

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 先端理工学研究科 数理・情報科学コース

計算科学特論 L16(2026-01-21 Wed)

最終更新: Time-stamp: "2026-01-21 Wed 17:38 JST hig"

今日の目標

- ギブスサンプリングを説明できる



L15-Q1

TA Prob and Sol: ベルヌイ分布のベイズ推定: 非一様事前分布

$X \sim B(1, y)$ の母数 y の事前分布を $f_Y(y) = 6y(1-y)u(y; 0, 1)$ とする.
 サイズ N の標本として, $x_1, \dots, x_N$ が得られた. ただし,
 $\sum_{i=1}^n x_i = k \leq N$ だった.
 母数 y の事後分布を求めよう.
 事後分布の確率密度関数が最大となる y を求めよう.

略解

$$\begin{aligned}
 f_{Y|X}(y|x_i) &= \frac{f_{X|Y}(x_1, \dots, x_N|y) f_Y(y)}{\int f_{X|Y}(x|y) f_Y(y) dy} \\
 &\propto \prod_i y^{x_i} (1-y)^{1-x_i} 6y(1-y)u(y; 0, 1) \\
 &\propto y^{k+1} (1-y)^{N-k+1} u(y; 0, 1)
 \end{aligned}$$

事後分布の確率密度関数が最大となるのは, $y = \frac{k+1}{N+2}$.

ここまで来たよ

15 MCMC・詳細つりあいの条件・メトロポリス・ヘイスティングズ法

16 ギブスサンプリング

- ギブスサンプリング
- ベイズ推定におけるギブスサンプリングの例
- ギブスサンプリングの一般論

ギブスサンプリング Replace 鎌谷 MCS5

ギブスサンプリング (Gibbs Sampling) は、ほしい多変数同時分布から乱数を生成するのが難しいが、条件付きの 1 変数分布からの乱数生成が容易なときに、うまいマルコフ連鎖を (Metropolis-Hastings とは一見別の方法で) で乱数生成する MCMC.

Gibbs は統計力学の重要人物. ギブズに近い発音だが、日本語では「ギブス」で定着している.

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.

まず，観測値 $x_1, \dots, x_N$ がない，ただのギブスサンプリングで．上の問でいう推定したい母数 $y = (x, \theta)$.

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 $\left(\frac{4}{10}, \frac{6}{10}\right)$ 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．
- ③ このとき， x, θ が同時分布にしたがうことを納得しよう．

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てよう.

離散型確率変数の同時分布 $f(\theta, x) =$

$x \backslash \theta$	1	2
10	1/15	7/15
20	4/15	3/15

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 $\left(\frac{1}{3}, \frac{2}{3}\right)$ 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・詳細つりあいの条件・メトロポリス・ヘイスティングズ法

16 ギブスサンプリング

- ギブスサンプリング
- ベイズ推定におけるギブスサンプリングの例
- ギブスサンプリングの一般論

ベイズ推定におけるギブスサンプリングの例 I

- ① 観測値 $x_1, \dots, x_N$ が正規分布 $N(\mu, 1/v)$ に従うとする．精度 $v = 1/\sigma^2$ ．

$$p(x|\mu, v) = \frac{\sqrt{v}}{\sqrt{2\pi}} e^{-\frac{v}{2}(x-\mu)^2}.$$

- ② 事後分布を求めたいパラメタは， $y = (\mu, v)$ ．
 ③ 事前分布は， $\mu \sim N(\mu_0, 1/v_0)$ ， v はパラメタ $(\alpha, \beta) = (\frac{1}{2}, \beta_0)$ のガンマ分布．

$$p(\mu) \propto e^{-\frac{v_0}{2}(\mu-\mu_0)^2}, \quad p(v) \propto \sqrt{v} e^{-\beta_0 v}.$$

- ④ 尤度関数

$$p(x_1, \dots, x_N|\mu, v) = \prod_{i=1}^N p(x_i|\mu, v) = \left(\frac{\sqrt{v}}{\sqrt{2\pi}} \right)^N e^{-\frac{v}{2} \sum_{i=1}^N (x_i - \mu)^2}.$$

ベイズ推定におけるギブスサンプリングの例 II

5 事後分布

$$\begin{aligned} p(\mu, v | x_1, \dots, x_N) \\ &\propto p(x_1, \dots, x_N | \mu, v) p(\mu) p(v) \\ &\propto (\sqrt{v})^N e^{-\frac{v}{2} \sum_{i=1}^N (x_i - \mu)^2} e^{-\frac{v_0}{2} (\mu - \mu_0)^2} \sqrt{v} e^{-\beta_0 v}. \end{aligned}$$

この2変数の事後分布の同時分布からサンプルを得るのは容易ではない (メトロポリス・ヘイスティングズ法を使うことになる)。

条件付き分布の具体的な形

- μ の条件付き分布 $p(\mu|v, x_1, \dots, x_N)$ は正規分布

$$N\left(\frac{v_0\mu_0 + v\sum_i x_i}{v_0 + Nv}, \frac{1}{v_0 + Nv}\right).$$

- v の条件付き分布 $p(v|\mu, x_1, \dots, x_N)$ はガンマ分布

$$\text{Gamma}\left(\alpha = \frac{N}{2} + \frac{1}{2}, \beta = \beta_0 + \frac{1}{2} \sum_i (x_i - \mu)^2\right).$$

ここで、真実を告げると、1変数の正規分布やガンマ分布は、容易に（メトロポリス・ヘイスティングズ法を使わなくても）乱数生成することができる。

このことを利用して多変数分布からサンプルするのがギブスサンプリング。

ベイズ推定におけるギブスサンプリングのアルゴリズム

観測値 $\{x_i\}$.

事後分布からのサンプルを得たい2つの母数 μ, v

i : MCMC の '時間' t

- $v^{(0)}$ に適当な初期値を代入
- while(true)
 - ▶ $\mu^{(i+1)} \sim p(\mu|v^{(i)}, x_1, \dots, x_N)$ から $\mu^{(i+1)}$ を生成. 正規分布にしたがう乱数生成.
 - ▶ $v \sim p(v^{(i+1)}|\mu^{(i+1)}, x_1, \dots, x_N)$ から $v^{(i+1)}$ を生成. ガンマ分布にしたがう乱数生成.
 - ▶ $(\mu^{(i+1)}, v^{(i+1)})$ をサンプルに加える
 - ▶ $i++$

ここまで来たよ

15 MCMC・詳細つりあいの条件・メトロポリス・ヘイスティングズ法

16 ギブスサンプリング

- ギブスサンプリング
- ベイズ推定におけるギブスサンプリングの例
- ギブスサンプリングの一般論

ベイズ推定におけるギブスサンプリングの一般論

観測値 $\{x_i\}$.

事後分布からのサンプルを得たい2つの母数 Y_1, Y_2

- Y_2 に初期値を代入
- while(true)
 - ▶ $Y_1 \sim p(y_1|y_2, \{x_i\})$ から y_1 を更新
 - ▶ $Y_2 \sim p(y_2|y_1, \{x_i\})$ から y_2 を更新
 - ▶ 直近の (y_1, y_2) をサンプルに加える

(ベイズ推定に限定しない) ギブスサンプリング

観測値が1個もない場合 $\{x_i\} = \emptyset$ も, 同時分布 $p(y_1, y_2)$ から得られる条件つき分布 $p(y_1|y_2), p(y_2|y_1)$ からの生成さえできれば, 同時分布からの生成アルゴリズムとして機能する.

例: 多次元正規分布からのランダムベクトル生成

ギブスサンプリングのもっと一般論

観測値 $\{x_i\}$ (以下の疑似コードでは表示省略)
事後分布からのサンプルを得たい k 個の母数 Y_i .

- $y_2, \dots, y_k$ に適当な初期値を代入
- while(true)
 - ▶ $Y_1 \sim Y_1 | y_2, \dots, y_k$ からサンプルして y_1 を更新
 - ▶ $Y_2 \sim Y_2 | y_1, y_3, \dots, y_k$ からサンプルして y_2 を更新
 - ▶ $\vdots$
 - ▶ $Y_k \sim Y_k | y_1, y_2, \dots, y_{k-1}$ からサンプルして y_k を更新
 - ▶ 直近の $(y_1, \dots, y_k)$ をサンプルに加える

わざわざ人工的な変数を追加して、乱数生成を複数段階に分けて、ギブスサンプリングの形にすることがある → 潜在変数.