Homework 7 - Predictive Modeling in Finance and Insurance

Dennis Goldenberg

2024-03-25

```
library(MASS)
library(ggplot2)
library(leaps)
Boston$chas <- factor(Boston$chas)
```

1. Generalized Linear Model

a. Showing pdf belongs to exponential family

I can modify the p.d.f and show:

$$\begin{aligned} f(y;\theta) &= \theta y^{-\theta-1} \\ &= \theta e^{\ln(y^{-\theta-1})} \\ &= \theta e^{-\ln(y)*\theta-\ln(y)} \\ &= e^{\ln(\theta)} * e^{-\ln(y)*\theta-\ln(y)} \\ &= e^{-\ln(y)\theta+\ln(\theta)-\ln(y)} \end{aligned}$$

Thus, we have that $a(y) = -\ln(y)$, $b(\theta) = \theta$, $c(\theta) = \ln(\theta)$, $d(y) = -\ln(y)$.

b. Natural Exponential family or Exponential Dispersion Family

Note that $a(y) \neq y$; thus the Pareto distribution is NOT part of the natural exponential family. However, if you restructure the p.d.f:

$$f(y;\theta) = e^{-\ln(y)\theta+\ln(\theta)-\ln(y)} = e^{-(\theta+1)\ln(y)+(1)\ln(\theta)}$$

You can note that the Pareto distribution is also part of the Exponential dispersion family, with $\lambda = 1$.

c. Score statistic

To find the score statistic, I need log-likelihood function:

$$\ell(\theta;y) = \ln(f(y;\theta)) = -\ln(y) * \theta + \ln(\theta) - \ln(y)$$

d. Pareto distribution and Exponential Family

2. Logistic Regression

3. Logistic Regression Part 2

a. Coefficients from Models

b. Probability Estimates from models

c. Expeced Values from 3 Models

d. Pearson residuals

e. Goodness of fit statistics

4. Maximum Likelihood Estimation approximation

a. Plot of number of cases against time period

b. Plot against log i

c. Fitting GLM

5. Deviance

a. Calculating Deviance

b. Calculating Pearson residual

c. Calculating deviance residual

6. Nominal Regression

a. Odds ratio

b. Probability Calculation

7. Ordinal Regression