Notation

D. Erik Boonstra, MS

Summer 2023

The current version of the notation for this project is based on Pan 2001a with some modifications made by D. Erik Boonstra and Joseph E. Cavanuagh.

Data generation

Suppose we have a random sample of observations from N individuals. For each individual i, we have

- $\mathbf{y}_i = \begin{pmatrix} y_{i1} & \dots & y_{in_i} \end{pmatrix}^\top$,
- covariates $\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1}^\top & \dots & \mathbf{x}_{in_i}^\top \end{pmatrix}^\top$, where $\mathbf{x}_{ij}$ is a $p-$dimensional vector,
- conditional on the covariates the observations of $\mathbf{y}_i$ are correlated but $\mathbf{y}_i$ and $\mathbf{y}_k$ are independent for all $i \neq k$,
- $\mathcal{D} = \{(\mathbf{y}_1, \mathbf{X}_1), \dots, (\mathbf{y}_N, \mathbf{X}_N)\}$, and
- $g(\boldsymbol{\mu}_i) = \mathbf{X}_i\boldsymbol{\beta}$,

where $\mathbb{E}(\mathbf{y}_i \mid \mathbf{X}_i) = \boldsymbol{\mu}_i$, $g(\cdot)$ is a specified link function, and $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \dots & \beta_p \end{pmatrix}^\top$ is a vector of unknown regression parameters to be estimated.

Generalized estimating equations

For *generalized estimating equations* (GEEs), the estimates for $\boldsymbol{\beta}$ are obtained by using the following estimating equations (Liang and Zeger, 1986)

$$S(\boldsymbol{\beta}; \mathbf{R}, \mathcal{D}) = \sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = 0, \tag{1}$$

where $\mathbf{D}_i = \mathbf{D}_i(\boldsymbol{\beta}) = \partial\boldsymbol{\mu}_i(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top$ and $\mathbf{V}_i$ is the working covariance matrix of $\mathbf{y}_i$. This covariance matrix, $\mathbf{V}_i$, can be expressed as

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

with $\mathbf{R} = \mathbf{R}_i(\boldsymbol{\alpha})$ being the working correlation matrix and $\mathbf{A}_i$ is a diagonal matrix of marginal variance, $\mathbb{V}(y_{ij}) = \phi\nu(\mu_{ij})$, where ϕ is the dispersion parameter and $\nu(\cdot)$ is mean-to-variance relationship defined by link function, $g(\cdot)$. GEEs are based on the quasi-likelihood. So, for simplicity, suppose we have a scalar response variable, y. Then, the (log) quasi-likelihood function defined by McCullagh and Nelder, 1989 is

$$Q(\mu, \phi; y) = \int_y^\mu \frac{y - t}{\phi\nu(t)} dt, \tag{2}$$

where $\mu = \mathbb{E}(y)$ and $\mathbb{V}(y) = \phi\nu(\mu)$. With a $1 \times p$ covariate $\mathbf{x}$ and a specified regression model

$$\mathbb{E}(y) = \mu = g^{-1}(\mathbf{x}\boldsymbol{\beta}),$$

the quasi-likelihood may be written as

$$\begin{aligned}Q(\mu, \phi; y) &= Q(g^{-1}(\mathbf{x}\boldsymbol{\beta}), \phi; y)\\ &= Q(\boldsymbol{\beta}, \phi; (y, \mathbf{x})).\end{aligned}$$

Now if the working independence model, $\mathbf{R} = \mathbf{I}$ is used, the working assumption is that the paired observations $(y_{ij}, \mathbf{X}_{ij})$ in $\mathcal{D}$ are independent. Thus,

$$Q(\boldsymbol{\beta}, \phi; \mathbf{I}, \mathcal{D}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} Q\left[\boldsymbol{\beta}, \phi; (y_{ij}, \mathbf{X}_{ij})\right], \tag{3}$$

and it is easy to see that $S(\boldsymbol{\beta}; \mathbf{I}, \mathcal{D})$ is equivalent to $\partial Q(\boldsymbol{\beta}, \phi; \mathbf{I}, \mathcal{D})/\partial \boldsymbol{\beta}$. The likelihood equivalent to $Q(\boldsymbol{\beta}, \phi; \mathbf{I}, \mathcal{D})$ would be the log-likelihood, $\ell(\boldsymbol{\beta}, \phi; \mathcal{D})$, function. For the exponential dispersion family, $Q(\boldsymbol{\beta}, \phi; \mathbf{I}, \mathcal{D})$ and $\ell(\boldsymbol{\beta}, \phi; \mathcal{D})$ are equivalent up to a constant, c. Assuming ϕ is known or estimated by another set of estimating equations, the estimate for $\boldsymbol{\beta}$ is $\boldsymbol{\beta}(\mathbf{R})$ using the estimating equations in 1. Additionally, the $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ has a consistent robust or sandwich based estimator, $\hat{\mathbf{V}}_r$; and, the empirical or model-based estimator is

$$\hat{\boldsymbol{\Omega}}_r = \left.\frac{-\partial^2 Q(\boldsymbol{\beta}; \mathbf{R}, \mathcal{D})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}(\mathbf{R})}.$$

Data transformation

With the robust estimator of $\mathrm{Cov}(\widehat{\boldsymbol{\beta}})$ now being defined, suppose $\mathbf{y}_i^*$ and $\mathbf{X}_i^*$ are $\mathbf{y}_i$ and $\mathbf{X}_i$ transformed by $\mathbf{R}$ such that

$$\begin{aligned}\mathbf{y}_i^* &= \hat{\mathbf{V}}_r^{-1/2} \mathbf{y}_i\\ \mathbf{X}_i^* &= \hat{\mathbf{V}}_r^{-1/2} \mathbf{X}_i,\end{aligned}$$

and $\mathcal{D}^* = \{(\mathbf{y}_1^*, \mathbf{X}_1^*), \ldots, (\mathbf{y}_N^*, \mathbf{X}_N^*)\}$.

Information criterions

The table below is the current collection of the information criterions proposed and investigated. Using the notation defined previously and letting GOF abbreviate goodness-of-fit term, we have the following.

Name	Abbreviation	GOF	Penalty	Notes
"An" information criteria	AIC	$-2\ell(\widehat{\boldsymbol{\beta}}, \mathcal{D})$	$2p$	
Bayesian information criterion	BIC	$-2\ell(\widehat{\boldsymbol{\beta}}, \mathcal{D})$	$\log(N)p$	
Quasi-likelihood under independence model criterion	QIC(R)	$-2Q(\widehat{\boldsymbol{\beta}}(\mathbf{R}), \mathbf{I}, \mathcal{D})$	$2tr(\hat{\boldsymbol{\Omega}}_{\mathbf{I}}\hat{\mathbf{V}}_r)$	
	QICu(R)	$-2Q(\widehat{\boldsymbol{\beta}}(\mathbf{R}), \mathbf{I}, \mathcal{D})$	$2p$	$tr(\hat{\boldsymbol{\Omega}}_{\mathbf{I}}\hat{\mathbf{V}}_r) = p$ if $\hat{\boldsymbol{\Omega}}_{\mathbf{I}} \approxeq \hat{\mathbf{V}}_r$
	BQICu(R)	$-2Q(\widehat{\boldsymbol{\beta}}(\mathbf{R}), \mathbf{I}, \mathcal{D})$	$\log(N)p$	Completely made up by DEB
Correlation information criteria	CIC(R)		$2tr(\hat{\boldsymbol{\Omega}}_{\mathbf{I}}\hat{\mathbf{V}}_r)$	
	AIC*	$-2\ell(\widehat{\boldsymbol{\beta}}^*, \mathcal{D}^*)$	$2p$	
	IC*(R)	$-2\ell(\widehat{\boldsymbol{\beta}}^*, \mathcal{D}^*)$	$2tr(\hat{\boldsymbol{\Omega}}_{\mathbf{I}}\hat{\mathbf{V}}_r)$	
	QIC*(I)	$-2Q(\widehat{\boldsymbol{\beta}}^*(\mathbf{I}), \mathbf{I}, \mathcal{D}^*)$	$2tr(\hat{\boldsymbol{\Omega}}^*_{\mathbf{I}}\hat{\mathbf{V}}^*_r)$	
	QIC*(R)	$-2Q(\widehat{\boldsymbol{\beta}}^*(\mathbf{I}), \mathbf{I}, \mathcal{D}^*)$	$2tr(\hat{\boldsymbol{\Omega}}_{\mathbf{I}}\hat{\mathbf{V}}_r)$	Penalty is based on $\mathcal{D}$ not $\mathcal{D}^*$
	QICu*(I)	$-2Q(\widehat{\boldsymbol{\beta}}^*(\mathbf{I}), \mathbf{I}, \mathcal{D}^*)$	$2p$	
	BQICu*(I)	$-2Q(\widehat{\boldsymbol{\beta}}^*(\mathbf{I}), \mathbf{I}, \mathcal{D}^*)$	$\log(N)p$	

Through simulations it was observed that $-2\ell(\widehat{\boldsymbol{\beta}}^*, \mathcal{D}^*) = -2Q(\widehat{\boldsymbol{\beta}}^*(\mathbf{I}), \mathbf{I}, \mathcal{D}^*) + c$.