POL 350C, Homework 5

May 9th, 2017

Assigned: 5/9/2017
Due: 5/18/2017

In this problem set we're going to continue to analyze political speech and a new data set of cabinet duration.

1 Analysis of Political Language

Using the language data from the previous problem set, we're going to use a Poisson Regression and Negative Binomial regression to model the incidence of the phrase "government takeover".

As a reminder, in the data set `SpeechDat.RData` we have the following variables:

1) Year : year of press release (either 2009 or 2010, be sure to include this as a factor)

2) ICPSR: a unique identifier for the legislator

3) obamcare: number of times "obamacare" appears in legislators' press releases

4) govTakeover: number of times "government takeover" appears in legislators' press releases

5) dem: An indicator vector that equals 1 if the legislator is a Democrat

6) dw_nom: a one-dimensional measure of legislator ideology. Negative values are more liberal, positive values are conservative.

a) Suppose that we assume the number of times "government takeover" appeares in legislators' press releases is distributed according to a Negative Binomial distribution. $\text{govTakeover}_i \sim \text{Negative Binomial}(\lambda_i, \sigma^2)$. Where $\lambda_i = \exp(X_i^{'}\boldsymbol{\beta})$

   i) Write a function to calculate the log-likelihood for a vector of parameters $\boldsymbol{\beta}$ and a value of the dispersion parameter σ^2. `Hint:` We need to find the values of $\boldsymbol{\beta}$ and σ^2 that maximize the log-likelihood, but we need to make sure that $\sigma^2 > 1$. To do this, I recommend the first few lines of your function look like the following:

```
log_lik<- function(params, X, Y){
beta<- params[1:ncol(X)] ##extracts the coefficients
sigma2<- params[ncol(X) + 1] ##extracts the sigma2
sigma2<- exp(sigma2)  + 1 ##ensures that sigma2>1
.
.
.
}
```

ii) Using the function from (a) (ii), use `optim` to obtain Maximum likelihood estimates for the following negative binomial regression:

$$\begin{aligned} Y_i &\sim \text{Negative Binomial}(\lambda_i, \sigma^2) \\ \lambda_i &= \exp(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \end{aligned}$$

Where $\boldsymbol{X}_i = (1, \text{dem}, 2010, \text{dw\_nom})$ and 2010 is an indicator equal to 1 if the year is 2010. Report the coefficients and the estimate of dispersion parameter. (Hint: remember that if `optim` reports dispersion parameter value α the estimate is $\exp(\alpha)$). Use `glm.nb` in the `MASS` library to confirm your answer. (Note: that the value θ reported from `glm.nb` is the value such that $\text{Var}(Y_i|X_i) = \lambda_i + \frac{\lambda_i^2}{\theta}$, while the parameterization used in the lecture slides has $\text{Var}(Y_i|X_i) = \lambda_i\sigma^2$.)

iii) Using `glm` estimate the analogous Poisson regression. Report those coefficients in a table.

c) Using the Negative Binomial and Poisson regressions, we're going to examine model fit and infer differences between Republicans and Democrats.

1) Assessing model fit.

i) In a histogram, provide the observed data.

ii) Taking care to maintain the same horizontal axis, provide predictions from the Poisson and Negative Binomial regression in a histogram. Compare and contrast the three histograms: which model seems to "fit" the data better?

2) Differences between Democrats and Republicans

i) Obtain the median DW-Nominate score among Democrats and the median DW-Nominate score among Republicans.

ii) Using those obtained values, compare the rate the median Democrat uses the phrase Obamcare to the rate the median Republican uses the phrase "government takeover" using the maximum likelihood estimates from the negative binomial and poisson regressions. Calculate a separate difference in 2009 and 2010. Are there noticeable differences between the two regressions?

iii) Using the multivariate normal based-simulation, calculate uncertainty for this difference. How much larger is the rate for Republicans?

2 Cabinet Duration

In this problem we're going to examine the cabinet duration data from King, Alt, Burns, and Laver (1990). To obtain the data, install the `Zelig` library and run the command `data(coalition)` at the prompt. We're going to analyze how the legal requirement of "investiture" affects the expected duration.

a) Using `R's` density function, create a density plot of `duration`. Provide the plot with clearly labeled axes and an informative title.

b) Using `Zelig` fit an exponential and weibull regression of `duration` on `fract` (a measure of fractionalization of parties) and `invest` (an indicator if there is a legal requirement for investiture.) We will follow the original authors and use `ciep12` as a censoring variable, equal to 0 if the government is censored because of the electoral period.

   To fit the models in `Zelig` use the following syntax:

```
exp_model <- zelig(Surv(duration, ciep12)~ invest + fract, model = 'exp', data = coalition)
weib_model<- zelig(Surv(duration, ciep12)~ invest + fract, model = 'weibull', data = coalitio
```

c) We're going to examine the Average Treatment Effect of investiture on expected duration.

   i) Calculate each observation's expected duration. Note, that for the parameterizations used in `zelig` if $Y \sim \text{Exponential}(\lambda_i)$ then $E[Y] = \frac{1}{\lambda_i}$ and if $Y \sim \text{Weibull}(\lambda_i, \alpha)$ then $E[Y] = \lambda_i \Gamma(1 + \frac{1}{\alpha})$.

   ii) Calculate each observation's counterfactual expected duration, if the value of investiture was changed.

   iii) Using the quantities from i and ii report the point estimate for the ATE of investiture on expected duration for both the Exponential and Weibull regression.

   iv) Using the multivariate normal simulation, obtain 95 percent confidence intervals for the ATE. Does the choice of model substantially affect your estimate?

   v) Compare the result in (iv) to the estimated effect of investiture on duration using a linear regression of duration on investiture. What do you notice?