Modeling Overdispersion Using Finite Mixtures with a Regression Linked to the Mean

Andrew M. Raim* and Nagaraj K. Neerchal
Department of Mathematics and Statistics,
University of Maryland, Baltimore County
*www.umbc.edu/~araim1

UMBC
AN HONORS UNIVERSITY IN MARYLAND

Summary

- **Objective:** To explore the idea of linking a regression to the mean of a finite mixture.
- Current focus is on logistic regression. Here we discuss some early ideas for this model.
- Computations become more difficult, but model would allow extra variation beyond the standard logistic regression model.
- We consider a hierarchical model, where the probabilties of the binomial mixture are random effects drawn from a set which represents the link from the mixture mean to a regression.

Motivation

- Logistic regression is commonly used to model a discrete outcome (T successes out of m trials) where the probability of success depends on covariates $\boldsymbol{x}_i$

$$T_i \overset{\text{ind}}{\sim} \text{Bin}(m_i, p_i), \quad g(p_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}, \quad i = 1, \ldots, n$$

- Overdispersion frequently occurs when observed data show larger variability than the model can handle. Note that binomial variance is tied to the mean since

$$\text{E}(T_i) = m_i p_i, \quad \text{Var}(T_i) = m_i p_i (1 - p_i)$$

- Finite mixtures have been used to address overdispersion. One approach discussed in (Frühwirth-Schnatter, 2006) is the finite mixture of regressions, e.g.

$$T_i \overset{\text{ind}}{\sim} \text{BinMix}(m_i, \boldsymbol{p}_i, \boldsymbol{\pi}), \quad g(p_{ij}) = \boldsymbol{x}_i^T \boldsymbol{\beta}_j, \quad j = 1, \ldots, J$$

  Models a regression conditionally on each unobserved subpopulation

- In this work, we use a finite mixture to handle extra variation, but assume a single regression which is modeled marginally over **subpopulations**. This is similar to Generalized Linear Mixed Models vs. Generalized Estimating Equations, where the former models conditionally and the latter models marginally over **individuals**

Preliminaries

- The finite mixture of binomials, $T \sim \text{BinMix}(m, \boldsymbol{p}, \boldsymbol{\pi})$

$$f(t \mid m, \boldsymbol{p}) = \sum_{j=1}^{J} \pi_j \binom{m}{t} p_j^t (1 - p_j)^{m-t}, \quad \text{E}(T) = m \boldsymbol{p}^T \boldsymbol{\pi}$$

$$\boldsymbol{p} = (p_1, \ldots, p_J) \in [0, 1]^J, \quad \boldsymbol{\pi} = (\pi_1, \ldots, \pi_J) \in \mathcal{S}^J,$$

$$\mathcal{S}^J = \left\{ \boldsymbol{p} \in [0, 1]^J : \sum_{j=1}^{J} p_j = 1 \right\}, \text{ the probability simplex}$$

- Consider linking a linear regression to the mixture probability of success $\boldsymbol{p}^T \boldsymbol{\pi}$ using the inverse link function $G(\cdot)$. The set of all $\boldsymbol{p}$ that honor the link is denoted

$$A(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\pi}) = \{ \boldsymbol{p} \in [0, 1]^J : \boldsymbol{p}^T \boldsymbol{\pi} = G(\boldsymbol{x}^T \boldsymbol{\beta}) \}$$

- For an independent sample, the following model has nJ parameters in the $\boldsymbol{p}_i$'s

$$T_i \overset{\text{ind}}{\sim} \text{BinMix}(m_i, \boldsymbol{p}_i, \boldsymbol{\pi}), \quad \boldsymbol{p}_i \in A_i$$

- Instead, treat $\boldsymbol{p}_i$'s as random effects using a hierarchical model

$$T_i \mid \boldsymbol{p}_i \overset{\text{ind}}{\sim} \text{BinMix}(m_i, \boldsymbol{p}_i, \boldsymbol{\pi}), \quad \boldsymbol{p}_i \overset{\text{ind}}{\sim} f_{A_i}$$

  where f_{A_i} is a density on the set $A_i = A(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\pi})$

Distributions on the set A

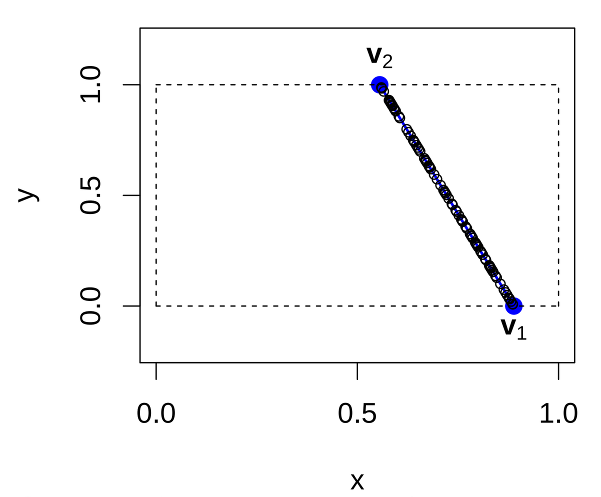
- A is a convex set — a hyperplane in a rectangle — so we can write

$$A(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\pi}) \equiv \left\{ \sum_{\ell=1}^{k} \lambda_\ell \boldsymbol{v}_\ell : \boldsymbol{\lambda} \in \mathcal{S}^k \right\}$$

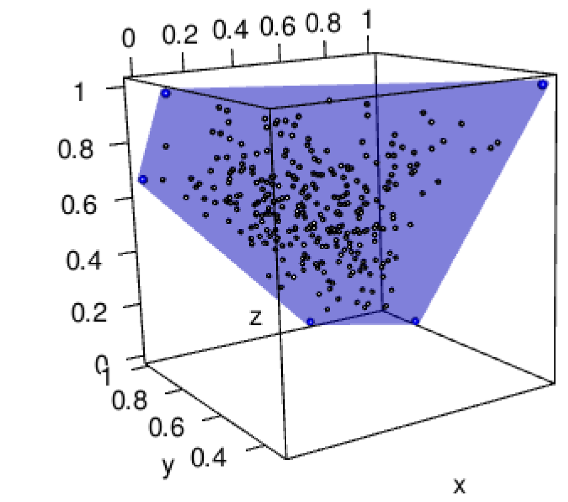
  for appropriate vertices $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$

- Consider a Dirichlet distribution on A

$$\boldsymbol{p} = \sum_{\ell=1}^{k} \lambda_\ell \boldsymbol{v}_\ell, \quad \boldsymbol{\lambda} \sim \text{Dirichlet}_k(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} = (1, \ldots, 1) \iff \text{Uniform}(A)$$



Sample of $n = 100$ on A with $J = 2$, $\boldsymbol{\pi} = (3/4, 1/4)$, $G = 2/3$.



Sample of $n = 300$ on A with $J = 3$, $\boldsymbol{\pi} = (1/4, 1/2, 1/4)$, $G = 2/3$.

- Danaher et al. (2012) recently proposed priors based on the Minkowski-Weyl decomposition to enforce (biologically motivated) polyhedral constraints in Bayesian analysis

Mixture Link Model

- We can now write the hierarchical model

$$\boxed{\text{Mixture Link:} \quad T \sim \text{MixLink}_J(\boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\pi})} \quad T_i \mid \boldsymbol{p}_i, \boldsymbol{\pi} \overset{\text{ind}}{\sim} \text{BinMix}(m_i, \boldsymbol{p}_i, \boldsymbol{\pi})$$

$$\boldsymbol{p}_i = \sum_{\ell=1}^{k_i} \lambda_\ell^{(i)} \boldsymbol{v}_\ell^{(i)}, \quad \text{where } \boldsymbol{v}_1^{(i)}, \ldots, \boldsymbol{v}_{k_i}^{(i)} \text{ are vertices of } A(\boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\pi})$$

$$\boldsymbol{\lambda}^{(i)} \overset{\text{ind}}{\sim} \text{Dirichlet}_{k_i}(1, \ldots, 1)$$

  For now $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\pi})$ will be treated as fixed/unknown (frequentist), but it can also be given a prior distribution (Bayesian)

- Likelihood is then a product of densities

$$f(t_i \mid \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j \int \text{Bin}(t_i \mid m_i, p_{ij}) \cdot f_{A_i}(p_{ij}) dp_{ij} \qquad (*)$$

- When $m = 1$ (single binomial trial), density collapses to usual logistic regression

$$f(t \mid \boldsymbol{\theta}) = G(\boldsymbol{x}^T \boldsymbol{\beta})^t [1 - G(\boldsymbol{x}^T \boldsymbol{\beta})]^{1-t}$$

- If $J = 2$, density has a computationally tractable form

$$f(t \mid \boldsymbol{\theta}) = \binom{m}{t} \sum_{j=1}^{2} \pi_j \frac{B_{v_{1j}}(t+1, m-t+1) - B_{v_{2j}}(t+1, m-t+1)}{v_{1j} - v_{2j}},$$

  where $B_x(\alpha, \beta)$ is the incomplete beta function and $\boldsymbol{v}_\ell = \begin{pmatrix} v_{\ell 1} \\ v_{\ell 2} \end{pmatrix}$ are vertices of A

- For general $\boldsymbol{\alpha}$, the density $(*)$ is an expectation of a linear combination of $\text{Dirichlet}_{k_i}(\boldsymbol{\alpha})$. Numerical methods for this distribution have been studied (Provost and Cheong, 2000; Kotz et al., 2000)

Mixture Link Distribution

Compare three variations on the binomial density which induce extra variation in a fundamentally different way

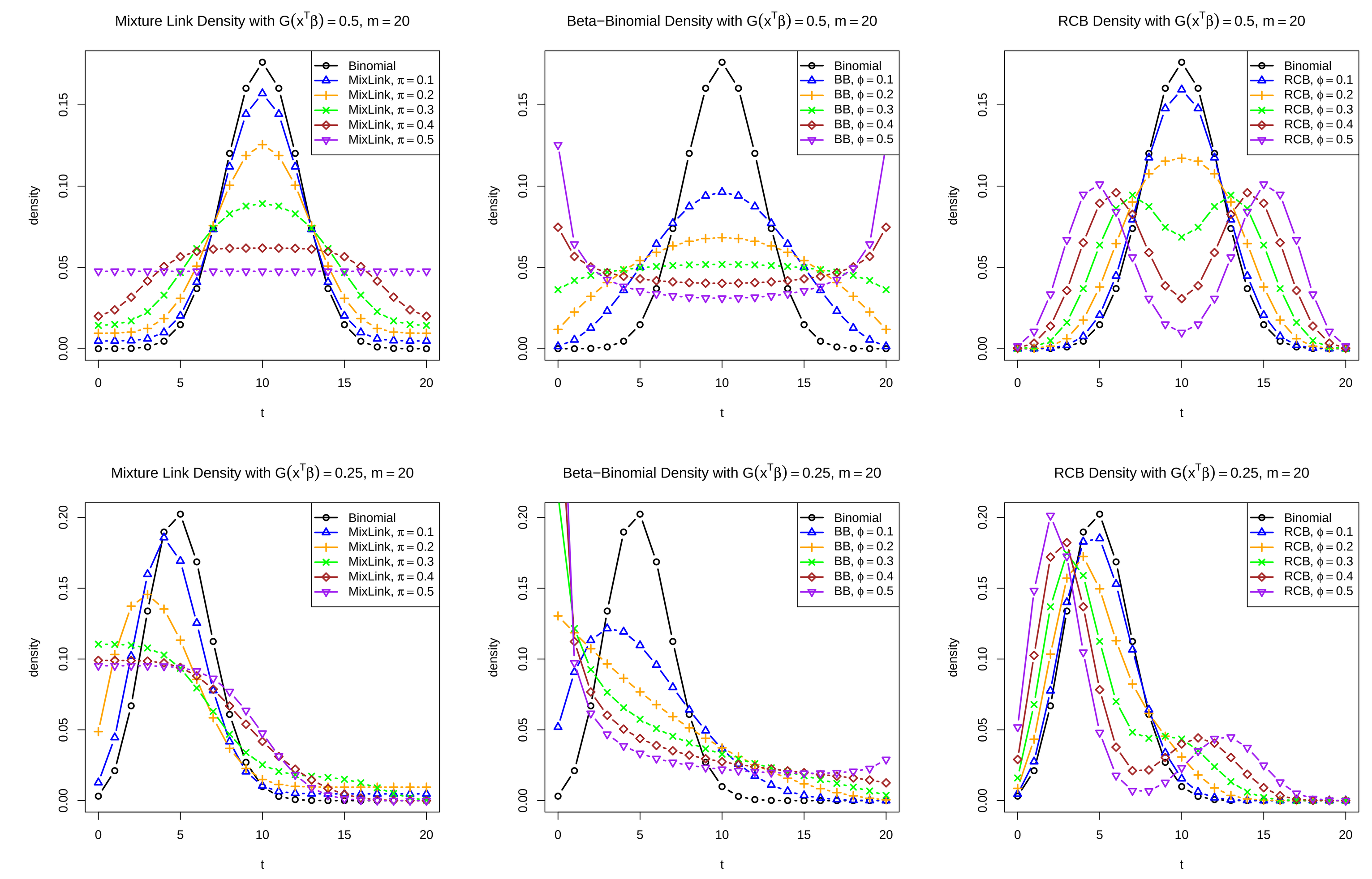
$$\boxed{\text{Beta-Binomial:} \quad T \sim \text{BB}(m, p, \phi)} \quad T \mid \mu \sim \text{Bin}(m, \mu), \quad \mu \sim \text{Beta}(\alpha, \beta)$$

$$\alpha = p\phi^{-2}(1 - \phi^2), \quad \beta = (1 - p)\phi^{-2}(1 - \phi^2)$$

$$\boxed{\text{Random-Clumped Binomial:} \quad T \sim \text{RCB}(m, p, \phi)} \quad T = NY + (X \mid N)$$

$Y \sim \text{Bin}(1, p)$ ← leader's decision (yes or no)
$N \sim \text{Bin}(m, \phi)$ ← number who follow the leader (indep. of Y)
$X \mid N \sim \text{Bin}(m - N, p)$ ← number who decide independently of leader

$p = G(\boldsymbol{x}^T \boldsymbol{\beta})$: probability of success, ϕ : overdispersion parameter



Concluding Remarks

Can fit the $J = 2$ mixture link model with standard optimization software

- We have used `optim` function (limited-memory BFGS method) in R
- Standard errors can be computed from the hessian at the solution
- Simulations show that the `optim` estimator for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\pi})$ is consistent

Next steps

- Find a reliable way to compute the density under general J
- Extend the model so that $\boldsymbol{p}_i$ need not be uniform on A_i (i.e. allow $\boldsymbol{\alpha} \neq \boldsymbol{1}$)
- Continue investigation of frequentist inference, and study Bayesian approaches

References

M. R. Danaher, A. Roy, Z. Chen, S. L. Mumford, and E. 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.

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

S. Kotz, N. Balakrishnan, and N. L. Johnson. *Continuous Multivariate Distributions, Volume 1, Models and Applications.* Wiley-Interscience, 2nd edition, 2000.

S. B. Provost and Y. Cheong. On the distribution of linear combinations of the components of a Dirichlet random vector. *Canadian Journal of Statistics*, 28(2):417–425, 2000.