frac{P(y|M_1)P(M_1)}{P(y|M_2)P(M_2)}$$

thus, it means,

$$posterior\ odds = Bayes\ factor \times prior\ odds$$

Bayes Factor:

$$BF_{12} = \frac{P(y|M_1)}{P(y|M_2)}$$

Bayes Factor describes the transformation of posterior odds from prior odds through consideration of the data. Without any prior knowledge, it normally assumes the prior preference of model is same that $P(M_1) = P(M_2) = 0.5$. Bayes factor is simply equal to the posterior odds in favor of model M_1.

The Marginal Likelihood is the Bayesian evidence and is defined as

$$m_i = P(y|M_i) = \int P(y|M_i, \theta_i)P(\theta_i|M_i)d\theta_i$$

θ_i is the vector of paramters in model M_i. $P(y|M_i, \theta_i)$ is the likelihood of model M_i measuring the probability of observing data y under the assumption that M_i is the true model. $P(\theta_i|M_i)$, or say $\pi(\theta_i|M_i)$, is the prior probability of the parameter of model M_i. In contrast, $P(\theta_i|y, M_i)$ is the posterior distribution of parameter of model M_i given data y. Briefly, Bayes Factor is expressed as

$$BF_{12} = \frac{m_1}{m_2} = \frac{\int P(y|M_1, \theta_1)P(\theta_1|M_1)d\theta_1}{\int P(y|M_2, \theta_2)P(\theta_2|M_2)d\theta_2}$$

Bayes Factor relies on the marginal likelihood which is the probability of the data given the model. Prior of the interest parameter θ carries important knowledge to Bayes Factor and should be spiked on the null hypothesis θ_0. Such very strong prior belief on parameter then will be modified by the observed data. In contrast to Bayesian Inference, the prior plays an important role in model selection such that the decision will be heavily affected by the selected prior even for a large sample. For such case, the noninformative priors are good choices for avoiding the difficulties of the dependence of Bayes Factor on the priors.

Bayes Factor have 'Ockham's razor' done automatically, that is, the penalty of including too much model structure is taken into account, and it will automatically prefer the simpler model. It is notorious that the marginal like-

lihoods are computationally difficult. However, they are easy to approximate with some numerical approach methods.

Interpretation of Bayes Factor:

Bayes Factor gives a measure of evidence between two models in terms of the support of the data, it should be noticed that none of the models is necessarily required to be true. In particular, Bayes Factor accesses two models based on the data in favor of model M_1. If the Bayes Factor is larger than 1, the evidence of data shows stronger support to model M_1 than M_2. Because the posterior odds is larger than the prior odds, in other words, the posterior probability of model M_1 is larger than its prior probability. If Bayes Factor is smaller than 1, model M_2 is more strongly supported by the data. Here is the scale for interpretation:

Bayes Factor	Strength of Evidence
$-\infty < BF \leq 0.1$	Strong against M_1
$0.1 < BF \leq (1/3)$	Substantial against M_1
$(1/3) < BF < 1$	Barely worth mentioning against M_1
$1 \leq BF < 3$	Barely worth mentioning for M_1
$3 \leq BF < 10$	Substantial for M_1
$10 \leq BF < \infty$	Strong for M_1

The frequentist method requires models to be nested for comparison analysis, and it is very difficult to do non-nested comparison in frequentist framework. Bayes Factor can access nested or non-nested models.

2.4 Comparison of multiple models

Bayes Factor is not only used to compare paired models, but it can do analysis of multiple models or hypotheses. Posterior Probability provides the numerical summaries of the strength of evidence in favor of models. When there are more than two candidate models for the data, it is more convenient

to use the Posterior Probability to measure the evidence of data for each model.

The Posterior Probability $P(M_i|y)$ for each model is given by

$$P(M_i|y) = \frac{P(y|M_i)P(M_i)}{P(y)} \qquad for\ i = 1, ..., n$$

where the number of total candidate models is n.
$P(y)$ is the unconditional probability of observing y, that is

$$P(y) = \sum_{i=1}^{n} P(y|M_i)P(M_i)$$

Also, all models satisfy $\sum_{i=1}^{n} P(M_i) = 1$

The model with the greatest Posterior Probability will be the best preference for the fitted data among the comparing models. When there is no prior knowledge of all models, it is common to initiate $P(M_i) = P(M_j)$ for all i, j. Then the Posterior Probability $P(M_i|y)$ becomes

$$P(M_i|y) = \frac{P(y|M_i)}{\sum_{i=1}^{n} P(y|M_i)} = \frac{m_i}{\sum_{i=1}^{n} m_i}$$

Posterior probability of given model is quite useful in understanding the difference among models for any given data. It simply provides the probability of each candidate model based on the information updated from the given data. The interpretation of posterior probability of model is simple and straightforward without any hidden confusing statistical concept which may cause misunderstanding in statistical analysis. Also, it does not require any

pre-knowledge about the model in order to calculate the test statistic. In general, it provides the basic comparison of all candidate models rather than any specific fitted ones.

2.5 Fractional Bayes factor

Fractional Bayes Factor (FBF) is developed, based on the Partial Bayes Factor, by *O'Hagan* (1995). It aims to solve the problem of the improper prior within the Bayes Factor via training the normalizing constant by part of data or part of likelihood. It is similar to the intrinsic Bayes factor method, but we only focus the fractional Bayes Factor here. It is known that the Bayes Factor is the ratio of the marginal likelihoods of two models. The marginal likelihood is

$$m_i = P(y|M_i) = \int \pi(\theta_i|M_i)P(y|M_i,\theta_i)d\theta$$

When the prior $\pi(\theta_i|M_i)$ is the improper probability, Bayes Factor is unable to be calculated.

$$\pi(\theta) \propto h(\theta)$$

where $h(\theta)$ diverges over the θ space. In this case, when we write the improper prior probability

$$\pi(\theta) = c\, h(\theta)$$

the normalizing constant c actually does not exist, but let us regard it as an unspecified constant here. The improper prior has no problem with analysis

of the posterior distribution of the parameters in Bayesian Inference, since

$$p(\theta|y) = \frac{P(y|\theta)\pi(\theta)}{P(y)} = \frac{P(y|\theta)h(\theta)}{\int P(y|\theta)h(\theta)d\theta}$$

the constant c cancels out from the existence in both numerator and denominator. Nevertheless, Bayes Factor can not handle in the same way for improper priors. Assume the prior probability for model M_1 is an improper prior, then

$$BF_{12} = \frac{P(y|M_1)}{P(y|M_2)} = c\,\frac{\int P(y|M_1,\theta_1)h(\theta_1|M_1)d\theta}{\int P(y|M_2,\theta_2)\pi(\theta_2|M_2)d\theta}$$

the unspecified constant c does not cancel out. Bayes Factor is uncertain due to the unspecified c. If the paramter priors of both models are all improper, then

$$BF_{12} = \frac{P(y|M_1)}{P(y|M_2)} = \frac{c_1}{c_2}\,\frac{\int P(y|M_1,\theta_1)h(\theta_1|M_1)d\theta}{\int P(y|M_2,\theta_2)h(\theta_2|M_2)d\theta}$$

the ratio of unspecified constants $\frac{c_1}{c_2}$ causes the uncertainty. It causes the Bayes Factor to take any value meaninglessly.

Let n be the number of observations in the data. The Fractional Bayes Factor try to reduce this sensitivity of Bayes Factor to the prior with a training sample of size m. $b = m/n$ is the training fraction.

$$BF_{12}^{b} = \frac{\int P(y|M_1,\theta_1)^b\pi(\theta_1|M_1)d\theta}{\int P(y|M_2,\theta_2)^b\pi(\theta_2|M_2)d\theta}$$

and defines the Fractional Bayes Factor as

$$FBF_{12} = \frac{BF_{12}}{BF_{12}^{b}}$$

The main practical question is how to decide the training sample size. The training fraction b should go to 0 as $n \to \infty$ to retain consistency of model choice. A minimal size m_0 satisifies this condition. Although there are debates over defining the minimal size m_0, the purpose of the minimal size is to leave as much data as possible to model comparison. Thus, the debates would not hinder the construction of FBF. For instance, if there was a doubt about whether two or three observations is appropriate as the minimal size, then the size three would be used.

But the small training sample size m will sacrifice the robustness to the misspecification of the prior. For these features of the FBF, the author suggests two alternatives of training sample size, $m = \log n$ and $m = \sqrt{n}$. The first increases very slowly with n, so it is keeping the training size very small while achieving robustness. The sencond increases faster than the first, and attaches more importance on robustness.

Therefore, here are three ways to set b:

(a) $b = \frac{m_0}{n}$, no concern with robustness

(b) $b = \frac{\log n}{n}$, intermediate option

(c) $b = \frac{1}{\sqrt{n}}$, serious concern with robustness

Normally, the minimal size m_0 is set up to be 2 or 3 in default. If there is any concern about the minimal size at $m_0 = 2$, it is recommended to simply adopt $m_0 = 3$. In our following study, the minimal size $m_0 = 3$ is used to calculate training fraction b.

Chapter 3

Markov Chain Monte Carlo and Importance Sampling

3.1 Markov chain Monte Carlo

In statistics, computation of integration is not always possible to obtain the analytical result. The reason is that integrand tends to be complex most of the time, or the integrand could be multivariate which causes the difficulty of simplification. These problems occur often in both frequentist and Bayesian studies, for instance, calculation of maximum likelihoods, Bayes factors, or bayesian inference, etc. Numerical solutions provide a direct and simple approach to these problems. A general solution is making use of computer simulation of Markov chain over the parameters space.

The Markov chain is a random process, a memoryless transition from one state to another, with the Markov property defining the transition probabilities for the next state which depends only on the current stage but not on the previous states. Let X_i denote the random variable at state i, x_i denote the

observed value of X_i at state i, and the state space is the range of possible X values.

$$P(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n)$$

The probability of a new observation $X_{n+1} = x$ given all observations $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is same as the probability of observing the value given only the previous one, that is, the draws of Markov chain are slightly dependent on the previous one. The chain walks around the parameters space and only remember the last place it has been.

The Monte Carlo method is based on the idea of drawing independent identically distributed samples from a desired distribution. The target distribution then can be approximated by the simulated samples. Once we have the Markov chain converged to the stationary distribution, it seems like we should be able to make use of Markov chain on finding the quantities of interest. Still, the Markov chain is not independent, the Monte Carlo Integration solves the dependence problem by strong law of large numbers (SLLN). Consequently, these samples can be used to approach the integrals on the target distribution.

$$\frac{1}{N} \sum_{i=1}^{N} g(x_i) \xrightarrow[N \to \infty]{a.s.} \int_{\chi} g(x) f(x) dx$$

where x_i is the ith sample of total N draws from the target distribution $f(x)$, $g(x)$ is a measurable function of X. This estimate is unbiased and almost surely converges to the integral by SLLN.

Markov chain Monte Carlo (MCMC) is the idea of sampling random samples from a probability distribution that is too complex to simulate directly. MCMC is based on Markov chain to produce the next sample values using the current sample values, and the produced samples all together is the Markov Chain. The constructed Markov Chain's stationary distribution is the target distribution. Therefore, for all the constructions of MCMC, the Markov chains will definitely converge to the stationary distribution no matter where the starting points are. The construction of MCMC is usually not difficult, but we often do not know how many steps the Markov chain needs to reach stationary status.

Burn-in Because the starting points are not the issue for MCMC sampling, it is flexible to assign any values as the starting points. However, the costing time to convergence is related to the starting points. Since the MCMC sampling takes time to reach stationary status, the steps walked before the stationary status do not represent the target distribution correctly. Those draws before the stationary status are called Burn-in, and they are thrown out from the overall draws in order to make the MCMC draws closer to the stationary distribution and less dependent on the starting points. A good chain will have rapid mixing featurereaching the stationary distribution quickly from an arbitrary position.

It is very difficult to check the convergence of MCMC chains, since all draws are slightly dependent and no idea to when the chain converges to the stationary distribution. One often takes heuristics to check the situation of

convergence. For example, initiate the starting points at different values and monitor whether the trace plots of each chain agree to each other.

3.2 The Metropolis-Hastings algorithm

Metropolis-Hastings algorithm is a Markov chain Monte Carlo method to obtain samples from a complex probability distribution which is not possible to draw samples directly. It is especially powerful at sampling from multi-dimensional distribution with high dimension. The generated sequence of samples can normally be used to approximate the target distribution or to approach an integral.

Suppose $P(x)$ is the desired distribution, $q(x^* \to x)$ is the proposal distribution that is used to change between any two states x to x*
the acceptance probability of Markov chain moving toward next state x^* is

$$A(x \to x^*) = min(1, \frac{(P(x^*)q(x^* \to x))}{(P(x)q(x \to x^*))})$$

this is also denoted as $A(x, x^*)$, and it is a measure of the probability of transition from x to x*. The range of it is between 0 and 1.

The Metropolis-Hastings algorithm is working in steps as follows:

1. initiate a arbitrary value as the start state $x^{(0)}$;
2. draw a random value u from uniform distribution $U(0, 1)$;
3. draw a random value x^* from proposal distribution $q(x^*|x^{(i)})$;

4. accept the draw according to acceptance probability $A(x^{(i)}, x^*)$.
   If $u < A(x^{(i)}, x^*)$, accepted $x^{(i+1)} = x^*$. Else, it remains at $x^{(i+1)} = x^{(i)}$ and so the state is same as the previous one;

5. go to 2 until enough states were generated;

The number of states generated must be chosen according to the proposal distribution. The variances of different proposals have a big impact on the results. If the proposal distribution is too narrow, Markov chain may be stuck at one mode of multimodal and can not walk over the entire region of the target distribution. If the proposal distribution is too wide, it suffers high rejection rate, and the Markov chain is highly correlated. A well mixing chain is able to visit all modes of the target distribution and has a high acceptance probability.

3.3 The Gibbs sampler

The Gibbs sampler is a special case of Metropolis-Hastings sampling such that draws are always accepted. It is designed for multivariate distribution simulation, the Gibbs sampler only consider univariate conditional distributions. Suppose the target distribution has n random variables and the full conditional distributions are

$$f(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) \quad \text{for } i = 1, ..., n$$

It is far easier to simulate n random variables sequentially from the full conditionals than generating a n-dimensional vector from the complex joint distribution one time.

The algorithm of the Gibbs sampler

1. initiate arbitrary values as the start state $(x_1^{(0)}, ..., x_n^{(0)})$;

2. for k=1 to N;

   draw random values $x^{(k)}$ from full conditional distributions;

   $x_1^{(k)} \sim f(x_1|x_2^{(k-1)}, ..., x_n^{(k-1)})$;

   $x_2^{(k)} \sim f(x_2|x_1^{(k)}, x_3^{(k-1)}, ..., x_n^{(k-1)})$;

   $\vdots$

   $x_i^{(k)} \sim f(x_i|x_1^{(k)}, ..., x_{i-1}^{(k)}, x_{i+1}^{(k-1)}, ..., x_n^{(k-1)})$;

   $\vdots$

   $x_n^{(k)} \sim f(x_n|x_1^{(k)}, ..., x_{n-1}^{(k)})$;

that is, each random variable is sampling from the updated full conditional distribution given the most recent values of the other random variables. It is still only depends on the previous observations but partially updated with the current state. One iteration of the sampling of all univariate distributions is called a scan of the sampler. The collection of scans is called the Gibbs sequence. After burn-in steps, the Gibbs sequence reaches the stationary distribution which is also our target distribution. Thus, the draws from the Gibbs sampler represent the desired distribution we want to simulate.

3.4 Simulation steps of multivariate distribution

For multivariate distribution simulation, Gibbs sampler method is able to be extended with the use of Methopolis-Hastings algorithm when the full conditional distributions of variables are not easy to sample from. Methopolis-Hastings algorithm can be used to sample from variables which are not possible to draw samples directly. Assume a multivariate distribution $f(x_1, x_2, ..., x_n)$ of n variables in the model, the full conditional distribution of variable x_i is $f(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$. In one iteration of simulation, the draw of variable x_i is sampling from a proposal distribution $q(x_i|x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$ which is similar to the full conditional distribution of the variable, and the draw is determined by acceptance probability of Metropolis-Hastings algorithm.

The Simulation steps of multivariate distribution:

1. initiate arbitrary values as the start state $(x_1^{(0)}, ..., x_n^{(0)})$;

2. for k=1 to N, iterate;

3. draw one random value of variables from proposal distributions;
   $x_1^* \sim q(x_1|x_2^{(k-1)}, ..., x_n^{(k-1)})$;
   $x_2^* \sim q(x_2|x_1^{(k)}, x_3^{(k-1)}, ..., x_n^{(k-1)})$;
   $\vdots$
   $x_i^* \sim q(x_i|x_1^{(k)}, ..., x_{i-1}^{(k)}, x_{i+1}^{(k-1)}, ..., x_n^{(k-1)})$;
   $\vdots$

$x_n^* \sim q(x_n|x_1^{(k)}, ..., x_{n-1}^{(k)})$;

4. draw a random value u from uniform distribution $U(0, 1)$;

5. accept the draw according to acceptance probability $A(x_i^{(k-1)}, x_i^*)$. for each variable x_i, $i = 1, ..., n$.
   If $u < A(x_i^{(k-1)}, x_i^*)$, accepted $x_i^{(k)} = x_i^*$. Else, it remains at $x_i^{(k)} = x_i^{(k-1)}$ and so the state is same as the previous one;

6. repeat from step 3, until obtain enough draws N;

In the simulation, it is assumed that full conditional distributions can not be simulated directly from, therefore proposal distributions for the full conditional distributions are defined to approximate a sequence of draws sampling from the true distribution. The values of other variables in the condition of proposal distribution are updated by the latest accepted values of those variables. The acceptance probability is defined in the Methopolis-Hastings algorithm.

This simulation method makes sampling possible for any multivariate distribution, and the proposal distribution of variable can be any known distribution which simplifies the simulation process. The calculation of acceptance probability may takes some computation time when the target distribution is complex, and it could be a problem especially when the number of iterations are huge. The mixing of simulation can be quick when the proposal distribution is quite close to the desired distribution, but the number of iterations necessarily needs to be large enough to make convergence to the true

distribution.

3.5 Importance sampling

Importance sampling is an efficient method to approach the integration which is too complicated to get the analytical solution. One advantage of importance sampling is the capability of reducing the variance of estimators such that it can greatly reduce the computation time by convergence with fewer MCMC iterations. It is widely utilized in posterior densities, Bayes factors, parameter estimation, bootstrap quantile problems, etc.

Importance sampling relies on importance functions generating samples rather than the distribution of interest. Let X be a random variable, h is a function of X, f is the distribtuion density of interest. Assume we want to know the quantity of expected value of function $h(x)$, but it would be unable to obtain the closed-form expression and the density $f(x)$ could be too complicated to simulate. Importance sampling makes use of simple probability densities in lieu of the complicated $f(x)$ in calculation of the expectation.

$$V = E_f[h(X)] = \int_{\chi} h(x)\frac{f(x)}{g(x)}g(x)dx = E_g[\frac{h(x)f(x)}{g(x)}]$$

that is, the expectation under the density $g(x)$, also named as importance function. $\frac{f(x)}{g(x)}$ is known as importance weight. The choice of the importance function $g(x)$ can be any probability density theoretically, but the purpose of importance sampling is to solve the difficulty of calculation, thus common

densities which are easy to simulate are the usual selections. In order to approximate the integral, a sample $X_1, X_2, ..., X_n$ of size n are generated from the density $g(x)$. By the law of large numbers,

$$\hat{V} = \frac{1}{n}\sum_{j=1}^{n} h(X_j)\frac{f(X_j)}{g(X_j)} \rightarrow E_f[h(X)]$$

This is called the importance sampling estimator. It is easy to calculate the mean of the function $h(x)f(x)/g(x)$. The average will almost surely converge to the desired expectation as the sample size n increases to infinity. Thus, importance sampling redistributes the law of X by importance function g so that its samples' frequencies are sorted directly according to their weights $h(X)f(X)/g(X)$.

The selection of importance functions g plays a important role in the estimation. A poor choice of the importance function may yield very poor outcomes. The performance of importance sampling is tested through the variance of the importance estimators

$$var(\hat{V}) = \frac{1}{n}\sigma^2$$

where

$$\sigma^2 = var_g[h(x)\frac{f(x)}{g(x)}] = \int \frac{h^2(x)f^2(x)}{g(x)}dx - V^2$$

In order to estimate the expectation, we need to estimate the variance of the

importance estimator for the confidence interval of V.

$$\sigma^2 = \int \frac{h^2(x)f^2(x)}{g(x)}dx - V^2 = \int \frac{(h(x)f(x) - Vg(x))^2}{g(x)}dx = E_g[\frac{(h(x)f(x) - Vg(x))^2}{g(x)^2}]$$

so, the estimated

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\frac{h(x)f(x)}{g(x)} - \hat{V})^2$$

Then 99% confidence interval for V is $\hat{V} \pm 2.58\hat{\sigma}/\sqrt{n}$

Importance sampling is a variance reduction technique that can be used in the Monte Carlo method. The idea behind importance sampling is that certain values of the input random variables in a simulation have more impact on the parameter being estimated than others. If these "important" values are emphasized by sampling more frequently, then the estimator variance can be reduced. Hence, the basic methodology in importance sampling is to choose a distribution which "encourages" the important values. This use of "biased" distributions will result in a biased estimator if it is applied directly in the simulation. However, the simulation outputs are weighted to correct for the use of the biased distribution, and this ensures that the new importance sampling estimator is unbiased. The weight is given by the likelihood ratio, that is, the RadonNikodym derivative of the true underlying distribution with respect to the biased simulation distribution.

3.6 Method review

In the following studies, Markov chain Monte Carlo method will be applied several times on different problems occurred due to the complexity of the models. In the distribution selection section, Gibbs sampler is embedded of Metropolis-Hastings algorithm for the purpose of simulation of all competing distributions which are not standard distributions. Importance sampling is implemented also in this section in order to solve the complex integrals of the marginal distributions of the Bayes factors. In the Bayesian inference of generalized Poisson distribution section, simulation of generalized Poisson distribution is carried out through Gibbs plus Metropolis-Hastings algorithm. The posterior distributions of parameters of generalized Poisson distribution are also calculated by numerical approach of importance sampling.

Chapter 4

Bayesian model selection of Poisson and related models without covariates

4.1 Poisson and related distributions

Poisson related distributions are developed based on the fundamental Poisson distribution. The distribution with more parameters extends the flexibility of modeling count data. However, more parameters means more complex the distribution is. That costs unnecessarily effort of fitting the distribution and losing the accuracy when the data actually is only from the simple distribution with fewer parameters. To avoid the problem of over fitting and to select the appropriate distribution for the count data, here, we propose the Bayesian method of distribution selection in Poisson related distributions.

The probability mass function of zero-inflated generalized Poisson distribution is:

$$f(y|\alpha,\phi,\lambda) = \Pr(N=y) = \phi \times \mathrm{I}_{[y=0]} + (1-\phi) \times \frac{(1+\alpha y)^{y-1}}{y!} \frac{(\lambda e^{-\alpha\lambda})^{y}}{e^{\lambda}}$$

where y = 0,1,2,..., $(1-e^{\lambda})^{-1} < \phi < 1, 0 < \lambda, 0 \leq \alpha < 1/\lambda$

In the zero-inflated generalized Poisson (ZIGP) distribution, the term ϕ plays the role of a zero-inflation factor, α controls the dispersion factor. Obviously, when $\alpha = 0$, the ZIGP distribution reduces to zero-inflated Poisson (ZIP) distribution with parameters ϕ and λ. When $\phi = 0$, the ZIGP distribution reduces to generalized Poisson (GP) distribution with parameters α and λ. When $\alpha = 0$ and $\phi = 0$, the ZIGP distribution reduces to Poisson (Poi) distribution with the parameter λ.

Reduced distributions:

When $\alpha = 0$, ZIGP reduces to ZIP distribution:

$$f(y|\phi,\lambda) = \phi\, \mathrm{I}_{[y=0]} + (1-\phi)\frac{\lambda^y}{y!e^{\lambda}}$$

When $\phi = 0$, ZIGP reduces to GP distribution:

$$f(y|\alpha,\lambda) = \frac{(1+\alpha y)^{y-1}(\lambda e^{-\alpha\lambda})^y}{y!e^{\lambda}}$$

When $\alpha = 0$ and $\phi = 0$, ZIGP reduces to Poisson distribution:

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

Among all models, there is a common parameter λ. Nevertheless, it needs to be cautious about the meaning of the parameter of the same symbol that is not necessarily same throughout the nested distributions. For instance, the

parameter λ is present in all distributions as a common parameter, but the meaning of it is different in each distribution. Within Poisson distribution, it is same as the value of mean and variance. But, in the other distributions, it does not indicate the mean of distribution any longer. Actually, the meaning of the parameter is changing among these distributions.

4.2 Model selection with conditionally uniform priors

Likelihoods of distributions over (ϕ, α, λ):

$$
\begin{aligned}
L^{\mathrm{ZIGP}}(\phi,\alpha,\lambda) &= \prod_{i=1}^{n} f(y_i|\phi,\alpha,\lambda) \\
&= [\phi + (1-\phi)e^{-\lambda}]^{n_0}[(1-\phi)e^{-\lambda}]^{n-n_0}\lambda^s e^{-s\alpha\lambda}\prod_{i=1}^{n}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \\
L^{\mathrm{ZIP}}(\phi,\lambda) &= \prod_{i=1}^{n} f(y_i|\phi,\lambda) = [\phi + (1-\phi)e^{-\lambda}]^{n_0}[(1-\phi)e^{-\lambda}]^{n-n_0}\frac{\lambda^s}{\prod_{i=1}^{n} y_i!} \\
L^{\mathrm{GP}}(\alpha,\lambda) &= \prod_{i=1}^{n} f(y_i|\alpha,\lambda) = (\lambda e^{-\alpha\lambda})^s e^{-\lambda n}\prod_{i=1}^{n}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \\
L^{\mathrm{Poi}}(\lambda) &= \prod_{i=1}^{n} f(y_i|\lambda) = \frac{\lambda^s e^{-\lambda n}}{\prod_{i=1}^{n} y_i!}
\end{aligned}
$$

The Bayesian analysis requires the prior information on parameters in distributions, and the results somewhat depend on the prior distribution assumption. If there is previous knowledge about the parameters, it is helpful to use the informative priors. Non-informative priors have been widely used especially when there is no prior knowledge on parameters. Here, because of lack of sufficient prior information, conditionally uniform priors are considered for

candidate models. For Poisson distribution, the Jeffreys prior for λ is used. For parameters of other distributions, the uniform priors are considered. These selected priors are listed in the following table.

Table 4.1: Conditionally uniform priors for model selection

Poisson	GP
$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}$ $0 < \lambda$	$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}$ $\alpha\|\lambda \sim Uniform(0, \lambda^{-1})$ $\pi(\alpha, \lambda) \propto \sqrt{\lambda}$ $0 \leq \alpha < 1/\lambda, 0 < \lambda$
ZIP	ZIGP
$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}$ $\eta\|\lambda \sim Uniform(0, 1)$ $\pi(\eta, \lambda) \propto \frac{1}{\sqrt{\lambda}}$ $0 < \eta < 1, 0 < \lambda$	$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}$ $\eta\|\lambda \sim Uniform(0, 1)$ $\alpha\|\lambda \sim Uniform(0, \lambda^{-1})$ $\pi(\eta, \alpha, \lambda) \propto \sqrt{\lambda}$ $0 < \eta < 1, 0 \leq \alpha < 1/\lambda, 0 < \lambda$

In the Conditionally uniform priors, η is a reparameterization of ϕ by $\eta = (1-\phi)(1-e^{-\lambda})$. Jeffreys prior for λ is not a proper prior, but other parameters are proper priors given λ.

The priors of zero-inflated generalized Poisson distribution have been used in *Angers and Biswas* (2003). The prior of λ that $\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}}$ is Jeffreys

prior for Poisson distribution which is developed from the square root of the determinant of the Fisher information matrix.

First, we make some notation for the simplification of the next calculations. Let n denote the number of observations in the data, n_0 be the number of zeros of the response variable Y, and s be the sum of the observed values of count Y that is $s = \sum_{i=1}^{n} y_i$.

Marginal likelihood is defined as

$$m_i = P(y|M_i) = \int P(y|M_i, \vec{\theta}) P(\vec{\theta}) d\vec{\theta}$$

where $\vec{\theta}$ is the vector of parameters in model M_i.

Due to the complexity of integration in computation of marginal likelihood, a transformation of parameter α is made by letting $\theta = \alpha\lambda$. Then the range of the parameter θ is $0 < \theta < 1$. Then the marginal likelihoods of Poisson family distributions are as follows.

$$m^{\text{ZIGP}} = \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!} \int_0^{\infty} \int_0^{1} \lambda^{n-0.5} \frac{e^{-s\theta}}{(e^{\lambda}-1)^{n-n_0}} \prod_{i=1}^{n} (\lambda + \theta y_i)^{y_i - 1} d\theta d\lambda$$

$$m^{\text{ZIP}} = \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!} \int_0^{\infty} \frac{\lambda^{s-0.5}}{(e^{\lambda}-1)^{n-n_0}} d\lambda$$

$$m^{\text{GP}} = \frac{1}{\prod_{i=1}^{n} y_i!} \int_0^{\infty} \int_0^{1} \lambda^{n-0.5} e^{-s\theta - n\lambda} \prod_{i=1}^{n} (\lambda + \theta y_i)^{y_i - 1} d\theta d\lambda$$

$$m^{\text{Poi}} = \frac{n^{-(s+0.5)}\Gamma(s+0.5)}{\prod_{i=1}^{n} y_i!}$$

These integrals are too complex to simplify to a analytical solution except Poisson distribution. However, the complex integral is approachable via the application of numerical method which is called importance sampling method. For the implementation of importance sampling, the quality of the process relies on the choice of importance sampling density. To cover the whole domain of parameters and make it as close to the integrand as possible, the following distributions are selected as the importance sampling densities for the calculation of marginal likelihoods.

The importance sampling densities:

$$\begin{aligned}
&\theta \sim U(0,1) \\
&h(\lambda) = \frac{1}{c}, \quad where\ \lambda \in (0,\infty) \\
&h^{\mathrm{ZIGP}}(\theta,\lambda) = h(\theta)h(\lambda) = \frac{1}{c} \\
&h^{\mathrm{ZIP}}(\lambda) = h(\lambda) = \frac{1}{c}
\end{aligned}$$

These densities have the same range of parameters as the marginal integrals and they are standard distributions which are straightforward to simulate. Because the integrands in the marginal integrals are changing over different data, it is difficult to detect the shape of the integrand so the flat density is suitable to approach the sampling. The importance sampling estimate of one integral $\int f(x)dx$ given the importance density $h(x)$ is

$$\int f(x)dx \approx \frac{1}{m}\sum_{i=1}^{m}\frac{f(x_i)}{h(x_i)} = \frac{c}{m}\sum_{i=1}^{m} f(x_i) \quad \text{for large m}$$

It is based on central limit theory that the approximation requires a large number of simulated values of variable X to converge to the target integral by the law of large numbers. The approximation of the marginal likelihoods based on importance sampling densities above are:

$$m^{\text{ZIGP}} \approx \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\lambda_j^{n-0.5}\frac{e^{-s\theta_j}}{(e^{\lambda_j}-1)^{n-n_0}}\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$$

$$m^{\text{ZIP}} \approx \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\frac{\lambda_j^{s-0.5}}{(e^{\lambda_j}-1)^{n-n_0}}$$

$$m^{\text{GP}} \approx \frac{1}{\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\lambda_j^{n-0.5}e^{-s\theta_j-n\lambda_j}\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$$

where the subscript j represents the jth step in the simulation. λ_j and θ_j are the jth simulated values on the jth step. y_i is the ith observation of response variable Y. By applying importance sampling approach, the complex integrals are turned into simple average math problem. The difficulty here is the programming complexity, and there is loop-in-loop calculation caused by the production part $\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$ sitting in the summation on j.

The posterior probability of each distribution given data y is then computed over the marginal likelihoods

$$P(M_i|y) = \frac{P(y|M_i)}{\sum_{i=1}^{n} P(y|M_i)} = \frac{m_i}{\sum_{i=1}^{n} m_i}$$

where M_i is one of the competitive ZIGP, ZIP, GP, and Poisson distributions, and the prior probability of each distribution is assumed to be equal. m_i is

the approximated marginal likelihood corresponding to the distribution M_i. The largest value of the posterior probability implies the best choice among the competitive distributions, and the sum of posterior probabilities of all competitive distributions equals one. However, the large posterior probability, ie. close to one, does not mean the appropriateness of the distribution, since it is only considered with the other given distributions based on the prior knowledge.

4.3 Simulation of data from model

Except Poisson distribution, the other distributions are not simple standard distributions that can be directly simulated from any package of software. The cumulative distribution function of a discrete distribution is easy to obtain by the sum of probability mass function,

$$F(y) = \sum_{i=0}^{y} f(i)$$

In the quantile method(or inverse distribution transform method), we use the quantile function Q_Y of the random variable Y such that $Q_Y(p)$ is the pth quantile value of Y. It is associated with the cumulative distribution function F as follows

$$Q_Y(F_Y(a)) = a,\ F_Y(Q_Y(p)) = p$$

The cumulative distribution function is in the range of $[0, 1]$, we generate a uniform random number $U \sim Uniform(0, 1)$ and feed it as the argument to

quantile function $Q_Y(U)$. Then, the function $Q_Y(U)$ is same as the variable Y following the same cumulative distribution function.

$$Pr(Q_Y(U) \leq y) = Pr(F_Y(Q_Y(U)) \leq F_Y(y)) = Pr(U \leq F_Y(y)) = F_Y(y)$$

To generate the random variable with these discrete distributions, we can apply the uniform distribution to find the corresponding quantile value of the desired distribution.

The process of simulation

1. fit the distribution with the initial values of parameters
2. initiate the start state $y = 0$
3. draw a random value u from uniform distribution $U(0, 1)$
4. accept the draw according to the probability $F(Y = y)$
5. If $u \leq F(Y = y)$, accepted and return the value of y
6. Else, let $y = y + 1$, repeat step 3 and 4

This process is to simulate one possible value from the distribution, in order to simulate more than one value, run the same process multiple times to obtain the desired number of simulated values.

4.4 Simulation results using conditionally uniform priors

The simulation analysis of comparative distributions is studied for the distribution selection with conditionally uniform priors. Different values of parameters of each distribution are defined for simulation analysis and 100 data points are randomly simulated from each defined distribution. Posterior probability of all candidate distributions are calculated over the simulated data using the conditionally uniform prior.

The flat densities $\lambda \sim Uniform(0, 15)$ and $\theta \sim Uniform(0, 1)$ are selected as importance sampling densities in comparing distributions in Poisson family. They are working well on approaching the marginal likelihoods of posterior probability with just 2000 samples from importance sampling densities.

The overall posterior probability of each comparative distribution is concluded from 100 iterations of simulation analysis over 100 simulated data sets from the same distribution. The results of 100 iterations include the mean of posterior probability and variance of posterior probability over 100 simulated data.

First, Poisson model is studied based on simulation analysis of the posterior probability of comparative models using conditionally uniform priors. The simulation results are given in table 4.2. The first column of the table gives the information of Poisson model parameter, where α and ϕ are set to be zero because they are parameters of other models. From the outcome. it is seen that the method is able to find the true model, Poisson distribution, for even

small values of parameter λ. The probability of choosing the true model is very high even when λ is small.

Table 4.2: Posterior probability of models using conditionally uniform priors

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0,0,1)	0.067038	0.195052	0.077599	0.66031	0.094287	0.087121	0.060059	0.142795
(0,0,2)	0.030922	0.147421	0.092679	0.728978	0.0542	0.115232	0.084049	0.152204
(0,0,3)	0.017801	0.08989	0.08799	0.80432	0.068511	0.078741	0.070929	0.134994
(0,0,4)	0.009425	0.055265	0.1218	0.813509	0.014067	0.061589	0.133794	0.159872
(0,0,5)	0.004727	0.03015	0.110382	0.854742	0.010115	0.021689	0.122598	0.132939

True model is Poisson distribution determined by λ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Next, generalized Poisson model is studied from simulation analysis using conditionally uniform priors, for these cases, parameter ϕ is always zero. Since the GP model parameter α is limited by the λ in the form of $\alpha < 1/\lambda$, the values of parameter α are studied depending on the value of parameter λ. Basically, four different parameter values of α are studied for each given value of λ. The results of simulated data are given in table 4.3, and from the results, posterior probability works well on selecting GP model even for small value of α. The probability of choosing true model is high except the case of small values of both α and λ that is $(\alpha = 0.2, \lambda = 1)$.

Zero-inflated Poisson model is also tested through the simulation process using conditionally uniform priors. The model parameter ϕ is studied for values ranged from 0.1 to 0.8. The results from simulation analysis in table 4.4 indicate that the posterior probability always has high probability of ZIP model when the parameter λ is larger than two. When λ is small, the

Table 4.3: Posterior probability of models using conditionally uniform priors

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.2,0,1)	0.247998	0.220556	0.389476	0.14197	0.179483	0.224408	0.202353	0.197698
(0.3,0,1)	0.320756	0.076561	0.590581	0.012102	0.131026	0.143602	0.157846	0.047159
(0.5,0,1)	0.353015	8.08E-05	0.646905	1.24E-08	0.130814	0.000448	0.130946	8.99E-08
(0.8,0,1)	0.377618	3.22E-36	0.622382	7.01E-60	0.179864	3.20E-35	0.179864	7.01E-59
(0.1,0,2)	0.143031	0.104033	0.621241	0.131695	0.085484	0.140927	0.213711	0.178933
(0.2,0,2)	0.209497	0.000281	0.790217	6.12E-06	0.144472	0.001757	0.144763	5.18E-05
(0.3,0,2)	0.216771	1.19E-15	0.783229	6.40E-21	0.160625	1.19E-14	0.160625	6.40E-20
(0.4,0,2)	0.195514	5.05E-66	0.804486	2.79E-101	0.09874	5.05E-65	0.09874	2.79E-100
(0.06,0,3)	0.089688	0.063248	0.624388	0.222676	0.098413	0.141266	0.288686	0.270051
(0.1,0,3)	0.111022	0.00392	0.877043	0.008015	0.096974	0.016265	0.101772	0.028958
(0.2,0,3)	0.131328	1.82E-21	0.868672	3.92E-34	0.117304	1.82E-20	0.117304	3.83E-33
(0.3,0,3)	0.137257	0	0.862743	0	0.153767	0	0.153767	0
(0.05,0,4)	0.060195	0.029542	0.745033	0.165229	0.07431	0.099143	0.264632	0.246322
(0.1,0,4)	0.070906	7.04E-07	0.929041	5.17E-05	0.063566	6.29E-06	0.063525	0.000512
(0.15,0,4)	0.085839	8.82E-28	0.914161	1.36E-28	0.132745	7.54E-27	0.132745	1.36E-27
(0.2,0,4)	0.063084	8.72E-153	0.936916	1.58E-158	0.049834	8.72E-152	0.049834	1.58E-157
(0.04,0,5)	0.032403	0.013482	0.776426	0.177688	0.064211	0.038852	0.259957	0.247574
(0.08,0,5)	0.03428	1.05E-05	0.964735	0.000975	0.032556	0.000103	0.033291	0.009682
(0.1,0,5)	0.046172	5.63E-14	0.953828	5.87E-12	0.065131	5.60E-13	0.065131	5.87E-11
(0.15,0,5)	0.040123	1.59E-116	0.959877	9.66E-111	0.04758	1.59E-115	0.04758	9.66E-110

True model is generalized Poisson distribution determined by λ and α. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

posterior probability prefer ZIP model when the model parameter ϕ is larger than 0.2, otherwise the preference of posterior probability tends to vary between ZIP and Poisson model for small ϕ.

Table 4.4: Posterior probability of models using conditionally uniform priors

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$Var[P(ZIGP\|y)]$	$Var[P(ZIP\|y)]$	$Var[P(GP\|y)]$	$Var[P(Poi\|y)]$
(0,0.1,1)	0.082236	0.217873	0.138614	0.561277	0.094289	0.108284	0.110117	0.202956
(0,0.2,1)	0.107777	0.317948	0.189809	0.384465	0.076689	0.184424	0.112745	0.246518
(0,0.5,1)	0.171548	0.537395	0.198211	0.092847	0.079088	0.204824	0.130097	0.150571
(0,0.8,1)	0.215555	0.472999	0.251621	0.059824	0.080208	0.194515	0.145345	0.124713
(0,0.1,2)	0.066435	0.372138	0.17795	0.383477	0.058589	0.241688	0.15529	0.271284
(0,0.2,2)	0.119115	0.585676	0.183726	0.111484	0.075643	0.253878	0.181019	0.169726
(0,0.5,2)	0.174026	0.797662	0.028294	1.71E-05	0.117632	0.171977	0.061875	8.19E-05
(0,0.8,2)	0.237883	0.713738	0.048377	1.86E-06	0.11183	0.167815	0.07017	1.21E-05
(0,0.1,3)	0.075748	0.594525	0.121042	0.208686	0.056656	0.298791	0.152541	0.262756
(0,0.2,3)	0.124213	0.8473	0.026973	0.001514	0.086474	0.136152	0.071252	0.004766
(0,0.5,3)	0.167577	0.831854	0.000568	6.51E-13	0.136797	0.138703	0.003621	5.56E-12
(0,0.8,3)	0.248959	0.734501	0.01654	1.14E-12	0.15278	0.186885	0.069469	1.13E-11
(0,0.1,4)	0.124645	0.765782	0.058782	0.050791	0.114179	0.245411	0.126876	0.142152
(0,0.2,4)	0.167475	0.829575	0.002936	1.48E-05	0.161519	0.168109	0.014387	0.000133
(0,0.5,4)	0.150368	0.849631	1.23E-06	1.31E-23	0.127526	0.12753	7.85E-06	9.41E-23
(0,0.8,4)	0.226854	0.772745	0.000401	1.42E-23	0.151529	0.152113	0.001422	1.18E-22
(0,0.1,5)	0.121404	0.85038	0.007079	0.021138	0.142444	0.170352	0.025924	0.093463
(0,0.2,5)	0.138268	0.86173	2.55E-06	1.14E-08	0.1254	0.125401	2.44E-05	9.15E-08
(0,0.5,5)	0.151525	0.848475	7.18E-11	1.29E-33	0.124996	0.124996	3.57E-10	1.24E-32
(0,0.8,5)	0.221124	0.778747	0.000129	4.03E-24	0.149818	0.149943	0.000678	4.03E-23

True model is zero-inflated Poisson distribution determined by λ and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

In the end, the full model zero-inflated generalized Poisson model is considered in simulation analysis using conditionally uniform priors. Simulation study of ZIGP model is showed in the following tables 4.5, 4.6, 4.7. The re-

sulst show our method is able to find true model when $\lambda \geq 3$ even when model paramters α and ϕ are both small. When $\lambda = 1$, our method is in favor of GP model when zero-inflated parameter ϕ is small at 0.1 , and it finds ZIP model better when ϕ is increased to 0.5 and α is smaller than 0.3, otherwise ZIGP, ZIP and GP models all are candidate models but ZIGP has larger probability of selection. For $\lambda = 2$, GP model has large probability when α and ϕ are all close to 0. When ϕ increases but α is still small, the probability of ZIP is larger compared with other models.

These simulation results demonstrate our method successfully chooses the correct distribution every time when λ is larger than 2, even if the values of distribution parameters are quite small. When λ is small, the shape of these distributions are similar to each other as the distribution parameters are close to zero. It is because Bayesian method automatically includes the penalty of complexity of model, so it will choose simpler distribution when comparative distributions are similar to each other, in this case, the parameters are close to zero when λ is small. Therefore posterior probability tends to prefer the distribution with less parameters when the difference between competing distributions is not great. Meanwhile these distributions stay away from each other even for small values of parameters when λ is large.

Table 4.5: Posterior probability of models using conditionally uniform priors: $\lambda = 1$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.2,0.1,1)	0.263724	0.242179	0.426081	0.068015	0.136111	0.210464	0.168771	0.135549
(0.3,0.1,1)	0.366172	0.121541	0.510512	0.001775	0.123177	0.205803	0.205144	0.008749
(0.4,0.1,1)	0.397287	0.039518	0.563088	0.000108	0.136981	0.115414	0.1731	0.001021
(0.5,0.1,1)	0.428239	0.011019	0.560742	3.58E-10	0.151344	0.064013	0.169855	2.10E-09
(0.6,0.1,1)	0.447374	9.10E-07	0.552625	1.70E-23	0.168485	6.93E-06	0.168487	1.59E-22
(0.7,0.1,1)	0.466866	4.05E-11	0.533134	2.92E-33	0.198102	4.04E-10	0.198102	2.92E-32
(0.8,0.1,1)	0.427839	9.75E-29	0.572161	5.48E-61	0.154458	9.63E-28	0.154458	5.48E-60
(0.2,0.5,1)	0.320538	0.447504	0.23093	0.001029	0.148777	0.295002	0.199787	0.005119
(0.3,0.5,1)	0.434403	0.272507	0.292361	0.000729	0.146926	0.284255	0.203278	0.007271
(0.4,0.5,1)	0.572598	0.115986	0.311414	1.80E-06	0.175997	0.191489	0.203445	1.80E-05
(0.5,0.5,1)	0.60588	0.025784	0.368335	3.57E-14	0.185414	0.098235	0.200011	2.45E-13
(0.6,0.5,1)	0.629396	0.011259	0.359346	1.21E-20	0.195046	0.057445	0.205967	9.09E-20
(0.7,0.5,1)	0.670187	0.000153	0.32966	2.39E-17	0.193218	0.000984	0.193307	2.39E-16
(0.8,0.5,1)	0.646627	1.26E-06	0.353371	1.84E-38	0.219336	1.26E-05	0.219337	1.84E-37
(0.2,0.8,1)	0.30394	0.425318	0.257005	0.013737	0.105552	0.2031	0.151442	0.052541
(0.3,0.8,1)	0.334276	0.430333	0.232116	0.003275	0.132242	0.265407	0.18101	0.023452
(0.4,0.8,1)	0.431189	0.280043	0.287461	0.001307	0.149034	0.253937	0.197462	0.012938
(0.5,0.8,1)	0.521717	0.175605	0.302676	2.90E-06	0.156695	0.223965	0.185452	2.56E-05
(0.6,0.8,1)	0.543618	0.116726	0.33965	6.47E-06	0.196078	0.220348	0.217984	6.37E-05
(0.7,0.8,1)	0.632781	0.042544	0.324675	2.38E-11	0.196053	0.126126	0.201634	2.15E-10
(0.8,0.8,1)	0.636544	0.019238	0.344219	4.08E-15	0.201465	0.096324	0.202508	3.53E-14

True model is zero-inflated generalized Poisson distribution determined by λ,α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.6: Posterior probability of models using conditionally uniform priors: $\lambda = 2, 3$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.1,0.1,2)	0.255944	0.193998	0.529915	0.020143	0.134921	0.237401	0.253665	0.077564
(0.2,0.1,2)	0.406849	0.002096	0.591055	1.02E-09	0.25888	0.01838	0.262653	5.80E-09
(0.4,0.1,2)	0.411213	5.86E-61	0.588787	1.09E-125	0.264091	5.86E-60	0.264091	1.09E-124
(0.1,0.5,2)	0.478536	0.455196	0.066268	4.17E-10	0.271912	0.330622	0.116753	3.90E-09
(0.2,0.5,2)	0.884601	0.04039	0.07501	1.93E-21	0.150848	0.107653	0.130106	1.91E-20
(0.4,0.5,2)	0.923785	1.55E-18	0.076215	4.73E-91	0.124866	1.55E-17	0.124866	4.73E-90
(0.1,0.8,2)	0.386627	0.496111	0.117262	1.43E-07	0.176701	0.281005	0.1619	1.43E-06
(0.2,0.8,2)	0.627821	0.272305	0.099875	8.75E-13	0.254647	0.30383	0.125651	8.26E-12
(0.4,0.8,2)	0.90202	0.001343	0.096636	4.64E-45	0.125721	0.007451	0.126485	4.64E-44
(0.1,0.1,3)	0.609116	0.040794	0.35009	5.21E-07	0.281417	0.114313	0.303098	2.89E-06
(0.2,0.1,3)	0.615664	3.29E-16	0.384336	4.68E-36	0.312669	3.29E-15	0.312669	4.68E-35
(0.3,0.1,3)	0.676429	0	0.323571	0	0.320074	0	0.320074	0
(0.1,0.5,3)	0.825189	0.173851	0.00096	1.36E-24	0.26282	0.263347	0.007418	1.33E-23
(0.2,0.5,3)	0.999402	1.75E-09	0.000598	2.99E-73	0.002626	1.46E-08	0.002626	2.99E-72
(0.3,0.5,3)	0.998972	5.09E-131	0.001028	0	0.004525	5.09E-130	0.004525	0
(0.1,0.8,3)	0.621163	0.363059	0.015777	2.06E-16	0.281714	0.292762	0.051075	1.94E-15
(0.2,0.8,3)	0.949123	0.033324	0.017553	2.35E-36	0.113257	0.102189	0.059418	2.35E-35
(0.3,0.8,3)	0.978115	4.83E-07	0.021884	5.57E-65	0.054332	4.83E-06	0.054332	5.57E-64

True model is zero-inflated generalized Poisson distribution determined by λ,α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.7: Posterior probability of models using conditionally uniform priors: $\lambda = 4, 5$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.05,0.1,4)	0.635631	0.25506	0.108375	0.000934	0.284433	0.28054	0.20028	0.006501
(0.1,0.1,4)	0.899197	0.001043	0.099761	3.04E-15	0.19218	0.007196	0.192591	1.99E-14
(0.2,0.1,4)	0.908301	7.79E-161	0.091699	0	0.197559	7.79E-160	0.197559	0
(0.05,0.5,4)	0.664946	0.335053	5.60E-07	9.82E-29	0.303643	0.303644	1.87E-06	9.82E-28
(0.1,0.5,4)	0.982244	0.017755	1.28E-06	5.95E-58	0.08589	0.085891	7.82E-06	4.42E-57
(0.2,0.5,4)	0.999997	4.21E-35	3.22E-06	7.00E-216	2.31E-05	4.21E-34	2.31E-05	0
(0.05,0.8,4)	0.401427	0.598121	0.000452	4.42E-28	0.241045	0.241852	0.002498	3.60E-27
(0.1,0.8,4)	0.88189	0.115154	0.002956	2.01E-46	0.17289	0.17392	0.017866	1.34E-45
(0.2,0.8,4)	0.99867	6.83E-09	0.00133	4.97E-111	0.005957	4.18E-08	0.005957	4.93E-110
(0.05,0.1,5)	0.8978	0.069579	0.03262	3.72E-08	0.201158	0.171714	0.124322	3.29E-07
(0.1,0.1,5)	0.968515	1.58E-12	0.031485	5.15E-21	0.120996	1.55E-11	0.120996	5.15E-20
(0.15,0.1,5)	0.956565	4.46E-108	0.043435	1.66E-152	0.157401	4.46E-107	0.157401	1.66E-151
(0.05,0.5,5)	0.762383	0.237617	3.67E-10	2.20E-56	0.279402	0.279402	2.03E-09	1.94E-55
(0.1,0.5,5)	0.999999	1.10E-06	5.50E-11	1.69E-101	5.62E-06	5.61E-06	1.95E-10	1.69E-100
(0.15,0.5,5)	1	1.09E-55	3.81E-10	0	2.37E-09	1.09E-54	2.39E-09	0
(0.05,0.8,5)	0.565335	0.434644	2.09E-05	2.17E-32	0.300953	0.300973	8.29E-05	2.17E-31
(0.1,0.8,5)	0.951723	0.048261	1.63E-05	1.27E-67	0.124136	0.124142	8.72E-05	1.24E-66
(0.15,0.8,5)	0.999973	1.65E-07	2.72E-05	2.32E-161	0.000119	1.61E-06	0.000119	1.63E-160

True model is zero-inflated generalized Poisson distribution determined by λ,α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

4.5 Derivation of Jeffreys priors

As before, the notation is defined that the number of observations in the data n, the number of zeros n_0, and the sum of the observations s. As the reparameterization in *Gupta et al.* (1996), let $\eta = (1-\phi)(1-e^{-\lambda})$. Then the parameters η and λ are orthogonal, and the prior density of η is uniform on (0,1) independently of λ. $P(Y=0) = 1-\eta$ and $P(Y>0) = \eta$. The likelihoods of models are then the following.

$$
\begin{aligned}
L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda) &= \prod_{i=1}^{n} f(y_i|\eta,\alpha,\lambda) \\
&= (1-\eta)^{n_0}\eta^{n-n_0}\frac{e^{-s\alpha\lambda}}{(e^{\lambda}-1)^{n-n_0}}\lambda^{s}\prod_{i=1}^{n}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \\
L^{\mathrm{ZIP}}(\eta,\lambda) &= \prod_{i=1}^{n} f(y_i|\eta,\lambda) = (1-\eta)^{n_0}\eta^{n-n_0}\frac{1}{(e^{\lambda}-1)^{n-n_0}}\frac{\lambda^{s}}{\prod_{i=1}^{n} y_i!} \\
L^{\mathrm{GP}}(\alpha,\lambda) &= \prod_{i=1}^{n} f(y_i|\alpha,\lambda) = (\lambda e^{-\alpha\lambda})^{s}e^{-\lambda n}\prod_{i=1}^{n}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!} \\
L^{\mathrm{Poi}}(\lambda) &= \prod_{i=1}^{n} f(y_i|\lambda) = \frac{\lambda^{s}e^{-\lambda n}}{\prod_{i=1}^{n} y_i!}
\end{aligned}
$$

Only ZIGP and ZIP distributions contain the reparameterized parameter, and GP and Poi likelihoods are the same as before. Then the fact that the parameters η and other parameters are orthogonal is proved in the following Fisher information matrix.

The Fisher information for ZIP distribution over parameters (η, λ) is

$$I^{\text{ZIP}}(\eta,\lambda) = \begin{bmatrix} -E[\frac{d^2}{d\eta^2}\ln L^{\text{ZIP}}(\eta,\lambda)] & -E[\frac{d^2}{d\eta d\lambda}\ln L^{\text{ZIP}}(\eta,\lambda)] \\ -E[\frac{d^2}{d\eta d\lambda}\ln L^{\text{ZIP}}(\eta,\lambda)] & -E[\frac{d^2}{d\lambda^2}\ln L^{\text{ZIP}}(\eta,\lambda)] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\eta(1-\eta)} & 0 \\ 0 & \frac{\eta(1-e^{-\lambda}-\lambda e^{-\lambda})}{\lambda(1-e^{-\lambda})^2} \end{bmatrix}$$

The Jeffreys prior of ZIP distribution derived from Fisher information is

$$\pi_J^{\text{ZIP}}(\eta,\lambda) \propto \sqrt{\frac{1-e^{-\lambda}-\lambda e^{-\lambda}}{(1-\eta)(1-e^{-\lambda})^2\lambda}}$$

where $0 < \eta < 1$, $0 < \lambda$.

the Fisher information of ZIP distribution is diagonal that the parameter η is orthogonal with λ. It is easy to see that the Jeffreys prior of λ in ZIP is different with Poisson distribution $\pi_J^{\text{Poi}}(\lambda) \propto \frac{1}{\sqrt{\lambda}}$, and the relationship is

$$\pi_J^{\text{ZIP}}(\lambda) \propto K_J^{\text{ZIP}}(\lambda)\pi_J^{\text{Poi}}(\lambda)$$
$$where \quad K_J^{\text{ZIP}}(\lambda) = \frac{\sqrt{1-e^{-\lambda}-\lambda e^{-\lambda}}}{1-e^{-\lambda}}$$

After the orthogonalization, the Jeffreys prior for the common parameter λ

is different in ZIP and Poisson distribution which is quite rare. *Bayarri et al.* (2008) discusses that the λ also enters into the definition of the nested model, through $\eta = 1 - e^{-\lambda}$. There is no clear guidance of which of the Jeffreys priors is the choice, in view of the indeterminacy issue in the Jeffreys priors.

Based on the fact that the Jeffreys prior of λ in ZIP is the Jeffreys prior in Poisson multiplying a function of λ which is $K_J^{\text{ZIP}}(\lambda)$, even though it is not clear which Jeffreys prior should be the one used for the common parameter λ, *Bayarri et al.* (2008) demonstrated Jeffreys prior of λ in ZIP is near the Jeffreys prior of λ in Poisson and is approximating it as λ goes to infinity. The ratio of these two Jeffreys priors, $K^{\text{ZIP}}(\lambda)$, is strictly increasing over the domain and bounded in a small range close to 1. By applying L'Hopital's rule to evaluate limits involving indeterminate forms.

$$\lim_{\lambda \to 0} \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{(1 - e^{-\lambda})^2} = \lim_{\lambda \to 0} \frac{\lambda}{2(1 - e^{-\lambda})} = \lim_{\lambda \to 0} \frac{1}{2e^{-\lambda}} = \frac{1}{2}$$
$$\lim_{\lambda \to \infty} \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{(1 - e^{-\lambda})^2} = 1$$

Thus, the boundary of $K_J^{\text{ZIP}}(\lambda)$ is

$$\sqrt{\frac{1}{2}} < \frac{\sqrt{1 - e^{-\lambda} - \lambda e^{-\lambda}}}{1 - e^{-\lambda}} < 1$$

These two Jeffreys priors for the common parameter λ are quite close, therefore we can take the Jeffreys prior of λ in Poisson model as an approximate Jeffreys prior of the common parameter λ in ZIP model for comparing these two models by means of Bayes factor. In this way, the approximate Jeffreys prior of ZIGP

(A-J prior) can be written as

$$\pi^{\text{ZIP}}_{A-J}(\eta,\lambda) \propto K^{\text{ZIP}}_{A-J}(\eta,\lambda)\pi^{\text{Poi}}_{J}(\lambda)$$
$$where \quad K^{\text{ZIP}}_{A-J}(\eta,\lambda) = \frac{1}{\sqrt{1-\eta}}$$

Jeffreys had recommended to use Jeffreys prior for the common parameter (λ), and proper prior for the parameter orthogonal with the remaining parameters. For the parameter η, because it is orthogonal to the common parameter λ, it is reasonable to propose a proper prior for the extra parameter. Modification of Jeffreys prior of η is considered by using uniform prior over its domain $(0,1)$.

$$\pi(\eta|\lambda) = \mathrm{I}_{(0<\eta<1)}$$

Modified Jeffreys prior of ZIP becomes

$$\pi^{\text{ZIP}}_{M-J}(\lambda) \propto K^{\text{ZIP}}_{M-J}(\lambda)\pi^{\text{Poi}}_{J}(\lambda)$$
$$where \quad K^{\text{ZIP}}_{M-J}(\lambda) = \mathrm{I}_{(0<\eta<1)}$$

Thus, the ratio K of Jeffreys priors in ZIP over Poisson being considered is summarized as follows.

$$K^{\text{ZIP}}_{l}(\eta,\lambda) = \begin{cases} \frac{1}{\sqrt{1-\eta}}\frac{\sqrt{1-e^{-\lambda}-\lambda e^{-\lambda}}}{1-e^{-\lambda}} & l = \text{J} \\ \frac{1}{\sqrt{1-\eta}} & l = \text{A-J} \\ \mathrm{I}_{(0<\eta<1)} & l = \text{M-J} \end{cases}$$

where l denotes the choice among defined priors for ZIP model. The Jeffreys prior for ZIP is now

$$\pi_l^{\mathrm{ZIP}}(\eta,\lambda) \propto K_l^{\mathrm{ZIP}}(\eta,\lambda)\pi_J^{\mathrm{Poi}}(\lambda)$$

where $0<\eta<1$, $0<\lambda$. $l = \mathrm{J}, \mathrm{A\text{-}J}, \mathrm{M\text{-}J}$.

Generalized Poisson distribution is another distribution developed on the basis of Poisson distribution with one more parameter to adjust the variance independently with the mean. To find the Jeffreys prior in GP distribution, the Fisher information for GP distribution is

$$I(\alpha,\lambda) = \begin{bmatrix} -E[\frac{d^2}{d\alpha^2}\ln L^{\mathrm{GP}}(\alpha,\lambda)] & -E[\frac{d^2}{d\alpha d\lambda}\ln L^{\mathrm{GP}}(\alpha,\lambda)] \\ -E[\frac{d^2}{d\alpha d\lambda}\ln L^{\mathrm{GP}}(\alpha,\lambda)] & -E[\frac{d^2}{d\lambda^2}\ln L^{\mathrm{GP}}(\alpha,\lambda)] \end{bmatrix}$$

$$= \begin{bmatrix} \frac{n\lambda^2(\lambda+2)}{(1+2\alpha)(1-\alpha\lambda)} & \frac{n\lambda}{1-\alpha\lambda} \\ \frac{n\lambda}{1-\alpha\lambda} & \frac{n}{\lambda(1-\alpha\lambda)} \end{bmatrix}$$

The Jeffreys prior for GP distribution is

$$\pi_J^{\mathrm{GP}}(\alpha,\lambda) \propto \frac{\sqrt{\lambda}}{\sqrt{(1+2\alpha)(1-\alpha\lambda)}}$$

where $0 \le \alpha < 1/\lambda$, $0<\lambda$

As the Jeffreys prior in ZIP, the relationship of GP and Poisson distribution is studied and expressed in the following.

$$\pi_J^{\mathrm{GP}}(\alpha,\lambda) \propto K_J^{\mathrm{GP}}(\alpha,\lambda)\pi_J^{\mathrm{Poi}}(\lambda)$$
$$where \quad K_J^{\mathrm{GP}}(\alpha,\lambda) \propto \frac{\lambda}{\sqrt{(1+2\alpha)(1-\alpha\lambda)}}$$

The parameter α is related to the parameter λ, and the normalizing constant is important to Bayes factor, because the normalizing constant can be treated as the multiplicative constant to the Jeffreys prior in Poisson distribution. By transformation of $x = \alpha - \frac{2-\lambda}{4\lambda}$, the integral of the conditional prior turns out to be

$$\int_0^{1/\lambda} \frac{\lambda}{\sqrt{(1+2\alpha)(1-\alpha\lambda)}} d\alpha = \frac{\sqrt{8}\lambda^{3/2}}{\lambda+2}[arcsin(1) - arcsin(\frac{\lambda-2}{\lambda+2})]$$

The limit of the integral is easy to get as λ goes to zero, but the other limit as λ goes to infinity contains indeterminate forms. Apply L'Hopital's rule,

$$lim_{\lambda\to\infty} \frac{\sqrt{8}\lambda^{3/2}[arcsin(1) - arcsin(\frac{\lambda-2}{\lambda+2})]}{\lambda+2} = 8$$
$$lim_{\lambda\to 0} \frac{\sqrt{8}\lambda^{3/2}[arcsin(1) - arcsin(\frac{\lambda-2}{\lambda+2})]}{\lambda+2} = 0$$

The range of the normalizing constant of the conditional prior of parameter α given λ is from 0 to 8. The function K_J^{GP} as a function of lambda is integrable, but the normalizing constant is not equal to one, and so we cannot consider $K_J^{\mathrm{GP}}(\alpha)$ as a conditional density of alpha given lambda. The

normalizing constant also depends on lambda, and has a range that includes the interval (0,8). This makes the Jeffreys prior not appropriate for testing hypothesis, as it would cause an a priori bias in the Bayes factor. One way to overcome this and still use the Jeffreys prior is to use partial Bayes factors instead such as Intrinsic or Fractional Bayes factors.

Next, the ZIGP Jeffreys prior is going to be developed and the relationship between Jeffreys priors of ZIGP and GP is studied. It shows that the relationship of ZIGP and GP is similar with the one of ZIP and Poisson, and we can rewrite the Jeffreys prior of ZIGP in the same way as ZIP to reduce the indeterminacy caused by the arbitrary normalizing constant. Let $\eta = (1-\phi)(1-e^{-\lambda})$, then the parameter η is orthogonal with (α, λ). The Fisher information for ZIGP distribution is

$$
\begin{aligned}
I(\eta,\alpha,\lambda) &= \begin{bmatrix} -E[\frac{d^2}{d\eta^2}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] & -E[\frac{d^2}{d\eta d\alpha}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] & -E[\frac{d^2}{d\eta d\lambda}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] \\ -E[\frac{d^2}{d\eta d\alpha}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] & -E[\frac{d^2}{d\alpha^2}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] & -E[\frac{d^2}{d\alpha d\lambda}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] \\ -E[\frac{d^2}{d\eta d\lambda}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] & -E[\frac{d^2}{d\alpha d\lambda}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] & -E[\frac{d^2}{d\lambda^2}\ln L^{\mathrm{ZIGP}}(\eta,\alpha,\lambda)] \end{bmatrix} \\
&= \begin{bmatrix} \frac{n}{(1-\eta)\eta} & 0 & 0 \\ 0 & \frac{n\eta\lambda^2(\lambda+2)}{(1-e^{-\lambda})(1+2\alpha)(1-\alpha\lambda)} & \frac{n\eta\lambda}{(1-e^{-\lambda})(1-\alpha\lambda)} \\ 0 & \frac{n\eta\lambda}{(1-e^{-\lambda})(1-\alpha\lambda)} & \frac{n\eta e^{\lambda}(e^{\lambda}+\alpha\lambda^2-\lambda-1)}{\lambda(1-\alpha\lambda)(e^{\lambda}-1)^2} \end{bmatrix}
\end{aligned}
$$

From Fisher information matrix, since the Fisher information cells of parameter η with α, λ are all zeros, the parameter η is orthogonal with the other two parameters (η, α).

The Jeffreys prior for ZIGP distribution is

$$\pi_J^{\text{ZIGP}}(\eta, \alpha, \lambda) \propto \sqrt{\frac{\eta\lambda(2e^\lambda - \lambda^2 - 2\lambda - 2)e^{2\lambda}}{(1-\eta)(1+2\alpha)(1-\alpha\lambda)(e^\lambda - 1)^3}}$$

where $0 < \eta < 1$, $0 \leq \alpha < 1/\lambda$, $0 < \lambda$

Since the Jeffreys prior for GP distribution is

$$\pi_J^{\text{GP}}(\alpha, \lambda) \propto \frac{\sqrt{\lambda}}{\sqrt{(1+2\alpha)(1-\alpha\lambda)}}$$

where $0 \leq \alpha < 1/\lambda$, $0 < \lambda$

The ratio of Jeffreys priors ZIGP over GP is

$$\frac{\pi_J^{\text{ZIGP}}(\eta, \alpha, \lambda)}{\pi_J^{\text{GP}}(\alpha, \lambda)} \propto \sqrt{\frac{\eta}{1-\eta}\frac{(2e^\lambda - \lambda^2 - 2\lambda - 2)e^{2\lambda}}{(e^\lambda - 1)^3}}$$

Since the parameter η is orthogonal to the common parameters α, λ of ZIGP and GP model, the ratio of Jeffreys prior of the joint parameters α, λ is a function of λ, $K_J^{\text{ZIGP}}(\alpha, \lambda) = \sqrt{\frac{(2e^\lambda - \lambda^2 - 2\lambda - 2)e^{2\lambda}}{(e^\lambda - 1)^3}}$. Its range determined by the domain of λ is $\frac{1}{\sqrt{3}} < K(\lambda) < \sqrt{2}$. Since $K(\lambda)$ is bounded in a small range around one, Jeffreys prior of α, λ in ZIGP model can be approximated by the Jeffreys prior of GP model. This approximate Jeffreys prior can be written as

$$\pi_{A-J}^{\text{ZIGP}}(\eta, \alpha, \lambda) \propto K_{A-J}^{\text{ZIGP}}(\eta, \alpha, \lambda)\pi_J^{\text{GP}}(\alpha, \lambda)$$
$$\textit{where} \quad K_{A-J}^{\text{ZIGP}}(\eta, \alpha, \lambda) \propto \sqrt{\frac{\eta}{1-\eta}}$$

Due to the orthogonality of the parameter η with the other parameters (α, λ),

a reasonable choice of proper prior on η is a uniform prior such that

$$\pi^{\text{ZIGP}}(\eta|\alpha,\lambda) = \text{I}_{(0<\eta<1)},\ 0<\eta<1$$

Therefore modified Jeffreys prior for parameters (η,α,λ) in ZIGP distribution can be defined as

$$\pi^{\text{ZIGP}}_{M-J}(\eta,\alpha,\lambda) \propto K^{\text{ZIGP}}_{M-J}(\eta,\alpha,\lambda)\pi^{\text{GP}}_{J}(\alpha,\lambda)$$
$$where \quad K^{\text{ZIGP}}_{M-J}(\eta,\alpha,\lambda) \propto \text{I}_{(0<\eta<1)}$$

Thus, the ratio K of Jeffreys priors in ZIGP over GP can be defined from the following choices.

$$K^{\text{ZIGP}}_{l}(\eta,\alpha,\lambda) = \begin{cases} \sqrt{\frac{\eta}{1-\eta}\frac{(2e^{\lambda}-\lambda^2-2\lambda-2)e^{2\lambda}}{(e^{\lambda}-1)^3}} & l = \text{J} \\ \sqrt{\frac{\eta}{1-\eta}} & l = \text{A-J} \\ \text{I}_{(0<\eta<1)} & l = \text{M-J} \end{cases}$$

where l denotes the choice among defined priors for ZIP model. The Jeffreys prior for ZIP is now

$$\pi^{\text{ZIGP}}_{l}(\eta,\alpha,\lambda) \propto K^{\text{ZIGP}}_{l}(\eta,\alpha,\lambda)\pi^{\text{GP}}_{J}(\alpha,\lambda)$$

where $0<\eta<1$, $0\leq\alpha<1/\lambda$, $0<\lambda$. $l = \text{J}, \text{A-J}, \text{M-J}$.

The modified Jeffreys priors (M-J priors) are considered in next section for the model selection of Poisson-related models. Jeffreys prior is not necessarily

Table 4.8: Jeffreys priors for distribution selection

Poisson
$\pi_J^{\text{Poi}}(\lambda) = \frac{1}{\sqrt{\lambda}}$ $0 < \lambda$
ZIP
$\pi_J^{\text{ZIP}}(\eta, \lambda) = K_l^{\text{ZIP}}(\eta, \lambda)\pi_J^{Poi}(\lambda)$ $0 \leq \eta < 1, 0 < \lambda$ $K_l^{\text{ZIP}}(\eta, \lambda) = \begin{cases} \frac{1}{\sqrt{1-\eta}} \frac{\sqrt{1-e^{-\lambda}-\lambda e^{-\lambda}}}{1-e^{-\lambda}} & l = \text{J} \\ \frac{1}{\sqrt{1-\eta}} & l = \text{A-J} \\ \text{I}_{(0<\eta<1)} & l = \text{M-J} \end{cases}$
GP
$\pi_J^{\text{GP}}(\alpha, \lambda) = K_J^{\text{GP}}(\alpha, \lambda)\pi_J^{\text{Poi}}(\lambda)$ $0 \leq \alpha < 1/\lambda, 0 < \lambda$ $K_J^{\text{GP}}(\alpha, \lambda) = \frac{\lambda}{\sqrt{(1+2\alpha)(1-\alpha\lambda)}}$
ZIGP
$\pi_J^{\text{ZIGP}}(\eta, \alpha, \lambda) = K_1^{\text{ZIGP}}(\eta, \alpha, \lambda)\pi_J^{GP}(\alpha, \lambda)$ $0 < \eta < 1, 0 \leq \alpha < 1/\lambda, 0 < \lambda$ $K_l^{\text{ZIGP}}(\eta, \alpha, \lambda) = \begin{cases} \sqrt{\frac{\eta}{1-\eta} \frac{(2e^{\lambda}-\lambda^2-2\lambda-2)e^{2\lambda}}{(e^{\lambda}-1)^3}} & l = \text{J} \\ \sqrt{\frac{\eta}{1-\eta}} & l = \text{A-J} \\ \text{I}_{(0<\eta<1)} & l = \text{M-J} \end{cases}$

proper priors, and the undefined normalizing constant does cause uncertainty in Bayes factor ratio. Among four Poisson family models, the priors of the common parameter are improper, but the normalizing constant cancels out in Bayes factor, such that the parameter λ through all Poisson family models. However, the Jeffreys prior for α is improper prior. The arbitrary constant in the prior of uncommon parameter makes indeterminacy issues in calculation Bayes factor, due to it only presents in either denominator or numerator of Bayes factor. For such cases, when the Jeffreys priors of uncommon parameters are improper, Bayes factor is biased in testing but several adjusted method have been developed for such cases. These methods are called fractional Bayes factor and Intrinsic Bayes Factors.

4.6 Model selection using modified Jeffreys priors

The modified Jeffreys priors (M-J priors) are selected for model selection using posterior probability. Among all these models, the Jeffreys prior containing parameter α in GP distribution is improper prior. The arbitrary normalizing constant on parameter α given λ cause the indeterminacy trouble in Bayes factor. Thus, the fractional Bayes factor method is applied in order to train the Jeffreys prior into proper prior with partial data.

Fractional Bayes factor is in the form of

$$BF_{01}^{f}(x) = \frac{q_0(b, x)}{q_1(b, x)}$$

where $\pi_i(\theta_i) = c_i h_i(\theta_i)$, c_i is the unspecified constant, i represents the ith model.

$$q_i(b,x) = \frac{\int \pi_i(\theta_i) f_i(x|\theta_i) d\theta_i}{\int \pi_i(\theta_i) f_i(x|\theta_i)^b d\theta_i} = \frac{\int h_i(\theta_i) f_i(x|\theta_i) d\theta_i}{\int h_i(\theta_i) f_i(x|\theta_i)^b d\theta_i}$$

here b is the power and its value is $b = \frac{n_1}{n}$ and n_1 is partial of sample n that $n = n_1 + n_2$. Let $m_i^b = \int \pi_i(\theta_i) f_i(x|\theta_i)^b d\theta_i$, so $q_i(b,x) = \frac{m_i}{m_i^b}$

$$BF_{01}^f(x) = \frac{q_0(b,x)}{q_1(b,x)} = \frac{m_0 m_1^b}{m_1 m_0^b}$$

Now the unspecified constants cancel out.

The posterior probability of each distribution given data y is then computed over the marginal likelihoods

$$P(M_i|y) = \frac{P(y|M_i)}{\sum_{i=1}^{n} P(y|M_i)} = \frac{m_i}{\sum_{i=1}^{n} m_i}$$

where M_i is one of the competitive ZIGP, ZIP, GP, and Poisson distributions, and the prior probability of each distribution is assumed to be equal. m_i is the approximated marginal likelihood corresponding to the distribution M_i. The largest value of the posterior probability implies the best choice among the competitive distributions, and the sum of posterior probabilities of all competitive distributions equals one. However, the large posterior probability, ie. close to one, does not mean the appropriateness of the distribution, since it is only considered with the other given distributions based on the prior

knowledge.

For Multiple model comparison, test is implemented by approximate posterior probability of model M_i.

$$P(M_i|y) = \frac{m_i}{\sum_{i=1}^{n} m_i} = \frac{BF_{i,Poi}}{\sum_{i=1}^{4} BF_{i,Poi}}$$
$$\approx \frac{\frac{m^i}{m_b^i}}{\frac{m^{\text{Poi}}}{m_b^{\text{Poi}}} + \frac{m^{\text{GP}}}{m_b^{\text{GP}}} + \frac{m^{\text{ZIP}}}{m_b^{\text{ZIP}}} + \frac{m^{\text{ZIGP}}}{m_b^{\text{ZIGP}}}}$$

where i represents one of comparative distributions, $i \in$ {ZIGP,ZIP,GP,Poi}. Bayes factors of posterior probability are approached by fractional Bayes factors.

Marginal likelihood is defined as

$$m_i = P(y|M_i) = \int P(y|M_i, \theta)P(\theta)d\theta$$

where θ is the vector of parameters in model M_i.

Due to the complexity of integration in computation of marginal likelihood, a transformation of parameter α is made by letting $\theta = \alpha\lambda$. Then the range of the parameter θ is $0 < \theta < 1$. Then the marginal likelihoods of Poisson family distributions are as follows.

$$m_{\text{M-J}}^{\text{ZIGP}} = \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!} \int_0^{\infty} \int_0^{1} \lambda^{n-0.5} \frac{e^{-s\theta}}{(e^{\lambda}-1)^{n-n_0}} \prod_{i=1}^{n} (\lambda+\theta y_i)^{y_i-1} d\theta d\lambda$$

$$m_{\text{M-J}}^{\text{ZIP}} = \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!} \int_0^{\infty} \frac{\lambda^{s-0.5}}{(e^{\lambda}-1)^{n-n_0}} d\lambda$$

$$m_{\text{J}}^{\text{GP}} = \frac{1}{\prod_{i=1}^{n} y_i!} \int_0^{\infty} \int_0^{1} \lambda^{n-0.5} e^{-s\theta-n\lambda} \prod_{i=1}^{n} (\lambda+\theta y_i)^{y_i-1} d\theta d\lambda$$

$$m_{\text{J}}^{\text{Poi}} = \frac{n^{-(s+0.5)}\Gamma(s+0.5)}{\prod_{i=1}^{n} y_i!}$$

These integrals are too complex to simplify to a analytical solution except Poisson distribution. Importance sampling is utilized to approach the value of the integrals via numerical iteration. For the implementation of importance sampling, the quality of the process relies on the choice of importance sampling density. To cover the whole domain of parameters and make it as close to the integrand as possible, the following distributions are selected as the importance sampling densities for the calculation of marginal likelihoods.

The importance sampling densities:

$$\theta \sim U(0,1)$$

$$h(\lambda) = \frac{1}{c}, \quad where\ \lambda \in (0,\infty)$$

$$h^{\text{ZIGP}}(\theta,\lambda) = h(\theta)h(\lambda) = \frac{1}{c}$$

$$h^{\text{ZIP}}(\lambda) = h(\lambda) = \frac{1}{c}$$

These densities have the same range of parameters as the marginal integrals and they are standard distributions which are straightforward to simulate. Because the integrands in the marginal integrals are changing over different data, it is difficult to detect the shape of the integrand so the flat density is suitable to approach the sampling. The importance sampling estimate of one integral $\int f(x)dx$ given the importance density $h(x)$ is

$$\int f(x)dx \approx \frac{1}{m}\sum_{i=1}^{m}\frac{f(x_i)}{h(x_i)} = \frac{c}{m}\sum_{i=1}^{m} f(x_i) \quad \text{for large m}$$

The approximation requires a large number of simulated values of variable X to converge to the target integral by the law of large numbers. The approximation of the marginal likelihoods based on importance sampling densities above are:

$$m_{\text{M-J}}^{\text{ZIGP}} \approx \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\lambda_j^{n-0.5}\frac{e^{-s\theta_j}}{(e^{\lambda_j}-1)^{n-n_0}}\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$$

$$m_{\text{M-J}}^{\text{ZIP}} \approx \frac{n_0!(n-n_0)!}{(n+1)!\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\frac{\lambda_j^{s-0.5}}{(e^{\lambda_j}-1)^{n-n_0}}$$

$$m_{J}^{\text{GP}} \approx \frac{1}{\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\lambda_j^{n-0.5}e^{-s\theta_j-n\lambda_j}\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$$

where the subscript j represents the jth step in the simulation. λ_j and θ_j are the jth simulated values on the jth step. y_i is the ith observation of response variable Y. By applying importance sampling approach, the complex integrals are turned into simple average math problem. The difficulty here is the programming complexity, and there is loop-in-loop calculation caused by the production part $\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$ sitting in the summation on j.

4.7 Simulation results using modified Jeffreys priors

The simulation is implemented in the same way as the distribution selection method using conditionally uniform prior. For the same settings of distribution parameters, 100 dataset are simulated for each defined distribution and the data size is 100. Importance sampling is also applied to help computation of integration of marginal likelihoods with the importance sampling densities $\lambda \sim Uniform(0, 15)$ and $\theta \sim Uniform(0, 1)$. The convergence of the approach reaches target value within 2000 sampling size. Overall results from 100 iterations of each distribution are concluded in terms of mean and variance of posterior probability.

First, Poisson model is studied based on simulation for the posterior probability of comparative models. The simulation results are given in table 4.9. The first column of the table gives the information of ZIGP model parameters, when α and ϕ are set to be zero, it reduces to Poisson distribution. From the outcome. it is seen that the method is able to find the right model, that is Poisson distribution, for even small values of parameter λ. The error variance is quite small which means the estimation of posterior probability is accurate and stable enough.

Generalized Poisson model and zero-inflated Poisson model are studied over simulation analysis using modified Jeffreys priors. The corresponding results are given in table 4.10 and 4.11. Simulation results of zero-inflated generalized Poisson model are listed in the following tables 4.12, 4.13, 4.14. With multiple settings of parameters in each model, the simulation results

Table 4.9: Posterior probability of models using modified Jeffreys prior

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0,0,1)	0.074304	0.212412	0.098991	0.614293	0.09057	0.089097	0.074236	0.139961
(0,0,2)	0.047534	0.219295	0.11377	0.619402	0.066584	0.134178	0.088859	0.160124
(0,0,3)	0.032512	0.183361	0.102157	0.68197	0.084191	0.107151	0.06624	0.139076
(0,0,4)	0.025196	0.135783	0.14177	0.69725	0.030302	0.086485	0.12312	0.161289
(0,0,5)	0.014391	0.089281	0.131848	0.76448	0.021213	0.045868	0.113048	0.1316

True model is Poisson distribution determined by λ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

give very reasonable probability of models. The probability of choosing the true model is increasing when the model parameter becomes larger. For generalized Poisson model, posterior probability works well even for small values of α, for example, posterior probability of GP is largest at (α=0.2,ϕ=0,λ=1). As for zero-inflated Poisson model, our method is in favor of simple Poisson model when parameters are close to 0, for instance, $P(Poi|y) = 0.539175$ when (α=0,ϕ=0.1,λ=1), but it works well on ZIP model, when $\lambda > 1$. For zero-inflate generalized Poisson model, the outcome displays our method can find the true model except small λ. These three models, ZIGP, ZIP and GP model, are close to each other when $\lambda = 1$. GP model is a good choice when zero-inflated parameter $\phi \leq 0.1$ and $\lambda \leq 2$.

The values of posterior probabilities of distributions using modified Jeffreys prior are similar to the results for conditionally uniform prior. For small value of λ, the method choose the simpler distribution if competing distributions are similar (distribution parameters are close to zero), otherwise posterior probability select the right one from which the tested data is simulated. For large

Table 4.10: Posterior probability of models using modified Jeffreys prior

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.2,0,1)	0.24363	0.231976	0.392459	0.131934	0.159125	0.220824	0.183115	0.179561
(0.3,0,1)	0.310726	0.083862	0.594553	0.010859	0.128795	0.145802	0.164995	0.042365
(0.5,0,1)	0.334554	0.000162	0.665283	2.29E-08	0.117391	0.000963	0.117767	1.78E-07
(0.8,0,1)	0.342977	3.90E-35	0.657023	3.45E-58	0.159974	3.88E-34	0.159974	3.45E-57
(0.1,0,2)	0.185847	0.140972	0.563272	0.109909	0.104477	0.17148	0.209871	0.150335
(0.2,0,2)	0.270507	0.000582	0.728904	6.76E-06	0.147847	0.003682	0.148584	5.59E-05
(0.3,0,2)	0.279067	5.62E-15	0.720933	1.99E-20	0.16129	5.62E-14	0.16129	1.99E-19
(0.4,0,2)	0.259647	3.82E-64	0.740353	1.60E-98	0.112204	3.82E-63	0.112204	1.60E-97
(0.06,0,3)	0.1514	0.098084	0.571349	0.179167	0.120192	0.157844	0.266485	0.224591
(0.1,0,3)	0.194434	0.007295	0.790961	0.00731	0.120488	0.027137	0.127048	0.026889
(0.2,0,3)	0.220339	1.11E-20	0.779661	2.76E-33	0.127252	1.11E-19	0.127252	2.66E-32
(0.3,0,3)	0.230834	0	0.769166	0	0.171585	0	0.171585	0
(0.05,0,4)	0.121616	0.052008	0.687535	0.138841	0.093524	0.116751	0.243192	0.214385
(0.1,0,4)	0.157729	2.77E-06	0.842208	6.01E-05	0.096987	2.45E-05	0.096923	0.000596
(0.15,0,4)	0.168621	8.63E-27	0.831379	5.23E-28	0.154066	7.33E-26	0.154066	5.23E-27
(0.2,0,4)	0.146708	1.87E-147	0.853292	7.00E-154	0.079378	1.87E-146	0.079378	7.00E-153
(0.04,0,5)	0.080011	0.029462	0.741094	0.149433	0.083617	0.06209	0.239376	0.212773
(0.08,0,5)	0.09386	2.97E-05	0.905354	0.000756	0.064901	0.000291	0.064679	0.00748
(0.1,0,5)	0.11488	4.31E-13	0.88512	1.02E-11	0.104944	4.29E-12	0.104944	1.02E-10
(0.15,0,5)	0.108208	5.35E-112	0.891792	2.53E-107	0.08292	5.35E-111	0.08292	2.53E-106

True model is generalized Poisson distribution determined by λ and α. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.11: Posterior probability of models using modified Jeffreys prior

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0,0.1,1)	0.082496	0.221708	0.156622	0.539175	0.080745	0.101607	0.114373	0.194785
(0,0.2,1)	0.102064	0.325431	0.197164	0.375341	0.069117	0.191402	0.119799	0.234203
(0,0.5,1)	0.137178	0.538709	0.218351	0.105762	0.057369	0.205473	0.138573	0.160393
(0,0.8,1)	0.149464	0.499155	0.279496	0.071885	0.057069	0.196467	0.162555	0.134855
(0,0.1,2)	0.076089	0.414904	0.179442	0.329565	0.06386	0.227757	0.148923	0.245086
(0,0.2,2)	0.122289	0.608157	0.174845	0.094709	0.07115	0.237728	0.169771	0.145439
(0,0.5,2)	0.156028	0.814449	0.029499	2.45E-05	0.101609	0.160343	0.068408	0.000122
(0,0.8,2)	0.191275	0.753252	0.05547	3.68E-06	0.086495	0.148051	0.076973	2.47E-05
(0,0.1,3)	0.087236	0.651268	0.100689	0.160806	0.059413	0.25838	0.126073	0.220567
(0,0.2,3)	0.12501	0.852452	0.021456	0.001082	0.082606	0.122228	0.061823	0.003096
(0,0.5,3)	0.146184	0.85307	0.000746	1.30E-12	0.11239	0.11519	0.005664	1.10E-11
(0,0.8,3)	0.201157	0.777472	0.021371	3.06E-12	0.127884	0.17921	0.076154	3.03E-11
(0,0.1,4)	0.133983	0.78824	0.042028	0.035749	0.11276	0.204547	0.100762	0.110156
(0,0.2,4)	0.159598	0.838757	0.001636	8.85E-06	0.14257	0.146565	0.007181	7.74E-05
(0,0.5,4)	0.131593	0.868405	1.21E-06	5.01E-23	0.102057	0.102061	7.88E-06	3.52E-22
(0,0.8,4)	0.178861	0.820303	0.000837	7.87E-23	0.137927	0.139225	0.003389	6.36E-22
(0,0.1,5)	0.13409	0.849001	0.00489	0.012018	0.133754	0.14697	0.016825	0.060356
(0,0.2,5)	0.14313	0.856869	1.44E-06	7.89E-09	0.123196	0.123197	1.35E-05	6.10E-08
(0,0.5,5)	0.136275	0.863725	1.29E-10	8.93E-33	0.107583	0.107583	7.00E-10	8.55E-32
(0,0.8,5)	0.168056	0.831687	0.000257	2.07E-23	0.126782	0.127202	0.001034	2.07E-22

True model is zero-inflated Poisson distribution determined by λ and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.12: Posterior probability of distributions using modified Jeffreys prior: $\lambda = 1$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.2,0.1,1)	0.247358	0.253994	0.430546	0.068103	0.116364	0.213641	0.181875	0.131791
(0.3,0.1,1)	0.34537	0.132971	0.519845	0.001813	0.116221	0.215886	0.206182	0.008696
(0.4,0.1,1)	0.376597	0.044331	0.578916	0.000156	0.126981	0.129112	0.168538	0.001507
(0.5,0.1,1)	0.390124	0.013471	0.596405	6.27E-10	0.14551	0.074729	0.169641	3.44E-09
(0.6,0.1,1)	0.414193	1.98E-06	0.585805	1.04E-22	0.180502	1.57E-05	0.180506	9.88E-22
(0.7,0.1,1)	0.410015	9.76E-11	0.589985	2.63E-32	0.189258	9.71E-10	0.189258	2.63E-31
(0.8,0.1,1)	0.380722	4.94E-28	0.619278	2.74E-59	0.157058	4.85E-27	0.157058	2.74E-58
(0.2,0.5,1)	0.261838	0.472583	0.264309	0.001271	0.114115	0.292338	0.222823	0.006124
(0.3,0.5,1)	0.355478	0.298704	0.344953	0.000864	0.123204	0.29369	0.23777	0.008608
(0.4,0.5,1)	0.473918	0.133203	0.392877	2.02E-06	0.189188	0.211332	0.246859	2.02E-05
(0.5,0.5,1)	0.509385	0.030806	0.45981	1.17E-13	0.207635	0.109327	0.232238	8.00E-13
(0.6,0.5,1)	0.518404	0.013176	0.468419	5.92E-20	0.224792	0.063724	0.241054	4.45E-19
(0.7,0.5,1)	0.53504	0.000222	0.464737	7.43E-17	0.236807	0.001399	0.236926	7.43E-16
(0.8,0.5,1)	0.540638	1.89E-06	0.45936	3.69E-37	0.24214	1.89E-05	0.242142	3.69E-36
(0.2,0.8,1)	0.21901	0.457961	0.303873	0.019156	0.084148	0.212493	0.187471	0.069381
(0.3,0.8,1)	0.252743	0.46315	0.278976	0.005131	0.11402	0.270776	0.215096	0.036872
(0.4,0.8,1)	0.328385	0.314909	0.354283	0.002423	0.133597	0.269478	0.23923	0.023955
(0.5,0.8,1)	0.396905	0.206113	0.396975	7.73E-06	0.165008	0.252103	0.24879	6.69E-05
(0.6,0.8,1)	0.408287	0.13241	0.459294	8.61E-06	0.202708	0.24076	0.269577	8.40E-05
(0.7,0.8,1)	0.484702	0.052059	0.463239	7.21E-11	0.218266	0.145798	0.245623	6.37E-10
(0.8,0.8,1)	0.503577	0.02432	0.472102	2.64E-14	0.247136	0.112712	0.258991	2.34E-13

True model is zero-inflated generalized Poisson distribution determined by λ, α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.13: Posterior probability of distributions using modified Jeffreys prior: $\lambda = 2, 3$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.1,0.1,2)	0.285824	0.230492	0.467207	0.016477	0.143532	0.247498	0.247491	0.063551
(0.2,0.1,2)	0.440179	0.00199	0.557831	1.68E-09	0.241001	0.016277	0.244548	9.41E-09
(0.4,0.1,2)	0.431407	3.42E-59	0.568593	4.19E-122	0.256501	3.42E-58	0.256501	4.19E-121
(0.1,0.5,2)	0.445116	0.483716	0.071168	7.44E-10	0.26105	0.327863	0.122651	6.84E-09
(0.2,0.5,2)	0.85074	0.051155	0.098105	8.92E-21	0.185317	0.127039	0.166585	8.81E-20
(0.4,0.5,2)	0.894757	3.97E-18	0.105243	1.76E-88	0.160103	3.97E-17	0.160103	1.76E-87
(0.1,0.8,2)	0.314743	0.543415	0.141841	2.63E-07	0.154399	0.286071	0.185474	2.63E-06
(0.2,0.8,2)	0.547238	0.308563	0.144198	2.84E-12	0.258646	0.323091	0.173699	2.67E-11
(0.4,0.8,2)	0.828534	0.002136	0.16933	1.60E-43	0.191413	0.011665	0.192691	1.60E-42
(0.1,0.1,3)	0.664637	0.044996	0.290366	4.61E-07	0.247879	0.118599	0.266179	2.81E-06
(0.2,0.1,3)	0.659074	8.38E-16	0.340926	2.62E-35	0.278333	8.38E-15	0.278333	2.62E-34
(0.3,0.1,3)	0.722038	0	0.277962	0	0.286159	0	0.286159	0
(0.1,0.5,3)	0.807374	0.191422	0.001204	5.63E-24	0.273803	0.274516	0.007964	5.45E-23
(0.2,0.5,3)	0.998965	4.27E-09	0.001035	3.84E-71	0.004539	3.64E-08	0.004539	3.84E-70
(0.3,0.5,3)	0.998589	1.04E-126	0.001411	0	0.005623	1.04E-125	0.005623	0
(0.1,0.8,3)	0.553642	0.421991	0.024367	7.48E-16	0.293929	0.311036	0.072372	7.03E-15
(0.2,0.8,3)	0.919619	0.045671	0.03471	3.51E-35	0.153856	0.129087	0.10046	3.51E-34
(0.3,0.8,3)	0.960863	9.63E-07	0.039136	7.63E-63	0.092282	9.63E-06	0.092282	7.63E-62

True model is zero-inflated generalized Poisson distribution determined by λ, α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.14: Posterior probability of distributions using modified Jeffreys prior: $\lambda = 4, 5$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.05,0.1,4)	0.660103	0.264998	0.074258	0.000641	0.270329	0.276982	0.15246	0.0049
(0.1,0.1,4)	0.924933	0.000976	0.074091	3.75E-15	0.152623	0.006608	0.152957	2.38E-14
(0.2,0.1,4)	0.925346	0	0.074654	0	0.170969	0	0.170969	0
(0.05,0.5,4)	0.622405	0.377594	1.31E-06	4.97E-28	0.310464	0.310466	5.05E-06	4.97E-27
(0.1,0.5,4)	0.978441	0.021557	2.21E-06	3.14E-56	0.093896	0.093896	1.24E-05	2.23E-55
(0.2,0.5,4)	0.99999	1.11E-33	9.81E-06	1.38E-209	6.46E-05	1.11E-32	6.46E-05	0
(0.05,0.8,4)	0.335219	0.663664	0.001117	3.85E-27	0.231041	0.233354	0.007431	3.12E-26
(0.1,0.8,4)	0.837603	0.155983	0.006413	6.34E-45	0.21197	0.214179	0.031655	4.11E-44
(0.2,0.8,4)	0.998229	2.53E-08	0.001771	1.38E-107	0.006535	1.68E-07	0.006535	1.37E-106
(0.05,0.1,5)	0.905786	0.070998	0.023216	2.65E-08	0.183761	0.168139	0.09362	2.25E-07
(0.1,0.1,5)	0.977959	3.80E-12	0.022041	9.54E-21	0.093249	3.73E-11	0.093249	9.54E-20
(0.15,0.1,5)	0.968101	1.39E-104	0.031899	1.91E-148	0.129963	1.39E-103	0.129963	1.91E-147
(0.05,0.5,5)	0.732945	0.267055	1.60E-09	8.92E-55	0.291737	0.291737	1.01E-08	7.81E-54
(0.1,0.5,5)	0.999998	1.96E-06	1.79E-10	1.32E-98	8.78E-06	8.78E-06	5.51E-10	1.32E-97
(0.15,0.5,5)	1	7.24E-54	8.33E-10	0	3.84E-09	7.24E-53	3.90E-09	0
(0.05,0.8,5)	0.493888	0.506068	4.41E-05	2.21E-31	0.306366	0.306406	0.00014	2.21E-30
(0.1,0.8,5)	0.93065	0.069195	0.000155	1.71E-65	0.161849	0.16191	0.001194	1.67E-64
(0.15,0.8,5)	0.999883	6.32E-07	0.000117	2.00E-156	0.000457	6.18E-06	0.000457	1.43E-155

True model is zero-inflated generalized Poisson distribution determined by λ, α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

value of λ, posterior probability is still sensitive to small value of distribution parameters. Conditionally uniform prior and modified Jeffreys prior both work well, posterior probabilities do not differ much based on simulation study.

4.8 Model selection using approximate Jeffreys priors

The approximate Jeffreys priors (A-J priors) are used in posterior probability comparison for Poisson-related model selection. Still, the Jeffreys prior containing parameter α in GP distribution is improper. The arbitrary normalizing constant on parameter α given λ cause the indeterminacy trouble in Bayes factor. Thus, the fractional Bayes factor method is applied in order to train the Jeffreys prior into proper prior with partial data.

The posterior probability of model M_i which is approached by fractional Bayes factor is

$$\begin{aligned} P(M_i|y) &= \frac{m_i}{\sum_{i=1}^{4} m_i} = \frac{BF_{i,Poi}}{\sum_{i=1}^{4} BF_{i,Poi}} \\ &\approx \frac{\frac{m^i}{m_b^i}}{\frac{m^{\text{Poi}}}{m_b^{\text{Poi}}} + \frac{m^{\text{GP}}}{m_b^{\text{GP}}} + \frac{m^{\text{ZIP}}}{m_b^{\text{ZIP}}} + \frac{m^{\text{ZIGP}}}{m_b^{\text{ZIGP}}}} \end{aligned}$$

where i represents one of comparative distributions, $i \in \{\text{ZIGP,ZIP,GP,Poi}\}$. Bayes factors of posterior probability are approached by fractional Bayes factors.

Marginal likelihood is defined as

$$m_i = P(y|M_i) = \int P(y|M_i, \theta)P(\theta)d\theta$$

where θ is the vector of parameters in model M_i.

Due to the complexity of integration in computation of marginal likelihood, a transformation of parameter α is made by letting $\theta = \alpha\lambda$. Then the range of the parameter θ is $0 < \theta < 1$. Then the marginal likelihoods of Poisson family distributions are as follows.

$$m_{\text{A-J}}^{\text{ZIGP}} = \frac{\Gamma(n_0 + 0.5)\Gamma(n - n_0 + 1.5)}{\Gamma(n+2)\prod_{i=1}^{n} y_i!} \int_0^{\infty}\int_0^{1} \lambda^{n-0.5} \frac{e^{-s\theta}}{(e^{\lambda}-1)^{n-n_0}} \prod_{i=1}^{n} (\lambda + \theta y_i)^{y_i - 1} d\theta d\lambda$$

$$m_{\text{A-J}}^{\text{ZIP}} = \frac{\Gamma(n_0 + 0.5)\Gamma(n - n_0 + 1)}{\Gamma(n+1.5)\prod_{i=1}^{n} y_i!} \int_0^{\infty} \frac{\lambda^{s-0.5}}{(e^{\lambda}-1)^{n-n_0}} d\lambda$$

$$m_{\text{J}}^{\text{GP}} = \frac{1}{\prod_{i=1}^{n} y_i!} \int_0^{\infty}\int_0^{1} \lambda^{n-0.5} e^{-s\theta - n\lambda} \prod_{i=1}^{n} (\lambda + \theta y_i)^{y_i - 1} d\theta d\lambda$$

$$m_{\text{J}}^{\text{Poi}} = \frac{n^{-(s+0.5)}\Gamma(s+0.5)}{\prod_{i=1}^{n} y_i!}$$

Importance sampling is utilized to approach the value of the integrals via numerical iteration. Since the integrands of marginal likelihoods are not changed, the same importance sampling densities are used for these approximate Jeffreys priors.

The importance sampling densities:

$$\theta \sim U(0,1)$$
$$h(\lambda) = \frac{1}{c}, \quad where\ \lambda \in (0,\infty)$$
$$h^{\text{ZIGP}}(\theta,\lambda) = h(\theta)h(\lambda) = \frac{1}{c}$$
$$h^{\text{ZIP}}(\lambda) = h(\lambda) = \frac{1}{c}$$

The approximation requires a large number of simulated values of variable X to converge to the target integral by the law of large numbers. The approximation of the marginal likelihoods based on importance sampling densities above are:

$$m_{\text{A-J}}^{\text{ZIGP}} \approx \frac{\Gamma(n_0+0.5)\Gamma(n-n_0+1.5)}{\Gamma(n+2)\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\lambda_j^{n-0.5}\frac{e^{-s\theta_j}}{(e^{\lambda_j}-1)^{n-n_0}}\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$$
$$m_{\text{A-J}}^{\text{ZIP}} \approx \frac{\Gamma(n_0+0.5)\Gamma(n-n_0+1)}{\Gamma(n+1.5)\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\frac{\lambda_j^{s-0.5}}{(e^{\lambda_j}-1)^{n-n_0}}$$
$$m_{\text{J}}^{\text{GP}} \approx \frac{1}{\prod_{i=1}^{n} y_i!}\frac{c}{m}\sum_{j=1}^{m}\lambda_j^{n-0.5}e^{-s\theta_j-n\lambda_j}\prod_{i=1}^{n}(\lambda_j+\theta_j y_i)^{y_i-1}$$

where the subscript j represents the jth step in the simulation. λ_j and θ_j are the jth simulated values on the jth step. y_i is the ith observation of response variable Y.

4.9 Simulation results using approximate Jeffreys priors

The simulation is implemented in the same way as the distribution selection method using conditionally uniform priors and modified Jeffreys priors. For the same settings of distribution parameters, 100 dataset are simulated for each defined distribution and the data size is 100. Importance sampling is also applied to help computation of integration of marginal likelihoods with the importance sampling densities $\lambda \sim Uniform(0, 15)$ and $\theta \sim Uniform(0, 1)$. The convergence of the approach reaches target value within 2000 sampling size. Overall results from 100 iterations of each distribution are concluded in terms of mean and variance of posterior probability.

Table 4.15: Posterior probability of models using approximate Jeffreys prior

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0,0,1)	0.064398	0.206482	0.089659	0.639461	0.070849	0.087594	0.05933	0.1218
(0,0,2)	0.053969	0.2132	0.108763	0.624068	0.065363	0.10899	0.078755	0.145865
(0,0,3)	0.043927	0.26128	0.108749	0.586045	0.038803	0.142831	0.087767	0.161186
(0,0,4)	0.036825	0.254936	0.100155	0.608083	0.038364	0.13215	0.0778	0.150839
(0,0,5)	0.039392	0.229348	0.124529	0.606731	0.036021	0.075503	0.132268	0.135229

True model is Poisson distribution determined by λ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Generalized Poisson model and zero-inflated Poisson model are studied over simulation analysis using approximate Jeffreys priors. The corresponding results are given in table 4.16 and 4.17. Simulation results of zero-inflated generalized Poisson model are listed in the following tables 4.18, 4.19, 4.20. The probability of choosing the true model is increasing when the value of the

model parameter increases. The values of posterior probabilities of distributions using approximate Jeffreys prior are similar to the results of conditionally uniform prior and modified Jeffreys prior. For small value of λ, the method choose the simpler distribution if competing distributions are similar (distribution parameters are close to zero), otherwise posterior probability select the right one from which the tested data is simulated. For large value of λ, posterior probability is still sensitive to small value of distribution parameters. These three types of priors perform equivalently well based on simulation study and they are all good choices for Poisson-related model selection.

Table 4.16: Posterior probability of models using approximate Jeffreys prior

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.2,0,1)	0.217422	0.179535	0.476598	0.126444	0.104091	0.182529	0.199207	0.186184
(0.3,0,1)	0.276309	0.144817	0.565116	0.013758	0.120146	0.228058	0.215167	0.046109
(0.5,0,1)	0.350152	0.000439	0.649409	2.23E-09	0.151395	0.002716	0.1514	1.70E-08
(0.8,0,1)	0.345976	7.13E-39	0.654024	5.70E-78	0.150612	7.13E-38	0.150612	5.70E-77
(0.1,0,2)	0.21428	0.122555	0.525202	0.137963	0.141479	0.183483	0.243633	0.204064
(0.2,0,2)	0.298874	0.00113	0.699992	4.30E-06	0.151302	0.006626	0.152078	3.77E-05
(0.3,0,2)	0.298279	3.65E-18	0.701721	5.86E-26	0.133984	3.56E-17	0.133984	5.06E-25
(0.4,0,2)	0.307104	1.58E-53	0.692896	7.98E-71	0.15467	1.58E-52	0.15467	7.98E-70
(0.06,0,3)	0.221663	0.10053	0.524778	0.153029	0.170395	0.12651	0.214751	0.180977
(0.1,0,3)	0.265592	0.009953	0.718178	0.006276	0.142643	0.050792	0.157795	0.02916
(0.2,0,3)	0.287448	8.53E-25	0.712552	1.34E-30	0.145749	8.39E-24	0.145749	9.79E-30
(0.3,0,3)	0.284935	0	0.715065	0	0.144897	0	0.144897	0
(0.05,0,4)	0.234818	0.045781	0.633489	0.085912	0.114843	0.066996	0.173281	0.144747
(0.1,0,4)	0.264855	1.66E-06	0.735143	8.37E-07	0.112269	1.45E-05	0.112268	7.58E-06
(0.15,0,4)	0.315986	2.63E-27	0.684014	9.87E-33	0.181414	2.62E-26	0.181414	9.86E-32
(0.2,0,4)	0.314677	0	0.685323	0	0.148645	0	0.148645	0
(0.04,0,5)	0.21432	0.045398	0.618663	0.121619	0.093688	0.063714	0.173138	0.167225
(0.08,0,5)	0.281518	8.67E-07	0.718481	4.45E-07	0.112542	7.95E-06	0.112541	3.22E-06
(0.1,0,5)	0.262607	2.04E-15	0.737393	7.55E-15	0.078526	1.43E-14	0.078526	6.67E-14
(0.15,0,5)	0.287767	1.50E-107	0.712233	9.51E-112	0.122184	1.50E-106	0.122184	9.51E-111

True model is generalized Poisson distribution determined by λ and α. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.17: Posterior probability of models using approximate Jeffreys prior

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0,0.1,1)	0.07514	0.248084	0.144089	0.532687	0.076888	0.150561	0.0897	0.197399
(0,0.2,1)	0.078481	0.298156	0.197617	0.425746	0.045394	0.162682	0.116361	0.213801
(0,0.5,1)	0.132795	0.530745	0.245054	0.091407	0.046931	0.200599	0.145372	0.148268
(0,0.8,1)	0.104082	0.528168	0.314177	0.053573	0.033782	0.197253	0.174184	0.130492
(0,0.1,2)	0.071851	0.416015	0.171186	0.340949	0.051903	0.239963	0.134931	0.239739
(0,0.2,2)	0.10797	0.646121	0.135759	0.11015	0.062382	0.239324	0.152266	0.164137
(0,0.5,2)	0.162994	0.790013	0.046978	1.50E-05	0.109035	0.19203	0.104561	8.79E-05
(0,0.8,2)	0.158494	0.762348	0.079158	3.87E-07	0.061395	0.156035	0.113594	1.67E-06
(0,0.1,3)	0.110426	0.657009	0.105574	0.12699	0.077885	0.234204	0.12418	0.195613
(0,0.2,3)	0.132989	0.845355	0.020814	0.000842	0.08918	0.148006	0.074233	0.005386
(0,0.5,3)	0.154303	0.845368	0.000329	4.06E-13	0.133119	0.133916	0.001407	3.69E-12
(0,0.8,3)	0.179152	0.783857	0.036991	3.99E-09	0.131352	0.194374	0.102432	3.95E-08
(0,0.1,4)	0.130112	0.810115	0.025575	0.034198	0.088554	0.150922	0.053192	0.092689
(0,0.2,4)	0.149374	0.850399	0.000226	8.90E-07	0.133352	0.133753	0.000772	4.32E-06
(0,0.5,4)	0.150614	0.849383	2.71E-06	6.21E-23	0.11748	0.117489	1.59E-05	5.00E-22
(0,0.8,4)	0.160643	0.837181	0.002176	8.45E-19	0.133667	0.137421	0.007676	8.45E-18
(0,0.1,5)	0.117311	0.869918	0.009092	0.003679	0.06972	0.1106	0.051478	0.029998
(0,0.2,5)	0.117447	0.882552	1.43E-06	4.06E-09	0.074566	0.074569	9.31E-06	3.67E-08
(0,0.5,5)	0.138213	0.861787	1.85E-09	1.05E-30	0.133627	0.133627	1.07E-08	1.04E-29
(0,0.8,5)	0.123355	0.876476	0.000169	7.32E-23	0.091941	0.092319	0.000886	7.32E-22

True model is zero-inflated Poisson distribution determined by λ and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.18: Posterior probability of distributions using approximate Jeffreys prior: $\lambda = 1$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.2,0.1,1)	0.224204	0.278074	0.422852	0.07487	0.108803	0.247579	0.213377	0.148125
(0.3,0.1,1)	0.305201	0.140731	0.547073	0.006995	0.107334	0.219081	0.203565	0.034205
(0.4,0.1,1)	0.369822	0.03144	0.598674	6.41E-05	0.156884	0.107344	0.195013	0.000405
(0.5,0.1,1)	0.369777	0.002767	0.627456	5.32E-10	0.143488	0.018651	0.148852	5.27E-09
(0.6,0.1,1)	0.348338	6.60E-05	0.651596	7.04E-15	0.143009	0.000656	0.143112	7.04E-14
(0.7,0.1,1)	0.337288	1.26E-14	0.662712	4.85E-27	0.161605	1.06E-13	0.161605	4.85E-26
(0.8,0.1,1)	0.346317	2.38E-29	0.653683	1.90E-72	0.151377	2.37E-28	0.151377	1.90E-71
(0.2,0.5,1)	0.260786	0.372321	0.365323	0.00157	0.0997	0.297266	0.249909	0.010012
(0.3,0.5,1)	0.34498	0.287154	0.367828	3.85E-05	0.144693	0.280781	0.247322	0.000353
(0.4,0.5,1)	0.416204	0.191622	0.392175	5.18E-09	0.15929	0.254156	0.263456	4.21E-08
(0.5,0.5,1)	0.414869	0.056874	0.528257	2.64E-09	0.206061	0.152379	0.25231	2.28E-08
(0.6,0.5,1)	0.517727	0.004597	0.477676	1.42E-15	0.234523	0.03212	0.239288	1.42E-14
(0.7,0.5,1)	0.496657	0.003172	0.500171	3.41E-25	0.245356	0.023956	0.249986	3.40E-24
(0.8,0.5,1)	0.497904	1.43E-09	0.502096	1.02E-29	0.238946	1.08E-08	0.238946	1.02E-28
(0.2,0.8,1)	0.162497	0.487158	0.344359	0.005986	0.064653	0.247705	0.223073	0.033934
(0.3,0.8,1)	0.216879	0.402664	0.376011	0.004446	0.113708	0.266009	0.248594	0.03357
(0.4,0.8,1)	0.245209	0.345906	0.405076	0.003809	0.113177	0.296739	0.271744	0.037135
(0.5,0.8,1)	0.315577	0.258042	0.42603	0.000352	0.174785	0.297207	0.287742	0.003516
(0.6,0.8,1)	0.332804	0.134217	0.532978	1.46E-06	0.214386	0.239525	0.293904	9.21E-06
(0.7,0.8,1)	0.44016	0.053106	0.506733	3.83E-07	0.246001	0.165514	0.272962	3.83E-06
(0.8,0.8,1)	0.408226	0.027511	0.564262	2.14E-10	0.234854	0.107285	0.253419	2.14E-09

True model is zero-inflated generalized Poisson distribution determined by λ, α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.19: Posterior probability of distributions using approximate Jeffreys prior: $\lambda = 2, 3$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.1,0.1,2)	0.256285	0.199196	0.525681	0.018838	0.112942	0.2596	0.249697	0.057879
(0.2,0.1,2)	0.451382	0.0031	0.545517	4.33E-07	0.221421	0.014831	0.226015	2.97E-06
(0.4,0.1,2)	0.449808	6.39E-71	0.550192	1.40E-121	0.235453	6.39E-70	0.235453	1.40E-120
(0.1,0.5,2)	0.447884	0.446292	0.105824	6.68E-09	0.215789	0.299886	0.156775	5.87E-08
(0.2,0.5,2)	0.868956	0.040503	0.090541	6.46E-22	0.181735	0.135071	0.147477	4.69E-21
(0.4,0.5,2)	0.900457	6.05E-27	0.099543	6.31E-100	0.165449	5.75E-26	0.165449	6.30E-99
(0.1,0.8,2)	0.277164	0.595478	0.127358	2.01E-09	0.155362	0.281429	0.191199	1.63E-08
(0.2,0.8,2)	0.540112	0.266831	0.193057	2.01E-09	0.252244	0.299619	0.210197	2.00E-08
(0.4,0.8,2)	0.735328	0.000674	0.263998	8.19E-43	0.245551	0.006741	0.245643	8.19E-42
(0.1,0.1,3)	0.647574	0.059964	0.292455	6.35E-06	0.248348	0.157055	0.259624	3.35E-05
(0.2,0.1,3)	0.722163	6.55E-13	0.277837	1.11E-27	0.25831	6.55E-12	0.25831	1.11E-26
(0.3,0.1,3)	0.759983	0	0.240017	0	0.241708	0	0.241708	0
(0.1,0.5,3)	0.846294	0.150751	0.002954	1.10E-27	0.229508	0.23083	0.016732	8.90E-27
(0.2,0.5,3)	0.998246	3.19E-11	0.001754	8.74E-73	0.007877	1.88E-10	0.007877	8.74E-72
(0.3,0.5,3)	0.998608	1.13E-99	0.001392	0	0.005403	1.13E-98	0.005403	0
(0.1,0.8,3)	0.539062	0.42673	0.034207	1.54E-20	0.299297	0.328114	0.084565	1.54E-19
(0.2,0.8,3)	0.928764	0.039575	0.031661	3.64E-39	0.145884	0.12472	0.087789	3.63E-38
(0.3,0.8,3)	0.980083	4.16E-14	0.019917	1.25E-124	0.055437	4.15E-13	0.055437	1.25E-123

True model is zero-inflated generalized Poisson distribution determined by λ, α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

Table 4.20: Posterior probability of distributions using approximate Jeffreys prior: $\lambda = 4, 5$

(α, ϕ, λ)	$\bar{P}(ZIGP\|y)$	$\bar{P}(ZIP\|y)$	$\bar{P}(GP\|y)$	$\bar{P}(Poi\|y)$	$SD[P(ZIGP\|y)]$	$SD[P(ZIP\|y)]$	$SD[P(GP\|y)]$	$SD[P(Poi\|y)]$
(0.05,0.1,4)	0.696007	0.248439	0.055544	9.27E-06	0.250273	0.268376	0.121914	3.81E-05
(0.1,0.1,4)	0.928379	0.000149	0.071472	1.82E-15	0.142883	0.000949	0.142956	1.75E-14
(0.2,0.1,4)	0.92089	7.71E-130	0.07911	0	0.152697	7.60E-129	0.152697	0
(0.05,0.5,4)	0.610686	0.389304	9.91E-06	4.73E-35	0.31879	0.318802	8.57E-05	4.35E-34
(0.1,0.5,4)	0.987153	0.012846	1.61E-06	1.42E-53	0.060353	0.060353	8.88E-06	1.37E-52
(0.2,0.5,4)	0.999997	3.12E-57	2.53E-06	0	1.73E-05	3.12E-56	1.73E-05	0
(0.05,0.8,4)	0.34866	0.64413	0.00721	6.94E-21	0.273483	0.285459	0.047446	6.94E-20
(0.1,0.8,4)	0.750207	0.247154	0.002639	9.46E-40	0.300987	0.302772	0.012384	9.46E-39
(0.2,0.8,4)	0.995993	2.16E-09	0.004007	1.99E-124	0.020654	1.63E-08	0.020654	1.99E-123
(0.05,0.1,5)	0.912581	0.072478	0.014937	4.83E-06	0.172928	0.162418	0.072575	4.82E-05
(0.1,0.1,5)	0.996178	1.66E-11	0.003822	2.92E-26	0.014222	1.35E-10	0.014222	2.92E-25
(0.15,0.1,5)	0.985236	3.28E-102	0.014764	1.51E-144	0.073889	4.27E-101	0.073889	1.43E-143
(0.05,0.5,5)	0.727632	0.272368	1.96E-09	2.05E-51	0.287658	0.287658	1.72E-08	2.04E-50
(0.1,0.5,5)	0.999593	0.000407	3.27E-09	2.57E-92	0.004676	0.004676	1.77E-08	3.63E-91
(0.15,0.5,5)	1	3.59E-46	1.32E-09	0	1.11E-08	3.59E-45	1.11E-08	0
(0.05,0.8,5)	0.4415	0.558357	0.000142	6.68E-37	0.312248	0.312412	0.000684	6.68E-36
(0.1,0.8,5)	0.93437	0.065087	0.000544	4.48E-67	0.176442	0.176612	0.003255	6.34E-66
(0.15,0.8,5)	0.999598	6.29E-06	0.000396	6.58E-161	0.00263	6.16E-05	0.00263	6.53E-160

True model is zero-inflated generalized Poisson distribution determined by λ, α and ϕ. Each row contains posterior probabilities based on 100 simulations from a model defined by parameters given in the first column.

4.10 Comparison of conditionally uniform prior, modified Jeffreys prior and approximate Jeffreys prior

The posterior probabilities using conditionally uniform priors with Bayes factor, modified Jeffreys priors with Bayes factor, modified Jeffreys priors with fractional Bayes factor approach, approximate Jeffreys priors with Bayes factor and modified Jeffreys priors with fractional Bayes factor are plotted together to compare the difference among these priors. The comparison of these three types of prior for Poisson model is displayed on figure 4.1. There are multiple plots on the figure, and each is the mean and standard deviation plot for Poisson model of certain λ value which is denoted above each plot. It is easy to see the posterior probability for Poisson model is always very high, thus posterior probability is able to choose Poisson model no matter how small λ is.

The difference among conditionally uniform priors, modified Jeffreys priors and approximate Jeffreys priors for GP model is plotted on figure 4.2. Each row represents a specific number of λ ranged from 1 to 5. Columns are ranked in ascending order of α, and each column has different values of α which is depending on the value λ in each row. The exact values of parameter λ and α are given above plots, while ϕ is always zero for GP model. From the figure, posterior probability prefer GP model as the right model for all values of parameters.

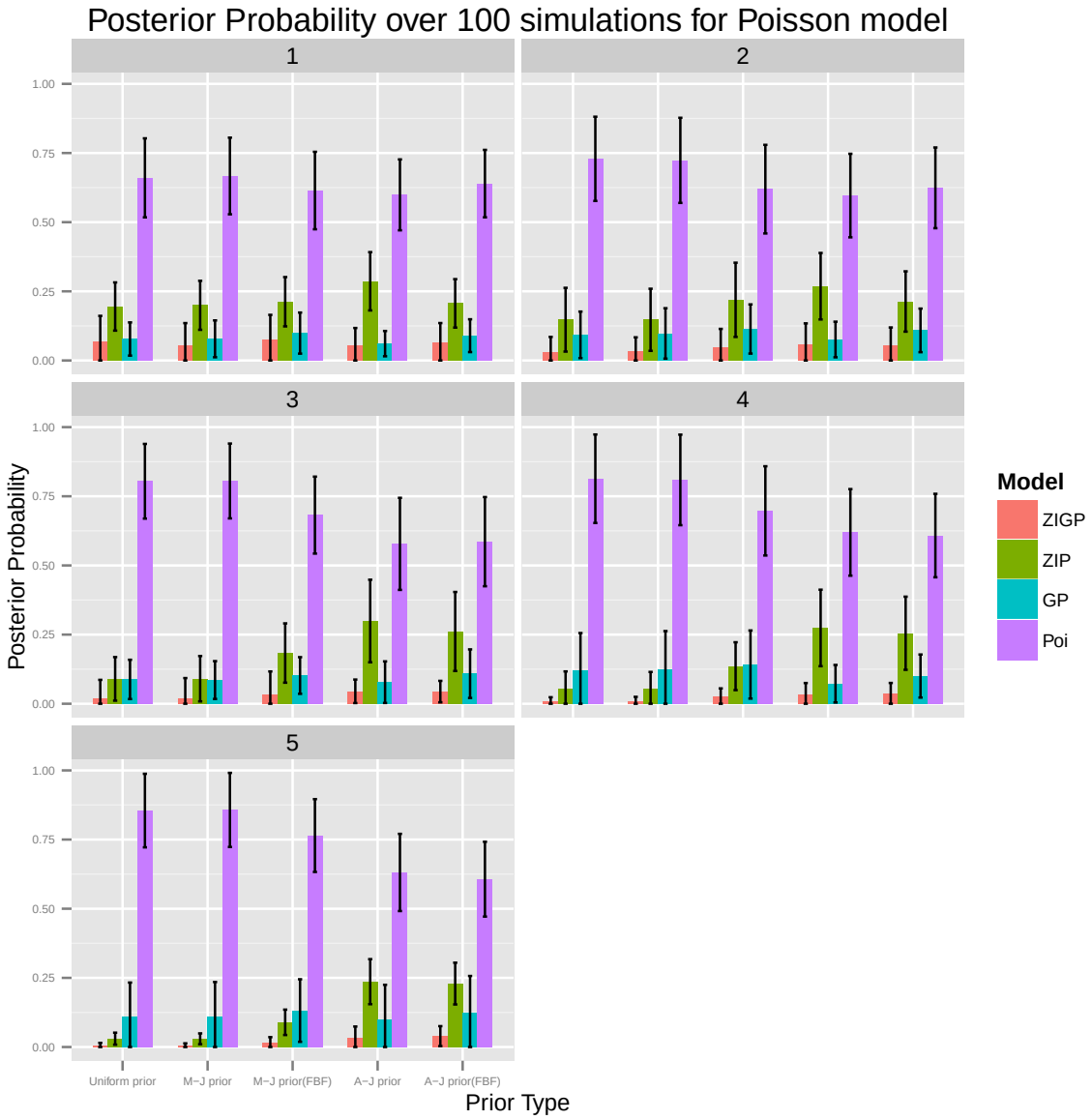
The comparison of these priors for ZIP model is given on figure 4.3. Columns give the value of ϕ, while rows are about the value of λ. Posterior probabil-

ity choose Poisson model for three sets of ZIP parameters, $(\phi = 0.1, \lambda = 1)$, $(\phi = 0.1, \lambda = 2)$ and $(\phi = 0.2, \lambda = 1)$. Except small parameters, posterior probability gives very high probability in support of ZIP model.

For ZIGP model, the results of posterior probability using these priors are provided in regard with the value of λ, the figures are 4.4, 4.5, 4.6, 4.7, 4.8. The columns are about zero-inflated parameter ϕ and rows are the values of over-dispersion parameter α. When $\lambda = 1$, these four comparative models are not outstanding from each other, therefore posterior probability is sensitive to the value of parameters. When $\lambda = 2$, posterior probability tends to support ZIGP model as both values of α and λ parameters increase. For $\lambda \geq 3$, posterior probability chooses ZIGP model except parameters are close to zero.

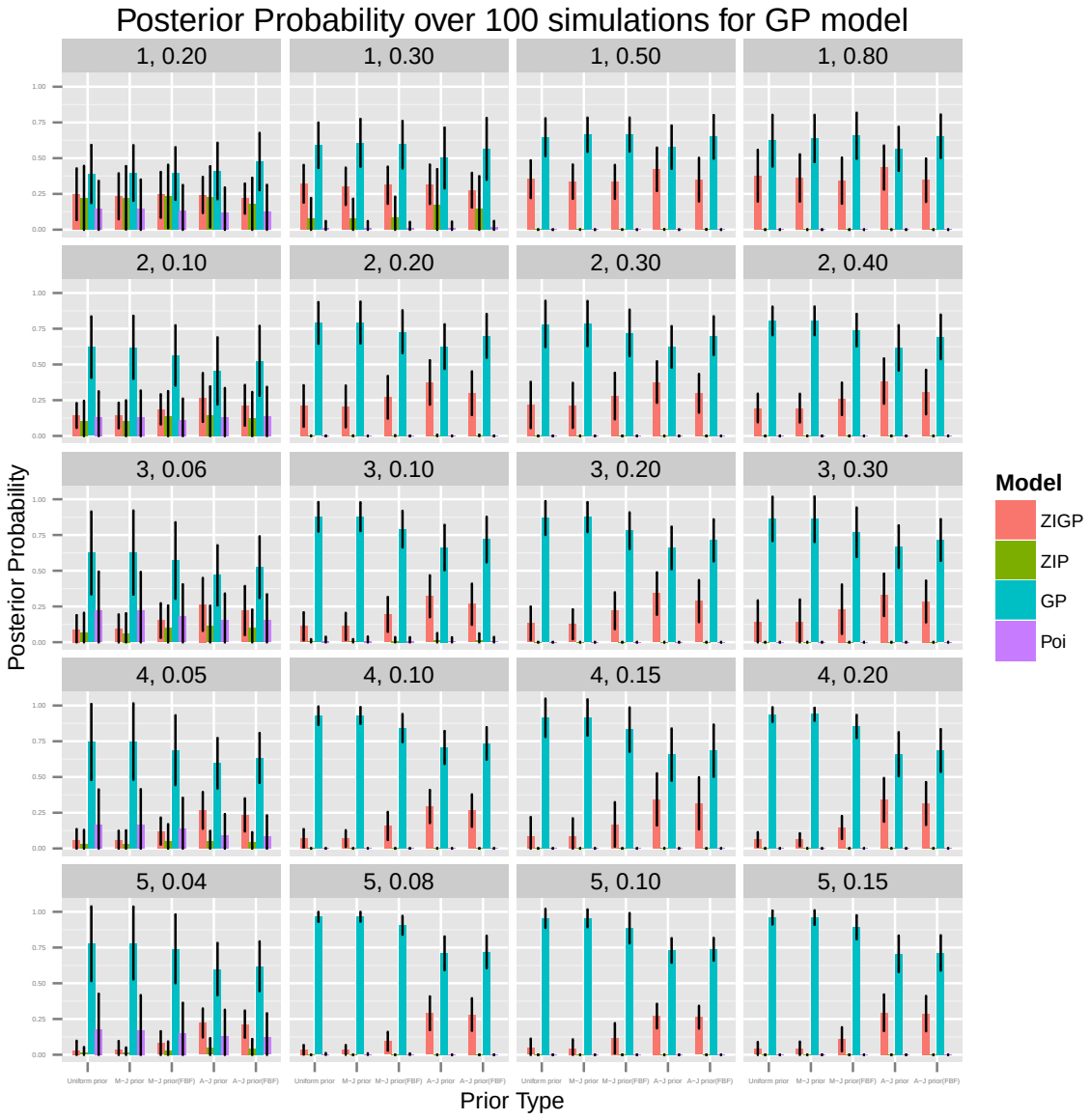
Based on these figures, we would say our method works well on model selection of ZIGP, ZIP, GP and Poisson distributions. It is able to give strong evidence to true model when its parameters are not very small. When the model parameters are too small to be distinguished from others, our method will choose simple model as the better model. Meanwhile, conditionally uniform priors, modified Jeffreys priors and approximate Jeffreys priors are all good choice for comparison of Poisson related distributions, and there is no big difference of selection results among them.

Figure 4.1: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:Poisson



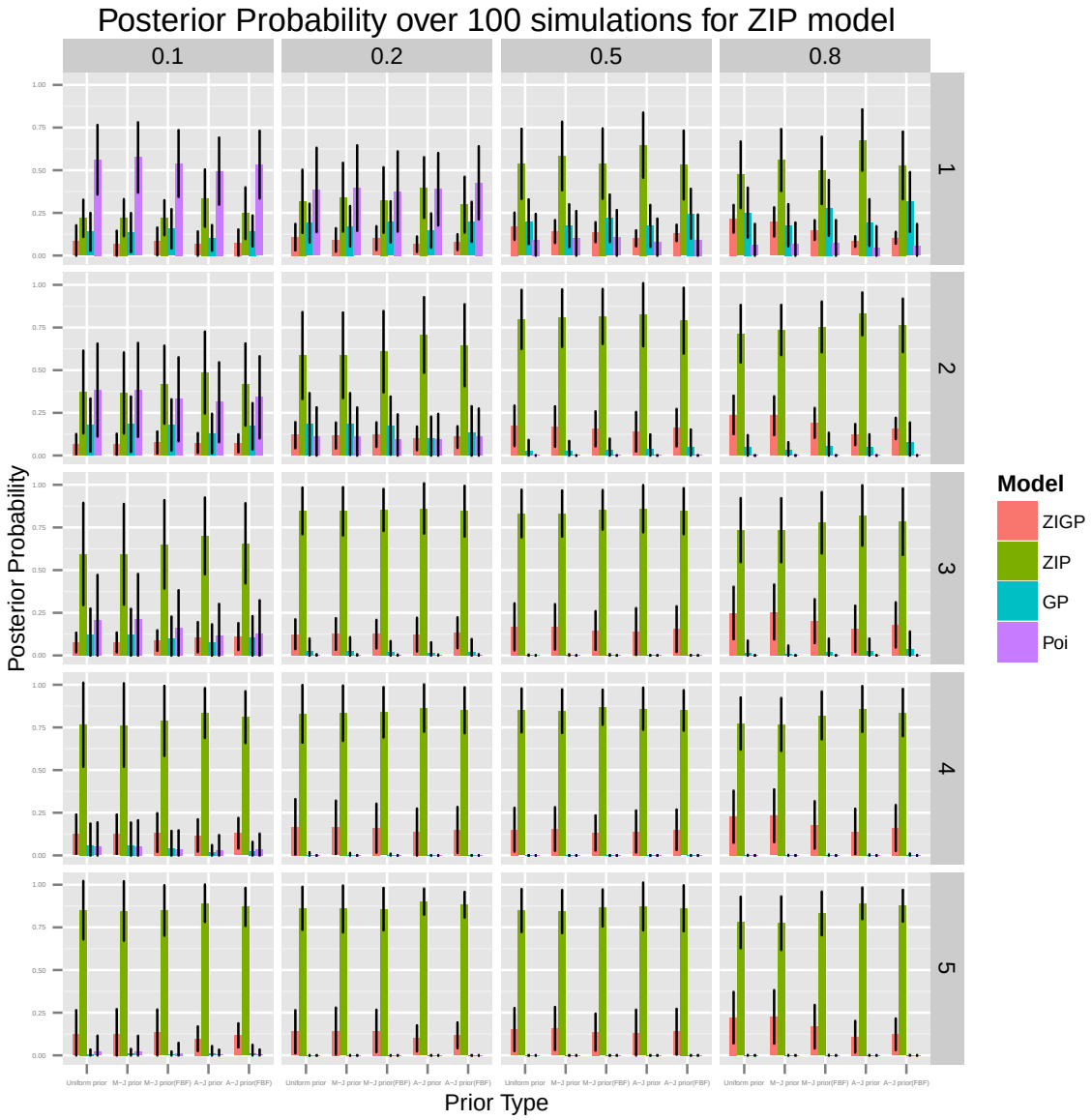
Posterior probabilities based on 100 simulations from a Poisson model defined by parameter λ given on the title of plots. $\lambda = 1, 2, 3, 4, 5$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.2: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:GP



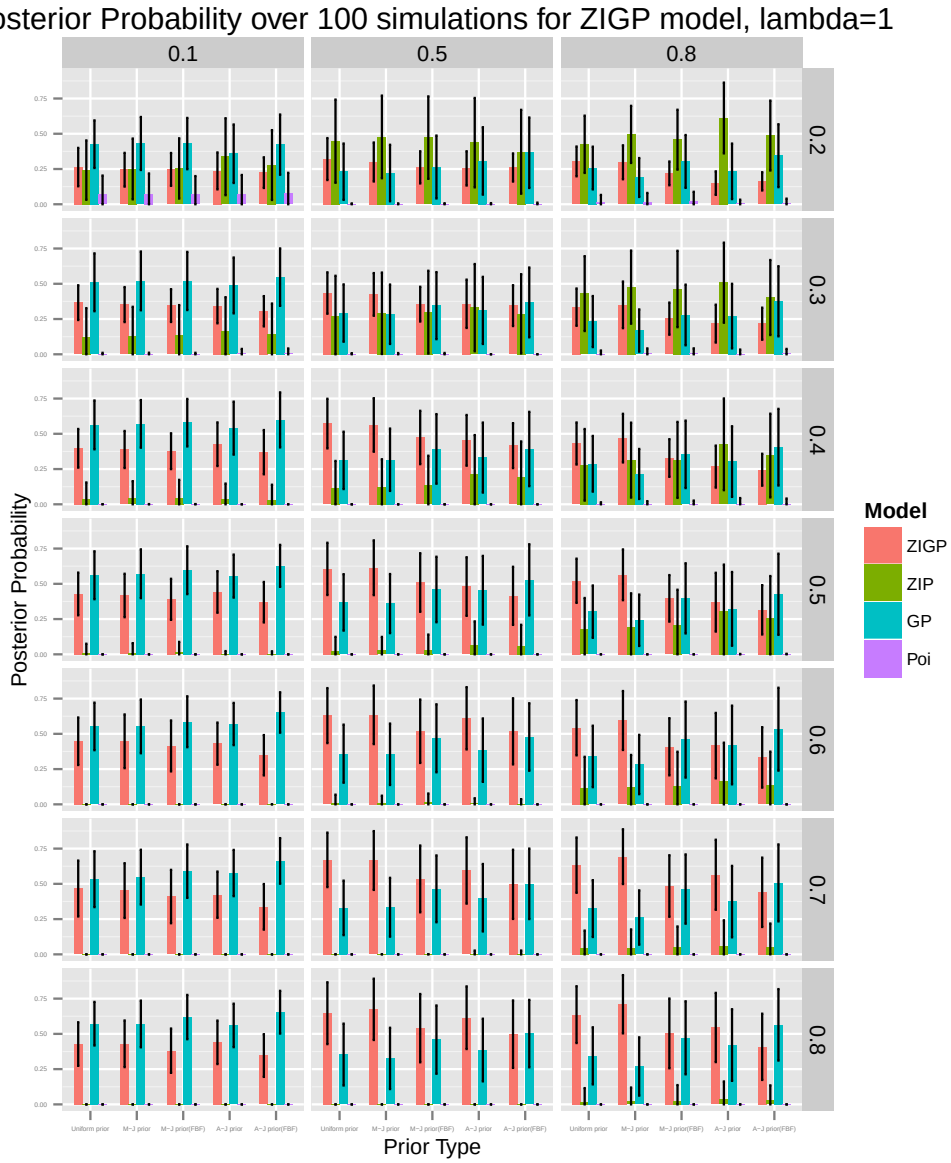
Posterior probabilities based on 100 simulations from a generalized Poisson model defined by parameters λ and α which are given on the title of each plot in order. $\lambda = 1, 2, 3, 4, 5$ and $0 < \alpha < 1/\lambda$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.3: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:ZIP



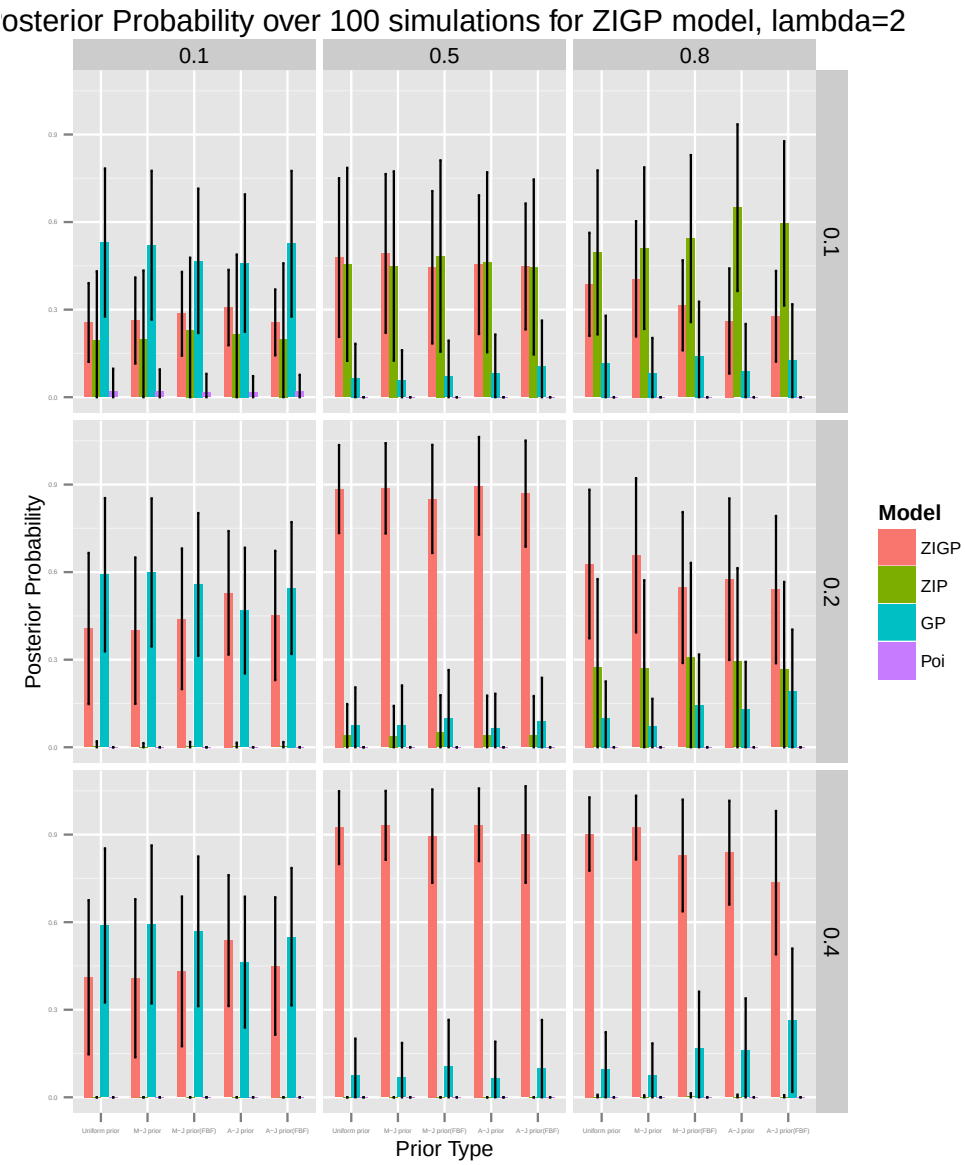
Posterior probabilities based on 100 simulations from a zero-inflated Poisson model defined by parameters λ and ϕ given on the title of rows and columns respectively. $\lambda = 1, 2, 3, 4, 5$ and $\phi = 0.1, 0.2, 0.5, 0.8$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.4: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:ZIGP($\lambda = 1$)



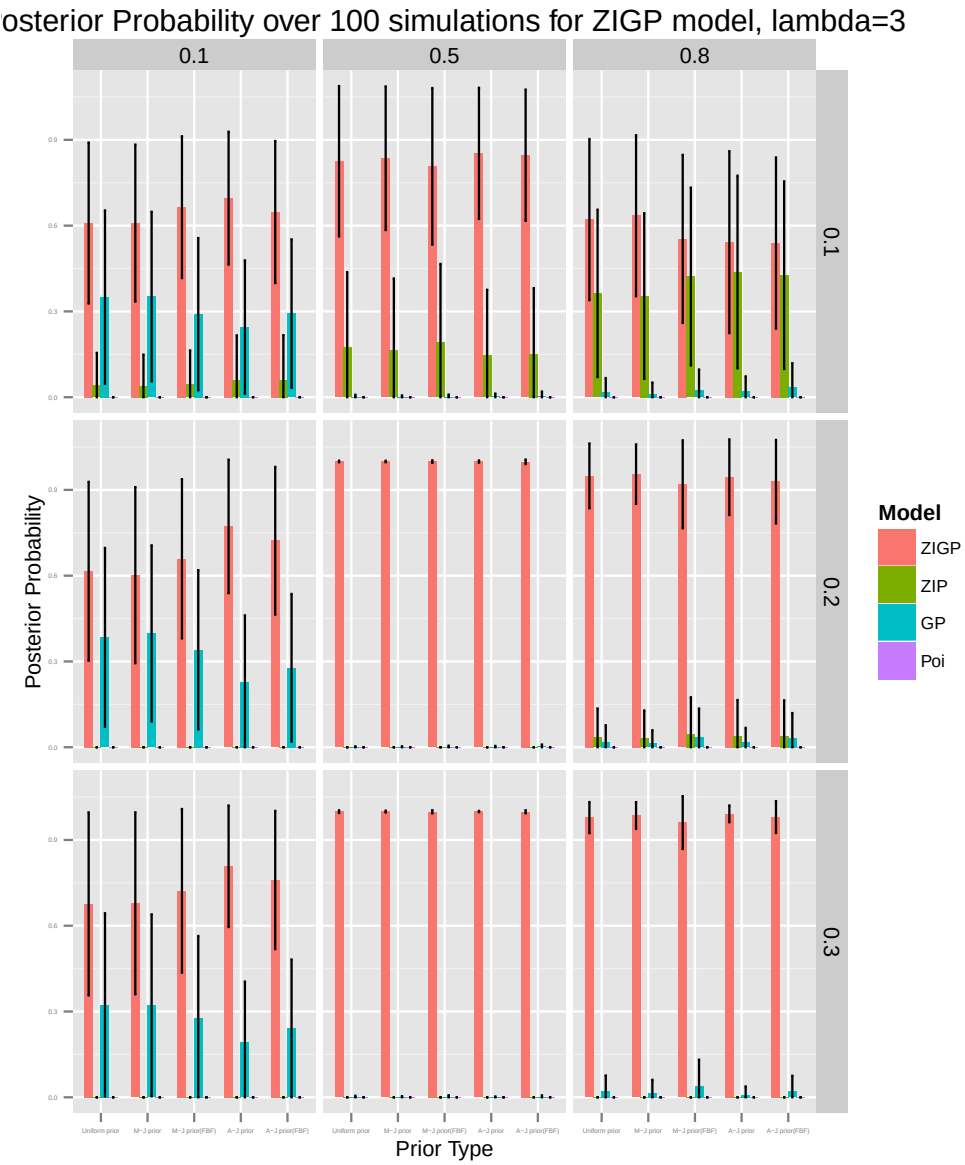
Posterior probabilities based on 100 simulations from a zero-inflated Poisson model defined by parameters $\lambda = 1$, ϕ and α, while ϕ and α are given on the title of columns and rows respectively. $\phi = 0.1, 0.5, 0.8$ and $\alpha = 0.2, 0.3, ..., 0.8$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.5: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:ZIGP($\lambda = 2$)



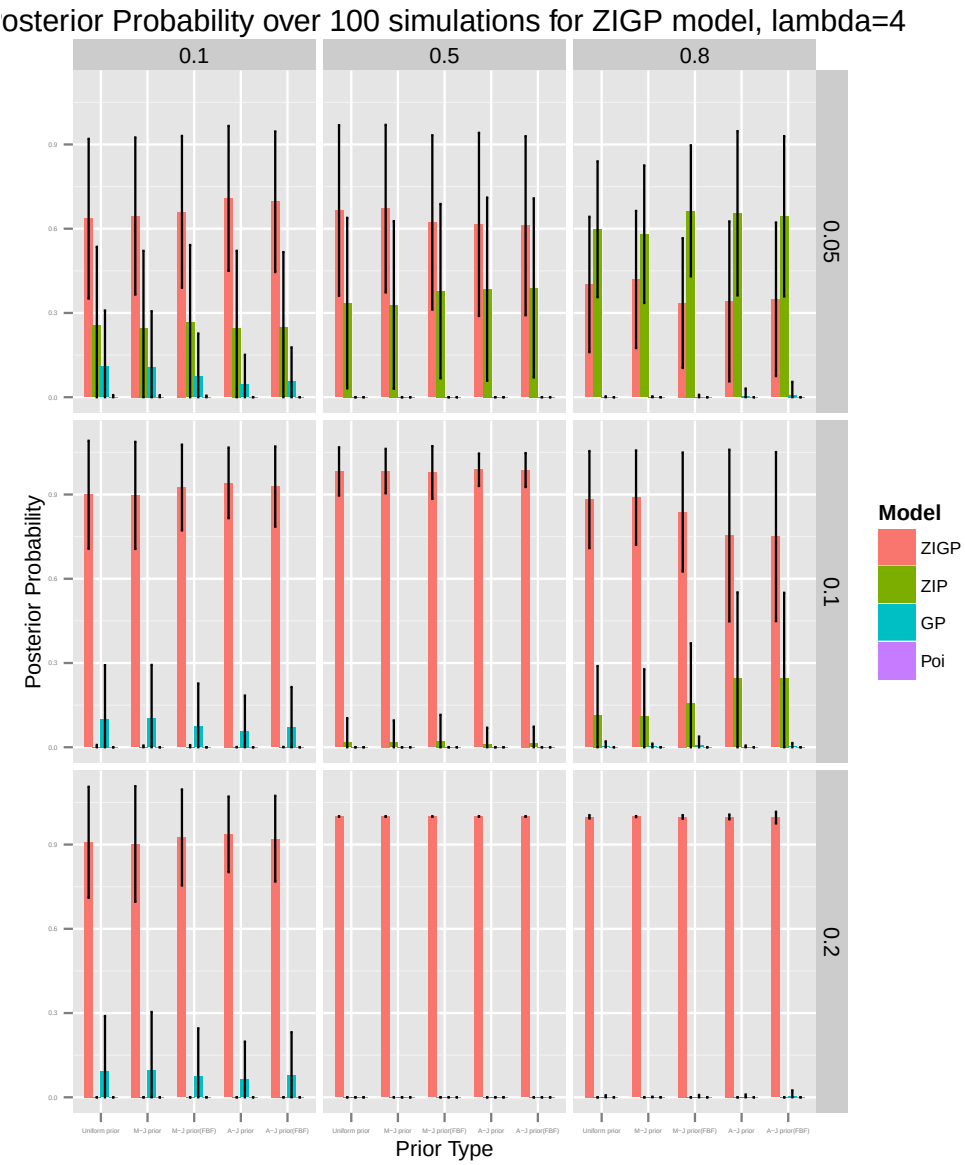
Posterior probabilities based on 100 simulations from a zero-inflated Poisson model defined by parameters $\lambda = 2$, ϕ and α, while ϕ and α are given on the title of columns and rows respectively. $\phi = 0.1, 0.5, 0.8$ and $\alpha = 0.1, 0.2, 0.4$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.6: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:ZIGP($\lambda = 3$)



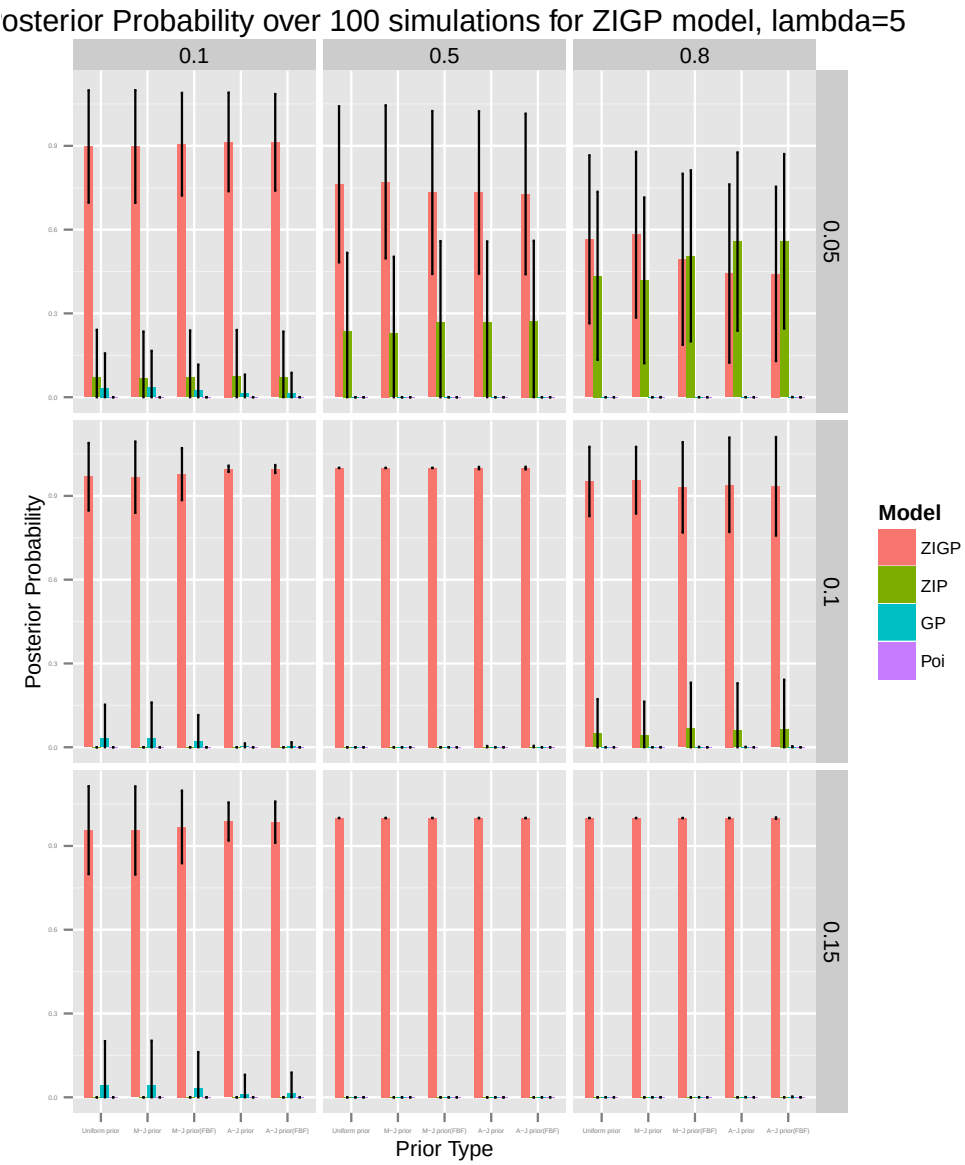
Posterior probabilities based on 100 simulations from a zero-inflated Poisson model defined by parameters $\lambda = 3$, ϕ and α, while ϕ and α are given on the title of columns and rows respectively. $\phi = 0.1, 0.5, 0.8$ and $\alpha = 0.1, 0.2, 0.3$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.7: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:ZIGP($\lambda = 4$)



Posterior probabilities based on 100 simulations from a zero-inflated Poisson model defined by parameters $\lambda = 4$, ϕ and α, while ϕ and α are given on the title of columns and rows respectively. $\phi = 0.1, 0.5, 0.8$ and $\alpha = 0.05, 0.1, 0.2$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

Figure 4.8: Comparison of posterior probability among cond-uniform priors, M-J priors and A-J priors, true model:ZIGP($\lambda = 5$)



Posterior probabilities based on 100 simulations from a zero-inflated Poisson model defined by parameters $\lambda = 5$, ϕ and α, while ϕ and α are given on the title of columns and rows respectively. $\phi = 0.1, 0.5, 0.8$ and $\alpha = 0.05, 0.1, 0.15$. Uniform prior represents conditionally uniform priors; M-J prior represents modified Jeffreys priors using Bayes factor; M-J prior(FBF) represents modified Jeffreys priors using fractional Bayes factor. A-J prior represents approximate Jeffreys priors using Bayes factor; A-J prior(FBF) represents approximate Jeffreys priors using fractional Bayes factor.

4.11 Real data analysis

Several real data projects are analyzed by our method and the results are examined with previous studies. The results from our method always give the same conclusion with other methods. The tested real data are Urinary tract infection data obtained from *Bayarri et al.* (2008), Micro-propagated shoots data from *Gupta et al.* (2005), Fetal movement data from *Gupta et al.* (1996).

4.11.1 Urinary tract infection data

This data has been studied by *Bayarri et al.* (2008) for testing Poisson against zero-inflated Poisson distribution by Bayes factor. The data are collected from 98 HIV-infected men treated at the Department of Internal Medicine at the Utrecht University hospital. The number of times they had a urinary tract infection was recorded in Table 4.21.

Table 4.21: Urinary tract infection data

Count	0	1	2	3
Frequency	81	9	7	1

Table 4.22: Summary of urinary tract infection data

Sum of Y	Number of obs	Number of zeros	Mean	Variance
26	98	81	0.2653	0.4031

The summary of urinary tract infection data is given in Table 4.22. Since the zero fraction value is 0.8265, it is apparent that zero-inflation is present in the observed data. The dispersion index is 1.5194, which is not a significant evidence for over-dispersion. Since the number of zeros is a large portion of data compared with the total number of observations, it is reasonable to assume that it might follow the zero-inflated Poisson distribution.

Bayarri et al. (2008) used the Bayes Factor to compare the Poisson(H0) and ZIP(H1) models, and the Bayes factor is

$$B_{10} = \frac{m^{ZIP}}{m^{Poi}} = 223.13$$

Also, The observed value of the score statistic is 15.34 and a pvalue 0.0001. The corresponding posterior probabilities for comparing only these two models are calculated. The results are listed in the following.

$$P(ZIP|y) = 0.9955383$$
$$P(Poi|y) = 0.0044617$$

The posterior probabilities give a strong evidence for supporting ZIP model in comparison with Poisson model. We take all Poisson-related models in comparison, the posterior probability of each model is given below.

The posterior probability of ZIP model is quite large compared with other models. Thus, it suggests ZIP model as the best model in comparison of Poisson-related models. Compared to previous results, our results give the same decision for comparing ZIP and Poisson distribution with the score statis-

Posterior probability of model for urinary tract infection data

Prior	$P(\text{ZIGP}\|y)$	$P(\text{ZIP}\|y)$	$P(\text{GP}\|y)$	$P(\text{Poi}\|y)$
Cond-uniform prior	0.154738	0.652959	0.187632	0.004669
M-J prior(FBF)	0.143587	0.700653	0.147410	0.008348
A-J prior(FBF)	0.119226	0.660690	0.210777	0.009306

tic. Meanwhile, from the overall testing on ZIGP, ZIP, GP and Poisson distribution, posterior probability shows that ZIP distribution still is the best one for the observed data.

4.11.2 Micro-propagated shoots data

The data consist of the number of roots produced by 270 micro-propogated shoots of the columnar apple cultivar Trajan. The roots had been produced under an 8- or 16-h photoperiod in culture systems that utilized one of four different concentrations of the cytokinin BAP in culture medium. Table 4.23 lists the detail of this data. *Gupta et al.* (2005) analyze this data to test zero-inflated generalized Poisson distribution, and their score test rejected zero-inflated Poisson distribution in favor of the zero inflated generalized Poisson distribution.

From the summary of data in table 4.24, the zero fraction is 0.2370, and dispersion index value is $D = 3.1045$. Since they are not outstanding compared with each other, no preference would be made on either ZIP or GP distribution. Posterior probability of model is calculated on Micro-propagated shoots data,

Table 4.23: Micro-propagated shoots data

Count	0	1	2	3	4	5	6	7
Frequency	64	10	13	15	21	18	24	21
Count	8	9	10	11	12	13	14	17
Frequency	23	21	17	12	5	2	3	1

Table 4.24: Summary of Micro-propagated shoots data

Sum of Y	Number of obs	Number of zeros	Mean	Variance
1366	270	64	5.0592	15.7065

and the results are given in the following table. It suggests ZIGP model with the posterior probability at 0.999942.

Table 4.25: Posterior probability of model for Micro-propagated data

Prior	$P(\text{ZIGP}\|y)$	$P(\text{ZIP}\|y)$	$P(\text{GP}\|y)$	$P(\text{Poi}\|y)$
Cond-uniform prior	0.999952	4.800551e-05	0	0
M-J prior(FBF)	0.999942	5.804387e-05	0	0
A-J prior(FBF)	0.999954	4.560269e-05	0	0

Gupta et al. (2005) tests the overdispersion and the zero-inflated factor of ZIGP distribution through score test method. The score statistic is 16.821 and a pvalue is 0.000041 for ZIP versus ZIGP distribution, thus ZIP distribution is rejected in favor of ZIGP distribution. Our posterior probabilities based on

only ZIP and ZIGP candidate distributions show that the posterior probability of ZIGP is almost 1 and the posterior probability of ZIP is close to 0. The score statistic is 159.669 and a pvalue is < 0.00001 for GP versus ZIGP distribution, thus GP distribution is rejected in favor of ZIGP distribution. Our posterior probabilities based on only GP and ZIGP candidate distributions show that $P(\text{ZIGP}|y) \approx 1$ and $P(\text{GP}|y) \approx 0$. The results are quite similar with the score test, but posterior probability gives a direct measure of the distribution and the meaning of the measure is straightforward.

4.11.3 Fetal movement data

This data includes numbers of body movements in fetal lambs observed through ultra sound. It is designed for a study of breathing and body movements in fetal lambs to examine the possible changes in the amount of movement pattern of fetal activity during the last two thirds of the gestation period. *Gupta et al.* (1996) analyzed one particular sequence of counts of the number of movements in 240 consecutive 5-s intervals. This data set is given in table 4.26 In the paper, the author provides the confidence interval of parameters based on the maximum likelihood estimation.

Table 4.26: Fetal movement data

Count	0	1	2	3	4	7
Frequency	182	41	12	2	2	1

The sequence of counts of the number of movements is a measurement in

240 consecutive 5-s intervals. There are 240 observations in the data set. Zero fraction is high at 0.7583, so it is highly zero-inflated. The dispersion index is $D = 1.835361$. The summary of the count data is in Table 4.27

Table 4.27: Summary of Fetal movement data

Sum of Y	Number of obs	Number of zeros	Mean	Variance
86	240	182	0.3583	0.6576

After running Bayesian method on distribution selection of Poisson, GP, ZIP, ZIGP distribution, the posterior probability of distribution is given in Table 4.28.

Table 4.28: Posterior probabilities of distribution for Fetal movement data

Prior	$P(\text{ZIGP}\|y)$	$P(\text{ZIP}\|y)$	$P(\text{GP}\|y)$	$P(\text{Poi}\|y)$
Cond-uniform prior	0.237391	7.988690e-03	0.754618	1.201905e-06
M-J prior(FBF)	0.207035	2.556833e-02	0.76739	4.742733e-06
A-J prior(FBF)	0.152251	2.119160e-02	0.826552	4.551867e-06

Gupta et al. (1996) estimates the parameters of ZIGP distribution in the frequentist way, the maximum likelihood estimation is used to test the significance of distribution parameters. The results of MLE and 95% CI are given in the following table.

Because the 95% confidence intervals of parameters λ and α all exclude zero, these two parameters are nonzero that they remain in the distribution.

Estimation of parameters by *Gupta et al.* (1996)

Parameters	ϕ	α	λ
MLE	- 0.3143	1.1254	0.2032
95% CI	(- 0.92433, 0.29578)	(0.83644, 1.67197)	(0.11161, 0.29478)

ϕ including zero in its 95% confidence interval and its asymptotic variance is small, but it is not clear enough to conclude the value of ϕ. Our results give the straightforward conclusion about the most proper distribution that is GP distribution with posterior probability at 0.754618 of conditionally uniform prior and 0.76739 using modified Jeffreys priors. This is a strong support for GP distribution.

4.12 Discussion

Bayesian method of model selection has its advantage of comparing models in complex form, it does not require any prior knowledge about models but updates belief through observed data. We developed the posterior probability for each candidate model. Posterior probability is based on Bayes factors but provides more straightforward interpretation compared with Bayes factor, while Bayes factor only allow to compare two models a one time, it can compare more than two models simultaneously. Non-informative prior, such as uniform prior and Jeffreys prior, is selected in the computation of the posterior probability, because little prior knowledge about models is normally available in practice. We derived Jeffreys priors for ZIGP, ZIP, GP and Poisson models, and further

modified Jeffreys priors of the models for model comparison, since the derived Jeffreys priors of GP and ZIGP models are not suitable priors in Bayes factors testing due to the uncertainty that improper prior brings in. For this reason, Fractional Bayes factors are introduced to replace Bayes factors for the purpose of removing the uncertainty of improper priors. The derived Jeffreys priors of ZIGP, ZIP, and GP models are approximated by modified Jeffreys prior and approximate Jeffreys prior which are defined to simplify the computation. The simulation studies are carried out by using conditionally uniform prior, modified Jeffreys prior and approximate Jeffreys prior separately, and Bayes factors and Fractional Bayes factors are implemented respectively to compare the results of the posterior probability of candidate models.

Our simulation demonstrates that the Bayesian method is very stable and accurate in finding the correct model. When the four models are different enough (e.g., when ZIP model with $\lambda \geq 3$ and $\phi \geq 0.1$ in Fig. 4.3), all three priors consistently pick out the correct model by a large margin in posterior probability. When the differences among the models are small, priors may favor either true model or a simpler one by a small margin in posterior probability. Such a discrepancy is not unreasonable, because under this circumstance, the simpler model characterize the data almost equally well as the true model, i.e. they have the same predictive power, hence a simpler model is better in the sense of Occams razor.

Uniform prior is easier to use than Jeffreys prior in calculation of posterior probabilities, therefore the conditionally uniform prior is recommended for the

use in comparison of Poisson-related models. Meanwhile, Jeffreys prior is a good choice when its feature of invariance under reparameterization is needed. Posterior probability method gives clear and straightforward explanation of results compared to other test methods for model comparison. Moreover, Bayesian approach to model comparison using the non-informative prior does not require prior information of model and can be used as an automatic or default method in any context.

Chapter 5

Bayesian model selection of Poisson and related regression models

In this chapter, distributions selection considers inclusion of covariates. The testing of generalized Poisson regression model versus Poisson regression model is addressed, and the comparison is extended to take the negative binomial regression model into consideration which is quite close to the generalized Poisson model.

5.1 Poisson, Generalized Poisson and Negative Binomial regression models

In some papers, the common parameter λ within GP regression model is treated in the same way with the one from Poisson regression model, but λ parameter is not the model mean in GP regression model. In fact, the meaning of λ changes among models, and the same log link assigned to the parameter λ of all models is not quite meaningful to carry over the effect of

covariate. To compare GP regression model versus Poisson regression model, the reparameterization of λ to the model mean μ in GP regression model is studied here.

Possion regression model

The Poisson model is written in the parameter notation of μ take the place of orginal λ.

The probability mass function of the Poisson Regression is

$$f(y_i|\mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \qquad y_i = 0, 1, 2, ...$$

where $\mu_i > 0$, it represents the value of the mean and variance of y_i. The covariates are included through the log link function that $\mu_i = exp(x_i\beta)$. $\beta = (\beta_1, \beta_2, ..., \beta_p)$ is a p-dimensional vector of unknown regression parameters to be estimated, and $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$ is the covariate vector having an effect on the population mean μ_i. In general, the log link function of covariate can be represented by $\mu = exp(X\beta)$, and X is a n by p covariate matrix, $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$ is the ith row of the covariate matrix X.

Likelihood of Poisson regression model:

$$\begin{aligned} L^{\mathrm{PR}}(\mu) &= \prod_{i=1}^{n} f(y_i|\mu_i) = \prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \\ &= \frac{1}{\prod_{i=1}^{n} y_i!} exp\left(\sum_{i=1}^{n} y_i x_i \beta - \sum_{i=1}^{n} e^{x_i\beta}\right) \end{aligned}$$

Generalized Poisson regression model

The probability mass function of the generalized Poisson (GP) regression is

$$f(y_i|\theta,\mu_i) = \frac{\mu_i(1-\theta)(\mu_i(1-\theta)+\theta y_i)^{(y_i-1)}e^{-((1-\theta)\mu_i+\theta y_i)}}{y_i!} \qquad y_i = 0,1,2,...$$

where $\mu_i > 0$, it is the mean of y_i. The variance of y_i is $\frac{\mu_i}{(1-\theta)^2}$, $0 \leq \theta < 1$. Since the parameter μ represents mean of both Poisson regression model and generalized Poisson regression model, the meaning of μ is same for both models. Therefore, the covariate link function can be defined in the same way in both models, commonly it is log link function $\mu_i = exp(x_i\beta)$.

Likelihood of generalized Poisson regression model:

$$\begin{aligned} L^{\text{GPR}}(\theta,\mu) &= \prod_{i=1}^{n} f(y_i|\theta,\mu_i) = \prod_{i=1}^{n} \frac{\mu_i(1-\theta)(\mu_i(1-\theta)+\theta y_i)^{(y_i-1)}e^{-((1-\theta)\mu_i+\theta y_i)}}{y_i!} \\ &= \frac{(1-\theta)^n}{\prod_{i=1}^{n} y_i!} exp\left(\sum_{i=1}^{n} x_i\beta - (1-\theta)\sum_{i=1}^{n} e^{x_i\beta} - \theta s\right) \prod_{i=1}^{n} \left((1-\theta)e^{x_i\beta}+\theta y_i\right)^{y_i-1} \end{aligned}$$

The GP model reduces to Poisson model as parameter θ goes to 0. Therefore, testing overdispersion for GP against Poi regression model comparison becomes a test of the hypothesis

$$H_0 : \theta = 0 \quad vs. \quad H_1 : 0 < \theta < 1$$

Bayes factor or posterior probability can be applied for the hypothesis testing. Importance sampling helps calculate marginal likelihoods of models.

Negative binomial regression model

The probability mass function of the Negative Binomail (NB) Regression is

$$f(y_i|\tau,\mu_i) = \frac{\Gamma(y_i+1/\tau)}{\Gamma(y_i+1)\Gamma(1/\tau)}\frac{(\tau\mu_i)^{y_i}}{(1+\tau\mu_i)^{y_i+1/\tau}} \qquad y_i = 0,1,2,...$$

$\tau > 0$ controls the degree of overdispersion. $\mu_i > 0$ is the parameter equal to the mean. Variance of y_i is $(1+\tau\mu_i)\mu_i$. Since the parameter μ is the mean of both Poisson and NB regression models, the meaning of μ are same in both models. Therefore, The covariates are included through the log link function that $\mu_i = exp(x_i\beta)$.

Likelihood of NB Regression Model

$$\begin{aligned} L^{\text{NBR}}(\tau,\mu) &= \prod_{i=1}^{n} f(y_i|\tau,\mu_i) = \prod_{i=1}^{n} \frac{\Gamma(y_i+1/\tau)}{\Gamma(y_i+1)\Gamma(1/\tau)}\frac{(\tau\mu_i)^{y_i}}{(1+\tau\mu_i)^{y_i+1/\tau}} \\ &= \prod_{i=1}^{n} \frac{\Gamma(y_i+1/\tau)}{\Gamma(y_i+1)\Gamma(1/\tau)}\frac{(\tau e^{x_i\beta})^{y_i}}{(1+\tau e^{x_i\beta})^{y_i+1/\tau}} \end{aligned}$$

This NB model reduces to Poisson model as τ goes to zero. Hence, testing Poisson regression model against NB regression model is equivalent to testing the hypothesis:

$$H_0 : \tau = 0 \quad vs. \quad H_1 : \tau > 0$$

5.2 Generalized Poisson vs. Poisson regression model

The bayesian approach to model comparison is based on Bayes Factor B_{10} of model M_1 to M_0 given by

$$B_{10} = \frac{m_1(y)}{m_0(y)} = \frac{\int f_1(y|\theta_1)\pi(\theta_1)d\theta_1}{\int f_0(y|\theta_0)\pi(\theta_0)d\theta_0}$$

It is common to make non-informative priors as the parameter priors $\pi(\theta_0)$ and $\pi(\theta_1)$ for the sake of the lack of pre-knowledge about the parameters.

The model Comparison of GP against Poisson Regression model is equivalent to testing the over-dispersion through the hypothesis:

$$H_0 : \theta = 0 \quad vs. \quad H_1 : 0 < \theta < 1$$

Assume equal prior probability of model: $P(M^{\mathrm{GPR}}) = P(M^{\mathrm{PR}}) = \frac{1}{2}$. The test statistic of Bayes factor is

$$BF_{\mathrm{GPR,PR}} = \frac{m^{\mathrm{GPR}}}{m^{\mathrm{PR}}}$$

Because covariate parameter vector β is a unknown vector of reals. The flat prior of multi-Normal distribution with large variance over all real domain is a reasonable prior for β.

$$\beta \sim N(0, 100 \times I_p)$$

For the overdispersion parameter θ, its range is over $(0,1)$, a suitable prior is simply the flat prior that is Uniform distribtuion over $(0,1)$.

$$\theta \sim Uniform(0,1)$$

The marginal likelihoods of Poisson and GP models are

$$\begin{aligned} m^{\mathrm{PR}} =& \int L^{\mathrm{PR}}(\beta)\pi(\beta)d\beta \\ =& \frac{1}{\sigma_\beta(2\pi)^{p/2}\prod_{i=1}^n y_i!}\int exp\left(\sum_{i=1}^n y_i x_i\beta - \sum_{i=1}^n e^{x_i\beta} - \frac{1}{2\sigma_\beta^2}\beta'\beta\right)d\beta \end{aligned}$$

$$\begin{aligned} m^{\mathrm{GPR}} =& \int L^{\mathrm{GPR}}(\theta,\beta)\pi(\theta,\beta)d\theta d\beta \\ =& \frac{1}{\sigma_\beta(2\pi)^{p/2}\prod_{i=1}^n y_i!}\int\int (1-\theta)^n exp\left(\sum_{i=1}^n x_i\beta - (1-\theta)\sum_{i=1}^n e^{x_i\beta}\right. \\ & \left. -\theta s - \frac{1}{2\sigma_\beta^2}\beta'\beta\right)\prod_{i=1}^n \left((1-\theta)e^{x_i\beta} + \theta y_i\right)^{(y_i-1)} d\theta d\beta \end{aligned}$$

The integration of marginal likelihoods of both models contain multiple variables that increase the complexity of the integrand, so it is impossible to obtain the analytical expression for the marginal likelihood. Thus, Importance sampling is utilized to approach the value of the integrals via numerical iteration.

Importance Sampling Approach of Marginal Likelihoods

While the number of variables present in the integral of marginal likeli-

hoods increases, the speed of convergence of importance sampling approach is dramatically increased. To reduce the dimension of variables in importance sampling method, the importance sampling density of β is possible to be estimated from the data in the first place. The estimated importance density concentrates on the true value of β with a small variance. It helps the Markov chain of importance sampling approach reach stationary status quickly.

The reason of that the coefficient β is able to adapt the given data to get knowledge about it is because β is associated with model mean μ, and model mean is same for all models given the data set. Therefore, the simple model, Poisson regression model, can be utilized to draw information about the estimated mean and variance of the coefficient β. Denote the estimated mean and variance as $\hat{\mu}_\beta$ and $\hat{\sigma}^2_\beta$ respectively. The importance sampling density for β is

$$\beta \sim N(\hat{\mu}_\beta, \hat{\sigma}^2_\beta \times I_p)$$

where $\hat{\mu}_\beta$ and $\hat{\sigma}^2_\beta$ are estimated values obtained from Poisson regression model (glm function) to increase the accuracy and efficiency of Importance Sampling.

The importance density of θ of GP model chooses the flat prior Uni-

form(0,1). Now, the importance sampling approach is

$$m^{\mathrm{PR}} \approx \frac{1}{m\prod_{i=1}^{n} y_i!}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta} exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j+\right.$$
$$\left.\sum_{i=1}^{n} y_i x_i\beta_j-\sum_{i=1}^{n} e^{x_i\beta_j}\right)$$

$$m^{\mathrm{GPR}} \approx \frac{1}{m\prod_{i=1}^{n} y_i!}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta}(1-\theta_j)^n exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j\right.$$
$$\left.+\sum_{i=1}^{n} x_i\beta_j-(1-\theta_j)\sum_{i=1}^{n} e^{x_i\beta_j}-\theta_j s\right)\prod_{i=1}^{n}\left(e^{x_i\beta_j}(1-\theta_j)+\theta_j y_i\right)^{(y_i-1)}$$

The Importance Sampling converges quickly after adopting the estimated beta parameter from Poisson Regression Model (glm), the number of iterations needed is more than 3000.

5.3 Negative Binomial vs. Poisson regression model

To test the hypothsis

$$H_0 : \tau = 0 \quad vs. \quad H_1 : \tau > 0$$

Assume equal prior probability of model: $P(M^{\mathrm{NBR}}) = P(M^{\mathrm{PR}}) = \frac{1}{2}$. The test statistic of Bayes factor is

$$BF_{\mathrm{NBR,PR}} = \frac{m^{\mathrm{NBR}}}{m^{\mathrm{PR}}}$$

The prior for the regression parameters β is same as Poisson regression model.

$$\beta \sim N(0, 100 \times I_p)$$

The NB regression parameter τ blongs to $(0, +\infty)$. Without any pre-knowledge of τ, it is common to have a simple flat prior over the domain.

$$\pi(\tau) = 1 \quad for\ \tau \in [0, +\infty)$$

However, the flat prior over the infinity domian is a improper prior. The Bayes factor is inappropriate for the improper prior on the uncommon parameter of competitive models. The uncertainty of improper prior in Bayes factor can cancel out through fractional Bayes factor.

The marginal likelihood of negative binomial model is

$$\begin{aligned} m^{\mathrm{NBR}} &= \int L^{\mathrm{NBR}}(\tau, \beta)\pi(\tau, \beta)d\tau d\beta \\ &= \int \prod_{i=1}^{n} \left(\frac{\Gamma(y_i + 1/\tau)(\tau e^{x_i\beta})^{y_i}}{\Gamma(y_i + 1)\Gamma(1/\tau)(1 + \tau e^{x_i\beta})^{y_i + 1/\tau}} \right) \cdot \frac{1}{\sigma_\beta (2\pi)^{p/2}} e^{-\frac{1}{2\sigma_\beta^2}\beta'\beta} d\tau d\beta \end{aligned}$$

Let the training fraction of FBF be $b = \frac{n_1}{n}$, where $n = n_1 + n_2$. The minimum size of training sample is $n_1 = 2$ by default. The fractional marginal

likelihoods of negative binomial and Poisson models are

$$
\begin{aligned}
m_b^{\mathrm{PR}} &= \int L^{\mathrm{PR}}(\beta)^b \pi(\beta) d\beta \\
&= \frac{1}{\sigma_\beta (2\pi)^{p/2} \left(\prod_{i=1}^n y_i!\right)^b} \int exp\left(b\sum_{i=1}^n y_i x_i \beta - b\sum_{i=1}^n e^{x_i\beta} - \frac{1}{2\sigma_\beta^2}\beta'\beta \right) d\beta
\end{aligned}
$$

$$
\begin{aligned}
m_b^{\mathrm{NBR}} &= \int L^{\mathrm{NBR}}(\tau,\beta)^b \pi(\tau,\beta) d\tau d\beta \\
&= \int \left[\prod_{i=1}^n \frac{\Gamma(y_i + 1/\tau)(\tau e^{x_i\beta})^{y_i}}{\Gamma(y_i+1)\Gamma(1/\tau)(1+\tau e^{x_i\beta})^{y_i+1/\tau}} \right]^b \cdot \frac{1}{\sigma_\beta (2\pi)^{p/2}} e^{-\frac{1}{2\sigma_\beta^2}\beta'\beta} d\tau d\beta
\end{aligned}
$$

The fractional Bayes factor is in the form of

$$
FBF_{\mathrm{NBR,PR}} = \frac{m^{\mathrm{NBR}} m_b^{\mathrm{PR}}}{m^{\mathrm{PR}} m_b^{\mathrm{NBR}}}
$$

The unspecified constant cancels out in the fractional Bayes factor through the training sample. The elements in computation of FBF are too complex to obtain the integral. As before, the importance sampling approach work again in approximation of the integration of these four marginal likelihoods. First in all, the importance sampling density is required to perform the algorithm. To reduce the variance of sampling in high dimension, the estimated distribution is applied to coefficient parameter β from the Poisson regression inference. The NB regression parameter τ has its range over $(0, \infty)$, but our interest is focus on the null hypothesis which is locate at origin point 0. In the context of testing purpose, a decreasing distribution from origin point to infinity is a proper choice for the importance density of τ. In this case, half-Cauchy

distribution is selected with large variance that is half-Cauchy(0,25). The distribution is peak at the origin zero, and right skewed to infinity. The shape of half-Cauchy(0,25) distribution is on figure 5.1.

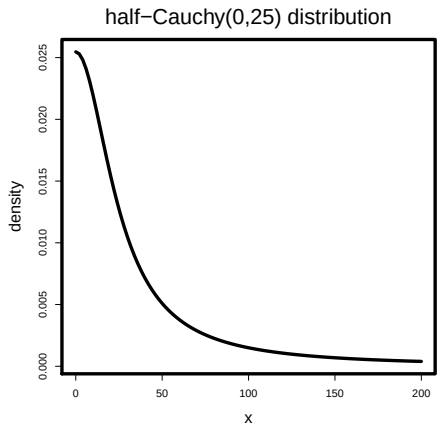


Figure 5.1: The half-Cauchy(0,25) distribution

The importance sampling densities for $FBF_{\text{NBR,PR}}$

$$\beta \sim N(\hat{\mu}_\beta, \hat{\sigma}^2_\beta \times I_p)$$

$$\tau \sim \text{half-Cauchy}(0, 25)$$

where $\hat{\mu}_\beta$ and $\hat{\sigma}^2_\beta$ are estimated values obtained from Poisson regression model (glm function) to increase the accuracy and efficiency of Importance Sampling.

Now, the importance sampling approach is

$$m^{\mathrm{PR}} \approx \frac{1}{m\prod_{i=1}^{n} y_i!}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta}exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j+\sum_{i=1}^{n}y_i x_i\beta_j-\sum_{i=1}^{n}e^{x_i\beta_j}\right)$$

$$m_b^{\mathrm{PR}} \approx \frac{1}{m\left(\prod_{i=1}^{n} y_i!\right)^b}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta}exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j+b\sum_{i=1}^{n}y_i x_i\beta_j-b\sum_{i=1}^{n}e^{x_i\beta_j}\right)$$

$$m^{\mathrm{NBR}} \approx \frac{1}{m}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta}exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j\right)\cdot\prod_{i=1}^{n}\left(\frac{\Gamma(y_i+1/\tau_j)(\tau_j e^{x_i\beta_j})^{y_i}}{\Gamma(y_i+1)\Gamma(1/\tau_j)(1+\tau_j e^{x_i\beta_j})^{y_i+1/\tau_j}}\right)\cdot\frac{\pi(\tau_j^2+25^2)}{50}$$

$$m_b^{\mathrm{NBR}} \approx \frac{1}{m}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta}exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j\right)\cdot\prod_{i=1}^{n}\left(\frac{\Gamma(y_i+1/\tau_j)(\tau_j e^{x_i\beta_j})^{y_i}}{\Gamma(y_i+1)\Gamma(1/\tau_j)(1+\tau_j e^{x_i\beta_j})^{y_i+1/\tau_j}}\right)^b\cdot\frac{\pi(\tau_j^2+25^2)}{50}$$

5.4 Posterior probability of model using flat priors

Until now, we have discussed the Bayes factor method on how to do paired test. It is well known that generalized Poisson model is quite similar to negative binomial model, which implies a common question of how to differentiate these

models simultaneously. Here, the posterior probability of model given data is used for comparison. The prior probability of model is assumed equal in default.

$$\begin{aligned} P(M_i|y) =& \frac{P(y|M_i)P(M_i)}{\sum_{j\in all} P(y|M_j)P(M_j)} = \frac{m_i}{m^{\mathrm{PR}} + m^{\mathrm{GPR}} + m^{\mathrm{NBR}}} \\ =& \frac{BF_i}{1 + BF_{\mathrm{GPR,PR}} + BF_{\mathrm{NBR,PR}}} \approx \frac{BF_i}{1 + BF_{\mathrm{GPR,PR}} + FBF_{\mathrm{NBR,PR}}} \end{aligned}$$

where m_i and BF_i are the marginal likelihood and Bayes factor of model from the competitive models, $i \in \{\mathrm{PR,GPR,NBR}\}$. $FBF_{\mathrm{NBR,PR}}$ is used to approximate $BF_{\mathrm{NBR,PR}}$ to reduce the uncertainty of improper prior. Posterior probability of model is convenient to compare more than two models in the frame of Bayesian method. It provides the probability of each model in selection, and the value of probability has explicit meaning and is easy for explanation.

5.5 Simulation of data from regression model

To simulate the data from a model with known values of parameters and covariate matrix in the model, the quantile method is applied for the multivariate regression model simulation. First, a set of parameter values $\beta^* = (\beta_1^*, ..., \beta_p^*)^T$ is given to define the model from which the simulated data comes. Second, a set of covariate value $\{x_{i(1)}, ..., x_{i(p)}\}$ is required for each simulated value $y_i, i = 1, ..., n$. Let n be the number of desired simulation data, the covariate

matrix $X_{n\times p}$ consists of n rows of covariate values. The first column of the covariate matrix $x_{.(1)}$ represents the index l of the intercept, and the other columns $x_{.(j)}, j = 2, ..., p$ are measure of the covariate variables respectively. The covariate matrix is supposed to be able to contain both fixed effects and random effects. For the mixed effects, the binary variable and continuous variable are combined to create the covariate matrix. Here, the continuous variable is assumed from the normal distribution with different mean and variance.

For example, the covariate matrix of two fixed effects and one random effect, where the random effect is composed of two categories which are from different normal distributions, is in the form that

$$X = \begin{bmatrix} 1\ 0\ 0 & x_1 \\ \vdots\ \vdots\ \vdots & \\ & \\ 1\ 0\ 0 & \vdots \\ 1\ 1\ 0 & \\ \vdots\ \vdots\ \vdots & x_k \\ 1\ 1\ 0 & x_{(k+1)} \\ & \\ 1\ 0\ 1 & \vdots \\ \vdots\ \vdots\ \vdots & \\ 1\ 0\ 1 & x_n \end{bmatrix}$$

where the $x_i \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_i \sim N(\mu_2, \sigma_2^2)\ for\ i = k + 1, ..., n$. The first column of the matrix is the index for intercept variable β_0 which is always equal to one. The second and third columns of the matrix are the qualitative variables identifying the the category for each subject. The last column is the quantitative random variable, each one of which is from a population.

Generally, the covariate matrix for multiple covariates can be written in

the convenient way of

$$X = \{X_{\cdot(1)}, X_{\cdot(2)}, ..., X_{\cdot(p)}\}_{n \times p}$$

where the $X_{\cdot(i)}$ is the ith covariate associated with the coefficient β_i. The first column vector $X_{\cdot(1)} = \mathbf{1}$ is always a vector of all ones in default. $X_{\cdot(i)} = (x_{1(i)}, x_{2(i)}, ..., x_{n(i)})^T$, and the variable can be either binary variable for the fixed effect or continuous variable for the random effect, meanwhile, there could be multiple different continuous distribution presenting in one covariate column data.

The process of simulation of quantile method:

1. fit the model with the initial values of parameters and covariate vector
2. initiate the start state $y = 0$;
3. draw a random value u from uniform distribution $U(0, 1)$;
4. accept the draw according to the probability $F(Y = y)$.
5. If $u \leq F(Y = y)$, accepted and return the value of y.
6. Else, let $y = y + 1$, repeat step 3 and 4.

Run the same process multiple times with the same setting of parameters and covariate values to obtain the desired number of simulated numbers from one model.

5.6 Simulation results of Fractional Bayes Factor

Simulation analysis is implemented to evaluate the Bayesian model selection method for Poisson regression model, generalized Poisson regression model, and negative binomial regression model. Posterior probability of model is the test statistic for model selection of these three models. For this simple initial analysis, we only learn the data simulated from negative binomial regression model.

Simulation analysis of two regression coefficients β is studied first. Assume there is only one covariate x, then regression coefficient β is 2-dimensional. The covariate matrix can be written as follows:

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_k \\ 1 & x_{(k+1)} \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

where $x_i \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_i \sim N(\mu_2, \sigma_2^2)\ for\ i = k+1, ..., n$. First, we just consider one set of coefficients (1,0.1), and make NBR model parameter τ change values as 0.1, 1, 5 and let GPR model parameter θ vary from 0 to 0.5 increasing by 0.1 unit. The covariate matrix simulated randomly from $x_i \sim N(3, 1)\ for\ i = 1, ..., k$ and $x_i \sim N(10, 2)\ for\ i = k+1, ..., n$. For each setting of parameters and coefficients, 100 datasets of size 100 are simulated from the negative binomial regression model. Importance sampling iterate 5000 times for each simulated data to draw conclusion of posterior probabilities of models. The mean value of posterior probability in each it-

eration of importance sampling is the final estimate of posterior probability. The results are displayed in table 5.1. Based on the simulation analysis, the fractional Bayes factor does not work well in approximation of Bayes factor in calculation of posterior probability of model. The results show the preference to GPR model even when the NBR model parameter τ is quite large at 5.

Table 5.1: Simulation results of one covariate using FBF

β	θ	τ	$\bar{P}(\text{PR}\|y)$	$\bar{P}(\text{GPR}\|y)$	$\bar{P}(\text{NBR}\|y)$	$SD[P(\text{PR}\|y)]$	$SD[P(\text{GPR}\|y)]$	$SD[P(\text{NBR}\|y)]$
(1,0.1)	0	na	0.826267	0.105379	0.068354	0.150765	0.092883	0.072922
(1,0.1)	0.1	na	0.494499	0.325749	0.179752	0.285439	0.196095	0.152314
(1,0.1)	0.2	na	0.149856	0.590846	0.259298	0.237879	0.226067	0.152471
(1,0.1)	0.3	na	0.004331	0.756727	0.238942	0.021468	0.202605	0.199938
(1,0.1)	0.5	na	6.61E-13	0.902208	0.097792	4.34E-12	0.15681	0.15681
(1,0.1)	na	0.1	0.149192	0.490464	0.360344	0.21794	0.21497	0.208682
(1,0.1)	na	1	1.96E-40	0.851846	0.148154	1.67E-39	0.23015	0.23015
(1,0.1)	na	5	4.67E-127	0.995	0.005	4.59E-126	0.029043	0.029043

Each row contains posterior probabilities based on 100 simulations from a model defined by parameters β, θ, τ and covariate matrix X. True model is determined by θ and τ, where $\tau =$ na means true model is GPR model, and $\theta =$ na says data is simulated from NBR model. Poisson model has parameter $\theta = 0, \tau = na$ particularly.

Joe and Zhu (2005) says GP and NB distributions are quite close to each other, when population mean is smaller than 5 or dispersion rate is small (close to 1). From this point of view, the testing of GP and NB models becomes sensitive to any approximation in the process. For this reason, FBF is not good approximation for the simulated data, because the mean of these data are all less than 5.

When the mean of data incrases greatly, the results of model selection do improve with application of FBF. For example, covariate matrix is in a form of

$$X = \begin{bmatrix} 1 & 0 & x_{31} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{3k} \\ 1 & 1 & x_{3(k+1)} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{3n} \end{bmatrix}$$

where $x_3 i \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_3 i \sim N(\mu_2, \sigma_2^2)\ for\ i = k+1, ..., n$.

These three covariates are all binary variables. The covariate matrix simulated randomly from $x_i \sim N(3, 1)\ for\ i = 1, ..., k$ and $x_i \sim N(10, 2)\ for\ i = k+1, ..., n$. The initial value set for covariate coefficient vector is $\beta = (1, 0.5, 0.1)^T$, of which the mean of the model is greater than 5. Based on *Joe and Zhu* (2005), it means the difference of GP and NB models is large enough to detect. Simulation is based on the simulated data from all candidate models with different model parameters. The statistics of simulation study is concluded from 100 simulated data for each setup model, and results are listed in the table 5.2.

The results of three covariate variables illustrate the same problem occurred in the first simulation study. The posterior probability is not able to find the correct NBR model when NBR model parameter τ is truly far away from zero. The application of FBF approach makes the posterior probability in favor of GPR model. When true model is GPR model, the posterior probability gives larger value for true model as the GPR parameter θ increases.

More complex covariate is studied to further understand the performance of our model testing method. One additional covariate is added and treated

Table 5.2: Simulation results of two covariates using FBF

β	θ	τ	$\bar{P}(\text{PR}\|y)$	$\bar{P}(\text{GPR}\|y)$	$\bar{P}(\text{NBR}\|y)$	$SD[P(\text{PR}\|y)]$	$SD[P(\text{GPR}\|y)]$	$SD[P(\text{NBR}\|y)]$
(1,0.5,0.1)	0	na	0.833559	0.098388	0.068053	0.11641	0.067369	0.073498
(1,0.5,0.1)	0.1	na	0.543488	0.312299	0.144213	0.280417	0.233034	0.157737
(1,0.5,0.1)	0.2	na	0.14348	0.658715	0.197804	0.211086	0.271381	0.207179
(1,0.5,0.1)	0.3	na	0.011178	0.878624	0.110199	0.061531	0.171926	0.160382
(1,0.5,0.1)	0.5	na	1.10E-09	0.938159	0.061841	1.10E-08	0.182948	0.182948
(1,0.5,0.1)	na	0.1	0.036628	0.425312	0.538061	0.127258	0.264881	0.286114
(1,0.5,0.1)	na	1	3.87E-52	0.684282	0.315718	2.95E-51	0.349251	0.349251
(1,0.5,0.1)	na	5	3.38E-181	0.998228	0.001772	0	0.010019	0.010019

Each row contains posterior probabilities based on 100 simulations from a model defined by parameters β, θ, τ and covariate matrix X. True model is determined by θ and τ, where τ = na means true model is GPR model, and θ = na says data is simulated from NBR model. Poisson model has parameter $\theta = 0, \tau = na$ particularly.

as binary variable. The covariate matrix is generated in the following way

$$X = \{X_{.(1)}, X_{.(2)}, X_{.(3)}, X_{.(4)}\}_{n\times 3}$$

where the $X_{.(i)}$ is the ith covariate vector associated with the coefficient β_i. For the simulation analysis, let $X_{.(1)} = [1]_{n\times 1}$ be the identity vector for intercept. $X_{.(2)} = (0, ..., 0, 1, ..., 1)^T$ and $X_{.(3)} = (0, ..., 0, 1, ..., 1, 0..., 0)^T$. $X_{.(4)}$ is a random vector with $x_i(4) \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_i(4) \sim N(\mu_2, \sigma_2^2)\ for\ i = k+1, ..., n$.

Testing values of model parameters θ and τ are kept same as before. The results from simualtion study are given in table 5.3.

Table 5.3: Simulation results of three covariates using FBF

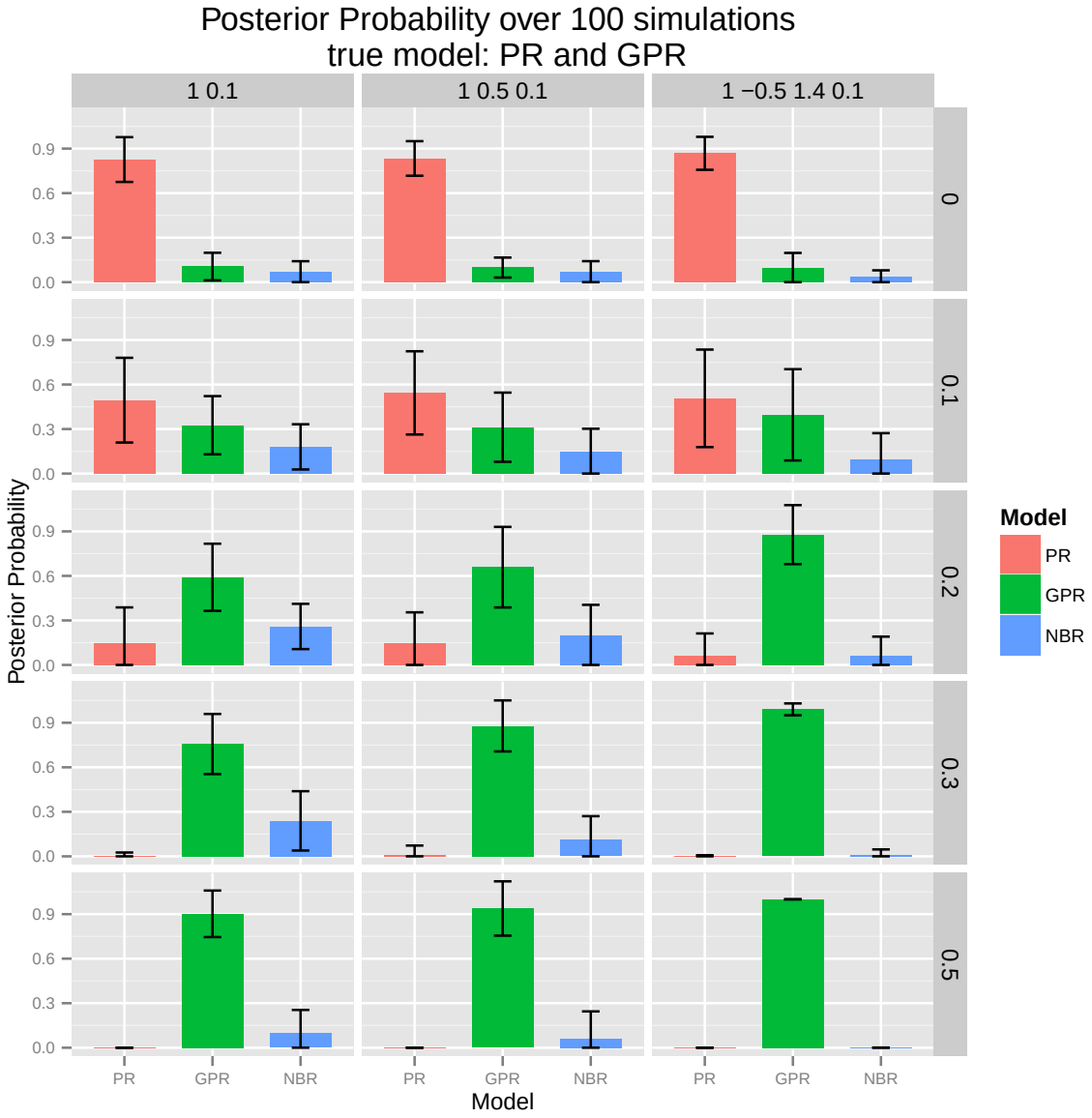
β	θ	τ	$\bar{P}$(PR$\|y$)	$\bar{P}$(GPR$\|y$)	$\bar{P}$(NBR$\|y$)	$SD[P$(PR$\|y)]$	$SD[P$(GPR$\|y)]$	$SD[P$(NBR$\|y)]$
(1,-0.5,1.4,0.1)	0	na	0.868192	0.09586	0.035947	0.111117	0.101186	0.043802
(1,-0.5,1.4,0.1)	0.1	na	0.50672	0.395764	0.097516	0.328551	0.307577	0.174752
(1,-0.5,1.4,0.1)	0.2	na	0.060035	0.878138	0.061827	0.152163	0.198809	0.128984
(1,-0.5,1.4,0.1)	0.3	na	0.001005	0.990331	0.008664	0.006478	0.039912	0.037884
(1,-0.5,1.4,0.1)	0.5	na	4.04E-17	0.999855	0.000145	3.98E-16	0.000839	0.000839
(1,-0.5,1.4,0.1)	na	0.1	3.40E-06	0.047342	0.952655	2.53E-05	0.171269	0.171273
(1,-0.5,1.4,0.1)	na	1	2.29E-137	0.092536	0.907464	2.19E-136	0.247901	0.247901
(1,-0.5,1.4,0.1)	na	5	0	0.861579	0.138421	0	0.306651	0.306651

Each row contains posterior probabilities based on 100 simulations from a model defined by parameters β, θ, τ and covariate matrix X. True model is determined by θ and τ, where τ = na means true model is GPR model, and θ = na says data is simulated from NBR model. Poisson model has parameter $\theta = 0, \tau = na$ particularly.

From the simulation analysis, the posterior probability does not give the right choice when true model is negative binomial regression model. The problem may be caused by the error introduced by fractional Bayes factor. FBF can not be used to take place of BF in posterior probability for this case. Graphic comparison is given by figure 5.2 and figure 5.3, and it is a more convenient way to summarize the overall performance of posterior probability with adjustment by fractional Bayes factor for improper prior.

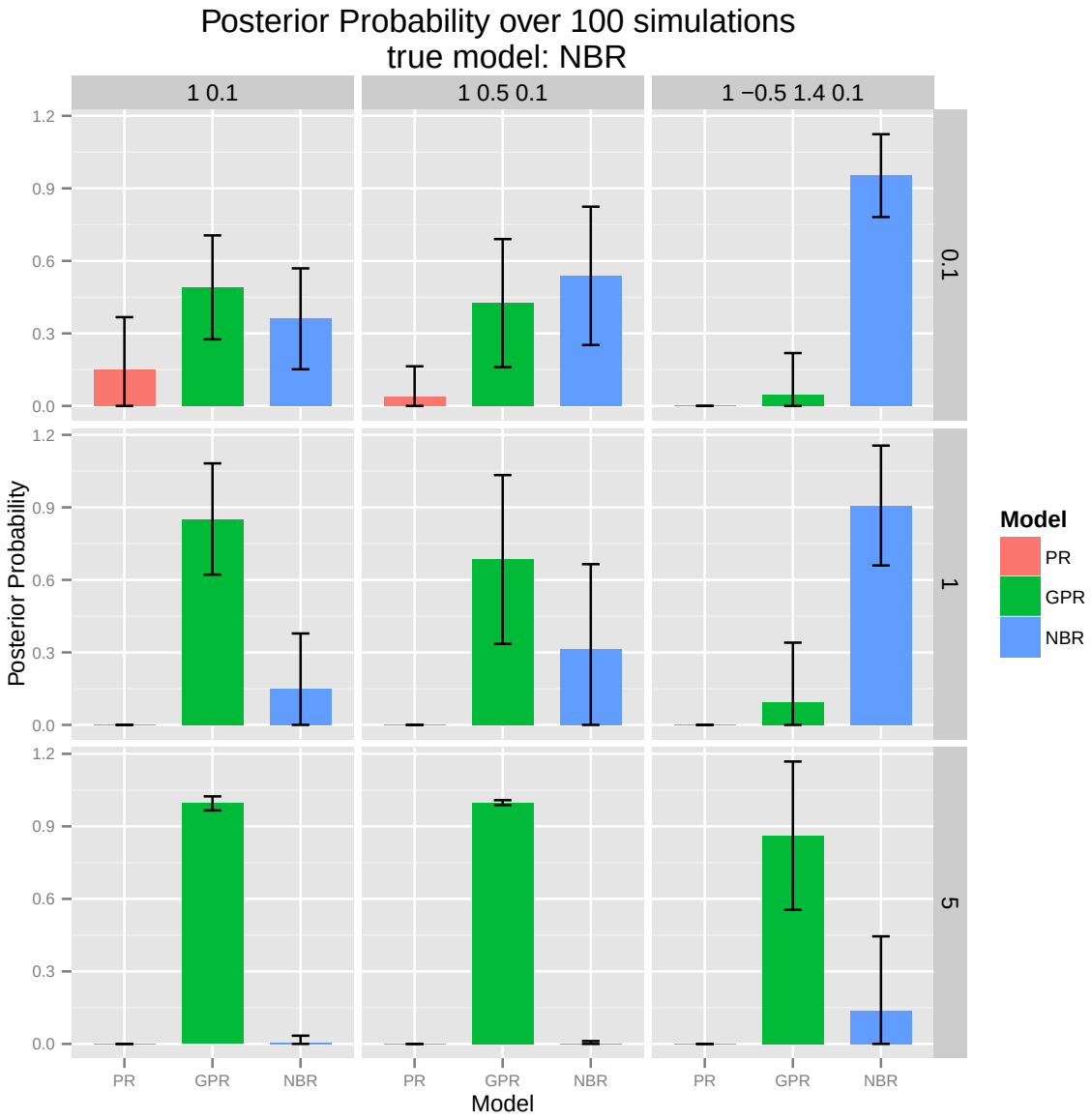
The convergence of the importance sampling is verified by iterating multiple times of importance sampling method over one simulated dataset. The iterated results of posterior probability of models are given in table 5.4.

Figure 5.2: Comparison of posterior probability using uniform priors, true model:PR and GPR



Posterior probability is computed from 100 simulations from a Poisson model($\theta = 0$) or generalized Poisson model ($\theta > 0$). Parameters θ is given on the title of rows and covariate coefficient vector β is given on the title of columns.

Figure 5.3: Comparison of posterior probability using uniform priors, true model:NBR



Posterior probability is computed from 100 simulations from a negative binomial model defined by parameters τ and covariate coefficient β. Parameters τ is given on the title of rows and coefficient vector β is given on the title of columns.

Table 5.4: Importance sampling convergence: FBF

β	θ	τ	$P(\text{PR}\|y)$	$P(\text{GPR}\|y)$	$P(\text{NBR}\|y)$
(1,0.1)	na	1	3.93E-69	0.982643	0.017357
(1,0.1)	na	1	4.41E-69	0.978277	0.021723
(1,0.1)	na	1	5.18E-69	0.978094	0.021906
(1,0.1)	na	1	4.28E-69	0.97742	0.02258
(1,0.1)	na	1	5.26E-69	0.976435	0.023565
(1,0.1)	na	1	5.16E-69	0.968112	0.031888
(1,0.5,0.1)	na	1	3.98E-122	0.975604	0.024396
(1,0.5,0.1)	na	1	1.66E-122	0.991139	0.008861
(1,0.5,0.1)	na	1	2.33E-122	0.968493	0.031507
(1,0.5,0.1)	na	1	6.65E-122	0.959337	0.040663
(1,0.5,0.1)	na	1	7.72E-123	0.990032	0.009968
(1,0.5,0.1)	na	1	4.34E-122	0.983729	0.016271
(1,-0.5,1.4,0.1)	na	1	2.40E-226	2.86E-07	1
(1,-0.5,1.4,0.1)	na	1	5.13E-226	5.74E-07	0.999999
(1,-0.5,1.4,0.1)	na	1	2.07E-226	1.40E-07	1
(1,-0.5,1.4,0.1)	na	1	1.02E-226	6.57E-08	1
(1,-0.5,1.4,0.1)	na	1	8.33E-226	4.97E-07	1
(1,-0.5,1.4,0.1)	na	1	1.18E-226	7.13E-05	0.999929

True model is NBR model. Each row contains posterior probabilities based on one simulation of data size 100 from a model defined by parameters β, θ, τ and covariate matrix X corresponding to β.

5.7 Posterior probability of model using proper non-informative priors

The application of fractional Bayes factor in order to approximate Bayes factor in negative binomial model using improper prior is not accurate enough based on the simulation results. Therefore, a proper non-informative prior for the negative binomial regression model is used to avoid the application of fractional Bayes factor.

The alternative prior for τ of NB regression model is a mixture half-normal distribution, which is a proper prior over the domain. The mixture half-normal prior for τ is defined as

$$\tau \sim \text{Half-Normal}(0, \sigma_\tau^2)$$
$$\sigma \sim \text{Inverse-Gamma}(0.5, 0.5)$$

The density of this mixture half-normal prior is

$$\pi(\tau) = \frac{2}{\sqrt{\pi}\Gamma(0.5)(\tau^2 + 1)}$$

refer to figure 5.4 for how the mixture half-normal distribution. It is more like flat prior with large spread shape and low density over the domain.

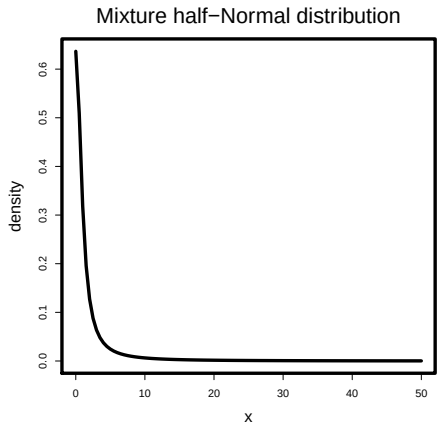


Figure 5.4: The mixture half-normal distribution

The marginal likelihood of NB model using mixture half-normal prior is

$$\begin{aligned} m^{\mathrm{NBR}} &= \int L^{\mathrm{NBR}}(\tau,\beta)\pi(\tau,\beta)d\tau d\beta \\ &= \int \prod_{i=1}^{n}\left(\frac{\Gamma(y_i+1/\tau)(\tau e^{x_i\beta})^{y_i}}{\Gamma(y_i+1)\Gamma(1/\tau)(1+\tau e^{x_i\beta})^{y_i+1/\tau}}\right)\cdot \\ &\frac{2}{\sqrt{\pi}\Gamma(0.5)(\tau^2+1)}\frac{1}{\sigma_\beta(2\pi)^{p/2}}e^{-\frac{1}{2\sigma_\beta^2}\beta'\beta}d\tau d\beta \end{aligned}$$

To approach this complex integral, the importance sampling densities are

$$\begin{aligned} \beta \sim & N(\hat{\mu}_\beta, \hat{\sigma}_\beta^2 \times I_p) \\ \tau \sim & \text{half-Cauchy}(0, 25) \end{aligned}$$

where $\hat{\mu}_\beta$ and $\hat{\sigma}_\beta^2$ are estimated values obtained from Poisson regression model (glm function) to increase the accuracy and efficiency of Importance Sampling.

Now, the importance sampling approach is

$$m^{\mathrm{NBR}} \approx \frac{1}{m}\sum_{j=1}^{m}\frac{\hat{\sigma}_\beta}{\sigma_\beta}exp\left(\frac{1}{2\hat{\sigma}_\beta^2}(\beta_j-\hat{\mu}_\beta)^T(\beta_j-\hat{\mu}_\beta)-\frac{1}{2\sigma_\beta^2}\beta_j^T\beta_j\right)\cdot$$
$$\prod_{i=1}^{n}\left(\frac{\Gamma(y_i+1/\tau_j)(\tau_j e^{x_i\beta_j})^{y_i}}{\Gamma(y_i+1)\Gamma(1/\tau_j)(1+\tau_j e^{x_i\beta_j})^{y_i+1/\tau_j}}\right)\cdot\frac{2}{\sqrt{\pi}\Gamma(0.5)(\tau_j^2+1)}\frac{\pi(\tau_j^2+25^2)}{50}$$

The marginal likelihoods of Poisson and GP model are same as before. Now, Bayes factor is appropriate for testing negative binomial vs. Poisson regression model. To test the hypothsis in NB model,

$$H_0:\tau=0 \quad vs. \quad H_1:\tau>0$$

The Bayes factor is

$$BF_{\mathrm{NBR,PR}}=\frac{m^{\mathrm{NBR}}}{m^{\mathrm{PR}}}$$

In the evaluation of fitting among Poisson, GP, NB regression models, the only change is the marginal likelihood of NB model compared with the fractional Bayes factor method. So the posterior probability of model can be directly computed through the marginal likelihoods given the assumption of equal prior probability of models.

$$P(M_i|y)=\frac{P(y|M_i)P(M_i)}{\sum_{j\in all}P(y|M_j)P(M_j)}=\frac{m_i}{m^{\mathrm{PR}}+m^{\mathrm{GPR}}+m^{\mathrm{NBR}}}$$

where $i\in\{\mathrm{PR,GPR,NBR}\}$

5.8 Simulation results of Bayes Factor

Several simulated data from one of these competitive models are tested by means of the posterior probabilities. The random covariate is taken into consideration for the cases of real world problems. Each model of fixed values of covariate matrix and parameters simulates 100 data sets, and each simulation contains 100 draws. Based on the simulated data sets, the posterior probability of each candidate model is estimated by importance sampling method. The number of iterations of importance sampling is 5000.

The simple condition of two regression coefficients is first studied. Assume there is only one covariate x, which is random variable, and the associated regression coefficient β is 2-dimensional. The covariate matrix X can be written as follows:

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_k \\ 1 & x_{(k+1)} \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

where $x_i \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_i \sim N(\mu_2, \sigma_2^2)\ for\ i = k+1, ..., n$.

Let $\mu_1 = 3, \sigma_1^2 = 1$ and $\mu_2 = 10, \sigma_2^2 = 2$. The results of posterior probability of model are given in the table 5.5.

The posterior probability makes the right choice for the simulated data when it is coming from the Poisson model. For GP and NB regression model, the probability of the right model is high when the model parameter is not close to 0. Because the models are quite close to each other when the model parameter is small, Bayesian method tends to prefer a simple model when the

Table 5.5: Simulation results of one covariate

β	θ	τ	$\bar{P}(\text{PR}\|y)$	$\bar{P}(\text{GPR}\|y)$	$\bar{P}(\text{NBR}\|y)$	$SD[P(\text{PR}\|y)]$	$SD[P(\text{GPR}\|y)]$	$SD[P(\text{NBR}\|y)]$
(1, 0.1)	0	na	0.830869	0.125648	0.043483	0.162043	0.114135	0.053154
(1, 0.1)	0.1	na	0.585887	0.318706	0.095407	0.281781	0.219571	0.091966
(1, 0.1)	0.2	na	0.138739	0.6317	0.229561	0.234575	0.230603	0.182202
(1, 0.1)	0.3	na	0.024038	0.719057	0.256905	0.094837	0.194309	0.183607
(1, 0.1)	0.5	na	9.74E-14	0.763703	0.236297	8.09E-13	0.229019	0.229019
(1, 0.1)	na	0.01	0.820604	0.138167	0.041229	0.126106	0.09309	0.04091
(1, 0.1)	na	0.1	0.123514	0.487383	0.389103	0.224898	0.220002	0.240793
(1, 0.1)	na	1	4.50E-34	0.149191	0.850809	4.50E-33	0.212914	0.212914
(1, 0.1)	na	5	1.78E-144	0.093891	0.906109	1.78E-143	0.180153	0.180153
(1, 0.1)	na	10	1.32E-95	0.077869	0.922131	1.32E-94	0.1491	0.1491

Each row contains posterior probabilities based on 100 simulations from a model defined by parameters β, θ, τ and covariate matrix X. True model is determined by θ and τ, where $\tau =$ na means true model is GPR model, and $\theta =$ na says data is simulated from NBR model. Poisson model has parameter $\theta = 0, \tau = na$ particularly.

performance over the given data of competing models have quite similar effect.

Consider adding one more binary covariate to the model, the covariate matrix X becomes

$$X = \begin{bmatrix} 1 & 0 & x_{31} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{3k} \\ 1 & 1 & x_{3(k+1)} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{3n} \end{bmatrix}$$

where $x_i \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_i \sim N(\mu_2, \sigma_2^2)\ for\ i = k+1, ..., n$.

Let $\mu_1 = 3, \sigma_1^2 = 1$ and $\mu_2 = 10, \sigma_2^2 = 2$. The results of various models are given in the table 5.6. The posterior probability correctly choose the right model even for the small value of τ, because GP and NB model are departing away from each other when the mean μ is larger than 5. For this case, the simulated data has a mean value larger than 5, so it is not affected by the approximation error.

Now, we add one more binary covariate to the model. The covariate matrix becomes 4 by n. In general, it can be expressed in the way of

$$X = \{X_{\cdot(1)}, X_{\cdot(2)}, X_{\cdot(3)}, X_{\cdot(4)}\}_{n \times 3}$$

where the $X_{\cdot(i)}$ is the ith covariate vector associated with the coefficient β_i. For the simulation analysis, let $X_{\cdot(1)} = [1]_{n \times 1}$ be the identity vector for intercept. $X_{\cdot(2)} = (0, ..., 0, 1, ..., 1)^T$ and $X_{\cdot(3)} = (0, ..., 0, 1, ..., 1, 0..., 0)^T$. $X_{\cdot(4)}$ is a random vector with $x_i(4) \sim N(\mu_1, \sigma_1^2)\ for\ i = 1, ..., k$ and $x_i(4) \sim N(\mu_2, \sigma_2^2)\ for\ i = k+1, ..., n$.

Table 5.6: Simulation results of two covariates

β	θ	τ	$\bar{P}(\text{PR}\|y)$	$\bar{P}(\text{GPR}\|y)$	$\bar{P}(\text{NBR}\|y)$	$SD[P(\text{PR}\|y)]$	$SD[P(\text{GPR}\|y)]$	$SD[P(\text{NBR}\|y)]$
(1, 0.5, 0.1)	0	na	0.858807	0.121077	0.020116	0.147187	0.13217	0.024475
(1, 0.5, 0.1)	0.1	na	0.603894	0.33187	0.064237	0.292562	0.252913	0.08637
(1, 0.5, 0.1)	0.2	na	0.155875	0.689751	0.154374	0.221149	0.257588	0.168836
(1, 0.5, 0.1)	0.3	na	0.006046	0.883086	0.110868	0.027004	0.139508	0.136915
(1, 0.5, 0.1)	0.5	na	1.53E-09	0.896272	0.103728	1.53E-08	0.189968	0.189968
(1, 0.5, 0.1)	na	0.01	0.793132	0.17051	0.036358	0.198191	0.1711	0.04145
(1, 0.5, 0.1)	na	0.1	0.048434	0.39943	0.552136	0.135747	0.266953	0.298271
(1, 0.5, 0.1)	na	1	7.37E-55	0.068647	0.931353	7.31E-54	0.164047	0.164047
(1, 0.5, 0.1)	na	5	1.22E-111	0.026392	0.973608	1.22E-110	0.093787	0.093787
(1, 0.5, 0.1)	na	10	3.32E-131	0.019126	0.980874	3.32E-130	0.062631	0.062631

Each row contains posterior probabilities based on 100 simulations from a model defined by parameters β, θ, τ and covariate matrix X. True model is determined by θ and τ, where $\tau =$ na means true model is GPR model, and $\theta =$ na says data is simulated from NBR model. Poisson model has parameter $\theta = 0, \tau = na$ particularly.

Table 5.7: Simulation results of three covariates

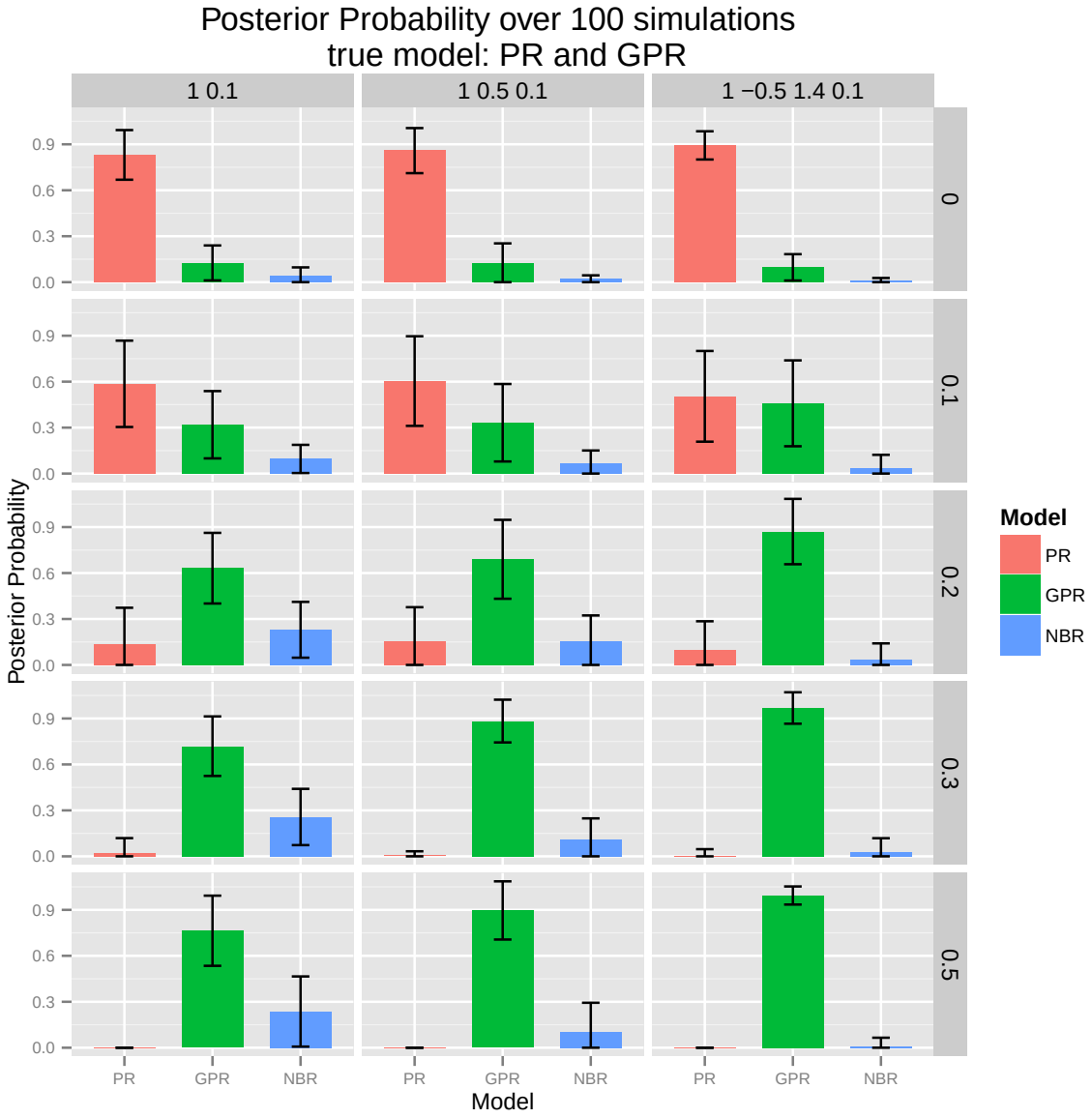
β	θ	τ	$\bar{P}(\text{PR}\|y)$	$\bar{P}(\text{GPR}\|y)$	$\bar{P}(\text{NBR}\|y)$	$SD[P(\text{PR}\|y)]$	$SD[P(\text{GPR}\|y)]$	$SD[P(\text{NBR}\|y)]$
(1, -0.5, 1.4, 0.1)	0	na	0.892932	0.096774	0.010294	0.092092	0.085991	0.017055
(1, -0.5, 1.4, 0.1)	0.1	na	0.504146	0.458505	0.037349	0.29606	0.280313	0.084906
(1, -0.5, 1.4, 0.1)	0.2	na	0.096259	0.870742	0.032998	0.188906	0.213272	0.107843
(1, -0.5, 1.4, 0.1)	0.3	na	0.00478	0.96844	0.026779	0.042008	0.102596	0.091589
(1, -0.5, 1.4, 0.1)	0.5	na	3.32E-19	0.99365	0.00635	3.31E-18	0.058994	0.058994
(1, -0.5, 1.4, 0.1)	na	0.01	0.666889	0.241965	0.091146	0.316791	0.250287	0.160703
(1, -0.5, 1.4, 0.1)	na	0.1	5.65E-06	0.043726	0.956268	5.42E-05	0.156761	0.156763
(1, -0.5, 1.4, 0.1)	na	1	1.06E-138	0.000127	0.999873	1.06E-137	0.000674	0.000674
(1, -0.5, 1.4, 0.1)	na	5	0	3.89E-05	0.999961	0	0.000274	0.000274
(1, -0.5, 1.4, 0.1)	na	10	0	0.000209	0.999791	0	0.001737	0.001737

Each row contains posterior probabilities based on 100 simulations from a model defined by parameters β, θ, τ and covariate matrix X. True model is determined by θ and τ, where $\tau =$ na means true model is GPR model, and $\theta =$ na says data is simulated from NBR model. Poisson model has parameter $\theta = 0, \tau = na$ particularly.

The simulation results of 100 data sets are displayed in table 5.7. The posterior probability shows strong evidence for Poisson model when the data is exactly simulated from it. As the GP parameter θ larger than 0.1, the posterior probability provides strong evidence in favor of GP model. When simulated data coming from NB model, the posterior probability works well for very small value of τ that our method is able to choose the right model with high certainty. Generally speaking, the posterior probability is capable to capture the difference among GPR, NBR and PR models even if the difference is small. It is also accurate about the probability for each model with small variance on the estimation. The comparison of simulation results is also plotted by true model, which refer to figure 5.5 and figure 5.6.

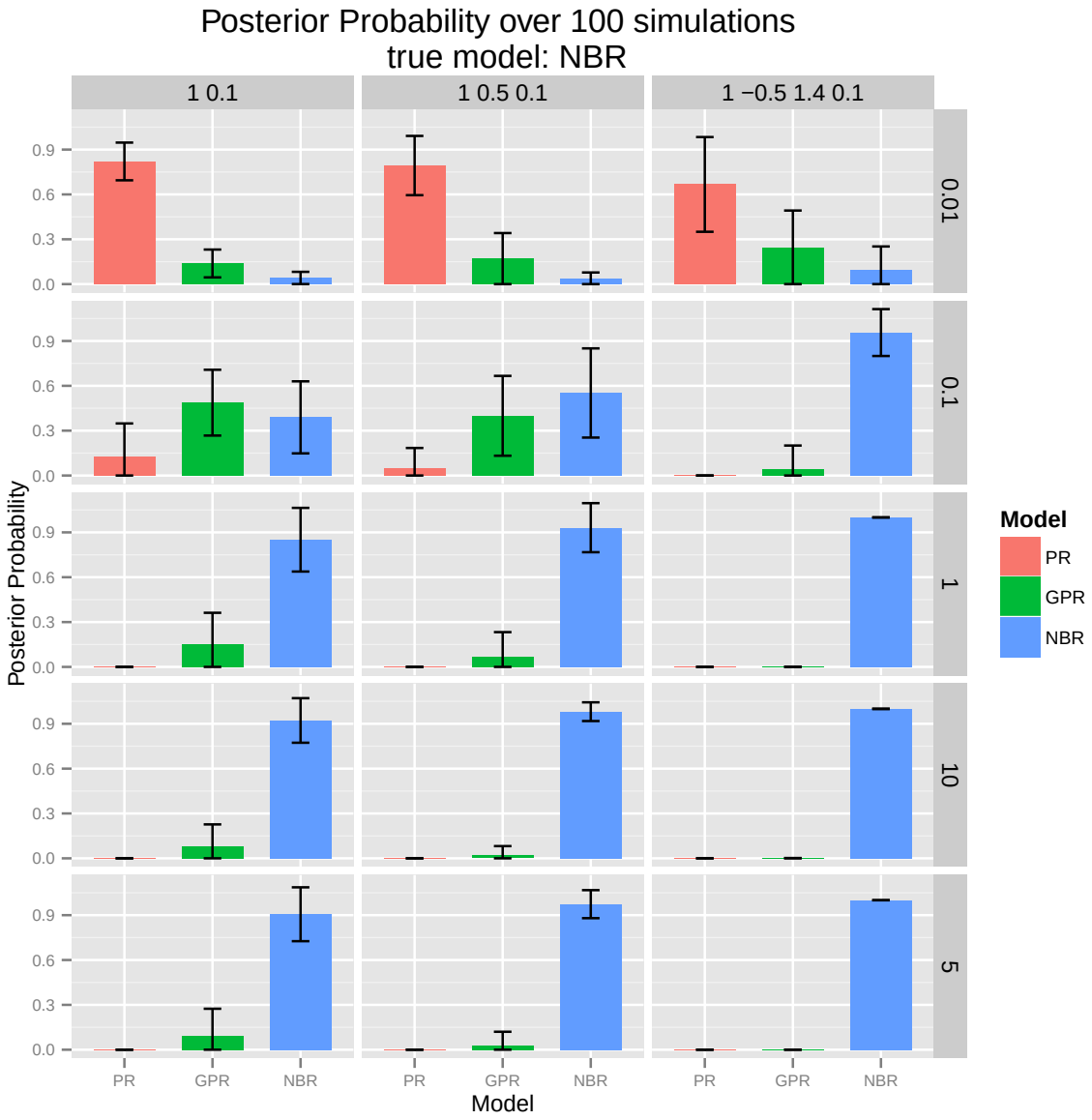
The convergence of importance sampling is verified through multiple iterations of importance sampling approach with different simulation from importance sampling density for one single dataset. Part of the iterations is displayed in table 5.8. It is easy to see the method converges pretty good with small variance for all models.

Figure 5.5: Comparison of posterior probability using flat priors, true model:PR and GPR



Posterior probability is computed from 100 simulations from a Poisson model($\theta = 0$) or generalized Poisson model ($\theta > 0$). Parameters θ is given on the title of rows and covariate coefficient vector β is given on the title of columns.

Figure 5.6: Comparison of posterior probability using flat priors, true model:NBR



Posterior probability is computed from 100 simulations from a negative binomial model defined by parameters τ and covariate coefficient β. Parameters τ is given on the title of rows and coefficient vector β is given on the title of columns.

Table 5.8: Importance sampling convergence: BF

β	θ	τ	$P(\text{PR}\|y)$	$P(\text{GPR}\|y)$	$P(\text{NBR}\|y)$
(1,0.1)	na	0.1	0.8403	0.134647	0.025052
(1,0.1)	na	0.1	0.841722	0.127793	0.030484
(1,0.1)	na	0.1	0.846044	0.130527	0.023429
(1,0.1)	na	0.1	0.838877	0.127948	0.033175
(1,0.1)	na	0.1	0.846603	0.128514	0.024883
(1,0.1)	na	0.1	0.843717	0.129717	0.026566
(1,0.5,0.1)	na	0.1	0.128297	0.804894	0.066809
(1,0.5,0.1)	na	0.1	0.130277	0.735939	0.133784
(1,0.5,0.1)	na	0.1	0.124268	0.736742	0.13899
(1,0.5,0.1)	na	0.1	0.119484	0.749205	0.13131
(1,0.5,0.1)	na	0.1	0.123536	0.772598	0.103866
(1,0.5,0.1)	na	0.1	0.116158	0.750493	0.133349
(1,-0.5,1.4,0.1)	na	0.1	4.79E-12	0.000424	0.999576
(1,-0.5,1.4,0.1)	na	0.1	4.78E-12	0.000638	0.999362
(1,-0.5,1.4,0.1)	na	0.1	2.71E-12	0.000222	0.999778
(1,-0.5,1.4,0.1)	na	0.1	4.17E-12	0.000615	0.999385
(1,-0.5,1.4,0.1)	na	0.1	2.85E-12	0.000254	0.999746
(1,-0.5,1.4,0.1)	na	0.1	2.02E-12	0.000192	0.999808

True model is NBR model. Each row contains posterior probabilities based on one simulation of data size 100 from a model defined by parameters β, θ, τ and covariate matrix X corresponding to β.

5.9 Real data analysis

5.9.1 Nurofibromatosis-2 (NF2) data

A spinal tumor count data for patients with the disease neurofibromatosis 2 (NF2) has been studied by *Joe and Zhu* (2005) and *Yang et al.* (2007) for model selection. The data contains 158 observations of tumor counts from patients. These data are heavily right skewed and have excess zeros at the same time. This data is summarized in Table 5.9.

Table 5.9: Nurofibromatosis-2 (NF2) dataset

Tumor Count	0	1	2	3	4	5	6	7	8	9	10
Frequency	70	13	15	6	7	5	9	9	1	2	5
Tumor Count	11	13	14	15	16	20	21	24	26	30	50
Frequency	1	1	1	1	1	3	1	3	1	1	2

The dispersion index (the ratio of variance to mean) is $D = 14.34792$, so the data are quite dispersed. The zero fraction (the proportion of zeros) is 0.443038, also the data is zero-inflated to great extent. The choice between GP and NB model is not straightforward based on the property of data set. *Joe and Zhu* (2005) studied this data set to show the difference between GP and NB distribution, and for a simple initial analysis, they did not use covariate information at that time. *Yang et al.* (2007) also studied this NF2 count data for testing Poisson, GP and NB regression model with only intercept coefficient in the covariate information. The author uses Vuong statistic to compare the

difference between GP and NB model. They all conclude that NB model is better than GP model for this observed data.

Bayesian posterior probability method is applied to this data, and the covairate $X = [1]_{n\times 1}$ is defined for including just intercept coefficient in the model. Our method is then able to perform on the NF2 data. Results are given in Table 5.10.

Table 5.10: Posterior probability for Nurofibromatosis-2 (NF2) data

$P(\text{PR}\|y)$	$P(\text{GPR}\|y)$	$P(\text{NBR}\|y)$
0	0.1588155	0.8411845

From the results, the posterior probability of NB model is 0.8411845 which is a strong evidence in favor of it. It is consistent with the previous studies, whereas they did not provide the strong evidence. *Joe and Zhu* (2005) calculated AIC of comparative models as $AIC_{\text{ZIGP}} = 743$, $AIC_{\text{NB}} = 744$ and $AIC_{\text{GP}} = 752.8$, which indicates that NB fits the data better than GP model, but zero-inflated GP is better than NB model. Whereas *Yang et al.* (2007) used the Vuong test for making a choice between the NB2 and GP model, and the Vuong test statistic is -2.218 which is less than -1.96 (95% confidence level), indicating the NB2 fits the data better than the GP. *Joe and Zhu* (2005) discussed that zero-inflated generalized Poisson would give a better fitting according to the presence of both over-dispersion and zero-inflation. Our posterior probability gives consistent conclusion with $P(\text{ZIGP}|y) = 0.762480$, $P(\text{GP}|y) = 0.228028$ and $P(\text{NB}|y) = 0.009490$.

5.9.2 Los Angeles school data

The Los Angeles school dataset is selected from two senior high school in the Los Angeles area. It includes several variables, our interest is focus on variables of gender, language NCE test scores, mathematics NCE test scores and days absent. *Yang et al.* (2007) analyzed the relation of days absent with the other variables. The summary of these variables are given in table 5.11. The dependent variable is days absent, others are used as indicator variable. Covariate matrix is a mixture of fixed and random values, where gender is binary variable, and others are continuous variables. The data includes 316 observations with 162 female and 154 male, and the mean of days absent is 5.810127 with variance equal 55.48764. The dispersion index is $D = 9.55$, thus the data are quite dispersed. The zero fraction is 0.1962025, so the data is zero-inflated too.

Table 5.11: Summary statistics for some variables in Los Angeles school data

Variable	Min	Max	Median	Mean	Standard deviation
Absent days	0	45	3	5.8	7.45
Langnce	1.0	99	50.0	50.1	17.94
Mathnce	1.0	99	48.9	48.8	17.88

There are 316 observations with 162 female and 154 male identified by gender. Dispersion index of absent days is $D = 9.55$ and zero fraction of absent days is 0.1962025.

Posterior probability results suggest GP regression model fits this data better than Poisson and NB regression model. *Yang et al.* (2007) compared

Table 5.12: Posterior probability of Los Angeles school data

$P(\text{PR}\|y)$	$P(\text{GPR}\|y)$	$P(\text{NBR}\|y)$
3.708740e-292	0.927237	0.072762

GP against Poisson regression model, they have the score statistic as 97.8164, the Wald statistic as 29.54 and their corresponding p-values are $p < 0.0001$. All the statistics suggest GP regression model is better than Poisson regression model. They tested NB versus Poisson regression model by the score statistic as 95.5677 and the Wald statistic as 10.47, which indicates NB regression model is better than Poisson regression model. They also found the better fitting of GP model with week evidence of Vuong test as 1.415 against NB model, since the score of Vuong test is not larger than 1.96 (95% confidence level), they suggest both models for fitting the data as a matter of similar prediction for the data. Since Vuong test depends on the maximum likelihoods, and the maximum likelihood needs the estimation of model parameters which maximizing the likelihood, it has restriction on the model comparison. Our method does not require any prior information about the comparative models, and it is only based on the basic structure of models with normally non-informative prior on model parameters. Overall, it is more basic and general testing method toward NB, GP, and Poisson regression model.

5.10 Discussion

The method of Bayes factor is limited to the comparison of two models at a time. When there are more than two candidate models, one has to go through all pairwise comparisons which is time consuming. To overcome this limitation, we developed a general Bayesian testing method in this chapter, so that more than two models can be compared simultaneously. Specifically, we sort all candidate models by their respective posterior probabilities computed from the input data, assuming priors are drawn from a non-informative distribution.

Our method serves as a direct guidance in model selection by providing a probability score for each candidate model. When the score of the leading candidate is higher than other candidates by a large margin, we verified that it is always the correct model. When several candidate models have very close scores, the models themselves turn out to be very close (in the sense that NBR and GPR models in Fig. 5.6), and the correct model is always one of them. Since this subset of candidates have almost equal predictive power, picking out a best one becomes a technicality and a matter of taste instead of the highest score model, for example, Occam would choose the one with the least number of parameters in favor of simplicity.

Compared with the frequentist approach, our method is more advantageous because it does not rely on the accuracy of the fitting procedure. In the frequentist approach, one has to compute, for each model, a set of optimal parameters from the input data. When comparing models using these parameters, it is difficult to quantify the effect of their deviation from the true values

(due to user or systematic errors, etc.) on the final result. In the Bayesian method, we allow each parameter of the candidate models to be drawn from a distribution, hence in effect average out such uncertainties. In this sense the Bayesian comparison method is more suitable for finding the proper model with little knowledge about the actual data.

Chapter 6

Bayesian estimation of parameters in generalized Poisson distribution

6.1 Introduction

In this chapter, we consider fitting the generalized Poisson distribution in Bayesian framework with interest in the Jeffreys priors for model parameters.

Generalized Poisson distribution

$$f(y|\alpha,\lambda) = \frac{(1+\alpha y)^{y-1}(\lambda e^{-\alpha\lambda})^y}{e^{\lambda} y!}$$

where $0 \leq \alpha < 1/\lambda$, $0 < \lambda$

First, change the parameters of the distribution by the transformation of $\theta = \alpha\lambda$. Then GP distribution becomes

$$f(y|\theta,\lambda) = \frac{\lambda(\lambda+\theta y)^{y-1} e^{-\theta y-\lambda}}{y!}$$

where $0 \leq \theta < 1$, $0 < \lambda$

Posterior density of joint parameters (θ, λ) of GP distribution is

$$p(\theta, \lambda|y) = \frac{p(y|\theta, \lambda)p(\theta, \lambda)}{p(y)} = \frac{L(\theta, \lambda)\pi(\theta\lambda)}{m(y)}$$

The marginal likelihood $m(y)$ is constant, we can simplify the calculation of posterior distribution by only focus on the form of density.

$$p(\alpha, \lambda|y) \propto L(\alpha, \lambda)\pi(\alpha, \lambda)$$

The likelihood of GP distribution is

$$L^{\mathrm{GP}}(\theta, \lambda) = \prod_{i=1}^{n} f(y_i|\theta, \lambda) = \frac{\lambda^n e^{-s\theta - n\lambda} \prod_{i=1}^{n} (\lambda + \theta y_i)^{y_i - 1}}{\prod_{i=1}^{n} y_i!}$$

The prior distribution of parameters is a key part of Bayesian inference, though a reasonable prior should have minor effect on posterior distribution with large sample size. Without any knowledge of parameters in the model, a non-informative prior is a appropriate choice. We will first discuss posterior inference with a simple set of priors and then Jeffreys priors.

6.2 Posterior distribution using conditionally uniform prior

It is quite common to select non-informative prior as the prior distribution for parameters in a model. As it does not require any information about the

parameter, and also it brings nearly none information to the inference of it. It is also convenient to implement, so let us first try simple conditionally uniform prior. The prior distributions for the parameters in GP distribution are defined as

$$\theta|\lambda \sim \text{Uniform}(0,1)$$

$$\pi(\lambda) \sim 1/\sqrt{\lambda} \qquad \lambda > 0$$

where α is dependent on λ from the definition of its domain in the distribution. The prior for θ is the uniform prior over its domain from 0 to 1, while the prior for λ is the Jeffreys prior derived from the Poisson distribution. The conditionally uniform prior (C-U) for the joint parameters of GP model is

$$\pi_{\text{C-U}}(\theta,\lambda) = 1/\sqrt{\lambda}$$

Posterior density of joint parameters in GP distribution

$$\begin{aligned} p_{\text{C-U}}(\theta,\lambda|y) &\propto L(\theta,\lambda)\pi_{\text{C-U}}(\theta,\lambda) \\ &= \frac{\lambda^{n-1/2}e^{-s\theta-n\lambda}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}}{\prod_{i=1}^{n} y_i!} \end{aligned}$$

The joint distribution is not explicitly tell what we want to know for the parameters. Due to the dependence between two parameters, we can not derive the posterior density for each parameter alone. But the full conditional posterior is available for further analysis which means inference based on simulation study. For each parameter, the full conditional distribution is the distribution of the parameter conditional on all the other parameters and the given data:

$p(\theta_j|\theta_{-j}, y)$

The full conditional posteriors of GP distribution are

$$p_{\text{C-U}}(\theta|\lambda, y) = \frac{p_{\text{C-U}}(\theta, \lambda|y)}{p_{\text{C-U}}(\lambda|y)} \propto \frac{\frac{\lambda^{n-1/2}e^{-s\theta-n\lambda}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}}{\prod_{i=1}^{n} y_i!}}{\int \frac{\lambda^{n-1/2}e^{-s\theta-n\lambda}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}}{\prod_{i=1}^{n} y_i!} d\theta}$$
$$\propto e^{-s\theta}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}$$

$$p_{\text{C-U}}(\lambda|\theta, y) = \frac{p_{\text{C-U}}(\theta, \lambda|y)}{p_{\text{C-U}}(\theta|y)} \propto \frac{\frac{\lambda^{n-1/2}e^{-s\theta-n\lambda}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}}{\prod_{i=1}^{n} y_i!}}{\int \frac{\lambda^{n-1/2}e^{-s\theta-n\lambda}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}}{\prod_{i=1}^{n} y_i!} d\lambda}$$
$$\propto \lambda^{n-1/2}e^{-n\lambda}\prod_{i=1}^{n}(\lambda+\theta y_i)^{y_i-1}$$

The full conditional posteriors are not any simple distribution. Metropolis-Hastings algorithm is applied to achieve simulation of posteriors in use of simulated values of other standard distributions. The proposal density of Metropolis-Hastings algorithm becomes effective with high acceptance rate when it is as close as possible to the target density. Proposal densities for full conditional posteriors using flat prior are defined below

$$\lambda_j \sim \text{Gamma}(shape = s, scale = \frac{\lambda_{j-1}}{s})$$
$$\theta_j \sim \text{Uniform}(\theta_{j-1} - 0.05, \theta_{j-1} + 0.05)$$

The proposal densities are dependent on the previous draw of the parameter, as the purpose of dependence is to improve the acceptance rate in simulation. The mean of proposal density is supposed to be the previous simulated value.

It ensures that the markov chain will quickly converge to the stationary status.

6.3 Posterior distribution using Jeffreys priors

The Jeffreys prior for parameters (α, λ) is

$$\begin{aligned}\pi_J^{\mathrm{GP}}(\alpha, \lambda) &= \pi_J(\alpha|\lambda)\pi_J^{\mathrm{Poi}}(\lambda) \\ &\propto \frac{\lambda}{\sqrt{(1+2\alpha)(1-\alpha\lambda)}}\frac{1}{\sqrt{\lambda}}\end{aligned}$$

where $0 \leq \alpha < 1/\lambda, 0 < \lambda$. Since Jeffreys prior is invariant under reparameterization of parameters $\theta = \alpha\lambda$

The Jeffreys prior for parameters (θ, λ) is

$$\begin{aligned}\pi_J^{\mathrm{GP}}(\theta, \lambda) &= \pi_J^{\mathrm{GP}}(\alpha, \lambda)\left|\frac{d\alpha}{d\theta}\right| \\ &\propto \frac{1}{\sqrt{(\lambda+2\theta)(1-\theta)}}\end{aligned}$$

where $0 \leq \theta < 1, 0 < \lambda$.

Posterior density of joint parameters using Jeffreys priors are

$$\begin{aligned}p_J(\theta, \lambda|y) &\propto L(\theta, \lambda)\pi_J(\theta, \lambda) \\ &= \frac{\lambda^n e^{-s\theta - n\lambda}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}\prod_{i=1}^n y_i!}\end{aligned}$$

The posterior density is too complex to draw statistical inference directly from the equation. Simulation is needed to carry out for obtaining the property of parameters. The full conditional posteriors are derived below for simulation

using Gibbs sampler method.

The full conditional posteriors of GP distribution are

$$p_J(\theta|\lambda,y) = \frac{p_J(\theta,\lambda|y)}{p_J(\lambda|y)} \propto \frac{\frac{\lambda^n e^{-s\theta-n\lambda}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}\prod_{i=1}^n y_i!}}{\int \frac{\lambda^n e^{-s\theta-n\lambda}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}\prod_{i=1}^n y_i!}d\theta}$$
$$\propto \frac{e^{-s\theta}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}}$$

$$p_J(\lambda|\theta,y) = \frac{p_J(\theta,\lambda|y)}{p_J(\theta|y)} \propto \frac{\frac{\lambda^n e^{-s\theta-n\lambda}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}\prod_{i=1}^n y_i!}}{\int \frac{\lambda^n e^{-s\theta-n\lambda}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}\prod_{i=1}^n y_i!}d\lambda}$$
$$\propto \frac{\lambda^n e^{-n\lambda}\prod_{i=1}^n(\lambda+\theta y_i)^{y_i-1}}{\sqrt{(\lambda+2\theta)(1-\theta)}}$$

Proposal densities for full conditional posteriors of GP distribution are defined to draw random numbers on the basis of Metropolis-Hastings algorithm.

$$\lambda_j \sim \text{Gamma}(shape = n, scale = \frac{\lambda_{j-1}}{n})$$
$$\theta_j \sim \text{Uniform}(\theta_{j-1} - 0.05, \theta_{j-1} + 0.05)$$

The proposal densities are dependent on the previous draw of the parameter, as the purpose of dependence is to improve the acceptance rate in simulation.

6.4 Simulation of data from posterior distribution

The simulation of the joint posterior distribution is based on the full conditional posterior densities through the Gibbs sampler. Sampling from the

full conditional posterior is accepting draws from a proposal distribution by means of Metropolis-Hastings algorithm. The simulated values are assessed by a acceptance probability.

The acceptance probability of Metropolis-Hastings algorithm is

$$A(x, x^*) = min(1, \frac{p(x^*|y)q(x|x^*)}{p(x|y)q(x^*|x)})$$

It is a measure of the probability of transition from x to x^*. Here x and x^* respectively denote the recently accepted value and new simulated value from the proposal density. $p(x^*|y)$ and $p(x|y)$ are the posterior density of interest at points x and x^*. $q(x^*|x)$ is the proposal density at x^* conditional on x.

The algorithm of sampling from joint posterior distribution

1. Initiate arbitrary values as the start state, for example $(\lambda_0 = 1, \theta_0 = 0.5)$;
2. for j=1 to m, iterate;
3. draw a random value of both λ and θ variables from proposal distributions;
   $\theta^* \sim q(\theta|\theta_{j-1})$;
   $\lambda^* \sim q(\lambda|\lambda_{j-1})$;
4. draw a random value from uniform distribution: $U \sim \text{Uniform}(0, 1)$
5. accept the draw according to acceptance probability $A(x, x^*)$;
   let $\theta_j = \theta^*$ if $A(\theta_{j-1}, \theta^*|\lambda_{j-1}) \geq U$, or $\theta_j = \theta_{j-1}$, otherwise;
   let $\lambda_j = \lambda^*$ if $A(\lambda_{j-1}, \lambda^*|\theta_j) \geq U$, or $\lambda_j = \lambda_{j-1}$, otherwise;

6. repeat the steps, until reach m draws.

The markov chain should converge quickly to a stationary status, normally burn-in having 500-1000 steps, after that sampling could be treated as the simulated data from the target distribution for analysis purpose. The acceptance probability also relies on the other parameter, so the most recent draw of the other parameter is a good choice for it.

6.5 Simulation results

Simulation is implemented with different settings of parameters in GP distribution. Each simualtion result is based on one simulated data from the same setting, and total number of simulation loops for each setting is $N = 100$ loops. Each loop consists of burn-in stage and stationary stage. Statistical results are then drawn from the stationary data.

For example, given a simulated data from the setting of GP($\theta = 0.5, \lambda = 1$). It is shown in table 6.1. There are 100 simulated numbers in the data, and the data is over-dispersed with large variance. Based on this data, we want to estimate parameters in the GP distribution. By applying Bayesian method, we can get the posterior densities for the parameters with proper prior assumptions.

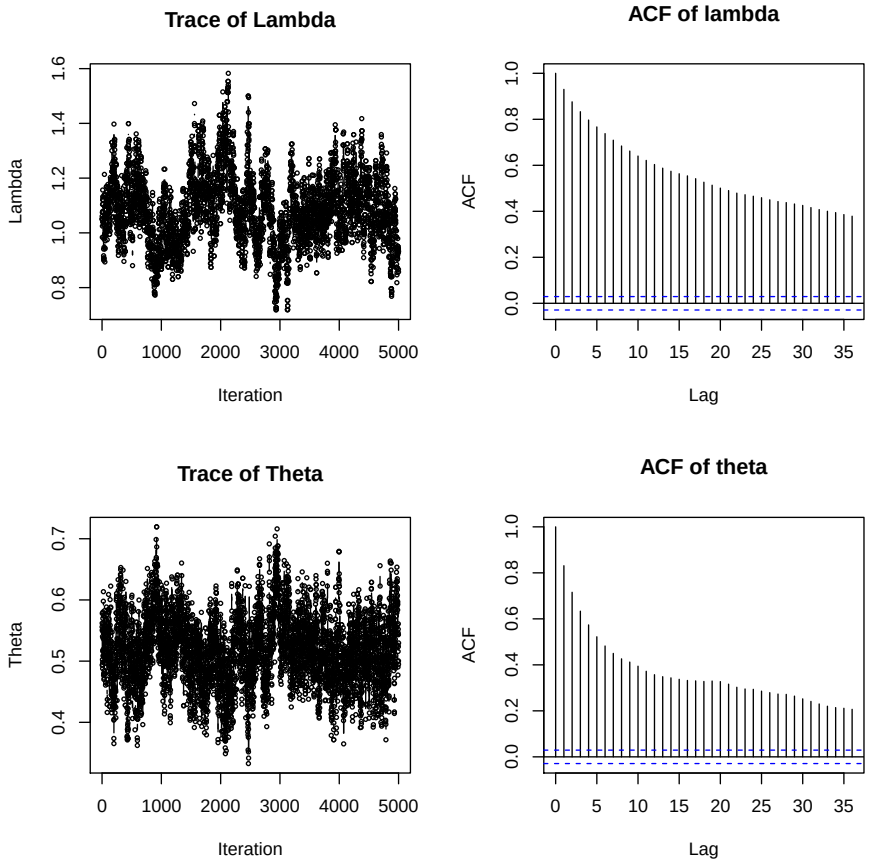
First, choose flat prior for θ and Jeffreys prior for λ. We have the full conditional posteriors of parameters, and then Markov chain Monte Carlo method can be used to obtain the simulated values from the posterior density.

Table 6.1: Simulated data of size 100 from GP($\theta = 0.5, \lambda = 1$)

Y	0	1	2	3	4	5	6	8	9	10	13	19
Frequency	33	24	15	6	8	4	2	3	1	2	1	1

Trace plot of Markov chain is plotted for each parameter in the figure 6.1. From the trace plot, it is easy to see that the chain reaches stationary quickly, and the size of 5000 iterations is enough for the MCMC simulation.

Figure 6.1: MC trace plot of lambda

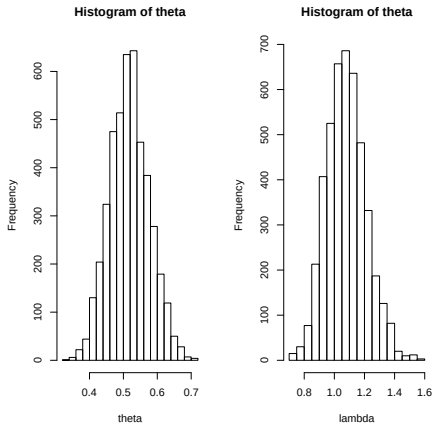


The ACF plots show autocorrelation which is reasonable because of the dependence of the proposal densities used in Markov chain. The improvement of using dependent proposal densities is significant in increasing the acceptance rate of Markov chain in Metropolis-Hastings algorithm. The acceptance rate is calculated by the equation of

$$\rho = \frac{\text{Number of acceptance of draws}}{\text{Total number of draws}}$$

The acceptance rate is the percentage of accepted draws in the Markov chain. A high value of acceptance rate indicates a good proposal density and a quick mixing in Markov chain. In this case, the acceptance rates are $\rho(\theta) \approx 0.80965$ and $\rho(\lambda) \approx 0.69885$.

Figure 6.2: Histogram of MCMC



Histogram plot shows the frequency of the values of parameter in the trace-plot, basically it stands for the posterior distribution of the parameter which

we want to know. The center of the histogram of parameter θ is close to 0.5 which is the true value of θ in the GP distribution, whereas the center of the histogram of parameter λ looks a litte bigger than 1. The actual results for parameters are

Table 6.2: Bayesian inference of flat prior

Parameter	True value	Estimate	95% CI
θ	0.5	0.5156234	(0.4084328, 0.6343092)
λ	1	1.084807	(0.8458981, 1.3584917)

However, the bias in the simulated values from MCMC is appropriate, because the simulated data from the GP distribution could contain bias during the process of simulation. The 95% credible interval correctly capture the true value of target parameters. Based on the results, the MCMC method works well on estimation of GP distribution given the flat prior and proposal densities.

Now, let us adopt Jeffreys priors for both λ and θ parameters. The Markov chain reaches stationary status quickly within 500 steps, the trace plot of parameters λ and θ are displayed in figure 6.3. The trace plot shows the values the parameter took during the runtime of the chain. From the trace plots and ACF plots, the simulated values of parameters are less autoregressive than the chain of flat prior.

Histogram of the simulated values of parameter, given in 6.4, indicates the posterior density of the parameter. From these plots, the results show better

Figure 6.3: MC trace plot of lambda

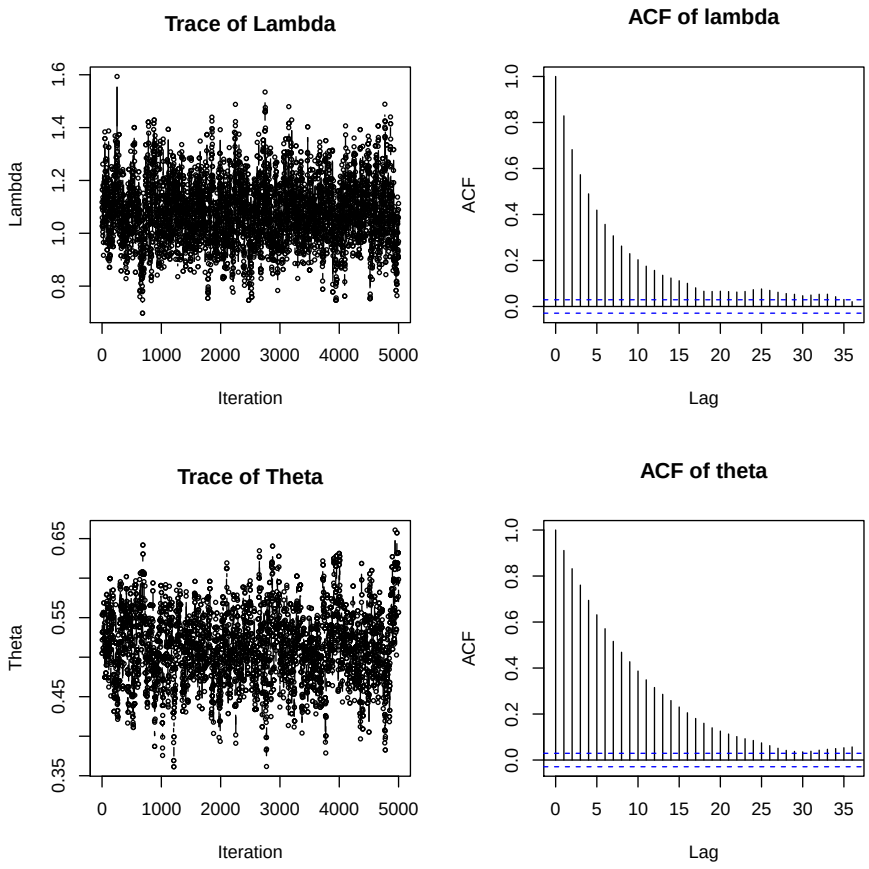
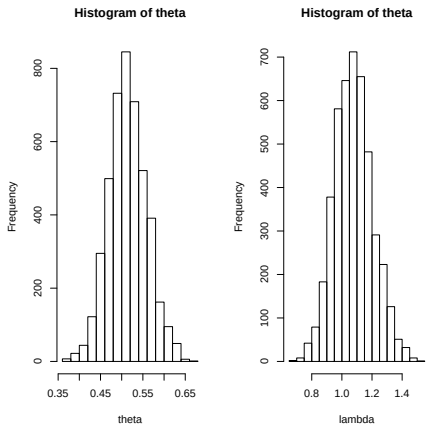


Figure 6.4: Histogram of MCMC



estimation of parameters compared with using flat prior for θ. The center of histogram of θ locates at 0.5. Although the center of λ is still a bit bigger than 1, the variance of the simulated values of λ is smaller than the one of flat prior. The actual results for parameters are

Table 6.3: Bayesian inference of Jeffreys priors

Parameter	True value	Estimate	95% CI
θ	0.5	0.5139565	(0.4295281, 0.6088664)
λ	1	1.077347	(0.8415542, 1.3358935)

The results show 95% CI correctly catch the true value of parameter. Also, the consistency of the center location of λ demonstrate that the bias is most likely intrinsic to the given data. Therefore, using Jeffreys priors give a better estimation of GP distribution. Next, we will implement this MCMC process for a set of 100 data simulated from the same setting of parameters in GP

distribution.

6.5.1 Validation of method

Several sets of parameter values in GP distribution are defined for the purpose of validation of our Bayesian inference method. For instance, $(0.2, 1.0)$, $(0.5, 1.0)$, $(0.8, 1.0)$, $(0.2, 2.0)$, etc. Validation is based on 100 data simulated from each set of parameters, Bayesian inference is drawn from the MCMC method for each simulated data set. The probability of the credible interval covering the true value of parameter is treated as a measure of the accuracy of our method, and it is called coverage probability. The higher the probability of coverage, the more accurate the method is. The variance of estimates is able to be estimated through a series of estimated values of parameter. Also, bias is given for use in comparison with variance. The bias is the difference between expectations of an estimator and the true value of the parameter being estimated. It is calculated by the way of

$$bias(\hat{\theta}) = E(\hat{\theta} - \theta)$$

The coverage probability (CProb) is a measure of accuracy of estimation of parameters. It is calculated by the percentage of estimated 95% CI containing the true parameter value over all 100 simulated data from the same GP distribution.

The results of estimation by Bayesian posterior probability method for a set of parameter values is listed in the table 6.4 for Jeffreys prior and table 6.5.

Each row is the result for one set of parameter values (θ, λ). The result of each row is concluded from Bayesian inference on 100 data simulated from the GP distribution with parameter values given by $(True\theta, True\lambda)$. The 'Est θ' and 'Est λ' are the mean of estimated values of 100 simulated data. 95% CI is calculated by taking average of 2.5th quantile and 97.5th quantile of estimator over all data. Variance, bias and coverage probability of estimator are computed by their definition over estimates of all simulated data.

The results display that both flat prior and Jeffreys prior work well on Bayesian inference of GP distribution. The coverage probability is high for all sets of parameter values, and variance and bias of the estimators are all low. It is quite accurate of using flat prior with very high coverage probability for each set of parameters.

Table 6.4: Bayesian inference simulation results for GP distribution of Jeffreys priors

True θ	Est of θ	95% CI of θ	MSE of θ	Bias of θ	CProb of θ	True λ	Est of λ	95% CI of λ	MSE of λ	Bias of λ	CProb of λ
0.2	0.186324	(0.078847, 0.301847)	0.004094	-0.01368	0.93	1	1.024005	(0.801987, 1.268538)	0.016539	0.024005	0.95
0.5	0.489838	(0.397908, 0.581909)	0.004284	-0.01016	0.82	1	1.02345	(0.797755, 1.274263)	0.010487	0.02345	0.99
0.8	0.784042	(0.724032, 0.845013)	0.002642	-0.01596	0.78	1	1.035772	(0.806162, 1.292035)	0.020924	0.035772	0.9
0.2	0.186372	(0.082147, 0.295136)	0.003847	-0.01363	0.92	2	2.02237	(1.662077, 2.408093)	0.035759	0.02237	0.96
0.4	0.386128	(0.295444, 0.474958)	0.003513	-0.01387	0.87	2	2.044773	(1.679573, 2.440108)	0.043411	0.044773	0.96
0.8	0.791689	(0.745495, 0.838022)	0.001156	-0.00831	0.87	2	2.040212	(1.66824, 2.444824)	0.041642	0.040212	0.94
0.3	0.296443	(0.197836, 0.391616)	0.002139	-0.00356	0.98	3	3.044564	(2.549659, 3.578656)	0.079954	0.044564	0.95
0.6	0.600124	(0.537718, 0.661684)	0.001853	0.000124	0.85	3	3.009035	(2.517303, 3.533387)	0.08722	0.009035	0.92
0.2	0.189497	(0.085609, 0.294157)	0.003	-0.0105	0.94	4	4.043036	(3.426539, 4.694808)	0.109726	0.043036	0.96
0.4	0.381829	(0.295783, 0.465228)	0.002853	-0.01817	0.9	4	4.0401	(3.42134, 4.701237)	0.110048	0.0401	0.95
0.25	0.245631	(0.142706, 0.345038)	0.002483	-0.00437	0.96	5	5.077634	(4.320703, 5.888748)	0.184523	0.077634	0.94
0.5	0.484138	(0.412397, 0.554049)	0.001876	-0.01586	0.89	5	5.15756	(4.403152, 5.956752)	0.17775	0.15756	0.95
0.75	0.746595	(0.705619, 0.787115)	0.000769	-0.0034	0.86	5	5.047137	(4.292411, 5.839008)	0.214592	0.047137	0.94

Table 6.5: Bayesian inference simulation results for GP distribution of flat prior

True θ	Est of θ	95% CI of θ	MSE of θ	Bias of θ	CProb of θ	True λ	Est of λ	95% CI of λ	MSE of λ	Bias of λ	CProb of λ
0.2	0.198667	(0.077937, 0.338856)	0.004	-0.00133	0.95	1	1.012334	(0.787381, 1.26057)	0.015744	0.012334	0.95
0.5	0.496663	(0.379774, 0.621665)	0.004036	-0.00334	0.97	1	1.01599	(0.787427, 1.272246)	0.010187	0.01599	0.99
0.8	0.786445	(0.703449, 0.873078)	0.002491	-0.01355	0.91	1	1.02976	(0.800673, 1.288493)	0.02009	0.02976	0.91
0.2	0.194987	(0.082523, 0.318473)	0.003507	-0.00501	0.95	2	2.004648	(1.635166, 2.396859)	0.03355	0.004648	0.96
0.4	0.391139	(0.284233, 0.501612)	0.003339	-0.00886	0.95	2	2.034271	(1.655937, 2.447815)	0.042048	0.034271	0.97
0.8	0.793005	(0.730868, 0.856822)	0.001134	-0.007	0.95	2	2.034092	(1.653584, 2.450025)	0.042239	0.034092	0.94
0.3	0.301506	(0.193528, 0.411887)	0.002066	0.001506	0.99	3	3.027795	(2.511242, 3.582625)	0.077048	0.027795	0.97
0.6	0.601955	(0.524492, 0.681025)	0.001849	0.001955	0.91	3	3.003197	(2.489212, 3.555658)	0.087293	0.003197	0.92
0.2	0.199059	(0.088313, 0.313287)	0.002687	-0.00094	0.98	4	3.9984	(3.35973, 4.671121)	0.101093	-0.0016	0.97
0.4	0.38512	(0.288046, 0.48304)	0.002681	-0.01488	0.93	4	4.023151	(3.367803, 4.72317)	0.106892	0.023151	0.98
0.25	0.253474	(0.145374, 0.362279)	0.002247	0.003474	0.97	5	5.028189	(4.241684, 5.854769)	0.166786	0.028189	0.96
0.5	0.486393	(0.402818, 0.569595)	0.001758	-0.01361	0.95	5	5.141365	(4.333711, 6.005367)	0.1703	0.141365	0.98
0.75	0.745414	(0.693748, 0.797463)	0.000764	-0.00459	0.95	5	5.08043	(4.282299, 5.935896)	0.232832	0.08043	0.95

6.6 Real data analysis

6.6.1 Nurofibromatosis-2 (NF2) data

A spinal tumor count data for patients with the disease neurofibromatosis 2 (NF2) has been studied by *Joe and Zhu* (2005) for distribution selection of Poisson-related distributions. The data contains 158 observations of tumor counts from patients. The detail about NF2 tumor count data is given in table 6.6

Table 6.6: Nurofibromatosis-2 (NF2) dataset

Tumor Count	0	1	2	3	4	5	6	7	8	9	10
Frequency	70	13	15	6	7	5	9	9	1	2	5
Tumor Count	11	13	14	15	16	20	21	24	26	30	50
Frequency	1	1	1	1	1	3	1	3	1	1	2

This data is summarized in Table 6.7. These data are heavily right skewed and have excess zeros at the same time. This is data has been tested over generalized Poisson regression model, negative binomial regression model, and Poisson regression model. Our testing shows the negative binomial regression model fits NF2 data more properly.

Now, this data is going to be used to demonstrate the accuracy of our Bayesian method. *Joe and Zhu* (2005) studied the distribution testing without covariate information, and also suggests negative binomial distribution is better than generalized Poisson distribution. In the paper, the author provided

Table 6.7: Summary statistics for Nurofibromatosis-2 (NF2)

Min	Max	Median	Mean	Variance
0	50	1	4.34	62.4

the values of estimation of fitted generalized Poisson distribution by maximum likelihood estimation method.

Table 6.8: Estimation of GP distribution for Nurofibromatosis-2 (NF2)

Prior	$\hat{\theta}$	$\hat{\lambda}$	SE of $\hat{\theta}$	SE of $\hat{\lambda}$
Jeffreys prior	0.7888138	0.9183445	0.02595482	0.09603467
non-informative prior	0.7879164	0.938116	0.0360353	0.09041671
MLE	0.789	0.913	0.036	0.095

Bayesian method for model estimation is implemented on the observed NF2 tumor data. The results based on Jeffreys prior and non-informative prior are listed in table 6.8, whereas, in the table, the maximum likelihood estimation is obtained from paper *Joe and Zhu* (2005). From the results, it is obvious that Bayesian method gives similar estimation compared to maximum likelihood estimation. In this case, Jeffreys prior works slightly better than flat prior in accuracy aspect, but they perform equivalently well in general.

6.7 Discussion

We present the Bayesian inference using Jeffreys prior for generalized Poisson distribution. Our interest is on the effect of Jeffreys prior on Bayesian estimation of generalized Poisson distribution. The Bayesian inference based on Jeffreys prior is compared with the Bayesian inference using conditionally uniform prior through simulation study. The study shows they both provide accurate estimates with high coverage probability for distribution parameters, but conditionally uniform prior has slightly higher coverage probability of true parameters than Jeffreys prior. A real data is studied to evaluate the Bayesian method by comparing the estimation with maximum likelihood estimation. In the real data case, GP distribution is fitted through our Bayesian method and the parameter estimation gives a evidence that Bayesian inference using either Jeffreys prior or conditionally uniform prior provides the results consistent with the frequentist inference by the maximum likelihood estimation.

Our Bayesian inference of generalized Poisson distribution help understand the difference between Jeffreys prior and conditionally uniform prior. Both provide accurate estimation of model parameters. Conditionally uniform prior is simple and easy to understand, while Jeffreys prior has its benefit of invariance under reparameterization of the parameter. The choice of prior between Jeffreys prior and conditionally uniform prior is made up to researcher's interest.

Bibliography

Bibliography

Angers, J.-F., and A. Biswas (2003), A bayesian analysis of zero-inflated generalized poisson model, *Computational statistics & data analysis*, *42*(1), 37–46.

Bayarri, M., J. O. Berger, G. S. Datta, et al. (2008), Objective bayes testing of poisson versus inflated poisson models, in *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, pp. 105–121, Institute of Mathematical Statistics.

Bhattacharya, A., B. S. Clarke, G. S. Datta, et al. (2008), A bayesian test for excess zeros in a zero-inflated power series distribution, in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, pp. 89–104, Institute of Mathematical Statistics.

Consul, P., and F. Famoye (1992), Generalized poisson regression model, *Communications in Statistics-Theory and Methods*, *21*(1), 89–109.

Consul, P. C. (1989), *Generalized Poisson distributions: properties and applications*, M. Dekker.

Consul, P. C., and G. C. Jain (1973), A generalization of the poisson distribution, *Technometrics*, *15*(4), 791–799.

Gupta, P. L., R. C. Gupta, and R. C. Tripathi (1996), Analysis of zero-adjusted count data, *Computational Statistics & Data Analysis*, *23*(2), 207–218.

Gupta, P. L., R. C. Gupta, and R. C. Tripathi (2005), Score test for zero inflated generalized poisson regression model, *Communications in Statistics-Theory and Methods*, *33*(1), 47–64.

Joe, H., and R. Zhu (2005), Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution, *Biometrical Journal*, *47*(2), 219–229.

Lambert, D. (1992), Zero-inflated poisson regression, with an application to defects in manufacturing, *Technometrics*, *34*(1), 1–14.

Lerner, B., A. Lone, and M. Rao (1997), On generalized poisson distributions, *PROBABILITY AND MATHEMATICAL STATISTICS-WROCLAW UNIVERSITY*, *17*, 377–385.

O'Hagan, A. (1995), Fractional bayes factors for model comparison, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 99–138.

Tuenter, H. J. (2000), On the generalized poisson distribution, *Statistica Neerlandica*, *54*(3), 374–376.

Xie, F.-C., B.-C. Wei, and J.-G. Lin (2009), Score tests for zero-inflated generalized poisson mixed regression models, *Computational Statistics & Data Analysis*, *53*(9), 3478–3489.

Yang, Z., J. W. Hardin, C. L. Addy, and Q. H. Vuong (2007), Testing approaches for overdispersion in poisson regression versus the generalized poisson model, *Biometrical journal*, *49*(4), 565–584.