Objective Bayes Testing of Poisson versus Inflated Poisson Models

M.J. Bayarri
University of Valencia

J.O. Berger
Duke University and SAMSI

G.S. Datta
University of Georgia

Abstract

The Poisson distribution is often used as a standard probability model for data involving counts. Quite often, however, such datasets are not well fit by a Poisson model because they contain more zero counts than are compatible with the Poisson model. For these situations, a zero-inflated Poisson (ZIP) distribution is often proposed. This article addresses testing a Poisson versus a ZIP model, using Bayesian methodology based on suitable objective priors. Specific choices of objective priors are justified and their properties investigated. The methodology is extended to testing Poisson versus ZIP regression models which include covariates. Several applications are given.

1 Introduction

The Poisson distribution is often used as a standard probability model for count data. For example, a production engineer may count the number of defects in items randomly selected from a production process. Quite often, however, such datasets are not well fit by a Poisson model because they contain more zero counts than are compatible with the Poisson model. An example is again provided by the production process; indeed, according to Ghosh et al. (2006), when some production processes are in a near perfect state, zero defects will occur with a high probability. However, random changes in the manufacturing environment can lead the process to an imperfect state, producing items with defects. The production process can move randomly back and forth between the perfect and the imperfect states. For this type of production process many items will be produced with zero defects, and this excess might be better modeled by a zero-inflated Poisson distribution than a Poisson distribution.

The zero-inflated Poisson $\mathrm{ZIP}(\lambda, p)$ distribution is given by the probability mass function

$$f_1(x \mid \lambda, p) = p\, I(x=0) + (1-p)\, f_0(x \mid \lambda), \quad x = 0, 1, 2, \dots\ , \tag{1}$$

where $0 \le p \le 1$, $\lambda > 0$, $I(\cdot)$ is the indicator function, and

$$f_0(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2, \dots\ , \tag{2}$$

is the Poisson probability function. The parameter p is referred to as the *zero-inflation parameter*.

The Poisson distribution is a special case of the the power series distribution and the ZIP distribution is a special case of the zero-inflated power series distribution. See Johnson, Kotz and Kemp (1992, pp. 312-318) for discussion of the zero-inflated power series distribution.

There are many references in the literature that utilize the ZIP distribution to model count data, with and without covariates. An example of a frequentist analysis in an industrial application can be found in Lambert (1992). Ghosh et al. (2006) used a Bayesian approach to fit a ZIP regression model to another industrial dataset.

While the aforementioned authors used the ZIP model to analyze their data, a number of authors have addressed the problem of checking whether a standard ZIP model is needed to model the data. From the frequentist perspective, Broek (1995) and Deng and Paul (2000) developed score tests for testing the hypothesis $\mathcal{H}_0 : p = 0$ vs. $\mathcal{H}_1 : p \neq 0$ in a ZIP regression model. From the Bayesian perspective, Bhattacharya et al. (2007) presented a Bayesian method to test $p \leq 0$ versus the alternative $p > 0$ by computing a certain posterior probability of the alternative hypothesis. As in Broek (1995) and Deng and Paul (2000), Bhattacharya et al. (2007) allow p to be negative in their model, as long as $p + (1 - p)e^{-\lambda} \geq 0$.

In this paper, we consider Bayesian testing of

$$M_0 : \; X_i \overset{i.i.d.}{\sim} f_0(\cdot \mid \lambda), \; i = 1, \ldots, n, \tag{3}$$

versus

$$M_1 : \; X_i \overset{i.i.d.}{\sim} f_1(\cdot \mid \lambda, p), \; i = 1, \ldots, n, \tag{4}$$

where f_0, f_1 are the probability functions of the Poisson and ZIP distributions as given in (1) and (2) respectively. Note that, as opposed to the situations in the papers mentioned above, values of $p < 0$ are not possible here. Indeed, we can alternatively formulate the problem as that of testing, within the ZIP model,

$$\mathcal{H}_0 : p = 0 \quad \text{versus} \quad \mathcal{H}_1 : p > 0.$$

Unlike the analysis in Bhattacharya et al. (2007), $p = 0$ (i.e., the Poisson model) is assumed to have priori believability (e.g. prior probability 1/2).

In Section 2 we develop the suggested objective testing of Poisson versus ZIP models when at least one of the counts is non zero. If all of the counts are zero, the ZIP distribution is not identifiable, and proper priors are required for all parameters; we address this case in Section 5. Section 3 is devoted to some comparative examples. We consider inclusion of covariates in Section 4, where we address the testing of Poisson versus ZIP regression models and give an example involving AIDS deaths. Most of the Proofs and technical details are relegated to an Appendix.

2 Formulation of the problem

When choosing between two models for the data, the Bayesian methodology is conceptually very simple (see, for instance, Berger 1985): one assesses prior probabilities of each model, prior distributions for the model parameters, and computes the posterior probabilities of each model. These posterior probabilities can be computed directly from the prior probabilities and the *Bayes Factor*, an (integrated) likelihood ratio for the models which is becoming a very popular report in Bayesian testing and model selection.

In many situations, it is not possible (for lack of time or resources) to carefully assess in a subjective manner all the needed priors. In these situations, very satisfactory answers are provided by *objective Bayesian analyses*, that is, Bayesian analyses that do not use external information other than that required to formulate the problem (see Berger 2006). In Section 2.1 we review the basics of Bayesian model selection and some difficulties with objective Bayes analysis. In Section 2.2 we justify the objective prior we choose for testing the Poisson versus the ZIP models and study some of its properties; Section 2.3 derives the corresponding Bayes Factor and its properties.

2.1 Bayesian model selection and Bayes factors

Suppose we are comparing two models, M_0 and M_1 for the data $\boldsymbol{X} = (X_1, \ldots, X_n)$,

$$M_i : \boldsymbol{X} \text{ has density } \; f_i(\boldsymbol{x} \mid \boldsymbol{\theta}_i),$$

where $\boldsymbol{\theta}_i$ are unknown model parameters in model M_i, $i = 0, 1$. Let $\pi_i(\boldsymbol{\theta}_i)$ denote the prior density of $\boldsymbol{\theta}_i$ under model M_i, and $m_i(\boldsymbol{x})$ denote the marginal or predictive density of $\boldsymbol{X}$ under the model M_i, that is,

$$m_i(\boldsymbol{x}) = \int f_i(\boldsymbol{x} \mid \boldsymbol{\theta}_i)\, \pi_i(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i, \;\; i = 0, 1 \; .$$

For given prior model probabilities $Pr(M_0)$ and $Pr(M_1) = 1 - Pr(M_0)$, the posterior probability of, say, M_0 can be expressed as:

$$Pr(M_0 \mid \boldsymbol{x}) = \left[1 + B_{10} \, \frac{Pr(M_0)}{Pr(M_1)} \right]^{-1} \quad , \tag{5}$$

where B_{10} is the *Bayes factor* (or integrated likelihood ratio) of M_1 to M_0 given by

$$B_{10} = \frac{m_1(\boldsymbol{x})}{m_0(\boldsymbol{x})} = \frac{\int f_1(\boldsymbol{x} \mid \boldsymbol{\theta}_1) \pi_1(\boldsymbol{\theta}_1) d\boldsymbol{\theta}_1}{\int f_0(\boldsymbol{x} \mid \boldsymbol{\theta}_0) \pi_0(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0} \; . \tag{6}$$

It is becoming common practice in Bayesian hypothesis testing to report Bayes factors since one can then simply use (5) and compute ones own posterior probabilities (assuming one agrees with the prior densities $\pi_i(\boldsymbol{\theta}_i)$). In objective Bayes analyses $\pi_i(\boldsymbol{\theta}_i)$ is chosen in an objective or conventional fashion and the hypotheses would be assumed to be equally likely a priori, that is $Pr(M_0) = Pr(M_1) = 1/2$.

Use of objective priors have a long history in Bayesian inference (see, for example, Berger and Sun, 2008, and Kass and Wasserman, 1996 for justifications and references). They are, however, typically improper (they integrate to infinity), and so are only defined up to an arbitrary multiplicative constant. This is not a problem in deriving the posterior distribution, since the same constant appears in both the numerator and denominator of Bayes theorem and so cancels. In model selection and hypothesis testing, however, it can be seen from (6) that when at least one of the priors $\pi_i(\boldsymbol{\theta}_i)$ is improper, the arbitrary constant does not cancel, so that the Bayes factor (and hence the posterior probabilities) is then arbitrary (and hence not defined). An important exception to this arises in invariant situations for parameters occurring in all of the models; Berger et al. (1998) show that use of the (improper) right Haar invariant prior is then permissible.

There are several ways to address this difficulty. One possibility is to try to directly 'fix' the Bayes factor by appropriately choosing the multiplicative constant, as in Ghosh and Samanta (2002). Popular methods (the *intrinsic Bayes factor* of Berger and Pericchi, 1996, and the *fractional Bayes factor* of O'Hagan, 1995) for fixing this constant arise as a consequence of 'training' the improper priors into proper priors based on part of the data or of the likelihood. We refer to Berger and Pericchi (2001) for a review, references and comparisons. Another possibility is to directly derive appropriate 'objective' but proper distributions $\pi_i(\boldsymbol{\theta}_i)$ to use in model selection. This venue was pioneered by Jeffreys (1961); see Bayarri and Garcia-Donato (2007) for methods and references. This is the approach taken in this paper (with a slight exception in Section 5.1.2).

2.2 Specification and justification of the objective priors

Returning to the testing of the Poisson (M_0) versus the ZIP (M_1) models, that is, testing

$$\begin{aligned} M_0 &: X_i \overset{i.i.d.}{\sim} f_0(\cdot \mid \lambda),\ i = 1, \ldots, n, \\ M_1 &: X_i \overset{i.i.d.}{\sim} f_1(\cdot \mid \lambda, p),\ i = 1, \ldots, n, \end{aligned} \tag{7}$$

the key issue is choice of the priors $\pi_0(\lambda)$ and $\pi_1(\lambda, p) = \pi_1(\lambda)\,\pi_1(p \mid \lambda)$.

A frequent simplifying procedure (both for subjective and objective methods) is to take $\pi_0(\lambda)$ equal to $\pi_1(\lambda)$, that is, to give the same prior to parameters occurring in all models under consideration. This, however, may be inappropriate, since λ might have entirely different meanings under model M_0 and under model M_1; the fact that we have used the same label does not imply that they have the same meanings. This frequent mistake is discussed, for example, in Berger et al. (1998).

Jeffreys (1961) and Kass and Vaidyanathan (1992) argue that, if the common parameters are *orthogonal* to the remaining parameters in each model (that is, the Fisher information matrix is block diagonal), then they can be assigned the same prior distribution. In this case, improper priors can be used, since the arbitrary constant would cancel in the Bayes factor.

Unfortunately, p and λ in the ZIP (M_1) model are not orthogonal, so we first reparameterize the original model as follows. We rewrite $f_1(x \mid \lambda, p)$ as

$$f_1^*(x \mid \lambda, p^*) = p^* I(x = 0) + (1 - p^*) f^T(x \mid \lambda), \quad x = 0, 1, 2, \ldots, \tag{8}$$

where $p^* = p + (1 - p)e^{-\lambda}$, and $f^T(x \mid \lambda)$ is the zero-truncated version of the standard Poisson distribution with parameter λ. Note that $p^* \geq e^{-\lambda}$. We can trivially express the Poisson (M_0) model as:

$$f_0^*(x \mid \lambda) = e^{-\lambda} I(x = 0) + (1 - e^{-\lambda}) f^T(x \mid \lambda), \quad x = 0, 1, 2, \ldots, \tag{9}$$

and now it can intuitively be seen that λ has the same meaning in both f_1^* and f_0^*. Indeed the Fisher Information matrix for p^* and λ can be checked to be diagonal.

After making an orthogonal reparameterization, Jeffreys (1961) recommended (i) using the *Jeffreys prior* (square root of Fisher information) for the 'common' parameters; (ii) using a reasonable proper prior for the extra parameters in the more complex model.

The situation here is very unusual, however, in that the Jeffreys prior for the 'common' λ is different for each model. The *Jeffreys prior* for λ in the Poisson model is well known to be $\pi_J^0 = 1/\sqrt{\lambda}$, whereas the Jeffreys prior for the orthogonalized ZIP model is easily shown to be the same as the Jeffreys prior for the truncated distribution $f^T(x \mid \lambda)$, which is

$$\pi_J^1(\lambda) = \frac{k(\lambda)}{\sqrt{\lambda}} \ ,$$

where

$$k(\lambda) = \frac{\{1 - (\lambda + 1)e^{-\lambda}\}^{1/2}}{1 - e^{-\lambda}} \ .$$

That these priors are different after orthogonalization is highly unusual and can be traced to the fact that λ also enters into the definition of the nested model, through $p^* = e^{-\lambda}$. In any case, we are left without clear guidance as to whether π_J^0 or π_J^1 should be used as the prior for λ. (Note that, in computing the Bayes factor, the same prior for λ must be used in both the numerator and the denominator; otherwise one is facing the indeterminacy issues discussed earlier.)

Under the orthogonalized ZIP model, we also need to specify a proper prior for p^* given λ, which we propose to take uniform over the interval $(e^{-\lambda}, 1)$, that is:

$$\pi_1(p^* \mid \lambda) = \frac{I(e^{-\lambda} < p^* \leq 1)}{1 - e^{-\lambda}} \ .$$

We can thus write the overall priors being considered for the two models $f_0^*(x \mid \lambda)$ and $f_1^*(x \mid \lambda, p^*)$ as, respectively,

$$\pi_0^l(\lambda) = \frac{k(\lambda)^l}{\sqrt{\lambda}}, \quad \pi_1^l(\lambda, p^*) = \frac{k(\lambda)^l}{\sqrt{\lambda}} \frac{I(e^{-\lambda} < p^* \leq 1)}{1 - e^{-\lambda}} ,$$

where l is 0 or 1 as we utilize one or the other of the two Jeffreys priors for λ.

It is computationally more convenient to work in the original (p, λ) parameterization. A change of variables above then results in the priors

$$\pi_0^l(\lambda) = \frac{k(\lambda)^l}{\sqrt{\lambda}}, \quad \pi_1^l(\lambda, p) = \frac{k(\lambda)^l}{\sqrt{\lambda}} \; I(0 < p \leq 1) \; , \tag{10}$$

which we will henceforth consider (for l equal to 0 or 1).

We are not aware of any desiderata that would suggest a preference for either the $l = 0$ or $l = 1$ priors, but luckily the two yield almost the same answers. Indeed, it can be checked by simple algebra that $k(\lambda)$ is a strictly increasing function of λ and that

$$\inf \; k(\lambda) = \frac{1}{\sqrt{2}} = 0.71 \; , \quad \text{and} \quad \sup \; k(\lambda) = 1. \tag{11}$$

Thus $k(\lambda)$ is quite flat as a function of λ, so that $k(\lambda)^1$ and $k(\lambda)^0 = 1$ are very similar. An immediate consequence for the Bayes factors B_{10}^l, $l = 0, 1$ is that

$$B_{10}^0/\sqrt{2} \; \leq B_{10}^1 \leq \sqrt{2}\, B_{10}^0 \; ,$$

so that the two Bayes factors can only differ by a modest amount (and in practice the difference is much smaller than this).

It is obviously a bit simpler to work with the $l = 0$ prior, so we drop the l superscript and henceforth utilize the prior

$$\pi_0(\lambda) = \frac{1}{\sqrt{\lambda}}, \quad \pi_1(p, \lambda) = \frac{1}{\sqrt{\lambda}} \; I(0 < p \leq 1) \; . \tag{12}$$

2.3 Objective Bayes factor for Poisson versus ZIP models

Recall that the model M_0 is the standard Poisson model and the model M_1 is the ZIP model. For a sample of n counts $X_1, \ldots, X_n$, let $\boldsymbol{X}$ denote the sample, $k = \sum_{i=1}^n I(X_i = 0)$ be the number of zero counts, and $s = \sum_{i=1}^n X_i$ be the total count. Note that $k = n$ is equivalent to $s = 0$. For given data $\boldsymbol{x}$, the densities $f_0(\boldsymbol{x} \mid \lambda)$ and $f_1(\boldsymbol{x} \mid \lambda, p)$ under the two models are given by

$$f_0(\boldsymbol{x} \mid \lambda) = \frac{e^{-n\lambda}\lambda^s}{\prod_{i=1}^n x_i!}, \quad f_1(\boldsymbol{x} \mid \lambda, p) = \frac{[p + (1-p)e^{-\lambda}]^k (1-p)^{n-k} e^{-(n-k)\lambda}\lambda^s}{\prod_{i=1}^n x_i!} \; .$$

For $s > 0$ (i.e., the counts are not all zero),

$$m_0(\boldsymbol{x}) = \int f_0(\boldsymbol{x} \mid \lambda)\pi_0(\lambda)d\lambda = \frac{\Gamma(s+\frac{1}{2})}{n^{s+\frac{1}{2}}\prod x_i!}\,.$$

Using the binomial expansion of $[p + (1-p)e^{-\lambda}]^k$,

$$\begin{aligned} m_1(\boldsymbol{x}) &= \int f_1(\boldsymbol{x} \mid \lambda, p)\pi_1(p,\lambda)dp\; d\lambda \\ &= \frac{1}{\prod x_i!}\sum_{j=0}^{k}\frac{k!}{j!(k-j)!}\int_0^\infty\int_0^1 p^j(1-p)^{n-j}e^{-(n-j)\lambda}\lambda^{s-\frac{1}{2}}dpd\lambda \\ &= \frac{k!}{(n+1)!\prod x_i!}\sum_{j=0}^{k}\frac{(n-j)!}{(k-j)!}\Gamma(s+\frac{1}{2})(n-j)^{-(s+\frac{1}{2})}. \end{aligned}$$

Both $m_0(\boldsymbol{x})$ and $m_1(\boldsymbol{x})$ are finite and the Bayes factor $B_{10}(\boldsymbol{x}) = m_1(\boldsymbol{x})/m_0(\boldsymbol{x})$ is

$$B_{10}(\boldsymbol{x}) = \frac{k!}{(n+1)!}\sum_{j=0}^{k}\frac{(n-j)!}{(k-j)!}(1-\frac{j}{n})^{-(s+1/2)}\,. \tag{13}$$

Note that, as intuitively expected, for any given n the Bayes factor is increasing in s (the number of zero's) for any fixed k (total count), and is increasing in k for any fixed s. We use this expression to calculate the Bayes factors for the three examples discussed in Section 3.

When $s = 0$ or equivalently all counts are zero ($\boldsymbol{x} = \mathbf{0}$), there is a problem. While $m_0(\mathbf{0}) = \Gamma(1/2)/\sqrt{n}$ remains finite, it is easy to see that $m_1(\mathbf{0})$ is infinite. Indeed for *any* prior of the form $h(p)\pi(\lambda)$, where $\pi(\lambda)$ is improper and $h(\cdot)$ is a proper density (as is required for testing), the marginal density $m_1(\mathbf{0})$ will be infinite. This is because, for $\boldsymbol{x} = \mathbf{0}$, the density $f_1(\boldsymbol{x} \mid \lambda, p) \geq p^n$ implying $m_1(\mathbf{0}) \geq \int_0^1 p^n h(p)dp \int_0^\infty \pi(\lambda)d\lambda = \infty$. We discuss what to do for this case in Section 5.

3 Applications

In this section we apply the methodology to three datasets to detect if zero-inflation is present in the data. These examples have been analyzed for zero-inflation previously using both frequentist and Bayesian procedures. Since there are non zero counts in all three examples, the Bayes factors are computed using (13).

Example 1. The first dataset is the Urinary Tract Infection (UTI) data used in Broek (1995), which used a score test to detect zero-inflation in a Poisson model. 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 as X. The data are recorded in Table 1. Merely by looking at the data it is apparent that zero-inflation is present.

X	0	1	2	3	Total
Frequency	81	9	7	1	98

Table 1: UTI Data

Equation (13) yields a Bayes factor $B_{10} = 223.13$ in favor of model M_1 versus model M_0; if the models were believed to be equally likely apriori, the resulting posterior model probabilities would be $Pr(M_1 \mid \boldsymbol{x}) = 0.995$ and $Pr(M_0 \mid \boldsymbol{x}) = 0.005$. This is indeed strong evidence in favor of the ZIP model.

In Bayesian testing of $\mathcal{H}_0 : p \leq 0$ versus $\mathcal{H}_1 : p > 0$, Bhattacharya et al. (2007) obtained $Pr(p > 0 \mid \boldsymbol{x}) = .999$. Broek (1995) reported the observed value of the score statistic as 15.34, yielding $p-$value 0.0001. All three analyses present strong evidence in favor of the ZIP model, but notice that the p-value seems to suggest stronger evidence against the Poisson null than the Bayesian analysis, and the point null Bayesian analysis suggests weaker evidence than the interval Bayesian test.

Example 2. The next dataset we consider is the Terrorism data from Conigliani, Castro and O'Hagan (2000). Table 2 gives the number of incidents of international terrorism per month (X) in the United States between 1968 and 1974. It is not intuitively clear whether or not there is zero-inflation in this data set.

The Bayes factor here is $B_{10} = 0.28$, yielding objective posterior probability $Pr(M_1 \mid \boldsymbol{x}) = 0.219$, which actually supports the Poisson model. The analysis of Bhattacharya et al. (2007) found $Pr(p > 0 \mid \boldsymbol{x}) = 0.507$, an indeterminate value. The observed value of the score statistic is 0.04, with a $p-$value 0.83. Conigliani et al (2000) test a Poisson null model against a nonparametric alternative, finding a fractional Bayes factor B^F_{10} of 0.0089 of the nonparametric alternative to the Poisson; the apparent strength of this conclusion, compared with the other results, is rather puzzling.

Example 3. The third data set we analyzed is the Cholera data first analyzed by McKendrick (1926). Table 3 shows the number of patients per household suffering

X	0	1	2	3	4	Total
Frequency	38	26	8	2	1	75

Table 2: Terror Data

X	0	1	2	3	4	Total
Frequency	168	32	16	6	1	223

Table 3: Cholera Data

from cholera in a village in India in the 1920's. Again, the observed high zero counts relative to the positive counts strongly suggests zero-inflation. While the Bayes factor is $B_{10} = 238090$, very strong evidence for zero-inflation, similarly strong evidence was reported by Bhattacharya et al. (2007) through the statistic $P(p > 0 \mid \boldsymbol{x}) = .9999$. The observed value of the score statistic is 30.56, effectively giving a $p-$value of 0.

4 Model selection in ZIP regression

Many applications involve count data where covariate information is available; see, for example Lambert (1992) and Ghosh et al. (2006). In this section we consider selection between Poisson regression and ZIP regression models, M_0^R and M_1^R respectively, given by

$$M_0^R : \quad X_i \overset{ind}{\sim} Poisson(\lambda_i), \; i = 1, \ldots, n, \tag{14}$$

$$M_1^R : \quad X_i \overset{ind}{\sim} ZIP(\lambda_i, p), \; i = 1, \ldots, n, \tag{15}$$

where the λ_i follow the log-linear relationship

$$log(\lambda_i) = a_{0i} + \boldsymbol{a}_i^T \boldsymbol{\beta},$$

with a_{0i} being a known off-set variable, $\boldsymbol{a}_i$ being a $q \times 1$ vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^T$. We assume that the matrix $\boldsymbol{A}^T = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ is of rank q. Let k denote the number of zero counts in the data, $0 \leq k \leq n$. For simplicity of notation, we index the observations in such a way that the first k counts are the zeros.

4.1 Objective priors for model selection

Generalizing the argument in Section 2.2 to the regression case is easy in one case, but difficult in the other. If we choose to base the analysis on the Jeffreys prior for $\boldsymbol{\beta}$ under the Poisson regression model M_0^R, the generalization is straightforward: the Jeffreys prior is easily computed as

$$\pi_0^R(\boldsymbol{\beta}) = |\sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T|^{1/2}. \tag{16}$$

Note that this prior is positive since the rank of $\boldsymbol{A}$ is q. Also utilizing this prior for $\boldsymbol{\beta}$ under model M_1^R, along with the independent uniform prior for p, results in the following priors to be utilized to compute B_{10}:

$$\pi_0^0(\boldsymbol{\beta}) = |\sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T|^{1/2}, \quad \pi_1^0(\boldsymbol{\beta}, p) = |\sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T|^{1/2} I(0 < p \leq 1)\,. \tag{17}$$

The generalization to the regression case of the second prior considered in Section 2.2 is much more difficult, because the Jeffreys prior under the ZIP regression model is very complicated. In Section 2.2, the derivation of the corresponding Jeffreys prior was essentially done by ignoring the non-zero counts, utilizing only the truncated Poisson distribution. This suggests modifying (16) by removing the terms corresponding to the zero counts, resulting in

$$\pi_1^R(\boldsymbol{\beta}) = |\sum_{i=k+1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T|^{1/2}. \tag{18}$$

From another intuitive perspective, the zero accounts arising from the inflation factor are clearly irrelevant in fitting the log linear model to the λ_i and, since we do not know which zero counts arise from the inflation factor, dropping them all from the Jeffreys prior has appeal.

The resulting overall prior for use in computing B_{10} is then

$$\pi_0^1(\boldsymbol{\beta}) = |\sum_{i=k+1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T|^{1/2}, \quad \pi_1^1(\boldsymbol{\beta}, p) = |\sum_{i=k+1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T|^{1/2} I(0 < p \leq 1)\,. \tag{19}$$

The first basic issue in use of these priors is whether or not they yield finite marginal distributions. This is addressed in the following theorems, the first of which deals with the marginal density under the Poisson regression model.

Theorem 4.1. For the Poisson regression model and either the Jeffreys prior ($j = 0$) or modified Jeffreys prior ($j = 1$),

$$m_0^R(\boldsymbol{x}) = \int_{R^q} \prod_{i=1}^{n} \{\frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}\} \pi_j^R(\boldsymbol{\beta}) d\boldsymbol{\beta} < \infty. \tag{20}$$

Proof. See the Appendix.

Note that, when there is more than one covariate, there is typically no closed-form expression for $m_0^R(\boldsymbol{x})$, which thus needs to be evaluated by numerical or Monte Carlo integration.

For the ZIP regression model, the marginal density $m_1^R(\boldsymbol{x})$, under an arbitrary improper prior $\pi(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$ and an independent uniform prior for p, is given by

$$m_1^R(\boldsymbol{x}) = \int_{R^q} \int_0^1 f_1(\boldsymbol{x} \mid \boldsymbol{\beta}, p)\, \pi(\boldsymbol{\beta})\, dp\ d\boldsymbol{\beta}, \tag{21}$$

where the density of $\boldsymbol{x}$, under model M_1^R, is given by

$$f_1(\boldsymbol{x} \mid \boldsymbol{\beta}, p) = \prod_{i=1}^{k} \{p + (1-p)e^{-\lambda_i}\}(1-p)^{n-k} \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}.$$

Again, as for $m_0^R(\boldsymbol{x})$, there is usually no closed-form expression for $m_1^R(\boldsymbol{x})$ and the marginal needs to be computed via numerical or Monte Carlo integration.

To investigate the finiteness of $m_1^R(\boldsymbol{x})$, note first that

$$p^k(1-p)^{n-k} \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!} \leq f_1(\boldsymbol{x} \mid \boldsymbol{\beta}, p) \leq \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}\,. \tag{22}$$

In view of this inequality and the independent uniform prior for p, the marginal $m_1^R(\boldsymbol{x})$ is finite if and only if

$$\int_{R^q} \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\ \lambda_i^{x_i}}{x_i!}\ \pi(\boldsymbol{\beta})\ d\boldsymbol{\beta} < \infty\,. \tag{23}$$

Theorem 4.2 below gives sufficient conditions for this to be finite under the priors (16) and (18) respectively. For use in the theorem, recall that we have k zero counts in the sample and they correspond to the first k observations. Also define $\boldsymbol{A}_+ = (\boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n)^T$. A key condition will be that this matrix has rank q which implies that $n \geq k+q$ (analogous to the condition of at least one positive count for the case of no covariate treated in Section 2).

Theorem 4.2.
Using $\pi_0^R(\boldsymbol{\beta})$*:* Suppose that, for the observation $X_j, j = 1, \ldots, k$, corresponding to the zero counts, the corresponding covariate vector $\boldsymbol{a}_j$ is such that

$$\boldsymbol{a}_j = \sum_{m=k+1}^{n} c_{mj}\,\boldsymbol{a}_m \ \text{ with }\ c_{mj} \geq 0, \ \text{ for }\ j = 1, \ldots, k, \ \text{ and } m = k+1, \ldots, n. \tag{24}$$

Then the marginal $m_1^R(\boldsymbol{x})$ is finite.

Using $\pi_1^R(\boldsymbol{\beta})$*:* If $\boldsymbol{A}_+$ has rank q, the marginal $m_1^R(\boldsymbol{x})$ is finite.

Proof. See the Appendix.

Clearly the condition under which $m_1^R(\boldsymbol{x})$ is finite is more general and much easier to check for $\pi_1^R(\boldsymbol{\beta})$ than for $\pi_0^R(\boldsymbol{\beta})$. This, together with the intuitive appeal of $\pi_1^R(\boldsymbol{\beta})$, leads us to recommend its use in practice. (Note that either of the two priors reduces to the prior recommended in Section 2 for the non-regression case.)

Remark 4.1. If the sufficient condition (24) does not hold, the marginal density $m_1^R(\boldsymbol{x})$ based on the Jeffreys prior may be infinite. For example, consider $n = 3$ and $q = 2$, with

$\lambda_1 = \lambda_2^{c_1}\lambda_3^{c_2}$, $\lambda_2 = exp(\beta_1)$, $\lambda_3 = exp(\beta_2)$ for suitable nonzero c_1, c_2 to be chosen later. It can be checked that the determinant of information matrix for $\boldsymbol{\beta}$ is then given by

$$|\boldsymbol{I}(\boldsymbol{\beta})| = \lambda_2\lambda_3 + c_1^2\lambda_2^{c_1}\lambda_3^{c_2+1} + c_2^2\lambda_2^{c_1+1}\lambda_3^{c_2}\,,$$

so that $|\boldsymbol{I}(\boldsymbol{\beta})|^{1/2} \geq |c_1|\lambda_2^{c_1/2}\lambda_3^{(c_2+1)/2}$. If a sample yields the values $X_1 = 0$, $X_2 = x_2$ and $X_3 = x_3$, then

$$\begin{aligned} m_1^R(\boldsymbol{x}) &\geq \frac{|c_1|}{2}\int_{R^2} \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}\frac{e^{-\lambda_3}\lambda_3^{x_3}}{x_3!}\lambda_2^{c_1/2}\lambda_3^{(c_2+1)/2}d\boldsymbol{\beta} \\ &= \frac{|c_1|}{x_2!x_3!2}\int_0^\infty e^{-\lambda_2}\lambda_2^{x_2-1+.5c_1}d\lambda_2\int_0^\infty e^{-\lambda_3}\lambda_3^{x_3-1+.5c_2+.5}d\lambda_3 = \infty\,, \end{aligned}$$

providing that $x_2 \leq -.5c_1$ or that $x_3 \leq -.5 - .5c_2$. For example, if $c_1 = -5$ and a sample produces $x_2 = 2$, then $m_1^R(\boldsymbol{x}) = \infty$. Note that here $\boldsymbol{a}_1 = -5\boldsymbol{a}_2 + c_2\boldsymbol{a}_3$, with $\boldsymbol{a}_2 = (1,0)^T$ and $\boldsymbol{a}_3 = (0,1)^T$, so that the condition (24) does not hold.

4.2 An illustrative application

We apply the methodology recommended in Section 4.1 to a dataset involving the number of deaths in men suffering from AIDS. The data provides the number of deaths for 598 census tracts in a large city of Spain over a period of eight years. The dataset, which was supplied to us by Dr. M.A.M. Beneyto, has a large number of tracts with zero deaths (actually, 303, which is k in our notation). Along with the number of deaths, the dataset also provides, for each census tract, the expected number of deaths E from AIDS (adjusting for the population and the distribution of ages in each tract) and an auxiliary variable W (continuous in nature) measuring the social status of each census tract.

In our application and for the ith census tract, we take $log(E_i)$ as the offset a_{0i} and propose a log-linear regression for λ_i with $q = 2$ and $\boldsymbol{a}_i = (1, W_i)^T$. First, we will ignore the covariate W and compute the Bayes factor taking $q = 1$ and $\boldsymbol{a}_i = 1$ based on the Jeffreys' prior. This model modifies the common mean model of Section 2.2 by incorporating the offset variable in the mean, which is here given by $E_i\lambda$ with $log\lambda = \beta_1$. The Bayes factor B_{10} including the offset has a rather cumbersome form and is given in the Appendix. For the specific data here, $B_{10} = 22,975$ which gives overwhelming evidence in favor of the ZIP model.

Epidemiologists who are knowledgeable about this study believed that the large number of zero counts in the data could be explained by the covariate measuring the social status and, indeed, suspected that a ZIP regression model would not be needed if the covariate were incorporated into the analysis. The Bayes factor in favor of the ZIP regression model versus the Poisson regression model (with $q = 2$) is given by 7.25. While this Bayes factor provides a moderate amount of evidence in favor of the ZIP regression model, it is much smaller than $22,975$, indicating that, indeed, the covariate can explain most of the excess zero counts.

5 Analysis with insufficient positive counts

5.1 All zero counts in the non-regression case

As noted in Section 2, the marginal density under model M_1 based on an improper prior for λ is not finite when all counts are zeros, and hence the Bayes factor is not well-defined. This is not a difficulty of only model selection; in this situation, it is also not possible to make inferences about the parameters of the ZIP model, since the joint posterior of the parameters (under the ZIP model) is improper. Indeed, when all counts are zero, the ZIP model parameters are not identifiable, and the data do not provide enough information to estimate the parameters. Since objective Bayes methods are typically based on information from the data alone, it is not surprising that problems are encountered.

We could simply invoke this argument and refrain from considering the case when all of the counts x_i are zero. However, it is interesting to explore several methodologies that have been proposed for difficult testing situations, partly to judge the success of the methodologies and partly to try to provide a reasonable answer to the case $\boldsymbol{x} = \mathbf{0}$. We continue, throughout the section, to assume that $p \sim Un(0, 1)$.

5.1.1 Directly choosing a proper prior for λ

If a proper prior is needed to define the Bayes factor for the situation of all zero counts, the most direct approach is to find a proper prior that seems compatible with certain behavior that we expect of the Bayes factor in this situation. A natural proper prior to consider for λ is a Gamma $Ga(a, b)$ conjugate prior under the Poisson model (M_0) given by the gamma $g(\lambda \mid a, b)$ density

$$g(\lambda \mid a, b) = \frac{b^a e^{-b\lambda} \lambda^{a-1}}{\Gamma(a)},$$

where a, b are suitably chosen positive constants. Of course, one is welcome to simply make subjective choices here, but we will argue for a certain choice (or choices) based on rather neutral thinking.

First, we assume that the *same* gamma prior is appropriate for λ, both under the Poisson and the ZIP models. This can be justified by the orthogonalization argument used in Section 2.1. With the uniform density for p and the $Ga(a, b)$ prior for λ, the resulting Bayes factor for arbitrary data $\boldsymbol{x}$ can be computed to be

$$B_{10}(\boldsymbol{x}) = \frac{k!}{(n+1)!} \sum_{j=0}^{k} \frac{(n-j)!}{(k-j)!} \left(1 - \frac{j}{n+b}\right)^{-(s+a)}, \tag{25}$$

by a similar argument to that leading to (13). This Bayes factor includes as an special case the objective Bayes factor in (13); indeed the Jeffrey's prior used there was a limiting case of the $g(\lambda \mid a, b)$ for $a = 1/2$ and $b = 0$. Note that the Bayes factor (25) is increasing in s, k and a, and decreasing in b.

For the special case $\boldsymbol{x} = \mathbf{0}$ (that is $s = 0$ and $k = n$), it can be checked that

$$B_{10}(\mathbf{0}) = \frac{(n+b)^a}{n+1} \sum_{j=0}^{n} \frac{1}{(j+b)^a} \geq 1\,. \tag{26}$$

This is reasonable: when a long stream of *only* zeros is observed, it is entirely natural to say that the data favor the ZIP model. But the degree of favoritism depends on a and b, and we turn to rather speculative desiderata to narrow the choice. Recall that the mean of the gamma(a, b) distribution for λ is ab^{-1} and the variance is ab^{-2}.

In order for the prior not to be too sharp, it is reasonable to require the prior standard deviation to be at least as large as the prior mean. This implies $a \leq 1$. It also seems reasonable to require the prior mean to be at least 1, so that small values of λ do not receive excessive prior probability. This leads to $b \leq a$. Since the Bayes factor is decreasing in b, the smallest Bayes factor satisfying the above constraints (that is, the one lending the most support for the Poisson model M_0) is then obtained by taking $b = a$ (this gives a prior mean 1); it is not unreasonable to select that prior from a reasonable class which is most favorable to the null model. Finally, one might judge it to be unappealing to utilize a prior for λ which is not bounded near zero (for $a < 1$ the gamma density is decreasing with an asymptote at $\lambda = 0$) which implies that a should be at least 1. Thus we end up with the choice $a = b = 1$. Note that $a = 1$ is the upper limit of $a \leq 1$ and the choice $a = 1$ now counterbalances the Bayes factor in favor of M_1 (whereas $b = a$ in the range $b \leq a$ tilts the Bayes factor in favor of M_0). This reasoning is all rather speculative and, of course, the result is a particular prior, which may not reflect actual prior beliefs. Nevertheless it is instructive to study the behavior of the Bayes factor when this prior is used.

For $a = b = 1$, that is, the Exponential(1) distribution, it can be checked that

$$B_{10} = \sum_{j=0}^{n} \frac{1}{j+1}\,,$$

which is thus our recommended default Bayes factor when observing only zero counts. Note that $B_{10}(\mathbf{0}) \approx log(n+1)$ for large n; indeed $\sum_{j=0}^{n} 1/(j+1) \leq 1 + \sum_{j=1}^{n} \int_{j}^{j+1} dx/x =$

$1 + log(n+1)$. Also, $\sum_{j=0}^{n} 1/(j+1) \geq \sum_{j=0}^{n-1} \int_j^{j+1} dx/(x+1) = log(n+1)$. Thus B_{10} is bounded between $\log(n+1)$ and $\log(n+1)+1$. So a *large* string of all zero counts in a sample will lead to a Bayes factor approaching infinity at the slow rate of $\log(n)$. The large sample behavior of the Bayes factor for this type of sample seems intuitively reasonable.

5.1.2 Training the improper prior

Another approach to the problem of obtaining a default proper prior is the intrinsic Bayes factor (IBF) approach of Berger and Pericchi (1996). This approach is based on the utilization of training samples or, more precisely, minimal training samples. A training sample is a portion of the data used to convert an improper prior to a proper posterior, which can then be used to combine with the remaining data to calculate the marginal density (of the remaining data). A minimal training sample (MTS) is the smallest sample for which the posterior (based on the MTS) is proper. So that the marginal density, and hence the Bayes factor, does not depend on a particular MTS selected, Berger and Pericchi (1996) recommended averaging the Bayes factors over all possible MTS's. A related alternative is the fractional Bayes factor of O'Hagan (1995).

Developing training samples for mixture models (as in the ZIP model) is not as clear as in many other situations, as was discussed in Pérez and Berger (2001). Since the first component of the mixture does not involve any parameters and the inflation parameter p has a proper distribution, following their recommendation here would result in the minimal training sample being a single observation, considered to be from the Poisson component of the mixture. This was independently suggested by Professor J.K. Ghosh (2006). Thus, we update the improper prior $\pi_1^I(p,\lambda) = 1/\lambda^{1/2}$ to a proper posterior by treating one of the zeros as coming from the Poisson(λ) distribution *under model* M_1. The resulting posterior, that is the 'trained' prior, is then

$$\pi_1(\lambda, p) = 2(1-p)e^{-\lambda}\lambda^{-1/2}/\Gamma(1/2)\,.$$

(Note that now the data is $x_1 = 0$ and "x_1 comes from the Poisson component".) This corresponds to assuming that, independently, $\lambda \sim Ga(1/2, 1)$ and $p \sim Beta(1,2)$. The prior mean for λ ia now 0.5 (and not 1 as before) and the prior mean for p is 1/3 (and not 1/2 as before). We utilize the the same $Ga(1/2, 1)$ prior for λ under model M_0 (noting that this prior also results from a training sample consisting of a 0 under model M_0).

Utilizing these prior specifications for the $n-1$ zero's left in the sample, we compute the Bayes factor $B_{10}(\mathbf{0})$ to be

$$B_{10} = \frac{2}{n+1}\sum_{j=0}^{n-1}(1-\frac{j}{n})^{1/2}\,.$$

Similarly to the results in Section 5.1.1, it is easy to see that $B_{10}(\mathbf{0}) \geq 1$. However, for large n the Bayes factor is approximately $2\int_0^1 (1-u)^{1/2}du = 4/3$, which only slightly favors the ZIP model. This result is much different, and intuitively less convincing, than the $\log n$ behavior seen in the previous subsection. The discrepancy perhaps arises from the rather artificial 'assignment' of the training sample to the Poisson part of the ZIP model.

TO JIM: THE BAYES FACTOR WHEN ONE OF THE COUNTS IS NOT ZERO AND THE REST IS ZERO IS

$$\frac{1}{n+1}\sum_{j=0}^{n-1}(1-\frac{j}{n})^{-1/2} \approx \int_0^1 x^{-1/2} = 2 \ \ \textbf{for large} \ \ n$$

I'D LIKE TO INCLUDE HERE A PICTURE OF THE BAYES FACTORS FOR THE 'CONVENIENT PROPER PRIOR', THE TRAINING DATA PRIOR, AND THE PRIOR FOR ONE COUNT OF 1. ALSO, IF RELEVANT THE ALTERNATIVE TRAINED PRIOR OF THE NEXT PARAGRAPH. I'LL DO THAT

TO JIM: THIS IS THE PART THAT I AM NOT SO SURE TO INCLUDE; THE TRAINING IS NOT SO CLEAR TO ME, THE BAYES FACTOR IS NOT CLOSE FORM AND FOR LARGE N IS THE SAME 4/3 RESULT AS BEFORE

TO GAURI FROM JIM: THE UPDATING YOU DO ABOVE IS NOT THE STANDARD ONE FOR THIS PROBLEM. WHEN WE HAVE DONE THIS FOR MIXTURE MODELS, WE ASSIGN THE TRAINING OBSERVATIONS TO MIXTURE COMPONENTS THAT NEED TRAINING, BUT DO NOT ALLOW THE ASSIGNMENTS TO ALSO AFFECT THE MIXTURE WEIGHTS (HERE p). THUS $1/\sqrt{\lambda}$ WOULD BE UPDATED BY ONE ZERO POISSON OBSERVATION, BUT p WOULD STAY UNIFORM. THE NOTION IS THAT THE ASSIGNMENT IS ARTIFICIAL, AND SO ONE SHOULD MINIMIZE ITS IMPACT. AS THIS IS THE STANDARD METHOD, IT SHOULD BE GIVEN FIRST. THE OTHER UPDATES COULD ALSO BE GIVEN, ALTHOUGH I AM NOT AT ALL SURE WHAT THE UPDATE BELOW IS ABOUT; YOU'LL NEED TO GIVE SOME JUSTIFICATION FOR DOING THIS.

On the other hand, if we update the prior $\pi_1(p,\lambda) = 1/\lambda^{1/2}$ by considering one of the zeros as a zero from the Poisson distribution, we can do so by multiplying this prior by $(1-p)e^{-\lambda}/\{p+(1-p)e^{-\lambda}\}$ = Probability that the zero is from the Poisson distribution given that there is a zero, then the resulting posterior given by $(1-p)e^{-\lambda}\lambda^{-1/2}/[C\{p+$

$(1-p)e^{-\lambda}\}]$ where the normalizing constant C is

$$C = \int_0^\infty \int_0^1 \frac{(1-p)e^{-\lambda}\lambda^{-1/2}}{p+(1-p)e^{-\lambda}} dpd\lambda = 1.2932.$$

Treating the above posterior as the prior for p, λ to find the marginal $m_1(\mathbf{0})$, it can be checked that

$$m_1(\mathbf{0}) = \frac{\Gamma(1/2)}{n(n+1)C} \sum_{j=0}^{n-1} (n-j)^{1/2}.$$

For large n it can be checked that $m_1(\mathbf{0}) \approx (1/C)\{\Gamma(1/2)\sqrt{n}/(n+1)\}(2/3)$. Again, as before, if we use the marginalized version of this prior for λ to get $m_0(\mathbf{0})$, then

$$m_0(\mathbf{0}) = \frac{1}{C\sqrt{n}} \int_0^1 \int_0^\infty \frac{(1-p)e^{-u}u^{-1/2}}{p+(1-p)e^{-\frac{u}{n}}} dudp.$$

Applying the Monotone Convergence theorem, it can be checked that $\sqrt{n}m_0(\mathbf{0})$ converges to $\Gamma(1/2)/(2C)$. Then, for large n, once again, the Bayes factor $B_{10}(\mathbf{0}) \approx 4/3$.

5.2 Insufficient positive counts in the regression case

In the regression situation of Section 4, it was necessary to have sufficient positive counts so that the conditions of Theorem 4.2 were satisfied. We will restrict discussion here to the situation involving the prior specifications in (19), for which the key condition needed for the marginal to be finite was that the matrix $\boldsymbol{A}_+((n-k) \times q)$ should be of rank q. If the number of positive counts $n-k$ is insufficient so that t, the rank of $\boldsymbol{A}_+$, is less than q, this solution will not work.

Remark 5.3. Indeed, neither the prior for $\boldsymbol{\beta}$ given by (16) nor by (18) guarantees a finite marginal density, as is shown in the Appendix.

We call this situation one of rank deficiency, with the rank deficiency of $\boldsymbol{A}_+$ equal to $q-t$. The situation is analogous to the case of all zero counts without covariates discussed in subsection 5.1. (In the setup of that section, $q = 1$ and rank $\boldsymbol{A}_+$ less than 1 means that $k = n$, i.e., no positive counts.) We could again merely recognize that this type of data is just not informative enough to allow for objective Bayes analysis. We shall however propose a prior that yields finite marginal densities, following similar reasoning to that used in subsection 5.1.

We continue to use a uniform $(0, 1)$ prior for p and focus on proposing suitable priors for $\boldsymbol{\beta}$. A discussion similar to that in Section 3 shows that this prior has to be at least, partially proper.

Note that, instead of specifying a prior on $\boldsymbol{\beta}$, we can specify a prior on q independent parametric functions of $\boldsymbol{\beta}$; our specific proposal is to carefully choose these functions such that t of them are well identified by the data with positive counts while the remaining $q-t$ are not. We then propose to use a version of Jeffreys prior on the former t functions, and a proper prior on the latter $q-t$ functions

Specifically, let $\boldsymbol{A}_0$ denote the $k \times q$ matrix whose k rows are $\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_k^T$. Rank of $\boldsymbol{A} = q$ and rank of $\boldsymbol{A}_+ = t$ imply that rank of $\boldsymbol{A}_0 \geq q-t$. Let $V_+ \subseteq R^q$ denote the vector space of dimension t formed by the columns of $\boldsymbol{A}_+^T$. Suppose $\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_r}$ are all of the vectors from $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ corresponding to the zero counts which are in V_+. Note that $0 \leq r \leq k-(q-t)$. These vectors are linear combinations of of the vectors $\boldsymbol{a}_{j_1}, \ldots, \boldsymbol{a}_{j_t}$ and the corresponding $\lambda_{i_1}, \ldots, \lambda_{i_r}$ are functions of $\lambda_{j_1}, \ldots, \lambda_{j_t}$. From the set of $\{\lambda_j : j \in \{1, \ldots, k\} - \{i_1, \ldots, i_r\}\}$ we select $q-t$ λ's, $\lambda_{l_1}, \ldots, \lambda_{l_{q-t}}$ such that $\{\boldsymbol{a}_{j_1}, \ldots, \boldsymbol{a}_{j_t}, \boldsymbol{a}_{l_1}, \ldots, \boldsymbol{a}_{l_{q-t}}\}$ is linearly independent.

Note that there is an $(n-k) \times t$ matrix $\boldsymbol{C}$ of rank t such that

$$(\boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n) = (\boldsymbol{a}_{j_1}, \ldots, \boldsymbol{a}_{j_t})\boldsymbol{C}^T.$$

The information matrix for $\lambda_{j_1}, \ldots, \lambda_{j_t}$ based on the Poisson model for the observations $k+1, \ldots, n$ is given by

$$\boldsymbol{I}(\lambda_{j_1}, \ldots, \lambda_{j_t}) = Diag(\lambda_{j_1}^{-1}, \ldots, \lambda_{j_t}^{-1})\boldsymbol{C}^T Diag(\lambda_{k+1}, \ldots, \lambda_n)\boldsymbol{C} Diag(\lambda_{j_1}^{-1}, \ldots, \lambda_{j_t}^{-1}), \quad (27)$$

We define partial Jeffreys' prior for $\lambda_{j_1}, \ldots, \lambda_{j_t}$ by

$$\pi_{PJ}(\lambda_{j_1}, \ldots, \lambda_{j_t}) = \{\prod_{i=1}^{t} \lambda_{j_i}^{-1}\} |\boldsymbol{C}^T Diag(\lambda_{k+1}, \ldots, \lambda_n)\boldsymbol{C}|^{1/2}. \quad (28)$$

Let $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{q-t}\}$ denote an orthonormal basis of the space spanned by $\boldsymbol{a}_{l_1}, \ldots, \boldsymbol{a}_{l_{q-t}}$. Define $\xi_w = e^{\boldsymbol{b}_w^T \boldsymbol{\beta}}$, $w = 1, \ldots, q-t$. Note that $\lambda_{l_w}, w = 1, \ldots, q-t$ can be expressed in terms of $\xi_1, \ldots, \xi_{q-t}$. Indeed,

$$log(\lambda_{l_w}) = a_{0l_w} + \sum_{h=1}^{q-t} d_{wh} log(\xi_h), \quad w = 1, \ldots, q-t,$$

where $d_{wh} = \boldsymbol{b}_h^T \boldsymbol{a}_{l_w}$. Finally, we assign independent exponential distributions with mean 1 to $\xi_{l_1}, \ldots, \xi_{l_{q-t}}$. This prior will induce a proper distribution on $\lambda_{l_w}, w = 1, \ldots, q-t$ with a density which we denote by $\pi_{prop}(\lambda_{l_1}, \ldots, \lambda_{l_{q-t}})$. The final prior used to calculate the marginal density under model M_1^R is then given by

$$\pi(\lambda_{j_1}, \ldots, \lambda_{j_t}, \lambda_{l_1}, \ldots, \lambda_{l_{q-t}}) = \pi_{PJ}(\lambda_{j_1}, \ldots, \lambda_{j_t})\pi_{prop}(\lambda_{l_1}, \ldots, \lambda_{l_{q-t}})\,;$$

this is partially Jeffreys prior and partially proper. The corresponding prior density on $\boldsymbol{\beta}$ is, of course, obtained through transformation. Further, along the line of the proof of

Theorem 4.2, it can be checked that the marginal density $m_1^R(\boldsymbol{x})$ will be finite. We omit the details to save space.

While there is arbitrariness in the specific choice of $\lambda_{l_1}, \ldots, \lambda_{l_{q-t}}$ to assign subjective prior distribution based on exponential distribution, the partial Jeffreys prior in (28) remains invariant to the choice of t independent λ's from $\lambda_{k+1}, \ldots, \lambda_n$. This solution thus seems reasonable for small $q - t$.

To avoid the arbitrariness, we could consider all possible selections of $q - t$ λ's from $\lambda_1, \ldots, \lambda_k$ so that these $q - t$ λ's and t λ's from $\lambda_{k+1}, \ldots, \lambda_n$ define a reparameterization of $\boldsymbol{\beta}$. For each selection we can calculate the Bayes factor, and in the spirit of intrinsic Bayes factor we can take a suitable average over all these Bayes factors. If the rank deficiency of $\boldsymbol{A}_+$ is 1, we will have $k - r$ Bayes factors to average.

Acknowledgments

The authors would like to thank Dr. M.A.M. Beneyto, Dr. Andres Nolasco and Dr. Inma Melchor for the data set used in Section 5, and Prof. Xiao-Li Meng for sharing the Cholera data. They thank Dr. Archan Bhattacharya for help with computing. Part of this research was conducted while G.S. Datta and M.J. Bayarri were visiting SAMSI/Duke University, whose support is gratefully acknowledged. This research is supported in part by NSF under grants DMS-0103265, SES-0241651, and AST-0507481, and by the Spanish Ministry of Science and Technology, under Grant MTM2004-03290.

References

[1] Bayarri, M.J. and García-Donato, G. (2007). Extending conventional priors for testing general hypotheses in linear models. *Biometrika*, **94**, 135152.

[2] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag.

[3] Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, **1**, 385-402.

[4] Berger, J.O. and Pericchi, L.R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91**, 109122.

[5] Berger, J.O. and Pericchi, L.R. (2001). Objective Bayesian methods for model selection: introduction and comparison (with discussion). In *Model Selection, Institute of Mathematical Statistics Lecture Notes- Monograph Series,*, **38**, Ed. P. Lahiri, pp. 135207, Beachwood Ohio: Institute of Mathematical Statistics.

[6] Pérez, J.M. and Berger, J. (2001). Analysis of mixture models using expected posterior priors, with application to classification of gamma ray bursts. In *Bayesian Methods, with applications to science, policy and official statistics*, E. George and P. Nanopoulos, eds., Official Publications of the European Communities, Luxembourg, 401–410.

[7] Berger, J., Pericchi, L. and Varshavsky, J. (1998). Bayes factors and marginal distributions in invariant situations. *Sankya A*, **60**, 307321.

[8] Berger, J. and Sun, D. (2008). Objective priors for a bivariate normal model with multivariate generalizations. To appear in *Annals of Statistics.*

[9] Bhattacharya, A., Clarke, B.S. and Datta, G.S. (2007). A Bayesian test for excess zeros in a zero-inflated power series distribution. Preprint.

[10] Broek, J.V.D. (1995). A score test for zero inflation on a Poisson distribution. *Biometrics*, **51**, 738-743.

[11] Conigliani, C., Castro, J. I. and O'Hagan, A. (2000). Bayesian assessment of goodness of fit against nonparametric alternatives. *Canadian Journal of Statistics*, **28**, 327-342.

[12] Deng, D. and Paul, S.R. (2000). Score test for zero inflation in generalized linear models. *Canadian Journal of Statistics*, **28**, 563-570.

[13] Ghosh, J.K. (2006). Personal communication.

[14] Ghosh, J.K. and Samanta, T. (2002). Nonsubjective Bayes testing - an overview. *Journal of Statistical Planning and Inference*, **103**, 205-223.

[15] Ghosh, S.K., Mukhopadhyay, P. and Lu, J.C. (2006). Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, **136**, 1360-1375.

[16] Jeffreys, H. (1961). *Theory of Probability, 3rd ed.* London: Oxford University Press.

[17] Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions.* Second edition. John Wiley & Sons Inc.

[18] Kass, R.E. and Vaidyanathan, S. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statististical Society B*, **54**, 12944.

[19] Kass, R.E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, **91**, 1343-1370.

[20] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1-14.

[21] McKendrick, A.G. (1926). Application of mathematics to medical problems. *Proc. Edin. Math. Soc.*, **44**, 98-130.

[22] Noble, B. (1969). *Applied Linear Algebra.* Prentice-Hall, New York.

[23] O'Hagan, A. (1995). Fractional Bayes factors for model comparisons. *Journal of the Royal Statistical Society, Ser. B*, **57**, 99-138.

Appendix

We outline here some of the proofs of Theorems and results appearing in the paper.

Proof of Theorem 5.1

Let $\boldsymbol{i}$ denote the indices $(i_1, \ldots, i_q)$ and $\boldsymbol{A}(\boldsymbol{i})$ denote a $q \times q$ submatrix of $\boldsymbol{A}$ based on rows $i_1, \ldots, i_q$. Then by Binet-Cauchy expansion of determinant (cf. Noble, 1969, p. 226) it can be shown that

$$|\sum_{i=1}^{n} \lambda_i \boldsymbol{a}_i \boldsymbol{a}_i^T| = \sum (\lambda_{i_1} \ldots \lambda_{i_q}) |\boldsymbol{A}(\boldsymbol{i}) \boldsymbol{A}(\boldsymbol{i})^T|, \tag{A1}$$

where the summation is over all submatrices of order $q \times q$. Dropping all the terms from the above summation for which $|\boldsymbol{A}(\boldsymbol{i})\boldsymbol{A}(\boldsymbol{i})^T| = 0$ we get from equations (16) and (A1) that

$$\pi_0^R(\boldsymbol{\beta}) \leq \sum^{*} (\lambda_{i_1} \ldots \lambda_{i_q})^{1/2} |\boldsymbol{A}(\boldsymbol{i}) \boldsymbol{A}(\boldsymbol{i})^T|^{1/2}, \tag{A2}$$

where $\sum^*$ above denotes summation over all $q \times q$ matrices for which $|\boldsymbol{A}(\boldsymbol{i})\boldsymbol{A}(\boldsymbol{i})^T| > 0$.

Since $e^{-\lambda_i}\lambda_i^{x_i}/x_i! < 1$, from (20) and (A2) we get

$$m_0^R(\boldsymbol{x}) \leq \sum^{*} \int_{R^q} \prod_{j=1}^{q} \{ \frac{e^{-\lambda_{i_j}} \lambda_{i_j}^{x_{i_j}}}{x_{i_j}!} \} (\lambda_{i_1} \ldots \lambda_{i_q})^{1/2} |\boldsymbol{A}(\boldsymbol{i}) \boldsymbol{A}(\boldsymbol{i})^T|^{1/2} d\boldsymbol{\beta}. \tag{A3}$$

Recall that $log(\lambda_i) = a_{0i} + \boldsymbol{a}_i^T \boldsymbol{\beta}$. Now transforming $\boldsymbol{\beta}$ to $(\lambda_{i_1}, \ldots, \lambda_{i_q})$ and using the Jacobian of transformation $(\lambda_{i_1} \ldots \lambda_{i_q})^{-1} |\boldsymbol{A}(\boldsymbol{i})\boldsymbol{A}(\boldsymbol{i})^T|^{-1/2}$, we get from (A3) that

$$m_0^R(\boldsymbol{x}) \leq \sum^{*} \prod_{j=1}^{q} \int_0^{\infty} \frac{e^{-\lambda_{i_j}} \lambda_{i_j}^{x_{i_j} - .5}}{x_{i_j}!} d\lambda_{i_j} < \infty, \tag{A4}$$

since each of the integrals in the right hand side of (A4) is finite. This completes the proof of Theorem 5.1.

Proof of Theorem 5.2

First, as in (A1) and (A2), it can be shown that for some positive c (not depending on parameters) less than 1

$$c \sum^{*} (\lambda_{i_1} \ldots \lambda_{i_q})^{1/2} |\boldsymbol{A}(\boldsymbol{i}) \boldsymbol{A}(\boldsymbol{i})^T|^{1/2} \leq \pi_0^R(\boldsymbol{\beta}) \leq \sum^{*} (\lambda_{i_1} \ldots \lambda_{i_q})^{1/2} |\boldsymbol{A}(\boldsymbol{i}) \boldsymbol{A}(\boldsymbol{i})^T|^{1/2}. \tag{A5}$$

In view of the above inequality and (23), the marginal $m_1^R(\boldsymbol{x})$ is finite if and only if

$$\int_{R^q} \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}(\lambda_{i_1}\ldots\lambda_{i_q})^{1/2}|\boldsymbol{A}(\boldsymbol{i})\boldsymbol{A}(\boldsymbol{i})^T|^{1/2}d\boldsymbol{\beta} < \infty \qquad \text{(A6)}$$

for each $\boldsymbol{i} = (i_1, \ldots, i_q)$ for which $|\boldsymbol{A}(\boldsymbol{i})\boldsymbol{A}(\boldsymbol{i})^T| > 0$.

Note that the sufficient condition stated in the theorem and the condition that rank of $\boldsymbol{A}$ is q imply that the regression matrix $\boldsymbol{A}_+^T = (\boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n)$ corresponding to the set of positive counts has rank q.

Suppose, with no loss of generality, $i_1 < \cdots < i_q$ in (A6). Also, suppose $i_1 < \cdots < i_t \le k < i_{t+1} < \cdots < i_q$. It is possible that t may be 0 or may be q. By the assumed condition that for $j = 1, \ldots, k$, $\boldsymbol{a}_j$ can be expressed as a linear combination of $\boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n$ with nonnegative coefficients, it follows that

$$\lambda_{i_j} = h_{i_j} \prod_{m=k+1}^{n} \lambda_m^{c_{mi_j}}, \quad j = 1, \ldots, t,$$

where $c_{mi_j} \ge 0$ and $h_{i_j} > 0$. Then

$$\prod_{j=1}^{t} \lambda_{i_j} = f \prod_{m=k+1}^{n} \lambda_m^{b_m},$$

where $b_m = \sum_{j=1}^{t} c_{mi_j} \ge 0$ and $f > 0$ are free from parameters.

Then the integrand (without $|\boldsymbol{A}(\boldsymbol{i})\boldsymbol{A}(\boldsymbol{i})^T|^{1/2}$) in (A6) can be simplified as

$$\begin{aligned}\prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}(\lambda_{i_1}\ldots\lambda_{i_q})^{1/2} &= \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i+\frac{1}{2}b_i}}{x_i!}(\lambda_{i_{t+1}}\ldots\lambda_{i_q})^{1/2} \\ &= [\prod_{j=t+1}^{q} \frac{e^{-\lambda_{i_j}}\lambda_{i_j}^{x_{i_j}+\frac{1}{2}b_{i_j}+\frac{1}{2}}}{x_{i_j}!}][\prod_{l=1}^{n+t-k-q} \frac{e^{-\lambda_{\alpha_l}}\lambda_{\alpha_l}^{x_{\alpha_l}+\frac{1}{2}b_{\alpha_l}}}{x_{\alpha_l}!}], \text{(A7)}\end{aligned}$$

where $\{\alpha_1, \ldots, \alpha_{n+t-k-q}\} = \{k+1, \ldots, n\} - \{i_{t+1}, \ldots, i_q\}$.

Suppose $\{s_1, \ldots, s_q\} \subset \{k+1, \ldots, n\}$ is such that $\{\boldsymbol{a}_{s_1}, \ldots, \boldsymbol{a}_{s_q}\}$ is a linearly independent set (such a set exists since $\boldsymbol{A}_+$ is of rank q). Note that for $y > 0$ the function $g(u) = e^{-u}u^y$ is maximized at $u = y$ implying

$$e^{-u}u^y \le e^{-y}y^y \qquad \text{(A8)}$$

for all $u > 0$.

By (A8) we get from (A7) that

$$\prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}(\lambda_{i_1}\ldots\lambda_{i_q})^{1/2} \leq D(\prod_{j=1}^{q} e^{-\lambda_{s_j}}\lambda_{s_j}^{d_{s_j}}), \tag{A9}$$

where $D > 0$ is a constant independent of the parameters and $d_{s_j} = x_{s_j} + \frac{1}{2}b_{s_j} + \frac{1}{2}$ if $s_j \in \{i_{t+1}, \ldots, i_q\}$, and $d_{s_j} = x_{s_j} + \frac{1}{2}b_{s_j}$ if $s_j \in \{\alpha_1, \ldots, \alpha_{n+t-k-q}\}$. Note that $d_{s_j} \geq 1$.

Since the Jacobian of transformation from $\boldsymbol{\beta}$ to $\lambda_{s_1}, \ldots, \lambda_{s_q}$ is $H/(\lambda_{s_1}\ldots\lambda_{s_q})$ for some $H > 0$ not depending on the parameters, by (A9) we have

$$\int_{R^q}\prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}(\lambda_{i_1}\ldots\lambda_{i_q})^{1/2}d\boldsymbol{\beta} \leq HD\prod_{j=1}^{q}\int_0^{\infty} e^{-\lambda_{s_j}}\lambda_{s_j}^{d_{s_j}-1}d\lambda_{s_j} < \infty \tag{A10}$$

since $d_{s_j} \geq 1$ for $j = 1, \ldots, q$. By (A10) and (A6) we conclude that $m_1^R(\boldsymbol{x})$ is finite. This completes the proof of Theorem 5.2.

Bayes factor in Section 4.2

The Bayes factor for testing Poisson versus ZIP models with (known) offsets E_i, that is for choosing between models

$$M_0^o : X_i \overset{\text{ind.}}{\sim} f_0(\cdot \mid E_i\,\lambda),\ i = 1, \ldots, n,$$
$$M_1^o : X_i \overset{\text{ind.}}{\sim} f_1(\cdot \mid E_i\,\lambda, p),\ i = 1, \ldots, n,$$

is $B_{10} = m_1(\boldsymbol{x})/m_0(\boldsymbol{x})$ where $m_1(\boldsymbol{x})$ and $m_1(\boldsymbol{x})$ are given respectively in (A12) and (A11) below.

Under model M_0^o, the likelihood is $C\ \lambda^s e^{-\lambda\sum_{i=1}^n E_i}$ where $C = \prod_{i=1}^n E_i^{x_i}/\prod_{i=1}^n x_i!$, so that, the marginal under the prior $\pi_0(\lambda) = 1/\sqrt{\lambda}$ is

$$m_0(\boldsymbol{x}) = C\ \frac{\Gamma(s+.5)}{(\sum_{i=1}^n E_i)^{s+.5}}\ . \tag{A11}$$

Under model M_1^o, the likelihood is

$$C\ (1-p)^{n-k}\lambda^s e^{-\lambda\sum_{t=k+1}^n E_t}\big[p^k + \sum_{j=1}^{k}\sum_{1\leq i_1<\ldots<i_j\leq k} p^{k-j}(1-p)^j e^{-\lambda\sum_{l=1}^j E_{i_l}}\big]$$

which, integrated with respect to the prior $\pi_1(p,\lambda) = I(0 < p \leq 1)/\sqrt{\lambda}$ gives the marginal:

$$m_1(\boldsymbol{x}) = C\,\Gamma(s{+}.5)\left[\frac{Be(k+1, n-k+1)}{(\sum_{t=k+1}^n E_t)^{s+.5}} + \sum_{j=1}^{k}\sum_{1\leq i_1<\ldots<i_j\leq k}\frac{Be(k-j+1, n-k+j+1)}{\big(\sum_{l=1}^j E_{i_l} + \sum_{t=k+1}^n E_t\big)^{s+.5}}\right], \tag{A12}$$

where $Be(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function.

Proof of Remark 5.3

It is possible to choose t linearly independent row vectors from the rows of $\boldsymbol{A}_+$. Let us denote these t rows by $j_1, \ldots, j_t$ where $\{j_1, \ldots, j_t\} \subseteq \{k+1, \ldots, n\}$. Note that all $\lambda_i, i = k+1, \ldots, n$ are functions of $\lambda_{j_1}, \ldots, \lambda_{j_t}$. Thus $\prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}$ is a function of $\lambda_{j_1}, \ldots, \lambda_{j_t}$. From (23), (A5) and (A6) it follows that Jeffreys' prior (16) will give an infinite marginal density under model M_1^R. This is because there is an index vector $\boldsymbol{i} = (i_1, \ldots, i_q)$ given by $(j_1, \ldots, j_t, l_1, \ldots, l_{q-t})$ so that the left hand integral of (A10) given by

$$\begin{aligned} &\int_{R^q} \prod_{i=k+1}^{n} \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!} (\lambda_{i_1} \ldots \lambda_{i_q})^{1/2} d\boldsymbol{\beta} \\ = \quad & J \int_0^\infty \ldots \int_0^\infty h(\lambda_{j_1}, \ldots, \lambda_{j_t}) d\lambda_{j_1} \ldots d\lambda_{j_t} \times \prod_{u=1}^{q-t} \int_0^\infty \lambda_{l_u}^{-1/2} d\lambda_{l_u} = \infty, \end{aligned}$$

where J is the Jacobian of transformation and $h(\lambda_{j_1}, \ldots, \lambda_{j_t})$ is a function of $\lambda_{j_1}, \ldots, \lambda_{j_t}$. On the other hand, the prior given by (**??**) is zero since rank of $\boldsymbol{A}_+$ is less than q, which completes the proof.