

ベイズ統計学のMCMCとの出会い



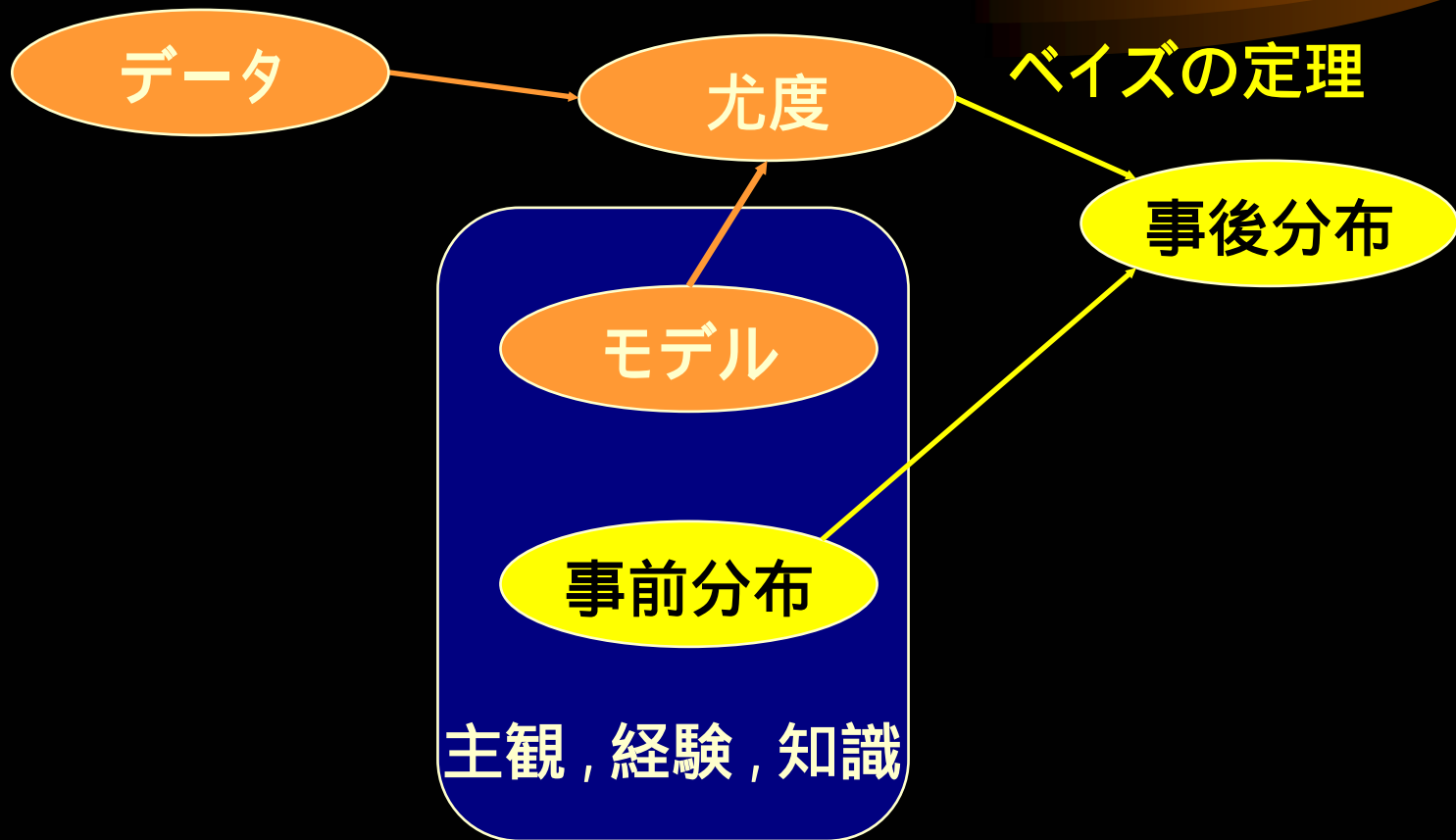
2004年10月15日
関西SASユーザー会

塩野義製薬(株)
町田光陽, 長谷川貴大, 田崎武信

ご紹介内容

1. ベイズ統計のMCMCとの出会い
2. ベイズ統計で最尤推定
3. ベイズ統計と混合効果モデルの関係
4. ベイズ論文の読みかた(私案)

ベイズ統計学



ベイズの定理

$$p(\theta | x) = \frac{p(x | \theta)p(\theta)}{\int p(x | \theta)p(\theta)d\theta}$$

$p(x | \theta)$ は統計学一般で利用される密度である． θ についての関数としてとらえると尤度にあたる．

右辺の $p(\theta)$ は事前(データがとられる前の)確率密度とよばれる．これはベイズ統計学でのみ登場する．

左辺の $p(\theta | x)$ は事後確率密度とよばれる．

分母の積分計算は困難な場合が多い．

MCMC (マルコフ連鎖モンテカルロ)

複雑な多次元分布から乱数を発生させる方法

MCMCの代表的なもの

Metropolis-Hastingsアルゴリズム

密度関数 $\pi(\cdot)$ をもつ分布に従う乱数を生成する Metropolis-Hastingsアルゴリズム

- ・初期値を $x^{(0)}$ とする.
- ・ $j = 1, 2, \dots, N$ について以下を繰り返す.

1. $q(x^{(j-1)}, \cdot)$ から y , $U(0,1)$ から u を生成する.

2. $u \leq \alpha(x^{(j-1)}, y)$ であれば $x^{(j)} = y$ とおく.

ここに $\alpha(x^{(j-1)}, y) = \min\left\{1, \frac{\pi(y)q(y, x^{(j-1)})}{\pi(x^{(j-1)})q(x^{(j-1)}, y)}\right\}$ である (数学的に厳密な表現はあえて避けた).

3. そうでなければ $x^{(j)} = x^{(j-1)}$ とおく.

- ・結果 $\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ を得る.

このアルゴリズムは候補発生密度 $q(x^{(j-1)}, \cdot)$ により特定される.

事後密度 $p_{\text{posterior}}(\theta | x) = \frac{p(x | \theta) p_{\text{prior}}(\theta)}{\int p(x | \theta) p_{\text{prior}}(\theta) d\theta}$ をもつ分布に従う乱数を生成したい。

このときMetropolis-Hastingsアルゴリズムが利用できる。

- ・初期値を $\theta^{(0)}$ とする。
- ・ $j = 1, 2, \dots, N$ について以下を繰り返す。

1. $q(\theta^{(j-1)}, \cdot)$ から $\theta^{\text{candidate}}$, $U(0,1)$ から u を生成する。

2. $u \leq \alpha(\theta^{(j-1)}, \theta^{\text{candidate}})$ であれば $\theta^{(j)} = \theta^{\text{candidate}}$ とおく。

ここに、 $\alpha(\theta^{(j-1)}, \theta^{\text{candidate}}) = \min \left\{ 1, \frac{p(x | \theta^{\text{candidate}}) p_{\text{prior}}(\theta^{\text{candidate}}) q(\theta^{(j-1)}, \theta^{\text{candidate}})}{p(x | \theta^{(j-1)}) p_{\text{prior}}(\theta^{(j-1)}) q(\theta^{\text{candidate}}, \theta^{(j-1)})} \right\}$ である。

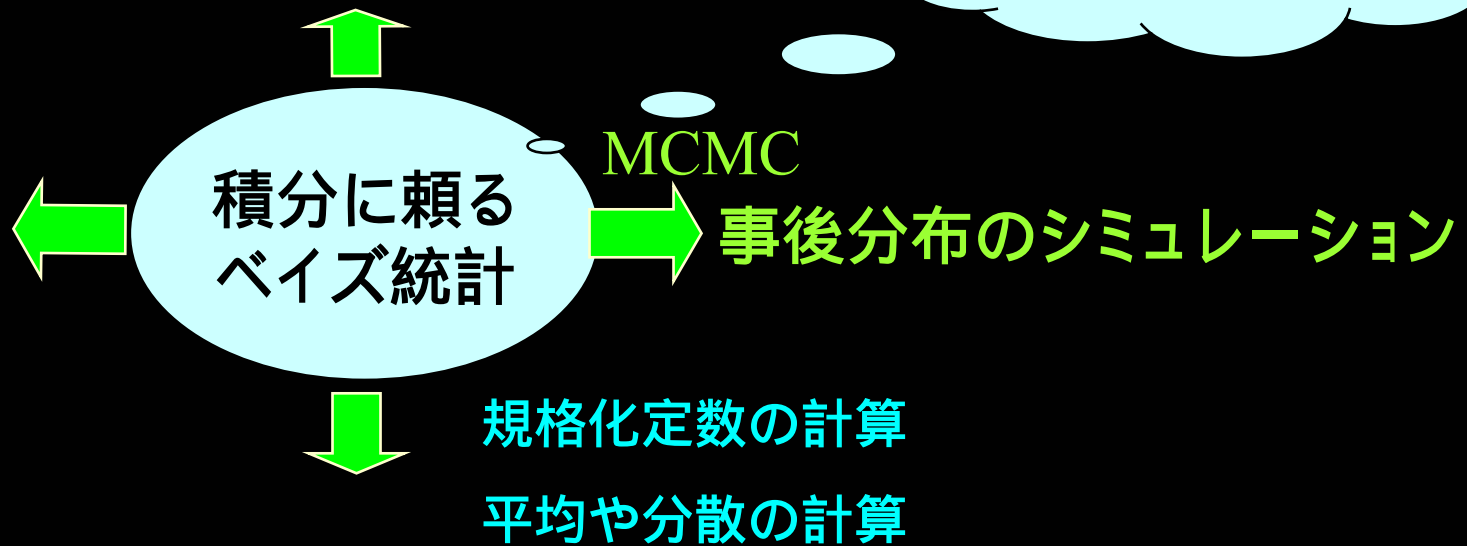
3. そうでなければ $\theta^{(j)} = \theta^{(j-1)}$ とおく。

・結果 $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)}\}$ を得る。

事後密度の規格化定数
(分母の積分の項)
が必要でない。

このアルゴリズムは候補発生密度 $q(\theta^{(j-1)}, \cdot)$ により特定される。

MCMCがベイズ統計の世界を広げた



θ がベクトル (θ_1, θ_2) の場合, 事後密度 $p_{posterior}(\theta | x)$ をもつ分布に従う乱数を生成したい.

θ_1, θ_2 の事前分布はそれぞれ独立, すなわち

$$p_{prior}(\theta_1, \theta_2) = p_{prior}(\theta_1)p_{prior}(\theta_2)$$

と仮定する. このとき同時事後密度は,

$$p_{posterior}(\theta_1, \theta_2 | x) = \frac{p(x | \theta_1, \theta_2)p_{prior}(\theta_1)p_{prior}(\theta_2)}{\iint p(x | \theta_1, \theta_2)p_{prior}(\theta_1)p_{prior}(\theta_2)d\theta_1d\theta_2}$$

となる.

Metropolis-Hastingsアルゴリズムを利用する際の注意点

Metropolis-Hastingsアルゴリズムで、 y はベクトルであってもよい。しかし、 y がベクトルであれば、現実には、 $q(x^{(j-1)}, \cdot)$ から y を直接生成する、というステップが可能になる例は少ないはずである。

したがって、Metropolis-Hastingsアルゴリズムを利用するためには、 y をスカラーのように扱う工夫が必要である。



MCMCの代表的なもの

Metropolis-Hastingsアルゴリズム

↓ y をスカラのように扱う工夫

- 1 . Single Component Metropolis-Hastingsアルゴリズム
- 2 . Gibbsサンプリング

Single Component Metropolis-Hastingsアルゴリズムを利用する場合

・初期値を $\begin{pmatrix} \theta_1^{(0)} \\ \theta_2^{(0)} \end{pmatrix}$ とする.

・ $j = 1, 2, \dots, N$ について以下を繰り返す.

1. $\theta_1^{(j)}$ を発生させる.

(1) $q_1 \left\{ \begin{pmatrix} \theta_1^{(j-1)} \\ \theta_2^{(j-1)} \end{pmatrix}, \cdot \right\}$ から $\theta_1^{candidate}$, $U(0,1)$ から u を発生する.

(2) $u \leq \alpha_1(\theta_1^{(j-1)}, \theta_1^{candidate})$ であれば, $\theta_1^{(j)} = \theta_1^{candidate}$ とおく.

$$\text{ここに, } \alpha_1(\theta_1^{(j-1)}, \theta_1^{candidate}) = \min \left\{ 1, \frac{p_{\text{posterior}}(\theta_1^{candidate} | x, \theta_2^{(j-1)}) q_1 \left\{ \begin{pmatrix} \theta_1^{candidate} \\ \theta_2^{(j-1)} \end{pmatrix}, \theta_1^{(j-1)} \right\}}{p_{\text{posterior}}(\theta_1^{(j-1)} | x, \theta_2^{(j-1)}) q_1 \left\{ \begin{pmatrix} \theta_1^{(j-1)} \\ \theta_2^{(j-1)} \end{pmatrix}, \theta_1^{candidate} \right\}} \right\} \text{である.}$$

(数学的に厳密な表現はあえて避けた).

フルの条件付き事後分布の密度関数

(3) そうでなければ $\theta_1^{(j)} = \theta_1^{(j-1)}$ とおく.

2. $\theta_2^{(j)}$ を発生させる.

(1) $q_2 \left\{ \begin{pmatrix} \theta_1^{(j)} \\ \theta_2^{(j-1)} \end{pmatrix}, \cdot \right\}$ から $\theta_2^{\text{candidate}}$, $U(0,1)$ から u を発生する.

(2) $u \leq \alpha_2(\theta_2^{(j-1)}, \theta_2^{\text{candidate}})$ であれば, $\theta_2^{(j)} = \theta_2^{\text{candidate}}$ とおく.

$$\text{ここに, } \alpha_2(\theta_2^{(j-1)}, y) = \min \left\{ 1, \frac{p_{\text{posterior}}(\theta_2^{\text{candidate}} | x, \theta_1^{(j)}) q_2 \left\{ \begin{pmatrix} \theta_1^{(j)} \\ \theta_2^{\text{candidate}} \end{pmatrix}, \theta_2^{(j-1)} \right\}}{p_{\text{posterior}}(\theta_2^{(j-1)} | x, \theta_1^{(j)}) q_2 \left\{ \begin{pmatrix} \theta_1^{(j)} \\ \theta_2^{(j-1)} \end{pmatrix}, \theta_2^{\text{candidate}} \right\}} \right\} \text{である.}$$

(数学的に厳密な表現はあえて避けた). **フルの条件付き事後分布の密度関数**

(3) そうでなければ $\theta_2^{(j)} = \theta_2^{(j-1)}$ とおく.

・ 結果 $\left\{ \begin{pmatrix} \theta_1^{(1)} \\ \theta_2^{(1)} \end{pmatrix}, \begin{pmatrix} \theta_1^{(2)} \\ \theta_2^{(2)} \end{pmatrix}, \dots, \begin{pmatrix} \theta_1^{(N)} \\ \theta_2^{(N)} \end{pmatrix} \right\}$ を得る.

このアルゴリズムは候補発生密度 $q_1 \left\{ \begin{pmatrix} \theta_1^{(c)} \\ \theta_2^{(c)} \end{pmatrix}, \cdot \right\}$ と $q_2 \left\{ \begin{pmatrix} \theta_1^{(c)} \\ \theta_2^{(c)} \end{pmatrix}, \cdot \right\}$ により特定される.

フルの条件付き事後分布の密度関数の導出

$$\begin{aligned} p_{\text{posterior}}(\theta_1 | x, \theta_2) &= \frac{p_{\text{posterior}}(\theta_1, \theta_2 | x)}{\int p_{\text{posterior}}(\theta_1, \theta_2 | x) d\theta_1} \\ &= \frac{p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) p_{\text{prior}}(\theta_2)}{\int p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) p_{\text{prior}}(\theta_2) d\theta_1} \\ &= \frac{p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) p_{\text{prior}}(\theta_2)}{p_{\text{prior}}(\theta_2) \int p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) d\theta_1} \\ &= \frac{p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1)}{\int p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) d\theta_1} \\ &\propto p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) \end{aligned}$$

$p_{\text{prior}}(\theta_1, \theta_2) = p_{\text{prior}}(\theta_1) p_{\text{prior}}(\theta_2)$
のとき

α_1 を計算する際に約分される
のでこの積分は必要ない。

尤度関数と θ_1 の事前密度の積に
比例する。
これは、 θ_1 と θ_2 の事前分布
を独立と仮定したためである。

同様に、

$$p_{\text{posterior}}(\theta_2 | x, \theta_1) \propto p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_2)$$

あるパラメタの、他のすべてのパラメタに条件付きの分布はフルの条件付き分布とよばれる。

フルの条件付き分布を利用することの正当性

$$\begin{aligned}
 \frac{p_{\text{posterior}}(\theta_1^{\text{candidate}}, \theta_2^{(j-1)} | x)}{p_{\text{posterior}}(\theta_1^{(j-1)}, \theta_2^{(j-1)} | x)} &= \frac{p(x | \theta_1^{\text{candidate}}, \theta_2^{(j-1)}) p_{\text{prior}}(\theta_1^{\text{candidate}}) p_{\text{prior}}(\theta_2^{(j-1)})}{\int \int p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) p_{\text{prior}}(\theta_2) d\theta_1 d\theta_2} \\
 &= \frac{p(x | \theta_1^{(j-1)}, \theta_2^{(j-1)}) p_{\text{prior}}(\theta_1^{(j-1)}) p_{\text{prior}}(\theta_2^{(j-1)})}{\int \int p(x | \theta_1, \theta_2) p_{\text{prior}}(\theta_1) p_{\text{prior}}(\theta_2) d\theta_1 d\theta_2} \\
 \theta_1 \text{ と } \theta_2 \text{ の同時事後密度の比} &= \frac{p(x | \theta_1^{\text{candidate}}, \theta_2^{(j-1)}) p_{\text{prior}}(\theta_1^{\text{candidate}})}{p(x | \theta_1^{(j-1)}, \theta_2^{(j-1)}) p_{\text{prior}}(\theta_1^{(j-1)})} \\
 &= \frac{p_{\text{posterior}}(\theta_1^{\text{candidate}} | x, \theta_2^{(j-1)})}{p_{\text{posterior}}(\theta_1^{(j-1)} | x, \theta_2^{(j-1)})} \quad \theta_2 \text{ に条件付きの } \theta_1 \text{ の事後密度の比} \\
 & \quad (\theta_1 \text{ についてのフルの条件付き事後密度の比})
 \end{aligned}$$

Gibbsサンプリングを利用する場合

- ・初期値を $\begin{pmatrix} \theta_1^{(0)} \\ \theta_2^{(0)} \end{pmatrix}$ とする.
- ・ $j = 1, \dots, N$ について以下を繰り返す .
 - 1 . $p_{\text{posterior}}(\cdot | x, \theta_2^{(j-1)})$ から $\theta_1^{(j)}$ を発生させる .
 - 2 . $p_{\text{posterior}}(\cdot | x, \theta_1^{(j)})$ から $\theta_2^{(j)}$ を発生させる .

結果 $\left\{ \begin{pmatrix} \theta_1^{(1)} \\ \theta_2^{(1)} \end{pmatrix}, \begin{pmatrix} \theta_1^{(2)} \\ \theta_2^{(2)} \end{pmatrix}, \dots, \begin{pmatrix} \theta_1^{(N)} \\ \theta_2^{(N)} \end{pmatrix} \right\}$ を得る .

Gibbsサンプリングの位置づけと特徴

< Single Component Metropolis-Hastingsアルゴリズムとの関係 >

GibbsサンプリングはSingle Component M-Hアルゴリズムでの候補発生分布にフルの条件付き分布を設定した場合に相当する。

< Gibbsサンプリングの長所 >

移動確率 α が必ず1になるため、手順が簡単でかつサンプリング効率がよい方法である。

< Gibbsサンプリングの短所 >

フルの条件付き事後分布に従う乱数を発生させる必要がある。フルの条件付き分布から乱数を直接発生させることが困難な場合は、棄却サンプリングなどを利用することが考えられる。

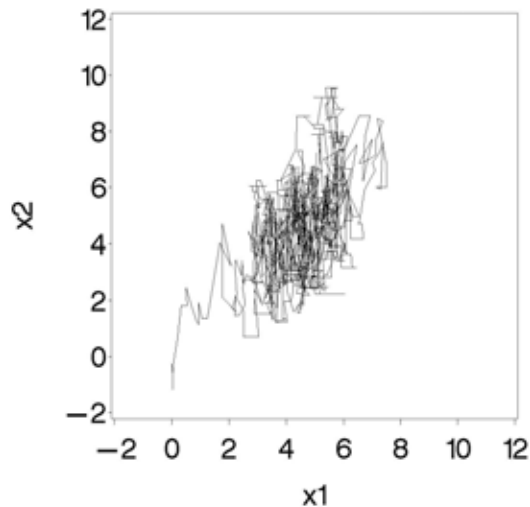
棄却サンプリングなどを利用してフルの条件付き分布に従う乱数を生成することが困難な場合は、Gibbsサンプリング以外の手法に頼らざるをえないと考えられる。

事例: 2変量正規分布のシミュレーション

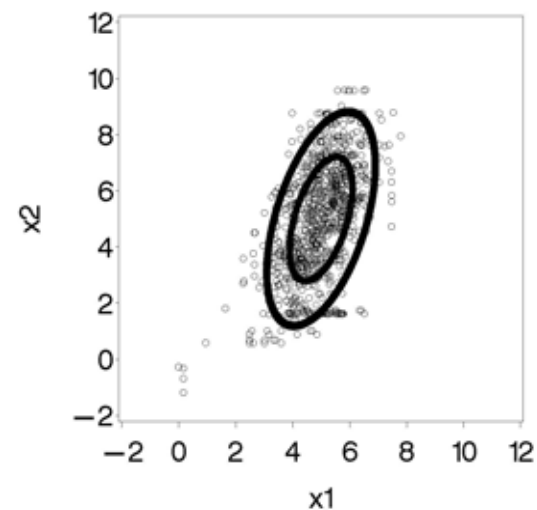
Single Component M-Hアルゴリズムによる, 平均ベクトル $\begin{pmatrix} 5 \\ 5 \end{pmatrix}$,
分散共分散行列 $\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$ をもつ2変量正規分布からのサンプリング

(設定) 初期値 $\begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $q_1\left\{\begin{pmatrix} x_1^{(j-1)} \\ x_2^{(j-1)} \end{pmatrix}, \cdot\right\}$ は, 平均 $x_1^{(j-1)}$, 標準偏差1の正規分布
 $q_2\left\{\begin{pmatrix} x_1^{(j)} \\ x_2^{(j-1)} \end{pmatrix}, \cdot\right\}$ は, 平均 $x_2^{(j-1)}$, 標準偏差1の正規分布

最初の1000回の点の軌跡



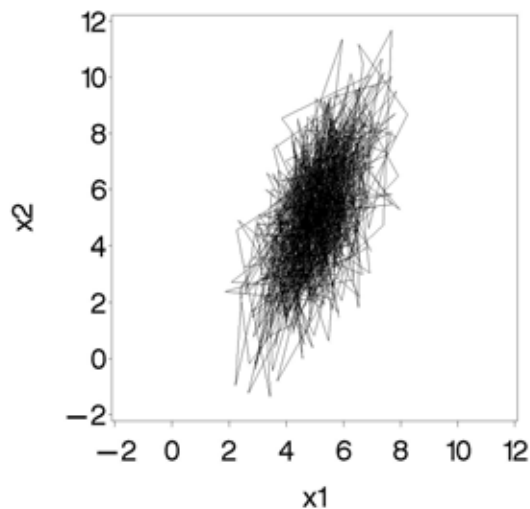
点と発生させたい2変量正規分布の等高線



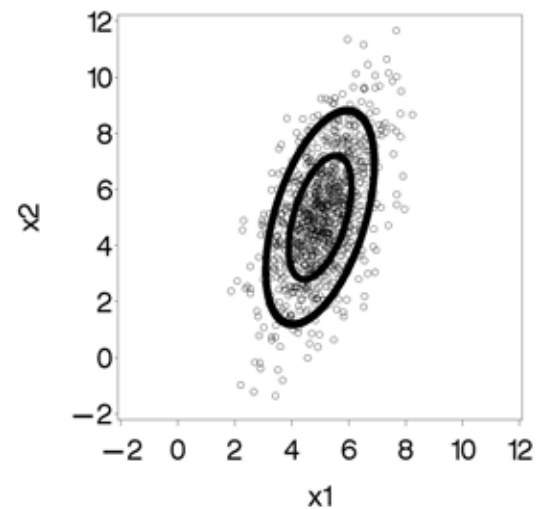
Gibbsサンプリングによる, 平均ベクトル $\begin{pmatrix} 5 \\ 5 \end{pmatrix}$, 分散共分散行列 $\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$ をもつ2変量正規分布からのサンプリング

(設定) 初期値 $\begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

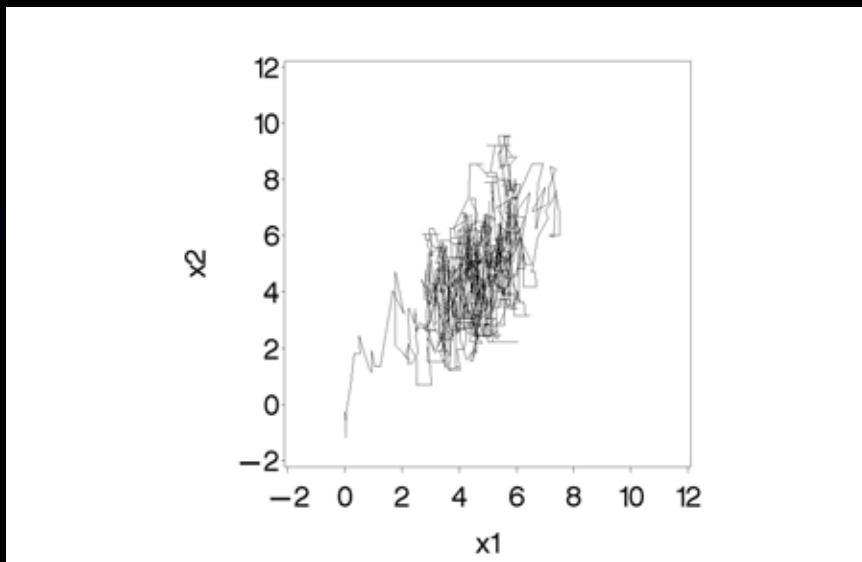
最初の1000回の点の軌跡



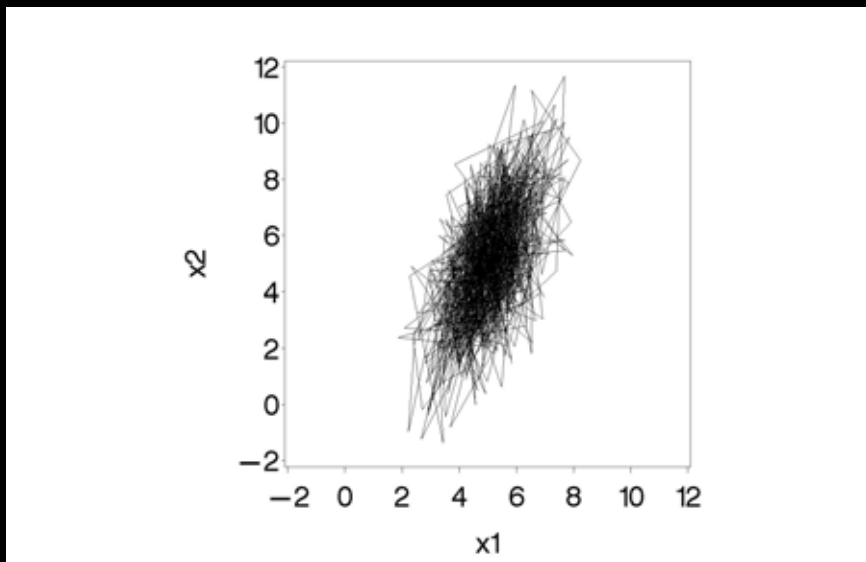
点と発生させたい2変量正規分布の等高線



Single Component M-Hアルゴリズム

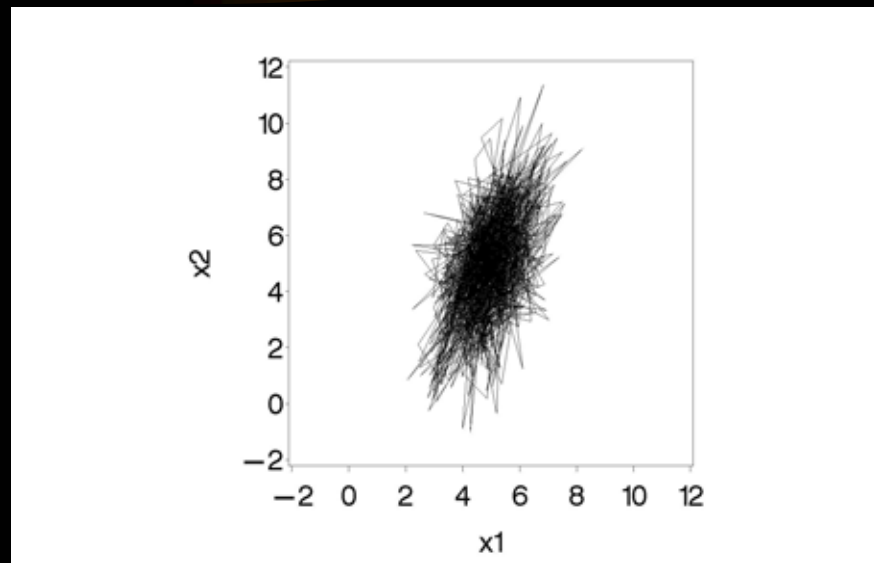


Gibbsサンプリング



最初の1000回の
点の軌跡の比較

従来法(コレスキー分解の応用)



ご紹介内容

1. ベイズ統計のMCMCとの出会い
2. **ベイズ統計で最尤推定**
3. ベイズ統計と混合効果モデルの関係
4. ベイズ論文の読みかた(私案)

ベイズ統計で最尤推定

ベイズ統計で無情報事前分布を設定した場合，得られる事後分布の密度関数は尤度関数に比例する．

$$p_{posterior}(\theta | x) = \frac{L(x | \theta) p_{prior}(\theta) d\theta}{\int L(x | \theta) p_{prior}(\theta) d\theta}$$

ここで $p_{prior}(\theta)$ が， θ に依存しない定数であれば

$$\begin{aligned} &= \frac{L(\theta | x)}{\int L(\theta | x) d\theta} \\ &\propto L(\theta | x) \end{aligned}$$

とくに，最尤推定値はベイズ統計で無情報事前分布を設定した場合の事後密度のモードに一致する．

事例

確率変数 X にパラメタ μ をもつポアソン分布 ,
確率変数 Y にパラメタ $\mu\theta$ をもつポアソン分布を想定したモデルをあてはめる .

$$X_i \sim Po(\mu), i = 1, \Lambda, m$$

$$Y_j \sim Po(\mu\theta), j = 1, \Lambda, n$$

このとき , μ の最尤推定量は \bar{X} , θ の最尤推定量は $\frac{\bar{Y}}{\bar{X}}$ となる .

$x = \{2, 4, 6, 6, 7\}$, $y = \{0, 0, 1, 1, 1, 1, 2, 2\}$ のとき ,

近似的な無情報事前分布を用いたベイズ統計の枠組みで最尤推定を行った .

μ および θ の各事前分布は $U(0, 100)$ とした .

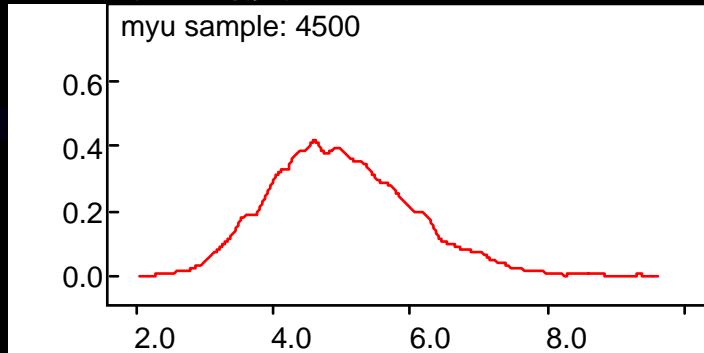
事後分布をMCMCでシミュレーション

Burn-in回数: 500回

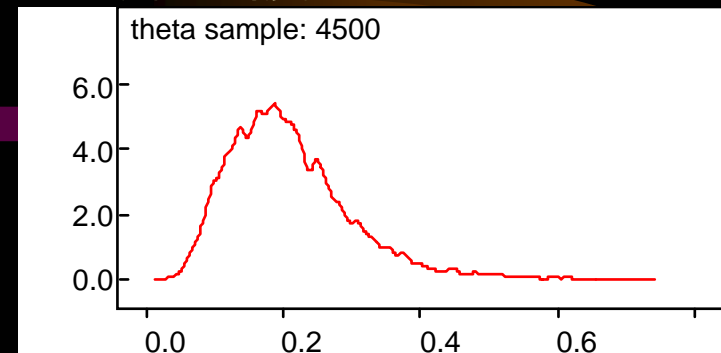
収束後のサンプリング回数: 4500回

事後分布(尤度関数に比例)のシミュレーション

μ の周辺密度

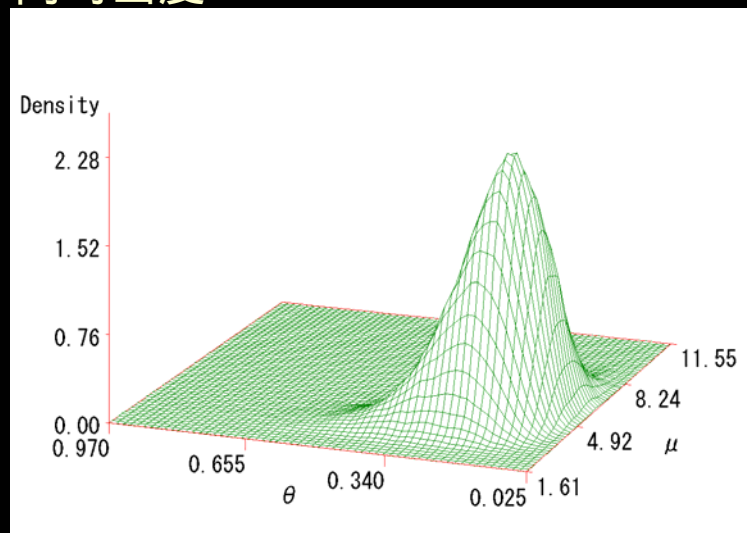


θ の周辺密度



密度の最も高いところが、近似的な最尤推定値に相当する。
下側2.5%点 ~ 97.5%点 が95% credible intervalである。これは最尤法による95%信頼区間に相当する。

同時密度



ご紹介内容

1. ベイズ統計のMCMCとの出会い
2. ベイズ統計で最尤推定
3. **ベイズ統計と混合効果モデルの関係**
4. ベイズ論文の読みかた(私案)

ベイズ統計と混合効果モデルの関係

Demidenko, E. (2004)の考え

混合効果モデル論 =

頻度論

+

ベイズ論

Demidenko, E. (2004). A compromise between Bayesian and frequentist approaches, *Mixed Models Theory and Applications*, Section 1.4, pp.8-11. John Wiley & Sons.

私たちの考え

頻度論

w

混合効果モデル論

$1-w$

ベイズ論

Demidenko, E. (2004)の考え

頻度論

$$y | \theta \sim L(y | \theta)$$

階層モデル

ベイズ論

$$y | \theta \sim L(y | \theta)$$

$$\theta \sim G(\theta)$$

混合効果モデル論

$$y | \theta \sim L(y | \theta, \tau)$$

$$\theta \sim G(\theta, \tau)$$

θ : 個体効果

τ : 集団効果

私たちの考え

利用できるデータが少ない経時データに対して個体ごとに切片と傾きが異なる直線をあてはめる場合で、モデルの推定可能性を検討した。

i は個体 ($i = 1, 2, 3$) , j は時点を示す。

頻度論

$$Y_{ij} = \alpha_i + \beta_i \cdot x_j + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

混合効果モデル論

$$Y_{ij} = \alpha_i + \beta_i \cdot x_j + \varepsilon_{ij} \quad \alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2) \quad \beta_i \sim N(\mu_\beta, \sigma_\beta^2) \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

ベイズ論

$$Y_{ij} = \alpha_i + \beta_i(x_j - \bar{x}) + \varepsilon_{ij} \quad \alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2) \quad \beta_i \sim N(\mu_\beta, \sigma_\beta^2) \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

事前分布

$$\mu_\alpha \sim N(0, 100^2) \quad \mu_\beta \sim N(0, 100^2) \quad \tau_\alpha = 1/\sigma_\alpha^2 \sim Ga(0.001, 0.001)$$

$$\tau_\beta = 1/\sigma_\beta^2 \sim Ga(0.001, 0.001) \quad \tau_\varepsilon = 1/\sigma_\varepsilon^2 \sim Ga(0.001, 0.001)$$

n_i は個体 i で得られた観測値数とする。

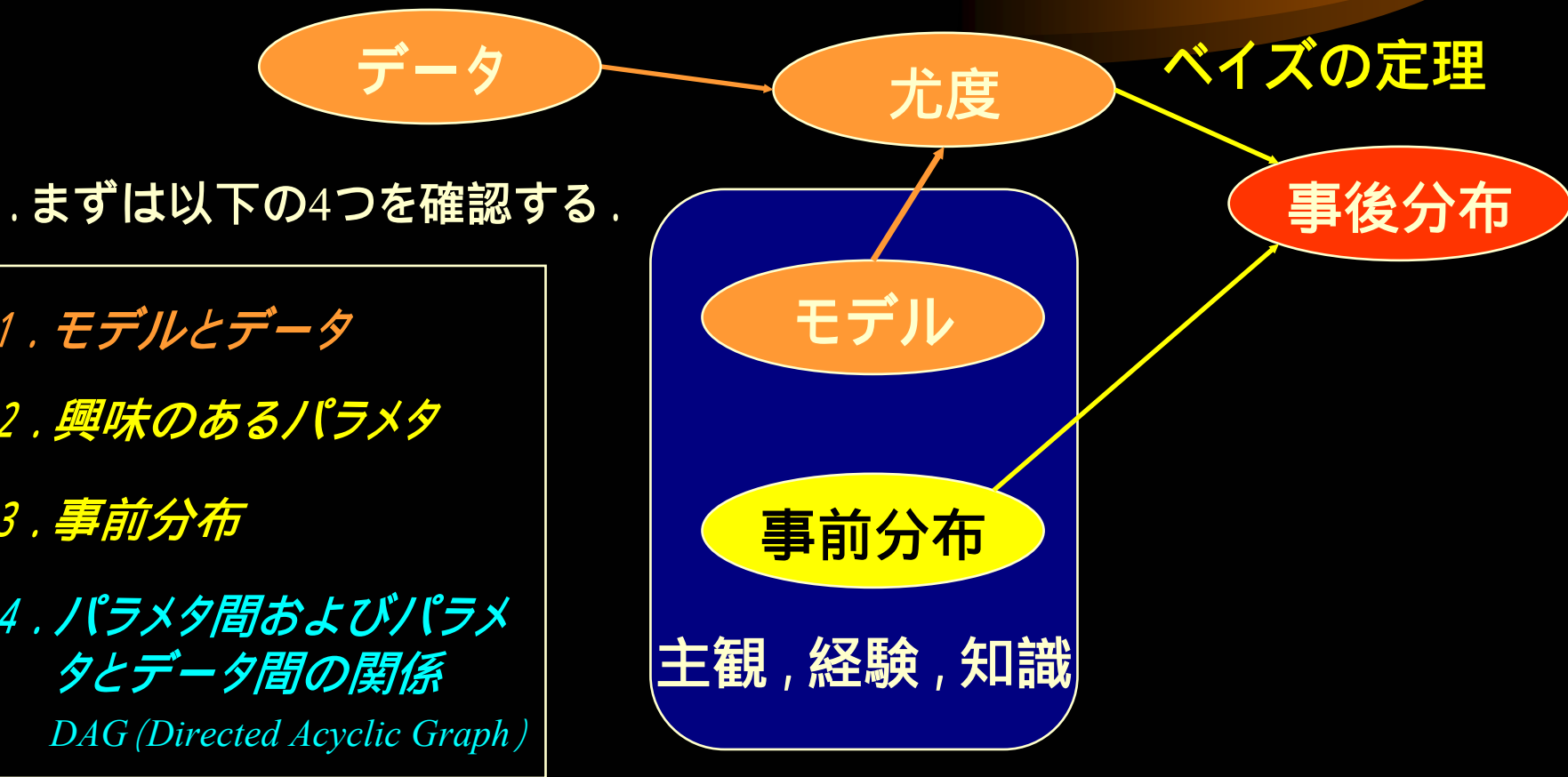
利用可能な観測値数の組み合わせごとのモデル式の推定可能性

	(n_1, n_2, n_3)							
	(1,1,1)	(2,1,1)	(2,2,1)	(3,2,1)	(2,2,2)	(3,2,2)	(3,3,2)	(3,3,3)
頻度論	×	×	×	×	×			
混合効果モデル	×	×						
ベイズ統計								

ご紹介内容

1. ベイズ統計のMCMCとの出会い
2. ベイズ統計で最尤推定
3. ベイズ統計と混合効果モデルの関係
4. ベイズ論文の読みかた(私案)

ベイズ統計が登場する文献を読むためのコツ(1/2)





パラメタ間およびパラメタとデータ間の関係

- DAG (Directed Acyclic Graph) に描くと頭の整理になる .
- BUGSではDAGを描くと , 計算プロセスが自動的にコーディングされる .

BUGS: ベイズ統計の考えに基づく計算を行う
フリーのソフトウェア

事例 . ロジスティック回帰分析

データ (毒性試験)

Log ₁₀ (用量)	標本サイズ	死亡数
1.691	59	4
1.724	60	10
1.755	62	19
1.784	56	31
1.811	63	52
1.837	59	53
1.861	62	60
1.844	60	60

第 k 用量の常用対数変換値... d_k

d_k を標準化した値... x_k

第 k 用量における動物数... n_k

第 k 用量における死亡数... m_k

第 k 用量における死亡率... θ_k

モデル

第 k 用量における死亡数 m_k に二項分布 $Binomial(\theta_k, n_k)$ を仮定し,

ロジスティック回帰モデル $\log \frac{\theta_k}{1-\theta_k} = \alpha + \beta x_k$ をあてはめる.

事前分布

α に正規分布 $N(0, 1^2)$ を設定する.

β に正規分布 $N(0, 3^2)$ を設定する.

事例 . ロジスティック回帰分析

用量の常用対数変換値 d_k を基準化した値を x_k とする .

DAG

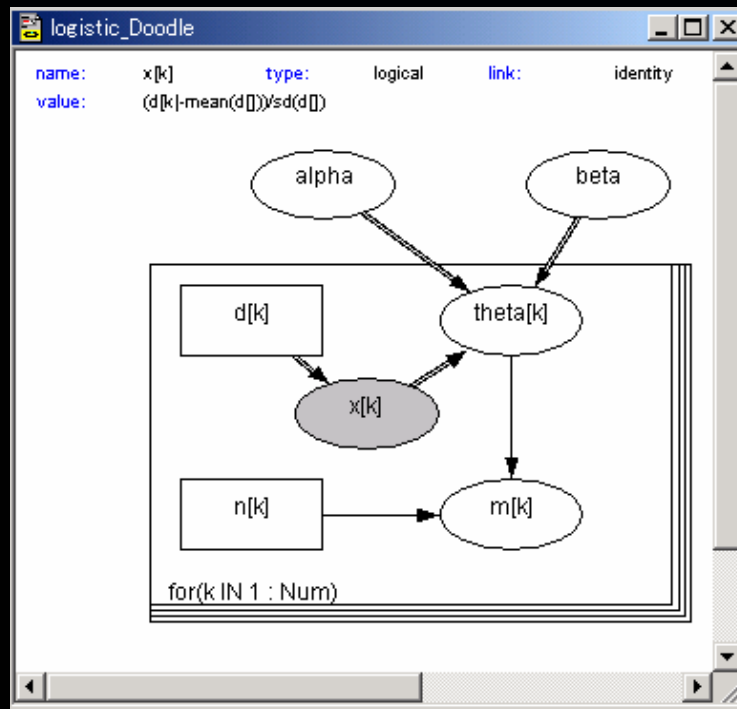
$$\left\{ \begin{array}{l} m_k \sim \text{Binomial}(\theta_k, n_k) \\ \log \frac{\theta_k}{1-\theta_k} = \alpha + \beta x_k \end{array} \right.$$

α の事前分布

$$\alpha \sim N(0, 1^2)$$

β の事前分布

$$\beta \sim N(0, 3^2)$$



- 固定定数
- 観測される変数 (データ)
- 観測されない変数
- 確率的な依存関係
- 論理的な関数

ベイズ統計が登場する文献を読むためのコツ(2/2)

次に以下を確認する。

MCMCを利用

NO

確認内容

共役事前分布
の利用

NO

YES

BUGSを利用

NO

確認内容

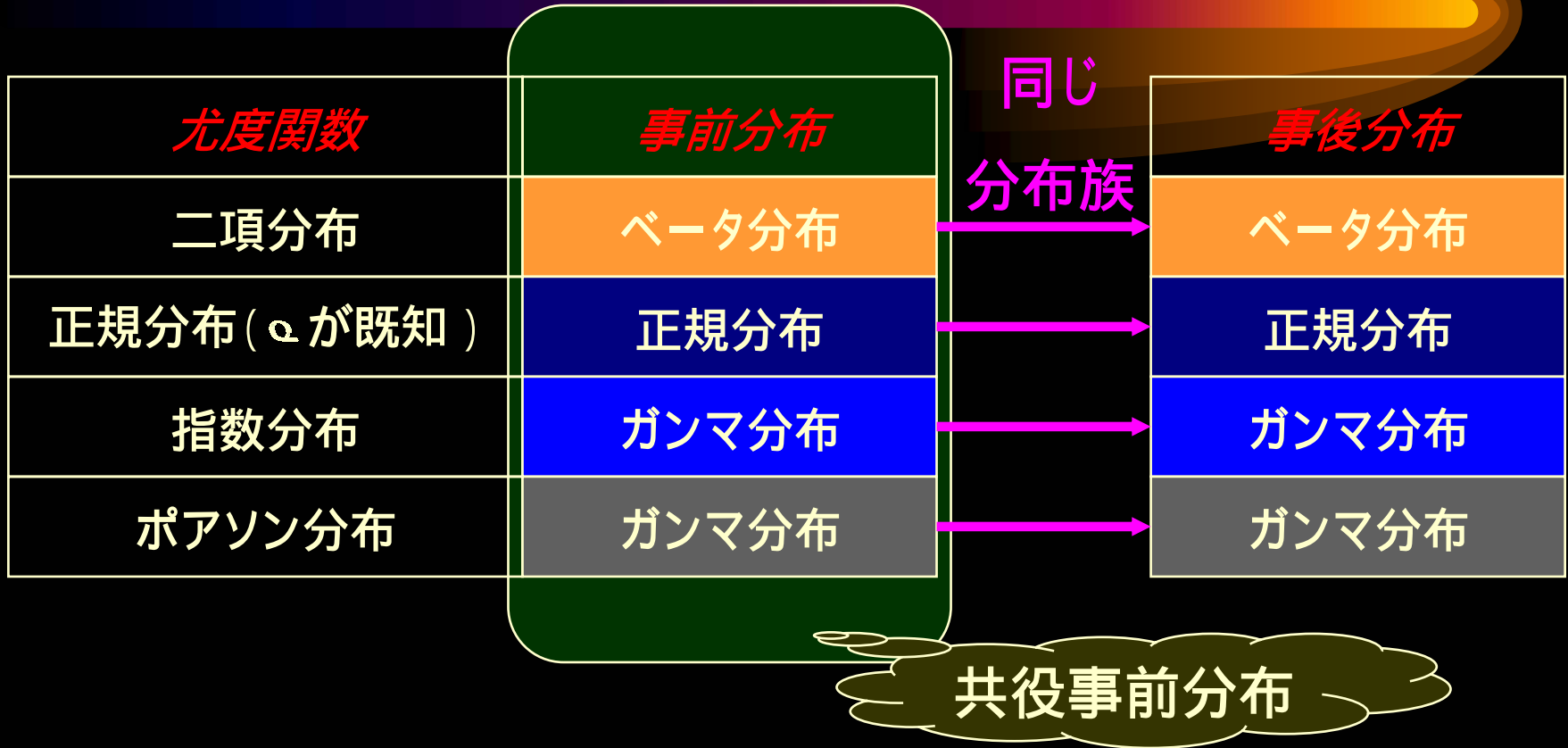
1. MCMCアルゴリズム(詳細確認)
2. フルの条件付き事後分布(パラメタが複数個ある場合)
3. ソフトウェア
4. Burn-in回数と収束後のサンプリング回数

YES

確認内容

1. Gibbsサンプリング(自明)
2. フルの条件付き事後分布
3. BUGS(自明)
4. Burn-in回数と収束後のサンプリング回数

共役事前分布



代表的なMCMCアルゴリズム：確認すべき内容

BUGS以外でMCMCを利用している場合
利用アルゴリズムと以下の内容を確認にする。

MCMCアルゴリズム	確認内容
M-Hアルゴリズム	候補発生分布
Single Component M-Hアルゴリズム	候補発生分布 フルの条件付き分布
Gibbsサンプリング	フルの条件付き分布 フルの条件付き分布からの1変量乱数発生方法

ご紹介内容

1. ベイズ統計のMCMCとの出会い
2. ベイズ統計で最尤推定
3. ベイズ統計と混合効果モデルの関係
4. ベイズ論文の読みかた(私案)

詳細釣り合い, 階層モデルについても言及したかったのですが, 今回は時間の都合もあり割愛させていただきました.

Metropolis-Hastingsアルゴリズムを私たちなりに解剖したスライドをご参考までに添付いたします.

Metropolis-Hastings アルゴリズム

密度関数 $\pi(\cdot)$ をもつ分布に従う乱数を生成する。

- ・初期値を $x^{(0)}$ とする。
- ・ $j = 1, 2, \dots, N$ について以下を繰り返す。

1. $q(x^{(j-1)}, \cdot)$ から y , $U(0,1)$ から u を生成する。

2. $u \leq \alpha(x^{(j-1)}, y)$ であれば, $x^{(j)} = y$ とおく。

ここに, $\alpha(x^{(j-1)}, y) = \min\left\{1, \frac{\pi(y)q(y, x^{(j-1)})}{\pi(x^{(j-1)})q(x^{(j-1)}, y)}\right\}$ である (数学的に厳密な表現はあえて避けた)。

3. そうでなければ $x^{(j)} = x^{(j-1)}$ とおく。

・結果 $\{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ を得る。

このアルゴリズムは候補発生密度 $q(x^{(j-1)}, \cdot)$ により特定される。

Metropolis-Hastings アルゴリズムの骨子

密度関数 $\pi(\cdot)$ をもつ分布に従う乱数を生成する。

1. x_{before} に依存する密度 $q(x_{before}, \cdot)$ をもつ分布から y を発生させ、確率 $\alpha(x_{before}, y)$ で $x_{after} = y$ とする。
2. そうでなければ、 $x_{after} = x_{before}$ とする。

ここに、 $\alpha(x_{before}, y) = \min \left\{ 1, \frac{\pi(y)q(y, x_{before})}{\pi(x_{before})q(x_{before}, y)} \right\}$ である (数学的に厳密な表現はあえて避けた)。

「 x_{before} から y への推移」の密度: $q(x_{before}, y)\alpha(x_{before}, y)$

詳細釣り合い (detailed balance)

次の詳細釣り合いの等式を満たすとき、密度 $\pi(\cdot)$ をもつ分布は、
「 x_{before} から y への推移」の密度 $R(x_{before}, y)$
で決まるマルコフ連鎖の定常分布である。

$$\pi(x_{before})R(x_{before}, y) = \pi(y)R(y, x_{before})$$

x_{before} の密度 「 x_{before} から y へ
の推移」の密度

Metropolis-Hastingsアルゴリズムは詳細釣り合いを満たす。

Metropolis-Hastingsアルゴリズムでは、

「 x_{before} から y への推移」の密度 $R(x_{before}, y)$ は $q(x_{before}, y)\alpha(x_{before}, y)$ である。

$$\begin{aligned}\pi(x_{before})R(x_{before}, y) &= \pi(x_{before})q(x_{before}, y)\alpha(x_{before}, y) \\ &= \pi(x_{before})q(x_{before}, y) \min \left\{ 1, \frac{\pi(y)q(y, x_{before})}{\pi(x_{before})q(x_{before}, y)} \right\} \\ &= \min \left\{ \pi(x_{before})q(x_{before}, y), \pi(y)q(y, x_{before}) \right\} \\ &= \pi(y)q(y, x_{before}) \min \left\{ \frac{\pi(x_{before})q(x_{before}, y)}{\pi(y)q(y, x_{before})}, 1 \right\} \\ &= \pi(y)R(y, x_{before})\end{aligned}$$

すなわち、密度 $\pi(\cdot)$ をもつ分布は、「 x_{before} から y への推移」の密度が $q(x_{before}, y)\alpha(x_{before}, y)$ であるマルコフ連鎖の定常分布である。

詳細釣り合いが満足されていれば、
あるサンプルが定常分布からいったんサンプリングされると、
その後のサンプルはすべてその定常分布からのものとなる。

詳細釣り合い

定常分布が存在すればそれが $\pi(\cdot)$ である