Overdispersion in glms

For a well fitting model:

$$\text{Residual Deviance} \approx \text{Residual d.f.}$$

What if Residual Deviance $\gg$ Residual d.f.?

(i) Badly fitting model

- omitted terms/variables
- incorrect relationship (link)
- outliers

(ii) variation greater than predicted by model:
$\Longrightarrow$ **Overdispersion**

- count data: $\text{Var}(Y) > \mu$
- counted proportion data:

$$\text{Var}(Y) > n\pi(1-\pi)$$

Causes of Overdispersion

- variability of experimental material
  – individual level variability
- correlation between individual responses
  e.g. litters of rats
- cluster sampling
  e.g. areas; schools; classes; children
- aggregate level data
- omitted unobserved variables
- excess zero counts (structural and sampling zeros)

Consequences

With correct mean model we have consistent estimates of $\boldsymbol{\beta}$ but:

- incorrect standard errors
- selection of overly complex models

Germination of Orobanche seed

Table 1: Orobanche seed germination data (table entries y_i/m_i).

O. aegyptiaca 75		*O. aegyptiaca 73*	
Bean	Cucumber	Bean	Cucumber
10/39	5/6	8/16	3/12
23/62	53/74	10/30	22/41
23/81	55/72	8/28	15/30
26/51	32/51	23/45	32/51
17/39	46/79	0/4	3/7
	10/13		

Simple Binomial Model

- No of seeds germinating y_i as y-variable
- Binomial logit model
- `Species * Extract` interaction model

Deviance(`Species.Extract`)	=	6.41 on 1 df
Residual Deviance	=	33.28 on 17 df

- Interaction significant
- Some evidence of overdispersion?

Worldwide Airline Fatalities, 1976-85

Year	Fatal accidents	Passenger deaths	Passenger miles (100 million)
1976	24	734	3863
1977	25	516	4300
1978	31	754	5027
1979	31	877	5481
1980	22	814	5814
1981	21	362	6033
1982	26	764	5877
1983	20	809	6223
1984	16	223	7433
1985	22	1066	7107

Simple Models

- Passenger miles (m_i) as exposure variable
- Poisson log-linear model
- Linear time trend

$$\boxed{\begin{aligned} Y_i \sim &\quad \text{Pois}(m_i \lambda_i) \\ \log \lambda_i = &\quad \beta_0 + \beta_1 \texttt{year} \end{aligned}}$$

Fatal accidents:

Deviance(`time trend`) = 20.68

Residual Deviance = 5.46 on 8 d.f.

Passenger deaths:

Deviance(`time trend`) = 202.1

Residual Deviance = 1051.5 on 8 d.f.

$\Longrightarrow$ compounding with aircraft size

Models for Overdispersion

Two broad categories

- assume some more general form for the variance function, possibly with additional parameters.

  Estimation using *moment methods, quasi-likelihood, extended quasi-likelihood, pseudo-likelihood, . . .*

- assume a two-stage model for the response with the response model parameter following some distribution.

  Maximum likelihood estimation (conjugate distribution models) or approximate methods (e.g. using first two moments as above)

  Full hierarchical model – Bayesian methods

Mean-variance Models

Overdispersed Proportion Data

Y_i successes out of m_i trials, $i = 1, \ldots, n$.

Model expected proportions π_i with link function g and

$$g(\pi_i) = \boldsymbol{\beta}'\mathbf{x}_i$$

- constant overdispersion

$$\boxed{\text{Var}(Y_i) = \phi m_i \pi_i (1 - \pi_1)}$$

- A general variance function:
  Overdispersion allowed to depend upon both m_i and π_i.

$$\begin{aligned}\text{Var}(Y_i) \quad &= \quad m_i \pi_i (1 - \pi_i) \times \\ & \quad\quad \left[1 + \phi(m_i - 1)^{\delta_1} \{\pi_i(1 - \pi_i)\}^{\delta_2}\right]\end{aligned}$$

Count data

Random variables Y_i represent counts with means μ_i.

- Constant overdispersion

$$\boxed{\mathsf{Var}(Y_i) = \phi\mu_i}$$

  can arise through a simple compounding process.
  Suppose that $N \sim \mathsf{Pois}(\mu_N)$ and $T = \sum_{i=1}^{N} X_i$, X_i are iid random variables.

$$\mathbf{E}[T] = \mu_T = \mathbf{E}_N(\mathbf{E}[T|N]) = \mu_N\mu_X$$

$$\begin{aligned}\mathsf{Var}(T) &= \mathbf{E}_N[\mathsf{Var}(T|N)] + \mathsf{Var}_N(\mathbf{E}[T|N]) \\ &= \mu_T\left(\frac{\sigma_X^2}{\mu_X} + \mu_X\right) = \mu_T\frac{\mathbf{E}[X^2]}{\mathbf{E}[X]}\end{aligned}$$

- A general variance function

$$\boxed{\mathsf{Var}(Y_i) = \mu_i\left\{1 + \phi\mu_i^{\delta}\right\}}$$

Two-stage Models – Binomial

Beta-Binomial

$$Y_i|P_i \sim \text{Bin}(m_i, P_i)$$

$$\mathbf{E}(P_i) = \pi_i \ \text{Var}(P_i) = \phi\pi_i(1 - \pi_i)$$

Unconditionally, $\mathbf{E}(Y_i) = m_i\pi_i$ and

$$\boxed{\text{Var}(Y_i) = m_i\pi_i(1 - \pi_i)[1 + (m_i - 1)\phi]}$$

Taking $P_i \sim \text{Beta}(\alpha_i, \beta_i)$, with $\alpha_i + \beta_i$ fixed, gives beta-binomial distribution for Y_i with the same variance function.

The same variance function results from assuming that individual binary responses are not independent but have a constant correlation.

Writing $Y_i = \sum_{j=1}^{m_i} R_{ij}$, where R_{ij} are Bernoulli random variables with

$$\mathbf{E}[R_{ij}] = \pi_i \text{ and } \mathsf{Var}(R_{ij}) = \pi_i(1-\pi_i)$$

then, assuming a constant correlation ρ between the R_{ij}'s for $j \neq k$, we have

$$\mathsf{Cov}(R_{ij}, R_{ik}) = \rho\pi_i(1-\pi_i)$$

and

$$\begin{aligned}
\mathbf{E}[Y_i] &= m_i\pi_i \\
\mathsf{Var}(Y_i) &= \sum_{j=1}^{m_i} \mathsf{Var}(R_{ij}) + \sum_{j=1}^{m_i}\sum_{k\neq j} \mathsf{Cov}(R_{ij}, R_{ik}) \\
&= m_i\pi_i(1-\pi_i) + m_i(m_i-1)[\rho\pi_i(1-\pi_i)] \\
&= m_i\pi_i(1-\pi_i)[1+\rho(m_i-1)],
\end{aligned}$$

Logistic-normal and related models

Random effect in the linear predictor

$$\eta_i = \boldsymbol{\beta}'\mathbf{x}_i + \sigma z_i$$

- assume $z_i \sim N(0,1)$
  - Probit-normal model - a convenient interpretation as a threshold model for a normally distributed latent variable (McCulloch, 1994).
  - Logistic-normal using EM algorithm with Gaussian quadrature.
  - Approximate approach using a Williams type III model with

$$\boxed{\mathsf{Var}(Y_i) = m_i\pi_i(1-\pi_i)[1+\phi(m_i-1)\pi_i(1-\pi_i)]}$$

- make no specific distributional assumption about z - estimate a discrete mixing distribution by non-parametric maximum likelihood (NPML).

Considered as the two-stage model, the logit(P_i) have a normal distribution with variance σ^2, i.e. logit$(P_i) \sim \mathsf{N}(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2)$. Writing

$$U_i = \text{logit}(P_i) = \log\frac{P_i}{(1-P_i)} \Rightarrow P_i = \frac{e^{U_i}}{(1+e^{U_i})}$$

and using Taylor series for P_i, around $U_i = \mathbf{E}[U_i] = \mathbf{x}_i^T\boldsymbol{\beta}$, we have

$$P_i = \frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{(1+e^{\mathbf{x}_i^T\boldsymbol{\beta}})} + \frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{(1+e^{\mathbf{x}_i^T\boldsymbol{\beta}})^2}(U_i - \mathbf{x}_i^T\boldsymbol{\beta}) + o(U_i - \mathbf{x}_i^T\boldsymbol{\beta}).$$

Then

$$\mathbf{E}(P_i) \approx \frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{(1+e^{\mathbf{x}_i^T\boldsymbol{\beta}})} := \pi_i$$

and

$$\text{Var}(P_i) \approx \left[\frac{e^{\mathbf{x}_i^T\boldsymbol{\beta}}}{(1+e^{\mathbf{x}_i^T\boldsymbol{\beta}})^2}\right]^2 \text{Var}(U_i) = \sigma^2\pi_i^2(1-\pi_i)^2$$

Consequently the variance function for the logistic-normal model can be approximated by

$$\text{Var}(Y_i) \approx m_i\pi_i(1-\pi_i)[1+\sigma^2(m_i-1)\pi_i(1-\pi_i)]$$

which Williams (1982) refers to as a type III variance function.

Two-stage models – Count data

Negative Binomial Type Variance

- Variation in Poisson rate parameter:

$$Y_i|\theta_i \sim \text{Pois}(\theta_i), \quad \theta_i \sim \Gamma(k, \lambda_i)$$

  leads to negative binomial distribution with

$$\mathbf{E}[Y_i] = \mu_i = k/\lambda_i$$

  and

$$\boxed{\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{k}}$$

  For known k, in the 1-parameter exponential family so still in glm framework.

- Different assumptions for the Γ-distribution lead to different parameterizations with different overdispersed variance functions, e.g. $\theta_i \sim \Gamma(k_i, \lambda)$ gives

$$\boxed{\text{Var}(Y_i) = \mu_i \left(1 + \frac{1}{\lambda}\right) = \phi\mu_i}$$

Negative binomial distribution

$$Y_i|\theta_i \sim \text{Pois}(\theta_i)$$
$$\theta_i \sim \text{Gamma}(k, \lambda_i),\ i = 1, \ldots, n$$

This leads to a **negative binomial distribution** for the Y_i with

$$f_{Y_i}(y_i; \mu_i, k) = \frac{\Gamma(k + y_i)}{\Gamma(k) y_i!} \frac{\mu_i^{y_i} k^k}{(\mu_i + k)^{k+y_i}}, \quad y_i = 0, 1, \ldots$$

and

$$\mathbf{E}(Y_i) = k/\lambda_i = \mu_i$$

$$\begin{aligned} \text{Var}(Y_i) &= \mathbf{E}_{\theta_i}[\text{Var}(Y_i|\theta_i)] + \text{Var}_{\theta_i}(\mathbf{E}[Y_i|\theta_i]) \\ &= \mathbf{E}[\theta_i] + \text{Var}(\theta_i) = \frac{k}{\lambda_i} + \frac{k}{\lambda_i^2} \end{aligned}$$

$$\boxed{\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{k}}$$

Poisson-normal and related models

Individual level random effect in the linear predictor

$$\eta_i = \boldsymbol{\beta}'\mathbf{x}_i + \sigma Z_i$$

- assume $Z_i \sim N(0,1)$, so

  $$Y_i|Z_i \sim \text{Pois}(\lambda_i) \quad \text{with} \quad \log \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma Z_i$$

  where $Z_i \sim N(0,1)$, which gives

  $$\begin{aligned} \mathbf{E}[Y_i] &= \mathbf{E}_{Z_i}(\mathbf{E}[Y_i|Z_i]) = \mathbf{E}_{Z_i}[e^{\mathbf{x}_i^T \boldsymbol{\beta} + \sigma Z_i}] \\ &= e^{\mathbf{x}_i^T \boldsymbol{\beta} + \frac{1}{2}\sigma^2} := \mu_i \\ \text{Var}(Y_i) &= \mathbf{E}_{Z_i}[\text{Var}(Y_i|Z_i)] + \text{Var}_{Z_i}(\mathbf{E}[Y_i|Z_i]) \\ &= e^{\mathbf{x}_i^T \boldsymbol{\beta} + \frac{1}{2}\sigma^2} + \text{Var}_{Z_i}(e^{\mathbf{x}_i^T \boldsymbol{\beta} + \sigma Z_i}) \\ &= e^{\mathbf{x}_i^T \boldsymbol{\beta} + \frac{1}{2}\sigma^2} + e^{2\mathbf{x}_i^T \boldsymbol{\beta} + \sigma^2}(e^{\sigma^2} - 1). \end{aligned}$$

  i.e. a variance function of the form

  $$\boxed{\text{Var}(Y_i) = \mu_i + k'\mu_i^2}$$

- make no specific distributional assumption about Z - estimate a discrete mixing distribution by NPML.

Estimation Methods

Maximum Likelihood

Under the negative-binomial model we have the following expression for the log-likelihood:

$$
\begin{aligned}
\ell(\boldsymbol{\mu}, k; \mathbf{y}) = \sum_{i=1}^{n} \Big\{ & y_i \log \mu_i \\
& + k \log k - (k + y_i) \log(k + \mu_i) \\
& + \log \frac{\Gamma(k + y_i)}{\Gamma(k)} - \log y_i! \Big\} \\
= \sum_{i=1}^{n} \Big\{ & y_i \log \mu_i + k \log k \\
& - (k + y_i) \log(k + \mu_i) \\
& + \mathsf{dlg}(y_i, k) - \log y_i! \Big\}
\end{aligned}
$$

Notice here that for fixed values of k we have a linear exponential family model and consequently a generalized linear model.

Modelling the μ_i's with a linear predictor $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ and link function $g(\mu_i) = \eta_i$ we obtain the following score equations for mle:

$$\begin{aligned}\frac{\partial \ell}{\partial \beta_j} &= \sum_{i=1}^{n} \left\{ \frac{y_i}{\mu_i} - \frac{k + y_i}{k + \mu_i} \right\} \frac{\partial \mu_i}{\partial \beta_j} \\ &= \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\mu_i(1 + \frac{\mu_i}{k})} \frac{1}{g\prime(\mu_i)} x_{ij}\end{aligned}$$

$$\frac{\partial \ell}{\partial k} = \sum_{i=1}^{n} \Big\{ \mathsf{ddg}(y_i, k) - \log(\mu_i + k) \\ - \frac{k + y_i}{k + \mu_i} + \log k + 1 \Big\}$$

The score equations for $\boldsymbol{\beta}$ are the usual quasi-score equations for a glm with

$$V(\mu) = \mu(1 + \frac{\mu}{k})$$

and

$$g(\boldsymbol{\mu}) = \eta$$

and so provide a simple approach for fitting negative binomial regression models using a Gauss-Seidel approach:

1. For a fixed value of k, estimate $\boldsymbol{\beta}$ using a standard glm fit (IRLS) with a variance function $V(\mu) = \mu + \mu^2/k$;

2. for fixed $\boldsymbol{\beta}$, and hence $\boldsymbol{\mu}$, estimate k using a Newton-Raphson iterative scheme

$$k^{(m+1)} = k^{(m)} - \left(\frac{\partial \ell}{\partial k} \Big/ \frac{\partial^2 \ell}{\partial k^2} \right) \Bigg|_{k^{(m)}}$$

3. iterating over 1 and 2 until convergence.

Maximum Quasi-Likelihood

For constant overdispersion models with a variance of the form

$$\boxed{\mathrm{Var}(Y_i) = \phi V_i(\mu_i)}$$

to estimate the regression parameters maximize the quasi-likelihood

$$Q = -\tfrac{1}{2} \sum_{i=1}^{n} \left\{ \frac{D(y_i, \mu_i)}{\phi} \right\}$$

where D is the deviance function

$$D(y, \mu) = -2 \int_{y}^{\mu} \frac{(y-t)}{V(t)} dt$$

$\hat{\boldsymbol{\beta}}$ is the same as the m.l.e. for the exponential

family with variance function $V(\mu)$, e.g. for $V(\mu) = \mu$ – Poisson

The overdispersion parameter ϕ is estimated by equating the Pearson X^2 statistic to the residual degrees of freedom.

1. overdispersed binomial model

$$\widetilde{\phi} = \frac{1}{(n-p)} \sum_{i=1}^{n} \frac{(y_i - m_i\hat{\pi}_i)^2}{m_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

2. overdispersed Poisson model

$$\widetilde{\phi} = \frac{1}{(n-p)} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

The standard errors of the $\widehat{\boldsymbol{\beta}}$ will be as for the non-dispersed model inflated by $\sqrt{\widetilde{\phi}}$.

Extended Quasi-Likelihood (EQL)

For $\text{Var}(Y_i) = \phi_i(\boldsymbol{\gamma})V_i(\mu_i, \boldsymbol{\lambda})$, the extended quasi-likelihood (EQL) criterion (Nelder and Pregibon 1987) involves maximizing

$$Q^+ = -\tfrac{1}{2}\sum_{i=1}^{n}\left\{\frac{D(y_i,\mu_i)}{\phi_i} + \log\left(2\pi\phi_i V_i(y_i)\right)\right\},$$

where D is the deviance function

$$D(y,\mu) = -2\int_y^\mu \frac{(y-t)}{V_i(t)}dt.$$

Beta-binomial variance function

$$V_i(t) = t(1 - t/m_i)$$

and

$$\phi_i = 1 + (m_i - 1)\phi$$

giving a binomial deviance function, $D_B(y_i, \mu_i)$ and

$$Q^+ = -\tfrac{1}{2}\sum_{i=1}^{n}\left\{\frac{D_B(y_i,\mu_i)}{\phi_i} + \log\left[2\pi\phi_i y_i\left(1 - \frac{y_i}{m_i}\right)\right]\right\}$$

EQL Estimation

The quasi-score equation for β_j is

$$\frac{\partial Q^+}{\partial \beta_j} = \sum_{i=1}^{n} \frac{1}{\phi_i} \frac{y_i - \mu_i}{\mu_i(1 - \frac{\mu_i}{m_i})} \frac{1}{g\prime(\mu_i)} x_{ij}$$

- estimation of $\boldsymbol{\beta}$ for known ϕ – usual binomial fit with weights
$$1/\phi_i = 1/[1 + (m_i - 1)\phi]$$
- estimation of ϕ – solve
$$\sum_{i=1}^{n} \left\{ \frac{D(y_i, \mu_i)}{\phi_i} - 1 \right\} \frac{d \log(\phi_i)}{d\phi} = \sum_{i=1}^{n} \left\{ \frac{D(y_i, \mu_i) - \phi_i}{\phi_i^2} \right\} \frac{d\phi_i}{d\phi} = 0$$
  Fit a gamma model to deviance components from the binomial fit using an identity link, fixed intercept of 1 and $m_i - 1$ as explanatory variable.
- Iterate over two estimation steps until convergence.

Pseudo-Likelihood

- Estimate $\boldsymbol{\beta}$ by quasi-likelihood as in EQL.
- Estimate additional parameters in the variance function by maximizing

$$P = -\tfrac{1}{2}\sum_{i=1}^{n}\left\{\frac{(y_i-\mu_i)^2}{\phi_i V_i(\mu_i)} + \log\left(2\pi\phi_i V_i(\mu_i)\right)\right\}.$$

Here, to estimate ϕ we need to solve

$$\sum_{i=1}^{n}\left\{\frac{(y_i-\mu_i)^2}{\phi_i V_i(\mu_i)} - 1\right\}\frac{d\log(\phi_i)}{d\phi} = \sum_{i=1}^{n}\left\{\frac{r_i^2-\phi_i}{\phi_i^2}\right\}\frac{d\phi_i}{d\phi} = 0$$

Procedure same as for EQL, but now using squared Pearson residuals

$$r_i = (y_i-\mu_i)/\sqrt{V_i(\mu_i)}$$

rather than deviances.

Example: Negative Binomial

The estimating equation for k is

$$\sum_{i=1}^{n} \left\{ \frac{(y_i - \mu_i)^2}{\mu_i(1 + \mu_i/k)} - 1 \right\} \frac{d \log V_k(\mu_i)}{dk} = 0$$

where $V_k(\mu_i) = \mu_i(1 + \mu_i/k)$.

Set this up as a gamma estimating equation using

- squared Poisson Pearson residuals as the y-variable
- an identity link
- a linear model with
  - a fixed offset of 1
  - the current estimate of μ_i as the explanatory variable.

Testing Overdispersion

- Goodness of fit test for overdispersed models is problematic
  - overdispersion parameters act as some form of scale parameter
  - effectively estimated from residual deviance
  - scaled deviance and Pearson X^2 typically approx equal to df.
- Test of overdispersion by comparison to the standard model fit.
  - For a two stage model, testing overdispersion frequently reduces to testing an additional parameter.
    Testing the negative binomial against Poisson distributions corresponds to testing $\theta = \widehat{\theta}$ against $\theta \to \infty$.
    This involves testing a value on the boundary of the parameter space - usual asymptotic theory does not apply.
    Appropriate asymptotic distribution for this statistic under the null hypothesis is one which has a probability mass of $\frac{1}{2}$ at 0 and a $\frac{1}{2}\chi^2_{(1)}$ distribution above 0.

Score Tests

Dean (1992) gives score tests for various overdispersion models.

Binary Data

- Beta-binomial

$$\frac{\sum\{[\hat{p}_i(1-\hat{p}_i]^{-1}[(y_i - m_i\hat{p}_i)^2 + \hat{p}_i(y_i - m_i\hat{p}_i) - y_i(1-\hat{p}_i)]\}}{\{2\sum m_i(m_i-1)\}^{1/2}}$$

- Logistic-Normal

$$\frac{\sum\{(y_i - m_i\hat{p}_i)^2 - m_i\hat{p}_i(1-\hat{p}_i)\}}{\hat{V}}$$

Count Data

- Constant overdispersion

$$\frac{1}{\sqrt{2n}} \sum \left\{ \frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} \right\}$$

- Negative Binomial

$$\frac{\sum\{(y_i - \hat{\mu}_i)^2 - y_i\}}{\{2 \sum \hat{\mu}_i^2\}^{1/2}}$$

- Poisson-lognormal

$$\frac{\sum\{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\}}{\{2 \sum \hat{\mu}_i^2\}^{1/2}}$$

She also gives adjusted versions which take account of the estimation of the mean required to compute the statistics.

Half-normal Plots with simulation envelopes

- Fit a model and calculate, $d_{(i)}$, the ordered absolute values of some diagnostic.
- Simulate 19 samples for the response variable using the fitted model and the same values for the explanatory variables.
- Refit the model to each sample and calculate the ordered absolute values of the diagnostic of interest, $d^*_{j(i)}$, $j = 1, \ldots, 19$, $i = 1, \ldots, n$.
- For each i, calculate the mean, minimum and maximum of the $d^*_{j(i)}$.
- Plot these values and the observed $d_{(i)}$ against the half-normal order statistics.

Model selection

From Jorgensen (1997), when:

- ϕ is known, in particular for natural exponential families, inference may be based on
$$D_2 - D_1 \sim \chi^2$$
- ϕ is unknown, inference may be based on the following $F-$statistic,
$$F = \frac{(D_2 - D_1)/(f_2 - f_1)}{D_1/f_1} \sim F(f_2 - f_1, f_1)$$
if ϕ is small.

For overdispersed models a test statistic for two nested models is

$$F = \frac{(D_2 - D_1)/(f_2 - f_1)}{\widehat{\phi}} \sim F_{f_2 - f_1, f_3}$$

where $\widehat{\phi}$ is estimated from preferably a maximal model, with f_3 df.

Then for an overdispersion parameter estimated from fitting the full model:

- fix its value
- weight the observations by $w_i = \frac{1}{\phi_i}$
- fit different sub-models
- the difference in the deviance for two alternative (nested) models is compared with percentage points of the F;
  non-significant result $\Rightarrow$ the two models cannot be distinguished

- Many interesting comparisons involve non-nested models
- Use of Akaike Information Criterion (AIC) or Bayes Information Criterion (BIC) for model selection

AIC $= -2 \log \mathsf{L} + 2$ (number of fitted parameters)
BIC $= -2 \log \mathsf{L} + \log n$ (number of fitted parameters)

Models for Orobanche Data

Binomial:

$$\boxed{\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i)}$$

- significant interaction
- overdispersion

Quasi-likelihood:

$$\boxed{\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i)}$$

- constant overdispersion $\widetilde{\phi} = 1.862$
- only marginal evidence of interaction
- extract only important factor

Williams:

$$\boxed{\text{Var}(Y_i) = m_i \pi_i (1 - \pi_i)[1 + \phi(m_i - 1)]}$$

- moment estimate $\widetilde{\phi} = 0.0249$
- only marginal evidence of interaction
- extract only important factor
- similarity to QL, even though m_i not equal.

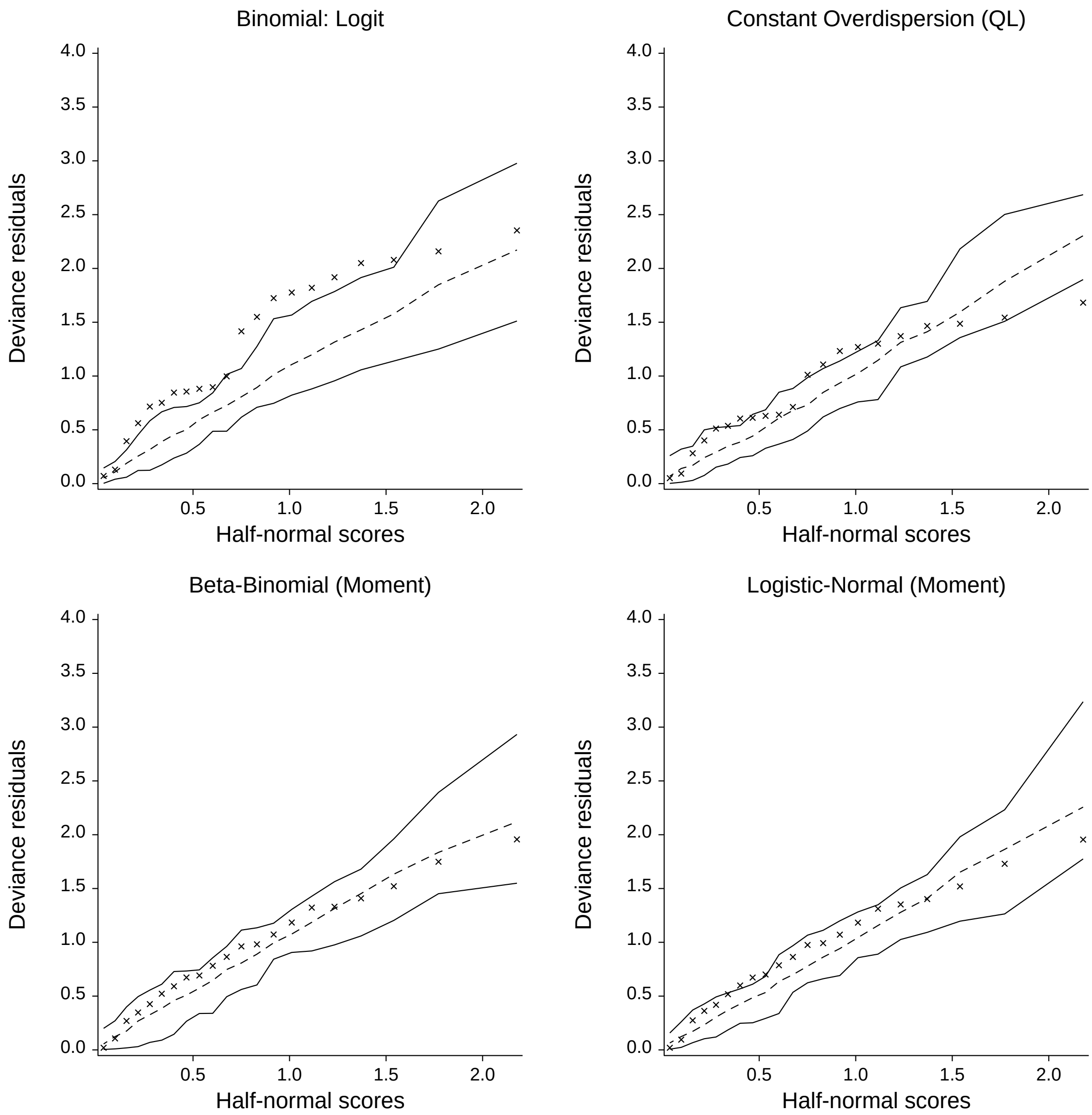

Figure 1: Orobanche Data. Half-normal plots: $\times$ – data; — simulated envelope

Quine's School Absence Data

Aitkin et al (1989) describe a data set on absence from school from a Sociological study of Australian Aboriginal and white children.
The response variable of interest is the number of `days` absent from school cross-classified by age (`A`, 4 levels), sex (`S`, 2 levels), cultural group (`C`, 2 levels) and learning rate (`L`, 2 levels).

- Using a Poisson model the residual deviance is very large even for the maximal model
  Deviance=1173.9 on 118 df.
  $\Rightarrow$ strong evidence of overdispersion.
- The negative binomial distribution provides a possible overdispersion model for this data
  Deviance=167.4 on 118 df.

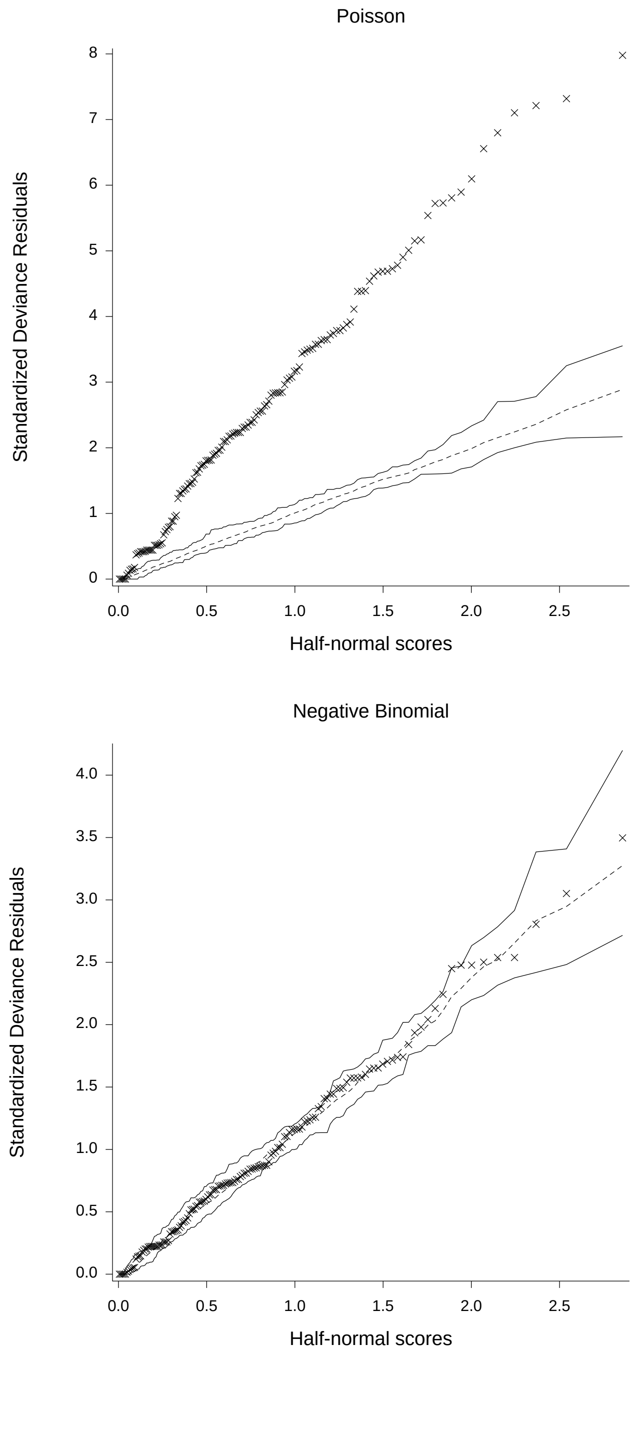

Figure 2: Quine data. Half-normal plots: × – data; — simulated envelope

Extended overdispersion models

Random effect models

- In many applications the overdispersion mechanism is assumed to be the same for all of the observations.
- However, in some applications it is quite conceivable that the overdispersion may be different in different subgroups of the data.
- Explicit models for the variance, and hence overdispersion, are easily handled by an additional model for the scale parameter of the form

$$h(\phi_i) = \boldsymbol{\gamma}^T \mathbf{z}_i$$

  - link function h, usually the identity or the log.
  - vector of explanatory variables $\mathbf{z}_i$ may include covariates in the mean model giving great flexibility for joint modelling of the mean and dispersion.
  - estimation can proceed by either EQL or PL using a gamma estimating equation for γ

Generalized linear mixed models

Another natural way to extend the category of two-stage models is to add more complex random effects structures in the linear predictor, taking

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

where $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{\gamma}$ is a vector of random effects and $\mathbf{x}_i$ and $\mathbf{z}_i$ are corresponding vectors of explanatory variables.

- Assuming that these random effects are normally distributed gives a direct generalization of the standard linear mixed model for normally distributed responses to what is commonly called the generalized linear mixed model (GLMM).
- Estimation within this family is non-trivial and a number of different approaches have been proposed, including penalised quasi-likelihood, restricted maximum likelihood and Bayesian methods using Markov chain Monte Carlo.
- In some simple models with nested random effects, maximum likelihood estimation is possible.

- In many situations the assumption of normality for the random effects is neither natural nor computationally convenient and Lee & Nelder (1996) propose an extension of GLMMs to hierarchical generalized linear models.
  - the random components can come from an arbitrary distribution, although they particularly favour the use of a distribution conjugate to that of the response.
  - estimation is based on h-likelihood, a generalization of the restricted maximum likelihood method used for standard normal linear mixed model.
  - such models are also easily handled within the Bayesian paradigm using Markov chain Monte Carlo methods
  - the non-parametric maximum likelihood approach can also be extended to these more complex models

Models for counts with excess zeros

- Mixed Poisson distributions
- Zero-modified distributions
- Hurdle models
- Semi-parametric hurdle models
- Birth process models
- Threshold models