第10章　計算機集約的手法

10.1　乱数の発生法

Monte Carlo 法は乱数生成法のことである．区間 (0, 1) 上の一様乱数の列を発生させる．この列は決定的なアルゴリズムにより生成されるので真の意味ではランダム性を持たないので, 疑似乱数とよばれるものである．R では高速で安定した一様疑似乱数を発生するアルゴリズムである Mersenne twister 法を利用している．一様乱数を使って代表的な分布に従う疑似乱数を生成する．ここでは, 一様論数からある分布に従う (疑似) 乱数をどのように生成させるかを説明する． まず, c.d.f や p.d.f.(p.m.f.) の性質を直接利用して, 目標の分布に従う乱数の生成法 (確率積分法とよぶことにする) をまず説明する．つぎに, 本質的に計算機を利用すること (計算機集中的な手法) で目標の疑似乱数を生成する方法である受容・棄却法やマルコフ連鎖法を説明する．

以下の議論では, 開区間 (0, 1) 上の一様疑似乱数は用意に生成し, そのように目的の分布の疑似乱数を生成するかを説明していく．以下では, (0, 1) 上の一様分布に対する疑似乱数のことを簡単に一様疑似乱数とよぶことにする．すなわち, 分布 P の疑似乱数を作成したとき, 一様疑似乱数から数 $x_1, x_2, \ldots, x_n$ を作成する．もちろん, これらは互いに関連しない数[1] である．これらが任意の開集合 $A \subset \mathbb{R}$ に対して

$$\mathsf{P}(A) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(x_i)$$

となるようにする．上記のような関係をみたす数 $x_1, x_2, \ldots, x_n$ を P の疑似乱数ということにする．

10.1.1　ベルヌーイ分布

$0 < \theta < 1$ なる θ に対して, Ber(θ) に従う乱数を発生させる場合には

- 一様乱数 u を発生させる．

[1]確率変数の言葉でいえば, 独立となるが, 数に対して独立を定義できないであろう．

- $0 < u \leq \theta \Longrightarrow x = 1$, $\theta < u < 1 \Longrightarrow x = 0$ とおく.

このとき x は $\mathsf{Ber}(\theta)$ の疑似乱数となる.

10.1.2　2 項分布

$n \in \mathbb{N}$, $0 < \theta < 1$ とする.

- $\mathsf{Ber}(\theta)$ の疑似乱数 $x_1, x_2, \ldots, x_n$ を発生させる.
- $s = x_1 + x_2 + \cdots + x_n$ とする.

このとき s は $\mathsf{Bin}(n, \theta)$ の疑似乱数となる.

10.1.3　標準正規分布

Box-Müller 法を用いる.

- 一様乱数 u_1, u_2 を発生させる.
- $r = \sqrt{-2\log u_1}$, $\theta = 2\pi u_2$ とする.
- $x = r\cos\theta$, $y = r\sin\theta$ とする.

このとき x, y は　$\mathsf{N}(0, 1)$ の 2 つの疑似乱数となる.

これは以下のことからわかる. 大文字が確率変数で小文字はその実現値とする. たとえば, r は確率変数 R の実現値とみる. まず R の p.d.f. は

$$\mathsf{p}^R(r) = \begin{cases} r\exp\left(-\dfrac{r^2}{2}\right) & (r > 0) \\ 0 & (r \leq 0) \end{cases}$$

となることに注意する. したがって (R, Θ) の同時 p.d.f. は

$$\mathsf{p}^{(R,\Theta)}(r, \theta) = \begin{cases} \dfrac{1}{2\pi} r\exp\left(-\dfrac{r^2}{2}\right) & (r > 0,\ 0 < \theta < 2\pi) \\ 0 & (\text{その他の場合}) \end{cases} \tag{10.1}$$

である. ここで変換

$$\boldsymbol{h} = (h_1, h_2) : (r, \theta) \mapsto (r\cos\theta, r\sin\theta) = (x, y)$$

を考える.

$$r = \sqrt{x^2 + y^2},\ \cos\theta = \frac{x}{\sqrt{x^2 + y^2}},\ \sin\theta = \frac{y}{\sqrt{x^2 + y^2}}$$

である. 変換 $\boldsymbol{h}$ の Jacobian は

$$J_{\boldsymbol{h}}(r,\,\theta) = \det\begin{pmatrix} \dfrac{\partial h_1}{\partial r} & \dfrac{\partial h_1}{\partial \theta} \\ \dfrac{\partial h_2}{\partial r} & \dfrac{\partial h_2}{\partial \theta} \end{pmatrix} = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r$$

となるので

$$J_{\boldsymbol{h}^{-1}}(x,\,y) = \frac{1}{r} = \frac{1}{\sqrt{x^2+y^2}}.$$

(10.1) に対して p.d.f. の変換公式を用いれば

$$\begin{aligned}
\mathsf{p}^{(X,Y)}(x,\,y) &= p^{(R,\Theta)}\left(\sqrt{x^2+y^2},\,\arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right)\right)\frac{1}{\sqrt{x^2+y^2}} \\
&= \frac{1}{2\pi}\sqrt{x^2+y^2}\exp\left(-\frac{x^2+y^2}{2}\right)\frac{1}{\sqrt{x^2+y^2}} \\
&= \frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)
\end{aligned}$$

となることよりわかる.

10.1.4 正規分布

$-\infty < \mu < \infty,\, 0 < \sigma < \infty$ とする. $\mathsf{N}(\mu,\,\sigma^2)$ に従う疑似乱数を生成させるには

- $\mathsf{N}(0,\,1)$ の疑似乱数 z を発生させる.
- $x = \mu + \sigma z$ とおく.

このとき x は $\mathsf{N}(\mu,\,\sigma^2)$ の疑似乱数となる.

10.1.5 確率の積分変換

連続型分布の c.d.f. F に対してその逆関数 F^{-1} が陽にわかれば F に従う疑似乱数を発生させることができる.

- 一様疑似乱数 u を発生させる.
- $x = \mathsf{F}^{-1}(u)$ とおく. ただし, c.d.f. F は, $\mathsf{supp}(\mathsf{F}) := \overline{\{x \in \mathbb{R};\, \mathsf{F}(x) > 0\}}$ 上で連続かつ狭義単調増加であることを仮定している.

このとき, x は F の疑似乱数となる. なぜならば, $X \sim \mathsf{F}$ としたとき, $\forall x \in \{x \in \mathbb{R};\, \mathsf{F}(x) > 0\}$ に対して

$$\mathsf{Pr}\big(X \le x\big) = \mathsf{Pr}\big(\mathsf{F}^{-1}(U) \le x\big) = \mathsf{Pr}\big(U \le \mathsf{F}(x)\big) = \mathsf{F}(x)$$

よりわかる.

10.1.6 受容・棄却法

疑似乱数をこれまでに説明してきた確率積分法等で発生させるのが難しい (疑似) 乱数の生成法を考える． 生成した乱数が従う分布の p.d.f. を p とする． さらに, 容易に乱数が生成できる分布の p.d.f. を q とする．この q はつぎのように取ることができたとする．

ある定数 $C \geq 1$ が存在して $\forall x$ に対して

$$\mathsf{p}(x) \leq C\mathsf{q}(x)$$

をみたす．この q に従う疑似乱数を用いて p に従う疑似乱数を発生させることができる．

- q の疑似乱数 x を発生させる．
- $\mathsf{Unif}\big(0,\, C\mathsf{q}(x)\big)$ の疑似乱数 u を発生させる．
- $\mathsf{p}(x) \geq u \Rightarrow \widetilde{x} = x$, $\mathsf{p}(x) < u \Rightarrow x$ を捨てる．

このとき $\widetilde{x}$ は p の疑似乱数となる．

このことは以下からわかる．まず

$$X \sim \mathsf{q}, \quad U|\, X = x \sim \mathsf{Unif}\big(0,\, C\mathsf{q}(x)\big)$$

に注意する．$-\infty < x < \infty$, $0 < u < C\mathsf{q}(x)$ のとき, $(X,\, U)$ の同時 p.d.f. $\mathsf{p}^{(X,\,U)}$ は

$$\mathsf{p}^{(X,\,U)}(x,\, u) = \mathsf{p}_{U|\,X}(u|\, x)\mathsf{q}(x) = \frac{1}{C\mathsf{q}(x)}\mathsf{q}(x) = \frac{1}{C}$$

で与えられることに注意する．よって

$$\mathsf{p}^{(X,\,U)}(x,\, u) = \begin{cases} \dfrac{1}{C} & (-\infty < x < \infty,\, 0 < u < C\mathsf{q}(x)) \\ 0 & (\text{その他の場合}) \end{cases}$$

である[2]. $t \in \mathsf{supp}(\mathsf{p}) \subset \mathbb{R}$ に対して

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\bigl(\widetilde{X} \le t\bigr) &= \frac{d}{dt}\mathsf{Pr}\bigl(X \le t \bigm| \mathsf{p}(X) \ge U\bigr) \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}\bigl(X \le t,\, \mathsf{p}(X) \ge U\bigr)}{\mathsf{Pr}\bigl(\mathsf{p}(X) \ge U\bigr)} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_0^t \left\{\int_0^{\mathsf{p}(x)} \mathsf{p}^{(X,U)}(x,\, u)\,\mathrm{d}u\right\}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\left\{\int_0^{\mathsf{p}(x)} \mathsf{p}^{(X,U)}(x,\, u)\,\mathrm{d}u\right\}\mathrm{d}x} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^t \left\{\int_0^{\mathsf{p}(x)} \frac{1}{C}\,\mathrm{d}u\right\}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\left\{\int_0^{\mathsf{p}(x)} \frac{1}{C}\,\mathrm{d}u\right\}\mathrm{d}x} \\
&= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^t \frac{\mathsf{p}(x)}{C}\,\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty} \frac{\mathsf{p}(x)}{C}\,\mathrm{d}x} \\
&= \frac{\dfrac{\mathsf{p}(t)}{C}}{\dfrac{1}{C}\displaystyle\int_{-\infty}^{\infty}\mathsf{p}(x)\,\mathrm{d}x} = \mathsf{p}(t)
\end{aligned}
$$

からわかる.

注意 10.1. $C \ge 1$ としているが, $1/C$ の割合で X を採用しているので, C は 1 に近いほど乱数生成の計算効率はよいことがわかる. □

例 10.2. $\alpha = 1.3$, $\lambda = 5.6$ としたときの $\mathsf{Ga}(\alpha,\, \lambda)$ に従う乱数を発生させよう. すなわち

$$
\mathsf{p}(x) = \begin{cases} \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & (x \ge 0) \\ 0 & (x < 0) \end{cases}
$$

に従う疑似乱数を生成する. ただし $\Gamma(\alpha) = \displaystyle\int_0^\infty x^{\alpha-1}e^{-x}\,\mathrm{d}x$ である. 簡単

[2]もちろん, $\mathsf{p}^{(X,U)}$ は同時 p.d.f. の条件をみたしている. なぜならば

$$
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathsf{p}^{(X,U)}(x,\, u)\,\mathrm{d}x\,\mathrm{d}u = \frac{1}{C}\int_{-\infty}^{\infty}\left\{\int_0^{C\mathsf{q}(x)} \mathrm{d}u\right\}\mathrm{d}x = \int_{-\infty}^{\infty}\mathsf{q}(x)\,\mathrm{d}x = 1
$$

からわかる.

に生成できる乱数の p.d.f. q として $\mathsf{Exp}(4)$ を用いる. すなわち

$$\mathsf{p}(x) = \begin{cases} 4e^{-4x} & (x \geq 0) \\ 0 & (x < 0). \end{cases}$$

である. $C = 1.2$ とすれば $x \geq 0$ のとき

$$\mathsf{p}(x) \leq C\mathsf{q}(x)$$

となることが計算機で確認することができる. $\mathsf{Exp}(4)$ の c.d.f. G は

$$\mathsf{G}(x) = \begin{cases} 1 - e^{-4x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

となるので, $0 < u < 1$ に対して

$$\mathsf{G}^{-1}(u) = -\frac{\log(1-u)}{4}$$

となる. さらに $1 - U \sim \mathsf{Unif}(0,\,1)$ であることに注意すれば

$$U \sim \mathsf{Unif}(0,\,1) \Rightarrow X = -\frac{\log U}{4} \sim \mathsf{Exp}(4)$$

となることがわかる.

よって, 以上の議論を踏まえると分布 p の疑似乱数の生成アルゴリズムは以下のようになる.

- 一様疑似乱数 v_1 を発生させる.
- $x = -\dfrac{\log v_1}{4}$ とおく.
- 別の一様疑似乱数 v_2 を発生させる. $u = C\mathsf{q}(x)v_2$ とおく.
- $u \leq \mathsf{p}(x) \Rightarrow \widetilde{x} = x$ とする. $u > \mathsf{p}(x)$ のときには x を捨てる.

この操作を繰り返せば, p に従う疑似乱数列を得る. □

10.2 Markov 連鎖 Monte Carlo 法 (MCMC 法)

μ を $\mathbb{R}^n$ 上の確率測度とする. すなわち $\mu : \mathcal{B}(\mathbb{R}^n) \ni A \mapsto \mu(A) \in [0,\,1]$ は以下の条件をみたす.

- $\mu(\mathbb{R}^n) = 1.$

- $A_1, A_2, \ldots \in \mathcal{B}(\mathbb{R}^n)$, $A_i \cap A_j = \varnothing\,(i \neq j)$ に対して

$$\mu\Bigl(\bigcup_{j=1}^{\infty} A_j\Bigr) = \sum_{j=1}^{\infty} \mu(A_j).$$

ただし, $\mathcal{B}(\mathbb{R}^n)$ は $\mathbb{R}^n$ 上の Borel 集合族である.
$f : \mathbb{R}^n \Rightarrow \mathbb{R}$ は μ 可積分関数としたとき積分

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

を求めた.
求積法では台 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^N$ と重み $w_1, w_2, \ldots, w_n\,(N \in \mathbb{N})$ をうまく選んで

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) \approx \sum_{j=1}^{N} w_j f(\boldsymbol{x}_j)$$

と近似することを目指す.
Monte Carlo 積分では, 台 $\boldsymbol{x}_j$ を確率測度 μ からランダムに生成させ, 重み w_j は分布 μ によって決める. X を μ に従う確率ベクトルとしたとき

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x}) = \mathsf{E}[f(X)] \approx \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{x}_j) \tag{10.2}$$

となることが期待される. ただし $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ は μ からのランダム標本の実現値である. MCMC 法では (10.2) をみたすような標本を生成する体系的な方法である.

10.2.1 基本的な考え方

$\mathcal{B} := \mathcal{B}(\mathbb{R}^d)$ を $\mathbb{R}^d$ 上の Borel 集合族とする. 写像

$$P : \mathbb{R}^d \times \mathcal{B} \longrightarrow [0,\,1]$$

は次の条件をみたす**確率推移核** (probability transition kernel) とよぶ.

(1) 各 $B \in \mathcal{B}$ に対して写像

$$\mathbb{R}^d \ni \boldsymbol{x} \mapsto P(\boldsymbol{x},\,B) \in [0,\,1]$$

は可測関数.

(2) 各 $\boldsymbol{x} \in \mathbb{R}^d$ に対して写像

$$\mathcal{B} \ni B \mapsto P(\boldsymbol{x},\, B) \in [0,\, 1]$$

は確率測度.

離散時間確率過程は確率変数 $\boldsymbol{X}_j \in \mathbb{R}^d\,(j = 1,\, 2,\, \ldots)$ の添え字集合 $\mathbb{N}$ で順序付けられた集合 $\{\boldsymbol{X}_j\}_{j=1}^{\infty}$ である. 推移核 P をもつ時間等質的な Markov 連鎖とは確率過程 $\{\boldsymbol{X}_j\}_{j=1}^{\infty}$ で次の性質を持つものである.

$$\begin{aligned}\mu^{\boldsymbol{X}_{j+1}}(B_{j+1}|\,\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{x}_j) = \mu^{\boldsymbol{X}_{j+1}}(B_{j+1}|\,\boldsymbol{x}_j) = P(\boldsymbol{x}_j,\, B_{j+1})\\ (B_{j+1} \in \mathcal{B},\, \boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{x}_j \in \mathbb{R}^d).\end{aligned}$$

ただし $\boldsymbol{X}_j$ の確率測度を $\mu_{\boldsymbol{X}_j}$ と書き, $\boldsymbol{X}_1 = \boldsymbol{x}_1,\, \boldsymbol{X}_2 = \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{X}_j = \boldsymbol{x}_j$ を与えたときの $\boldsymbol{X}_{j+1}$ の条件付き確率測度を $\mu_{\boldsymbol{X}_{j+1}}(\,\cdot\,|\,\boldsymbol{x}_1,\, \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{x}_j)$ と書いた. すなわち $\boldsymbol{X}_1 = \boldsymbol{x}_1,\, \boldsymbol{X}_2 = \boldsymbol{x}_2,\, \ldots,\, \boldsymbol{X}_j = \boldsymbol{x}_j$ を与えたときの $\boldsymbol{X}_{j+1}$ の条件付き分布は $\boldsymbol{X}_j = \boldsymbol{x}_j$ にのみ依存する. そして $\boldsymbol{X}_{j+1}$ の分布 $\mu_{\boldsymbol{X}_{j+1}}$ は時間に関して等質である. すると $\boldsymbol{X}_{j+1}$ の分布 $\mu_{\boldsymbol{X}_{j+1}}$ は次のように表現される.

$$\mu^{\boldsymbol{X}_{j+1}}(B_{j+1}) = \int_{\mathbb{R}^d} P(\boldsymbol{x}_j,\, B_{j+1})\mu_{\boldsymbol{X}_{j+1}}(\mathrm{d}\boldsymbol{x}_j) =: \mu_{\boldsymbol{X}_{j+1}} P(B_{j+1}).$$

さらに 3 つの概念を導入する. $\mu P(\,\cdot\,) = \int_{\mathbb{R}^d} P(\boldsymbol{x},\,\cdot\,)\,d\mu(\boldsymbol{x})$ とおく.

- 測度 μ は推移核 $P(\boldsymbol{x}_j,\, B_{j+1})$ の**不変測度**であるとは

  $$\mu P = \mu \iff \int_{\mathbb{R}^d} P(\boldsymbol{x},\, B)\,\mathrm{d}\mu(\boldsymbol{x}) = \mu(B) \qquad (B \in \mathcal{B})$$

  をみたすときをいう.

- 推移核 P は μ に関して**既約**であるとは各 $\boldsymbol{x} \in \mathbb{R}^d$ と $B \in \mathcal{B}$ で $\mu(B) > 0$ なるものに対して, ある正の整数 k が存在して

  $$P^{(k)}(\boldsymbol{x},\, B) > 0$$

  をみたすときをいう. ただし

  $$\begin{aligned}P^{(k)}(\boldsymbol{x}, B) &= \int_{\mathbb{R}^d} P(\boldsymbol{y},\, B) P^{(k-1)}(\boldsymbol{x},\, \mathrm{d}\boldsymbol{y})\\ P^{(1)}(\boldsymbol{x},\, B) &= P(\boldsymbol{x},\, B) \quad (\boldsymbol{x} \in \mathbb{R}^d,\, B \in \mathcal{B})\end{aligned}$$

  である.

- P は既約な推移核とする. P は**周期的**であるとはある整数 $m \geq 2$ が存在して互いに異なる空でない集合 $E_j\ (j = 1, 2, \ldots, m)$ と $\boldsymbol{x} \in E_j$ に対して
$$P(\boldsymbol{x}, E_{j+1(\bmod m)}) = 1 \qquad (\forall x \in E_j)$$
であるときをいう. P が周期的でないとき **非周期的**であるという.

命題 10.3. μ は $\mathbb{R}^d$ 上の確率測度とし, $\{\boldsymbol{X}_j\}_{j=1}^{\infty}$ は推移核 P を持つ時間等質な Markov 連鎖とする. さらに μ は推移核 P に関して不変測度とし, P は既約で非周期的とする. このとき $\forall \boldsymbol{x} \in \mathbb{R}^d$ に対し

$$\lim_{N\to\infty} P^{(N)}(\boldsymbol{x}, B) = \mu(B) \qquad (B \in \mathcal{B}) \tag{10.3}$$

と μ 可積分な関数 f に対して

$$\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} f(\boldsymbol{X}_j) = \int_{\mathbb{R}^d} f(\boldsymbol{x})\,\mathrm{d}\mu(\boldsymbol{x})$$

がほとんど確実に成立する.

Proof. 信じることにする. □

注意 10.4. 生成する乱数の従う分布を μ と考えることになる. 既約で非周期的な推移確率核 P がわかると μ は (10.3) でつくることができるわけである. P の構成の仕方として, 次の節で詳解する Metropolitan-Hasting 法が有名である. □

10.2.2 推移核の Metropolis-Hastings による構成

μ は $\mathbb{R}^d$ 上の確率測度とし $\mathbb{R}^d$ 上の Lebesgue 測度 $\mathsf{m}(\cdot)$ に関して絶対連続と仮定する. このことより $\mathbb{R}^d$ 上の非負関数 $\pi(\cdot)$ が存在して

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mathsf{m}}(\boldsymbol{x}) = \pi(\boldsymbol{x})$$

と書ける.

P は任意の推移核とする. 点 $\boldsymbol{x} \in \mathbb{R}^d$ が与えられたとき推移核 P は点 $\boldsymbol{x}$ を異なる点 $\boldsymbol{y}$ に移動させるか, そのまま留める. このことから推移核 P は以下のように分解できると仮定する.

$$P(\boldsymbol{x}, B) = \int_B K(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{y}) + r(\boldsymbol{x})\mathbb{1}_B(\boldsymbol{x}).$$

ただし

$$\mathbb{1}_B(\boldsymbol{x}) = \begin{cases} 1 & (\boldsymbol{x} \in B) \\ 0 & (\boldsymbol{x} \notin B) \end{cases}$$

である．核 $K(\boldsymbol{x},\,\boldsymbol{y}) : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0,\,\infty)$ に対して, $\boldsymbol{x}$ を $\boldsymbol{y}$ を含む微小な領域 $d\mathsf{m}(\boldsymbol{y})$ に移す確率は $K(\boldsymbol{x},\,\boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{y})$ とし, $\boldsymbol{x}$ に留めおく確率を $r(\boldsymbol{x})$ である．条件 $P(\boldsymbol{x},\,\mathbb{R}^d) = 1$ から

$$r(\boldsymbol{x}) = 1 - \int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\,d\mathsf{m}(\boldsymbol{y}) \tag{10.4}$$

を得る．

次に μ が P に関して不変になるための条件を K と π で表現してみよう．

まず $B \in \mathcal{B}$ に対して

$$\begin{aligned}
\mu P(B) &= \int_{\mathbb{R}^d} P(\boldsymbol{x},\,B)\,\mathrm{d}\mu(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} P(\boldsymbol{x},\,b)\frac{d\mu}{dm}(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) = \int_{\mathbb{R}^d} P(\boldsymbol{x},\,B)\,\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} \left\{\int_B K(\boldsymbol{x},\,\boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{y}) + r(\boldsymbol{x})\mathbb{1}_B(\boldsymbol{x})\right\}\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) \\
&= \int_{\mathbb{R}^d} \left\{\int_B K(\boldsymbol{x},\,\boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{y})\right\}\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) + \int_{\mathbb{R}^d} r(\boldsymbol{x})\pi(\boldsymbol{x})\mathbb{1}_B(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) \\
&= \int_B \left\{\int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,d\mathsf{m}(\boldsymbol{x})\right\}\mathrm{d}\mathsf{m}(\boldsymbol{y}) + \int_B r(\boldsymbol{y})\pi(\boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{y}) \\
&\qquad (2\text{ 項目の積分の変数を } \boldsymbol{x} \text{ から } \boldsymbol{y} \text{ に変更}) \\
&= \int_B \left\{\int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + r(\boldsymbol{y})\pi(\boldsymbol{y})\right\}\mathrm{d}\mathsf{m}(\boldsymbol{y}) \qquad (10.5)
\end{aligned}$$

と書きかえられることに注意する．すると $\pi(\boldsymbol{x})\,d\mathsf{m}(\boldsymbol{x})$ が P に関して不変測度になるためには

$$\begin{aligned}
\mu P = \mu &\iff \mu P(B) = \mu(B) \qquad (\forall B \in \mathcal{B}) \\
&\iff \int_B \left\{\int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) + r(\boldsymbol{y})\pi(\boldsymbol{y})\right\}\mathrm{d}\mathsf{m}(\boldsymbol{y}) \\
&\qquad\quad = \int_B \pi(\boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{y}) \qquad (\forall B \in \mathcal{B}) \\
&\iff \int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) + r(\boldsymbol{y})\pi(\boldsymbol{y}) = \pi(\boldsymbol{y}) \qquad (\forall \boldsymbol{y} \in \mathbb{R}^d) \\
&\iff \int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) = \pi(\boldsymbol{y})(1 - r(\boldsymbol{y})) \qquad (\forall \boldsymbol{y} \in \mathbb{R}^d)
\end{aligned} \tag{10.6}$$

となる．(10.4) を (10.6) の最右辺に代入 ($\boldsymbol{x}$ を $\boldsymbol{y}$ に変更した式) すると

$$\begin{aligned}
&\int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) = \pi(\boldsymbol{y})\int_{\mathbb{R}^d} K(\boldsymbol{y},\,\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) \\
&\iff \int_{\mathbb{R}^d} K(\boldsymbol{x},\,\boldsymbol{y})\pi(\boldsymbol{x})\,\mathrm{d}\mathsf{m}(\boldsymbol{x}) = \int_{\mathbb{R}^d} K(\boldsymbol{y},\,\boldsymbol{x})\pi(\boldsymbol{y})\,\mathrm{d}\mathsf{m}(\boldsymbol{x})
\end{aligned} \tag{10.7}$$

を得る．この式を**釣合方程式** (balance equation) とよぶ．K が

$$\pi(\boldsymbol{x})K(\boldsymbol{x},\,\boldsymbol{y}) = \pi(\boldsymbol{y})K(\boldsymbol{y},\,\boldsymbol{x}) \quad (\forall \boldsymbol{x},\,\boldsymbol{y} \in \mathbb{R}^d) \tag{10.8}$$

をみたすとき K は**詳細的釣合方程式** (detailed balance equation) をみたすという．

Metropolis-Hastings アルゴリズムでは (10.8) をみたす推移核 K を構成する．

$q : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0,\,\infty)$ は任意の関数で

$$\int q(\boldsymbol{x},\,\boldsymbol{y})\,\mathrm{dm}(\boldsymbol{y}) = 1$$

をみたす．この関数 q を用いて推移核 Q を

$$Q(\boldsymbol{x},\,B) = \int_B q(\boldsymbol{x},\,\boldsymbol{y})\,\mathrm{dm}(\boldsymbol{y}) \qquad (\forall B \in \mathcal{B})$$

で定める．q が詳細的釣合方程式をみたし

$$K(\boldsymbol{x},\,\boldsymbol{y}) = q(\boldsymbol{x},\,\boldsymbol{y}), \quad r(\boldsymbol{x}) = 0$$

と仮定する．みたしていなければ乗数倍で修正すればよいので, 一般性を失わずに上記のことを仮定してよいことが以下の考察からわかる．まず

$$K(\boldsymbol{x},\,\boldsymbol{y}) = \alpha(\boldsymbol{x},\,\boldsymbol{y})q(\boldsymbol{x},\,\boldsymbol{y}) \tag{10.9}$$

とおく．この α の選び方はいかのようにする． ある $\boldsymbol{x},\,\boldsymbol{y} \in \mathbb{R}^d$ に対して

$$\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x}) < \pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})$$

ならば

$$\pi(\boldsymbol{y})\alpha(\boldsymbol{y},\,\boldsymbol{x})q(\boldsymbol{y},\,\boldsymbol{x}) = \pi(\boldsymbol{x})\alpha(\boldsymbol{x},\,\boldsymbol{y})q(\boldsymbol{x},\,\boldsymbol{y})$$

となるように α を選ぶ．これは

$$\alpha(\boldsymbol{y},\,\boldsymbol{x}) = 1, \qquad \alpha(\boldsymbol{x},\,\boldsymbol{y}) = \frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})} < 1$$

とすればよいことがわかる．したがって， $\boldsymbol{x}$ と $\boldsymbol{y}$ を入れ替えることで

$$\begin{aligned} \alpha(\boldsymbol{x},\,\boldsymbol{y}) &= \min\left\{1,\,\frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\,\boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\,\boldsymbol{y})}\right\}, \\ K(\boldsymbol{x},\,\boldsymbol{y}) &= \alpha(\boldsymbol{x},\,\boldsymbol{y})q(\boldsymbol{x},\,\boldsymbol{y}) \end{aligned}$$

とすると K は詳細的釣合方程式をみたすことがわかる．実際, q が釣合方程式を満たしているときは, $\alpha = 1$ なので自明．そうでないときは $\alpha(\boldsymbol{x},\, \boldsymbol{y}) = \dfrac{\pi(\boldsymbol{y})q(\boldsymbol{y},\, \boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\, \boldsymbol{y})}$ なので

$$
\begin{aligned}
\pi(\boldsymbol{x})K(\boldsymbol{x},\, \boldsymbol{y}) &= \pi(\boldsymbol{x})\alpha(\boldsymbol{x},\, \boldsymbol{y})q(\boldsymbol{x},\, \boldsymbol{y}) \\
&= \pi(\boldsymbol{x})q(\boldsymbol{x},\, \boldsymbol{y})\frac{\pi(\boldsymbol{y})q(\boldsymbol{y},\, \boldsymbol{x})}{\pi(\boldsymbol{x})q(\boldsymbol{x},\, \boldsymbol{y})} \\
&= \pi(\boldsymbol{y})q(\boldsymbol{y},\, \boldsymbol{x})
\end{aligned}
$$

からわかる．

候補分布 q が対称とする．すなわち

$$q(\boldsymbol{x},\, \boldsymbol{y}) = q(\boldsymbol{y},\, \boldsymbol{x}) \qquad (\forall \boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^d)$$

とする．すると

$$\alpha(\boldsymbol{x},,\, \boldsymbol{y}) = \min\left\{1,\, \frac{\pi(\boldsymbol{y})}{\pi(\boldsymbol{x})}\right\} \tag{10.10}$$

となる．対称な q の例としては

$$
\begin{aligned}
&q(\boldsymbol{x},\, \boldsymbol{y}) = g(\boldsymbol{x} - \boldsymbol{y}), \\
&\quad g : \mathbb{R}^d \longrightarrow [0,\, \infty) \text{ で } g(\boldsymbol{x}) = g(-\boldsymbol{x})
\end{aligned}
$$

をみたすものがある．

以上の議論から繰り返し回数 K のアルゴリズムは以下のようになる．

Step 1 : 初期値 $\boldsymbol{x}_1 \in \mathbb{R}^d$ を選び, $k = 1$ とする．

Step 2 : $q(\boldsymbol{x}_k,\, \boldsymbol{y})$ に従う疑似乱数 $\boldsymbol{y}$ を生成する．

Step 3 : $\alpha(\boldsymbol{x},\, \boldsymbol{y}) = \min\left\{1,\, \dfrac{\pi(\boldsymbol{y})q(\boldsymbol{y},\, \boldsymbol{x}_k)}{\pi(\boldsymbol{x}_k)q(\boldsymbol{x}_k,\, \boldsymbol{y})}\right\}$ を計算する．

Step 4 : $\mathsf{Unif}(0,\, 1)$ に従う疑似乱数 t を生成する．

Step 5 :

$$
\begin{aligned}
\alpha(\boldsymbol{x}_k,\, \boldsymbol{y}) \geq t \;\Rightarrow\; \boldsymbol{x}_{k+1} = \boldsymbol{y}. \\
\alpha(\boldsymbol{x}_k,\, \boldsymbol{y}) < t \;\Rightarrow\; \boldsymbol{x}_{k+1} = \boldsymbol{x}_k
\end{aligned}
$$

とおく．

Step 6 : $k = K$ ならば $\boldsymbol{x}_k$ を求めたい乱数として出力する．$k < K$ ならば $k \longrightarrow k + 1$ として, Step 2 に戻る．

をみたすものがある.

注意 10.5. 上記アルゴリズムの Step 4 で得られた $\boldsymbol{X}_{k+1}$ の分布を確認する. 以後は, 簡単のために $d=1$ とする. $X_k = x_k$ のときの条件付き p.d.f. を $\mathsf{q}(y|\, x_k)$ とする. 以下では $d\mathsf{m}(y)$ を簡単に dy と書くことにする. $Y \sim \mathsf{q}(y|\, x_k)$ かつ $U \sim \mathsf{Unif}(0,\, 1)$ で Y と U は独立とする. このとき $t \in \mathbb{R}$ に対して

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{Pr}(X_{k+1} \le t) &= \frac{\mathrm{d}}{\mathrm{d}t}\,\mathsf{Pr}\bigl(Y \le t \bigm| \alpha(X_k,\, Y) \ge U\bigr) \qquad (\because \text{Step 5 より}) \\ &= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\,\mathsf{Pr}\bigl(Y \le t,\, \alpha(X_k,\, Y) \ge U)\bigr)}{\mathsf{Pr}\bigl(\alpha(X_k,\, Y) \ge U\bigr)} \\ &= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t}\left\{\int_0^{\alpha(x_k,\, y)} \mathrm{d}u\right\} q(x_k,\, y)\,\mathrm{d}y}{\displaystyle\int_{-\infty}^{\infty}\left\{\int_0^{\alpha(x_k,\, y)} \mathrm{d}u\right\} q(x_k,\, y)\,\mathrm{d}y} \\ &= \frac{\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_{-\infty}^{t} \alpha(x_k,\, y) q(x_k,\, y)\,\mathrm{d}y}{\displaystyle\int_{-\infty}^{\infty} \alpha(x_k,\, y) q(x_k,\, y)\,\mathrm{d}y} \\ &= \frac{\alpha(x_k,\, t) q(x_k,\, t)}{\displaystyle\int_{-\infty}^{\infty} \alpha(x_k,\, y) q(x_k,\, y)\,\mathrm{d}y}\end{aligned}$$

となり, (10.9) から, $X_k = x_k$ を与えたときの X_{k+1} の条件付き分布は

$$X_{k+1} \sim \frac{K(x_k,\, y)}{\displaystyle\int_{-\infty}^{\infty} K(x_k,\, y)\,\mathrm{d}y}$$

となることがわかる. □

例 10.6. p.d.f.

$$\pi(x) = Cx^{-4}\bigl(1 + |x|^3\bigr) \qquad (x \in \mathbb{R}) \tag{10.11}$$

から乱数を発生させることを考える. ただし

$$C^{-1} = \int_{-\infty}^{\infty} x^{-4}\bigl(1 + |x|^3\bigr)\,\mathrm{d}x$$

である. 候補 p.d.f. として

$$Y|\, x_k \sim \mathsf{N}(x_k,\, 1)$$

を採用する．すなわち

$$q(x,\,y) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}(y-x)^2\right\}$$

である．ここで π は円周率である．すると (10.10) と (10.11) から

$$\alpha(x,\,y) = \min\left\{1,\, e^{-y^4+x^4}\frac{1+|y|^3}{1+|x|^3}\right\}$$

となる．

以上の議論から繰り返し回数 K のアルゴリズムは以下のようになる．

Step 1 : 初期値 $x_1 \in \mathbb{R}$ を選び, $k=1$ とする．

Step 2 : $\mathsf{N}(x_k,\,1)$ からの疑似乱数 y を生成する．

Step 3 : $\alpha(x_k,\,y) = \min\left\{1,\, e^{-y^4+x_k^4}\dfrac{1+|y|^3}{1+|x_k|^3}\right\}$ を計算する．

Step 4 : $\mathsf{Unif}(0,\,1)$ からの疑似乱数 t を生成する．

Step 5 :

$$\alpha(x_k,\,y) \ge t \Longrightarrow x_{k+1} = y.$$
$$\alpha(x_k,\,y) < t \Longrightarrow x_{k+1} = x_k$$

とおく．

Step 6 : $k=K$ ならば $\boldsymbol{x}_k$ を求めたい乱数として出力する．$k<K$ ならば $k \longrightarrow k+1$ として, Step 2 に戻る．

□

10.3 Gibbs サンプリング法

10.4 ブートストラップ法

ブートストラップ法は推定量の標準誤差を推定する方法である．F を未知の c.d.f. とし, $X_1,\,X_2,\,\ldots,\,X_n \overset{\text{i.i.d.}}{\sim} \mathsf{F}$ とする．未知の母数には依存し

ないある関数 g があって $S_n = g(X_1, X_2, \ldots, X_n)$ と書く．すなわち, S_n は統計量である．このときに

$$T(\mathsf{F}) = \mathsf{Var}_{\mathsf{F}}\big[S_n\big]$$

を推定したいとする．一般にこの量は未知の分布 F に依存する．たとえば $S_n = \overline{X}_n$ のとき,

$$\mathsf{Var}_{\mathsf{F}}\big[S_n\big] = \frac{\sigma^2}{n}$$

であり, 未知の母数 σ に依存する．ただし $\sigma^2 = \mathsf{E}\big[(X_1 - \mu)^2\big]$, $\mu = \mathsf{E}[X_1]$ である．$T(\mathsf{F})$ の推定量として, 差し込み推定量を考える．

$$T(\widehat{\mathsf{F}}_n) = \mathsf{Var}_{\widehat{\mathsf{F}}_n}\big[S_n\big].$$

ここで 大数の法則を思い出す． $B \in \mathbb{N}$ とし, G を c.d.f. とする．$Y_1, Y_2, \ldots, Y_B \overset{\text{i.i.d.}}{\sim} \mathsf{G}$ とする．$B \to \infty$ のとき

$$\overline{Y}_B = \frac{1}{B}\sum_{j=1}^{B} Y_j \xrightarrow{P} \mathsf{E}[Y]$$

となる．ただし $Y \sim \mathsf{G}$ である．さらに, 関数 h は有限な期待値 $h(Y)$ をもつとする．すると $B \to \infty$ のとき

$$\frac{1}{B}\sum_{j=1}^{B} h(Y_j) \xrightarrow{P} \mathsf{E}\big[h(Y)\big].$$

$\mathsf{Var}_{\widehat{\mathsf{F}}_n}\big[S_n\big]$ をシミュレーションで近似することを考える．$\mathsf{Var}_{\widehat{\mathsf{F}}_n}\big[S_n\big]$ は「データが分布 $\widehat{\mathsf{F}}_n$ に独立同一にしたがっているときの S_n の分散」と考えることができる．$X_1^*, X_2^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \widehat{\mathsf{F}}_n$ とし, $S_n^* = g(X_1^*, X_2^*, \ldots, X_n^*)$ を計算する．

- 現実の世界: $\mathsf{F} \Longrightarrow X_1, X_2, \ldots, X_n \Longrightarrow S_n = g(X_1, X_2, \ldots, X_n)$.
- ブートストラップの世界: $\widehat{\mathsf{F}}_n \Longrightarrow X_1^*, X_2^*, \ldots, X_n^* \Longrightarrow S_n^* = g(X_1^*, X_2^*, \ldots, X_n^*)$.

では, $X_1^*, X_2^*, \ldots, X_n^*$ を $\widehat{\mathsf{F}}_n$ からどのように発生させるのだろうか? $\widehat{\mathsf{F}}_n$ は各 $X_j\,(j = 1, 2, \ldots, n)$ に確率 $1/n$ をのせたものであったことを思い出そう．

$\widehat{\mathsf{F}}_n$ に従う確率変数 $X_j^*\,(j = 1, 2, \ldots, n)$ はもとのデータ
$X_1, X_2, \ldots, X_n$ から無作為抽出されたものと同じ分布になる

したがって $X_1^*, X_2^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \widehat{\mathsf{F}}_n$ を発生させるためには, $X_1, X_2, \ldots, X_n$ からデータを復元抽出すればよい．手続きをまとめると以下のようになる．

- $X_1^*, X_2^*, \ldots, X_n^* \overset{\text{i.i.d.}}{\sim} \widehat{\mathsf{F}}_n$ を発生させる. $X_1, X_2, \ldots, X_n$ から復元抽出をするだけでよい.
- $S_n^* = g(X_1^*, X_2^*, \ldots, X_n^*)$ を計算する.
- 上のふたつのステップを B 回繰り返し, $S_{n,1}^*, S_{n,2}^*, \ldots, S_{n,B}^*$ を求める.
-
$$\mathsf{v}_{\text{boot}} = \frac{1}{B}\sum_{j=1}^{B}\left(S_{n,j}^* - \frac{1}{B}\sum_{k=1}^{B} S_{n,k}^*\right)^2$$
を計算する.

このとき
$$\mathsf{Var}_F(S_n) \overbrace{\approx}^{\text{そん色ない}} \mathsf{Var}_{\widehat{F}_n}(S_n) \overbrace{\approx}^{\text{小さい}} \mathsf{v}_{\text{boot}}.$$
$0 < \alpha < 1$ に対して, $z_{\alpha/2}$ を標準正規分布の両側 α 点とすれば, $T(\mathsf{F}) = \mathsf{E}[S_n]$ とし
$$\widehat{\mathsf{se}}_{\text{boot}} := \sqrt{\mathsf{v}_{\text{boot}}}$$
としたとき
$$S_n \pm z_{\alpha/2}\widehat{\mathsf{se}}_{\text{boot}}$$
は $T(\mathsf{F})$ の近似 $(1-\alpha)\%$ 信頼区間となる. すなわち
$$\mathsf{Pr}\left(S_n - z_{\alpha/2}\widehat{\mathsf{se}}_{\text{boot}} \leq T(F) \leq S_n + z_{\alpha/2}\widehat{\mathsf{se}}_{\text{boot}}\right) \approx 1-\alpha$$
となることが期待される. 実際, 適当な仮定のもとで上記の主張を理論的に正当化できることが知られている.