# Generalized Estimating Equations (GEE)

In order to estimate generalized linear marginal models (GLMM) we proposed the generalized estimating equations (GEE) estimators. The idea is that a GLMM is specified by

1. A **link function** and conditional mean, $g(\mu_{ij}) = X_{ij}'\beta$.
2. A **variance function** such that $\text{var}(Y_{ij}) = \phi V(\mu_{ij})$.
3. A **pairwise association function** which gives $\text{cor}(Y_{ij}, Y_{ij}) = \mathbf{R}_i(\rho)$.

This allowed us to specify the complete mean and associations of a longitudinal model, without the need for *any* distributional assumptions. We then said that $\beta$ could be estimated by solving a system of estimating equations given by

$$U(\beta) = \sum_{i=1}^{n} D_i' V_i^{-1} \left( Y_i - g^{-1}(X_i\beta) \right).$$

Here $D_i$ is a derivative matrix, $D_i = \frac{\partial}{\partial\beta} g^{-1}(X_i\beta)$ (dimension is going to be $k \times p$), $V_i$ is the (working) covariance matrix $V_i = A_i^{1/2}\mathbf{R}_i(\rho)A_i^{1/2}$ (dimension is going to be $k \times k$), and $g^{-1}()$ is the inverse of the specified link function (dimension of the difference term is $k \times 1$). $\beta$ is as such a $p$-dimensional vector of parameters, and is the core focus of our interest.

We have claimed that this is an M-estimator, and as such, results in consistent and asymptotically normal estimates. We will show that these estimates are unbiased momentarily, but to begin, let's consider three commonly used GEE procedures for different types of data to solidify the concepts!

# Continuous Data with Identity Link

When presenting linear marginal models, we made the assumption that $Y_i \sim \text{MVN}(X_i\beta, \Sigma_i)$, where we considered $\Sigma_i = \sigma^2\mathbf{R}(\rho)$, predominantly, but also considered a more general matrix for the variance. The downside to this specification for continuous data was that it relied on the normality assumption, which is not desirable. We could specify the same type of structure, estimated through a GEE, allowing for robustness to distributional assumptions. If we take $Y_{ij}$ to be a continuous variate, we can define

$$\mu_{ij} = X_{ij}'\beta \quad \text{and} \quad \phi V(\mu_{ij}) = \phi.$$

Additionally, we can take $\mathbf{R}_i(\rho)$ to be our favourite correlation pattern matrix.

With this, we have that

$$\mu_i = X_i\beta \implies D_i = \frac{\partial}{\partial\beta}\mu_i = X_i.$$

Then, our GEE becomes

$$\begin{aligned}U(\beta) &= \sum_{i=1}^{n} D_i'V_i^{-1}\left(Y_i - \mu_i\right) \\ &= \sum_{i=1}^{n} X_i'[\phi\mathbf{R}_i(\rho)]^{-1}\left(Y_i - X_i\beta\right) \\ &= \frac{1}{\phi}\sum_{i=1}^{n} X_i'\mathbf{R}_i(\rho)^{-1}\left(Y_i - X_i\beta\right).\end{aligned}$$

Solving $U(\widehat{\beta}) = 0$ gives us the GEE estimators for $\beta$, which we can see **solve the exact same system of equations as with the LMM we have already seen!** As a result, just like with linear regression, the assumption of normality is not required!

# Binary Longitudinal Data

If we have that $Y_{ij}$ are binary variables, and we wish to model $E[Y_{ij}|\cdot]$ (or, equivalently, the probability that $Y_{ij} = 1$) then the natural choice is to consider *logistic regression*. Recall that generally for binary data we take the logistic link function, $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = x'\beta$, and the variance of binary data is $\mu(1-\mu)$. Taking these as inspiration, we are tempted to set

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = X_{ij}'\beta \implies \mu_{ij} = \left(1 + \exp\left[-X_{ij}'\beta\right]\right)^{-1} \quad \text{and} \quad \phi V(\mu_{ij}) = \phi\mu_{ij}(1-\mu_{ij}).$$

The function $(1+\exp(-x))^{-1}$ is referred to as the inverse-logistic, or **expit** function. For correlation, while any pattern matrix would once again work, it often will make sense to specify an unstructured relationship $\mathbf{R}(\rho) = [\rho_{j\ell}]_{\forall j\neq\ell}$. With this specification, we can consider

$$\begin{aligned}D_i &= \frac{\partial}{\partial\beta}\begin{pmatrix}\text{expit}(X_{i1}'\beta) \\ \vdots \\ \text{expit}(X_{ik}'\beta)\end{pmatrix} \\ &= \begin{pmatrix} X_{i11}\frac{\exp(-X_{i1}'\beta)}{(1+\exp(-X_{i1}'\beta))^2} & X_{i12}\frac{\exp(-X_{i1}'\beta)}{(1+\exp(-X_{i1}'\beta))^2} & \cdots & X_{i1p}\frac{\exp(-X_{i1}'\beta)}{(1+\exp(-X_{i1}'\beta))^2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik1}\frac{\exp(-X_{ik}'\beta)}{(1+\exp(-X_{ik}'\beta))^2} & X_{ik2}\frac{\exp(-X_{ik}'\beta)}{(1+\exp(-X_{ik}'\beta))^2} & \cdots & X_{ikp}\frac{\exp(-X_{ik}'\beta)}{(1+\exp(-X_{ik}'\beta))^2}\end{pmatrix} \\ &= \begin{pmatrix} X_{i11}\mu_{i1}(1-\mu_{i1}) & X_{i12}\mu_{i1}(1-\mu_{i1}) & \cdots & X_{i1p}\mu_{i1}(1-\mu_{i1}) \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik1}\mu_{ik}(1-\mu_{ik}) & X_{ik2}\mu_{ik}(1-\mu_{ik}) & \cdots & X_{ikp}\mu_{ik}(1-\mu_{ik})\end{pmatrix} \\ &= \begin{pmatrix}\mu_{i1}(1-\mu_{i1})X_{i1}' \\ \vdots \\ \mu_{ik}(1-\mu_{ik})X_{ik}'\end{pmatrix}\end{aligned}$$

Here we have used the fact that $\mu_{ij} = \text{expit}(X_{ij}'\beta)$ and that

$$\begin{aligned}\frac{\exp(-x)}{(1+\exp(-x))^2} &= \frac{1}{(1+\exp(-x))}\cdot\frac{1}{\exp(x)(1+\exp(-x))}\\ &= \text{expit}(x)\frac{1}{\exp(x)+1}\\ &= \text{expit}(x)(1-\text{expit}(x)).\end{aligned}$$

If we define

$$\mathbf{A}_i = \begin{pmatrix}\mu_{i1}(1-\mu_{i1}) & 0 & \cdots & 0 \\ 0 & \mu_{i2}(1-\mu_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \mu_{ik}(1-\mu_{ik})\end{pmatrix},$$

then $D_i = \mathbf{A}_i X_i$. Moreover, $A_i$ represents the variance terms and so our working covariance structure is given by

$$V_i = \mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}.$$

We can then combine all of these terms and get that the GEE for $\beta$ is given by the solution to

$$\begin{aligned}U(\beta) &= \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - \mu_i)\\ &= \sum_{i=1}^{n} X_i'\mathbf{A}_i'\left(\mathbf{A}_i^{1/2}\mathbf{R}_i\mathbf{A}_i^{1/2}\right)^{-1}(Y_i - \text{expit}(X_{ij}'\beta)) = \mathbf{0}.\end{aligned}$$

# Count Longitudinal Data

When we have (univariate) count data we commonly will use a Poisson regression, which has a log-link function and where $\text{var}(Y) = E[Y] = \lambda$. In the longitudinal case, we can keep this inspiration, and specify our model as

$$\log(\mu_{ij}) = X_{ij}'\beta \implies \mu_{ij} = \exp(X_{ij}'\beta) \quad \text{and} \quad \phi V(\mu_{ij}) = \phi\mu_{ij}.$$

Again, while in theory any correlation pattern is fine, it is often advisable to start with an unstructured pattern, $\mathbf{R}_i(\rho) = [\rho_{j\ell}]_{\forall j\neq\ell}$. Here we can compute

$$D_i = \frac{\partial}{\partial\beta}\begin{pmatrix}\exp(X_{i1}'\beta)\\ \vdots \\ \exp(X_{ik}'\beta)\end{pmatrix} = \begin{pmatrix}\exp(X_{i1}'\beta)X_{i1}'\\ \vdots \\ \exp(X_{ik}'\beta)X_{ik}'\end{pmatrix}.$$

Defining

$$\mathbf{A}_i = \exp(X_i\beta) = \mu_i,$$

we get that

$$\begin{aligned} U(\beta) &= \sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - \mu_i) \\ &= \sum_{i=1}^{n} X_i' \mathbf{A}_i' (\mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2})^{-1}(Y_i - \exp(X_i\beta)). \end{aligned}$$

Then, solving $U(\widehat{\beta}) = 0$ produces our GEE estimator for $\beta$.

If you recall, in a log-linear model for count data (in the standard GLM case) we often wish to scale our mean by the exposure time through the use of an offset. That is, instead of specifying that $E[Y] = \mu$, we would specify that $E[Y] = \mu t$, where $t$ is the exposure time. That way we can accommodate data which have differing observation periods. The same principle is going to be used for longitudinal data, generally speaking, where we will actually want $\log(\mu_{ij}) = X_{ij}'\beta + \log(\Delta t_{ij})$, where $\Delta t_{ij}$ is the length of the observed interval (for individual $i$ and time $j$). If we have intervals which are evenly spaced then $\Delta t_{ij}$ is constant across all $j$, and including this offset is not necessary (giving us the derived estimators above). However, in the (common) situation where the length of observations differ, then we will want to ensure that this offset is included to properly calibrate our estimates! This will not materially change the analysis as described, but it does mean that $\mu_i = \exp(X_i\beta)\Delta t_i$.

# Nuisance Parameters and Estimation Procedure

Once we have estimated $\beta$, we can define the corresponding residuals of our estimates as $Y_{ij} - \widehat{\mu}_{ij}$. Just as is common with regression analysis, it is typically useful to scale these residuals by the standard deviation of our estimates, which give us the **Pearson residuals**,

$$\widehat{r}_{ij} = \frac{Y_{ij} - \widehat{\mu}_{ij}}{\sqrt{V(\widehat{\mu}_{ij})}}.$$

We can use these residuals to estimate the values of the nuisance parameters, $\rho$ and $\phi$. While it is occasionally the case that we have direct interest in $\rho$ or $\phi$, this will seldom be of scientific interest where our focus will tend to instead be on the mean pattern. Still, if we need, we can use

$$\begin{aligned} \widehat{\phi} &= \frac{1}{n-p} \sum_{i=1}^{n} \sum_{j=1}^{k_i} \widehat{r}_{ij}^2 \\ \widehat{\rho}_{jk} &= \frac{1}{\widehat{\phi}(n-p)} \sum_{i=1}^{n} \widehat{r}_{ij}\widehat{r}_{ik}. \end{aligned}$$

These estimators will not generally be of much interest to us in this course.

Except in special cases the GEE estimators will need to be solved numerically and iteratively, since $\widehat{\beta}$ will depend on $\rho$ and $\phi$, and vice-versa. The general process will then be:

1. Initial a starting value for $\widehat{\beta}$.
2. Use $\widehat{\beta}$ to compute $(\widehat{\phi}, \widehat{\rho})$.
3. Update $\widehat{\beta}$ based on these estimated values, and $U(\beta)$.

Then steps (2) and (3) are repeated until our estimates converge.

# Unbiasedness of Estimating Equations

In order to apply the theory of M-estimators, we need to be able to conclude that (at the true value of $\beta$), $E[U(\beta)] = 0$. In order to do this, suppose that we know that $E[Y_i|X_i] = g^{-1}(X_i\beta)$, exactly as specified. Then

$$\begin{aligned} E\left[U(\beta)\right] &= E\left[\sum_{i=1}^{n} D_i'V_i^{-1}(Y_i - g^{-1}(X_i\beta))\right] \\ &= \sum_{i=1}^{n} E\left[D_i'V_i^{-1}(Y_i - g^{-1}(X_i\beta))\right] \\ &= \sum_{i=1}^{n} E\left[D_i'V_i^{-1}(E[Y_i - g^{-1}(X_i\beta)|X_i])\right] \\ &= \sum_{i=1}^{n} E\left[D_i'V_i^{-1}(E[Y_i|X_i] - g^{-1}(X_i\beta))\right] \\ &= 0. \end{aligned}$$

Here we have used the fact that $D_i$ and $V_i$ are both non-random conditional on $X_i$, and $E[\cdot] = E[E[\cdot|X_i]]$. In particular, this means that the only component which needs to be correctly specified is the mean model! The values for $V_i$ and $D_i$ do not impact the unbiasedness of $U(\beta)$.

We can then apply our standard M-estimator theory which gives asymptotic variance of the estimator as $J^{-1}\Gamma J^{-1'}$, where

$$J = \sum_{i=1}^{n} D_i'V_i^{-1}D_i \quad \text{and} \quad \Gamma = \sum_{i=1}^{n} D_i'V_i^{-1}\operatorname{var}(Y_i)V_i^{-1}D_i.$$

# Model-Based versus Sandwich Variance

The theory of M-estimation gives us the sandwich variance estimator being valid, no matter $D_i$ or $V_i$, which permits us robustness against the misspecification of our working covariance matrix. However, if $V_i$ is correctly specified, such that $\operatorname{var}(Y_i) = V_i$, then note that $\Gamma = \sum_{i=1}^{n} D_i'V_i^{-1}D_i$, and so the variance of our estimators becomes $J^{-1}$.

Because of this simplification, we refer to $J^{-1}$ as the **model based variance estimator**. It is only valid when our work covariance matrix is correctly specified, which we will not typically assume, but this estimator does show-up in the literature and is valid whenever our model is correct.

# Additional Considerations and Conclusions

The estimation of GLMM through GEEs provides a flexible and robust mechanism for estimating longitudinal models for any type of outcome data. There are a few considerations to keep in mind with this process.

1. Parameters in a GLMM are interpreted as population averages. There are **no** individual effects measured, and instead, $\beta_j$ refers to the expected change *over the population.*

2. The linear structure of the marginal model implicitly assumes that $Y_{ij} \perp X_{ik} | X_{ij}$ for all $j \neq k$. This will not generally be the case if $X_{ij\ell}$ is a variable that varies randomly over time, and as a result **GLMMs can only accommodate time-invariant variates or deterministic covariates.**

3. The asymptotic normality of our estimators is a large-sample property. It provides us with the capacity for Wald-type statistics, but relies on large $n$ and the behaviour of these estimates in small samples may be less predictable (particularly when $V_i$ is incorrectly specified).

4. GEE is **not** a likelihood based procedure. As a result **you cannot use likelihood ratio testing** on models fit using GEE. This also means that we cannot use `AIC` **or** `BIC` directly either. Instead, a modified version of `QIC` (quasi information criteria) can be defined and used in place.

5. We can also use a **generalized Score statistic**, which is useful for testing $H_0 : C\beta = 0$ for some constant $C$, where

$$U(\widetilde{\beta}_G)'\mathbf{G}_m C'[C\mathbf{G}_r C']^{-1}C\mathbf{G}_m U(\widetilde{\beta}_G) \overset{H_0}{\sim} \chi^2_r,$$

   where $\widetilde{\beta}_G$ is the GEE estimator under the null, $\mathbf{G}_m = \widehat{J}^{-1}$ and $\mathbf{G}_r = \widehat{J}^{-1}\widehat{\Gamma}\widehat{J}^{-1}$. Here, $r$ is the rank of $C$. As a general rule, this is not implemented in statistical software, so practitioners tend to use the Wald statistic instead.