Bayesian Analytics
What, How, and Application

LG Aimers

김재환
고려대학교 경영대학

Bayesian Analytics

- **Bayesian:** *Principle and Structure*
  - Example
  - Principles and Structure

- **Bayesian:** *Estimation Algorithm*
  - MCMC
  - Gibbs Sampler | Metropolis-Hastings Algorithm

- **Bayesian:** *Solving Real Problem*
  - Predicting heterogeneous consumer response

베이지안 • Bayesian

Conditional Probability

Bayes 정리

Sampling ?

사후확률

Markov Chain

주관적 ?

Joint Probability

사전확률

Marginal Probability

베이지안 네트워크

Algorithm

Bayesian Analytics

$$P(궁금한대상|Data)$$

궁금한 것 ?

변수

예 : **GPA**

3.75
4.12
4.30
2.41
3.01
:

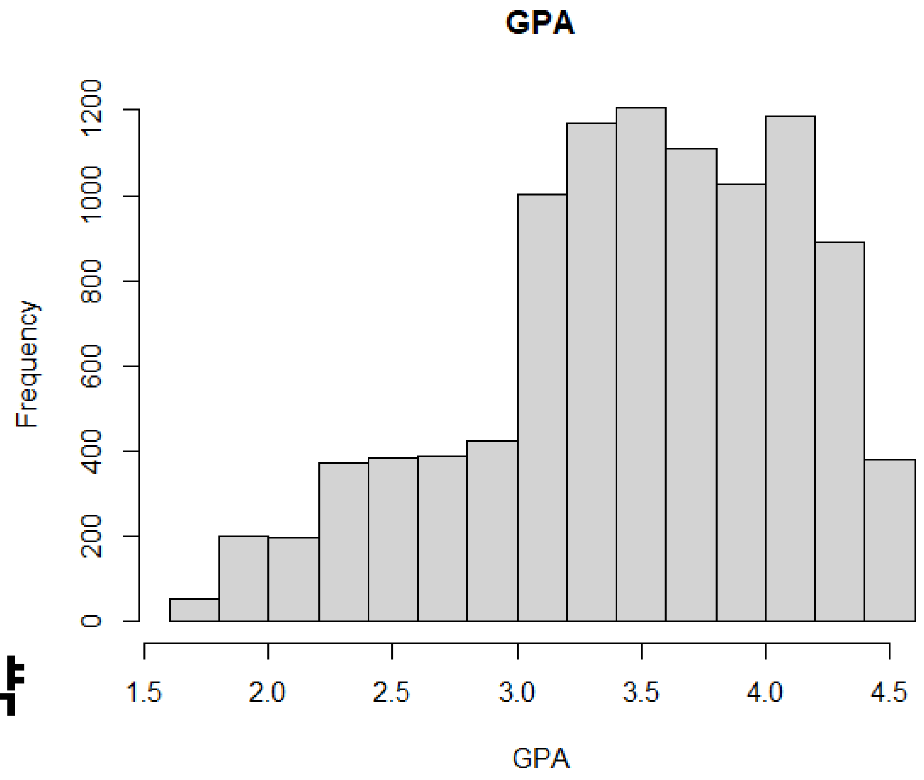
궁금한 것 ?

변수

예 : GPA

3.75
4.12
4.30
2.41
3.01
:

- 분포



- 분포 요약

평균 : 3.48
중위수: 3.54
표준편차: 0.65

궁금한 것 ?

변수

예 : GPA

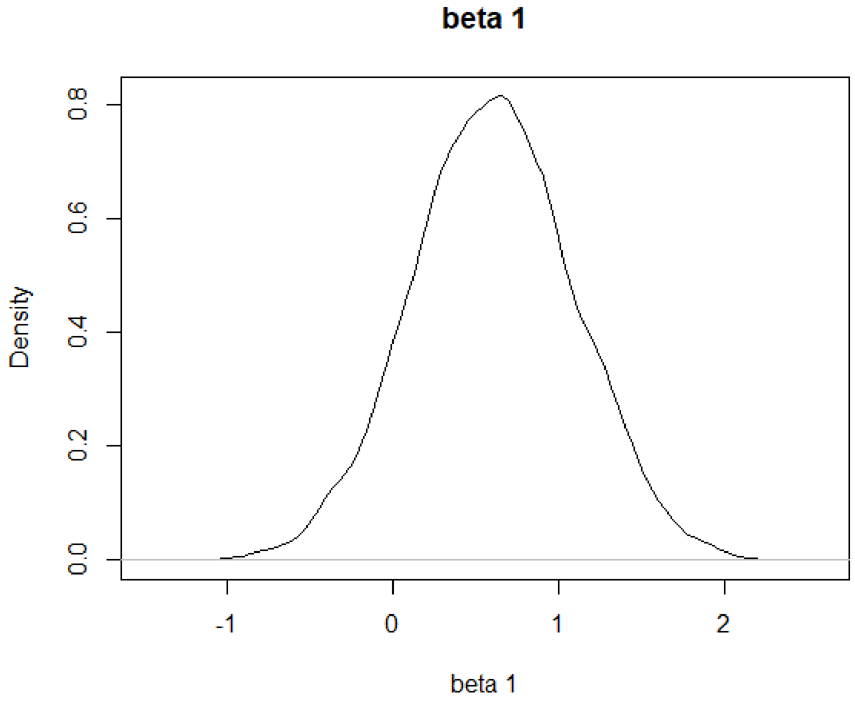
3.75
4.12
4.30
2.41
3.01
⋮

모수

예 : $\mathbf{Salary} = \beta \cdot \mathbf{GPA} + \varepsilon$

0.45
0.01
1.24
1.13
0.75
⋮

궁금한 것 ?



평균 : 0.60
중위수: 0.60
표준편차: 0.49
95%C.I : [-0.36, 1.56]

- **분포**
- **분포 요약**

모수

예 : $\text{Salary} = \beta \cdot \text{GPA} + \varepsilon$

0.45
0.01
1.24
1.13
0.75
:

궁금한 것 ?

변수

예 : GPA

3.75
4.12
4.30
2.41
3.01
:

- 분포

- 분포 요약

모수

예 : $\textbf{Salary} = \beta \cdot \textbf{GPA} + \varepsilon$

0.45
0.01
1.24
1.13
0.75
:

궁금한 것 ?

변수

예 : GPA

3.75
4.12
4.30
2.41
3.01
:

- 분포

- 분포 요약

모수

예 : $\mathbf{Salary} = \beta \cdot \mathbf{GPA} + \varepsilon$

0.45
0.01
1.24
1.13
0.75
:

Example

Question

X의 y 에 대한 예측력

Data: (y, x)

obs	y	x
1	16.3	0
2	11.5	1
3	8.3	1
4	16.2	0
5	13.6	0
:	:	:

Model

$$y = \beta_0 + \beta_1 X + \varepsilon$$

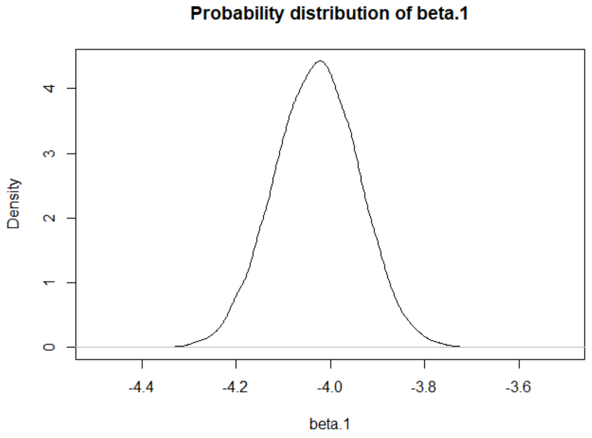
$$\varepsilon \sim N(0, \sigma^2)$$

Results

$\beta_0 = ?\ \ \beta_1 = ?\ \sigma^2 = ?$

> Answers

	beta.0	beta.1	s.sq
mean	15.04	-4.03	2.11
2.5%	14.91	-4.22	1.91
97.5%	15.17	-3.84	2.32



Example

Question

Data: (y, x)

Model

Results

Example

Question

$\beta_0 =? \ \ \beta_1 =? \ \ \sigma^2 =?$

Results

$P(\beta_0|Data)$
$P(\beta_1|Data)$
$P(\sigma^2|Data)$

Data: (y, x)

Model

Example

Question

$\beta_0 =?\ \ \beta_1 =?\ \sigma^2 =?$

Model

$$y = \beta_0 + \beta_1 X + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

Data: (y, x)

obs	y	x
1	16.3	0
2	11.5	1
3	8.3	1
4	16.2	0
5	13.6	0
:	:	:

Results

$P(\beta_0|Data)$
$P(\beta_1|Data)$
$P(\sigma^2|Data)$

Example

Question

$\beta_0 =? \;\; \beta_1 =? \;\; \sigma^2 =?$

$\boldsymbol{parameters}$

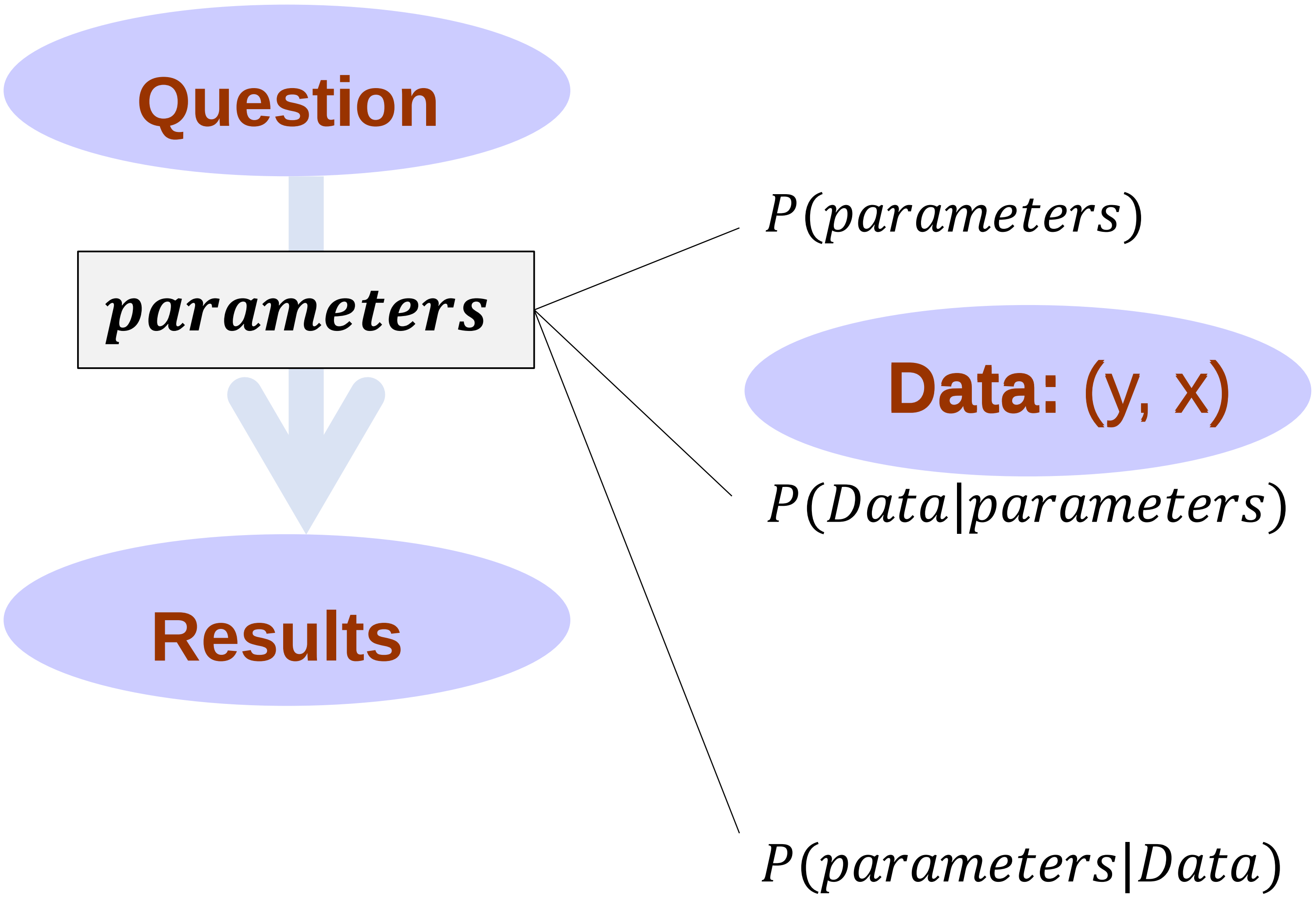
Results

$P(\beta_0|Data)$

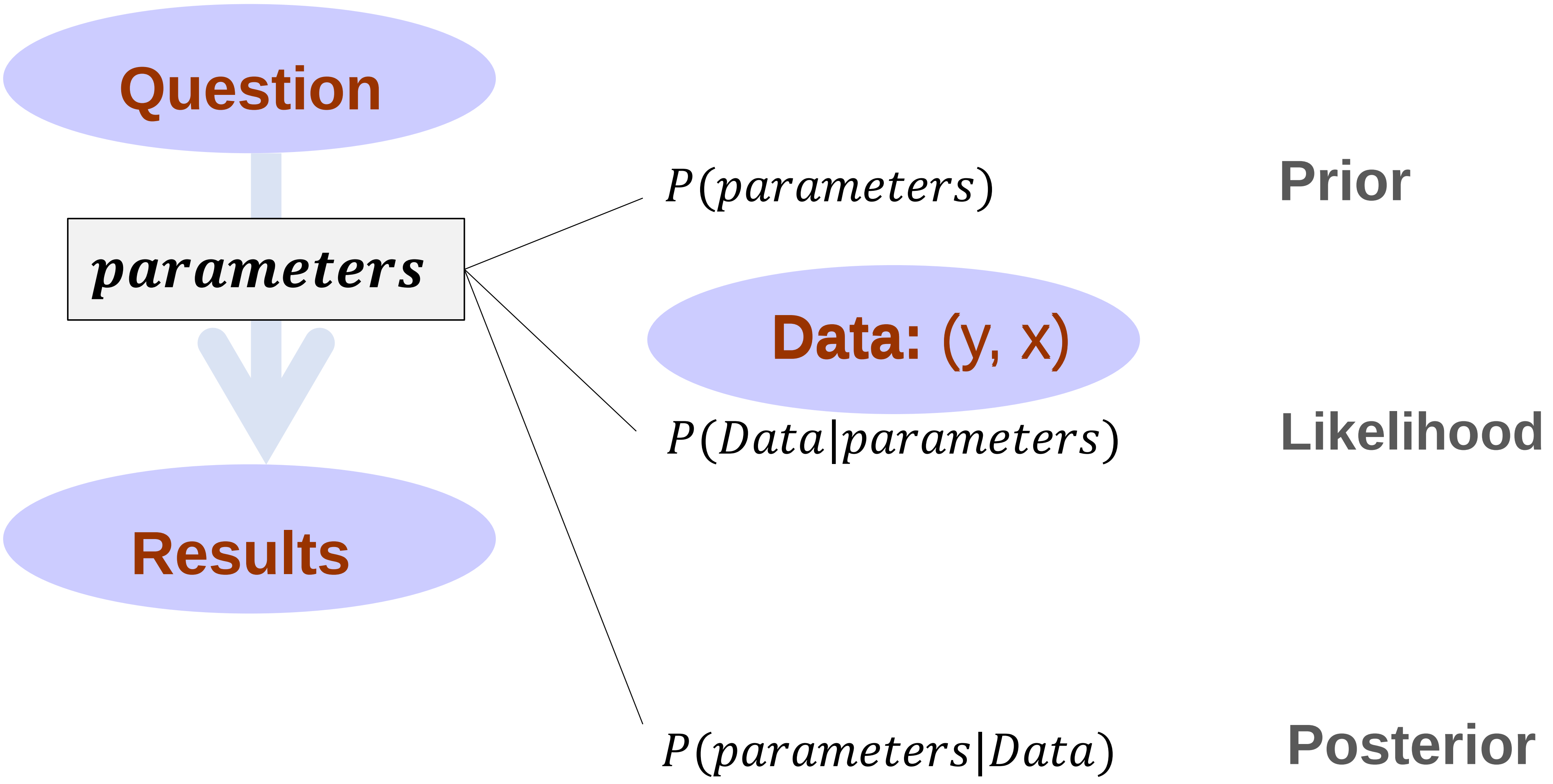
$P(\beta_1|Data)$

$P(\sigma^2|Data)$

Example


Question
parameters
Results
$P(parameters)$
Data: (y, x)
$P(Data|parameters)$
$P(parameters|Data)$

Example


Question
parameters
Results
$P(parameters)$
Prior
Data: (y, x)
$P(Data|parameters)$
Likelihood
$P(parameters|Data)$
Posterior

Example

$parameters$

$P(parameters)$ — Prior

Data: (y, x)

$P(Data|parameters)$ — Likelihood

$P(parameters|Data)$ — Posterior

Example

$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$\boxed{\boldsymbol{parameters}}$

$P(\beta_0), P(\beta_1), P(\sigma^2)$

$P(parameters)$

Prior

Data: (y, x)

$P(y|\beta_0, \beta_1, \sigma^2, X)$

$P(Data|parameters)$

Likelihood

$P(\beta_0|y, X), P(\beta_1|y, X),$
$P(\sigma^2|y, X)$

$P(parameters|Data)$

Posterior

Example

$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}\,, \boldsymbol{\beta} = (\beta_0, \beta_1)'$

$\boxed{\boldsymbol{parameters}}$

$P(\beta_0), P(\beta_1), P(\sigma^2)$	$P(parameters)$	**Prior**
	Data: (y, x)	
$P(y\|\beta_0, \beta_1, \sigma^2, X)$	$P(Data\|parameters)$	**Likelihood**
$P(\beta_0\|y, X), P(\beta_1\|y, X),$ $P(\sigma^2\|y, X)$	$P(parameters\|Data)$	**Posterior**

Example

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \boldsymbol{\beta} = (\beta_0, \beta_1)'$ $\boxed{\boldsymbol{parameters}}$

$P(\boldsymbol{\beta}), P(\sigma^2)$ $P(parameters)$ **Prior**

Data: (y, x)

$P(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X})$ $P(Data|parameters)$ **Likelihood**

$P(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}), P(\sigma^2|\mathbf{y}, \mathbf{X})$ $P(parameters|Data)$ **Posterior**

Example

$$\boldsymbol{parameters}$$

$$\boldsymbol{\beta} \sim N\left(\boldsymbol{\beta}_{prior}, \Sigma_{prior}\right)$$
$$\sigma^2 \sim IG(\mathrm{a}_{prior}, \mathrm{b}_{prior})$$

$P(parameters)$

Prior

Data: (y, x)

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

$P(Data|parameters)$

Likelihood

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_{post}, \Sigma_{post})$$
$$\sigma^2 \sim IG(\mathrm{a}_{post}, \mathrm{b}_{post})$$

$P(parameters|Data)$

Posterior

원리, 구성요소

- **Prior:** $\boldsymbol{\beta} \sim N\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_0\right)$

- **Likelihood:** $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

- **Posterior:** $\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}, \sigma^2 \sim N(\boldsymbol{\beta}_1, \boldsymbol{\Sigma}_1) \propto$ *Prior x Likelihood*

원리, 구성요소

- **Prior:** $\boldsymbol{\beta} \sim N\left(\tilde{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_0\right)$

$$P(\boldsymbol{\beta}) = \left(2\pi^K|\Sigma_0|\right)^{-1/2} exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}\right)'\Sigma_0^{-1}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}\right)\right\}$$

- **Likelihood:** $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$

$$P(\mathbf{y}|\mathbf{X},\boldsymbol{\beta}) \equiv L(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \left(2\pi\sigma^2\right)^{-n/2} exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

- **Posterior:** $\boldsymbol{\beta}|\mathbf{y},\mathbf{X},\sigma^2 \sim N(\boldsymbol{\beta}_1, \boldsymbol{\Sigma}_1) \propto$ *Prior x Likelihood*

$$P(\boldsymbol{\beta}|\mathbf{y},\mathbf{X},\sigma^2) = \left(2\pi^K|\Sigma_1|\right)^{-1/2} exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_1)'\Sigma_1^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_1)\right\}$$

$$\boxed{\begin{aligned}\boldsymbol{\beta}_1 &= \boldsymbol{\Sigma}_1\left(\boldsymbol{\Sigma}_0^{-1}\tilde{\boldsymbol{\beta}} + \sigma^{-2}\mathbf{X}'\mathbf{y}\right)\\ \boldsymbol{\Sigma}_1 &= \left(\boldsymbol{\Sigma}_0^{-1} + \sigma^{-2}\mathbf{X}'\mathbf{X}\right)^{-1}\end{aligned}}$$

$cf.$

$$\boldsymbol{\beta}_{MLE} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$Var(\boldsymbol{\beta}_{MLE}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Bayesian ?

Bayesian ?

$$Posterior\ Probability = \frac{Prior\ Probability \times Likelihood}{Marginalized\ Probability}$$

$$\Leftrightarrow \quad P(\beta|Data) = \frac{P(\beta) \cdot P(Data|\beta)}{P(Data)}$$

Bayes 정리

$$= \frac{P(\beta) \cdot P(Data|\beta)}{\int P(\beta) \cdot P(Data|\beta)\, d\beta}$$

적분

Bayesian ?

$$Posterior\ Probability = \frac{Prior\ Probability \times Likelihood}{Marginalzed\ Probability}$$

$$\Leftrightarrow \quad P(\beta|Data) = \frac{P(\beta) \cdot P(Data|\beta)}{P(Data)}$$

Bayes 정리

$$= \frac{P(\beta) \cdot P(Data|\beta)}{\int P(\beta) \cdot P(Data|\beta)\, d\beta}$$

적분

$$\Leftrightarrow \quad \underbrace{P(\beta|Data)}_{\text{Posterior}} \propto \underbrace{P(\beta)}_{\text{Prior}} \times \underbrace{P(Data|\beta)}_{\text{Likelihood}}$$

Summary

$$P(\text{궁금한대상}|Data)$$

$$P(\boldsymbol{parameter}|Data)$$

Summary

- **What is Bayesian Analytics?**

  Question is about "parameter" and answer is about "parameter".

  For unknown parameter, assessing uncertainty through the **probability** distribution of parameter **after** data were seen.

$$P(\text{궁금한대상}|Data)$$

- **Data Scientist's job (Components)**
  - Model : Parameter + Data
  - Computation: Prior $\times$ Likelihood $\propto$ Posterior
  - **Output**: Posterior distribution of parameters

Bayesian Analytics
What, How, and Application

LG Aimers

김재환
고려대학교 경영대학

Bayesian Analytics

- **Bayesian:** *Principle and Structure*
  - Example
  - Principles and Structure
- **Bayesian:** *Estimation Algorithm*
  - MCMC
  - Gibbs Sampler | Metropolis-Hastings Algorithm
- **Bayesian:** *Solving Real Problem*
  - Predicting heterogeneous consumer response

Posterior Distribution of Parameters: How ?

Posterior

∝ Likelihood x Prior

$P(Data|\beta)$ $P(\beta)$

Case 1

Easy

Case 2

Not easy

Posterior Distribution of Parameters: How ?

Posterior

∝ Likelihood x Prior

$P(Data|\beta)$ $\quad$ $P(\beta)$

<u>Case 1</u>

Easy

$P(Data|\beta)$ $\quad$ $P(\beta)$ $\quad$ $P(\beta|Data)$

$$Poisson(\beta) \times Gamma(a,b) \propto Gamma(a_1,b_1)$$

Posterior Distribution of Parameters: How ?

Posterior

∝ Likelihood x Prior

$P(Data|\beta) \quad P(\beta)$

Case 2

Not easy

$P(Data|\beta) \qquad P(\beta) \qquad P(\beta|Data)$

$logistic(\beta) \times Normal(\tilde{\beta}, \tilde{\sigma}^2) \propto \; ?$

Posterior Distribution of Parameters: How ?

Posterior

$\propto$ Likelihood x Prior

$P(Data|\beta)$ $P(\beta)$

Case 1

Easy

Solution

Simulation

Markov Chain Monte Carlo (MCMC)

Case 2

Not easy

Estimation Algorithm

- **Monte Carlo?**

Drawing samples 'many times' from a probability distribution

예: $\theta^{(i)} \sim f(\theta_{fix}) \quad i = 1, \dots R(big\ number)$

- **Markov Chain + Monte Carlo ?**

Drawing samples 'many times' from a probability distribution that is ***dependent on the last sample***

예: $\theta^{(i)} \sim f\left(g(\theta^{(i-1)})\right) \quad i = 1, \dots R(big\ number)$

Algorithm 1: Gibbs Sampler

- **Goal: Estimate Joint Posterior Distribution of Parameters**

= *Reaching the joint probability distribution via sampling from conditional distribution*

<u>Example</u>

Model: $y = X\theta_1 + \varepsilon \qquad \varepsilon \sim N(0, \theta_2)$

Output: $P(\theta_1, \theta_2 | y, X)$

When: It is "not easy" to draw from the joint posterior distribution, $P(\theta_1, \theta_2 | y, X)$, but "easy" to draw from the conditional distributions, $P(\theta_1 | \theta_2, y, X)$, $P(\theta_2 | \theta_1, y, X)$.

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)} \\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)} \\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)} \\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)} \\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

Step 0. Pick an initial $\theta_2^{(0)}$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

Step 0. Pick an initial $\theta_2^{(0)}$

Step 1-(1). One draw from $P\left(\theta_1 | \theta_2^{(0)} y, X\right)$ $\rightarrow$ Keep it as $\theta_1^{(1)}$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

Step 0. Pick an initial $\theta_2^{(0)}$

Step 1-(1). One draw from $P\left(\theta_1 | \theta_2^{(0)} y, X\right)$ $\rightarrow$ Keep it as $\theta_1^{(1)}$

Step 1-(2). One draw from $P\left(\theta_2 | \theta_1^{(1)} y, X\right)$ $\rightarrow$ Keep it as $\theta_2^{(1)}$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

Step 0. Pick an initial $\theta_2^{(0)}$

Step 1-(1). One draw from $P\left(\theta_1 | \theta_2^{(0)} y, X\right)$ $\rightarrow$ Keep it as $\theta_1^{(1)}$

Step 1-(2). One draw from $P\left(\theta_2 | \theta_1^{(1)} y, X\right)$ $\rightarrow$ Keep it as $\theta_2^{(1)}$

Step 2-(1). One draw from $P\left(\theta_1 | \theta_2^{(1)} y, X\right)$ $\rightarrow$ Keep it as $\theta_1^{(2)}$

…

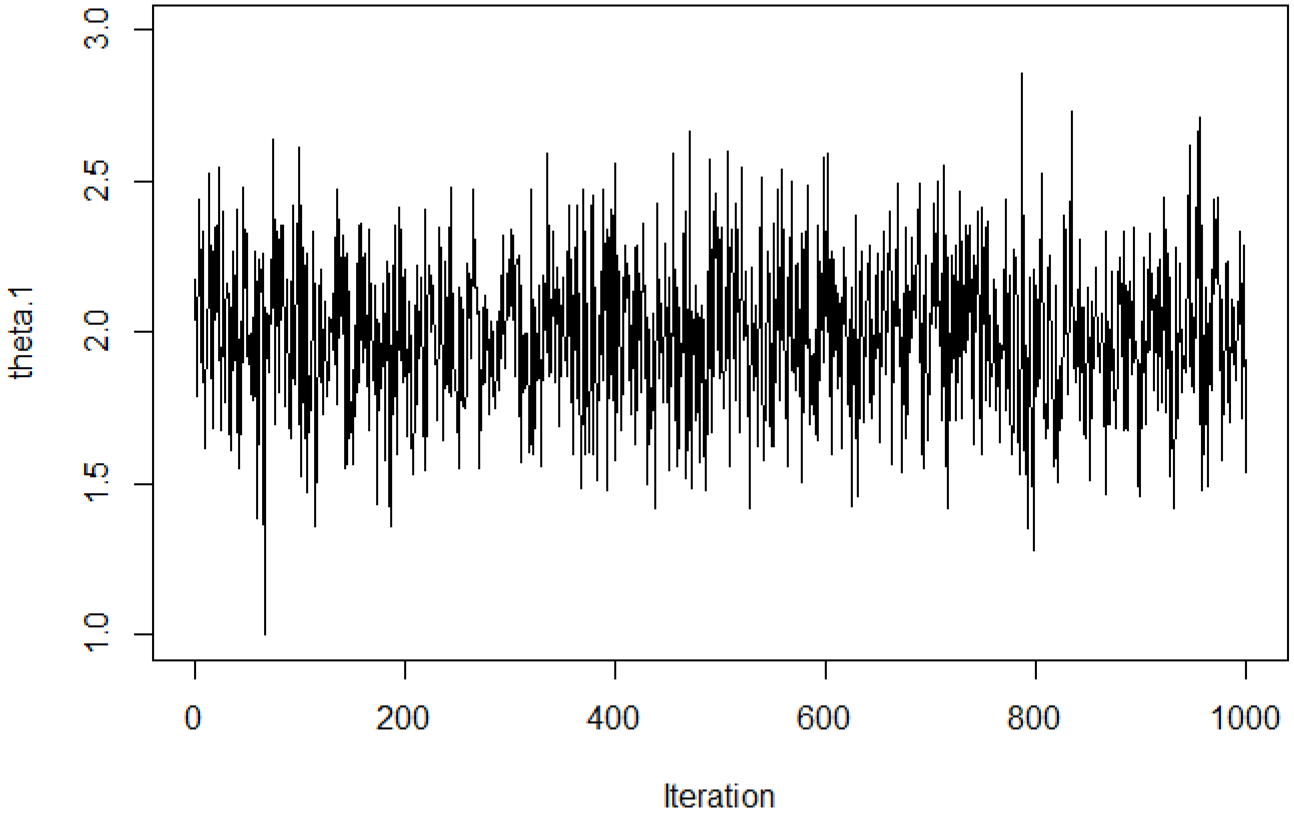
Step R-(2). One draw from $P\left(\theta_1 | \theta_1^{(R)} y, X\right)$ $\rightarrow$ Keep it as $\theta_2^{(R)}$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

	θ_1	θ_2
1	$\theta_1^{(1)}$	$\theta_2^{(1)}$
2	$\theta_1^{(2)}$	$\theta_2^{(2)}$
3	$\theta_1^{(3)}$	$\theta_2^{(3)}$
:		
:		
R	$\theta_1^{(R)}$	$\theta_2^{(R)}$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

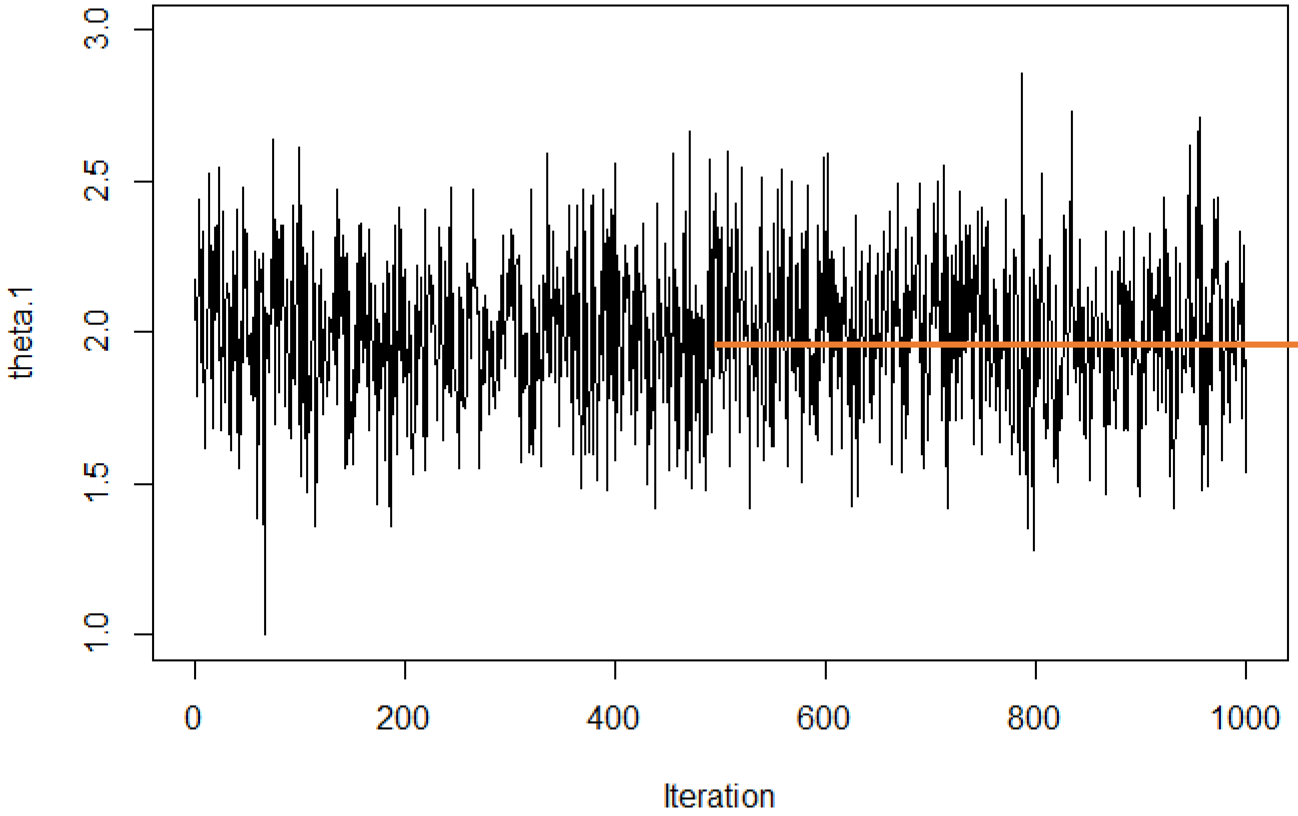
Trace plot $(\theta_1 | y, X)$



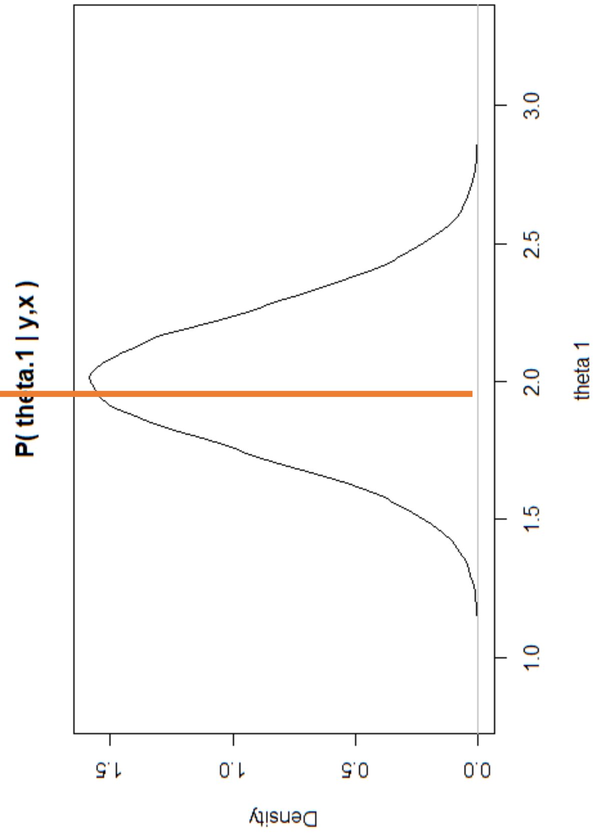
	θ_1	θ_2
1	$\theta_1^{(1)}$	$\theta_2^{(1)}$
2	$\theta_1^{(2)}$	$\theta_2^{(2)}$
3	$\theta_1^{(3)}$	$\theta_2^{(3)}$
:		
R	$\theta_1^{(R)}$	$\theta_2^{(R)}$

Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

Trace plot $(\theta_1 | y, X)$

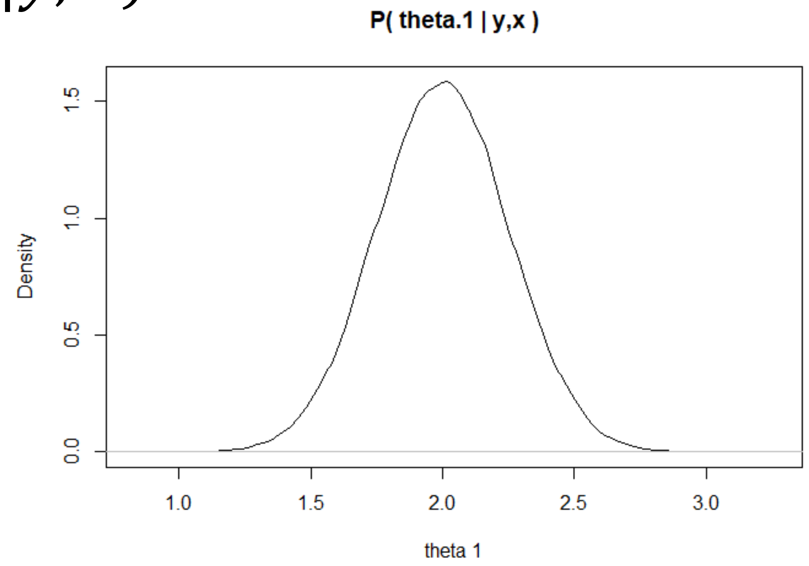


$P(\theta_1 | y, X)$



Goal: Draw $\begin{pmatrix}\theta_1^{(1)}\\ \theta_2^{(1)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(2)}\\ \theta_2^{(2)}\end{pmatrix}, \begin{pmatrix}\theta_1^{(3)}\\ \theta_2^{(3)}\end{pmatrix}, \cdots, \begin{pmatrix}\theta_1^{(R)}\\ \theta_2^{(R)}\end{pmatrix}$ from $P(\theta_1, \theta_2 | y, X)$

$P(\theta_1 | y, X)$



	θ_1	θ_2
1	$\theta_1^{(1)}$	$\theta_2^{(1)}$
2	$\theta_1^{(2)}$	$\theta_2^{(2)}$
3	$\theta_1^{(3)}$	$\theta_2^{(3)}$
:		
R	$\theta_1^{(R)}$	$\theta_2^{(R)}$

$\text{Mean}(\theta_1 | y, X) =$

$\text{Std}(\theta_1 | y, X) =$

$95\%\ \text{C.I}(\theta_1 | y, X) = (\ 1.51\ ,\ \ 2.49\ )$

Algorithm 2: Metropolis-Hastings

- **Goal: Estimate Joint Posterior Distribution of Parameters**

  When: It is “not easy” to draw from the posterior distribution.

Algorithm 2: Metropolis-Hastings

- **Goal: Estimate Joint Posterior Distribution of Parameters**

When: It is "not easy" to draw from the posterior distribution.

<u>Example</u>

Model: $U = X\theta + \varepsilon \qquad \varepsilon \sim logistic(0,1)$

$$y = 1 \;\; if \; U > 0$$

$$y = 0 \;\; otherwise$$

Likelihood: $\mathrm{P}(y|\theta, X) = \left[\frac{exp(X\theta)}{1+exp(X\theta)}\right]^{y} \left[\frac{1}{1+exp(X\theta)}\right]^{1-y}$

Prior: $\theta \sim N(\mu_\theta, \sigma_\theta)$

Output: $P(\theta|y, X)$ = ?

Goal: Draw $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \cdots, \theta^{(R)}$ from $P(\theta|y, X)$

- Idea:

(i) "A candidate draw" is sampled from an auxiliary density from which we can easily take samples.

(ii) Either "Accept or Reject" the candidate draw <u>based on some rules</u> that are <u>related to the posterior distributions</u>.

(iii) The accepted draws are used to summarize the posterior distribution.

Goal: Draw $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \cdots, \theta^{(R)}$ from $P(\theta|y, X)$

Step 0. Pick an initial $\theta^{(0)}$

Step 1-(1). One draw, δ, from $N(0, \tau^2)$ $\rightarrow$ Let $\theta_{candidate} = \theta^{(0)} + \delta$

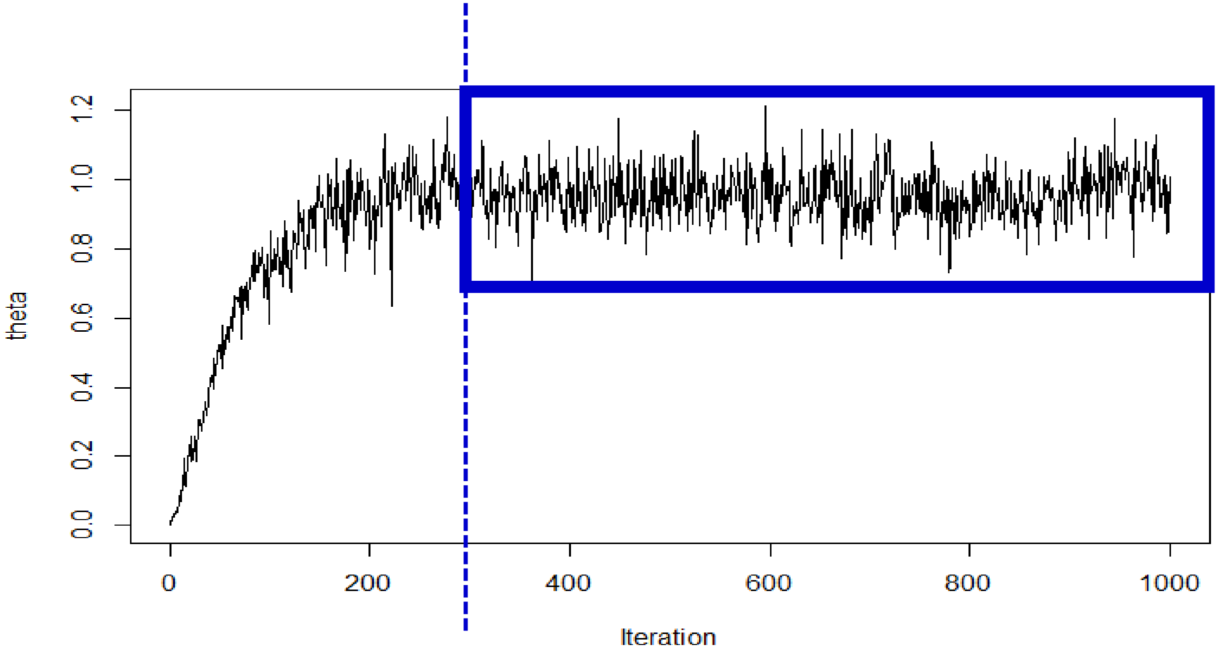
Step 1-(2). Evaluate $a = \dfrac{Likelihood(\theta_{candidate}) \times Prior(\theta_{candidate})}{Likelihood(\theta^{(0)}) \times Prior(\theta^{(0)})} \dfrac{\times N(\theta^{(0)} | \theta_{candidate})}{\times N(\theta_{candidate} | \theta^{(0)})}$

Step 1-(3). One draw, u, from $Uniform(0,1)$ $\rightarrow$ Accept $\theta_{candidate}$ as $\theta^{(1)}$ $if\ a > u$
Keep $\theta^{(0)}$ as $\theta^{(1)}$ $otherwise$

:

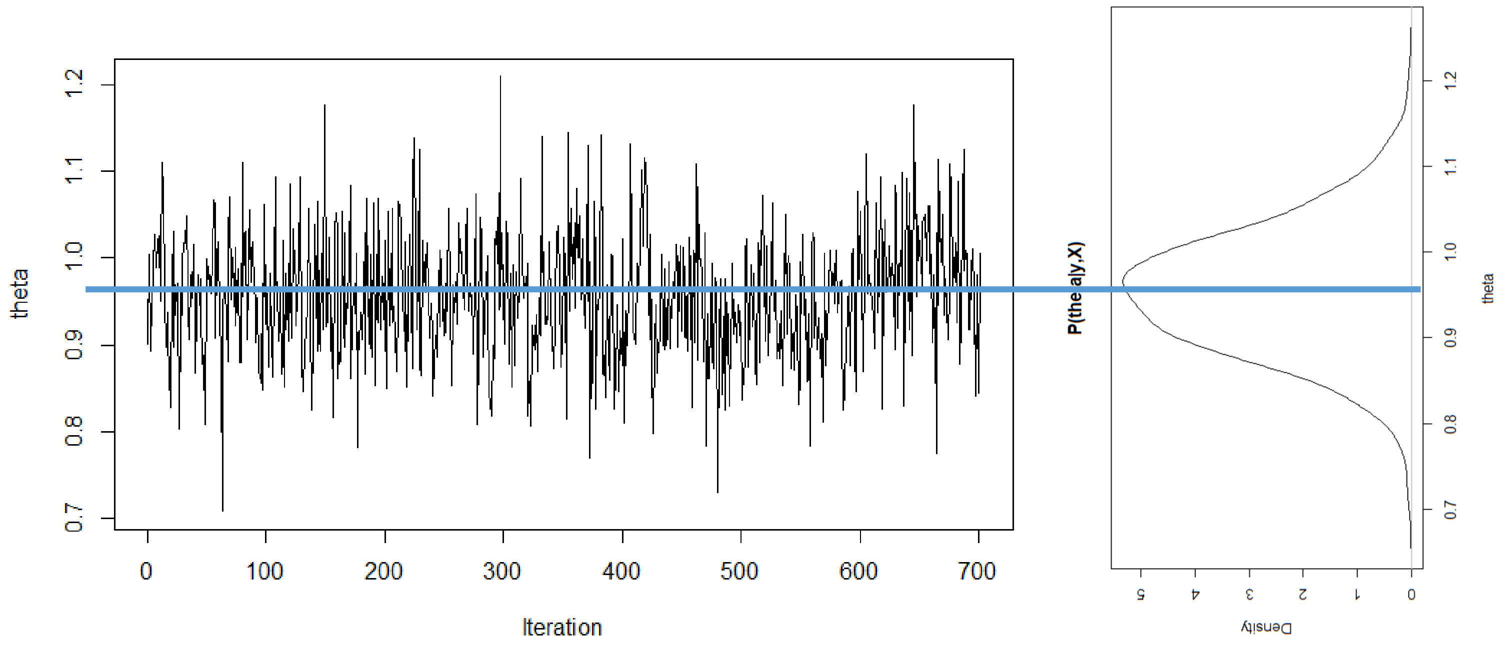
Repeat until last iteration

Goal: Draw $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \cdots, \theta^{(R)}$ from $P(\theta|y, X)$



	θ
1	$\theta^{(1)}$
2	$\theta^{(2)}$
3	$\theta^{(3)}$
:	
:	
R	$\theta^{(R)}$

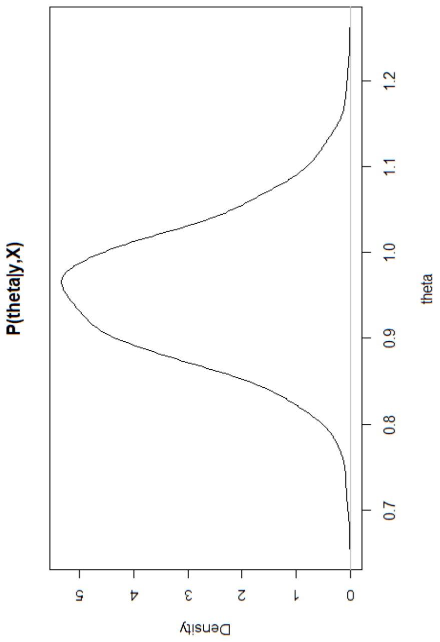
Goal: Draw $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \cdots, \theta^{(R)}$ from $P(\theta|y, X)$



	θ
1	$\theta^{(1)}$
2	$\theta^{(2)}$
3	$\theta^{(3)}$
:	
:	
R	$\theta^{(R)}$

Goal: Draw $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \cdots, \theta^{(R)}$ from $P(\theta|y, X)$

$$\text{Mean}(\theta|y,X) = 0.96$$
$$\text{Std}(\theta|y,X) = 0.07$$
$$95\%\ \text{C.I}(\theta|y,X) = (\,0.82\,,\ 1.11\,)$$



	θ
1	$\theta^{(1)}$
2	$\theta^{(2)}$
3	$\theta^{(3)}$
:	
:	
R	$\theta^{(R)}$

Summary

- **How to estimate posterior distribution of parameters?**

- **MCMC**
  - Recursive sampling from the conditional distributions over the long iterations will give the joint posterior distribution of parameters.
  - **Gibbs Sampling**
    - When we know the conditional distribution:$\rightarrow$ all draws are ‘accepted’
  - **Metropolis Hastings** (Random Walk MH)
    - When we don’t know the conditional distribution: $\rightarrow$ some draws are ‘accepted’
- **Data Scientist’s job:**
  - **Input :** Data, Prior, Likelihood, Starting values
  - **Output**: Trace plot, Distribution, Summary statistics

Summary

- **Data Scientist's job (cont'd)**
  - Check list for Metropolis algorithm:
    1. Convergence
    2. Multiple starting points
    3. Burn-in
    4. Autocorrelation – Acceptance rate, ACF

Step 0. Pick an initial $\theta^{(0)}$

Step 1-(1). One draw, δ, from $N(0, \tau^2)$ $\rightarrow$ Let $\theta_{candidate} = \theta^{(0)} + \delta$

Step 1-(2). Evaluate $a = \dfrac{Likelihood(\theta_{candidate}) \times Prior(\theta_{candidate})}{Likelihood(\theta^{(0)}) \times Prior(\theta^{(0)})} \dfrac{\times N(\theta^{(0)} | \theta_{candidate})}{\times N(\theta_{candidate} | \theta^{(0)})}$

Step 1-(3). One draw, u, from $Uniform(0,1) \rightarrow$ Accept $\theta_{candidate}$ as $\theta^{(1)}$ $if\ a > u$
Keep $\theta^{(0)}$ as $\theta^{(1)}$ $otherwise$

Summary

- **Data Scientist's job (cont'd)**
  - Check list for Metropolis algorithm:
    1. Convergence
    2. Multiple starting points
    3. Burn-in
    4. Autocorrelation – Acceptance rate, ACF
       - 30% acceptance rate
       - Every __th draws are kept in the final output

Bayesian Analytics
What, How, and Application

LG Aimers

김재환
고려대학교 경영대학

Bayesian Analytics

- **Bayesian:** *Principle and Structure*
  - Example
  - Principles and Structure

- **Bayesian:** *Estimation Algorithm*
  - MCMC
  - Gibbs Sampler | Metropolis-Hastings Algorithm

- **Bayesian:** *Solving Real Problem*
  - Predicting heterogeneous consumer response

Application

Hierarchical Bayesian Analysis for Heterogeneous Consumer Behavior

- **Prediction :** 수요예측

- **Heterogeneity:** 개인 고객별로 parameter가 다른 상태

- **Conjoint Method:** 현재 시장내 제품에는 없는, 새로운 feature 도입되었을 시의 수요에 대한 예측

Application

Hierarchical Bayesian Analysis for Heterogeneous Consumer Behavior

00 기업

Want : 현재 사업 지역 → 다른 지역으로 확장
As-is : 낯선 경쟁사의 서비스를 좋아할 리는 없을 것. 승산은 없는가?
Prediction/Decision: 어느 고객에게 어떤 식으로 제안하면 그래도 승산 있을까?

$$r \times m - c > 0$$

Application

Hierarchical Bayesian Analysis for Heterogeneous Consumer Behavior

Decision Problem [D]

- Whom to target
- What to offer

Information Need [I]

- Who is likely to respond to our service offering ?

Data Scientist's job

- Offering answers to [D] and [I]

via providing posterior probability of parameters

Application

Hierarchical Bayesian Analysis for Heterogeneous Consumer Behavior

비교

Likelihood : $y = x\beta + \varepsilon$

Prior $\beta \sim N(\beta_{prior}, \sigma^2_{prior})$

$y_h = x_h\beta_h + \varepsilon_h$ **: Likelihood**

$\beta_h \sim N(\bar{\beta}, \mathrm{v})$ **Heterogeneity**

Application

Hierarchical Bayesian Analysis for Heterogeneous Consumer Behavior

- Data
- Model - Output and Input
- Estimation
- Output processing
- Managerial Decision

Data

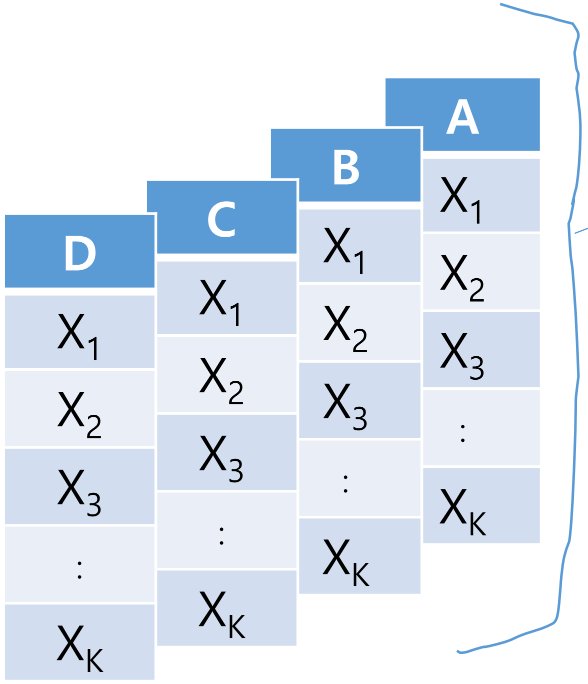
Transaction

Time	Buy	Customer
오전 11:10:14	C	ID 00001
오후 9:04:35	C	ID 00001
오후 8:15:43	A	ID 00001
오전 10:41:49	B	ID 13247
오후 1:07:35	C	ID 13247
오후 2:22:59	B	ID 13247
오후 2:04:27	B	ID 13247
오후 4:57:06	C	ID 00586
오후 3:10:53	C	ID 00586
:	:	:

$\{ y \}$

Data

Service offering



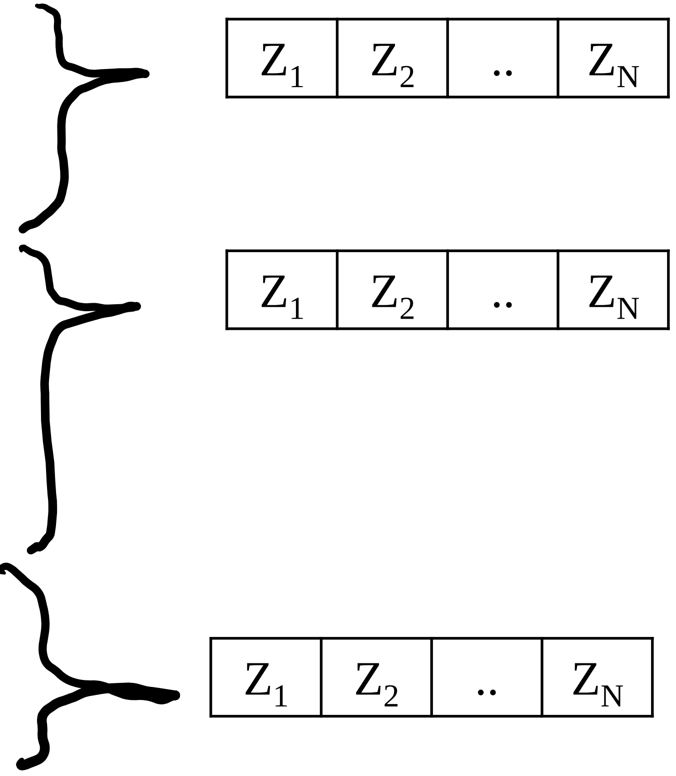
{ X }

Transaction

Time	Buy	Customer
오전 11:10:14	C	ID 00001
오후 9:04:35	C	ID 00001
오후 8:15:43	A	ID 00001
오전 10:41:49	B	ID 13247
오후 1:07:35	C	ID 13247
오후 2:22:59	B	ID 13247
오후 2:04:27	B	ID 13247
오후 4:57:06	C	ID 00586
오후 3:10:53	C	ID 00586
:	:	:

{ y }

Consumer



{ Z }

Model

1 . Among the competing services (i = A, B, C, D), an individual(h)'s purchase of a service/product(i) is dependent on the attributes,$\mathbf{x}_{h,i} = (x_{1,i}, \cdots, x_{K,i})$ and how much she values those attributes,$\boldsymbol{\beta}_h = (\beta_{1,h}, \cdots, \beta_{K,h})$; She purchases one that she likes most.

$$P\left(y_{h,t} = \mathrm{i} \mid \boldsymbol{\beta}_h, \mathbf{X}_{h,t}\right) = \frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}$$

$$h = 1,..,H \;, \quad i = A,..,D, \quad t = 1,..,T_h$$

Model

2 . Such attribute importance is 'different' across individuals ($\boldsymbol{\beta}_h \neq \boldsymbol{\beta}_{h'}$)

$$P\left(y_{h,t} = i \mid \boldsymbol{\beta}_h, \mathbf{X}_{h,t}\right) = \frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}$$

$$\boldsymbol{\beta}_1 = (\beta_{1,h}, \cdots, \beta_{K,h})$$
$$\boldsymbol{\beta}_2 = (\beta_{1,h}, \cdots, \beta_{K,h})$$
$$\vdots$$
$$\boldsymbol{\beta}_H = (\beta_{1,h}, \cdots, \beta_{K,h})$$

h	$\beta_{1,h}$	$\cdots$	$\beta_{K,h}$
1	2 ,	$\cdots$	, -1
2	0 ,	$\cdots$	, -3
:			
H	-3 ,	$\cdots$	, -0.1

Model

2 . Such attribute importance is 'different' across individuals ($\boldsymbol{\beta}_h \neq \boldsymbol{\beta}_{h'}$)

3 . Individual-specific parameter may be related to that individuals observed characteristics such as demo-, techno-, socio-, psycho-graphic characteristics, $\mathbf{z}_h = (z_{1,h}, \cdots, z_{N,h})$

$$P(y_{h,t} = \mathrm{i} \,|\, \boldsymbol{\beta}_h, \mathbf{X}_{h,t}) = \frac{\exp(\boldsymbol{\beta}_h{}' \mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}' \mathbf{x}_{h,j,t})}$$

$$\boldsymbol{\beta}_1 = (\beta_{1,h}, \cdots, \beta_{K,h})$$
$$\boldsymbol{\beta}_2 = (\beta_{1,h}, \cdots, \beta_{K,h})$$
$$\vdots$$
$$\boldsymbol{\beta}_H = (\beta_{1,h}, \cdots, \beta_{K,h})$$

h	$\beta_{1,h}$	$\cdots$	$\beta_{K,h}$
1	2 ,	$\cdots$	, -1
2	0 ,	$\cdots$	, -3
:			
H	-3 ,	$\cdots$	, -0.1

$Z_{1,h}$	$\cdots$	$Z_{N,h}$
1 ,	$\cdots$	, 0
1 ,	$\cdots$	, 2
4 ,	$\cdots$	, 1

Model

2 . Such attribute importance is 'different' across individuals ($\boldsymbol{\beta}_h \neq \boldsymbol{\beta}_{h'}$)

3 . Individual-specific parameter maybe related to that individuals observed characteristics such as demo-, techno-, socio-, psycho-graphic characteristics, $\mathbf{z}_h = (z_{1,h}, \cdots, z_{N,h})$ which is 'mean-corrected'.

$$P(y_{h,t} = i \mid \boldsymbol{\beta}_h, \mathbf{X}_{h,t}) = \frac{\exp(\boldsymbol{\beta}_h{}' \mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}' \mathbf{x}_{h,j,t})}$$

$$\underset{(k \times 1)}{\boldsymbol{\beta}_h} = \underset{(k \times N)}{\Gamma} \underset{(N \times 1)}{\mathbf{z}_h} + \underset{(K \times 1)}{\zeta_h} \qquad \zeta_h \sim N(\mathbf{0}, \mathbf{V}_\beta)$$

Model

- Parameter : $\{\boldsymbol{\beta}_h, h = 1,..,H\}, \quad , \mathbf{V}_\beta$
- Data : $\{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\} \;\; ; h = 1,..,H; \;\; t = 1,..T_h$

- Likelihood: $P\left(y_{h,t} = \mathrm{i} \,\middle|\, \boldsymbol{\beta}_h, \mathbf{X}_{h,t}\right) = \dfrac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}$

- Heterogeneity: $\underset{k\cdot 1}{\boldsymbol{\beta}_h} = \overset{k\cdot n}{\Gamma} \underset{N\cdot 1}{\mathbf{z}_h} + \zeta_h \quad \Leftrightarrow \quad \boldsymbol{\beta}_h \sim N(\Gamma \mathbf{z}_h, \underbrace{\mathbf{V}_\beta}_{k\cdot K})$

- Prior: $P\left(vec(\Gamma) | \mathbf{V}_\beta\right) = N(\Gamma_0, \mathbf{V}_\beta \otimes \underbrace{\mathbf{A}^{-1}}_{n\cdot n})$

  $P\left(\mathbf{V}_\beta\right) = Invert\ Wishart(d_0, \mathbf{D}_0)$

Variable description

Conjoint Data: y, X, z

H = 946; Total observations = 14,799

Y: At each choice task, **two** alternatives are offered: i=1, 2

X : Attributes (service features)

X	Attributes
X_1	Low fixed rate
X_2	Low annual fee
X_3	Out-of-state
:	:
X_{14}	High Credit limit

Z: customer characteristics

Z	Characteristics
1	Age
2	Income
3	Gender

Case source

Allenby, G. and J. Ginter, "Using Extremes to Design Products and Segment Markets" 1995.

Output and Input

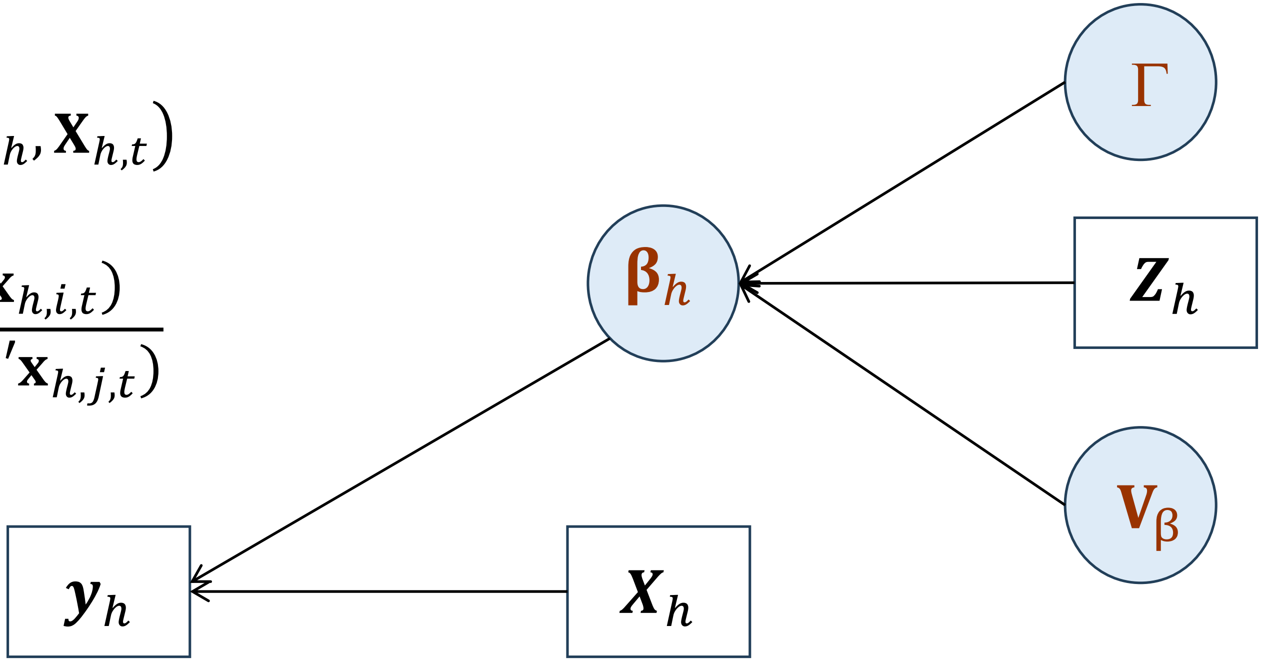
- Parameter : $\{\boldsymbol{\beta}_h, h = 1, .., H\}, \quad , \mathbf{V}_{\beta}$
- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_{\beta} \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

Output and Input

- Parameter : $\{\boldsymbol{\beta}_h, h = 1, \ldots, H\}, \quad , \mathbf{V}_\beta$
- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

$$P\left(y_{h,t} = \mathrm{i} \mid \boldsymbol{\beta}_h, \mathbf{X}_{h,t}\right)$$

$$= \frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}$$



$$\boldsymbol{\beta}_h \sim N(\Gamma \mathbf{z}_h, \mathbf{V}_\beta)$$

Estimation

- Parameter : $\{\boldsymbol{\beta}_h, h = 1, \ldots, H\}, \quad , \mathbf{V}_\beta$
- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

$$\{\boldsymbol{\beta}_h\} \sim P(\{\boldsymbol{\beta}_h\} \mid \Gamma, \mathbf{V}_\beta, \{y_{h,t}\}, \{\mathbf{X}_{h,t}\})$$

$$\Gamma \sim P(\Gamma \mid \{\boldsymbol{\beta}_h\}, \mathbf{V}_\beta, \{\mathbf{z}_h\})$$

$$\mathbf{V}_\beta \sim P\left(\mathbf{V}_\beta \mid \{\boldsymbol{\beta}_h\}, \Gamma, \{\mathbf{z}_h\}\right)$$

Estimation

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

---

1. $\{\boldsymbol{\beta}_h\} \sim P(\{\boldsymbol{\beta}_h\} \mid \Gamma, \mathbf{V}_\beta, \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}) \propto \prod_t \prod_i \left[\frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}\right]^{d_{h,i,t}} \times N(\Gamma \mathbf{z}_h, \mathbf{V}_\beta)$

$\boxed{d_{h,i,t} = I(y_{h,t} = i)}$

For h=1,

draw, δ, from $N(0, \tau^2) \rightarrow \boldsymbol{\beta}_{candidate}^{(i)} = \boldsymbol{\beta}_h^{(i-1)} + \delta$

$$log.posterior(candidate) = \left[\sum_t d_{h,i,t} \ln\left(\frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}\right) + \ln N(\boldsymbol{\beta}_h | \Gamma \mathbf{z}_h, \mathbf{V}_\beta)\right]_{\boldsymbol{\beta}_{candidate}^{(i)}}$$

$$log.posterior(current) = \left[\sum_t d_{h,i,t} \ln\left(\frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}\right) + \ln N(\boldsymbol{\beta}_h | \Gamma \mathbf{z}_h, \mathbf{V}_\beta)\right]_{\boldsymbol{\beta}_h^{(i-1)}}$$

Estimation

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

---

1. $\{\boldsymbol{\beta}_h\} \sim P(\{\boldsymbol{\beta}_h\} \mid \Gamma, \mathbf{V}_\beta, \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}) \propto \prod_t \prod_i \left[\frac{\exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,i,t})}{\sum_j \exp(\boldsymbol{\beta}_h{}'\mathbf{x}_{h,j,t})}\right]^{d_{h,i,t}} \times N(\Gamma \mathbf{z}_h, \mathbf{V}_\beta)$

$$a = \exp(log.posterior(candidate) - log.posterior(current))$$

One draw, u, from $Uniform(0,1) \rightarrow$ $\boldsymbol{\beta}_h^{(i)} = \boldsymbol{\beta}_{candidate}^{(i)} \; if \; a > u$

$\boldsymbol{\beta}_h^{(i)} = \boldsymbol{\beta}_h^{(i-1)} \quad otherwise$

$\rightarrow$ Repeat for h=2,…..H

Estimation

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

---

2. $\Gamma \sim P(\Gamma \mid \{\boldsymbol{\beta}_h\}, \mathbf{V}_\beta, \{\mathbf{z}_h\})$

$$vec(\Gamma) \sim N(\boldsymbol{\gamma}^*, \mathbf{G})$$

$$\mathbf{G} = (\mathbf{Z}^{*\prime}(\mathbf{I}\otimes\mathbf{V}_\beta^{-1})\mathbf{Z}^*)^{-1}$$

$$\boldsymbol{\gamma}^* = \mathbf{G}(\mathbf{Z}^{*\prime}(\mathbf{I}\otimes\mathbf{V}_\beta^{-1})\mathbf{Z}^*\tilde{\mathbf{v}})$$

$$\tilde{\mathbf{v}} = (\mathbf{Z}^{*\prime}(\mathbf{I}\otimes\mathbf{V}_\beta^{-1})\mathbf{Z}^*)^{-1}\ (\mathbf{Z}^{*\prime}(\mathbf{I}\otimes\mathbf{V}_\beta^{-1})\boldsymbol{\beta})$$

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_H')'$$

$$\mathbf{Z}^* = (\mathbf{I}\otimes\mathbf{z}_{\mathrm{h=1}}, \mathbf{I}\otimes\mathbf{z}_{\mathrm{h=2}}, \cdots, \mathbf{I}\otimes\mathbf{z}_{\mathrm{h=H}})'$$

Estimation

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

---

3. $\mathbf{V}_\beta \sim P\left(\mathbf{V}_\beta \mid \{\boldsymbol{\beta}_h\}, \Gamma, \{\mathbf{z}_h\}\right)$

$$\mathbf{V}_\beta \sim Inverted\ Whishart(d_1, \mathbf{D}_1)$$

$$d_1 = d_0 + H$$

$$\mathbf{D}_1 = \mathbf{D}_0 + \sum_h (\boldsymbol{\beta}_h - \Gamma \mathbf{z}_h)(\boldsymbol{\beta}_h - \Gamma \mathbf{z}_h)'$$

Outputs

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

$$\{\boldsymbol{\beta}_h\} \sim P(\{\boldsymbol{\beta}_h\} \mid \Gamma, \mathbf{V}_\beta, \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}) \qquad h = 1, .., H$$

$(H \times K) \quad \times \text{no. iterations}$

$$\Gamma \sim P(\Gamma \mid \{\boldsymbol{\beta}_h\}, \mathbf{V}_\beta, \{\mathbf{z}_h\})$$

$(K \times N) \quad \times \text{no. iterations}$

$$\mathbf{V}_\beta \sim P\left(\mathbf{V}_\beta \mid \{\boldsymbol{\beta}_h\}, \Gamma, \{\mathbf{z}_h\}\right)$$

$(K \times K) \quad \times \text{no. iterations}$

$\rightarrow \frac{K \times (K+1)}{2}$

- ✓ **1st, Trace plots**
- ✓ **2nd, Posterior distribution**
- ✓ **3rd, Summary**

Outputs

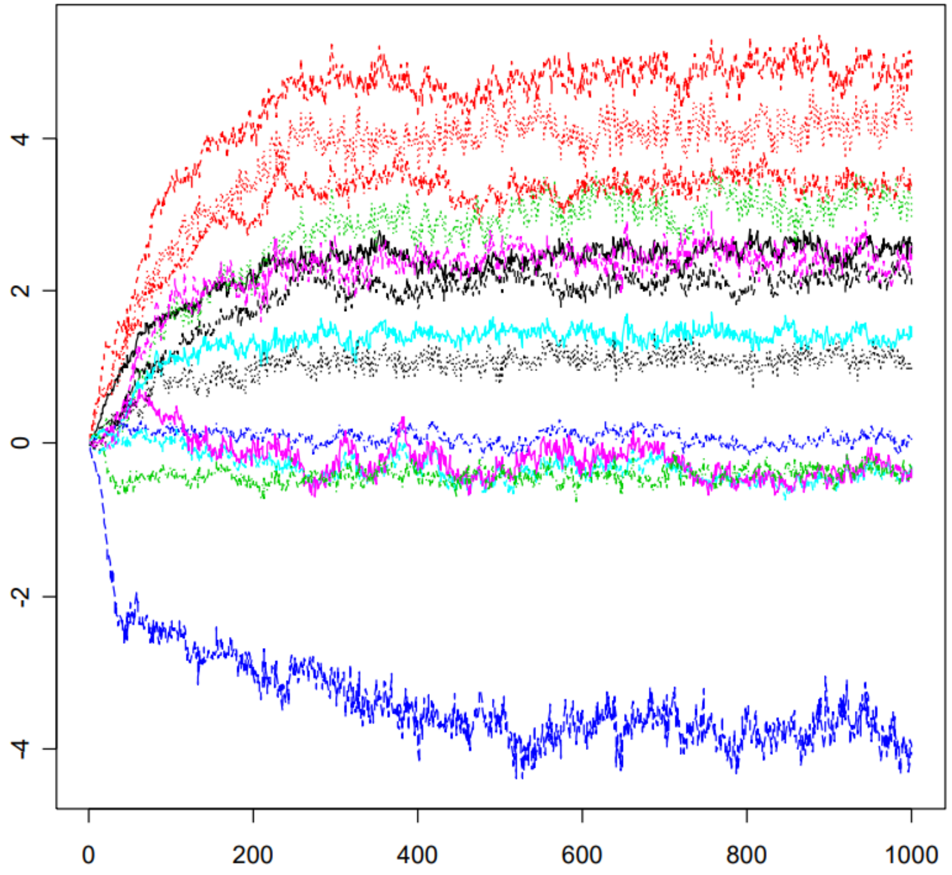
- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

$$\underset{(K \times N)}{\Gamma} \sim \underset{\times \text{ no. iterations}}{P(\Gamma \mid \{\boldsymbol{\beta}_h\}, \mathbf{V}_\beta, \{\mathbf{z}_h\})}$$

✓ **1st, Trace plots**

2nd, Posterior distribution

3rd, Summary



Outputs

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

$$\underset{(K \times N)}{\Gamma} \sim P(\Gamma \mid \{\boldsymbol{\beta}_h\}, \mathbf{V}_\beta, \{\mathbf{z}_h\})$$

1st, Trace plots

2nd, Posterior distribution

✓ **3rd, Summary**

Customer feature / **Product feature**	$\mathbf{z}_1$ Intercept	$\mathbf{z}_2$ Age	$\mathbf{z}_3$ Income	$\mathbf{z}_4$ Gender
$(\boldsymbol{\beta}_1)$ Low Fixed Interest	4.383	-0.025	0.021	0.324
$(\boldsymbol{\beta}_2)$ Low Annual Fee	4.158	-0.010	0.004	1.302
$(\boldsymbol{\beta}_3)$ Out-of-State Bank	-3.758	-0.003	0.013	-0.054
:				
$(\boldsymbol{\beta}_K)$ High Credit Limit	1.116	-0.010	-0.003	0.368

Outputs

- Posterior : $P(\{\boldsymbol{\beta}_h\}, \Gamma, \mathbf{V}_\beta \mid \{y_{h,t}\}, \{\mathbf{X}_{h,t}\}, \{\mathbf{z}_h\})$

$$\underset{(K \times K)}{\mathbf{V}_\beta} \sim P\left(\mathbf{V}_\beta \mid \{\boldsymbol{\beta}_h\}, \Gamma, \{\mathbf{z}_h\}\right)$$

1st, Trace plots

2nd, Posterior distribution

✓ **3rd, Summary**

Covariance / ***Correlation***

	β_1	β_2	β_3	$\ldots$	β_K
(β_1) Low Fixed Interest	8.8	*0.32*	*0.18*		*0.16*
(β_2) Low Annual Fee	3.4	13.5	*0.55*		*0.47*
(β_3) Out-of-State Bank	2.1	8.1	15.9		*0.39*
$\vdots$				$\ddots$	
(β_K) High Credit Limit	1.2	4.3	3.9	$\ldots$	6.3

Managerial Decision

Whom (to target) **and What** (to offer)

---

00 기업

Want : 현재 사업 지역 → 다른 지역으로 확장

As-is : 낯선 경쟁사의 서비스를 좋아할 리는 없을 것. 승산은 없는가?

(β_3) Out-of-State Bank	-3.758

Prediction/Decision: 어느 고객에게 어떤 식으로 제안하면 그래도 승산을 있을까?

Managerial Decision

Whom (to target) **and What** (to offer)

- To identify incentives so that a regional bank can successfully offer credit cards to out-of-state customers.

(β_3) Out-of-State Bank	**-3.758**

- **Two incentives** to overcome this penalty:

  Option 1: low fixed interest (β_1)

  Option 2: low annual fee (β_2)

Managerial Decision

Whom (to target) **and What** (to offer)

- To identify incentives so that a regional bank can successfully offer credit cards to out-of-state customers.

($\boldsymbol{\beta}_3$) Out-of-State Bank	**-3.758**

Customer feature **Product feature**	$\mathbf{z_1}$ Intercept	β_1	β_2	β_3	$\cdots$	β_K
($\boldsymbol{\beta}_1$) Low Fixed Interest	4.383	8.8	*0.32*	*0.18*		*0.16*
($\boldsymbol{\beta}_2$) Low Annual Fee	4.158	3.4	13.5	*0.55*		*0.47*
($\boldsymbol{\beta}_3$) Out-of-State Bank	-3.758	2.1	8.1	15.9		*0.39*
:					$\ddots$	
($\boldsymbol{\beta}_K$) High Credit Limit	1.116	1.2	4.3	3.9	$\cdots$	6.3

Managerial Decision

Whom (to target) **and What** (to offer)

- To identify incentives so that a regional bank can successfully offer credit cards to out-of-state customers.

(β_3) Out-of-State Bank	**-3.758**

- **Two incentives** to overcome this penalty:

  Option 1: low fixed interest (β_1) $\rightarrow$ on average $(\beta_1 - |\beta_3|) = 4.38$

  Option 2: low annual fee (β_2) $\rightarrow$ on average $(\beta_2 - |\beta_3|) = 4.16$

➔ (1) Relying on '**average**', approximately the same increase in consumer utility

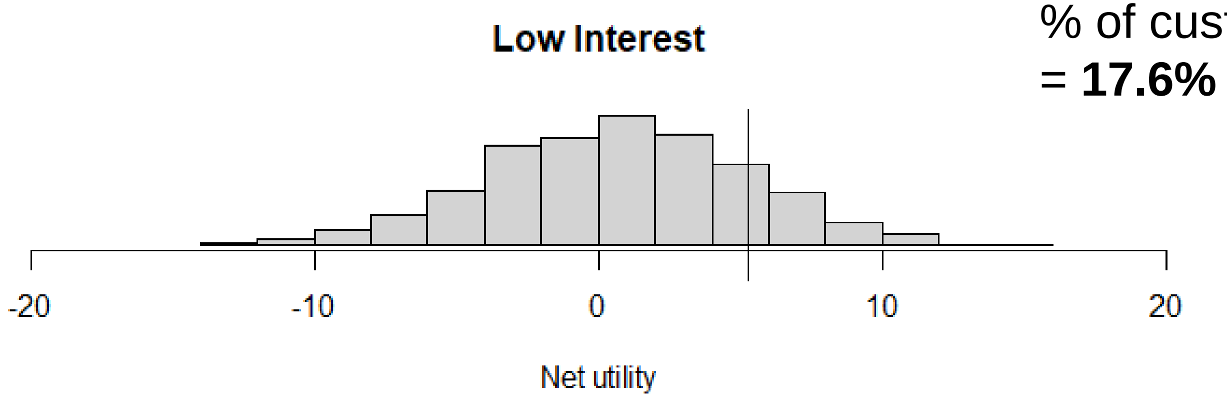
(2) Accounting for '**heterogeneity(variance)**', they are not the same.

Managerial Decision

Whom (to target) **and What** (to offer)

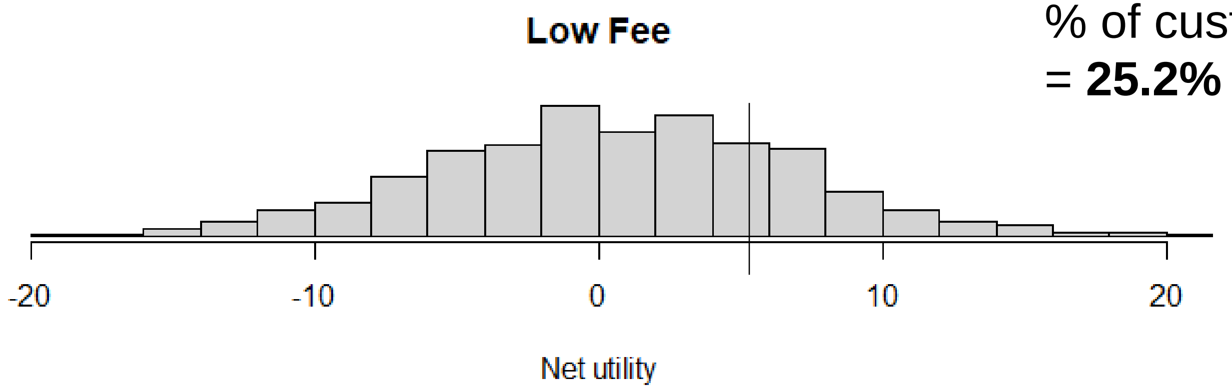
Heterogeneity of net utility across customers

$(\beta_1 - |\beta_3|)$



% of customers ($(\beta_1 - |\beta_3|) > 5$)
= **17.6%**

$(\beta_2 - |\beta_3|)$



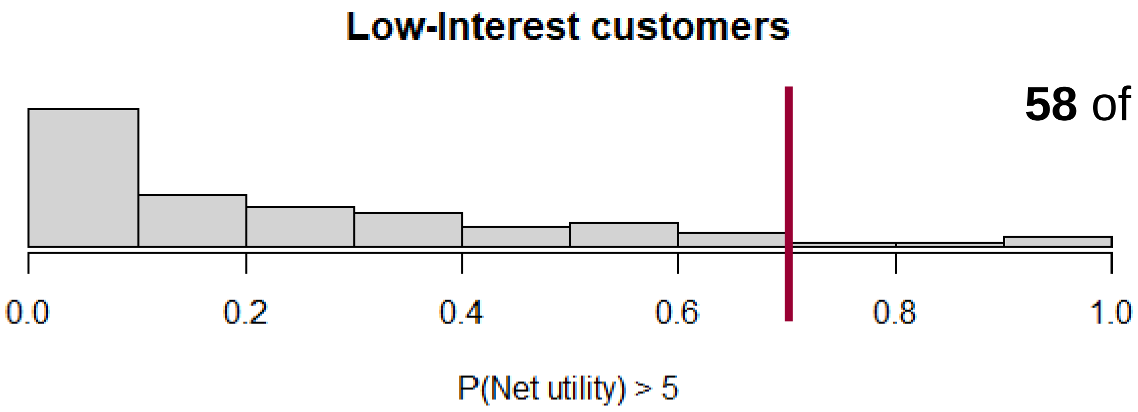
% of customers ($(\beta_2 - |\beta_3|) > 5$)
= **25.2%**

Managerial Decision

Whom (to target) **and What** (to offer)

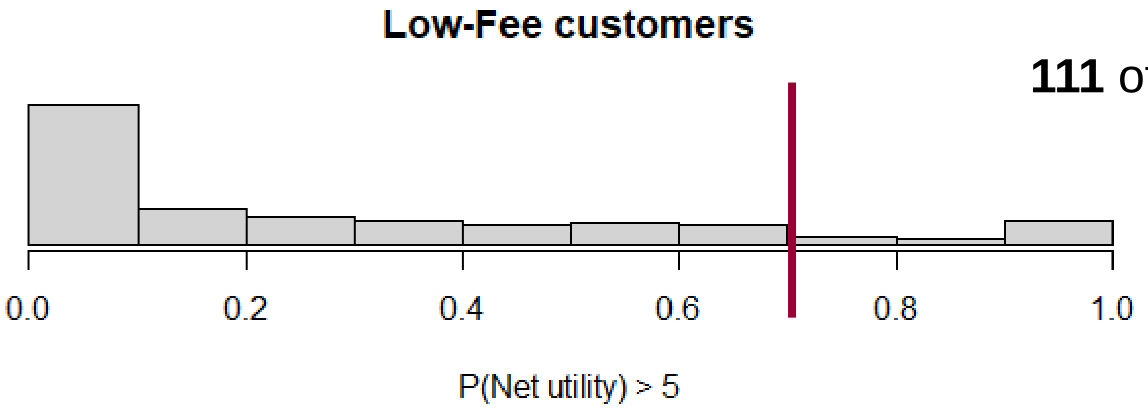
Expected size of customers?

$$\frac{\sum \mathrm{I}\left[\beta_1^{(r)} - |\beta_3^{(r)}| > 5\right]}{R}$$



58 of 946 customers

$$\frac{\sum \mathrm{I}\left[\beta_2^{(r)} - |\beta_3^{(r)}| > 5\right]}{R}$$



111 of 946 customers

Managerial Decision

Whom (to target) **and What** (to offer)

- **Two incentives** to overcome this penalty:

  Option 1: low fixed interest (β_1)

  Option 2: low annual fee (β_2)

Summary

- **Decision problem: Whom to target and what to offer**

- **Understanding 'heterogenous' response propensity is the key.**

  Posterior distribution of Individual-specific parameter

- **Bayesian framework meets heterogeneity.**

  "Likelihood – Prior" setting mimics "Likelihood – Heterogeneity"

  $P(\mathrm{y}|\beta_h) - \beta_h \sim \beta_h|\beta_o, \mathrm{V}_o$ $\qquad$ $P(\mathrm{y}|\beta_h) - \beta_h \sim \beta_h|\bar{\beta}, \mathrm{V}_\beta$

  $$\bar{\beta} \sim \bar{\beta} \mid \beta_h, \mathrm{V}_\beta$$

  $$\mathrm{V}_\beta \sim \mathrm{V}_\beta \mid \beta_h, \bar{\beta}$$