

Integration in the GLMM

As with the NLMEM there are a number of possible approaches for integrating out the random effects including:

- Analytical approximations, including Laplace, and the closely-related penalized quasi-likelihood approach.
- Gaussian quadrature.
- Importance sampling Monte Carlo

268

Penalized Quasi-Likelihood

Breslow and Clayton (1993) introduced the method of Penalized Quasi-Likelihood (PQL) which was an attempt to extend quasi-likelihood to GLMMs. One justification of the method is a first-order Laplace approximation.

PQL is very poor for binary data but may be OK for binomial and Poisson data (as long as the counts are not too small).

Within the `lme4` package the `lmer` function may be used to fit GLMMs using MLE/REML; the required integrals can be approximated using penalized quasi-likelihood, Laplace, or adaptive Gaussian quadrature.

269

GLMMs for the Seizure Data

PQL standard error for β_1 looks off here (doesn't tie in with later analyses).

Adaptive quadrature option is not available for this model.

```
> library(lme4) # Need Matrix package version 0.995-5
> lmermod1 <- lmer(y ~ x1+x2+x3+(1|ID)+offset(log(time)),family=poisson,
  data=seiz,method="PQL")
> summary(lmermod1)
Generalized linear mixed model fit using PQL
Random effects:
   Groups             Name         Variance   Std.Dev.
   ID (Intercept)      0.20035    0.44761
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.076279   0.092852 11.5914 < 2e-16 ***
x1           -0.019602   0.128149  -0.1530  0.87843
x2            0.110798   0.046888   2.3630  0.01813 *
x3           -0.103681   0.065055  -1.5937  0.11099

> lmermod2 <- lmer(y ~ x1+x2+x3+(1|ID)+offset(log(time)),family=poisson,
  data=seiz,method="Laplace")
> summary(lmermod2)
Generalized linear mixed model fit using Laplace
```

270

```
Formula: y ~ x1 + x2 + x3 + (1 | ID) + offset(log(time))
Data: seiz
Family: poisson(log link)
      AIC      BIC   logLik deviance
970.2882 988.7231 -480.1441 960.2882
Random effects:
   Groups             Name         Variance   Std.Dev.
   ID (Intercept)      0.60832    0.77995
# of obs: 295, groups: ID, 59
Estimated scale (compare to 1) 1.671041
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.032640   0.152524  6.7703 1.285e-11 ***
x1           -0.023848   0.210494  -0.1133  0.90980
x2            0.110798   0.046895   2.3627  0.01814 *
x3           -0.103681   0.065065  -1.5935  0.11105
```

The Laplace approach gives significantly different (and more reliable estimates).

271

Random intercepts and slopes

We may also allow the treatment effect to vary between individuals.

```
> lmermod4 <- lmer(y ~ x1+x2+x3+(1+x2|ID)+offset(log(time)),
  family=poisson,data=seiz,method="Laplace")
> summary(lmermod4)
Generalized linear mixed model fit using Laplace
802.2693 828.0782 -394.1347 788.2693
Random effects:
Groups Name      Variance Std.Dev. Corr
ID      (Intercept) 0.49990  0.70704
      x2            0.23189  0.48155  0.166
# of obs: 295, groups: ID, 59
Estimated scale (compare to 1) 1.403177
Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.0712501  0.1398516  7.6599 1.861e-14 ***
x1            0.0494975  0.1927053  0.2569  0.79729
x2           -0.0023708  0.1078657 -0.0220  0.98246
x3           -0.3072281  0.1501527 -2.0461  0.04075 *
```

272

Bayesian Inference for GLMMs

A Bayesian approach to inference for a GLMM adds a prior distribution for β, α , to the likelihood $L(\beta, \alpha)$. Again a proper prior is required for the matrix \mathbf{D} . In general a proper prior is not required for β – the exponential family and linear link lead to a likelihood that is well-behaved. Closed-form inference is unavailable, but MCMC is almost as straightforward as in the linear mixed model case. The joint posterior is

$$p(\beta, \mathbf{W}, \mathbf{b} \mid \mathbf{y}) \propto \prod_{i=1}^m \{p(\mathbf{y}_i \mid \beta, \mathbf{b}_i)p(\mathbf{b}_i \mid \mathbf{W})\} \pi(\beta)\pi(\mathbf{W}).$$

Suppose we have priors:

$$\begin{aligned} \beta &\sim N_{q+1}(\beta_0, \mathbf{V}_0) \\ \mathbf{W} &\sim W_{q+1}(r, \mathbf{R}^{-1}) \end{aligned}$$

The conditional distributions for β, τ, \mathbf{W} are unchanged from the linear case. There is no closed form conditional distribution for β , or for \mathbf{b}_i , but Metropolis-Hastings step can be used (or adaptive rejection sampling can be utilized, the conditional is log concave).

273

Priors for β and α in the GLMM

Lognormal Priors

It is convenient to specify lognormal priors for positive parameters θ , since one may specify two quantiles of the distribution, and directly solve for the two parameters of the prior. In a GLMM we can often specify priors for more meaningful parameters than elements of β . For example, e^{β_1} is the relative risk/rate in a log linear model, and is the odds ratio in a logistic model.

Suppose we wish to specify a lognormal prior for a generic parameter θ .

Denote by $\text{LN}(\mu, \sigma)$ the lognormal distribution with $E[\log \theta] = \mu$ and $\text{var}(\log \theta) = \sigma^2$, and let θ_1 and θ_2 be the q_1 and q_2 quantiles of this prior.

Then

$$\mu = \log(\theta_1) \left(\frac{z_{q_2}}{z_{q_2} - z_{q_1}} \right) - \log(\theta_2) \left(\frac{z_{q_1}}{z_{q_2} - z_{q_1}} \right), \quad \sigma = \frac{\log(\theta_1) - \log(\theta_2)}{z_{q_1} - z_{q_2}}. \quad (44)$$

As an example, suppose that for θ we believe there is a 50% chance that the relative risk is less than 1 and a 95% chance that it is less than 5; with $q_1 = 0.5, \theta_1 = 1.0$ and $q_2 = 0.95, \theta_2 = 5.0$, we obtain lognormal parameters $\mu = 0$ and $\sigma = \log 5/1.96 = 0.98$.

274

Gamma Priors

Consider the random intercepts model with $b_i \sim_{iid} N(0, \sigma^2)$.

It is not straightforward to specify a prior for σ , which represents the standard deviation of the residuals on the linear predictor scale, and is consequently not easy to interpret.

We specify a gamma prior $\text{Ga}(a, b)$ for the precision $\tau = 1/\sigma^2$, with parameters a, b specified *a priori*. The choice of a gamma distribution is convenient since it produces a marginal distribution for the residuals in closed form.

Specifically the two-stage model

$$b_i | \sigma \sim_{iid} N(0, \sigma^2), \quad \tau = \sigma^{-2} \sim \text{Ga}(a, b)$$

produces a marginal distribution for b_i which is $t_d(0, \lambda^2)$, a Student's t distribution with $d = 2a$ degrees of freedom, location zero, and scale $\lambda^2 = b/a$.

We now consider a log link, in which case the above is equivalent to the residual relative risks following a log t distribution.

We specify the range $\exp(\pm R)$ within which the residual relative risks will lie with probability q , and use the relationship $\pm t_{q/2}^d \lambda = \pm R$, where $t_{q/2}^d$ is the q -th quantile of a Student t random variable with d degrees of freedom, to give $a = d/2, b = R^2 d/2 (t_{q/2}^d)^2$.

275

For example, if we assume *a priori* that the residual relative risks follow a log Student t distribution with 2 degrees of freedom, and that 95% of these risks fall in the interval (0.5,2.0) then we obtain the prior, $\text{Ga}(1, 0.0260)$.

In terms of σ this results in (2.5%, 97.5%) quantiles of (0.084,1.01) with posterior median 0.19.

It is important to assess whether the prior allows all reasonable levels of variability in the residual relative risks, in particular small values should not be excluded.

The prior $\text{Ga}(0.001,0.001)$ which has previously been used (e.g. in the WinBUGS manual) should be avoided for this very reason (this corresponds to relative risks which follow a log Student t distribution with 0.002 degrees of freedom).

276

Bayesian Inference for the Seizure Data

We fit various models and begin with a discussion of prior specification.

We fit four models to the seizure data.

Model 1 Random intercepts only, $\pi(\boldsymbol{\beta}) \propto 1$, $\tau \sim \text{Ga}(1, 0.260)$ – corresponds to Student t_2 residuals and 95% $\in (0.5, 2.0)$.

Model 2 Random intercepts only, $\pi(\boldsymbol{\beta}) \propto 1$, $\tau \sim \text{Ga}(2, 1.376)$ – corresponds to Student t_4 residuals and 95% $\in (0.1, 10.0)$.

Model 3 Random effects for intercept and for x_2 .

Model 4 We allow a bivariate Student t distribution for the pair of random effects introduced in Model 3.

Model 5 We introduce “measurement error” into the model.

277

	Estimates (standard deviations)				
	Model 1	Model 2	Model 3	Model 4	Model 5
β_0	1.03 (0.15)	1.03 (0.15)	1.08 (0.13)	0.92 (0.15)	1.00 (0.18)
β_1	-0.024 (0.21)	-0.034 (0.21)	0.042 (0.19)	0.16 (0.20)	0.091 (0.24)
β_2	0.11 (0.047)	0.11 (0.047)	0.0045 (0.11)	-0.030 (0.11)	0.012 (0.10)
β_3	-0.11 (0.065)	-0.10 (0.065)	-0.31 (0.15)	-0.32 (0.15)	-0.30 (0.14)
σ_0	0.64 (0.13)	0.66 (0.13)	0.71 (0.072)	0.71 (0.10)	0.82 (0.084)
σ_1	–	–	0.473 (0.062)	0.399 (0.078)	
ρ	–	–	0.19 (0.16)	0.21 (0.21)	
σ_e	–	–	–	–	0.39 (0.033)

Table 10: Posterior means and standard deviations for Bayesian analysis of seizure data; σ_0 is the standard deviation of the random intercepts, σ_1 is the standard deviation of the random period effect, and ρ is the correlation between these random effects; σ_e is the standard deviation of the measurement error.

280

Poisson Model with a “nugget” effect

Recall the model

$$\begin{aligned} Y_{ij}|b_i &\sim \text{Poisson}(t_{ij} \exp(\mathbf{x}_{ij}\beta + b_i)) \\ b_i &\sim \text{N}(0, \sigma_0^2) \end{aligned}$$

has a single parameter only, σ_0 to allow for excess-Poisson variability *and* between-individual variability.

In the LMEM model we have

$$\begin{aligned} \text{E}[Y_{ij}|b_i] &= \mathbf{x}_{ij}\beta + b_i + \epsilon_{ij} \\ b_i &\sim \text{N}(0, \sigma_0^2) \\ \epsilon_{ij} &\sim \text{N}(0, \sigma_e^2) \end{aligned}$$

with b_i and ϵ_{ij} independent.

281

By analogy we might consider the model:

$$\begin{aligned} Y_{ij}|b_i, b_{ij} &\sim \text{Poisson}(t_{ij} \exp(\mathbf{x}_{ij}\boldsymbol{\beta} + b_i + b_{ij})) \\ b_i &\sim \text{N}(0, \sigma_0^2) \\ b_{ij} &\sim \text{N}(0, \sigma_e^2) \end{aligned}$$

with b_i and b_{ij} independent.

We now two parameters to allow for between-individual variability, σ_0 , and excess-Poisson variability, σ_e .

Unfortunately there is no simple marginal interpretation of σ_0 and σ_e :

$$\begin{aligned} \text{E}[Y_{ij}] &= t_{ij} \exp(\mathbf{x}_{ij}\boldsymbol{\eta} + \sigma_e^2/2 + \sigma_0^2) = \mu_{ij} \\ \text{var}(Y_{ij}) &= \mu_{ij} + \mu_{ij}^2(e^{\sigma_e^2} - 1)(e^{\sigma_0^2} - 1) \\ \text{cov}(Y_{ij}) &= t_{ij}t_{ik} \exp(\mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{x}_{ik}\boldsymbol{\beta})e^{\sigma_0^2}(e^{\sigma_e^2} - 1) \end{aligned}$$

Another possibility would be to start with a negative binomial distribution, and then introduce a random effect, b_i . This reveals the “heaven and hell” of mixed-effects models — we have a lot of flexibility in the models we can fit, but many formulations that are similar produce different marginal mean and covariance structures, and often there is no obvious “right” choice.

282

WinBUGS for Model 1

```
# Model 3 - Poisson lognormal for nugget also
model
{
for (i in 1:n){
  for (j in 1:k){
    Y[i,j] ~ dpois(mu[i,j])
    log(mu[i,j]) <- log(t[j])+beta0+beta1*x1[i]+beta2*x2[j]+
                    beta3*x1[i]*x2[j]+b[i]+be[i,j]
    be[i,j] ~ dnorm(0,taue)
  }
  b[i] ~ dnorm(0,tau)
}
taue ~ dgamma(1,0.26)
tau ~ dgamma(1,0.26)
sigma <- sqrt(1/tau)
sigmae <- 1/sqrt(taue)
beta0 ~ dflat()
beta1 ~ dflat()
beta2 ~ dflat()
beta3 ~ dflat()
}
```

283

Seizure Data

Patient 49 had counts 151,102,65,72,63 under progabide — very surprising.

In DHLZ dropping this individual gave a parameter of interest of -0.30.

Posterior medians of b_{ij} for this individual ($i = 49, j = 0, 1, 2, 3, 4$) are:

-0.61, 0.61, 0.18, 0.27, 0.65, 0.15

Conclusions: there is evidence of a statistically significant treatment effect, under Model 4 the 95% credible interval on β_3 is (-0.60,-0.28).

Under model 5 the 95% credible interval on β_3 is (-0.59,-0.030).

284

Generalized Estimating Equations (GEEs)

Liang and Zeger (1986, Biometrika), and Zeger and Liang (1986, Biometrics) considered GLMs with dependence within individuals (in the context of longitudinal data).

Theorem (Liang and Zeger, 1986): the estimator $\hat{\beta}$ that satisfies

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

where $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \beta}$, $\mathbf{W}_i = \mathbf{W}_i(\beta, \alpha)$ is the working covariance model, $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\beta)$ and $\hat{\alpha}$ is a consistent estimator of α , is such that

$$\mathbf{V}_\beta^{-1/2} (\hat{\beta} - \beta) \rightarrow_d \mathbf{N}(\mathbf{0}, \mathbf{I}),$$

where \mathbf{V}_β is given by

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

In practice an empirical estimator of $\text{cov}(\mathbf{Y}_i)$ is substituted to give $\widehat{\mathbf{W}}_\beta$.

285

Choice of Working Covariance Models

As in the linear case, various assumptions about the form of the working covariance may be assumed (what is a natural choice?); we write

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

where $\mathbf{\Delta}_i = \text{diag}[\text{var}(Y_{i1}), \dots, \text{var}(Y_{in_i})]^T$ and \mathbf{R}_i is a working correlation model, for example, independence, exchangeable, AR(1), unstructured.

- For small m the sandwich estimator will have high variability and so model-based variance estimators may be preferable (but would we trust asymptotic normality if m were small anyway?).
- Model-based estimators are more efficient if the model is correct.

Published comments:

- Liang and Zeger (1986): “little difference when correlation is moderate”.
- McDonald (1993): “The independence estimator may be recommended for practical purposes”.
- Zhao, Prentice and Self (1992): Assuming independence “can lead to important losses of efficiency”.
- Fitzmaurice, Laird and Rotnitsky (1993): “important to obtain a close approximation to $\text{cov}(\mathbf{Y}_i)$ in order to achieve high efficiency”.

286

GEE for the Seizure Data

We have the log-linear model is given

$$\log E[Y_{ij}] = \log \mu_{ij} = \log t_{ij} + \beta_0^* + \beta_1 x_{i1} + \beta_2 x_{ij2} + \beta_3 x_{i1} x_{ij2}$$

and $\text{var}(Y_{ij}) = \alpha \mu_{ij}$. Recall β_1 is baseline comparison of rates, β_2 is period effect and β_3 is treatment effect.

Both quasi-likelihood and working independence GEE have estimating equation

$$\mathbf{G}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}) = \sum_{i=1}^m \mathbf{x}_i^T (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0},$$

but differ in the manner in which the standard errors are calculated.

	Estimates (standard errors)				
	Poisson	Quasi-Lhd	GEE Ind	GEE Exch	GEE AR(1)
β_0^*	1.35 (0.034)	1.35 (0.15)	1.35 (0.16)	1.35 (0.16)	1.31 (0.16)
β_1	0.027 (0.047)	0.027 (0.21)	0.027 (0.22)	0.027 (0.22)	0.015 (0.21)
β_2	0.11 (0.047)	0.11 (0.21)	0.11 (0.12)	0.11 (0.12)	0.16 (0.11)
β_3	-0.10 (0.065)	-0.10 (0.29)	-0.10 (0.22)	-0.10 (0.22)	-0.13 (0.27)
α_1, α_2	1.0, 0	19.7, 0	19.4, 0	19.4, 0.78	20.0, 0.89

Table 11: Parameter estimates and standard errors under various models; α_1 is a variance parameter, and α_2 a correlation parameter.

The point estimates under Poisson, quasi-likelihood and GEE working independence will always agree. The Poisson standard errors are clearly much too small. The quasi-likelihood standard errors are increased by $\sqrt{19.7} = 4.4$, but do not acknowledge dependence on observations on the same individual (it is as if we have 59×5 independent observations). The standard errors of estimated parameters that are associated with time-varying covariates (β_2 and β_3) are reduced under GEE, since within-person comparisons are being made. The coincidence of the estimates and standard errors for independence and exchangeability is a quirk of the balanced design.

288

Interpretation of Marginal and Conditional Coefficients

In a marginal model (which we consider under GEE), we have

$$E[Y | x] = \exp(\gamma_0 + \gamma_1 x)$$

in which case e^{γ_1} is the change in the average response when we increase x by 1 unit in the population under consideration.

Under the conditional (mixed effects) model the interpretation of regression coefficients is conditional on the value of the random effect.

For the model

$$E[Y | x, b] = \exp(\beta_0 + \beta_1 x + b),$$

with $b \sim_{iid} N(0, \sigma^2)$, the marginal mean is given by:

$$E[Y | x] = E_{b|\sigma^2} \{E[Y | x, b]\} = \exp(\beta_0 + \sigma^2/2 + \beta_1 x).$$

Hence for the log-linear model, e^{β_1} has the same marginal interpretation to e^{γ_1} (the marginal intercept is $\gamma_0 = \beta_0 + \sigma^2/2$), though estimation of the latter via GEE produces a consistent estimator in more general circumstances (though there is an efficiency loss if the random effects model is correct).

289

In the model

$$E[Y | x, \mathbf{b}] = \exp\{\beta_0 + b_{0i} + (\beta_1 + b_{1i})x_i\}$$

e^{β_1} is the relative risk between two populations with the same \mathbf{b} but whose x values differ by one unit, that is:

$$\exp(\beta_1) = \frac{E[Y | x, \mathbf{b}]}{E[Y | x - 1, \mathbf{b}]}$$

An alternative interpretation is to say that it is the expected change between two “typical individuals”, that is, individuals with random effects, $\mathbf{b} = \mathbf{0}$.

With $\mathbf{b} \sim_{iid} N(\mathbf{0}, \mathbf{D})$ we have the marginal mean

$$E[Y | x] = \exp\{\beta_0 + D_{00}/2 + x(\beta_1 + D_{01}) + x^2 D_{11}/2\}$$

so that there is no marginal mean interpretation of $\exp(\beta_1)$ (the latter is the marginal median).

290

Stochastic Covariates

In some longitudinal situations, the response at time t on individual i may depend on not just the current covariates, but also previous values.

For example, in an investigation into the health effects of recent air pollution we may believe that the response depends on not just today’s exposure, but also the preceding days.

In such situations, obtaining the correct form of the model will in general be difficult, and instead we might decide to estimate the association for a simpler model.

As an example, suppose that we have a single covariate, and we decide to examine the *cross-sectional* association:

$$\mu_{ij} = E[Y_{ij} | X_{ij}]. \quad (45)$$

In such a situation great care must be taken to obtain a consistent estimator.

We demonstrate with a GEE approach, though the pitfalls of estimation apply equally to likelihood and Bayesian approaches.

291

With a GLM:

$$\eta_{ij} = g(\mu_{ij}) = \mathbf{x}_{ij}\boldsymbol{\beta},$$

and assume for simplicity $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$. The generalized estimating function is given by

$$\mathbf{G}(\boldsymbol{\beta}) = \sum_{i=1}^m \mathbf{D}_i^\top \mathbf{W}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i)$$

which has second row

$$G_2(\boldsymbol{\beta}) = \sum_{i=1}^m \left[\sum_{j=1}^{n_i} \sum_{k=1}^{n_i} X_{ij} W_{ijk}^* (Y_{ik} - \mu_{ik}) \right]$$

where (45) is the assumed model, i.e. $\mu_{ik} = \text{E}[Y_{ij} | X_{ik}]$, and $W_{ijk}^* = \frac{\partial \mu_{ij}}{\partial \eta_{ij}} W_i^{jk}$ with W_i^{jk} the (j, k) -th element of \mathbf{W}_i^{-1} .

292

To obtain consistency we require

$$\text{E}[\mathbf{G}(\boldsymbol{\beta})] = \mathbf{0}.$$

Previously we have seen that if the mean specification is correct then we obtain consistency of $\hat{\boldsymbol{\beta}}$.

Since now the estimating function depends on the random variables

$\mathbf{X} = (X_1, \dots, X_m)^\top$ the expectation is with respect to \mathbf{X} and \mathbf{Y} . Specifically

$$\text{E}_{Y, X}[G_2(\boldsymbol{\beta})] = \sum_{i=1}^m \text{E}_{Y_i, X_i} \left[\sum_{j=1}^{n_i} \sum_{k=1}^{n_i} X_{ij} W_{ijk}^* (Y_{ik} - \mu_{ik}) \right]$$

and

$$\begin{aligned} \text{E}_{Y_i, X_i} [X_{ij} W_{ijk}^* (Y_{ik} - \mu_{ik})] &= \text{E}_{X_i} \left\{ \text{E}_{Y_i | X_i} [X_{ij} W_{ijk}^* (Y_{ik} - \mu_{ik})] \right\} \\ &= \text{E}_{X_i} \left\{ X_{ij} W_{ijk}^* (\text{E}[Y_{ik} | X_{i1}, \dots, X_{in_i}] - \mu_{ik}) \right\} \end{aligned}$$

Hence to ensure an unbiased estimating function, and hence consistency of our estimator, we require

$$\text{E}[Y_{ik} | X_{i1}, \dots, X_{in_i}] = \mu_{ik} = \text{E}[Y_{ik} | X_{ik}],$$

otherwise we have bias.

293

Suppose we assume working independence, the above simplifies to

$$G_2(\boldsymbol{\beta}) = \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij} W_{ijj}^* (Y_{ij} - \mu_{ij}),$$

so that we only need to consider the expectation with respect to $\mathbf{Y}_i, \mathbf{X}_i$:

$$\mathbb{E}_{Y_{ij}|X_{ij}}[\mathbf{G}(\boldsymbol{\beta})] = \sum_{i=1}^m \sum_{j=1}^{n_i} \mathbb{E}_{X_i} [X_{ij} W_{ijj}^* (\mathbb{E}[Y_{ij} | X_{ij}] - \mu_{ij})] = \mathbf{0},$$

and we obtain a consistent estimator.

For more details see DHLZ, Section 12.3.1.

Cross-Sectional Versus Longitudinal Studies

Consider modeling $Y = \text{FEV}_1$ as a function of age. We might envisage that Y changes both over time within an individual, and that individuals may have different baseline levels of Y from which they begin, due to “cohort” effects. A birth cohort is a group of individuals who were born in the same year.

Cohort effects include the effects of environmental pollutants, and differences in lifestyle choices and medical treatment.

In a cross-sectional study a group of individuals are measured at a single time point. A great advantage of longitudinal studies, over cross-sectional studies is that both cohort and aging effects can be estimated.

To illustrate, consider the model:

$$E[Y_{ij} | \text{Age}_{ij}] = \beta_0 + \beta_c \text{Age}_{i1} + \beta_l (\text{Age}_{ij} - \text{Age}_{i1})$$

where Y_{ij} is the j -th FEV₁ measurement on individual i and Age_{ij} is the age of the individual when that measurement was taken, with Age_{i1} begin the age on a certain day (so that all the individuals are comparable).

Parameter interpretation:

We have

$$E[Y_{i1} | \text{Age}_{i1}] = \beta_0 + \beta_c \text{Age}_{i1},$$

so that β_c is the average change in Y between two populations who differ by one unit in their baseline ages; said another way we are examining the differences in Y between two birth cohorts a year apart.

Also

$$E[Y_{ij} | \text{Age}_{ij}] - E[Y_{i1} | \text{Age}_{i1}] = \beta_l (\text{Age}_{ij} - \text{Age}_{i1})$$

so that β_l is the longitudinal effect, that is the change in the average response between two populations who are in the same birth cohort, and whose ages differ by one year. In a cohort study with $E[Y_{ij} | x_{ij}] = \beta_0 + \beta_1 x_{ij}$ we have

$$E[\hat{\beta}_1] = \beta_l + \frac{\sum_{i=1}^m n_i (x_{i1} - \bar{x}_{i1})(\bar{x}_i - \bar{x})}{\sum_{i=1}^m \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2} (\beta_c - \beta_l).$$