W07 Wed. 1018

cf { latent variable.
latent class, strata.

Topic 5. Random Effects Models (GLM)

Random effects GLM -

elegant design

cf

ui 未知

(a) $Y_{ij} \mid (x_{ij}, u_i)\text{'s} \overset{independent}{\sim} f_Y(y_{ij} \mid x_{ij}, u_i, \beta, \phi),\ j = 1, \cdots, m_i.$

strong assumption

(b) $U_i\text{'s} \overset{iid}{\sim} f_U(u \mid G)$.

zipcode, latent

eg. family effect, repeating measurement.

model 的異質性的來源, 承認 latent 的存在, 但是先不處理,

effects

Estimation for generalized linear mixed models -

Conditional maximum likelihood estimation:

To simplify the discussion, let $a(\phi) = 1$ and $\theta_{ij} = x_{ij}^T \beta + d_{ij}^T u_i$.

Treating $\underset{n\times q}{U} = (U_1, \cdots, U_n)^T$ as fixed, the likelihood function for β, ϕ, and U is

$$L(\beta, \phi, U) = \prod_{i=1}^n \prod_{j=1}^{m_i} f_Y(y_{ij} \mid u_i, \beta) = \prod_{i=1}^n \prod_{j=1}^{m_i} \exp\{\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{a(\phi)} + c(y_{ij}; \phi)\}$$

上下互消的結構.

$$\propto \exp(\beta^T \sum_{i=1}^n T_{1i}(\underset{\sim}{y_i}) + \sum_{i=1}^n u_i^T T_{2i}(\underset{\sim}{y_i}) - \sum_{i=1}^n \sum_{j=1}^{m_i} \psi(\theta_{ij})),$$

看不到 看得到

where $T_{1i}(\underset{\sim}{y_i}) = \sum_{j=1}^{m_i} x_{ij} y_{ij}$ and $T_{2i}(\underset{\sim}{y_i}) = \sum_{j=1}^{m_i} d_{ij} y_{ij}$.

看不明白.

It implies that

conditional likelihood

$$f_{y_i|T_{2i}}(y_{i1}, \cdots, y_{im_i} \mid t_{2i}, \beta) = \frac{f_{y_i, T_{2i}}(y_{i1}, \cdots, y_{im_i}, t_{2i} \mid \beta)}{f_{T_{2i}}(t_{2i} \mid \beta)} = \frac{\exp(\beta^T T_{1i}(\underset{\sim}{y_i}) + U_i^T t_{2i})}{\sum_{\{y_i : T_{2i} = t_{2i}\}} \exp(\beta^T T_{1i}(\underset{\sim}{y_i}) + U_i^T t_{2i})}$$

clustering 的 β

$$= \frac{\exp(\beta^T T_{1i}(\underset{\sim}{y_i}))}{\sum_{\{y_i : T_{2i} = t_{2i}\}} \exp(\beta^T T_{1i}(\underset{\sim}{y_i}))}$$

僅和 fix effect 的 β 有关.

or $f_{T_{1i}|T_{2i}}(t_{1i} \mid t_{2i}, \beta) = \dfrac{\sum_{\{y_i : T_{1i} = t_{1i}, T_{2i} = t_{2i}\}} \exp(\beta^T t_{1i})}{\sum_{\{y_i : T_{2i} = t_{2i}\}} \exp(\beta^T T_{1i}(\underset{\sim}{y_i}))}, i = 1, \cdots, n.$

Thus, the conditional likelihood for β is defined to be the maximizer of

$$L_c(\beta) = \prod_{i=1}^n \frac{\sum_{\{y_i : T_{1i} = t_{1i}, T_{2i} = t_{2i}\}} \exp(\beta^T t_{1i})}{\sum_{\{y_i : T_{2i} = t_{2i}\}} \exp(\beta^T T_{1i}(\underset{\sim}{y_i}))}, \text{ i.e. } \hat{\beta}_c = \arg\max_\beta L_c(\beta).$$

Maximum likelihood estimation:

latent variable　　multi-dim Covariance matrix.

Assume further that $u_i \sim N_q(0,G)$.

Let $\delta = (\beta, G)$. The likelihood function of δ is

q維的 integration.

$$L(\delta | \underset{\sim}{y}) = \prod_{i=1}^{n} \int (\prod_{j=1}^{m_i} f(y_{ij} | x_{ij}, u_i)) f_U(u_i | G) du_i$$

要考慮計算成本.

$$\propto \prod_{i=1}^{n} \int (\prod_{j=1}^{m_i} \exp(\beta^T x_{ij} y_{ij} + u_i^T d_{ij} y_{ij} - \psi(\theta_{ij}))) |G|^{-\frac{1}{2}} \exp(\frac{-u_i^T G^{-1} u_i}{2}) du_i.$$

The score function for β, based on the complete data $(\underset{\sim}{y}, U)$, is

link function, response function

$$S_\beta(\delta | \underset{\sim}{y}, U) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} x_{ij}(y_{ij} - \mu_{ij}(u_i)) = 0, \text{where } \mu_{ij}(u_i) = h^{-1}(x_{ij}^T \beta + d_{ij}^T u_i).$$

Similarly, the score function for G is $S_G(\delta | \underset{\sim}{y}, U) = \frac{1}{2}[G^{-1}(\sum_{i=1}^{n} u_i u_i^T) G^{-1} - nG^{-1}] = 0$.

HW04 Q1
對 G 微分得到

The observed score functions are then defined to be

$$\begin{cases} S_\beta(\delta | \underset{\sim}{y}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} x_{ij}(y_{ij} - E[\mu_{ij}(u_i) | y_{ij}]) \cancel{=0} \\ S_G(\delta | \underset{\sim}{y}) = \frac{1}{2}[G^{-1} \sum_{i=1}^{n} E[u_i u_i^T | y_{ij}] G^{-1} - nG^{-1}] \cancel{=0} \end{cases}.$$

Methods to solve for the MLE of δ:

Expectation – maximization

(1) EM algorithm (Dempster et al.(1977))

(2) Monte Carlo integration methods. (For example, Gibbs sampling (See Appendix).)

(3) Approximated score functions without computing the integrations. (The central idea of the approximated score functions is to use conditional modes rather than conditional means and approximate the conditional distribution of U_i given $\underset{\sim}{y}_i$ via a Gaussian distribution with the same mode and curvature.)

mean

(3*)Let the surrogate response variable $Z_{ij} = h(\mu_{ij}) + h'(\mu_{ij})(y_{ij} - \mu_{ij}) = x_{ij}^T \beta + d_{ij}^T u_{ij} +$ ①

② $h'(\mu_{ij})(y_{ij} - \mu_{ij})$ with $y_{ij} - \mu_{ij} \triangleq \varepsilon_{ij}$. It implies that $Z_i = (z_{i1}, \cdots, z_{im_i})^T \sim (X_i \beta, V_i)$, where

mean

$$X_i = \begin{pmatrix} x_{i1}^T \\ x_{i2}^T \\ \vdots \\ x_{im_i}^T \end{pmatrix} \text{and } V_i = D_i G D_i^T + Q_i \text{with } D_i = \begin{pmatrix} d_{i1}^T \\ d_{i2}^T \\ \vdots \\ d_{im_i}^T \end{pmatrix}, Q_i = Diag(E[V(y_{ij} | u_i)(h'(\mu_{ij}))^2]).$$

from ①, ②

Z_{ij}

$h(Y_{ij}) \approx h(\mu_{ij}) + h'(\mu_{ij})(y_{ij} - \mu_{ij})$

first order expansion

covariance matrix

Approximate $\begin{pmatrix} Z_i \\ u_i \end{pmatrix}$ by $N_{m_i+q}\left(\begin{pmatrix} X_i\beta \\ 0 \end{pmatrix}, \begin{pmatrix} V_i & D_iG \\ GD_i^T & G \end{pmatrix}\right)$, one has $E[u_i \mid z_i] = GD_i^TV_i^{-1}(z_i - x_i^T\beta)$ and $V[u_i|z_i] = G - GD_i^TV_i^{-1}D_iG$.

For ~~fixed~~ given $\hat{G}^{(k)}$, updated values of β and U are obtained by iterative solving

$\hat{\beta}^{(k+1)} = (\sum_{i=1}^n X_i^T \hat{V}_i^{(k)-1} X_i)^{-1}(\sum_{i=1}^n X_i^T \hat{V}_i^{(k)-1} \hat{Z}_i^{(k)})$ and $\hat{u}_i^{(k+1)} = \hat{G}^{(k)} D_i^T \hat{V}_i^{(k)-1}(\hat{z}_i^{(k)} - x_i\hat{\beta}^{(k)})$.

Using the score function $S_G(\delta \mid \underset{\sim}{y}) = \frac{1}{2}[G^{-1}\sum_{i=1}^n E[u_iu_i^T \mid y_{ij}]G^{-1} - nG^{-1}] = 0$, $\hat{G}^{(k+1)}$ can be ~~estimated by~~ obtained as

u 的 cond. second moment

$\hat{G}^{(k+1)} = \frac{1}{n}\sum_{i=1}^n E[u_iu_i^T|\underset{\sim}{y}_i] = \frac{1}{n}\sum_{i=1}^n E[u_i|\underset{\sim}{y}_i](E[u_i|\underset{\sim}{y}_i])^T + \frac{1}{n}\sum_{i=1}^n V[u_i|\underset{\sim}{y}_i]$ ~~with~~ $E[u_i|\underset{\sim}{y}_i]$ ~~and~~ $= \hat{u}_i^{(k)}$,

$V[u_i|\underset{\sim}{y}_i]$ ~~being separated u_i and~~ $= \hat{G}^{(k)} - \hat{G}^{(k)} D_i^T \hat{V}_i^{(k)-1} D_i \hat{G}^{(k)} \cancel{= (D_i^T Q_i^{-1} D_i + G^{-1})^{-1}}$.

Logistic regression for binary responses -

$L(\beta, Y_1, \ldots, Y_n) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} P(Y_{ij}=1 | U_i, X_i)(1 - P(Y_{ij}=1 | U_i, X_i))$

Consider $\operatorname{logit} P(Y_{ij} = 1 \mid U_i) = \beta_0 + U_i + X_{ij}^T\beta$: random intercept logistic model.

Let $r_i = \beta_0 + U_i$. The joint likelihood function for β and r_i is

$L(\beta, r_1, \cdots, r_n) = \prod_{i=1}^n \exp[r_i\sum_{j=1}^{m_i} y_{ij} + (\sum_{j=1}^{m_i} y_{ij}x_{ij}^T)\beta - \sum_{j=1}^{m_i}\ln(1 + \exp(r_i + x_{ij}^T\beta))]$.

The conditional likelihood, which is equivalent to that derived in stratified

Outcome = 1 or 0

case-control studies, is $L_c(\beta) = \prod_{i=1}^n \left[\dfrac{\exp(\sum_{j=1}^{m_i} y_{ij}x_{ij}^T\beta)}{\sum_{\{y_i : \sum_{l=1}^{m_i} y_{il} = t_{i2}\}} \exp(\sum_{l=1}^{m_i} y_{il}x_{ij}^T\beta)}\right]$.

Example: 2×2 cross-over trial

Group	(1,1)	(1,0)	(0,1)	(0,0)
placebo-treatment (0,1)	a_1	b_1	c_1	d_1
treatment-placebo (1,0)	a_2	b_2	c_2	d_2

Show: $L_c(\beta) = \left(\dfrac{\exp(\beta_1)}{1 + \exp(\beta_1)}\right)^{b_2 + c_1}\left(\dfrac{1}{1 + \exp(\beta_1)}\right)^{b_1 + c_2}$.

HW04. Q2

$m_i = 2$

Remark.

(1) Conventionally, zero cell is replaced with 0.5 in calculation.

(2) $(a_1 + d_1 + a_2 + d_2)$ pairs are uninformative. Consequently, standard errors of regression estimates tend to be larger than in a marginal or random effects analysis.

Beta-binomial distribution:

$u_i \in [0,1]$

(1) $Y_{ij}\text{'s} \mid u_i \overset{iid}{\sim} Bernoulli(u_i),\ j = 1, \cdots, m_i.$

(2) $u_i\text{'s} \overset{iid}{\sim} Beta(a,b)$ with $E[u_i] = \dfrac{a}{a+b} \triangleq \mu$ and $V[u_i] = \mu(1-\mu)\dfrac{1}{a+b+1} \triangleq \mu(1-\mu)\delta$.

Let $Y_i = Y_{i1} + \cdots + Y_{im_i}, i = 1, \cdots, n.$

$$P(Y_i = y \mid m_i) = \binom{m_i}{y} \int_0^1 \frac{u_i^{\,y}(1-u_i)^{m_i - y} u_i^{\,a-1}(1-u_i)^{b-1}}{B(a,b)} du_i$$

$$= \binom{m_i}{y} \frac{\prod_{i=0}^{y-1}(a+i)\prod_{i=0}^{m_i-y-1}(b+i)}{\prod_{i=0}^{m_i-1}(a+b+i)} = \binom{m_i}{y} \frac{\prod_{i=0}^{y-1}(\mu+ri)\prod_{i=0}^{m_i-y-1}(1-\mu+ri)}{\prod_{i=0}^{m_i-1}(1+ri)},$$

where $a = \mu r^{-1}, b = (1-\mu)r^{-1}$, and $r = \dfrac{\delta}{1-\delta}$.

It implies that $\mu + (m_i - 1)r \ge 0$ and $(1-\mu) + (m_i - 1)r \ge 0$,

$$\text{or } r \ge \max\{\frac{-\mu}{(m_i - 1)}, \frac{-(1-\mu)}{(m_i - 1)}\}$$

$$\text{or } \delta \ge \max\{\frac{-\mu}{(m_i - \mu - 1)}, \frac{-(1-\mu)}{(m_i + \mu)}\}$$

$$\ge \max\{\frac{-\mu}{(n_0 - \mu - 1)}, \frac{-(1-\mu)}{(n_0 + \mu)}\}, \text{ where } n_0 = \min\{m_1, \cdots, m_n\}.$$

Rosaer (1984) extended the beta-binomial to allow the covariates to vary within clusters as

不合理.

$$\text{logit} := \ln\frac{P(Y_{ij}=1|\cdots)}{1-P(Y_{ij}=1|\cdots)}$$

$$\log it P(Y_{ij} = 1 \mid y_{i1}, \ldots, y_{ij-1}, y_{ij+1}, \ldots, y_{im_i}, x_{ij}) = \log(\frac{\theta_{i1} + (y_i - y_{ij})\theta_{i2}}{1 - \theta_{i1} + ((m_i - 1) - (y_i - y_{ij}))\theta_{i2}}) + x_{ij}^{\ T}\beta,$$

$j = 1, \cdots, m_i$, where θ_{i1} is the intercept parameter and θ_{i2} characterizes the association between pairs of response for the same cluster.

limitations: β measures the effect of x_{ij} on Y_{ij}, which cannot first be explained by the other responses in the cluster.

Logistic models with Gaussian random effects:

$$L(\beta,G|\underset{\sim}{y}) \propto \prod_{i=1}^{n} \int \exp[\beta^T \sum_{j=1}^{m_i} x_{ij}y_{ij} + u_i^T \sum_{j=1}^{m_i} d_{ij}y_{ij} - \sum_{j=1}^{n} \ln(1+\exp(x_{ij}^T\beta + d_{ij}^T u_i))]$$

$$\cdot |G|^{\frac{-q}{2}} \exp(\frac{-u_i^T G^{-1} u_i}{2}) du_i. \qquad (\beta,G) = \underset{(\beta,G)}{\operatorname{argmax}}\, L(\beta,G|\underset{\sim}{y}).$$

Counted responses -

Consider $\ln(E[Y_{ij}|U_i]) = \beta_0 + U_i + x_{ij}^T\beta + \ln(t_{ij})$: random intercept log-linear model

for count data where $\beta_0 + U_i \triangleq r_i$.

The conditional likelihood approach:
The joint likelihood function for β and $(r_1,\cdots,r_n)$ is

$$L(\beta,r_1,\cdots,r_n) = \prod_{i=1}^{n} \exp\{r_i \sum_{j=1}^{m_i} y_{ij} + \beta^T \sum_{j=1}^{m_i} y_{ij}x_{ij} + \sum_{j=1}^{n_i} y_{ij}\ln(t_{ij}) - \sum_{j=1}^{m_i} t_{ij}\exp(r_i + x_{ij}^T\beta)\}.$$

The conditional likelihood is

$$L_c(\beta) = \prod_{i=1}^{n} \frac{\exp(\beta^T \sum_{j=1}^{m_i} y_{ij}x_{ij} + \sum_{j=1}^{m_i} y_{ij}\ln(t_{ij}))}{\sum_{\{\sum_{j=1}^{m_l} y_{lj} = y_i\}} \exp(\beta^T \sum_{j=1}^{m_l} y_{lj}x_{lj} + \sum_{j=1}^{m_l} y_{lj}\ln(t_{lj}))}.$$

Example:

HW.04 Q3

(1) Y_{ij}'s $|u_i \overset{indep.}{\sim} Poisson(u_i)$. (2) $u_i \overset{iid}{\sim} Gamma(\alpha,\beta)$, where $\alpha\beta \triangleq \mu$ and $\alpha\beta^2 \triangleq \phi\mu^2$.

show

It implies that Y_{ij} is Negative-binomial with $E[Y_{ij}] = \mu$ and $V[Y_{ij}] = \mu + \phi\mu^2$.

沒有考慮 Covariance

(Extension 1)

(1) Y_{ij}'s $|x_{ij},u_i \overset{independent}{\sim} (u_i\exp(x_{ij}^T\beta), \phi u_i \exp(x_{ij}^T\beta))$, where U_i is a latent variable.

(2) $U_i \overset{iid}{\sim} (1,\sigma^2)$.

(Extension 2)

(1) Y_{ij}'s $|x_{ij},u_i \overset{independent}{\sim} Poisson(x_{ij}^T\beta + d_{ij}^T u_i)$

(2) U_i's $\overset{iid}{\sim} f(u_i|G)$

線性，但不合理？
会有負值？

Appendix

Gibbs Sampler – (A Monte Carlo method for estimating the desired posterior distributions)

<u>Premise</u>: Consider three variables (U, V, W) and the conditional distributions of each given the remainder has a simple form while the joint distribution is more complicated.

Let $[U, V, W]$ represent the joint distribution, and $[U \mid V, W]$, $[V \mid U, W]$, and $[W \mid U, V]$ denote the conditional distributions.

The Gibbs Sampler is a method for generating a random variable from $[U, V, W]$ as below.

Step 0: Given arbitrary starting values $U^{(0)}, V^{(0)}, W^{(0)}$.

Step 1: Generate $U^{(1)} \sim [U \mid V^{(0)}, W^{(0)}]$, $V^{(1)} \sim [V \mid U^{(1)}, W^{(0)}]$, and $W^{(1)} \sim [W \mid U^{(1)}, V^{(1)}]$.

$$\vdots$$

Step B :Generate $U^{(B)} \sim [U \mid V^{(B-1)}, W^{(B-1)}]$, $V^{(B)} \sim [V \mid U^{(B)}, W^{(B-1)}]$, and $W^{(B)} \sim [W \mid U^{(B)}, V^{(B)}]$.

靠近真实的分配

Under some regularity conditions, Geman and Geman (1984) showed that $[U^{(B)}, V^{(B)}, W^{(B)}] \xrightarrow{d} [U, V, W]$ at an exponential rate as $B \to \infty$.

The distribution $[U, V, W]$ can be approximated by the empirical distribution of the M values $[U^{(B+k)}, V^{(B+k)}, W^{(B+k)}]$, $k = 1, \ldots, M$, where B is large enough and M is chosen to give sufficient precision to the empirical distribution of interest.

Gibbs Sampling approach for GLM with random effects -

Consider $f(y_{ij} \mid u_i, \beta) = e^{\left(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{a(\phi)} - c(y_{ij};\phi)\right)}$ with $g(u_i \mid G) = (2\pi)^{\frac{-q}{2}} \mid G \mid^{\frac{-1}{2}} e^{\frac{-u_i^T G^{-1} u_i}{2}}$, and $h(\mu_{ij}) = x_{ij}^T \beta + z_{ij}^T u_i$.

The likelihood function of (β, G) is

$$L(\beta, G \mid \underset{\sim}{y}) \propto \prod_{i=1}^{n} \int \prod_{j=1}^{m_i} f(y_{ij} \mid u_i, \beta) \mid G \mid^{\frac{-1}{2}} \exp\left(\frac{-u_i^T G^{-1} u_i}{2}\right) du_i.$$

視為隨機的

In a Bayesian approach to analyzing the random effects GLM, the parameters (β, G) are random variables and are treated symmetrically with the longitudinal measurements and unobserved latent variables. Thus, the random effects GLM is an example of a hierarchical Bayes model.

conditional dis.
看不到

Assumptions: $[\beta \mid G, U, \underset{\sim}{y}] = [\beta \mid U, \underset{\sim}{y}]$, $[G \mid \beta, U, \underset{\sim}{y}] = [G \mid U]$ and $[U \mid \beta, G, \underset{\sim}{y}]$.

1. Assume that β has a flat prior function. Then,

$[\beta \mid U^{(k)}, \underset{\sim}{y}] \propto \prod_{i=1}^{n} \prod_{j=1}^{m_i} f(y_{ij} \mid U_i^{(k)}, \beta) \approx N(\beta^{(k)}, V_\beta^{(k)})$, as $n \to \infty$, where $\beta^{(k)}$ is the

逼近

maximum likelihood estimator and $V_\beta^{(k)}$ is the inverse of the Fisher information.

Adjustment for smaller samples - "Rejection sampling" (Ripley, 1987)

Let $f(\beta \mid U^{(k)}, \underset{\sim}{y})$ and $\phi(\beta \mid \beta^{(k)}, V_\beta^{(k)})$ denote separately the true density and Gaussian density. Choose a constant $c \geq 1$ such that $c\phi(\beta \mid \beta^{(k)}, V_\beta^{(k)}) \geq f(\beta \mid U^{(k)}, \underset{\sim}{y})$.

uniform

Step1: Generate $\beta^* \sim \phi(\beta \mid \beta^{(k)}, V_\beta^{(k)})$ and $u \sim U(0,1)$.

Step2: If $\dfrac{f(\beta^* \mid b^{(k)}, \underset{\sim}{y})}{c\phi(\beta^* \mid \beta^{(k)}, V_\beta^{(k)})} < u$, $\beta^{(k+1)} = \beta^*$. Otherwise, the process returns to Step1.

2. $[G \mid U^{(k)}]$

Assume that $\pi(G) \propto |G|^{-1}$: non-informative prior (see Box and Tiao, (1973)). Then,

$[G \mid U^{(k)}] \sim$ Inverted Wishart$(S^{(k)}, n-q+1)$, where $S^{(k)} = \sum_{i=1}^{n} U_i^{(k)} U_i^{(k)T}$.

Remark:

If $A \sim$ Wishart$(\Sigma_{p\times p}, n)$, the p.d.f of A is $f_{\mathrm{A}}(\mathrm{A}) \propto |\mathrm{A}|^{\frac{-1}{2}(n-p-1)} e^{\frac{-1}{2} tr \Sigma^{-1} \mathrm{A}}$. It implies that

$\mathrm{B} = \mathrm{A}^{-1} \sim$ Inverted Wishart(Σ^{-1}, n) with p.d.f. $f_{\mathrm{B}}(\mathrm{B}) \propto |\mathrm{B}|^{\frac{-1}{2}(n+p+1)} e^{\frac{-1}{2} tr \Sigma^{-1} \mathrm{B}^{-1}}$. Thus,

$\pi(G \mid U^{(k)}) \propto |G|^{\frac{-1}{2}(n+2)} e^{\frac{-1}{2} tr(S^{(k)} G^{-1})}$, i.e., $[G \mid U^{(k)}] \sim$ Inverted Wishart$(S^{(k)}, n-q+1)$.

3.$[U|\beta^{(k)},G^{(k)},\underset{\sim}{y}]$

Using $f(U_i|\hat{\beta}^{(k)},G^{(k)},\underset{\sim}{y}_i) \propto f(\underset{\sim}{y}_i\,|U_i,\hat{\beta}^{(k)})g(U_i\,|\,G^{(k)}) \triangleq f_n(U_i\,|\,\hat{\beta}^{(k)},G^{(k)},\underset{\sim}{y}_i)$, we can find the mode and curvature of $f_n(U_i\,|\,\hat{\beta}^{(k)},G^{(k)},\underset{\sim}{y}_i)$, which matches a Gaussian density.

Using the surrogate response $Z_i^* = X_i\beta + D_iU_i + Diag(h'(\mu_i))(\underset{\sim}{y}_i - \mu_i)$, the maximum value of $f_n(U_i\,|\,\hat{\beta}^{(k)},G^{(k)},\underset{\sim}{y}_i)$ occurs at $U_i = (D_i^T Q_i^{-1} D_i + G^{(k)^{-1}})^{-1} D_i^T Q_i^{-1}(Z_i^* - X_i\beta^{(k)})$ $= G^{(k)}D_i(D_iG^{(k)}D_i^T + Q_i)^{-1}(Z_i^* - X_i\hat{\beta}^{(k)})$ and its curvature is $V_i = (D_i^T Q_i^{-1} D_i + G^{(k)^{-1}})^{-1}$.

Similar to the method in (3), $U_i^{(k)}$ can be obtained.