Topics for today

- *Generalized linear mixed models (GzLMM)*

Related reading: Sections 5 in Non-normal notes.

5 Generalized linear mixed models (GzLMM)

- GzLMMs combine generalized linear model and linear mixed model theory. There are greater complexities in fitting GzLMMs, due to the nonlinearity involved with the model. Fitting of the models generally involves approximations of some sort. This section outlines some of the basic approaches to fitting a GzLMM.

- When extending GzLM theory to longitudinal data, we consider the mean link function in terms of both subject (*i*) and time (*j*): $g\left(\mu_{ij}\right)=\mathbf{X}_{ij}^{r}\boldsymbol{\beta}$, where $\mathbf{X}_{ij}^{r}$ denotes the j^{th} row of $\mathbf{X}_i$ (if considering the subject-specific model) or the $(ij)^{\text{th}}$ row of $\mathbf{x}$ (if considering the full data model).

- We could extend the model to other types of clustered data but for now we'll just focus on longitudinal data.

- Adding the 'mixed' component, $\mathbf{Z}_{ij}^{r}\mathbf{b}_i$, to the mean link function for a longitudinal GzLM yields a GzLMM:

$$g\left(\mu_{ij}\right) = \mathbf{X}_{ij}^{r}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{r}\mathbf{b}_i$$

  where g is a link as previously discussed (e.g., log link for counts, logit link for binary outcomes), $\mu_{ij} = E\left(Y_{ij} \mid \mathbf{b}_i, \mathbf{x}_{ij}\right)$, $\mathbf{b}_i \sim N\left(\mathbf{0}, \mathbf{G}_i\right)$, and $\mathbf{Z}_{ij}^{r}$ is the jth row of $\mathbf{Z}_i$, the covariate matrix for subject i, associated with random effects $\mathbf{b}_i$.

- The left side of a GzLMM looks like a GzLM and the right side looks much like an LMM. The mean is often simplified to $\mu_{ij} = E\left(Y_{ij} \mid \mathbf{b}_i\right)$ in the literature (or my notes), where conditioning on $\mathbf{x}_{ij}$ is implied.

- GzLMMs are not fit using GEEs. One approach to fit the model is to employ nonlinear mixed modeling techniques (e.g., PROC NLMIXED in SAS), or to use an approach that involves iterative fits of a linear mixed model to approximate the true model (e.g., PROC GLIMMIX in SAS).

- Interpretation of effects associated with GzLMMs are different than those based on GzLM/GEEs, which will be discussed more later.

- The linear predictor is similar but generalized from the GzLM case in the same way that random effects are added to a general linear model to get a linear mixed model: $\eta_{ij} = \mathbf{X}_{ij}^{r}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{r}\mathbf{b}_i$.

- We can express the model in 'complete data' form as

$$g\left(\boldsymbol{\mu}\right) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

  where $\boldsymbol{\mu} = E\left(\mathbf{Y} \mid \mathbf{b}, \mathbf{x}\right)$ (an $r_{tot}\times 1$ vector) and quantities on the right-hand side of the equation are defined as in the early part of the LMM notes. An expression of the model above that will be useful for estimation discussed later is

$$\boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}).$$

- Let $h(\mathbf{b}_i)$ and $f(\mathbf{y}_i)$ denote the pdf's of the random effects and responses for subject i, respectively. Also, let $l(\mathbf{y}_i \mid \mathbf{b}_i)$ denote the conditional pdf of of the responses given the random effects that is a member of the exponential family (e.g., Poisson, binomial, geometric, gamma). Then, we can express the density of the responses as

$$f(\mathbf{y}_i) = \int l(\mathbf{y}_i \mid \mathbf{b}_i)\mathrm{h}(\mathbf{b}_i)d\mathbf{b}_i \qquad \text{for subjects } i=1,\ldots,n.$$

  This will be useful in setting up a likelihood equation for estimation of parameters.

5.1 Fitting the GzLMM by approximating the likelihood function

- When subjects are assumed to be independent (the standard case), then the likelihood function is $L = \prod_{i=1}^{n} f(\mathbf{y}_i)$.

- For normal outcomes, the likelihood could be expressed in closed form because the integral in the likelihood function involves only normal distributions, but numerical techniques were required to optimize the function.

- For non-normal outcomes, the function cannot even be written in closed form.

- However, we can approximate the log-likelihood function using a technique such as quadrature, which essentially approximates integrals of quantities in the likelihood that are difficult to evaluate with sums of rectangle areas (i.e., like a histogram approximation). The approximated likelihood can then be maximized using numerical techniques to determine (approximate) maximum likelihood parameter estimates. A Laplace method can also be used to approximate the likelihood instead of adaptive quadrature. See the SAS Help Documentation under PROC GLIMMIX for more detail.

5.2 Fitting the GzLMM using linearization methods

- An alternative to the approach above is to create pseudo data (**P**) using the framework of the GzLMM and original responses (**Y**) that can be modeled with a standard LMM.

- The transformed 'pseudo-data' **P** is approximately normally distributed and can be fit with the linear mixed model $\mathbf{P} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$ using pseudo-ML or pseudo-REML estimation to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$. [The likelihood (restricted likelihood) for the pseudo data is referred to as the pseudo likelihood (pseudo restricted likelihood).]

- Next, the pseudo data is recomputed using the formula above using the new parameter and random effects estimates, and the process is repeated iteratively until estimates converge.

- This is a doubly-iterated procedure, since we update estimates after each linear mixed model fitting, and within each linear mixed model fitting we use an iterative procedure as well. The benefit of this approach is that we can take advantage of what standard LMM theory has to offer.

- For example, we can fit a model that has both random effects and higher level structures for **R**, or that has multiple-level or complex (e.g. crossed) random effects. [For GEE, we could specify a working covariance structure for **R** but not include random effects; for GzLMM methods that use techniques to approximate the likelihood, we can specify random effects but cannot have non-simple **R** matrices, or random effects at multiple levels.]

- One drawback to the linearization method is estimator bias that has been reported (see the SAS Help Documentation: *Notes on Bias of Estimators* page). However, for larger samples, the bias should diminish.

- SAS needs initial values of **P** to start the iterative process. If no specification is made, the GLIMMIX output indicates what was used (e.g., 'Starting from: GLM estimates' or '…data'). [I believe the 'GLM' they are referring to are what we call GzLM.] The method used depends on what types of covariance parameters are specified in the model (**R**-side or **G**-side).

- For more information on the linearization method, see the SAS Help Documentation under PROC GLIMMIX, or see the following journal article: Wolfinger, R.D. and O'Connell, M. (1993) Generalized linear mixed models: a pseudo-likelihood approach. *J Stat Comp and Sim* 48, 233-243.

5.3 *Software to fit GzLMMs*

- There are two procedures available to estimate a GzLMM using methods of approximating the likelihood.

  - PROC NLMIXED, as discussed and demonstrated in the 'Non-normal' lecture notes. This procedure uses adaptive Gaussian quadrature to approximate the true likelihood, and then optimization is carried out using a dual Quasi-Newton method.

  - PROC GLIMMIX also has the ability to approximate the true likelihood function, using either adaptive quadrature or a Laplace approximation. When method=quad is specified, as below, a Gauss-Hermite Quadrature method is used to approximate the likelihood, and the dual Quasi-Newton method is again used for optimization.

- To demonstrate the different procedures, exacerbation data from the Kunsberg kids / air pollution study was used. In this case, the 2003-04 study year was used; otherwise data are similar to that presented in the 'Non-normal' notes. The code and abbreviated output follow.

```
proc nlmixed data=y5dat_red;
 parms b0=0.5          b_poll=0.05
       b_day=0.005     b_wkend=-0.9
       b_holiday=-0.8  b_friday=0.3
       s2u=2;
 eta = b0 + b_poll*pm25cen02
   + b_day*day + b_wkend*weekend
   + b_holiday*holiday + u;
 Expeta = exp(eta);
 p=expeta/(1+expeta);
 model exacerb~binary(p);
 random u~normal(0,s2u) subject=id;run;

The NLMIXED Procedure

Specifications

Data Set                 WORK.Y5DAT_RED
Dependent Variable       exacerb
Dist. for Dependent Var. Binary
Random Effects           u
Dist. for Random Effects Normal
Subject Variable         id
Optimization Technique   Dual Quasi-Newton
Integration Method       Adaptive Gaussian
                            Quadrature
```

```
proc glimmix data=test method=quad noreml;
 model exacerb(event='1')
   = pm25cen02 day weekend holiday
  / solution distribution=binary;
 random intercept / subject=id; run;

The GLIMMIX Procedure

Model Information

Data Set                     WORK.Y5DAT_RED
Response Variable            exacerb
Response Distribution        Binary
Link Function                Logit
Variance Function            Default
Variance Matrix Blocked By   id
Estimation Technique         Maximum Likelihood
Likelihood Approximation     Gauss-Hermite
                                Quadrature

Degrees of Freedom Method    Containment
Optimization Information
```

```
Dimensions

Observations Used                  6923
Observations Not Used              1634
Total Observations                 8557
Subjects                             43
Max Obs Per Subject                 162
Parameters                            6
Quadrature Points                     1

NOTE: GCONV convergence criterion satisfied.
```

```
Optimization Technique        Dual Quasi-Newton
Parameters in Optimization    6
Lower Boundaries              1
Upper Boundaries              0
Fixed Effects                 Not Profiled
Starting From                 GLM estimates
Quadrature Points             7
Number of Observations Used   6923

Dimensions

G-side Cov. Parameters           1
Columns in X                     5
Columns in Z per Subject         1
Subjects (Blocks in V)          43
Max Obs per Subject            162

The GLIMMIX procedure is modeling the probability
that exacerb='1'.

Response Profile

Ordered                      Total
Value    exacerb         Frequency
1        0                    6449
2        1                     474

Convergence criterion (GCONV=1E-8) satisfied.
```



```
Fit Statistics

-2 Log Likelihood               2142.2
AIC (smaller is better)         2154.2
AICC (smaller is better)        2154.2
BIC (smaller is better)         2164.7

Parameter Estimates

                   Std
Parameter Estimate Error   DF   t Value  Pr>|t|
b0      -5.6106   0.5809   42    -9.66   <.0001
b_poll  -0.00613  0.01037  42    -0.59   0.5574
b_day    0.009643 0.001357 42     7.11   <.0001
b_wkend -0.2338   0.1392   42    -1.68   0.1004
b_holiday -0.1299 0.1946   42    -0.67   0.5082
s2u       7.5648 2.5298   42     2.99   0.0046
```

```
Fit Statistics

-2 Log Likelihood           2141.39
AIC  (smaller is better)    2153.39
AICC (smaller is better)    2153.41
BIC  (smaller is better)    2163.96
CAIC (smaller is better)    2169.96
HQIC (smaller is better)    2157.29

Fit Statistics for Conditional Distribution

-2 log L(exacerb | r. effects)     1995.76
Pearson Chi-Square                 4398.25
Pearson Chi-Square / DF               0.64

Covariance Parameter Estimates
                                      Standard
Cov Parm     Subject    Estimate       Error
Intercept    id           7.5773      2.4621

Solutions for Fixed Effects

                         Standard
Effect      Estimate      Error      DF  t Value Pr>|t|
Intercept   -5.5356     0.5556       42  -9.96  <.0001
pm25cen02   -0.00614    0.01038    6876  -0.59  0.5542
day          0.009645   0.001357   6876   7.11  <.0001
weekend     -0.2339     0.1392     6876  -1.68  0.0929
holiday     -0.1299     0.1947     6876  -0.67  0.5045
```

- The slight differences in estimates might be attributable to different default approximation methods used in the respective procedures. However notice also that the specified DF does not match for the slopes estimates of the predictors.

- For comparison, let's examine PROC GLIMMIX using the linearization method.

- On the left is the same model being fit, just using the linearization method; on the right is the addition of a statement that will model repeated measures within subjects over time.

- Note that instead of using the REPEATED statement here (which doesn't exist in PROC GLIMMIX), we add another RANDOM statement with the key word _residual_. You can really think of this as a REPEATED statement, since it specifies the R matrix. The output on the left shows pretty decent similarity to the previous results based on quadrature. The model that uses the AR(1) structure (lower right) does have a substantially lower -2 log likelihood, but there are currently no commonly accepted goodness-of-fit statistics to compare models (even nested ones).


```
proc glimmix data=y5dat_red method=mspl;
 model exacerb(event='1')
   = pm25cen02 day weekend holiday
      / solution distribution=binary;
 random intercept / subject=id; run;


The GLIMMIX Procedure

Model Information

Data Set                      WORK.Y5DAT_RED
Response Variable             exacerb
Response Distribution         Binary
Link Function                 Logit
Variance Function             Default
Variance Matrix Blocked By    id
Estimation Technique          PL
Degrees of Freedom Method     Containment
Number of Observations Used        6923


Dimensions

G-side Cov. Parameters           1
Columns in X                     5
Columns in Z per Subject         1
Subjects (Blocks in V)          43
Max Obs per Subject            162
```

```
proc glimmix data=y5dat_red method=mspl;
 model exacerb(event='1')
   = pm25cen02 day weekend holiday
      / solution distribution=binary;
 random intercept / subject=id;
 random _residual_ / subject=id
 type=ar(1); run;

The GLIMMIX Procedure

Model Information

Data Set                      WORK.Y5DAT_RED
Response Variable             exacerb
Response Distribution         Binary
Link Function                 Logit
Variance Function             Default
Variance Matrix Blocked By    id
Estimation Technique          PL
Degrees of Freedom Method     Containment
Number of Observations Used        6923


Dimensions

G-side Cov. Parameters           1
R-side Cov. Parameters           2
Columns in X                     5
Columns in Z per Subject         1
Subjects (Blocks in V)          43
Max Obs per Subject            162
```

```
Optimization Information

Optimization Technique  Newton-Raphson
                        with Ridging
Parameters in Optimization    1
Lower Boundaries              1
Upper Boundaries              0
Fixed Effects                 Profiled
Starting From                 Data

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics

-2 Log Pseudo-Likelihood    47864.26
Generalized Chi-Square       4341.53
Gener. Chi-Square / DF          0.63

Covariance Parameter Estimates

                                    Standard
Cov Parm     Subject    Estimate       Error
Intercept    id           4.7531      1.2455


Solutions for Fixed Effects

                  Standard
Effect   Estimate    Error    DF   t Value Pr>|t|
Intercept -4.9398   0.4275     42  -11.55  <.0001
pm25cen02 -0.00606  0.01032  6876   -0.59  0.5576
day       0.009547  0.001349 6876    7.08  <.0001
weekend    -0.2313  0.1384   6876   -1.67  0.0948
holiday    -0.1285  0.1937   6876   -0.66  0.5072
```

```
Optimization Information

Optimization Technique  Newton-Raphson with Ridging
Parameters in Optimization    2
Lower Boundaries              2
Upper Boundaries              1
Fixed Effects                 Profiled
Residual Variance             Profiled
Starting From                 Data

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics

-2 Log Pseudo-Likelihood    38935.51
Generalized Chi-Square       4302.00
Gener. Chi-Square / DF          0.62

Covariance Parameter Estimates

                                    Standard
Cov Parm     Subject    Estimate       Error
Intercept    id           3.0102      0.8374
AR(1)        id           0.6911    0.008856
Residual                  0.6214     0.01808

Solutions for Fixed Effects

                       Standard
Effect     Estimate    Error    DF t Value Pr>|t|
Intercept   -4.3896   0.4103     42 -10.70 <.0001
pm25cen02   -0.00713  0.008261  6876 -0.86 0.3881
day          0.009147 0.002195  6876  4.17 <.0001
weekend     -0.2085   0.07911   6876 -2.64 0.0084
holiday     -0.00787  0.1267    6876 -0.06 0.9505
```


- There are several functions written for R software in that can be used to fit GzLMMs. The glmer function in the LME4 package will use a Laplace approximation of the likelihood; the glmmPQL function in the MASS package will do the linearization method and Pseudo-likelihood estimation.

- With the data above, I have obtained estimates with these functions using default specifications. For glmer, the estimates are not close to that of SAS, but there is a warning message that iteration limit was reached without convergence.

- For glmmPQL, estimates of fixed effects are closer to that of SAS for the model with only random intercept (not sure why the variances are much bigger); but estimates are not as close for the model that adds the AR(1) structure. From what I can tell, SAS estimates seem more reliable. Also, the run time seems a bit shorter with SAS. The R code and output for the model with random intercept follow (compare with SAS output in upper left).

```
library(MASS)
gm1 <- glmmPQL(fixed=exacerb ~ pm25cen02 + day + weekend + holiday,
random=~1 | id, family = binomial,data=dat)
> gm1
Linear mixed-effects model fit by maximum likelihood
  Data: dat
  Log-likelihood: NA
  Fixed: exacerb ~ pm25cen02 + day + weekend + holiday
 (Intercept)    pm25cen02          day      weekend      holiday
-4.651413433 -0.007761101  0.009174357 -0.226429397 -0.137666135

Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    1606.884  4562551

Variance function:
 Structure: fixed weights
 Formula: ~invwt
Number of Observations: 6923
Number of Groups: 43
```