An Extension of Generalized Linear Models to Finite Mixture Outcomes

Andrew M. Raim

U.S. Census Bureau

Center for Statistical Research and Methodology

andrew.raim@gmail.com

UMBC Statistics Seminar

Feb 19, 2016

Joint work with **Nagaraj K. Neerchal** (UMBC) and **Jorge G. Morel** (UMBC).

Overview

- Overdispersion occurs when a given statistical model can not capture the variability observed in the data. It is commonly encountered in the analysis of categorical and count data.
- The Mixture Link Binomial distribution was proposed in Raim (2014, Ph.D. Thesis) as a model for overdispersed binomial data.
- In this work, we extend the idea to continuous and count outcomes.
- Inference is carried out with Bayesian statistics using Gibbs and Metropolis-Hastings samplers.

Regression in a Heterogeneous Population

- Suppose there are J possible regression functions

$$\mathbf{x}^T \boldsymbol{\beta}^{(1)}, \quad \ldots, \quad \mathbf{x}^T \boldsymbol{\beta}^{(J)}.$$

- Suppose $Y_i \stackrel{\text{ind}}{\sim} \text{Bin}(m_i, G(\mathbf{x}_i^T \boldsymbol{\beta}^{(Z_i)}))$, given a latent subpopulation label

$$Z_i = \begin{cases} 1 & \text{w.p. } \pi_1 \\ \vdots & \\ J & \text{w.p. } \pi_J. \end{cases}$$

  where G is an inverse link function such as the Logistic$(0, 1)$ CDF.

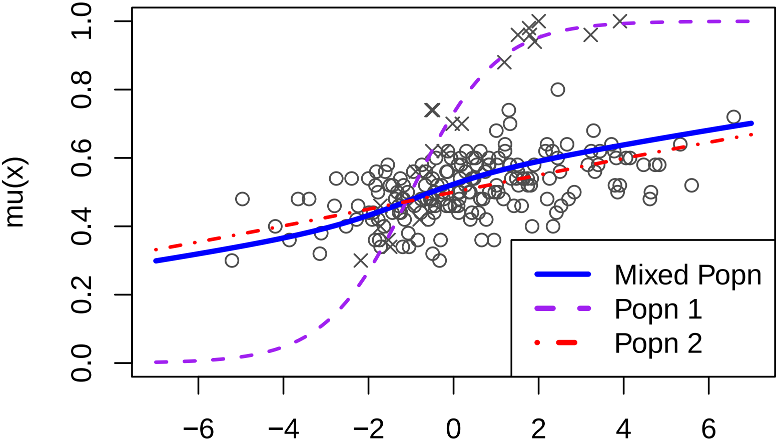
- The overall success probability of a single trial is

$$\mathrm{E}\left(Y/m \mid \mathbf{x}\right) = \sum_{j=1}^{J} \pi_j G(\mathbf{x}^T \boldsymbol{\beta}^{(j)}).$$

Example

$$Y_i \overset{\text{ind}}{\sim} \begin{cases} \text{Bin}[50, \mu_1(x_i)] & \text{w.p. } \pi_1 = 0.1, \\ \text{Bin}[50, \mu_2(x_i)] & \text{w.p. } \pi_2 = 0.9, \end{cases} \quad i = 1, \ldots, 200,$$

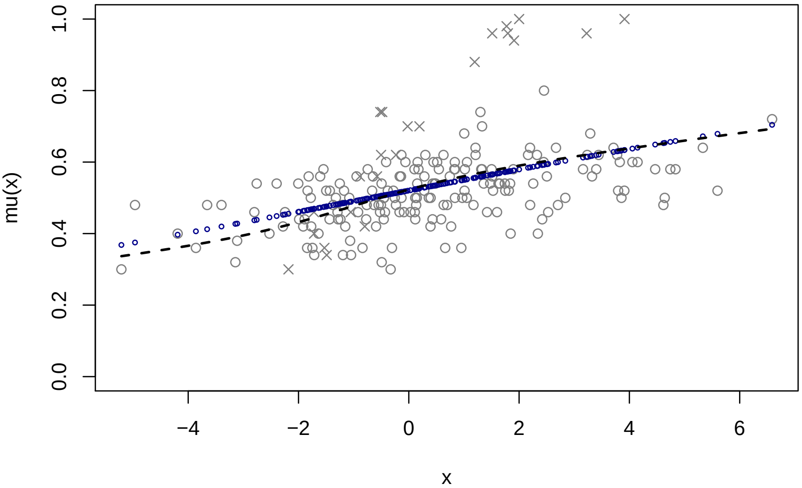
$$\mu_1(x) = G(1 + x), \quad \mu_2(x) = G(0 + 0.1x), \quad \mu(x) = \pi_1\mu_1(x) + \pi_2\mu_2(x)$$



Example

Logistic Regression

	Mean	SD	2.5%	50%	97.5%
β_0	0.0818	0.0205	0.0421	0.0819	0.1198
β_1	0.1194	0.0102	0.0997	0.1193	0.1398



Randomized Quantile Residuals

- Dunn and Smyth (1996) propose randomized quantile residuals for diagnostics on GLMs and other non-normal models.
- Interpretation is similar to OLS residuals on a standard normal scale.
- For y_i independently drawn from a continuous distribution,

$$r_i = \Phi^{-1}\{F(y_i \mid \hat{\boldsymbol{\theta}})\}.$$

- For y_i independently drawn from a discrete distribution,

$$r_i = \Phi^{-1}\{u_i\}, \quad u_i \stackrel{\text{ind}}{\sim} U(a_i, b_i), \quad a_i = \lim_{\varepsilon \downarrow 0} F(y_i - \varepsilon \mid \hat{\boldsymbol{\theta}}), \quad b_i = F(y_i \mid \hat{\boldsymbol{\theta}}).$$

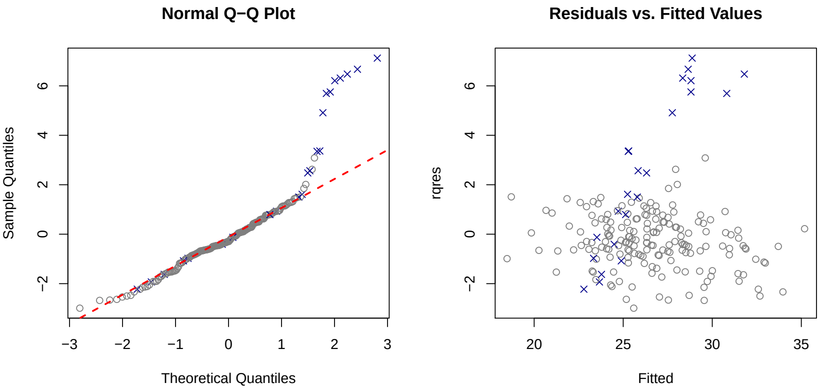
- A Bayesian version using draws $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$ from posterior is

$$r_i = \frac{1}{R} \sum_{r=1}^{R} \Phi^{-1}\{u_i^{(r)}\}, \quad \text{where} \quad u_i^{(r)} \stackrel{\text{ind}}{\sim} U(a_i^{(r)}, b_i^{(r)}),$$

$$a_i^{(r)} = \lim_{\varepsilon \downarrow 0} F(y_i - \varepsilon \mid \boldsymbol{\theta}^{(r)}), \quad \text{and} \quad b_i^{(r)} = F(y_i \mid \boldsymbol{\theta}^{(r)}).$$

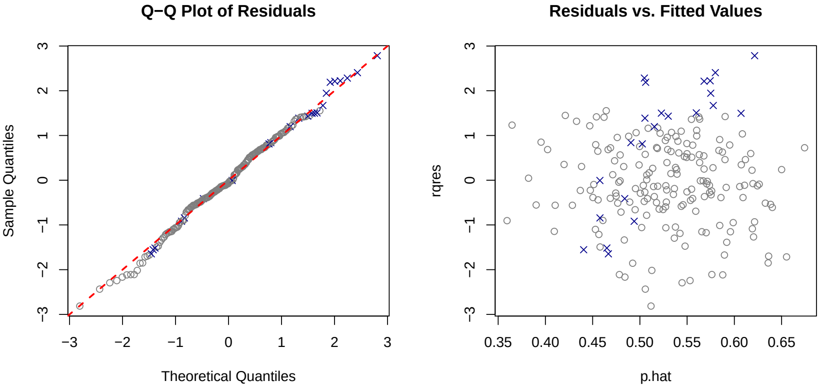
Example

Residuals from Binomial Regression



Example

Residuals from Correct Model (Mixture of Logistic Regressions)



Some Approaches for Overdispersion in GLMs

- Likelihoods which support overdispersion using latent random variables.
  1. Beta-Binomial (Otake and Prentice, 1984) and Random-Clumped Binomial (Morel and Nagaraj, 1993).
  2. Negative-Binomial (Hilbe, 2011)
  3. t-distribution (Liu and Rubin, 1995).
- Quasi-likelihood methods.
  1. Dispersion multiplier (Agresti, 2002, §4.7).
  2. Generalized Estimating Equations (Liang and Zeger, 1986).
- Generalized Linear Mixed Models (McCulloch, Searle, and Neuhaus, 2008).
- Finite mixtures of regressions (Frühwirth-Schnatter, 2006).
- (Bayesian) Generalized link function (Basu and Mukhopadhyay, 2000a,b).
- (Bayesian) Generalized exponential families (Dey and Ravishanker, 2000).

Mixture Link Model

- Consider a random variable Y with density

$$f(y \mid \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j g(y \mid \boldsymbol{\theta}_j).$$

- Mixing proportions $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$ in $\mathcal{S}^J \stackrel{\text{def}}{=} \{\boldsymbol{\pi} \in [0, \infty)^J : \mathbf{1}^T \boldsymbol{\pi} = 1\}$.

- Densities $g(y \mid \boldsymbol{\theta}_j)$ belong to a common family parameterized by $\boldsymbol{\theta}_j = (\mu_j, \phi_j)$.
  1. μ_j is expected value under g.
  2. ϕ_j is remaining parameter of g.

- Overall expected value is $\mathrm{E}(Y) = \sum_{j=1}^{J} \pi_j \mu_j = \boldsymbol{\pi}^T \boldsymbol{\mu}$.

- Application may naturally restrict μ_j to a subset of $\mathbb{R}$.
  1. If y is an outcome of OLS then $\mu_j \in \mathbb{R}$.
  2. If y is a count then $\mu_j \in [0, \infty)$.
  3. if y is Bernoulli or Binomial then $\mu_j \in [0, 1]$.

- Denote space of μ_j as $\mathcal{M}$, so that $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J) \in \mathcal{M}^J$.

Mixture Link Model

- Suppose we observe a random sample $Y_1, \dots, Y_n$ with

$$Y_i \sim f(y_i \mid \boldsymbol{\theta}_i) = \sum_{j=1}^{J} \pi_j g(y_i \mid \mu_{ij}, \phi_{ij}).$$

- Here, $\mathrm{E}(y_i) = \boldsymbol{\pi}^T \boldsymbol{\mu}_i$ where $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iJ}) \in \mathcal{M}^J$.
- Suppose also that each Y_i has a (fixed) predictor $\mathbf{x}_i \in \mathbb{R}^d$.
- Let $\vartheta_i \stackrel{\text{def}}{=} G(\mathbf{x}_i^T \boldsymbol{\beta})$ denote an inverse-linked regression function of interest.
- As in traditional GLM, we wish to link $\mathrm{E}(y_i)$ to ϑ_i.
- To do this, we will consider the set

$$A(\vartheta_i, \boldsymbol{\pi}) = \{\boldsymbol{\mu} \in \mathcal{M}^J : \boldsymbol{\mu}^T \boldsymbol{\pi} = \vartheta_i\}.$$

- If we restrict ourselves to $\boldsymbol{\mu}_i \in A(\vartheta_i, \boldsymbol{\pi})$, then we enforce the link.
- Approach will be to take $\boldsymbol{\mu}_i$ as a random effect in $A(\vartheta_i, \boldsymbol{\pi})$.

Mixture Link Model

- Once a distribution over $A(\vartheta, \boldsymbol{\pi})$ has been determined, we obtain the density

$$
\begin{aligned}
f(y_i \mid \boldsymbol{\beta}, \boldsymbol{\pi}, \phi_i) &= \int \sum_{j=1}^{J} \pi_j g(y_i \mid \mu_{ij}, \phi_{ij}) \cdot f_{A^{(i)}}(\boldsymbol{\mu}_i) d\boldsymbol{\mu}_i \\
&= \sum_{j=1}^{J} \pi_j \int g(y_i \mid w, \phi_{ij}) \cdot f_{A_j^{(i)}}(w) dw.
\end{aligned}
$$

- Here, $f_{A^{(i)}}$ represents the J-dimensional density over $A(\vartheta_i, \boldsymbol{\pi})$ and $f_{A_j^{(i)}}$ represents the marginal density of the jth coordinate.

- Evaluating density requires computating J univariate integrals.

- By construction, $\mathrm{E}(Y_i) = \vartheta_i$, but variance and other properties depend on g and distribution of $\boldsymbol{\mu}_i$.

- Density is invariant to permutations of labels $\{1, \ldots, J\}$.

Probability-Valued Means

- Consider $\mathcal{M} = [0, 1]$, as in the case of Binomial regression.
- Notice $A(\vartheta, \boldsymbol{\pi}) = \{\boldsymbol{\mu} \in [0, 1]^J : \boldsymbol{\mu}^T \boldsymbol{\pi} = \vartheta\}$ is bounded and convex.
- We can obtain the decomposition

$$A(\vartheta_i, \boldsymbol{\pi}) = \left\{ \sum_{\ell=1}^{k_i} \lambda_\ell \mathbf{v}_\ell^{(i)} : \boldsymbol{\lambda} \in \mathcal{S}^{k_i} \right\} = \left\{ \mathbf{V}^{(i)} \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathcal{S}^{k_i} \right\}.$$

- $\mathbf{V}^{(i)} = (\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{k_i}^{(i)})$ is a $J \times k_i$ matrix which forms the convex hull.
- $\boldsymbol{\lambda}^{(i)}$ belongs to the probability simplex $\mathcal{S}^J$.
- A related approach was taken by Danaher et al. (2012). They used priors based on the Minkowski-Weyl decomposition to enforce (biologically motivated) polyhedral constraints for parameters.

Probability-Valued Means

Random effects distribution

- A natural choice for a random effects distribution on $\mathcal{S}^J$ is $\boldsymbol{\lambda}^{(i)} \stackrel{\text{ind}}{\sim} \text{Dirichlet}_{k_i}(\boldsymbol{\alpha})$.
- However, this leads to each component of $\boldsymbol{\mu}_i = \mathbf{V}^{(i)}\boldsymbol{\lambda}^{(i)}$ following a linear-combination-of-Dirichlet distribution; its density has no known closed form for general k_i (Provost and Cheong, 2000).
- Instead, we consider a more practical form using Beta random effects with first and second moments matched to Dirichlet random effects.
- This ensures, e.g., that $\text{E}(\boldsymbol{\mu}_i) \in A(\vartheta_i, \boldsymbol{\pi})$ so that

$$\text{E}(Y_i) \equiv \boldsymbol{\pi}^T \, \text{E}(\boldsymbol{\mu}_i) \quad \text{reduces to} \quad \vartheta_i.$$

- The linear-combination-of-Dirichlet density can differ substantially from the moment-matched Beta density, but the Mixture Link density with Dirichlet vs. moment-matched Beta are very close (Raim, 2014).

Probability-Valued Means

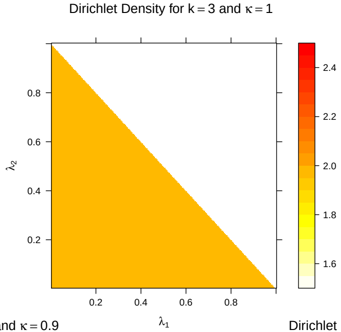
Random effects distribution

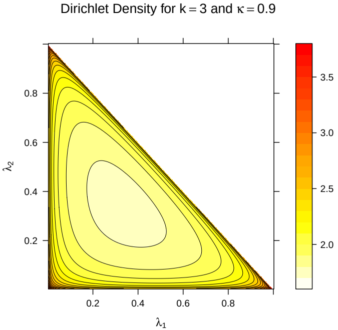
- Model with Dirichlet random effects is

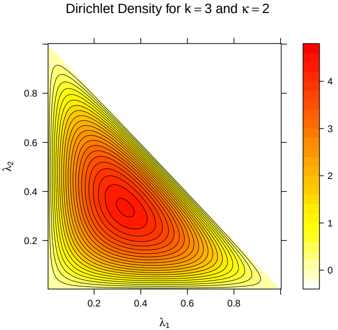
$$
\begin{aligned}
Y_i &\stackrel{\text{ind}}{\sim} \sum_{j=1}^{J} \pi_j g(y_i \mid \mu_{ij}, \phi_{ij}) \\
\boldsymbol{\mu}_i &= \mathbf{V}^{(i)} \boldsymbol{\lambda}^{(i)}, \quad \text{where } \mathbf{V}^{(i)} \text{ contains vertices of } A(\vartheta_i, \boldsymbol{\pi}), \\
\boldsymbol{\lambda}^{(i)} &\stackrel{\text{ind}}{\sim} \text{Dirichlet}_{k_i}(\alpha_1^{(i)}, \ldots, \alpha_{k_i}^{(i)}).
\end{aligned}
$$

- We restrict $(\alpha_1^{(i)}, \ldots, \alpha_{k_i}^{(i)}) = \kappa \mathbf{1}$ ("Symmetric Dirichlet") for several reasons.
  1. The dimension k_i can vary with i so that an arbitrary $\boldsymbol{\alpha}$ will not be compatible with all observations.
  2. The ordering of the vertices in $\mathbf{V}^{(i)}$ is arbitrary, and no clear correspondence between those vertices and the elements of $\boldsymbol{\alpha}$.

Symmetric Dirichlet Density







Probability-Valued Means

Random effects distribution

- Model with Beta random effects is

$$
\begin{aligned}
Y_i &\stackrel{\text{ind}}{\sim} \sum_{j=1}^{J} \pi_j g(y_i \mid \mu_{ij}, \phi_{ij}) \\
\mu_{ij} &= (u_{ij} - \ell_{ij})\psi_{ij} + \ell_{ij}, \quad j = 1, \ldots, J \\
\psi_{ij} &\sim \text{Beta}(a_{ij}, b_{ij}),
\end{aligned}
$$

- ℓ_{ij} and u_{ij} are min and max elements of $\mathbf{v}_{j\cdot}^{(i)}$ (the jth row of $\mathbf{V}^{(i)}$).

- Dirichlet random effects would have moments

$$
\text{E}(\mathbf{v}_{j\cdot}^{(i)T}\boldsymbol{\lambda}) = \bar{v}_{j\cdot}^{(i)} \quad \text{and} \quad \text{Var}(\mathbf{v}_{j\cdot}^{(i)T}\boldsymbol{\lambda}) = \frac{k_i \mathbf{v}_{j\cdot}^{(i)T}\mathbf{v}_{j\cdot}^{(i)} - (k_i \bar{v}_{j\cdot}^{(i)})^2}{k_i^2(1 + k_i\kappa)},
$$

- To obtain a_{ij} and b_{ij}, equate

$$
\text{E}(\mu_{ij}) = (u_{ij} - \ell_{ij})\frac{a_{ij}}{a_{ij} + b_{ij}} + \ell_{ij} \text{ and } \text{Var}(\mu_{ij}) = \frac{(u_{ij} - \ell_{ij})^2 a_{ij} b_{ij}}{(a_{ij} + b_{ij})^2(a_{ij} + b_{ij} + 1)}
$$

  to the Dirichlet moments and solve for a_{ij} and b_{ij}.

Probability-Valued Means

Random effects distribution

Model is therefore

$$
\begin{aligned}
Y_i &\overset{\text{ind}}{\sim} \sum_{j=1}^{J} \pi_j g(y_i \mid \mu_{ij}, \phi_{ij}) \\
\mu_{ij} &= (u_{ij} - \ell_{ij})\psi_{ij} + \ell_{ij}, \quad j = 1, \ldots, J \\
\psi_{ij} &\sim \text{Beta}(a_{ij}, b_{ij}),
\end{aligned}
$$

where

$$
\begin{aligned}
&\ell_{ij} \text{ and } u_{ij} \text{ are min and max elements of the } j\text{th row of } \mathbf{V}^{(i)}, \\
&a_{ij} = (\bar{v}_{j.}^{(i)} - \ell_{ij})^2 \left(\frac{k_i \mathbf{v}_{j.}^{(i)T} \mathbf{v}_{j.}^{(i)} - (k_i \bar{v}_{j.}^{(i)})^2}{k_i^2(1 + k_i \kappa)} \right)^{-1} \frac{u_{ij} - \bar{v}_{j.}^{(i)}}{u_{ij} - \ell_{ij}} - \frac{\bar{v}_{j.}^{(i)} - \ell_{ij}}{u_{ij} - \ell_{ij}}, \\
&b_{ij} = a_{ij} \left(\frac{u_{ij} - \bar{v}_{j.}^{(i)}}{\bar{v}_{j.}^{(i)} - \ell_{ij}} \right).
\end{aligned}
$$

Probability-Valued Means

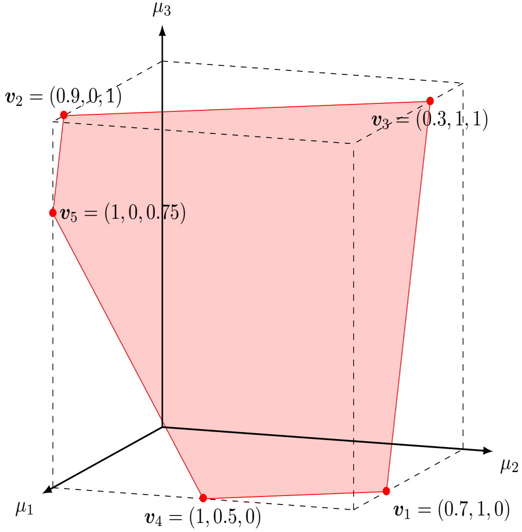


Figure : $A(\vartheta, \boldsymbol{\pi})$ with $\boldsymbol{\pi} = (0.5, 0.3, 0.2)$ and $\vartheta = 0.65$.

Probability-Valued Means

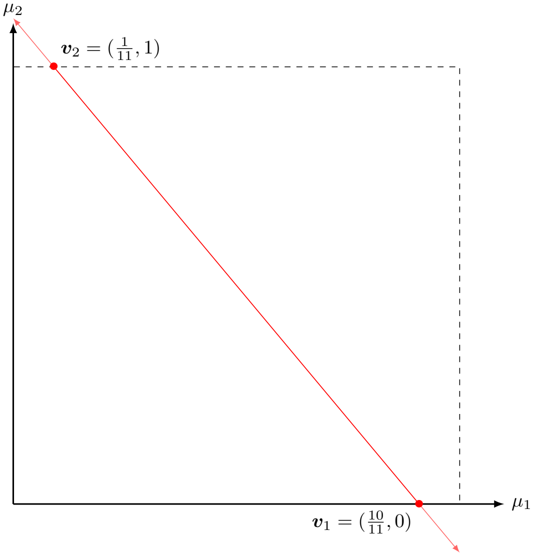


Figure : A with $\boldsymbol{\pi} = (\frac{11}{20}, \frac{9}{20})$ and $\vartheta = \frac{1}{2}$.

Probability-Valued Means

Computation of Vertices

Lemma

Suppose $\mathbf{v} = (v_1, \ldots, v_J)$ *is a point in* A *with two or more coordinates* $v_j \notin \{0, 1\}$. *Then* $\mathbf{v}$ *is not an extreme point of* A.

Algorithm

function $\textsc{FindVertices}(\vartheta, \boldsymbol{\pi})$
  $\mathcal{V} \leftarrow \varnothing$
  for $j = 1, \ldots, J$ **do**
    if $\pi_j > 0$ **then**
      for all $\boldsymbol{\mu}_{-j} \in \{0, 1\}^{J-1}$ **do**
        $\mu_j^* \leftarrow \pi_j^{-1} \left[\vartheta - \boldsymbol{\mu}_{-j}^T \boldsymbol{\pi}_{-j}\right]$
        $\mathbf{v}^* \leftarrow (\mu_1, \ldots, \mu_{j-1}, \mu_j^*, \mu_{j+1}, \ldots, \mu_J)$
        $\mathcal{V} \leftarrow \mathcal{V} \cup \mathbf{v}^*$ if $\mathbf{v}^* \in A(\vartheta, \boldsymbol{\pi})$
  return Matrix $\mathbf{V}$ with columns $\mathbf{v}^* \in \mathcal{V}$

Number of steps is $J \cdot 2^{J-1}$.

Probability-Valued Means

Mixture Link Binomial

- Suppose $g(y_i \mid w, \phi_{ij}) = \text{Bin}(y_i \mid m_i, w)$ so that

$$
\begin{aligned}
&Y_i \overset{\text{ind}}{\sim} \sum_{j=1}^{J} \pi_j \binom{m_i}{y_i} \mu_{ij}^{y_i} (1 - \mu_{ij})^{m_i - y_i} \\
&\mu_{ij} = (u_{ij} - \ell_{ij}) \psi_{ij} + \ell_{ij}, \quad j = 1, \ldots, J \\
&\psi_{ij} \sim \text{Beta}(a_{ij}, b_{ij}).
\end{aligned}
$$

- Expectation is $\text{E}(Y) = m\vartheta$ and variance is

$$
\text{Var}(Y) = m\vartheta\,(1 - m\vartheta) + m(m-1) \sum_{j=1}^{J} \pi_j \frac{\mathbf{v}_{j\cdot}^T \mathbf{v}_{j\cdot} + \kappa (k \bar{v}_{j\cdot})^2}{k(1 + \kappa k)}.
$$

- **Remark:** For the case $m_i = 1$, Mixture Link Binomial simplifies to usual logistic regression model with $Y_i \overset{\text{ind}}{\sim} \text{Ber}(m_i, \vartheta_i)$.

Probability-Valued Means

Bayesian Mixture Link Binomial

- For a Bayesian specification, we may assume priors

$$\begin{aligned}
\boldsymbol{\beta} &\sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\beta}}), \\
\boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\gamma}), \\
\kappa &\sim \text{Gamma}(a_\kappa, b_\kappa).
\end{aligned}$$

- A reasonably fast MCMC algorithm can be obtained.
  1. Take ψ_{ij} as augmented data.
  2. Use a Metropolis-within-Gibbs sampler.
  3. Use simple Random Walk Metropolis Hastings to propose draws for parameters and augmented data.

- All steps rely on repeated computation of

$$Q_i = \sum_{j=1}^{J} \pi_j \text{Bin}\Big(y_i \mid m_i, (u_{ij} - \ell_{ij})\psi_{ij} + \ell_{ij}\Big) \mathcal{B}(\psi_{ij} \mid a_{ij}, b_{ij});$$

`R` implementation of MCMC benefits from writing this part in `C/C++`.

Positive Means

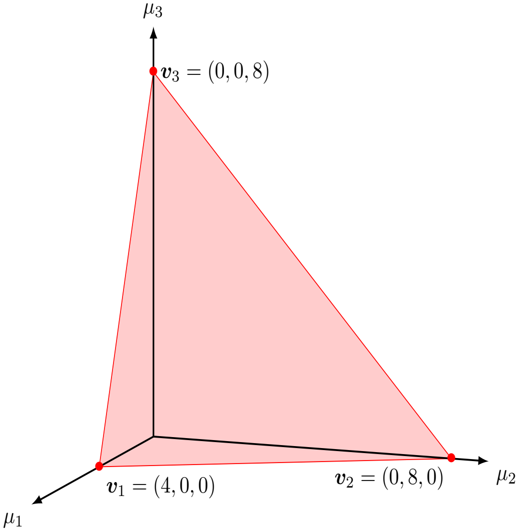


Figure : $A(\vartheta, \boldsymbol{\pi})$ with $\boldsymbol{\pi} = (0.5, 0.25, 0.25)$ and $\vartheta = 2$.

Positive Means

- Very similar to case of probability-valued means, except vertex computation differs (is much simpler).

- $A(\vartheta, \boldsymbol{\pi}) = \{\boldsymbol{\mu} \in [0, \infty)^J : \boldsymbol{\mu}^T \boldsymbol{\pi} = \vartheta\}$ is still bounded and convex.

Lemma

Suppose $\mathbf{v} = (v_1, \ldots, v_J)$ *is a point in* A *with two or more coordinates* $v_j > 0$. *Then* $\mathbf{v}$ *is not an extreme point of* A.

- Therefore, we explicitly have that $\mathbf{V}^{(i)} = \text{Diag}\left(\vartheta_i/\pi_1, \ldots, \vartheta_i/\pi_J\right)$.

Positive Means

- Now with $g(y_i \mid w, \phi_{ij}) = \mathcal{P}(y_i \mid w)$, Poisson Mixture Link can be formulated exactly as Binomial Mixture Link.

$$
\begin{aligned}
Y_i &\overset{\text{ind}}{\sim} \sum_{j=1}^{J} \pi_j \frac{e^{-\mu_{ij}} \mu_{ij}^{y_i}}{y_i!} \\
\mu_{ij} &= (u_{ij} - \ell_{ij})\psi_{ij} + \ell_{ij}, \quad j = 1, \ldots, J \\
\psi_{ij} &\sim \text{Beta}(a_{ij}, b_{ij}).
\end{aligned}
$$

- Expected value of Y is $\text{E}(Y) = \vartheta$ with variance

$$
\text{Var}(Y) = \vartheta + \left[\sum_{j=1}^{J} \pi_j \bar{v}_{j.}^2 - \vartheta^2 \right] + \sum_{j=1}^{J} \pi_j \frac{k \mathbf{v}_{j.}^T \mathbf{v}_{j.} - (k \bar{v}_{j.})^2}{k^2(1 + \kappa k)}
$$

- MCMC can also be done the same way as Binomial setting.

Real-Valued Means

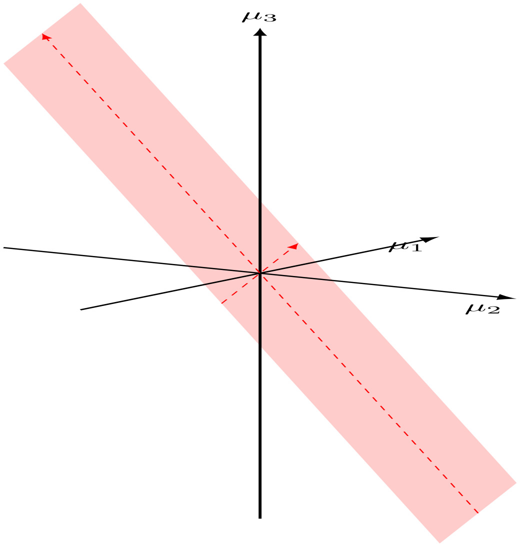


Figure : $A(\vartheta, \boldsymbol{\pi})$ with $\boldsymbol{\pi} = (0.5, 0.3, 0.2)$ and $\vartheta = 0$.

Real-Valued Means

- We can decompose

$$\begin{aligned} A(\vartheta, \boldsymbol{\pi}) &= \{\boldsymbol{\mu} \in \mathbb{R}^J : \boldsymbol{\mu}^T \boldsymbol{\pi} = \vartheta\} \\ &= \{\tilde{\boldsymbol{\mu}} \in \mathbb{R}^J : \tilde{\boldsymbol{\mu}}^T \boldsymbol{\pi} = 0\} + \vartheta \mathbf{1}. \end{aligned}$$

- For any $\tilde{\boldsymbol{\mu}}$ in $\{\tilde{\boldsymbol{\mu}} \in \mathbb{R}^J : \tilde{\boldsymbol{\mu}}^T \boldsymbol{\pi} = 0\}$ we can write

$$\tilde{\mu}_J = -\pi_J^{-1}(\pi_1 \tilde{\mu}_1 + \cdots + \pi_{J-1} \tilde{\mu}_{J-1})$$

  with $\tilde{\mu}_1, \ldots, \tilde{\mu}_{J-1}$ unrestricted.

- A basis for subspace $\{\tilde{\boldsymbol{\mu}} \in \mathbb{R}^J : \tilde{\boldsymbol{\mu}}^T \boldsymbol{\pi} = 0\}$ is the $J \times (J-1)$ matrix

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & 1 \\ -\pi_1/\pi_J & -\pi_2/\pi_J & \cdots & -\pi_{J-1}/\pi_J \end{pmatrix}.$$

- We can therefore represent any $\boldsymbol{\mu} \in A(\vartheta, \boldsymbol{\pi})$ as

$$\boldsymbol{\mu} = \mathbf{V}\boldsymbol{\lambda} + \vartheta \mathbf{1} \quad \text{for some } \boldsymbol{\lambda} \in \mathbb{R}^{J-1}.$$

Real-Valued Means

Random effects distribution

- We can take $\lambda_1, \ldots, \lambda_{J-1} \stackrel{\text{iid}}{\sim} \mathsf{N}(0, \kappa^2)$ so that

$$\begin{aligned}
\boldsymbol{\mu} = \mathbf{V}\boldsymbol{\lambda} + \vartheta\mathbf{1} \sim \mathsf{N}(\vartheta\mathbf{1}, \kappa^2\mathbf{V}\mathbf{V}^T), \quad &\text{where} \\
&\mathbf{V}\mathbf{V}^T = \begin{pmatrix} \mathbf{I} & -\pi_J^{-1}\boldsymbol{\pi}_{-J} \\ -\pi_J^{-1}\boldsymbol{\pi}_{-J}^T & \pi_J^{-2}\boldsymbol{\pi}_{-J}^T\boldsymbol{\pi}_{-J} \end{pmatrix}, \\
&\mathbf{I} \text{ is the } (J-1) \times (J-1) \text{ identity matrix}, \\
&\boldsymbol{\pi}_{-J} = (\pi_1, \ldots, \pi_{J-1}).
\end{aligned}$$

- Now the Mixture Link density becomes

$$f(y_i \mid \boldsymbol{\beta}, \boldsymbol{\pi}, \phi_i, \kappa) = \sum_{j=1}^{J} \pi_j \int g(y_i \mid w, \phi_{ij}) \cdot \mathsf{N}(w \mid \vartheta_i, \kappa^2 a_j) dw, \quad \text{where}$$

$$a_j = \begin{cases} 1 & \text{for } j = 1, \ldots, J-1 \\ \pi_J^{-2}\boldsymbol{\pi}_{-J}^T\boldsymbol{\pi}_{-J} & \text{for } j = J. \end{cases}$$

Real-Valued Means

Normal Mixture Link

- Suppose $g(y_i \mid w, \phi_{ij}) = \mathrm{N}(y_i \mid w, \sigma_j^2)$. Here, Mixture Link density explicitly integrates to

$$f(y_i \mid \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma_1^2, \ldots, \sigma_J^2, \kappa) = \sum_{j=1}^{J} \pi_j \mathrm{N}(y_i \mid \vartheta_i, \kappa^2 a_j + \sigma_j^2).$$

- If the J subpopulations have a common variance, this simplifies to

$$f(y_i \mid \boldsymbol{\beta}, \boldsymbol{\pi}, \sigma^2, \kappa) =$$
$$(1 - \pi_J)\mathrm{N}(y_i \mid \vartheta_i, \kappa^2 + \sigma^2) + \pi_J \mathrm{N}(y_i \mid \vartheta_i, \kappa^2 \pi_J^{-2} \boldsymbol{\pi}_{-J}^T \boldsymbol{\pi}_{-J} + \sigma^2).$$

- If $J = 2$, then $\pi_J^{-2} \boldsymbol{\pi}_{-J}^T \boldsymbol{\pi}_{-J} = [(1 - \pi_J)/\pi_J]^2$.
  1. Recall $\mathbf{V}$ was constructed selecting Jth component as constrained.
  2. To avoid identifiability/label switching problems, enforce $\pi_J < 1/2$.
  3. Then small π_J yields a rare "contamination group" with large variance.

- The overall mean is $\mathrm{E}(Y_i) = \vartheta_i$, and

$$\mathrm{Var}(Y_i) = \kappa^2 \frac{1 - \pi_J}{\pi_J} + \sigma^2$$

Real-Valued Means

Bayesian Normal Mixture Link

- May need additional constraints on variance parameters for usable statistical model (work in progress).

- MCMC is simpler than previous cases — do not need augmented data to avoid integration.

Back to Example

Table : Binomial

	Mean	SD	2.5%	50%	97.5%
β_0	0.0818	0.0205	0.0421	0.0819	0.1198
β_1	0.1194	0.0102	0.0997	0.1193	0.1398

Table : Mixture Link Binomial $J = 2$ with basic Random Walk Metropolis Hastings

	Mean	SD	2.5%	50%	97.5%
β_0	0.0124	0.0218	-0.0318	0.0125	0.0544
β_1	0.0815	0.0103	0.0610	0.0815	0.1014
π_1	0.0756	0.0169	0.0463	0.0747	0.1085
κ	0.5699	0.2096	0.2229	0.5479	1.0351

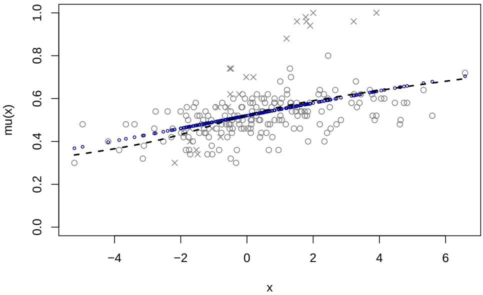


Figure : Binomial

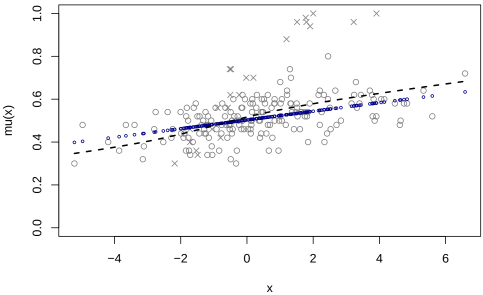
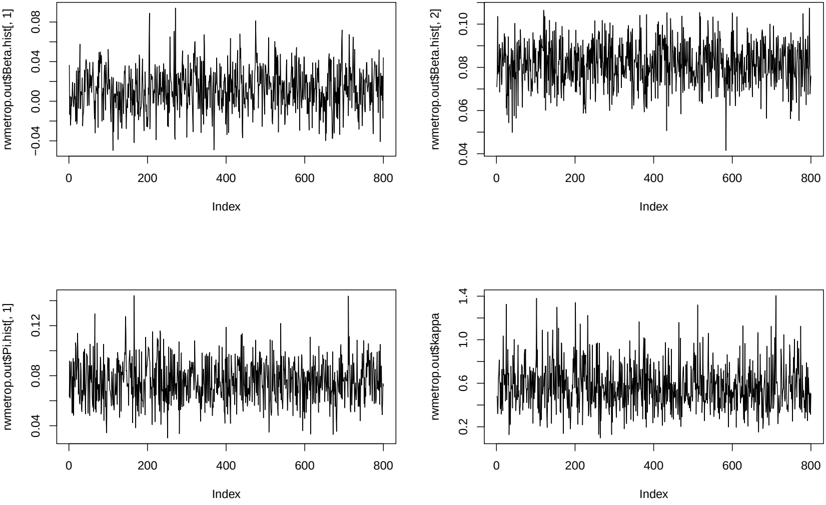
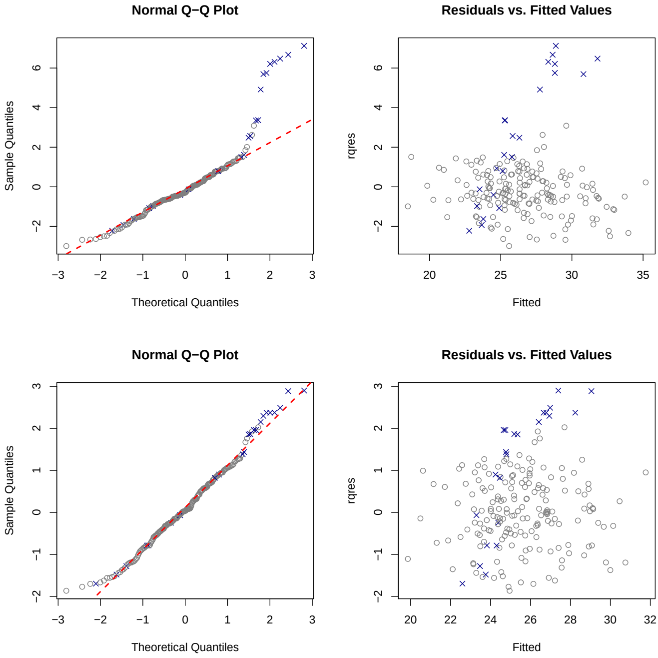


Figure : Mixture Link Binomial

Back to Example

Trace Plots




Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
Residuals vs. Fitted Values
rqres
Fitted
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
Residuals vs. Fitted Values
rqres
Fitted

Back to Example

Posterior Predictive Distribution

- The posterior predictive distribution for a new sample $\tilde{\mathbf{y}}$ given the observed sample $\mathbf{y}$ is

$$\begin{aligned} f(\tilde{\mathbf{y}} \mid \mathbf{y}) &= \int f(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}, \mathbf{y}) dF(\boldsymbol{\theta} \mid \mathbf{y}) \\ &= \int f(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}) dF(\boldsymbol{\theta} \mid \mathbf{y}). \end{aligned}$$

- Then to sample from $f(\tilde{\mathbf{y}} \mid \mathbf{y})$:
  1. Sample $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(R)}$ from posterior $f(\boldsymbol{\theta} \mid \mathbf{y})$.
  2. Sample $\tilde{\mathbf{y}}^{(r)}$ from $f(\tilde{\mathbf{y}} \mid \boldsymbol{\theta}^{(r)})$ for $r = 1, \ldots, R$.

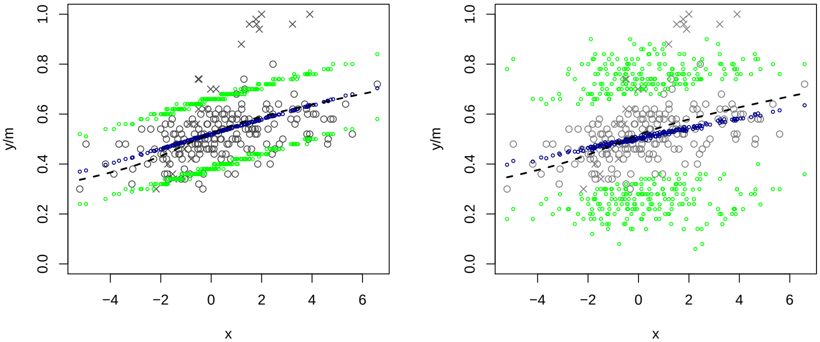
  Now $(\tilde{\mathbf{y}}^{(1)}, \ldots, \tilde{\mathbf{y}}^{(R)})$ is our predictive sample.

- A prediction for ith observation is $\frac{1}{R} \sum_{r=1}^{R} \tilde{y}_i^{(r)}$.

- A 95% prediction interval for ith observation is given by 2.5% and 97.5% quantiles of $(\tilde{y}_i^{(1)}, \ldots, \tilde{y}_i^{(R)})$.

Back to Example

95% Posterior Prediction Intervals



Conclusions and Future Work

Conclusions

- Proposed an extension of GLM using finite mixture distribution for the response.
- Fully likelihood-based.
- Involves a special random effects structure to link regression to mixture mean.
- Formulated model for real-valued means, positive means, and probability-valued means.
- Examples show benefits of extra variation through quantile residuals and posterior predictive intervals.

Future Work

- Study statistical properties.
- Apply to other datasets.
- Compare to other overdispersion models.

References I

Alan Agresti. *Categorical Data Analysis*. Wiley-Interscience, 2nd edition, 2002.

James H. Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

Sanjib Basu and Saurabh Mukhopadhyay. Bayesian analysis of binary regression using symmetric and asymmetric links. *Sankhya: The Indian Journal of Statistics, Series B*, 62(3):372–387, 2000a.

Sanjib Basu and Saurabh Mukhopadhyay. Binary response regression with normal scale mixture links. In Bani K. Mallick Dipak K. Dey, Sujit K. Ghosh, editor, *Generalized Linear Models: A Bayesian Perspective*, pages 231–242. CRC Press, 2000b.

Michelle R. Danaher, Anindya Roy, Zhen Chen, Sunni L. Mumford, and Enrique F. Schisterman. Minkowski-Weyl priors for models with parameter constraints: An analysis of the biocycle study. *Journal of the American Statistical Association*, 107(500):1395–1409, 2012.

Dipak K. Dey and Nalini Ravishanker. Bayesian approaches for overdispersion in generalized linear models. In Bani K. Mallick Dipak K. Dey, Sujit K. Ghosh, editor, *Generalized Linear Models: A Bayesian Perspective*, pages 73–88. CRC Press, 2000.

References II

Peter K. Dunn and Gordon K. Smyth. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244, 1996.

Sylvia Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, 2006.

Joseph M. Hilbe. *Negative Binomial Regression*. Cambridge University Press, 2nd edition, 2011.

Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Chuanhai Liu and Donald B. Rubin. ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica*, 5:19–39, 1995.

Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus. *Generalized, Linear, and Mixed Models*, volume 2. Wiley-Interscience, 2nd edition, 2008.

Jorge G. Morel and Neerchal K. Nagaraj. A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, 80(2):363–371, 1993.

Masanori Otake and Ross L. Prentice. The analysis of chromosomally aberrant cells based on beta-binomial distribution. *Radiation Research*, 98(3):456–470, 1984.

Serge B. Provost and Young-Ho Cheong. On the distribution of linear combinations of the components of a dirichlet random vector. *Canadian Journal of Statistics*, 28(2):417–425, 2000.

References III

Andrew M. Raim. Computational methods in finite mixtures using approximate information and regression linked to the mixture mean. Ph.D. Thesis, Department of Mathematics and Statistics, University of Maryland, Baltimore County, 2014.

Andrew M. Raim, Marissa N. Gargano, Nagaraj K. Neerchal, and Jorge G. Morel. Bayesian analysis of overdispersed binomial data using mixture link regression. In *JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association*, pages 2794–2808, 2015.

Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, 2nd edition, 2010.