Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Statistik
Prof. Claudia Czado, Ph.D. (cczado@ma.tum.de)
Ariane Hanebeck (ariane.hanebeck@tum.de)

WS 2022/2023

Exercise Sheet 6
Generalized Linear Models

Discussion of the tutorial exercises on November 28, 2022.

There will be no exercise session on Thursday, December 1st, so please participate in the sessions on Monday, preferably the session from 16:15-17:45.

Preparations We study a survey that was conducted to understand the relationship between characteristics of a bank loan and characteristics of the lendee. A subset of the variables consists of the loan amount `lamount` and the variable `fixed` that indicates whether the loan used a fixed rate of interest or was allowed to vary with the Consumer's Price Index. Download the data set `loan.dat` from the moodle platform.

Problem 1 Load the `loan.dat` data set. In the following, we denote the centered and scaled observations of the predictor variable `lamount` by $\mathbf{x} = (x_1, \dots, x_n)^\top$. Let us define

$$Y_i = \begin{cases} 1, & \text{if loan } i \text{ has a fixed interest rate,} \\ 0, & \text{if loan } i \text{ has a varying interest rate.} \end{cases}$$

Then, $Y_i|X_i = x_i \sim Bin(1, p_i)$ for $i = 1, \dots, n$. We consider the model

$$p_i := p(x_i) = P(Y_i = 1|X_i = x_i) = \Phi(\beta \cdot x_i)\,, \quad \beta \in \mathbb{R}\,, \quad \text{for} \quad i = 1, \dots, n, \tag{1}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, i.e., we use $\Phi^{-1}(\cdot)$ as the link function, aka probit. Note that we have one covariate, and we do not have an intercept in this model.

a) Center and scale the predictor variable `lamount`, i.e., subtract its mean from all observations and divide them by the standard deviation of `lamount` using `R`.

b) Write down the log-likelihood for β given n pairs of observations (x_i, y_i). Then write an `R`-code for the calculation of this log-likelihood.

c) Plot the log-likelihood for β for a range of β-values using `R`, e.g., $\beta \in [-3, 3]$. Determine the maximum likelihood estimate (MLE) $\widehat{\beta}$ of β based on this plot.

d) Now, we denote the unscaled predictor variable `lamount` by $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)^\top$. Write down the formula for the estimated probabilities p_i in terms of $\tilde{x}_i$ and $\widehat{\beta}$.

e) Plot the estimate for the probability of a loan having a fixed interest rate as a function of the unscaled predictor variable `lamount`, where the function domain is $[0, 300\,000]$.

f) What is the estimated probability of a loan with `lamount` equal to 25 000 to have a varying interest rate?

g) What is the estimated probability of a loan with `lamount` equal to 180 000 to have a fixed interest rate?

h) Write down the formula for the ratio of the probability that the loan i has a fixed interest rate to the probability that the loan i has a varying interest rate, i.e., *odds* of having a fixed interest rate.

Now, instead of the model given in Equation (1), consider the following model

$$p_i := p(x_i) = P(Y_i = 1|X_i = x_i) = \frac{exp(\beta \cdot x_i)}{1 + exp(\beta \cdot x_i)}, \quad \beta \in \mathbb{R}, \quad \text{for} \quad i = 1, \dots, n, \tag{2}$$

where we use the logit link function.

i) Write down the formula for the *odds* of having a fixed interest rate, considering the model in Equation (2).

j) If the predictor variable `lamount` increases by one unit, can you quantify how much log-odds increase using probit and logit link functions in Equations (1) and (2), respectively, in terms of β.

Problem 2 (Additional) Write down the iteratively weighted least squares algorithm for the generalized linear model defined in Equation (1) of Problem 1.

Problem 3 (Additional) Let $o(1) := \frac{p(1)}{1-p(1)}$ and $o(0) := \frac{p(0)}{1-p(0)}$ be the odds of $p(1) = P(Y = 1|X = 1)$ and $p(0) = P(Y = 1|X = 0)$, respectively. The odds ratio is defined as

$$\Psi := \frac{o(1)}{o(0)} = \frac{\frac{p(1)}{1-p(1)}}{\frac{p(0)}{1-p(0)}}.$$

Prove that if the odds ratio is equal to 1 ($\Psi = 1$), the binary response Y and the dichotomous covariate X are independent. Show that the other direction works as well: if Y and X are independent, then the odds ratio is equal to 1.