Generalized Linear Models

Drew - Davison STA 310: Homework 5

1. The probability density function (PDF) of the Pareto Type I distribution is given by:

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \mathbf{1}_{\{x \geq x_m\}}$$

where:

- $\mathbf{1}_{\{x \geq x_m\}}$ is the indicator function, which is 1 if $x \geq x_m$ and 0 otherwise.
- $\alpha > 0$ is the shape parameter.
- $x_m > 0$ is the scale parameter (minimum possible value) of the data.

Is this a member of the exponential family (in canonical form), where both parameters are unknown? If so, identify the components in canonical form. If not, explain why not.

To determine whether the Pareto Type I distribution belongs to the exponential family, we compare it to the general form of an exponential family distribution:

$$f(x|\theta) = h(x) \exp\left(\eta(\theta)T(x) - A(\theta)\right). \tag{1}$$

Rewriting the given PDF:

$$f(x) = \alpha x_m^\alpha \exp\left(-(\alpha+1)\ln x\right) \mathbb{1}_{\{x \geq x_m\}}. \tag{2}$$

Comparing with the exponential family form:

- The term $\exp(-(\alpha+1)\ln x)$ suggests that the sufficient statistic is $T(x) = \ln x$.
- The coefficient $(\alpha + 1)$ corresponds to the natural parameter $\eta(\theta)$.
- The function $h(x)$ should be independent of the parameter.

A key requirement for a distribution to belong to the exponential family is that the support of x must not depend on the parameter. However, in the Pareto Type I distribution, the support is restricted to $x \geq x_m$, where x_m is a parameter.

Since the support of x depends on x_m, the Pareto Type I distribution **does not** belong to the exponential family in its canonical form. The dependency of the support on a parameter violates a fundamental property required for membership in the exponential family.

2. Assume a logistic regression model (i.e., binary regression under canonical link). Assume that y_i is binary with success probability p^1 and $i = 1, \ldots, n$. Specifically, in (simplified) logistic regression, we have that

$$\log\left(\frac{p}{1-p}\right) = X_i^T \boldsymbol{\beta}_{p \times 1},$$

Note that X_i^T is a row vector where it's dimension is $1 \times p$. a. Provide the log-likelihood function as a function of the regression parameters. That is, show that

The log-likelihood function for logistic regression is

$$\mathcal{L}(p) = \prod_{i=1}^{n} f(y_i)$$

$$= \prod_{i=1}^{n} p^{y_i} (1-p)^{1-y_i},$$

$$\log \mathcal{L}(p) = \sum_{i=1}^{n} \left[y_i \log(p) + \log(1-p) - y_i \log(1-p) \right]$$

$$= \sum_{i=1}^{n} \left[y_i \, \log\left(\frac{p}{1-p}\right) + \log(1-p) \right].$$

Recall that $X_i^T \boldsymbol{\beta}_{p \times 1}$ so the log-likelihood becomes

$$\log \mathcal{L}(p) = \sum_{i=1}^{n} \left[y_i \, \mathbf{X}_i^T \boldsymbol{\beta} - \log\left(1 + \exp\left(\mathbf{X}_i^T \boldsymbol{\beta}\right)\right) \right].$$

In matrix notation, we can write the expression as

$$\log \mathcal{L}(\beta) = \boldsymbol{y}^T X \boldsymbol{\beta} - \mathbf{1}^T \log(\mathbf{1} + \exp(X\boldsymbol{\beta}))$$

(a) Log-Likelihood Function

The probability of success in a logistic regression model is given by:

$$\log\left(\frac{p}{1-p}\right) = X\beta, \tag{3}$$

where X is the design matrix and β is the vector of regression parameters. The likelihood function is:

$$L(\beta) = \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{1-y_i}. \tag{4}$$

Taking the natural logarithm, the log-likelihood function becomes:

$$\log L(\beta) = \sum_{i=1}^{n} \left[y_i \log p_i + (1-y_i) \log(1-p_i) \right]. \tag{5}$$

---

[1]This technically should be p_i, but for simplicity of the derivation, I am supressing this notation.

Substituting $\log\left(\frac{p_i}{1-p_i}\right) = X_i\beta$ into the equation,

$$\log L(\beta) = \sum_{i=1}^{n} \left[y_i X_i \beta - \log(1 + e^{X_i\beta}) \right]. \tag{6}$$

In matrix notation:

$$\log L(\beta) = y^T X \beta - 1^T \log(1 + e^{X\beta}). \tag{7}$$

b. Show (derive) the score functions of the log-likelihood with respect to the regression parameters is as follows:

$$\nabla_{\boldsymbol{\beta}} \log \mathcal{L}(p) = \sum_{i=1}^{n} \left[y_i \, \mathbf{X}_i - \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})} \, \mathbf{X}_i \right].$$

You may also write the expression in matrix notation as follows:

Define $\pi_i = \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1+\exp(\mathbf{X}_i^T \boldsymbol{\beta})}$. Then,

$$\nabla_{\boldsymbol{\beta}} \log \mathcal{L}(p) = X^T(\boldsymbol{y} - \boldsymbol{\pi})$$

(b) Score Function

The score function, which is the gradient of the log-likelihood with respect to β, is given by:

$$\nabla_{\beta} \log L(\beta) = \sum_{i=1}^{n} \left(y_i X_i - \frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} X_i \right). \tag{8}$$

Define $\pi_i = \frac{e^{X_i^T \beta}}{1+e^{X_i^T \beta}}$, then in matrix notation:

$$\nabla_{\beta} \log L(\beta) = X^T(y - \pi). \tag{9}$$

c. Show (derive) the Hessian function of the log-likelihood with respect to the regression parameters is as follows:

$$\nabla_{\boldsymbol{\beta}}^2 \log \mathcal{L}(p) = -\sum_{i=1}^{n} \left[\frac{1}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})} \cdot \frac{\exp(\mathbf{X}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i^T \boldsymbol{\beta})} \, \mathbf{X}_i \, \mathbf{X}_i^T \right].$$

You may also write the expression in matrix notation as follows:

Let $W = diag\{\pi_1(1 - \pi_1), \ldots, \pi_n(1 - \pi_n)\}$. It follows that the Hessian can be written as $-X^T W X$.

(c) Hessian Matrix

The Hessian, which is the second derivative of the log-likelihood, is given by:

$$\nabla_{\beta}^2 \log L(\beta) = -\sum_{i=1}^{n} \left(\frac{1}{1 + e^{X_i^T \beta}} \cdot \frac{e^{X_i^T \beta}}{1 + e^{X_i^T \beta}} X_i X_i^T \right). \tag{10}$$

Using the weight matrix $W = \text{diag}\{\pi_i(1 - \pi_i)\}$, we can write the Hessian in matrix notation as:

$$\nabla_{\beta}^2 \log L(\beta) = -X^T W X. \tag{11}$$

d. Using the previous steps to write out the Newton-Raphson Algorithm in matrix form.

(d) Newton-Raphson Algorithm

The Newton-Raphson update for β is given by:

$$\beta^{(t+1)} = \beta^{(t)} - \left(\nabla_\beta^2 \log L(\beta)\right)^{-1} \nabla_\beta \log L(\beta). \tag{12}$$

Expanding,

$$\beta^{(t+1)} = \beta^{(t)} + (X^T W X)^{-1} X^T (y - \pi). \tag{13}$$

   e. A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/don't admit, is a binary variable, which can be read in using the command below.

Consider running a logistic regression using the glm() function in R below. Write your own function to implement the Newton Raphson algorithm to output the regression coefficients, and verify that the coefficients match those from the glm() function.

```r
df <- read.csv('https://stats.idre.ucla.edu/stat/data/binary.csv')
mylogit <- glm(admit ~ gre + gpa + rank, data = df, family = "binomial"(link = "logit"))
summary(mylogit)
```

```
##
## Call:
## glm(formula = admit ~ gre + gpa + rank, family = binomial(link = "logit"),
##     data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.449548   1.132846  -3.045  0.00233 **
## gre          0.002294   0.001092   2.101  0.03564 *
## gpa          0.777014   0.327484   2.373  0.01766 *
## rank        -0.560031   0.127137  -4.405 1.06e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 459.44  on 396  degrees of freedom
## AIC: 467.44
##
## Number of Fisher Scoring iterations: 4
```

```r
y<-df[,1]
# add column of 1's for the intercept
x<-as.matrix(cbind(rep(1,400),df[,c(2,3,4)]))

Newton_Raphson_logistic <- function(x, y, b.init, tol=1e-8) {
  beta <- b.init
  repeat {
    pi <- 1 / (1 + exp(-x %*% beta))
```

4

```r
    W <- diag(as.vector(pi * (1 - pi)))
    grad <- t(x) %*% (y - pi)
    H <- -t(x) %*% W %*% x
    beta_new <- beta - solve(H) %*% grad
    if (sum(abs(beta_new - beta)) < tol) break
    beta <- beta_new
  }
  return(beta)
}
```

```r
b.init <- rep(0, ncol(x))
coefficients <- Newton_Raphson_logistic(x, y, b.init)
print(coefficients)
```

```
##                      [,1]
## rep(1, 400) -3.44954840
## gre           0.00229396
## gpa           0.77701357
## rank         -0.56003139
```

This function iteratively updates β using the Newton-Raphson method until convergence. The estimated coefficients can be compared with those from `glm()` to verify correctness.

3. Propose a homework exercise that you believe would help reinforce the class content to students in this class. You may not repeat any exercise explicitly that was used in class and this exercise should be done individually.

Write and propose the question(s), include a rubric, and include a solution.

You may utilize online resource, but please do cite them if you use them and provide details.

Problem Description:

In this exercise, you are tasked with fitting a logistic regression model for predicting a binary outcome (success or failure) based on two predictors: the study time (in hours) and the prior GPA of a student. You will perform the following tasks, which will help you understand the underlying mechanics of logistic regression, assess model validity, and interpret the results.

Consider the following setup for a fictional dataset of 100 students:

- **Response Variable (Y)**: Whether a student passes or fails the exam (1 = pass, 0 = fail).

- **Predictor Variables:**

  - **Study Time (X1)**: Number of hours spent studying.
  - **GPA (X2)**: Prior GPA of the student (on a scale from 0 to 4).

The relationship between the response variable and predictors will be modeled using logistic regression.

Tasks:

1. **Identify the Binomial Random Variable**

   a. Define the response variable and assess if it follows a binomial distribution. What assumptions are required to use the binomial distribution for modeling this type of data? How would you check these assumptions?

2. **Write the Logistic Regression Model**

   a. Write the logistic regression model for predicting the probability of passing the exam using study time and GPA as predictors.

   b. Express the logistic regression model in two forms: (1) as a function of the logit (log-odds) and (2) as a function of the probability p.

3. **Fit the Logistic Regression Model**

   a. Conceptually describe the fitting procedure of logistic regression. How does it differ from fitting a linear least squares regression (LLSR) model, which you might use for continuous outcomes?

4. **Interpret the Coefficients**

   a. Suppose the estimated coefficients for the logistic regression model are:

   $$\hat{\beta}_0 = -2, \quad \hat{\beta}_1 = 0.3, \quad \hat{\beta}_2 = 0.5$$

   Interpret these coefficients in terms of odds ratios. What do these coefficients suggest about the relationship between the predictors (study time, GPA) and the probability of passing the exam?

5. **Use Residual Deviance for Model Evaluation**

   a. Define residual deviance in the context of logistic regression.

   b. How would you compare the model with predictors (study time and GPA) to a null model (model with no predictors)? What role does the residual deviance play in this comparison?

6. **Model Evaluation: Unusual Observations and Transformations**

   a. Describe how residual analysis can be used to identify unusual observations or influential points in the logistic regression model. What are the next steps you might take if you find such observations?

   b. How could transformations of the predictors (study time and GPA) affect the model fit? What types of transformations might you consider, if any?

7. **Binary vs. Binomial Responses**

   a. Discuss the difference between fitting a logistic regression model with a binary response (such as the pass/fail variable) and a binomial response.

   b. If the response variable were binomial (e.g., the number of passes out of a fixed number of students), how would the logistic regression model change? Provide a conceptual description of how you would model such a response.

Rubric:

Criterion	Points	Description
Identification of Binomial Random Variable	15	Correctly identifies the response variable as binomial and assesses the validity of the binomial assumptions.
Logistic Regression Model	20	Accurately writes the logistic regression model in both logit and p forms.
Fitting Procedure	15	Clearly explains the difference between fitting logistic regression and fitting linear regression for binary outcomes.
Interpretation of Coefficients	20	Accurately interprets the estimated coefficients in terms of odds ratios, explaining the impact of each predictor on the response variable.
Model Evaluation Using Residual Deviance	20	Correctly defines and applies residual deviance for model comparison and evaluation.
Residual Analysis and Transformations	5	Provides a clear discussion of how to use residual analysis to detect issues and potential transformations of predictors.
Binary vs. Binomial Responses	5	Demonstrates a clear understanding of the difference between binary and binomial logistic regression models.
Total	100	

Solution Outline:

1. **Identify the Binomial Random Variable**

   - The response variable 'Y' is binary: 1 = pass, 0 = fail. The assumptions of the binomial distribution are: - The trials (students) are independent. - Each trial results in either success or failure (pass or fail). - The probability of success (passing the exam) is constant across trials. These assumptions can be checked using exploratory data analysis and graphical methods.

2. **Write the Logistic Regression Model**

   - The logistic regression model is:

   $$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot \text{study time} + \beta_2 \cdot \text{GPA}$$

   The corresponding model in terms of probability is:

   $$p = \frac{1}{1 + \exp-(\beta_0 + \beta_1 \cdot \text{study time} + \beta_2 \cdot \text{GPA})}.$$

3. **Fit the Logistic Regression Model**

- Logistic regression uses maximum likelihood estimation (MLE) to find the coefficients that best fit the model. Unlike linear regression, which uses ordinary least squares, logistic regression estimates the parameters by maximizing the likelihood function.

4. **Interpret the Coefficients**

- β_0 is the log-odds of passing when study time and GPA are zero.
- β_1 (0.3): For each additional hour of study time, the odds of passing increase by a factor of $\exp(0.3) \approx 1.35$.
- β_2 (0.5): For each additional point in GPA, the odds of passing increase by a factor of $\exp(0.5) \approx 1.65$.

5. **Use Residual Deviance for Model Evaluation**

- Residual deviance is used to evaluate the model's fit. The residual deviance for a model is the difference between the log-likelihood of the fitted model and the log-likelihood of the saturated model (perfect fit). Comparing this value with the null model's deviance tells us how much better the model fits the data.

6. **Model Evaluation: Unusual Observations and Transformations**

- Residuals can be plotted to check for influential observations. If any residuals are unusually large, they may indicate outliers or influential data points. Consider transformations like logarithms or polynomial terms if the relationship between the predictors and outcome is non-linear.

7. **Binary vs. Binomial Responses**

- **Binary Response**: A binary logistic regression models the probability of one of two outcomes (e.g., pass/fail).
- **Binomial Response**: If we had counts of successes (e.g., number of passes out of a group of students), the logistic regression model would model the probability of success as a binomial distribution (number of successes out of a fixed number of trials).