

# 認知情報解析学 | ベイズ統計 & MCMC

# BAYES' RULE

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

$$p(x|y)p(y) = p(x,y)$$

$$p(y|x)p(x) = p(x,y)$$

$$p(x|y)p(y) = p(y|x)p(x)$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\sum_i p(y|x_i)p(x_i)}$$

# BAYES' RULE - 例題

$$p(x|y) = \frac{p(y|x)p(x)}{\sum_i p(y|x_i)p(x_i)}$$

$p(D|+test)$

$$= \frac{p(+test|D)p(D)}{p(+test|D)p(D) + p(+test|\neg D)p(\neg D)}$$

$$\begin{aligned} p(D|+test) &= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times (1 - 0.001)} \\ &= 0.019 \end{aligned}$$

病気Dの発症率 : 0.001

試薬の正しい陽性反応 : 0.99

試薬の間違った陽性反応 : 0.05

陽性反応があった場合の病気Dを発症している確率は？

# モデルへの応用

$$p(x|y) = \frac{p(y|x)p(x)}{\sum_i p(y|x_i)p(x_i)}$$

$$p(Param|Data) = \frac{p(Data|Param)p(Param)}{\sum_i p(Data|Param_i)p(Param_i)} = \frac{p(D|\theta)p(\theta)}{\sum_i p(D|\theta_i)p(\theta_i)}$$

$p(\theta|D)$  事後分布 posterior

$p(D|\theta)$  尤度 likelihood

$p(\theta)$  事前分布 prior

$p(D) = \sum_i p(D|\theta_i)p(\theta_i)$  evidence (for the model)

# データの順序の影響

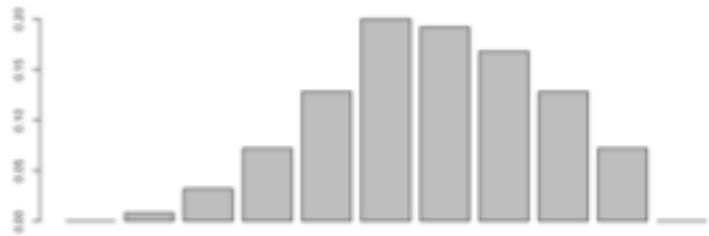
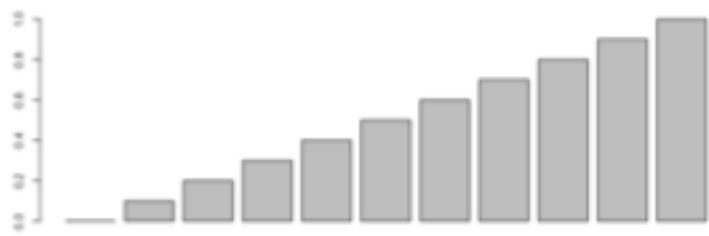
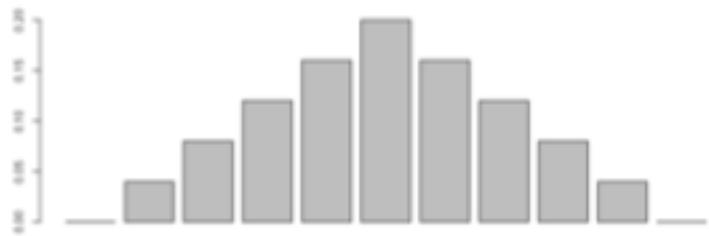
$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\sum_i p(D|\theta_i)p(\theta_i)}$$

$$\begin{aligned} p(\theta|D_1, D_2) &= \frac{p(D_1, D_2|\theta)p(\theta)}{\sum_i p(D_1, D_2|\theta_i)p(\theta_i)} \\ &= \frac{p(D_1|\theta)p(D_2|\theta)p(\theta)}{\sum_i p(D_1|\theta_i)p(D_2|\theta_i)p(\theta_i)} \\ &= \frac{p(D_2|\theta)p(D_1|\theta)p(\theta)}{\sum_i p(D_2|\theta_i)p(D_1|\theta_i)p(\theta_i)} = p(\theta|D_2, D_1) \end{aligned}$$

# 例 — 1□のコイン

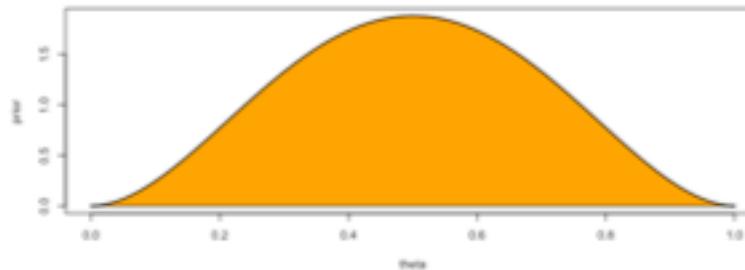
Likelihood – Bernoulli dist.

$$p(y|\theta) = \theta^y(1 - \theta)^{(1-y)}$$

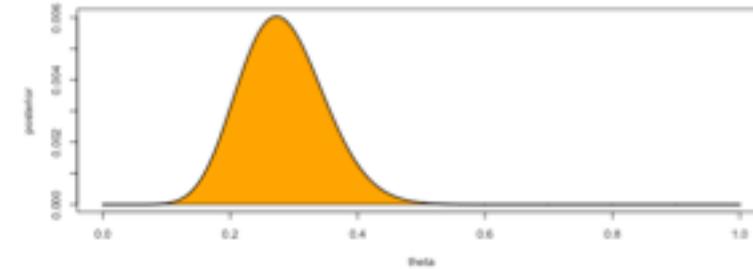
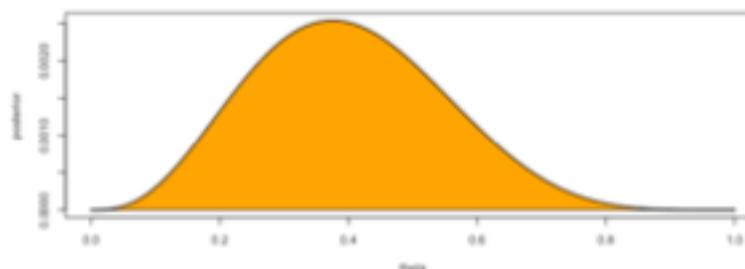
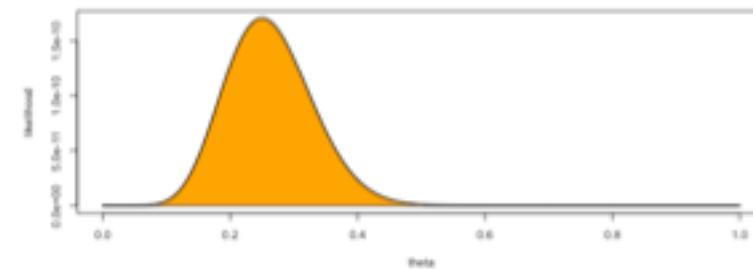
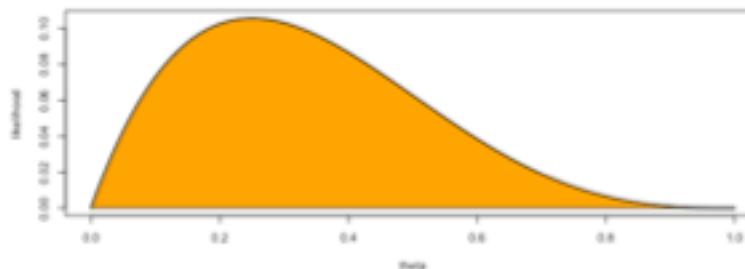
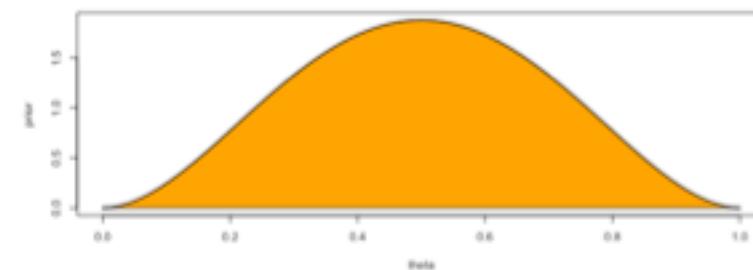


# DATAの数の影響

$N=4, Z = 1$

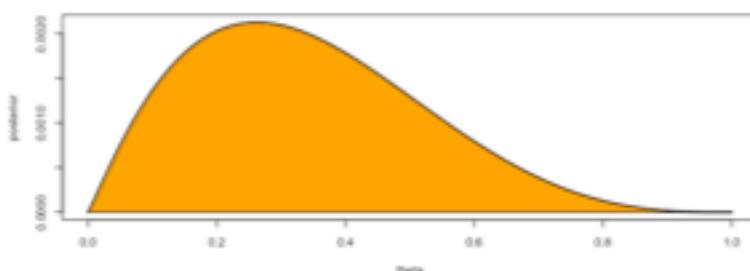
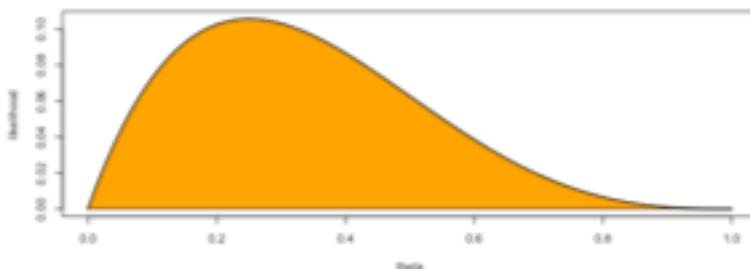
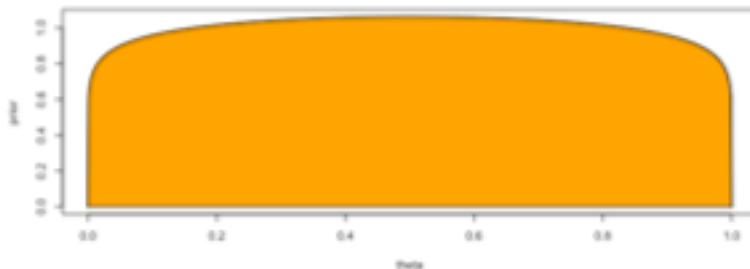


$N=40, Z = 10$

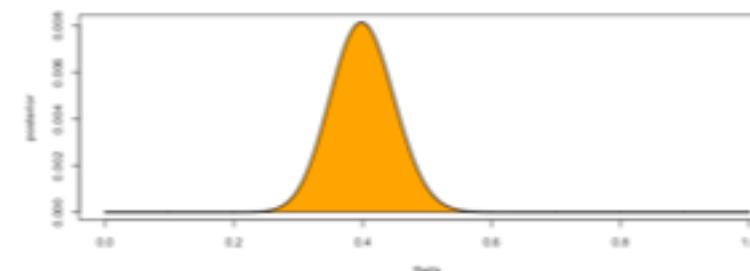
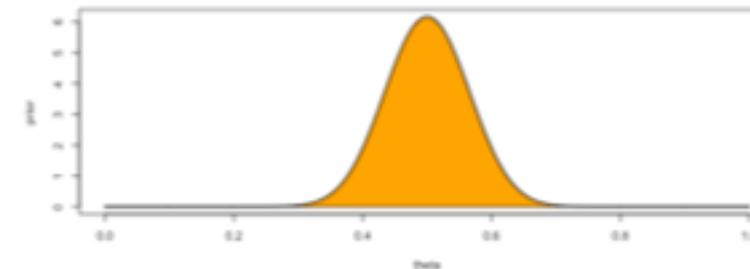


# PRIORの影響

$N=4, Z=1$



$N=40, Z=10$



# LIKELIHOOD 再び

Likelihood – Bernoulli dist.

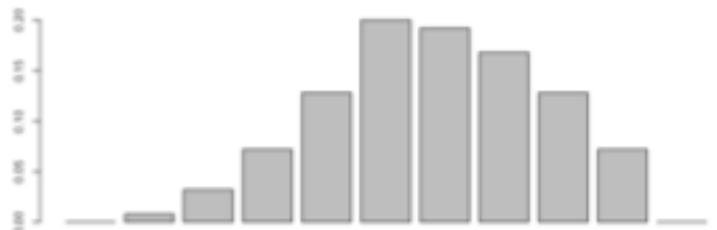
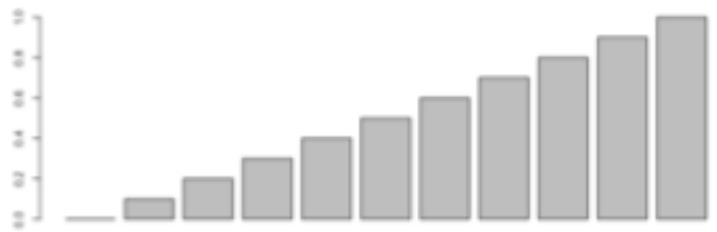
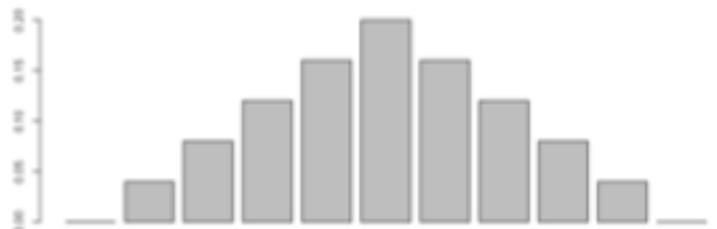
$$p(y|\theta) = \theta^y(1-\theta)^{1-y}$$

Likelihood – Binomial dist.

$$p(\mathbf{y}|\theta) = \prod_i p(y_i|\theta)$$

$$= \prod_i \theta^{y_i}(1-\theta)^{1-y_i}$$

$$= \theta^{\sum y_i}(1-\theta)^{\sum(1-y_i)} = \theta^z(1-\theta)^{N-z}$$

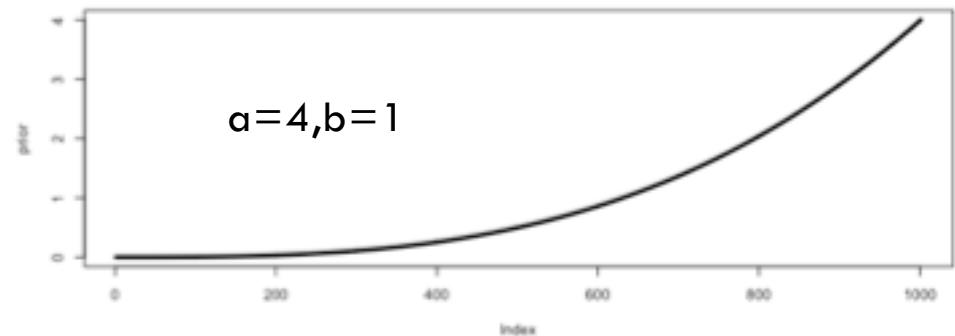
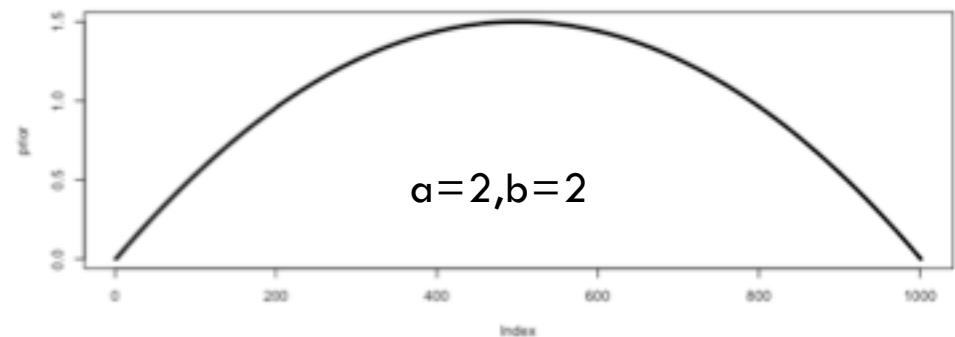
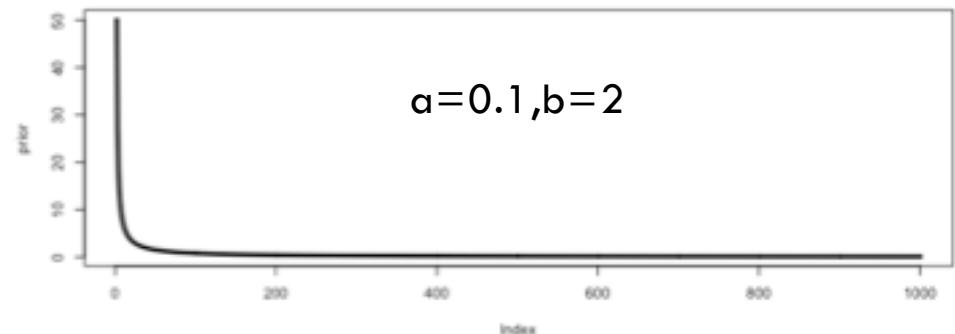


# ベータ分布(PRIOR)

$$p(\theta|a,b) = \text{beta}(\theta|a,b)$$

$$= \frac{\theta^{(a-1)}(1-\theta)^{(b-1)}}{B(a,b)}$$

$$B(a,b) = \int_0^1 \theta^{(a-1)}(1-\theta)^{(b-1)} d\theta$$



# POSTERIOR

$$\begin{aligned} p(\theta|z, N) &= \frac{p(z, N|\theta)p(\theta)}{p(z, N)} \\ &= \frac{\theta^z(1-\theta)^{(N-z)}\frac{\theta^{(a-1)}(1-\theta)^{(b-1)}}{B(a,b)}}{p(z,N)} \\ &= \frac{\theta^z(1-\theta)^{(N-z)}\theta^{(a-1)}(1-\theta)^{(b-1)}}{B(a,b)p(z,N)} \\ &= \frac{\theta^{(z+a-1)}(1-\theta)^{(N-z+b-1)}}{B(a,b)p(z,N)} \\ &= \frac{\theta^{(z+a-1)}(1-\theta)^{(N-z+b-1)}}{\theta^{(z+a-1)}(1-\theta)^{(N-z+b-1)}} \\ &= \frac{B(z+a, N-z+b)}{B(z+a, N-z+b)} \end{aligned}$$

$$\begin{aligned} p(\theta|z, N) &= \frac{\theta^{(z+a-1)}(1-\theta)^{(N-z+b-1)}}{B(z+a, N-z+b)} \\ &= beta(\theta|z+a, N-z+b) \end{aligned}$$

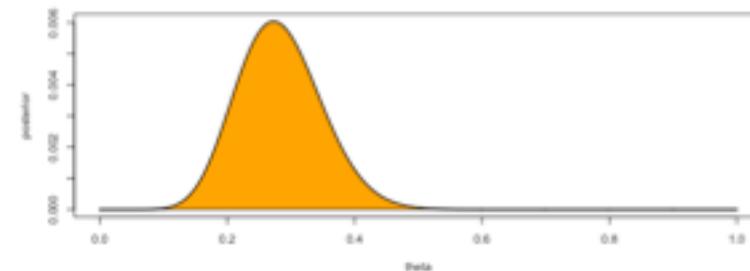
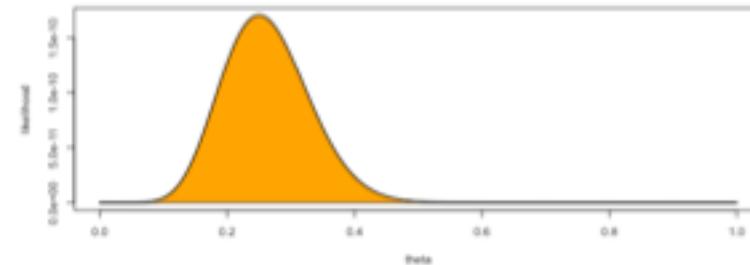
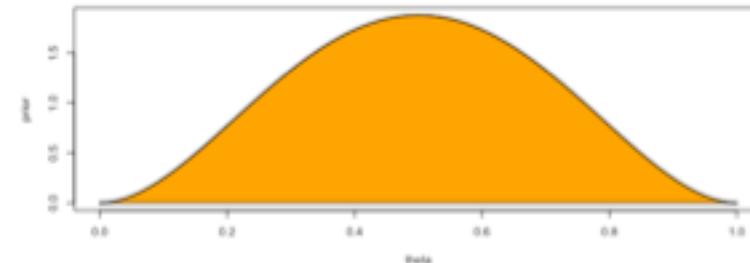
# 実習1：

Beta Prior

Bernoulli/binomial Likelihood

Posterior

の関係を可視化する関数を作りましょう



# ベイズ統計の問題点

分析的に解けない  $P(D)$

Grid search ではパラメターが増えるごとに計算量が指数関数的に増える

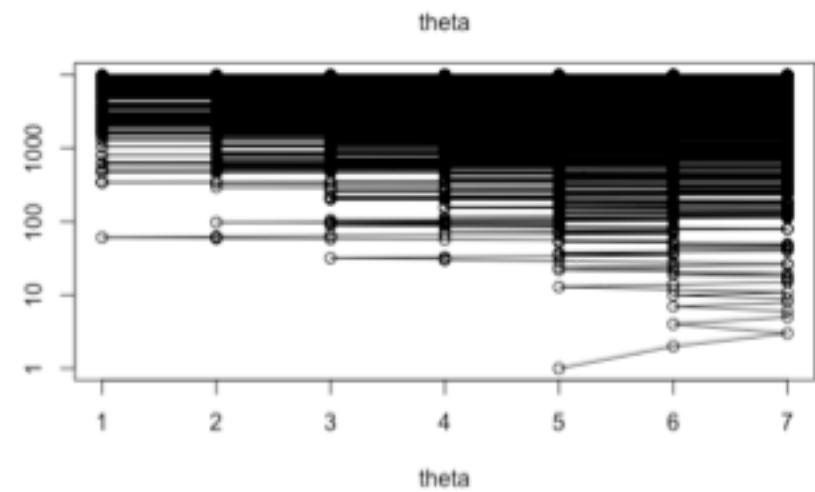
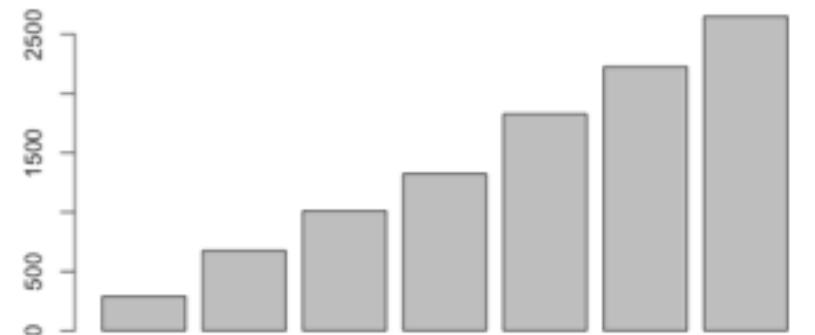
解決策

Markov Chain Monte Carlo (MCMC) posterior distribution をシミュレーションを用いて近似する

# 実習2 MCMCのちょっと前

設定:

- 政治家が7つの島へ政治活動を行う
- 人口は、1、2、3、4、5、6、7とする
- 隣の島へのみ移動することができる
- 右隣の島、左隣の島を候補にする確率は0.5
- 実際に移動するか否かは以下の確率で決定する
- $p(\text{move}) = \frac{\text{population}_{\text{proposed}}}{\text{population}_{\text{current}}}$
- もし  $\text{UNIF}_{\text{random}} < p(\text{move}) \rightarrow \text{move}$



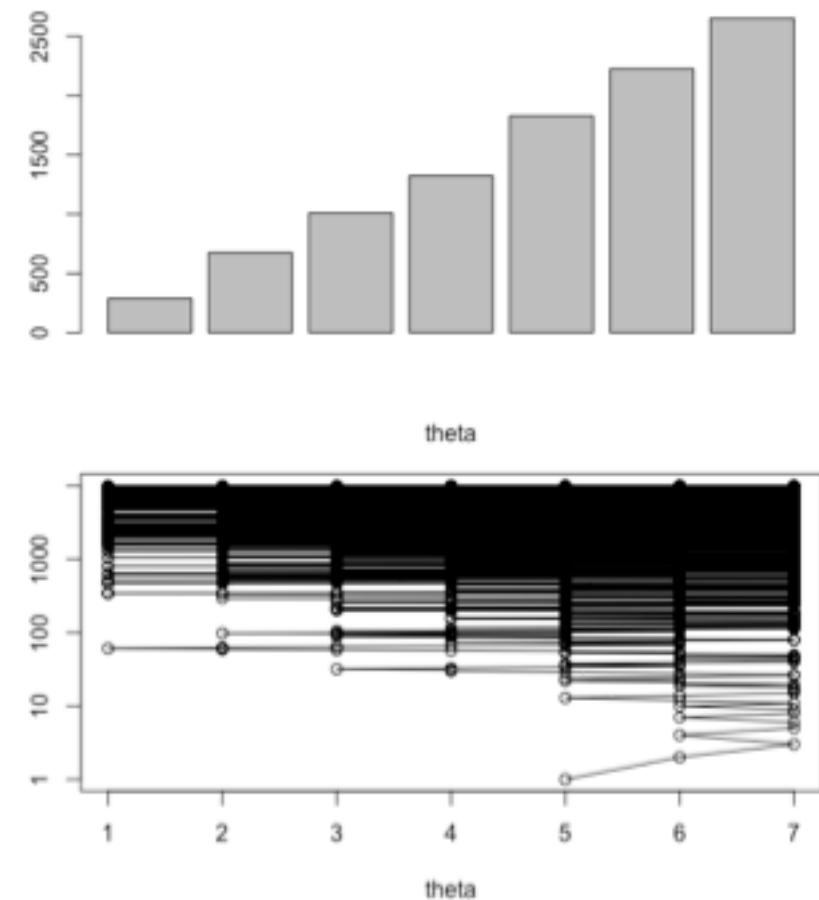
# METROPOLIS ALGORITHM

$$p(\text{move}) = \min\left(\frac{\text{population}_{\text{proposed}}}{\text{population}_{\text{current}}}, 1\right)$$

If UNIF<sub>random</sub> <  $p(\text{move}) \rightarrow \text{move}$

必要なrandom-walkでのプロセス:

- 新たなパラメター値を提案するための乱数の生成
- ターゲット分布での評価
  - i. e., Compute POP(theta\_proposed)/POP(theta\_current)
- uniform distributionに従う乱数を生成、採択・棄却を決定



# WHY IT WORKS

Suppose we are at position  $\theta$

Probability of moving to  $\theta+1 = p(\theta \rightarrow \theta + 1)$

$$p(\theta \rightarrow \theta + 1) = p(\text{propose } \theta + 1) \times p(\text{accept } \theta + 1)$$

$$p(\theta \rightarrow \theta + 1) = 0.5 \times \min\left(\frac{\text{POP}(\theta+1)}{\text{POP}(\theta)}, 1\right)$$

$$p(\theta + 1 \rightarrow \theta) = 0.5 \times \min\left(\frac{\text{POP}(\theta)}{\text{POP}(\theta+1)}, 1\right)$$

$$\frac{p(\theta \rightarrow \theta + 1)}{p(\theta + 1 \rightarrow \theta)} = \frac{0.5 \times \min\left(\frac{\text{POP}(\theta+1)}{\text{POP}(\theta)}, 1\right)}{0.5 \times \min\left(\frac{\text{POP}(\theta)}{\text{POP}(\theta+1)}, 1\right)} = \begin{cases} \frac{1}{\text{POP}(\theta)/\text{POP}(\theta+1)} & \text{if } \text{POP}(\theta + 1) > \text{POP}(\theta) \\ \frac{\text{POP}(\theta)/\text{POP}(\theta+1)}{1} & \text{if } \text{POP}(\theta + 1) < \text{POP}(\theta) \end{cases}$$
$$= \frac{\text{POP}(\theta+1)}{\text{POP}(\theta)}$$

相対的な移動確率は一ゲット分布の相対的な値と全く同じとなる

# WHY IT WORKS

$$T = \begin{bmatrix} 0.5 \times \left[ 1 - \min\left(\frac{P(\theta-2)}{P(\theta-1)}, 1\right) \right] + 0.5 \times \left[ 1 - \min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) \right] & 0.5 \times \min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) & 0 \\ 0.5 \times \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right) & 0.5 \times \left[ 1 - \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right) \right] + 0.5 \times \left[ 1 - \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right) \right] & 0.5 \times \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right) \\ 0 & 0.5 \times \min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) & 0.5 \times \left[ 1 - \min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) \right] + 0.5 \times \left[ 1 - \min\left(\frac{P(\theta+2)}{P(\theta+1)}, 1\right) \right] \end{bmatrix}$$

$$w = [\dots, P(\theta - 1), P(\theta), P(\theta + 1), \dots] / Z$$

$$Z = \sum P(\theta)$$

$$\begin{aligned} \sum_r w T_{r\theta} &= \frac{P(\theta-1)}{Z} 0.5 \min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) \\ &\quad + \frac{P(\theta)}{Z} 0.5 \left[ 1 - \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right) \right] + 0.5 \left[ 1 - \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right) \right] \\ &\quad + \frac{P(\theta+1)}{Z} 0.5 \min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) \end{aligned}$$

# WHY IT WORKS

$$\begin{aligned}\sum_r wT_{r\theta} &= \frac{P(\theta-1)}{Z} 0.5 \min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) \\ &\quad + \frac{P(\theta)}{Z} 0.5 \times \left[1 - \min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right)\right] + 0.5 \times \left[1 - \min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right)\right] \\ &\quad + \frac{P(\theta+1)}{Z} 0.5 \min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right)\end{aligned}$$

$P(\theta - 1) < P(\theta) \& P(\theta + 1) < P(\theta)$  の場合

- $\sum_r wT_{r\theta} = \frac{P(\theta-1)}{Z} 0.5 + \frac{P(\theta)}{Z} 0.5 \left[1 - \frac{P(\theta-1)}{P(\theta)}\right] + 0.5 \left[1 - \frac{P(\theta+1)}{P(\theta)}\right] + \frac{P(\theta+1)}{Z} 0.5 = \frac{P(\theta)}{Z}$

$P(\theta - 1) < P(\theta) \& P(\theta) < P(\theta + 1)$  の場合

- $\sum_r wT_{r\theta} = \frac{P(\theta-1)}{Z} 0.5 + \frac{P(\theta)}{Z} 0.5 \left[1 - \frac{P(\theta-1)}{P(\theta)}\right] + 0.5 [1 - 1] + \frac{P(\theta+1)}{Z} \frac{P(\theta)}{P(\theta+1)} = \frac{P(\theta)}{Z}$

$P(\theta) < P(\theta + 1) \& P(\theta - 1) = N/A$  の場合

- $\sum_r wT_{r\theta} = \frac{P(\theta)}{Z} 0.5 + \frac{P(\theta+1)}{Z} 0.5 \frac{P(\theta)}{P(\theta+1)} = \frac{P(\theta)}{Z} 0.5 + \frac{P(\theta)}{Z} 0.5 = \frac{P(\theta)}{Z}$

# METROPOLIS、一般的には

Target distribution  $P(\theta)$

Proposal distribution は様々な形をとれる

Accept a new position at prob =  $\frac{P(\theta_{target})}{P(\theta_{current})}$

A random uniform number is used for decision

# 実習 : MCMC BERNoulli

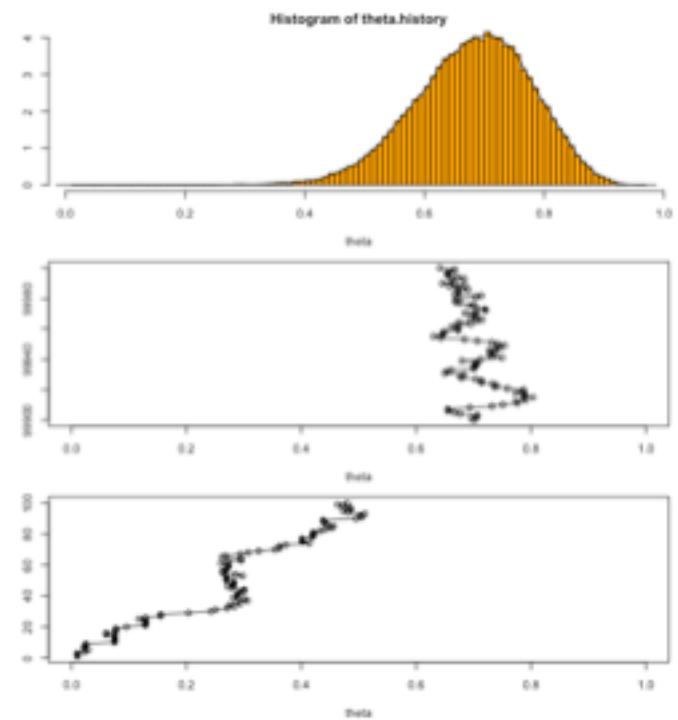
1. Randomly generate a proposed jump  $\Delta\theta \sim N(0, \sigma)$

$$1. \quad \theta_{pro} = \theta_{cur} + \Delta\theta$$

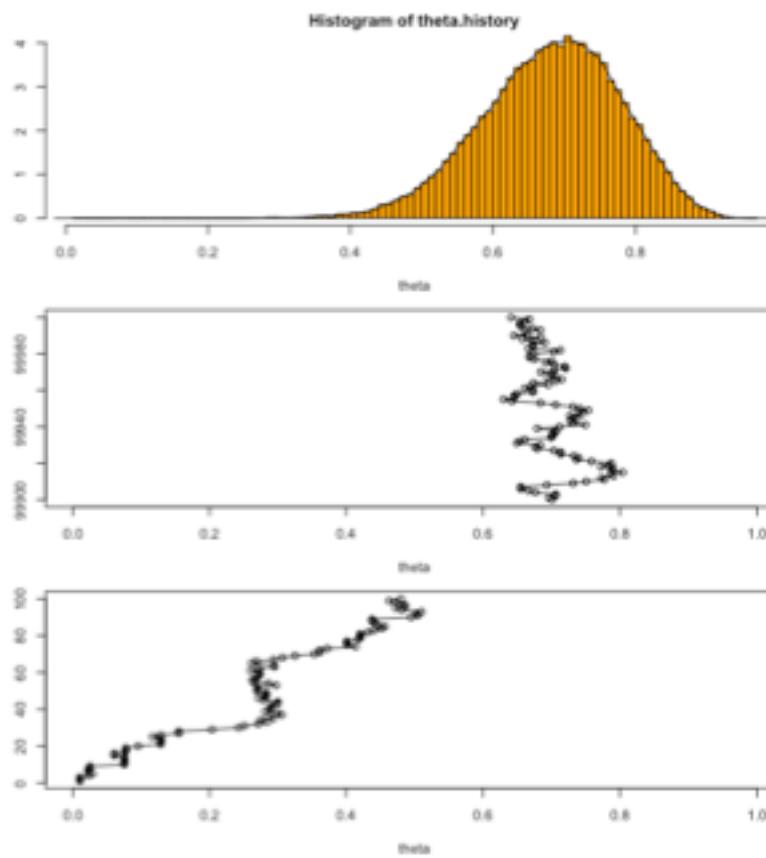
2. Compute probability of moving to the proposed value

$$\begin{aligned} p_{move} &= \min \left( 1, \frac{P(\theta_{pro})}{P(\theta_{cur})} \right) = \min \left( 1, \frac{p(D|\theta_{pro})p(\theta_{pro})}{p(D|\theta_{cur})p(\theta_{cur})} \right) \\ &= \min \left( 1, \frac{Ber(z, ND|\theta_{pro})beta(\theta_{pro}|a,b)}{Ber(z, ND|\theta_{cur})beta(\theta_{cur}|a,b)} \right) \\ &= \min \left( 1, \frac{\theta_{pro}^z (1-\theta_{pro})^{(N-z)} \theta_{pro}^{(a-1)} (1-\theta_{pro})^{(b-1)} / B(a,b)}{\theta_{cur}^z (1-\theta_{cur})^{(N-z)} \theta_{cur}^{(a-1)} (1-\theta_{cur})^{(b-1)} / B(a,b)} \right) \end{aligned}$$

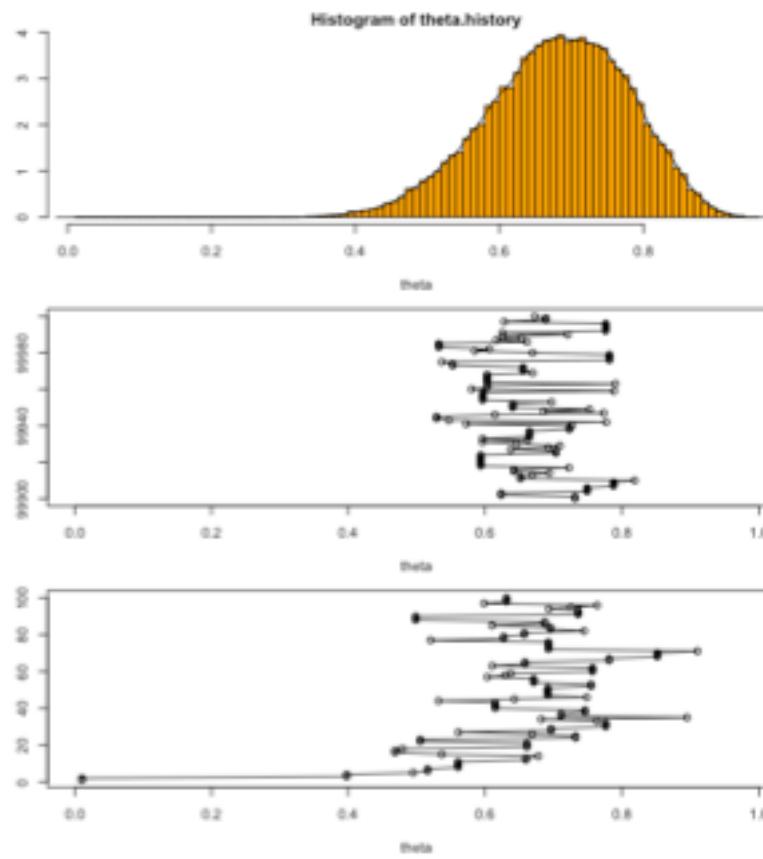
3. Accept new value of runif < pmove



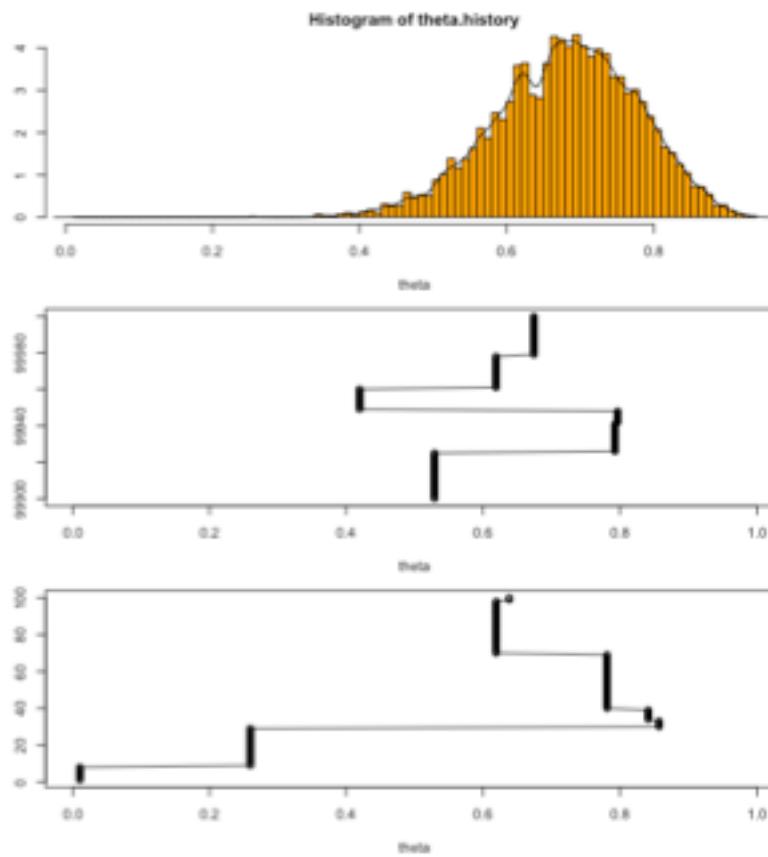
# EXAMPLE SIGMA=0.02



# EXAMPLE SIGMA=0.2



# EXAMPLE SIGMA=2



# 2つのコインがある場合

前提：2つのコインフリップは独立である

likelihood

$$\begin{aligned} p(D|\theta_1, \theta_2) &= \prod_{coin1} p(y_i|\theta_1, \theta_2) \prod_{coin2} p(y_j|\theta_1, \theta_2) \\ &= \theta_1^{z_1} (1 - \theta_1)^{(N_1 - z_1)} \theta_2^{z_2} (1 - \theta_2)^{(N_2 - z_2)} p(\theta_1, \theta_2) \end{aligned}$$

$$p(\theta_1, \theta_2|D) = \frac{p(D|\theta_1, \theta_2)p(\theta_1, \theta_2)}{p(D)} = \frac{p(D|\theta_1, \theta_2)p(\theta_1, \theta_2)}{\int \int p(D|\theta_1, \theta_2)p(\theta_1, \theta_2)d\theta_1 d\theta_2}$$

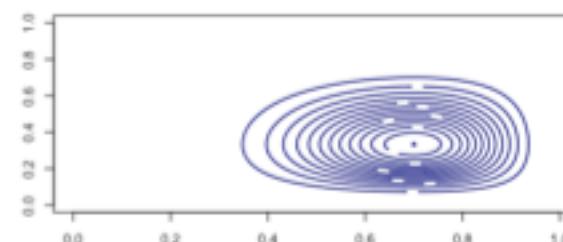
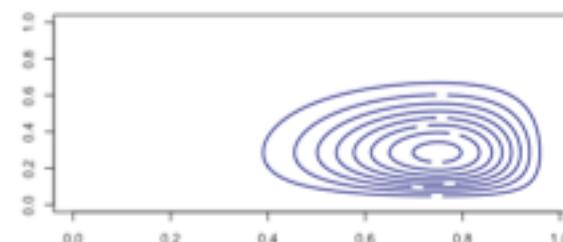
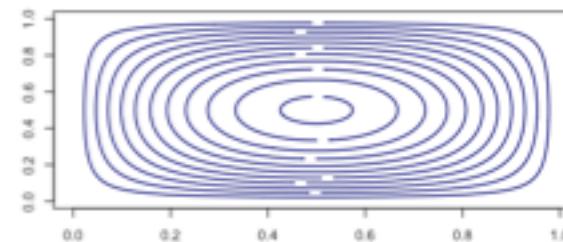
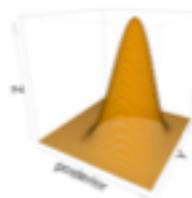
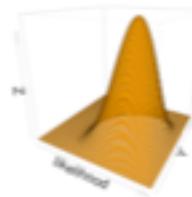
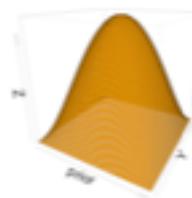
# 分析的：POSTERIOR

$$\begin{aligned} p(\theta_1, \theta_2 | D) &= \frac{p(D|\theta_1, \theta_2)p(\theta_1, \theta_2)}{p(D)} \\ &= \frac{\theta_1^{z_1}(1-\theta_1)^{(N_1-z_1)}\theta_2^{z_2}(1-\theta_2)^{(N_2-z_2)}p(\theta_1, \theta_2)}{p(D)} \\ &= \frac{\theta_1^{z_1}(1-\theta_1)^{(N_1-z_1)}\theta_2^{z_2}(1-\theta_2)^{(N_2-z_2)}\theta_1^{a_1-1}(1-\theta_1)^{(b_1-1)}\theta_2^{a_2-1}(1-\theta_2)^{(b_2-1)}}{p(D)B(a_1, b_1)B(a_2, b_2)} \end{aligned}$$

$$p(D)B(a_1, b_1)B(a_2, b_2) = B(z_1 + a_1, N_1 - z_1 + b_1)B(z_2 + a_2, N_2 - z_2 + b_2)$$

# ANALYTICAL SOLUTION

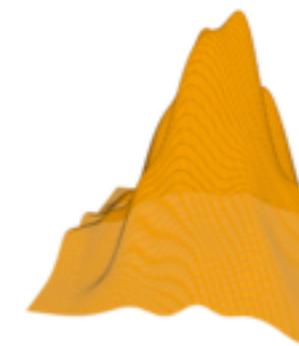
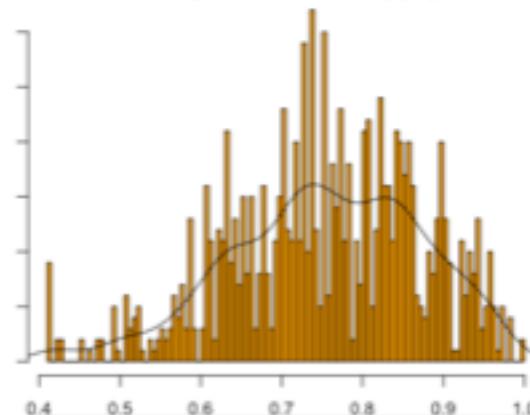
$z_1 = 6, N_1 = 8$   
 $z_2 = 2, N_2 = 7$



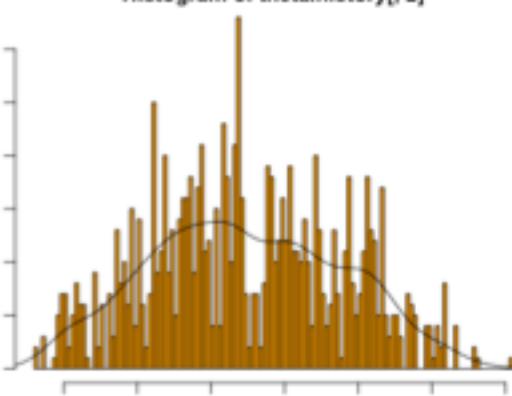
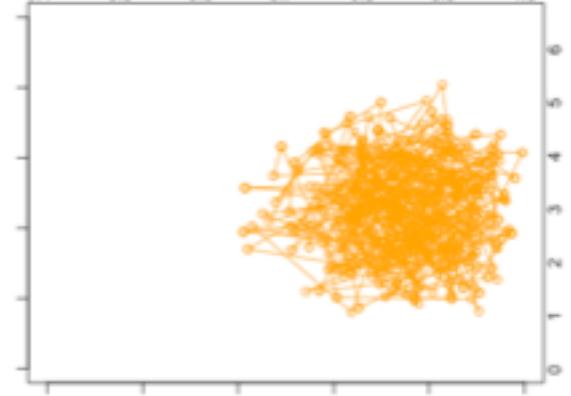
# METROPOLIS

$z_1 = 6, N_1 = 8$   
 $z_2 = 2, N_2 = 7$

Histogram of theta.history[, 1]



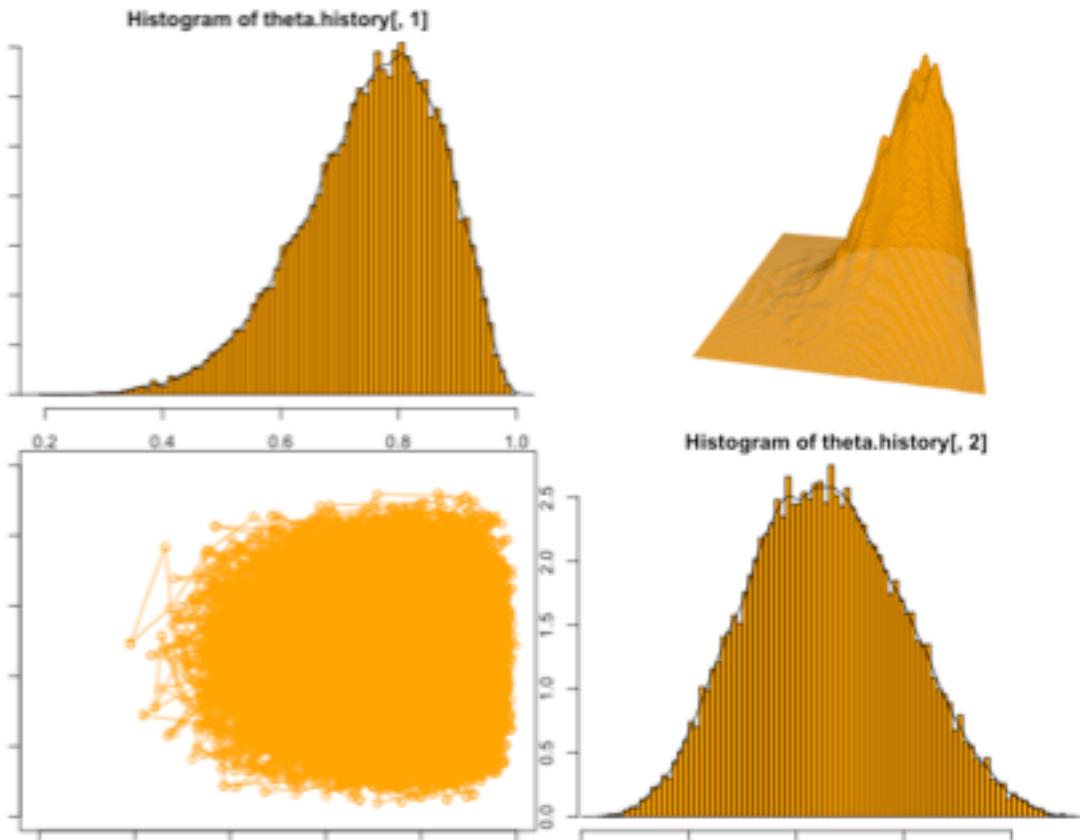
Histogram of theta.history[, 2]



1000 steps

# METROPOLIS

$z_1 = 6, N_1 = 8$   
 $z_2 = 2, N_2 = 7$



100000 steps

# GIBBS SAMPLING

Metropolis – 全てのパラメタを一括採択・棄却

- proposed all parameters is accepted or rejected together

$$p(\theta_1, \theta_2 | D) = \frac{p(D|\theta_1, \theta_2)p(\theta_1, \theta_2)}{p(D)} = \frac{\theta_1^{z_1}(1-\theta_1)^{(N_1-z_1)}\theta_2^{z_2}(1-\theta_2)^{(N_2-z_2)}\theta_1^{a_1-1}(1-\theta_1)^{(b_1-1)}\theta_2^{a_2-1}(1-\theta_2)^{(b_2-1)}}{p(D)B(a_1, b_1)B(a_2, b_2)}$$

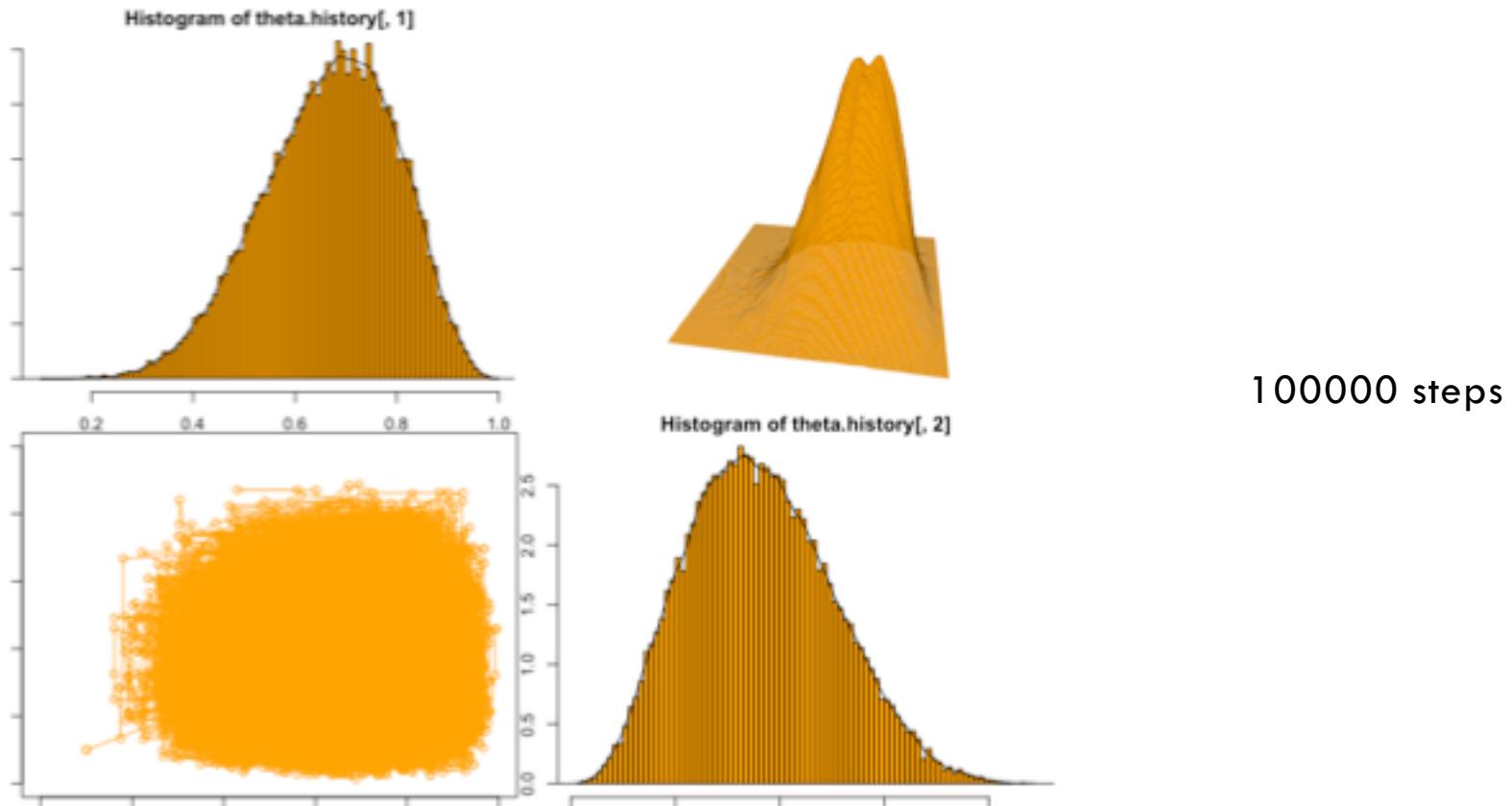
Gibbs sample – 1つ1つを個別に採択・棄却

- 提案されたパラメタは他のパラメタの値で条件づけて採択・棄却  $p(\theta_i, |\{\theta_{j \neq i}\}, D)$

$$\begin{aligned} p(\theta_1 | \theta_2, D) &= \frac{p(\theta_1, \theta_2 | D)}{p(\theta_2 | D)} = \frac{p(\theta_1, \theta_2 | D)}{\int p(\theta_1, \theta_2 | D) d\theta_1} \\ &= \frac{\text{beta}(\theta_1 | z_1 + a_1, N_1 - z_1 + b_1) \text{beta}(\theta_2 | z_2 + a_2, N_2 - z_2 + b_2)}{\int \text{beta}(\theta_1 | z_1 + a_1, N_1 - z_1 + b_1) \text{beta}(\theta_2 | z_2 + a_2, N_2 - z_2 + b_2) d\theta_1} \\ &= \frac{\text{beta}(\theta_1 | z_1 + a_1, N_1 - z_1 + b_1) \text{beta}(\theta_2 | z_2 + a_2, N_2 - z_2 + b_2)}{\text{beta}(\theta_2 | z_2 + a_2, N_2 - z_2 + b_2) \int \text{beta}(\theta_1 | z_1 + a_1, N_1 - z_1 + b_1) d\theta_1} \\ &= \text{beta}(\theta_1 | z_1 + a_1, N_1 - z_1 + b_1) \end{aligned}$$

$$p(\theta_2 | \theta_1, D) = \text{beta}(\theta_2 | z_2 + a_2, N_2 - z_2 + b_2)$$

# 実習4 GIBBS SAMPLING



# 3 GOALS OF MCMC

## Posteriorの適正

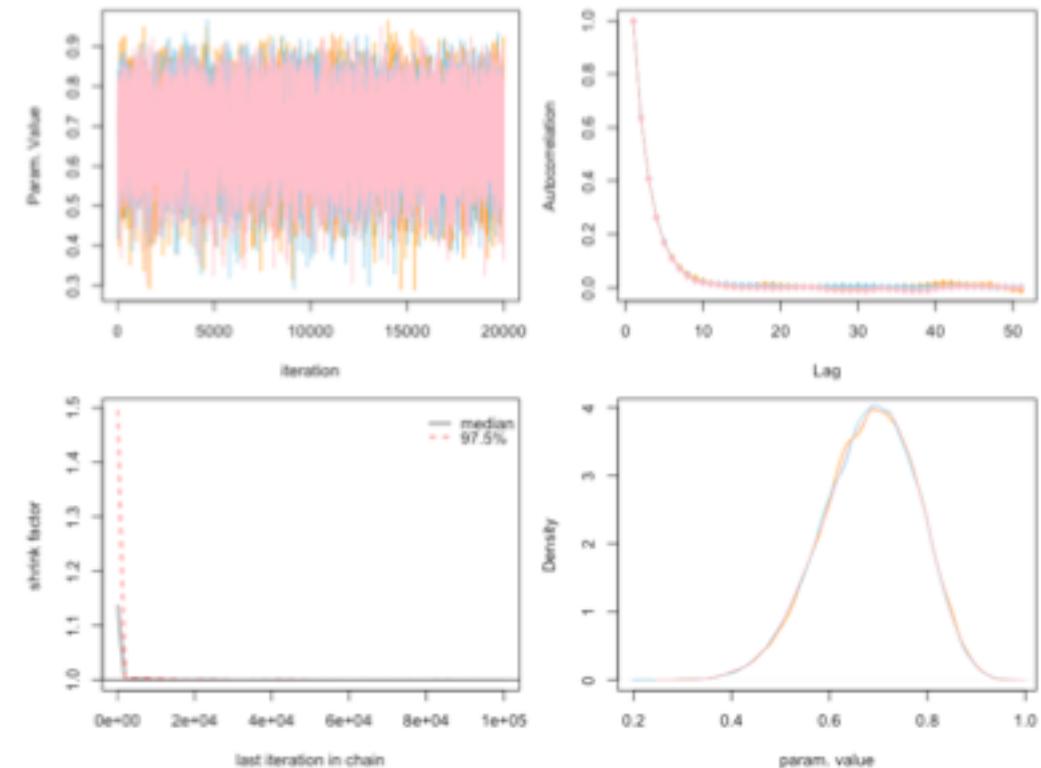
- 初期値に依存しなく、十分な探索が必要
- 複数のMCMCの結果を比較
- Shrink factor (chain内分散とchain間分散)

## 正確性と安定性

- Autocorrelationをチェックする
- Effective sample size (ESS)
  - サンプル数をautocorrelationの値で補正
- Compare HDIs
- Monte Carlo Standard Error:  $SD/\sqrt{ESS}$

## 効率

- Parallel processing
- sampling methodを修正
- Modelsの修正



# 3 GOALS OF MCMC

## Posteriorの適正

- 初期値に依存しなく、十分な探索が必要
- 複数のMCMCの結果を比較
- Shrink factor (chain内分散とchain間分散)

## 正確性と安定性

- Autocorrelationをチェックする
- Effective sample size (ESS)
  - サンプル数をautocorrelationの値で補正
- Compare HDIs
- Monte Carlo Standard Error:  $SD/\sqrt{ESS}$

## 効率

- Parallel processing
- sampling methodを修正
- Modelsの修正

