SUBJECT-SPECIFIC GENERALIZED LINEAR MODELS FOR LONGITUDINAL DATA

Outline:

- Overview of Subject-specific Models
- Generalized Linear Mixed Effects Model (GLMM)
  - ML method and numerical techniques
  - Measure of dependence for logistic-normal model
- Subject-specific versus Population Average Models
- Estimation of Random Effects
- Generalized Linear Fixed Effects Model

(Review) Linear Mixed Effects Model

- We have discussed how to use random effects to model the dependence among subjects in the linear model settings.
- This idea of random effects can be extended to generalized linear models (GLM) for describing discrete and non-Gaussian continuous responses.
- In (subject-specific) linear mixed effects model for longitudinal data: focus was on **mean response** of Y_{ij} as a function of covariates $\boldsymbol{x}_{ij}$ and $\boldsymbol{d}_{ij}$ and conditional on a set of **random effects** $\boldsymbol{U}_i$:

$$\mathrm{E}(Y_{ij}|\boldsymbol{U}_i, X_i, D_i) = \mu_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{d}_{ij}'U_i$$

(D_i is almost always contained in X_i)

- **Example:**

$$\mathrm{E}(Y_{ij}|\boldsymbol{U}_i, X_i) = (\beta_0 + U_{i1}) + (\beta_1 + U_{i2})x_{ij}$$

  - $\beta_0 + U_{i1}$ is the **subject-specific** intercept (for the ith subject)
  - β_0 is the mean of subject-specific intercepts
  - β_1 is the mean of subject-specific slopes with respect to x_{ij}

- The β's have a subject-specific interpretation

- Now, lets do the same thing for generalized linear models

Subject-specific Generalized Linear Model

- Key components of a **generalized linear mixed (or fixed) effects model** (conditional model) for longitudinal response $\boldsymbol{Y}_i$:
  - **Linear predictor** conditional on **subject-specific effects** $\boldsymbol{U}_i$:

$$\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i$$

    same as for the linear mixed model

  - **Link function:** Conditional mean is connected to conditional linear predictor via link function $h(\cdot)$:

$$h(\mu_{ij}) = h(\mathrm{E}(Y_{ij}|\boldsymbol{U}_i, X_i, D_i)) = \eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i$$

- Conditional **distribution** of Y_{ij} given μ_{ij} (i.e., given $(\boldsymbol{U}_i, X_i)$)
  * The distribution for mixed models replaces the variance function we used in the GEE models
  * Given $\boldsymbol{U}_i$, the Y_{ij}'s are **assumed independent** of one another
    – Analogous to linear mixed effects model $Y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}U_i + Z_{ij}$, where Z_{ij} are independent of each other
    $\Rightarrow$ Conditional on U_i of subject i, Y_{ij}'s are independent of one another

- **Important note**:
  μ_{ij} and η_{ij} are **not** the same here as in the marginal models (GEE)!
  - Because here μ_{ij} and η_{ij} depend on $\boldsymbol{U}_i$, are subject i's own **personal** linear predictor and mean
  - Therefore, the coefficients $\boldsymbol{\beta}$ are not the same either
  - In the **GEE/marginal** model case, μ_{ij} and η_{ij} were **population average** quantities

- For a given subject i, all μ_{ij}'s and η_{ij}'s share random effect $\boldsymbol{U}_i$:
  - $\boldsymbol{U}_i$ accounts for the **natural heterogeneity** between subjects due to unmeasured factors ($\boldsymbol{U}_i$ is not observed)
  - $\boldsymbol{U}_i$ accounts for the observed correlation (association) among the repeated measures Y_{ij} comprising $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$
  - $\boldsymbol{U}_i$'s are independent across subjects

- In a **generalized linear fixed effects model**, the $\boldsymbol{U}_i$'s are treated as fixed quantities

  – $\boldsymbol{U}_i$'s can be used to control subject-level confounders (just as with linear fixed effects models)

- In a **generalized linear mixed effects model**, the $\boldsymbol{U}_i$'s are treated as random
  - $\boldsymbol{U}_i$'s are independent across subjects and a distribution is assumed for $\boldsymbol{U}_i$
  - generally, $\boldsymbol{U}_i \sim F$ (usually normal distribution) and $\boldsymbol{U}_i$ is **independent** of X_i
  - so $\boldsymbol{U}_i$ cannot control subject-level confounders
  - usually,

$$\boldsymbol{U}_i \sim \text{MVN}(0, G)$$

    where G depends on a set of variance-covariance parameters $\boldsymbol{\gamma}$ just as with linear mixed models

Example
Logistic Regression Model for Binary Data

- **Example:** Consider the data on 878 mothers in Georgia, each giving birth to five children. Define low birth weight (LBW) as being ≤ 2500 gm

- $Y_{ij} = I(\text{low birth weight}_{ij})$, a 0 or 1 variable, for child j of mother i

- Our model is

$$\mu_{ij} = \mathrm{E}(Y_{ij}|U_i, X_i) = \Pr(Y_{ij} = 1|U_i, X_i)$$

$$\mathrm{logit}(\mu_{ij}) = \eta_{ij} = (\beta_0 + U_i) + \beta_1 \mathtt{mage}_{ij}$$

where $\mathtt{mage}_{ij}$ is the maternal age for child j of mother i

- $b_{0i} = \beta_0 + U_i$ is subject (mother)-specific intercept on logit scale

- β_1 is the subject-specific **log odds ratio** relating maternal age to the probability of LBW infants
  - A mother's odds of a LBW infant are multiplied by e^{β_1} when this mother's maternal age increases by 1.
  - Even though β_1 has a subject-specific interpretation, it is **assumed** that the effect of maternal age is constant across subjects
  - ie, the slope w.r.t. maternal age on logit scale is same for all subjects
  - No marginal interpretation for β_1 (details later)
- A more general model:

$$\eta_{ij} = (\beta_0 + U_{i1}) + (\beta_1 + U_{i2})\texttt{mage}_{ij}$$

  - β_1 is **average subject-specific** log odds ratio for LBW and maternal age
  - $(\beta_1 + U_{i2})$ is subject i-specific log odds ratio relating LBW to maternal age

Random/Mixed Effects Model Estimation
Maximum Likelihood

- Recall Y_{ij}'s are **independent** of one another given $\boldsymbol{U}_i, X_i$
- Suppose the density of Y_{ij} given $\boldsymbol{U}_i, X_i$ is $f_y(y_{ij}|\boldsymbol{U}_i, X_i, \boldsymbol{\beta})$
- Then the **likelihood** (conditional on $\boldsymbol{U}_i$) for subject i is

$$L_i(\boldsymbol{\beta}, \boldsymbol{U}_i) = f_{\boldsymbol{y}}(\boldsymbol{y}_i|\boldsymbol{U}_i, X_i, \boldsymbol{\beta}) = \prod_{j=1}^{n_i} f_y(y_{ij}|\boldsymbol{U}_i, \boldsymbol{x}_{ij}, \boldsymbol{\beta})$$

- Under random effects model, we assume

$$\boldsymbol{U}_i \sim \text{MVN}(0, G)$$

and $\boldsymbol{U}_i$ is independent of covariates X_i

- Estimation is then accomplished via **maximum marginal likelihood**:
  - Likelihood $L_i(\boldsymbol{\beta}, \boldsymbol{U}_i)$ contains $\boldsymbol{U}_i$, which is unobserved
  - Integrate over $\boldsymbol{U}_i$ based on the density $f_{\boldsymbol{U}}(\boldsymbol{u}_i; G)$ of $\boldsymbol{U}_i$, we obtain the **marginal likelihood** for subject i

$$\begin{aligned} L_i^M(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= \int_{\boldsymbol{u}_i} L_i(\boldsymbol{\beta}, \boldsymbol{u}_i) f_{\boldsymbol{U}}(\boldsymbol{u}_i; \boldsymbol{\gamma})\, d\boldsymbol{u}_i \\ &= \int_{\boldsymbol{u}_i} \left\{ \prod_{j=1}^{n_i} f_y(y_{ij} | \boldsymbol{u}_i, \boldsymbol{x}_{ij}, \boldsymbol{\beta}) \right\} f_{\boldsymbol{U}}(\boldsymbol{u}_i; \boldsymbol{\gamma})\, d\boldsymbol{u}_i \end{aligned}$$

    where $\boldsymbol{\gamma}$ is the parameter for G (variance-covariance matrix of $\boldsymbol{U}_i$)
  - Then full marginal (over $\boldsymbol{U}_i$) likelihood function for entire data set is

$$L^M(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_i L_i^M(\boldsymbol{\beta}, \boldsymbol{\gamma})$$

    which is then maximized jointly with respect to $(\boldsymbol{\beta}, \boldsymbol{\gamma})$

- **Note:** ReML is generally not available for non-linear models

- In linear mixed models, the integral required to obtain L_i^M is available in **closed form**

- However, for most other generalized linear mixed models, the integral can only be obtained **numerically**.

Numerical Techniques in Estimation of GLMM

- Recall that the **marginal likelihood** for subject i is

$$L_i^M(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \int_{\boldsymbol{u}_i} L_i(\boldsymbol{\beta}, \boldsymbol{u}_i) f_{\boldsymbol{u}}(\boldsymbol{u}_i; G)\, d\boldsymbol{u}_i$$

- However, numerical integration is only practical when the number of random effects is small.
- Due to the difficulty of numerical integration, two basic approaches for making inference in GLMM have been proposed:
  - Approximate objective functions: Integral Approximation methods
  - Approximate model: Linearization methods

Integral Approximation: Gauss-Hermite Quadrature Methods

- Numerical integration replaces this incomputable integral with the approximation

$$L_i^M(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \int_{\boldsymbol{u}_i} L_i(\boldsymbol{\beta}, \boldsymbol{u}_i) f_{\boldsymbol{U}}(\boldsymbol{u}_i; G)\, d\boldsymbol{u}_i \approx \sum_{q=1}^{d} L_i(\boldsymbol{\beta}, \boldsymbol{u}_q^*) w(\boldsymbol{u}_q^*)$$

– sums over the set of **nodes** $\boldsymbol{u}_q^*$
– positive **weights** $w(\boldsymbol{u}_q^*)$ sum to 1
– $d =$ the number of quadrature points (nodes) in the numerical approximation

- Notes:
  - the more nodes one uses, the better the approximation (but the longer the computation time)
  - problem becomes **exponentially** more difficult with the **number** of random effects in the model
    - two is usually the most that is practical
- **Adaptive** Gauss-Hermite quadrature improves on GH quadrature by trying to choose nodes to make approximation more accurate
  - centers quadrature points with respect to the mode of the function being integrated and scales them according to estimated curvature at the mode.

– can drastically reduce number of quadrature points needed to approximate the integrals and improve the speed of integration

– Adaptive GHQ is state-of-the-art

- Laplace method is a special case with quadrature point $d=1$
- The approximated likelihood can then be maximized with standard algorithms (eg, Newton-Raphson) to estimate parameters of GLMM
- With integral approximation, the likelihood function can be approximated, and likelihood based inference procedures can be employed, eg

  – Estimate $\mathrm{var}(\hat{\boldsymbol{\beta}})$ by inverting the negative second derivative of the likelihood function

  – For complex models, second derivative may be computed numerically rather than analytically

  – LRT to compare nested models

Linearization Methods

- Linearization-based methods use a relatively simple form of the linearized model
- Recall matrix form of model

$$h(\boldsymbol{\mu}) = h(\mathrm{E}(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{X}, D)) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + D\boldsymbol{U}$$

  where $h(\cdot)$ is applied to each element of $\boldsymbol{\mu}$

- A first order Taylor expansion of μ about $\boldsymbol{\beta}^0$ and $\boldsymbol{U}^0$ yields

$$\begin{aligned} \boldsymbol{\mu} &= h^{-1}(\boldsymbol{\eta}) \\ &= h^{-1}(\boldsymbol{\eta}^0) + \left(\frac{\partial h^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)_{\boldsymbol{\beta}^0, \boldsymbol{U}^0} \times \boldsymbol{X}(\boldsymbol{\beta} - \boldsymbol{\beta}^0) \\ &\quad + \left(\frac{\partial h^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)_{\boldsymbol{\beta}^0, \boldsymbol{U}^0} \times D(\boldsymbol{U} - \boldsymbol{U}^0) \end{aligned}$$

- Define $\Delta^0 = \left(\frac{\partial h^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)_{\boldsymbol{\beta}^0, \boldsymbol{U}^0}$
  – Recall vector-by-vector derivative

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

  – Δ^0 is a diagonal matrix

- Rearranging terms we get

$$(\Delta^0)^{-1}\{\boldsymbol{\mu} - h^{-1}(\boldsymbol{\eta}^0)\} + \boldsymbol{X}\boldsymbol{\beta}^0 + D\boldsymbol{U}^0 = \boldsymbol{X}\boldsymbol{\beta} + D\boldsymbol{U}$$

- Define pseudo response

$$\widetilde{\boldsymbol{Y}} = (\Delta^0)^{-1}\{\boldsymbol{\mu} - h^{-1}(\boldsymbol{\eta}^0)\} + \boldsymbol{X}\boldsymbol{\beta}^0 + D\boldsymbol{U}^0$$

$\Rightarrow$ We can consider a linear mixed effects model

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{\beta} + D\boldsymbol{U} + \boldsymbol{Z}$$

where $\mathrm{var}(\boldsymbol{Z}) = \mathrm{var}(\widetilde{\boldsymbol{Y}}|\boldsymbol{X}, \boldsymbol{U}) = (\Delta^0)^{-1}\mathrm{var}(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{U})(\Delta^0)^{-1}$

- Double iteration scheme:
  - Estimate $\boldsymbol{\beta}$ and $\boldsymbol{U}$ based on

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{X}\boldsymbol{\beta} + D\boldsymbol{U} + \boldsymbol{Z}$$

    where $\mathrm{var}(\boldsymbol{Z}) = (\Delta^0)^{-1}\mathrm{var}(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{U})(\Delta^0)^{-1}$
  - Pseudo-response and error variance of the linearized model are recomputed
  - This process repeats until convergence.

- Advantages of Linearization method:
  - Typically can be fit based on only mean and variance in linearized form.
  - Best when
    - joint distribution is difficult to ascertain
    - errors are correlated
    - number of random effects is large
    - random effects are crossed

- Disadvantages of Linearization method:
  - Potentially biased estimates, especially for binary data.
  - The objective function to be optimized after each linearization update is dependent on the current pseudo-data.
  - True likelihood not known $\rightarrow$ no likelihood ratio tests
  - The convergence is easy to fail

- Integral Approximation method:

  Disadvantages:

  - Limited random effects structures.
  - Number of random effects should be relatively small.

  Advantages:

  - Adaptive methods can achieve any desired accuracy
  - Likelihood ratio testing accessible

- Software to fit GLMM:
  (Different software will give you different answers because different numerical techniques are used)
  - Stata:
    - Commands `xtlogit, re` and `xtpoisson, re normal` use adaptive GH quadrature by default, with GH quadrature as an option
    - Commands `xtmelogit` and `xtmepoisson` use adaptive GH
  - SAS `proc glimmix` options:
    - Pseudo-likelihood estimation based on linearization approximation: `method=RSPL` (default), and variation `method=RMPL`
    - Integral approximation using Laplace method: `method=Laplace`
    - Adaptive quadrature: `method=QUAD`

- SAS `proc nlmixed` is more general in model specification
  * Provide options for adaptive GH quadrature and other methods to approximate integral
  * Allow non-normal random effects
  * Need to write out the model, link function, etc
- R:
  * `glmer()` (package `lme4`) uses adaptive Gauss-Hermite quadrature, currently implemented only for a single scalar random effect
  * Other package: `nlme`, etc.

Example
Logistic Mixed Effects Model for Binary Data

- Consider the data on 878 mothers in Georgia, each giving birth to five children. Define LBW as being ≤ 2500 gm

- Consider the subject-specific model

$$\text{logit}(\mu_{ij}) = \text{logit}\left(\Pr(Y_{ij} = 1 | U_i, X_i)\right) = (\beta_0 + U_i) + \beta_1 \texttt{mage}$$

  - $Y_{ij} = I(\text{LBW}_{ij})$
  - U_i is normally distributed with mean 0 and variance ν^2
  - U_i is independent of X_i
  - Given U_i, Y_{ij} are independent

- SAS code to load, examine data, and divide the mother's age by 10 (so it represent age in decades).

```
data birthwt;
infile 'birthwt.raw';
input id birthorder birthwt momage momage_avg momage_dev;
run;

data birthwt;
set birthwt;
if(birthwt ne .)then lbw=(birthwt<=2500);
mage=momage/10;
run;

proc freq data=birthwt;
table lbw;
run;
```

lbw	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	3934	89.61	3934	89.61
1	456	10.39	4390	100.00

- Fit random effects logistic model using quadrature method in SAS:

```
Proc glimmix data=birthwt method=quad (qpoints=50);
 class id;
 model lbw =mage/dist=binomial link=logit s;
 random int/ subject=id;
 title1 'GLMM, quadrature method';
run;
```

- SAS Note:
  - `method=quad (qpoints=50)` option indicates adaptive quadrature method with 50 nodes to fit GLMM.
  - `MODEL` statement specifies fixed effects, distribution function for the response variable, and link function that links the conditional means with the linear predictors.
  - `s` or `SOLUTIONS` option requests fixed-effects parameter estimates

- Note for Stata: you can fit the same model by
  `xtlogit lbw mage, re`

- SAS results:

```
    -2 Log Likelihood             2789.18

       Covariance Parameter Estimates

                                             Standard
Cov Parm      Subject     Estimate          Error

Intercept     id            1.7288         0.2675


                  Solutions for Fixed Effects

                              Standard
Effect         Estimate          Error         DF    t Value    Pr > |t|

Intercept       -2.0396         0.3254        877      -6.27      <.0001
mage            -0.3326         0.1448       3511      -2.30      0.0217
```

- Interpretation:
  - $\widehat{\beta}_0 = -2.04$ is the **average** mother-specific intercept for the log odds of having a LBW baby as a function of maternal age
  - $\widehat{\nu}^2 = 1.73$ is the between-subject variance of subject-specific intercepts; it reflects the variation in the propensity of mothers for LBW babies
  - $\widehat{\beta}_1 = -0.33$ is the subject-specific log odds ratio for LBW for a one-unit (ie, 10 years) difference in maternal age

- we can do LRT for whether the random effect is significant by testing

$$H_0 : \nu^2 = 0 \quad \text{vs.} \quad H_A : \nu^2 > 0$$

- SAS code:

```
Proc glimmix data=birthwt method=quad (qpoints=50);
 class id;
 model lbw =mage/dist=binomial link=logit s;
 title1 'No random effect';
run;
```

```
                      -2 Log Likelihood            2917.99
```

- We have

$$\chi^2 = 2917.99 - 2789.18 = 128.81$$

- **Note:** Recall for hypothesis tests that $\nu^2 = 0$, the same one-sidedness applies
  $\rightarrow$ test is conservative using DF=1
  $\rightarrow$ should divide p-value by 2

- R code:

```
> 1-pchisq(128.81,1)
[1] 0
> (1-pchisq(128.81,1))/2
[1] 0
```

- The P-value is very small even with conservative test
  $\rightarrow$ we would reject the hypothesis that there is no across subject variation in intercepts in this model
  (ie, random intercept is necessary)

- Another way to fit model using SAS PROC NLMIXED (need specify the model yourself):

```
proc nlmixed data=birthwt;
 eta = b0 + b_mage*mage + u;
 p = exp(eta)/(1+exp(eta));
 model lbw~binary(p);
 random u~normal(0,s2u) subject=id;
run;
```

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits		Gra
b0	-1.9511	0.3155	877	-6.19	<.0001	-2.5702	-1.3320	-2.3
b_mage	-0.3402	0.1416	877	-2.40	0.0165	-0.6182	-0.06225	1.3
s2u	1.4560	0.2175	877	6.69	<.0001	1.0290	1.8829	-6.3

- Results are a little bit different

- Fit the model using linearization method:

```
Proc glimmix data=birthwt method= RSPL NOCLPRINT NOITPRINT;
class id;
model lbw =mage/dist=binomial link=logit s;
random int/ subject=id;
title1 'GLMM, default method RSPL';
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	id	0.9677	0.1258

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-1.6167	0.2805	877	-5.76	<.0001
mage	-0.3095	0.1292	3511	-2.40	0.0166

- When fitting GLMM for binary data:

  – Quadrature method and linearization method gave very different results

  – because linearization method tends to have large bias for binary data

Measure of dependence for the logistic-normal model

- **Latent response formulation**: Underlying the observed dichotomous response Y, there is an latent continuous response Y^*: $Y_{ij} = 1$ if $Y_{ij}^* > 0$ and $Y_{ij} = 0$ otherwise

- For latent response, we specify a linear model with a random intercept:

$$Y_{ij}^* = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + U_i + Z_{ij}$$

where $U_i \sim N(0, \nu^2)$,
Z_{ij} has a standard logistic distribution:

$$\Pr(Z_{ij} < a) = \frac{\exp(a)}{1 + \exp(a)}$$

which has mean zero and variance of $\pi^2/3$

- Thus

$$
\begin{aligned}
\Pr(Y_{ij} = 1 | \boldsymbol{x}_{ij}, U_i) &= \Pr(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i + Z_{ij} > 0) \\
&= \Pr\{Z_{ij} > -(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i)\} \\
&= 1 - \frac{\exp\{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i)\}}{1 + \exp\{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i)\}} \\
&= \frac{1}{1 + \exp\{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i)\}}
\end{aligned}
$$

$$\Rightarrow \mathrm{logit}\{\Pr(Y_{ij} = 1 | \boldsymbol{x}_{ij}, U_i)\} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i$$

- With latent response

$$Y_{ij}^* = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + U_i + Z_{ij}$$

  – $\mathrm{var}(U_i) = \nu^2$ (between-subject variance)
  – $\mathrm{var}(Z_{ij}) = \pi^2/3$ (within-subject variance)

- Recall, (residual) intraclass correlation of latent responses is

$$\rho = \mathrm{corr}(Y_{ij}^*, Y_{ik}^*|X_i) = \frac{\nu^2}{\nu^2 + \pi^2/3}$$

- Example: the model we just fitted could be rewritten as:

$$Y_{ij}^* = (\beta_0 + U_i) + \beta_1\texttt{mage} + Z_{ij}$$

  we have

$$\hat{\rho} = \frac{\hat{\nu}^2}{\hat{\nu}^2 + \pi^2/3} = \frac{1.73}{1.73 + 3.29} = 0.345$$

Generalized Linear Mixed Effects Models

Example with Random Intercept and Slope Poisson Regression Model for Count Data

- Epileptic patients $(m = 59)$ were enrolled in a clinical trial of a treatment to control seizures
- Each subject was randomized to treatment with either progabide or placebo
- Y_{ij} is the number of seizures at each of a series of eight-week (baseline) or two-week periods for two treatment groups of patients
- Conditional link function is the logarithm:

$$h(\mu_{ij}) = \log(\mu_{ij}) = \eta_{ij}, \qquad \mathrm{E}(Y_{ij}|X_i, \boldsymbol{U}_i) = \mu_{ij}$$

- The conditional distribution of Y_{ij} is

$$Y_{ij}|\boldsymbol{U}_i, X_i \sim \text{Poisson}(\mu_{ij})$$

where "Poisson(μ_{ij})" means "Poisson with mean μ_{ij}"

Note: No overdispersion here — these models require a fully-specified probability distribution (unless you use linearization method)

- The linear predictor in our example is

$$\begin{aligned}\eta_{ij} &= \beta_0 + \beta_1 \texttt{tx}_i + \beta_2 \texttt{post}_{ij} + \beta_3 \texttt{txpost}_{ij} + U_{i1} + U_{i2}\texttt{post}_{ij} + \log(\texttt{length}) \\ &= (\beta_0 + \beta_1 \texttt{tx}_i + U_{i1}) + (\beta_2 + \beta_3 \texttt{tx}_i + U_{i2})\texttt{post}_{ij} + \log(\texttt{length})\end{aligned}$$

  - Random effects $\boldsymbol{U}_i = (U_{i1}, U_{i2})' \sim N(0, G)$
  - β_0 is the mean subject-specific pre-treatment log-seizure rate for subjects on placebo

- $\beta_0 + \beta_1$ is the rate for those on treatment
- β_2 is the mean subject-specific post- versus pre-treatment log rate ratio of seizures for a subject on placebo
(this is a **within-subject** comparison)
- β_3 is the relative effect of post- versus pre-treatment for progabide subjects as compared to placebo subjects
(it compares the mean **subject-specific** post- versus pre-treatment log rate ratio for treatment to that for placebo)
- $\log(\texttt{length})$ is an **offset** term used to account for different lengths of exposure time for different observations

- SAS code to fit the model using default linearization method:

```
data seizure;
set seizure;
post = 1*(time >= 1);
txtime = tx*post;
loglength=log(length);
run;

proc glimmix data=seizure;
  class id;
  model seiz=tx post tx*post /dist=poisson offset=loglength link=log s;
  random int post/type=un subject=id;
  title1 'random intercept and slope, RSPL';
run;
```

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	id	0.5067	0.1032
UN(2,1)	id	0.05359	0.05680
UN(2,2)	id	0.2364	0.06135

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	1.0885	0.1407	57	7.73	<.0001
tx	0.04976	0.1939	177	0.26	0.7978
post	0.002198	0.1086	57	0.02	0.9839
tx*post	-0.3032	0.1512	177	-2.00	0.0465

- If fit by quadrature method in SAS:

```
Proc glimmix data=birthwt method=quad (qpoints=50);
 class id;
 model lbw =mage/dist=binomial link=logit s;
 random int/ subject=id;
 title1 'GLMM, quadrature method';
run;
```

```
                  -2 Log Likelihood            1848.90

              Covariance Parameter Estimates

       Cov                                      Standard
       Parm        Subject      Estimate           Error

       UN(1,1)     id             0.5010          0.1010
       UN(2,1)     id            0.05631         0.05590
       UN(2,2)     id             0.2333         0.06070
```

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	1.0711	0.1405	57	7.62	<.0001
tx	0.04959	0.1931	177	0.26	0.7976
post	-0.00230	0.1097	57	-0.02	0.9834
tx*post	-0.3075	0.1513	177	-2.03	0.0436

- Quadrature and linearization method provide very similar results.
  – In general the linearization method works better for count data when the estimated rate is high.
  – Quadrature method takes more time.

- **Key interpretation** (based on results using quadrature method):
  - $\hat{\beta}_2 = -.002$ is log-rate ratio ($\exp(\hat{\beta}_2) = .998$ is rate ratio) comparing post-treatment to pre-treatment for a **typical** subject on placebo
  - $\hat{\beta}_3 = -.308$ is the log-ratio of rate ratios ($\exp(\hat{\beta}_3) = .74$ is the ratio of rate ratios) comparing the post- versus pre-treatment effects for **typical** subjects on treatment versus those on placebo
  - $\hat{\text{var}}(U_{i1}) = 0.50$ is the between-subject variance in log-seizure rate at baseline for each treatment group
  - $\hat{\text{var}}(U_{i1}) = 0.23$ is the between-subject variance in subject-specific log-rate ratios comparing post- to pre-treatment seizure in either the treatment or placebo groups

- **Note:** We can also use SAS PROC NLMIXED to fit this model:

```
proc nlmixed data=seizure;
 eta = b0 + b_tx*tx + b_post*post + b_txtime*txtime
    + loglength + u1 + u2*post;
 mu = exp(eta);
 model seiz~POISSON(mu);
 random u1 u2 ~ normal([0,0],[s21,cov12,s22]) subject=id;
run;
```

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits		Gra
b0	1.0712	0.1404	57	7.63	<.0001	0.7901	1.3524	-0.
b_tx	0.04950	0.1929	57	0.26	0.7984	-0.3367	0.4357	-0.
b_post	-0.00238	0.1095	57	-0.02	0.9827	-0.2216	0.2168	-0.
b_txtime	-0.3072	0.1510	57	-2.04	0.0465	-0.6095	-0.00492	-0.
s21	0.4999	0.1007	57	4.96	<.0001	0.2982	0.7016	0.0
cov12	0.05656	0.05567	57	1.02	0.3139	-0.05491	0.1680	0.0
s22	0.2319	0.06028	57	3.85	0.0003	0.1112	0.3526	0.0

- Results are similar

- Note for Stata:
  - you can fit the model with random intercept and random slope by
    `xtmepoisson seiz tx post txpost || id:  post, cov(uns) exposure(length)`
  - Random intercept-only models can be fitted with
    `xtpoisson, re normal`
  - The advantages of `xtmepoisson` and `xtmelogit` over
    `xtpoisson, re normal` and `xtlogit, re` are:
    - Ability to include more than random intercepts
    - Empirical Bayes estimation of random effects $\boldsymbol{U}_i$

- We can test whether the random post- versus pre- effects term is necessary
  - remove the "random slope" and doing a LRT
  - Need to use likelihoods from quadrature method

- Fit a model with random intercept only in SAS:

```
proc glimmix data=seizure method=quad (qpoints=50);
  class id;
  model seiz=tx post tx*post /dist=poisson offset=loglength link=log s;
  random int/subject=id;
  title1 'random intercept only, quadrature method';
run;
```

```
                -2 Log Likelihood            2021.15
```

```
                    Solutions for Fixed Effects

                              Standard
Effect         Estimate          Error        DF    t Value    Pr > |t|

Intercept        1.0326         0.1527        57       6.76      <.0001
tx             -0.02385         0.2107       234      -0.11      0.9099
post             0.1108        0.04689       234       2.36      0.0189
tx*post         -0.1037        0.06505       234      -1.59      0.1123
```

- $\chi^2 = 2021.15 - 1848.90 = 172.25$
- R code to obtain p-value:

```
> 1-pchisq(172.25,1)
[1] 0
> (1-pchisq(172.25,1))/2  #more accurate way
[1] 0
```

- Even with conservative test (p-value not divided by 2), it clearly indicates that the data support a random slope model

Subject-specific versus Population Average Models

- **Subject-specific:** random, mixed or fixed effects models

  Population average: marginal (GEE) models

- In a marginal model: we model $\mathrm{corr}(Y_{ij}, Y_{ik})$

  In a subject-specific model: correlation arises from $\boldsymbol{U}_i$

- Recall that with the linear model, the **mixed model** is, generally

$$\mathrm{E}(Y_{ij}|\boldsymbol{U}_i, X_i, D_i) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i$$

  and the induced marginal model version could then be written as

$$\mathrm{E}(Y_{ij}|X_i) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} \quad \text{because} \quad \mathrm{E}(\boldsymbol{d}'_{ij}\boldsymbol{U}_i|X_i) = 0$$

  giving $\boldsymbol{\beta}$ both a subject specific and a population average interpretation

- This dual interpretation does not hold in a non-linear models

Example: Random Intercept Logistic Model

- Random intercept logistic model is

$$\Pr(Y_{ij} = 1 | U_i, X_i) = \frac{1}{1 + \exp\{-(U_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta})\}}$$

where $\eta_{ij} = U_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ is the "subject-specific" or "conditional" linear predictor

- If we integrate out U_i, we obtain

$$\Pr(Y_{ij} = 1 | X_i) = \int_u \frac{1}{1 + \exp\{-(u + \boldsymbol{x}'_{ij}\boldsymbol{\beta})\}} f_u(u; G)\, du$$

where $f_u(\cdot; G)$ is the Gaussian density of U_i with variance-covariance G (here $G = \nu^2$)

- This can be approximated by marginal model (ignoring the difference among subjects)

$$\Pr(Y_{ij} = 1|X_i) \approx \frac{1}{1 + \exp\{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta}^*)\}},$$

but is **not** equal to

$$\frac{1}{1 + \exp\{-(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\}},$$

– $\boldsymbol{\beta}^*$ is for marginal model: describes the ratio of populations odds
– $\boldsymbol{\beta}$ is for conditional model: describes the ratio of an individual's odds

- Zeger et al. (1988) showed that:
  For a logistic-normal model with only a random intercept

$$U_i \sim N(0, \nu^2)$$

  then

$$\boldsymbol{\beta}^* \approx (c^2\nu^2 + 1)^{-1/2}\boldsymbol{\beta}$$

  where $c = 16\sqrt{3}/(15\pi)$

- Therefore, $\boldsymbol{\beta}$ from the subject-specific logistic model does not inherit the population average interpretation in a logistic model
- The marginalized version of the random intercept model is approximately logistic, but the β's are **not** the same; rather they are **attenuated** toward zero
- The difference between β and β^* increases with ν^2.

- The figure shows subject-specific curves for $\Pr(Y_{ij} = 1|U_i)$ for several subjects, and the average of theses as the marginal mean.

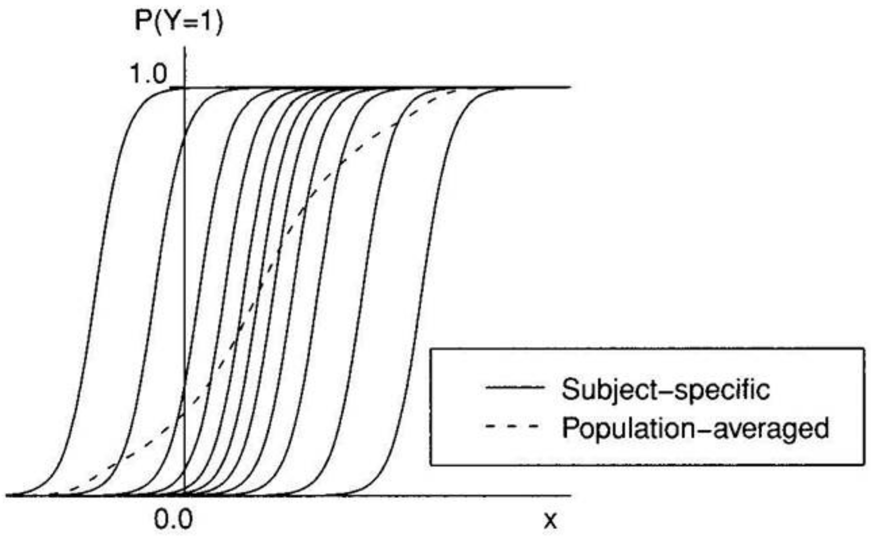

Figure 13.1 Logistic random-intercept model, showing its subject-specific curves and the population-averaged marginal curve averaging over these.

- Illustrates why the marginal effect is smaller than the conditional effect, for a single explanatory variable X.

- **Example**: Compare results from a population average (PA) model fit to our random intercept model fit:

```
*PA exhangeable model;
proc genmod data=birthwt descending;
  class id;
  model lbw =mage/ dist=binomial link=logit;
  repeated subject=id/type=exch;
  title1 'GEE';
run;
```

```
              Analysis Of GEE Parameter Estimates
             Empirical Standard Error Estimates

                   Standard   95% Confidence
Parameter Estimate    Error       Limits            Z Pr > |Z|

Intercept  -1.6368   0.2569  -2.1403  -1.1332   -6.37   <.0001
mage       -0.2426   0.1209  -0.4795  -0.0056   -2.01   0.0448
```

- Compared to our previous GLMM results:

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	-2.0396	0.3254	877	-6.27	<.0001
mage	-0.3326	0.1448	3511	-2.30	0.0217

Notes:

  - β coefficients are different; PA β^* is attenuated towards zero
  - Z or t statistics are similar in two models

Example: Random Intercept Log-linear Model

- In random intercept log-linear models for count data

$$\mathrm{E}(Y_{ij}|U_i, X_i) = \exp(U_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta})$$

where $\eta_{ij} = U_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ is the "subject-specific" or "conditional" linear predictor

- If we integrate out U_i, we obtain

$$\begin{aligned}
\mathrm{E}(Y_{ij}|X_i) &= \int_u \exp(u + \boldsymbol{x}'_{ij}\boldsymbol{\beta}) f_u(u; G)\, du \\
&= \int_u \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) \exp(u) f_u(u; G)\, du \\
&= \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) \exp(\frac{\nu^2}{2}) = \exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{\nu^2}{2})
\end{aligned}$$

where $f_u(\cdot; G)$ is the Gaussian density of U_i with variance-covariance $G = \nu^2$

- Thus,
  - Intercept $\beta_0^* = \beta_0 + \frac{\nu^2}{2}$
  - Slope $\beta_k^* = \beta_k$ (if $k > 0$)

- **Conclusion**: For log-linear model with random intercepts, all parameters for marginal model except the intercept will have the same value and interpretation as in random intercept model.

- **Example**: Compare results from a population average (PA) model fit to our random intercept model fit:

```
*PA exhangeable model;
proc genmod data=seizure;
class id;
model seiz=tx post txtime /dist=poisson offset=loglength scale=Pearson;
repeated subject=id / type=exch covb corrb corrw modelse;
run;
```

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.3476	0.1574	1.0392	1.6560	8.56	<.0001
tx	0.0265	0.2219	-0.4083	0.4613	0.12	0.9049
post	0.1108	0.1161	-0.1168	0.3383	0.95	0.3399
txtime	-0.1037	0.2136	-0.5223	0.3150	-0.49	0.6274

- Compared to our previous random intercept log-linear model:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
Intercept	id	0.6090	0.1170

Solutions for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	1.0326	0.1527	57	6.76	<.0001
tx	-0.02385	0.2107	234	-0.11	0.9099
post	0.1108	0.04689	234	2.36	0.0189
tx*post	-0.1037	0.06505	234	-1.59	0.1123

- Note:
  - Predict β_0^* in marginal model by $\hat{\beta}_0$ in conditional model by theory:

$$\hat{\beta}_0^* = \hat{\beta}_0 + \frac{\nu^2}{2} = 1.0326 + 0.6090/2 = 1.3371$$

    which is close to $\hat{\beta}_0^* = 1.3476$ estimated from marginal model
  - Other β parameters are very close from two models
    (except coefficient for `tx` are close to 0 in both models)

Estimation of Random Effects

- Just as with LMM, estimation of random effects can be accomplished with GLMM via the **empirical Bayes** method
- Recall, **Bayes** estimate of $\boldsymbol{U}_i$ is

$$\mathrm{E}(\boldsymbol{U}_i|\boldsymbol{Y}_i, X_i) = \frac{\int_{\boldsymbol{u}_i} \boldsymbol{u}_i L_i(\boldsymbol{\beta}, \boldsymbol{u}_i) f_{\boldsymbol{U}}(\boldsymbol{u}_i; G)\,\mathrm{d}\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i} L_i(\boldsymbol{\beta}, \boldsymbol{u}_i) f_{\boldsymbol{U}}(\boldsymbol{u}_i; G)\,\mathrm{d}\boldsymbol{u}_i}$$

  where

  - $L_i(\boldsymbol{\beta}, \boldsymbol{U}_i)$ is the likelihood for $\boldsymbol{U}_i$
  - $f_{\boldsymbol{U}}(\boldsymbol{u}_i; G)$ is the density of $\boldsymbol{U}_i$

- **Empirical Bayes** estimate is this same thing with estimates plugged in for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ (recall $\boldsymbol{\gamma}$ is the parameter governing G)
- These integrals can be computed in an approximate fashion such as Gaussian-Hermite quadrature

- Posterior distribution of $(\boldsymbol{U}_i|\boldsymbol{Y}_i, X_i)$ is **not normal**:
  - Mean is not the only central value that could be used as an estimate
  - Often estimate $\boldsymbol{U}_i$ using empirical posterior **mode** of the posterior distribution of $(\boldsymbol{U}_i|\boldsymbol{Y}_i, X_i)$
  - Then compute predicted values as

$$\hat{\eta}_{ij} = \boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{d}'_{ij}\hat{\boldsymbol{U}}_i$$

  and

$$\hat{\mathrm{E}}(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{U}_i) = \hat{\mu}_{ij} = h^{-1}(\hat{\eta}_{ij})$$

- SAS code for seizure example:

```
proc glimmix data=seizure method=quad (qpoints=50);
  class id;
  model seiz=tx post tx*post /dist=poisson offset=loglength link=log s;
  random int post/type=un subject=id;
  title1 'random intercept and slope, quadrature method';
  output out=out1  pred(ilink)=predicted;
run;

proc print data=out1(where=(id=201) keep=id seiz predicted) noobs;run;
proc print data=out1(where=(id=207) keep=id seiz predicted) noobs;run;
```

- SAS note:
  - `output out=out1 pred(ilink)=predicted`: creates a dataset `out1` with predicted values
  - `pred(ilink)=`: predict $\mathrm{E}(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{U}_i) = \mu_{ij} = h^{-1}(\eta_{ij})$
  - If use `pred=` without `ilink` option: predict $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i$

- Results:

id	seiz	predicted
201	18	18.1388
201	4	4.0821
201	4	4.0821
201	6	4.0821
201	2	4.0821

id	seiz	predicted
207	151	151.220
207	102	74.635
207	65	74.635
207	72	74.635
207	63	74.635

- It appears as if GLMM is doing an excellent job of capturing the variability in the data, even for the outlier subject number 207

- Note for Stata:
  Posterior estimates are available after estimation with `xtmepoisson` and `xtmelogit`:

```
xtmepoisson seiz tx post txpost || id: post , cov(uns) exposure(length)

predict muhat , mu nooffset
```

  – `mu` option: request predicted values of μ_{ij}
  – `nooffset` option: do not include offset term in prediction

Fixed Effect Model Estimation

- Purpose: avoid following assumptions for $\boldsymbol{U}_i$
  (analogous to linear fixed effects model)
  – distribution (eg, normality)
  – independence of X_i

- Application:
  - **only** interested in the effects of covariates varying within subjects
  - are not interested in estimating $\boldsymbol{U}_i$
  - model only has a random intercept U_i (in most common applications)
  - Often used for matched data in observational studies.
    $*$ stratification or matching (ie, cluster) is used to control for confounding

- model is from the **canonical** exponential family, e.g.:
  - normal Y_{ij} with linear link
  - binomial with logit link (logistic regression)
  - Poisson with log link (Poisson regression)

  But not (for simplicity in estimation):
  - probit regression
  - Poisson with square-root link

- Idea: Treat $\boldsymbol{U}_i$ as a **fixed quantity**, a nuisance parameter, and try to eliminate it from the problem
  (estimate $\boldsymbol{\beta}$ using conditional likelihood given sufficient statistics for $\boldsymbol{U}_i$)
  - suppose the density of Y_{ij} given $\boldsymbol{U}_i, X_i$ is $f_y(y_{ij}|\boldsymbol{U}_i, X_i, \boldsymbol{\beta})$
  - recall Y_{ij}'s are **independent** of one another given $\boldsymbol{U}_i, X_i$

– then the **likelihood** (conditional on $\boldsymbol{U}_i$) from subject i data is

$$L_i = f_{\boldsymbol{y}}(\boldsymbol{y}_i|\boldsymbol{U}_i, X_i, \boldsymbol{\beta}) = \prod_{j=1}^{n_i} f_y(y_{ij}|\boldsymbol{U}_i, \boldsymbol{x}_{ij}, \boldsymbol{\beta})$$

- Recall density function of Y from a scaled exponential distribution:

$$f(y;\theta,\phi) = exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\}$$

and we have

$$\mathrm{E}(Y_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{U}_i) = b'(\theta_{ij}) = \mu_{ij} = h^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i)$$

- Likelihood function for subject i is (consider $\boldsymbol{\beta}$ and $\boldsymbol{U}_i$ as parameters):

$$L_i(\boldsymbol{\beta}, \boldsymbol{U}_i) \;\; = \;\; \texttt{constant} \times \exp\left[\sum_j \{y_{ij}\theta_{ij} - b(\theta_{ij})\}\right] + \texttt{constant}$$

- If $h()$ is canonical link function $\Rightarrow \theta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i$
- Thus, we have (ignore constant)

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{U}_i) &= \prod_i \exp\left[\sum_j \left\{y_{ij}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{d}'_{ij}\boldsymbol{U}_i) - b(\theta_{ij})\right\}\right] \\
&= \prod_i \exp\left[\boldsymbol{\beta}' \sum_j \boldsymbol{x}_{ij} y_{ij} + \boldsymbol{U}'_i \sum_j \boldsymbol{d}_{ij} y_{ij} - \sum_j b(\theta_{ij})\right]
\end{aligned}
$$

- By factorization theorem, $\sum_j \boldsymbol{d}_{ij} y_{ij}$ is a **sufficient statistic** for $\boldsymbol{U}_i$ given fixed $\boldsymbol{\beta}$
- In the simple case of a random intercept $b_{0i} = (\beta_0 + U_i)$, the sufficient statistic for b_{0i} is just $\sum_j y_{ij}$
  - $\sum_j y_{ij}$ contains **all** of the information in the data $\boldsymbol{Y}_i$ about b_{0i}

- Hence, we use the likelihood conditioning on $\sum_j y_{ij}$:
  – discard data information about b_{0i} (nuisance)
  – only use data which does not contain information about b_{0i}

- This is called the **conditional likelihood**:

$$L_i^C = f_{\boldsymbol{y}}(\boldsymbol{y}_i | \sum_j y_{ij}, \boldsymbol{\beta}, U_i, X_i) = f_{\boldsymbol{y}}(\boldsymbol{y}_i | \sum_j y_{ij}, \boldsymbol{\beta}, X_i)$$

  and it does not contain $\boldsymbol{U}_i$

- For simple cases such as the random intercept model, the conditional likelihood is reasonably easy to maximize

- FE estimator can be viewed as a conditional version of RE estimator

Example: Conditional logistic regression

- Example: We are interested in the mother-specific effect of maternal age on baby's probability of low birth weight. Our model is

$$\text{logit}(\mu_{ij}) = \text{logit}\left\{\Pr(Y_{ij} = 1|U_i, X_i)\right\} = \underbrace{(\beta_0 + U_i)} + \beta_1 \texttt{mage}$$

  where U_i is considered as fixed.

  Interest is on β_1, the **mother-specific log odds ratio** relating maternal age to low birth weight

- **Note:** $b_{0i} = (\beta_0 + U_i)$ could also contain **between-subject** (i.e., between-mother) effects
  – mother's literacy level, her socio-economic status, etc.
  – get folded into U_i

- In SAS, we can use PROC LOGISTIC to fit this conditional logistic regression:

```
PROC LOGISTIC DATA=birthwt DESC;
 MODEL lbw = mage;
 STRATA id;
RUN;
```

```
                      Response Variable                  lbw
                      Number of Response Levels          2
                      Number of Strata                   878
                      Number of Uninformative Strata     580
                      Frequency Uninformative            2900
```

Analysis of Conditional Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
mage	1	0.0627	0.1956	0.1028	0.7485

```
                  Odds Ratio Estimates

                      Point          95% Wald
Effect             Estimate      Confidence Limits

mage                  1.065       0.726       1.562
```

- Interpretation: For a given mother
  – Estimated decrease in log odds of LBW 0.063 per 1-unit of maternal age (ie, 10 years).
  – Odds ratio for LBW for two babies born 10 years about **to the same mother** is

$$\widehat{\mathsf{OR}} = e^{0.063} = 1.065$$

- However, this is not at all significant and stands in sharp contrast to the results from the random intercept model
  - This is a "subject-specific" odds ratio
  - Adjusted for all subject-level variables, observed or unobserved
- Note for software:
  – In Stata, you can fit this model by
  `xtlogit lbw mage, fe`
  – In R, you can fit conditional logistic regression by
  `clogit(lbw ~ mage + strata(id))`
  (in package `survival`)

- There are 580 subjects "dropped" (ie, 580 Uninformative Strata)
  - suppose a subject (mother) has 5 observations
  - suppose that $\sum_j y_{ij} = 5$
  - then $y_{ij} = 1$ for each j

- therefore, for these subjects, conditioning on $\sum_j y_{ij}$ eliminates all information in the data, and these subjects are uninformative for β_1
- intuitively: no within-subject variation in y_{ij}

- Conditional likelihood is likelihood-based method, so that all the usual likelihood-based inferential tools are available, eg
  - likelihood ratio tests
  - Wald-tests and Wald-based confidence intervals

Generalized Linear Mixed Models (GLMM) Summary

- Four components of GLMM:
  - linear predictor (same as in LMM)
  - link function
  - distribution of Y_{ij} conditional on $\boldsymbol{U}_i$
  - distribution of $\boldsymbol{U}_i$ (same as in LMM)
- Hierarchical model formulation and interpretation:

  Example: For subject i, the probability of breastfeeding as a function of age is given by a **subject-specific** logistic model:

$$\text{logit}(\mu_{ij}) = \eta_{ij} = b_{0i} + b_{1i}\texttt{age}$$

  – **subject-specific** intercept: $b_{0i} = \beta_0 + U_{i1}$ (or, perhaps, $b_{0i} = \beta_0 + \beta_2 \texttt{sex}_i + U_{i1}$)

– **subject-specific** slope: $b_{1i} = \beta_1 + U_{i2}$

Note: Because this is a logistic regression model,

- "subject-specific intercept" is really the subject-specific log odds of breastfeeding at age 0 years (or at whatever value of age has been chosen as the center)
- "subject-specific slope" really means the subject-specific log odds ratio of breastfeeding for a unit difference in one-year of age

And:

- β_0 is the average subject-specific intercept for boys ($\texttt{sex}_\texttt{i} = 0$)
- $\beta_0 + \beta_2$ is the average subject-specific intercept for girls ($\texttt{sex}_\texttt{i} = 1$)
- β_1 is the average subject-specific slope

- Often, models are fitted with only a random intercept:
  - this sets $\text{var}(U_{i2}) = 0$, which sets $U_{i2} = 0$ for all i
  - β_1 still has the interpretation as a "**subject-specific** slope with respect to age", assuming this slope is constant across subjects
  - U_{i1} captures unobserved heterogeneity due to differences across subjects
  - Heterogeneity could be due to unobserved **confounding factors** $\rightarrow$ use **fixed effects** model estimation
    (only works for covariates that vary within-subject)
- Mixed effects models have a **subject-specific** interpretation
  - approximate formulae exist for re-expressing subject-specific models as population-average models
  - these approximations can be very useful for model checking, as fitted (approximate) marginal model means (proportions) can be compared to observed means (proportions) in the data

- Choice between a subject-specific (random effects) and a population average (marginal) model
  – Often depends on the question being asked than the data
  – both types of models can fit the data equally well

- With missing data (more details later):
  - GEE and GLMM methods will use all observations.
    * Subjects with some missing observations also contribute.
  - If the model is correctly specified
    * GEE are consistent when the responses are missing completely at random (MCAR)
    * GLMM are consistent when the responses are missing at random (MAR)

- Fitting of models with random $\boldsymbol{U}_i$'s requires numerical integration via Gauss-Hermit quadrature or linearization approximation.

- How to model $\boldsymbol{U}_i$?
  - As a **fixed effect:**

    – do not have assume $\boldsymbol{U}_i$ independent of X_i

    – do not get between-subject effects

    Uses longitudinal data to **control unmeasured between-subject confounders**

  - As a **random effect:**

    – must assume $\boldsymbol{U}_i$ independent of X_i

    – can estimate between-subject effects

    – can do empirical bayes estimation of $\boldsymbol{U}_i$'s

    Uses longitudinal data to **model individual trajectories**

- Motivations and justifications for using **random effect** model estimation (in contrast to the case of fixed effects)

  Suppose:

  - we are interested in both within- and between-subject effects (e.g., the effects of age and of sex of child)
  - it is not unreasonable to assume that $\boldsymbol{U}_i$ is independent of X_i
  - it is reasonable to assume that $\boldsymbol{U}_i$ is (at least approximately) normally distributed
  - we are interested in more than just a random intercept
  - we would like to estimate the subject-level random effects $\boldsymbol{U}_i$

- Some resources for R:

  – About GLMM

  – About conditional logistic regression and GLMM for binary data