

First Extension to GEE

To formalize the method of moments estimator for $\boldsymbol{\alpha}$, Prentice (1988) proposed using a second set of estimating equations for $\boldsymbol{\alpha}$. In the context of data with $\text{var}(Y_{ij}) \propto v(\mu_{ij})$:

$$\begin{aligned}\mathbf{G}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^m \mathbf{D}_i^T \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) \\ \mathbf{G}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^m \mathbf{E}_i^T \mathbf{H}_i^{-1} (\mathbf{T}_i - \boldsymbol{\Sigma}_i)\end{aligned}$$

where $R_{ij} = \{Y_{ij} - \mu_{ij}\}/v(\mu_{ij})^{1/2}$, to give “data”

$$\mathbf{T}_i^T = (R_{i1}R_{i2}, \dots, R_{in_i-1}R_{in_i}, R_{i1}^2, \dots, R_{in_i}^2),$$

$\boldsymbol{\Sigma}_i(\boldsymbol{\alpha}) = \text{E}[\mathbf{T}_i]$ is a model for the correlations and variances of the standardized residuals, $\mathbf{E}_i = \frac{\partial \boldsymbol{\Sigma}_i}{\partial \boldsymbol{\alpha}}$, and $\mathbf{H}_i = \text{cov}(\mathbf{T}_i)$ is the working covariance model. \mathbf{G}_1 is of dimension $(k+1) \times 1$ and \mathbf{G}_2 is $a \times 1$ where $k+1 = \dim(\boldsymbol{\beta})$ and $a = \dim(\boldsymbol{\alpha})$.

The vector \mathbf{T}_i has $n_i(n_i-1)/2 + n_i$ (so $\boldsymbol{\Sigma}$ is a model for the covariances and variances, and depends on parameters $\boldsymbol{\alpha}$) elements in general — the working covariance model \mathbf{H}_i is in general complex.

306

Advantages:

- It is straightforward to incorporate a regression model, i.e. $\boldsymbol{\alpha} = f(\boldsymbol{x})$, for the variance-covariance parameters. In `geepack` there are three estimating equations, one each for the mean, scale and correlation parameters.
- If \mathbf{G}_2 is correctly specified then there will be efficiency gains.

Disadvantages:

- If $\text{E}[\mathbf{T}] \neq \boldsymbol{\Sigma}$ then we will not get a consistent estimate of the true variance model — but consistency of $\boldsymbol{\beta}$ through \mathbf{G}_1 is guaranteed, so long as $\hat{\boldsymbol{\alpha}}$ converges to something (intuition: the weight matrix \mathbf{W} is not optimal, but we get consistency of $\boldsymbol{\beta}$ for any choice of non-singular \mathbf{W} — GLS is consistent for any fixed \mathbf{W}).
- For general \mathbf{H} we will require the estimation of fourth order statistics, i.e. $\text{var}(\mathbf{T})$, which is highly unstable unless we have lots of data; $\mathbf{H}_i = \mathbf{I}$, i.e. working independence may be used as an alternative (suggested by DHLZ).

307

Example: Assume for simplicity $n_i = n = 3$ so that

$$\mathbf{T}_i^T = [R_{i1}R_{i2} \ R_{i1}R_{i3} \ R_{i2}R_{i3} \ R_{i1}^2 \ R_{i2}^2 \ R_{i3}^2]$$

with an exchangeable model:

$$\Sigma(\boldsymbol{\alpha})^T = \mathbf{E}[\mathbf{T}_i^T] = [\alpha_1\alpha_2 \ \alpha_1\alpha_2 \ \alpha_1\alpha_2 \ \alpha_1 \ \alpha_1 \ \alpha_1]$$

so that α_1 is the “excess” variance, and α_2 is the correlation on observations on the same unit.

With an identity \mathbf{H}_i the EF for $\boldsymbol{\alpha}$ is

$$\mathbf{G} = \sum_{i=1}^m \begin{bmatrix} \alpha_2 & \alpha_2 & \alpha_2 & 1 & 1 & 1 \\ \alpha_1 & \alpha_1 & \alpha_1 & 0 & 0 & 0 \end{bmatrix} \left(\begin{bmatrix} R_{i1}R_{i2} \\ R_{i1}R_{i3} \\ R_{i2}R_{i3} \\ R_{i1}^2 \\ R_{i2}^2 \\ R_{i3}^2 \end{bmatrix} - \begin{bmatrix} \alpha_1\alpha_2 \\ \alpha_1\alpha_2 \\ \alpha_1\alpha_2 \\ \alpha_1 \\ \alpha_1 \\ \alpha_1 \end{bmatrix} \right)$$

307-1

Hence we need to simultaneously solve the two equations:

$$\begin{aligned} \sum_{i=1}^m \hat{\alpha}_2 \left[\sum_{j<k} R_{ij}R_{ik} - \hat{\alpha}_1\hat{\alpha}_2 \right] + \sum_{j=1}^3 (R_{ij}^2 - \hat{\alpha}_1) &= 0 \\ \sum_{i=1}^m \hat{\alpha}_1 \left[\sum_{j<k} R_{ij}R_{ik} - \hat{\alpha}_1\hat{\alpha}_2 \right] &= 0 \end{aligned}$$

Dividing the second of these by $\hat{\alpha}_1$ gives

$$\hat{\alpha}_1\hat{\alpha}_2 = \frac{1}{3m} \sum_{i=1}^m \sum_{j<k} R_{ij}R_{ik}$$

and substituting this into the first equation gives

$$\hat{\alpha}_1 = \frac{1}{3m} \sum_{i=1}^m \sum_{j<k} R_{ij}^2$$

to give a pair of method-of-moment estimators.

307-2

Second Extension to GEE: Connected Estimating Equations, GEE2^a

In GEE2, there are a *connected* set of joint estimating equations for β and α . Such an approach was proposed by Zhao and Prentice (1990), and Prentice and Zhao (1991).

This approach is particularly appealing if the variance-covariance model is of interest.

To motivate such a pair, consider the following model for a single individual with n *independent* observations:

$$Y_i | \beta, \alpha \sim_{ind} N \{ \mu_i(\beta), \Sigma_i(\beta, \alpha) \},$$

where, for example, we may have $\Sigma_i(\beta, \alpha) = \alpha \mu_i(\beta)^2$, $i = 1, \dots, n$.

We have the likelihood

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

^aGEE1 is the method in which we have a single estimating equation, and a consistent estimator of α .

The score equations are given by

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= -\frac{1}{2} \sum_{i=1}^n \left(\frac{\partial \Sigma_i}{\partial \beta} \right)^T \frac{1}{\Sigma_i} + \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^T \frac{(Y_i - \mu_i)}{\Sigma_i} + \frac{1}{2} \sum_{i=1}^n \left(\frac{\partial \Sigma_i}{\partial \beta} \right)^T \frac{(Y_i - \mu_i)^2}{\Sigma_i^2} \\ &= \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^T \frac{(Y_i - \mu_i)}{\Sigma_i} + \sum_{i=1}^n \left(\frac{\partial \Sigma_i}{\partial \beta} \right)^T \frac{[(Y_i - \mu_i)^2 - \Sigma_i]}{2\Sigma_i^2} \end{aligned} \quad (47)$$

and

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= -\frac{1}{2} \sum_{i=1}^n \left(\frac{\partial \Sigma_i}{\partial \alpha} \right)^T \frac{1}{\Sigma_i} + \frac{1}{2} \sum_{i=1}^n \left(\frac{\partial \Sigma_i}{\partial \alpha} \right)^T \frac{(Y_i - \mu_i)^2}{\Sigma_i^2} \\ &= \sum_{i=1}^n \left(\frac{\partial \Sigma_i}{\partial \alpha} \right)^T \frac{[(Y_i - \mu_i)^2 - \Sigma_i]}{2\Sigma_i^2} \end{aligned} \quad (48)$$

This pair of quadratic estimating functions, are unbiased given correct specification of the first two moments – so note that if the variance model is wrong, we are no longer guaranteed a consistent estimator of β .

If it's correct, however, there will be a gain in efficiency.

Let

$$S_i = (Y_i - \mu_i)^2$$

with, under the model,

$$\begin{aligned} E[S_i] &= \Sigma_i \\ \text{var}(S_i) &= E[S_i^2] - E[S_i]^2 = 3\Sigma_i^2 - \Sigma_i^2 = 2\Sigma_i^2 \end{aligned}$$

Hence we can rewrite (47) and (48)

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \mathbf{D}_i^T \mathbf{V}_i^{-1} (Y_i - \mu_i) + \sum_{i=1}^n \mathbf{E}_i \mathbf{W}_i^{-1} (S_i - \Sigma_i) \\ \frac{\partial l}{\partial \boldsymbol{\alpha}} &= \sum_{i=1}^n \mathbf{F}_i \mathbf{W}_i^{-1} (S_i - \Sigma_i) \end{aligned}$$

where $\mathbf{D}_i = \partial \mu_i / \partial \boldsymbol{\beta}$, $\mathbf{E}_i = \partial \Sigma_i / \partial \boldsymbol{\beta}$ and $\mathbf{F}_i = \partial \Sigma_i / \partial \boldsymbol{\alpha}$.

This can be compared with the usual estimating equation specification:

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \mathbf{D}_i^T \mathbf{V}_i^{-1} (Y_i - \mu_i).$$

310

GEE2 Continued

The general form of estimating equations, in the dependent data setting, is given by

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

where $\mathbf{D}_i = \partial \mu_i / \partial \boldsymbol{\beta}$, $\mathbf{E}_i = \partial \Sigma_i / \partial \boldsymbol{\beta}$ and $\mathbf{F}_i = \partial \Sigma_i / \partial \boldsymbol{\alpha}$, and we have “working” variance-covariance structure

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

311

When $\mathbf{C}_i = \mathbf{0}$ we obtain

$$\begin{aligned}\mathbf{G}_1(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) + \sum_{i=1}^m \mathbf{E}_i \mathbf{W}_i^{-1} (\mathbf{S}_i - \boldsymbol{\Sigma}_i) \\ \mathbf{G}_2(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^m \mathbf{F}_i \mathbf{W}_i^{-1} (\mathbf{S}_i - \boldsymbol{\Sigma}_i)\end{aligned}$$

which are the dependent data version of the normal score equations we obtained earlier.

Prentice and Zhao show that these equations arise from a so-called *quadratic exponential model*:

$$p(\mathbf{Y}_i | \boldsymbol{\theta}_i, \boldsymbol{\lambda}_i) = k_i^{-1} \exp[\mathbf{Y}_i^T \boldsymbol{\theta}_i + \mathbf{S}_i^T \boldsymbol{\lambda}_i + c_i(\mathbf{Y}_i)].$$

For example, $c_i = 0$ gives the multivariate normal.

For consistency of $\hat{\boldsymbol{\beta}}$ we require models for both \mathbf{Y}_i and \mathbf{S}_i to be correct — increased efficiency if models are correct.

312

CHAPTER 10: BINARY DATA MODELS

We devote an entire chapter to binary data since such data are challenging, both in terms of modeling the dependence, and parameter interpretation.

We again consider mixed effects models approached (with inference from likelihood or Bayes perspectives), and GEE.

Motivating Data - Indonesian Child Health

DHLZ describe a dataset in which the response is the absence/presence of respiratory illness in 275 children, with multiple measurements being collected over time – quarterly measurements were taken for up to six consecutive quarters. Age also recorded.

Question of interest:

Is the prevalence of respiratory infection higher amongst children who suffer from xerophthalmia, an ocular manifestation of chronic vitamin A deficiency?

Figure 36 shows the infection indicator versus time for the first 49 children.

313

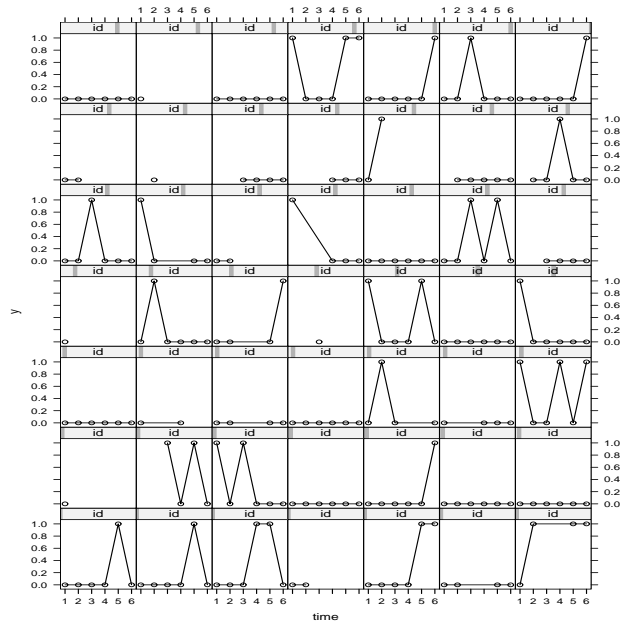


Figure 36: Respiratory infection versus time for the first 49 children.

314

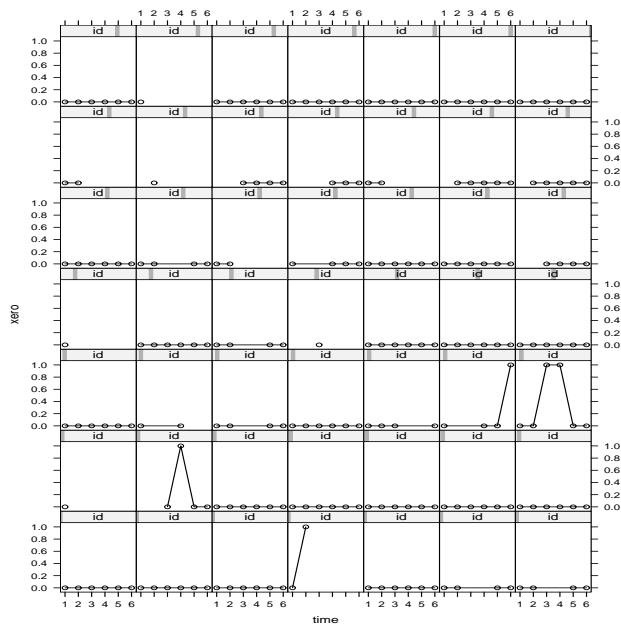


Figure 37: Xerophthalmia versus time for the first 49 children.

315

Exploratory Summaries

With binary responses and covariates, plots are not so informative - instead look at tables.

```
> table(n) # no of Y=1 counts per child (ie 22 children one y=1),1200 in total
  1  2  3  4  5  6
22 32 29 55 15 122
> xtabs(~xerocross+ycross) # cross-sectional comparison of respiratory
      ycross          # event (ycross) and xerophthalmia (xerocross)
xerocross  0  1          # at first time point
          0 234 31
          1  9  1
> 234/(9*31)
[1] 0.8387097
> summary(glm(cbind(ycross,1-ycross) ~ xerocross, family="binomial"))
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.0213      0.1911 -10.576  <2e-16 ***
xerocross    -0.1759      1.0712  -0.164    0.87
> exp(-.1759)
[1] 0.8387018
```

316

Now add in age (potential confounder).

```
> modglm <- glm(cbind(ycross,1-ycross) ~ agecross+xerocross, family="binomial")
> summary(modglm) # Cross-sectional analyses
Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.17017      0.21785  -9.962  <2e-16 ***
agecross     -0.02395      0.01059  -2.262   0.0237 *
xerocross     0.22630      1.09367   0.207   0.8361
Null deviance: 197.79  on 274  degrees of freedom
Residual deviance: 192.17  on 272  degrees of freedom
> modglmq <- glm(cbind(ycross,1-ycross) ~ agecross+xerocross,family=quasibinomial())
> summary(modglmq)
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.17017      0.21493 -10.097  <2e-16 ***
agecross     -0.02395      0.01045  -2.293   0.0226 *
xerocross     0.22630      1.07900   0.210   0.8340
(Dispersion parameter for quasibinomial family taken to be 0.9733635)
Null deviance: 197.79  on 274  degrees of freedom
Residual deviance: 192.17  on 272  degrees of freedom
```

317

Overview of Modeling Clustered Binary Data

For a single binary variable, Y , all moments are determined by $p = E[Y]$. Specifically $E[Y^r] = p$ for $r \geq 1$.

Consider n observations from a single individual $\mathbf{Y} = (Y_1, \dots, Y_n)^T$.

The saturated model has $2^n - 1$ parameters (in contrast to n means, n variances and $n(n-1)/2$ correlations for the saturated normal model).

There is no obvious or natural manner by which multivariate binary data can be specified. We begin with a random effects formulation that induces correlations.

318

Logistic Mixed Effects Models

A GLMM for binary data takes the binomial exponential family, with canonical link being logistic.

We have

Stage 1: $Y_{ij} \sim_{ind} \text{Binomial}(n_{ij}, p_{ij})$ with

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_{ij}\mathbf{b}_i$$

Stage 2: $\mathbf{b}_i \sim_{iid} N(\mathbf{0}, \mathbf{D})$.

Marginal moments are not available in closed form.

319

Conjugate model

$$\begin{aligned} Y | p &\sim \text{Binomial}(n, p) \\ p &\sim \text{Beta}(a, b) \end{aligned}$$

gives

$$\Pr(Y = y) = \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+y)\Gamma(b+n-y)}{\Gamma(a+b+n)}, \quad y = 0, 1, \dots, n.$$

Marginal moments:

$$\mathbb{E}[Y] = n\mathbb{E}[p] = n \frac{a}{a+b}, \quad \text{var}(Y) = n\mathbb{E}[p](1 - \mathbb{E}[p]) \frac{a+b+n}{a+b+1}$$

Limitations: likelihood is complex, extension to covariates increases complexity, and no tractable way to allow random effect slopes.

320

Parameter Interpretation

For the random intercepts model the conditional parameters β^c and marginal parameters β^m are approximately linked through

$$\begin{aligned} \mathbb{E}[Y] &= \frac{\exp(\mathbf{x}\beta^m)}{1 + \exp(\mathbf{x}\beta^m)} = \mathbb{E}_b\{E[Y|b]\} \\ &= \mathbb{E}_b \left[\frac{\exp(\mathbf{x}\beta^c + b)}{1 + \exp(\mathbf{x}\beta^c + b)} \right] \approx \frac{\exp(\mathbf{x}\beta^c / [c^2\sigma^2 + 1]^{1/2})}{1 + \exp(\mathbf{x}\beta^c / [c^2\sigma^2 + 1]^{1/2})} \end{aligned}$$

where $c = 16\sqrt{3}/(15\pi)$. Hence the marginal coefficients are attenuated towards zero; Figure 38 illustrates for particular values of $\beta_0, \beta_1, \sigma_0^2$.

For the model

$$\log \left(\frac{\mathbb{E}[Y | \mathbf{b}]}{1 - \mathbb{E}[Y | \mathbf{b}]} \right) = \mathbf{x}\beta^c + \mathbf{z}\mathbf{b}$$

where $\mathbf{b} \sim_{iid} N_{q+1}(\mathbf{0}, \mathbf{D})$ we obtain

$$\mathbb{E}[Y] \approx \frac{\exp(\mathbf{x}\beta^c | c^2 \mathbf{D} \mathbf{z} \mathbf{z}^T + \mathbf{I}_{q+1} |^{-(q+1)/2})}{1 + \exp(\mathbf{x}\beta^c | c^2 \mathbf{D} \mathbf{z} \mathbf{z}^T + \mathbf{I}_{q+1} |^{-(q+1)/2})}$$

so that

$$\beta^m \approx |c^2 \mathbf{D} \mathbf{z} \mathbf{z}^T + \mathbf{I}_{q+1} |^{-(q+1)/2} \beta^c$$

321

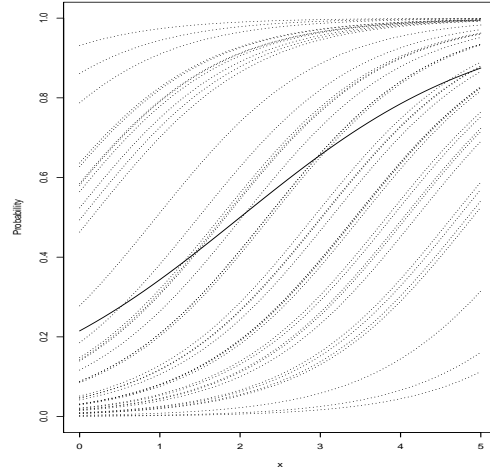


Figure 38: Individual-level curves (dotted lines) from random intercepts logistic GLMM with $\log(\text{E}[Y | b]) / (1 - \text{E}[Y | b]) = \beta_0 + \beta_1 x$, with $\beta_0 = -2, \beta_1 = 1$ and $b \sim_{iid} N(0, 2^2)$, along with marginal curve (solid curve). Approximate attenuation is 1.54.

322

```
> library(lme4)
> memmod1 <- lmer(cbind(y,1-y) ~ age+xero+(1|id),family=binomial,method="PQL")
> summary(memmod1)
Random effects:
      Groups      Name      Variance      Std.Dev.
      id (Intercept)      3.7334      1.9322
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.6920622  0.1795893 -14.9901 < 2.2e-16 ***
age          -0.0229009  0.0088506  -2.5875  0.009667 **
xero         0.1219975  0.5379760   0.2268  0.820602
memmod2 <- lmer(cbind(y,1-y) ~ age+xero+(1|id),family=binomial,method="Laplace")
> summary(memmod2)
Random effects:
      Groups      Name      Variance      Std.Dev.
      id (Intercept)      3.7334      1.9322
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.6920622  0.1810916 -14.8658 < 2e-16 ***
age          -0.0229009  0.0089756  -2.5515  0.01073 *
xero         0.1219975  0.5380319   0.2267  0.82062
```

323