Lognormal and Gamma Mixed Negative Binomial Regression

Mingyuan Zhou

Joint work with Lingbo Li, David Dunson, Lawrence Carin

Duke University, Durham NC, USA

ICML, Edinburgh, June 27 2012

Count Data

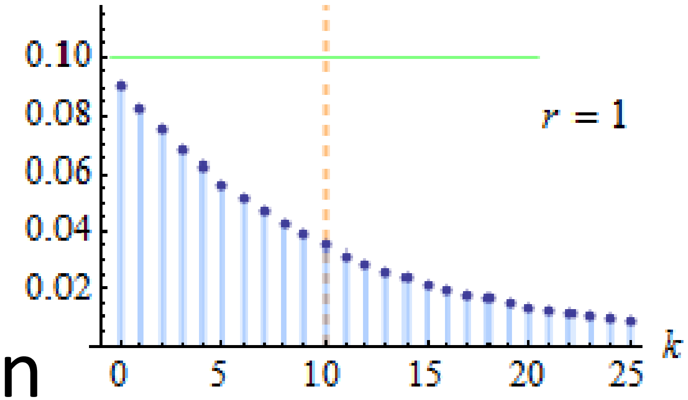
- Count data
  - Number of auto issuance claims
  - Next generation sequencing
  - **Number of points in a cluster (mixture model)**
  - **Number of words in document *j* assigned to topic *k* (topic model, mixture membership model)**
- Overdispersion: Variance > Mean
  - Heterogeneity: difference between individuals
  - Contagion: dependence between the occurrence of events
- Poisson distribution (variance = mean)
- Negative binomial distribution (variance $\geq$ mean)

Negative binomial distribution

$$X \sim \mathrm{NB}(r, p)$$

- Gamma Poisson mixture distribution

$$\begin{aligned} f_X(k) &= \int_0^\infty \mathrm{Pois}(k;\lambda)\mathrm{Gamma}\left(\lambda; r, \frac{p}{1-p}\right) d\lambda \\ &= \frac{\Gamma(r+k)}{k!\Gamma(r)}(1-p)^r p^k \end{aligned}$$



- Compound Poisson distribution

$$X = \sum_{\ell=1}^{L} Y_\ell, \quad L \sim \mathrm{Pois}(-r\log(1-p)), \quad Y_\ell \sim \mathrm{Log}(p)$$

- Variance $\mu + r^{-1}\mu^2 \geq$ Mean μ

The figure is from http://en.wikipedia.org/wiki/File:Negbinomial.gif

Poisson regression

- Poisson regression
  - Model:

$$y_i \sim \mathrm{Pois}(\lambda_i), \quad \lambda_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

$$\boldsymbol{x}_i = [1, x_{i1}, \cdots, x_{iP}]^T$$

  - Model assumption (equal-dispersion):

$$\mathbb{E}[y_i | \boldsymbol{x}_i] = \mathrm{Var}[y_i | \boldsymbol{x}_i] = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

Poisson regression with random effect

- Multiplicative random effect:

$$y_i \sim \mathrm{Pois}(\lambda_i), \quad \lambda_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})\epsilon_i$$

$$\mathbb{E}[y_i|\boldsymbol{x}_i] = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})\mathbb{E}[\epsilon_i] \quad \mathrm{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \frac{\mathrm{Var}[\epsilon_i]}{\mathbb{E}^2[\epsilon_i]}\mathbb{E}^2[y_i|\boldsymbol{x}_i]$$

- Negative binomial regression (gamma random effect):

$$\epsilon_i \sim \mathrm{Gamma}(r, 1/r) = \frac{r^r}{\Gamma(r)}{\epsilon_i}^{r-1}e^{-r\epsilon_i}$$

$$\mathrm{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \phi\mathbb{E}^2[y_i|\boldsymbol{x}_i] \qquad \phi = r^{-1}$$

- Lognormal-Poisson regression (lognormal random effect):

$$\epsilon_i \sim \ln\mathcal{N}(0, \sigma^2)$$

$$\mathrm{Var}[y_i|\boldsymbol{x}_i] = \mathbb{E}[y_i|\boldsymbol{x}_i] + \left(e^{\sigma^2} - 1\right)\mathbb{E}^2[y_i|\boldsymbol{x}_i]$$

Lognormal & gamma mixed NB regression

- Lognormal-gamma mixed NB regression

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

$$r \sim \mathrm{Gamma}(a_0, 1/h), \qquad \epsilon_i \sim \ln \mathcal{N}(0, \varphi^{-1})$$

$$\mathrm{logit}(p_i) = \ln \frac{p_i}{1-p_i}$$

- Lognormal-gamma-gamma-Poisson

$$y_i \sim \mathrm{Pois}(\lambda_i), \quad \lambda_i \sim \mathrm{Gamma}\left(r, \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})\epsilon_i\right)$$

$$r \sim \mathrm{Gamma}(a_0, 1/h), \qquad \epsilon_i \sim \ln \mathcal{N}(0, \varphi^{-1})$$

LGNB regression

- Properties:

$$\mathbb{E}[y_i|\boldsymbol{x}_i] = \mathbb{E}_{\epsilon_i}[\mathbb{E}[y_i|\boldsymbol{x}_i, \epsilon_i]] = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta} + \sigma^2/2 + \ln r)$$

$$\begin{aligned}\mathrm{Var}[y_i|\boldsymbol{x}_i] &= \mathbb{E}_{\epsilon_i}[\mathrm{Var}[y_i|\boldsymbol{x}_i,\epsilon_i]] + \mathrm{Var}_{\epsilon_i}[\mathbb{E}[y_i|\boldsymbol{x}_i,\epsilon_i]] \\ &= \mathbb{E}[y_i|\boldsymbol{x}_i] + \left(e^{\sigma^2}(1+r^{-1}) - 1\right)\mathbb{E}^2[y_i|\boldsymbol{x}_i].\end{aligned}$$

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

Default Bayesian Analysis using Data Augmentation

- Inferring the NB dispersion parameter *r*

$$y_i \stackrel{iid}{\sim} \mathrm{NB}(r,p), \;\; r \sim \mathrm{Gamma}(a, 1/b)$$

  - Compound Poisson representation of $y \sim \mathrm{NB}(r,p)$

$$y = \sum_{\ell=1}^{L} u_\ell, \;\; L \sim \mathrm{Pois}(-r\ln(1-p)), \;\; u_\ell \stackrel{iid}{\sim} \mathrm{Log}(p)$$

  - Conjugate updates under augmentation

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

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

Inferring the regression coefficients

- Inferring the regression coefficients

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

  - Polya-Gamma random variable

$$\omega_i \sim \mathrm{PG}(y_i + r, 0)$$

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

  - Likelihood

$$\mathcal{L}(\psi_i) \propto \frac{(e^{\psi_i})^{y_i}}{(1+e^{\psi_i})^{y_i+r}} = \frac{2^{-(y_i+r)}\exp(\frac{y_i-r}{2}\psi_i)}{\cosh^{y_i+r}(\psi_i/2)}$$

$$\propto \exp\left(\frac{y_i - r}{2}\psi_i\right)\mathbb{E}_{\omega_i}\left[\exp(-\omega_i\psi_i^2/2)\right].$$

Inferring the regression coefficients

- Conditional posterior of ψ

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

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

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

- Conditional posterior of ω_i

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

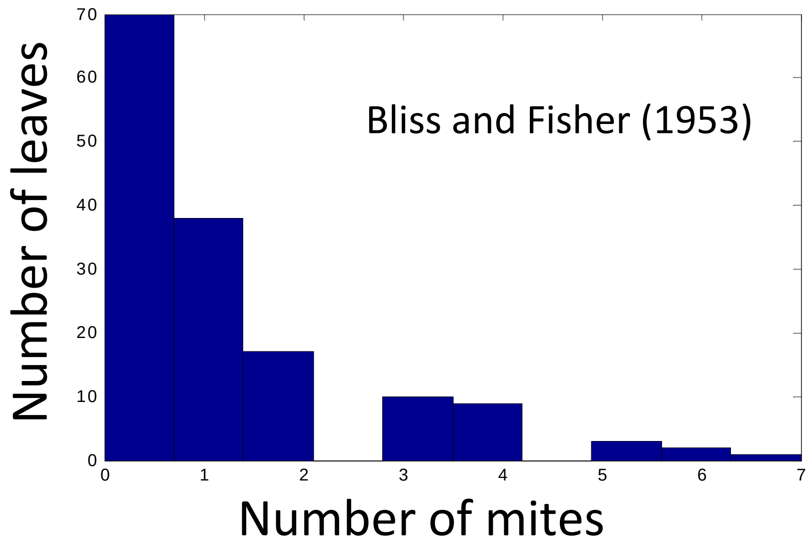
$$(\omega_i|-) \sim \mathrm{PG}(y_i + r, \psi_i)$$

Inference

- Gibbs Sampling
  - L_i, Multinomial
  - r , Gamma
  - ω_i, Polya-gamma
  - $\boldsymbol{\psi}$, Normal
  - $\boldsymbol{\beta}$, Normal
  - Hyper-parameters

- Variational Bayes

Example results: univariate count analysis



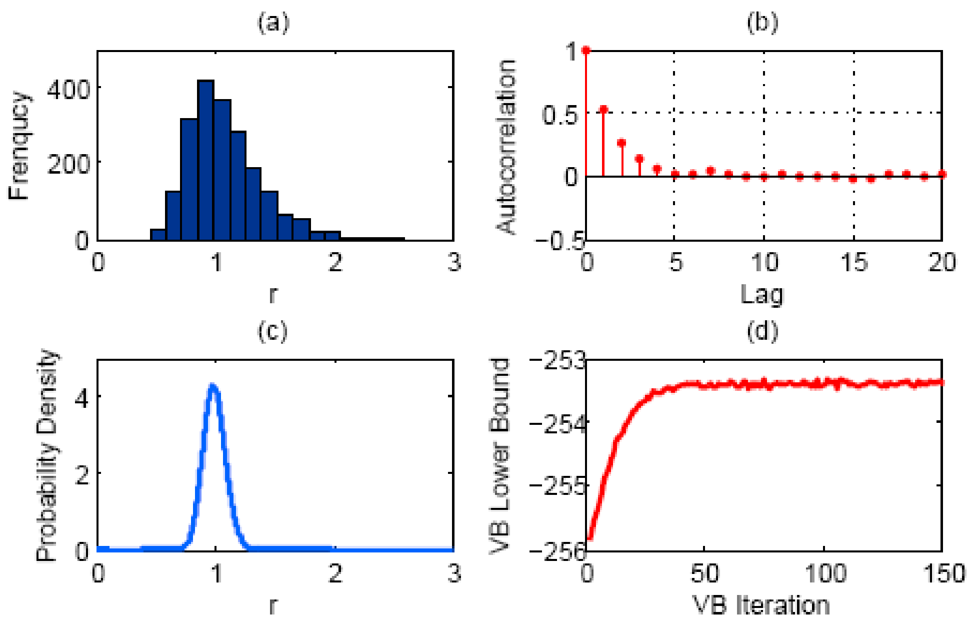
$$x_i \overset{iid}{\sim} \mathrm{NB}(r, p), \;\; i = 1, \cdots, N$$

$$r \sim \mathrm{Gamma}(a, 1/b)$$

$$p \sim \mathrm{Beta}(\alpha, \beta).$$

Gibbs sampling

Variational Bayes



Count regression for NASCAR

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

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

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

Test of goodness of fit

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

$$E = \sum_{i=1}^{N} e_i^2, \quad e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i(1 + \hat{\kappa}\hat{\mu}_i)}}$$

Posterior of the quasi-dispersion

Gibbs sampling Variational Bayes

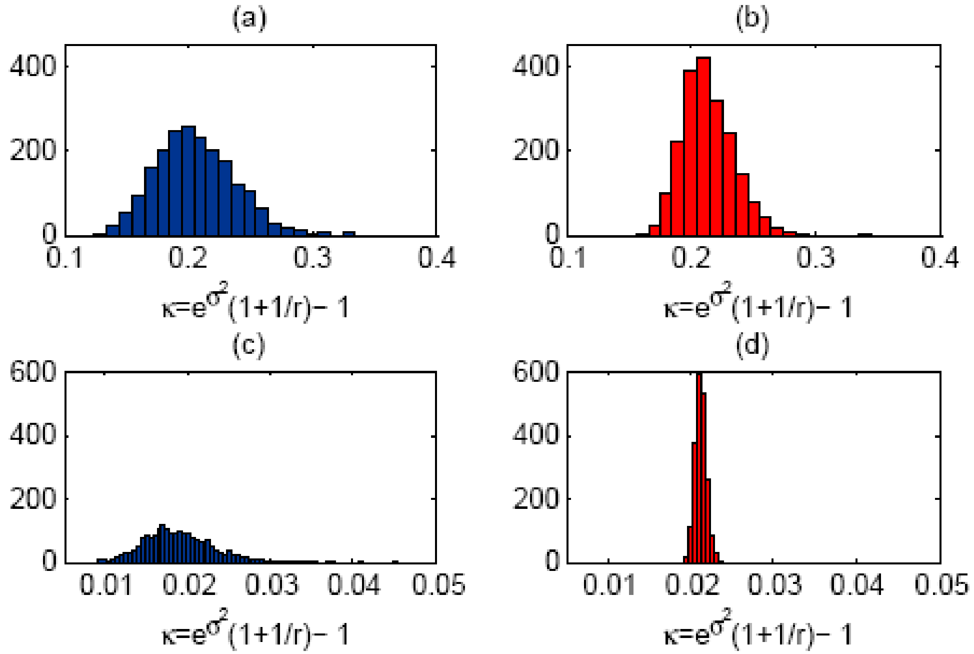


Figure 2. The histograms of the quasi-dispersion $\kappa = e^{\sigma^2}(1+1/r)-1$ based on (a) the 2000 collected Gibbs samples for NASCAR, (b) the 2000 simulated samples using the VB Q functions for NASCAR, (c) the 2000 collected Gibbs samples for MotorIns, and (d) the 2000 simulated samples using the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

Conclusions

- Lognormal & gamma mixed NB regression
- Compound Poisson, Polya-Gamma
- Closed-form Gibbs sampling and VB
- Future directions under the lognormal-gamma-NB framework:
  - Multivariate count regression
  - Log Gaussian process
  - Mixture modeling, topic modeling