Generalized Linear Models

Rosario Barone*

*Email: rosario.barone@uniroma2.it

Introduction

A generalized linear model (GLM) is a flexible generalization of ordinary linear regression. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

Behind the GLM

Let suppose to be in the ordinary linear regression framework, such that:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

The described model predict the expected value $E(Y|X)$ as a linear combination of a set of observed values $(X_1, \ldots, X_k)$, implying that a constant change in one of the predictors leads to a constant change in the response variable.

Behind the GLM

This results appropriate for response variables which can vary indefinitely in either direction, or more generally for any quantity that only varies by a relatively small amount compared to the variation in the predictive variables.

However, in cases where the response variable Y is expected to be always positive and varying over a wide range, constant input changes lead to exponentially varying, rather than constantly varying, output changes. Therefore, these assumptions turn out to be inappropriate.

Wedderburn's solution

Wedderburn (with Nelder) in 1972 (he was only 25!!!) proposed a Generalization of linear models for situations in which the outcome is not Gaussian, summarized as follows:

- specify distribution for the dependent variable $f(Y|\theta)$;
- specify a link function $g(\cdot)$;
- specify a linear predictor;
- a model for the variance of the outcome (usually) automatically follows, hence heteroscedasticity.

Wedderburn's solution

The idea was to model a parameter rather than the outcome Y itself. With the linear model, this is equivalent since $E[Y] = \mu$ and $Y_i = \mu + \varepsilon_i$. In general, Y_i can not be simply expressed as a function of $E[Y]$ and a random shock.

In other words, GLMs allows:

- for response variables that have arbitrary distributions (Binomial, Multinomial, Poisson).
- arbitrary function of the response variable (the link function) to vary linearly with the predictors.

Assumption on the distribution of Y

Let $\mathbf{Y}$ represent the dependent variable and $\mathbf{X}$ represent the regressors. In the GLM, the distribution of the dependent variable $f(\mathbf{Y}|\theta)$ is assumed to belong to the exponential family. The exponential family is a parametric set of probability distributions. Some examples:

- Normal
- Poisson
- binomial (with fixed n)
- multinomial (with fixed n)
- negative binomial (with fixed number of failures).

*Note tath the parameters which must be fixed determine a limit on the size of observations.

Model definition

We define the distribution $f(Y|X)$, with mean $\boldsymbol{\mu}$ of the depending on the independent variables, X, through:

$$\mathrm{E}(Y|X) = \boldsymbol{\mu} = g^{-1}(X\boldsymbol{\beta})$$

where:

- $E(Y|X)$ is the expected value of Y conditional on X;
- $X\boldsymbol{\beta}$ is the linear predictor;
- g is the link function.

The variance is typically a function, V, of the mean:

$$\mathrm{var}(Y|X) = \nu(g^{-1}(X\boldsymbol{\beta})).$$

However, by choosing ν as a distribution of the exponential family we get a more flexible model.

Summary as overview:

In order to define a GLM we need to specify three elements:

- A probability distributions on Y .
- A linear predictor η.
- A link function g.

Linear Predictor

Let X be the matrix of observed covariates, with $(x_{i1}, \ldots, x_{ip})$ beeing the set of covariates for the i-th individual. The *systematic* component of the GLM relates the $\{\eta_i\}$ to the X using the *linear predictor*:

$$\eta_i = \sum_j \beta_j x_{ij} \quad \text{for } i = 1, \ldots, n.$$

In matrix form, it can be expressed as

$$\eta = X\boldsymbol{\beta}.$$

In other words, it is the quantity which incorporates the information about the independent variables into the model.

Link function

The GLM links η_i to $\mu_i = E(Y_i|X_i)$ by a link function $g(\cdot)$. Therefore, the link function provides the relationship between the linear predictor and the mean of the distribution function.

$$\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij} \quad \text{for } i = 1, \ldots, n.$$

The link function that transforms the mean to the natural parameter is called the *canonical link*, i.e $g(\cdot)$ is such that $g(\mu_i) = \theta_i$ and

$$\theta_i = \sum_j \beta_j x_{ij}.$$

Link function

There are several desirable statistical properties of using the canonical link:

- $\sum_i x_{ij} y_i$ for $j = 1, \ldots, p$ is the sufficient statistics;
- the Netwon Method and the Fisher scoring for finding the MLE coincide;
- the derivation of the MLE is semplified;
- the sum of the residuals is ensured to be 0.

Likelihood Inference

For n independent observations, the likelihood function is:

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log(f(y_i; \theta_i, \psi))$$

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_{i=1}^{n} c(y_i, \phi)$$

After some analytics, we get the *likelihood equations*:

$$\frac{\mathcal{L}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\text{var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0.$$

Likelihood Inference

- $\boldsymbol{\beta}$ does not appear in these equations, however $\mu_i = g^{-1}(\sum_j \beta_j x_{ij})$;
- the likelihood equations depend on the distribution of Y_i only trough μ_i and var(Y_i);
- the variance depends on the mean through a functional form var$(Y_i) = \nu(\mu_i)$.
- For most GLMs the likelihood equations are nonlinear functions of $\boldsymbol{\beta}$: we need an iterative method to solve nonlinear equations and determine the maximum of a likelihood function.

Likelihood Inference: saturated GLM

A **saturated** GLM is a model with a separate parameter for each observation: it has a perfect fitting. However, it is not helpful:

- not parsimonius
- hard interpretation

We use it as baseline for other models, such as for checking model fit!

Likelihood Inference: Deviance of the model

Let $\tilde{\mu}$ be the maximum likelihood estimate of the saturated model and let $\hat{\mu}$ be the maximum likelihood estimate of the unsaturated model we want to check.

$$-2 \log \left(\frac{\mathcal{L}(\hat{\mu}, Y)}{\mathcal{L}(\tilde{\mu}, Y)} \right) = -2 \left[\ell(\hat{\mu}, Y) - \ell(\tilde{\mu}, Y) \right]$$

After some calculations we get:

$$= \mathsf{D}(Y, \hat{\mu})/\psi,$$

which is called *scaled deviance*: the smaller is this value, the better is the fit.

Likelihood inference: Likelihood ratio

Essentially, the deviance is the likelihood-ratio statistic for testing the null hypothesis that the model holds against the alternative that a more general model holds.

Suppose we want two models, M_0 and M_1 with MLE beeing respecrively $\hat{\mu}_0$ and $\hat{\mu}_1$. Then, the likelihood-ratio statistic is

$$-2\left[\ell(\hat{\mu}_0, Y) - \ell(\hat{\mu}_1, Y)\right] = \mathsf{D}(Y, \hat{\mu}_0) - \mathsf{D}(Y, \hat{\mu}_1)$$

This statistic is large when M_1 fits better compared to M_0.

GLM residuals

For the GLM we consider two type of residuals:

- *deviance residuals*: based on the idea of evaluating the distance of the fitted model from the perfect fitting model;
- *Pearson residuals*: based on the idea of subtracting off the mean and dividing by the standard deviation.

Each of these types of residuals can be used to create an RSS-like statistic.

Quasi-Likelihood

The quasi-likelihood estimation an alternative approach proposed by Wedderburn (1974), which assumes only a mean variance relationship rather than a specific distribution for Y_i.
It has a link function and linear predictor of the usual GLM form, but instead of assuming a distributional type for Y_i; it assumes only

$$\text{var}(Y_i) = \nu(\mu_i)$$

for some chosen variance function ν.

- The equations that determine QML estimates are the same as the likelihood equations for GLMs.
- QMLE are MLE if and only if Y_i is assumed to belong to the natural exponential family (distributional assumption).