ON HIERARCHICAL GENERALIZED LINEAR MODELS

ZHONG-ZHAN ZHANG

NA XIA

College of Applied Sciences
Beijing University of Technology

ZZHANG@bjut.edu.cn

- ***Introduction***
- ***Definitions and Implementation***
- ***Statistical Inference?***
- ***A Special Model***
- ***Discussions***

1. Introduction

- Key Words: heterogeneity and dispersion modelling
- Background in practice: heterogeneity appears in clinical trials, economy research, financial engineering etc.
- Mechanism causing heterogeneity: unobservable or latent random variable
- Methodology: put random effect terms into model

Trace of Development:

Linear Model

$$y = X\beta + \varepsilon$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

$\Rightarrow$

Linear Mixed Model

$$y = X\beta + Zu + \varepsilon$$
$$u, \varepsilon \sim N(0, *)$$

$\Downarrow$ (Linear Model → GLM)

GLM

$$E(y) = \mu$$
$$g(\mu) = X\beta$$

$\Rightarrow$

$\Downarrow$ (Linear Mixed Model → GLMM)

GLMM

$$E(y|u) = \mu'$$
$$g(\mu') = X\beta + Zu$$
$$u \sim N(0, *)$$

$\Downarrow$

Hierarchical Generalized Linear Model

Generalized Linear Mixed Models and Related Works

- There were many researches since 1980's
- GLMM(Breslow and Clayton, 1993, JASA):

  $E(y|v) = \mu' = h(X\beta + Zv)$,

  $var(y|v) = \phi a_i V(\mu')$

  $v \sim N(0, D(\theta))$

  Approximation method:

  Penalized Quasi-Likelihood

  Marginal Quasi-Likelihood

  Breslow and Lin (1995, Biometrika)

  Vonesh et al(2002, JASA)

- Multilevel models (Goldstein, 1995) etc
- Longitudinal Data (Diggle, Liang and Zeger, 1994) etc
- Bayesian (Wakefield et al 1994, Appl Stat) etc
- Applications: biostatistics, economy, quality control, etc

2. Definitions and Implementation

Hierarchical Generalized Linear Models
(Lee, Nelder and others, 1996-2005)

- Motivation:
  - Extending normal distri. of random components to arbitrary
  - Using Hierarchical likelihood instead of marginal distri. with integration
- Questions: how to make inference? how to compute? how good?

2.1. HGLM

- Model:
  - Conditional (log-)likelihood:

$$l(\theta', \phi; y|u) = [y\theta' - b(\theta')]/\phi + c(y, \phi)$$

  - Link function:

$$E(y|u) = h(X\beta + v), \;\; v = v(u) : \text{monotone}$$

  - Distri. of u: various including Gamma, Inverse Gamma, LogNormal, etc.
  - Conjugate distri. of u:

$$\theta' = \theta(\mu) + \theta(u)$$

$v = \theta(u)$ has a log-likelihood like $a_1(\alpha)v - a_2(\alpha)b(v)$

- h-likelihood:

$$h = l(\beta, \phi; y|v) + l(\alpha; v)$$

where $l(\beta, \phi; y|v)$ the log density for $y|v$, $l(\alpha; v)$ that for v.

- MHLE: estimators got from maximizing h by solving

$$\partial h/\partial \beta = 0, \;\; \partial h/\partial v = 0$$

- Choice of v: the scale on which the random effect of u is linear
- When dispersion parameters are not known, a profile likelihood method is suggested.

- For conjugate u: for data like y_{ij}, where i is the subject index, and j that of the replicated observation, h-likelihood becomes

$$\sum_{ij}[y_{ij}\theta'_{ij} - b(\theta'_{ij})]/\phi + \sum_{i}[a_1(\alpha)v_i - a_2(\alpha)b(v_i)]$$

$\partial b(\theta')/\partial\theta' = \mu'$ leads to $\partial b(v)/\partial v = u$,

$$\partial h/\partial v_i = \Big[\sum_j (y_{ij} - \mu'_{ij}) + \phi a_1(\alpha)\Big]/\phi - a_2(\alpha)u_i$$

and the estimator $\hat{u}_i$ satisfies

$$\hat{u}_i = \frac{\sum_j [y_{ij} - \mu'_{ij}] + \phi a_1(\alpha)}{\phi a_2(\alpha)}$$

2.2. Fitting Implementation For Conjugate HGLMs

- Main Idea:

(1)α known, Iterative weighted least square

$$\begin{pmatrix} \boldsymbol{XWX}^\tau & \boldsymbol{XWZ}^\tau \\ \boldsymbol{ZWX}^\tau & \boldsymbol{ZWZ}^\tau + \boldsymbol{U} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} + \delta\boldsymbol{\beta} \\ \boldsymbol{v} + \delta\boldsymbol{v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{XWw} \\ \boldsymbol{ZWw} + \boldsymbol{R} \end{pmatrix}$$

for estimating $\boldsymbol{\beta}$ and $\boldsymbol{v}$;

where

$$\boldsymbol{W} = \text{diag}(\phi^{-1}(\boldsymbol{X}_{11}^{\tau}\boldsymbol{\beta})\mu'_{11}, \cdots, \phi^{-1}(\boldsymbol{X}_{tn_t}^{\tau}\boldsymbol{\beta})\mu'_{tn_t}),$$
$$\boldsymbol{\mu}'_i = (\mu'_{i1}, \cdots, \mu'_{in_i})^{\tau},$$
$$\boldsymbol{\mu}' = (\boldsymbol{\mu}_1'^{\tau}, \cdots, \boldsymbol{\mu}_t'^{\tau});$$
$$\boldsymbol{U} = \alpha\ \text{diag}(u_1, \cdots, u_t),$$
$$\boldsymbol{\mathcal{U}} = \text{diag}(\boldsymbol{\mathcal{U}}_1, \cdots, \boldsymbol{\mathcal{U}}_t),$$
$$\boldsymbol{\mathcal{U}}_i = \text{diag}(\mu'_{i1}, \mu'_{i2}, \cdots, \mu'_{in_i}),$$
$$\boldsymbol{w} = \log \boldsymbol{\mu}' + \boldsymbol{\mathcal{U}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}'),$$
$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{v} + (\alpha\boldsymbol{1} - \alpha\boldsymbol{e}^{\boldsymbol{v}}),$$
$$\boldsymbol{1} = (1, \cdots, 1)^{\tau}.$$

$Z = (z_{ijk})$ $n \times t$ matrix with $z_{ijk} = 1$ if $i = k$, $z_{ijk} = 0$ otherwise, $i, k = 1, \cdots, t, j = 1, \cdots, n_i$; $\boldsymbol{\delta\beta}$ and $\boldsymbol{\delta v}$ are the adjustment of $\boldsymbol{\beta}$ and $\boldsymbol{v}$ respectively; $\boldsymbol{w}$ is the adjusted dependent variable.

(2) α unknown,

$$\sum_{i=1}^{t} \{v_i + \log \alpha + 1 - u_i - \psi(\alpha)\}$$

$$-\frac{1}{2} tr \{\boldsymbol{K} diag(u_1, \cdots, u_t)\} = 0$$

to estimate it, where ψ is digamma function, $\boldsymbol{K}$ is determined from

$$\begin{pmatrix} \boldsymbol{XWX}^{\tau} & \boldsymbol{XWZ}^{\tau} \\ \boldsymbol{ZWX}^{\tau} & \boldsymbol{ZWZ}^{\tau} + \boldsymbol{U} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{K} \end{pmatrix}.$$

3. Inference

- Point Estimation: get from the iterative multiple interconnected WLS.
- Interval Estimation and Test: Lee and Nelder employ the methods of GLM in principle
- Properties: for some special examples, intuitively good.
- Further Investigations Needed: Lihua Wang (doctoral thesis, University of Georgia, 2004)

4. A Special Model

Data:

y: event count, $\boldsymbol{X}$: covariate vector

$(y_{ij}, \boldsymbol{X_{ij}})$: observed from the jth subject in the ith group

u_i: unobservable common random component for group i

Model: multiplicative Poisson-Gamma type model

$y_{ij}|u_i$: with mean $\mu'_{ij} = \mu_{ij}u_i$ and variance $\phi\boldsymbol{\mu}'_i$

$u_i \sim$ Gamma(α, α), with mean 1 and a known variance parameter α.

- $\boldsymbol{\mu}'_i = E(\boldsymbol{Y}_i|\boldsymbol{v}) = \boldsymbol{e}^{(\boldsymbol{X}_i^\tau \boldsymbol{\beta} + \boldsymbol{v}_i)}$
- $\boldsymbol{\mu}_i = \boldsymbol{e}^{(X_i^\tau \beta)}, \ u_i = e^{v_i}$
- $\boldsymbol{\theta}'_i = \log \boldsymbol{\mu}'_i = (\boldsymbol{X_i^\tau \beta} + \boldsymbol{v_i})$

The kernel of the H-likelihood for P-G type model:

$$H = \sum_{i=1}^{t} \{(\boldsymbol{Y}_i^\tau \boldsymbol{\theta}_i' - \boldsymbol{1}^\tau \boldsymbol{e}^{\boldsymbol{\theta}_i'})/\phi\} + \sum_{i=1}^{t} \left\{\alpha v_i - \alpha e^{v_i}\right\} \quad (3.1)$$

Consider the over-dispersion factor which depends on covariate variables through known functions $\phi(\boldsymbol{X}_{ij}, \boldsymbol{\beta})$, that is, for the jth subjects in the ith group, the over-dispersion factor is $\phi_{ij} = \phi(\boldsymbol{X}_{ij}, \boldsymbol{\beta})$. This leads to estimating equation

$$\sum_{i=1}^{t} \boldsymbol{X}_i \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta})(\boldsymbol{Y}_i - e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} u_i) = 0 \quad (3.2)$$

$$\{(\boldsymbol{Y}_i - \boldsymbol{\mu}_i u_i)^\tau \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta})\boldsymbol{1}\} + \alpha - \alpha u_i = 0 \quad (3.3)$$

where $\boldsymbol{\Phi}_i(\boldsymbol{\beta}) = diag(\phi_{i1}, \cdots, \phi_{in_i})$.

Remark:

One explanation of $\phi_{ij} = \phi(\boldsymbol{X}_{ij}, \boldsymbol{\beta})$ is, there is another multiplicative subject-wise random factor η_{ij} with mean 1 and constant variance, independent of u_i, such that

$$y_{ij}|u_i, \eta_{ij} \sim P(e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} u_i \eta_{ij})$$

which leads to

$$E(y_{ij}|u_i, \eta_{ij}) = e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} u_i \eta_{ij}$$

and

$$var(y_{ij}|u_i) = e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} u_i [1 + e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} u_i var(\eta_{ij})]$$

It holds that

- $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{e}^{\boldsymbol{x}},\ \ \boldsymbol{b}(\boldsymbol{\theta}'_i) = \boldsymbol{e}^{\boldsymbol{\theta}'_i},\ \ b(v_i) = e^{v_i} = u_i.$
- $\mu''_{ij} = \mu_{ij} u_i \eta_{ij} = E(y_{ij}|u, \eta)$
- $\mu'_{ij} = \mu_{ij} u_i = E(y_{ij}|\boldsymbol{u});\ \mu_{ij} = e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}};$
- $\boldsymbol{COV}(\boldsymbol{Y}_i|\boldsymbol{v}) = \boldsymbol{\phi}(\boldsymbol{X}_i^{\tau}\boldsymbol{\beta})\ddot{\boldsymbol{b}}(\boldsymbol{\theta}'_i) = \boldsymbol{\phi}(\boldsymbol{X}_i^{\tau}\boldsymbol{\beta})\boldsymbol{\mathcal{U}}_i;$
- $E(u_i) = 1;\quad a_1(\alpha) = a_2(\alpha) = \alpha.$
- $E(\eta_{ij}) = 1;\quad var(\eta_{ij}) < \infty.$

From (3.3), we have

$$\hat{u}_i = \frac{\boldsymbol{Y}_i^\tau \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta})\mathbf{1} + \alpha}{\boldsymbol{\mu}_i^\tau \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta})\mathbf{1} + \alpha}. \tag{3.4}$$

Substituting u_i with $\hat{u}_i$ in (3.2), we obtain the equation

$$\sum_{i=1}^{t} \boldsymbol{X}_i \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta}) \Big(\boldsymbol{\mathcal{E}}_i - e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} \mathbf{1}^\tau \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta}) \boldsymbol{\mathcal{E}}_i R_i(\boldsymbol{\beta})$$

$$+ e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} (u_i - 1) \alpha R_i(\boldsymbol{\beta}) \Big) = 0 \tag{3.5}$$

where $\boldsymbol{\mathcal{E}}_i = \boldsymbol{Y_i} - e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} u_i$,

$$R_i(\boldsymbol{\beta}) = 1/[(e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})^\tau \boldsymbol{\Phi}_i^{-1}(\boldsymbol{\beta})\mathbf{1} + \alpha].$$

Replace the $\mathbf{\Phi}_{\boldsymbol{i}}^{-1}(\boldsymbol{\beta})$ behind $\boldsymbol{X}_i$ by

$$\mathbf{\Lambda}_{\boldsymbol{i}}(\boldsymbol{\beta}) = diag(e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})(I - R_i(\boldsymbol{\beta})\mathbf{\Phi}_i^{-1}(\boldsymbol{\beta})\mathbf{1}(e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})^\tau)V_i(\boldsymbol{\beta})$$

where $V_i(\boldsymbol{\beta}) > 0$ is a given weight matrix, $i = 1, \cdots, t$. This leads to a GEE

$$L_t(\boldsymbol{\beta}) \equiv \sum_{i=1}^{t} \boldsymbol{X}_i \mathbf{\Lambda}_i(\boldsymbol{\beta}) \Big[\boldsymbol{\mathcal{E}}_i - e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} \mathbf{1}^\tau \mathbf{\Phi}_i^{-1}(\boldsymbol{\beta}) \boldsymbol{\mathcal{E}}_i R_i(\boldsymbol{\beta}) + e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}}(u_i - 1)\alpha R_i(\boldsymbol{\beta}) \Big] = 0. \qquad (3.6)$$

We discuss the properties of solution $\hat{\boldsymbol{\beta}}_t$ of (3.6)

Suppose that $\beta_0 \in \mathcal{B}^p$, where $\mathcal{B}^p$ is a bounded subset of p-dimensional Euclid space. Some additional assumptions are given as follows.

A1. $\{\boldsymbol{X}_i, i \geq 1\}$ *has boundedness,* $\lambda_t \geq ct^\delta$ *for sufficiently large* t *and* $\delta \in (3/4, 1]$, *where* λ_t *is the smallest eigenvalue of the symmetric matrix* $\sum_{i=1}^{t} \boldsymbol{X}_i \boldsymbol{X}_i^\tau$;

A2. $E\left[\boldsymbol{COV}_{\boldsymbol{\beta}_0}(\boldsymbol{Y}_i|u_i)\right] \geq cI, i = 1, 2, \cdots,$

$\sup_{i\geq 1} \boldsymbol{E}_{\boldsymbol{\beta}_0} \|\boldsymbol{Y}_i\|^{\bar{p}} < \infty, \bar{p} = 8/3;$

A3. $\phi(\cdot,\cdot) > c > 0$ *is twice continuously differentiable, moreover,* $\phi(\cdot)$ *and its first and second order partial derivative have boundedness in arbitrary bounded subset;*

A4. *for all* i, $\boldsymbol{V}_i(\boldsymbol{\beta}) > cI$ *for all* $\boldsymbol{\beta} \in \mathcal{B}^p$, *where* $c > 0$ *is a constant independent of* i; *and all elements of* $\boldsymbol{V}_i(\boldsymbol{\beta})$ *have continuous second order partial derivatives; moreover, the elements of* $\boldsymbol{V}_i(\boldsymbol{\beta})$, *their first and second order partial derivatives are bounded in arbitrary subset of* $\mathcal{B}^p$.

Theorem 1 (Consistency) *Suppose the assumptions A1∼A4 hold, then there exists an estimator $\hat{\boldsymbol{\beta}}_t$ of $\boldsymbol{\beta}_0$, such that*

$$P\Big(\boldsymbol{L}_t(\hat{\boldsymbol{\beta}}_t) = 0, \textit{for all sufficiently large } t\Big) = 1, \tag{3.7}$$

$$\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0 = O(t^{-(\delta-1/2)}(\log\log t)^{1/2}) \quad a.s. \tag{3.8}$$

In the important special case of $\delta = 1$, *it holds*

$$\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0 = O(t^{-1/2}(\log\log t)^{1/2}), \tag{3.9}$$

which is the rate of law of the iterated logarithm for iid series.

The main idea of the proof of Th. 1

The existence and properties of $\hat{\boldsymbol{\beta}}_t$ based on the fact that if

$$\inf\left\{\|\boldsymbol{L_t}(\boldsymbol{\beta}) - \boldsymbol{L_t}(\boldsymbol{\beta_0})\| : \boldsymbol{\beta} \in \bar{S}_t\right\} > 2\|\boldsymbol{L_t}(\boldsymbol{\beta_0})\|,$$

where $\bar{S}_t$ is the surface of sphere S_t centered at $\boldsymbol{\beta_0}$, then

$$\inf\left\{\|\boldsymbol{L_t}(\boldsymbol{\beta})\| : \boldsymbol{\beta} \in \bar{S}_t\right\} > \|\boldsymbol{L_t}(\boldsymbol{\beta_0})\|,$$

thus $\|\boldsymbol{L_t}(\boldsymbol{\beta})\|$ has a local minimum point $\hat{\boldsymbol{\beta}}_t$ in the interior of S_t. Moreover, $\boldsymbol{L_t}(\hat{\boldsymbol{\beta}}_t) = 0$ is to be verified.

Denote $\mathcal{U}_i = diag(\mu_{i1}, \cdots, \mu_{in_i}), \mathcal{U}'_i = diag(\mu'_{i1}, \cdots, \mu'_{in_i})$

$$\boldsymbol{Q}_t(\boldsymbol{\beta}) = \boldsymbol{L}_t(\boldsymbol{\beta})/t$$

$$\begin{aligned}\boldsymbol{W_t}(\boldsymbol{\beta}) &= COV(\sqrt{t}\boldsymbol{Q}_t(\boldsymbol{\beta})) \\ &= \frac{1}{t}\sum_{i=1}^{t} \boldsymbol{X}_i E\Big\{\Lambda_i(\boldsymbol{\beta})[(I - A_i)^\tau \Phi_i(\boldsymbol{\beta})\mathcal{U}'_i(I - A_i) \\ &\quad + \boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}}(\boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})^\tau \alpha^2 E[(u_i - 1)^2 R_i^2(\boldsymbol{\beta})]]\Lambda_i(\boldsymbol{\beta})\Big\}\boldsymbol{X}_i\end{aligned}$$

where $A_i = \Phi_i^{-1}(\boldsymbol{\beta})\mathbf{1}(\boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})^\tau R_i(\boldsymbol{\beta})$;

$\boldsymbol{\Phi_i}(\boldsymbol{\beta})$, $\boldsymbol{R_i}(\boldsymbol{\beta})$ are the same as above.

Theorem 2 (Normality) Further suppose that

A5: $\lim_{t\to\infty} \boldsymbol{W_t}(\boldsymbol{\beta}) = \boldsymbol{W}(\boldsymbol{\beta})$;

A6: $\lim_{t\to\infty} \boldsymbol{E_\beta}(-\frac{\partial \boldsymbol{Q_t}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\tau}) = \boldsymbol{F}(\boldsymbol{\beta})$, *where* $\boldsymbol{F}(\boldsymbol{\beta})$ *is a positive definite matrix.*

A7: $\{n_i : i = 1, \cdots, t\}$ *is bounded.*

Then

1.
$$\sqrt{t}\, \boldsymbol{Q_t}(\boldsymbol{\beta_0}) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{W}(\boldsymbol{\beta_0})) \qquad (3.10)$$

2.
$$\sqrt{t}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta_0}) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{F}^{-1}(\boldsymbol{\beta_0})\boldsymbol{W}(\boldsymbol{\beta_0})(\boldsymbol{F}^{-1}(\boldsymbol{\beta_0}))^\tau) \qquad (3.11)$$

3. $F(\beta_0)$ and $W(\beta_0)$ may be consistently estimated by plug-in method.

Simulation Results:

- Sample size $t = 50, n_i = 20.$

  5000 replications

- Two models:

$$\phi_{ij} = e^{-\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} + 1,$$

$$\phi_{ij} = \frac{u_i e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}}{1+e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}} + 1.$$

Simulation Results(For L-N estimator):

Table 1: $\phi_{ij} = e^{-\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} + 1$, α known

α	β	MSE($\hat{\boldsymbol{\beta}}_{\boldsymbol{t}}$)	$K > 1.36$	$K > 1.23$
0.5	0.5	4.7580×10^{-7}	0.0440	0.0870
1	0.2	1.4823×10^{-5}	0.0386	0.0784
	0.5	4.3893×10^{-7}	0.0404	0.0826
	0.8	1.0756×10^{-8}	0.0452	0.0858
1.5	0.5	4.3345×10^{-7}	0.0408	0.0832

$*\dagger$ K: Kolmogorov Statistics for $\hat{u}_i$;

1.36 and 1.24: 0.05 and 0.1 upper quantiles of Kolmogorov Distribution.

Table 2: $\phi_{ij} = e^{-\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}} + 1$, α unknown

α	β	MSE($\hat{\boldsymbol{\beta}}_t$)	MSE($\hat{\alpha}$)	$K > 1.36$	$K > 1.23$
0.5	0.5	4.7580×10^{-7}	0.0078	0.0440	0.0870
1	0.2	1.4823×10^{-5}	0.0377	0.0386	0.0784
	0.5	4.4434×10^{-7}	0.0390	0.0464	0.0874
	0.8	1.0774×10^{-8}	0.0382	0.0416	0.0814
1.5	0.5	4.3047×10^{-7}	0.0921	0.0454	0.0872

Table 3: $\phi_{ij} = \frac{u_i e^{\boldsymbol{X}_{ij}^{\mathsf{T}}\boldsymbol{\beta}}}{1+e^{\boldsymbol{X}_{ij}^{\mathsf{T}}\boldsymbol{\beta}}} + 1$, α known

α	β	MSE($\hat{\boldsymbol{\beta}}_t$)	$K > 1.36$	$K > 1.23$
0.5	0.5	1.2698×10^{-6}	0.0432	0.0846
1	0.2	1.4823×10^{-5}	0.0336	0.0716
	0.5	1.0421×10^{-6}	0.0368	0.0822
	0.8	2.5440×10^{-8}	0.0408	0.0860
1.5	0.5	9.7594×10^{-7}	0.0442	0.0826

Table 4: $\phi_{ij} = \frac{u_i e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}}{1+e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}} + 1$, α unknown

α	β	MSE($\hat{\boldsymbol{\beta}}_t$)	MSE($\hat{\alpha}$)	$K > 1.36$	$K > 1.23$
0.5	0.5	1.2668×10^{-6}	0.0079	0.0460	0.0852
1	0.2	2.6069×10^{-5}	0.0372	0.0300	0.0656
	0.5	1.0490×10^{-6}	0.0377	0.0418	0.0848
	0.8	2.5740×10^{-8}	0.0375	0.0424	0.0856
1.5	0.5	9.1786×10^{-7}	0.0932	0.0378	0.0804

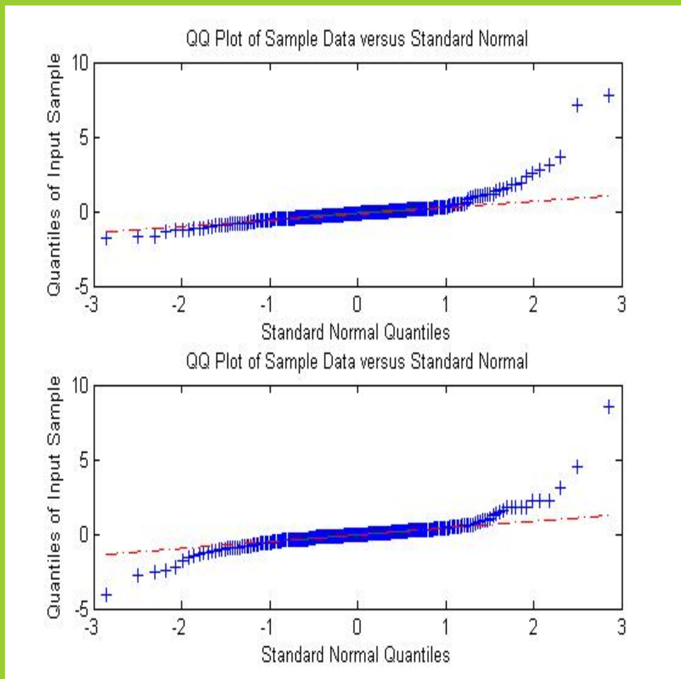
An Example:

- Data: (Thall and Vail, 1990) 59 epileptics
- response: seizure counts during two weeks before each of four visits to clinic.
- covariates: T=treatment(progabide or placebo), A=log(age), B=log(base record), T$*$B, V=Visit,V_4=the fourth visit
- The model described in page 17 might include a random dispersion.

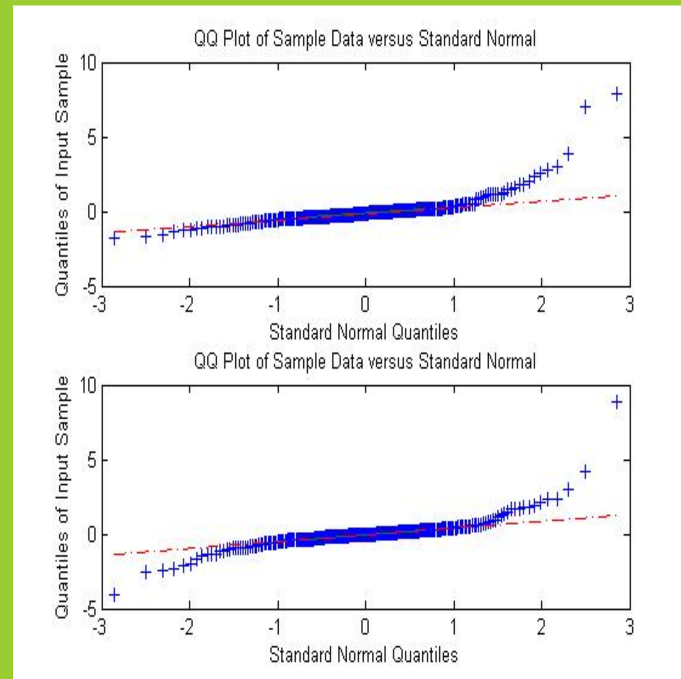
Model 1: $\rho_{ij} = u_i e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}, \quad \phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2.$(constant dispersion)

Model 2: $\rho_{ij} = e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}, \quad \phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1+u_i.$(random dispersions)

- we first give out the Q-Q plots of the standardized residuals $\tilde{r}_{ij1} = (\hat{r}_{ij1} - \bar{r}_1)/\hat{\sigma}_1$ and the standardized conditional residuals $\tilde{r}_{ij2} = (\hat{r}_{ij2} - \bar{r}_2)/\hat{\sigma}_2$ of the complete data for Model 1 and Model 2 respectively in Figure 1 and Figure 2, where $\hat{r}_{ij1} = y_{ij} - e^{\boldsymbol{X}_{ij}^{\tau}\hat{\boldsymbol{\beta}}}, \hat{r}_{ij2} = y_{ij} - e^{\boldsymbol{X}_{ij}^{\tau}\hat{\boldsymbol{\beta}}}\hat{u}_i, \bar{r}_h = \frac{1}{236}\sum_{i,j}\hat{r}_{ijh}, \hat{\sigma}_h = \sqrt{\frac{1}{235}\sum_{i,j}(\hat{r}_{ijh} - \hat{r}_{ijh})^2}, h = 1, 2$. In each figure, the above plot is of the standardized residuals, and the below one is of the standardized conditional residuals.

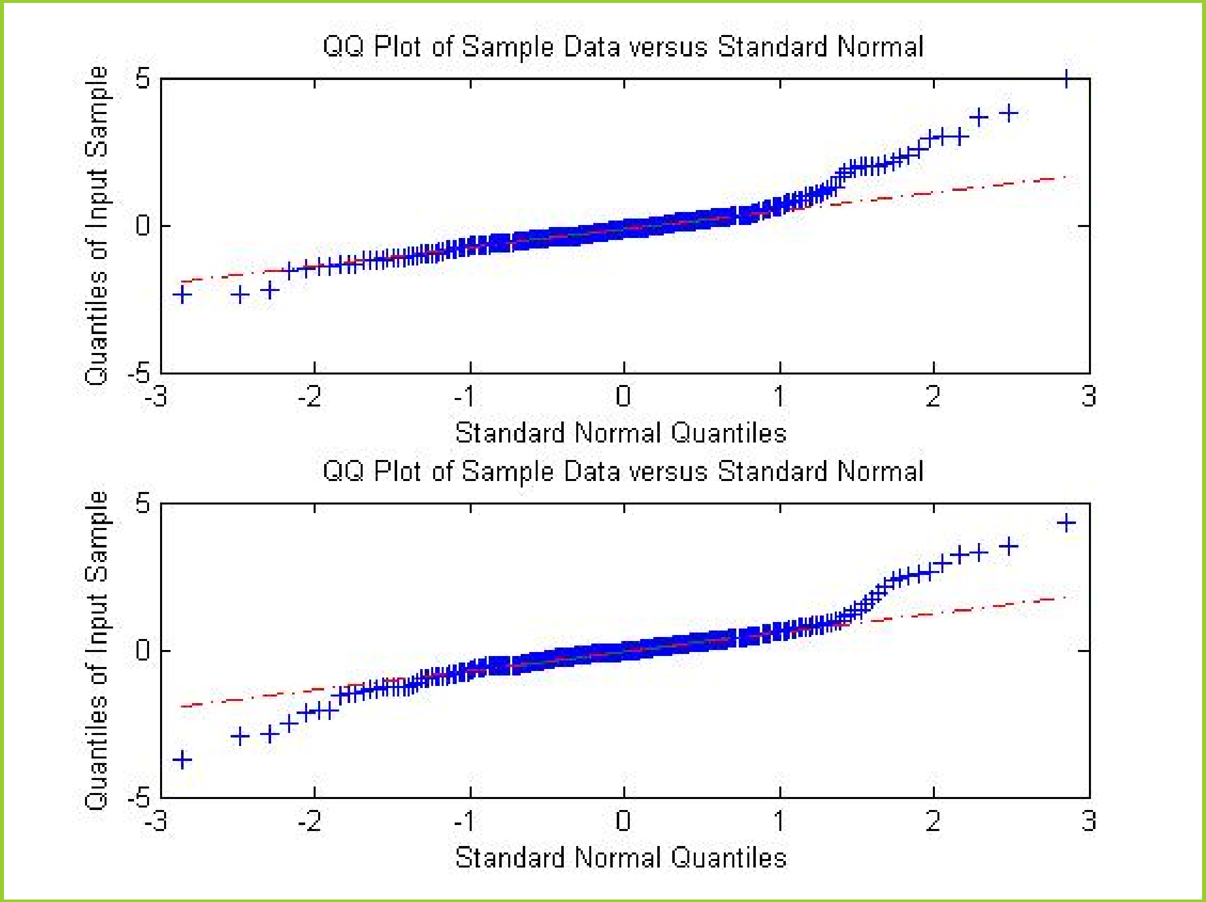


$\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2$, complete data.



$\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$, complete data.

- For the purpose of comparison with respect to conventional analysis, we delete the data of patient 227 and 207, although this might not be necessary.
- No more unusual observations are found in the plots.



227 and 207 are deleted.

Table 5 $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2$,

summaries of analyses for the eplieptics data

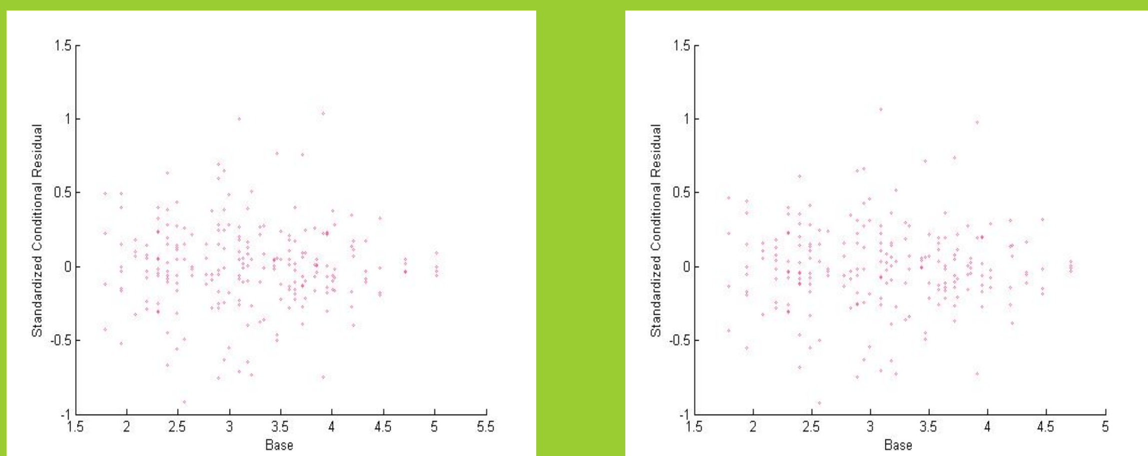
Parameter	$\hat{\boldsymbol{\beta}}_t$	a.s.d.	p value
$\boldsymbol{\beta}_0$	-1.2512	1.5482	0.4190
	(-1.1737)	(1.5307)	(0.4433)
$\boldsymbol{\beta}_T$	-0.8802	0.5372	0.1013
	(-0.7268)	(0.5555)	(0.1907)
$\boldsymbol{\beta}_A$	0.4899	0.4525	0.2790
	(0.4787)	(0.4469)	(0.2841)
$\boldsymbol{\beta}_B$	0.8743	0.1736	4.7735×10^{-7}
	(0.7975)	(0.1747)	(4.9617×10^{-6})
$\boldsymbol{\beta}_V$	-0.1481	0.2904	0.6099
	(-0.3747)	(0.2946)	(0.2034)
$\boldsymbol{\beta}_{T*B}$	0.3383	0.2733	0.2158
	(0.2706)	(0.2916)	(0.3536)
$\boldsymbol{\beta}_{V_4}$	-0.1015	0.1547	0.5116
	(0.0302)	(0.1571)	(0.8476)

Table 6 $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$,

summaries of analyses for the eplieptics data

Parameter	$\hat{\boldsymbol{\beta}}_t$	a.s.d.	p value
$\boldsymbol{\beta}_0$	-1.2913	0.7094	0.0687
	(-1.1983)	(0.4631)	(0.0097)
$\boldsymbol{\beta}_T$	-0.8958	0.1190	5.1292×10^{-14}
	(-0.7388)	(0.1606)	(4.2067×10^{-6})
$\boldsymbol{\beta}_A$	0.4888	0.0486	0
	(0.4758)	(0.0583)	(4.4409×10^{-16})
$\boldsymbol{\beta}_B$	0.8807	0.0294	0
	(0.8046)	(0.0285)	(0)
$\boldsymbol{\beta}_V$	-0.2913	5.0888	0.9544
	(-0.4372)	(2.8956)	(0.8800)
$\boldsymbol{\beta}_{T*B}$	0.3419	0.0410	0
	(0.2717)	(0.0636)	(1.9626×10^{-5})
$\boldsymbol{\beta}_{V_4}$	-0.0069	2.7286	0.9980
	(0.0866)	(1.5422)	(0.9553)

- At last, we give the scatter plots of standardized conditional residuals, which is defined by $[y_{ij} - \hat{u}_i\hat{\mu}_{ij}]/[\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta})\hat{\mu}_{ij}\hat{u}_{ij}]$, for the complete data and the data with patients 227 and 207 deleted respectively in Figure 5 and Figure 6 for model 2 on the log baseline counts at four occasion. The figures show the dispersion structure is needed.



For the complete data   For the data with 227 and 207 deleted

Residual scatter plots of Model 2

Conclusions for P-G Type Models:

- Good convergence rate based on the order to infinity of $X'X$
- Asymptotic normality leads to validity of inferences based on normal distribution
- Simulation shows good behaviors for moderate sample sizes
- Easy implemented
- The conclusions have been proved for typical conjugate HGLMs (constant dispersion and random dispersion)

5. Discussions

- Bias: the estimators of random effects might suffer from bias. This bias may leads to the estimating equation biased.
- Efficiency: still unknown, especially for hypothesis test.

- Further theoretical studies needed. Very important!

  A lot of examples: some intuitive method wrong

- Too large to make a unified theoretical studies.

Thanks !