PLSC 503: "Multivariate Analysis for Political Research"

Exercise Seven
April 2, 2020

Part I

The *exponential distribution* is a simple one-parameter distribution with PDF:

$$\Pr(X = x) = \lambda \exp(-\lambda x)$$

where $x \in [0, \infty)$ and $\lambda > 0$. You can read more about it here. Using the exponential distribution as an example, your assignment[1] is to address:

1. *Consistency*: Using a Newton-Raphson optimization method, show via simulation that $\hat{\lambda}_{MLE} \rightarrow \lambda$ as $N \rightarrow \infty$.

2. *Invariance to reparameterization*: Reparameterize the exponential such that $\phi = 1/\lambda$, and show via simulation that $\hat{\phi}_{MLE} \rightarrow \phi$ as $N \rightarrow \infty$.

3. *Optimization*: Choose up to three alternative optimization algorithms, and discuss the differences (if any) in $\hat{\lambda}_{MLE}$ obtained by them. How do those differences change (if at all) as $N \rightarrow \infty$?

Part II

In the latter part of this exercise, we'll compare parameter estimates from a simple, two-variable linear regression model obtained via OLS to those estimated through maximum likelihood. The data for this exercise (available as a `.csv` file on the course github site) are state-level data from 2009 ($N = 50$). They contain four variables:

- `State` is – unsurprisingly – the name of the state,
- `Obesity` is the percentage of adults in the state with a body mass index (BMI) greater than 30,
- `Temp` is the mean annual temperature for the state,
- `Beer` is the average per capita beer consumption in the state, in gallons.

The straightforward idea is to estimate the marginal associations of temperature and beer consumption with obesity; that is, treat `Obesity` as your "dependent variable." In particular:

[1] Hints: (a) you can use `rexp` to make random draws from an exponential distribution; (b) I am *not* asking you to write your own optimizer; you should use `maxLik` or `optim` or one of the other built-in optimizers.

1. Begin with the linear model:

$$\texttt{Obesity}_i = \beta_0 + \beta_1 \texttt{Temp}_i + \beta_1 \texttt{Beer}_i + u_i \tag{1}$$

   and assume $u \sim N(0, \sigma^2)$. Write a short program to estimate β_0, β_1, β_2, and σ^2 via maximum likelihood. Discuss your substantive "findings" to the best of your ability.

2. Estimate the same model via OLS, and compare the OLS and MLE results. What can you say about the various coefficient estimates obtained via the two methods? Their standard errors? The variance parameter σ^2? What can you say about the relative "goodness" of the two estimates?

3. Evaluate the following hypotheses for both the OLS- and MLE-estimated models:

   (a) $\beta_1 = 0.3$

   (b) $\beta_2 = \beta_1$

   (c) $\sigma^2 = 10$

   and discuss the implications of those evaluations.

4. Reestimate your MLE results, choosing a different optimization routine; so, if you used some form of Newton algorithm the first time around, switch to BHHH, or DFP, etc. Briefly discuss the differences (if any) between these "new" MLEs and those reported above.

5. Finally, reestimate your model under the assumption that the stochastic component follows a logistic distribution (that is, $u \sim \text{logistic}(0, s^2)$). You can read more about the logistic distribution in Evans et al., or here. Report those estimates as well.

This assignment is due (in the usual electronic fashion) at 5:00 p.m. EST on Friday, April 10, 2020 and is worth 50 points.