Statistical Methods for Correlated Data

Generalized Estimating Equations for General Regression

Michele Guindani

Department of Biostatistics
UCLA

Marginal models

- We have already presented the basics of the GEE approach. The extension to GLM mean models is conceptually straightforward, since all that is required is specification of a mean model and a working covariance model.

Marginal models

- We have already presented the basics of the GEE approach. The extension to GLM mean models is conceptually straightforward, since all that is required is specification of a mean model and a working covariance model.
- The defining feature of marginal models is a regression model relating the mean response at each occasion, via a suitable link function, to the covariates.
- With a marginal model, the main focus is on making inferences about population means.
- As a result, marginal models for longitudinal data separately model the mean response and the within-subject association among the repeated responses. In a marginal model, the goal is to make inferences about the former, whereas the latter is regarded as a nuisance characteristic of the data that must be taken into account in order to make correct inferences about changes in the population mean response over time.

Marginal models

- A marginal model for longitudinal data has the following three-part specification:
  1. The conditional expectation of each response, $E\left(Y_{ij}|\boldsymbol{X}_{ij}\right)=\mu_{ij}$, is assumed to depend on the covariates through a known link function $h^{-1}(\cdot)$, e.g., $\operatorname{logit}\left(\mu_{ij}\right)$ or $\log\left(\mu_{ij}\right)$
  $$h^{-1}\left(\mu_{ij}\right)=g(\mu_{ij})=\eta_{ij}=\boldsymbol{X}_{ij}'\boldsymbol{\beta}$$
  2. The conditional variance of each Y_{ij}, given $\boldsymbol{X}_{ij}$, is assumed to depend on the mean
  $$\operatorname{Var}\left(Y_{ij}\right)=\phi v\left(\mu_{ij}\right)$$
  3. (LDA-specific component) The conditional within-subject association among the vector of repeated responses, given the covariates, is a function of a vector of association parameters, $\boldsymbol{\alpha}$ (and also depends upon the means, μ_{ij})(e.g., the components of $\boldsymbol{\alpha}$ might represent the pairwise correlations or log odds ratios among the repeated responses).

A Marginal model for count data

Counts are often model with a log-link function ("Poisson")

- mean model

$$\log(\mu_{ij}) = \eta_{ij} = X_{ij}^{r}\beta$$

- variance model

$$\text{Var}(Y_{ij}|X_{ij}) = \phi\mu_{ij}$$

  with $\phi =$ over-dispersion parameter

- within-subject association

$$\text{Corr}(y_{ij}, y_{ik}|x_{ij}, x_{ik}) = \alpha_{jk}$$

  The within-subject association is specified in terms of an unstructured pairwise correlation pattern. Of course, other choices for the link and variance functions are possible; similarly other models for the correlation (e.g., first-order autoregressive correlation pattern) are also possible.

A Marginal model for a binary response

- Suppose that y_{ij} is a binary response $\begin{cases} 0 & \text{failure} \\ 1 & \text{success} \end{cases}$
- The mean model is specified by

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \eta_{ij} = X_{ij}^\top \beta$$

- The variance model is specified by

$$\mathrm{Var}\left(Y_{ij}|X_{ij}\right) = \mu_{ij}\left(1-\mu_{ij}\right) \quad (\phi = 1)$$

- The within-subject association among the vector of repeated responses is assumed to have an unstructured pairwise log odds ratio pattern,

$$\log \mathrm{OR}\left(Y_{ij}, Y_{ik}|X_{ij}, X_{ik}\right) = \alpha_{jk}$$

where

$$\mathrm{OR}\left(Y_j, Y_k\right) = \frac{\Pr\left(Y_j = 1, Y_k = 1\right)\Pr\left(Y_j = 0, Y_k = 0\right)}{\Pr\left(Y_j = 1, Y_k = 0\right)\Pr\left(Y_j = 0, Y_k = 1\right)}$$

Marginal Models

- A crucial aspect of marginal models is that the mean response and within-subject association are modeled separately.
- This separation of the modeling of the mean response and the association among responses has important implications for interpretation of the regression parameters in the model for the mean response.
- Regression parameters, $\boldsymbol{\beta}$, in the marginal model have so-called population-averaged interpretations. That is, they describe how the mean response in the population is related to the covariates

- The previous examples demonstrate how the specification of the three components of a marginal model might differ according to the type of response variable.
- However, they should not be considered prescriptions for constructing marginal models; in principle, any suitable link function can be chosen and other assumptions about the variances and within-subject associations can be made.
- The choices for the three components of a marginal model should reflect statistical and subject-matter considerations.

GEE estimation

- Let's consider a "working" covariance matrix

$$\boldsymbol{W}_i = \boldsymbol{\Delta}_i^{1/2} \boldsymbol{R}_i(\boldsymbol{\alpha}) \boldsymbol{\Delta}_i^{1/2}$$

  where $\boldsymbol{\Delta}_i = \mathrm{diag}\left[\mathrm{var}\left(Y_{i1}\right), \ldots, \mathrm{var}\left(Y_{in_i}\right)\right]^{\mathrm{T}}$ and $\boldsymbol{R}_i$ is a working correlation model.

- Then, GEE takes the estimator the estimator $\widehat{\beta}$ that satisfies

$$\boldsymbol{G}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}) = \sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \left(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i\right) = \mathbf{0}$$

  where

$$\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$$

- The estimator $\widehat{\beta}$ will not be of closed form unless the link is linear

GEE estimation

- $\sqrt{m}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \mathrm{N}_{k+1}\left(\mathbf{0}, \boldsymbol{V}_\beta\right)$ where V_β takes the sandwich form

$$\left(\sum_{i=1}^{m} D_i^T W_i^{-1} D_i\right)^{-1} \left[\sum_{i=1}^{m} D_i^T W_i^{-1} \operatorname{cov}\left(Y_i\right) W_i^{-1} D_i\right] \left(\sum_{i=1}^{m} D_i^T W_i^{-1} D_i\right)^{-}$$

- In practice, an empirical estimator of $\operatorname{cov}\left(\boldsymbol{Y}_i\right)$ is substituted to obtain $\widehat{\boldsymbol{V}}_\beta$.
- This produces a consistent estimator of the standard error of $\widehat{\beta}$, so long as we have independence between units $i \neq i', i, i' = 1, \ldots, m$.

GEE estimation

- For small m, the variance estimator may be unstable and GLMMs model-based estimation may be preferable.

- Similarly as previously discussed, the efficiency of the estimator is improved when the chosen variance-covariance matrix is "closer" to the true dependence structure of the data (similar discussion as for linear GEEs)
- However, for discrete data, there is often no natural choice since, in this setting, the correlation is not an intuitive measure of dependence.

Extensions of the Standard GEE methods

- Recall, the GEE estimation originally proposed by Liang and Zeger (1986) is based on estimating $\rho_{ist}(\boldsymbol{\alpha}) = \text{Corr}\left(Y_{is}, Y_{it}\right)$ using at each iteration "pseudo-data" for the correlation structure:
$$U_{ist}(\boldsymbol{\beta}) = \frac{\left(Y_{is} - \mu_{is}\right)\left(Y_{it} - \mu_{it}\right)}{\phi\left\{v\left(\mu_{is}\right)v\left(\mu_{it}\right)\right\}^{1/2}}$$
letting $\boldsymbol{U}_i(\boldsymbol{\beta}) = \left(U_{i12}, U_{i13}, \ldots, U_{in_{i-1}n_i}\right)'$ and considering a second set of estimating equations
$$u_{\alpha}(\boldsymbol{\alpha}) = \sum_{i=1}^{N} E_i' W_i^{-1}\left\{\boldsymbol{U}_i(\boldsymbol{\beta}) - \boldsymbol{\rho}_i(\boldsymbol{\alpha})\right\} = \mathbf{0}$$
where $W_i \approx \text{Cov}\left(\boldsymbol{U}_i\right)$ and $E_i = \partial\boldsymbol{\rho}_i(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$. The working covariance matrix for $\boldsymbol{U}_i$ is typically specified as $\text{diag}\left\{\text{Var}\left(U_{ist}\right)\right\}$. In their original paper on GEE, Liang and Zeger (1986) let W_i be the $\left(n_i \times n_i - 1\right)/2 \times \left(n_i \times n_i - 1\right)/2$ identity matrix, whereas Prentice (1988) suggested letting W_i be a diagonal matrix with the approximate variances of U_{ist} along the diagonal.

Extensions of the Standard GEE methods

- As previously discussed, when V_i is correctly specified in the standard GEE approach, the resulting estimator of $\boldsymbol{\beta}$ is efficient in the sense of having the smallest variance among all estimators in the class of GEE estimators considered (GM theorem).
- However, the estimators of the correlation parameters $\boldsymbol{\alpha}$ are not efficient (no GM).

Extensions of the Standard GEE methods

- As previously discussed, when V_i is correctly specified in the standard GEE approach, the resulting estimator of $\boldsymbol{\beta}$ is efficient in the sense of having the smallest variance among all estimators in the class of GEE estimators considered (GM theorem).
- However, the estimators of the correlation parameters $\boldsymbol{\alpha}$ are not efficient (no GM). In addition,
  - Crowder (Biometrika, 1995) discusses issues with inconsistent estimation of the within-subject correlations under a mis-specified "working" correlation structure based on asymptotic theory
  - Moment estimators are designed for continuous random variables, and there is no guarantee that the estimates produced will lie within the admissible range

- The estimated correlation matrix may not be a positive definite matrix in certain cases.
- Chaganty and Joe (2006), Sabo and Chaganty (Stats in Medicine, 2010) discuss further limits on the estimation of the correlation that are not immediately addressed by GEE. For example, if the random vector consists of a series of binary observations then there are added restrictions imposed on the correlations by the univariate marginal probabilities.
- This has led researchers to develop develop alternative estimating equations for α
- The standard set of estimating equations that we have considered so far is often referred to as first-order GEE or "GEE1".

Multivariate normal estimating equations

- One set of estimating equations for α that has generated some interest in the statistical literature is that based on the notion of "Gaussian" estimation (see, for example, Lipsitz, Laird, and Harringtor, 1992; Lee, Laird, and Johnstor, 1999; Hall and Severini, 1998 ; Lipsitz et al., 2000 ; Fitzmaurice, Lipsitz, and Molenberghs, 2001 ; Wang and Cares; 2003).
- The key idea here is to base estimation of α on the multivariate normal estimating equations for the correlations.

Multivariate normal estimating equations

- One set of estimating equations for α that has generated some interest in the statistical literature is that based on the notion of "Gaussian" estimation (see, for example, Lipsitz, Laird, and Harringtor, 1992; Lee, Laird, and Johnstor, 1999; Hall and Severini, 1998 ; Lipsitz et al., 2000 ; Fitzmaurice, Lipsitz, and Molenberghs, 2001 ; Wang and Cares; 2003).
- The key idea here is to base estimation of α on the multivariate normal estimating equations for the correlations.
- Let
$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_i(\boldsymbol{\beta}) = \frac{1}{\sqrt{\phi}} A_i^{-1/2} \left(\mathbf{Y}_i - \boldsymbol{\mu}_i\right)$$
be the vector of standardized residuals, where A_i is a diagonal matrix with diagonal elements $v\left(\mu_{ij}\right)$.
- A second set of (moment) estimating equations can be obtained as the score equations for $\boldsymbol{\alpha}$ under the assumption that $\boldsymbol{\xi}_i \sim N\left(\mathbf{0}, R_i(\boldsymbol{\alpha})\right)$

Multivariate normal estimating equations

- Specifically, the estimating equations for $\boldsymbol{\alpha}$ are given by

$$u_\alpha(\boldsymbol{\alpha}) = \left[\frac{\partial}{\partial \boldsymbol{\alpha}} \left\{ \sum_{i=1}^{N} \log |R_i(\boldsymbol{\alpha})| - \sum_{i=1}^{N} \boldsymbol{\xi}_i'(\boldsymbol{\beta}) R_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{\xi}_i(\boldsymbol{\beta}) \right\} \right] = \mathbf{0}$$

  The r-th component of $\boldsymbol{u}_\alpha(\boldsymbol{\alpha})$ equals

$$\sum_{i=1}^{N} \operatorname{tr} \left[R_i^{-1}(\boldsymbol{\alpha}) \left\{ \boldsymbol{\xi}_i(\boldsymbol{\beta}) \boldsymbol{\xi}_i'(\boldsymbol{\beta}) - R_i(\boldsymbol{\alpha}) \right\} R_i^{-1}(\boldsymbol{\alpha}) \dot{\mathrm{R}}_{ir}(\boldsymbol{\alpha}) \right]$$

  with $\dot{\mathbf{R}}_{ir}(\boldsymbol{\alpha}) = \partial R_i(\boldsymbol{\alpha}) / \partial \alpha_r$.
- The multivariate normal estimating equations is particularly appealing with unbalanced data: it ensure that the algorithm leads to an estimated correlation matrix $R_i(\widehat{\boldsymbol{\alpha}})$, which is non-negative definite
- As further modifications, Wang and Carey (2004) proposed estimating the correlation parameters by differentiating the Cholesky decomposition of the working correlation matrix; a similar approach was also used by Ye and Pan (2006). Similar Gaussian or "quadratic" estimating equations have been proposed by Crowder (1995) and Qu, Lindsay, and Li (2000).

Second-order generalized estimating equations (GEE2)

- Another approach considers a second-order extension of generalized estimating equations, and it is hereafter referred to as GEE2 (Zhao and Prentice, 1990; Liang, Zeger, and Qaqish, 1992).
- The approach considers a *connected set of joint estimating equations* for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$

Second-order generalized estimating equations (GEE2)

- Another approach considers a second-order extension of generalized estimating equations, and it is hereafter referred to as GEE2 (Zhao and Prentice, 1990; Liang, Zeger, and Qaqish, 1992).
- The approach considers a *connected set of joint estimating equations* for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$
- To motivate a pair of estimating equations, consider the following model for a single individual with n independent observations:

$$Y_i|\boldsymbol{\beta},\boldsymbol{\alpha} \sim_{\text{ind}} \mathrm{N}\left[\mu_i(\boldsymbol{\beta}), \Sigma_i(\boldsymbol{\beta},\boldsymbol{\alpha})\right]$$

  For simplicity, assume $\Sigma_i()$ is a scalar (n indep. obs.). Extension of the argument to the non-independent case is straightforward.
- The log-likelihood is

$$l(\boldsymbol{\beta},\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{i=1}^{n} \log\left(\Sigma_i\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(Y_i - \mu_i\right)^2}{\Sigma_i}$$

GEE2

- Differentiation gives the score equations as

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Sigma_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}}\frac{1}{\Sigma_i} + \sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}}\frac{(Y_i - \mu_i)}{\Sigma_i} + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Sigma_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}}\frac{(Y_i - \mu_i)^2}{\Sigma_i^2}$$
$$= \sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}}\frac{(Y_i - \mu_i)}{\Sigma_i} + \sum_{i=1}^{n}\left(\frac{\partial \Sigma_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}}\frac{\left[(Y_i - \mu_i)^2 - \Sigma_i\right]}{2\Sigma_i^2}$$

$$\frac{\partial l}{\partial \alpha} = -\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Sigma_i}{\partial \alpha}\right)^{\mathrm{T}}\frac{1}{\Sigma_i} + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Sigma_i}{\partial \alpha}\right)^{\mathrm{T}}\frac{(Y_i - \mu_i)^2}{\Sigma_i^2}$$
$$= \sum_{i=1}^{n}\left(\frac{\partial \Sigma_i}{\partial \alpha}\right)^{\mathrm{T}}\frac{\left[(Y_i - \mu_i)^2 - \Sigma_i\right]}{2\Sigma_i^2}$$

- $\frac{\partial \Sigma_i}{\partial \boldsymbol{\beta}}$ is not zero; hence, the GEE2 estimating equations for $\boldsymbol{\beta}$ are quite different from the GEE1 estimation equations for $\boldsymbol{\beta}$.

- More in general, for dependent data, the general form of estimating equations is

$$\sum_{i=1}^{m} \begin{bmatrix} \boldsymbol{D}_i & \mathbf{0} \\ \boldsymbol{E}_i & \boldsymbol{F}_i \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{V}_i & \boldsymbol{C}_i \\ \boldsymbol{C}_i^{\mathrm{T}} & \boldsymbol{W}_i \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Y}_i - \boldsymbol{\mu}_i \\ \boldsymbol{S}_i - \boldsymbol{\Sigma}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

  where $\boldsymbol{S}_i = (\boldsymbol{Y}_i - \boldsymbol{\mu}_i)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)^T$, $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}, \boldsymbol{E}_i = \partial \boldsymbol{\Sigma}_i / \partial \boldsymbol{\beta}$, and $\boldsymbol{F}_i = \partial \boldsymbol{\Sigma}_i / \partial \alpha$ and we have "working" variance-covariance structure

$$\begin{aligned} \boldsymbol{V}_i &= \operatorname{var}(\boldsymbol{Y}_i) \\ \boldsymbol{C}_i &= \operatorname{cov}(\boldsymbol{Y}_i, \boldsymbol{S}_i) \\ \boldsymbol{W}_i &= \operatorname{var}(\boldsymbol{S}_i) \end{aligned}$$

- When $C_i = 0$, we obtain

$$\boldsymbol{G}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + \sum_{i=1}^{m} \boldsymbol{E}_i \boldsymbol{W}_i^{-1} (\boldsymbol{S}_i - \boldsymbol{\Sigma}_i)$$

$$\boldsymbol{G}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \boldsymbol{F}_i \boldsymbol{W}_i^{-1} (\boldsymbol{S}_i - \boldsymbol{\Sigma}_i)$$

  These equations can be seen as arising from a quadratic exponential model (see Wakefield and HLDA)

GEE2

- This pair of quadratic estimating functions is unbiased given correct specification of the first two moments (normality of the data is not required)
- GEE2 can yield substantial bias in estimators of both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ (Fitzmaurice, Lipsitz, and Molenberghs, 2001) if assumptions about second moments are mis-specified. This is in contrast to GEE1 where a consistent estimator of $\boldsymbol{\beta}$ is obtained provided only that the mean of $\boldsymbol{Y}_i$ has been correctly specified.
- In the GEE2 approach, if the variance model is wrong, we are no longer guaranteed a consistent estimator of $\boldsymbol{\beta}$.
- The main appeal of GEE2 is that - if the second-order assumptions are correct - it is almost fully efficient for both the marginal regression parameters, $\boldsymbol{\beta}$, and the within-subject associations, $\boldsymbol{\alpha}$.
- It is notable that none of the major statistical software packages currently have options for GEE2.

Quadratic Inference Functions

- Quadratic inference functions (QIF) provide another quasi-likelihood inference in the Marginal Generalized Linear models.
- In comparison to the GEE approach, QIF has the following advantages:
  - The application of QIF does not require more model assumptions than the GEE.
  - Qu et al. (2000) showed that the QIF estimator of $\boldsymbol{\beta}$ in the MGLM is more efficient than the GEE estimator, when the working correlation is misspecified, but equally efficient when the working correlation is correctly specified. This efficiency gain is due to the fact that QIF does not need to estimate the parameters in a given correlation structure.

- Moreover,
  - The QIF estimators are robust against unduly large outliers or contaminated data points, whereas the GEE is not robust and very sensitive to influential data cases.Refer to Qu and Song (2004) for more details.
- The formulation of the QIF is based on the fact that the inverse of the working correlation, say $\boldsymbol{R}(\boldsymbol{\alpha})$, can be written s a linear combination of basis coefficients. More specifically, we can write,

$$\mathbf{R}^{-1}(\gamma) = \sum_{l=1}^{m} \gamma_l M_l$$

where $M_1, \ldots, M_m$ are known matrices and $\gamma_1, \ldots, \gamma_m$ are unknown coefficients, functions of the correlation parameters $\boldsymbol{\alpha}$.

QIF

- Example 1: The compound symmetry working correlation matrix $\boldsymbol{R}$ givbes rise to $\boldsymbol{R}^{-1}(\boldsymbol{\gamma}) = \gamma_1 \, M_1 + \gamma_2 \, M_2$, where γ_l, $l = 1, 2$ are both functions of $\boldsymbol{\alpha}$. The two basis matrices are $M_1 = \mathbf{I}_{n_i}$, the n_i -dimensional identity matrix, and M_2, a matrix with 0 on the diagonal and 1 off the diagonal.
- Example 2: The AR-1 working correlation structure can be written as $\mathbf{R}^{-1}(\gamma) = \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 M_3$, where $M_1 = \mathbf{I}_{n_i}$, M_2 with 1 on the two main off-diagonals and 0 elsewhere, and M_3 with 1 on the corners (1,1) and (n_i, n_i) , and 0 elsewhere.

- The decomposition of the correlation matrix as such is not unique. However, the important point is that it allows to use different moment conditions, which can be used instead of the usual GEE equations.
- More specifically, consider the (GEE) estimating equation, where we assume $\boldsymbol{W}_i = \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{A}_i^{\frac{1}{2}}$ for some working correlation structure $\boldsymbol{R}(\boldsymbol{\alpha})$,

$$\sum_{i=1}^{N} \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \left(\gamma_1 M_1 + \ldots + \gamma_m M_m\right) \mathbf{A}_i^{-1/2} \mathbf{r}_i = \mathbf{0}$$

where $\mathbf{D}_i = \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}\right)$, $\mathbf{r}_i = \boldsymbol{Y}_i - \boldsymbol{\mu}_i$,
$\mathbf{A}_i = \operatorname{diag}\left\{\operatorname{Var}\left(Y_{i1}\right), \cdots, \operatorname{Var}\left(Y_{in_i}\right)\right\}$.

QIF

- The previous expression is in fact a linear combination of elements of the following vector of estimating equations (inference functions), each being related to one basis matrix,

$$\begin{aligned}
\varphi(\mathbf{Y};\boldsymbol{\beta}) &= \frac{1}{N}\sum_{i=1}^{N}\varphi_i\left(\mathbf{Y}_i;\boldsymbol{\beta}\right) \\
&= \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N}\mathbf{D}_i^T\mathbf{A}_i^{-1/2}M_1\mathbf{A}_i^{-1/2}\mathbf{r}_i \\ \vdots \\ \sum_{i=1}^{N}\mathbf{D}_i^T\mathbf{A}_i^{-1/2}M_m\mathbf{A}_i^{-1/2}\mathbf{r}_i \end{bmatrix}
\end{aligned}$$

where the coefficients γ_l's are not involved. If an independence working correlation is assumed, the φ reduces to the GEE, with $m = 1$ and $M_1 = I_{n_i}$.

QIF

- Since the number of components in inference function φ is greater than the dimension of $\boldsymbol{\beta}$, it is impossible to directly solve $\varphi(\boldsymbol{\beta}) = 0$ for $\boldsymbol{\beta}$. This is beause the $\boldsymbol{\beta}$ is overidentified.
- Following a Generalized Methods of Moment approach, one may follow a slightly different approach and optimize a quadratic distance function (Hansen 1982) of the φ

  $$Q(\boldsymbol{\beta}) = \boldsymbol{\varphi}^T(\mathbf{Y};\boldsymbol{\beta})\mathbf{C}(\boldsymbol{\beta})^{-1}\boldsymbol{\varphi}(\mathbf{Y};\boldsymbol{\beta})$$

  and take its minimizer as the estimator of $\boldsymbol{\beta}$.
- Note that here the matrix $\mathbf{C}(\boldsymbol{\beta}) = \left(1/K^2\right)\sum_{i=1}^{K}\varphi_i(\boldsymbol{\beta})\varphi_i(\boldsymbol{\beta})^T$ is a consistent estimate of the variance of $\varphi(\mathbf{Y};\boldsymbol{\beta})$
- This objective function Q is referred to as the quadratic inference function (QIF).
- Under some regularity conditions, e.g. unbiased estimating equations, the solver of the quadratic objective function yields a consistent estimator (Lee, 1996).
- Under some regularity conditions, the minimizer is also unique.

- Qu et al. (2000) showed that the QIF estimator $\widehat{\beta}$ is consistent and asymptotically normal, and more efficient than GEE estimators

Biometrika (2000), **87,** 4, *pp.* 823–836

Printed in Great Britain

Improving generalised estimating equations using quadratic inference functions

By ANNIE QU

Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

qu@stat.orst.edu

BRUCE G. LINDSAY and BING LI

Department of Statistics, 326 Thomas Building, The Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.

bgl@psu.edu bing@stat.psu.edu

Summary

Generalised estimating equations enable one to estimate regression parameters consistently in longitudinal data analysis even when the correlation structure is misspecified. However, under such misspecification, the estimator of the regression parameter can be inefficient. In this paper we introduce a method of quadratic inference functions that does not involve direct estimation of the correlation parameter, and that remains optimal even if the working correlation structure is misspecified. The idea is to represent the inverse of the working correlation matrix by the linear combination of basis matrices, a representation that is valid for the working correlations most commonly used. Both asymptotic theory and simulation show that under misspecified working assumptions these estimators are more efficient than estimators from generalised estimating equations. This approach also provides a chi-squared inference function for testing nested models and a chi-squared regression misspecification test. Furthermore, the test statistic follows a chi-squared distribution asymptotically whether or not the working correlation structure is correctly specified.

Some key words: Generalised estimating equation; Generalised method of moments; Linear approximate inverse; Longitudinal data; Quadratic inference function; Quasilikelihood.

- Finally, the objective function Q can be used to define certain model selection criteria such as Akaike's information criterion (AIC) and the traditional Bayes information criterion (BIC), respectively, as follows,

$$\mathrm{AIC} = Q(\widehat{\boldsymbol{\beta}}) + 2(m-1)p$$
$$\mathrm{BIC} = Q(\widehat{\boldsymbol{\beta}}) + \{(m-1)p\} \ln(K)$$

- In addition, a score-type test for a nested model can be derived
- Suppose the hypothesis of interest is $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$, under a partition of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ in the full model,

$$g(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} = \mathbf{x}_{1ij}^T \boldsymbol{\beta}_1 + \mathbf{x}_{2ij}^T \boldsymbol{\beta}_2$$

Here $\dim(\boldsymbol{\beta}_1) = p_1, \dim(\boldsymbol{\beta}_2) = p_2$, and $p_1 + p_2 = p$. Then under the H_0, the difference of QIF

$$\mathrm{DQIF} = Q\left(\widehat{\boldsymbol{\beta}}_1\right) - Q(\widehat{\boldsymbol{\beta}}) \overset{asy}{\sim} \chi^2_{p_2}$$