Mixture Link Models for Binomial Data with Overdispersion

Andrew M. Raim

Department of Mathematics and Statistics
University of Maryland, Baltimore County
Baltimore, MD, U.S.A.

U.S. Census Bureau
November 21, 2013

Joint work with Nagaraj K. Neerchal

Overview of the Talk

- Binomial data often exhibit **overdispersion** or **extra variation** relative to the standard Logistic regression model.

- We will present a new model to account for the extra variation in a general way. The idea is to link the regression to the marginal probability of success of a finite mixture of binomials, rather than a single binomial distribution.

- A goodness-of-fit study carried out on a real dataset shows promising results.

Hiroshima Example

An illustrative dataset is from (Morel and Neerchal, 2012), originally from Sofuni et al. (1978).

Chromosome abberations were studied in Hiroshima atomic bomb surivivors 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 abberations
- d_i: total radiation dose (T65-gamma + T65-neutron, in rads) received by the ith subject
- $z_i = \frac{d_i - \overline{d}}{\sqrt{\frac{1}{n}\sum_{\ell=1}^{n}(d_\ell - \overline{d})^2}}$: standardized radiation dose

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

Qn: What is the effect of radiation dose on the number of chromosome abberations?

Hiroshima Example

Logistic Regression

- A standard technique for this kind of analysis is logistic regression

$$T_i \stackrel{\text{ind}}{\sim} \text{Bin}(m_i, p_i), \quad p_i = G(\beta_1 x_{i1} + \cdots + \beta_d x_{id}) = G(\boldsymbol{x}_i^T \boldsymbol{\beta})$$

$$f(t_i \mid m_i, p_i) = \binom{m_i}{t_i} p_i^{t_i} (1 - p_i)^{m_i - t_i} \cdot I(t_i \in \{0, \dots, m_i\})$$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \binom{m_i}{t_i} G(\boldsymbol{x}_i^T \boldsymbol{\beta})^{t_i} \left[1 - G(\boldsymbol{x}_i^T \boldsymbol{\beta})\right]^{m_i - t_i}$$

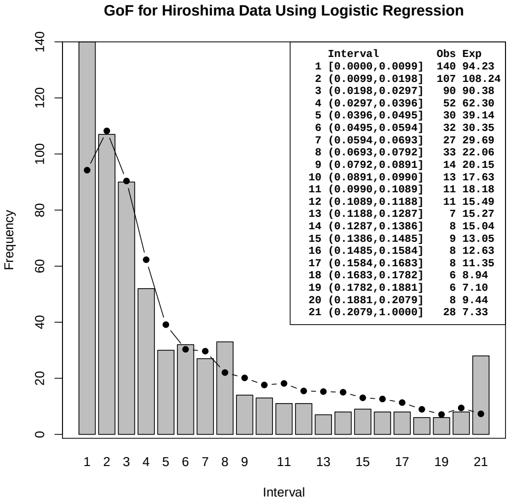
where $G(x) = 1/(1 + e^{-x})$.

- Estimation of $\boldsymbol{\beta}$, and ensuing inference, is usually carried out by (numerically) solving $\frac{\partial}{\partial \boldsymbol{\beta}} \log L(\boldsymbol{\beta}) = \mathbf{0}$.

- Logistic regression is a special case of the Generalized Linear Model (GLM) framework.

Hiroshima Example

Goodness of Fit



$$T_i \overset{\text{ind}}{\sim} \text{Bin}(m_i, p_i),$$

$$g(p_i) = \beta_0 + \beta_1 z_i + \beta_2 z_i^2,$$

$$i = 1, \dots 648$$

	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

Hiroshima Example

Problems with Model Fit

Which assumptions might be violated?

- Binomial implicitly assumes that lymphocytes within a subject are **independent**, and have the **same probability** of abberation.

- Other covariates beside radiation exposure might be important for determining abberations.

- Therefore, we suspect there may be **overdispersion** in this example — extra variation in the data which is not addressed by the model.

Handling Overdispersion in Logistic Regression

Quasi-Likelihood

- Introduce a dispersion parameter ϕ to inflate the variance $\text{Var}(T_i) = \phi m_i p_i (1 - p_i)$.
- Generalized Estimating Equations (GEE) considers ungrouped binary data $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{in_i})$ for the ith subject, and allows analyst to assume a correlation structure within.
- Both techniques make use of score-like equations which may not correspond to a true likelihood.

Handling Overdispersion in Logistic Regression

Binomial with Extra Variation

- Zero-inflated binomial (ZIB) selects zero w.p. ϕ and binomial w.p. $1-\phi$

$$\text{P}(T = t \mid m, p, \phi) = \phi I(t = 0) + (1-\phi)\text{Bin}(t \mid m, p).$$

  Similarly, any of $0, 1, \ldots, m$ may be selected to be inflated.

- Random-clumped binomial (RCB) (Morel and Nagaraj, 1993) considers the inflated value to be random. $T = NY + (X \mid N)$ with

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

- Beta-binomial (BB) assumes a hierarchy

$$T \mid \mu \sim \text{Bin}(m, \mu), \quad \mu \sim \text{Beta}(\alpha, \beta), \qquad p = \frac{\alpha}{\alpha+\beta},\ \phi = \frac{1}{\alpha+\beta+1},$$

  so that $\text{Var}(T) = mp(1-p)\{1 + \phi(m-1)\}$.

Handling Overdispersion in Logistic Regression

GLMMs and Finite Mixtures

- Generalized Linear Mixed Model (GLMM) adds flexibility of random effects. But computation may be difficult depending on the structure selected by the analyst.

- Logistic regression with a random intercept (e.g. Follmann and Lambert, 1989; Aitkin, 1996). Can use nonparametric maximum likelihood (NPMLE) to avoid assumptions about random effect distribution.

- Finite mixtures can model multiple latent subpopulation. E.g.

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

  Can be extended to a finite mixture of regressions (Frühwirth-Schnatter, 2006) by linking $\mu_j = G(\boldsymbol{x}^T \beta_j)$.

Mixture Link Model

- **Objective of this work:** Link a regression $G(\boldsymbol{x}^T \beta)$ from the outcome

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

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J) \in \mathcal{S}^J, \qquad \boldsymbol{\mu} = (\mu_1, \ldots, \mu_J) \in [0,1]^J,$$

to mixture success probability of a single trial $\boldsymbol{\mu}^T \boldsymbol{\pi} = \pi_1 \mu_1 + \cdots + \pi_J \mu_J$.

- $\mathcal{S}^J = \left\{ \boldsymbol{x} \in [0,1]^J : \sum_{j=1}^J x_j = 1 \right\}$ is the probability simplex in $\mathbb{R}^J$.

- The link $g(\boldsymbol{\mu}^T \boldsymbol{\pi}) = \boldsymbol{x}^T \beta$ leads us to consider the set

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

of all $\boldsymbol{\mu}$ satisfying the link. For regression, we take $p = G(\boldsymbol{x}^T \beta)$.

Optimization Approach

- Consider an independent sample

$$T_i \stackrel{\text{ind}}{\sim} \text{BinMix}(m_i, \boldsymbol{\mu}_i, \boldsymbol{\pi}), \quad \boldsymbol{\mu}_i \in A_i, \quad i = 1, \ldots, n,$$
$$\text{where } A_i = A(p_i, \boldsymbol{\pi}) \text{ and } p_i = G(\boldsymbol{x}_i^T \beta)$$

  and $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$ are fixed and unknown.

- MLE approach would suggest maximizing

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} \pi_j \binom{m_i}{t_i} \mu_{ij}^{t_i} (1 - \mu_{ij})^{m_i - t_i} \right\},$$
$$\text{subject to} \quad g(\boldsymbol{\mu}_i^T \boldsymbol{\pi}) = \boldsymbol{x}_i^T \beta, \text{ for } i = 1, \ldots, n.$$

- β enters optimization only through the constraint. This suggests a profile likelihood approach

$$Q(\beta) = \sup_{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n, \boldsymbol{\pi}} \left\{ \log L(\boldsymbol{\theta}) : g(\boldsymbol{\mu}_i^T \boldsymbol{\pi}) = \boldsymbol{x}_i^T \beta, \text{ for } i = 1, \ldots, n \right\}.$$

- But dimension of parameter space grows with n.

Random Effects Approach

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

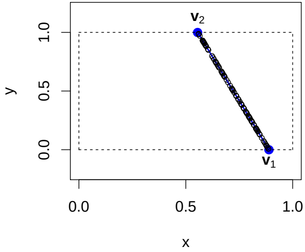
$$A_i = \text{conv}(\boldsymbol{v}_1^{(i)}, \ldots, \boldsymbol{v}_{k_i}^{(i)}) = \left\{ \sum_{\ell=1}^{k_i} \lambda_\ell \boldsymbol{v}_\ell^{(i)} : \boldsymbol{\lambda} \in \mathcal{S}^{k_i} \right\} = \left\{ \boldsymbol{V}^{(i)} \boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathcal{S}^{k_i} \right\}.$$

- $\boldsymbol{V}^{(i)}$ can vary for each observation when A_i depends on a covariate $\boldsymbol{x}_i$. The number of vertices k_i can also vary.
- We will consider $\boldsymbol{\mu}_i = \boldsymbol{V}^{(i)} \boldsymbol{\lambda}^{(i)} \in A_i$ with $\boldsymbol{\lambda}^{(i)} \stackrel{\text{ind}}{\sim} \text{Dirichlet}_{k_i}(\boldsymbol{\alpha})$

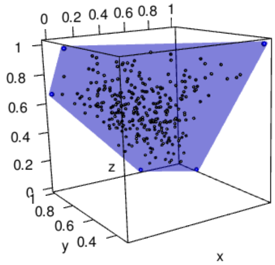
$$f(\boldsymbol{\lambda} \mid \boldsymbol{\alpha}) = \frac{\lambda_1^{\alpha_1 - 1} \cdots \lambda_k^{\alpha_k - 1}}{\text{B}(\boldsymbol{\alpha})} \cdot I(\boldsymbol{\lambda} \in \mathcal{S}^k), \quad \text{where } \text{B}(\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)}{\Gamma(\alpha_1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

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

Drawing a Sample on A



(a)

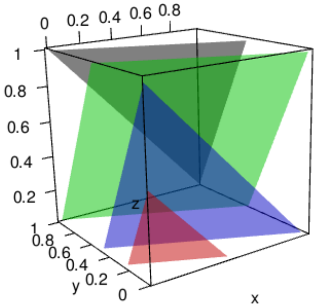


(b)

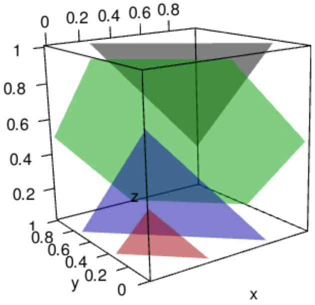
Figure: A sample drawn from A:
(a) $n = 100$ with $J = 2$, $\pi = (\frac{3}{4}, \frac{1}{4})$, $p = \frac{2}{3}$,
(b) $n = 300$ with $J = 3$, $\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, $p = \frac{2}{3}$.

(Number of vertices k depends on orientation of hyperplane)

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

Hierarchical Mixture Link Model

We can now write the model as

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

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

- k_i can vary between observations. E.g. p_i vary due to a regression.
- Difficult to maintain correspondence between $\boldsymbol{v}_\ell^{(i)}$ and $\alpha_\ell^{(i)}$.

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

Parameterized by: $\boldsymbol{\theta} = \begin{cases} (p, \boldsymbol{\pi}, \kappa) \in \mathbb{R}^{1+(J-1)+1}, & \text{if } T_i \text{ are iid, or} \\ (\boldsymbol{\beta}, \boldsymbol{\pi}, \kappa) \in \mathbb{R}^{d+(J-1)+1}, & \text{in the case of a regression} \end{cases}$

Symmetric Dirichlet Density

$k = 3$

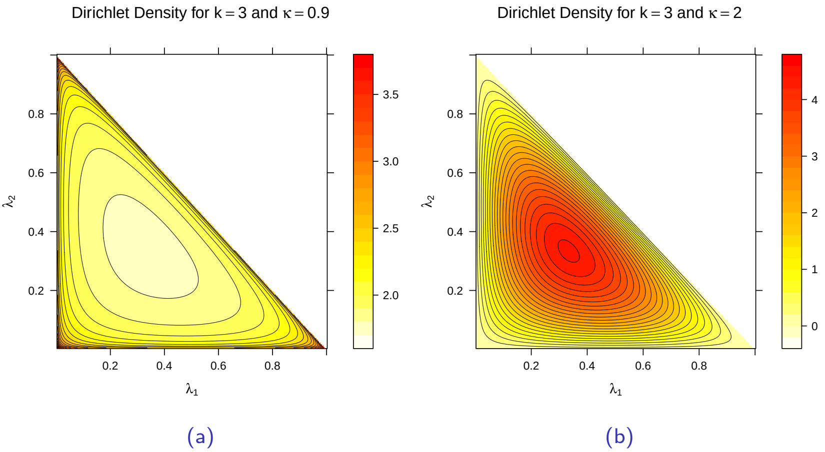


Figure: $\text{Dirichlet}_3(\boldsymbol{\lambda} \mid \kappa, \kappa, \kappa)$ density.

Moments of T

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

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

- Some moments of T can be obtained as

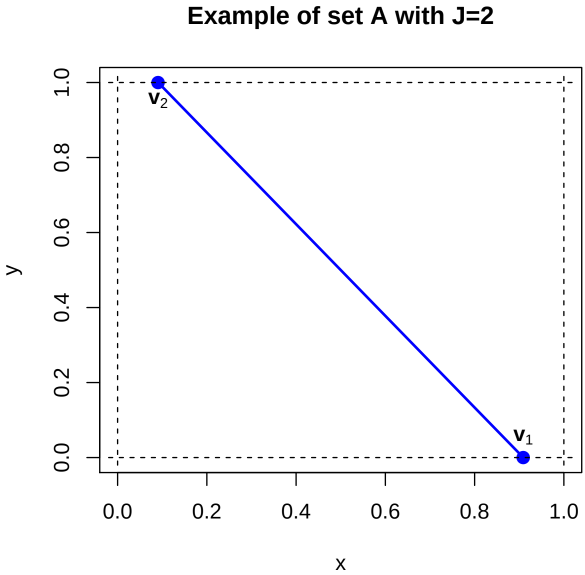
$$\text{E}(T) = m \sum_{j=1}^{J} \pi_j \bar{v}_{j\cdot}, \qquad \text{E}(T^2) = \text{E}(T) + m(m-1) \sum_{j=1}^{J} \pi_j \frac{\boldsymbol{v}_{j\cdot}^T \boldsymbol{v}_{j\cdot} + \kappa(k\bar{v}_{j\cdot})^2}{k(1+\kappa k)},$$

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

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

where $\boldsymbol{v}_{j\cdot}$ is the jth row of $\boldsymbol{V}$ and $\bar{v}_{j\cdot}$ is its mean.

Computing the Vertices of A



Computing the Vertices of A

Lemma

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

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

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

where $\pi_2 = 1 - \pi_1$.

Lemma (Characterization of Extreme Points of A)

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

Computing the Vertices of A

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

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

- 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 $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}$.
- Algorithm checks $J \cdot 2^{J-1}$ points, and therefore impractical for large J.

Computing the Density

Trivial case: $m = 1$

- The Bernoulli Mixture Link model simplifies in a trivial way

$$
\begin{aligned}
&\int \left\{ \sum_{j=1}^{J} \pi_j \mu_j^t (1 - \mu_j)^{1-t} \right\} f_A(\boldsymbol{\mu}) d\boldsymbol{\mu} \\
&= \begin{cases} \int \boldsymbol{\mu}^T \boldsymbol{\pi} f_A(\boldsymbol{\mu}) d\boldsymbol{\mu} = G(\boldsymbol{x}^T \beta), & \text{if } t = 1, \\ \int (1 - \boldsymbol{\mu}^T \boldsymbol{\pi}) f_A(\boldsymbol{\mu}) d\boldsymbol{\mu} = 1 - G(\boldsymbol{x}^T \beta) & \text{if } t = 0. \end{cases} \\
&= \left[G(\boldsymbol{x}^T \beta) \right]^t \left[1 - G(\boldsymbol{x}^T \beta) \right]^{1-t}
\end{aligned}
$$

- Here the mixture link model is equivalent to usual logistic regression.
- However, if $m \geq 2$ so that there is more than one Bernoulli trial, the two models no longer coincide.

Computing the Density

Simple case: $J = 2$, $\kappa = 1$

- Suppose there are two vertices, say $\boldsymbol{v}_1 = \begin{pmatrix} v_{11} \\ v_{12} \end{pmatrix}$ and $\boldsymbol{v}_2 = \begin{pmatrix} v_{21} \\ v_{22} \end{pmatrix}$.
- Then $A = \{\lambda \boldsymbol{v}_1 + (1-\lambda)\boldsymbol{v}_2 : \lambda \in [0,1]\}$ and $\boldsymbol{\mu} \in A$ can be written as

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \lambda v_{11} + (1-\lambda) v_{21} \\ \lambda v_{12} + (1-\lambda) v_{22} \end{pmatrix} \quad \text{where } \lambda \sim \mathrm{U}(0,1).$$

  We may then write the density as

$$\binom{m}{t} \sum_{j=1}^{2} \pi_j \int_0^1 [\lambda v_{1j} + (1-\lambda) v_{2j}]^t [1 - (\lambda v_{1j} + (1-\lambda) v_{2j})]^{m-t} d\lambda$$

  Let $w = \lambda v_{1j} + (1-\lambda) v_{2j}$, which gives

$$= \binom{m}{t} \sum_{j=1}^{2} \pi_j \int_{v_{2j}}^{v_{1j}} w^t (1-w)^{m-t} dw$$

$$= \binom{m}{t} \sum_{j=1}^{2} \pi_j \frac{\mathrm{B}_{v_{1j}}(t+1, m-t+1) - \mathrm{B}_{v_{2j}}(t+1, m-t+1)}{v_{1j} - v_{2j}}.$$

Computing the Density

Density of Linear Combination of Dirichlet

- The Mixture Link density depends on density of $\boldsymbol{v}_{j\cdot}^T \boldsymbol{\lambda}$. 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 $\boldsymbol{c}^T \boldsymbol{\lambda}$ as

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

Computing the Density

Density of Linear Combination of Dirichlet

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

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

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

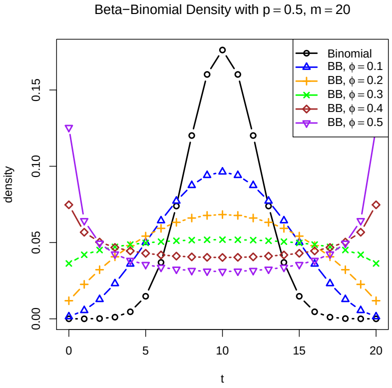
- Therefore the probability $\mathrm{P}\left(\boldsymbol{c}^T \boldsymbol{\lambda} \le x\right)$ can be computed by

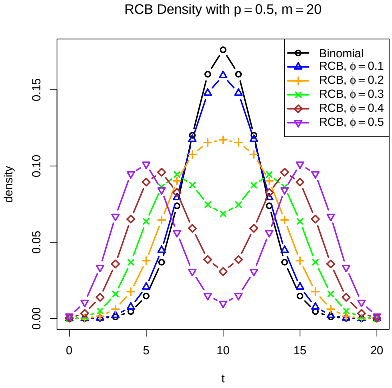
$$
F_{\boldsymbol{c}^T \boldsymbol{\lambda}}(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin\left(\sum_{j=1}^k \alpha_j \arctan\{(c_j - x)u\}\right)}{u \prod_{j=1}^k \left(1 + (c_j - x)^2 u^2\right)^{\alpha_j/2}} du
$$

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

RCB and BB Density

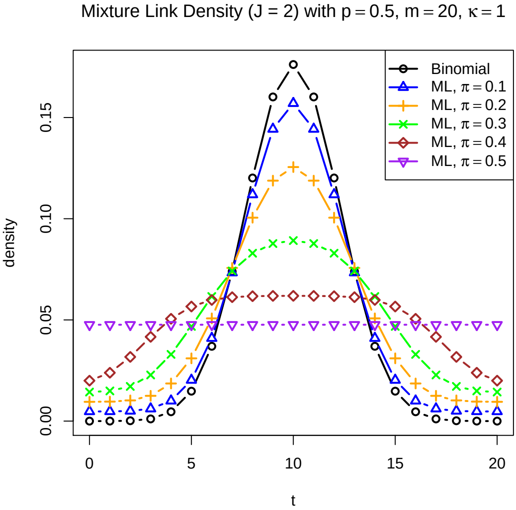
For Comparison





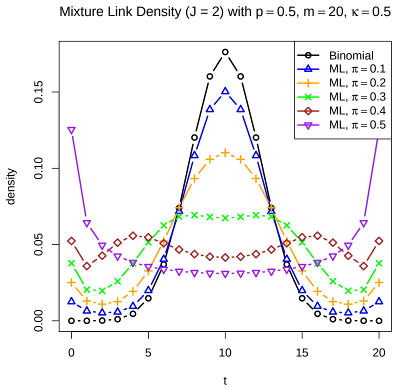
Mixture Link Density

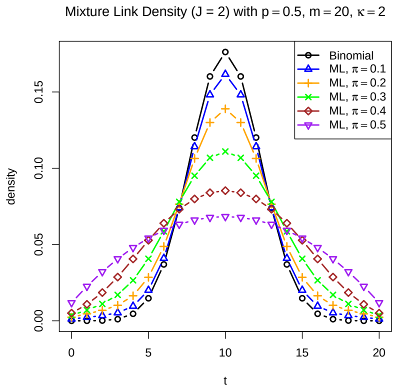
$\kappa = 1$



Mixture Link Density

$\kappa \in \{0.5, 2\}$





Hiroshima GOF

To compare models for goodness-of-fit on the Hiroshima dataset:

- 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 a variation on the usual Pearson chi-square GOF test statistic to allow varying m_i (Neerchal and Morel, 1998; Sutradhar et al., 2008).

This study is presented in (Raim and Neerchal, 2013).

Hiroshima GOF

GOF Test for Varying m_i

- To test a binomial model for GOF

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

- 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 $E_\ell(\boldsymbol{\theta}) \geq 5$.

Hiroshima GOF

GOF Test for Varying m_i

Sutradhar et al. (2008) shows 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 $r-1$.

Hiroshima GOF

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 are in parentheses.)

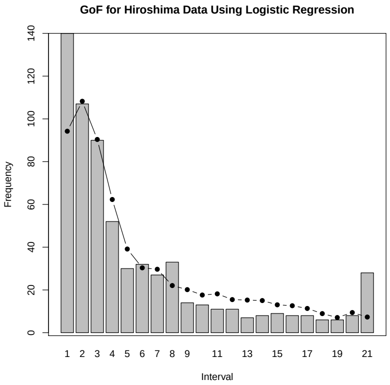
Hiroshima GOF

Model Comparison Statistics

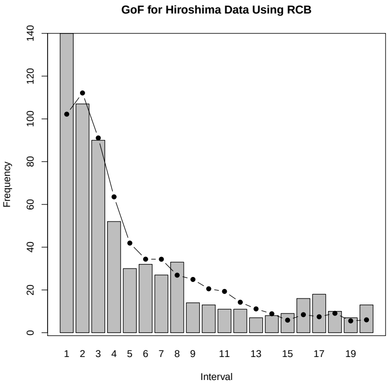
Model	LogLik	q	AIC	BIC	statistic	GOF df range	p-value
Logistic	-1814.19	3	3634.40	3647.80	110.38	[17,20]	$< 10^{-13}$
RCB	-1567.50	4	3143.00	3160.90	68.25	[15,19]	$< 10^{-6}$
BB	-1487.92	4	2983.85	3001.74	93.79	[12,18]	$< 10^{-11}$
RCB-Reg	-1546.61	6	3105.22	3132.07	63.96	[18,22]	$< 10^{-5}$
BB-Reg	-1429.61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

Hiroshima GOF

Observed vs. Expected Counts



(a) Logistic.

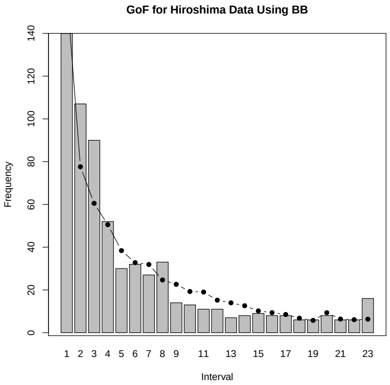


(b) RCB.

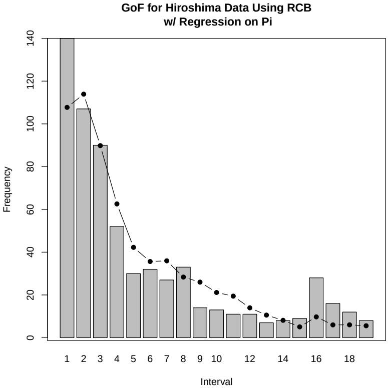
Figure: GOF plots for observed vs. expected counts. The grey bars represent the observed counts for a given interval, and the black dots are the expected counts under the MLE. Note that the choice of intervals varies between models.

Hiroshima GOF

Observed vs. Expected Counts



(a) BB.

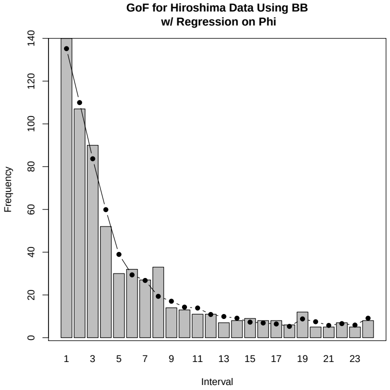


(b) RCB-Reg.

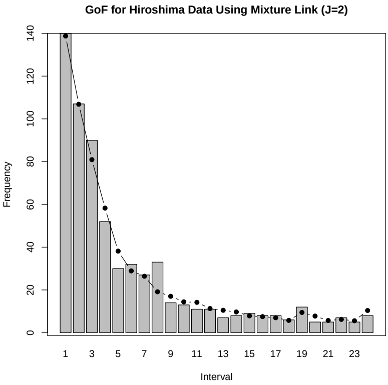
Figure: GOF plots for observed vs. expected counts. The grey bars represent the observed counts for a given interval, and the black dots are the expected counts under the MLE. Note that the choice of intervals varies between models.

Hiroshima GOF

Observed vs. Expected Counts



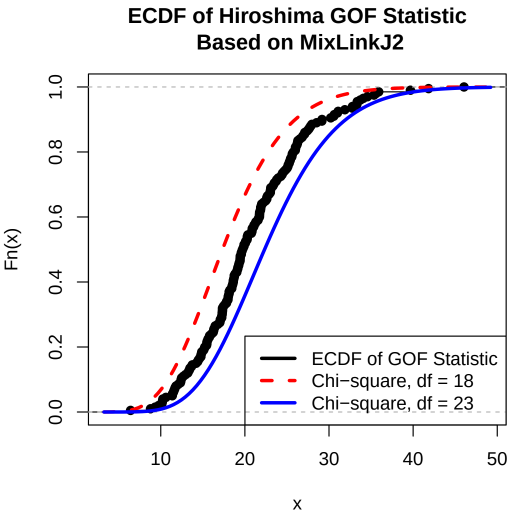
(a) BB-Reg.



(b) MixLinkJ2.

Figure: GOF plots for observed vs. expected counts. The grey bars represent the observed counts for a given interval, and the black dots are the expected counts under the MLE. Note that the choice of intervals varies between models.

Hiroshima GOF



Empirical CDF computed from 200 parametric bootstrap samples

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

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.
- Computational methods are given to evaluate the likelihood so that standard numerical techniques such as MLE and MCMC can be applied.
- Good fit for Hiroshima data compared to some standard binomial models for extra variation.

Future Work

- Effect of increasing J?
- Bayesian infererence? Other frequentist estimation methods besides numerical optimization?
- Estimator properties?
- Extend to other outcome types: Normal, Poisson, etc.

References I

Murray Aitkin. A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6:251–262, 1996.

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.

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.

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.

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

References II

J. G. Morel and N. K. Neerchal. *Overdispersion Models in SAS*. SAS Institute, 2012.

N. K. Neerchal and J. G. Morel. Large cluster results for two parametric multinomial extra variation models. *Journal of the American Statistical Association*, 93(443):1078–1087, 1998.

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.

Andrew M. Raim. *Computational Methods for Finite Mixtures using Approximate Information and Regression Linked to the Mixture Mean*. PhD thesis, University of Maryland, Baltimore County, 2013. (In progress).

Andrew M. Raim and Nagaraj K. Neerchal. Modeling overdispersion in binomial data with regression linked to a finite mixture probability of success. In *JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association*, 2013.

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.

References III

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.

Acknowledgement: The computational resources used for this work were provided by the High Performance Computing Facility at the University of Maryland, Baltimore County (UMBC). See `www.umbc.edu/hpcf` for information on the facility and its uses. Andrew additionally thanks the facility for financial support as an RA.