

ギブスサンプリング

樋口さぶろお <https://hig3.net>

龍谷大学大学院 先端理工学部研究科 数理・情報科学コース

計算科学特論 L16(2025-01-22 Wed)

最終更新: Time-stamp: "2025-01-22 Wed 10:40 JST hig"

今日の目標

- ギブスサンプリングを説明できる



ここまで来たよ

15 MCMC によるベイズ推定

16 ギブスサンプリング

- 混合ガウス分布の事後分布からのギブスサンプリング
- ギブスサンプリングの一般論

ギブスサンプリング

Hoff ベイズ §6

鎌谷 MCS5

ギブスサンプリング (Gibbs Sampling) は、ほしい分布（ベイズ推定ならパラメタの事後分布）が定常分布になるような、うまいマルコフ連鎖を (Metropolis-Hastings とは一見別の方法で) つくる MCMC.

定義 (ギブスサンプリング)

モデルが θ, X の同時分布を定めていて、同時分布からサンプルするのは難しいが、条件付き分布 $\theta|X, X|\theta$ からサンプルすることはできるとき、これを交互に繰り返すマルコフ連鎖を考え、極限分布を求めるアルゴリズム.

混合ガウス分布 GMM=Gaussian mixture model

計算科学 I(2024)L02

正規分布 normal distribution = ガウス分布 Gaussian distribution

岩薩林 確率・統計 §4.5

混合する = mix

混合ガウス分布を定義する準備として、次の離散-連続混合型分布 $f(x, y)$ を考える。

$\pi_0 + \pi_1 = 1, \pi_x \geq 0, \pi_{\text{他}} = 0$ とする。

$$f(x, y) = \pi_x \frac{1}{(2\pi\sigma_x^2)^{1/2}} e^{-\frac{(y-\mu_x)^2}{2\sigma_x^2}} = \begin{cases} \pi_0 \cdot \frac{1}{(2\pi\sigma_0^2)^{1/2}} e^{-\frac{(y-\mu_0)^2}{2\sigma_0^2}} & (x = 0) \\ \pi_1 \cdot \frac{1}{(2\pi\sigma_1^2)^{1/2}} e^{-\frac{(y-\mu_1)^2}{2\sigma_1^2}} & (x = 1) \\ 0 & (x \neq 0, 1) \end{cases}$$
$$= \delta_{x0} \pi_0 \frac{1}{(2\pi\sigma_0^2)^{1/2}} e^{-\frac{(y-\mu_0)^2}{2\sigma_0^2}} + \delta_{x1} \pi_1 \frac{1}{(2\pi\sigma_1^2)^{1/2}} e^{-\frac{(y-\mu_1)^2}{2\sigma_1^2}}.$$

X の周辺分布は離散型でベルヌイ分布 $B(1, \pi_1)$

$$f_X(x) = \int f(x, y) dy = \int_{-\infty}^{+\infty} \pi_x \cdot \frac{1}{(2\pi\sigma_x^2)^{1/2}} e^{-\frac{(y-\mu_x)^2}{2\sigma_x^2}} dy = \pi_x.$$

Y の周辺分布は連続型で GMM

定義 (GMM=混合ガウス分布)

連続型確率変数 Y がパラメタ $(\pi_0, \pi_1, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$ の GMM に従うとは、次の確率密度関数にしたがうことをいう。

$$f_Y(y) = \sum_x f(x, y) = \pi_0 \cdot \frac{1}{(2\pi\sigma_0^2)^{1/2}} e^{-\frac{(y-\mu_0)^2}{2\sigma_0^2}} + \pi_1 \cdot \frac{1}{(2\pi\sigma_1^2)^{1/2}} e^{-\frac{(y-\mu_1)^2}{2\sigma_1^2}}.$$

まず X, Y の同時分布を考えた後, その Y の周辺分布を考え, $X = 0, 1$ (コインの裏表) は和を取って消して混合ガウス分布を得た.

観測されたデータ $Y_1, \dots, Y_N$.

ベイズ推定したいパラメタ $\pi_0 = \theta, \pi_1 = 1 - \theta$.

計算の便利のため, X を復活させ, X, θ の同時分布を考えて θ をギブスサンプリングする.

ギブスサンプリング (θ のマルコフ連鎖) でサンプルを出力

Listing 1: Gibbs Sampling

```
1 double theta=thetainitial; // これをサンプルする
2 int n=サンプルサイズ;
3 int x[n]; // 0 or 1, コインの裏表を勝手に想定
4 double y[]={,,}; // n個の測定値
5
6 while(true){
7     // X
8     for(int i=0; x<n; i++){
9         x[i]=getx(theta); //  $\backslash(theta\backslash)$  の条件のもとでの  $x[i]$  の分布
10    }
11
12    // theta
13    theta=gettheta(x); //  $X[]$  の条件のもとでの  $theta$  の分布はベータ
14
15    print("%f\n",theta);
16 }
```

次は、どちらも（モンテカルロ不要で）実行できる

`int gety(double theta)` 次の条件付き確率で 0 または 1 をランダムに返す.

$$P(X = x_i | \theta) = \begin{cases} \frac{\theta f(y_i; \mu_1, \sigma_1)}{\theta f(y_i; \mu_1, \sigma_1) + (1-\theta) f(y_i; \mu_0, \sigma_0)} & (x_i = 1) \\ \frac{(1-\theta) f(y_i; \mu_0, \sigma_0)}{\theta f(y_i; \mu_1, \sigma_1) + (1-\theta) f(y_i; \mu_0, \sigma_0)} & (x_i = 0) \end{cases}$$

`double gettheta(int x[])` 次のベータ分布の確率密度関数にしたがって $\theta \in [0, 1]$ をランダムに返す. $k = \sum_i x_i$.

$$f(\theta) = \frac{1}{Z} \theta^k (1 - \theta)^{n-k}$$

(ベータ分布にしたがう乱数は $U(0, 1)$ 擬似乱数だけから作れる)

L16-Q1

Quiz(ギブスサンプリング)

乱数やサイコロは使えず, コイン (ただし, 表の確率 p は自由に設定できる), すなわち $B(1, p)$ の乱数だけが使える状況を考えよう.

離散型確率変数の同時分布 $f(\theta, x) =$

$x \backslash \theta$	1	2
10	1/10	2/10
20	3/10	4/10

を考える.

- ① 初期値が $\theta = 1$ だったとき, ギブスサンプリングのループの 1 回目が終わった時点で得られる標本の分布を求めよう.
- ② θ の分布 $\left(\frac{4}{6/10}\right)$ が, ギブスサンプリングの定めるマルコフ連鎖の定常分布になっていることを確かめよう.

L16-Q2

Quiz(ギブスサンプリング)

乱数やサイコロは使えず, コイン (ただし, 表の確率 p は自由に設定できる), すなわち $B(1, p)$ の乱数だけが使える状況を考えよう.

離散型確率変数の同時分布 $f(\theta, x) =$

$x \backslash \theta$	1	2
10	1/15	4/15
20	2/15	8/15

を考える.

- ① 初期値が $\theta = 1$ だったとき, ギブスサンプリングのループの 1 回目が終わった時点で得られる標本の分布を求めよう.
- ② θ の分布 $\left(\begin{smallmatrix} 3/15 \\ 12/15 \end{smallmatrix} \right)$ が, ギブスサンプリングの定めるマルコフ連鎖の定常分布になっていることを確かめよう.

ここまで来たよ

15 MCMC によるベイズ推定

16 ギブスサンプリング

- 混合ガウス分布の事後分布からのギブスサンプリング
- ギブスサンプリングの一般論

ギブスサンプリングの一般論

2つの確率変数 X, Θ

- θ に初期値を代入
- while(1)
 - ▶ $X \sim X|\theta, \{y_i\}$ からサンプルして x を更新
 - ▶ $\Theta \sim \Theta|x$ からサンプルして θ を更新

ギブスサンプリングのもっと一般論

k 個の確率変数 X_i . (さっきの例は $k = 2, X_1 = X, X_2 = \theta$)

- $x_1, \dots, x_k =$ 初期値を代入
- while(true)
 - ▶ $X_1 \sim X_1 | x_2, \dots, x_k$ からサンプルして x_1 を更新
 - ▶ $X_2 \sim X_2 | x_1, x_3, \dots, x_k$ からサンプルして x_2 を更新
 - ▶ $\vdots$
 - ▶ $X_k \sim X_k | x_1, x_2, \dots, x_{k-1}$ からサンプルして x_k を更新