Generalized Estimating Equations

Suppose we assume

$$\mathrm{E}[\boldsymbol{Y}_i \mid \boldsymbol{\beta}] = \boldsymbol{x}_i\boldsymbol{\beta},$$

and consider the $n_i \times n_i$ *working* variance-covariance matrix:

$$\mathrm{var}(\boldsymbol{Y}_i \mid \boldsymbol{\beta}, \boldsymbol{\alpha}) = \boldsymbol{W}_i.$$

To motivate GEE we begin by assuming that $\boldsymbol{W}_i$ is known. In this case the GLS estimator minimizes

$$\sum_{i=1}^{m} (\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^{\mathrm{T}} \boldsymbol{W}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}),$$

and is given by the solution to the estimating function

$$\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}),$$

which is

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{x}_i \right)^{-1} \sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{Y}_i.$$

We now examine the properties of this estimator.

We have

$$\mathrm{E}[\widehat{\boldsymbol{\beta}}] = \left(\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{x}_i \right)^{-1} \sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \mathrm{E}[\boldsymbol{Y}_i] = \boldsymbol{\beta},$$

so long as the mean is correctly specified.

If the information about $\boldsymbol{\beta}$ grows with increasing m, then $\widehat{\boldsymbol{\beta}}$ is consistent.

The variance, $\mathrm{var}(\widehat{\boldsymbol{\beta}})$, is given by

$$\left(\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{x}_i \right)^{-1} \left(\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \mathrm{var}(\boldsymbol{Y}_i) \boldsymbol{W}_i^{-1} \boldsymbol{x}_i \right) \left(\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{x}_i \right)^{-1}.$$

Notice that if the assumed variance-covariance matrix is correct, i.e. $\mathrm{var}(\boldsymbol{Y}_i) = \boldsymbol{W}_i$, then

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{x}_i \right)^{-1},$$

and a Gauss-Markov Theorem shows that, in this case, the estimator is efficient amongst linear estimators (see Exercises).

If m is large then a multivariate central limit theorem shows that $\widehat{\boldsymbol{\beta}}$ is asymptotically normal.

We now suppose that $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{W}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is of known form, where $\boldsymbol{\alpha}$ are parameters in the variance-covariance model, which we begin by assuming are known.

The regression parameters are contained in $\boldsymbol{W}_i$ to allow, mean-variance relationships, for example,

$$\begin{aligned} \text{var}(Y_{ij} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \alpha_1 \mu_{ij}^2 \\ \text{cov}(Y_{ij}, Y_{ik} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \alpha_1 \alpha_2^{|t_{ij}-t_{ik}|} \mu_{ij}\mu_{ik} \end{aligned}$$

where $\mu_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta}$, α_1 is the variance (which is assumed constant across time and across individuals), and α_2 is the correlation (which is assumed to be the same for all individuals), and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$.

For known $\boldsymbol{\alpha}$ we would minimize

$$\sum_{i=1}^m (\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^{\text{T}} \boldsymbol{W}_i^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}),$$

with solution given by the root of the estimating function

$$\sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}).$$

In general finding the roots of this equation is not available in closed form.

However, if $\boldsymbol{W}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{W}_i(\boldsymbol{\alpha})$ we have

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i(\boldsymbol{\alpha})^{-1} \boldsymbol{x}_i \right)^{-1} \sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i^{-1}(\boldsymbol{\alpha}) \boldsymbol{Y}_i.$$

Finally, suppose that $\boldsymbol{\alpha}$ is unknown but we have a procedure by which a consistent estimator $\widehat{\boldsymbol{\alpha}}$ is produced.

We then solve the estimator function

$$\boldsymbol{G}(\boldsymbol{\beta}) = \sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i(\widehat{\boldsymbol{\alpha}}, \boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}).$$

In general iteration is needed to simultaneously estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Let $\widehat{\boldsymbol{\alpha}}^{(0)}$ be an initial estimate. Then set $t = 0$ and iterate between

1. Solve $\boldsymbol{G}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}^{(t)}) = \mathbf{0}$ to give $\widehat{\boldsymbol{\beta}}^{(t+1)}$,
2. Estimate $\widehat{\boldsymbol{\alpha}}^{(t+1)}$ with $\widehat{\mu}_i = \mu_i \left(\widehat{\boldsymbol{\beta}}^{(t+1)} \right)$. Set $t \to t+1$ and return to 1.

We have

$$\text{var}(\widehat{\boldsymbol{\beta}})^{1/2} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \text{N}_{k+1}\left(\mathbf{0}, \boldsymbol{I}\right),$$

where

$$\begin{aligned} \widehat{\text{var}}(\widehat{\boldsymbol{\beta}}) &= \left(\sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i^{-1}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) \boldsymbol{x}_i \right)^{-1} \\ &\times \left(\sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i^{-1}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) \text{var}(\boldsymbol{Y}_i) \boldsymbol{W}_i^{-1}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) \boldsymbol{x}_i \right) \\ &\times \left(\sum_{i=1}^m \boldsymbol{x}_i^{\text{T}} \boldsymbol{W}_i^{-1}(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) \boldsymbol{x}_i \right)^{-1}. \end{aligned}$$

We have assumed that $\text{cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_{i'}) = 0$ for $i \neq i'$.

The final element of GEE is sandwich estimation of $\text{var}(\widehat{\boldsymbol{\beta}})$. In particular $\text{cov}(\boldsymbol{Y}_i)$ is estimated by

$$(\boldsymbol{Y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{Y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}),$$

Empirical would be a better word than *robust* for the estimator of the variance – not robust to sample size, could be highly unstable.

We can write the $(k+1) \times 1$ estimating function as

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{W}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{\beta}) \quad (2)$$

$$\sum_{i=1}^{m} \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{W}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta}) \quad (3)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}[\boldsymbol{x}_{i1} \cdots \boldsymbol{x}_{in_i}] \begin{bmatrix} W_i^{11} & \cdots & W_i^{1n_i} \\ \cdots & \cdots & \cdots \\ W_i^{n_i1} & \cdots & W_i^{n_in_i} \end{bmatrix} \begin{bmatrix} Y_{i1} - \boldsymbol{x}_{i1}\boldsymbol{\beta} \\ \cdots \\ Y_{in_i} - \boldsymbol{x}_{in_i}\boldsymbol{\beta} \end{bmatrix} \quad (4)$$

where W_i^{ij} denotes entry (i,j) of the inverse $\boldsymbol{W}_i$.

We emphasize (4) since this form emphasizes that the basic unit of replication is indexed by i.

For example, the asymptotics depend on $m \to \infty$.

Example: Suppose for simplicity that we have a balanced design, with $n_i = n$ for all i, and

$$\begin{aligned} \text{var}(Y_{ij}) &= \mathrm{E}[(Y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\beta})^2] = \mathrm{E}[\epsilon_{ij}^2] = \alpha_1 \\ \text{cov}(Y_{ij}, Y_{ik}) &= \mathrm{E}[(Y_{ij} - \boldsymbol{x}_{ij}\boldsymbol{\beta})(Y_{ik} - \boldsymbol{x}_{ik}\boldsymbol{\beta})] = \mathrm{E}[\epsilon_{ij}\epsilon_{ik}] = \alpha_1\alpha_{2jk}, \end{aligned}$$

for $i = 1, ..., m; j, k = 1, ..., n; j \neq k$. Hence we have $n + n(n-1)/2$ elements of $\boldsymbol{\alpha}$.

Letting

$$e_{ij} = Y_{ij} - \boldsymbol{x}_{ij}\widehat{\boldsymbol{\beta}},$$

method-of-moments estimators are given by

$$\widehat{\alpha}_1 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} e_{ij}^2,$$

and

$$\widehat{\alpha}_1\widehat{\alpha}_{2jk} = \frac{1}{m}\sum_{i=1}^{m} e_{ij}e_{ik}.$$

Generalized Estimating Equation (GEE) Summary

We have:

- Regression parameters (of primary interest) $\boldsymbol{\beta}$ and,
- Variance-covariance parameters $\boldsymbol{\alpha}$.

We have considered the GEE

$$\boldsymbol{G}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0},$$

where

- $\mu_i = \mu_i(\boldsymbol{\beta}) = \boldsymbol{x}_i \boldsymbol{\beta}$.
- $\boldsymbol{D}_i = \boldsymbol{D}_i(\boldsymbol{\beta}) = \frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \boldsymbol{x}_i^{\mathrm{T}}$,
- $\boldsymbol{W}_i = \boldsymbol{W}_i(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the “working” covariance model,

Three important ideas:

1. Separate estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.
2. Sandwich estimation of $\mathrm{var}(\widehat{\boldsymbol{\beta}})$.
3. Replication across units in order to estimate covariances – so we have assumed that observations on different units are independent.

Notes:

- We have seen the first and second ideas in independent data situations – e.g. estimation of the α parameter in the quadratic negative binomial model.
- We may use method of moments estimators for $\boldsymbol{\alpha}$ (or set up another estimating equation, see later).
- In a dependent data situation we could just follow 1, and go with model-based standard errors:

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right)^{-1}. \quad (5)$$

The sandwich estimator of $\mathrm{var}(\widehat{\boldsymbol{\beta}})$ is given by

$$\left(\sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right)^{-1} \left\{ \sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \mathrm{cov}(\boldsymbol{Y}_i) \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right\} \left(\sum_{i=1}^{m} \boldsymbol{D}_i^{\mathrm{T}} \boldsymbol{W}_i^{-1} \boldsymbol{D}_i \right)^{-1}$$

Substitution of $\mathrm{cov}(\boldsymbol{Y}_i) = \boldsymbol{W}_i = \boldsymbol{V}_i$ (where $\boldsymbol{V}_i$ is the “true” covariance model) in the above gives (5).

To implement the GEE we need to, in general, iterate between estimation of $\boldsymbol{\beta} | \widehat{\alpha}$ and $\boldsymbol{\alpha} | \widehat{\beta}$.

If we have an independence working model ($\boldsymbol{W}_i = \boldsymbol{I}$) then no iteration necessary (since no $\boldsymbol{\alpha}$ in the GEE) – in this case we’d want to use sandwich estimation, however.

Dental Example

We analyze the dental data using LMEs and GEE.

First we plot the data using a "trellis" plot.

```
> library(nlme)
> data(Orthodont)
> Orthgirl <- Orthodont[Orthodont$Sex=="Female",]
> trelldat <- groupedData( distance ~ age | Subject, data=Orthgir]
> plot(trelldat)
```

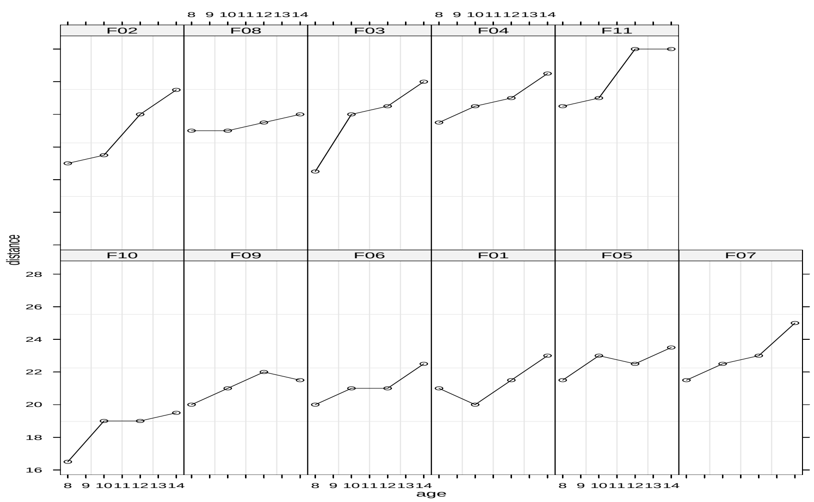

Figure 6: Length versus age (in years) for 11 girls.

Generalized Estiamting Equations

Look at various estimators of $\boldsymbol{\beta}$ for girls only. Note here that we might doubt the asymptotics for GEE since we only have replication across $m = 11$ units (girls).

Start with ordinary least squares – unbiased estimator for $\boldsymbol{\beta}$, but standard errors are wrong because independence is assumed.

```
> summary(lm(distance~age,data=Orthgirl))

Call:
lm(formula = distance ~ age, data = Orthgirl)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.3727     1.6378  10.608 1.87e-13 ***
age           0.4795     0.1459   3.287  0.00205 **
Residual standard error: 2.164 on 42 degrees of freedom
Multiple R-Squared: 0.2046,     Adjusted R-squared: 0.1856
F-statistic:  10.8 on 1 and 42 DF,  p-value: 0.002053
```

Now implement GEE with working independence – the following is an R implementation (in Splus we would use gee()).

```
install.packages("geepack",lib="/home/faculty/jonno/teaching/571/2005/notes/examples/geepack")
.libPaths("/home/faculty/jonno/teaching/571/2005/notes/examples/geepack")
library(geepack)
```

```
> summary(geese(distance~age,id=Subject,data=Orthgirl,
corstr="independence"))
Call:
geese(formula = distance ~ age, id = Subject, data = Orthgirl,
    corstr = "independence")

Mean Model:
 Mean Link:                 identity
 Variance to Mean Relation: gaussian
 Coefficients:
              estimate    san.se      wald            p
(Intercept) 17.3727273 0.7819784 493.56737 0.000000e+00
age          0.4795455 0.0666386  51.78547 6.190604e-13
Scale Model:
 Scale Link:                identity
 Estimated Scale Parameters:
            estimate   san.se     wald           p
(Intercept) 4.470403 1.373115 10.59936 0.001131270
Correlation Model:
 Correlation Structure:     independence
Returned Error Value:    0
Number of clusters:   11   Maximum cluster size: 4
```


Next we examine an exchangeable correlation structure in which all pairs of observations on the same unit have a common correlation:

```
> summary(geese(distance~age,id=Subject,data=Orthgirl,
corstr="exchangeable"))
Call:
geese(formula = distance ~ age, id = Subject, data = Orthgirl,
    corstr = "exchangeable")
Mean Model:
 Mean Link:                 identity
 Variance to Mean Relation: gaussian
 Coefficients:
              estimate    san.se      wald            p
(Intercept) 17.3727273 0.7819784 493.56737 0.000000e+00
age          0.4795455 0.0666386  51.78547 6.190604e-13
Scale Model:
 Scale Link:                identity
 Estimated Scale Parameters:
            estimate   san.se     wald           p
(Intercept) 4.470403 1.373115 10.59936 0.001131270
Correlation Model:
 Correlation Structure:     exchangeable
 Correlation Link:          identity
 Estimated Correlation Parameters:
       estimate    san.se     wald            p
alpha 0.8680178 0.1139327 58.04444 2.564615e-14
Returned Error Value:    0
Number of clusters:   11   Maximum cluster size: 4
```

Notes:

- Independence estimates are always identical to OLS because we have assumed working independence, which means that the estimating equation is the same as the normal equations.
- Standard errors are smaller because regressor (time) is changing within an individual.
- Here we obtain the same estimates for exchangeable as working independence but only because balanced and complete (i.e. no missing) data.

Finally we look at AR(1) errors – this time we see slight differences in estimates and standard errors.

```
> summary(gee(gdistance~gage,id=gSubject,corstr="ar1"))
Call:
geese(formula = distance ~ age, id = Subject, data = Orthgirl,
    corstr = "ar1")
Mean Model:
 Mean Link:                 identity
 Variance to Mean Relation: gaussian
 Coefficients:
             estimate     san.se      wald            p
(Intercept) 17.3049830 0.85201953 412.51833 0.000000e+00
age          0.4848065 0.06881228  49.63692 1.849965e-12
Scale Model:
 Scale Link:                identity
 Estimated Scale Parameters:
            estimate   san.se   wald            p
(Intercept) 4.470639 1.341802 11.101 0.0008628115

Correlation Model:
 Correlation Structure:     ar1
 Correlation Link:          identity
 Estimated Correlation Parameters:
       estimate     san.se     wald p
alpha 0.9298023 0.07164198 168.4403 0
Returned Error Value:    0
Number of clusters:   11   Maximum cluster size: 4
```

Now delete last two observations from girl 11 to illustrate that identical answers before were consequence of balance and completeness of data.

```
> Orthgirl2<-Orthgirl[1:42,]
> summary(lm(distance~age,data=Orthgirl2))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.0713     1.5102  11.966 8.56e-15 ***
age           0.3963     0.1357   2.921  0.00571 **
Residual standard error: 1.964 on 40 degrees of freedom


> summary(geese(distance~age,id=Subject,data=Orthgirl2,
corstr="independence"))
Coefficients:
              estimate     san.se      wald            p
(Intercept) 18.0713312 0.82603439 478.61250 0.000000e+00
age          0.3962971 0.06934195  32.66253 1.096304e-08
Scale Model:
 Scale Link:                identity
 Estimated Scale Parameters:
            estimate   san.se     wald           p
(Intercept) 3.674926 1.317669 7.778294 0.005287771
Correlation Model:
 Correlation Structure:     independence
Returned Error Value:    0
Number of clusters:   11   Maximum cluster size: 4
```



```
> summary(geese(distance~age,id=Subject,data=Orthgirl2,
corstr="exchangeable"))
Call:
geese(formula = distance ~ age, id = Subject, data = Orthgirl2,
    corstr = "exchangeable")
Mean Model:
 Mean Link:                 identity
 Variance to Mean Relation: gaussian
 Coefficients:
              estimate     san.se      wald            p
(Intercept) 17.6050097 0.79007168 496.52320 0.000000e+00
age          0.4510122 0.06641218  46.11913 1.112765e-11
Scale Model:
 Scale Link:                identity
 Estimated Scale Parameters:
            estimate   san.se    wald           p
(Intercept) 3.706854 1.320019 7.88589 0.004982194
Correlation Model:
 Correlation Structure:     exchangeable
 Correlation Link:          identity
 Estimated Correlation Parameters:
       estimate     san.se     wald p
alpha 0.7968515 0.09367467 72.36198 0
Returned Error Value:    0
Number of clusters:   11   Maximum cluster size: 4
```

REML

Recall that in the linear model for independent data the MLE for σ^2 has finite sample bias since there is no degrees of freedom adjustment for estimation of $\boldsymbol{\beta}$.

This is also true in the dependent data case. One remedy to this is known as Restricted Maximum Likelihood (REML).

Last quarter we saw a justification for this in terms of marginal likelihood, we now provide another.

Suppose we place a flat prior on $\boldsymbol{\beta}$, i.e. $\pi(\boldsymbol{\beta}) \propto 1$, and then integrate out $\boldsymbol{\beta}$ to obtain the "likelihood":

$$p(\boldsymbol{y}|\sigma_\epsilon^2, \sigma_0^2) = \int_{\boldsymbol{\beta}} p(\boldsymbol{y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_0^2)\pi(\boldsymbol{\beta})\ d\boldsymbol{\beta}.$$

which may be maximized with respect to $\sigma_\epsilon^2, \sigma_0^2$.

Simple Example of REML

Consider the linear regression for independent data: $\boldsymbol{Y}|\boldsymbol{\beta}, \sigma^2 \sim N(\boldsymbol{x\beta}, \boldsymbol{I}_n\sigma^2)$, with $\dim(\boldsymbol{\beta}) = k+1$.

Consider

$$p(\boldsymbol{y}|\sigma^2) = \int p(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\beta})d\boldsymbol{\beta},$$

and assume $\pi(\boldsymbol{\beta}) \propto 1$ so that

$$\begin{aligned} p(\boldsymbol{y}|\sigma^2) &= \int (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{x\beta})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x\beta})\right] d\boldsymbol{\beta} \\ &= (2\pi\sigma^2)^{-n/2} \int \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{x}\widehat{\boldsymbol{\beta}}+\boldsymbol{x}\widehat{\boldsymbol{\beta}}-\boldsymbol{x\beta})^{\mathrm{T}}\right. \\ &\times \left.(\boldsymbol{y}-\boldsymbol{x}\widehat{\boldsymbol{\beta}}+\boldsymbol{x}\widehat{\boldsymbol{\beta}}+\boldsymbol{x\beta})\right] d\boldsymbol{\beta} \\ &= (2\pi\sigma^2)^{-(n-k-1)/2} \exp\left[-\frac{RSS}{2\sigma^2}\right] |\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}|^{-1/2} \end{aligned}$$

where the residual sum of squares

$$RSS = (\boldsymbol{y}-\boldsymbol{x}\widehat{\boldsymbol{\beta}})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}\widehat{\boldsymbol{\beta}}).$$

Maximization of $l(\sigma^2) = p(\boldsymbol{y}|\sigma^2)$ yields the unbiased estimator

$$\widehat{\sigma}^2 = \frac{RSS}{n-k-1}.$$

LME Example of REML

Again obtain the distribution of the data as a function of $\boldsymbol{\alpha}$ only, by integrating $\boldsymbol{\beta}$ from the model, and assuming an improper flat prior for $\boldsymbol{\beta}$.

We have

$$p(\boldsymbol{y}|\boldsymbol{\alpha}) = \int_{\boldsymbol{\beta}} p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}) \times \pi(\boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{\beta},$$

leading to

$$\begin{aligned} l(\boldsymbol{\alpha}) &= \log p(\boldsymbol{y}|\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{i=1}^{m} \log |\boldsymbol{V}_i(\boldsymbol{\alpha})| \\ &- \frac{1}{2}\sum_{i=1}^{m} \log |\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{V}_i(\boldsymbol{\alpha}) \boldsymbol{x}_i| \\ &- \frac{1}{2}\sum_{i=1}^{m} (\boldsymbol{y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}})^{\mathrm{T}} \boldsymbol{V}^{-1}(\boldsymbol{\alpha})(\boldsymbol{y}_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}), \end{aligned}$$

which differs from the "usual" likelihood by the term

$$-\frac{1}{2}\sum_{i=1}^{m} \log |\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{V}_i(\boldsymbol{\alpha}) \boldsymbol{x}_i|.$$

This expression as the same as that which results from the maximization of the distribution of the residuals.

In nearly all cases MLE of $\boldsymbol{\alpha}$ are not available in closed form – hence use (for example) `lme()` in `R`.

Estimates of $\boldsymbol{\beta}$ change since they are a function of $\widehat{\boldsymbol{\alpha}}$.

Linear Mixed Effects Models

We fit using MLE and REML.

```
> remlelm <- lme( distance ~ age, data = Orthgirl,
random = ~1 | Subject )
> summary(remlelm)
Linear mixed-effects model fit by REML
Random effects:
 Formula: ~1 | Subject
        (Intercept)  Residual
StdDev:     2.06847 0.7800331
Fixed effects: distance ~ age
                Value Std.Error DF   t-value p-value
(Intercept) 17.372727 0.8587419 32 20.230440       0
age          0.479545 0.0525898 32  9.118598       0
> mlelm <- lme( distance ~ age, data = Orthgirl,
random = ~1 | Subject, method = "ML" )
```

```
> summary(mlelm)
Linear mixed-effects model fit by maximum likelihood
Random effects:
 Formula: ~1 | Subject
        (Intercept)  Residual
StdDev:    1.969870 0.7681235
Fixed effects: distance ~ age
                Value Std.Error DF   t-value p-value
(Intercept) 17.372727 0.8506287 32 20.423397       0
age          0.479545 0.0530056 32  9.047078       0
```