Discussion of

Structural Estimation of the Effect of Out-of-Stocks
by Eric Bradlow

A Bayesian Clusterwise Variable Selection Procedure
for Market Segmentation
by Feng Liang

Ed George

OBayes09
University of Pennsylvania
June 2009

0. What do these papers have in common?

- Both are driven by interesting, important real problems
- Both are about estimating probability consumer i makes choice j
- Both model these probabilities as functions of individual utilities
- Both model utilities as linear models of covariates

1. Out-of-Stock Paper by Bradlow

- Standard Multinomial Logit Choice Model

$$\Pr(i \text{ chooses } j) = \frac{\exp(\beta_i X_j + \xi_j)}{1 + \sum_k \exp(\beta_i X_k + \xi_k)}$$

  - Consumers $i = 1, \ldots, n$
  - Alternatives $j = 1, \ldots, J \quad (j = 0$ is the no-purchase option)

- Ideal data:
  - Complete purchase history
  - Full product availability during all purchases

- Actual data:
  - Total number of customers
  - Total sales of each j
  - Starting and ending inventories I^j and $\hat{I}^j$
- Suppose $I^j > \hat{I}^j = 0$. Uh-oh!
  j became unavailable at some unknown point.
- Solution: Modify model to

$$\Pr(i \text{ chooses } j) = \frac{a_i^j \cdot \exp(\beta_i X_j + \xi_j)}{1 + \sum_k a_i^k \cdot \exp(\beta_i X_k + \xi_k)}$$

  Then augment data with complete customer purchase history w and implied availabilities a (a_i^j is availability of j for i)
- MCMC simulation can now be used for Bayes estimation of ALL unknowns

Comments

- Cool things that you can do with the simulated estimates
  - Estimate lost sales due to stock-outs
  - Measure effectiveness of using promotions to recapture lost sales
- The strategy for moving around the w space is to partition into pairs and then to consider all pairwise swamps. Is this optimal in any sense? What about random pairwise swaps? Or random permutations?
- As you show, your approach is clearly better than ignoring stock outs. Are there other better straw men?
- In future, will complete data on w be available? How might your models be elaborated if you had such complete data? Would you use the same modeling assumptions?

2. Segmentation Based Variable Selection by Liang

- Probit Choice Model

$$\Pr(i \text{ chooses product based on profile } \ell) = \Phi(x_{i\ell}\beta_i + \epsilon_{i\ell})$$

  - Consumers $i = 1, \ldots, n$
  - Profiles $\ell = 1, \ldots, m$
  - $x_{i\ell}$ is $1 \times p$, with p LARGE

- Limitation: Most consumers base choices on only a few $x_{i\ell}$ components.

- Idea: Use auxiliary data to partition consumers into homogenous segments such that

$$\beta_{ij} \mid [i \text{ in segment } s] \sim N(\mu_{sj}, \tau^2)$$

- Variable selection prior:

$$\mu_{sj} \sim (1-p)\delta_0 + pN(0,v)$$

- Segmentation prior:

$$\pi(H_1,\ldots,H_n) \propto \exp\left(\lambda \sum e_{ii'}\, 1_{([H_i = H_{i'}]}\right)$$

where $e_{ii'}$ measures "similarity" of i and i'

Comments

- Applaud the general strategy of local modeling
  - Partition global heterogeneous data into local homogenous subsets
  - Now use simple models for local subsets
- I vote for treating λ as a tuning constant to be used for EDA.
- Why not do model estimation and partitioning jointly? Similarity of $\beta_i' s$ seems relevant for for partitioning?
- Alternative segmentation strategies: Treed modeling based on CART? Soft clustering?
- Ignoring natural constraints on the regression coefficients may be unwise.