

Generalized Estimating Equations

Sergio Olmos Pardo

29/3/2018

Framework

The GEE approach developed by Liang and Zeger (1986) can be seen as a generalization of the estimation method of quasi-likelihood proposed by Wedderburn (1974), which allows for a separate specification of the mean and variance structure. Moreover, it is not necessary that these specifications correspond to a proper likelihood function. It suffices to specify the correct mean structure, together with a *working* variance structure, and to define parameter estimates as the solutions of a quasi-score function or *generalized estimating equation*.

Consider a sample of independent multivariate observations $Y_i = (Y_{i1}, \dots, Y_{it}, \dots, Y_{in_i})$ of $i = 1, 2, \dots, K$ clusters each with n_i observations. Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{it}, \dots, \mathbf{x}_{in_i})^T$ be the $n_i \times p$ matrix of covariate values in cluster i . The expectations $E(Y_{it}) = \mu_{it}$ are related to the p dimensional regressor vector \mathbf{x}_{it} by the mean link function g

$$g(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}.$$

Let

$$\text{VAR}(Y_{it}) = \phi a_{it},$$

where ϕ is a common scale parameter and $a_{it} = a(\mu_{it})$ is a known variance function.

Let $\mathbf{R}_i(\boldsymbol{\alpha})$ be a working correlation matrix completely described by the parameter vector $\boldsymbol{\alpha}$ of length m , that reflects the pattern of correlation among observations in cluster i . Let

$$\mathbf{V}_i = \phi \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2}$$

be the corresponding working covariance matrix of \mathbf{Y}_i , where \mathbf{A}_i is the diagonal matrix with entries a_{it} .

For given estimates $(\hat{\phi}, \hat{\boldsymbol{\alpha}})$ of $(\phi, \boldsymbol{\alpha})$ the estimate $\hat{\boldsymbol{\beta}}$ is the solution of the generalized estimating equation

$$\sum_{i=1}^K \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}.$$

Liang and Zeger (1986) suggest to use consistent moment estimates for ϕ and $\boldsymbol{\alpha}$ and then use an iterative process between a modified Fisher scoring algorithm for $\boldsymbol{\beta}$ and the moment estimation of ϕ and $\boldsymbol{\alpha}$. This scheme yields a consistent estimate for $\boldsymbol{\beta}$ even if the working correlation matrices $\mathbf{R}_i(\boldsymbol{\alpha})$ are misspecified.

In ordinary maximum likelihood estimation, estimates of $\boldsymbol{\beta}$ are consistent and asymptotically normal, and the variance of the estimator can be consistently estimated by the inverse of the observed information matrix given the hypothesis of conditional independence between the observations. In the context of clustered data, however, the independence assumption may not be valid and the inverse of the observed information matrix can result in inconsistent estimates of the asymptotic variance of $\hat{\boldsymbol{\beta}}$ (Liang and Zeger 1986; Zeger and Liang 1986). Accordingly, Liang and Zeger (1986) and Zeger and Liang (1986) propose a robust variance-covariance estimator to adjust for intra-cluster correlation.

According to Liang and Zeger (1986) and Zeger and Liang (1986), given the estimate $\hat{\beta}$, it follows that $K^{1/2}(\hat{\beta} - \beta)$ is asymptotically multivariate normally distributed with zero mean and variance-covariance matrix

$$\Sigma = K \Sigma_0^{-1} \Sigma_1 \Sigma_0^{-1} \quad (1)$$

where

$$\Sigma_0 = \sum_{i=1}^K \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta^T},$$

$$\Sigma_1 = \sum_{i=1}^K \frac{\partial \mu_i^T}{\partial \beta} V_i^{-1} \text{COV}(Y_i) V_i^{-1} \frac{\partial \mu_i}{\partial \beta^T}.$$

Σ_0 is the conventional, model-based Fisher information matrix, and its inverse can be seen as the naive variance estimator based on the naive model which assumes that observations within a cluster are conditionally independent. On the other hand, Σ_1 is the covariance matrix of the score statistic and accounts for intra-cluster correlation (Liu 2015). The estimate $\hat{\Sigma}$ is a consistent estimate of Σ even if the working correlation matrices $R_i(\alpha)$ are misspecified (Liang and Zeger 1986; Zeger and Liang 1986).

Replacing β, ϕ and α by the consistent estimates of the iterative algorithm, and the variance-covariance matrix $\text{COV}(Y_i)$ by $(Y_i - \mu_i)(Y_i - \mu_i)^T$ in equation 1 yields the so called *sandwich* variance estimator $\hat{\Sigma}$ of Σ .

Therefore, given this empirical adjustment on dependence among measurements in the same cluster, the clustered data can be assumed to be conditionally independent and a valid Wald score can be derived to perform hypothesis testing on parameter estimates (Liu 2015).

Extensions

This approach to GEE only estimates the mean structure, while the association structure is treated as nuisance. This framework is often called GEE1. In many applications, however, the association structure is, in fact, of primary interest. By adding a second set of estimating equations to the GEE framework, Prentice (1988) extends the GEE1 approach allowing the simultaneous estimation of both the mean and the association structure. Because the approach of Prentice (1988) uses the first two moments it is sometimes referred to as GEE2. The estimation of β and α is obtained by an iterative procedure, giving statistically efficient, consistent and robust estimates (Prentice 1988; Liang, Zeger, and Qaqish 1992). Zhao and Prentice (1990) further broaden the GEE approach by allowing for joint estimation of β and α .

Nonetheless, the solution of GEE2 is computationally difficult when cluster size is large, restricting the applicability of the approach. Moreover, if intra-cluster correlations are incorrectly specified neither β nor α are estimated consistently (Liang, Zeger, and Qaqish 1992). For these reasons the GEE2 approach will not be considered further.

The `geepack` package

The `geepack` package implements the GEE approach in R.

References

- Liang, Kung-Yee, and Scott L. Zeger. 1986. “Longitudinal Data Analysis Using Generalized Linear Models.” *Biometrika* 73 (1). [Oxford University Press, Biometrika Trust]: 13–22. <http://www.jstor.org/stable/2336267>.
- Liang, Kung-Yee, Scott L. Zeger, and Bahjat Qaqish. 1992. “Multivariate Regression Analyses for Categorical Data.” *Journal of the Royal Statistical Society. Series B (Methodological)* 54 (1). [Royal Statistical Society, Wiley]: 3–40. <http://www.jstor.org/stable/2345947>.
- Liu, Xian. 2015. *Methods and Applications of Longitudinal Data Analysis*. 1st ed. Academic Press.
- Prentice, Ross L. 1988. “Correlated Binary Regression with Covariates Specific to Each Binary Observation.” *Biometrics* 44 (4). [Wiley, International Biometric Society]: 1033–48. <http://0-www.jstor.org.catalog.uoc.edu/stable/2531733>.
- Wedderburn, R. W. M. 1974. “Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method.” *Biometrika* 61 (3). [Oxford University Press, Biometrika Trust]: 439–47. <http://www.jstor.org/stable/2334725>.
- Zeger, Scott L., and Kung-Yee Liang. 1986. “Longitudinal Data Analysis for Discrete and Continuous Outcomes.” *Biometrics* 42 (1). [Wiley, International Biometric Society]: 121–30. <http://www.jstor.org/stable/2531248>.
- Zhao, Lue Ping, and Ross L. Prentice. 1990. “Correlated Binary Regression Using a Quadratic Exponential Model.” *Biometrika* 77 (3): 642–48. doi:10.1093/biomet/77.3.642.