

マーケティング・リサーチ特論 ～階層ベイズモデルとMCMC法～

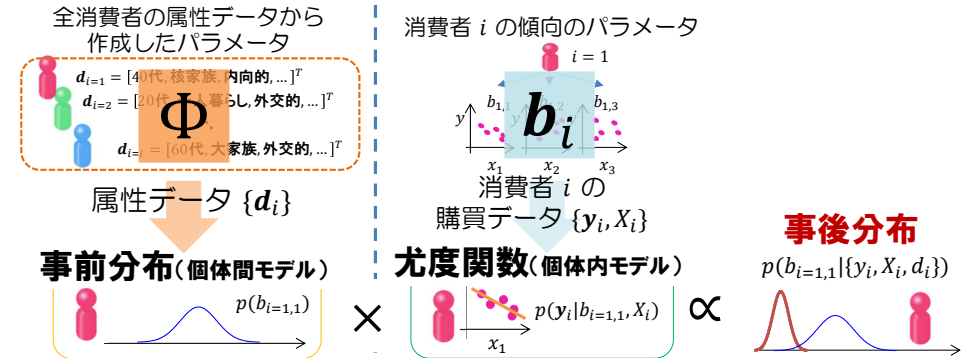
2024年度1学期： 水曜3限
担当教員： 石垣 司

復習：階層ベイズ線形回帰モデルでやりたい事

階層モデル

個体間モデル $p(\mathbf{b}_i) : \mathbf{b}_i = \Phi \mathbf{d}_i + \boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_i \sim N_p(\mathbf{0}, V)$

個体内モデル $y_i = X_i \mathbf{b}_i + e_i, e_i \sim N(0, \sigma_i^2 I)$



個人の係数 $\{\mathbf{b}_i\}$ をベイズモデルで安定的に推定する仕組み

階層ベイズ線形回帰モデル

回帰モデルと事前分布の階層性的関係図

IW: 逆ウィシャート分布
IG: 逆ガンマ分布

全消費者のパラメータ

Φ の事前分布

V の事前分布

$$\text{vec}(\Phi) \sim N_{(P+1)(L+1)}(\text{vec}(\tilde{M}_\Phi), \tilde{\Sigma}_\Phi) \quad V \sim IW(\tilde{\Psi}, \tilde{\nu})$$

各消費者の i の推定対象

σ_i^2 の事前分布

$$\mathbf{b}_i \text{ の事前分布 } \mathbf{b}_i \sim N_{P+1}(\Phi \mathbf{d}_i, V) \quad \sigma_i^2 \sim IG(\tilde{\alpha}, \tilde{\beta})$$

$$\text{観測データ } y_i = X_i \mathbf{b}_i + e_i, e_i \sim N(0, \sigma_i^2 I)$$

事後分布: $p(\{\mathbf{b}_i\}, \{\sigma_i^2\}, \Phi, V | \{y_i\}, \{X_i\}, \{d_i\}) \propto$

$$p(\{\sigma_i^2\}) p(\Phi, V) \prod_{i=1}^I p(\{\mathbf{b}_i\} | \Phi, V, \{d_i\}) p(\{y_i\} | \{\mathbf{b}_i\}, \{\sigma_i^2\}, \{X_i\})$$

σ_i^2 の事前分布
 Φ, V の事前分布

\mathbf{b}_i の事前分布

尤度関数

この先の話の流れの整理

各パラメータ $\{\mathbf{b}_i\}, \{\sigma_i^2\}, \Phi, V$ をベイズ推定したい

– しかしながら、事後分布 $p(\{\mathbf{b}_i\}, \{\sigma_i^2\}, \Phi, V | \{y_i\}, \{X_i\}, \{d_i\})$ の解析的な導出は困難

1. 今回: マルコフ連鎖モンテカルロ法(MCMC法)を用いた乱数による事後分布の近似アルゴリズムを利用する
2. 次回: MCMC法の効率的なアルゴリズムである Gibbs Sampling 法を用いるため、4つの完全条件付き事後分布を解析的に導出する

$$p(\mathbf{b}_i | \{\mathbf{b}_{-i}\}, \{\sigma_i^2\}, \Phi, V, \{y_i\}, \{X_i\}, \{d_i\})$$

$$p(\sigma_i^2 | \{\sigma_{-i}^2\}, \{\mathbf{b}_i\}, \Phi, V, \{y_i\}, \{X_i\}, \{d_i\})$$

$$p(\Phi | \{\mathbf{b}_i\}, \{\sigma_i^2\}, V, \{y_i\}, \{X_i\}, \{d_i\})$$

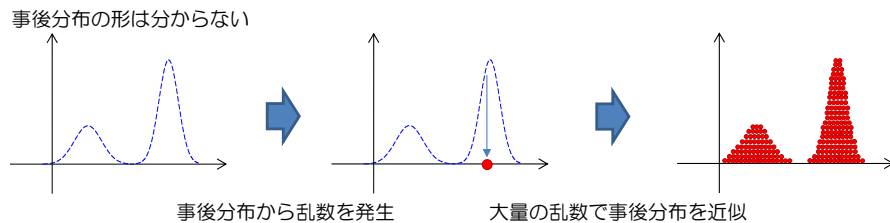
$$p(V | \{\mathbf{b}_i\}, \{\sigma_i^2\}, \Phi, \{y_i\}, \{X_i\}, \{d_i\})$$

マルコフ連鎖モンテカルロ法 #1

Markov chain Monte Carlo (MCMC) Method

マルコフ連鎖を用いることで特定の確率分布からの乱数を発生させる方法の総称

- MCMC法はほとんどの事後分布の推定で利用できる
共役事前分布を利用した事後分布の推定は特定の場合のみ
- 乱数の数の増大で目標とする事後分布に一致する
- 本授業: Metropolis-Hastings (M-H法), Gibbs sampling 法



5

マルコフ連鎖モンテカルロ法 #2

Markov chain Monte Carlo (MCMC) Method

モンテカルロ法

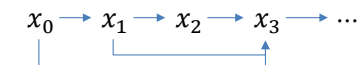
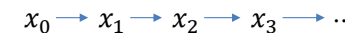
- 乱数を用いた数値計算やシミュレーション手法の総称
- 一様分布に従う疑似乱数は数論的手法によりコンピュータ上で生成できる
Python や R の疑似乱数生成は Mersenne Twister 法
正規分布に従う乱数(正規乱数)も一様乱数を変換し生成する

マルコフ連鎖

- 確率変数の列(正確には確率過程) $\{x_0, x_1, \dots, x_T\}$ がすべての t に対して $p(x_t|x_{t-1}) = p(x_t|x_0, x_1, \dots, x_{t-1})$ を満たすとき、確率変数の列 $\{x_0, x_1, \dots, x_T\}$ はマルコフ連鎖であるという

マルコフ連鎖

非マルコフ連鎖



6

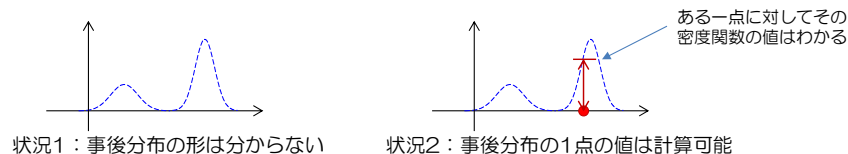
本授業での事後分布推定の状況

状況1: 事後分布が所属する分布のクラスは分かるが、その確率密度関数の具体的な形は分からない

- 例: $N(b|\mu^?, \Sigma^?)$ は多変量正規分布の確率密度関数。しかし、その平均ベクトルと分散共分散行列の値は不明

状況2: 変数が固定されたら事後分布の確率密度関数(または、そのカーネル部分)の値は分かる

- 例: $\mu^?, \Sigma^?$ の値が $\hat{\mu}, \hat{\Sigma}$ などと固定されれば $N(b|\hat{\mu}, \hat{\Sigma})$ の確率密度関数の値は容易に計算できる

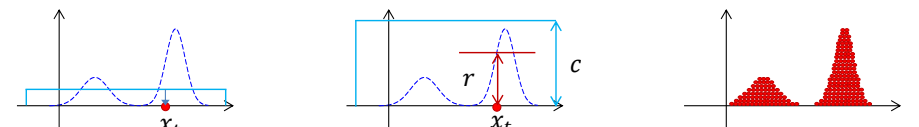


7

MCMC法を導入する必要性

事後分布の推定には効率的なアルゴリズムが必要

- 例: 棄却法(状況1&2の下で実行できる単純な近似推定法)



1. 正規分布や一様分布など簡単に乱数を生成できる分布を用いて、乱数 x_t を発生させる。(図では一様分布)
2. 事後分布をすべて覆うように一様分布を定数倍する。 c と r の値を計算。一様分布 $U[0,1]$ から値を1つ生成し、その値が r/c よりも小さければ乱数 x_t を採択。大きければ棄却
3. 1と2の操作を大量に繰り返して、採択された乱数のヒストグラムで事後分布を近似

- しかし、多変量の確率分布を扱う場合には、変数の数の指数乗のオーダーで計算量と棄却率が增大する(次元の呪い)
- 効率的な乱数の生成法がMCMC法

8

マルコフ連鎖の状態推移 #1

マルコフ連鎖の状態推移とその確率

- $\pi(x)$: 推定したい事後分布の確率密度関数(目標分布)
 $x_t \sim \pi(x)$ となるマルコフ連鎖を構成したい
- $f(x|x_t)$: 状態推移の確率密度関数(推移カーネル密度)
 $x_t \sim \pi(x)$ となるマルコフ連鎖は $f(x|x_t)$ に従い状態推移する
 f 自体は t で変化しない関数であることに注意
- \mathcal{X} : 状態空間(各状態を取りうる値の集合)
 $\mathcal{X} = \mathbb{R}, \mathcal{X} = \mathbb{R}^n, \mathcal{X} = \mathbb{R}^{n \times m}$ などを考えるとイメージしやすい
- マルコフ連鎖の状態推移確率(推移カーネル)

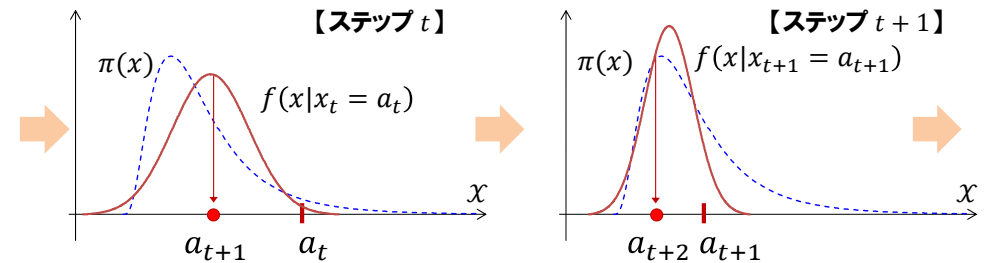
$$p(x_{t+1} \in A | x_t = a_t) = \int_A f(x|x_t = a_t) dx \quad (a_t, x \in \mathcal{X}, A \subset \mathcal{X})$$

x_t の実現値が a_t のときに x_{t+1} のとる値が範囲 A の中に推移する確率

マルコフ連鎖の状態推移 #2

マルコフ連鎖の状態推移と実現値の生成イメージ

- 乱数で近似したいのは $\pi(x)$ (図中の ---) の形
- もしも, $f(x|x_t)$ が既知かつ乱数生成可能ならば, 下記の繰り返しで効率的に $\pi(x)$ を近似するためのマルコフ連鎖を生成できる



- そのような $f(x|x_t)$ は存在するのか?

図中の $f(x|x_t)$ の関数はイメージ。正確な推移カーネル密度の形ではない

マルコフ連鎖の状態推移 #3

定義: 次式を満たす $\pi(x)$ は定常分布(不変分布)とよばれる

$$\pi(y) = \int f(y|x)\pi(x)dx, x, y \in \mathcal{X}$$

定理: 斉次的, 既約, 非周期的なマルコフ連鎖はただ一つの定常分布をもつ

- かつ, 任意の初期分布から $t \rightarrow \infty$ で定常分布に収束する
- 斉次的: (大雑把に言うと)マルコフ連鎖の推移確率 $f(x|x_t)$ が時点 t で変化しない性質
- 既約性: (大雑把に言うと)どの状態からマルコフ連鎖をスタートしても \mathcal{X} のすべての領域を有限回の推移で動き回ることができる性質
- 非周期性: (大雑把に言うと)マルコフ連鎖が一定の決まった周期である状態に必ず戻ってくるということはない性質

つまり, $\pi(x)$ を定常分布とする $f(x|x_t)$ を設計できると, マルコフ連鎖 $x_t \sim \pi(x)$ を生成できることが保証されている

(大雑把に言うと)の部分の数学的な説明は割愛。参考書などを参考にしてほしい

マルコフ連鎖の状態推移 #4

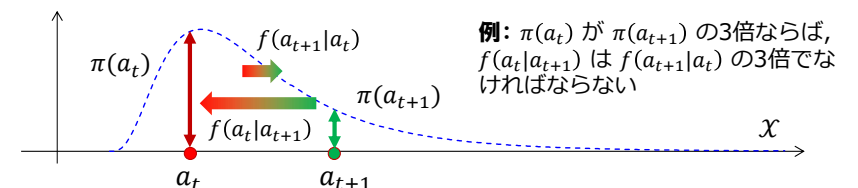
詳細つり合い条件

- $\pi(x)$ が定常分布となるための十分条件

$$\pi(a_t)f(a_{t+1}|a_t) = \pi(a_{t+1})f(a_t|a_{t+1})$$

$$\because \pi(x)f(y|x) = \pi(y)f(x|y) \Rightarrow \int \pi(x)f(y|x)dx = \int \pi(y)f(x|y)dx = \pi(y)$$

\mathcal{X} 上のすべての2点でこのバランスが成立している状態は定常分布



- 詳細つり合い条件を満たすように $f(x|x_t)$ を決めれば, $x_t \sim \pi(x)$ となるマルコフ連鎖を生成できる

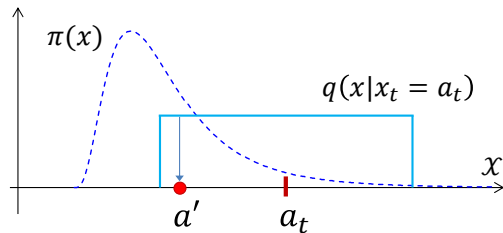
Metropolis-Hastings (M-H) 法 #1

提案分布

- $q(x|x_t)$: 提案分布の確率密度関数
簡単に乱数を生成できる確率分布なら何でもよい
- a' : 提案分布から発生させた乱数。 x_{t+1} の実現値の候補

M-H法のストラテジー

- $f(x|x_t)$ は分からないが、提案分布から発生させた乱数 a' を推移確率 $f(x|x_t)$ に従って発生させた乱数とするためのアルゴリズム



13

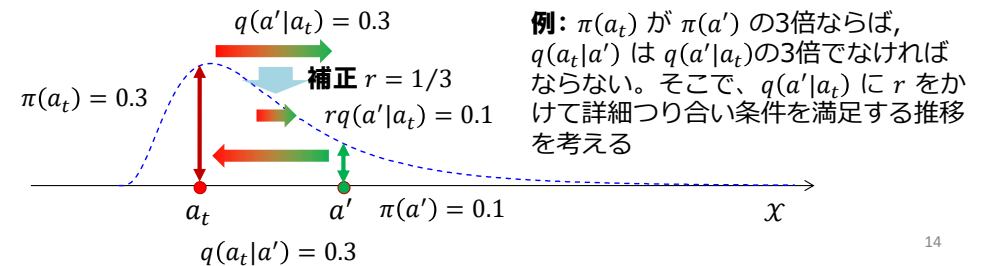
Metropolis-Hastings (M-H) 法 #2

提案分布から発生させた乱数の詳細つり合い条件

$$\pi(a_t)r q(a'|a_t) = \pi(a')q(a_t|a'), r = \frac{\pi(a')q(a_t|a')}{\pi(a_t)q(a'|a_t)}$$

- 比率 $\frac{\pi(a')}{\pi(a_t)} = \frac{r q(a'|a_t)}{q(a_t|a')}$ を保った推移の定常分布は $\pi(x)$ となる

M-H 法ではこの比を保つために、 $\pi(x)$ の値が大きい方から小さい方への移動に対して r で補正する



14

Metropolis-Hastings (M-H) 法 #3

M-H 法のアルゴリズム

- 適当な初期値 $x_{t=0} = a_0$ を与える
 $t = 0, 1, \dots, T$ において以下を繰り返す

 1. 提案分布 $q(x|x_t)$ から a' を発生させ $r = \frac{\pi(a')q(a_t|a')}{\pi(a_t)q(a'|a_t)}$ を計算
 2. 一様分布から $U[0,1]$ から u を発生させる
 3. $\min(1, r) \geq u$ なら $a_{t+1} = a'$, それ以外は $a_{t+1} = a_t$ とする
 4. 1 へ戻る

$\pi(a_t) > \pi(a')$ のとき、 $a' \rightarrow a_t$ へ推移するかしないかを確率 r で補正。
 $\pi(a') > \pi(a_t)$ のときは補正無しで $a_t \rightarrow a'$ へ推移する

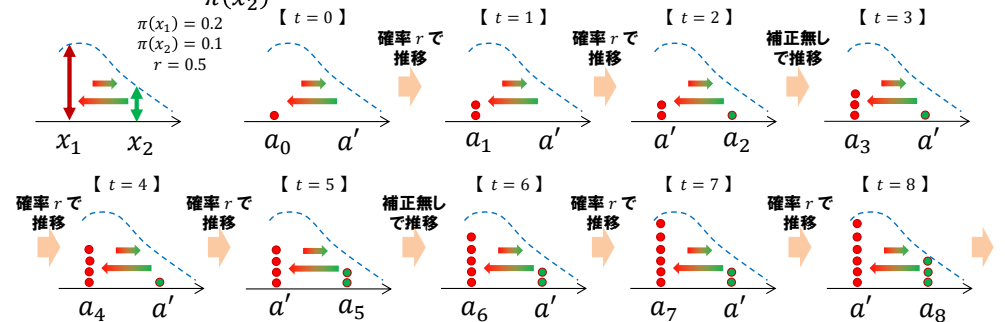
 - 十分大きな T に対してマルコフ連鎖のサンプルパスの部分列 $\{a_{T-M}, \dots, a_{T-1}, a_T\}$ (等) のヒストグラムを確率密度関数 $\pi(x)$ の近似とする

15

Metropolis-Hastings (M-H) 法 #4

M-H 法のアルゴリズムのイメージ

- 例: $\frac{\pi(x_1)}{\pi(x_2)} = 2$ で x_1 と x_2 にしか動かないときのイメージ



- \mathcal{X} 上のすべての2点でこのバランスが成立するように確率 r で状態推移を調整している

この例は離散マルコフ連鎖 $\pi = [2/3 \ 1/3], T = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$ と同様。左の π は T の定常分布

16

M-H 法と提案分布の例

ランダムウォーク連鎖(Metropolis法)

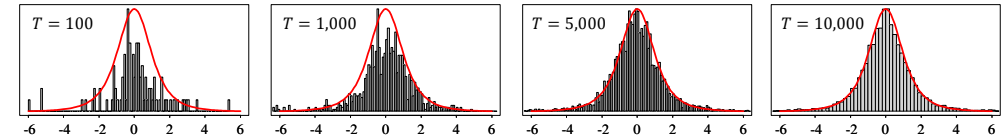
- 提案分布に左右対称な分布(例:正規分布)を採用し,
 $a' = a_t + e_t, e_t \sim N(0, \alpha)$ として候補を作成
- $q(a_t|a') = q(a'|a_t)$ なので $r = \pi(a')/\pi(a_t)$ と簡略化
- 分散パラメータ α (ステップ・サイズとよばれる)の値が小さいと採択率は大きいが \mathcal{X} 全体を動くのに時間がかかる。値が大きいと採択率が小さくなる

独立連鎖

- 提案分布を $q(x|x_t)$ から $q(x)$ へ簡略化
- $\pi(x)$ の形状に近い $q(x)$ を用いると採択率が高い効率的な連鎖となる

MCMC法の実行

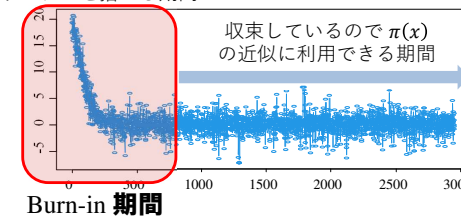
連鎖の長さ T と $\pi(x)$ 近似精度(例: 自由度 3 の t 分布の近似)



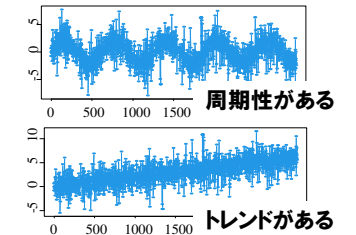
- マーケティング・リサーチでは数万程度の T が用いられる傾向

収束と Burn-in 期間

定常分布に収束していないので
 サンプルを捨てる期間



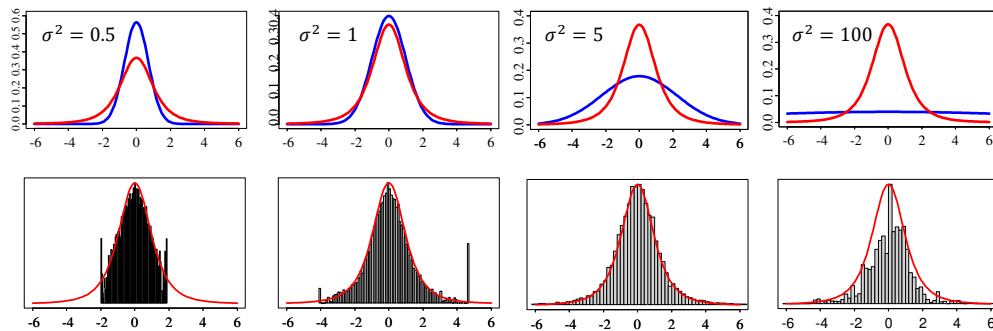
このような結果は利用できない
 望ましいマルコフ連鎖ではない例



M-H法と提案分布

独立連鎖の提案分布(例: $T = 30,000$ での自由度 3 の t 分布の近似)

- 提案分布 $N(0, \sigma^2)$ — 目標分布 — 提案分布



両裾の形状に対応できていない

ちょうど良い
 提案分布

採択率が低く
 $T = 30,000$ では足りない
 ($T =$ 百万程度では収束)

M-H法では適切な連鎖を得るための試行錯誤が必要

Gibbsサンプリング

完全条件付き分布を利用したMCMC法

- 推定対象が多変量、かつ、完全条件付き分布を解析的に導出できる場合に利用できる
- M-H法の一つで採択率 $r = 1$ となるアルゴリズム
 利用できる条件はM-H法と比べて限られるが、利用できれば効率的なアルゴリズム
- ランダムウォーク連鎖のステップサイズの値や独立連鎖の提案分布など、M-H法で必要な試行錯誤が不要
- 本授業の本題は、階層ベイズ線形回帰モデルの事後分布推定のためのGibbsサンプリング法の導出

Gibbsサンプリングと完全条件付き分布

完全条件付き分布

- あるパラメータに対して他のパラメータで条件付けしたときの条件付き分布
- 例: 3変数 $x = [x, y, z]$ の同時分布 $f(x, y, z)$ に対して, $f(x|y, z), f(y|x, z), f(z|x, y)$ がそれぞれ完全条件付き分布

“完全条件付き分布を解析的に導出できる”

⇒ $\pi(x|x_t)$ を提案分布 $q(x|x_t)$ として利用できる

⇒ 採択率が 1

$$r = \frac{\pi(a')q(a_t|a')}{\pi(a_t)q(a'|a_t)} = \frac{\pi(a')\pi(a_t|a')}{\pi(a_t)\pi(a'|a_t)} = \frac{\pi(a', a_t)}{\pi(a_t, a')} = 1$$

21

Gibbsサンプリング

Gibbs サンプリングのアルゴリズム

- 推定対象を $x_{1t}, x_{2t}, \dots, x_{pt}$ とし, 適当な初期値 $(x_{1t=0}, x_{2t=0}, \dots, x_{pt=0}) = (a_{10}, a_{20}, \dots, a_{p0})$ を与える
- $t = 0, 1, \dots, T$ において以下を繰り返す

$$\begin{cases} a_{1,t+1} \sim \pi(x_1 | a_{2t}, \dots, a_{pt}) \\ a_{2,t+1} \sim \pi(x_2 | a_{1,t+1}, a_{3t}, \dots, a_{pt}) \\ \vdots \\ a_{p,t+1} \sim \pi(x_p | a_{1,t+1}, \dots, a_{p-1,t}) \end{cases} \quad x_1, \dots, x_p \in \mathcal{X}$$

- 各乱数は常に採択されるため効率がよいマルコフ連鎖
- 次回の授業内容: 階層ベイズ線形回帰モデルへGibbsサンプリングを適用するための条件付き事後分布の導出と実例

22

M-H法とGibbsサンプリング混合法

M-H法とGibbs サンプリングの混合アルゴリズム

- 推定対象を $x_{1t}, \dots, x_{pt}, y_{1t}, \dots, y_{pt}$ とし, x_{it} を完全条件付き分布が導出可能な変数, y_{it} を導出できない変数とする
- 適当な初期値の下で, $t = 0, 1, \dots, T$ において以下を繰り返す

$$\begin{array}{l} \text{Gibbsサンプリング} \\ \text{による乱数生成} \\ \uparrow \\ \left(\begin{array}{l} a_{x_1,t+1} \sim \pi(x_1 | a_{x_2t}, \dots, a_{y_{pt}}) \\ \vdots \\ a_{x_{pt+1}} \sim \pi(x_p | a_{x_1t+1}, \dots, a_{x_{p-1}t+1}, a_{y_1t}, \dots, a_{y_{pt}}) \\ \downarrow \\ a_{y_1,t+1} \leftarrow q(y_1 | a_{x_1t+1}, \dots, a_{x_{pt+1}}, a_{y_2t}, \dots, a_{y_{pt}}) \\ \vdots \\ a_{y_{p,t+1}} \leftarrow q(y_p | a_{x_1t+1}, \dots, a_{y_{p-1}t+1}) \end{array} \right) \\ \text{M-Hアルゴリズム} \\ \text{による乱数生成} \end{array} \quad \left[\leftarrow \right] \text{はM-Hアルゴリズムで発生させた乱数の意味}$$

$x_1, \dots, x_p, y_1, \dots, y_p \in \mathcal{X}.$

M-H法もGibbsサンプリングも詳細つり合い条件を満たしている推移であることからこの混合アルゴリズムも定常分布に収束する

23