Robustness of Estimators When Fitting Mixed Poisson Regression Models

Kimberly S. Weems
North Carolina Central University, Durham, North Carolina

Paul J. Smith
University of Maryland, College Park, Maryland

The LmB Conferences on Multivariate Count Data
Besançon, France
July 4-6, 2018

Outline

- Mixed Poisson Models
- Influence Functions
- Robustness of MLE's
  - Poisson-gamma (P-Γ)
  - Poisson-inverse Gaussian (P-IG)
  - Poisson-lognormal (P-LN)
- Simulation Results and Example

Mixed Poisson Model (for a single covariate X_i)

Poisson mixed models are often used to analyze count data that exhibit over-dispersion.

- $Y_i | X_i, U_i \sim \text{Poisson}(U_i \mu_i)$
- $\mu_i = \mu(X_i) = \exp(\beta_0 + \beta_1 X_i)$ is the mean of Y_i
- Random effects $U_i \sim F$
- $\mathbb{E}(U_i) = 1$ and $\text{Var}(U_i) = \tau$
- U_i and X_i are independent
- Let $\theta = (\beta_0, \beta_1, \tau)$. Marginal density of Y_i is

$$P_{y_i} = P(Y_i = y_i; X_i, \theta) = \int_0^\infty \frac{(u\mu_i)^{y_i} \exp(-u\mu_i)}{y_i!} f(u) du,$$

for $y_i = 0, 1, 2, \cdots$ with marginal mean μ_i and variance $\mu_i(1 + \mu_i \tau)$.

Motivation

- The distribution of the random effects, known as the mixing distribution, is often chosen for computational convenience.
- This assumption motivated us to consider the robustness of the maximum likelihood estimators (MLEs) of θ when the mixing distribution is contaminated (i.e., an $\epsilon-$contamination of a specified parametric family).
- Different perspectives:
  - McCulloch and Neuhaus (2013), Heagerty and Zeger (2000) – minimal impact
  - Litière et al. (2008) – greater sensitivity

Statement of Problem

- Suppose that the nominal distribution function F is perturbed by a contaminating distribution function G, so that

$$U_i \sim (1-\epsilon)F + \epsilon G.$$

- Our goal is to determine the effect of mixing distribution misspecification on the MLE $\hat{\theta}$, where $\hat{\theta}$ solves

$$\nabla_\theta \log L(\theta; \mathbf{Y}) = 0,$$

$L(\theta; \mathbf{Y}) = \prod_{i=1}^{n} P_{y_i}$ for $\mathbf{Y} = (Y_1, \cdots, Y_n)$.

The Influence Function

- A Gâteaux derivative which characterizes the local stability of an estimator in the presence of perturbations.
- For MLE's

$$\mathbf{IF}(u;\hat{\theta},F) = I^{-1}(\theta)s(\theta;u)$$

where I is Fisher's information matrix, s is the conditional expected score matrix, and

$$\mathbf{IF}(u;\hat{\theta},F) = \begin{bmatrix} IF(u;\hat{\beta}_0,F) \\ IF(u;\hat{\beta}_1,F) \\ IF(u;\hat{\tau},F) \end{bmatrix}.$$

The Influence Function for MLEs

- We extend the infinitesimal approach based on the Influence Function of Hampel (1974) and of Gustafson (1996).

- Let $\ell = \log L(\theta; \mathbf{Y})$. Consider the Fisher information matrix

$$I(\theta) = -\mathbb{E}\left[\nabla_{\theta}^{\otimes 2}\ell(\theta; \mathbf{Y})\right]$$

  and the conditional expected score matrix

$$s(\theta; u) = \mathbb{E}\left[\nabla_{\theta}\ell(\theta; \mathbf{Y}) \middle| U_i = u\right],$$

  where all expectations are taken with respect to the nominal distribution F.

- The MLE $\hat{\theta}$ is robust against mixing distribution misspecification if

$$\int_0^{\infty} \mathbf{IF}(u; \hat{\theta}, F) G(du) = \int_0^{\infty} I^{-1}(\theta) s(\theta; u) G(du)$$

  is uniformly bounded in G.

The Integral of the Influence Function

$$\int_0^\infty \mathbf{IF}(u;\hat{\theta},F)G(du) = \int_0^\infty I^{-1}(\theta)s(\theta;u)G(du)$$

- The integrand $I^{-1}(\theta)s(\theta;u)$ is an influence function for misspecification of the mixing distribution F
- integrating this quantity captures the effect of a contaminating distribution G on the MLE of θ.
- gives a first-order approximation to the asymptotic bias in estimating θ that is introduced by the ϵ-contamination of F by a distribution G.

The Integral of the Influence Function

In order to bound $I^{-1}(\theta)s(\theta;u)G(du)$, we bound $\int_0^\infty s(\theta;u)G(du)$ by using properties of well-known functions:

- Poisson-gamma............................Digamma Function
- Poisson-inverse Gaussian...............Bessel Function
- Poisson-lognormal........................Lambert W Function

Assumptions

A-1. Let $F, G \in \mathcal{F}$, where

$$\mathcal{F} = \{F|F \text{ is a cdf on}(0,\infty), \int uF(du) = 1, \int u^2 F(du) = 1+\tau\}.$$

A-2. $n^{-1}I \longrightarrow I^*$ as $n \longrightarrow \infty$, where I^* is positive definite.

A-3. The covariates $X_1, X_2, \cdots, X_N$ are an i.i.d. sample from a nondegenerate distribution whose support is a compact region of $\mathbb{R}^p$. Additionally, Cov($\mathbf{X}$) is positive definite, where $\mathbf{X} = (X_1, X_2, \cdots, X_n)^T$.

A-4. The interchange of expectation and differentiation of the log-likelihood and its derivatives is permitted.

Poisson-Inverse Gaussian Model

$$\mathbf{IF}(u;\hat{\theta},F) = I^{-1}(\theta)s(\theta;u),$$

where $s(\theta;u) =$

$$\begin{pmatrix} \mathbb{E}_{Y|X,U}\left[Y - (Y+1)\frac{P_{y+1}}{P_y}\middle| U=u\right] \\ X\,\mathbb{E}_{Y|X,U}\left[Y - (Y+1)\frac{P_{y+1}}{P_y}\middle| U=u\right] \\ \tau^{-2}(1+\tau\mu)\,\mathbb{E}_{Y|X,U}\left[\mu^{-1}(Y+1)\frac{P_{y+1}}{P_y} - 1\middle| U=u\right] \end{pmatrix}.$$

Bessel Function Approach

- Let $\nu = \{\tau^{-1}(\tau^{-1} + 2\mu)\}^{1/2}$. Then

$$(y+1)\frac{P_{y+1}}{P_y} = (1+2\tau)^{-1/2}\frac{K_{y+\frac{1}{2}}(\nu)}{K_{y-\frac{1}{2}}(\nu)},$$

  where $K_k(\nu)$ denotes the modified spherical Bessel function of the third kind of order $k \in \mathbb{R}$.
- We use properties of Bessel functions and asymptotic results to find bounds for the P-IG probability ratios.

Asymptotics: Definition and Notation

- A positive function L, defined on $[0, \infty)$, varies slowly at infinity if, for all $c > 0$,

$$\lim_{x \to \infty} L(cx)/L(x) = 1$$

.

- The notation $a(x) \sim b(x)$, $x \to \infty$, means

$$\lim_{x \to \infty} a(x)/b(x) = 1$$

.

Asymptotic Behavior of P-IG Probability Ratios

Theorem (Willmot 1990)

Let P_y denote probabilities of a mixed Poisson distribution so that

$$P_y = \int_0^\infty \frac{(\lambda x)^y \exp(-\lambda x)}{y!} f(x) dx.$$

Suppose that $f(x) \sim C(x) x^\alpha \exp(-\beta x), \quad x \to \infty$, *where* $C(x)$ *is a locally bounded function on* $(0, \infty)$ *which varies slowly at infinity,* $\beta \geq 0$, *and* $-\infty < \alpha < \infty$ *(with* $\alpha < -1$ *if* $\beta = 0$*). Then* P_y *satisfies*

$$P_y \sim \frac{C(y)}{(\lambda + \beta)^{\alpha+1}} \left(\frac{\lambda}{\lambda + \beta} \right)^y y^\alpha, \quad y \to \infty.$$

Asymptotic Behavior of P-IG Probability Ratios

Proposition

For the P-IG$(1, 1/\tau)$ distribution

$$\lim_{y \to \infty} \frac{P_{y+1}}{P_y} = \frac{2\mu\tau}{2\mu\tau + 1}.$$

This result follows from (Willmot 1990).

Bounds for the Ratios of P-IG Probabilities

Using properties of Bessel functions, we compute

- A lower bound

$$(y+1)\frac{P_{y+1}}{P_y} \geq (1+2\tau)^{-1/2}\left(\frac{1}{1+2\nu}\right)^y$$

- An upper bound

$$(y+1)\frac{P_{y+1}}{P_y} \leq (1+2\tau)^{-1/2}(1+\nu^{-1})(2y-1)$$

Poisson-Inverse Gaussian Model

Theorem

Let $\mu = \mu(X)$ and $\nu = +\sqrt{\tau^{-1}(\tau^{-1} + 2\mu)}$. The integral of the conditional expected score function for β_0, given by

$$\int_0^\infty \mathbb{E}_{Y|X,U}\left\{ Y - (Y+1)\frac{P_{Y+1}}{P_Y} \middle| U = u \right\} G(du),$$

is bounded above by

$$\mu - (1+2\tau)^{-1/2}\left[1 + \mathcal{L}_G\left(\frac{2\nu\mu}{1+2\nu}\right)\right]$$

and below by

$$\mu - (1+2\tau)^{-1/2}(1+\nu^{-1})(2\mu - 1),$$

where $\mathcal{L}_G(\cdot)$ denotes the Laplace transform of G.

Similar results hold for β_1 and τ.

Simulation Study

- Generate 1000 *iid* covariates $\{X_i\} \sim \mathrm{N}(0, 1)$ that are fixed throughout the study and form $\mu_i = \exp\{\beta_0 + X_i\beta_1\}$.
- Generate 1000 random effects $\{U_i\}$, from either the IG or a misspecified mixing distribution distribution, such that $\mathbb{E}(U_i) = 1$ and $\mathrm{Var}(U_i) = \tau$.
- Generate 1000 sample responses $\{Y_i\}$, conditionally on U_i and X_i, such that

$$\mathbb{E}(Y_i|X_i, U_i) \sim U_i \exp\{\beta_0 + X_i\beta_1\}.$$

- Use `gamlss` (Stasinopoulos et al. 2017) to estimate β_0, β_1, and τ via ML under an assumed IG mixing distribution.

Table 1: True and estimated parameter values under assumed inverse Gaussian mixing distribution, $\epsilon = .01$ and $\epsilon = 1$.

Parameter	β_0	β_1 $\epsilon = 0.01$	τ	β_0	β_1 $\epsilon = 1$	τ
True	0.5	1	0.25	0.5	1	0.25
Estimated	0.4987	1.0003	0.2404	0.5021	0.9966	0.2609
S.E.($\times 10^2$)	(3.25)	(2.92)	(3.47)	(3.26)	(3.12)	(3.76)
True	0.5	1	0.5	0.5	1	0.5
Estimated	0.4993	0.9998	0.4859	0.4982	0.9970	0.5651
S.E.($\times 10^2$)	(3.65)	(3.51)	(5.71)	(3.77)	(3.72)	(6.59)
True	0.5	1	1	0.5	1	1
Estimated	0.4980	1.0022	0.9626	0.5005	0.9811	1.2965
S.E.($\times 10^2$)	(4.26)	(4.30)	(10.39)	(4.67)	(4.46)	(14.05)
True	0.5	1	2	0.5	1	2
Estimated	0.4953	1.0003	1.9093	0.4985	0.9424	3.1598
S.E.($\times 10^2$)	(5.29)	(5.26)	(20.87)	(6.38)	(5.76)	(36.91)

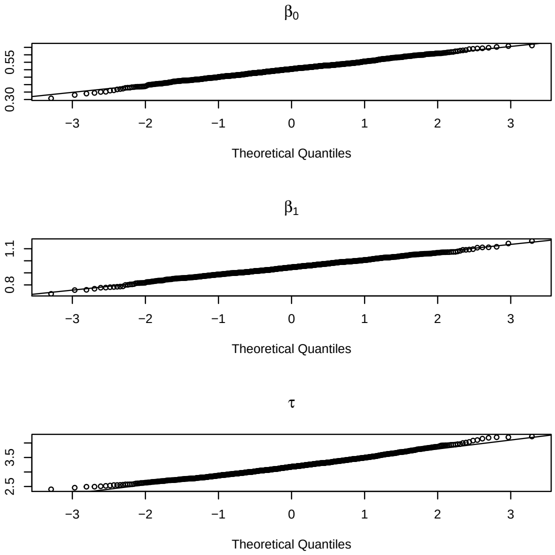

Figure 1: QQ plots of 1000 simulated MLEs for Poisson-gamma data under assumed IG mixing distribution with $\beta_0 = 0.5$, $\beta_1 = 1$, and $\tau = 2$.

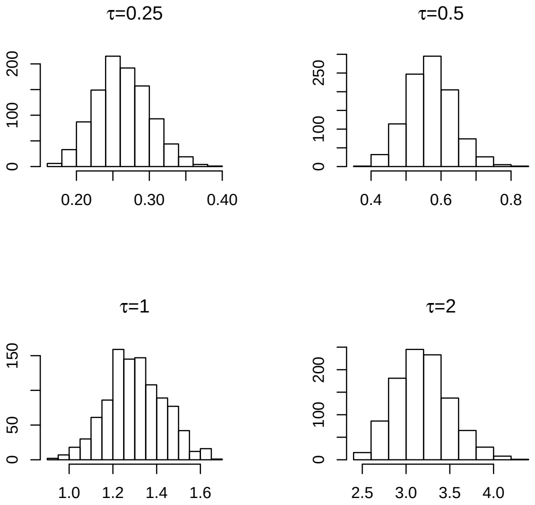

Figure 2: Plots of 1000 simulated MLEs of τ for Poisson-gamma data under assumed PIG model, $\beta_0 = 0.5$, $\beta_1 = 1$, $\tau = (0.25, 0.5, 1, 2)$.

Example: Crash Data Analysis

- Response data on the number of crashes at a particular intersection
- Two covariates X_1 and X_2 (unnamed for privacy reasons)
- $n = 868$
- Dispersion index $= 8.7$, indicating overdispersion
- Poisson-Gamma (i.e., negative binomial) and Poisson-IG regression models were fitted

Data provided by D. Lord and S. Geedipally of the Texas A&M Transportation Institute, College Station, Texas, USA.

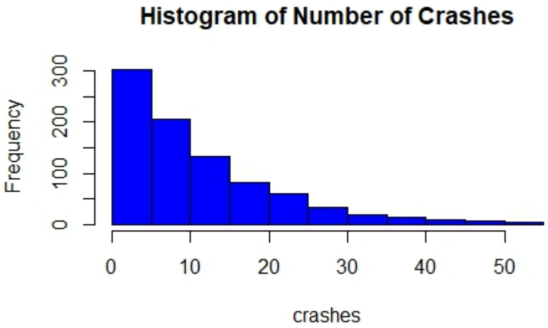

Figure 3: Histogram of number of crashes at a particular area.

Crash Data Analysis

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}$
P-IG	2.9263	1.0000	1.0000	0.1980
	(0.0261)	(0.0000)	(0.0000)	(0.0003)
				AIC=5211.12
NB	2.9039	1.0000	1.0000	0.1852
	(0.0265)	(0.0000)	(0.0000)	(0.0002)
				AIC=5204.23

Summary for P-IG Model

- For the regression parameter estimates, our study supports conclusions of those who argue that the choice of the mixing distribution has little impact on the MLEs.
- However, for the MLE of the variance of the random effects distribution, our simulation results suggest that mixing distribution misspecification can have a substantial impact
- One can be reasonably confident when using ML to estimate regression parameters. However, if one believes that modeling assumptions may be incorrect, alternatives to MLE should be considered, particularly when estimating the random effects variance.

Poisson-lognormal Model

$$\mathbf{IF}(u;\hat{\theta},F) = I^{-1}(\theta)s(\theta;u),$$

where

$$s(\beta_0;u) = \mathbb{E}_{Y|X,U}\left\{ Y-(Y+1)\frac{P_{y+1}}{P_y}\right| U=u\Bigg\},$$

$$s(\beta_1;u) = \mathbb{E}_{Y|X,U}\left\{ X\left(Y-(Y+1)\frac{P_{y+1}}{P_y}\right)\right| U=u\Bigg\}$$

and

$$\begin{aligned} s(\tau;u) &= (1+\tau)^{-1}\mathbb{E}_{Y|X,U}\Big\{(Y^2-Y/2)-(Y+1)(2Y+1/2)\tfrac{P_{y+1}}{P_y} \\ &\qquad + (Y+1)(Y+2)\tfrac{P_{y+2}}{P_y}|U=u\Big\} \end{aligned}$$

Poisson-lognormal Model

Let $\sigma^2 = \log(1+\tau)$. Then

$$P_y = \frac{\mu^y}{y!} \exp\{\sigma^2 y(y-1)/2\} H(y),$$

where

$$H(y) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma^2}(\sigma^2 v - w)^2 - \mu e^w\right\} dw,$$

$w = \log(u)$, and $v = y - 1/2$.

Poisson-lognormal Model

Ratios of P-LN probabilities can be written as follows:

$$\frac{P_{y+1}}{P_y} = \frac{\mu}{y+1} \exp\left\{\sigma^2 y\right\} \frac{H(y+1)}{H(y)}$$

and

$$\frac{P_{y+2}}{P_y} = \frac{\mu^2}{(y+1)(y+2)} \exp\left\{\sigma^2(2y+1)\right\} \frac{H(y+2)}{H(y)}.$$

Poisson-lognormal Model

- Saddlepoint techniques were used by de Bruijn (1953) to study the behavior of a function similar to H.
- The results of de Bruijn (1961) served as the foundation for the asymptotic theory of the Lambert W function, defined as

$$W(z) \exp\{W(z)\} = z.$$

- We study the asymptotic behavior of H using the methods of de Bruijn.

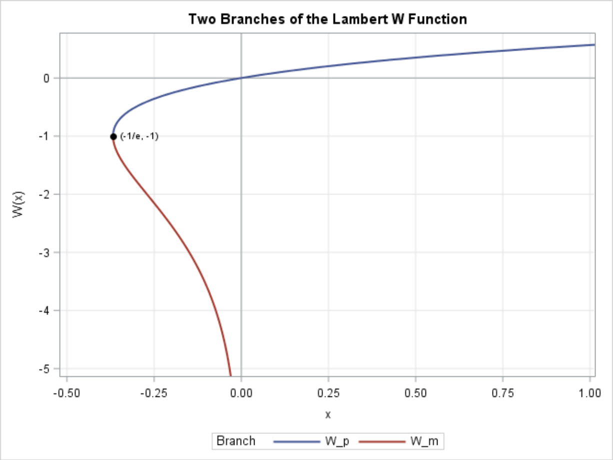

Figure 4: From https://blogs.sas.com/content/iml/2016/08/31/lambert-w-function-sas.html.

Poisson-lognormal Model

Approximating $H(y+\rho)/H(y)$ [sketch of proof]

- Use a power series expansion to obtain an asymptotic representation for H in terms of its saddlepoint.
- Obtain a first-order Taylor series approximation for the saddlepoint and substitute into H.
- Obtain an asymptotic representation for $H(y+\rho)/H(y)$ and determine the rate at which the ratios decrease.

Poisson-lognormal Model

Theorem

Let ρ be fixed. As $y \longrightarrow \infty$,

$$\log \frac{H(y+\rho)}{H(y)} = -\rho \left\{ \sigma^2 y - \log y + \frac{\log y}{\sigma^2 y} \right\} + O\left(\frac{1}{y}\right).$$

Corollary

There exist constants y_0 and C that depend only on μ and σ^2, such that if $y > y_0$,

$$\frac{H(y+\rho)}{H(y)} \leq C \exp\{-\rho\sigma^2 y\}.$$

Poisson-lognormal Model

Using Corollary 1, we have that

$$
\begin{aligned}
(y+1)\frac{P_{y+1}}{P_y} &= (y+1)\left[\frac{\mu}{y+1}\exp\left\{\sigma^2 y\right\}\frac{H(y+1)}{H(y)}\right] \\
&= \mu\exp\left\{\sigma^2 y\right\}\frac{H(y+1)}{H(y)} \\
&\leq \mu\exp\left\{\sigma^2 y\right\}C\exp\left\{-\sigma^2 y\right\} \\
&= \mu C
\end{aligned}
$$

Poisson-lognormal Model

Theorem

The integral of the conditional expected score function for β_0*, given by*

$$\int_0^\infty \mathbb{E}_{Y|X,U}\left\{ Y - (Y+1)\frac{P_{Y+1}}{P_Y} \,\middle|\, U = u \right\} G(du),$$

is uniformly bounded over $G \in \mathcal{F}$*.*

Similar results hold for β_1 and τ.

Summary of Results

- The integral of the conditional expected score matrix $s(\theta; u)$ is bounded uniformly in G.
- The integral of the Influence Function

$$\mathbf{IF}(u; \hat{\theta}, F) = I^{-1}(\theta) s(\theta; u)$$

  is bounded uniformly in G.
- The MLE $\hat{\theta}$ is robust against mixing distribution misspecification (small perturbations).
- Practically, the regression parameter estimators are robust; the random effects variance estimator may not be robust.

Selected References

Gustafson P. (1996) The Effect of Mixing-distribution Misspecification in Conjugate Mixture Models. *Canad J Stat* 24:307–318.

Litière S., Alonso A. & Molenberghs G. (2008) The Impact of a Misspecified Random-effects Distribution on the Estimation and the Performance of Inferential Procedures in Generalized Linear Mixed Models. *Stat Med,* 27(16):3125–3144.

McCulloch, C. E. & Neuhaus, J. M. (2011). Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. *Statistical Science*, 26(3), 388-402.

Stasinopoulos, D. & Rigby, R. A. (2007). Generalized Additive Models for Location, Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, 23(7), 1-46.

Weems, K. S. & Smith, P. J. (2018). Assessing the Robustness of Estimators when Fitting Poisson Inverse Gaussian Models. *Metrika.* https://doi.org/10.1007/s00184-018-0664-1.

Willmot G. E. (1990) Asymptotic Tail Behaviour of Poisson Mixtures with Applications. *Adv Appl Probab* 22:147–159.