Stat238: Lab 9

November 4

Problems

1. Further discussion of Harm's applied problem:

   I am trying to build a model for predicting media slant for financial news. The prediction is that investors might have heterogeneous preferences for what financial news they want to read. For various reasons people might prefer to read news in accordance with their prior beliefs. There is some empirical support for this happening with political news, where studies found pretty significant media slant in news articles about the same congressional hearings, where the direction of the slant (to the left or right) could be well predicted by the political demography of the zip code where the newspaper circulated. The mechanism is quite clear: If the news market is competitive, then what should happen is that news will target different audiences, audiences that are basically clusters across some preference dimensions (like preference for left or right ideology). Whether this happens for financial news is not quite clear however, because we are not sure to what extent people have heterogeneous preferences for financial news. To analyze these questions, I managed to get hold a large dataset of online news about different company earnings announcements by different news sources.

   What I would like to do is estimate the financial sophistication and degree of optimism for the target audiences for each news source where these quantities are not directly observable but could impact the language of the news posts and whether a given earnings announcement is covered by a post. The observable variables I can construct are noisy measures of writing style and tone of the news posts. On top of that we can observe a bunch of company characteristics which should in theory also affect news worthiness and interact with the sophistication and optimism dimensions of target groups. And I have the text of the earnings announcement as published by the company itself as benchmark for slant. So, the goal of the exercise is to derive some estimate of each news source's unobserved choice of target group in terms of their average sophistication and optimism.

2. For problem 2a on PS4, I ask you to compute $p(y)$ for the normal and skew-normal models. For lab I would like to compare the estimates of $p(y)$ that different people get so we can have some idea of whether you are getting robust estimates. So please choose one or more of the computational approaches for estimating $p(y)$ from either the prior or the posterior and get estimates for different numbers of draws from the prior or posterior (try this for the simple normal model first). We'll then compare values on the board near the end of lab.