6440

FINAL PROJECT PRESENTATION

Hierarchical Models of Consumers' Purchase Frequency

Report: Muqian Ye

Contents

PART 1

Background

The existing methods / theories

Background

Purchase Count

Follows a Poisson distribution. The timing of purchases is somewhat random, but the rate (counts / unit time) is constant.

Priors on latent parameters

Assume the long-run average purchase rates of individuals follow a Gamma distribution.

Negative Binomial Distribution Theory

NBD model enables the prediction of repeat purchase from data on the purchase from data on the purchase frequency of the brand for any given reporting period.

Log linear model

Poisson log linear-model is for n observations that take integer count values. Each Yi is modeled as an independent Poisson(λi) random variable, where log(λi) is a linear combination of the covariates corresponding to the i th observation.

PART 2

Data and Variables

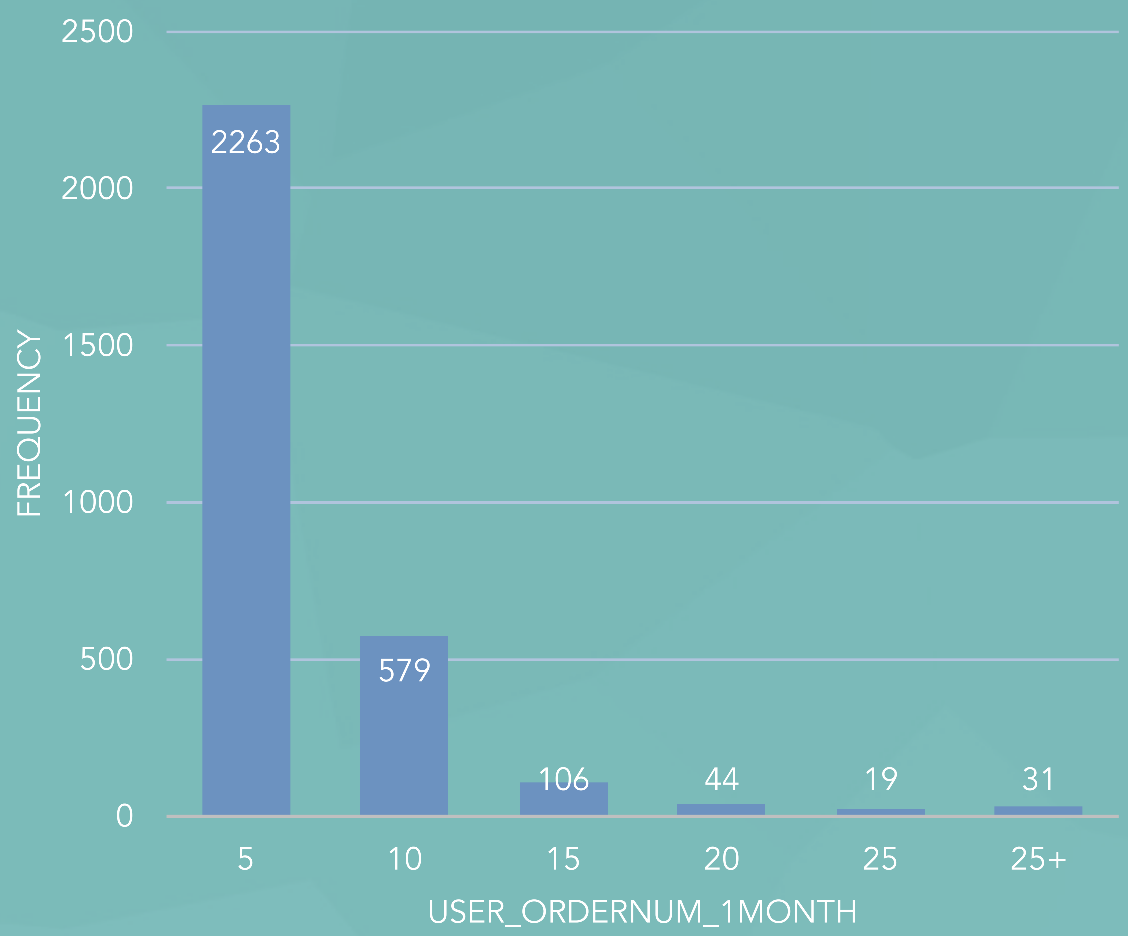
Where the data comes from / What the data look like

Data and Variables

- Data comes from *Ctrip*, a provider of travel services
- Dataset includes characteristics of products and consumers, also part of their purchase history
- 3042 consumers who placed orders at 2017-06-06 are selected
- Dependent variable is total amounts of orders placed by consumers in last month
- Except for the ID of consumers, five covariates are chosen as explanatory variable to build the model

Variable	Description
uid	ID of the consumers
user_avgadvanceddate	Average advanced date to place orders
user_avgstar	Average level of hotel reservation
user_activation	Level of activity
user_citynum	Number of cities involved by past orders
user_avgdealprice	Average price of dealt orders
user_ordernum_1month	Number of orders placed in last month

Data and Variables



PART 3

Hierarchical Model

Model specification / Priors of parameters / Conditional posterior distribution

Hierarchical Model

Basic model & Priors

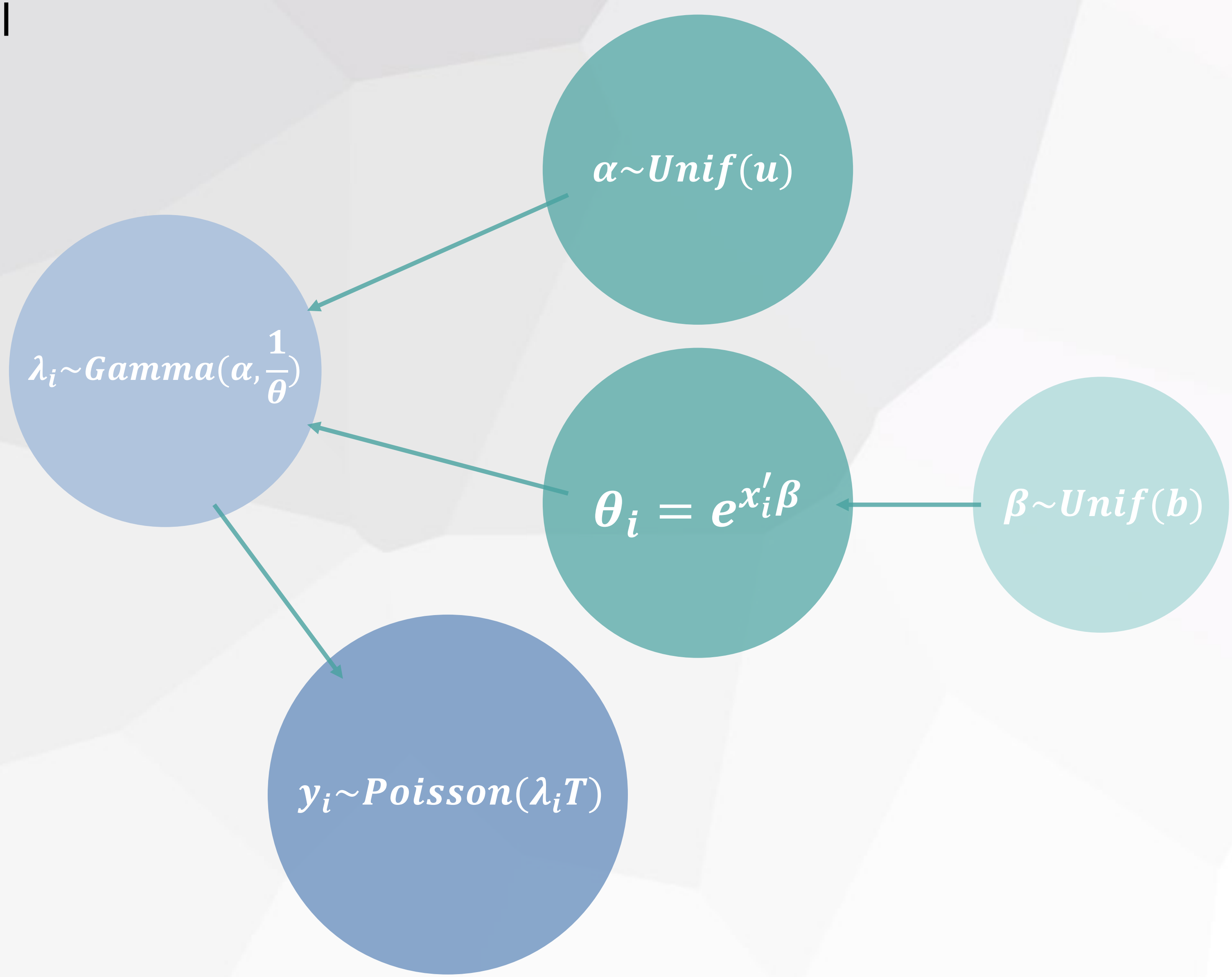
$$\lambda_i \sim Gamma(\alpha, \frac{1}{\theta})$$

$$y_i | \alpha, \theta \sim NegBin(\alpha, \frac{\theta}{\theta + 1})$$

$$y_i \sim Poisson(\lambda_i T)$$

Hierarchical Model

Basic model & Priors



Hierarchical Model

Conditional posterior distributions

For $\boldsymbol{\lambda_i}$

$$\lambda_i|\alpha,\theta_i \sim Gamma(y_i+\alpha, T+\theta_i{}^{-1})$$

Denote $e^{\beta_k}=\varphi_k$

$$\theta_i = e^{x_i'\beta} = \prod_{k=1}^{K}(e^{\beta_k})^{x_{ik}} = \prod_{k=1}^{K}\varphi_k^{x_{ik}}$$

Assume all covariates are dummy variables

$$\varphi_k|\alpha,\lambda_i \sim Inv-Gamma\left(\alpha\sum_{i=1}^{n}\boldsymbol{x_{ik}}-1, \sum_{i\in I_{(k)}}^{n}\left(\lambda_i\prod_{\substack{j=1,\\ j\neq k}}^{K}\boldsymbol{\varphi_j}^{-\boldsymbol{x_{ik}}}\right)\right)$$

$$I_{(k)}=\{i|x_{ik}=1\}$$

For $\boldsymbol{\alpha}$

$$p(\alpha|\lambda_i,\theta_i) \propto \prod_{i=1}^{n}\frac{(\lambda_i/\theta_i)^{\alpha}}{\Gamma(\alpha)\lambda_i}e^{-\lambda_i/\theta_i}$$

$$Poisson(\lambda_i/\theta_i)$$

PART 4

MCMC Simulation

Gibbs Sampler / Metropolis Hasting

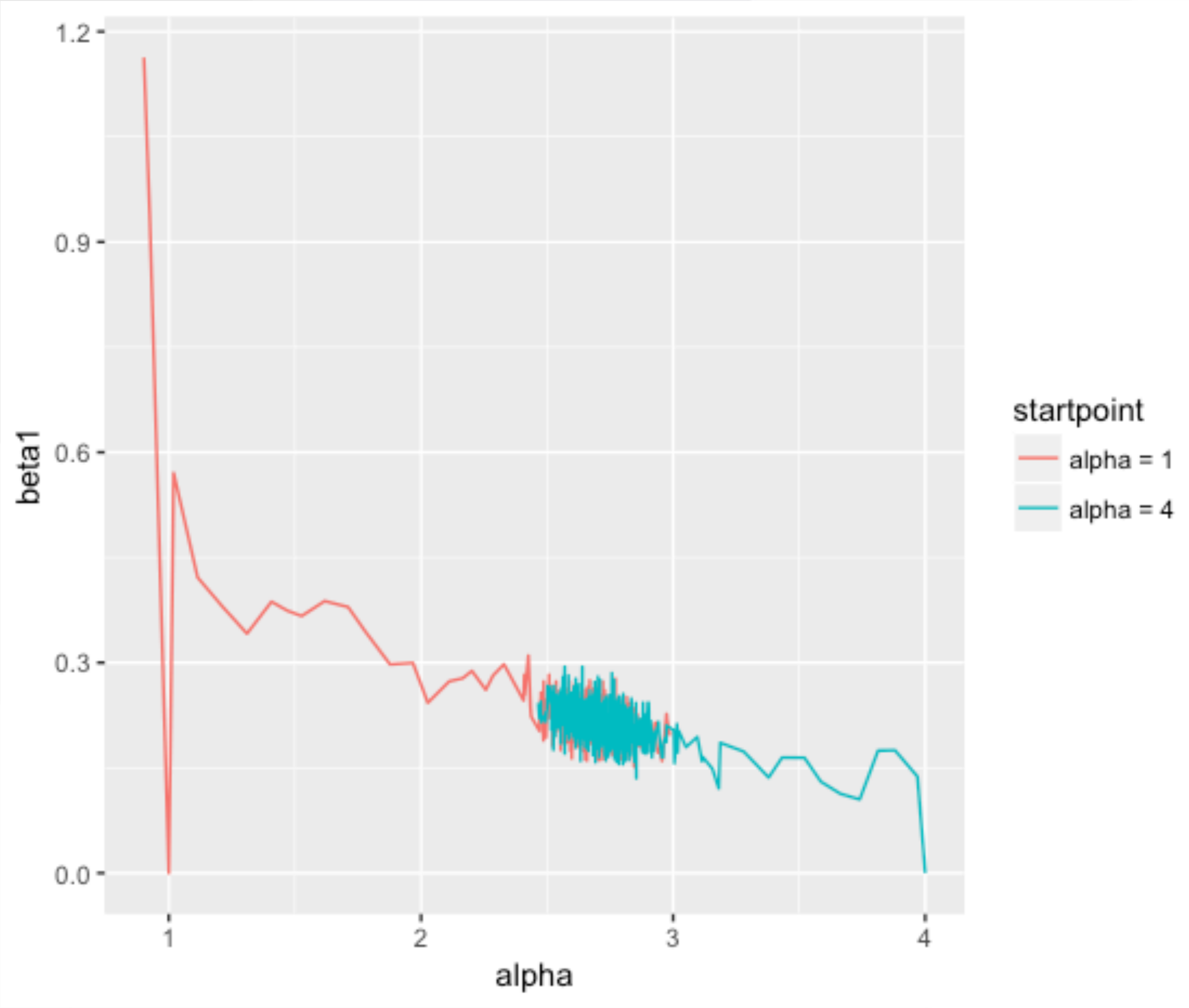
MCMC Simulation

Gibbs Sampler

- Transfer covariates into dummy variables

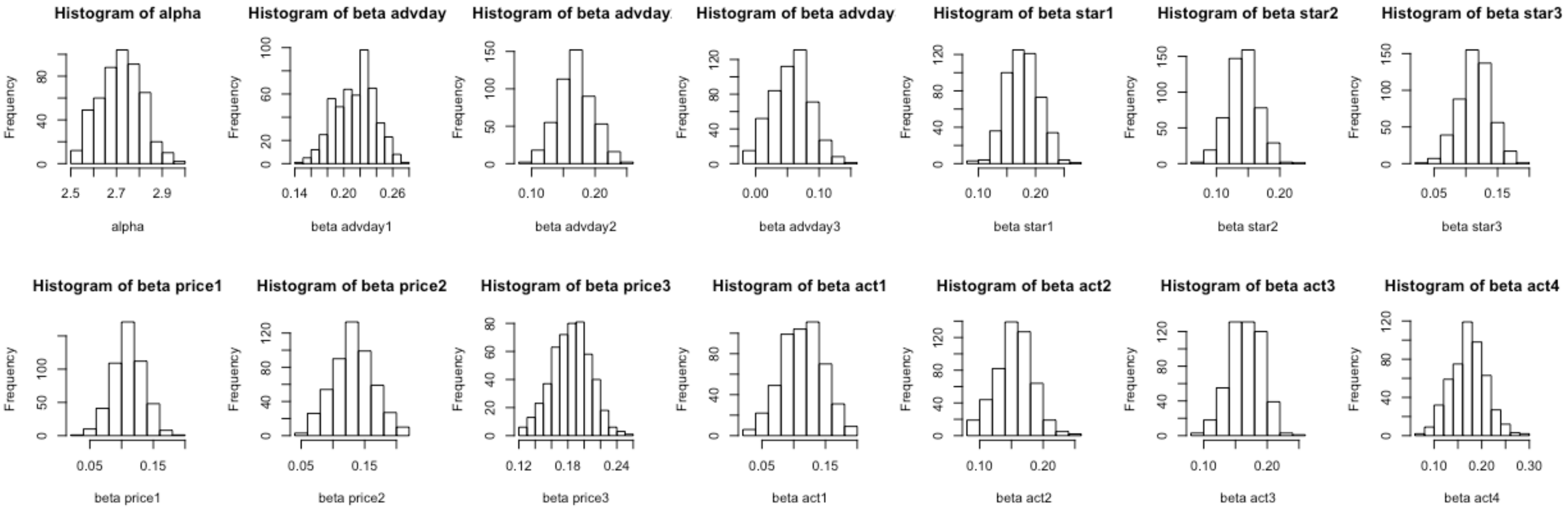
user_avgadvanceddate	<2	2~5	>=5	
user_avgstar	<7	7~9	>=9	
user_activation	<2000	2000~4000	4000~6000	>=6000
user_citynum	<6	6~20	>=20	
user_avgdealprice	<1000	1000~1500	>=1500	

- Start with different points
- Iterate 1000 times
- The sequences reach approximate converge
- The coefficients are difficult to interpret



MCMC Simulation

Gibbs Sampler

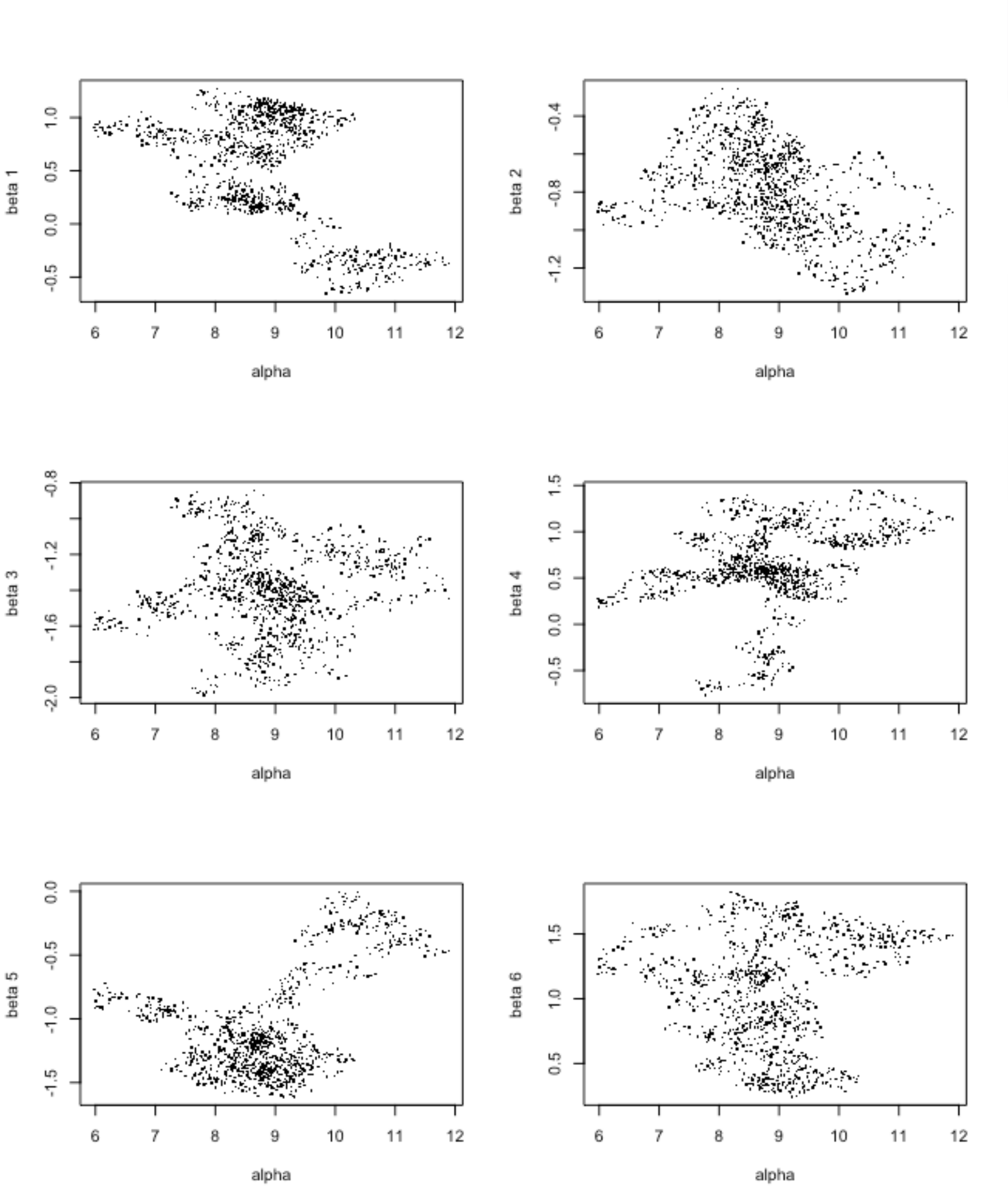


MCMC Simulation

Metropolis Hasting

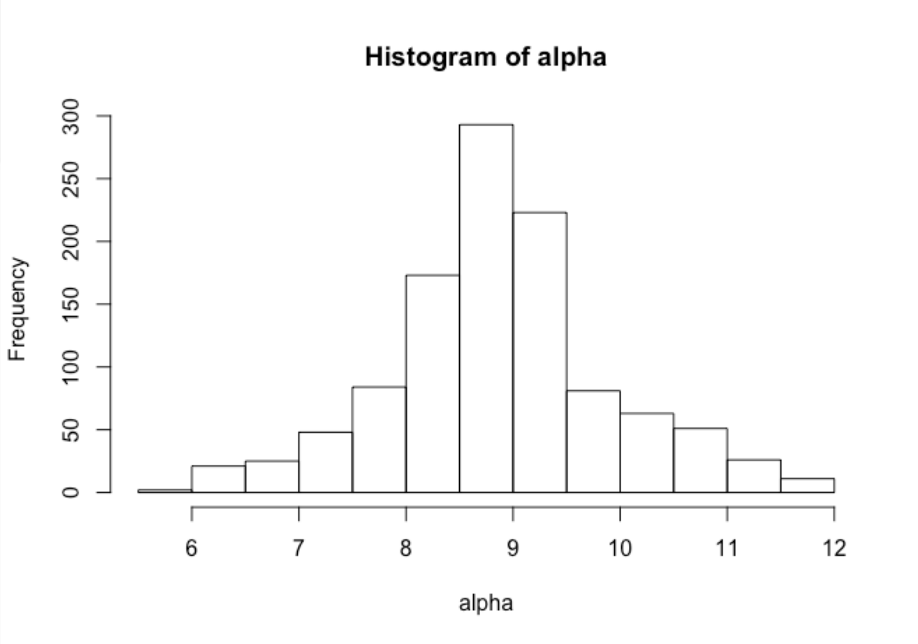
- Covariates could be numerical, better adaptation to individual differences
- Iterate 80000 times, 3102 pairs of parameters were accepted
- The sequences reach approximate converge
- Keep the parameters in last 1000 iterations
- The coefficients tells: **the consumers who travel to more cities, and place the orders with higher price more decisively are more likely to have higher purchase frequencies.**

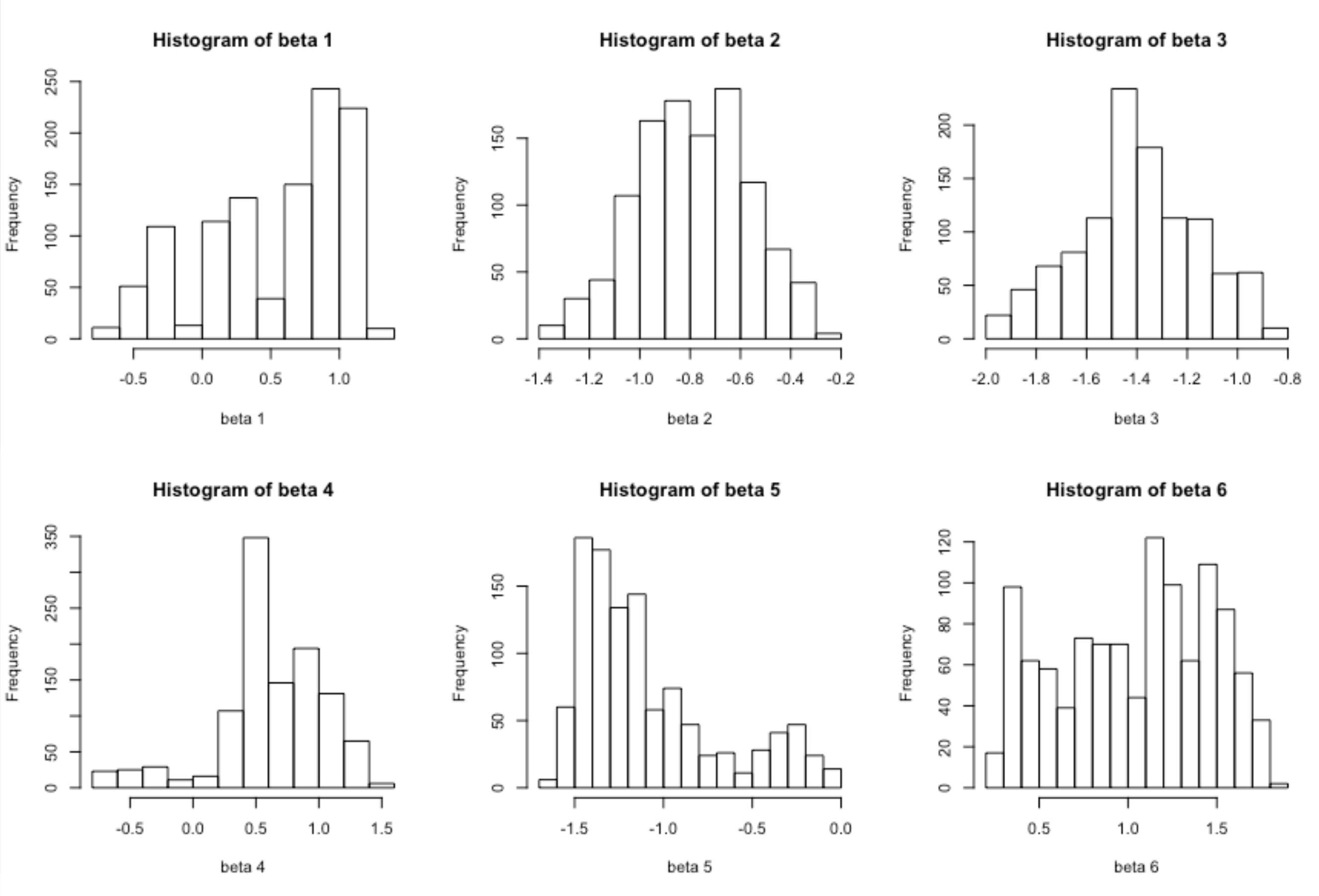
Variable	Min	Median	Mean	Max
alpha	5.983	8.823	8.872	11.889
beta_constant	-0.6547	0.7229	0.5340	1.2732
beta_avgadvanceddate	-1.3345	-0.7851	-0.7882	-0.2550
beta_avgstar	-1.9870	-1.4052	-1.3969	-0.8405
beta_avgdealprice	-0.7675	0.5963	0.6235	1.4495
beta_activation	-1.6122	-1.2154	-1.0889	-0.0043
beta_citynum	0.2418	1.1241	1.0391	1.8253



MCMC Simulation

Gibbs Sampler





THANKS!