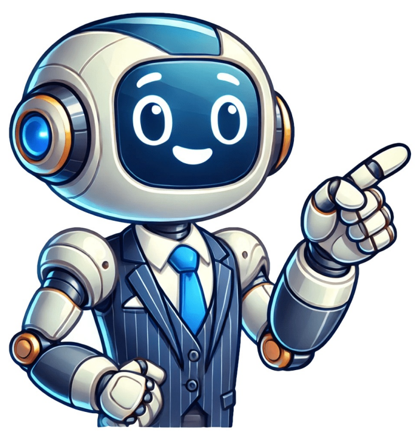
Continue



BackgroundGeneralized linear mixed models (or GLMMs) are an extension of linearmixed models to allow response variables from different distributions,such as binary responses. Alternatively, you could think of GLMMs asan extension of generalized linear models (e.g., logistic regression)to include both fixed and random effects (hence mixed models). Thegeneral form of the model (in matrix notation) is:$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$Where \(\mathbf{y}\) is a \(N \times 1\) column vector, the outcome variable;\(\mathbf{X}\) is a \(N \times p\) matrix of the \(p\) predictor variables;\(\boldsymbol{\beta}\) is a \(p \times 1\) column vector of the fixed-effects regressioncoefficients (the \(\beta\)s); \(\mathbf{Z}\) is the \(N \times q\) design matrix forthe \(q\) random effects (the random complement to the fixed \(\mathbf{X})\);\(\mathbf{u}\) is a \(q \times 1\) vector of the randomeffects (the random complement to the fixed \(\boldsymbol{\beta})\);and \(\boldsymbol{\varepsilon}\) is a \(N \times 1\)column vector of the residuals, that part of \(\mathbf{y}\) that is not explained bythe model, \(\boldsymbol{X\beta} + \mathbf{Zu}\). To recap:$$\overbrace{\mathbf{y}}^{\mbox{N x 1}} \quad = \quad\overbrace{\underbrace{\mathbf{X}}_{\mbox{N x p}} \quad \underbrace{\boldsymbol{\beta}}_{\mbox{p x 1}}}^{\mbox{N x 1}} \quad + \quad\overbrace{\underbrace{\mathbf{Z}}_{\mbox{N x q}} \quad \underbrace{\mathbf{u}}_{\mbox{q x 1}}}^{\mbox{N x 1}} \quad + \quad\overbrace{\boldsymbol{\varepsilon}}^{\mbox{N x 1}}$$To make this more concrete, lets consider an example from asimulated dataset. Doctors (\(q = 407\)) indexed by the \(j\)subscript each see \(n_{j}\) patients. So our grouping variable is thedoctor. Not every doctor sees the same number of patients, rangingfrom just 2 patients all the way to 40 patients, averaging about21. The total number of patients is the sum of the patients seen byeach doctor$$N = \sum_{j}^q n_j$$In our example, \(N = 8525\) patients were seen by doctors.Our outcome, \(\mathbf{y}\) is a continuous variable,mobility scores. Further, suppose we had 6 fixed effects predictors,Age (in years), Married (0 = no, 1 = yes),Sex (0 = female, 1 = male), Red Blood Cell (RBC) count, andWhite Blood Cell (WBC) count plus a fixed intercept andrandom intercept for every doctor. For simplicity, we are only goingto consider random intercepts. We will let every other effect befixed for now. The reason we want any random effects is because weexpect that mobility scores within doctors may becorrelated. There are many reasons why this could be. For example,doctors may have specialties that mean they tend to see lung cancerpatients with particular symptoms or some doctors may see moreadvanced cases, such that within a doctor,patients are more homogeneous than they are between doctors.To put this example back in our matrix notation, we would have:$$\overbrace{\mathbf{y}}^{\mbox{8525 x 1}} \quad = \quad\overbrace{\underbrace{\mathbf{X}}_{\mbox{8525 x 6}} \quad \underbrace{\boldsymbol{\beta}}_{\mbox{6 x 1}}}^{\mbox{8525 x 1}} \quad + \quad\overbrace{\underbrace{\mathbf{Z}}_{\mbox{8525 x 407}} \quad \underbrace{\mathbf{u}}_{\mbox{407 x 1}}}^{\mbox{8525 x 1}} \quad + \quad\overbrace{\boldsymbol{\varepsilon}}^{\mbox{8525 x 1}}$$$$\mathbf{y} = \left[\begin{array}{l} \text{mobility} \\ 2 \\ 2 \\ \ldots \\ 3 \end{array} \right] \begin{array}{l} n_{ij} \\ 1 \\ 2 \\ \ldots \\ 8525 \end{array} \quad \mathbf{X} = \left[\begin{array}{llllll} \text{Intercept} & \text{Age} & \text{Married} & \text{Sex} & \text{WBC} & \text{RBC} \\ 1 & 64.97 & 0 & 1 & 6087 & 4.87 \\ 1 & 53.92 & 0 & 0 & 6700 & 4.68 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 56.07 & 0 & 1 & 6430 & 4.73 \\ \end{array} \right] $$$$\boldsymbol{\beta} =\left[\begin{array}{c}4.782 \\.025 \\.011 \\.012 \\0 \\-.009\end{array}\right]$$Because \(\mathbf{Z}\) is so big, we will not write out the numbershere. Because we are only modeling random intercepts, it is aspecial matrix in our case that only codes which doctor a patientbelongs to. So in this case, it is all 0s and 1s. Each column is onedoctor and each row represents one patient (one row in thedataset). If the patient belongs to the doctor in that column, thecell will have a 1, 0 otherwise. This also means that it is a sparsematrix (i.e., a matrix of mostly zeros) and we can create a picturerepresentation easily. Note that if we added a random slope, thenumber of rows in \(\mathbf{Z}\) would remain the same, but thenumber of columns would double. This is why it can becomecomputationally burdensome to add random effects, particularly whenyou have a lot of groups (we have 407 doctors). In all cases, thematrix will contain mostly zeros, so it is always sparse. In thegraphical representation, the line appears to wiggle because thenumber of patients per doctor varies.In order to see the structure in more detail, we could also zoom inon just the first 10 doctors. The filled space indicates rows ofobservations belonging to the doctor in that column, whereas thewhite space indicates not belonging to the doctor in that column.If we estimated it, \(\mathbf{u}\) would be a columnvector, similar to \(\boldsymbol{\beta}\). However, in classicalstatistics, we do not actually estimate \(\mathbf{u}\).Instead, we nearly always assume that for the j th element of vector \(\mathbf{u}\):$$u_j \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$$Which is read: \(u_j\) is distributed as normal with mean zero andvariance G. Where \(\mathbf{G}\) is the variance-covariance matrixof the random effects. Because we directly estimated the fixedeffects, including the fixed effect intercept, random effectcomplements are modeled as deviations from the fixed effect, so theyhave mean zero. The random effects are just deviations around thevalue in \(\boldsymbol{\beta}\), which is the mean. So what is leftto estimate is the variance. Because our example only had a randomintercept, \(\mathbf{G}\) is just a \(1 \times 1\) matrix, the variance ofthe random intercept. However, it can be larger. For example,suppose that we had a random intercept and a random slope, then$$\mathbf{G} =\begin{bmatrix}\sigma^{2}_{int} & \sigma^{2}_{int,slope} \\\sigma^{2}_{int,slope} & \sigma^{2}_{slope}\end{bmatrix}$$Because \(\mathbf{G}\) is a variance-covariance matrix, we know thatit should have certain properties. In particular, we know that it issquare, symmetric, and positive semidefinite. We also know that this matrix hasredundant elements. For a \(q \times q\) matrix, there are\(\frac{q(q+1)}{2}\) unique elements. To simplify computation byremoving redundant effects and ensure that the resulting estimatematrix is positive definite, rather than model \(\mathbf{G}\)directly, we estimate \(\boldsymbol{\theta}\) (e.g., a triangularCholesky factorization \(\mathbf{G} = \mathbf{LDL^{T}}\)).\(\boldsymbol{\theta}\) is not always parameterized the same way,but you can generally think of it as representing the randomeffects. It is usually designed to contain non redundant elements(unlike the variance covariance matrix) and to be parameterized in away that yields more stable estimates than variances (such as takingthe natural logarithm to ensure that the variances arepositive). Regardless of the specifics, we can say that$$\mathbf{G} = \sigma(\boldsymbol{\theta})$$In other words, \(\mathbf{G}\) is some function of\(\boldsymbol{\theta}\). So we get some estimate of\(\boldsymbol{\theta}\) which we call \(\hat{\boldsymbol{\theta}}\).Various parameterizations and constraints allow us to simplify themodel for example by assuming that the random effects areindependent, which would imply the true structure is$$\mathbf{G} =\begin{bmatrix}\sigma^{2}_{int} & 0 \\0 & \sigma^{2}_{slope}\end{bmatrix}$$The final element in our model is the variance-covariance matrix of theresiduals, \(\mathbf{\varepsilon}\) or the conditional covariance matrix of\(\mathbf{y} | \boldsymbol{X\beta} + \boldsymbol{Zu}\).The most common residual covariance structure is$$\mathbf{R} = \boldsymbol{I\sigma^2_{\varepsilon}}$$where \(\mathbf{I}\) is the identity matrix (diagonal matrix of 1s)and \(\sigma^2_{\varepsilon}\) is the residual variance. Thisstructure assumes a homogeneous residual variance for all(conditional) observations and that they are (conditionally)independent. Other structures can be assumed such as compoundsymmetry or autoregressive. The \(\mathbf{G}\) terminology is commonin SAS, and also leads to talking about G-side structures for thevariance covariance matrix of random effects and R-side structuresfor the residual variance covariance matrix.So the final fixed elements are \(\mathbf{y}\), \(\mathbf{X}\),\(\mathbf{Z}\), and \(\boldsymbol{\varepsilon}\). The final estimatedelements are \(\hat{\boldsymbol{\beta}}\),\(\hat{\boldsymbol{\theta}}\), \(\hat{\mathbf{G}}\), and\(\hat{\mathbf{R}}\). The final model depends on the distributionassumed, but is generally of the form:$$\mathbf{y} | \boldsymbol{X\beta} + \boldsymbol{Zu} \sim\mathcal{F}(\mathbf{0}, \mathbf{R})$$We could also frame our model in a two level-style equation forthe \(i\)-th patient for the \(j\)-th doctor. There we areworking with variables that we subscript rather than vectors asbefore. The level 1 equation adds subscripts to the parameters\(\beta\)s to indicate which doctor they belong to. Turning to thelevel 2 equations, we can see that each \(\beta\) estimate for a particular doctor,\(\beta_{pj}\), can be represented as a combination of a mean estimate for that parameter, \(\gamma_{p0}\), and a random effect for that doctor, (\(u_{pj}\)).In this particular model, we see that only the intercept(\(\beta_{0j}\)) is allowed to vary across doctors because it is the only equationwith a random effect term, (\(u_{0j}\)). The other \(\beta_{pj}\) are constant across doctors.$$\begin{array}{l l}L1: & Y_{ij} = \beta_{0j} + \beta_{1j}Age_{ij} + \beta_{2j}Married_{ij} + \beta_{3j}Sex_{ij} + \beta_{4j}WBC_{ij} + \beta_{5j}RBC_{ij} + e_{ij} \\L2: & \beta_{0j} = \gamma_{00} + u_{0j} \\L2: & \beta_{1j} = \gamma_{10} \\L2: & \beta_{2j} = \gamma_{20} \\L2: & \beta_{3j} = \gamma_{30} \\L2: & \beta_{4j} = \gamma_{40} \\L2: & \beta_{5j} = \gamma_{50}\end{array}$$Substituting in the level 2 equations into level 1, yields themixed model specification. Here we grouped the fixed and randomintercept parameters together to show that combined they give theestimated intercept for a particular doctor.$$Y_{ij} = (\gamma_{00} + u_{0j}) + \gamma_{10}Age_{ij} + \gamma_{20}Married_{ij} + \gamma_{30}SEX_{ij} + \gamma_{40}WBC_{ij} + \gamma_{50}RBC_{ij} + e_{ij}$$Generalized Linear Mixed ModelsUp to this point everything we have said applies equally to linearmixed models as to generalized linear mixed models. Now lets focusin on what makes GLMMs unique.What is different between LMMs and GLMMs is that the responsevariables can come from different distributions besides gaussian. Inaddition, rather than modeling the responses directly,some link function is often applied, such as a log link. Wewill talk more about this in a minute. Let the linear predictor,\(\eta\), be the combination of the fixed and random effectsexcluding the residuals.\[\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma}\]The generic link function is called \(g(\cdot)\). The link functionrelates the outcome \(\mathbf{y}\) to the linear predictor\(\eta\). Thus:\[\begin{array}{l}\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} \\g(\cdot) = \text{link function} \\h(\cdot) = g^{-1}(\cdot) = \text{inverse link function}\end{array}\]So our model for the conditional expectation of \(\mathbf{y}\)(conditional because it is the expected value depending on the levelof the predictors) is:\[g(E(\mathbf{y})) = \boldsymbol{\eta}\]We could also model the expectation of \(\mathbf{y}\):\[E(\mathbf{y}) = h(\boldsymbol{\eta}) = \boldsymbol{\mu}\]With \(\mathbf{y}\) itself equal to:\[\mathbf{y} = h(\boldsymbol{\eta}) + \boldsymbol{\varepsilon}\]Link Functions and FamiliesSo what are the different link functions and families? There aremany options, but we are going to focus on three, link functions andfamilies for binary outcomes, count outcomes, and then tie it backin to continuous (normally distributed) outcomes.For a binary outcome, we use a logistic link function and theprobability density function, or PDF, for the logistic. Theseare:\[\begin{array}{l}g(\cdot) = log_{e}(\frac{p}{1 p}) \\h(\cdot) = \frac{e^{(\cdot)}}{1 + e^{(\cdot)}} \\PDF = \frac{e^{-\left(\frac{x \mu}{s}\right)}}{s \left(1 + e^{-\left(\frac{x \mu}{s}\right)}\right)^{2}} \\\text{where } s = 1 \text{ which is the most common default (scale fixed at 1)} \\PDF = \frac{e^{-(x \mu)}}{\left(1 + e^{-(x \mu)}\right)^{2}} \\E(X) = \mu \\Var(X) = \frac{\pi^{2}}{3} \\\end{array}\]Count OutcomesFor a count outcome, we use a log link function and the probabilitymass function, or PMF, for the poisson. Note that we call this aprobability mass function rather thanprobability density function because the support isdiscrete (i.e., for positive integers). These are:\[\begin{array}{l}g(\cdot) = log_{e}(\cdot) \\h(\cdot) = e^{(\cdot)} \\PMF = Pr(X = k) = \frac{\lambda^{k}e^{-\lambda}}{k!} \\E(X) = \lambda \\Var(X) = \lambda \\\end{array}\]Continuous OutcomesFor a continuous outcome where we assume a normal distribution, themost common link function is simply the identity. In this case,there are some special properties that simplify things:\[\begin{array}{l}g(\cdot) = \cdot \\h(\cdot) = \cdot \\g(\cdot) = h(\cdot) \\g(E(X)) = E(X) = \mu \\g(Var(X)) = Var(X) = \Sigma^2 \\PDF(X) = \left(\frac{1}{\Sigma \sqrt{2 \pi}}\right) e^{\frac{-(x \mu)^{2}}{2 \Sigma^{2}}}\end{array}\]So you can see how when the link function is the identity, itessentially drops out and we are back to our usual specification ofmeans and variances for the normal distribution, which is the modelused for typical linear mixed models. Thus generalized linear mixedmodels can easily accommodate the specific case of linear mixedmodels, but generalize further.InterpretationThe interpretation of GLMMs is similar to GLMs; however, there isan added complexity because of the random effects. On the linearizedmetric (after taking the link function), interpretation continues asusual. However, it is often easier to back transform the results tothe original metric. For example, in a random effects logisticmodel, one might want to talk about the probability of an eventgiven some specific values of the predictors. Likewise in a poisson(count) model, one might want to talk about the expected countrather than the expected log count. These transformationscomplicate matters because they are nonlinear and so even randomintercepts no longer play a strictly additive role and instead canhave a multiplicative effect. This section discusses this concept inmore detail and shows how one could interpret the model results.Suppose we estimated a mixed effects logistic model, predictingremission (yes = 1, no = 0) from Age, Married (yes = 1, no = 0), andIL6 (continuous). We allow the intercept to vary randomly by eachdoctor. We might make a summary table like this for the results.Mixed Effects Logistic for Remission StatusParameterEst.SEp-valueORIntercept1.467.274