Review of An's Project

Yinan Zhu

April 14, 2017

Hi, An. It is nice to read your analysis on the cheese data. Overall your model is clear. Your coding style is very reader-friendly. You are also good at visualizing data. Reading your work really inspires myself in a lot of aspects. However I have some concerns about your model theoretically, maybe it's just some points I didn't really get. Also a little suggestion on data analysis excecution.

1 Model setup and derivation

This section is based on the pdf file which explains the model and derives the posterior distribution.

1. Consider including a term $log(price)_i * disp_i * \mu_{si}$, where μ is another coefficient you can fit in the similar way as β. Effectively, in the case that a store display the cheese, the elacity will be $\beta_{si} + \mu_{si}$ while in the absense of display, the elacity is β_{si}. In other words, this term answers the question: Will the elacity be affected by the adds? Does this effect varies among store? This is definitely a worth while thing to study.

2. I guess ϵ_i is a random variable with some normal distribution? But it seems that it didn't play any role in the posterior distribution you gave.

   Maybe you include its variance in α_{si}? But there are different random effects here, one is on the observable and the other is on the intercept (which varies across stores). By doing this, you scrambled these two effects together.

2 Hierachical model using lmer package

This section is based on "cheese.r"

1. It would be more illuminating to visualize the models in various ways. Such as an analysis on the random effect of each store, the correlation of different factors, and the best fit line of a specific store (instead of stacking them together).

2. You have tried a large variety of models which is cool. More can be studied across different models except the RMS. For example, m3 and m4 seems to have similar RMS, does that mean the term $(lprice|disp)$, which is the main difference between them, have no effect ?

3 Hierachical model using MCMC

This section is based on "cheese_mc.r"

1. The runtime is unexpectedly long, considering your gibbs sampling is relatively simple (no multivariate normal or Wishart distribution). I think one of the reasons is that in the loops that update sum_a or similar, you compute quantity like $tauv * lprice^2$ every single iteration. You can compute $lprice^2$ once for all and store this result as a vector. Also when using rnorm you can make the mean and sd as vector to generate a vector of random variables to reduce the number of calling functions.

2. I am interested in how this model compares with result of lmer, may be you can add some analysis in that aspect.