Chin. Ann. Math.
26B:3(2005),467–474.

ASYMPTOTIC NORMALITY OF QUASI MAXIMUM LIKELIHOOD ESTIMATE IN GENERALIZED LINEAR MODELS***

YUE Li* CHEN Xiru**

Abstract

For the Generalized Linear Model (GLM), under some conditions including that the specification of the expectation is correct, it is shown that the Quasi Maximum Likelihood Estimate (QMLE) of the parameter-vector is asymptotic normal. It is also shown that the asymptotic covariance matrix of the QMLE reaches its minimum (in the positive-definte sense) in case that the specification of the covariance matrix is correct.

Keywords Quasi likelihood estimate, Generalized linear model, Asmptotically normal, Asymptotic normality
2000 MR Subject Classification 62J05

§ 1 . Introduction and Main Results

Much studies have been made for the MLE and QMLE of the parameters of GLM (see [1–6]), among others. Let $Y_1, \cdots, Y_n$ be independent observations of the q-dimensional response vector Y, where Y_i obeys an exponential distribution of the form

$$\exp(y'\theta - b(\theta))d\mu(y)|_{\theta=\theta^{(i)}},$$

where $\theta = (\theta_1, \cdots, \theta_q)'$ and

$$\theta^{(i)} = \dot{b}^{-1}(h(Z_i'\beta_0)).$$

Here $\dot{b}(\theta) = \Big(\dfrac{\partial b}{\partial \theta_1}, \cdots, \dfrac{\partial b}{\partial \theta_q}\Big)'$, $\dot{b}^{-1}$ is its inverse. $h(t) \equiv h(t_1, \cdots, t_q) \equiv (h_1(t), \cdots, h_q(t))'$ is the inverse of link function g, and $Z_i = Z(X_i)$ is a known $p \times q$ matrix generated by the i-th observation X_i of independent variable X. β_0 is the true value of the unknown p-vector parametre β. With these notations, the log-likelihood equation assumes the form

$$\sum_{i=1}^{n} Z_i H(Z_i'\beta)\Sigma^{-1}(Z_i'\beta)(y_i - h(Z_i'\beta)) = 0, \tag{1.1}$$

Manuscript received December 23, 2003.
*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China.
E-mail: yueli8024@sina.com
**Graduate School of the Chinese Academy of Sciences, Beijing 100049, China.
E-mail: xczhu@public.bta.net.cn
***Project supported by the National Natural Science Foundation of China.

where $\Sigma(t) = \ddot{b}(\dot{b}^{-1}(h(t)))$, $\ddot{b}(\theta)$ is the covariance matrix of the distribution $\exp(y'\theta - b(\theta)d\mu(y))$ and its (j,k)-element is $\dfrac{\partial b}{\partial\theta_j\partial\theta_k}$, $j,k=1,\cdots,q$, $H(t)=(\dot{h}_1(t),\cdots,\dot{h}_q(t))$.

Since Equation (1.1) only involves the expectation and covariance matrix of Y, this suggests us to construct Equation (1.1) and try to use its root to estimate β_0 in case that we do not know the distribution of Y_i but know its expectation and covariance matrix. This is the so-called QMLE. Further investigation reveals that it is even not necessary to know the covariance matrix of Y_i, we can just replace it by a suitably-chosen matrix. Thus we may consider a more general form of the quasi-likelihood equation as follows

$$L_n(\beta) \equiv \sum_{i=1}^n Z_i H(Z_i'\beta)\Lambda_i(\beta)(y_i - h(Z_i'\beta)) = 0, \tag{1.2}$$

where $\Lambda_i(\beta)$ is a suitably-chosen positive definite q-matrix to replace $\Sigma^{-1}(Z_i'\beta)$ in (1.1).

The first problem is to confirm that there exists a root $\hat{\beta}_n$ of (1.2). One method in dealing with this problem is to construct a so-called Quasi-likelihood function $U(\beta)$ such that $\dot{U}(\beta)$ is just the right-hand side of (1.2). Then if $\hat{\beta}_n$ is a local maximum point of U, $\hat{\beta}_n$ must be a root of (1.2). Hence we need only to verify that there exists such a local maximum point in the vicinity of the true parameter β_0. This idea was first advanced in [9] for the case $q=1$. [5] extends the method for $q>1$ by choosing an exponential distribution $\exp(y'\theta - b(\theta))d\mu(y)$, link function $g=h^{-1}$ and $(\ddot{b}(\dot{b}^{-1}(h(\cdot))))^{-1}$ for $\Lambda_i(\beta)$ in (1.2). This approach has the merit in that we may borrow the method developed in case where Y_i has an exponential distribution. But in so doing we lose in some degree the freedom of choosing $\Lambda_i(\beta)$ in (1.2), and this may result in loss of efficiency of the QMLE.

To avoid this difficulty, we proposed in [7] a new method in proving the existence of a solution of (1.2), which does not resort to constructing a quasi likelihood function, and thus maintains completely the freedom of choice of $\Lambda_i(\beta)$ in (1.2). The strong consistency was studied in [7]. In this paper we shall prove the asymptotic normality of the solution. The main results are formulated in the following two theorems.

Theorem 1.1. *Suppose that the following conditions are satisfied:*

(1) $\{Z_i, i\ge 1\}$ *is bounded.* $\underline{\lambda}_n \ge cn^\delta$ *for n sufficiently large and some* $\delta\in(4/5,1]$, $c>0$, $\underline{\lambda}_n$ *and* $\overline{\lambda}_n$ *are the smallest and largest eigenvalues of* $S_n=\sum_{i=1}^n Z_iZ_i'$ *respectively.*

(2) $\sup_{i\ge 1} E\|Y_i\|^{\bar{p}}<\infty$, $\bar{p}=17/7$. $\mathrm{COV}(Y_i)>cI$, $i\ge 1$ *for some* $c>0$.

(3) *The partial derivatives of* 2*-th orders of* $h_1,\cdots,h_q$ *exist and are continuous,* $\det(H(t)) \neq 0$.

(4) $\Lambda_i(\beta)>0$*, the* 2*-th partial derivatives of each element of* $\Lambda_i(\beta)$ *exist and are continuous. The element of* $\Lambda_i(\beta)$*, together with its* 1*-th and* 2*-th partial derivatives, is uniformly bounded on any bounded set of* β *for arbitrarily* $i\ge 1$.

Then with probability one (1.2) *has a solution* $\hat{\beta}_n$ *for n sufficiently large, and*

$$B_n^{-1/2}Q_n(\hat{\beta}_n-\beta_0)\xrightarrow{d} N(0,I_p), \tag{1.3}$$

where

$$B_n=\sum_{i=1}^n Z_iH_i\Lambda_i\Sigma_i\Lambda_i'H_i'Z_i', \qquad Q_n=\sum_{i=1}^n Z_iH_i\Lambda_iH_i'Z_i'$$

with $H_i = H(Z_i'\beta_0),\ \Lambda_i = \Lambda_i(\beta_0),\ \Sigma_i = \mathrm{COV}(y_i)$.

From (1.3), $\hat{\beta}_n - \beta_0$ has asymptotic covariance matrix $(Q_n B_n^{-1} Q_n)^{-1}$, which depends on the selection of Λ_i. We hope to choose Λ_i in such a way so that $Q_n B_n^{-1} Q_n$ is maximized. In this respect, we have the following theorem.

Theorem 1.2. *When* $\Lambda_i = \Sigma_i^{-1}, i \geq 1$, $Q_n B_n^{-1} Q_n$ *reaches maximum, i.e.,*

$$Q_n B_n^{-1} Q_n \leq \sum_{i=1}^{n} Z_i H_i \Sigma_i^{-1} H_i' Z_i'. \tag{1.4}$$

(1.4) *indicates that if we make a correct specification of the covariance matrix of* Y_i, *we obtain a QMLE of the parameters with minimum asymptotic covariance matrix, which means greater efficiency of the estimator.*

§ 2 . Proof of Theorem 1.1

Take unit vector λ. Denote

$$\xi_n = \lambda' B_n^{-1/2} L_n(\beta_0) = \sum_{i=1}^{n} \xi_{ni}, \quad \xi_{ni} = \lambda' B_n^{-1/2} Z_i H_i \Lambda_i e_i, \qquad 1 \leq i \leq n,$$

where $e_i = y_i - h(Z_i'\beta_0)$. It is easy to see that $E(\xi_{ni}) = 0,\ 1 \leq i \leq n$, and

$$\mathrm{Var}(\xi_n) = \sum_{i=1}^{n} \mathrm{Var}(\xi_{ni}) = \lambda' B_n^{-1/2} \sum_{i=1}^{n} Z_i H_i \Lambda_i \Sigma_i \Lambda_i H_i' Z_i' B_n^{-1/2} \lambda = \lambda'\lambda = 1.$$

Hence in order to prove

$$\xi_n \xrightarrow{d} N(0,1), \tag{2.1}$$

we only need to prove

$$g_n(\varepsilon) \equiv \sum_{i=1}^{n} E[|\xi_{ni}|^2 I(|\xi_{ni}| > \varepsilon)] \to 0, \quad \forall \varepsilon > 0, \qquad n \to \infty.$$

Let

$$e_i^* = \Sigma_i^{-1/2} e_i, \quad a_{ni}' = \lambda' B_n^{-1/2} Z_i H_i \Lambda_i \Sigma_i^{1/2}, \qquad 1 \leq i \leq n.$$

Then $\xi_{ni} = a_{ni}' e_i^*$. Let $\eta_{ni} = \xi_{ni}/\|a_{ni}\|$. Then $|\xi_{ni}| \leq \|a_{ni}\| \|e_i^*\|$. So we have

$$\|e_i^*\| \geq |\xi_{ni}|/\|a_{ni}\| = |\eta_{ni}|. \tag{2.2}$$

From the boundedness of Z_i, we know that $\{Z_i'\beta_0\}$ is bounded. By the assumption (4) there exists $c > 0$ such that $\Lambda_i \geq cI$, $i \geq 1$. Also from the boundedness of $\{Z_i'\beta_0\}$ and the assumption (3), we know that there exists $c > 0$ such that $|\det(H_i)| \geq c$ for $i \geq 1$. Moreover, $\Sigma_i = \mathrm{COV}(y_i) \geq cI$. Combining these facts, it follows that there exits a constant $c > 0$ such that $H_i \Lambda_i \Sigma_i \Lambda_i H_i' \geq cI,\ i \geq 1$. From this we know $B_n \geq c \sum\limits_{i=1}^{n} Z_i Z_i'$. This, together with the assumption (1), gives $\underline{\lambda}(B_n) \ge cn^{\delta}$. Hence $\lambda_{\min} B_n \to \infty$. By the assumptions (1), (2) and (4), we know that $\max\limits_{1\le i\le n} \|a_{ni}\| \to 0$, $n \to \infty$, and

$$\sum_{i=1}^{n} a'_{ni} a_{ni} = \lambda' B_n^{-1/2} \sum_{i=1}^{n} Z_i H_i \Lambda_i \Sigma_i \Lambda_i H'_i Z'_i B_n^{-1/2} \lambda = 1. \tag{2.3}$$

Denote

$$g_n(\varepsilon) = \sum_{i=1}^{n} E[|\xi_{ni}|^2 I(|\xi_{ni}| > \varepsilon)] = \sum_{i=1}^{n} \|a_{ni}\|^2 E[|\eta_{ni}|^2 I(|\eta_{ni}| > \varepsilon/\|a_{ni}\|)].$$

From the assumption (2), taking $\alpha = \bar{p} - 2 = 3/7$, we have

$$g_n(\varepsilon) \le \sum_{i=1}^{n} \|a_{ni}\|^2 \|a_{ni}\|^{\alpha} \varepsilon^{-\alpha} E[|\eta_{ni}|^{\bar{p}}].$$

This equation and (2.2), (2.3) give

$$g_n(\varepsilon) \le \max \|a_{ni}\|^{\alpha} \varepsilon^{-\alpha} \sup_{i\ge 1} E\|e_i^*\|^{\bar{p}} \to 0$$

for any given $\varepsilon > 0$. Hence we proved (2.1). Since (2.1) holds for any unit vector, we have

$$B_n^{-1/2} L_n(\beta_0) \xrightarrow{d} N(0, I). \tag{2.4}$$

Denote $H_{ni} = H(Z'_i \hat{\beta}_n)$, $\Lambda_{ni} = \Lambda_i(\hat{\beta}_n)$, $1 \le i \le n$. As $L_n(\hat{\beta}_n) = 0$, we have

$$\begin{aligned} L_n(\beta_0) &= L_n(\beta_0) - L_n(\hat{\beta}_n) \\ &= \sum_{i=1}^{n} Z_i (H_i \Lambda_i - H_{ni} \Lambda_{ni}) e_i + \sum_{i=1}^{n} Z_i H_{ni} \Lambda_{ni} (h(Z'_i \hat{\beta}_n) - h(Z'_i \beta_0)) \\ &= \sum_{i=1}^{n} Z_i (H_i \Lambda_i - H_{ni} \Lambda_{ni}) e_i + \sum_{i=1}^{n} Z_i H_{ni} \Lambda_{ni} \widetilde{H}'_{ni} Z'_i (\hat{\beta}_n - \beta_0) \\ &= \sum_{i=1}^{n} Z_i (H_i \Lambda_i - H_{ni} \Lambda_{ni}) e_i + \sum_{i=1}^{n} Z_i (H_{ni} \Lambda_{ni} \widetilde{H}'_{ni} - H_i \Lambda_i H'_i) Z'_i (\hat{\beta}_n - \beta_0) \\ &\quad + \sum_{i=1}^{n} Z_i H_i \Lambda_i H'_i Z'_i (\hat{\beta}_n - \beta_0) \equiv J_{1n} + J_{2n} + J_{3n}, \end{aligned} \tag{2.5}$$

where $\widetilde{H}_{ni} = (\dot{h}_1(Z'_1 \beta_{n1}), \cdots, \dot{h}_q(Z'_1 \beta_{nq}))$, $1 \le i \le n$, $\beta_{n1}, \cdots, \beta_{nq}$ lie all on the line-segment with end-points β_0 and $\hat{\beta}_n$, and $J_{3n} = Q_n(\hat{\beta}_n - \beta_0)$.

Now we proceed to show that

$$B_n^{-1/2} J_{kn} = o_p(1), \qquad k = 1, 2. \tag{2.6}$$

If (2.6) can be proved, then from (2.4)–(2.6) we obtain (1.3).

For a matrix $A = (a_{ij})$, denote $\max\limits_{i,j} |a_{ij}|$ by $|A|$. From the assumptions (3) and (4), we have

$$|H_{ni} \Lambda_{ni} \widetilde{H}'_{ni} - H_i \Lambda_i H'_i| = O_p(\hat{\beta}_n - \beta_0) \text{ accordingly for } 1 \le i \le n.$$

Hence, in view of the result $\hat{\beta}_n - \beta_0 = O_p(n^{-(\delta-1/2)})$, proved in [7], we have

$$J_{2n} = nO_p(\|\hat{\beta}_n - \beta_0\|^2) = O_p(n^{2-2\delta}). \tag{2.7}$$

Further, by $B_n \geq cS_n$ and the assumption (1), we have $|B_n^{-1/2}| = O(n^{-\delta/2})$. Thus from (2.11), we get $B_n^{-1/2}J_{2n} = O_p(n^{2-5\delta/2})$. Since $\delta > 4/5$, we get

$$B_n^{-1/2}J_{2n} = o_p(1). \tag{2.8}$$

Now consider J_{1n}. Take $r \in (2/(3\delta-1), 10/7)$, r sufficiently near $\bar{p}-1$. Let $\bar{e}_i = e_i I(\|e_i\| \leq i^{1/r}),\ i = 1, 2, \cdots$. Since $r < \bar{p} - 1 < \bar{p}$, we have

$$\sum_{i=1}^{\infty} P(\bar{e}_i \neq e_i) = \sum_{i=1}^{\infty} P(\|e_i\| > i^{1/r}) \leq \sup_{i\geq 1} E(\|e_i\|^{\bar{p}}) \sum_{i=1}^{\infty} i^{-\bar{p}/r} < \infty.$$

Hence from Borel-Cantelli lemma, we have $\bar{e}_n = e_n$ with probability one for n large enough. So instead of J_{1n} we need only to consider

$$\overline{J}_{1n}(\hat{\beta}_n) \equiv \sum_{i=1}^{n} Z_i(H_i\Lambda_i - H_i(\hat{\beta}_n)\Lambda_i(\hat{\beta}_n))\bar{e}_i. \tag{2.9}$$

Since $\|E(\bar{e}_i)\| = \|E(e_i - \bar{e}_i)\| \leq i^{-(\bar{p}-1)/r}K\ \big(K = \sup_{i\geq 1} E\|e_i\|^{\bar{p}}\big)$, we have

$$\Big\|\sum_{i=1}^{n} Z_i(H_i\Lambda_i - H_i(\hat{\beta}_n)\Lambda_i(\hat{\beta}_n))E\bar{e}_i\Big\| \leq c\sum_{i=1}^{\infty} K i^{-(\bar{p}-1)/r} < \infty.$$

Hence instead of $\overline{J}_{1n}(\hat{\beta}_n)$, we need only to consider

$$\widetilde{J}_{1n}(\hat{\beta}_n) \equiv \sum_{i=1}^{n} Z_i(H_i\Lambda_i - H_i(\hat{\beta}_n)\Lambda_i(\hat{\beta}_n))\tilde{e}_i, \tag{2.10}$$

where $\tilde{e}_i = \bar{e}_i - E\bar{e}_i$. We have $E(\tilde{e}_i) = 0,\ \sup_{i\geq 1} E\|\tilde{e}_i\|^{\bar{p}} < \infty,$ and $\sup_{1\leq i\leq n} |\tilde{e}_i| \leq 2n^{1/r}$.

Taking $a \geq 3/10$, we find that $M = [n^{pa}]$ points to the sphere $S \equiv \{\gamma : \|\gamma - \beta_0\| \leq cn^{-(\delta-1/2)}\}$. For any $\gamma \in S$, we can find j such that $\|\gamma_j - \gamma\| \leq cn^{-d},\ d = \delta - 1/2 + a$. Denote

$$\widetilde{J}_{1nj} \equiv \sum_{i=1}^{n} Z_i(H_i\Lambda_i - H_i(\gamma_j)\Lambda_i(\gamma_j))\tilde{e}_i \equiv \sum_{i=1}^{n} e_{ij}. \tag{2.11}$$

Consider its l-th element, denoted by $\widetilde{J}^l_{1nj} \equiv \sum_{i=1}^{n} e^l_{ij}$.

Now we make use of Bernstein inequality: Suppose that $X_i, \cdots, X_n$ are independent random variables, $EX_i = 0,\ 1 \leq i \leq n$, and there exists a finite constant b such that $|X_i| \leq b, 1 \leq i \leq n$. Then for any given $\varepsilon > 0$, we have

$$P\Big(\Big|\sum_{i=1}^{n} X_i/n\Big| \geq \varepsilon\Big) \leq 2\exp(-n\varepsilon^2/(2b\varepsilon + 2\bar{\sigma}^2)), \quad \bar{\sigma}^2 = \sum_{i=1}^{n} \mathrm{Var}(X_i)/n.$$

Since

$$P(|\widetilde{J}^l_{1nj}| \geq \varepsilon_0 n^{\delta/2}) = P\Big(\Big|\frac{1}{n}\sum_{i=1}^n e^l_{ij}\Big| \geq \varepsilon_0 n^{\delta/2-1}\Big), \tag{2.12}$$

we employ Bernstein inequality to the right-hand side of (2.12). Note that here we have $b = cn^{-(\delta-1/2)+1/r}$, $\bar{\sigma}^2 \leq cn^{-(2\delta-1)}$, $\varepsilon = \varepsilon_0 n^{\delta/2-1}$, $b\varepsilon = cn^{-(\delta/2-1/2)+1/r}$, $n\varepsilon^2 = cn^{\delta-1}$, where $c > 0$ is a constant which may assume different values in each of its appearance. From the choice of variable r, we have $(\delta-1)-(-(\delta/2+1/2)+1/r) > 0$. Also, $(\delta-1)-(-(2\delta-1)) = 3\delta - 2 > 0$. From these facts we see that for (2.12), $\frac{n\varepsilon^2}{2(b\varepsilon+\bar{\sigma}^2)} \geq cn^\alpha$ for some $c > 0$, $\alpha > 0$, where c and α do not depend on $j = 1, \cdots, M$ and $l = 1, \cdots, q$. This gives

$$P\Big(\max_{1\leq j\leq M} \|\widetilde{J}_{1nj}\| \geq \sqrt{q}\varepsilon_0 n^{\delta/2}\Big) \leq n^{pa}\exp(-cn^\alpha). \tag{2.13}$$

Since $\sum\limits_{n=1}^{\infty} n^{pa}\exp(-cn^\alpha) < \infty$, by Borel-Cantelli lemma, we conclude $\max\limits_{1\leq i\leq M} \|\widetilde{J}_{1nj}\| = o(n^{\delta/2})$, a.s.

Now take arbitrarily $\gamma \in S$. Find j such that $\|\gamma_j - \gamma\| \leq cn^{-d}$. Denoting $\tilde{\gamma} = \gamma_j$, we have

$$\begin{aligned}\|\widetilde{J}_{1n}(\hat{\beta}_n)\| &\leq \max_{1\leq l\leq M} \|\widetilde{J}_{1nl}\| + \sum_{i=1}^n \|Z_i\| \sup_\gamma \|H_i(\gamma)\Lambda_i(\gamma) - H_i(\tilde{\gamma})\Lambda_i(\tilde{\gamma})\|\|\tilde{e}_i\| \\ &\equiv K_1 + K_2.\end{aligned} \tag{2.14}$$

Earlier we proved $K_1 = o_p(n^{\delta/2})$. Also, $K_2 \leq cn^{-d}\sum\limits_{i=1}^n \|\tilde{e}_i\| \leq cn^{1-d+1/r}$. In view of the choice of a, we have $1 - d \leq \delta/2$. So $K_2 = o_p(n^{\delta/2})$. Combining these two results, we conclude $\tilde{J}_{1n}(\tilde{\beta}_n) = o_p(n^{\delta/2})$. As noticed earlier, this deduces $J_{1n} = o_p(n^{\delta/2})$. Hence

$$B_n^{-1/2} J_{1n} = o_p(1). \tag{2.15}$$

Combining (2.8) and (2.15), we prove (2.6). As stated earlier, this completes the proof of Theorem 1.1.

§ 3 . Proof of Theorem 1.2

Denote $W_i = H_i' Z_i'$. Then (1.4) can be written as

$$\Big(\sum_{i=1}^n W_i'\Lambda_i W_i\Big)\Big(\sum_{i=1}^n W_i'\Lambda_i\Sigma_i\Lambda_i W_i\Big)^{-1}\Big(\sum_{i=1}^n W_i'\Lambda_i W_i\Big) \leq \sum_{i=1}^n W_i'\Sigma_i^{-1}W_i. \tag{3.1}$$

Let $D_i = \Sigma_i^{1/2}\Lambda_i\Sigma_i^{1/2}$. Then (3.1) reduces to

$$\begin{aligned}&\Big(\sum_{i=1}^n W_i'\Sigma_i^{-1/2}D_i\Sigma_i^{-1/2}W_i\Big)\Big(\sum_{i=1}^n W_i'\Sigma_i^{-1/2}D_i^2\Sigma_i^{-1/2}W_i\Big)\Big(\sum_{i=1}^n W_i'\Sigma_i^{-1/2}D_i\Sigma_i^{-1/2}W_i\Big) \\ &\leq \sum_{i=1}^n W_i'\Sigma_i^{-1}W_i.\end{aligned} \tag{3.2}$$

Letting $U_i = \Sigma_i^{-1/2} W_i \Big(\sum\limits_{i=1}^n W_i' \Sigma_i^{-1} W_i \Big)^{-1/2}$, we have

$$\sum_{i=1}^n U_i' U_i = I. \tag{3.3}$$

Then (3.2) reduces to

$$\Big(\sum_{i=1}^n U_i' D_i U_i \Big)^2 \leq \sum_{i=1}^n U_i' D_i^2 U_i. \tag{3.4}$$

In fact, denote $T_n = \sum\limits_{i=1}^n W_i' \Sigma_i^{-1} W_i$. Then (3.2) can be written as

$$T_n^{1/2} \sum_{i=1}^n U_i' D_i U_i T_n^{1/2} \Big(T_n^{1/2} \sum_{i=1}^n U_i' D_i^2 U_i T_n^{1/2} \Big)^{-1} T_n^{1/2} \sum_{i=1}^n U_i' D_i U_i T_n^{1/2} \leq T_n.$$

Obviously, this is equivalent to (3.4).

In order to prove (3.4), we take unit vector λ, denote $a_i = U_i \lambda$, and proceed to show that

$$\sum_{i=1}^n a_i' D_i^2 a_i \geq \Big\| \sum_{i=1}^n a_i' D_i U_i \Big\|^2. \tag{3.5}$$

Denote $\mu_i = D_i a_i,\ 1 \leq i \leq n$. Then (3.5) can be written as

$$\sum_{i=1}^n \mu_i' \mu_i \geq \Big\| \sum_{i=1}^n \mu_i' U_i \Big\|^2. \tag{3.6}$$

Denote $\mu' = (\mu_1', \cdots, \mu_n'),\ V' = (U_1', \cdots, U_n')$. Then the right-hand side of (3.6) equals $\|\mu' V\|^2$. Since $V'V = I$, the rows of V are mutually orthogonal unit vectors. Hence $\|\mu' V\|^2 \leq \|\mu\|^2 = \sum\limits_{i=1}^n \mu_i' \mu_i$. This gives (3.6), hence (3.4). As noted ealier, it ends the proof of Theorem 1.2.

Remark 3.1. To make (1.4) an equality, one choice is, as we just show,

$$\Lambda_i = \Sigma_i^{-1}. \tag{3.7}$$

The problem is whether or not (3.7) is the only choice. We may easily give an example, in which the answer is in the negative. For example, in the case of $q \geq p$ and $Z_1 = Z_2$, we can take $\Lambda_1 = \Sigma_2^{-1},\ \Lambda_2 = \Sigma_1^{-1},\ \Lambda_i = \Sigma_i^{-1},\ i \geq 3$. Is it possible to construct an example other than such trivial cases? We cannot yet give a definite answer for this problem. However, we can deduce formally a necessary and sufficient condition ensuring the equality in (1.4).

Obviously, to make (1.4), or (3.6), an equality, it is necessary and sufficient that $\mu \in \mathcal{M}(V)$, where $\mathcal{M}(V)$ is the linear subspace spanned by the columns of V. That is to say, for any given unit q-vector, there must exist a q-vector b such that

$$\mu_i = D_i U_i \lambda = U_i b, \qquad 1 \leq i \leq n. \tag{3.8}$$

Denote $T = \sum_{i=1}^{n} Z_i H_i \Sigma_i^{-1} H_i' Z_i'$. Then (3.8) can be written as

$$D_i \Sigma_i^{-1/2} H_i' Z_i' T^{-1/2} \lambda = U_i b, \qquad 1 \le i \le n. \tag{3.9}$$

Since $T > 0$, from (3.9) we see that for any q-vector $\lambda^*(= T^{-1/2}\lambda)$, there exists a q-vector b, such that

$$D_i \Sigma_i^{-1/2} H_i' Z_i' \lambda^* = U_i b, \qquad 1 \le i \le n. \tag{3.10}$$

From (3.10), it follows that there exists a $q \times p$ matrix A such that

$$D_i \Sigma_i^{-1/2} H_i' Z_i' = U_i A, \qquad 1 \le i \le n. \tag{3.11}$$

From (3.3) and (3.11), we have

$$A = \sum_{i=1}^{n} U_i' D_i \Sigma_i^{-1/2} H_i' Z_i' = T^{-1/2} \sum_{i=1}^{n} Z_i H_i \Lambda_i H_i' Z_i' \equiv T^{-1/2} T^*, \qquad T^* = \sum_{i=1}^{n} Z_i H_i \Lambda_i H_i' Z_i'.$$

Hence

$$\Sigma_i^{1/2} \Lambda_i H_i' Z_i' = \Sigma_i^{-1/2} H_i' Z_i' T^{-1/2} T^{-1/2} T^* = \Sigma_i^{-1/2} H_i' Z_i' T^{-1} T^*, \qquad 1 \le i \le n. \tag{3.12}$$

This is the necessary and sufficient condition that $\{\Lambda_i\}$ must satisfy in order to make (1.4) an equality. But the involved form of this condition makes it difficult in applying it to specific cases.

References

[1] Haberman, S. J., Maximum likelihood estimates in exponential response models. *Ann. Statist.*, **5**(1997), 1148–1169.

[2] Nordberg, L., Asymptotic normality of maximum likelihood estimates based on independent,unequally distributed observations in exponential family models, *Scand. J. Statist.*, **7**(1980), 27–32.

[3] Fahrmeir, L. & Kanfmann, H., Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models, *Ann. Statist.*, **13**(1985), 342–368.

[4] Fahrmeir, L. & Kanfmann, H., Asymptotic inference in discrete response models, *Statistical Papers*, **27**(1986), 179–205.

[5] Fahrmeir, L., Maximum likelihood estimation in misspecified generalized linear models, *Ststistics*, **21**(1990), 487–502.

[6] Chen, K. et al., Strong consistency of maximum Quasi-likelihood estimators in generalized linear models with fixed and adaptive design, *Ann. Statist.*, **27**(1999), 1155–1163.

[7] Yue, L. & Chen, X. R., Rate of strong consistency of quasi maximum likelihood estimate in generalized linear models, *Science in China, Ser. A*, **47:**6(2004), 882–893.

[8] Petrov, V. V., Sum of Independent Variables, Springer-Verlag, Berlin, New York, 1975.

[9] McCullagh, P. & Nelder, J. A., Generlized Linear Models, Chapman and Hall, London, 1989.