Generalized linear mixed model

1   Generalized linear model

As the name suggests it is a generalization of the basic linear model. Specifically, it is a generalization of the relation between $E(y_i)$ and the linear term $\mathbf{x}_i^T \boldsymbol{\beta}$. In linear regression, we set $E(y_i) = \mathbf{x}_i^T \boldsymbol{\beta}$.

However, $\mathbf{x}_i^T \boldsymbol{\beta}$ is unrestricted as it can be any value in $(-\infty, \infty)$, but $E(y_i)$ may have restrictions depending on it's type. Like if y_i is binary, the term $E(y_i)$ can only be between 0 and 1. Specifically, $E(y_i) = P(y_i = 1)$. Hence, we need to find ways to relate $E(y_i)$ and $\mathbf{x}_i^T \boldsymbol{\beta}$.

We aim for such a relationship as it would help us to examine how different predictors in $\mathbf{x}_i$ are affecting y_i in expectation. This is helpful for statistical inferences. However, in most GLM models, the predictors happen to control both $E(y_i)$ and $V(y_i)$. This often creates problems in inference. But this is in general unavoidable. For the linear regression model, $V(y_i)$ is controlled by a separate σ parameter (which is also called the error variance). Thus such an issue is not there for linear regression models.

1.1   Some of the standard GLM models

In GLM, the models have two parts, 1) the base probability model, and 2) the link to relate the parameters of the assumed probability model with $\mathbf{x}_i^T \boldsymbol{\beta}$. For each part, one may have one or multiple choices.

Basic requirements for the base probability model is that it should support the outcome y_i. For the **link function**, we must have the mapping between $\mathbf{x}_i^T \boldsymbol{\beta}$ and $E(y_i)$ **bijective** i.e. say we set $E(y_i) = g(\mathbf{x}_i^T \boldsymbol{\beta})$, then g should be such that for each choice of $\mathbf{x}_i^T \boldsymbol{\beta}$, there is only one possible value of $E(y_i)$ and also for each choice of $E(y_i)$, there is only one possible value of $\mathbf{x}_i^T \boldsymbol{\beta}$. The link function has to be bijective as otherwise, the estimate of $\boldsymbol{\beta}$ would not be unique and thus not useful for inference.

The base probability model is also called the **error model**, as it induces the randomness or uncertainty in y_i.

Some examples:

- When y_i is binary: The choice for base probability is only Bernoulli i.e. $y_i \sim \text{Bernoulli}(p_i)$, the parameter is subject-specific as it will be associated with $\mathbf{x}_i^T \boldsymbol{\beta}$. For the link, there are multiple choices. Two of the most popular ones are 1) logistic i.e. $p_i = \frac{1}{1 + e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}$

or $\frac{e^{-\mathbf{x}_i^T\boldsymbol{\beta}}}{1+e^{-\mathbf{x}_i^T\boldsymbol{\beta}}}$; 2) probit i.e. $p_i = \Phi(\mathbf{x}_i^T\boldsymbol{\beta})$, where Φ is Cumulative distribution function for standard Gaussian (this also opens the door for many other choices); 3) log-log link: $-\log(-\log(p_i)) = \mathbf{x}_i^T\boldsymbol{\beta}$ or $\pi_i = \exp(-\exp(-\mathbf{x}_i^T\boldsymbol{\beta}))$ (popular in survival analysis for a different context).

- When y_i is count-valued: There are multiple choices for the probability model. One of the most popular choice is the Poisson distribution. In that case, our probability model is $y_i \sim \text{Poisson}(\lambda_i)$ and this requires λ_i to be positive and to respect this requirement the most popular link is $\lambda_i = \exp(\mathbf{x}_i^T\boldsymbol{\beta})$. Just for clarity, e could have been replaced any other positive valued number, but due to some computational advantages, e is the usual choice. However, $\lambda_i = (\mathbf{x}_i^T\boldsymbol{\beta})^2$ is **not a possible choice for a link** as the square function is not bijective and for each value of λ_i, there are two possible values of $\mathbf{x}_i^T\boldsymbol{\beta}$.

- When y_i is continuous, but positive valued: Normal error model can still be considered. However, other models are Gamma-regression. Here the base probability / error model is set as Gamma as $y_i \sim \text{Gamma}(\alpha_i, \beta_i)$ where $E(y_i) = \alpha_i\theta_i$ and then set α_i and θ_i in such a way that $\alpha_i\theta_i$ will be some function of $\mathbf{x}_i^T\boldsymbol{\beta}$. Unfortunately, there is no fixed choice for this step.

- When y_i is bounded, but continuous ($a \le y_i \le b$): In this case, one can take a logit transformation (i.e. $\log(\frac{y_i'}{1-y_i'})$ where $y_i' = \frac{y_i-a}{b-a}$) of the data and fit the traditional linear model with $\log(\frac{y_i'}{1-y_i'})$ as outcome since its support is $(-\infty, \infty)$. We are generally hesitant to take non-linear transformations like logit on the data. Thus one can consider Beta-regression which relies on a linear transformation of the data. If $a \le y_i \le b$, then $0 \le y_i' = \frac{y_i-a}{b-a} \le 1$ and thus we can assume $y_i' \sim \text{Beta}(\alpha_i, \theta_i)$ and relate $\frac{\alpha_i}{\alpha_i+\theta_i}$ with $\mathbf{x}_i^T\boldsymbol{\beta}$.

- When y_i is bounded and discrete: There are two possibilities. The discrete values can be either ordinal or categorical. In the case of categorical data (like types of cars), we can fit a multinomial logistic model which is a generalization of the binary case. For ordinal data (like those discrete values can mean the severity of a disease), there are some standard models for such outcome with logistic and probit links.

In a nutshell, GLM models are built on linking the expectation of the observed data with the $\mathbf{x}_i^T\boldsymbol{\beta}$. Except for normal, the mean and variance for most other distributions share one or more parameters. Thus, $\mathbf{x}_i^T\boldsymbol{\beta}$ often influence the $V(y_i)$ too. Like if $y_i \sim \text{Poisson}(e^{\mathbf{x}_i^T\boldsymbol{\beta}})$, then $E(y_i) = V(y_i) = e^{\mathbf{x}_i^T\boldsymbol{\beta}}$. It is thus often important to examine the mean and variance of the data while specifying a GLM model. Specifically, for count data, the application of the Poisson model is thus restrictive due to the theoretical mean and variance being equal. Count data (or counts of the number of successes) often have variability that far exceeds that predicted by Poisson (or binomial) distribution (*Overdispersion*). In the case of overdispersed count data, negative-binomial is a popular choice. In case of 'number of successes', beta-binomial model is a popular choice.

R package `glm` has several of the above options and also there other packages for some specific links.

2 Generalized linear mixed model

Like in the linear mixed model (LME), the generalized linear mixed model puts some subject-specific terms as random effects. Similar to LME, this model is only applicable to situations with multiple measurements from a given subject like in a longitudinal design. The random effects terms serve as both individual-specific contributions and as a tool to induce dependence among repeated observations from a given subject.

In case of LME, the subject-specific conditional expectations are $E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i$. Hence, for GLMM, the goal is to set a link between $E(\mathbf{y}_i)$ and $\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i$. Again depending on the data type, we pick a 'bijective' link. Similar to LME, We again assume $\boldsymbol{\eta}_i \sim \text{MVN}(\mathbf{0}, \mathbf{G})$. If the predictors in $\mathbf{Z}_i$ are not all included in $\mathbf{X}_i$, the expectation of $\boldsymbol{\eta}_i$ may not be assumed to be $\mathbf{0}$.

2.1 Implementation

`glmer` function in `lme4` supports a wide range of GLMM models, but not all. Specifically, `glmer` cannot implement the models from Section 2.2 It is often better to learn optimization tools to implement these methods. Also, Bayesian approaches are often preferable for implementing these methods. R package `brms` has some options.

2.2 Explicit modeling of time-dependence

2.2.1 Gridded observation times

The random effect part of the mixed model induces dependence among the observations, but we may want a more explicit formulation of time dependence. Like in autoregressive model, the expected outcome at time t conditionally depends on the outcome at $t-1$. In a balanced setting, let $\mathbf{y}_i = \{y_{i,1}, \ldots, y_{i,n}\}$ with n longitudinal visits and we can set $E(y_{i,j})$ depends on $\mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \theta y_{i,j-1}$. Like in a Poisson regression model, we can model $y_{i,j} \sim \text{Poisson}(e^{\lambda_{i,j}})$ and $\log(\lambda_{i,j}) = \mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \phi y_{i,j-1}$. Sometimes $y_{i,j-1}$ is replaced by $\log(y_{i,j-1})$ too.

Fitting the above model using standard software is easy (unless you are imposing any additional constraint on ϕ to address causality etc.). You just need to append $\{y_{i,n-1}, \ldots, y_{i,1}, 0\}$ with $\mathbf{X}_i$.

A moving average (MA) adaptation will look like $y_{i,j} \sim \text{Poisson}(e^{\lambda_{i,j}})$ and $\log(\lambda_{i,j}) = \mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \theta\lambda_{i,j-1}$. Sometimes $\lambda_{i,j-1}$ is replaced by $\log(\lambda_{i,j-1})$ too.

An Autoregressive moving average (ARMA) will be $y_{i,j} \sim \text{Poisson}(e^{\lambda_{i,j}})$ and $\log(\lambda_{i,j}) = \mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \phi y_{i,j-1} + \theta\lambda_{i,j-1}$. Both $y_{i,j-1}$ and $\lambda_{i,j-1}$ can be replaced by their logs.

However, the above models should be applied with caution as they may modulate the variability unexpectedly. The computation first needs to specify the likelihood as $\prod_i f(y_{i,1} \mid \mathbf{x}_{i,1}, \mathbf{z}_{i,1})f(y_{i,2} \mid \mathbf{x}_{i,2}, \mathbf{z}_{i,2}, y_{i,1}) \ldots f(y_{i,n} \mid \mathbf{x}_{i,n}, \mathbf{z}_{i,n}, y_{i,n-1})$ i.e. product of conditionals.

2.2.2 Irregular time

As in LME, the explicit time dependence modeling for irregular observation times can be achieved with the help of exponential or Gaussian covariance as in the following model, $y_{i,j} \sim$

Poisson($e^{\lambda_{i,j}}$) for $j = t_{i,1}, \ldots, t_{i,n_i}$ and $\log(\lambda_{i,j}) = \mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \gamma_{i,j}$ and $\boldsymbol{\gamma}_i = \{\gamma_{i,t_{i,1}}, \ldots, \gamma_{i,t_{i,n_i}}\} \sim$ MVN$(0, \boldsymbol{\Sigma}_i)$.

$$\boldsymbol{\Sigma}_i = \sigma^2 \begin{bmatrix} 1 & f(|t_{i,1} - t_{i,2}|) & \ldots & f(|t_{i,1} - t_{i,n_i}|) \\ f(|t_{i,1} - t_{i,2}|) & 1 & \ldots & f(|t_{i,2} - t_{i,n_i}|) \\ \vdots & \vdots & \ddots & \vdots \\ f(|t_{i,1} - t_{i,n_i}|) & f(|t_{i,1} - t_{i,n_i-1}|) & \ldots & 1 \end{bmatrix} \tag{1}$$

where $f(x) = \exp(-x/\rho)$ for exponential kernel and $f(x) = \exp(-x^2/\rho)$ is for Gaussian kernel.

The computation will again start with writing down the likelihood. In this case, one can directly write it for the above Poisson error model as,

$$\prod_i \left[\prod_{j=t_{i,1}}^{t_{i,n_i}} \frac{e^{-\mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \gamma_{i,j}}(\mathbf{x}_{i,j}^T\boldsymbol{\beta} + \mathbf{z}_{i,j}^T\boldsymbol{\eta}_i + \gamma_{i,j})^{y_{i,j}}}{y_{i,j}!} \right] \frac{1}{\sqrt{2\pi}\det(\boldsymbol{\Sigma}_i)^{t_{i,n_i}/2}} \exp(-\boldsymbol{\gamma}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\gamma}_i/2)$$