Modern Bayesian Inference in Zero-Inflated Poisson Models

Erlis Ruli and Laura Ventura

Abstract The Zero-Inflated Poisson (ZIP) distribution, typically assumed for modeling count data with excess of zeros, assumes that with probability p the only possible observation is zero, and with probability $1-p$ a $Poisson(\psi)$ random variable is observed. Both the probability p and the mean ψ may depend on covariates. In this paper we discuss and apply Bayesian inference based on matching priors and on higher-order asymptotics to perform accurate inference on ψ only, even for small sample sizes.

Key words: Asymptotic expansions, Count data, Matching prior, Modified profile likelihood, Nuisance parameter, Tail area probability, ZIP regression.

1 Introduction

Count data often show more zeros than what would be expected from a Poisson distribution. ZIP models (and ZIP regression models) represents a useful class of models for such data. In recent years there has been considerable interest in these models in various application areas (see [6]). Usually, statistical inference is based on Expectation-Maximization (see, e.g., [4]) or on maximum likelihood. On the contrary, Bayesian inference for ZIP models remains relatively unexplored.

In this paper we discuss some recent advances in Bayesian inference based on higher-order asymptotics (see, e.g., [5] and [2]) and matching priors (see [7]) to ZIP models to perform accurate inference on the parameter of interest ψ, even for small sample sizes. We also extend the problem to the ZIP regression model. The pro-

Erlis Ruli
Department of Statistics, University of Padova, Italy, e-mail: ruli@stat.unipd.it

Laura Ventura
Department of Statistics, University of Padova, Italy, e-mail: ventura@stat.unipd.it

posed approach avoids elicitation on the nuisance parameters and multidimensional integral computations. The accuracy of the proposed methodology is illustrated both by numerical studies and by a real-life dataset concerning clinical studies.

The paper is organized as follows. In Section 2, we review some recent advances on Bayesian inference on ψ. In Section 3 the ZIP model, both in the scalar and regression context, is considered. Section 4 illustrates the numerical studies along with an application to a real dataset.

2 Background

Let $y = (y_1, \ldots, y_n)$ be a random sample of size n from a random variable $Y \sim p(y;\theta)$, with $\theta = (\psi, \lambda) \in \Theta \subset \mathbb{R}^d$, ψ the interest parameter and λ the nuisance parameter ($\dim(\psi) = p$, $\dim(\lambda) = q$ and $p + q = d$). Let $L(\theta) = L(\psi, \lambda)$ be the likelihood, $\ell(\theta) = \ell(\psi, \lambda)$ the loglikelihood, $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$ the maximum likelihood estimate (MLE) of θ. Moreover, let $\hat{\lambda}_\psi$ be the constrained MLE of λ and let $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$. Given the likelihood function and a prior $\pi(\psi, \lambda)$, Bayesian inference on ψ is based on

$$\pi(\psi|y) \propto \int L(\psi, \lambda)\pi(\psi, \lambda)\,\mathrm{d}\lambda. \tag{1}$$

The computation of (1) requires the elicitation of θ and multidimensional numerical integration. These may be avoided using recent advances in Bayesian inference ([7] and [9]) where λ is eliminated through a suitable pseudo-likelihood of ψ only, with properties similar to a genuine likelihood. In this respect, Bayesian inference on ψ may be based on

$$\pi_M(\psi|y) \propto \pi(\psi) L_M(\psi), \tag{2}$$

where $L_M(\psi)$ is the modified profile likelihood of [1] and $\pi(\psi)$ is a suitable prior on ψ only. Let $i_{\psi\psi}(\theta)$, $i_{\psi\lambda}(\theta)$ and $i_{\lambda\lambda}(\theta)$ be the blocks of the expected information $i(\theta)$. For scalar ψ, it can be shown that (2) based on the matching prior

$$\pi^*(\psi) \propto i_{\psi\psi.\lambda}(\hat{\theta}_\psi)^{1/2}, \tag{3}$$

where $i_{\psi\psi.\lambda}(\theta) = i_{\psi\psi}(\theta) - i_{\psi\lambda}(\theta) i_{\lambda\lambda}(\theta)^{-1} i_{\lambda\psi}(\theta)$ is the partial information, is a genuine posterior distribution (see [7]). When ψ is multidimensional, a matching prior for ψ only to be used in (2) is not available. Nevertheless, a Jeffreys' type prior of the form $\pi^*(\psi) \propto |i_{\psi\psi.\lambda}(\hat{\theta}_\psi)|^{1/2}$ may be considered (see [8]) in (2). Applications of the resulting posterior are not available at present in the statistical literature and will be discussed in this contribution.

3 Application to ZIP models

ZIP model. Consider a random sample of size n from $Y \sim ZIP(\psi, p)$. It is of interest to make inference on ψ. The probability density function of Y is given by

$$p(y;\psi,\lambda) = pa(y) + (1-p)\frac{e^{-\psi}\psi^y}{y!},\ y = 0,1,\ldots,\ \psi > 0,\ \frac{1}{1-e^{\psi}} < p < 1, \quad (4)$$

where $a = a(y)$ is equal to 1 if $y = 0$, and is equal to 0 if $y > 0$. Let us reparametrize the model with $\tau = (\psi, \lambda)$, with $\lambda = \lambda(\psi, p) = \frac{1-p}{1-1/(1-e^{\psi})}$, so that $0 < \lambda < 1$. This parametrization may be useful because with (ψ, λ) the model has separable parameters. The likelihood for (ψ, λ) is

$$L(\psi,\lambda) = (1-\lambda)^{s0}\lambda^{n-s0}\frac{\psi^s}{(e^{\psi}-1)^{n-s0}} = L(\lambda)L(\psi), \quad (5)$$

with $s_0 = \sum_{i=1}^{n} a_i$, $s_0 \leq n$ and $s = \sum_{i=1}^{n}(1-a_i)y_i$. Since $L(\tau) = L(\psi)L(\lambda)$, inference on ψ should be carried out using $L(\psi) = \psi^s/(e^{\psi}-1)^{n-s0}$. Moreover, we have that $L(\psi) = L_P(\psi) = L_M(\psi)$. Note that $L(\psi)$ is equal to the likelihood function derived from the truncated Poisson distribution (see [10]). It can be shown that the matching prior (3) for ψ is

$$\pi^*(\psi) = i_{\psi\psi.\lambda}(\hat{\tau}_{\psi})^{1/2} \propto \left[\frac{e^{\psi}(e^{\psi}-1) - \psi\hat{\lambda}e^{\psi}}{\psi(e^{\psi}-1)^2}\right]^{1/2}, \quad (6)$$

and the posterior distribution (2) for ψ is thus

$$\pi^*(\psi|s,s_0) \propto \frac{\psi^{s-1/2}e^{\psi/2}(e^{\psi}-1-\hat{\lambda}\psi)^{1/2}}{(e^{\psi}-1)^{n+1-s0}}. \quad (7)$$

Regression model. Now, let us assume $Y_i \sim ZIP(\psi_i, \lambda_i)$, for $i = 1,\ldots,n$, where ψ_i and λ_i are modelled by the following common link functions (see, e.g., [4]) $\psi_i = e^{\mathbf{B}_i\beta}$ and $\lambda_i = e^{\mathbf{G}_i\gamma}/(1+e^{\mathbf{G}_i\gamma})$, where $\mathbf{B}$ and $\mathbf{G}$ are non random design matrices, β is the vector of interest parameters and γ is vector of nuisance parameters. The likelihood for β is independent from $L(\gamma)$ and $L_M(\beta) = L(\beta)$. Since $\dim(\beta) > 1$ a matching prior for β is not easily available. However, using results in [8] the Jeffreys'-type prior for β may be considered, which is given by

$$\pi^*(\beta) = \left|\sum_i \left\{\mathbf{B}_i e^{\mathbf{B}_i\beta}e^{\mathbf{G}_i\hat{\gamma}}\exp(e^{\mathbf{B}_i\beta})\frac{1+e^{\mathbf{B}_i\beta}+\exp(e^{\mathbf{B}_i\beta})}{(1+e^{\mathbf{G}_i\hat{\gamma}})\left[\exp(e^{\mathbf{B}_i\beta})-1\right]^2}\right\}\right|^{1/2}. \quad (8)$$

In this case the posterior distribution (2) is given by the prior (8) times $L(\beta)$.

4 Numerical studies and real data application

We investigate the empirical coverage of Bayesian credible sets obtained from (7) and (1) based on $\pi(\psi) \sim Ga(a,b)$ with $a = b = 0.01$ and $\lambda \sim U(0,1)$. The posteriors (1) and (7) can be approximated and then integrated to give marginal tail area probabilities (see [5] and [9] for details).

Numerical studies. We analysed the behaviour of (7) and (1) under the ZIP model through 50.000 Monte Carlo trials. In particular, using marginal tail area probabilities obtained from (7) and (1) we compare the empirical frequentist coverage for 95% HPD sets and the lower and upper 5% tail. For a sample of size $n = 10$ the matching prior achieves coverages of `0.9422`, `0.0343` and `0.0234`, respectively. The same quantities for the Gamma-Uniform prior, are `0.9394`, `0.0429` and `0.0117`. The matching prior achieves better coverage probabilities, but the difference in performance between the two priors is more evident when left and right tail coverages are compared.

Example. Now we apply this methodology to our dataset (see [3]). The variables considered are the number of spots or silica particles (`NS`) and the number of positive zones (`NPZ`) in which there is at least one spot in the lung tissues. We obtained estimates of ψ along with its 95% HPD, under the ZIP model, with the matching prior and with the gamma-uniform prior. We estimate ψ also by maximising its profile likelihood. Both priors compared with the MLEs give very similar results.

As a final remark, we note that simulation studies and related application with real data for the ZIP regression model are under investigation.

References

1. Barndorff-Nielsen, O.E.: On a formula for the distribution of the maximum likelihood estimator. Biometrika **70**, 343–365 (1983)
2. Brazzale, A. R., Davison, A. C., Reid, N.: Applied Asymptotics. Cambridge University Press, Cambdidge (2007)
3. Fassina, A., Corradin, M., El Mazloum, R., Murer, B., Furlan, C., Montisci, M., Guolo, A., Ventura, L.: Silica levels and lung cancer: results of an ESEM study. Inhal. Toxicol., **21**, 133-140.
4. Lambert, D.: Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics **34**, 1–14 (1992)
5. Reid, N.: Asymptotics and the theory of inference. Ann. Statist. **31**(6), 1695–1731 (2003)
6. Ridout, M., Demetrio, C.G.B., Hinde, J.: Models for count data with many zeros. International Biometric Conference, Cape Town (1998)
7. Ventura, L., Cabras, S., Racugno, W.: Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. J. Am. Statist. Assoc. **104**, 768–774 (2009)
8. Ventura, L., Cabras, S., Racugno, W.: Default prior distributions from quasi- and quasi-profile likelihoods. J. Statist. Plann. Inf. **43**, 2937–2942 (2010)
9. Ventura, L., Racugno, W.: Recent advances on Bayesian inference for $P(X < Y)$. Bayesian Anal. **6**, 411–428 (2011)
10. Yip, P.: Inference about the mean of a Poisson distribution in the presence of a nuisance parameter. Aust. J. Statist. **30**(3), 299–306 (1988)