Generalized Linear Models and Maximum Likelihood

Applied Quantitative Methods (AQM)

January 16, 2019

Model

Consider observations (X, Y) where Y is the dependent variable of interest. X denotes a set of p predictors $x_1, ..., x_p$.

A **generalized linear model** (GLM) comprises three ingredients:

i. A Linear Predictor: For observation i,

$$\eta_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{1p} = \beta \cdot \mathrm{X}$$

ii. A Link Function: A link function g relates the mean of the dependent variable of interest, $E(Y_i) = \mu_i$ to the linear predictor value, η_i.

$$E(Y_i) = g^{-1}(\eta_i)$$

iii. Probability Distribution: The noise or stochasticity in the response process is modelled using a distribution in the *exponential family.*

GLM

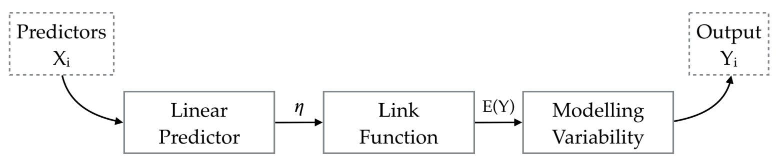

Figure 1: Schematic of the Generalized Linear Model.

Exponential Family

What is the exponential family of distributions?

The **exponential family of distributions** have a probability density given as follows, where θ is the canonical parameter and ϕ is the dispersion parameter:

$$f(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right\}$$

Exercise 1. Common distributions such as the normal distribution, binomial and Poisson distributions all belong to the exponential family. Show the equivalence mathematically by computing the specific instantiations of the various parameters.

Relation to Linear Regression

A generalized linear model generalizes the **standard linear regression** model.

$$Y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_p x_{1p} + \epsilon_i = \beta \cdot \mathsf{X} + \epsilon$$

$$\epsilon_i \sim N(0, \sigma^2)$$

Exercise 2. Identify the link function, distribution and the linear predictor components.

Poisson Regression

Poisson Regression. Poisson regression models count data, $Y_i \in \{0, 1, 2...\}$ in terms of explanatory variables $x \in \mathbb{R}^n$.

Poisson regression models Y with a Poisson distribution with the logarithm of its expected value being a linear combination of the predictor variables, x.

Exercise 3. Write down a Poisson regression model. Identify the link function, distribution and the linear predictor components for binomial and Poisson regression.

Linear Model Estimation

Consider the following linear regression model:

$$y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \epsilon = \beta \cdot \mathsf{X} + \epsilon$$

Given observations $(X_1, y_1), (X_2, y_2), ..., (X_n, y_n)$, how do we **estimate** the above model parameters?

Our objective is to identify parameter estimates, $\hat{\beta}_0$, ..., $\hat{\beta}_p$ such that the linear model fits the observed model *closely*.

Model Estimation: Least Squares Approach

Suppose that our parameter estimates are $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$.

Consider the **residuals**, $e_i = y_i - \hat{y}_i$, that is the difference between the actual response and the model's predicted response.

The **least squares approach** chooses the parameters $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$, such that it minimizes the sum of squared residuals.

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

Model Estimation: Least Squares Approach

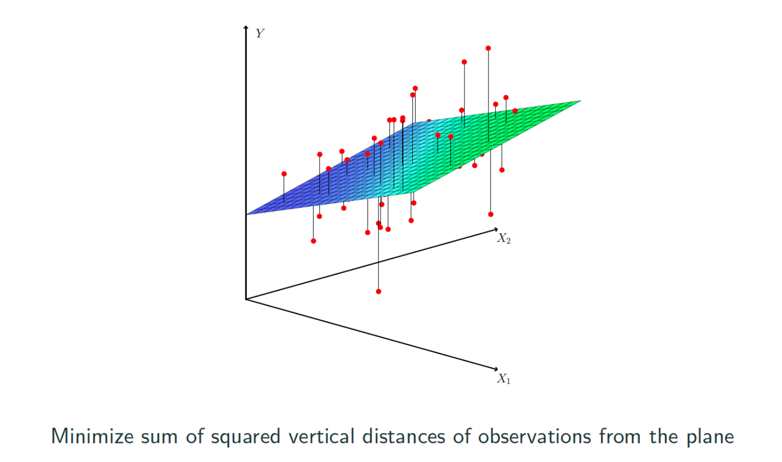

Figure 2: Graphical depiction of the least squares approach to linear regression.

Model Estimation: Least Squares Approach

Exercise 4. Consider the linear regression model, $Y = \beta \cdot \mathsf{X} + \epsilon$, where y and ϵ are $n \times 1$ vectors and X is a $n \times p$ design matrix of regressors. Obtain the OLS estimators $\hat{\beta}$ analytically (make suitable assumptions on the errors as needed).

Model Estimation: Maximum Likelihood Approach

Consider the following linear regression model:

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon = \beta \cdot \mathrm{X} + \epsilon$$

An alternative approach to estimate the model parameters is via the **maximum likelihood estimation** procedure.

Assume that the errors, $\epsilon_i \sim N(0, \sigma^2)$, are independent of X (homoskedastic) and independent across observations (i.i.d).

Model Estimation: Maximum Likelihood Approach

What is the conditional probability, $p(y_i|X = x_i, \beta, \sigma^2)$, of observing a response y_i for predictors X_i?

Then, what is the likelihood of observing, $(X_1, y_1), (X_2, y_2), ..., (X_n, y_n)$?

The maximum likelihood approach chooses the parameters such that they maximize the (log) likelihood of observing the given data.

Exercise 5. Write down the log likelihood function for a linear regression model. Analytically, maximize the log likelihood function to obtain the ML estimators of a linear regression.

In-Session Exercise

Linear Regression Implementation

Implement a linear regression model from scratch by building an optimization routine (gradient descent) that minimizes the sum of squared residuals or maximizes the log-likelihood.

Exercises

I. Linear Regression and Interpretation

1. Import the advertising dataset.
2. Run a simple linear regression on sales and newspaper advertising. Plot Sales vs Newspaper and overlay the predicted relationship. Is there a significant relationship?[1]
3. Run a multiple linear regression on sales and all three advertising media. What about the relationship between sales and newspaper advertising now?
4. Why would a significant predictor with a simple regression disappear with multiple predictors?

[1] What do the p-values represent statistically?

Exercises

II. Linear Regression: Interpretation and Extensions

1. How good is the model fit? What is a good measure of model fit?[2]
2. Obtain the 95% confidence intervals around the model parameter estimates. What is the distinction between prediction interval and confidence intervals?
3. Suppose our data has certain qualitative variables, gender, ethnicity etc., how would we incorporate these into a linear regression model?
4. Incorporate some interaction effects in your model.[3]

---

[2] How to interpret the R^2 value statistically?
[3] Did you adhere to the hierarchical principle?

Exercises

III. Linear Regression: Assumptions and Diagnostics

1. What are the assumptions made in the linear regression model with respect to the errors?
2. The assumptions can be verified by plotting the residuals e_i versus the predictor x_i. Ideally, how should the residual plot look like?
3. The normality of errors can be verified by plotting the quantile-quantile (Q-Q) plot.
4. Obtain the Q-Q plot and the residual plots for the above linear regression models.