A Very Brief Introduction to Generalized Estimating Equations

Gesine Reinert

Department of Statistics

University of Oxford

1. GEEs in the GLM context

Idea: extend generalized linear models (GLMs) to accommodate the modeling of correlated data

Examples: Whenever data occur in clusters (panel data): Patient histories, insurance claims data (collected per insurer), etc.

Often people would fit a linear model to such data and only then adjust the standard errors to account for the clustering; the problem is that this post-hoc approach does not affect the parameter estimates in the model. Instead use GEEs:

GEE for GLMs in a nutshell:

1. Estimate a straightforward GLM, calculate the matrix of scaling values.

2. The scaling matrix adjusts the Hessian in the next iteration.

   Each subsequent iteration updates the parameter estimates, the adjusted Hessian matrix, and a matrix of scales.

The matrix of scales can be parametrized to allow user control over the structure of dependence in the data.

2. A Review of GLMs

For the *exponential family*, the likelihood may be expressed as

$$\exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Example: Poisson:

$$f(y,\mu) = \frac{e^{-\mu}\mu^y}{y!} = \exp\left\{\frac{yln(\mu) - \mu}{1} - ln\Gamma(y+1)\right\}$$

Other examples include normal, binomial, gamma, inverse Gaussian, geometric

Denote the mean by μ, then we use the parametrization

$$\theta = g(\mu)$$

where g is a monotone function called the *canonical link function*; g may include covariates.

With this parametrization,

$$E(y) = b'(\theta) = \mu$$

$$V(y) = b''(\theta)a(\phi)$$

Often the variance and the mean are dependent.

The function

$$V(\mu) = b''(\theta(\mu))$$

is also called the *variance function*.

Generalized linear regression model:

$$\eta_i = g(\mu_i) = \mathbf{X}\beta$$

Estimating equation: ℓ is the log likelihood,

$$\frac{\partial \ell}{\partial \theta} = 0$$

gives maximum-likelihood estimates

often use Newton-Raphson or Fisher scoring recursion to solve

By chain rule, treating the dispersion $a(\phi)$ as ancillary,

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \left[\left(\frac{\partial \ell}{\partial \theta}\right)\left(\frac{\partial \theta}{\partial \mu}\right)\left(\frac{\partial \mu}{\partial \eta}\right)\left(\frac{\partial \eta}{\partial \beta_j}\right)\right]_{p\times 1} \\
&= \left[\sum_i \left(\frac{y_i - b'(\theta_i)}{a(\phi)}\right)\frac{1}{V(\mu_i)}\left(\frac{\partial \mu}{\partial \eta}\right)_i x_{ji}\right]_{p\times 1} \\
&= \left[\sum_i \frac{y_i - \mu_i}{a(\phi)V(\mu_i)}\left(\frac{\partial \mu}{\partial \eta}\right)_i x_{ji}\right]_{p\times 1}
\end{aligned}$$

This leads to the *estimating equation*

$$\left[\left\{\frac{\partial \ell}{\partial \beta_j} = \sum_i \frac{y_i - \mu_i}{a(\phi)V(\mu_i)} \left(\frac{\partial \mu}{\partial \eta}\right)_i x_{ji}\right\}_{j=1,\ldots,p}\right]_{p\times 1}$$
$$= [0]_{p\times 1}$$

The variance is usually estimated by the observed Hessian (matrix of second derivatives) or the expected Hessian (Fisher information)

$$\widehat{V}_H(\widehat{\beta}) \;=\; (E)\left\{\left(-\frac{\partial^2 \ell}{\partial\beta_u \partial\beta_v}\right)\right\}^{-1}_{p\times p}$$

Problem: Generalized linear model assumes independent observations

Alternatively we can use a *sandwich estimate*: Let

$$\Psi(\beta) = \sum_{i=1}^{n} \Psi_i(\mathbf{x}_i, \beta)$$

with

$$\Psi_i(\mathbf{x}_i, \beta) = \left(\frac{\partial \ell}{\partial \eta}\right)_i = \left(\frac{\partial \ell}{\partial \mu}\right)_i \left(\frac{\partial \mu}{\partial \eta}\right)_i$$

being the *estimating equation* for the ith observation (in abuse of notation)

The sandwich estimate is of the form

$$A^{-1}BA^{-T}$$

where

$$A = \widehat{V}_H(\widehat{\beta}) = \left\{ E\left(\frac{\partial \Psi(\beta)}{\partial \beta} \right) \right\}^{-1}$$

is the usual estimate of the variance, and

$$B = E\Psi(\beta)^T \Psi(\beta)$$

is the correction term.

In the GLM,

$$\widehat{\Psi}_i(\mathbf{x}_i, \widehat{\beta}) \;=\; \mathbf{x}_i^T \left(\frac{y_i - \widehat{\mu}_i}{V(\widehat{\mu})_i} \right) \left(\frac{\partial \mu}{\partial \eta} \right)_i \widehat{\phi}$$

and

$$\widehat{B}(\widehat{\beta}) \;=\; \left[\sum_{i=1}^{n} \mathbf{x}_i^T \left\{ \frac{y_i - \widehat{\mu}_i}{V(\widehat{\mu})_i} \left(\frac{\partial \mu}{\partial \eta} \right)_i \widehat{\phi} \right\}^2 \mathbf{x}_i \right]_{p \times p}$$

Assume that $(a(\phi))^{-1}$ is estimated by $\widehat{\phi}$

The sandwich estimate combines the variance estimate from the specified model with a variance matrix constructed from the data

The sandwich estimate can be modified to take panel data into account

Is relatively robust to model misspecification

3. Generalized Estimating Equations

Assume n panels, n_i correlated observations in panel i; vector $\mathbf{x}$ of covariates to explain observations

exponential family, for observation t in panel i

$$\exp\left\{\frac{y_{it}\theta_{it} - b(\theta_{it})}{a(\phi)} + c(y_{it}, \phi)\right\}$$

Generalized Estimating Equations (GEEs) introduce second-order variance components directly into an estimating equation: ad-hoc rather than post-hoc

Include the panel effect in the estimating equation: solve

$$\begin{aligned}\Psi(\beta) &:= \left\{\sum_{i=1}^{n} \mathbf{x}_{ji}^{T} Diag\left(\frac{\partial \mu}{\partial \eta}\right)[V(\mu_i)]^{-1}\left(\frac{\mathbf{y}_i - \mu_i}{a(\phi)}\right)\right\} \\ &= 0\end{aligned}$$

with

$$V(\mu_i) = (Diag(V(\mu_{it})))^{\frac{1}{2}}\, R(\alpha)\, (Diag(V(\mu_{it})))^{\frac{1}{2}}$$

being an $n_i \times n_i$-matrix

Here, $R(\alpha)$ is the correlation matrix within panels, estimated through the parameter α

Liang and Zeger (1986) showed asymptotic normality.

Choice of $R(\alpha)$:

- Independent
- Exchangeable
- Autoregressive
- Unstructured
- Free specification

Example: (Hardin + Hilbe) Insurance claims data: payout y for car insurance claims given the car group (car1, car2, car3) and vehicle age group (value1, value2, value3); covariates for the interaction of the car and vehicle age group indicators

Panels defined by the policy holder's age group

Assume exchangeable correlation structure

Population-averaged (PA) model: include the within-panel dependence by averaging effects over all panels

Subject-specific model: include subject-specific panel-level components

Example: Subject-specific: estimate the odds of a child having respiratory illnes if the mother smokes compared to the odds of the same child having respiratory illness if the mother does not smoke

Population-averaged: estimate the odds of an average child with respiratory illness and a smoking mother to the odds of an average child with respiratory illness and a nonsmoking mother.

Population-averaged models are often included in statistical software (R, SAS, S-PLUS, Stata)

Subject-specific models require specification of the randomness for each subject, and therefore additional calculation and/or programming

4. Example: Geomorphological data

joint work with Stephan Harrison (Geography)

Certain landscape features are recorded in a river valley and its tributary streams

The (circular) data for the valley come in stretches, and are recorded both up-stream and down-stream

There are 692 observations, 370 of these indicate presence of the feature

The feature occurs in clumps; the longest clump is of length 42

The data is decomposed into 6 stretches (panels), the smallest has 45 observations, the largest has 205 observations

The underlying question is whether or not there is a preferred orientation of these features

We treat each stretch as a cluster, and assume that the clusters are independent

There is clearly autocorrelation in the data

Use logistic regression model

To avoid the assumption that all panels have the same autocorrelation, we model the autoregressive dependence explicitly by including the binary covariate of presence of the feature at the previous considered location

Result: the estimated correlation parameter for the autoregressive model of order 1 is not significant (p-value 0.322); taking the previous observation into account already takes care of the first order autocorrelation

Therefore we repeated the analysis with an independent error structure:

	Estimate	S.E.	P-value
Intercept	-1.832	0.131	0.000
Cosine	0.290	0.290	0.021
Previous obs.	3.849	0.225	0.000

Regression on sine alone gave no significance for the sine component

Also we regressed on the the product of sine and cosine, with no significant result

Conclusion: Positive cosines are significantly favoured

Had we ignored the dependence in the data: the sine contribution would turn out to be significant at level 9.68×10^{-8}, with -1.262 as coefficient for the sine, indicating erroneously that negative signs, i.e. westerly orientations, would be preferred.

5. Last Remarks

One can use some *working* correlation structure that may be wrong, but the resulting regression coefficient estimate is still consistent and asymptotically normal; but selection of an appropriate correlation structure improves efficiency

There are further extensions of the model:

multinomial data

models where some covariates are measured with error

robust version

missing data: under development

Residual analysis and tests for coefficients in the model: available

References

Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data.* Oxford University Press.

Godambe, V.P. ed. (1991). *Estimating Functions.* Oxford University Press.

*Hardin, J.W. and Hilbe, J.M. (2003). *Generalized Estimating Equations.* Chapman and Hall, Boca Raton etc.

Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression.* Second Edition. Wiley, New York etc.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

(and references therein)