

メトロポリス・ヘイスティングス 法

Sihan Wang

6/4/2021

マルコフ連鎖: 定常分布への収束

遷移行列 \mathbf{T} を知っている状態で, どんな定常分布 $p(\cdot)$ に収束するのか求めたい.

遷移行列

$$\mathbf{T} = \begin{pmatrix} p(X^{(t+1)} = \text{紋} | X^{(t)} = \text{紋}) & p(X^{(t+1)} = \text{紋} | X^{(t)} = \text{縞}) & p(X^{(t+1)} = \text{紋} | X^{(t)} = \text{玉}) \\ p(X^{(t+1)} = \text{縞} | X^{(t)} = \text{紋}) & p(X^{(t+1)} = \text{縞} | X^{(t)} = \text{縞}) & p(X^{(t+1)} = \text{縞} | X^{(t)} = \text{玉}) \\ p(X^{(t+1)} = \text{玉} | X^{(t)} = \text{紋}) & p(X^{(t+1)} = \text{玉} | X^{(t)} = \text{縞}) & p(X^{(t+1)} = \text{玉} | X^{(t)} = \text{玉}) \end{pmatrix}$$

マルコフ連鎖

$$\mathbf{p}^{(1)} = (p(X^{(1)} = \text{紋}) \quad p(X^{(1)} = \text{縞}) \quad p(X^{(1)} = \text{玉}))$$

$$\mathbf{p}^{(2)} = \mathbf{p}^{(1)} \mathbf{T}$$

$$\mathbf{p}^{(t+1)} = \mathbf{p}^{(t)} \mathbf{T}$$

要素ごとにみると

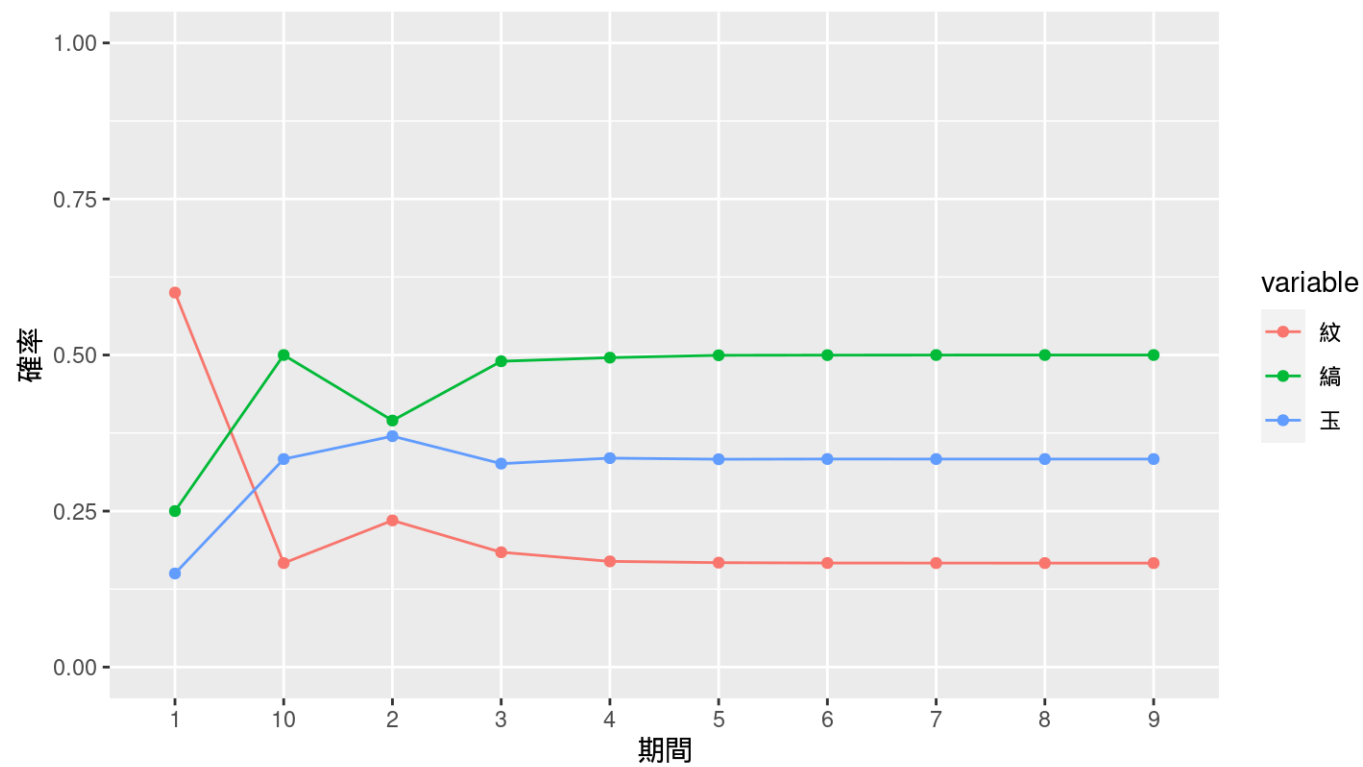
$$\begin{aligned}\mathbf{p}^{(t+1)} &= \mathbf{p}^{(t)} \mathbf{T} \\ p(X^{(t+1)} = \text{紋}) &= p(X^{(t+1)} = \text{紋} | X^{(t)} = \text{紋})p(X^{(t)} = \text{紋}) + \\ &\quad p(X^{(t+1)} = \text{紋} | X^{(t)} = \text{縞})p(X^{(t)} = \text{縞}) + \\ &\quad p(X^{(t+1)} = \text{紋} | X^{(t)} = \text{玉})p(X^{(t)} = \text{玉}) \\ p(X^{(t+1)} = j) &= \sum_{i=1}^3 p(X^{(t+1)} = j | X^{(t)} = i) p(X^{(t)} = i)\end{aligned}$$

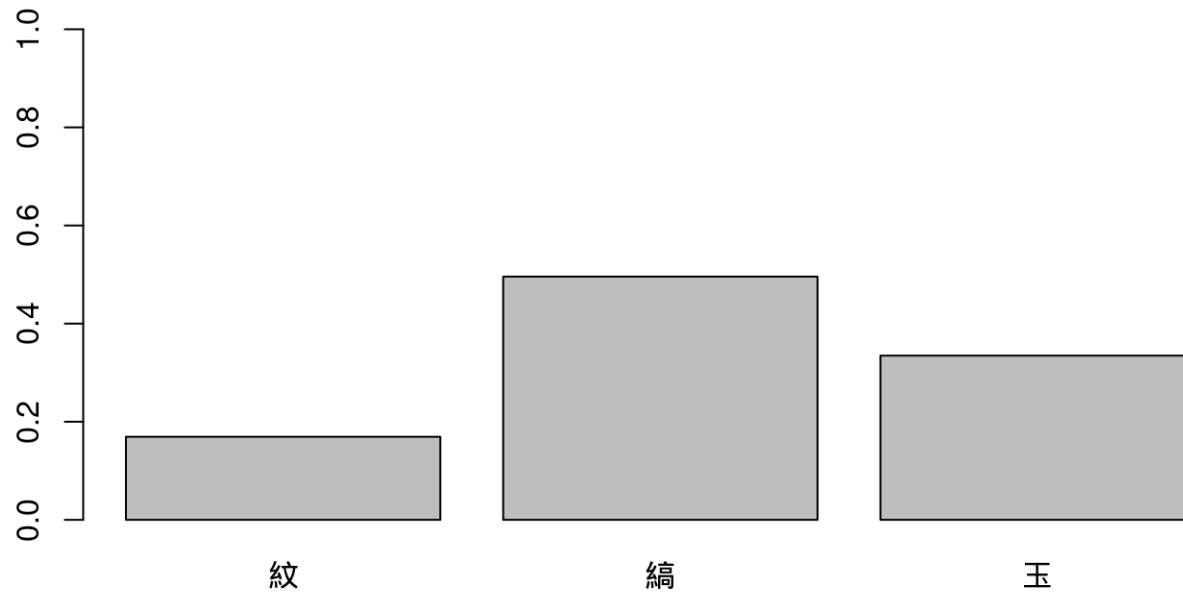
数值例:

$$\mathbf{T} = \begin{pmatrix} 0.3 & 0.3 & 0.4 \\ 0.1 & 0.5 & 0.4 \\ 0.2 & 0.6 & 0.2 \end{pmatrix}$$
$$\mathbf{p}^{(1)} = (0.6 \quad 0.25 \quad 0.15)$$

##		紋	縞	玉	t
## 1	0.6000000	0.2500000	0.1500000		1
## 2	0.2350000	0.3950000	0.3700000		2
## 3	0.1840000	0.4900000	0.3260000		3
## 4	0.1694000	0.4958000	0.3348000		4
## 5	0.1673600	0.4996000	0.3330400		5
## 6	0.1667760	0.4998320	0.3333920		6
## 7	0.1666944	0.4999840	0.3333216		7
## 8	0.1666710	0.4999933	0.3333357		8
## 9	0.1666678	0.4999994	0.3333329		9
## 10	0.1666668	0.4999997	0.3333334		10

マルコフ連鎖 確率分布の推移





定常分布は $\bar{\mathbf{p}} = (\bar{p}(X = \text{紋}), \bar{p}(X = \text{縞}), \bar{p}(X = \text{玉})) = (\frac{1}{6}, \frac{1}{2}, \frac{1}{3})$ くらいですね.

マルコフ連鎖: 詳細釣り合い条件

収束する先である定常分布 $\bar{p}(\cdot)$ を知っている状態で, どんな推移行列 \mathbb{T} かを求めたい.
(\mathbb{T} がわかれば, $\bar{p}(\cdot)$ に従う乱数が作れる!)

定常分布 $\bar{p}(\cdot)$ に収束している状態を数式で表すと,

$$\bar{p}(X' = j) = \sum_i \Pr(X' = j|X = i)\bar{p}(X = i) \text{ for all } j$$

$$(\bar{p}(X = i) = \sum_j \Pr(X = i|X' = j)\bar{p}(X' = j) \text{ for all } i)$$

なじみのある行列で書くと,

$$\mathbf{p}^{(\infty)} = \mathbf{p}^{(\infty)} \mathbf{T}$$

$$\overbrace{\bar{p}(X' = j)}^{\text{確率分布}\bar{p}(\cdot)\text{は一定}} = \sum_i \overbrace{\Pr(X' = j|X = i)}^{\text{未知の遷移核}} \overbrace{\bar{p}(X = i)}^{\text{確率分布}\bar{p}(\cdot)\text{は一定}} \quad \text{for all } j$$

行列で表すと

$$\overbrace{\bar{\mathbf{p}}}^{\text{確率分布一定}} = \overbrace{\bar{\mathbf{p}}}^{\text{確率分布一定}} \overbrace{\mathbf{T}}^{\text{未知の遷移行列}}$$

マルコフ過程が収束するための十分条件は,

$$\Pr(X = i|X' = j)\bar{p}(X' = j) = \Pr(X' = j|X = i)\bar{p}(X = i)$$

これを**詳細釣り合い条件**といいます.

証明は以下の通りです.

$$\Pr(X = i|X' = j)\bar{p}(X' = j) = \Pr(X' = j|X = i)\bar{p}(X = i)$$

両辺を*i*に関して和をとる

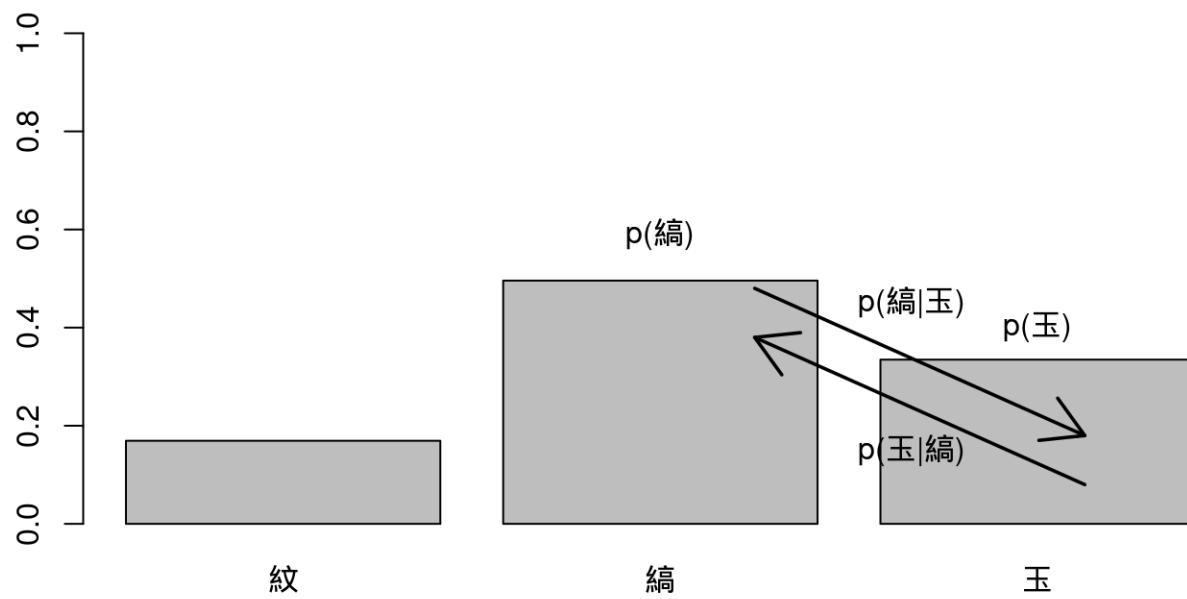
$$\sum_i \Pr(X = i|X' = j)\bar{p}(X' = j) = \sum_i \Pr(X' = j|X = i)\bar{p}(X = i)$$

$$\bar{p}(X' = j) \overbrace{\sum_i \Pr(X = i|X' = j)}^{\text{周辺化して1}} = \sum_i \Pr(X' = j|X = i)\bar{p}(X = i)$$
$$\bar{p}(X' = j) = \sum_i \Pr(X' = j|X = i)\bar{p}(X = i)$$

詳細釣り合い条件の意味は, 収束中1ステップあたりの

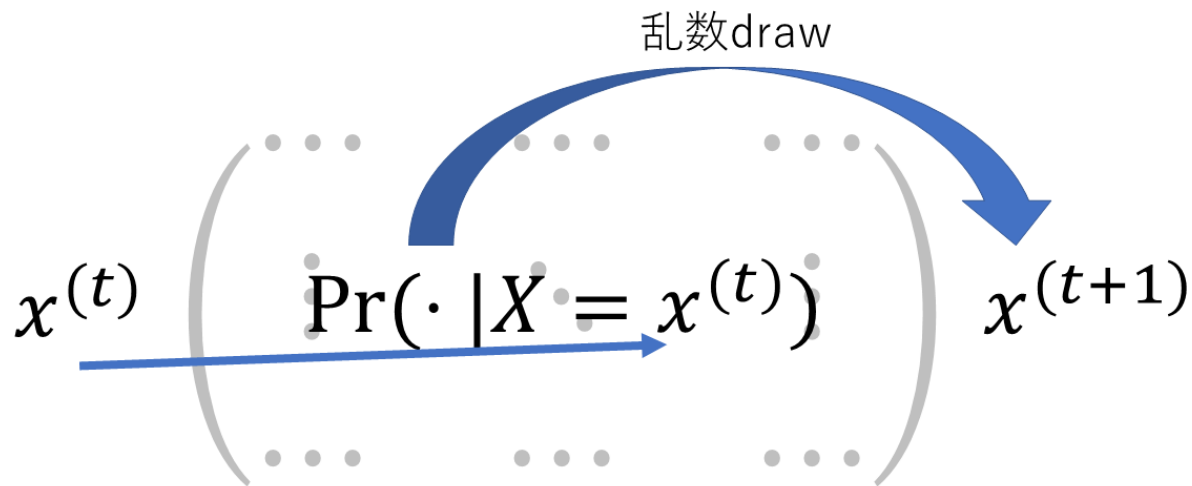
「今期 $X = i \rightarrow$ 1期後 $X' = j$ に遷移する量 $p(X' = j|X = i)p(X = i)$ 」と「1期前 $X' = j \rightarrow$ 今期 $X = i$ に遷移する量 $p(X = i|X' = j)p(X' = j)$ 」が等しい

$$\overbrace{\Pr(X = i|X' = j)\bar{p}(X' = j)}^{j \rightarrow i \text{の移動量}} = \overbrace{\Pr(X' = j|X = i)\bar{p}(X = i)}^{i \rightarrow j \text{の移動量}}$$



マルコフ連鎖: 定常分布の乱数発生

- 各要素が詳細釣り合い条件 $\Pr(X = i | X' = j) \bar{p}(X' = j) = \Pr(X' = j | X = i) \bar{p}(X = i)$ を満たす遷移核である遷移行列 \mathbf{A} を発見する.
- $x^{(t)}$ を遷移行列 \mathbf{T} にくぐらせる(条件付確率分布 $\Pr(\cdot | X = x^{(t)})$ から乱数発生する) と定常分布 $\bar{p}(\cdot)$ に従った $x^{(t+1)}$ が出てくる. (初期値は定義域なら何でもok)
- 更に $x^{(t+1)}$ を遷移行列 \mathbf{T} にくぐらせると... というふうに繰り返すと, 定常分布 $\bar{p}(\cdot)$ に従った状態が乱数列 $\mathbf{x} = \{x_0, x_1, \dots\}$ が得られる.



Markov Chain 乱数発生イメージ

$t = 1$ のとき

3

$\Pr(\cdot | x^{(0)} = j)$ から
乱数 $x^{(1)} = i$ を draw

Detailed balanced T

	$X' = i$	$X \sim j$	$X \sim k$
$X = i$	\sim		
$X = j$	0.1	0.5	0.4
$X = k$	0.2	0.6	0.2

2

遷移核(条件付確率分布)
 $\Pr(\cdot | x^{(0)} = j)$ を特定

1

(一様分布などから draw)
初期値 $x^{(0)} = j$ を用いて

4

サンプリング結果
 $x = \{j, i\}$

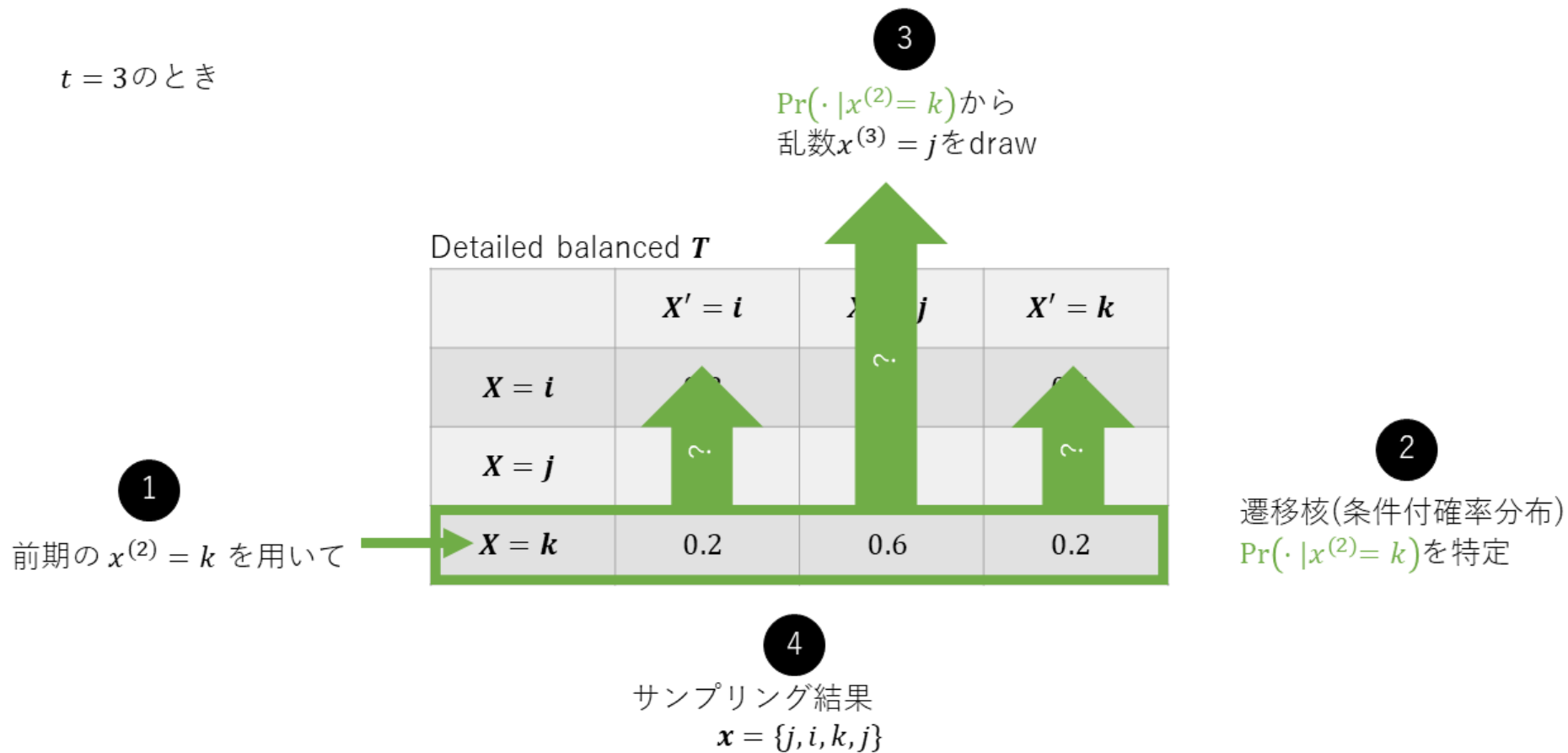
Markov Chain 乱数発生イメージ

$t = 2$ のとき



Markov Chain 乱数発生イメージ

$t = 3$ のとき



<例題>

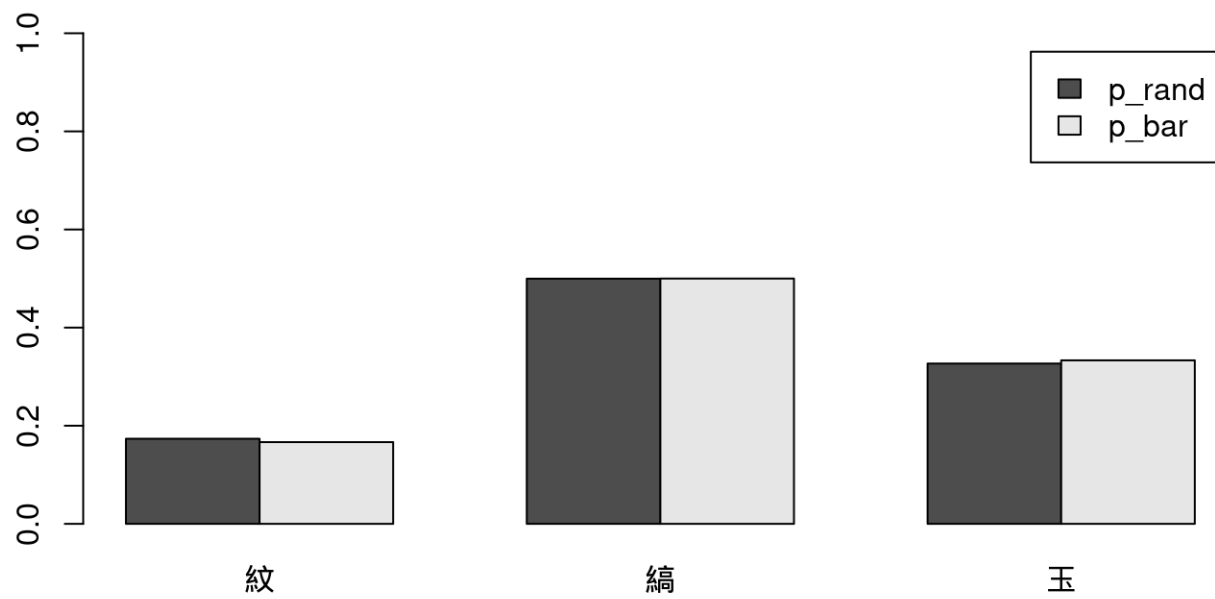
定常分布は $\bar{\mathbf{p}} = (\frac{1}{6}, \frac{1}{2}, \frac{1}{3})$ でした. 定常分布を元に詳細釣り合い条件を満たす遷移核を何とかして見つけておきました. 確かめてみてください.

	紋	縞	玉
紋	0.3	0.3	0.4
縞	0.1	0.5	0.4
玉	0.2	0.6	0.2

この遷移核を用いて, 定常分布に従う乱数を作ってみましょう.

```
## [1] "Generated Random Numbers"
```

```
##      [1] 1 3 2 3 1 3 1 1 2 3 2 2 2 3 2 3 2 3 1 3 3 1 3 3 2 2 2 2 3 2 2 3 2 2 3
##     [37] 3 3 2 3 2 3 3 3 2 3 3 2 2 3 3 2 3 2 2 2 2 2 3 3 2 1 3 2 2 3 2 2 3 2 3 2
##     [73] 3 2 2 2 3 1 3 3 1 2 3 2 2 3 2 3 2 2 2 2 2 3 1 1 1 2 2 3 3 3 2 2 1 2 2 2
##    [109] 2 1 2 1 1 3 2 2 2 3 2 2 3 1 2 2 2 3 2 3 2 2 3 2 2 2 3 1 2 3 2 1 2 3 3 2
##    [145] 2 3 2 2 3 2 3 3 2 2 2 3 2 3 2 2 2 3 2 3 1 2 2 2 3 2 3 1 1 3 2 3 3 3 3
##    [181] 1 3 2 3 2 3 2 3 3 1 2 3 2 2 1 3 2 2 3 2 3 2 2 3 2 2 3 2 2 2 3 1 3 3 1 1
##    [217] 3 3 1 1 3 2 2 2 3 1 2 2 2 2 3 2 2 2 3 2 3 3 2 2 2 2 3 1 3 2 2 2 2 1 1
##    [253] 1 2 3 2 2 2 2 2 1 2 2 2 3 3 2 2 3 3 2 2 2 2 2 2 2 2 2 2 3 2 3 1 3 2 2 2
##    [289] 2 2 2 2 2 1 3 2 3 2 2 2 1 2 2 2 2 2 1 3 1 3 2 2 2 3 3 2 2 2 2 1 2 3 2 2
##    [325] 1 1 1 1 1 3 3 1 3 2 3 2 2 2 2 3 1 1 3 2 2 3 2 2 2 2 3 2 2 3 2 1 3 1 3 1
##    [361] 2 2 2 3 2 3 1 3 3 3 2 2 3 1 3 2 2 2 3 2 2 2 3 3 1 3 2 3 1 3 1 3 2 1 2 2
##    [397] 2 3 2 2 2 3 2 1 1 1 1 3 3 2 1 1 3 3 3 3 2 2 2 3 1 2 3 3 2 2 2 2 1 1 1 1
##    [433] 3 2 3 2 3 3 2 2 1 3 2 3 3 2 2 2 1 3 2 1 2 2 2 2 3 2 1 2 2 2 2 2 2 1 1
##    [469] 3 2 3 2 1 1 3 2 2 3 2 2 3 3 2 3 1 2 3 2 2 3 2 2 1 3 2 2 2 2 3 1 3 2 3 3
##    [505] 3 2 1 2 3 3 2 3 2 2 3 3 2 3 2 2 3 1 1 3 2 2 1 1 1 1 1 2 2 2 2 2 2 2 2
##    [541] 3 2 1 2 2 2 2 1 3 2 3 1 2 2 1 1 3 2 2 3 1 2 2 2 3 1 2 2 3 2 1 3 2 2 2 3
##    [577] 2 3 2 3 3 2 1 2 2 3 2 2 3 1 1 2 2 3 3 1 2 2 2 2 2 3 1 3 3 1 1 3 2 1 3 1
##    [613] 3 2 3 3 2 2 2 3 2 3 2 3 2 2 3 2 3 2 1 1 3 2 3 2 2 2 3 3 1 3 1 3 3 2 2
##    [649] 3 3 1 2 3 2 2 3 2 2 3 3 1 3 3 2 3 2 3 1 3 2 3 3 2 3 2 3 2 2 2 2 3 3 2
##    [685] 3 2 3 2 2 2 1 3 2 3 2 2 3 3 2 1 2 1 1 1 2 3 2 2 3 1 1 1 2 2 3 2 2 2 2
##    [721] 3 2 3 1 1 3 1 3 2 2 2 3 2 3 2 3 2 2 3 2 2 2 3 3 2 2 2 3 2 3 2 1 2 2 2
##    [757] 2 2 2 1 2 3 2 3 2 2 2 3 3 1 1 2 3 2 1 2 2 2 3 2 1 1 1 1 2 3 2 3 2 2 2
##    [793] 2 1 3 2 3 2 3 1 3 1 3 1 1 3 3 2 1 3 2 2 2 2 1 3 2 1 1 3 3 2 3 3 1 1 3 2
##    [829] 2 2 3 3 1 1 1 2 3 2 3 1 2 1 3 2 2 2 2 2 3 3 2 1 1 3 1 3 2 2 3 2 2 2 3
##    [865] 2 2 2 2 3 3 2 2 1 1 3 2 2 2 2 3 1 3 1 1 1 2 3 1 2 3 2 2 2 2 2 2 3 2 2
##    [901] 2 3 3 2 2 2 2 2 2 2 2 3 3 3 2 2 2 2 2 3 3 1 3 1 1 3 1 1 1 3 2 2 3 1 3 2 3
##    [937] 3 1 2 3 2 2 1 3 2 2 2 3 3 1 1 3 2 3 1 2 2 3 2 2 2 2 2 2 3 2 2 2 3 1 1 1
##    [973] 2 2 1 3 2 2 3 3 1 3 2 1 3 2 2 3 2 1 2 2 2 3 2 2 3 3 2 2 2 3 3 2 1 3 2 3
##   [1009] 2 3 3 1 2 2 1 2 3 1 3 2 2 2 3 2 3 2 3 2 2 2 3 2 2 2 1 2 2 2 3 1 3 2 2 3
##   [1045] 2 2 2 3 2 2 1 3 2 2 1 1 3 3 2 3 2 2 1 1 3 2 1 3 2 1 2 2 2 3 2 1 2 1 3 2
##   [1081] 2 2 2 3 2 3 1 2 2 1 2 2 3 2 2 3 3 2 3 3 1 2 2 2 2 2 3 2 2 1 3 2 3 2 3 3
##   [1117] 2 2 3 2 3 2 3 2 3 3 1 3 2 2 2 3 2 2 2 2 2 2 3 1 2 2 2 3 2 3 2 2 2 3 2 3
##   [1153] 2 3 1 1 3 2 2 2 2 2 2 3 2 1 3 1 3 3 2 3 3 2 2 3 2 3 3 2 2 2 2 2 2 1 3 2
##   [1189] 3 2 2 3 1 3 3 2 1 1 1 3 2 2 3 1 2 3 1 1 1 2 3 2 3 2 2 2 1 1 3 2 2 3 2 3
```



詳細釣り合い条件を満たしている遷移行列 \mathbf{T} がわからない状態で,
どうやって定常分布にしたがう乱数を発生したらいいのでしょうか?

$$\overbrace{\Pr(X = i | X' = j)}^? \bar{p}(X' = j) = \overbrace{\Pr(X' = j | X = i)}^? \bar{p}(X = i)$$

メトロポリス・ヘイスティングス法 (Metropolis-Hastings: MH))

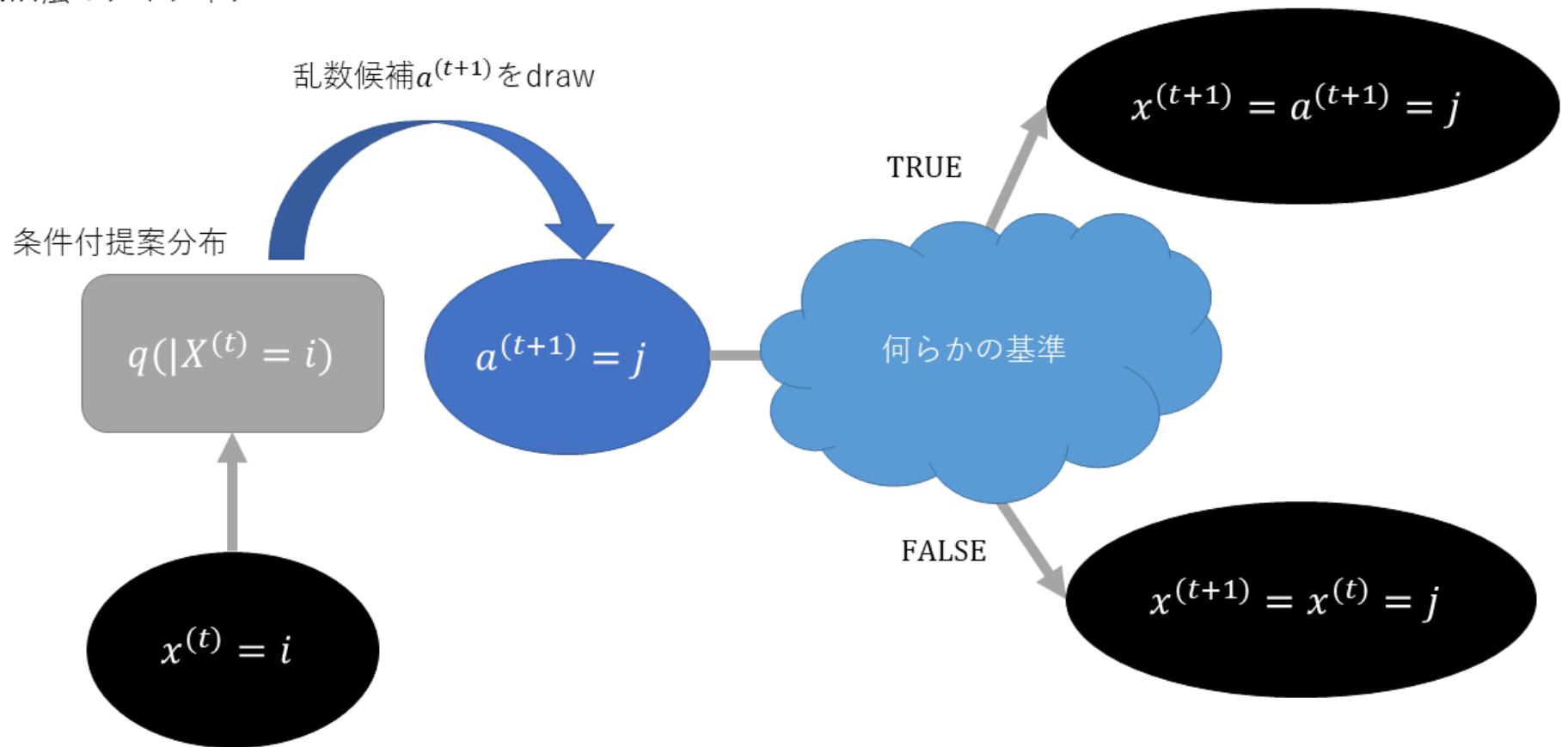
遷移核 $\Pr(|X = i)$ からdrawする代わりに,
適当な乱数生成が容易な分布(**提案分布**) $q(|X = i)$ から
乱数候補 $a^{(t+1)}$ をdrawします.
(遷移核も, 行列 A の各行で切ってみると条件付確率**分布**です)

drawした乱数候補 $a^{(t+1)}$ について, “ある基準”を満たせば採用し,
満たさなければ破棄(マルコフ連鎖なので前の状態を維持)します.

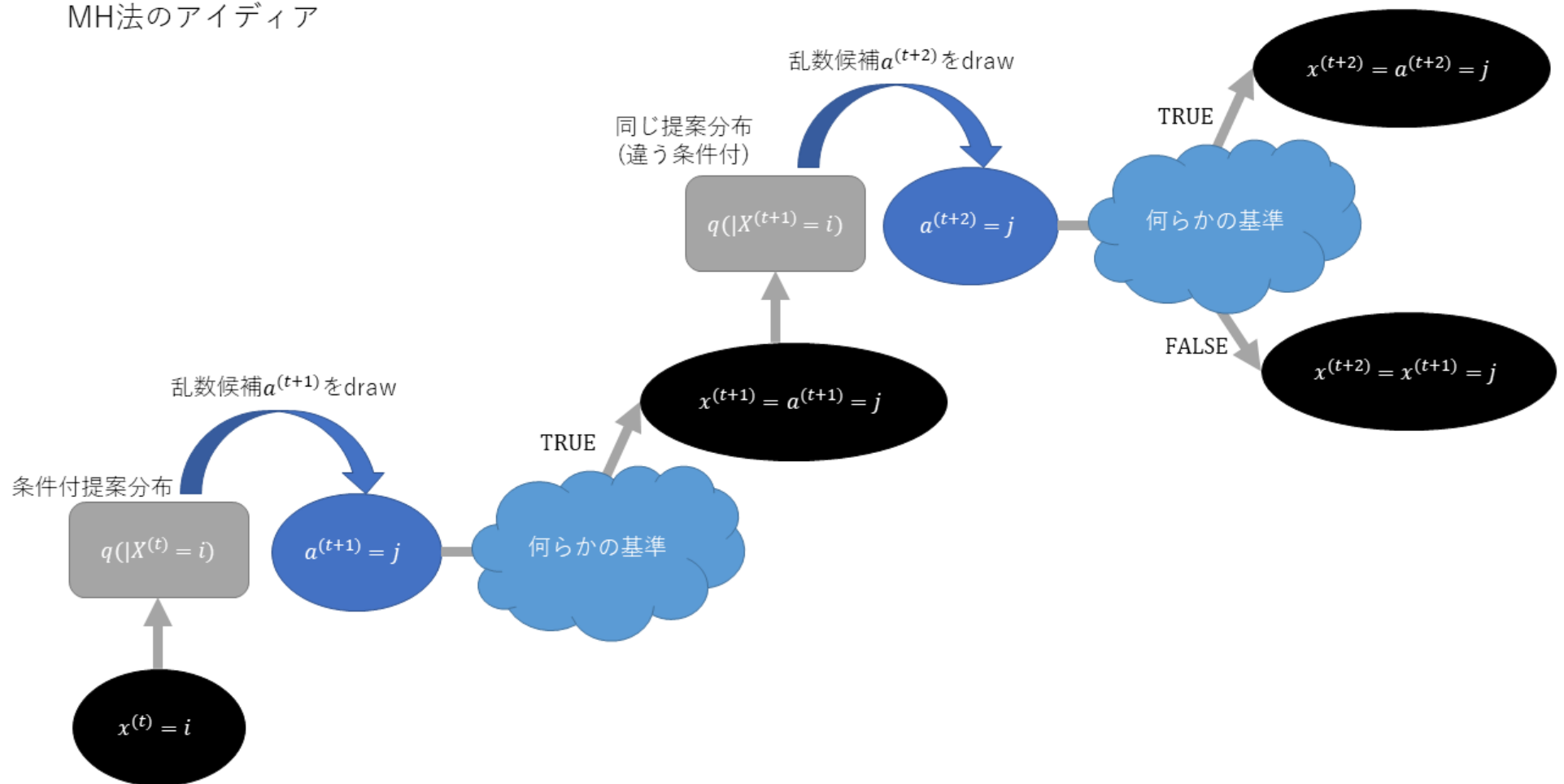
そうやって採用と破棄を繰り返せば, 定常分布(**目的分布**)に従う乱数が作れそうです.

これがMH法のざっくりしたアイデアとなります.

MH法のアイデア



MH法のアイデア



もっと正確にいうと,

drawした乱数候補 $a^{(t+1)}$ について, “ある基準”に基づいて, (マルコフ連鎖は確率的遷移なので)一定確率 A で採用し, 一定確率 $1 - A$ で前の状態を維持します.

このように, もとの遷移にさらに確率的な要素を加えて補正することを確率補正といいます.

それでは, MH法のポイントである“ある基準”が何なのか見てみましょう.

“ある基準”とは, 「詳細釣り合い条件が成り立つ」かどうかです.

(遷移に詳細釣り合い条件さえ成り立てば目的の定常分布に収束するので乱数を好きなだけ作れます)

そこで詳細釣り合い条件が成り立つように以下の確率 $A(a^{(t+1)}, X^{(t)})$ で提案分布を確率補正します.

$$A(a^{(t+1)}, X^{(t)}) = \min \left(1, \frac{q(X^{(t)} | a^{(t+1)}) \bar{p}(a^{(t+1)})}{q(a^{(t+1)} | X^{(t)}) \bar{p}(X^{(t)})} \right)$$

確率 $A(a^{(t+1)}, X^{(t)})$!順番注意! を提案分布 $q(a^{(t+1)}|X^{(t)})$ と組み合わせて確率補正すると,
補正された遷移核 $q(a^{(t+1)}|X^{(t)}) A(a^{(t+1)}, X^{(t)})$ は詳細釣り合い条件を満たします.
従って確率補正した遷移は目的の定常分布に収束します.

<証明>

$$\begin{aligned}
 q(a^{(t+1)}|X^{(t)}) A(a^{(t+1)}, X^{(t)}) \bar{p}(X^{(t)}) &= q(a^{(t+1)}|X^{(t)}) \min \left(1, \frac{q(X^{(t)}|a^{(t+1)})\bar{p}(a^{(t+1)})}{q(a^{(t+1)}|X^{(t)})\bar{p}(X^{(t)})} \right) \bar{p}(X^{(t)}) \\
 &= \min \left(q(a^{(t+1)}\bar{p}(X^{(t)})|X^{(t)}), q(X^{(t)}|a^{(t+1)})\bar{p}(a^{(t+1)}) \right) \\
 &= q(X^{(t)}|a^{(t+1)}) \min \left(\frac{q(a^{(t+1)}\bar{p}(X^{(t)})|X^{(t)})}{q(X^{(t)}|a^{(t+1)})\bar{p}(a^{(t+1)})}, 1 \right) \bar{p}(a^{(t+1)}) \\
 &= q(X^{(t)}|a^{(t+1)}) A(X^{(t)}, a^{(t+1)}) \bar{p}(a^{(t+1)})
 \end{aligned}$$

時点の前後は一般的には関係ないので, $a^{(t+1)} = X', X^{(t)} = X$ とすると,

$$q(X'|X) A(X', X) \bar{p}(X) = q(X|X') A(X, X') \bar{p}(X')$$

MH法のイメージ

提案分布 $q(X'|X)$ を適当に乱数発生が容易な分布にする

提案分布 $q(X'|X)$

	$X' = i$	$X' = j$	$X' = k$	
$X = i$	1/3	1/3	1/3	$q(X' X = i)$
$X = j$	1/3	1/3	1/3	$q(X' X = j)$
$X = k$	1/3	1/3	1/3	$q(X' X = k)$

MH法のイメージ

$t = 1$ のとき

3

$q(\cdot | X = x^{(0)} = j)$ から
乱数候補 $a^{(1)} = i$ を draw

4

コイン投げ:
確率 $A(a^{(1)}, x^{(0)})$ で $x^{(1)} = a^{(1)} = i$
確率 $1 - A(a^{(1)}, x^{(0)})$ で $x^{(1)} = x^{(0)} = j$

提案分布 $q(X' | X)$

	$X' = i$	$X' = j$	$X' = k$
$X = i$	\sim	\sim	\sim
$X = j$	1/3	1/3	1/3
$X = k$	1/3	1/3	1/3

1

(一様分布などから draw)
初期値 $x^{(0)} = j$ を用いて

2

提案分布
 $q(\cdot | X = x^{(0)} = j)$ を特定

5

サンプリング結果
 $x = \{j, i\}$

MH法のイメージ

$t = 2$ のとき

3

$q(\cdot | X = x^{(1)} = i)$ から
乱数候補 $a^{(2)} = k$ をdraw

4

コイン投げ:
確率 $A(a^{(2)}, x^{(1)})$ で $x^{(2)} = a^{(2)} = k$
確率 $1 - A(a^{(2)}, x^{(1)})$ で $x^{(2)} = x^{(1)} = i$

1

前期 $x^{(1)} = i$ を用いて

提案分布 $q(X'|X)$

	$X \sim i$	$X \sim j$	$X \sim k$
$X = i$	1/3	1/3	1/3
$X = j$	1/3	1/3	1/3
$X = k$	4/7	2/7	1/7

2

提案分布
 $q(\cdot | X = x^{(1)} = i)$ を特定

5

サンプリング結果
 $\mathbf{x} = \{j, i, i\}$

MH法のイメージ

$t = 3$ のとき

3

$q(\cdot | X = x^{(2)} = i)$ から
乱数候補 $a^{(3)} = k$ を draw

4

コイン投げ:
確率 $A(a^{(3)}, x^{(2)})$ で $x^{(3)} = a^{(3)} = k$
確率 $1 - A(a^{(3)}, x^{(2)})$ で $x^{(3)} = x^{(2)} = i$

1

前期 $x^{(2)} = i$ を用いて

提案分布 $q(X'|X)$

	$X \sim i$	$X \sim j$	$X \sim k$
$X = i$	1/3	1/3	1/3
$X = j$	1/3	1/3	1/3
$X = k$	4/7	2/7	1/7

2

提案分布
 $q(\cdot | X = x^{(2)} = i)$ を特定

5

サンプリング結果
 $\mathbf{x} = \{j, i, i, k\}$

<例題> 定常分布は $\bar{\mathbf{p}} = (\frac{1}{6}, \frac{1}{2}, \frac{1}{3})$ でした. 詳細釣り合い条件を満たす遷移核はわかりません. なので適当に提案分布を一様分布にしてみます.

- ・ 提案分布

	紋	縞	玉
紋	1/3	1/3	1/3
縞	1/3	1/3	1/3
玉	1/3	1/3	1/3

MH法で定常分布 $\bar{\mathbf{p}}$ に従う乱数を発生させてみましょう

```
## [1] "acceptance rate"
```

```
## [1] 0.7785455
```

```
## [1] "Run Time"
```

```
## Time difference of 2.377194 secs
```

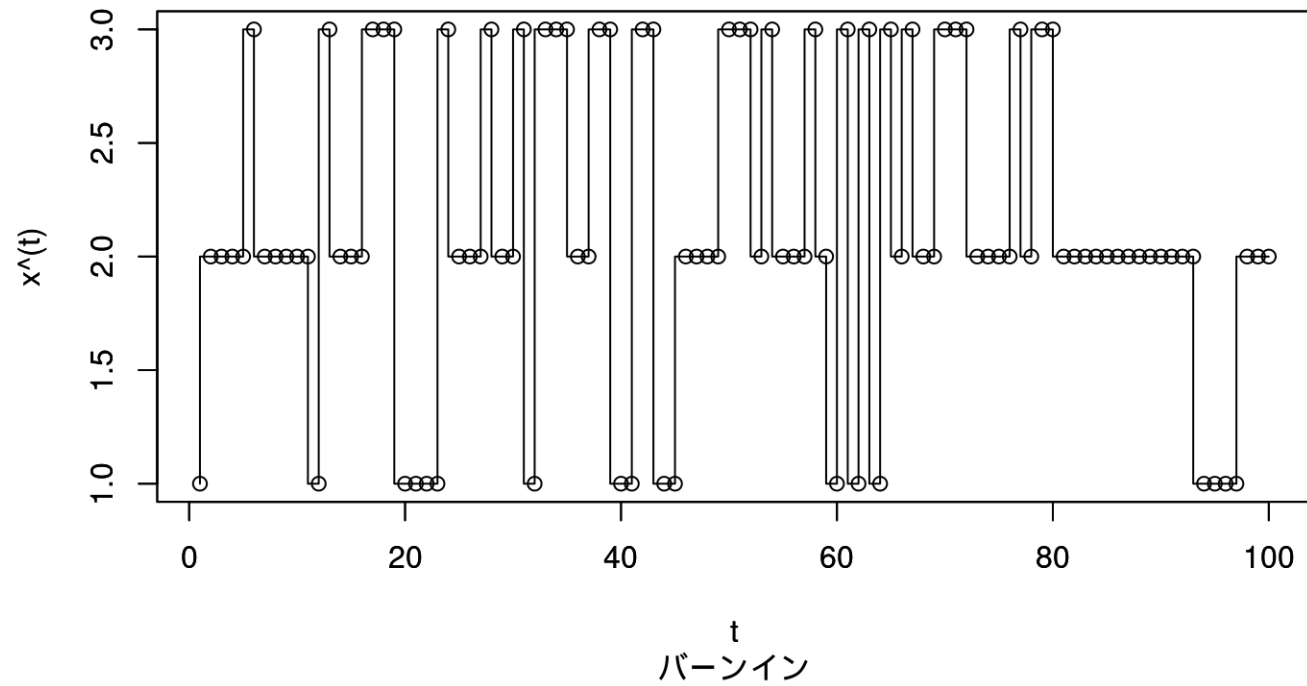
```
## [1] "sampled distribution"
```

```
## [1] 0.16751 0.49850 0.33399
```

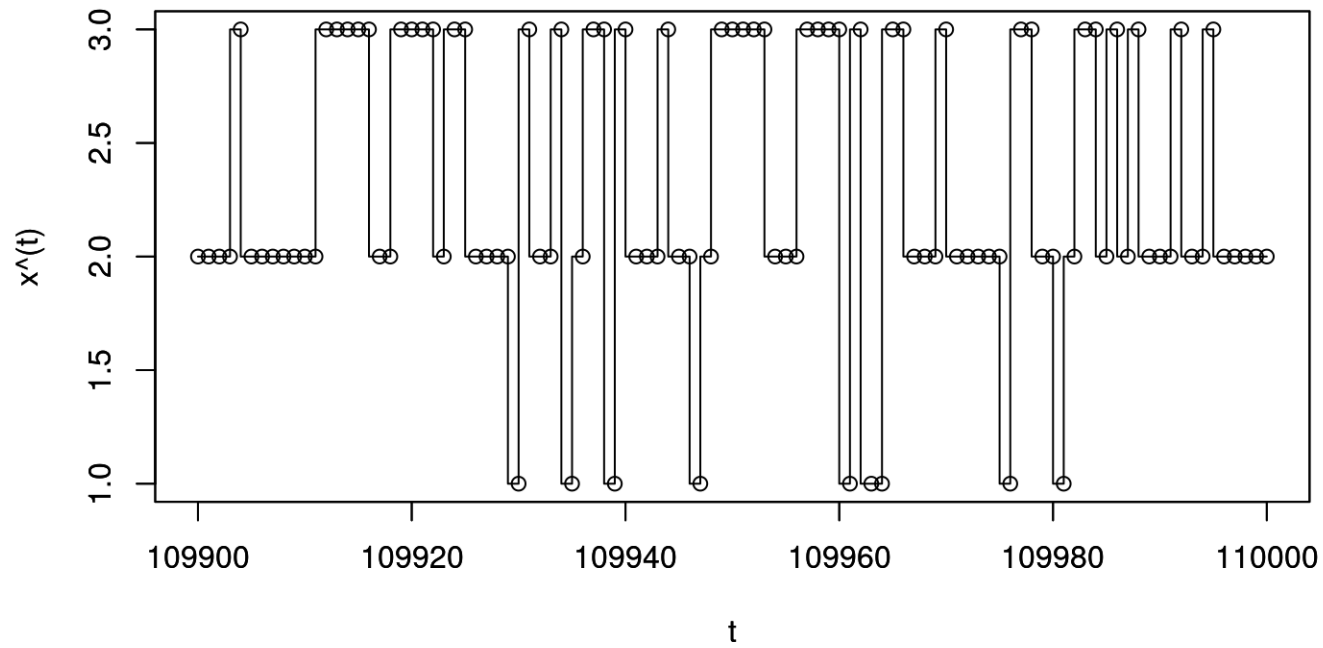
```
## [1] "invariant distribution"
```

```
## [1] 0.1666667 0.5000000 0.3333333
```

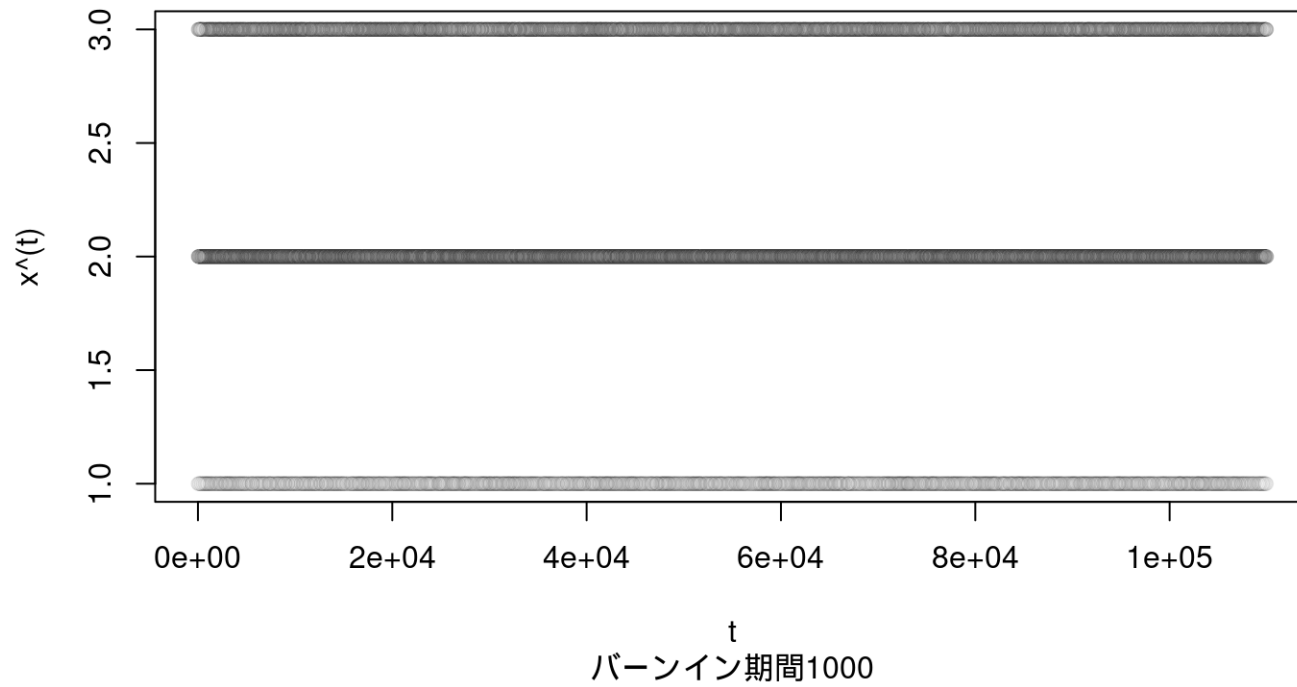
トレースプロット: $t=0\sim 100$

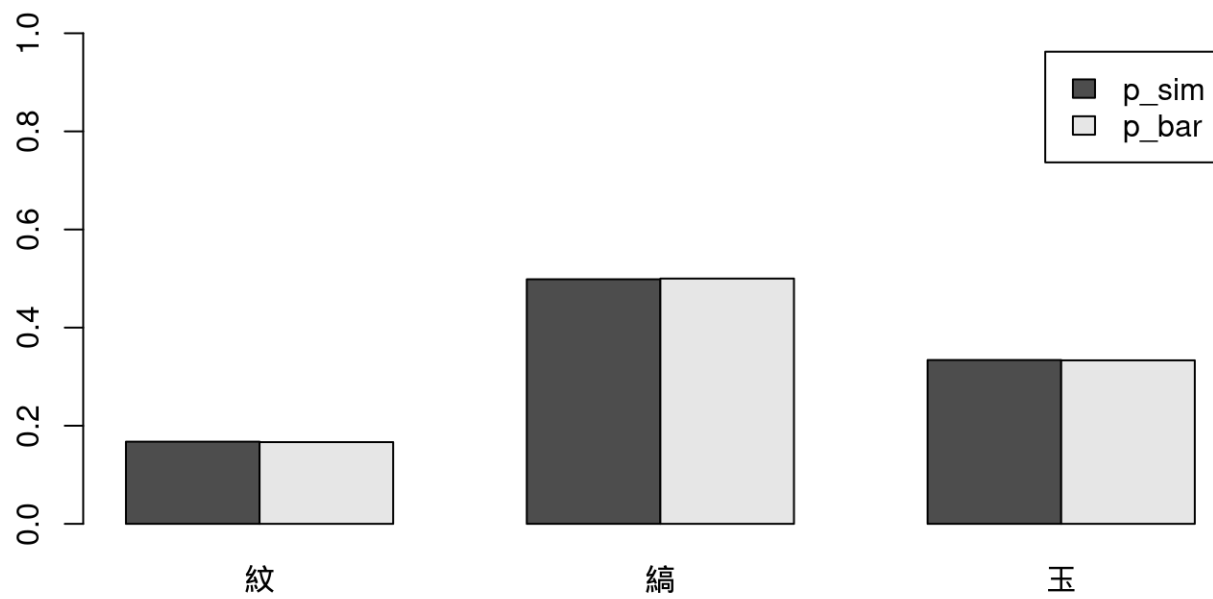


トレースプロット: t=99900~100000



トレースプロット





離散から連続へ

これまでは離散の例で, 確率変数 X のとりうる値は{紋, 縞, 玉}の3つのみでした.

私たちの目的は事後分布 $f(\theta|D)$ に従う連続の確率変数 θ (範囲は $(-\infty, \infty)$)の乱数を発生させることです.
離散を連続に置き換えていきましょう.

- ・ 定常分布

$$\bar{p}(X) \rightarrow f(\theta|D)$$

- ・ 提案分布

$$q(a^{(t+1)}|X^{(t)}) \rightarrow q(a^{(t+1)}|\theta^{(t)})$$

- ・ 補正確率

$$\begin{aligned}
A(a^{(t+1)}, X^{(t)}) &= \min \left(1, \frac{q(X^{(t)}|a^{(t+1)})\bar{p}(a^{(t+1)})}{q(a^{(t+1)}|X^{(t)})\bar{p}(X^{(t)})} \right) \\
&\quad \downarrow \\
A(a^{(t+1)}, \theta^{(t)}) &= \min \left(1, \frac{q(\theta^{(t)}|a^{(t+1)})f(a^{(t+1)}|D)}{q(a^{(t+1)}|\theta^{(t)})f(\theta^{(t)}|D)} \right) \\
&= \min \left(1, \frac{q(\theta^{(t)}|a^{(t+1)}) \frac{f(D|a^{(t+1)})f(a^{(t+1)})}{f(D)}}{q(a^{(t+1)}|\theta^{(t)}) \frac{f(D|\theta^{(t)})f(\theta^{(t)})}{f(D)}} \right) \\
&= \min \left(1, \frac{q(\theta^{(t)}|a^{(t+1)})f(D|a^{(t+1)})f(a^{(t+1)})}{q(a^{(t+1)}|\theta^{(t)})f(D|\theta^{(t)})f(\theta^{(t)})} \right)
\end{aligned}$$

MH法の特殊形

今まで一般的なMH法を紹介してきましたが, ここからはMH法の特殊形(よりシンプル)を紹介します.

独立メトロポリス・ヘイスティングス法 (Independence Metropolis-Hastings: IMH)

一般的なMH法の提案分布は条件付きでした. $q(\cdot|\theta^{(t)})$

IMH法は提案分布を1期前の乱数に依存しない無条件分布 $q(\cdot)$ とします. そうすると各ステップでdrawされる乱数の候補が互いに独立(*i. i. d.*)になります.

補正確率は以下ようになります

$$A(a^{(t+1)}, \theta^{(t)}) = \min \left(1, \frac{q(\theta^{(t)})f(a^{(t+1)}|D)}{q(a^{(t+1)})f(\theta^{(t)}|D)} \right)$$

<例題>

ポアソン分布のパラメータ(期待値)を推定します. ポアソン分布の尤度にガンマ分布の事前分布をかけて事後分布が $f(\theta|\alpha = 11, \lambda = 13)$ のガンマ分布となりました. これに従う乱数をIMHで発生してみましょう. (ガンマ分布なら期待値と分散は解析的に求まっているのでわざわざMH法使わなくてもいけるのですが教材としてやってみます)

目的分布である事後カーネルは,

$$f(D|\theta)f(\theta) = e^{-13\theta}\theta^{10}$$

提案分布 $q(\theta)$ は何でもいいですが, 簡単な正規分布 $\text{Normal}(1, 0.5)$ にしてみます. (実際ここの調整が難しいのですが)

解析解

Mean[1] 0.846

Var[1] 0.065

SD[1] 0.255

採択率[1] 0.41231

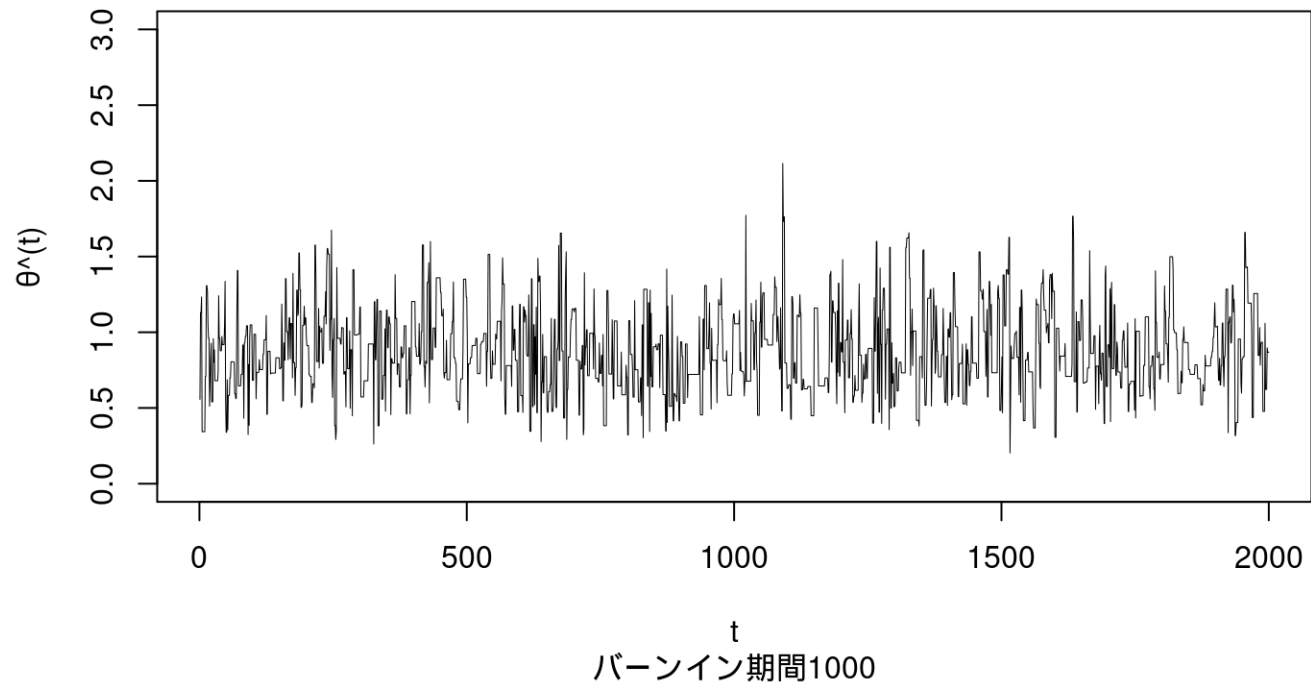
モンテカルロ積分

Mean[1] 0.846

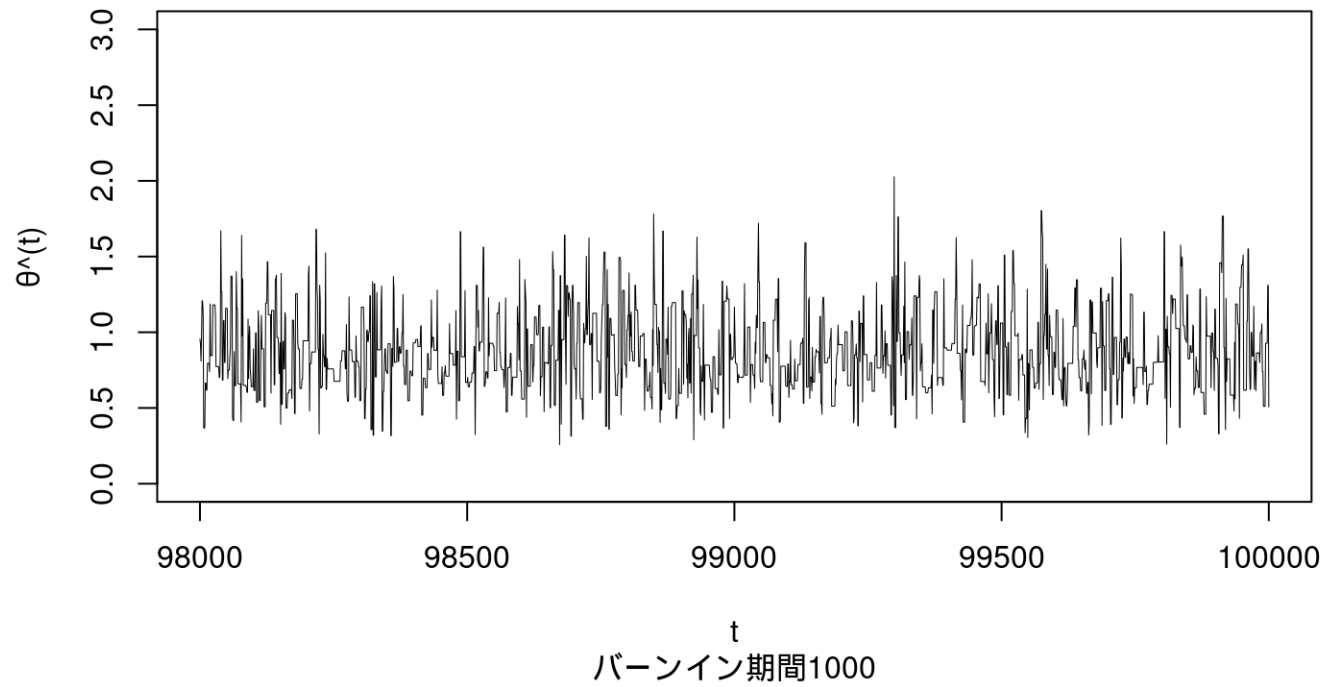
Var[1] 0.066

SD[1] 0.256

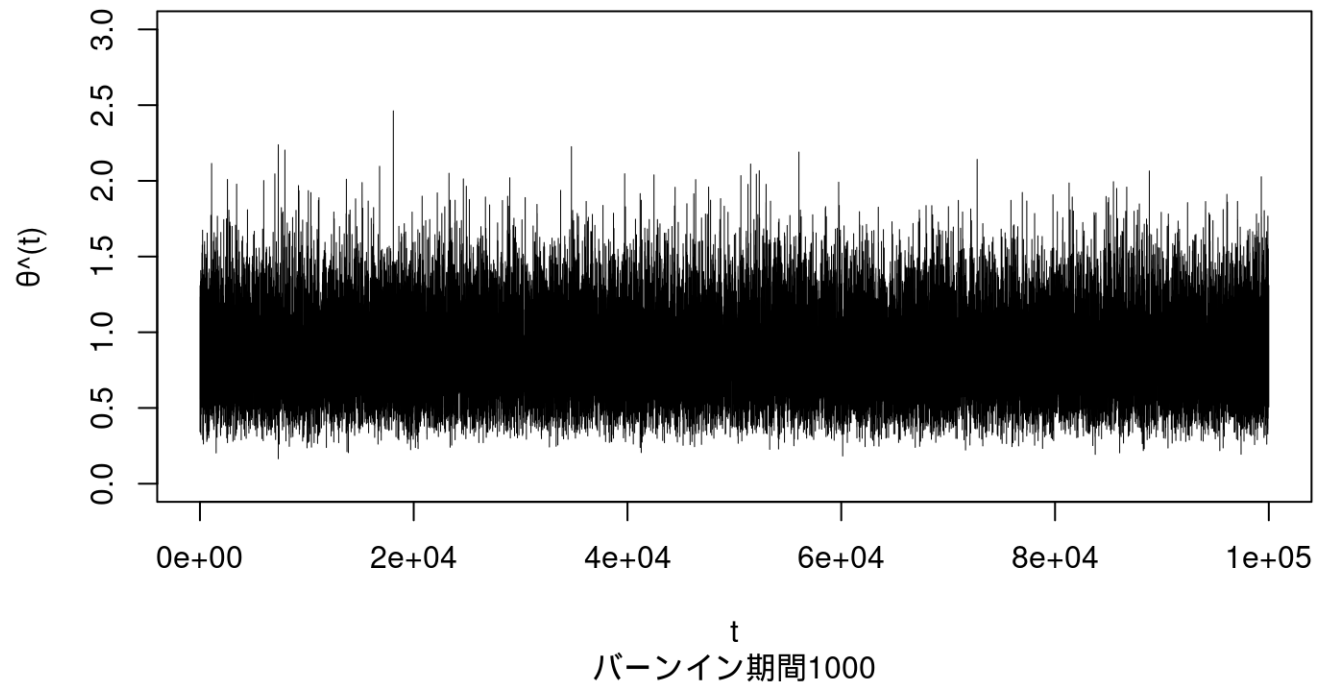
乱数 $\theta^{\wedge}(t)$ のトレースプロット



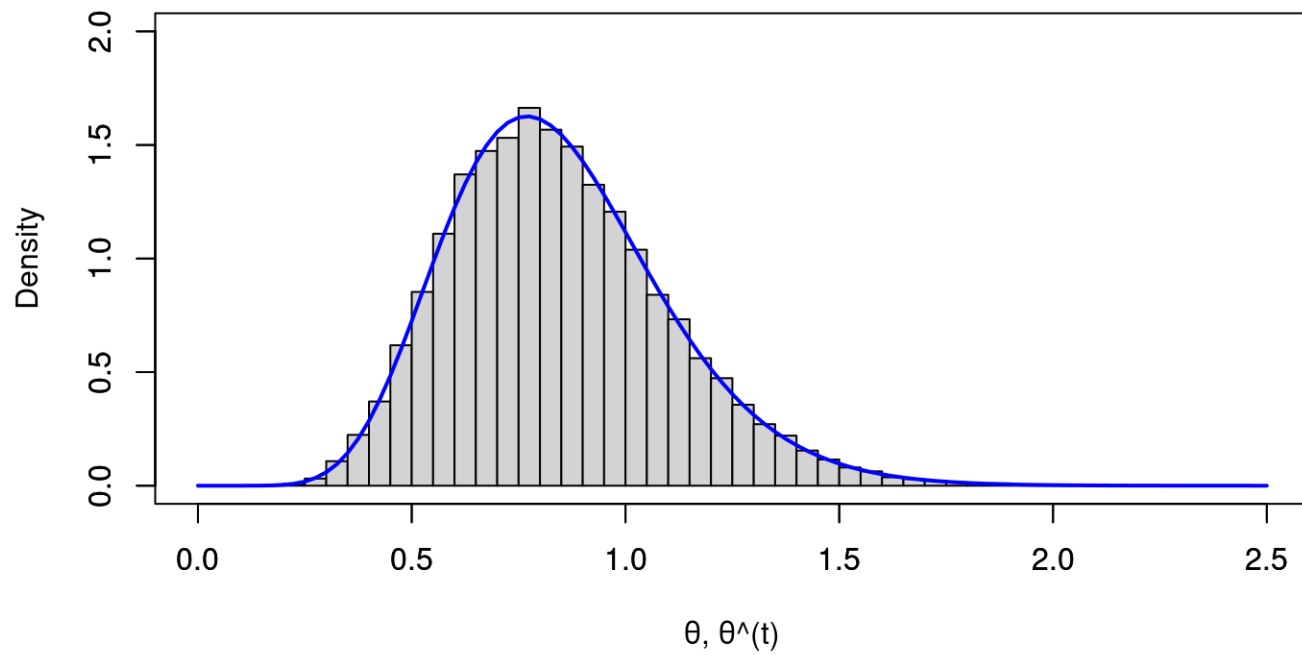
乱数 θ^t のトレースプロット



乱数 $\theta^{\wedge}(t)$ のトレースプロット



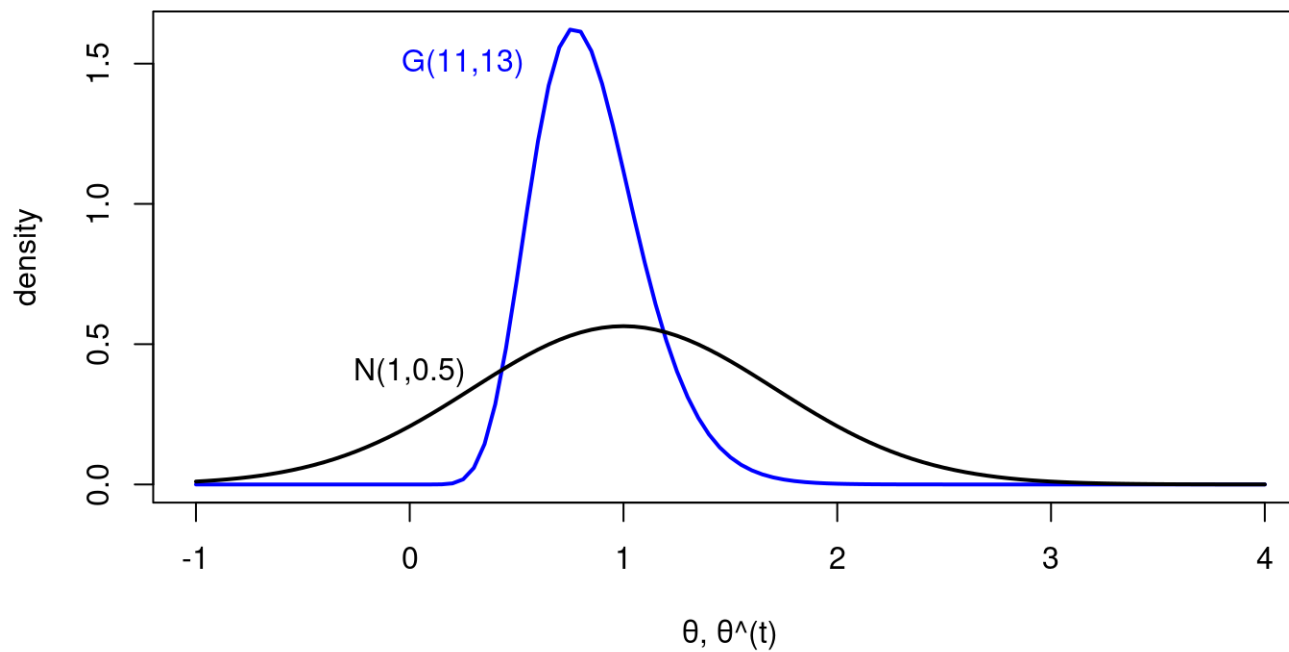
事後分布(青線)とサンプリング結果(ヒストグラム)



IMH法の弱点

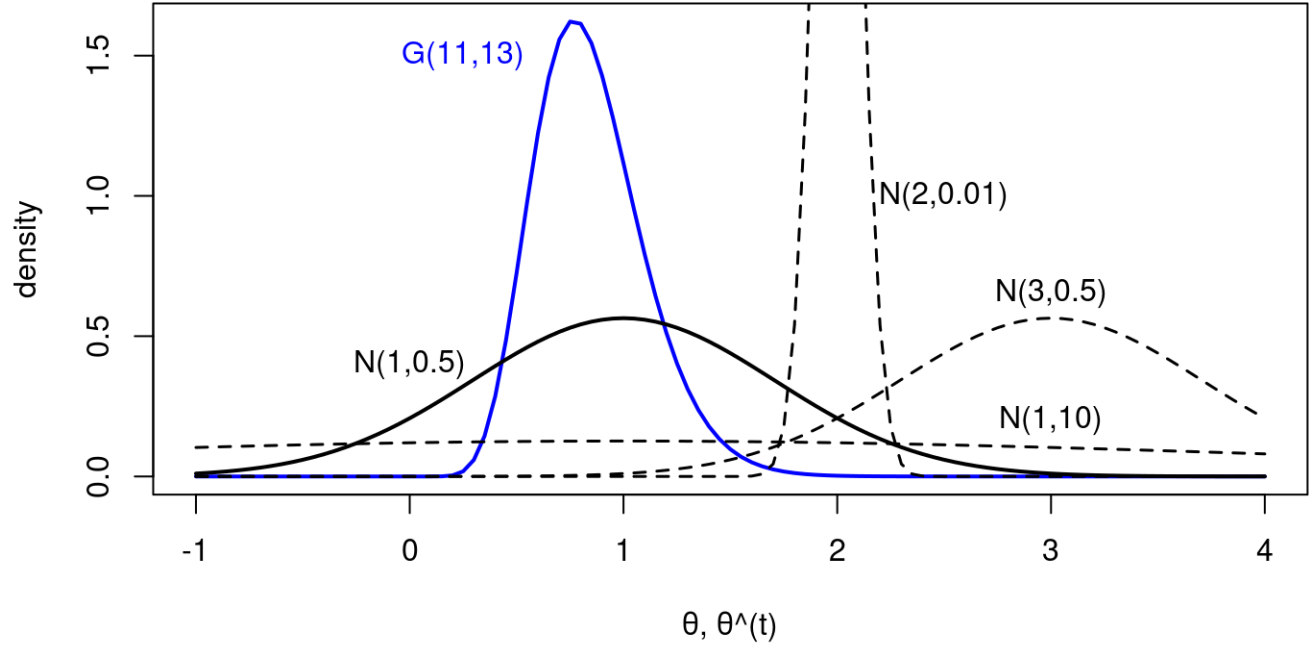
上の例では提案分布を $\text{Normal}(1, 0.5)$ としてみました. 結果的には良いMCMCサンプリングができと言えます. その理由は, 目的分布である事後分布 $\text{Gamma}(\alpha = 11, \lambda = 13)$ と結構重なっていたからです.

事後分布と提案分布

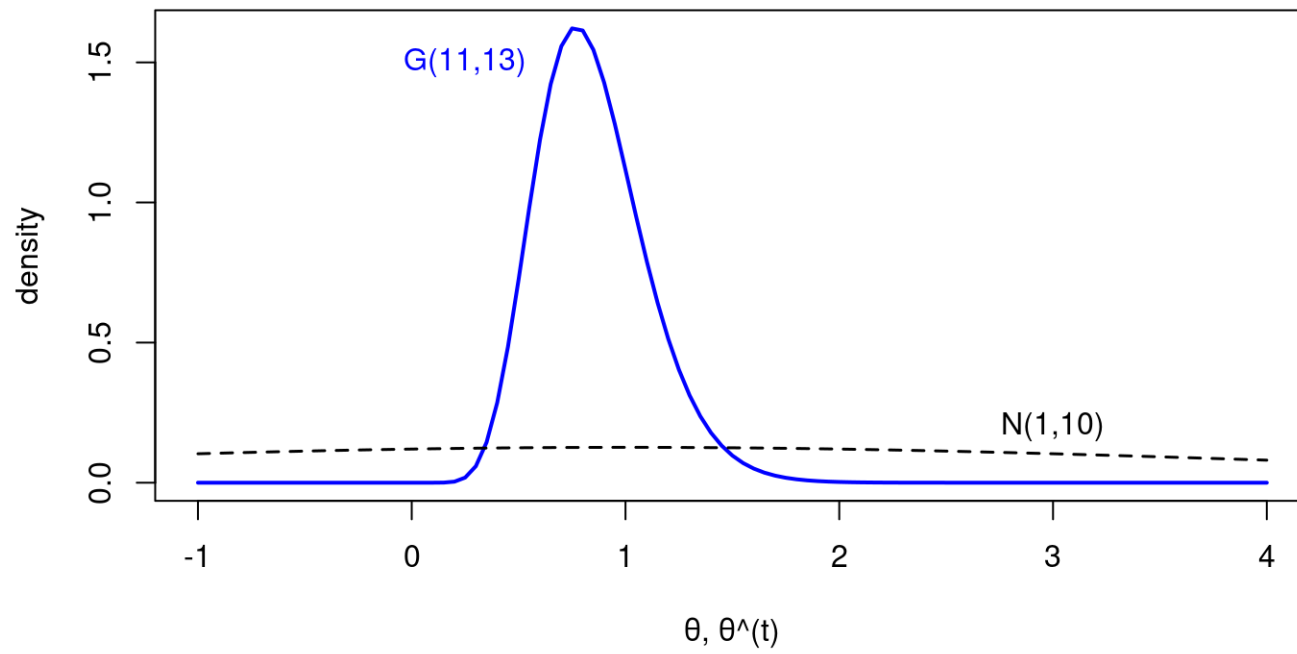


提案分布の期待値と分散を変えるとどうなるでしょう.

事後分布といろいろな提案分布



事後分布といろいろな提案分布



```
## 解析解
```

```
## Mean[1] 0.846
```

```
## Var[1] 0.065
```

```
## SD[1] 0.255
```

```
## [1] "proposal: N(1, 10)"
```

```
## 採択率[1] 0.09924
```

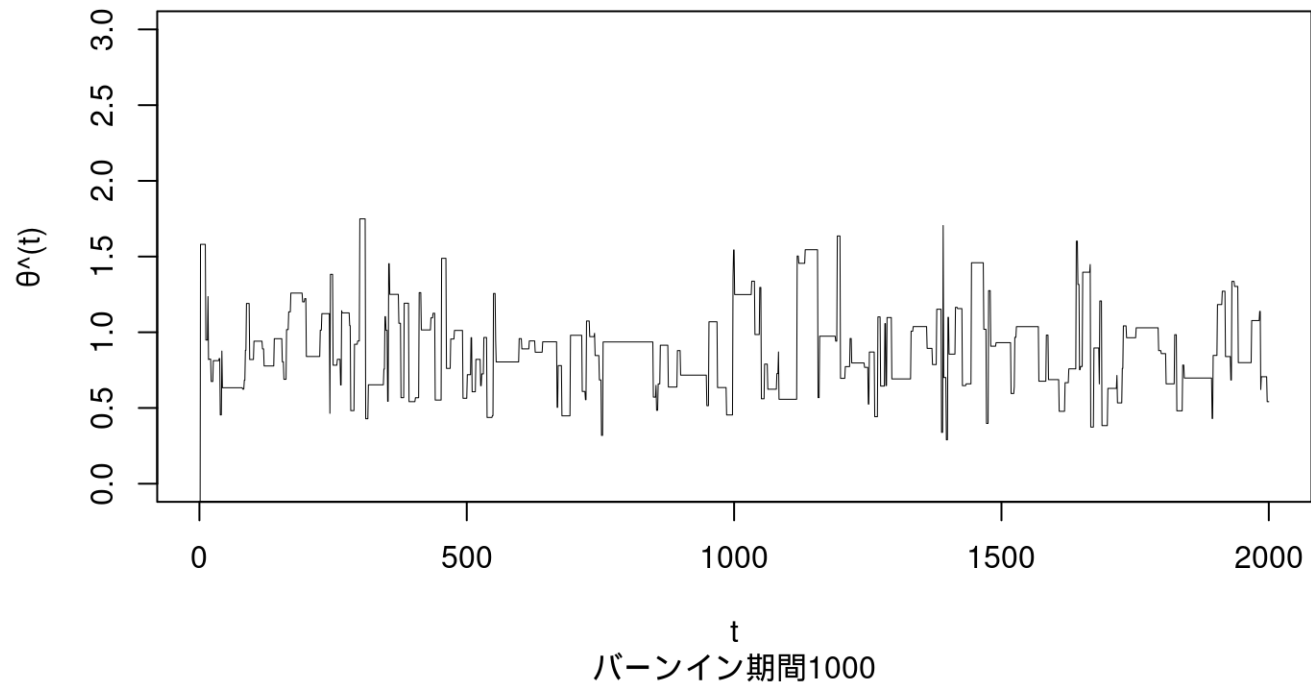
```
## モンテカルロ積分
```

```
## Mean[1] 0.845
```

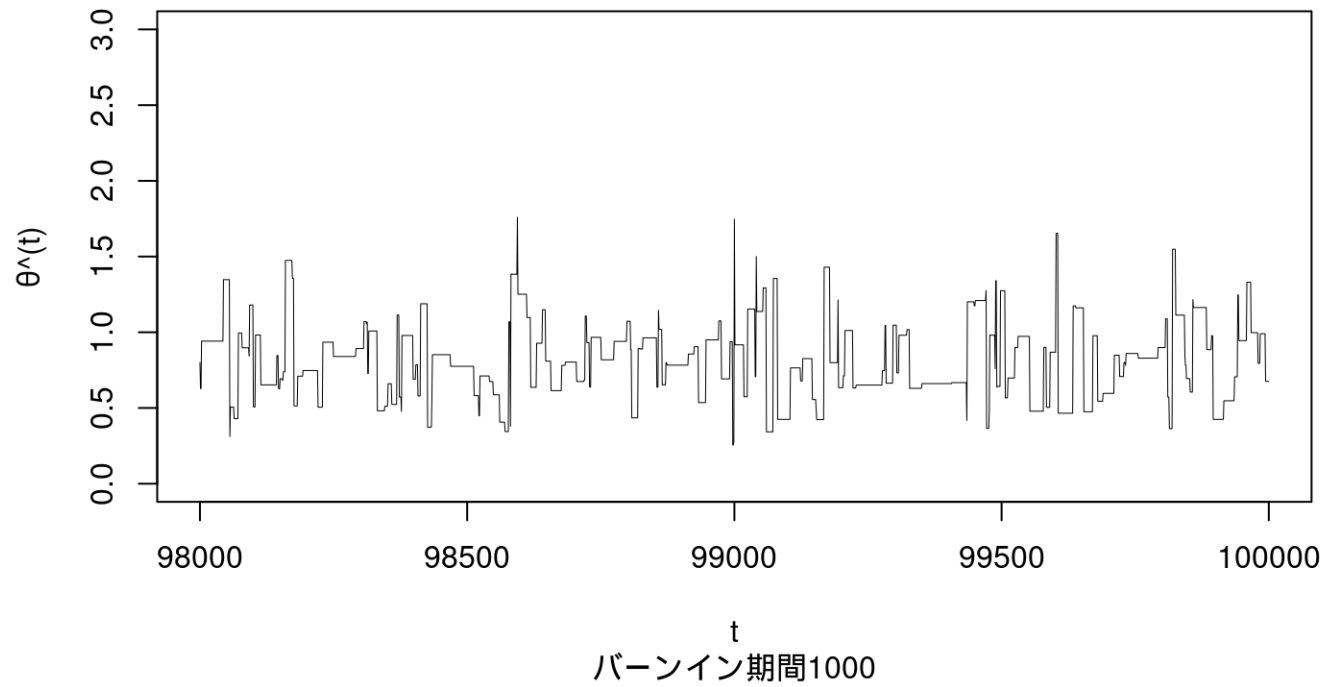
```
## Var[1] 0.065
```

```
## SD[1] 0.256
```

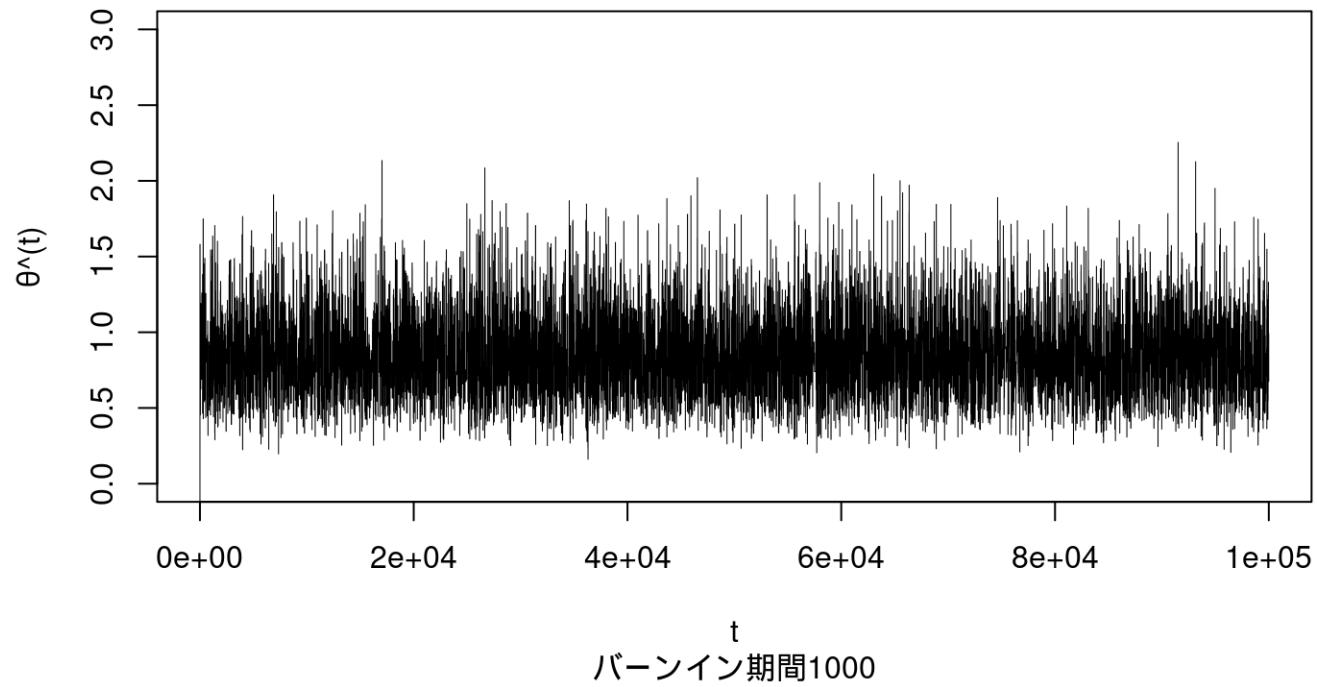
乱数 $\theta^{\wedge}(t)$ のトレースプロット



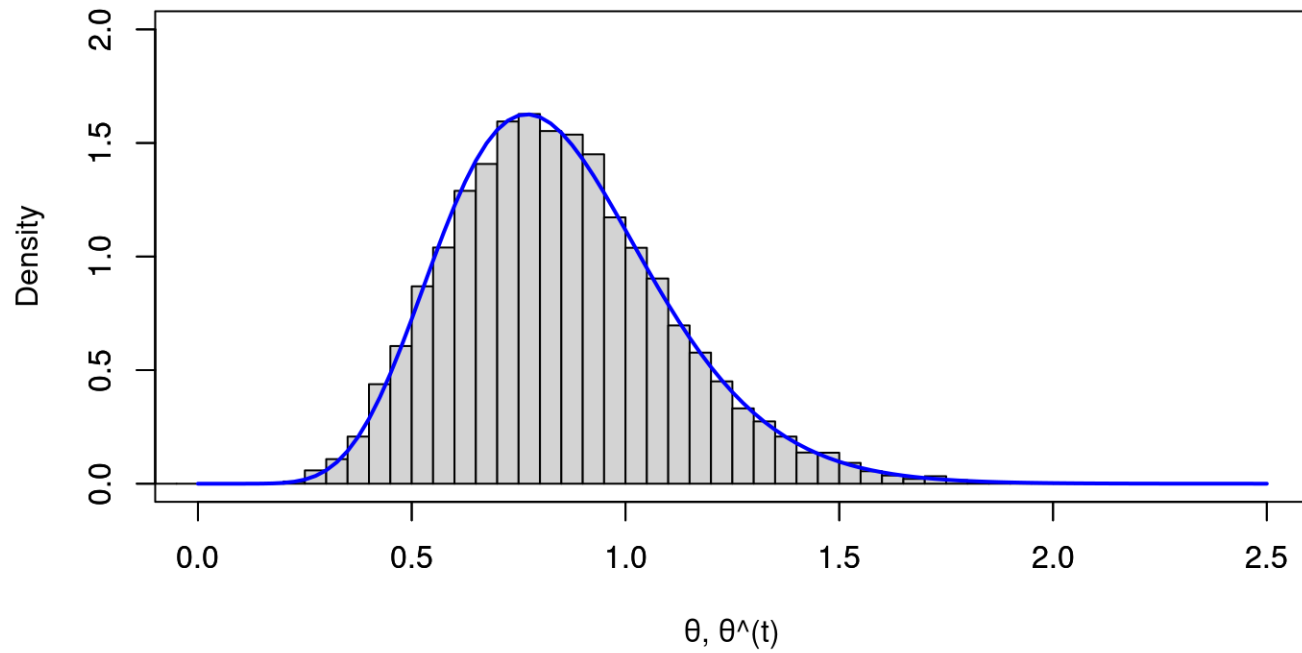
乱数 $\theta^{\wedge}(t)$ のトレースプロット



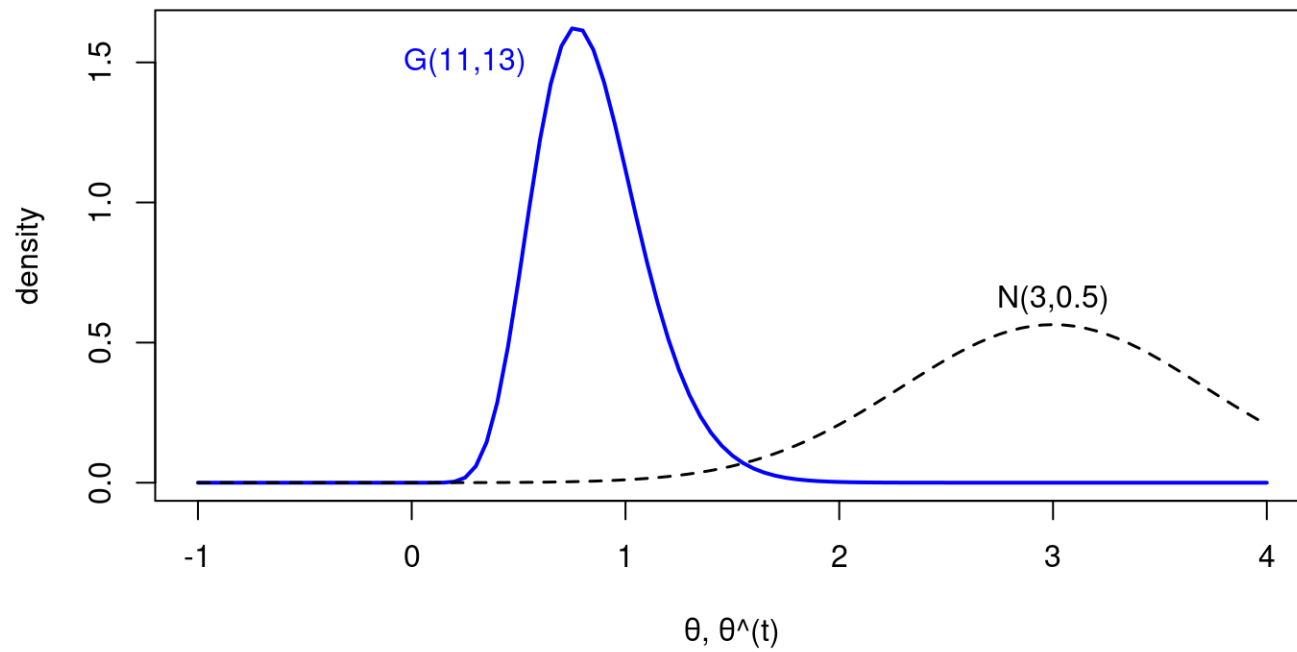
乱数 $\theta^{\wedge}(t)$ のトレースプロット



事後分布(青線)とサンプリング結果(ヒストグラム)



事後分布といろいろな提案分布



```
## 解析解
```

```
## Mean[1] 0.846
```

```
## Var[1] 0.065
```

```
## SD[1] 0.255
```

```
## [1] "proposal: N(3, 0.5)"
```

```
## 採択率[1] 0.00407
```

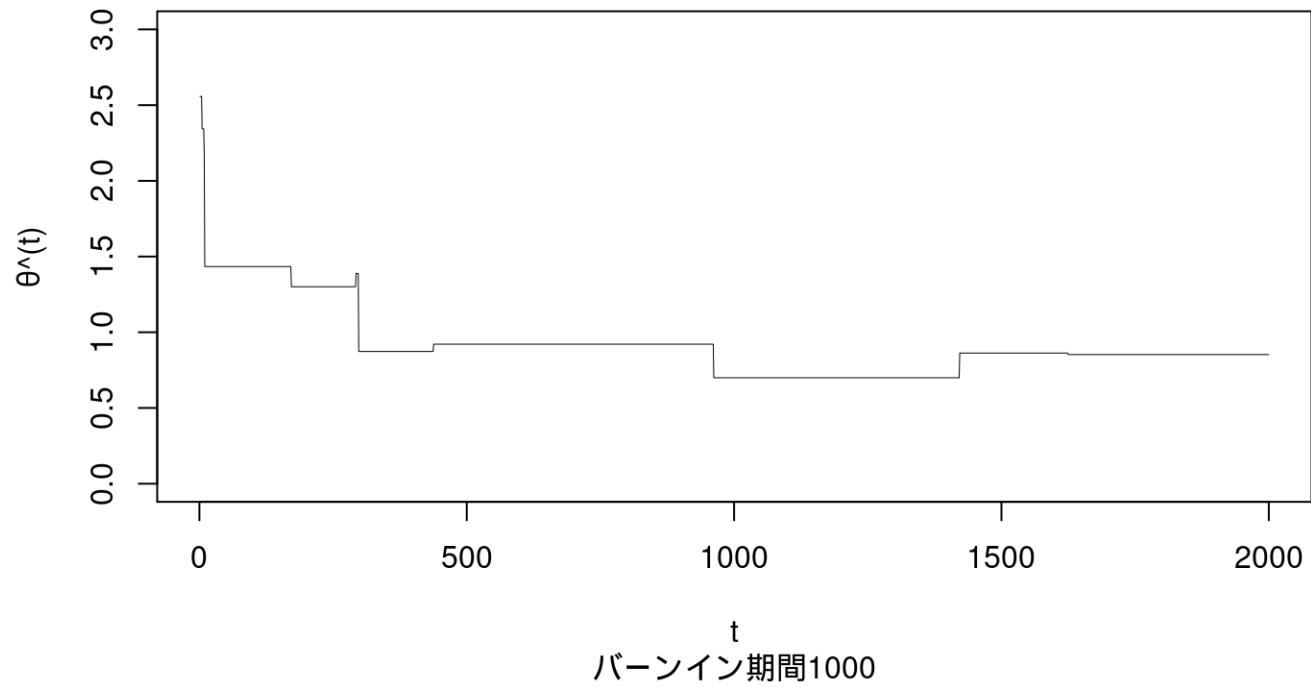
```
## モンテカルロ積分
```

```
## Mean[1] 0.847
```

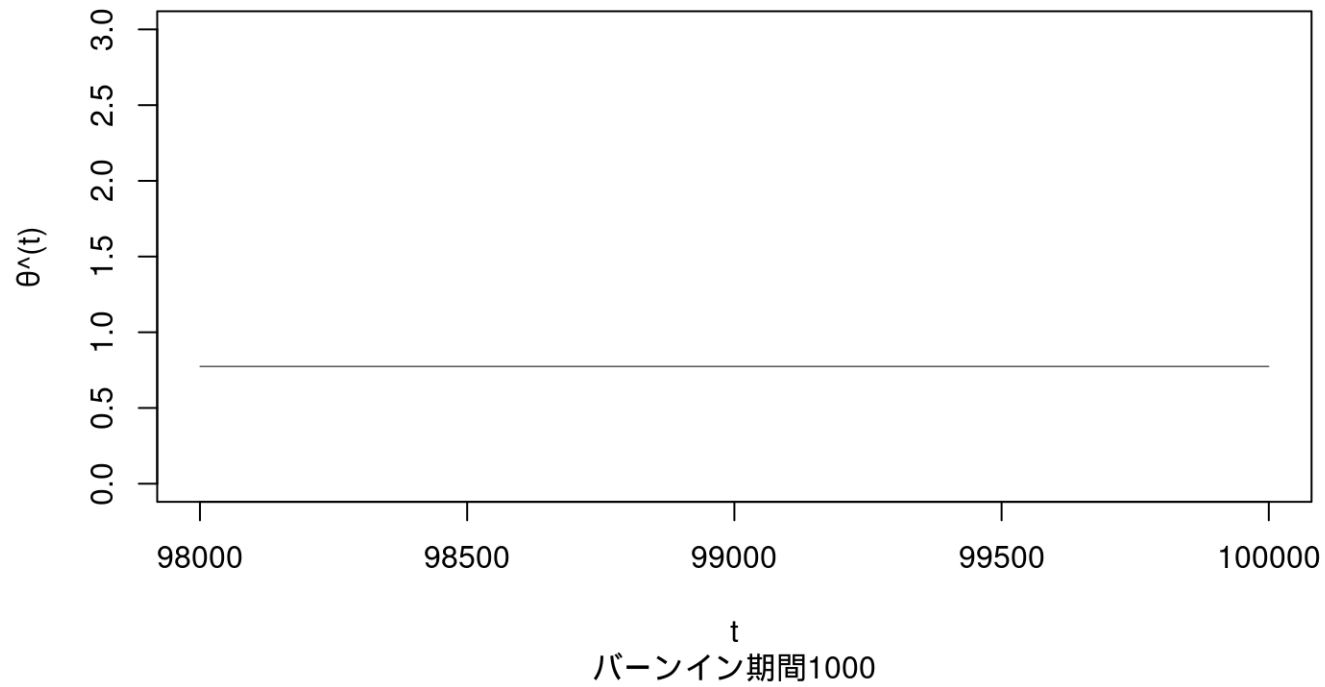
```
## Var[1] 0.053
```

```
## SD[1] 0.229
```

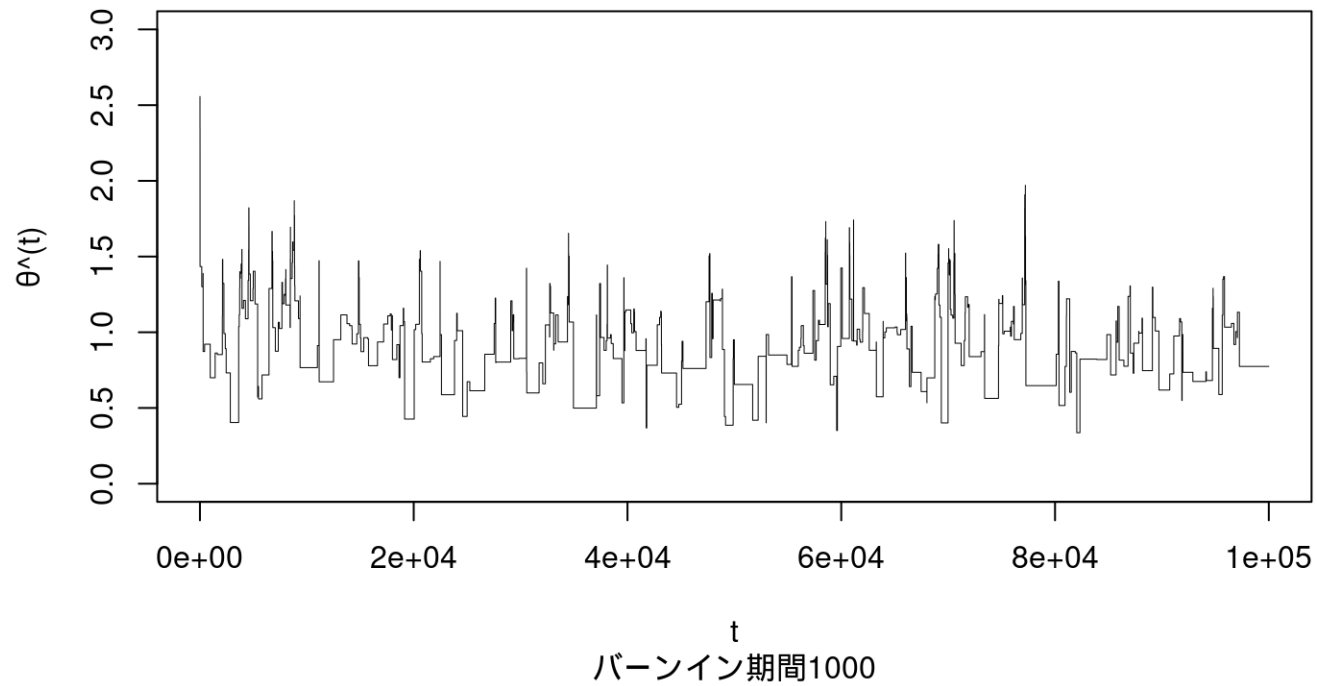
乱数 $\theta^{\wedge}(t)$ のトレースプロット



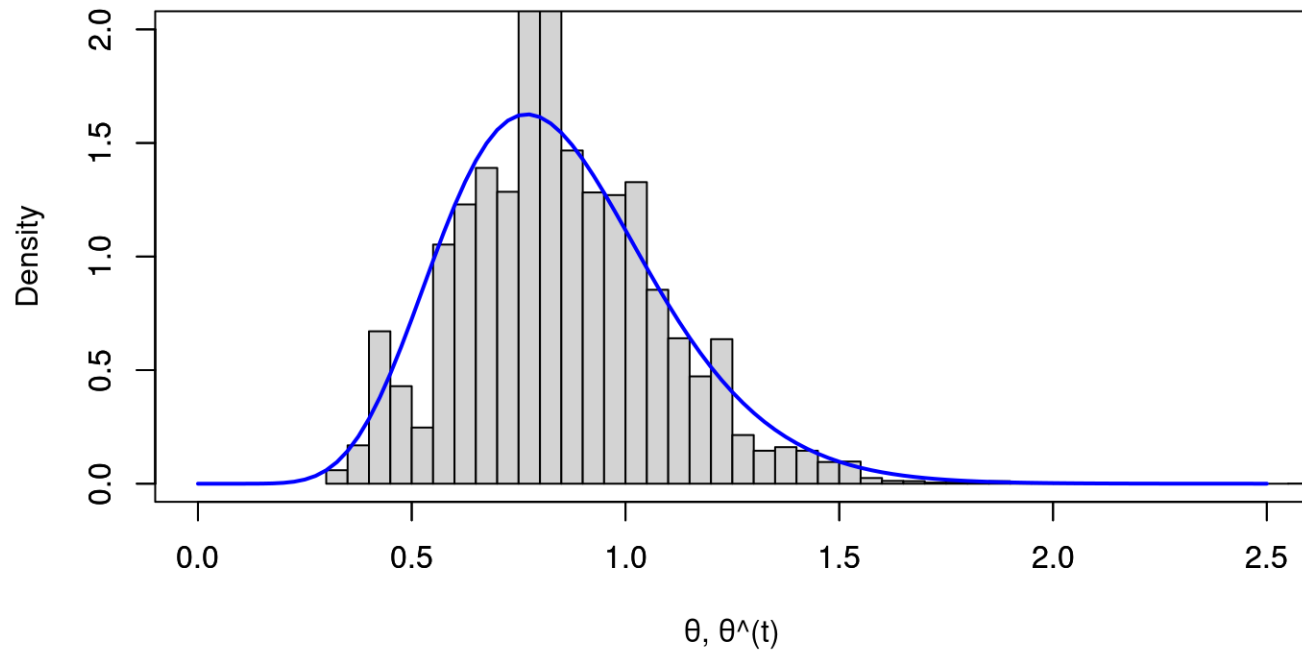
乱数 θ^t のトレースプロット



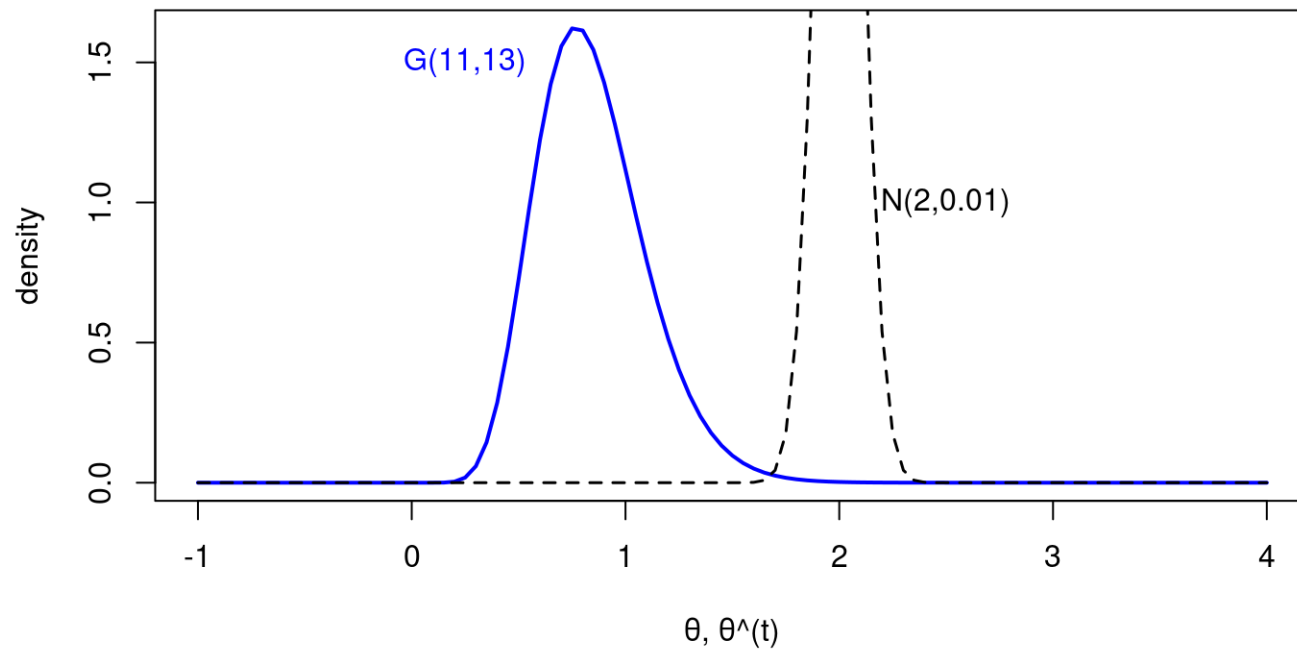
乱数 $\theta^{\wedge}(t)$ のトレースプロット



事後分布(青線)とサンプリング結果(ヒストグラム)



事後分布といろいろな提案分布



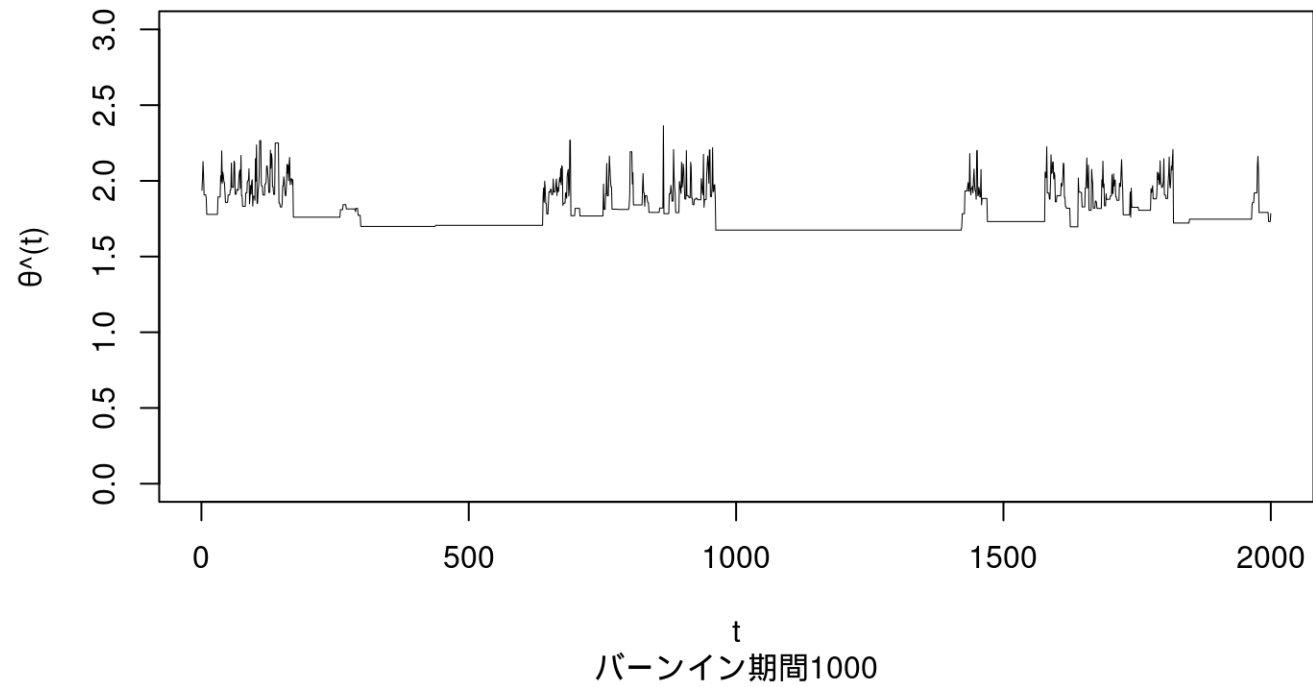

```
## 解析解
## Mean[1] 0.846
## Var[1] 0.065
## SD[1] 0.255

## [1] "proposal: N(2, 0.01)"

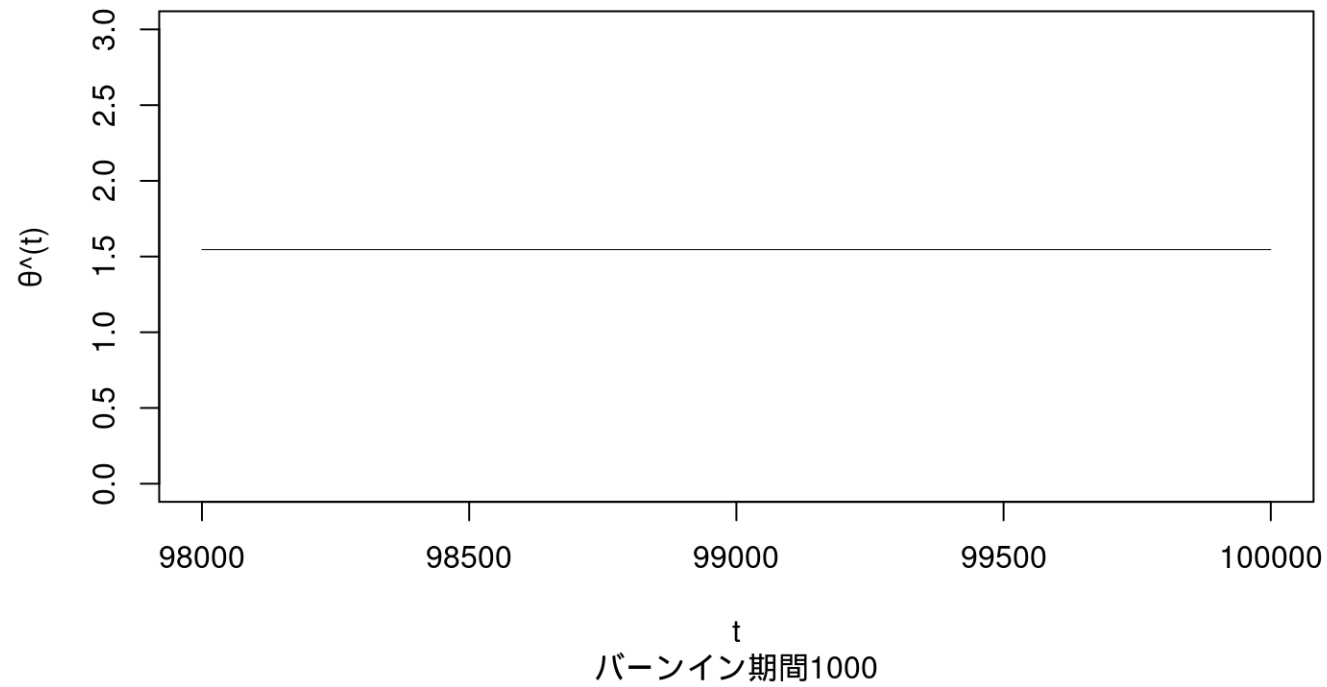
## 採択率[1] 0.05196

## モンテカルロ積分
## Mean[1] 1.632
## Var[1] 0.015
## SD[1] 0.123
```

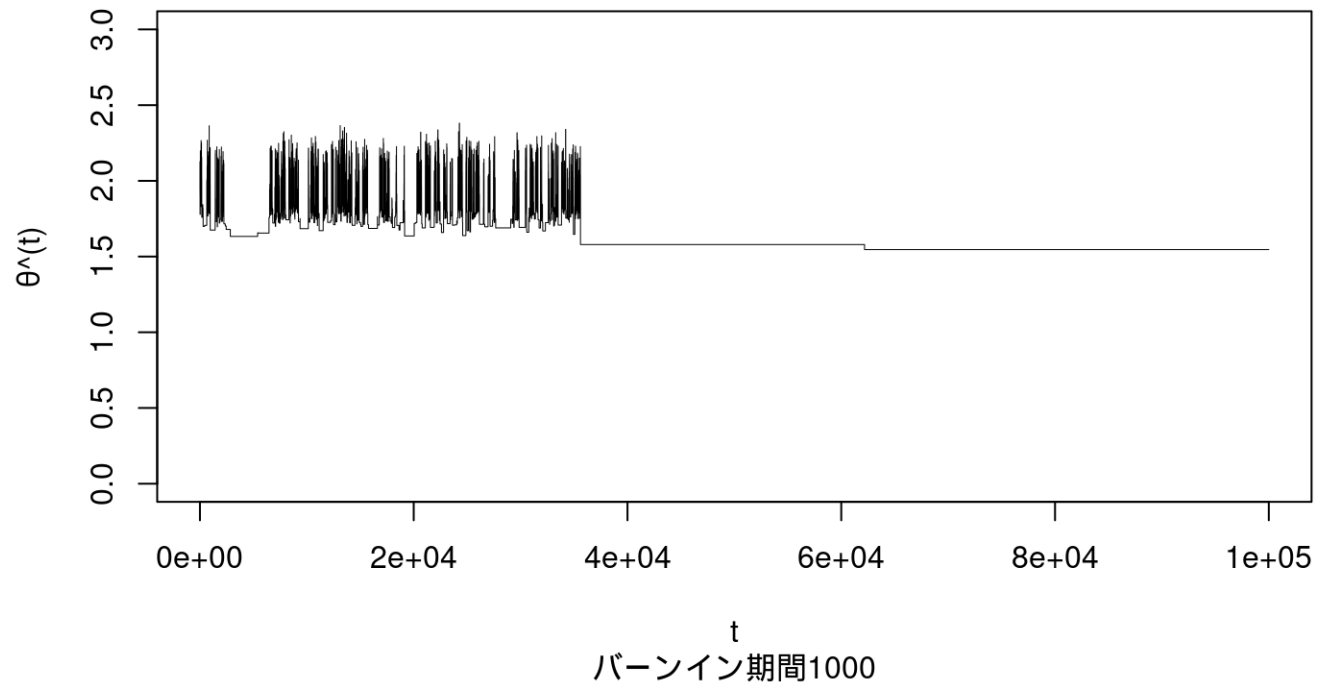
乱数 $\theta^{\wedge}(t)$ のトレースプロット



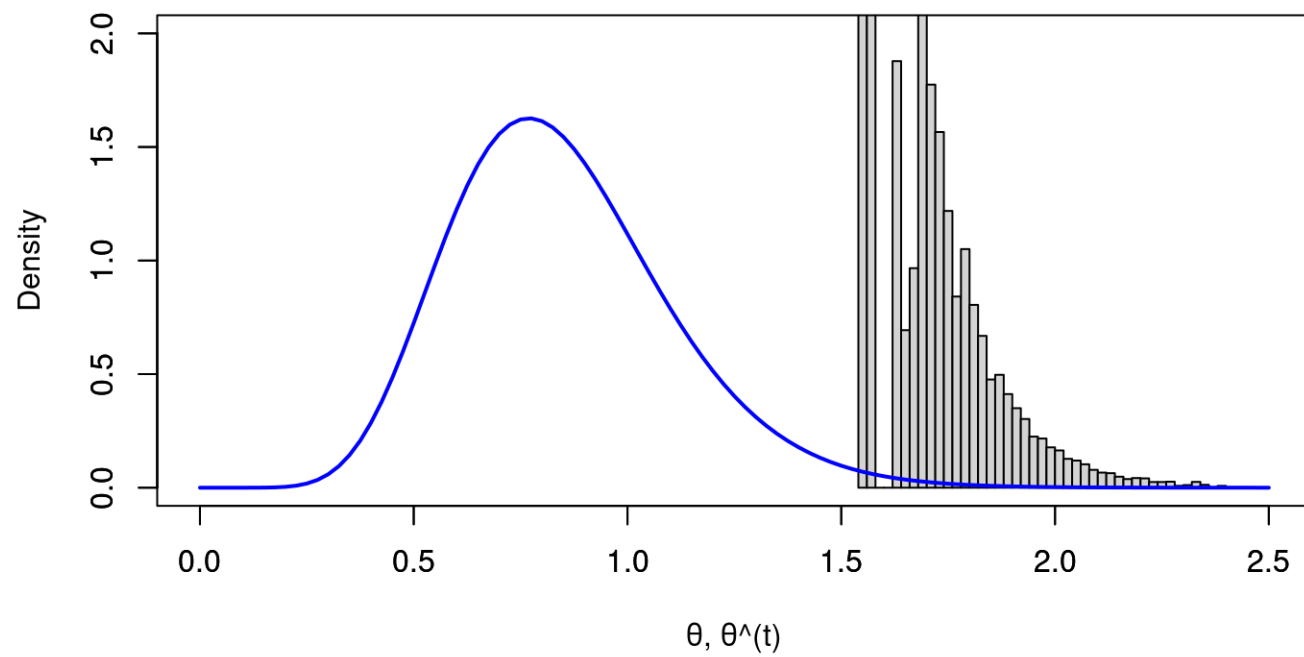
乱数 θ^t のトレースプロット



乱数 θ^t のトレースプロット



事後分布(青線)とサンプリング結果(ヒストグラム)



例えば提案分布が $\text{Normal}(1, 2)$ だと, θ の範囲はうまくカバーできてるものの事後分布と重なっている面積が小さいため受容率が低く収束も遅くなります. $\text{Normal}(3, 0.5)$ や $\text{Normal}(2, 0.01)$ はそもそも重なっている面積がほぼないので絶望的です.

仮に事後分布の位置や形がわかっていたら上手な提案分布を設定できるかもしれませんが, 実際の分析では事後分布の形を把握することが難しいので, 分析者がIMHの提案分布を設定するのは非常に困難です. そこでランダムウォークMHを導入します.

ランダムウォーク メトロポリス・ヘイスティングス法 (Random Walk Metropolis-Hastings: RWMH)

RWMHはパラメータの乱数候補 $a^{(t)}$ を各ステップで提案分布 $q(\cdot)$ から逐一drawするのではなく,

$$\begin{aligned} a^{(t+1)} &= \theta^{(t)} + e \\ e &\sim \text{Normal}(0, \delta_e^2) \text{ or } \text{Uniform}(-\delta'_e, \delta'_e) \end{aligned}$$

とします. e はノイズで, $E[e] = 0$ となるような左右対称な分布(平均0の正規分布や左右対称な一様分布)に従います. すると $a^{(t+1)}$ は1ステップ前の乱数 $\theta^{(t)}$ と平均0のノイズ e だけで決まるランダムウォークになります. ランダムウォークに動くということは, 仮に1ステップ前の乱数 $\theta^{(t)}$ が“良い”乱数であれば, 次の乱数候補 $a^{(t+1)}$ もそう悪くない乱数になるということです(期待値は1ステップ前の乱数 $E[a^{(t+1)}] = \theta^{(t)}$). これは当てずっぽうに毎ステップ同じ提案分布 $q(\cdot)$ から乱数候補をdrawするIMHよりも良さそうです.

RWMH法

RWMHの提案分布自体は正規分布や一様分布でなくても何でもいいですが, 対称な分布にすると便利です. 対称な提案分布にすると,

$$q(a^{(t+1)}|\theta^{(t)}) = q(\theta^{(t)}|a^{(t+1)})$$

となります. よってRWMHの補正確率は

$$A(a^{(t+1)}, \theta^{(t)}) = \min \left(1, \frac{f(a^{(t+1)}|D)}{f(\theta^{(t)}|D)} \right)$$

となり, 補正確率 r から提案分布 $q(\cdot)$ が消えて事後分布の比だけになります. あとはいつものMH法のようにサンプリングします.

<例題>

先のIMHの例題をRWMHでやってみます. 初期値は $\theta^{(0)} = 4$ としてわざと遠くにしてみます. ノイズ e の分散は $\delta_e^2 = 0.1$ とします(実際これは分析者が調節します).

解析解

Mean[1] 0.846

Var[1] 0.065

SD[1] 0.255

採択率[1] 0.6381918

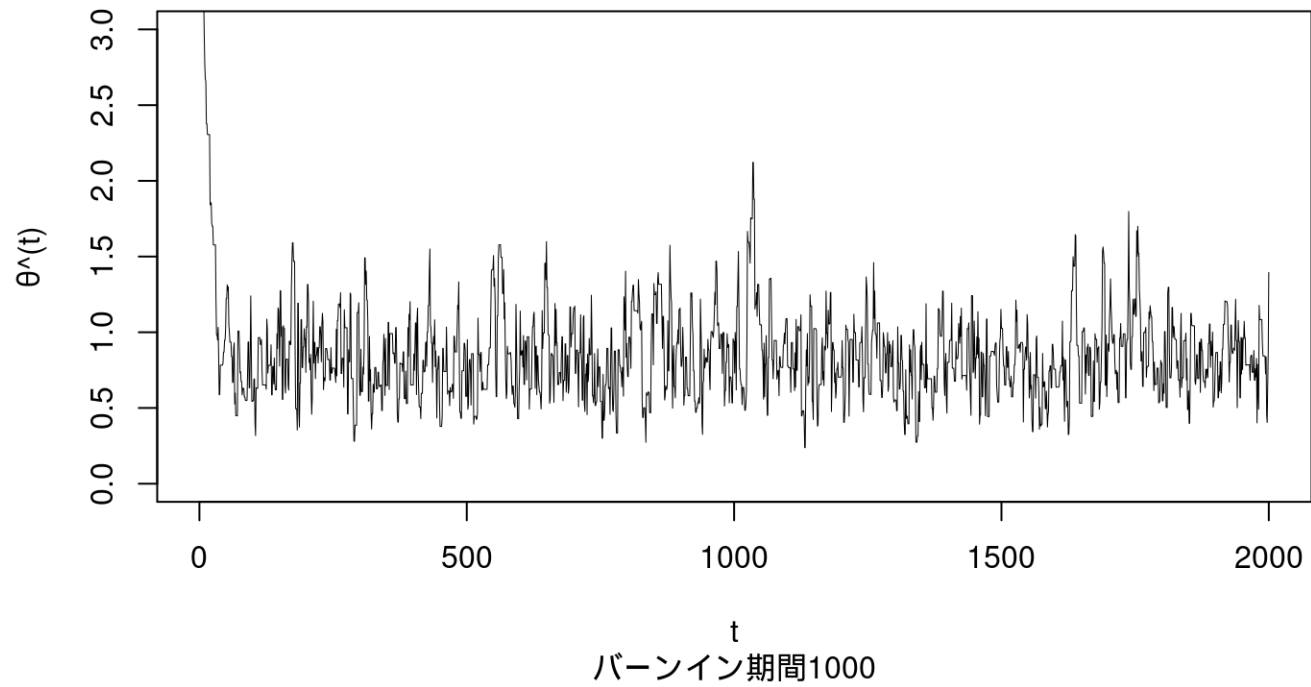
モンテカルロ積分

Mean[1] 0.849

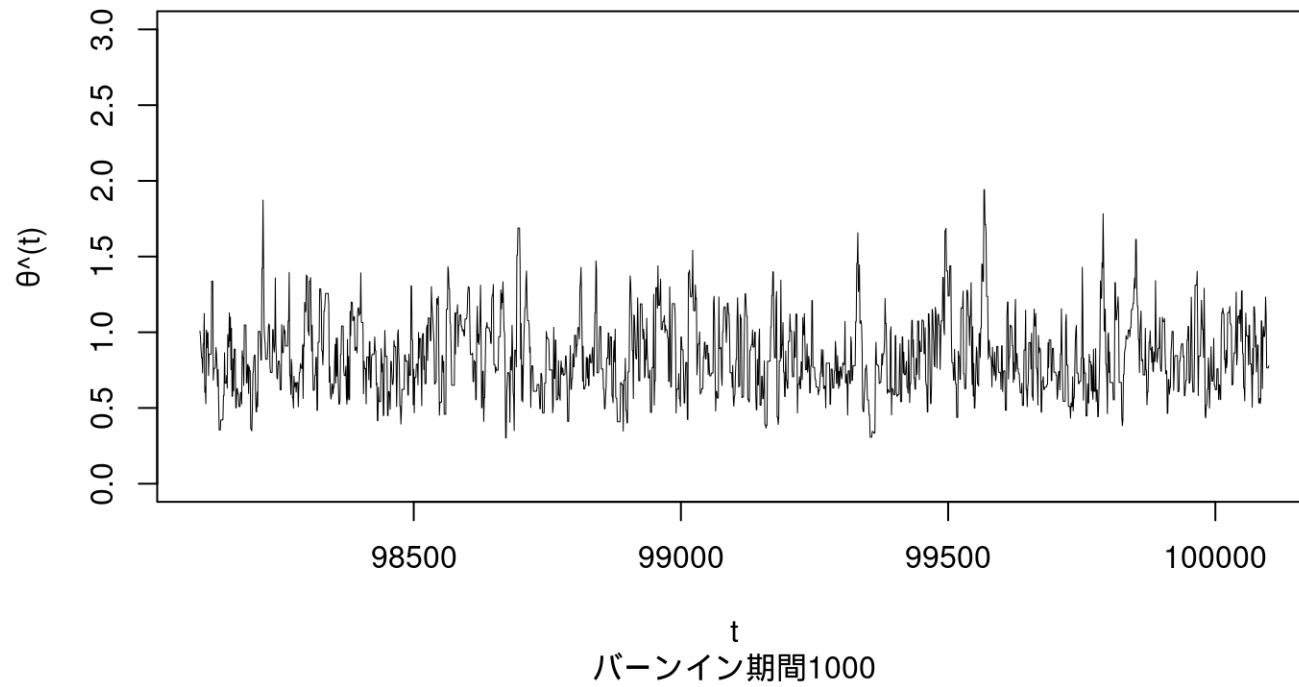
Var[1] 0.065

SD[1] 0.255

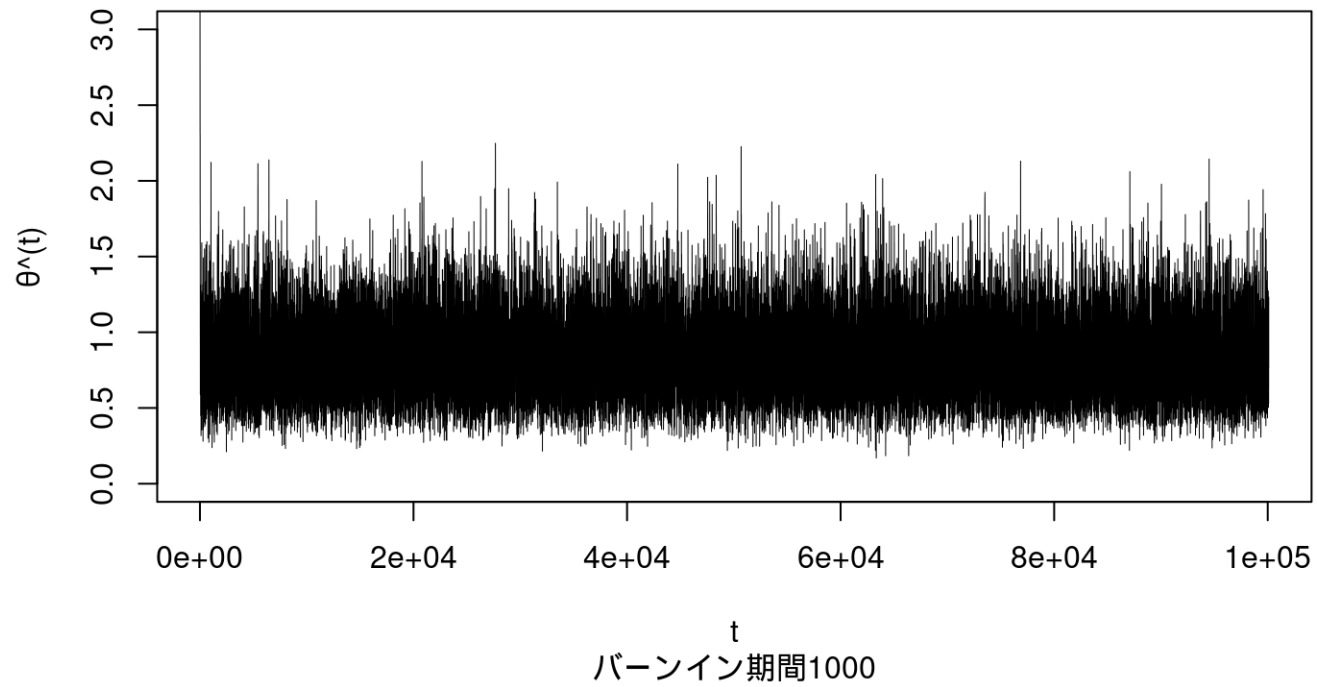
乱数 $\theta^{\wedge}(t)$ のトレースプロット



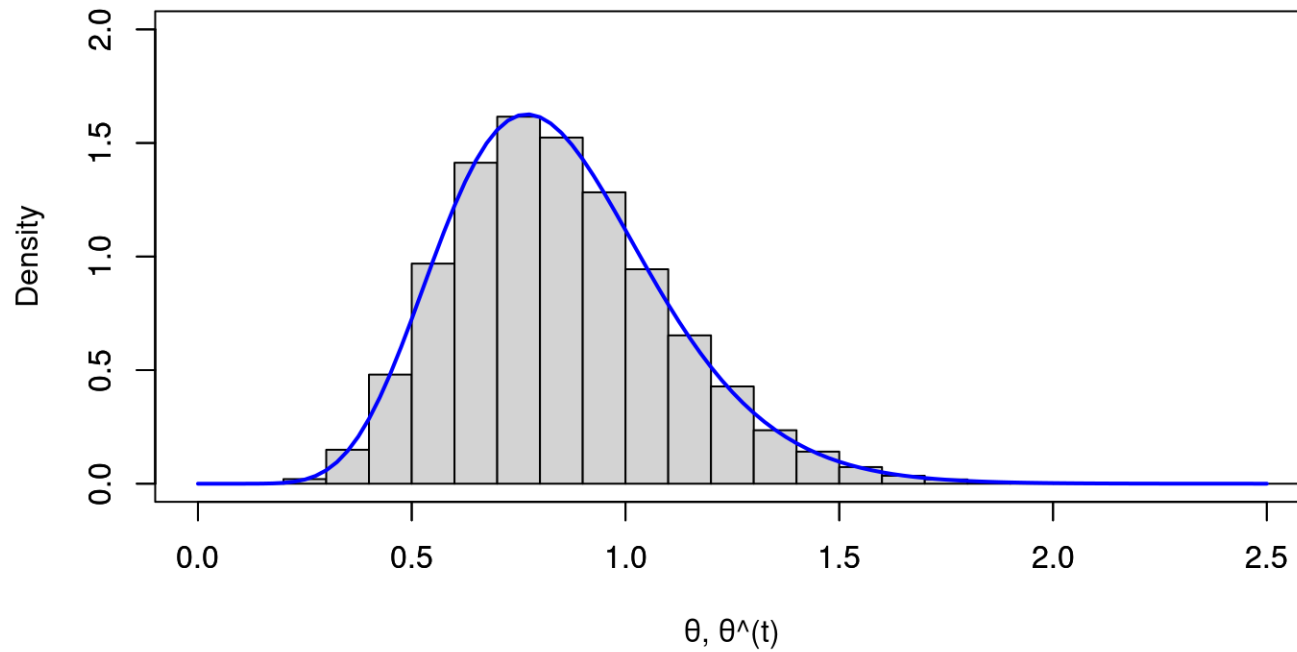
乱数 $\theta^{\wedge}(t)$ のトレースプロット



乱数 $\theta^{\wedge}(t)$ のトレースプロット



事後分布(青線)とサンプリング結果(ヒストグラム)



初期値がでたらめでも, IMHよりきれいに事後分布に重なったサンプリングができています(δ_e をうまく設定できればの話ですが).

RWMH法の弱点

RWMHの成功のカギは δ_e をうまく設定できるかどうかです. δ_e が大きすぎると, 1回ランダムウォークの移動距離が長く, 目標分布の確率密度が高い領域になかなか入れずに受容率が下がってしまいます.

反対に δ_e が小さすぎると, 1回ランダムウォークの移動距離が短く, 収束まで時間がかかってしまいます(乱数同士が相関). δ_e はハイパーパラメータなので分析者が地道にちょうどいい値を探すしかありません.

$\delta_e = 100000$: 1ステップの移動幅が超デカイ

解析解

Mean[1] 0.846

Var[1] 0.065

SD[1] 0.255

採択率[1] 0.001008991

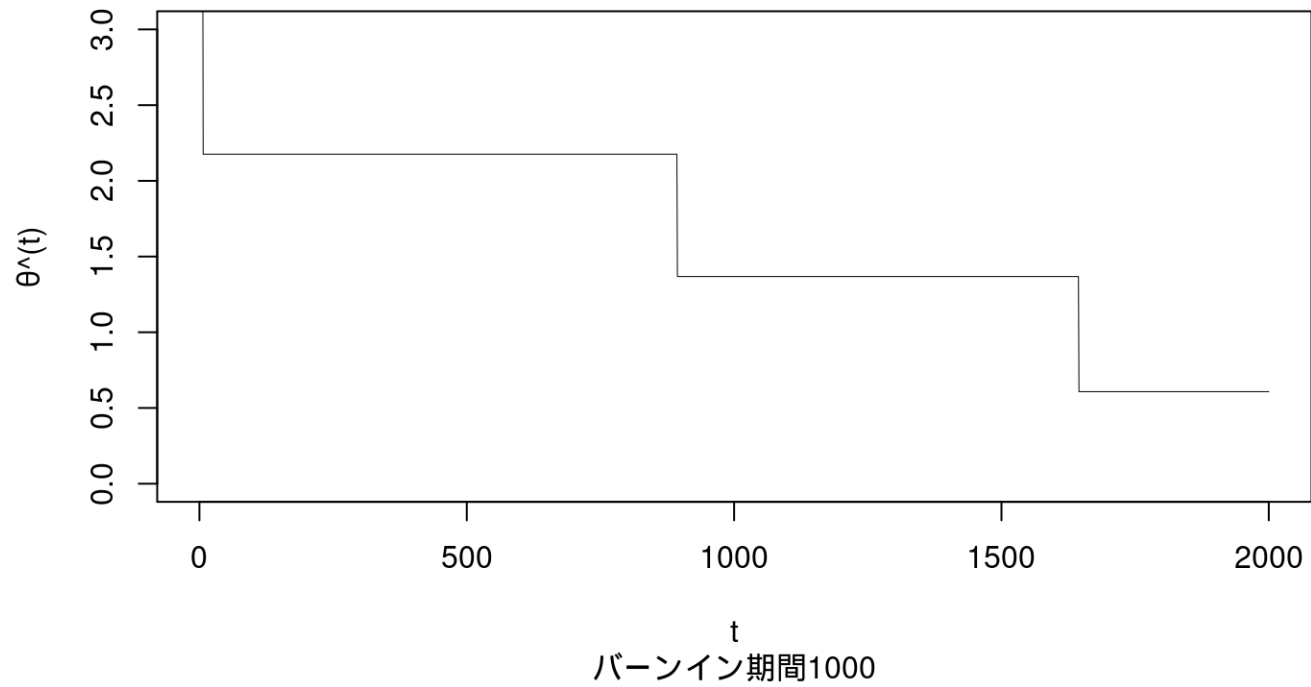
モンテカルロ積分

Mean[1] 0.852

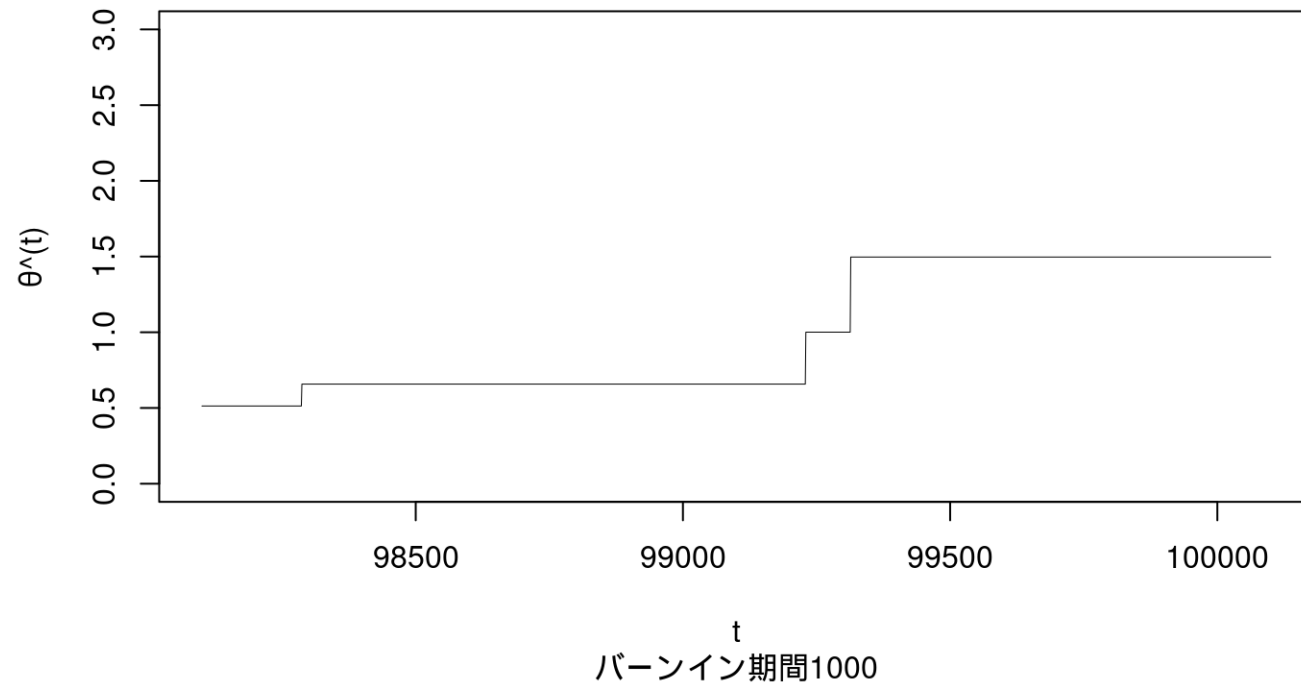
Var[1] 0.079

SD[1] 0.281

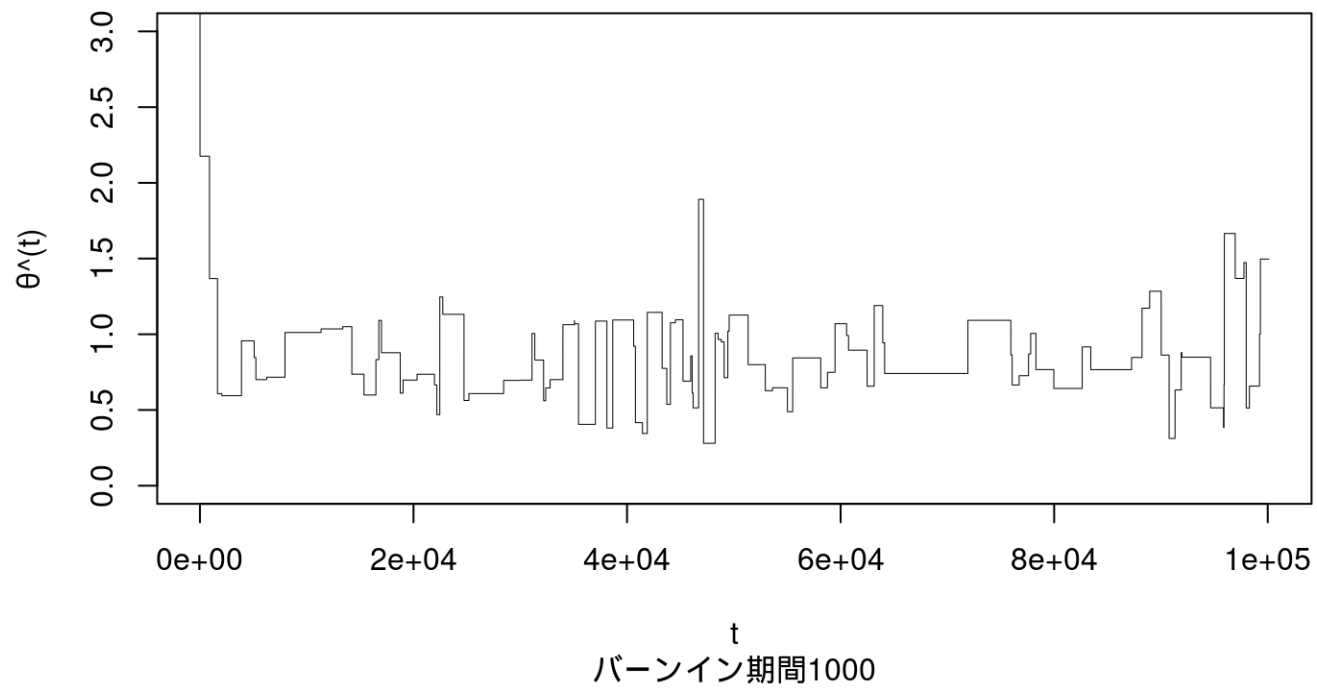
乱数 θ^t のトレースプロット



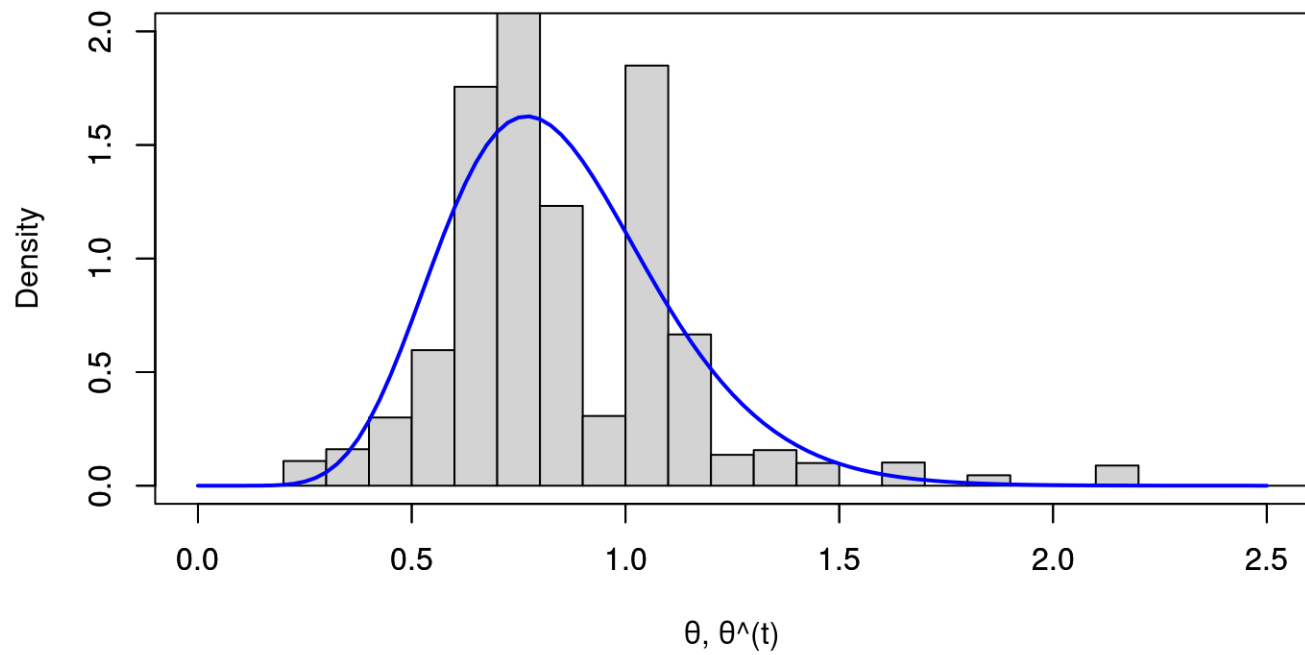
乱数 θ^t のトレースプロット



乱数 θ^t のトレースプロット



事後分布(青線)とサンプリング結果(ヒストグラム)



$\delta_e = 0.00001$: 1ステップの移動幅が超小さい

解析解

Mean[1] 0.846

Var[1] 0.065

SD[1] 0.255

採択率[1] 0.9922877

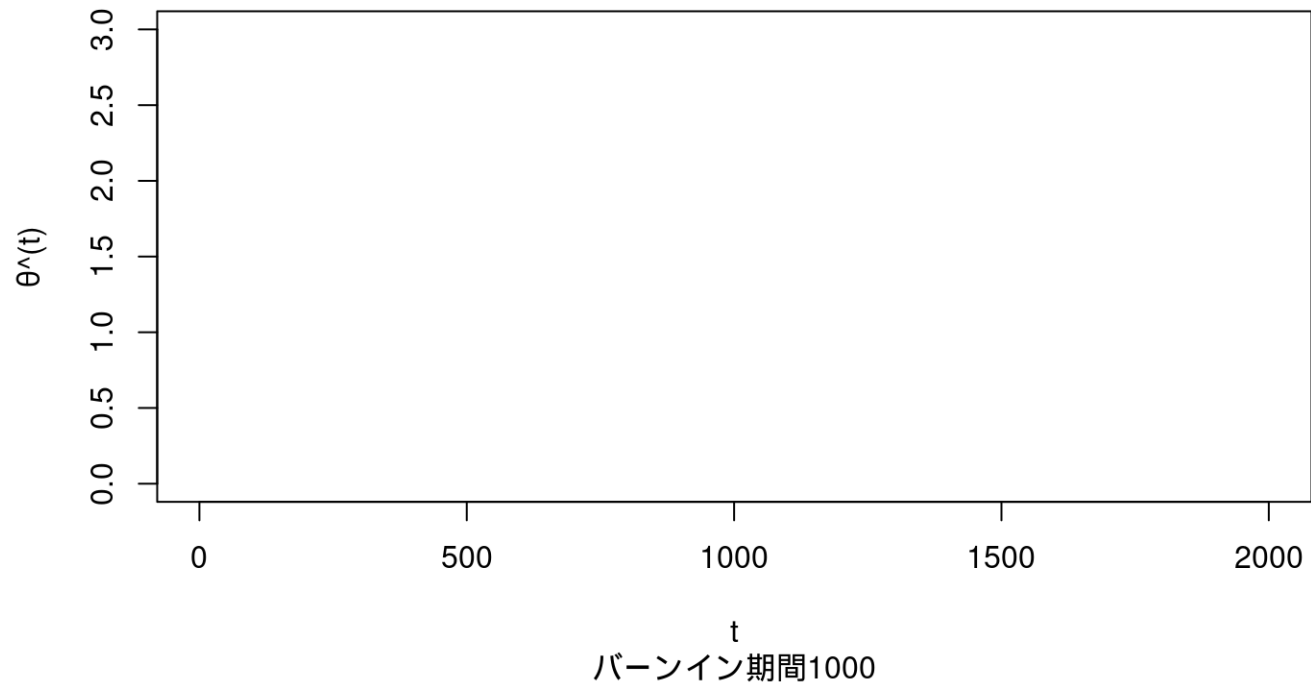
モンテカルロ積分

Mean[1] 1.805

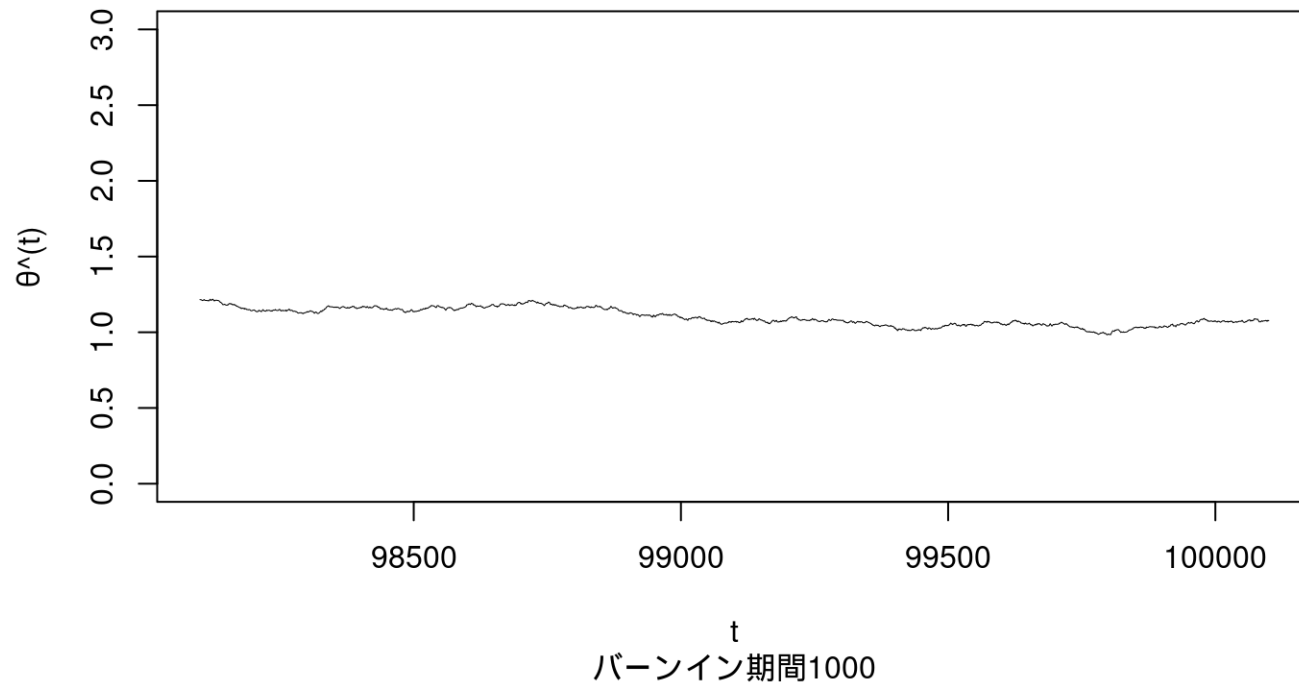
Var[1] 0.771

SD[1] 0.878

乱数 $\theta^{\wedge}(t)$ のトレースプロット



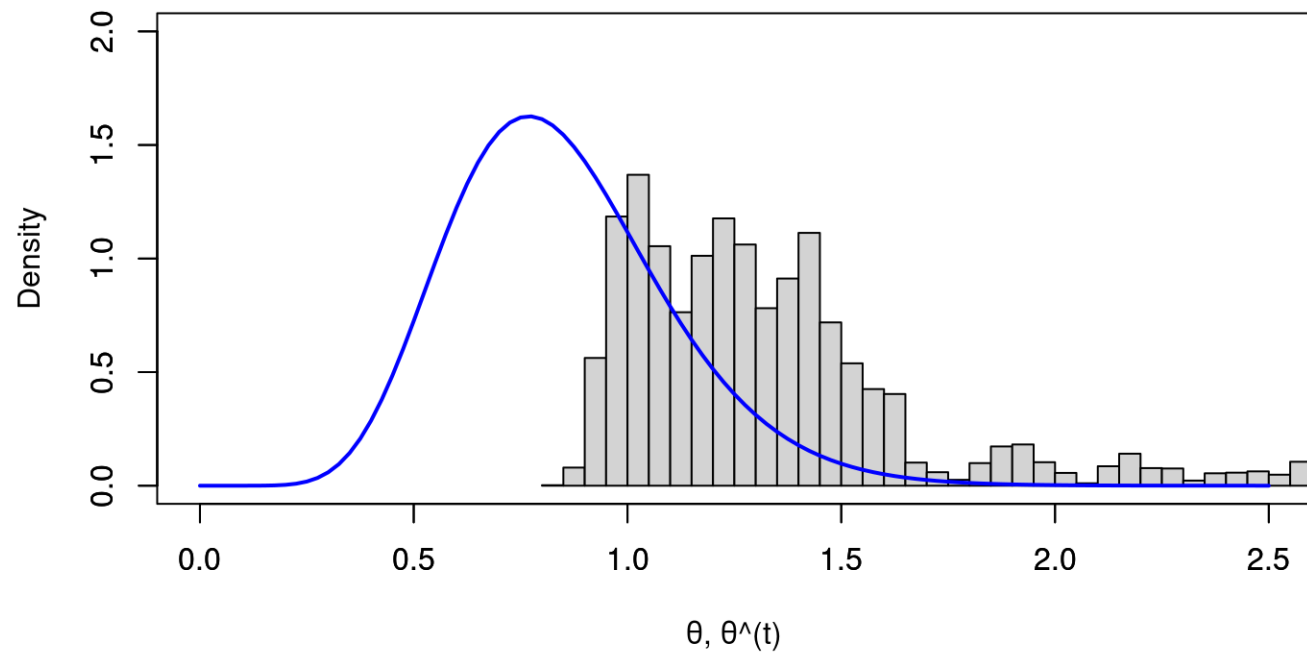
乱数 θ^t のトレースプロット



乱数 θ^t のトレースプロット



事後分布(青線)とサンプリング結果(ヒストグラム)



先の例題ではポアソン分布のパラメータ1つだけの推定でしたが、
回帰モデルを推定する際、回帰モデルには複数推定するパラメータがあります。
なのでRWMHはその分だけノイズの分散 $\delta_{e1}, \delta_{e2}, \dots, \delta_{ek}$ を設定し、別々に乱数をサンプリングしなければなりません。
こうなるとハイパーパラメータの設定はさらに難しくなります。

回帰モデル

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$
$$\iff y_i | \mathbf{x}_i, \beta_0, \beta_1, \dots, \beta_k, \sigma \sim N(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}, \sigma^2)$$

乱数候補のランダムウォーク

$$\begin{pmatrix} a_{\beta_0}^{(t)} \\ a_{\beta_1}^{(t)} \\ \vdots \\ a_{\beta_k}^{(t)} \\ a_{\sigma}^{(t)} \end{pmatrix} = \begin{pmatrix} \beta_0^{(t-1)} \\ \beta_1^{(t-1)} \\ \vdots \\ \beta_k^{(t-1)} \\ \sigma^{(t-1)} \end{pmatrix} + \begin{pmatrix} e_{\beta_0} \\ e_{\beta_1} \\ \vdots \\ e_{\beta_k} \\ e_{\sigma} \end{pmatrix}$$

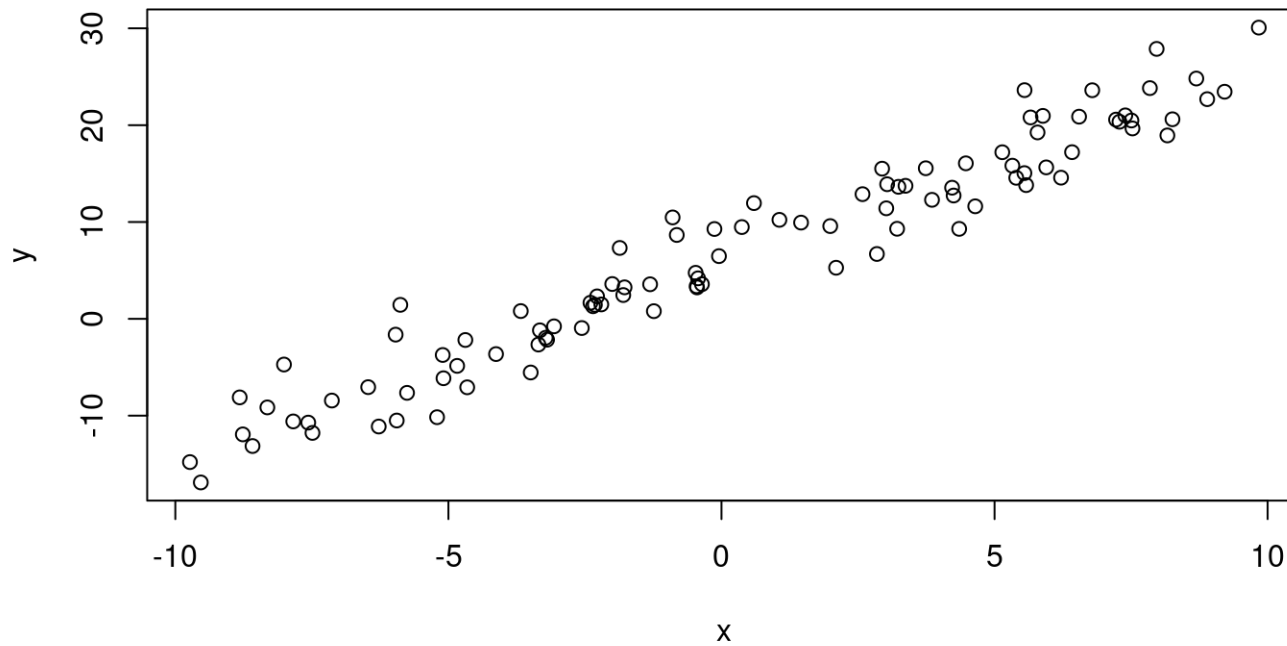
$$\begin{pmatrix} e_{\beta_0} \\ e_{\beta_1} \\ \vdots \\ e_{\beta_k} \\ e_{\sigma} \end{pmatrix} \sim N \begin{pmatrix} 0, \delta_{e_{\beta_0}}^2 \\ 0, \delta_{e_{\beta_1}}^2 \\ \vdots \\ 0, \delta_{e_{\beta_k}}^2 \\ 0, \delta_{e_{\sigma}}^2 \end{pmatrix}$$

$$\begin{pmatrix} e_{\beta_0} \\ e_{\beta_1} \\ \vdots \\ e_{\beta_k} \\ e_{\sigma} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_{e\beta_0}^2 & 0 & \cdots & 0 & 0 \\ 0 & \delta_{e\beta_1}^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \delta_{e\beta_k}^2 & 0 \\ 0 & 0 & \cdots & 0 & \delta_{e\sigma}^2 \end{pmatrix} \right)$$

RWMH法: 複数パラメータの場合

<例題>

以下のデータを回帰する.



推定するモデルは以下の通り.

$$y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$$

$$\iff y_i | X_i, \beta_0, \beta_1, \sigma \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

発生させる乱数は以下の通り

$$\begin{pmatrix} a_{\beta_0}^{(t)} \\ a_{\beta_1}^{(t)} \\ a_{\sigma}^{(t)} \end{pmatrix} = \begin{pmatrix} \beta_0^{(t-1)} \\ \beta_1^{(t-1)} \\ \sigma^{(t-1)} \end{pmatrix} + \begin{pmatrix} e_{\beta_0} \\ e_{\beta_1} \\ e_{\sigma} \end{pmatrix}$$

$$\begin{pmatrix} e_{\beta_0} \\ e_{\beta_1} \\ e_{\sigma} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \delta_{e\beta_0}^2 & 0 & 0 \\ 0 & \delta_{e\beta_1}^2 & 0 \\ 0 & 0 & \delta_{e\sigma}^2 \end{pmatrix} \right)$$

$$e_{\sigma} \geq 0$$

