Lognormal and Gamma Mixed Negative Binomial Regression

Mingyuan Zhou, Lingbo Li, David Dunson and Lawrence Carin

ECE and Statistics, Duke University, Durham, NC 27708, USA

Introduction

➢In regression analysis of counts, a lack of simple and efficient algorithms for posterior computation has made Bayesian approaches appear unattractive and thus underdeveloped.

➢We propose a lognormal and gamma mixed negative binomial (NB) regression model for counts, and present efficient closed-form Bayesian inference.

➢By placing a gamma distribution prior on the NB dispersion parameter *r*, and connecting a lognormal distribution prior with the logit of the NB probability parameter *p*, efficient Gibbs sampling and variational Bayes inference are both developed.

➢The closed-form updates are obtained by exploiting conditional conjugacy via both a compound Poisson representation and a Polya-Gamma distribution based data augmentation approach.

➢The proposed Bayesian inference can be implemented routinely, while being easily generalizable to more complex settings involving multivariate dependence structures.

Regression Models for Counts

❑Poisson and Negative binomial distributions

$$f_X(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad f_X(k) = \int_0^\infty \text{Pois}(k;\lambda)\text{Gamma}\left(\lambda; r, \frac{p}{1-p}\right)d\lambda = \frac{\Gamma(r+k)}{k!\Gamma(r)}(1-p)^r p^k$$

Overdispersion: Variance > Mean

Heterogeneity: difference between individuals

Contagion: dependence between the occurrence of events

❑Poisson regression

$$y_i \sim \text{Pois}(\lambda_i), \quad \lambda_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}) \qquad \mathbb{E}[y_i|\boldsymbol{x}_i] = \text{Var}[y_i|\boldsymbol{x}_i] = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$$

❑Poisson regression with random effect

$$y_i \sim \text{Pois}(\lambda_i), \quad \lambda_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\epsilon_i \qquad \text{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \frac{\text{Var}[\epsilon_i]}{\mathbb{E}^2[\epsilon_i]}\mathbb{E}^2[y_i|\boldsymbol{x}_i]$$

Negative binomial regression

$$\epsilon_i \sim \text{Gamma}(r, 1/r) = \frac{r^r}{\Gamma(r)}\epsilon_i^{r-1}e^{-r\epsilon_i} \qquad \text{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \phi\mathbb{E}^2[y_i|\boldsymbol{x}_i]$$

Lognormal-Poisson regression

$$\epsilon_i \sim \ln\mathcal{N}(0,\sigma^2) \qquad \text{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \left(e^{\sigma^2}-1\right)\mathbb{E}^2[y_i|\boldsymbol{x}_i]$$

LGNB Regression

❑Lognormal-gamma-gamma-Poisson regression

$$y_i \sim \text{Pois}(\lambda_i), \quad \lambda_i \sim \text{Gamma}\left(r, \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\epsilon_i\right), \quad r \sim \text{Gamma}(a_0, 1/h), \quad \epsilon_i \sim \ln\mathcal{N}(0,\varphi^{-1})$$

❑Lognormal gamma mixed NB regression

$$p_i = \frac{e^{\psi_i}}{1+e^{\psi_i}} = \frac{\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\epsilon_i}{1+\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})\epsilon_i}, \quad \text{logit}(p_i) = \ln\frac{p_i}{1-p_i}$$

$$y_i \sim \text{NB}(r, p_i), \quad \psi_i = \text{logit}(p_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \ln\epsilon_i, \quad r \sim \text{Gamma}(a_0, 1/h), \quad \epsilon_i \sim \ln\mathcal{N}(0,\varphi^{-1})$$

❑Properties

$$\mathbb{E}[y_i|\boldsymbol{x}_i] = \mathbb{E}_{\epsilon_i}[\mathbb{E}[y_i|\boldsymbol{x}_i,\epsilon_i]] = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma^2/2 + \ln r) \quad \text{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}_{\epsilon_i}[\text{Var}[y_i|\boldsymbol{x}_i,\epsilon_i]] + \text{Var}_{\epsilon_i}[\mathbb{E}[y_i|\boldsymbol{x}_i,\epsilon_i]] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \left(e^{\sigma^2}(1+r^{-1})-1\right)\mathbb{E}^2[y_i|\boldsymbol{x}_i]$$

❑Quasi-dispersion

NB $\kappa = \phi = r^{-1}$ **Lognormal-Poisson** $\kappa = \left(e^{\sigma^2}-1\right)$ **LGNB** $\kappa = \left(e^{\sigma^2}(1+r^{-1})-1\right)$

Inferring *r* under Compound Poisson

$y \sim \text{NB}(r,p)$ can be augmented as $y = \sum_{\ell=1}^{L} u_\ell, \ L \sim \text{Pois}(-r\ln(1-p)), \ u_\ell \overset{iid}{\sim} \text{Log}(p)$

$$y_i \overset{iid}{\sim} \text{NB}(r,p), \ r \sim \text{Gamma}(a, 1/b)$$

$$\text{Pr}(L_i = j|-) = R_r(y_i, j), \quad j = 0, \cdots, y_i.$$

$$R_r(m,j) = F(m,j)r^j \Big/ \sum_{j'=1}^{m} F(m,j')r^{j'} \qquad F(m,j) = \begin{cases} \frac{m-1}{m}F(m-1,j) + \frac{1}{m}F(m-1,j-1) & \text{if } 1 \le j \le m; \\ 0 & \text{otherwise.} \end{cases}$$

$$(r|-) \sim \text{Gamma}\left(a + \sum_{i=1}^{N} L_i, \frac{1}{b - N\ln(1-p)}\right)$$

Inferring β using Polya-Gamma

❑Polya-Gamma distribution $X \sim \text{PG}(a,c)$

$$X \overset{D}{=} \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{g_k}{(k-1/2)^2 + c^2/(4\pi^2)}, \quad g_k \sim \text{Gamma}(a,1)$$

❑Data augmentation

$$y_i \sim \text{NB}(r, p_i), \quad \psi_i = \text{logit}(p_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \ln\epsilon_i \quad \epsilon_i \sim \ln\mathcal{N}(0,\varphi^{-1})$$

$$\omega_i \sim \text{PG}(y_i + r, 0) \quad \mathbb{E}_{\omega_i}\left[\exp(-\omega_i\psi_i^2/2)\right] = \cosh^{-(y_i+r)}(\psi_i/2)$$

$$\mathcal{L}(\psi_i) \propto \frac{(e^{\psi_i})^{y_i}}{(1+e^{\psi_i})^{y_i+r}} = \frac{2^{-(y_i+r)}\exp(\frac{y_i-r}{2}\psi_i)}{\cosh^{y_i+r}(\psi_i/2)} \propto \exp\left(\frac{y_i-r}{2}\psi_i\right)\mathbb{E}_{\omega_i}\left[\exp(-\omega_i\psi_i^2/2)\right]$$

❑Gibbs sampling

$$(\boldsymbol{\psi}|-) \propto \mathcal{N}(\boldsymbol{\psi}; \mathbf{X}\boldsymbol{\beta}, \varphi^{-1}\mathbf{I})\prod_{i=1}^{N} e^{-\frac{\omega_i}{2}\left(\psi_i - \frac{y_i-r}{2\omega_i}\right)^2}$$

$$(\boldsymbol{\psi}|-) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \boldsymbol{\mu} = \boldsymbol{\Sigma}[(\boldsymbol{y}-r)/2 + \varphi\mathbf{X}\boldsymbol{\beta}] \quad \boldsymbol{\Sigma} = (\varphi\mathbf{I} + \boldsymbol{\Omega})^{-1}$$

$$(\omega_i|-) \propto \exp(-\omega_i\psi_i^2/2)\text{PG}(\omega_i; y_i + r, 0)$$

$$(\omega_i|-) \sim \text{PG}(y_i + r, \psi_i)$$

Model and Inference

$$y_i \sim \text{NB}(r, p_i), \quad \psi_i = \text{logit}(p_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \ln\epsilon_i$$

$$\epsilon_i \sim \ln\mathcal{N}(0,\varphi^{-1}), \quad \varphi \sim \text{Gamma}(e_0, 1/f_0)$$

$$\boldsymbol{\beta} \sim \prod_{p=0}^{P}\mathcal{N}(0,\alpha_p^{-1}), \quad \alpha_p \sim \text{Gamma}(c_0, 1/d_0)$$

$$r \sim \text{Gamma}(a_0, 1/h), \quad h \sim \text{Gamma}(b_0, 1/g_0)$$

Gibbs sampling

$$\text{Pr}(L_i = j|-) = R_r(y_i, j), \quad j = 0, \cdots, y_i$$

$$(r|-) \sim \text{Gamma}\left(a_0 + \sum_{i=1}^{N} L_i, \frac{1}{h - \sum_{i=1}^{N}\ln(1-p_i)}\right)$$

$$(\omega_i|-) \sim \text{PG}(y_i + r, \psi_i)$$

$$(\boldsymbol{\psi}|-) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (\boldsymbol{\beta}|-) \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$$

$$(h|-) \sim \text{Gamma}(a_0 + b_0, 1/(g_0 + r))$$

$$(\varphi|-) \sim \text{Gamma}\left(e_0 + \frac{N}{2}, \frac{1}{f_0 + \|\boldsymbol{\psi} - \mathbf{X}\boldsymbol{\beta}\|_2^2/2}\right)$$

$$(\alpha_p|-) \sim \text{Gamma}\left(c_0 + 1/2, 1/(d_0 + \beta_p^2/2)\right)$$

Variational Bayes

$$\tilde{a} = a_0 + \sum_{i=1}^{N}\langle L_i\rangle, \quad \tilde{h} = \langle h\rangle + \sum_{i=1}^{N}\langle\ln(1+e^{\psi_i})\rangle$$

$$\tilde{\boldsymbol{\Sigma}} = \left(\langle\varphi\rangle\mathbf{I} + \tilde{\boldsymbol{\Omega}}\right)^{-1}, \quad \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}\left[(\boldsymbol{y} - \langle r\rangle)/2 + \langle\varphi\rangle\mathbf{X}\boldsymbol{\beta}\right]$$

$$\tilde{\boldsymbol{\Sigma}}_\beta = \left(\langle\varphi\rangle\mathbf{X}^T\mathbf{X} + \langle\tilde{\mathbf{A}}\rangle\right)^{-1}, \quad \tilde{\boldsymbol{\mu}}_\beta = \langle\varphi\rangle\tilde{\boldsymbol{\Sigma}}_\beta\mathbf{X}^T\langle\boldsymbol{\psi}\rangle$$

$$\tilde{b} = a_0 + b_0, \quad \tilde{g} = \langle r\rangle + g_0, \quad \tilde{e} = e_0 + N/2$$

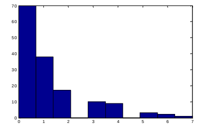
$$\tilde{f} = f_0 + \frac{\langle\boldsymbol{\psi}^T\boldsymbol{\psi}\rangle}{2} - \langle\boldsymbol{\psi}\rangle^T\mathbf{X}\langle\boldsymbol{\beta}\rangle + \frac{\text{tr}[\mathbf{X}\langle\boldsymbol{\beta}\boldsymbol{\beta}^T\rangle\mathbf{X}^T]}{2}$$

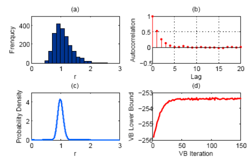
$$\tilde{c}_p = c_0 + 1/2, \quad \tilde{d}_p = d_0 + \langle\beta_p^2\rangle/2$$

$$\langle\omega_i\rangle = \mathbb{E}_{r,\psi_i}[\mathbb{E}[\omega_i|r,\psi_i,y_i]] = (y_i + \langle r\rangle)\left\langle\frac{\tanh(\psi_i/2)}{2\psi_i}\right\rangle$$

Experiments

❑Univariate count data analysis





❑Count regression

Table 1. The MLEs or posterior means of the lognormal variance parameter σ^2, NB dispersion parameter r, quasi-dispersion κ and regression coefficients $\boldsymbol{\beta}$ for the Poisson, NB and LGNB regression models on the NASCAR dataset, using the MLE, VB or Gibbs sampling for parameter estimations.

Model Parameters	Poisson (MLE)	NB (MLE)	LGNB (VB)	LGNB (Gibbs)
σ^2	N/A	N/A	0.1396	0.0289
r	N/A	5.2484	18.5825	6.0420
β_0	-0.4903	-0.5038	-3.5271	-2.1680
β_1 (Laps)	0.0021	0.0017	0.0015	0.0013
β_2 (Drivers)	0.0516	0.0597	0.0674	0.0643
β_3 (TrkLen)	0.6104	0.5153	0.4192	0.4200

Table 2. Test of goodness of fit with Pearson residuals.

Models (Methods)	NASCAR	MotorIns
Poisson (MLE)	655.6	485.6
NB (MLE)	138.3	316.5
IG-Poisson (MLE)	N/A	319.7
LGNB ($r \equiv 1000$, Gibbs)	**117.8**	296.7
LGNB(VB)	126.1	**275.5**
LGNB(Gibbs)	129.0	284.4

LGNB (VB) Correlation matrix for $(\beta_1, \beta_2, \beta_3)^T$

$$\begin{pmatrix} 1.0000 & -0.4824 & 0.8933 \\ -0.4824 & 1.0000 & -0.7171 \\ 0.8933 & -0.7171 & 1.0000 \end{pmatrix}$$

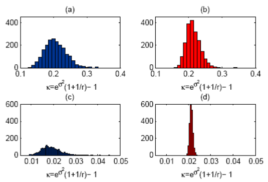


Figure 2. The histograms of the quasi-dispersion $\kappa = e^{\sigma^2}(1+1/r)-1$ based on (a) the 2000 collected Gibbs samples for NASCAR, (b) the 2000 simulated samples using the VB Q functions for NASCAR, (c) the 2000 collected Gibbs samples for MotorIns, and (d) the 2000 simulated samples using the VB Q functions for MotorIns.

Future work under the lognormal-gamma-NB framework

Multivariate count regression
Log Gaussian process
Mixture modeling, topic modeling