glms: a Transformative Paradigm for Statistical Practice and Education

John Hinde

Statistics Group,
School of Mathematics, Statistics and Applied Mathematics
National University of Ireland, Galway
john.hinde@nuigalway.ie

Research Supported by SFI Award 07/MI/012

Imperial College, London

28 March 2015

Summary

The Paper — 1972

The Paper — 1972

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

370

Generalized Linear Models

By J. A. Nelder and R. W. M. Wedderburn

Rothamsted Experimental Station, Harpenden, Herts

The Paper — 1972

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

370

Generalized Linear Models

By J. A. Nelder and R. W. M. Wedderburn

Rothamsted Experimental Station, Harpenden, Herts

- published in Series A

The Paper — 1972

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370 370

Generalized Linear Models

By J. A. Nelder and R. W. M. Wedderburn
Rothamsted Experimental Station, Harpenden, Herts

- published in Series A
- 15 pages long

The Paper — 1972

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

370

Generalized Linear Models

By J. A. Nelder and R. W. M. Wedderburn

Rothamsted Experimental Station, Harpenden, Herts

- published in Series A
- 15 pages long
- many examples — over half the paper

The Paper — 1972

J. R. Statist. Soc. A, 370
(1972), **135**, *Part* 3, p. 370

Generalized Linear Models

By J. A. NELDER and R. W. M. WEDDERBURN
Rothamsted Experimental Station, Harpenden, Herts

- published in Series A
- 15 pages long
- many examples — over half the paper
- *"useful way of unifying . . . unrelated statistical procedures"*

glms — the authors

John Nelder: 1924 — 2010

- Statistician at National Vegetable Research Station (NVRS), now Horticultural Research International, Wellesbourne — 1949-68
  - theory of general balance — unifying framework for the wide range of designs in agricultural experimentation
  - initial work on GenStat

glms — the authors

John Nelder: 1924 — 2010

- Statistician at National Vegetable Research Station (NVRS), now Horticultural Research International, Wellesbourne — 1949-68
  - theory of general balance — unifying framework for the wide range of designs in agricultural experimentation
  - initial work on GenStat
- Head of the Statistics Department at Rothamsted — 1968-1984
  - theory of generalized linear models, with the late Robert Wedderburn
  - Applied Statistics Algorithms in *Applied Statistics, JRSSC*
  - further development of GenStat, with NAG
  - development of GLIM, first released in 1974

glms — the authors

John Nelder: 1924 — 2010

- Statistician at National Vegetable Research Station (NVRS), now Horticultural Research International, Wellesbourne — 1949-68
  - theory of general balance — unifying framework for the wide range of designs in agricultural experimentation
  - initial work on GenStat
- Head of the Statistics Department at Rothamsted — 1968-1984
  - theory of generalized linear models, with the late Robert Wedderburn
  - Applied Statistics Algorithms in *Applied Statistics, JRSSC*
  - further development of GenStat, with NAG
  - development of GLIM, first released in 1974
- visiting Professor at Imperial College — 1972-2009
  - GLIMPSE "expert system" based on GLIM
  - theory of hierarchical generalized linear models (HGLMs), with Youngjo Lee

glms — the authors

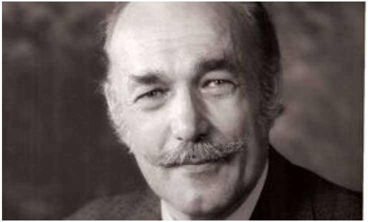
John Nelder: 1924 — 2010

- Statistician at National Vegetable Research Station (NVRS), now Horticultural Research International, Wellesbourne — 1949-68
    - theory of general balance — unifying framework for the wide range of designs in agricultural experimentation
    - initial work on GenStat
- Head of the Statistics Department at Rothamsted — 1968-1984
    - theory of generalized linear models, with the late Robert Wedderburn
    - Applied Statistics Algorithms in *Applied Statistics, JRSSC*
    - further development of GenStat, with NAG
    - development of GLIM, first released in 1974
- visiting Professor at Imperial College — 1972-2009
    - GLIMPSE "expert system" based on GLIM
    - theory of hierarchical generalized linear models (HGLMs), with Youngjo Lee

Robert Wedderburn: 1947 —1975

- Died aged 28 of anaphylactic shock from an insect bite.

John Nelder: 1924 — 2010



John at the Rothamsted Conference in 2004

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
  - Poisson count data: square-root transformation, $\sqrt{y}$
  - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
  - Exponential times: log transformation, $\log(y)$

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
    - Poisson count data: square-root transformation, $\sqrt{y}$
    - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
    - Exponential times: log transformation, $\log(y)$
- Probit analysis: Finney (1952) maximum likelihood for tolerance distribution in toxicology

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
  - Poisson count data: square-root transformation, $\sqrt{y}$
  - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
  - Exponential times: log transformation, $\log(y)$
- Probit analysis: Finney (1952) maximum likelihood for tolerance distribution in toxicology
- Dyke & Patterson (1952): logit model for analysis of proportions in factorial experiment

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
  - Poisson count data: square-root transformation, $\sqrt{y}$
  - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
  - Exponential times: log transformation, $\log(y)$
- Probit analysis: Finney (1952) maximum likelihood for tolerance distribution in toxicology
- Dyke & Patterson (1952): logit model for analysis of proportions in factorial experiment
- transformations to linearity

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
  - Poisson count data: square-root transformation, $\sqrt{y}$
  - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
  - Exponential times: log transformation, $\log(y)$
- Probit analysis: Finney (1952) maximum likelihood for tolerance distribution in toxicology
- Dyke & Patterson (1952): logit model for analysis of proportions in factorial experiment
- transformations to linearity
- Box-Cox transformation (1964)

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
  - Poisson count data: square-root transformation, $\sqrt{y}$
  - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
  - Exponential times: log transformation, $\log(y)$
- Probit analysis: Finney (1952) maximum likelihood for tolerance distribution in toxicology
- Dyke & Patterson (1952): logit model for analysis of proportions in factorial experiment
- transformations to linearity
- Box-Cox transformation (1964)
- Inverse polynomials, Nelder (1966)

glms — the background

- analysis of non-normal data — variance stabilising transformation of the response
  - Poisson count data: square-root transformation, $\sqrt{y}$
  - Binomial proportions: arc-sin-square-root, $\sin^{-1}(\sqrt{y})$
  - Exponential times: log transformation, $\log(y)$
- Probit analysis: Finney (1952) maximum likelihood for tolerance distribution in toxicology
- Dyke & Patterson (1952): logit model for analysis of proportions in factorial experiment
- transformations to linearity
- Box-Cox transformation (1964)
- Inverse polynomials, Nelder (1966)
- Nelder (1968): *. . . one transformation leads to a linear model and another to normal error.*

glms — the idea

glms — the idea

Gauss - You were one of the discussants of the Box-Cox 1964 paper and you also introduced the idea of inverse polinomials in 1966. How did you get the idea of Generalized Linear Models?

Nelder - That's an interesting question. I don't really think I know the answer to it. There is a paper that I wrote in 1970 which was published in Biometrics; in this I drew attention to the fact that there was a considerable similarity between a model with gamma errors and an inverse linear response curve and the model for Probit Analysis. I didn't understand at that time exactly what the connection was, though I could see there was one. Then in the subsequent two years somehow the idea jelled, so that Wedderburn and I could see what was common to these models. That's how it came about, but exactly how I did it I don't know. Similarly in the General Balance papers I first had the idea in a

glm Paper: contents

- Intro: background (**2 pages**)

glm Paper: contents

- Intro: background (**2 pages**)
  - random component: 1-parameter exponential family
  - linear predictor: $\eta = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$
  - link function: $g(\mu) = \eta$

glm Paper: contents

- Intro: background (**2 pages**)
  - random component: 1-parameter exponential family
  - linear predictor: $\eta = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$
  - link function: $g(\mu) = \eta$
- Model fitting: (**3 pages**)
  - maximum likelihood estimation using Fisher Scoring

Iteratively (Re)-Weighted Least Squares

glm Paper: contents

- Intro: background (**2 pages**)
  - random component: 1-parameter exponential family
  - linear predictor: $\eta = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$
  - link function: $g(\mu) = \eta$
- Model fitting: (**3 pages**)
  - maximum likelihood estimation using Fisher Scoring

    Iteratively (Re)-Weighted Least Squares

  - sufficient statistics — canonical links

glm Paper: contents

- Intro: background (**2 pages**)
  - random component: 1-parameter exponential family
  - linear predictor: $\eta = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$
  - link function: $g(\mu) = \eta$
- Model fitting: (**3 pages**)
  - maximum likelihood estimation using Fisher Scoring

    Iteratively (Re)-Weighted Least Squares

  - sufficient statistics — canonical links
  - Analysis of Deviance

    minimal $\leftrightarrow$ complete (saturated) models

glm Paper: contents

- Intro: background (**2 pages**)
  - random component: 1-parameter exponential family
  - linear predictor: $\eta = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$
  - link function: $g(\mu) = \eta$
- Model fitting: (**3 pages**)
  - maximum likelihood estimation using Fisher Scoring

    Iteratively (Re)-Weighted Least Squares

  - sufficient statistics — canonical links
  - Analysis of Deviance

    minimal $\leftrightarrow$ complete (saturated) models

- Special distributions, examples (**6 pages**)

glm Paper: contents

- Intro: background (**2 pages**)
  - random component: 1-parameter exponential family
  - linear predictor: $\eta = \beta_0 + \beta_1 x_1 + \cdots \beta_p x_p$
  - link function: $g(\mu) = \eta$
- Model fitting: (**3 pages**)
  - maximum likelihood estimation using Fisher Scoring

    Iteratively (Re)-Weighted Least Squares
  - sufficient statistics — canonical links
  - Analysis of Deviance

    minimal $\leftrightarrow$ complete (saturated) models
- Special distributions, examples (**6 pages**)
- Models in Teaching Statistics (**1 page**)

glm Paper: examples

- Normal: observations normal on log-scale; additive effects on inverse scale

glm Paper: examples

- Normal: observations normal on log-scale; additive effects on inverse scale
- Poisson: Fisher's tuberculin-test data — Latin square of counts

glm Paper: examples

- Normal: observations normal on log-scale; additive effects on inverse scale
- Poisson: Fisher's tuberculin-test data — Latin square of counts
- Poisson: multinomial distributions for contingency tables

glm Paper: examples

- Normal: observations normal on log-scale; additive effects on inverse scale
- Poisson: Fisher's tuberculin-test data — Latin square of counts
- Poisson: multinomial distributions for contingency tables
- Binomial: Probit & Logit models

glm Paper: examples

- Normal: observations normal on log-scale; additive effects on inverse scale
- Poisson: Fisher's tuberculin-test data — Latin square of counts
- Poisson: multinomial distributions for contingency tables
- Binomial: Probit & Logit models
- Gamma: estimation of variance components in incomplete block design

John Nelder & Statistical Computing

- Anti black-box packages

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal
- System should not allow stupid models — marginality

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal
- System should not allow stupid models — marginality
- Model specification using Wilkinson & Rogers formulæ

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal
- System should not allow stupid models — marginality
- Model specification using Wilkinson & Rogers formulæ
- All structures available to the user — input to other routines

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal
- System should not allow stupid models — marginality
- Model specification using Wilkinson & Rogers formulæ
- All structures available to the user — input to other routines
- system should be open — user extendible (GLIM, GenStat, S/R, . . .)

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal
- System should not allow stupid models — marginality
- Model specification using Wilkinson & Rogers formulæ
- All structures available to the user — input to other routines
- system should be open — user extendible (GLIM, GenStat, S/R, . . .)
- Requires user expertise/knowledge

John Nelder & Statistical Computing

- Anti black-box packages
- User should be in control
- Default output should be minimal
- System should not allow stupid models — marginality
- Model specification using Wilkinson & Rogers formulæ
- All structures available to the user — input to other routines
- system should be open — user extendible (GLIM, GenStat, S/R, . . .)
- Requires user expertise/knowledge

Principles embodied in GLIM

— a system specifically for fitting glms.

GLIM: Interactive package (A Fistful of $'s!!)

GLIM: Interactive package (A Fistful of $'s!!)

```
[i] ? $yvar days $error p $
[i] ? $fit A*S*C*L $
[o] scaled deviance =  1173.9 at cycle 4
[o]     residual df =   118
```

GLIM: Interactive package (A Fistful of $'s!!)

```
[i] ? $yvar days $error p $
[i] ? $fit A*S*C*L $
[o] scaled deviance =  1173.9 at cycle 4
[o]      residual df =   118
```

Or, in John's preferred style . . .

GLIM: Interactive package (A Fistful of $'s!!)

```
[i] ? $yvar days $error p $
[i] ? $fit A*S*C*L $
[o] scaled deviance =  1173.9 at cycle 4
[o]     residual df =   118
```

Or, in John's preferred style . . .

```
[i] ? $y days $e p $
[i] ? $f A*S*C*L $
```

Dissemination of glms

- Conferences — *"That's a glm!"*

Dissemination of glms

- Conferences — *"That's a glm!"*
- Nelder (1984) Models for Rates with Poisson Errors: *In a recent paper, Frome (1983) described the fitting of models with Poisson errors and data in the form of rates . . . fitted simply by GLIM . . . or the use of a program that handles iterative weighted least squares*

Dissemination of glms

- Conferences — *"That's a glm!"*
- Nelder (1984) Models for Rates with Poisson Errors: *In a recent paper, Frome (1983) described the fitting of models with Poisson errors and data in the form of rates . . . fitted simply by GLIM . . . or the use of a program that handles iterative weighted least squares*
- Nelder (1991) Generalized Linear Models for Enzyme-Kinetic Data: *Ruppert, Cressie, and Carroll (1989) discuss various models for fitting the Michaelis-Menten equations to data on enzyme kinetics. I find it surprising that they do not include, among the models they consider, generalized linear models (GLMs) with an inverse link*

Dissemination of glms

- Conferences — *"That's a glm!"*
- Nelder (1984) Models for Rates with Poisson Errors: *In a recent paper, Frome (1983) described the fitting of models with Poisson errors and data in the form of rates . . . fitted simply by GLIM . . . or the use of a program that handles iterative weighted least squares*
- Nelder (1991) Generalized Linear Models for Enzyme-Kinetic Data: *Ruppert, Cressie, and Carroll (1989) discuss various models for fitting the Michaelis-Menten equations to data on enzyme kinetics. I find it surprising that they do not include, among the models they consider, generalized linear models (GLMs) with an inverse link*
  The data-transformation approach suffers from the disadvantage that normality of errors and linearity of systematic effects are still being sought simultaneously

Generalized Linear Models — Monograph

The International Statistical Institute

HOME | ISI CONGRESS | MEMBERSHIP | ISI ASSOCIATIONS | ISI COMMITTEES | SPECIAL TOPICS | STATISTICAL SOCIETIES

Login

ABOUT ISI
PUBLICATIONS
GLOSSARY
PAYMENTS
INFO SERVICE
LATEST NEWS
FAQ
SITEMAP

Follow us

Font Size Changer

STATISTICS2013.ORG
INTERNATIONAL YEAR OF
STATISTICS
PARTICIPATING ORGANIZATION

International Statistical Institute (ISI)

2013 Karl Pearson Prize

The ISI's Karl Pearson Prize was established in 2013 to recognize a contemporary a research contribution that has had profound influence on statistical theory, methodology, practice, or applications. The contribution can be a research article or a book and must be published within the last three decades. The prize is sponsored by Elsevier B.V.

The inaugural Karl Pearson Prize is awarded to Peter McCullagh and John Nelder[1] for their monograph *Generalized Linear Models* (1983).

This book has changed forever teaching, research and practice in statistics. It provides a unified and self-contained treatment of linear models for analyzing continuous, binary, count, categorical, survival, and other types of data, and illustrates the methods on applications from different areas. The monograph is based on several groundbreaking papers, including "Generalized linear models," by Nelder and Wedderburn, JRSS-A (1972), "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," by Wedderburn, Biometrika (1974), and "Regression models for ordinal data," by P. McCullagh, JRSS-B (1980). The implementation of GLM was greatly facilitated by the development of GLIM, the interactive statistical package, by Baker and Nelder. In his review of the GLIM3 release and its manual in JASA 1979 (pp. 934-5), Peter McCullagh wrote that "It is surprising that such a powerful and unifying tool should not have achieved greater popularity after six or more years of existence." The collaboration between McCullagh and Nelder has certainly remedied this issue and has resulted in a superb treatment of the subject that is accessible to researchers, graduate students, and practitioners.

The prize will be presented on August 27, 2013 at the ISI World Statistics Congress in Hong Kong and will be followed by the Karl Pearson Lecture by Peter McCullagh.

Karl Pearson Lecture: Statistical issues in modern scientific research

Peter McCullagh
University of Chicago, USA

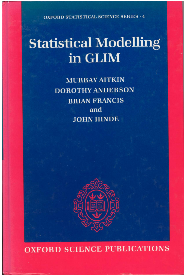
Statistical Modelling in GLIM (1989)

An applied *how to* text with integrated GLIM code.

Statistical Modelling in GLIM (1989)

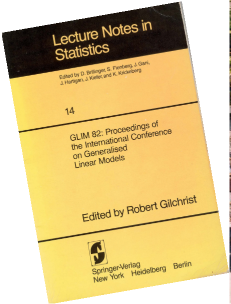
An applied *how to* text with integrated GLIM code.

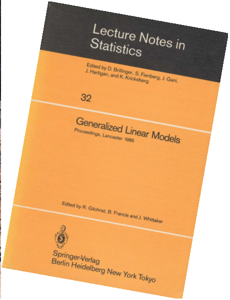
- normal models
  - regression
  - analysis of variance
- binomial responses
- multinomial and Poisson
  - count data
  - multiway tables
- survival models
  - parametric
  - Cox PH — piecewise exponential
  - discrete time



GLIM Conferences, IWSM, Statistical Modelling

- GLIM conferences — really on glms
- **IWSM**: **I**nternational **W**orkshop on **S**tatistical **M**odelling
- Eventually led to **Statistical Modelling Society**





Statistical Modelling Journal

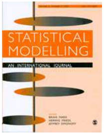
In 2000, founding of journal **Statistical Modelling**

availability of data and code with papers → *reproducible research*

Statistical Modelling: An International Journal http://stat.uibk.ac.at/SMIJ/

SMij

Statistical Modelling: An International Journal



Aims and Scope

Editorial Board

For Authors

Archives

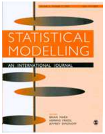
Modelling Society

STATISTICAL MODELLING

AN INTERNATIONAL JOURNAL

from

SAGE Publications



Statistical Modelling: An International Journal publishes original and high-quality articles that recognize statistical modelling as the general framework for the application of statistical ideas. Submissions must reflect important developments, extensions, and applications in statistical modelling. The journal also encourages submissions that describe scientifically interesting, complex or novel statistical modelling aspects from a wide diversity of disciplines, and submissions that embrace the diversity of applied statistical modelling.

Indexed by Science Citation Index Expanded, ISI Alerting Services, and CompuMath Citation Index, beginning with volume 3 (2003).

Extending the basic glm

- **response distribution**
  - multivariate vector of responses
  - exponential dispersion models
  - generalized distributions
  - quasi-distributions
  - mixtures
  - joint responses: *longitudinal* + *time to event*, ...

Extending the basic glm

- **response distribution**
  - multivariate vector of responses
  - exponential dispersion models
  - generalized distributions
  - quasi-distributions
  - mixtures
  - joint responses: *longitudinal* + *time to event*, ...
- **linear predictor**
  - smooth terms — gams, etc
  - random effects
  - multiple linear predictors — *modelling mean and dispersion*, gamlss, etc

Extending the basic glm

- **response distribution**
  - multivariate vector of responses
  - exponential dispersion models
  - generalized distributions
  - quasi-distributions
  - mixtures
  - joint responses: *longitudinal* + *time to event*, ...
- **linear predictor**
  - smooth terms — gams, etc
  - random effects
  - multiple linear predictors — *modelling mean and dispersion*, gamlss, etc
- **link function**
  - parametric links
  - composite link functions — *(Thompson & Baker, 1981)*
  - non-linear glms — `gnm` *(Turner & Firth, 2012)*

Normal Models

$$\mathbf{y} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\epsilon}$$

- single error term includes
  - individual observation/measurement error
  - *experimental* unit variability
  - unobserved covariates

Normal Models

$$\mathbf{y} = \boldsymbol{\beta}^T \mathbf{x} + \epsilon$$

- single error term includes
  - individual observation/measurement error
  - *experimental* unit variability
  - unobserved covariates
- for simplest data structures/designs use **normal linear model**

Normal Models

$$\mathbf{y} = \beta^T \mathbf{x} + \epsilon$$

- single error term includes
  - individual observation/measurement error
  - *experimental* unit variability
  - unobserved covariates
- for simplest data structures/designs use **normal linear model**
- more complex situations
  - structure in *experimental* unit variability
  - repeated measures/longitudinal observations
  - ...

Normal Mixed Model

$$\mathbf{y} = \beta^T \mathbf{x} + \gamma^T \mathbf{z} + \epsilon$$

- $\mathbf{z}$ unobserved **random effects**

Normal Mixed Model

$$\mathbf{y} = \beta^T \mathbf{x} + \gamma^T \mathbf{z} + \epsilon$$

- $\mathbf{z}$ unobserved **random effects**
- shared random effects
  - multi-level/variance components models
  - longitudinal observations
  - spatial structure

Normal Mixed Model

$$\mathbf{y} = \beta^T \mathbf{x} + \gamma^T \mathbf{z} + \epsilon$$

- $\mathbf{z}$ unobserved **random effects**
- shared random effects
  - multi-level/variance components models
  - longitudinal observations
  - spatial structure
- $\mathbf{z}$ normal
  - normal model with **structured covariance matrix**
- standard mixed model analyses – ML, REML
- widely available in standard software

Generalized Linear Models

Models for counts, proportions, times, ...

$$\mathbf{y} \sim F(\boldsymbol{\mu}) \qquad g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x}$$

- distributional assumption relates to the observation/measurement process
- how does this model incorporate
  - experimental/individual unit variability?
  - unobserved covariates?

Generalized Linear Models

Models for counts, proportions, times, ...

$$\mathbf{y} \sim F(\boldsymbol{\mu}) \qquad g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x}$$

- distributional assumption relates to the observation/measurement process
- how does this model incorporate
  - experimental/individual unit variability?
  - unobserved covariates?

It doesn't!

Generalized Linear Models

Models for counts, proportions, times, ...

$$\mathbf{y} \sim F(\boldsymbol{\mu}) \qquad g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x}$$

- distributional assumption relates to the observation/measurement process
- how does this model incorporate
  - experimental/individual unit variability?
  - unobserved covariates?

It doesn't!

hence overdispersion, etc

Random Effect Models

Include random effect(s) in the linear predictor

$$\boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z}$$

Random Effect Models

Include random effect(s) in the linear predictor

$$\boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z}$$

- single conjugate random effect at individual level – standard overdispersion models
  - negative binomial for count data
  - beta-binomial for proportions

Random Effect Models

Include random effect(s) in the linear predictor

$$\boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z}$$

- single conjugate random effect at individual level – standard overdispersion models
  - negative binomial for count data
  - beta-binomial for proportions
- $\mathbf{z}$ normal $\longrightarrow$ **generalized linear mixed models**

Random Effect Models

Include random effect(s) in the linear predictor

$$\boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x} + \boldsymbol{\gamma}^T \mathbf{z}$$

- single conjugate random effect at individual level – standard overdispersion models
  - negative binomial for count data
  - beta-binomial for proportions
- $\mathbf{z}$ normal $\longrightarrow$ **generalized linear mixed models**
- $\mathbf{z}$ unspecified $\longrightarrow$ **nonparametric maximum likelihood**

John's Approach (1984)

Additional random term in linear predictor

Suppose that the linear predictor η should be modelled by $\eta + \varepsilon$, where ε is a random variable with variance θ. θ is the between-group or 'higher level' component of variance. Taking expectations over the lower level error we have

$$E_1(y) = \mu = g(\eta + \varepsilon) \approx g(\eta) + g'(\eta)\,\varepsilon$$

Hence assuming independence between the two error components we have

$$\text{var}(y) \approx \text{var}_1(y) + g'^2\,\theta$$

$$\approx V(\mu) + \theta g'^2 \quad \text{for a GLM.}$$

In particular for a canonical link where $g' = V$, we have

$$\text{var}(y) \approx V(\mu) + \theta\, V^2(\mu) \qquad (1)$$

Motivating Application

- 4x2 factorial micropropagation experiment of the apple variety Trajan – a 'columnar' variety.
- Shoot tips of length 1.0-1.5 cm were placed in jars on a standard culture medium.
- 4 concentrations of cytokinin BAP added

  High concentrations of BAP often inhibit root formation during micropropagation of apples, but maybe not for 'columnar' varieties.
- Two growth cabinets, one with 8 hour photoperiod, the other with 16 hour.

  Jars placed at random in one of the two cabinets

Response variable: number of roots after 4 weeks culture at 22°C.

Motivating Application: Data

	Photoperiod							
	8				16			
BAP (μM)	2.2	4.4	8.8	17.6	2.2	4.4	8.8	17.6
No. of roots								
0	0	0	0	2	**15**	**16**	**12**	**19**
1	3	0	0	0	0	2	3	2
2	2	3	1	0	2	1	2	2
3	3	0	2	2	2	1	1	4
4	6	1	4	2	1	2	2	3
5	3	0	4	5	2	1	2	1
6	2	3	4	5	1	2	3	4
7	2	7	4	4	0	0	1	3
8	3	3	7	8	1	1	0	0
9	1	5	5	3	3	0	2	2
10	2	3	4	4	1	3	0	0
11	1	4	1	4	1	0	1	0
12	0	0	2	0	1	1	1	0
>12	13,17	13	14,14	14				
No. of shoots	30	30	40	40	30	30	30	40
Mean	5.8	7.8	7.5	7.2	3.3	2.7	3.1	2.5
Variance	14.1	7.6	8.5	8.8	16.6	14.8	13.5	8.5
Overdispersion index	1.42	-0.03	0.13	0.22	**4.06**	**4.40**	**3.31**	**2.47**

Dispersion

Second factorial cumulant

$$S(X) = \text{Var}(X) - \text{E}[X]$$

Useful summary:

- underdispersion: $-\text{E}[X] \leq S(X) < 0$
- equidispersion (Poisson): $S(X) = 0$
- overdispersion: $S(X) > 0$

Dispersion

Second factorial cumulant

$$S(X) = \text{Var}(X) - \text{E}[X]$$

Useful summary:

- underdispersion: $-\text{E}[X] \leq S(X) < 0$
- equidispersion (Poisson): $S(X) = 0$
- overdispersion: $S(X) > 0$

Fisher's dispersion index

$$D(X) = \frac{\text{Var}(X)}{\text{E}[X]} = 1 + \frac{S(X)}{\text{E}[X]}$$

Standard Models

Poisson (Po)

$$\mathrm{Var}(X) = \mu \qquad S(X) = 0$$

Standard Models

Poisson (Po)

$$\mathrm{Var}(X) = \boldsymbol{\mu} \qquad S(X) = 0$$

Negative binomial (NB2): Poisson-Gamma mixture

$$\mathrm{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^2 \qquad S(X) = \gamma\boldsymbol{\mu}^2$$

Note: Poisson-lognormal mixture has same variance function

Standard Models

Poisson (Po)

$$\text{Var}(X) = \boldsymbol{\mu} \qquad S(X) = 0$$

Negative binomial (NB2): Poisson-Gamma mixture

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^2 \qquad S(X) = \gamma\boldsymbol{\mu}^2$$

Note: Poisson-lognormal mixture has same variance function

Negative binomial (NB1): alternative Poisson-Gamma mixture

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu} = \phi\boldsymbol{\mu} \qquad S(X) = \gamma\boldsymbol{\mu}$$

same variance function as a **quasi-Poisson** model

Standard Models

Poisson (Po)

$$\text{Var}(X) = \boldsymbol{\mu} \qquad S(X) = 0$$

Negative binomial (NB2): Poisson-Gamma mixture

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^2 \qquad S(X) = \gamma\boldsymbol{\mu}^2$$

Note: Poisson-lognormal mixture has same variance function

Negative binomial (NB1): alternative Poisson-Gamma mixture

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu} = \phi\boldsymbol{\mu} \qquad S(X) = \gamma\boldsymbol{\mu}$$

same variance function as a **quasi-Poisson** model

Poisson-inverse Gaussian

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^3 \qquad S(X) = \gamma\boldsymbol{\mu}^3$$

Extended variance function

An natural generalization is

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^p \qquad S(X) = \gamma\boldsymbol{\mu}^p$$

for some general power p.

Extended variance function

An natural generalization is

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^p \qquad S(X) = \gamma\boldsymbol{\mu}^p$$

for some general power p.

Suggested by Hinde & Demétrio (1998) and Nelder (??).

Extended variance function

An natural generalization is

$$\text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^p \qquad S(X) = \gamma\boldsymbol{\mu}^p$$

for some general power p.

Suggested by Hinde & Demétrio (1998) and Nelder (??).

Class of Poisson mixtures, **Poisson-Tweedie models** $PT_p(\boldsymbol{\mu}, \gamma)$

$$Z \sim Tw_p(\boldsymbol{\mu}, \gamma), \quad X|Z \sim Po(Z) \Rightarrow X \sim PT_p(\boldsymbol{\mu}, \gamma)$$

has moments

$$\text{E}[X] = \text{E}[Z] = \boldsymbol{\mu} \qquad \text{Var}(Z) = \gamma\boldsymbol{\mu}^p \qquad \text{Var}(X) = \boldsymbol{\mu} + \gamma\boldsymbol{\mu}^p$$

Tweedie Models

Family	E[Z]	Var(Z)	Type	Support
Normal	μ	γ	Continuous	R
Poisson	μ	μ	Discrete	N_0
Non-central gamma	μ	$\gamma\mu^{3/2}$	Cont. + atom	R_0
Gamma	μ	$\gamma\mu^2$	Continuous	R_+
Inverse Gauss	μ	$\gamma\mu^3$	Continuous	R_+

Only Poisson distribution is discrete.

Poisson-Tweedie Models

Family	E[X]	$S(X)$	Disp. Type	$ZI(X)$
Poisson	μ	0	Equi	0
Hermite	μ	γ	Over	+
Neyman Type A (*Poisson-Poisson*)	μ	$\gamma\mu$	Over	+
Pólya-Aeppli Type A (*Poisson-compound Poisson*)	μ	$\gamma\mu^{3/2}$	Over	+
Negative binomial	μ	$\gamma\mu^2$	Over	+
Binomial	μ	$-\gamma\mu^2$	Under	+
Poisson-Inv. Gauss	μ	$\gamma\mu^3$	Over	−

Motivating Application: Data

	Photoperiod							
	8				16			
BAP (μM)	2.2	4.4	8.8	17.6	2.2	4.4	8.8	17.6
No. of roots								
0	0	0	0	2	**15**	**16**	**12**	**19**
1	3	0	0	0	0	2	3	2
2	2	3	1	0	2	1	2	2
3	3	0	2	2	2	1	1	4
4	6	1	4	2	1	2	2	3
5	3	0	4	5	2	1	2	1
6	2	3	4	5	1	2	3	4
7	2	7	4	4	0	0	1	3
8	3	3	7	8	1	1	0	0
9	1	5	5	3	3	0	2	2
10	2	3	4	4	1	3	0	0
11	1	4	1	4	1	0	1	0
12	0	0	2	0	1	1	1	0
>12	13,17	13	14,14	14				
No. of shoots	30	30	40	40	30	30	30	40
Mean	5.8	7.8	7.5	7.2	3.3	2.7	3.1	2.5
Variance	14.1	7.6	8.5	8.8	16.6	14.8	13.5	8.5
Overdispersion index	1.42	-0.03	0.13	0.22	**4.06**	**4.40**	**3.31**	**2.47**

Zero-inflated models

If Y_i has a *zero-inflated Poisson* (ZIP) distribution, given by

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\lambda_i} & y_i = 0 \\ (1-\omega_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & y_i > 0 \end{cases}$$

Zero-inflated models

If Y_i has a *zero-inflated Poisson* (ZIP) distribution, given by

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\lambda_i} & y_i = 0 \\ (1-\omega_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & y_i > 0 \end{cases}$$

Lambert (1992) considered models in which

$$\log(\lambda_i) = \mathbf{x}_i^T\boldsymbol{\beta} \quad \text{and} \quad \log\left(\frac{\omega_i}{1-\omega_i}\right) = \mathbf{z}_i^T\boldsymbol{\gamma}$$

where **x** and **z** are covariate vectors and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of parameters.

Zero-inflated models

If Y_i has a *zero-inflated Poisson* (ZIP) distribution, given by

$$\Pr(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\lambda_i} & y_i = 0 \\ (1-\omega_i)\dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} & y_i > 0 \end{cases}$$

Lambert (1992) considered models in which

$$\log(\lambda_i) = \mathbf{x}_i^T\beta \quad \text{and} \quad \log\left(\frac{\omega_i}{1-\omega_i}\right) = \mathbf{z}_i^T\gamma$$

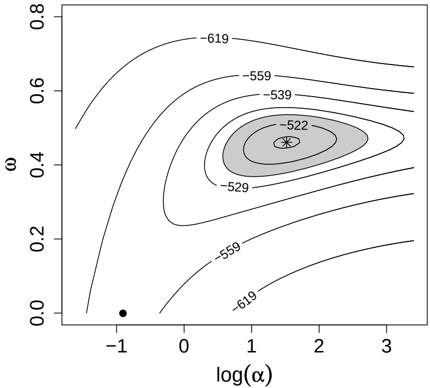
where **x** and **z** are covariate vectors and β and γ are vectors of parameters.

Similar mixture models are available for the negative binomial distribution (ZINB), etc.

Trajan apple cultivation data: fitted frequencies

No. of Roots	Observed	Fitted frequencies Poisson	Neg-bin	ZIP	ZINB	ZIGPD
0	62	7.4	55.8	62	62	62
1	7	21.3	19.8	1.6	5.1	4.8
2	7	30.4	12.2	4.4	7.6	7.6
3	8	29	8.6	7.9	8.9	9.1
4	8	20.8	6.4	10.8	9.1	9.3
5	6	11.9	4.9	11.8	8.4	8.5
6	10	5.7	3.9	10.7	7.2	7.2
7	4	2.3	3.1	8.3	5.8	5.8
8	2	0.8	2.5	5.7	4.5	4.5
9	7	0.3	2.1	3.4	3.4	3.4
10	4	0.1	1.7	1.9	2.5	2.4
11	2	0	1.4	0.9	1.8	1.7
≥ 12	3	0	5.8	0.7	3.6	3.7
$-2 \times$ log-lik		840.7	550.2	537.9	519.3	519.8
G^2		335.5	36.9	31.2	9.1	9.4

Trajan apple cultivation data: ZINB



Contour plot of 2×log-likelihood for α and ω with μ fixed at the sample mean: maximum likelihood estimates for ZINB ($*$) and negative binomial models ($\bullet$).

Trajan Apples: model fitting results

P	is a two level factor for photoperiod
H	is a four level factor for the BAP levels
Lin(H)	is a linear trend over the levels of H
	(on the log-concentration scale for BAP.)

	Models						
Description	λ	ω	α	$-2\log L$	df	AIC	BIC
Poisson	H*P	0	0	1556.9	262	1572.9	1601.7
	P	0	0	1571.9	268	1575.9	1583.1
Neg-Bin	H*P	0	const	1399.6	261	1417.6	1450.0
	H*P	0	P	1264.6	260	1284.6	1320.6
	H*P	0	H*P	1254.8	254	1286.8	1344.4
	Lin(H)*P	0	P	1270.1	264	1282.1	1303.7
	P	0	P	1272.4	266	**1280.4**	**1294.8**
	P	0	const	1403.9	267	1409.9	1420.7

Trajan Apples: model fitting results

Description	Models λ	ω	α	$-2\log$L	df	AIC	BIC
ZIP	H*P	const	0	1338.0	261	1356.0	1388.4
	H*P	P	0	1244.5	260	1264.5	1300.5
	H*P	H*P	0	1238.2	254	1270.2	1327.8
	Lin(H)*P	P	0	1250.2	264	**1262.2**	1283.8
	P	P	0	1261.3	266	1269.3	**1283.7**
	P	const	0	1355.2	267	1361.2	1372.0
ZINB	H*P	const	const	1324.8	260	1344.8	1380.8
	H*P	P	const	1232.5	259	1254.5	1294.1
	H*P	P	P	1226.3	258	1250.3	1293.5
	H*P	H*P	H*P	1205.6	246	1253.6	1340.0
	Lin(H)*P	P	P	1231.0	262	**1247.0**	1275.8
	P	P	P	1237.7	264	1249.7	**1271.3**
	P	P	const	1243.9	265	1253.9	1271.9
	P	const	const	1336.5	266	1344.5	1358.9
	const	P	const	1257.8	266	1265.8	1280.2

Dataset: Biological Pest Control

- Termite *Heterotermes tenuis*: an important pest of sugarcane in Brazil, causing damage of up to 10 metric tonnes/ha/year.

Dataset: Biological Pest Control

- Termite *Heterotermes tenuis*: an important pest of sugarcane in Brazil, causing damage of up to 10 metric tonnes/ha/year.
- Fungus *Beauveria bassiana*: a possible microbial control.

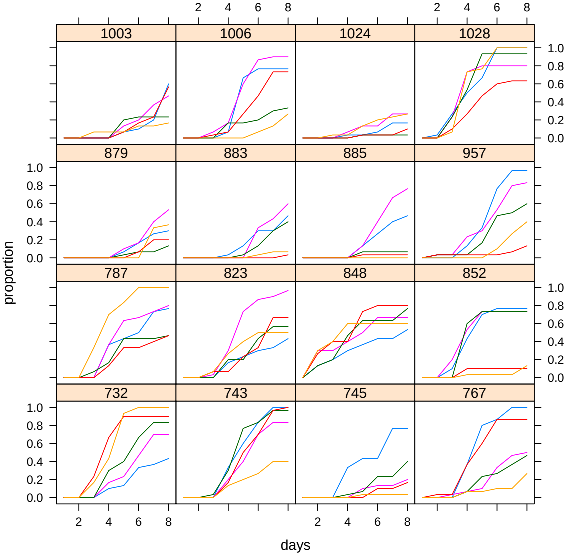
Dataset: Biological Pest Control

- Termite *Heterotermes tenuis*: an important pest of sugarcane in Brazil, causing damage of up to 10 metric tonnes/ha/year.
- Fungus *Beauveria bassiana*: a possible microbial control.
- Experiment: on the pathogenicity and virulence of 142 different isolates of *Beauveria bassiana*.
  - Completely randomized experiment: five replicates of each of the 142 isolates.
  - Solutions of the isolates applied to groups (clusters) of $n = 30$ termites kept in plastic Petri-dishes.
  - Mortality in the groups was measured daily for eight days

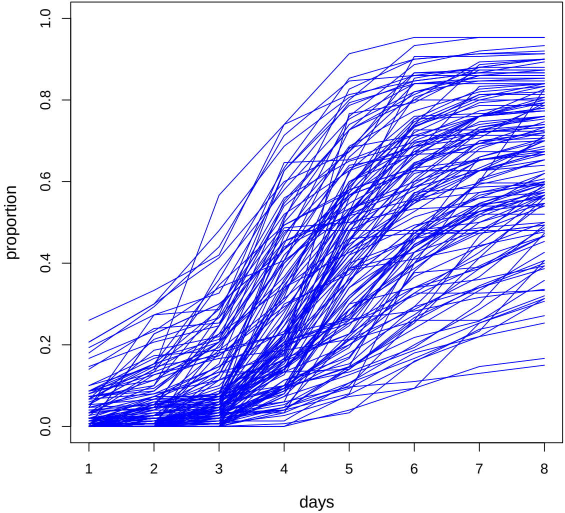
Dataset: Biological Pest Control

- Termite *Heterotermes tenuis*: an important pest of sugarcane in Brazil, causing damage of up to 10 metric tonnes/ha/year.
- Fungus *Beauveria bassiana*: a possible microbial control.
- Experiment: on the pathogenicity and virulence of 142 different isolates of *Beauveria bassiana*.
  - Completely randomized experiment: five replicates of each of the 142 isolates.
  - Solutions of the isolates applied to groups (clusters) of $n = 30$ termites kept in plastic Petri-dishes.
  - Mortality in the groups was measured daily for eight days
- **Data**: 710 ordered multinomial observations of length eight.

Cumulative Mortality: sample of isolates



Cumulative Mortality: spaghetti plot of all isolates



Multinomial Model: Cumulative Proportions

Because of natural time ordering consider models for the **cumulative proportions** (isolate i, replicate k)

$$R_{ik,d} = \text{proportion of insects dead by day } d,$$

$$\gamma_{ik,d} = \mathrm{E}(R_{ik,d}) = \text{probability an insect dies by day } d,$$

$$\mathbf{R}_{ik} = (R_{ik,1}, R_{ik,2}, \ldots, R_{ik,D})^T = \frac{1}{n}\mathbf{LY}_{ik}$$

Multinomial Model: Cumulative Proportions

Because of natural time ordering consider models for the **cumulative proportions** (isolate i, replicate k)

$$R_{ik,d} = \text{proportion of insects dead by day } d,$$

$$\gamma_{ik,d} = \mathrm{E}(R_{ik,d}) = \text{probability an insect dies by day } d,$$

$$\mathbf{R}_{ik} = (R_{ik,1}, R_{ik,2}, \ldots, R_{ik,D})^T = \frac{1}{n}\mathbf{L}\mathbf{Y}_{ik}$$

$$\mathrm{E}[\mathbf{R}_{ik}] = \mathbf{L}\boldsymbol{\pi}_{ik} = \boldsymbol{\gamma}_{ik}$$

$$\mathrm{Var}[\boldsymbol{R}_{ik}] = \frac{1}{n}\boldsymbol{L}[\mathrm{diag}\{\boldsymbol{\pi}_{ik}\} - \boldsymbol{\pi}_{ik}\boldsymbol{\pi}_{ik}^T]\boldsymbol{L}^T = \boldsymbol{V}(\gamma_{ik})$$

Multinomial Model(ctd)

- Use a glm with link function: $g(\gamma_{ik}) = \mathbf{X}_{ik}\beta_i$

Multinomial Model(ctd)

- Use a glm with link function: $g(\gamma_{ik}) = \mathbf{X}_{ik}\beta_i$
- Logit link function $\longrightarrow$ cumulative logistic model

$$g(\gamma_{ikj}) = \text{logit}(\gamma_{ikj}) = \log\left(\frac{\sum_{s=1}^{j} \pi_{ik,s}}{\sum_{s=j+1}^{D+1} \pi_{ik,s}}\right) = \eta_{ikj}$$

- alternative models: discrete survival models, other ordinal models

Multinomial Model(ctd)

- Use a glm with link function: $g(\gamma_{ik}) = \mathbf{X}_{ik}\beta_i$
- Logit link function $\longrightarrow$ cumulative logistic model

$$g(\gamma_{ikj}) = \text{logit}(\gamma_{ikj}) = \log\left(\frac{\sum_{s=1}^{j} \pi_{ik,s}}{\sum_{s=j+1}^{D+1} \pi_{ik,s}}\right) = \eta_{ikj}$$

- Linear predictor: isolate specific factors, time dependency, ...

Multinomial Model(ctd)

- Use a glm with link function: $g(\gamma_{ik}) = \mathbf{X}_{ik}\boldsymbol{\beta}_i$
- Logit link function $\longrightarrow$ cumulative logistic model

$$g(\gamma_{ikj}) = \text{logit}(\gamma_{ikj}) = \log\left(\frac{\sum_{s=1}^{j} \pi_{ik,s}}{\sum_{s=j+1}^{\text{D}+1} \pi_{ik,s}}\right) = \eta_{ikj}$$

- Linear predictor: isolate specific factors, time dependency, ...

  e.g. Isolate specific linear time effect, constant over replicates

$$\eta_{ikj} = \beta_{1i} + \beta_{2i} t_j,$$

Random Effect Models

Incorporate random effects in the linear predictor:

- Add random effect for each experimental unit (groups of insects).
  - simple time shifts
  - time dependent covariates with random coefficients
  - Replicate level random effect — accounts for overdispersion

Random Effect Models

Incorporate random effects in the linear predictor:

- Add random effect for each experimental unit (groups of insects).
  - simple time shifts
  - time dependent covariates with random coefficients
  - Replicate level random effect — accounts for overdispersion
- Model isolates as a random effect.

$$\eta_{ikj} = \mu + time_j + u_i + \epsilon_{ik}$$

Non-parametric maximum likelihood techniques give a finite mass-point distribution $\{\omega_k; z_k\}$ for the isolate effects u_i.
Using a small number of components may identify effective isolates – look at the posterior distribution of u_i.

Dirichlet-Multinomial Model

Additional variation across replicates $\longrightarrow$ overdispersion

Dirichlet-Multinomial Model

Additional variation across replicates $\longrightarrow$ overdispersion

- Allow variation in multinomial parameter $\boldsymbol{\pi}$ — two-stage model

$$\boldsymbol{Y}_{ik} \mid \boldsymbol{p}_{ik} \sim \text{Multinomial}(n; \boldsymbol{p}_{ik})$$

$$\boldsymbol{p}_{ik} = (p_{ik,1}, \ldots, p_{ik,\mathrm{D}}, p_{ik,\mathrm{D}+1})^T \text{ follows a Dirichlet distribution}$$

Dirichlet-Multinomial Model

Additional variation across replicates $\longrightarrow$ overdispersion

- Allow variation in multinomial parameter $\boldsymbol{\pi}$ — two-stage model

$$\boldsymbol{Y}_{ik} \mid \boldsymbol{p}_{ik} \sim \text{Multinomial}(n; \boldsymbol{p}_{ik})$$

$$\boldsymbol{p}_{ik} = (p_{ik,1}, \ldots, p_{ik,\mathrm{D}}, p_{ik,\mathrm{D}+1})^T \text{ follows a Dirichlet distribution}$$

- Dirichlet-multinomial model for **Y** and **R** with

$$E[\boldsymbol{R}_{ik}] = \gamma_{ik}$$

and covariance matrix given by

$$\text{Var}[\boldsymbol{R}_{ik}] = \boldsymbol{V}(\gamma_{ik})[1 + \rho_i(n-1)]$$

where ρ_i is an (isolate specific) overdispersion parameter

Dirichlet-Multinomial Model

Additional variation across replicates $\longrightarrow$ overdispersion

- Allow variation in multinomial parameter $\boldsymbol{\pi}$ — two-stage model

$$\boldsymbol{Y}_{ik} \mid \boldsymbol{p}_{ik} \sim \text{Multinomial}(n; \boldsymbol{p}_{ik})$$

$\boldsymbol{p}_{ik} = (p_{ik,1}, \ldots, p_{ik,\mathrm{D}}, p_{ik,\mathrm{D}+1})^T$ follows a Dirichlet distribution

- Dirichlet-multinomial model for **Y** and **R** with

$$E[\boldsymbol{R}_{ik}] = \gamma_{ik}$$

and covariance matrix given by

$$\text{Var}[\boldsymbol{R}_{ik}] = \boldsymbol{V}(\gamma_{ik})[1 + \rho_i(n-1)]$$

where ρ_i is an (isolate specific) overdispersion parameter

- Generalization of beta-binomial model

Random Intercept Model

- Model additional variation by including random effects in the linear predictor

$$g(q_{ikj}) = \eta_{ikj} + \xi_{ik} = \beta_{1i} + \beta_{2i} t_j + \xi_{ik}$$

where ξ_{ik} is a random effect with $E[\xi_{ik}] = 0$, $\text{Var}[\xi_{ik}] = \sigma_i^2$

Random Intercept Model

- Model additional variation by including random effects in the linear predictor

$$g(q_{ikj}) = \eta_{ikj} + \xi_{ik} = \beta_{1i} + \beta_{2i} t_j + \xi_{ik}$$

where ξ_{ik} is a random effect with $E[\xi_{ik}] = 0$, $\text{Var}[\xi_{ik}] = \sigma_i^2$

- Taylor series approximations give

$$E[\boldsymbol{R}_{ik}] = E[E(\boldsymbol{R}_{ik}|\boldsymbol{q}_{ik})] = E[\boldsymbol{q}_{ik}] \approx \gamma_{ik}$$

and

$$\text{Var}[\boldsymbol{R}_{ik}] \approx \boldsymbol{V}(\gamma_{ik}) + \left(1 - \frac{1}{n}\right) \sigma_i^2 [\boldsymbol{h}'(\boldsymbol{\eta}_{ik})][\boldsymbol{h}'(\boldsymbol{\eta}_{ik})]^T$$

where $\boldsymbol{h}$ is inverse link function with derivative $\boldsymbol{h'}$

Random Intercept Model

- Model additional variation by including random effects in the linear predictor

$$g(q_{ikj}) = \eta_{ikj} + \xi_{ik} = \beta_{1i} + \beta_{2i} t_j + \xi_{ik}$$

where ξ_{ik} is a random effect with $E[\xi_{ik}] = 0$, $\text{Var}[\xi_{ik}] = \sigma_i^2$

- Taylor series approximations give

$$E[\boldsymbol{R}_{ik}] = E[E(\boldsymbol{R}_{ik}|\boldsymbol{q}_{ik})] = E[\boldsymbol{q}_{ik}] \approx \gamma_{ik}$$

and

$$\text{Var}[\boldsymbol{R}_{ik}] \approx \boldsymbol{V}(\gamma_{ik}) + \left(1 - \frac{1}{n}\right) \sigma_i^2 [\boldsymbol{h}'(\boldsymbol{\eta}_{ik})][\boldsymbol{h}'(\boldsymbol{\eta}_{ik})]^T$$

where $\boldsymbol{h}$ is inverse link function with derivative $\boldsymbol{h}'$

- Analagous to approximate variance function for logistic-normal distribution

Random Intercept + Random Slope Model

- Extend to include correlated random effects for intercept and slope

$$g(q_{ikj}) = \beta_{1i} + \xi_{ik} + (\beta_{2i} + \zeta_{ik})t_j = \eta_{ikj} + \xi_{ik} + \zeta_{ik} t_j$$

where $(\xi_{ik}, \zeta_{ik})^T$ has $E[\xi_{ik}] = E[\zeta_{ik}] = 0$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \nu_i^2 & \lambda_i \nu_i \tau_i \\ \lambda_i \nu_i \tau_i & \tau_i^2 \end{bmatrix}$$

Random Intercept + Random Slope Model

- Extend to include correlated random effects for intercept and slope

$$g(q_{ikj}) = \beta_{1i} + \xi_{ik} + (\beta_{2i} + \zeta_{ik})t_j = \eta_{ikj} + \xi_{ik} + \zeta_{ik} t_j$$

where $(\xi_{ik}, \zeta_{ik})^T$ has $E[\xi_{ik}] = E[\zeta_{ik}] = 0$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \nu_i^2 & \lambda_i \nu_i \tau_i \\ \lambda_i \nu_i \tau_i & \tau_i^2 \end{bmatrix}$$

- Approximations now give

$$E[\boldsymbol{R}_{ik}] \approx \boldsymbol{\gamma}_{ik}$$

and

$$\text{Var}[\boldsymbol{R}_{ik}] \approx \boldsymbol{V}(\boldsymbol{\gamma}_{ik}) + \left(1 - \frac{1}{n}\right) \Big\{ \nu_i^2 [\boldsymbol{h}'(\boldsymbol{\eta}_{ik})][\boldsymbol{h}'(\boldsymbol{\eta}_{ik})]^T$$
$$+ \tau_i^2 [\boldsymbol{h}'(\boldsymbol{\eta}_{ik}) * \boldsymbol{t}_{ik}][\boldsymbol{h}'(\boldsymbol{\eta}_{ik}) * \boldsymbol{t}_{ik}]^T + \lambda_i \nu_i \tau_i [\boldsymbol{h}'(\boldsymbol{\eta}_{ik})][\boldsymbol{h}'(\boldsymbol{\eta}_{ik})]^T * [\boldsymbol{1}\boldsymbol{t}_{ik}^T + \boldsymbol{t}_{ik}\boldsymbol{1}^T] \Big\}$$

Results — Surprising Outcome?

Results — Surprising Outcome?

- Parameter estimates from all four models are **identical**

Results — Surprising Outcome?

- Parameter estimates from all four models are **identical**
- Robust se's from all four models are **identical**

Results — Surprising Outcome?

- Parameter estimates from all four models are **identical**
- Robust se's from all four models are **identical**
- Model based se's exhibit simple relationships

Results — Surprising Outcome?

- Parameter estimates from all four models are **identical**
- Robust se's from all four models are **identical**
- Model based se's exhibit simple relationships

- Numerous explanations posited by various colleagues, but . . .

All down to forms of models and matrix algebra

Influence on Teaching

- Extension of general linear model

Influence on Teaching

- Extension of general linear model
- Analysis of non-normal data

Influence on Teaching

- Extension of general linear model
- Analysis of non-normal data
- Likelihood based inference

Influence on Teaching

- Extension of general linear model
- Analysis of non-normal data
- Likelihood based inference
- Model selection, comparison, validation

Influence on Teaching

- Extension of general linear model
- Analysis of non-normal data
- Likelihood based inference
- Model selection, comparison, validation
- Iterative computational methods

Influence on Teaching

- Extension of general linear model
- Analysis of non-normal data
- Likelihood based inference
- Model selection, comparison, validation
- Iterative computational methods
- Extending model classes

Influence on Teaching

- Extension of general linear model
- Analysis of non-normal data
- Likelihood based inference
- Model selection, comparison, validation
- Iterative computational methods
- Extending model classes

Combination of theory & application

Bristol: Generalised Linear Models

Syllabus

- Overview of data analysis, motivating examples. Review of linear models. (1 lecture)
- Generalized linear models (GLMs). Exponential family model, sufficiency issues. Link function, canonical link. (5 lectures)
- Inference for generalized linear models, based on asymptotic theory: confidence intervals, hypothesis testing, goodness of fit. Results interpretation. Diagnostics. (4 lectures)
- Binary responses, logistic regression, residuals and diagnostics. (2 lectures)
- Introduction to survival analysis. Distribution theory: standard parametric models. Proportional odds model and connection to binomial GLM's. Inference assuming a parametric form for the baseline hazard. (4 lectures)

UCSC: Generalized Linear Models

- **Introduction to GLMs**
  Statistical modeling in the context of GLMs. Exponential dispersion family of distributions (definitions, properties, and examples). Components of a GLM, examples of GLMs.
- **Likelihood inference for GLMs**
  Likelihood estimation (iterative weighted least squares) and inference (asymptotic interval estimates). Model diagnostics (residuals for GLMs, model comparison criteria).
- **Regression models for categorical responses and count data**
  Models for binary responses (dose-response modeling, probit and logit models). Poisson regression and log-linear models. Basic ideas for modeling of contingency tables. Multinomial response models for nominal or ordinal responses.
- **Bayesian GLMs**
  General setting, examples, priors for GLMs. MCMC posterior simulation methods for GLMs. Bayesian residual analysis and model choice. Hierarchical GLMs, overdispersed GLMs, generalized linear mixed models.

Acknowledgements

- Norma Coffey
- Clarice Demétrio
- Jochen Einbeck
- Silvia de Freitas
- Emma Holian
- Naratip Jansakul
- Bent Jørgensen
- Marie-José Martinez
- Georgios Papageorgiou
- Martin Ridout
- Mariana Ragassi Urbano
- Afrânio Vieira