Mixture Link Models for Binomial Data with Overdispersion

Andrew M. Raim

(Now with U.S. Census Bureau)

Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD, U.S.A.
araim1@umbc.edu

9th Annual Probability and Statistics Day at UMBC

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

Overview

- Overdispersion occurs when a standard statistical model does not capture the variability observed in the data. Commonly encountered in the analysis of categorical and count data.

- We present initial work on the model from Raim (2014, Ph.D. Thesis), some estimators, and an application to Chromosome Aberration data.

- **Idea:** Suppose observations belong to J latent subpopulations with probabilities $\pi_1, \dots, \pi_J$. Consider logistic regression on the overall weighted probability of success

$$\pi_1 \mu_1 + \cdots + \pi_J \mu_J,$$

where $\mu_j = \text{P}(\text{Success} \mid j\text{th subpopulation})$.

Chromosome Aberration Example

An illustrative dataset used in (Morel and Neerchal, 2012), from Awa et al. (1978).

Chromosome aberrations were studied in Hiroshima atomic bomb survivors between Jan 1968 and Nov 1969

- $n = 648$ subjects
- m_i: number of circulating lymphocytes examined on the ith subject (between 30 and 100)
- t_i: number of chromosome aberrations
- d_i: total radiation dose (T65-gamma + T65-neutron, in rads) received by the ith subject
- $z_i = \frac{d_i - \bar{d}}{\sqrt{\frac{1}{n}\sum_{\ell=1}^{n}(d_\ell - \bar{d})^2}}$: standardized radiation dose

for $i = 1, \ldots, n$.

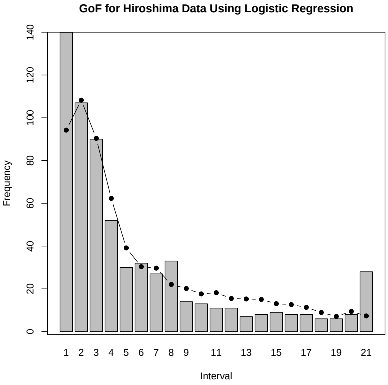
Qn: What is the effect of radiation dose on the probability of chromosome aberration?

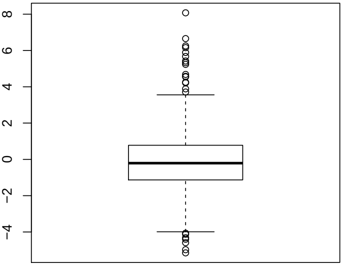
Chromosome Aberration Example

Logistic Regression

$$T_i \overset{\text{ind}}{\sim} \text{Bin}(m_i, p_i),$$
$$p_i = G(\beta_0 + \beta_1 z_i + \beta_2 z_i^2),$$
$$i = 1, \ldots, 648$$

	Estimate (SE)	z-value
β_0	-3.0306 (0.0246)	-123.42
β_1	1.3017 (0.0343)	37.98
β_2	-0.3071 (0.0158)	-19.40





Binomial Regression Models for Overdispersion

Some Established Approaches

- Likelihoods which support overdispersion using latent random variables.
  1. Beta-Binomial (Otake and Prentice, 1984),
  2. Zero-Inflated Binomial (Hall, 2000)
  3. Random-Clumped Binomial (Morel and Nagaraj, 1993).
- Quasi-likelihood methods.
  1. Dispersion multipler (Agresti, 2002, §4.7).
  2. Generalized Estimating Equations (Liang and Zeger, 1986).
- Generalized Linear Mixed Models (McCulloch, Searle, and Neuhaus, 2008).
- Follmann and Lambert (1989) assume a finite mixture to approximate logistic regression with a random intercept. Estimation by nonparametric MLE avoids assumptions on random effect distribution.
- Finite mixtures of regressions. (Frühwirth-Schnatter, 2006).

Mixture Link Binomial Model

Formulation

- Start with a finite mixture of binomial densities,

$$T \sim f(t \mid m, \boldsymbol{\theta}) = \sum_{j=1}^{J} \pi_j \binom{m}{t} \mu_j^t (1-\mu_j)^{m-t},$$

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J) \in \mathcal{S}^J \quad \text{(the probability simplex in } \mathbb{R}^J\text{)},$$

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J) \in [0,1]^J.$$

- Mixture success probability of a single trial is $\mathrm{E}(T/m) = \boldsymbol{\mu}^T \boldsymbol{\pi} = \pi_1 \mu_1 + \cdots + \pi_J \mu_J$.
- **Objective**: Link regression function $\mathbf{x}_i^T \boldsymbol{\beta}$ to $\boldsymbol{\mu}^T \boldsymbol{\pi}$ using logistic link,

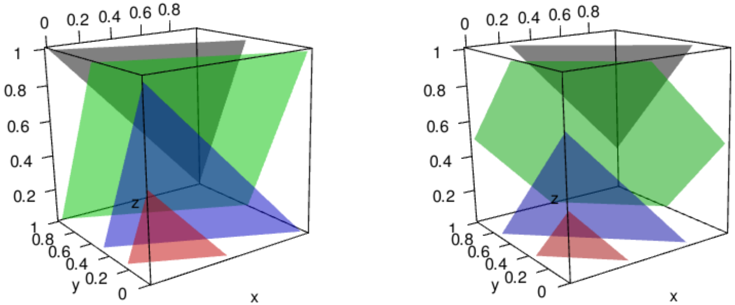
$$\boldsymbol{\mu}^T \boldsymbol{\pi} \stackrel{\text{link}}{=} p_i, \quad \text{where } p_i \stackrel{\text{def}}{=} G(\mathbf{x}_i^T \boldsymbol{\beta}).$$

- To allow the possiblity of a regression, suppose $\boldsymbol{\mu}_i$ varies with the observation as well. To enforce the link, $\boldsymbol{\mu}_i$ must be in the set

$$A(p_i, \boldsymbol{\pi}) = \{\boldsymbol{\mu} \in [0,1]^J : \boldsymbol{\mu}^T \boldsymbol{\pi} = p_i\}.$$

- For the no-regression case, take $p_i = p$.

Visualizing A with $J = 3$



(a) $\boldsymbol{\pi} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$

(b) $\boldsymbol{\pi} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

Figure: The set $A(p, \boldsymbol{\pi}) = \left\{ \boldsymbol{\mu} \in [0, 1]^3 : \mu_1 \pi_1 + \mu_2 \pi_2 + \mu_3 \pi_3 = p \right\}$. In each case, $p \in \{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$ is shown (from front to back).

Mixture Link Binomial Model

Random Effects Approach

- Take $\boldsymbol{\mu}_i$ as random effect to avoid dimensionality issue.
- $A_i = \{\boldsymbol{\mu} \in [0,1]^J : \boldsymbol{\mu}^T \boldsymbol{\pi} = p_i\}$ is a bounded convex set. Therefore we can find vertices $\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{k_i}^{(i)} \in \mathbb{R}^J$ such that

$$A_i = \text{conv}(\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{k_i}^{(i)}) = \left\{ \sum_{\ell=1}^{k_i} \lambda_\ell \mathbf{v}_\ell^{(i)} : \boldsymbol{\lambda} \in \mathcal{S}^{k_i} \right\} = \left\{ \mathbf{V}^{(i)} \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathcal{S}^{k_i} \right\}.$$

- $\mathbf{V}^{(i)}$ can vary for each observation. Number of vertices k_i can also vary.
- We will consider $\boldsymbol{\mu}_i = \mathbf{V}^{(i)} \boldsymbol{\lambda}^{(i)} \in A_i$ with $\boldsymbol{\lambda}^{(i)} \stackrel{\text{ind}}{\sim} \text{Dirichlet}_{k_i}(\boldsymbol{\alpha})$

$$f(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}) = \frac{\lambda_1^{\alpha_1 - 1} \cdots \lambda_k^{\alpha_k - 1}}{\text{B}(\boldsymbol{\alpha})} \cdot I(\boldsymbol{\lambda} \in \mathcal{S}^k), \quad \text{where } \text{B}(\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)}{\Gamma(\alpha_1 + \cdots + \alpha_k)}.$$

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

Mixture Link Binomial Model

Hierarchical Model

We can write the model as

$$
\begin{aligned}
&T_i \mid \boldsymbol{\mu}_i, \boldsymbol{\pi} \overset{\text{ind}}{\sim} \text{BinMix}(m_i, \boldsymbol{\mu}_i, \boldsymbol{\pi}) \\
&\boldsymbol{\mu}_i = \mathbf{V}^{(i)} \boldsymbol{\lambda}^{(i)}, \quad \text{where } \mathbf{V}^{(i)} = (\mathbf{v}_1^{(i)} \cdots \mathbf{v}_{k_i}^{(i)}) \text{ are vertices of } A(p_i, \boldsymbol{\pi}) \\
&\boldsymbol{\lambda}^{(i)} \overset{\text{ind}}{\sim} \text{Dirichlet}_{k_i}(\boldsymbol{\alpha}^{(i)}).
\end{aligned}
$$

Assume **Symmetric Dirichlet** with $\boldsymbol{\alpha}^{(i)} = (\kappa, \ldots, \kappa)$ for $\kappa > 0$, because:

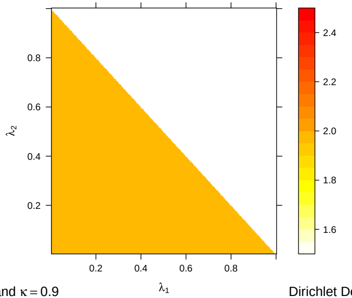
- k_i can vary between observations.
- Difficult to maintain correspondence between $\mathbf{v}_\ell^{(i)}$ and $\alpha_\ell^{(i)}$.

$$
\text{Density: } f(t \mid m, \boldsymbol{\theta}) = \binom{m}{t} \sum_{j=1}^{J} \pi_j \int w^t (1-w)^{m-t} \cdot f_{\mathbf{v}_{j\cdot}^T \boldsymbol{\lambda}}(w)\, dw
$$

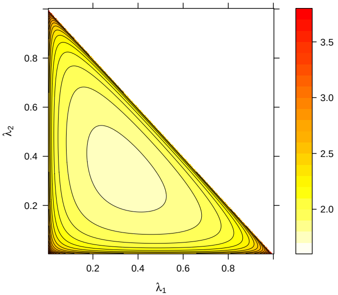
$$
\text{Parameterized by: } \boldsymbol{\theta} = \begin{cases} (p, \boldsymbol{\pi}, \kappa) \in \mathbb{R}^{1+(J-1)+1}, & \text{no-regression case,} \\ (\boldsymbol{\beta}, \boldsymbol{\pi}, \kappa) \in \mathbb{R}^{d+(J-1)+1}, & \text{regression case.} \end{cases}
$$

Symmetric Dirichlet Density

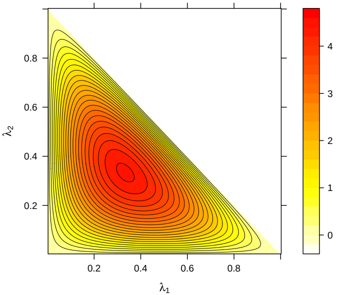
Dirichlet Density for $k = 3$ and $\kappa = 1$



Dirichlet Density for $k = 3$ and $\kappa = 0.9$



Dirichlet Density for $k = 3$ and $\kappa = 2$



Mixture Link Binomial Model

Expectation and Variance

- Recall moments of $\boldsymbol{\lambda} \sim \text{Dirichlet}(\boldsymbol{\alpha})$

$$\text{E}(\boldsymbol{\lambda}) = \frac{\boldsymbol{\alpha}}{\alpha_0}, \quad \text{Var}(\boldsymbol{\lambda}) = \frac{\alpha_0 \, \text{Diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}\boldsymbol{\alpha}^T}{\alpha_0^2(\alpha_0 + 1)}, \quad \text{E}(\boldsymbol{\lambda}\boldsymbol{\lambda}^T) = \frac{\text{Diag}(\boldsymbol{\alpha}) + \boldsymbol{\alpha}\boldsymbol{\alpha}^T}{\alpha_0(\alpha_0 + 1)}.$$

- The expectation and variance of $T \sim \text{MixLink}_J(m, p, \boldsymbol{\pi}, \kappa)$ can be obtained as

$$\text{E}(T) = m \sum_{j=1}^{J} \pi_j \bar{v}_{j.} \equiv mp,$$

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

where $\mathbf{v}_{j.}$ contains elements of the jth row of $\mathbf{V}$, and $\bar{v}_{j.}$ is its mean.

Computing the Vertices of A

Lemma

Suppose $J = 2$, $A = \{\mu \in [0,1]^2 : \mu_1\pi_1 + \mu_2\pi_2 = p\}$ has two distinct vertices $\mathbf{v}_1, \mathbf{v}_2$, and $0 < \pi_1 < 1$. Then the vertices of A are given by

$$\mathbf{v}_1 = \begin{cases} \left(\frac{1}{\pi_1}(p - \pi_2), 1\right), & \textit{if } \frac{1}{\pi_1}(p - \pi_2) \geq 0 \\ \left(0, \frac{1}{\pi_2}p\right), & \textit{o.w.}, \end{cases}$$

$$\mathbf{v}_2 = \begin{cases} \left(\frac{1}{\pi_1}p, 0\right), & \textit{if } \frac{1}{\pi_1}p \leq 1 \\ \left(1, \frac{1}{\pi_2}(p - \pi_1)\right), & \textit{o.w.}, \end{cases}$$

where $\pi_2 = 1 - \pi_1$.

Lemma (Characterization of Extreme Points of A)

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

Computing the Vertices of A

Algorithm: Find vertices of the set $A(p, \boldsymbol{\pi})$.

$$
\begin{array}{l}
\textbf{function } \textsc{FindVertices}(p, \boldsymbol{\pi}) \\
\quad \mathcal{V} \leftarrow \varnothing \\
\quad \textbf{for } j = 1, \ldots, J \textbf{ do} \\
\quad\quad \textbf{if } \pi_j > 0 \textbf{ then} \\
\quad\quad\quad \textbf{for all } \boldsymbol{\mu}_{-j} \in \{0,1\}^{J-1} \textbf{ do} \\
\quad\quad\quad\quad \mu_j^* \leftarrow \frac{1}{\pi_j}\left[p - \boldsymbol{\mu}_{-j}^T \boldsymbol{\pi}_{-j}\right] \\
\quad\quad\quad\quad \mathbf{v}^* \leftarrow (\mu_1, \ldots, \mu_{j-1}, \mu_j^*, \mu_{j+1}, \ldots, \mu_J) \\
\quad\quad\quad\quad \textbf{if } \mathbf{v}^* \in A \textbf{ then} \\
\quad\quad\quad\quad\quad \mathcal{V} \leftarrow \mathcal{V} \cup \{\mathbf{v}^*\} \\
\quad \text{Let } \mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_k) \text{ for all } \mathbf{v}_\ell \in \mathcal{V} \\
\quad \textbf{return } \mathbf{V}
\end{array}
$$

- In searching for extreme points, we must only consider those with at most one component not equal to 0 or 1.
- Special case of vertex finding for polyhedron $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$.
- Algorithm checks $J \cdot 2^{J-1}$ points, and therefore impractical for large J.

Computing the Density

Linear Combination of Dirichlet

$$\text{Recall: } f(t \mid m, \boldsymbol{\theta}) = \binom{m}{t} \sum_{j=1}^{J} \pi_j \int w^t (1-w)^{m-t} \cdot f_{\mathbf{v}_{j\cdot}^T \boldsymbol{\lambda}}(w)\, dw$$

- The Mixture Link density depends on density of $\mathbf{v}_{j\cdot}^T \boldsymbol{\lambda}$. Provost and Cheong (2000) relate this distribution to the linear combination of χ^2 random variables.

- If $X_j \stackrel{\text{ind}}{\sim} \chi^2_{v_j}$ for $j = 1, \ldots, k$, then (Kotz et al., 2000)

$$\left(\frac{X_1}{\sum_{j=1}^k X_j}, \ldots, \frac{X_k}{\sum_{j=1}^k X_j} \right) \sim \text{Dirichlet}_k(\boldsymbol{\alpha}), \quad \text{where } \alpha_j = v_j/2.$$

- Now if $\boldsymbol{\lambda} \sim \text{Dirichlet}_k(\boldsymbol{\alpha})$, we may write the distribution of a linear combination $\mathbf{c}^T \boldsymbol{\lambda}$ as

$$\mathrm{P}\left(\sum_{j=1}^k c_j \lambda_j \leq x \right) = \mathrm{P}\left(\sum_{j=1}^k c_j \frac{X_j}{\sum_{\ell=1}^k X_\ell} \leq x \right) = \mathrm{P}\left(\sum_{j=1}^k (c_j - x) X_j \leq 0 \right).$$

Computing the Density

Linear Combination of Dirichlet

- The cdf of $\mathbf{b}^T\mathbf{X}$ is obtained using inversion formula (Imhof, 1961)

$$\begin{aligned} F_{\mathbf{b}^T\mathbf{X}}(x) &= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathfrak{Im}\{e^{-iux}\phi(u)\}}{u}du \\ &= \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin(\frac{1}{2}\sum_{j=1}^k v_j \arctan(b_j u) - \frac{1}{2}xu)}{u\prod_{j=1}^k (1+b_j^2u^2)^{v_j/4}}du \end{aligned}$$

where $\phi_{\mathbf{b}^T\mathbf{X}}(t) = \prod_{j=1}^k (1-2b_j it)^{-v_j/2}$ is the characteristic function.

- Therefore the probability $\mathrm{P}\left(\mathbf{c}^T\boldsymbol{\lambda} \le x\right)$ can be computed by

$$F_{\mathbf{c}^T\boldsymbol{\lambda}}(x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\left(\sum_{j=1}^k \alpha_j \arctan\{(c_j - x)u\}\right)}{u\prod_{j=1}^k \left(1+(c_j-x)^2u^2\right)^{\alpha_j/2}}du$$

- See the `imhof` function from `CompQuadForm` package in `R`.

Beta Approximation to the Density

- We can approximate the linear combination of Dirichlet density by a simpler beta density; select parameters by moment matching.
- Suppose $B \sim \text{Beta}(a, b)$ so that $B^* = (u - \ell)B + \ell$ for $\ell < u$ has a shifted/scaled beta distribution on the interval (ℓ, u) with

$$\text{E}(B^*) = (u - \ell)\frac{a}{a+b} + \ell, \quad \text{Var}(B^*) = (u - \ell)^2 \frac{ab}{(a+b)^2(a+b+1)}.$$

- For $\boldsymbol{\lambda} \sim \text{Dirichlet}_k(\kappa \mathbf{1})$, we have

$$\xi = \text{E}(\mathbf{c}^T \boldsymbol{\lambda}) = \bar{c}, \quad \tau^2 = \text{Var}(\mathbf{c}^T \boldsymbol{\lambda}) = \frac{k\mathbf{c}^T\mathbf{c} - (k\bar{c})^2}{k^2(1 + k\kappa)}$$

- Equating $\text{E}(B^*) = \xi$ and $\text{Var}(B^*) = \tau^2$ and solving for a and b, we obtain that

$$a = \left(\frac{\xi - \ell}{\tau}\right)^2 \frac{u - \xi}{u - \ell} - \frac{\xi - \ell}{u - \ell}, \quad b = a\left(\frac{u - \xi}{\xi - \ell}\right)$$

Beta Approximation to the Density

- Now let (ℓ_j, u_j) represent the range of $\mathbf{v}_{j\cdot}^T \boldsymbol{\lambda}$

- To obtain the beta approximation to Mixture Link, let

$$\xi_j = \mathrm{E}(\mathbf{v}_{j\cdot}^T \boldsymbol{\lambda}) = \bar{v}_{j\cdot}, \quad \tau_j^2 = \mathrm{Var}(\mathbf{v}_{j\cdot}^T \boldsymbol{\lambda}) = \frac{k \mathbf{v}_{j\cdot}^T \mathbf{v}_{j\cdot} - (k \bar{v}_{j\cdot})^2}{k^2(1 + k\kappa)},$$

  so that $B_j^* \sim (u_j - \ell_j)\mathrm{Beta}(a_j, b_j) + \ell_j$ is moment-matched with

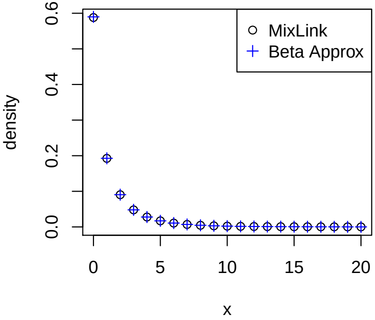
$$a_j = \left(\frac{\xi_j - \ell_j}{\tau_j} \right)^2 \frac{u_j - \xi_j}{u_j - \ell_j} - \frac{\xi_j - \ell_j}{u_j - \ell_j} \quad \text{and} \quad b_j = a_j \left(\frac{u_j - \xi_j}{\xi_j - \ell_j} \right).$$

- Now an approximation to the Mixture Link density may be computed as

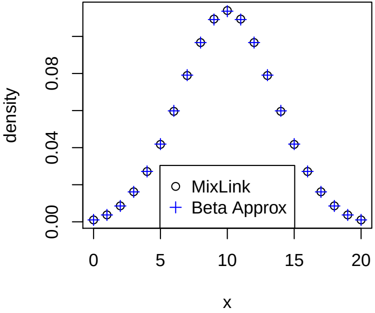
$$f(t \mid m, \boldsymbol{\theta}) \approx \binom{m}{t} \sum_{j=1}^{J} \pi_j \int_{\ell_j}^{u_j} w^t (1-w)^{m-t} \frac{1}{u_j - \ell_j} h \left(\frac{w - \ell_j}{u_j - \ell_j} \;\middle|\; a_j, b_j \right) dw$$

Beta Approximation to the Density

$\kappa = 0.5$



(a) $p = 0.05$



(b) $p = 0.50$

Figure: Comparison of exact Mixture Link density f and density g using beta approximation with $m = 20$ trials and $\boldsymbol{\pi} = (\frac{1}{20}, \frac{2}{20}, \frac{3}{20}, \frac{4}{20}, \frac{10}{20})$.

Estimator using Sample Proportions

No-Regression Case

- Consider a sample $T_i \stackrel{\text{ind}}{\sim} \text{MixLink}_J(m_i, p, \boldsymbol{\pi}, \kappa)$ for $i = 1, \ldots, n$.
- Recall that $\text{E}(T_i / m_i) = p$.
- An unbiased estimator of p is $\tilde{p} = \frac{1}{n} \sum_{i=1}^n T_i / m_i$.
- The variance of $\tilde{p}$ is

$$\text{Var}(\tilde{p}) = \frac{1}{n^2} \sum_{i=1}^n \left[\frac{1}{m_i} p(1 - m_i p) + \frac{m_i - 1}{m_i} \sum_{j=1}^J \pi_j \frac{\mathbf{v}_{j\cdot}^T \mathbf{v}_{j\cdot} + \kappa(k \bar{v}_{j\cdot})^2}{k(1 + \kappa k)} \right],$$

  a function of p, $\boldsymbol{\pi}$, and κ.
- If the sequence $\{m_n\}$ is bounded, we have

$$\frac{\tilde{p} - p}{\sqrt{\text{Var}(\tilde{p})}} \xrightarrow{\mathcal{L}} \text{N}(0, 1), \quad \text{as } n \to \infty.$$

Moment Estimator for κ

No-Regression Case, iid

- Consider an iid sample $T_i \stackrel{\text{ind}}{\sim} \text{MixLink}_J(m, p, \boldsymbol{\pi}, \kappa)$ for $i = 1, \ldots, n$.
- Can derive a moment estimator for κ, given p and $\boldsymbol{\pi}$, using moment

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

- Taking $W = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i(T_i-1)}{m(m-1)}$, consider

$$\tilde{\kappa}(W) = \frac{\sum_{j=1}^{J} \pi_j \mathbf{v}_{j.}^T \mathbf{v}_{j.} - kW}{k^2 W - \sum_{j=1}^{J} \pi_j (k\bar{v}_{j.})^2}.$$

- For large samples $\tilde{\kappa}$ is normal with mean κ and variance

$$\frac{1}{n}\left\{\frac{k\sum_{j=1}^{J}\pi_j \bar{v}_{j.}^2 - \sum_{j=1}^{J}\pi_j \mathbf{v}_{j.}^T\mathbf{v}_{j.}}{k^2\left[\mathrm{E}\left(\frac{T(T-1)}{m(m-1)}\right) - \sum_{j=1}^{J}\pi_j\bar{v}_{j.}^2\right]^2}\right\}^2 \mathrm{Var}\left[\frac{T(T-1)}{m(m-1)}\right].$$

- $\tilde{\kappa}$ need not be positive, but $\mathrm{P}(\tilde{\kappa} < 0) \to 0$ as $n \to \infty$.

Gauss-Newton Estimator

Regression Case

- Consider a sample $T_i \stackrel{\text{ind}}{\sim} \text{MixLink}_J(m_i, p_i, \boldsymbol{\pi}, \kappa)$, where $p_i = G(\mathbf{x}_i^T \boldsymbol{\beta})$.
- We may write
$$T_i / m_i = G(\mathbf{x}_i^T \boldsymbol{\beta}) + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{ind}}{\sim} (0, \sigma_i^2)$$
and estimate $\boldsymbol{\beta}$ by minimizing $Q(\boldsymbol{\beta}) = \sum_{i=1}^n \left[T_i / m_i - G(\mathbf{x}_i^T \boldsymbol{\beta}) \right]^2$.
- This yields Gauss-Newton iterations
$$\boldsymbol{\beta}^{(g+1)} = \boldsymbol{\beta}^{(g)} - (\mathbf{J}_{\boldsymbol{\beta}^{(g)}}^T \mathbf{J}_{\boldsymbol{\beta}^{(g)}})^{-1} \mathbf{J}_{\boldsymbol{\beta}^{(g)}}^T \mathbf{r}_{\boldsymbol{\beta}^{(g)}}, \quad \text{for } g = 0, 1, \ldots,$$
$\mathbf{J}_{\boldsymbol{\beta}}$ is the $n \times d$ matrix with entries $\Big(-\partial G(\mathbf{x}_i^T \boldsymbol{\beta}) \,/\, \partial \beta_j \Big)$,
$\mathbf{r}_{\boldsymbol{\beta}}$ is the $n \times 1$ matrix with entries $\Big(T_i / m_i - G(\mathbf{x}_i^T \boldsymbol{\beta}) \Big)$.
- If $\boldsymbol{\beta}^{(0)}$ is a consistent estimator for $\boldsymbol{\beta}$, then $\mathbf{V}_{\boldsymbol{\beta}}^{-1/2} (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \text{N}(\mathbf{0}, \mathbf{I}_d)$ as $n \to \infty$, with
$$\mathbf{V}_{\boldsymbol{\beta}} = (\mathbf{J}_{\boldsymbol{\beta}}^T \mathbf{J}_{\boldsymbol{\beta}})^{-1} \mathbf{J}_{\boldsymbol{\beta}}^T \mathbf{V}_{\tilde{p}} \mathbf{J}_{\boldsymbol{\beta}} (\mathbf{J}_{\boldsymbol{\beta}}^T \mathbf{J}_{\boldsymbol{\beta}})^{-1} \quad \text{and} \quad \mathbf{V}_{\tilde{p}} = \text{Diag}\{\text{Var}(T_i / m_i)\}.$$

Chromosome Aberration Data

Compare models for goodness-of-fit:

- Logistic: $T_i \stackrel{\text{ind}}{\sim} \text{Bin}(m_i, p_i)$,
- RCB: $T_i \stackrel{\text{ind}}{\sim} \text{RCB}(m_i, p_i, \phi)$,
- BB: $T_i \stackrel{\text{ind}}{\sim} \text{BB}(m_i, p_i, \phi)$,
- RCB-Reg: $T_i \stackrel{\text{ind}}{\sim} \text{RCB}(m_i, p_i, \phi_i)$,
- BB-Reg: $T_i \stackrel{\text{ind}}{\sim} \text{BB}(m_i, p_i, \phi_i)$,
- MixLinkJ2: $T_i \stackrel{\text{ind}}{\sim} \text{MixLink}_2(m_i, p_i, \boldsymbol{\pi}, \kappa)$,

with regressions

- $g(p_i) = \beta_0 + \beta_1 z_i + \beta_2 z_i^2$ for all models,
- $g(\phi_i) = \gamma_0 + \gamma_1 z_i + \gamma_2 z_i^2$ for the two "-Reg" models.

We consider two likelihood-dependent ways to evaluate model performance.

- A variation on the Pearson chi-square GOF test statistic to allow varying m_i in binomial setting (Sutradhar et al., 2008).
- Randomized quantile residuals (Dunn and Smyth, 1996).

Numerical MLE used for all models in this study.

Chromosome Aberration Data

GOF Test for Varying m_i

- To test a binomial model for GOF

$$H_0 : T_i \stackrel{\text{ind}}{\sim} f(t_i \mid m_i, \boldsymbol{\theta}) \text{ for some } \boldsymbol{\theta} \in \Theta \quad \text{vs.} \quad H_1 : \text{Not.}$$

- GOF test statistic

$$X(\boldsymbol{\theta}) = \sum_{\ell=1}^{r} \frac{[O_\ell - E_\ell(\boldsymbol{\theta})]^2}{E_\ell(\boldsymbol{\theta})}, \quad \text{where}$$

$$E_\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=0}^{m_i} f(t \mid m_i, \boldsymbol{\theta}) \, I\left(\frac{t}{m_i} \in \mathrm{I}_\ell\right) \quad \text{and} \quad O_\ell = \sum_{i=1}^{n} I\left(\frac{t_i}{m_i} \in \mathrm{I}_\ell\right)$$

and $\mathrm{I}_1, \ldots, \mathrm{I}_r$ are disjoint intervals that cover $[0, 1]$.

- Analyst is free to select I_ℓ, but it is suggested to follow the rule of thumb that all $E_\ell(\boldsymbol{\theta}) \geq 5$.

Chromosome Aberration Data

GOF Test for Varying m_i

Sutradhar et al. (2008) show that

- $X(\boldsymbol{\theta}) \sim \chi^2_{r-1}$ when all parameters are known.

- $X(\hat{\boldsymbol{\theta}}) \sim \chi^2_{r-1-q}$ when $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ is estimated by maximizing the *grouped* likelihood

$$L_g(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{\ell=1}^{r} \left[\mathrm{P}\left(\frac{t_i}{m_i} \in \mathrm{I}_\ell \;\middle|\; m_i, \boldsymbol{\theta} \right)^{I\left(\frac{t_i}{m_i} \in \mathrm{I}_\ell\right)} \right]$$

- **Recovery of df**. When $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q$ is estimated by maximizing the *ungrouped* likelihood

$$L_u(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i \mid m_i, \boldsymbol{\theta})$$

$X(\hat{\boldsymbol{\theta}})$ follows a χ^2_ν distribution with ν between $r-1-q$ and $r-1$.

Chromosome Aberration Data

Randomized Quantile Residuals

- Dunn and Smyth (1996) propose randomized quantile residuals for diagnostics on GLMs and other non-normal models.

- Interpretation of residuals 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,

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

Chromosome Aberration Data

Maximum Likelihood Estimates

	Logistic
β_0	-3.0306 (0.0246)
β_1	1.3017 (0.0343)
β_2	-0.3071 (0.0158)

	RCB
β_0	-2.9901 (0.0352)
β_1	1.2040 (0.0415)
β_2	-0.3429 (0.0242)
ϕ	0.1511 (0.0080)

	BB
β_0	-2.9487 (0.0445)
β_1	1.1144 (0.0550)
β_2	-0.2676 (0.0276)
ϕ	0.1661 (0.0076)

	RCB-Reg
β_0	-3.0699 (0.0338)
β_1	1.3010 (0.0444)
β_2	-0.3705 (0.0244)
γ_0	-2.3526 (0.0965)
γ_1	0.9331 (0.1569)
γ_2	-0.2365 (0.0565)

	BB-Reg
β_0	-3.0145 (0.0445)
β_1	1.3594 (0.0564)
β_2	-0.3449 (0.0332)
γ_0	-1.8611 (0.0737)
γ_1	0.7993 (0.1109)
γ_2	-0.1610 (0.0525)

	MixLinkJ2
β_0	-3.0061 (0.0441)
β_1	1.3656 (0.0562)
β_2	-0.3383 (0.0314)
π_1	0.3297 (0.0175)
κ	1.6293 (0.2472)

(Standard errors using Hessian are in parentheses.)

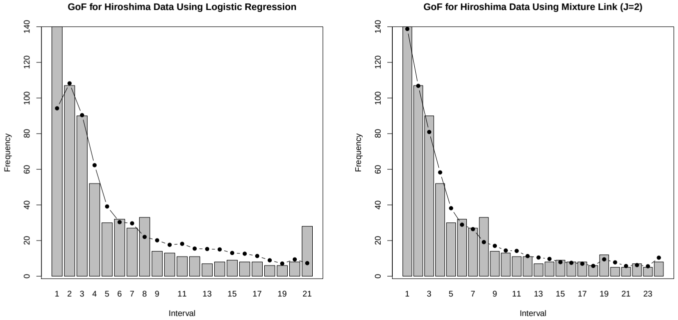
β_0	β_1	β_2	π_1	κ
0.0458	0.0520	0.0306	0.0170	0.2858

(Standard errors for MixLinkJ2 using 500 bootstrap samples.)

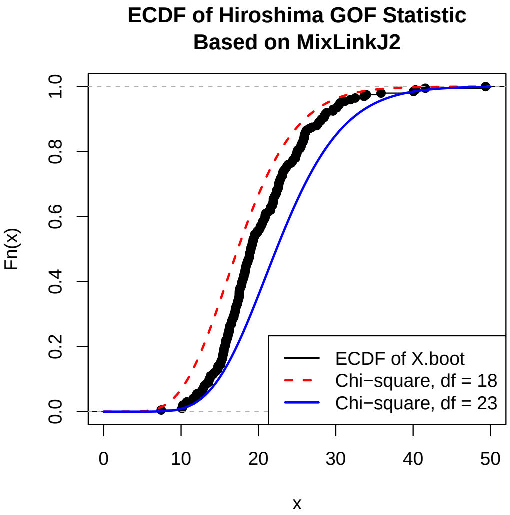
Chromosome Aberration Data

Model Comparison Statistics

Model	LogLik	q	AIC	BIC	GOF statistic	df range	p-value
Logistic	-1814.19	3	3634.40	3647.80	110.38	[17,20]	$< 10^{-13}$
RCB	-1567.50	4	3143.00	3160.90	68.25	[15,19]	$< 10^{-6}$
BB	-1487.92	4	2983.85	3001.74	93.79	[12,18]	$< 10^{-11}$
RCB-Reg	-1546.61	6	3105.22	3132.07	63.96	[18,22]	$< 10^{-5}$
BB-Reg	-1429.61	6	2871.21	2898.05	19.40	[17,23]	> 0.3063
MixLinkJ2	-1433.33	5	2876.66	2905.51	19.50	[18,23]	> 0.3615



Chromosome Aberration Data

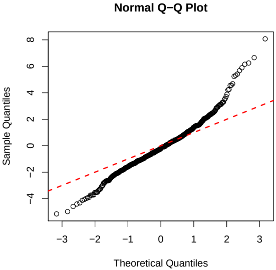


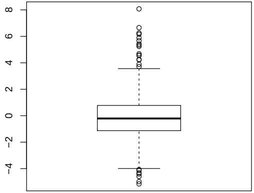
Empirical CDF computed from 200 parametric bootstrap samples

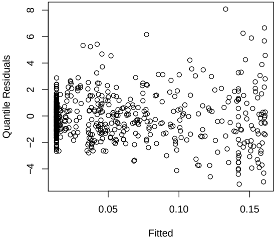
$$T_i^{(b)} \overset{\text{ind}}{\sim} \text{MixLink}_2(\mathbf{x}_i, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\pi}}, \hat{\kappa}) \text{ for } b = 1, \ldots, 200$$

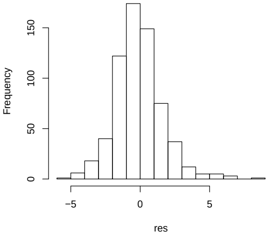
Chromosome Aberration Data

Quantile Residuals for Logistic



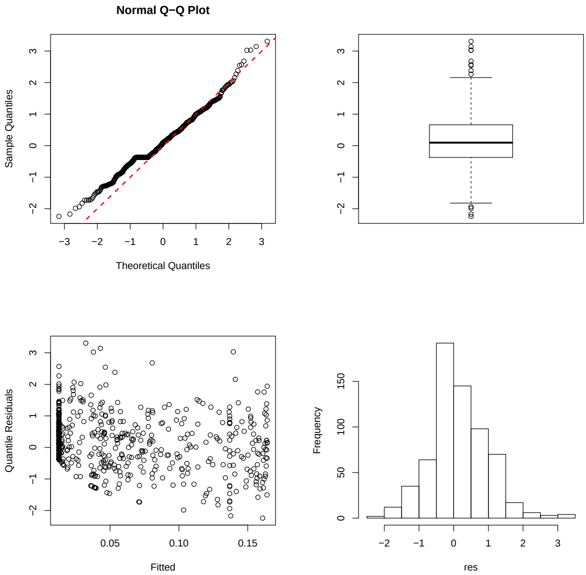






Chromosome Aberration Data

Quantile Residuals for MixLinkJ2



Conclusions and Future Work

Conclusions

- Starting from a finite mixture of binomials, we propose a model to link the mixture probability of success to a regression.
- The mixture success probabilities $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iJ})$ are treated as random effects on the set of all $\boldsymbol{\mu}$ where link to the regression holds.
- Density takes work to evaluate, but is easy to draw from.
- Model-dependent quantities such as GOF statistic and quantile residuals are adversely affected by overdispersion.

Future Work

- Further development of frequentist estimation and inference.
- Bayesian inference.
- Effect of increasing J?
- Extend to other outcome types: Normal, Poisson, etc.

References I

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

A. Awa, T. Sofuni, T. Honda, M. Itoh, S. Neriishi, and M. Otake. Relationship between the radiation dose and chromosome aberrations in atomic bomb survivors of Hiroshima and Nagasaki. *Journal of Radiation Research*, 19(2): 126–140, 1978.

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.

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

Dean A. Follmann and Diane Lambert. Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, 84(405): 295–300, 1989.

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

Daniel B. Hall. Zero-inflated poisson and binomial regression with random effects: A case study. *Biometrics*, 56(4):1030–1039, 2000.

References II

J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3):419–426, 1961.

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

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

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.

Jorge G. Morel and Nagaraj K. Neerchal. *Overdispersion Models in SAS*. SAS Institute, 2012.

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.

Santosh C. Sutradhar, Nagaraj K. Neerchal, and Jorge G. Morel. A goodness-of-fit test for overdispersed binomial (or multinomial) models. *Journal of Statistical Planning and Inference*, 138(5):1459–1471, 2008.

Table: Distance $D(f, g)$ between MixLink f and beta approx g with $m = 20$ trials.

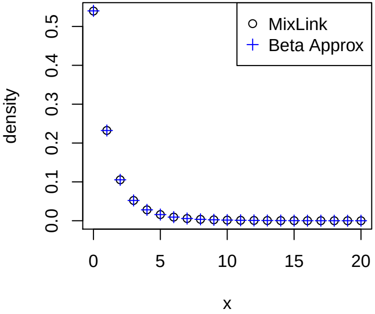
$\boldsymbol{\pi}$	p	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$
$(\frac{1}{2}, \frac{1}{2})$	0.05	1.390E−03	2.220E−16	1.665E−16
	0.1	1.595E−03	5.135E−16	5.829E−16
	0.5	4.767E−07	4.718E−16	6.800E−16
$(\frac{1}{4}, \frac{3}{4})$	0.05	9.483E−04	1.665E−16	3.331E−16
	0.1	1.332E−03	4.163E−16	3.608E−16
	0.5	1.014E−03	9.576E−16	1.193E−15
$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	0.05	1.716E−03	2.047E−06	4.290E−06
	0.1	1.214E−03	8.808E−07	2.207E−06
	0.5	3.488E−03	1.268E−03	3.766E−04
$(\frac{1}{6}, \frac{2}{6}, \frac{3}{6})$	0.05	1.825E−03	2.906E−06	4.546E−06
	0.1	1.125E−03	7.904E−07	2.529E−06
	0.5	1.935E−03	8.333E−04	3.257E−04
$(\frac{1}{10}, \frac{2}{10}, \frac{7}{10})$	0.05	2.092E−03	5.015E−06	4.519E−06
	0.1	1.396E−03	1.486E−06	2.663E−06
	0.5	5.658E−04	2.556E−04	5.106E−05
$(0.05, 0.1, 0.85)$	0.05	2.408E−03	8.446E−06	6.074E−06
	0.1	4.902E−04	1.828E−04	6.382E−05
	0.5	3.740E−04	7.331E−05	2.303E−06

Table: Distance $D(f, g)$ between MixLink f and beta approx g with $m = 20$ trials.

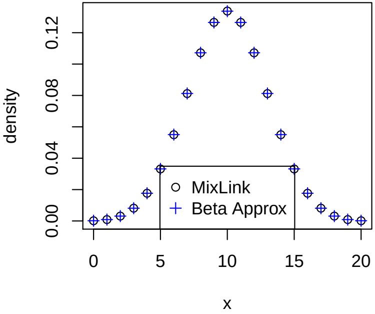
$\boldsymbol{\pi}$	p	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 2$
$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$	0.05	1.637E−03	2.766E−06	4.332E−06
	0.1	1.157E−03	7.914E−07	2.343E−06
	0.5	2.531E−07	1.200E−07	1.882E−07
$(\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10})$	0.05	1.925E−03	3.953E−06	4.724E−06
	0.1	1.142E−03	1.199E−06	3.008E−06
	0.5	5.745E−04	1.818E−04	4.667E−05
$(0.05, 0.1, 0.15, 0.7)$	0.05	2.490E−03	9.792E−06	5.821E−06
	0.1	3.946E−04	4.358E−04	2.040E−04
	0.5	2.026E−04	5.511E−05	9.627E−06
$(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$	0.05	1.995E−03	3.198E−06	4.050E−06
	0.1	1.152E−03	9.531E−07	2.010E−06
	0.5	9.851E−05	2.251E−05	4.207E−06
$(\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{4}{15}, \frac{5}{15})$	0.05	1.902E−03	4.634E−06	4.698E−06
	0.1	7.119E−03	3.804E−03	1.515E−03
	0.5	6.814E−06	3.633E−06	9.605E−07
$(\frac{1}{20}, \frac{2}{20}, \frac{3}{20}, \frac{4}{20}, \frac{10}{20})$	0.05	2.230E−03	7.487E−06	5.065E−06
	0.1	1.473E−03	9.554E−04	4.061E−04
	0.5	3.739E−04	9.918E−05	2.944E−05

Beta Approximation to the Density

$\kappa = 1$



(a) $p = 0.05$

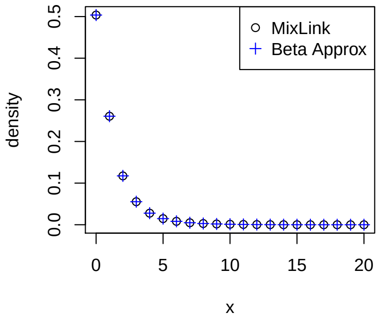


(b) $p = 0.50$

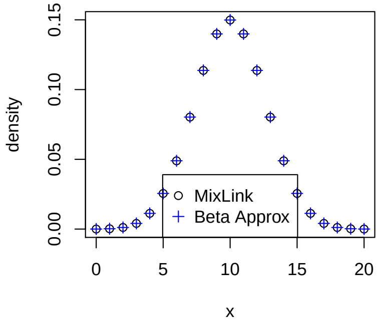
Figure: Comparison of exact Mixture Link density f and density g using beta approximation with $m = 20$ trials and $\boldsymbol{\pi} = (\frac{1}{20}, \frac{2}{20}, \frac{3}{20}, \frac{4}{20}, \frac{10}{20})$.

Beta Approximation to the Density

$\kappa = 2$



(a) $p = 0.05$



(b) $p = 0.50$

Figure: Comparison of exact Mixture Link density f and density g using beta approximation with $m = 20$ trials and $\boldsymbol{\pi} = (\frac{1}{20}, \frac{2}{20}, \frac{3}{20}, \frac{4}{20}, \frac{10}{20})$.