

マルコフ連鎖モンテカルロ法と詳細つりあいの条件

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 L14(2025-01-08 Wed)

最終更新: Time-stamp: "2025-01-08 Wed 12:53 JST hig"

今日の目標

- マルコフ連鎖モンテカルロ法を説明できる
- マルコフ連鎖の対称分布, 分布に関して詳細つりあいを満たすマルコフ連鎖を説明できる



L13-Q1

Quiz 解答: 逆関数法による擬似乱数生成

$$\begin{aligned}
 f_R(r)dr &= f_Y(y)dy \\
 -\frac{200}{21} \frac{1}{r^3} dr &= dy \\
 \frac{200}{21} \frac{1}{2} r^{-2} &= y + C
 \end{aligned}$$

$y = 0$ のとき $r = -2$ より, $C = \frac{4}{21}$. $r^2 = 100(4 + 21y)^{-1}$.

$r = \pm[100(4 + 21y)^{-1}]^{1/2}$.

$r(0) = -5, r(1) = -2$ に注意すると, $r = g(y) = -10(4 + 21y)^{-1/2}$

L13-Q2

Quiz 解答: 棄却法による乱数生成

Listing 1: 棄却法

```
1  extern double getuniform();
2  double f(double x);
3
4  double getRandom(){
5      double x,y;
6      double a=1.0, b=4.0;
7      double M=16/21.0; /* = max f = f(4) */
8
9      while(1){
10         x=a+(b-a)*getuniform(); /* U(a,b) */
11         y=M*getuniform(); /* U(0,M) */
12         if(y<f(x)){
13             break;
14         }
15     }
16     return x;
17 }
18
19 double f(double x){
```

```
20 |   return 1/21.0*x*x; /* assume a<=x<b*/  
21 | }
```

(x, y) は $(1, 4) \times (0, \frac{16}{21})$ の面積 $\frac{16}{7}$ の長方形に一様に分布する. このうち, $f(x) > y$ の面積 1 の領域の (x, y) に対しては x が返される. よって, 棄却は平均して, $0 \cdot \frac{7}{16} + 1 \cdot \frac{9}{16} \cdot \frac{7}{16} + 2 \cdot (\frac{9}{16})^2 \cdot \frac{7}{16} + \dots = \frac{9}{7}$ 回起きる. よって乱数 1 個の生成に `double getuniform()` の平均 $(1 + \frac{9}{7}) \times 2$ 回の呼び出しを要する.

ここまで来たよ

13 逆関数法・棄却法による乱数の生成

14 マルコフ連鎖モンテカルロ法と詳細つりあいの条件

- マルコフ連鎖モンテカルロ法
- メトロポリス・ヘイスティングス法
- 二項モデルのベイズ推定への適用例

マルコフ連鎖モンテカルロ法が求められる状況

目的 やりにくい確率 (密度) 関数 $f(x)$ にしたがう X の標本 (=乱数) を実際に抽出する. 標本期待値を求める. ベイズ推定する. 計算科学 I(2024)L03

確率の問題だけではない. 例えば, 図形の面積は何かの確率変数の母期待値として書けることがある

前提

- ① 計算機で, 一様分布 $U(0, 1)$ の独立同分布に「ほぼ」したがうような **擬似乱数列** (pseudo-random number) は比較的容易に生成できる.
 - ▶ `double getuniform()` 計算科学 II(2024)L12
- ② 任意の分布にしたがう確率分布にしたがう擬似乱数列を容易に作れるわけではない.
 - ▶ 作る工夫: 逆関数法, 棄却法, 個別の超絶技巧, ... 計算科学 II(2024)L13
- ③ $f(x)$ が定数倍だけ不明なことがある.
 - ▶ $f(x) = \frac{1}{Z}g(x)$ で, 定数 $Z = \int_{-\infty}^{+\infty} g(x)dx$ が計算できないことがある.
 - ▶ 同じことの別の言い方: 比 $f(x_1)/f(x_2)$ は計算できるけど, $f(x)$ そのものは計算できないことがある.

マルコフ連鎖モンテカルロ法 (MCMC=Markov Chain Monte Carlo)

設計問題: $g(x) = Zf(x) \geq 0$ が与えられたとき, $f(x) = \frac{1}{Z}g(x)$ がその極限分布であるようなマルコフ連鎖 M (で, かつ「いいやつ」) を作れ.

実行: このマルコフ連鎖をシミュレートして, 状態列 $X(t)$ として擬似乱数列を得る

Listing 2: MCMC

```
1 double x=xinitial; // 初期状態
2 for (n=0;n<サンプルサイズ;n++){
3     for (t=0;t<待ち時間;t++){
4         x=getnext(x); //  $U(0,1)$  一様乱数を使い,  $M$ にしたがう
5     }
6     // 定常分布に近づいただろう
7     printf( "%f\n", x );
8 }
```

- 「いいやつ」: $X(t)$ から $X(t+1)$ を得るのは, $U(0,1)$ 擬似乱数だけでできる
- 「いいやつ」: 極限分布が一意でない, 存在しない, 可約でない, 周期的でない
- 最初のそれなりの個数の出力は, 初期状態の影響を受けているかもしれないので=極限分布に十分に近づいていないかもしれないので捨てる (**burn-in**).
- 出力が定常分布にしたがっていても, 互いに独立になっていないかもしれないので, 待ち時間は十分大きくとる. 自己相関係数で判定できる. 計算科学 II(2024)L12
- 「いいやつ」: 自己相関係数の 0 への収束が速い → 待ち時間を大きくしなくても済む → 計算時間が短くて済む

極限分布が与えられたときのマルコフ連鎖設計作戦

定義 (対称分布)

マルコフ連鎖 (の転置推移確率行列) M_{xy} に対して, 次の条件を満たす分布 p_x を, M の対称分布 (symmetric distribution) という.

$$\forall x, y \in S \quad M_{xy}p_y = M_{yx}p_x$$

(上の式では, もちろん和はとってない)

- $P(X(0) = y \text{ かつ } X(1) = x) = P(X(0) = x \text{ かつ } X(1) = y)$ なので, 同時分布が $f_{X(0)X(1)}(x, y) = f_{X(0)X(1)}(y, x)$ だというのが '対称' の語源.
- 任意の 2 つの状態 x, y を結ぶ矢印は, 向きが逆のもの 2 本あるが, その上を通過する確率が同じ, という条件.

$\vec{p}$ が M の対称分布であることを, M は分布 p に関する詳細つりあいの条件 (detailed balance condition) を満たすともいう.

命題

対称分布は定常分布である.

p の次の時刻の分布 q を求める.

$$q_x = \sum_y M_{xy} p_y = \sum_y M_{yx} p_x = 1 \cdot p_x = p_x.$$

よって p_x は定常分布.

逆は真でない. (既約で非周期的なマルコフ連鎖に限ったとしても) 対称分布は定常分布の中の特別なクラス.

極限分布は定常分布に限られるのだった. 定常分布なら極限分布かも?

設計の方針

$\bar{p}$ を定常分布として持つ M を作る. さらに限定して, 対称分布として持つような特殊な M をわざわざ作る. というのも, 効率はともかく, ある程度機械的に作れるから (メトロポリス-ヘイスティングス法).

この方針にしたがわず, 対称でなく (irreversible) 「性能のよい」 M を得る方法も近年盛んに研究されている.

ここまで来たよ

13 逆関数法・棄却法による乱数の生成

14 マルコフ連鎖モンテカルロ法と詳細つりあいの条件

- マルコフ連鎖モンテカルロ法
- メトロポリス・ヘイスティングス法
- 二項モデルのベイズ推定への適用例

メトロポリス・ヘイスティングス法

与えられた $\vec{p}$ を対称分布として持つような M を作るうまい方法のひとつ.

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

$$M_{xy} = q_{xy} \min\left\{1, \frac{p_x q_{yx}}{p_y q_{xy}}\right\} + k_y \times \delta_{xy}$$

定数 $k_y \geq 0$ は, M_{xy} が転置確率行列となるように調整できる.

$\vec{p}$ は比 p_x/p_y の形でだけ現れている (比例定数が不要である) ことに注意.

q_{xy} **提案分布** 自由に決めていい転置推移確率行列. これにしたがって, 次に移る状態の候補を「提案」する.

メトロポリス・ヘイスティングス法のアルゴリズム

Listing 3: MH in MCMC

```
1 //確率  $q(y, x)$  にしたがって提案  $y$  を選ぶ.
2 double getproposal(double y, double x);
3 double getuniform(void); //  $[0, 1)$  一様擬似乱数
4
5 double getnext(double x){ //  $x$ : 現在の状態
6     int y; // 次の状態
7     int yproposal;
8     double u;
9     yproposal=getproposal(y,x); // 提案.
10    u=getuniform();
11    if(u< min(1, (p(x)*q(y,x))/(p(y)*q(x,y)) )){
12        y=yproposal; // 提案を採択
13    } else {
14        y=x; // 提案を棄却
15    }
16    return y; // 次の状態
17 }
```

転置推移確率行列であること

列の和が 1 になるように, (定数) で対角成分を調節できる.

対称であること

$p_x q_{yx} \geq p_y q_{xy}$ のとき

$$M_{xy} = q_{xy} \cdot 1, \quad M_{yx} = q_{yx} \cdot \frac{p_y q_{xy}}{p_x q_{yx}}$$

よって, $M_{xy} p_y = q_{xy} p_y$. $M_{yx} p_x = q_{yx} p_y$.

$p_x q_{yx} < p_y q_{xy}$ のとき

$$M_{xy} = q_{xy} \cdot \frac{p_x q_{yx}}{p_y q_{xy}}, \quad M_{yx} = q_{yx} \cdot 1$$

よって, $M_{xy} p_y = p_x q_{yx}$. $M_{yx} p_x = p_x q_{yx}$.

メトロポリス法 (q_{xy} が対称な場合)

例

- S が有限集合のとき, $q_{xy} = 1/|S|$ (全状態平等な提案),
- $x, y \in S = \mathbb{Z}$ のとき, $q_{xy} = \delta_{x, y-1} + \delta_{x, y+1}$ (近くの状態を提案)
- $x, y \in S = \mathbb{R}$ のとき, $q_{xy} = I_{[|x-y| \leq \frac{1}{2}]}(x)$ (近くの状態を提案)

このとき q は $\min$ の中から約分される (メトロポリス法という. こちらが先に提案された).

直観的には, q_{xy} で x が提案されたなら,

- p_x が p_y 以上のときは, かならず x に移る
- 以上でないときは, 移すかどうか比に応じて決める

L14-Q1

Quiz(メトロポリス-ヘイスティングス法・MCMC)

次の問に、短いけど完全な日本語の文章で答えよう。

- ① マルコフ連鎖モンテカルロ法とは何か。確率分布、定常分布、サンプル（標本）、マルコフ連鎖という言葉を使って答えよう（これらの言葉をさらに説明する必要はない）。
- ② マルコフ連鎖モンテカルロ法において、なぜ、burn-in が必要なのか、説明しよう。
- ③ マルコフ連鎖モンテカルロ法において、なぜ、内側のループで毎回サンプルしてはいけないのか、説明しよう。
- ④ メトロポリス法の転置推移確率行列において、なぜ $\min\{1, \dots\}$ という書き方にしないといけないのか、説明しよう。
- ⑤ メトロポリス法の転置推移確率行列において、なぜ $\frac{p(x)}{p(y)}$ であって、 $\frac{p(y)}{p(x)}$ ではないのか、後で自分が思い出すのに役立つ言い方で、説明しよう。

メトロポリス法を思いついた人の心理を想像

$M_{xy} = p_x$ ってしたいけど, p_x 小さいから効率悪いな… Z わからないから比で書きたいな… または密度だと 1 越えちゃうな…

$M_{xy} = \frac{p_x}{p_x + p_y}$ 熱浴法 heat-bath もあるな…

もっと効率よくしたいな… $\frac{p_x}{p_x + p_y} : \frac{p_y}{p_x + p_y}$ じゃなくて, $\frac{p_x}{p_y} : 1$ にしちゃえ.

ここまで来たよ

13 逆関数法・棄却法による乱数の生成

14 マルコフ連鎖モンテカルロ法と詳細つりあいの条件

- マルコフ連鎖モンテカルロ法
- メトロポリス・ヘイスティングス法
- 二項モデルのベイズ推定への適用例

二項モデル（複数のベルヌイ分布）

Hoff ベイズ §3.1

計算科学 I(2024)L03

値が不明なパラメタ θ の確率で表 $Y = 1$ が出るコインを N 回投げた結果を, $Y_1, \dots, Y_N = 0, 1$ とする.

Y_i は独立同分布にしたがう. $Y_i \stackrel{\text{i.i.d.}}{\sim} \text{B}(1, \theta)$ ($n = 1, p = \theta$ の二項分布).

θ は不明ということで, 分布をもつ確率変数 Θ と思って, 前回の Quiz と

- $Y_1, \dots, Y_N$ の組 $\leftrightarrow$ 柿の質量 X
- $\Theta \leftrightarrow$ 柿の品種 AB Y

と対応させ, θ の確率分布を $y_1, \dots, y_N$ の情報で更新していきたい.
 $N = 1$ のとき, $Y_1 = y_1$ を得る確率は (なぜか $\text{given } |$ を使って),

$$f_{Y|}(y_1|\theta) = \theta^{y_1}(1 - \theta)^{1-y_1}.$$

一般の N のとき,

$$f_{Y|}(y_1, \dots, y_N|\theta) = \theta^{\sum_i y_i} (1 - \theta)^{N - \sum_i y_i}.$$

θ の分布

θ についての主観的確率（信念）は，連続型確率変数 Θ で表現できる．
何の先入観も持たないときの事前確率密度は $\Theta \sim U(0, 1)$ ，すなわち，一様分布

$$f(\theta) = \begin{cases} 1 & (0 \leq \theta \leq 1) \\ 0 & (\text{他}) \end{cases}$$

と考えられる．

このとき， $f_{Y|}(y_1, \dots, y_N | \theta)$ は条件付き確率であることが納得できる．

ベイズの定理より, $Y_1 = y_1, \dots, Y_N = y_N$ という結果を得たときの θ の事後確率密度は,

$$\begin{aligned} f_{|Y}(\theta|y_1, \dots, y_N) &= \frac{f(y_1, \dots, y_N, \theta)}{f_Y(y_1, \dots, y_N)} \\ &= \frac{f_{Y|} (y_1, \dots, y_N|\theta) f(\theta)}{f_Y(y_1, \dots, y_N)} \\ &= \frac{\theta^{\sum_i y_i} (1-\theta)^{N-\sum_i y_i} \cdot 1}{f_Y(y_1, \dots, y_N)} \quad (0 \leq \theta \leq 1) \end{aligned}$$

~~分子だけ見ると θ 依存性はわかるが, $Z =$ 分母まで正確に求めよう.~~

$$f_Y(y_1, \dots, y_N) = \int_0^1 p(y_1, \dots, y_N, \theta) d\theta = \int_0^1 \theta^{\sum_i y_i} (1-\theta)^{N-\sum_i y_i} \cdot 1 d\theta = \dots$$

~~ベータ分布, ベータ関数の迷宮へ~~

計算科学 I(2024)L03

MCMC の言葉では

マルコフ連鎖の状態 $x, y = \theta \in S = [0, 1]$.

ややこしい…上の Y は別の確率変数なので、ここでは状態を $x = \theta, y = \phi$ と書く.

提案分布の確率密度は $q_{\theta\phi} = 1$ などでもよい. 対称なのでメトロポリス法でいい.

$$M_{\theta\phi} = \min\left\{1, \frac{\theta^{\sum_i y_i} (1 - \theta)^{N - \sum_i y_i}}{\phi^{\sum_i y_i} (1 - \phi)^{N - \sum_i y_i}}\right\}.$$

上の M の計算には何も難しいところがない. ϕ は今のパラメタ, θ は提案された次のパラメタ. $y_1, \dots, y_N$ は観測結果のデータで, MCMC の実行中は単なる整数定数.

- $q_{\theta\phi} = 5\mathbf{I}_{[|\theta-\phi|=0.1]}(\theta)$ などとしてもいいが, 端 $y = 0, 1$ では対称でなくなるのでメトロポリス-ヘイスティングス法で.