Notes on Generalized Linear Models

compiled by D.Gueorguiev 12/1/2024

# Introductory Notes

## The Problem of Looking at Data

Suppose we have a number of measurements or counts, together with some associated structural or contextual information, such as the order in which the data were collected, which measuring instruments were used, and other differences in the conditions under which the individual measurements were made. To interpret such data, we search for a pattern, for example that one measuring instrument has produced consistently higher readings than another.

Such systematic effects are likely to be blurred or overwhelmed by other variation of a more haphazard nature. The latter variation is usually described in statistical terms, no attempt being made to model or to predict the actual haphazard contribution to each observation.

Statistical models contain both of those elements which are *systematic effects* and *random effects* accordingly. The problem of looking and understanding data demands the formulation of patterns that are thought capable of describing succinctly not only the systematic variation in the data under study, but also for describing patterns in similar data.

## Theory as Pattern

We shall consider theories as generating patterns of numbers, which in some sense can replace the data, and can themselves be described in terms of a small number of quantities. These quantities are called *parameters*. By giving the parameters different values, specific patterns can be generated.

Take for example the following very simple model:

connecting the quantities and via the parameter pair and .

In practice, however, we never measure the ’s exactly so the relationship between and is only approximately linear.

Despite this lack of exactness, we still can choose values of and , say and , that in some suitable sense best describe the now approximatelylinear relation between and . The quantities , which we denote by , are *fitted values* generated by the model and the data. The pattern they represent approximates the data values and can be summarized by the pair .

## Model Fitting

The fitting of a simple linear relationship between the s and s requires us to choose from the set of all possible pairs of parameter values a particular pair that makes the patterned set closest to the observed data. In order to make this statement precise we need a measure of ‘closeness’ or , alternatively, of distance or discrepancy between the observed s and the fitted s.

Examples of such discrepancy functions include the norm

and the norm

Least squares use the norm or sum of squared deviations as a measure of discrepancy:

The appropriateness of norms () as measures of discrepancy depends on the stochastic independence and also on the assumption that the variance of each observation is independent of its mean value.

The discrepancy functions above can be justified with statistical considerations. For instance, the classical least squares criterion arises if we regard the -values as fixed or non-stochastic and the -values are assumed to have the Normal, or Gaussian distribution with mean in which

frequency of , given , (1)

where is linearly related to through the coefficients and . The scale factor is the standard deviation of , and describes the ‘width’ of the errors when measured about the mean value.

Two interpretations of (1):

(i) If we regard (1) as a function of for fixed , the function specifies the probability density of the observations.

(ii) For a given observation y, (1) can be regarded as a function of , giving the relative plausibility of different values of for a particular observed value . In this interpretation, (1) is known as the *likelihood function*.

Let us denote with the logarithm of the likelihood function (1). Then being equal to , is identical to the sum-of-squares criterion. The function of , , attains its minimum at .

For a more complicated model in which varies systematically from observation to observation, we define the closest set to be that whose values maximize the likelihood or, equivalently, minimize .

More generally, we can extend our interest beyond a single point that minimizes , to the shape of the likelihood surface in the neighborhood of the minimum. That shape is related to the *Fisher information* metric which tells us how much information concerning the parameters there is in the data.

Statement 1: Let are independent and satisfy the linear model

for given covariates and unknown parameters .

# References

[1] [Generalized Linear Models, P. McCullagh, J.A. Nelder, 2nd Edition, 1983](https://github.com/dimitarpg13/probabilistic_machine_learning/blob/main/applied_statistics/books/GeneralizedLinearModels/GeneralizedLinearModels_McCullagh_Nelder_1983.pdf)

[2] [Generalized Linear Model, Wikipedia](https://en.wikipedia.org/wiki/Generalized_linear_model)