Generalized and Multilevel Linear Models

Felipe José Bravo Márquez

October 6, 2021

- In this class we will learn two powerful extensions to the linear model, which we have discussed extensively throughout this course.
- The first extensions is the **Generalized Linear Model** (GLM) which allows the use of distributions other than Gaussian in the outcome variable.
- GLMs can be particularly useful when our outcome variable is binary or bounded to positive values.
- **Multilevel models** (also known as hierarchical or mixed effects models), on the other hand, are useful when there are predictors at different level of variation.
- For example, when studying student performance, we may have information at different levels: individual students (e.g., family background), class-level information (e.g., teacher), and school-level information (e.g., neighborhood) [Gelman et al., 2013].
- Multilevel models extend linear regression to include categorical input variable representing these levels, while allowing intercepts and possibly slopes to vary by level [Gelman and Hill, 2006].

Generalized Linear Models

- The linear regression models of previous classes worked by first assuming a Gaussian distribution over outcomes.
- Then, we replaced the parameter that defines the mean of that distribution, μ, with a linear model.
- This resulted in likelihood definitions of the sort:

$$\begin{aligned} y_i &\sim N(\mu_i, \sigma) && \text{[likelihood]} \\ \mu_i &= \beta_0 + \beta_1 x_i && \text{[linear model]} \end{aligned}$$

- When the outcome variable is either discrete or bounded, a Gaussian likelihood is not the most powerful choice.
- Consider for example a count outcome, such as the number of blue marbles pulled from a bag.
- Such a variable is constrained to be zero or a positive integer.

Generalized Linear Models

- The problem of using a Gaussian model with such a variable is that the model wouldn't know that counts can't be negative.
- So it would happily predict negative values, whenever the mean count is close to zero [McElreath, 2020].
- In linear regression we basically replace the parameter describing the shape of the Gaussian likelihood μ with a linear model.
- The the essence of a Generalized Linear Model (GLM) is to generalize this strategy to probability distributions other than the Gaussian.
- And it results in models that look like this:

$$\begin{aligned} y_i &\sim \text{Binomial}(n, p_i) && \text{[likelihood]} \\ f(p_i) &= \beta_0 + \beta_1 x_i && \text{[generalized linear model]} \end{aligned}$$

Generalized Linear Models

- The first change we can note is that likelihood is binomial instead of Gaussian.
- For a count outcome y for which each observation arises from n trials and with constant expected value $n * p$, the binomial distribution is the de facto choice.
- The function f represents a **link** function, to be determined separately from the choice of distribution.
- Generalized linear models need a link function, because rarely is there a "μ", a parameter describing the average outcome.
- Parameters are also rarely unbounded in both directions, like μ.
- For example, the shape of the binomial distribution is determined, like the Gaussian, by two parameters.

Generalized Linear Models

- But unlike the Gaussian, neither of these parameters is the mean.
- Instead, the mean outcome is $n * p$, which is a function of both parameters.
- Since n is usually known (but not always), it is most common to attach a linear model to the unknown part, p.
- But p is a probability, so p_i must lie between zero and one.
- But there's nothing to stop the linear model $\beta_0 + \beta_1 x_i$ from falling below zero or exceeding one.
- The **link** function f provides a solution to this common problem.
- The link function that is commonly used when working with binomial GLMs is the logit function.
- It maps a parameter that is defined as a probability p (i.e., $0 \leq p \leq 1$), onto a linear model that can take on any real value.

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i$$

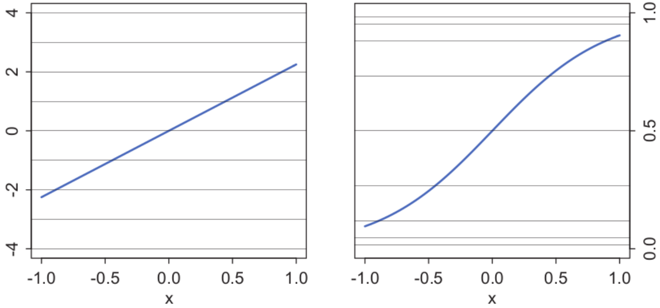
$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$$

- If we solve the logit equation for p_i we get:

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \tag{1}$$

which is known as the **sigmoid** function.

- It is also called the **logistic** or the **inverse-logit** function.



The logit link transforms a linear model (left) into a probability (right).

Generalized Linear Models

- There are two common flavors of GLM that use binomial likelihood functions and logit link functions:
  1. **Logistic regression**: when the data are organized into single-trial cases ($n = 1$), such that the outcome variable can only take values 0 and 1. In this case the likelihood function can also be represented with a Bernoulli distribution.
  2. **Aggregated binomial regression**: when the outcome can take the value zero or any positive integer up to n, the number of trials.
- Other distributions that can be used in GLMs are the Exponential, the Poisson and the Gamma distribution.
- These distributions along with the Gaussian and the Binomial are part of the exponential family, which is a set of probability distributions that can be parameterized with the following expression:

$$f(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta)) \tag{2}$$

Other types of GLMs

- There are two common flavors of GLM that use binomial likelihood functions:

Conclusions

- Blabla

References I

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013).
Bayesian data analysis.
CRC press.

Gelman, A. and Hill, J. (2006).
Data analysis using regression and multilevel/hierarchical models.
Cambridge university press.

McElreath, R. (2020).
Statistical rethinking: A Bayesian course with examples in R and Stan.
CRC press.