PhD Empirical Methods
Spring 2021

Problem Set 6 – MLE and Numerical Integration
Due 5/6/2021

In 1969, the popular magazine *Psychology Today* published a 101-question survey on extramarital affairs. Professor Ray Fair (1978) extracted a sample of 601 observations on men and women who were currently in their first marriage and analyzed their responses to the survey. He used the "Tobit" model as his estimation framework for this study. The dependent variable is a count of the number of affairs, so instead of a Tobit, a standard Poisson model may be a better choice. Download the data set from github (ps5.mat), and estimate the parameters to the model below using the methods that I describe below.

Data Description

- y - count data: number of affairs in the past year.
- $\boldsymbol{x}$ - constant term=1, age, number of years married, religiousness (1-5 scale), occupation (1-7 scale), self-rating of marriage (1-5 scale)

Assingment

Assume that the data were generated from the Poisson model from problem set 5, but with a single "random coefficient" on the variable *number of years married.* The random coefficient is distributed as a two-point discrete distribution G with the following form:

$$\beta_i^{\text{no. years}} = \begin{cases} 0 & \text{with prob. } \rho \\ \beta* & \text{with prob. } 1-\rho \end{cases}$$

The following is the data generating assumptions for the Poisson model, where j is the number of affairs:

$$Pr[y_i = j] = \frac{e^{-\lambda_i}\lambda_i^j}{j!} \tag{1}$$

$$log(\lambda_i) = \boldsymbol{x}_i'\boldsymbol{\beta_i} \tag{2}$$

$$E(y_i \mid x_i) = e^{\boldsymbol{x}_i'\boldsymbol{\beta_i}} \tag{3}$$

for some $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$.

The log-likelihood function is:

$$\begin{aligned}
\ln L &= \sum_{i=1}^{n} \ln f\left(y_i \left| x_i\right., \beta_i\right) \\
&= \sum_{i=1}^{n} \ln \int \frac{e^{-\lambda_i}\lambda_i^j}{j!} dG \\
&= \sum_{i=1}^{n} \left[\int -\lambda_i dG + \int y_i \ln \lambda_i dG - \ln y_i!\right] \\
&= \sum_{i=1}^{n} \left[\int -e^{\mathbf{x_i}'\beta_\mathbf{i}} dG + \int y_i x_i'\beta_i dG - \ln y_i!\right]
\end{aligned}$$

1. Estimate the parameter vector $(\boldsymbol{\beta}, \rho)$ using maximum likelihood. Use as the starting value a vector of zeros. Use your software's implementation of the QN BFGS algorithm. To evaluate the integral in the log-liklihood function, use Monte Carlo integration. In other words, sum over draws from the two-point distribution defined by $(\beta*, \rho)$. Present results where the number of draws, R, is 10 and 30. [So, for each observation in the dataset, you will need to draw 10(30) times from two-point distribution and compute the expected log-liklihood function.]