Selection Of Random Effects Distributions In Mixed Counts Models with Quasi-Likelihood Estimation

Jamie Riggs

Northwestern University

IAA COIN

CCS-6 Statistical Sciences Seminar Series
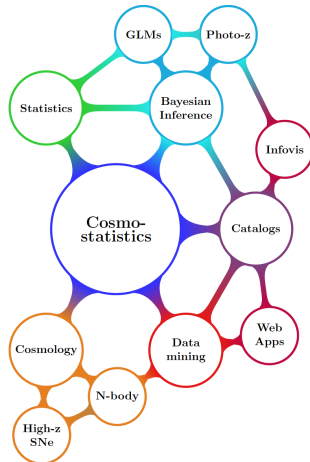**Los Alamos National Laboratory**

April 7, 2015

Outline

- Introduction
  - IAA/COIN
  - COIN progress
- Problem Genesis
- Models
- Proposition
- Estimation and Efficiency
- Conclusions
- Acknowledgments
  J.M. Hilbe, JPL
  T.L. Lalonde UNCO
  R.S. de Souza, MTA Eötvös University

IAA/COIN

International Astrostatistics Association (IAA)

- Researchers for statistical problems of astronomical data

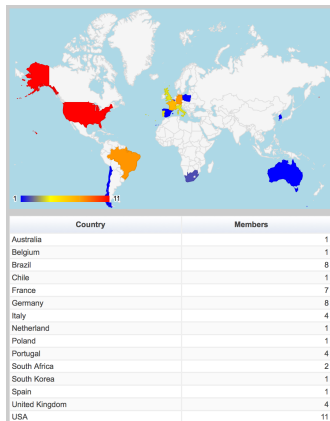 Astronomers   Physicists
  Statisticians   Computer Scientists

- https://asaip.psu.edu/organizations/iaa

Cosmostatistics Initiative (COIN)

- Objective: develop tools for exploring and modeling astronomical data

- Website: goo.gl/vCWA17

IAA/COIN

Country	Members
Australia	1
Belgium	1
Brazil	8
Chile	1
France	7
Germany	8
Italy	4
Netherlan	1
Poland	1
Portugal	4
South Africa	2
South Korea	1
Spain	1
United Kingdom	4
USA	11

	total	distinct		
nations	534	52		

area	Freq.	Percent	Cum	
Africa	50	9.36	9.36	
Asia	67	12.55	21.91	
Aust	14	2.62	24.53	
Europe	171	32.02	56.55	
NAmer	197	36.89	93.45	
SAmer	35	6.55	100.00	

subject	Freq	Percent	Cum	
stat	137	25.66	25.66	
astro	335	62.73	88.39	
compute	35	6.55	94.94	
physics	23	4.31	99.25	
other	4	0.75	100.00	

Gender	Freq	Percent	Cum	
male	425	79.59	79.59	
female	109	20.41	100.00	

IAA/COIN Completed Projects

https://asaip.psu.edu/organizations/iaa/iaa-working-group-of-cosmostatistics/the-cosmostatistics-initiative-coin

1. The Overlooked Potential of Generalized Linear Models in Astronomy - I: Binomial Regression.
   arxiv link: http://arxiv.org/abs/1409.7696.

2. The Overlooked Potential of Generalized Linear Models in Astronomy-II: Gamma regression and photometric redshifts.
   ASCL link: http://ascl.net/1408.018

IAA/COIN Currently Ongoing Projects

1. CosmoABC: Cosmological parameter estimation via PMC-ABC. To be submited to A&C
2. The Overlooked Potential of Generalized Linear Models in Astronomy-III: Bayesian Models for Count data
3. Gaussian Mixture Models for bi-modality studies in Globular Cluster
4. Multiple Survey Analysis and Supernova Cosmology
5. Unsupervised Learning and galactic stellar population

IAA/COIN

- Projects use https://www.overleaf.com for collaborative work on developing manuscripts. All members of a project have access to the Latex manuscript and can contribute as desired.
- COIN is part of the Astronomers Workbench (AWOB) http://awob.mpg.de/
- The AWOB is a collaborative forum and e-publication platform allowing project group members to share data, results, manuscripts and other project-related items throughout the life of the project.
- AWOB is itself a project of the Max Planck Digital Library (MPDL), with the Max Planck Institute for Astrophysics (MPA)) and the Max Planck Institute for Extraterrestrial Physics (MPE) as partners.
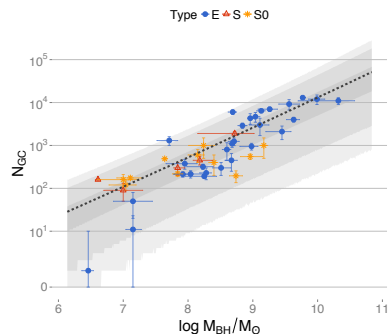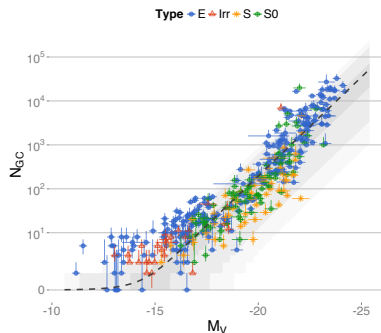
Problem Genesis

Sunspot Counts Cycles
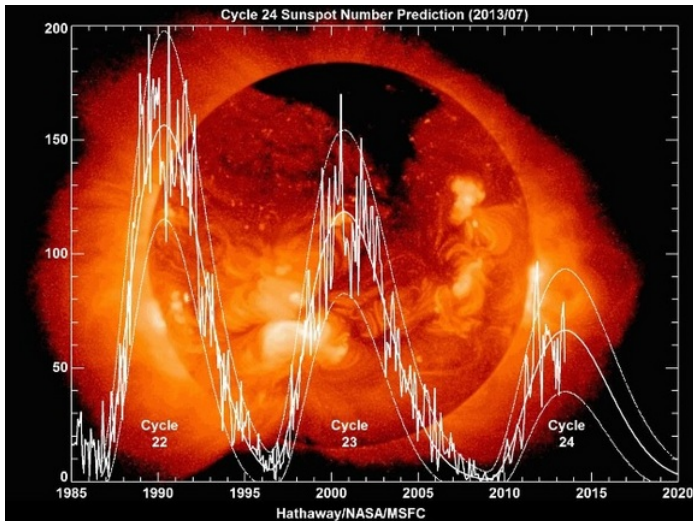
Mars Crater Counts (NASA/MOLA Science Team /D. Loizeau et al.)
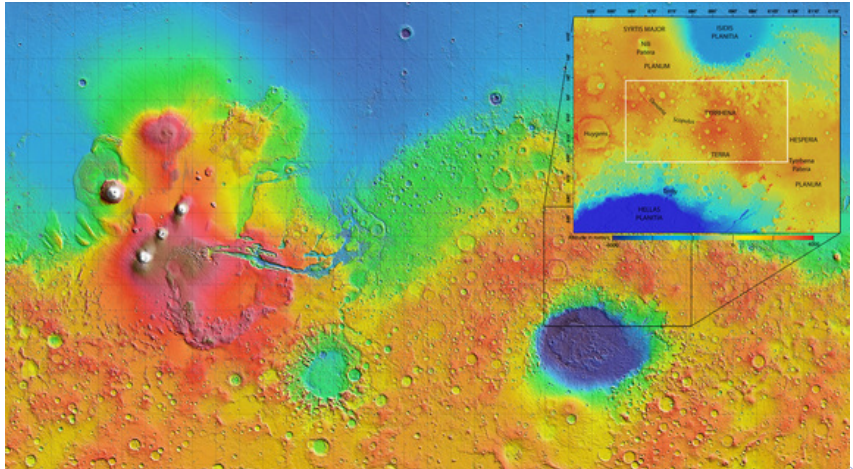
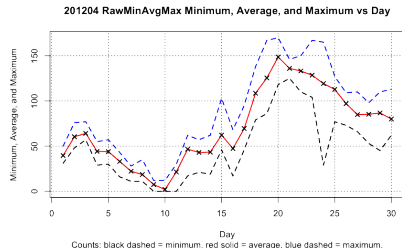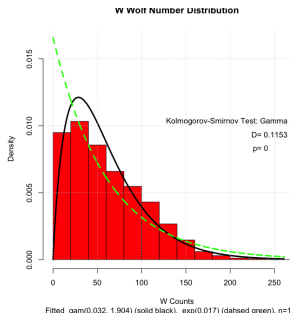# Globular Cluster Counts (COIN, in preparation)

Models

Sunspot Counts Models (partial cycle data)

- Sunspot counts, diurnal
- 60 observers in a month (random effect)
- Fixed effects
    - Seeing condition (E, G, F, P)
    - Location in sunspot cycle (year, month)
    - Observing experience (7 levels)

# Raw Sunspot Counts (typical monthly non-minima data)

W Wolf Number Distribution

Kolmogorov-Smirnov Test: Gamma
D= 0.1153
p= 0

Density

W Counts

Fitted gam(0.032, 1.904) (solid black), exp(0.017) (dahsed green), n=1

201204 RawMinAvgMax Minimum, Average, and Maximum vs Day

Minimum, Average, and Maximum

Day

Counts: black dashed = minimum, red solid = average, blue dashed = maximum.

$$y_{ij}^* \mid u_i \overset{\cdot}{\sim} \mathcal{N}(\mu_{ij}^*, \sigma_{ij}^{*2}), \quad u_i \overset{\cdot}{\sim} \mathcal{N}(\mu_{Ri}, \sigma_{Ri}^2)$$

For $Y|u \sim Poi(\boldsymbol{\mu}_{n \times 1})$, transform the response such that

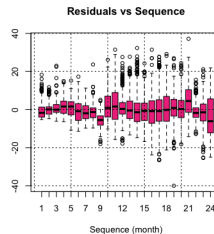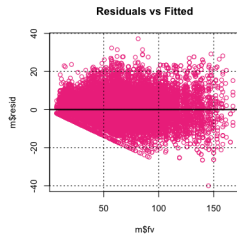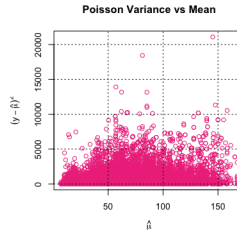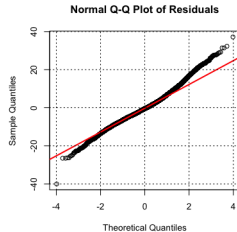| Effect	Mean Model	Dispersion Model
Fixed	$Y_{ij}^* \mid u_i \sim \mathcal{N}(\mu_{ij}^*, \sigma^{*2})$ $\mu_{ij}^* = x_{ij}^T \boldsymbol{\beta} + z_i^T u$ $y_{ij}^* = \ln(y_{ij} + 1)$	$\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^{*2})$
Random	$u_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_R^2)$	

$$y_{ij}^* \mid u_i \,\dot\sim\, \mathcal{N}(\mu_{ij}^*, \sigma_{ij}^{*2}), \quad u_i \,\dot\sim\, \mathcal{N}(\mu_{Ri}, \sigma_{Ri}^2), \quad \frac{Var(\hat{Y}_{ij}^*)}{\hat{Y}_{ij}^*} = 24.204$$
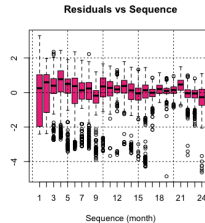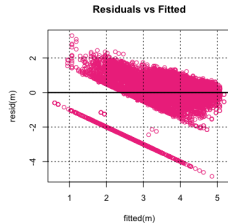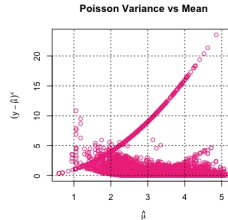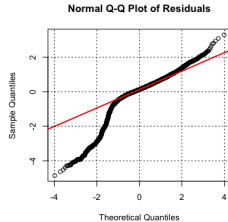
$$y_{ij} \mid u_i \overset{.}{\sim} \mathcal{P}oi(\mu_{ij}), \quad u_i \overset{.}{\sim} \mathcal{N}(\mu_{Ri}, \sigma_{Ri}^2)$$

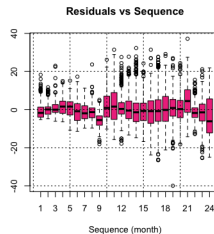| Effect	Mean Model	Dispersion Model
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij}\mu_{ij})$ $\eta_{ij} = x_{ij}^T \beta + z_i^T v$ $\eta_{ij} = g(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \gamma_0$ $\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim \mathcal{N}(\mu_{Ri}, \sigma_{Ri}^2)$ $v_i = g_R(u_i)$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$ $\eta_{dRi} = \delta_0$ $\eta_{dRi} = g_{dR}(\zeta_i)$

$$y_{ij} \mid u_i \overset{.}{\sim} \mathcal{P}oi(\mu_{ij}), \quad u_i \overset{.}{\sim} \mathcal{N}(\mu_{Ri}, \sigma_{Ri}^2), \quad \frac{Var(\hat{Y}_{ij})}{\hat{Y}_{ij}} = 20.010$$

$$y_{ij} \mid u_i \overset{\cdot}{\sim} \mathcal{P}oi(\mu_{ij}), \quad u_i \overset{\cdot}{\sim} gamma(\theta_i, \kappa_i)$$

Effect	Mean Model	Dispersion Model
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij}\mu_{ij})$ $\eta_{ij} = x_{ij}^T \beta + z_i^T v$ $\eta_{ij} = g(\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$ $\eta_{dij} = \gamma_0$ $\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim gamma(\theta_i, \kappa_i)$ $v_i = g_R(u_i)$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$ $\eta_{dRi} = \delta_0$ $\eta_{dRi} = g_{dR}(\zeta_i)$

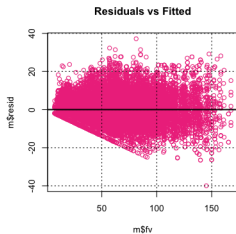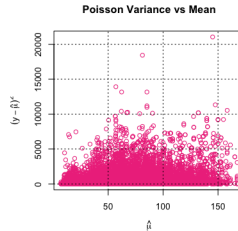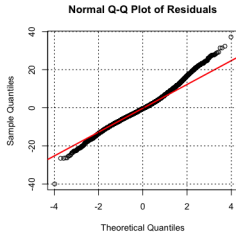$$y_{ij} \mid u_i \overset{\cdot}{\sim} \mathcal{P}oi(\mu_{ij}), \quad u_i \overset{\cdot}{\sim} gamma(\theta_i, \kappa_i), \quad \frac{Var(\hat{Y}_{ij})}{\hat{Y}_{ij}} = 20.010$$

Proposition

Proposition

Overdispersion of

- The conditional response mean model deviance residuals
- The random effects mean model deviance residuals

Contain information about random effects distributions

Propose

- A random effects mean-variance power-law function
- Estimate the mean-variance power-law function exponent ψ:

$$V_R(\mu_{Ri}) = \mu_{Ri}^{\psi} \tag{1}$$

from random effects deviance residuals (from HGLM IWLS estimation)

Rationale

- Counting (especially) large numbers of events is complex
- A stochastic component of this complex system is observer
- An exponential family mean-variance (first and second moment) specification is estimable via quasi-likelihood
- A mean-variance power-law function suggests stochastic process clustering structures
- Mean-variance power-law function is scale invariant thus accommodating small to large numbers of observers

HGLM random effects power-law mean-variance quasi distribution

Effect	Mean Model	Dispersion Model
Fixed	$Y_{ij} \mid u_i \sim (\mu_{ij}, \phi_{ij}\mu_{ij})$	$d_{ij} \sim (\phi_{ij}, 2\phi_{ij}^2)$
	$\eta_{ij} = x_{ij}^T \beta + z_i^T v$	$\eta_{dij} = \gamma_0$
	$\eta_{ij} = g(\mu_{ij})$	$\eta_{dij} = g_d(\phi_{ij})$
Random	$u_i \sim (\mu_{Ri}, \zeta_i \mu_{Ri}^{\psi})$	$d_{R_i} \sim (\zeta_i, 2\zeta_i^2)$
		$\eta_{dRi} = \delta_0$
	$v_i = g_R(u_i)$	$\eta_{dRi} = g_{dR}(\zeta_i)$

Random effects deviance residuals

From (Lee and Nelder, 1996):

- For a power-law exponent ψ
- The random effects i^{th} deviance residual d_{Ri} is defined as

$$d_{Ri} = \int_{\hat{\mu}_{Ri}}^{\xi_i} \frac{\xi_i - s}{s^\psi} ds \qquad (2)$$

- ξ_i is the ith random effects pseudo model response
- $\hat{\mu}_{Ri}$ is the ith estimated mean of the random effects
- s is the scaling variable of integration

Random effects deviance residual expected value

The expected value of the ith random effects deviance residual is

$$\mathcal{E}(d_{Ri}) = 2 \left[\frac{\mathcal{E}(\xi_i^{2-\psi}) - (2-\psi)\mu_{Ri}^{1-\psi}\mathcal{E}(\xi_i) + (1-\psi)\mu_{Ri}^{2-\psi}}{(1-\psi)(2-\psi)} \right] \quad (3)$$

Power normal distribution

Power normal distribution models random effects deviance residuals

$$d_{Ri} \overset{\cdot}{\sim} \mathcal{PN}(\lambda, f_R(\mu_{Ri}), V_R(\mu_{Ri})) \tag{4}$$

- Right skewed
- Left truncated (λ the cut point)
- Degenerates to normal as $\lambda \to 0$
- λ estimated from data as a Box-Cox transformation exponent
- PN deviance residuals model parameters characterize the random effects power-law function exponent

Power normal distribution expected value

The first moment of the power normal distribution is estimated as (Freeman and Modarres, 2006)

$$\mathcal{E}(d_{Ri}) = (\lambda d_{\lambda i}^{\lambda} + 1)^{\frac{1}{\lambda}}(1 - \lambda^2)$$
$$= d_{Ri}(1 - \lambda^2) \tag{5}$$

Where λ is the PN distribution truncation parameter

Parameter estimation

The data allow the estimation of

- The random effects deviance residual d_{Ri}
- The power normal distribution parameter λ
- The random effects mean μ_{Ri}
- The pseudo-response ξ_i

From which the random effects mean-variance power-law function exponent ψ may be estimated

Mathematics and Analysis

- Mathematical development

- Myers's and Montgomery's (2002) Rats data

- Sunspot counts data supplied by the American Association of Variable Star Observers Solar Section

Mathematical development

$$2\left[\frac{\mathcal{E}(\xi^{2-\psi}) - (2-\psi)\mu_{Ri}^{1-\psi}\mathcal{E}(\xi_i) + (1-\psi)\mu_{Ri}^{2-\psi}}{(1-\psi)(2-\psi)}\right] = d_{Ri}(1-\lambda^2) \quad (6)$$

- Equate RE deviance residuals to the power normal first moment
- Simplify Equation 6
- Show term-by-term estimations
- Use IWLS for parameter estimation

Taylor Approximation

- In Equation 6, estimate μ_{Ri} with $\mathcal{E}(\xi_i)$

$$2\left[\frac{\mathcal{E}(\xi^{2-\psi}) - (2-\psi)[\mathcal{E}(\xi_i)]^{2-\psi} + (1-\psi)[\mathcal{E}(\xi_i)]^{2-\psi}}{(1-\psi)(2-\psi)}\right] = d_{Ri}(1-\lambda^2)$$

(7)

- Estimate $\mathcal{E}(\xi_i^{2-\psi})$ with a Taylor approximation about $\hat{\mu}_{Ri}$

$$\mathcal{E}[H(\xi_i)] \doteq \mathcal{E}\left\{H(\hat{\mu}_{Ri}) + H'(\hat{\mu}_{Ri})(\xi_i - \hat{\mu}_{Ri}) + \frac{1}{2}H''(\hat{\mu}_{Ri})(\xi_i - \hat{\mu}_{Ri})^2\right\}$$
$$= \bar{\xi}^{-\psi}Var(\xi_i)$$

(8)

Closed Form Estimate of ψ

With this approximation, we have

$$\bar{\xi}^{-\psi} Var(\xi_i) = (1 - \lambda^2)\bar{d}_R, \tag{9}$$

$\bar{d}_R$ substituted for d_{Ri} as deviance residuals assumed independent. Therefore, the power-law exponent is

$$\hat{\psi} = \frac{\ln[Var(\xi_i)] - \ln(1 - \hat{\lambda}^2) - \ln(\bar{d}_R)}{\ln(\bar{\xi})}, \tag{10}$$

Empirical (Nonlinear Regression) Estimate of ψ

Uses Taylor approximation where the means $\bar{\xi}$ and $\bar{d}_R$ are substituted with ξ_i and d_{Ri} as assumed independent

$$\xi_i^{-\psi} Var(\xi_i) = (1 - \hat{\lambda}^2)d_{Ri}$$
$$d_{Ri} = \frac{Var(\xi_i)}{1 - \hat{\lambda}^2}\xi_i^{-\psi}, \quad -1 < \hat{\lambda} < 1$$
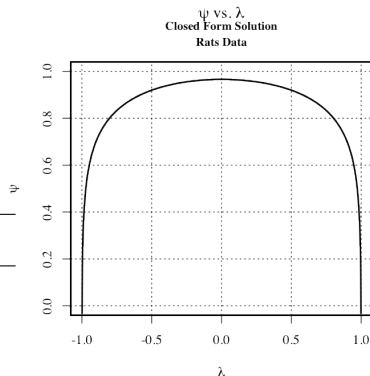$$d_{Ri} = a\xi_i^{-\psi}$$

where $a = Var(\xi_i)/(1 - \hat{\lambda}^2)$, a and ψ estimated

Estimation and Efficacy

Rat data estimates

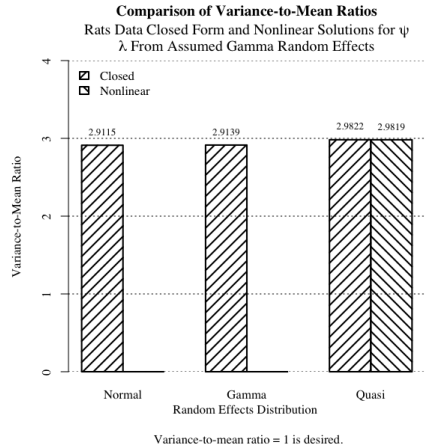| $\hat{\lambda}$	95% CI
0.1010	(-0.0288, 0.2718)

Solution	$\hat{\psi}$	95% CI
Closed form	0.9639	(0.8686, 1.0478)
Empirical	0.9701	(0.9001, 1.0396)
$\Delta \; \hat{\psi}$	0.6432%	

ψ vs. λ
Closed Form Solution
Rats Data

Rat data variance-to-mean ratio

- Closed form ratio 0.01% larger than nonlinear ratio

- Closed form ratio 2.43% larger than normal ratio

- Nonlinear ratio 2.35% larger than normal ratio

- Closed form ratios 2.35% gamma ratio

- Nonlinear ratio 2.3331% gamma ratio

Comparison of Variance-to-Mean Ratios
Rats Data Closed Form and Nonlinear Solutions for ψ
λ From Assumed Gamma Random Effects

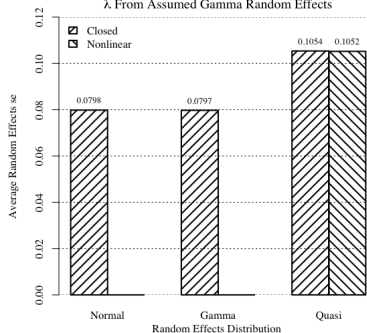Variance-to-mean ratio = 1 is desired.

Rat data standard errors

- Closed form estimate 0.19% greater than nonlinear estimate
- Closed form estimate 32.01% greater than normal estimate
- Nonlinear estimate 31.81% larger than normal estimate
- Closed form estimate 32.13% gamma estimate
- Nonlinear estimate 31.93% gamma estimate
- Normal and gamma standard errors often too small giving Type I errors
- Quasi standard errors more conservative accounting for overdispersion

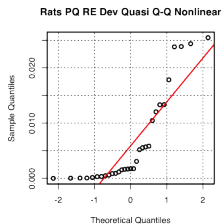| Solution	s.e.(RE)	95% CI
Closed form	0.1054	(0.0773, 0.1334)
Empirical	0.1052	(0.0771, 0.1333)
$\Delta s.e(RE)$	0.1898%	

Comparison of Random Effects Average Standard Errors
Rats Data Closed Form and Nonlinear Solutions for ψ
λ From Assumed Gamma Random Effects

Smallest average standard error is best.

Rat data normal Q-Q diagnostic plots
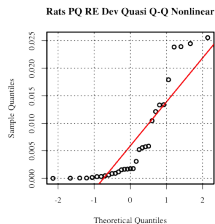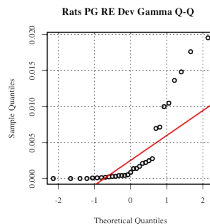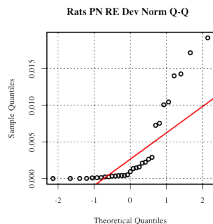
Rat data efficacy measures

- The variance-to-mean ratio

  No distribution is preferred

- Random effects standard errors

  Normal and gamma se often too small for overdispersion correction

- Normality diagnostics

  The quasi distribution for REs suggests a nonnormal residual distribution

Sunspot data estimates

Estimate	95% CI
-0.0202	(-0.0826, 0.0017)

Solution	$\hat{\psi}$	95% CI
Closed form	0.7734	(0.6077, 0.8882)
Empirical	0.9930	(0.9662, 1.0197)
$\Delta\ \hat{\psi}$	16.132%	

ψ vs. λ
Closed Form Solution
SSN Data
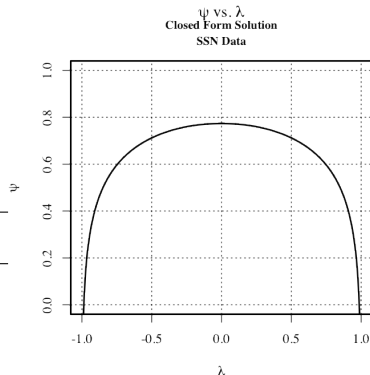
Sunspot data variance-to-mean ratio

- Closed form ratio 0.02% larger than nonlinear ratio

- Closed form ratio 0.28% larger than normal ratio

- Nonlinear ratio 0.26% larger than normal ratio

- Closed form ratios 0.28% gamma ratio

- Nonlinear ratio 0.26% gamma ratio

Comparison of Variance-to-Mean Ratios
SSN Data Closed Form and Nonlinear Solutions for ψ
λ From Assumed Gamma Random Effects

Variance-to-mean ratio = 1 is desired.

Sunspot data standard errors

- Closed form estimate 3.68% greater than nonlinear estimate
- Closed form estimate 48.26% greater than normal estimate
- Nonlinear estimate 42.85% larger than normal estimate
- Closed form estimate 49.77% gamma estimate
- Nonlinear estimate 44.30% gamma estimate
- Normal and gamma standard errors often too small giving Type I errors
- Quasi standard errors more conservative accounting for overdispersion

Solution	s.e.(RE)	95% CI	
Closed form	0.2312	(0.2024,	0.2599)
Empirical	0.2227	(0.1944,	0.2515)
$\Delta s.e(RE)$	3.6765%		

Comparison of Random Effects Average Standard Errors
SSN Data Closed Form and Nonlinear Solutions for ψ
λ From Assumed Gamma Random Effects
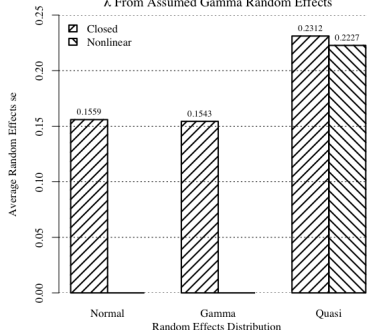
Smallest average standard error is best.

Sunspot data normal Q-Q diagnostic plots
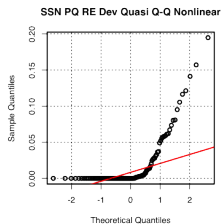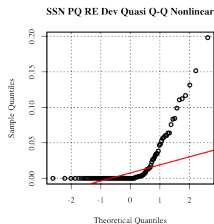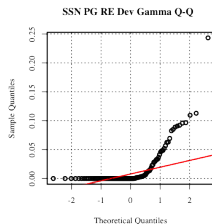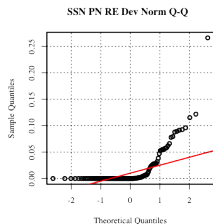
Sunspot data efficacy measures

- The variance-to-mean ratio

  No distribution is preferred

- Random effects standard errors

 Normal and gamma se often too small for overdispersion correction

- Normality diagnostics

  The quasi distribution for REs suggests a nonnormal residual distribution

Conclusions

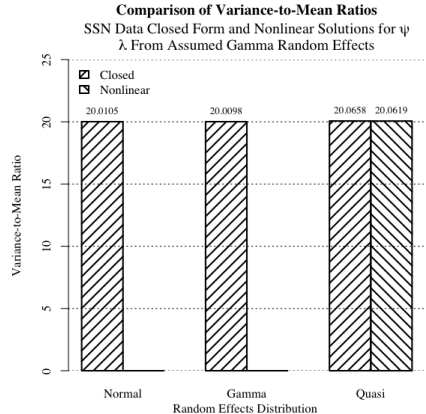## Proposed

- Expected value of the random effects deviance residuals
    - Assumed distributed as a power normal distribution (PN)
    - PN mean estimated from first moment approximation
    - PN truncation parameter λ estimated from the Box-Cox transformation on the random effects deviance residuals
    - Data set specific

- Power-law quasi distribution exponent ψ estimated
    - As a closed form solution
    - Empirical determination

Efficacy

- Overdispersion
  Quasi power-law no worse than normal and gamma

- Normality of standardized deviance residuals
  Quasi power-law no worse, sometimes better than normal and
  gamma

- Random effects standard errors
  Quasi power-law s.e. reflects overdispersion underestimated by
  normal and gamma standard errors

Future Research

- Power normal distribution
    - MLE rather than MME to avoid domain constraints
    - Examine effect on random effects overdispersion
- Equate expected value of random effects deviance residuals to other distributions
    - Account for truncation (truncated quasi distribution)
    - Right-skewness
- Profile estimation of ψ
- Random effects estimation of means model deviance residuals
    - Use the joint glm means model
    - Variance components estimation
    - Bayesian estimation
    - Simulation study

Conclusions

Overdispersion in counts data show more variation than the Poisson model assumes. Random effects standard errors are under-estimated by assuming normal or gamma distributions.

The power normal distribution estimates of the power-law quasi distribution exponent ψ on the selected data sets provide increased random effects standard error corrections commensurate with the overdispersed data.

References

References

- Box, G. & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society B, 26*(2), 211-252.

- Freeman, J. & Modarres, R. (2006a). Inverse box-cox: The power-normal distribution. *Statistics & Probability Letters*, 76:764–772.

- Freeman, J. & Modarres, R. (2006a). Properties of the power-normal distribution. Technical report, Department of Statisticals, George Washington University.

- Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society B*, 58(4):619-678.

- Maruo, K, Shirahata, S., & Goto, M. (2011). Underlying assumptions of the power-normal distribution. *Behaviormetrika, 38*(1), 85-95.

- Myer, P. and Montgomery, D. (2002). *Generalized Linear Models with Applications in Engineering and the Sciences*. John Wiley & Sons, Inc, New York.

Selection Of Random Effects Distributions In Mixed Counts Models with Quasi-Likelihood Estimation

Jamie Riggs

Northwestern University

IAA COIN

CCS-6 Statistical Sciences Seminar Series
Los Alamos National Laboratory

April 7, 2015

Appendix

Hierarchical Generalized Linear Model

Conditional Response Part

$$y \mid v \sim Poi(\boldsymbol{\mu})$$
$$\boldsymbol{\mu} = \text{mean vector of } y \mid v$$
$$g(\boldsymbol{\mu}) = X\boldsymbol{\beta} + Zv$$
$$X_{n \times p} = \text{fixed effects design matrix}$$
$$\boldsymbol{\beta}_{p \times 1} = \text{fixed effects parameter vector}$$

Hierarchical Generalized Linear Model

Random Effects Part

$$Z_{n \times q} = \text{random effects design matrix}$$
$$v = g_R(u), \text{ random effects param fct consistent with } g(\cdot)$$
$$u_{q \times 1} \sim \mathcal{EF}(\boldsymbol{\mu}_R, \boldsymbol{\zeta} V(\boldsymbol{\mu}_R))$$
$$\boldsymbol{\mu}_{R, q \times 1} = \text{mean vector of random effects } u$$
$$\zeta = \text{dispersion parameter}$$

Box-Cox transformation

Suppose X is a random variable $\ni x \in \{X : x \geq 0\}$

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda} & , \lambda \neq 0 \\ \ln(X) & , \lambda = 0 \end{cases} \tag{11}$$

The inverse of the normal random variable Y is

$$X = \begin{cases} (\lambda Y + 1)^{\frac{1}{\lambda}} & , \lambda \neq 0 \\ \exp(Y) & , \lambda = 0 \end{cases} \tag{12}$$

Truncated normal distribution

Y is more accurately represented as a truncated normal (TN) distribution than a normal distribution

$$Y = \begin{cases} \mathcal{TN}(\mu_Y, \sigma_Y^2, -\frac{1}{\lambda}) & , \lambda \neq 0 \\ \mathcal{N}(\mu_Y, \sigma_Y^2) & , \lambda = 0, \end{cases} \tag{13}$$

where $\frac{1}{\lambda}$ is the left or right truncation value.

Truncated normal distribution

The probability distribution function (pdf) of Y is

$$g\left(Y \mid \mu, \sigma^2, -\frac{1}{\lambda}\right) = \frac{1}{K(T)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(Y-\mu)^2\right]. \quad (14)$$

The constant $K(T)$ is

$$K(T) = \begin{cases} \Phi[\text{sgn}(\lambda)T] & , \lambda \neq 0 \\ 1 & , \lambda = 0, \end{cases} \quad (15)$$

where Φ is the cdf of the standard normal distribution.
$T = \frac{1}{\lambda\sigma} + \frac{\mu}{\sigma}$ makes $K(T)$ the normalizing constant. With only positive real number support for a random variable X,

$$Y(\lambda) = \begin{cases} \frac{X^\lambda - 1}{\lambda} \sim \mathcal{TN}(\mu, \sigma^2, -\frac{1}{\lambda}) & , \lambda \neq 0 \\ \ln(X) \sim \mathcal{N}(\mu, \sigma^2) & , \lambda = 0. \end{cases} \quad (16)$$

Power normal distribution

Redefine rv $X \ni X \sim \mathcal{PN}(\lambda, \mu, \sigma)$, a power normal distribution

$$f\left(X \mid \lambda, \mu, \sigma^2\right) = \frac{1}{K(T)} \frac{1}{\sqrt{2\pi\sigma^2}} X^{\lambda-1} \exp\left[-\frac{1}{2\sigma^2}(p_\lambda(X) - \mu)^2\right], X > 0, \tag{17}$$

where $K(T)$ is as above, and

$$p_\lambda(X) = Y. \tag{18}$$

ML estimators are not consistent for constant $K(T)$ (Maruo et. al., 2011)

Power normal parameter estimation

Freeman & Modares (2006a) and Freeman & Modares (2006b) provide moments estimators.

The rth moment of the pre-transform X is

$$\mathcal{E}X^r = \begin{cases} \int_{-\frac{1}{\lambda}}^{\infty} (\lambda y + 1)^{\frac{r}{\lambda}} \phi\left(\frac{y-\mu}{\sigma}\right) \frac{dy}{d\sigma} & , \lambda > 0 \\ \exp(r\mu + r^2\sigma^2/2) & , \lambda = 0. \end{cases} \quad (19)$$

In discrete form:

$$\mathcal{E}X^r = \begin{cases} \sum_{\text{even } k \geq 0} \frac{\sigma^k k!}{s^{\frac{k}{2}} \left(\frac{k}{2}\right)!} (\lambda Y + 1)^{\frac{r}{\lambda} - k} \prod_{l=1}^{k-1}(r - l\lambda) & , \lambda \neq 0 \\ \exp(r\mu + r^2\sigma^2/2) & , \lambda = 0. \end{cases} \quad (20)$$

Power normal parameter estimation

The series approximation of the first moment, μ, is

$$\mu = \mathcal{E}X^1 = \begin{cases} (\lambda Y + 1)^{\frac{1}{\lambda}}(1 - \lambda^2) & , \lambda \neq 0 \\ \exp(\mu + \sigma^2/2) & , \lambda = 0, \end{cases} \tag{21}$$

which is equated to the expected value of the random effects
deviance residuals.

DEQL Algorithm

1. Initialize starting values

2. Construct an augmented response vector $y_a = (y, \boldsymbol{\xi})^T$

3. Use a GLM to estimate initial values of the conditional response model parameters $\boldsymbol{\beta}$ and the random effects parameters v, given the conditional response dispersion parameter ϕ, and the random effects dispersion parameter ζ. Save the deviance components and leverages from this initial fitted model.

DEQL Algorithm

4. Use a $d_i \sim gamma(\cdot, \cdot)$ GLM to estimate β_d from the conditional response deviance residuals and their associated leverages q_i (IWLS leverages from, where these deviance components are from step 3). The gamma GLM is a random intercept model. Update the dispersion parameter by setting ϕ equal to the predicted response vector from this step 4 model

5. Use a similar GLM to step 4 to estimate ζ from the random effects deviance residuals again obtained from the step 3 GLM

6. Iterate steps 3 to 5 until convergence

Linear Models

- Some definitions
    - Fixed effects X has all possible observations levels
    - Random effects Z is a sample of all possible observations levels
- Classical General Linear Models (GLMs)
    - $g(\boldsymbol{\mu}) = \boldsymbol{\mu}$, the link is the identity for linearity in the parameters
    - The responses Y_i must follow a normal (Gaussian) distribution
    - The random effects parameters u_i must follow a normal distribution
- Generalized Linear Models
    - $g(\boldsymbol{\mu})$, the link the identity, log, logit, reciprocal, etc., to make linear in the parameters
    - The responses Y_i need not follow a normal distribution
    - The random effects parameters u_i need not follow a normal distribution

Random Effects (REs) in GLMs

- Current use
    - Allows u_i to follow normal, beta, gamma, inverse-gamma
    - Theory allows RE variance to be a scaled function of the RE mean, $V(\mu_i) = \phi \mu_R$
- Extended use (not used in practice)
    - Allow u_i to follow a power function function of the RE mean, $V(\mu_i) = \mu_R^\psi$
    - If $\psi = 0$, REs follow a normal distribution
    - If $\psi = 1, 2$, REs follow a gamma distribution
    - If $\psi = 1/2$, REs follow a beta distribution
    - If $\psi \neq 0, 1/2, 1, 2$, no named pdf
- Model standardized residuals must follow a normal distribution to have an adequate fit to the data

Selection Of Random Effects Distributions In Mixed Counts Models with Quasi-Likelihood Estimation

Jamie Riggs

Northwestern University

IAA COIN

CCS-6 Statistical Sciences Seminar Series
Los Alamos National Laboratory

April 7, 2015