

Introduction to the Generalized Mixed Model

Bahjat F. Qaqish
BIOS 667

The basic idea is to extend the linear mixed model:

- (1) Replace the identity link with other link functions.
- (2) Allow the conditional variance to depend on the conditional mean.

This brings about many non-trivial complications.

All the basic notation of the linear mixed model will be used.

The link function: $g(\cdot)$. This function will be applied to vectors componentwise.

The variance function: $V(\cdot)$.

Notation for the conditional mean (given the random effects) of one observation:

$$g(\nu_{ij}) = g(E[Y_{ij}|b_i]) = x_{ij}^\top \beta + z_{ij}^\top b_i$$

The conditional variance given b_i is

$$\text{var}(Y_{ij}|b_i) = \phi V(\nu_{ij}).$$

Notation for one cluster (e.g. one subject):

$$g(\nu_i) := g(E[Y_i|b_i]) = X_i \beta + Z_i b_i$$

The *residual* or *within-subject* variance:

Conditional independence of the components of Y_i given b_i will be assumed throughout. The reason will be explained in the comments section at the end.

The *between-subject* variance component, $\text{cov}(b_i) = G$. Further, b_i will be assumed to follow a multivariate normal distribution with $E[b_i] = 0$, $\text{cov}(b_i) = G$.

The matrix G can vary among groups of subjects. This will be illustrated in an example.

Notation for all the clusters:

$$g(\nu) = g(E[Y|b]) = X\beta + Zb$$

The construction of X , Z and b is exactly as it is in the linear mixed model.

$R := \text{cov}(Y|b)$ is diagonal with i th block R_i .

Random intercept logistic regression for Bernoulli outcomes:

$$\nu_{ij} = E[Y_{ij}|b_i] = P(Y_{ij} = 1|b_i),$$

$\text{logit}\nu_{ij} = b_i + x_{ij}^\top\beta$,
 $\nu_{ij} = 1/\{1 + \exp(-b_i - x_{ij}^\top\beta)\}$,
 $b_i \sim N(0, \sigma^2)$ (notation: σ^2 is g_{11}).

The marginal mean (by double expectation) is $\mu_{ij} = E[Y_{ij}] = P(Y_{ij} = 1) = E[\nu_{ij}]$,

$$\mu_{ij} = \int_{-\infty}^{\infty} \{1 + \exp(-u - x_{ij}^\top\beta)\}^{-1} \phi(u, 0, \sigma^2) du,$$

where $\phi(u, a, b)$ is the pdf of the $N(a, b)$ distribution evaluated at u . There is no closed form expression for this integral; it has to be evaluated numerically. Since Y_{ij} is a Bernoulli random variable, once μ_{ij} is available, the variance can be computed, $\text{var}(Y_{ij}) = \mu_{ij}(1 - \mu_{ij})$. Note that in general, $\text{logit}\mu_{ij}$ will be a complicated function of $(x_{ij}, \beta, \sigma^2)$. It will not be linear in the covariates, i.e. can't be written as $x_{ij}^\top\gamma$, except in trivial cases such as a single binary covariate $x_{ij} \in \{0, 1\}$ (any function of such a covariate can be written in linear form).

The joint conditional pmf of Y_i given b_i :

Because of the conditional independence assumption, this is a product,

$$P(Y_i = y_i | b_i) = \prod_{j=1}^{n_i} P(Y_{ij} = y_{ij} | b_i) = \prod_{j=1}^{n_i} \nu_{ij}^{y_{ij}} (1 - \nu_{ij})^{1-y_{ij}}.$$

The joint marginal pmf of Y_i , by double expectation, is:

$$P(Y_i = y_i) = E_{b_i} P(Y_i = y_i | b_i) = \int_{-\infty}^{\infty} P(Y_i = y_i | t) \phi(t, 0, \sigma^2) dt,$$

This pmf, viewed as a function of (β, σ^2) , is the likelihood contribution from the i th cluster. Again, it is evaluated numerically. The log-likelihood and its first (score vector) and second (information matrix) derivatives are evaluated numerically.

The marginal covariance: $E[Y_{ij}Y_{ik}] = E[E[Y_{ij}Y_{ik} | b_i]] = E[\nu_{ij}\nu_{ik}]$, the last equality following from the conditional independence assumption. Hence,

$$\text{cov}(Y_{ij}, Y_{ik}) = E[\nu_{ij}\nu_{ik}] - \mu_{ij}\mu_{ik}.$$

Clearly, even this fairly simple model is not easy to fit. The maximization is iterative.

Files: 6city*

Interpreting σ^2 : (1) Variation or heterogeneity in the population of subjects with the same x_{ij} .
 (2) Can be used to judge the covariate effects relative to the population heterogeneity.

The implied (or induced) marginal mean vector and covariance matrix can be computed using either numerical integration or simulation.

Numerical integration:

Functions `quad` in SAS/IML and `integrate` in R do (one-dimensional) numerical integration.

Simulation:

Simulate a large number of independent standard normal variates, Z , then compute $b_i = \sigma Z$,

$$\text{logit}\nu_{ij} = b_i + x_{ij}^\top\beta,$$

and ν_{ij} , then take the average (sample mean). That average is a simulation-based estimate of $E[\nu_{ij}] = \mu_{ij}$.

Similarly, for $E[\nu_{ij}\nu_{ik}]$, compute

$$\begin{aligned}\text{logit}\nu_{ij} &= b_i + x_{ij}^\top\beta, \\ \text{logit}\nu_{ik} &= b_i + x_{ik}^\top\beta,\end{aligned}$$

then take the average of $\nu_{ij}\nu_{ik}$.

Random effects logistic regression:

Models with more than one random effect (per subject) do not present any new conceptual issues. But the computations become more demanding. Models with q random effects will require, in each iteration, a q -dimensional numerical integration for each cluster, in order to compute the likelihood, score vector, etc.

In SAS, `proc nlmixed` fits generalized mixed models by maximum-likelihood, using numerical integration (also known as “quadrature”).

The intra-class correlation (ICC):

Some books give $\sigma^2/(\sigma^2 + \pi^2/3)$ as an “intraclass correlation (ICC)”, e.g. section 9.5.1 of the textbook. It must be noted that that ICC is *based entirely on latent variables*. A more relevant ICC index is one based on *observable variables*, Y_{ij} . This can be obtained via the usual variance decomposition

$$\begin{aligned}\text{var}(Y_{ij}) &= \mu_{ij}(1 - \mu_{ij}) \\ &= E[\text{var}(Y_{ij}|b_i)] + \text{var}(E[Y_{ij}|b_i]) \\ &= E[\nu_{ij}(1 - \nu_{ij})] + \text{var}(\nu_{ij}) \\ &= \text{within} + \text{between}.\end{aligned}$$

The resulting “intra-class correlation” is

$$\frac{\text{between}}{\text{total}} = \frac{\text{var}(\nu_{ij})}{\mu_{ij}(1 - \mu_{ij})}.$$

Even though this ICC is actually not a correlation, we’ll continue to refer to it as the ICC. It is clear that the decomposition will vary with the covariates. This is to be expected because the variance is related to the mean. That is, we don’t expect the same variance decomposition to apply across all covariate values. This complicates reporting a bit, but it is the nature of these models.

Either numerical integration or simulation is needed to compute the “within” and “between” terms. Simulation is quite simple and requires only minimal programming.

A numerical example:

In the 6-city study, suppose that the true parameter values are as follows: $\sigma^2 = 4.6867$ and

$$\text{logit}\nu_{ij} = b_i + \beta_1 + \beta_2 MS_i + \beta_3(AGE_{ij} - 9) = b_i - 3.1016 + 0.3986 MS_i - 0.1756(AGE_{ij} - 9).$$

Then, for a subject with $MS = 0$, the marginal mean vector μ_i (computed by numerical integration) is $(0.1614974, 0.1467228, 0.1329173, 0.120067)^\top$, and $\text{var}(\nu_{i1}) = 0.0485029$. At age 7, the total variance is $0.1614974(1 - 0.1614974) = 0.135416$; and the ICC is $0.0485029/0.135416 = 0.3581767$. Also, $E[Y_{i1}Y_{i4}] = E[\nu_{i1}\nu_{i4}] = 0.0603263$ and $\text{corr}(Y_{i1}, Y_{i4}) = 0.3422408$.

The interpretation of β is conditional on the random effect because contrasts are in ν . In within-subject contrasts, b_i as well as all subject-level covariates (e.g. MS_i) cancel out. For example, the slope for AGE is the contrast $\text{logit}\nu_{i4} - \text{logit}\nu_{i3} = -0.1756$, and this contrast does not involve the random effect b_i .

The covariate “mother smoking” varies only between clusters, not within clusters. For such covariates we consider *random pairs of subjects*. Suppose that subject A is a random subject drawn from the population of children with $MS = 1$, and independently subject B is a random subject drawn from the population of children with $MS = 0$. The contrast between the two subject-specific log odds of respiratory infection at the j th occasion is

$$\text{logit}\nu_{Aj} - \text{logit}\nu_{Bj} = (b_A - b_B) + \beta_2(MS_A - MS_B) = (b_A - b_B) + 0.3986.$$

The subject-specific random effects show up in the contrast. The contrast above is the log of the ratio of the two subject-specific odds of respiratory infection at the same occasion. It is a *random subject-pair-specific* log odds ratio. The coefficient 0.3986 is the expected value of the contrast. The expectation is taken over random pairs of subjects; one subject being a random draw from the population of children with $MS = 1$, the other being an independent random draw from the population of children with $MS = 0$. The distribution of the above contrast is $N(\beta_2, 2\sigma^2) = N(0.3986, 9.3734)$. Now we can ask some interesting questions. What is the probability that a random subject whose mother is a smoker (e.g. A) has a higher subject-specific risk of respiratory infection at the same age than a random subject whose mother is a non-smoker (e.g. B). The answer is simply the probability that a $N(0.3986, 9.3734)$ variate exceeds 0, which is $1 - \Phi(-0.3986/\sqrt{9.3734}) \approx 0.55$, only slightly over one half (0.5 would be the case if MS had no effect). The 5% and 95% quantiles of the $N(0.3986, 9.3734)$ are $(-4.6, 5.4)$, corresponding to odds ratios of $1/103$ and 229 , a very wide range. The contrast above is a log odds ratio. What about the odds ratio, the exponential of the above contrast? That odds ratio is distributed as log normal with mean $\exp(\beta_2 + 2\sigma^2/2) = \exp(0.3986 + 4.6867) \approx 162$ and median $\exp(\beta_2) = \exp(0.3986) \approx 1.5$. The distribution of the pair-specific odds ratio is highly right-skewed with mean much larger than the median. Its 5% and 95% quantiles are, of course, $(1/103, 229)$. The odds ratio is less than 1 with probability about 0.45. All of this points to large heterogeneity in the subject-specific risk. Even though mother smoking has an effect, it pales in comparison to the heterogeneity among children.

Note: Because of the skewness mentioned above care must be taken in using the mean as a single summary measure. As a hypothetical example, suppose that $\beta_2 = -1$ and $\sigma^2 = 2$. The pair-specific odds ratio has mean 2.7 and median 0.37. The mean is > 1 while the median is < 1 . Does MS increase the risk or reduce it? The probability that the odds ratio is < 1 is $\Phi(-\beta_2/\sqrt{2\sigma^2}) \approx 0.69$. That is, the odds ratio is > 1 in only 31% of the random pairs. Even though the mean of the pair-specific odds ratio is 2.7, the odds ratio is < 1 in the majority, 69%, of the pairs.

In the above model, $\text{logit}\nu_{ij}$ is linear in the covariates x_{ij} , but the marginal logit, $\text{logit}\mu_{ij}$ is

not. However, a linear approximation can be derived, and it has the form

$$\text{logit}\mu_{ij} \approx \frac{x_{ij}^\top \beta}{\sqrt{1 + t^2 \sigma^2}},$$

where $t = \sqrt{3}/\pi$ (or $t = (16/15)\sqrt{3}/\pi$, depending on some of the approximation details). Notice that the denominator is > 1 if $\sigma > 0$. This is known as the *attenuation* effect; the approximate marginal coefficients are smaller in magnitude than the conditional coefficients. Attenuation increases with σ . The attenuation factor, $1/\sqrt{1 + 3\sigma^2/\pi^2}$, is approximately 0.88 for $\sigma = 1$, 0.67 for $\sigma = 2$, 0.52 for $\sigma = 3$.

How well does the approximation work in estimates from the 6-city data? The mixed-model parameters given above for the 6-city study were actually estimates. Fitting a marginal model using GEE with exchangeable correlation produced the estimates

$$\text{logit}\hat{\mu}_{ij} = -1.8804 + 0.2651MS_i - 0.1134(AGE_{ij} - 9).$$

The observed ratios are $1.8804/3.1016 = 0.606$, $0.2651/0.3986 = 0.665$, $0.1134/0.1756 = 0.646$, while the factor $1/\sqrt{1 + 3\sigma^2/\pi^2} = 0.642$ and $1/\sqrt{1 + 3(16/15)^2\sigma^2/\pi^2} = 0.618$.

Another way to compute an attenuation factor is as follows. Using the numerical example given above for a subject with $MS = 0$, an OLS fit for $\text{logit}\mu_{ij}$ is

$$\text{logit}\mu_{ij} = -1.8762 - 0.1149(AGE_{ij} - 9).$$

So the attenuation factors are $1.8762/3.1016 = 0.60$ and $0.1149/0.1756 = 0.65$.

Prediction of b :

The posterior mode is commonly used since it does not require additional calculations after estimates have been found. It is needed in the quadrature step and so will have already been computed by the time the estimates are found. In contrast, computing the posterior mean would require extensive additional computations.

Hypothesis testing for variance components follows the same rules as in the linear mixed model.

A note on SAS proc nlmixed:

Some past versions of SAS had a bug that generated invalid standard error estimates if “empirical” and “replicate” were used in the same invocation of proc nlmixed. The bug appears to have been fixed in “SAS/STAT 15.1”.

Random effects log-linear regression:

Log-linear regression models are used for counts as well as for naturally positive outcomes such as height, weight, durations, etc. First, consider a model with only a random intercept.

$$\nu_{ij} = E[Y_{ij}|b_i],$$

$$\log \nu_{ij} = b_i + x_{ij}^\top \beta,$$

$$\nu_{ij} = \exp(b_i) \exp(x_{ij}^\top \beta),$$

The marginal mean (by double expectation) is $\mu_{ij} = E[Y_{ij}] = E[\nu_{ij}]$,

$$\mu_{ij} = E[\exp(b_i)] \exp(x_{ij}^\top \beta),$$

and

$$\log \mu_{ij} = \log E[\exp(b_i)] + x_{ij}^\top \beta.$$

Thus the log of the marginal mean follows nearly the same model as the conditional mean. The only difference is that the term $\log E[\exp(b_i)]$ is added to the intercept. That is, the intercept in the marginal model ($\log \mu_{ij}$) is different than the intercept in the conditional model ($\log \nu_{ij}$), but all the other coefficients are the same. So, elements of β other than the intercept have dual conditional and marginal interpretations (similar to the linear mixed model, but for a different reason).

Note that $\log E[\exp(b_i)] = \sigma^2/2$ since $b_i \sim N(0, \sigma^2)$.

Second, consider a model with several random effects. By analogy with the above, the following is clear. Covariates x_{ijk} that are not in z_{ij} (do not have random coefficients), and are not functionally related to z_{ij} , will factor out (just like $\exp(x_{ij}^\top \beta)$ above), and their coefficients will have dual conditional and marginal interpretations. But for covariates x_{ijk} that are in z_{ij} , or are functionally related to z_{ij} , no simplification occurs; their coefficients will have only a conditional interpretation. (example of “functionally related” are age and age²).

Why is the conditional independence assumption made? The reason is the same as in the marginal models. Other than the multivariate normal, there are no mathematically and numerically convenient multivariate distributions that allow easy modelling of the mean and covariance.