

サンプリング

October 19, 2023

サンプリングとは

▶ a

モンテカルロ法 (Monte Carlo method)

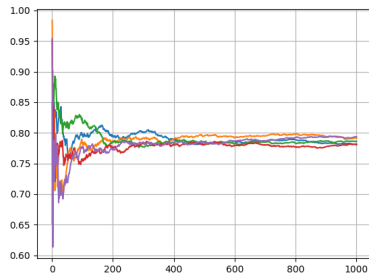
モンテカルロ法とは、乱数を用いた数値計算手法の総称。

▶ 例：定積分の計算

$$\int_0^1 \sqrt{1-x^2} dx = \frac{\pi}{4}$$

→ 乱数列 $X_1, X_2, \dots, X_n \sim \text{i.i.d. } U(0, 1)$ を生成し $f(x) = \sqrt{1-x^2}$ の平均 $\sum_{i=1}^n f(X_i)/n$ を計算すると積分 (期待値) が計算できる¹。

- * 実行例：シミュレーションを5回実行した時の結果
- * 横軸：サンプル数，縦軸：平均 $\sum_{k=1}^n f(x_i)/n$
- * サンプル数 n を大きくするにつれて，厳密値 $\pi/4 \simeq 0.785$ に収束する様子が見られる



¹大数の法則より次が成り立つ． $\sum_{i=1}^n f(X_i)/n \longrightarrow \mathbb{E}[f(x)] = 1/(b-a) \int_a^b f(x) dx \quad (n \rightarrow \infty)$

モンテカルロ法

モンテカルロ法の欠点

▶ 無駄が多い

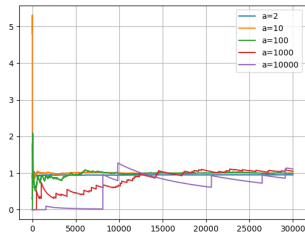
★ 例：次の積分計算は…

$$\int_{-a}^a \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\} dx$$

★ 見積もり： $|x|$ が大きい領域は積分値にほとんど寄与しない²

★ ところで、一様乱数 $X \sim U(-100, 100)$ が $|X| > 2$ となる確率は 98%.

→ 実質的なサンプル数はサンプル数 n に対して $n/100$ 程度しかない (ほぼ全て無駄³)



- * 各 a について上のガウス積分をモンテカルロ法で計算したときの結果
- * 横軸：サンプル数，縦軸：(推定した)積分値．最適値は 1 である．
- * a の値が大きい程，収束に時間がかかっていることが分かる

² $P(|X| \leq 2 \cdot \sqrt{1}) \simeq 0.95$

³高次元だとさらに無駄が増える．そもそも関数が複雑になり，見当すらつかない．

モンテカルロ法

モンテカルロ法この問題を解決するのが，マルコフ連鎖モンテカルロ法 (計算時間のほとんどを積分への寄与が大きな領域での計算に使える

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

- ▶ マルコフ連鎖モンテカルロ法とは，“確率分布を生成する”手法である。

- ★ 事後分布に関する期待値の計算

$p(w)$: 事後分布, $w_1, w_2, \dots, w_K \sim p(w)$ なる確率変数 (実現値)

$$\mathbb{E}_w[f(w)] = \int f(w)p(w) dw \approx \frac{1}{K} \sum_{k=1}^K f(w_k) \quad (\spadesuit)$$

- ★ 自由エネルギーの計算

$$F_n(1) = -\log Z_n(1) = -\sum_{k=0}^{J-1} \log \mathbb{E}_w^{(\beta_k)} \left[\exp\{-(\beta_{k+1} - \beta_k) \hat{H}(w)\} \right]$$

- ▶ マルコフ性を利用して $\{w_k\}_{k=1}^K$ ($w_1, w_2, \dots, w_K \sim p(w)$) を得る方法 ((♠) が成り立つような $\{w_k\}_{k=1}^K$ を得る方法) をマルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo method; MCMC) と呼ぶ。

- ♣ アイデア：極限分布 (定常分布) が目的の確率分布であるようなマルコフ連鎖を生成すること

MCMC

- ▶ 実現したい確率分布を π とする．極限分布が π であるようなマルコフ連鎖 $\{x^{(k)}\}_{k=1}^K$ を生成したい．
- ▶ マルコフ連鎖 $\{x^{(k)}\}_{k=1}^K$ が詳細釣り合い条件 (detailed-balance condition) を満たすように推移確率 P を設計することで達成できる⁴．

MCMC 法

推移確率密度 p をもつマルコフ連鎖 $\{x^{(k)}\}_{k=1}^K$ が次の条件を満たすならば，十分大きな K をとる時，(♠) が成り立つ．

- ▶ 詳細釣り合い条件が成り立つ⁵．

$$p(x_b|x_a)\pi(x_a) = p(x_a|x_b)\pi(x_b) \quad \forall x_a, \forall x_b \in S$$

- ▶ 状態空間の任意の要素 w の近傍に到達する確率が 0 ではない．

⁴ “十分大きな” K をとれば．後述する MCMC 法の注意点を参照．

⁵ 詳細釣り合い条件は (♠) が成り立つための十分条件であり必要条件ではない．cf. global balance

MCMC

- ▶ 状態空間が高々可算の場合，MCMC 法の枠組みは次のように述べることができる．

MCMC 法

推移確率 P をもつマルコフ連鎖 $\{x^{(k)}\}_{k=1}^K$ が次の条件を満たすならば，十分大きな K をとる時，(♠) が成り立つ．

- ▶ マルコフ連鎖 $\{x^{(k)}\}_{k=1}^K$ は既約かつ非周期的である．
- ▶ 詳細釣り合い条件が成り立つ⁶．

$$P(x'|x)\pi(x) = P(x|x')\pi(x') \quad \forall x, \forall x' \in S$$

⁶詳細釣り合い条件は (♠) が成り立つための十分条件であり必要条件ではない．

補足 1：極限分布の存在と一意性



補足2：詳細釣り合い条件と定常分布

- ▶ マルコフ連鎖 $\{x^{(k)}\}_{k=1}^K$ は定常分布をもつと仮定する（詳細釣り合い条件以外の条件から従う）。
- ▶ 詳細釣り合い条件を満たすマルコフ連鎖の定常分布は π であることを示す。

(証明) 次を示せばよい⁷。

$$\int_{x \in S} \pi(x) p(x, y) dx = \pi(y)$$

これは

$$(\text{左辺}) = \int_{x \in S} \pi(y) p(y, x) dx = \pi(y) \int_{x \in S} p(y, x) dx = \pi(y) \cdot 1$$

より従う。

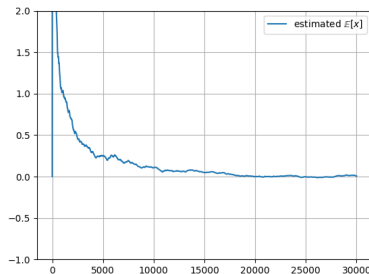
⁷状態空間が高々可算の場合には、 $\sum_{i \in S} \pi(i) P(i, j) = \pi(j)$ を示せばよい。

補足 3 : (♠) が成り立つことの証明



MCMC 法の注意点 1

- ▶ バーンイン：MCMC アルゴリズムの開始からしばらくの間は初期値の影響があるので目的とする分布からのサンプリングを考えることができない．初期値の影響が消えるまでの期間のことをバーンイン⁸ という．
 - * この期間のサンプリングは分布の近似には用いられない．
 - * バーンインの定量的な評価をするには $\rightarrow k \leq K_{\text{cut}}$ のサンプルを捨てた場合の期待値を K_{cut} の関数としてプロットし，期待値が一定になる十分大きな K_{cut} を選ぶ．



⁸ バーンインがどの程度の回数必要であるか，また，バーンインを過ぎてからどの程度の回数を繰り返す必要があるのか，という問題は重要であるが，難しい問題である．

MCMC 法の注意点 2

- ▶ 自己相関：MCMC で得られるサンプル列はマルコフ連鎖になっている
- 隣り合うサンプル $x^{(k)}$ と $x^{(k+1)}$ には相関がある．これを自己相関⁹と呼ぶ．
- ▶ 自己相関が強いと…
 - ★ 目的の確率分布になかなか到達しない
 - ★ 無駄の多いサンプルを集めてしまう
- ✓ 相関が強いサンプルを独立に扱ってしまうと誤差評価が不正確になってしまうので注意が必要
- ✓ 期待値を精度よく評価するためには十分な数の独立なサンプル¹⁰が必要

⁹自己相関を定量的に評価する方法の 1 つ：ジャックナイフ法．cf. 自己相関長，ジャックナイフ法

¹⁰素朴なモンテカルロ法で得られるサンプル列は独立であるのと異なり，MCMC で得られるそれは独立ではない

MCMC 法の注意点 3

- ▶ ポテンシャル障壁の問題： π を小さくする領域が二つ以上あって、その複数の領域の間を往復するために π が大きな場所を通過する必要があるとき、複数の領域からのサンプル回数を分布に応じて適切な割合にすることは難しいことが多い。領域と領域の間に偏りが生じやすく、場合によっては、サンプルされない領域が生じることもある。
- ポテンシャル障壁の問題という。

MCMC 法の注意点 4

▶ エントロピー障壁の問題：

→

MCMC 法の収束判定



主要な MCMC アルゴリズム

- ▶ メトロポリス法 (Metropolis method)
- ▶ ギブス・サンプリング (Gibbs sampling)
- ▶ ハイブリッドモンテカルロ法 (Hybrid Monte Carlo, Hamiltonian Monte Carlo)
- ▶ メトロポリス・ヘイスティングス法 (Metropolis Hastings algorithm; MH)
- ▶ ランジュバン・モンテカルロ法 (Langevin Monte Carlo)

MCMC アルゴリズムの応用

- ▶ simulated annealing; SA (焼きなまし法)
- ▶ レプリカ交換法

メトロポリス法 (Metropolis method)

メトロポリス法

次の形の確率分布 $p(x) := \exp\{-S(x)\}/Z$ からのサンプリングを考える．関数 $S(x)$ は既知であるとする．以下の手順に従ってサンプル $\{x^{(k)}\}_{k=1}^K$ ($x^{(k)} \sim p(x)$) を得る．

- ▶ 初期値 $x^{(1)}$ を決める．提案分布を $q(y|x)$ とする．ただし，提案分布 q は対称性 $\forall x, \forall y, q(y|x) = q(x|y)$ を満たすものとする．
- ▶ $t = 1, 2, \dots, K$ に対して以下を実行
 - ① 得られている $x^{(t)}$ から x' を条件付き確率 $q(x'|x^{(t)})$ に従って生成する．
 - ② $U \sim U(0, 1)$ を生成し， $x^{(t+1)}$ を更新する．

$$x^{(t+1)} = \begin{cases} x' & \text{if } U \leq \alpha(x, x'), \\ x^{(t)} & \text{if otherwise.} \end{cases}$$

ただし， $\alpha(x, x') := \min\left\{1, \frac{p(x')}{p(x)}\right\}$ である．

- ③ $t \leftarrow t + 1$

メトロポリス法 (概要)

- ▶ $\alpha(x, x')$ は受理確率 (acceptance prob.) と呼ばれるものであり,

$$\alpha(x, x') := \min \left\{ 1, \frac{p(x')}{p(x)} \right\} = \min \left\{ 1, \frac{\exp\{-S(x')\}}{\exp\{-S(x)\}} \right\}$$

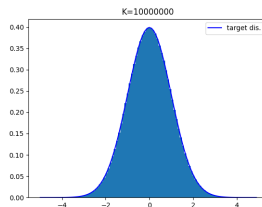
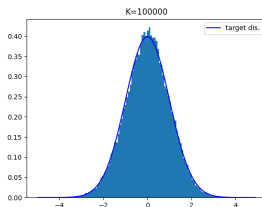
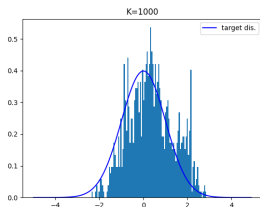
で定められる.

→ Z の値を計算する必要はない

- ▶ メトロポリス法は重点サンプリングの一種

メトロポリス法 (実行例 1)

- ▶ 例: $p(x) = \exp\{-x^2/2\}/Z$ ($x \in \mathbb{R}$) に従うサンプル列を生成する
 - ★ 初期値は $x_1 = 0$ とする.
 - ★ 提案分布は区間 $[-0.5, 0.5]$ 上の一様分布とする.
- ▶ サンプル数 $K = 10^3, 10^5, 10^7$ とした時の経験分布

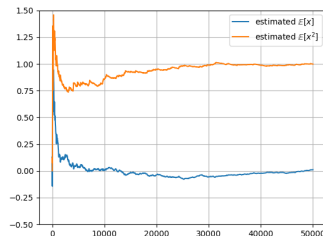


- ▶ K を大きくするにつれて目的の分布に収束していく様子が見られる

メトロポリス法 (実行例 1)

▶ サンプル数 $K = 1, \dots, 50000$ とした時の推定した期待値

- * 最適値は $\mathbb{E}[x] = 0, \mathbb{E}[x^2] = 1$.
- * K を大きくするにつれて最適値に収束していく様子が見られる
- *



メトロポリス法 (実行例 2)

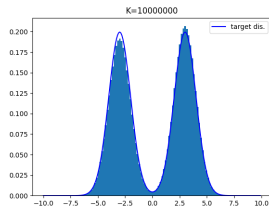
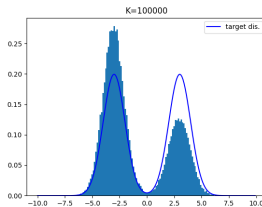
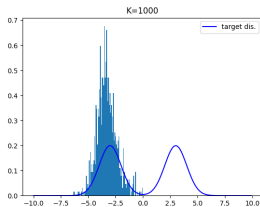
- ▶ 例：次で与えられる 2 つのガウス分布の重ね合わせからのサンプリングを行う

$$p(x) = \frac{\exp\{-(x-3)^2/2\} + \exp\{-(x+3)^2/2\}}{2\sqrt{2\pi}}$$

★ 初期値は $x_1 = 0$ とする．

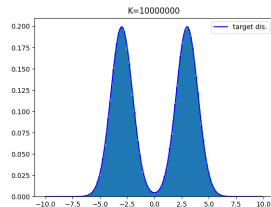
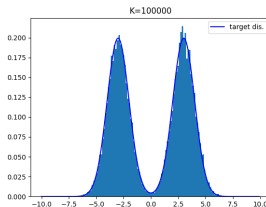
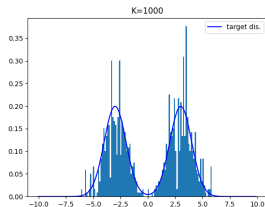
★ 提案分布は区間 $[-c, c]$ 上の一様分布とする．ここで， $c = 0.5, 5.0$

- ▶ case1. $c = 0.5$ ．サンプル数 $K = 10^3, 10^5, 10^7$ とした時の経験分布



メトロポリス法 (実行例 2)

▶ case2. $c = 5.0$. サンプル数 $K = 10^3, 10^5, 10^7$ とした時の経験分布



? 収束性能の差は何が要因か.

A. 重点サンプリングだから.

- * 重点サンプリングをするとは、確率の低い状態はできるだけ避けながらサンプリングをすること. そのため、多峰性 (multimodal) を有する分布からのサンプリングを考える場合、谷底付近からのサンプリングを避けようとする
- * メトロポリス法の場合、ステップ幅が峰を行き来できる程でなければ、片方の山からのサンプリングを繰り返すことになり、バランスよくサンプリングできない
→ 収束性能の悪化

メトロポリス法 (補足)

▶ メトロポリス法で生成されたマルコフ連鎖は詳細釣り合い条件を満たす

(証明) メトロポリス法によるマルコフ連鎖の遷移核密度 $\kappa(y|x)$ が詳細釣り合い条件を満たすことを言えばよい.

★ 遷移する場合: $q(y|x)\alpha(x, y)$

★ 遷移しない場合: 新しい場所が選択される確率を $Q(x)$ とすると

$$Q(x) = \int_{y \in S} q(y|x)\alpha(x, y) dy$$

これより元の場所 x に留まる確率は $1 - Q(x)$ なので, $\delta(y - x)(1 - Q(x))$ よって,

$$\kappa(y|x) = q(y|x)\alpha(x, y) + \delta(y - x)(1 - Q(x))$$

である¹¹.

¹¹確率密度であることに注意

メトロポリス法 (補足)

よって,

$$\begin{aligned}\kappa(y|x) \exp\{-S(x)\} &= q(y|x) \min\{\exp\{-S(x)\}, \exp\{-S(y)\}\} \\ &\quad + \delta(y-x)(1-Q(x)) \exp\{-S(x)\} \\ &= q(x|y) \alpha(y, x) \exp\{-S(y)\} \\ &\quad + \delta(x-y)(1-Q(y)) \exp\{-S(y)\} \\ &= (q(x|y) \alpha(y, x) + \delta(x-y)(1-Q(y))) \exp\{-S(y)\} \\ &= \kappa(x|y) \exp\{-S(y)\}\end{aligned}$$

より, 詳細釣り合いが従う.

ギブス・サンプリング (Gibbs sampling)

ギブス・サンプリング (systematic/coordinatewise Gibbs sampling)

A

ギブス・サンプリング (Gibbs sampling)

ギブス・サンプリング (random sweep & Grouped Gibbs sampling)

目的とする確率分布を $p(w)$ ($w \in \mathbb{R}^d$) とする．変数 w を二つの変数 w_1, w_2 に分割して $w = (w_1, w_2)$ とする．確率分布 $p(w) = p(w_1, w_2)$ から定まる条件付き確率を $p(w_1|w_2), p(w_2|w_1)$ とする．以下の手順に従ってサンプル $\{w^{(k)}\}_{k=1}^K$ ($w^{(k)} \sim p(w)$) を得る．

- ▶ 初期値 $w^{(1)}$ を決める．
- ▶ $t = 1, 2, \dots, K$ に対して以下を実行
 - ① $w^{(t)} = (w_1^{(t)}, w_2^{(t)})$ を用いて次のルールに従い $w' = (w'_1, w'_2)$ を生成する．

$$w' = \begin{cases} (w'_1, w'_2) \sim p(w'_1|w'_2)p(w'_2|w_1^{(t)}) & \text{w.p. } 1/2, \\ (w'_1, w'_2) \sim p(w'_2|w'_1)p(w'_1|w_2^{(t)}) & \text{w.p. } 1/2. \end{cases}$$

② $w^{(t+1)} \leftarrow w'$

③ $t \leftarrow t + 1$

ギブス・サンプリング (概要)

- ▶ ギブス・サンプリングは目的分布 (結合分布) の条件付き分布からのサンプリングが容易である場合に有効である
- ▶ ギブス・サンプリングはベイズ解析で利用される技術の 1 つである
- ▶ ギブス・サンプリングは，一方を止めて一方を確率的に選ぶ方法であるから，両方を同時に動かさないと移動しにくい場合には，サンプリング効率が悪くなることもある．

ギブス・サンプリング (実行例 1)



ギブス・サンプリング (実行例 1)



ギブス・サンプリング (補足)

▶ ギブス・サンプリングで生成されたマルコフ連鎖は詳細釣り合い条件を満たす
(証明) 条件付き確率 (遷移核密度) $\kappa(w'|w) = \kappa(w'_1, w'_2|w_1, w_2)$ について詳細釣り合い

$$\kappa(w'_1, w'_2|w_1, w_2)p(w_1, w_2) = \kappa(w_1, w_2|w'_1, w'_2)p(w'_1, w'_2)$$

が成り立つことを示せばよい.

$$\kappa(w'_1, w'_2|w_1, w_2) = \frac{1}{2} \{p(w'_1|w'_2)p(w'_2|w_1) + p(w'_2|w'_1)p(w'_1|w_2)\}$$

であるから

$$\begin{aligned} p(w'_1|w'_2)p(w'_2|w_1)p(w_1, w_2) &= \frac{p(w'_1, w'_2)}{p(w'_2)} \frac{p(w'_2, w_1)}{p(w_1)} p(w_1, w_2) \\ &= p(w'_1, w'_2) \frac{p(w'_2, w_1)}{p(w'_2)} \frac{p(w_1, w_2)}{p(w_1)} \\ &= p(w'_1, w'_2)p(w_1|w'_2)p(w_2|w_1) \end{aligned}$$

ギブス・サンプリング (補足)

となる．同様に

$$\begin{aligned} p(w'_2|w'_1)p(w'_1|w_2)p(w_1, w_2) &= p(w'_1, w'_2) \frac{p(w'_1, w_2)}{p(w'_1)} \frac{p(w_1, w_2)}{p(w_2)} \\ &= p(w'_1, w'_2)p(w_2|w'_1)p(w_1|w_2) \end{aligned}$$

なので

$$\begin{aligned} (\text{左辺}) &= \frac{1}{2} \{ p(w'_1, w'_2)p(w_1|w'_2)p(w_2|w_1) + p(w'_1, w'_2)p(w_2|w'_1)p(w_1|w_2) \} \\ &= \frac{1}{2} \{ p(w_2|w_1)p(w_1|w'_2) + p(w_1|w_2)p(w_2|w'_1) \} p(w'_1, w'_2) \\ &= \kappa(w_1, w_2|w'_1, w'_2)p(w'_1, w'_2) \end{aligned}$$

である．これより詳細釣り合いが得られた．

ハイブリッドモンテカルロ法 (Hybrid Monte Carlo; HMC)

- ▶ HMC 法とは：解析力学による物体の軌道のシミュレーションとメトロポリス法を組み合わせた手法．メトロポリス法の改良になっている．

♣ アイデア

- ★ 目的：確率分布 $p(w) = \exp\{-\beta H(w)\}/Z \propto \exp\{-\beta H(w)\}$ からのサンプリング
- ★ 確率の大きい点 (積分 (♠) に大きな寄与をする点) を重点的にサンプリングしたい
- ★ つまり， $H(w)$ の小さい領域 (谷底) からサンプリングしたい
- ★ そこで，次の物理的考察に注目：ポテンシャルエネルギー¹² を $H(w)$ とするハミルトニアン

$$\mathcal{H}(w, p) = \frac{1}{2} \|p\|_2^2 + \beta H(w) \quad (p: \text{運動量})$$

に従う物体は $H(w)$ の小さい領域を動き回る¹³．

- $\mathcal{H}(w, p)$ に従う物体を時間発展させた時の位置 w' を候補点として選ぶとよい¹⁴．
運動量 p はランダムに生成すればいい．

¹² 正確には違う．言葉遣いもあやしい．詳しくは解析力学へ．cf. ハミルトン方程式 etc

¹³ ハミルトニアン保存 $d\mathcal{H}/dt = 0$ による．エネルギー保存則のようなもの．

¹⁴ (後述するが) あとは受理確率を大きくできればいいが，理論的には \mathcal{H} の保存から 1 になる．

ハイブリッドモンテカルロ法

▶ アイデア ♣ を形にしたい.

→ 確率分布 $p(w, p) \propto \exp\{-\beta\mathcal{H}(w, p)\}$ に対して, サンプルの候補をアイデア ♣ によって決めると変更したメトロポリス法を行うことで実現.

- ★ 運動量 p とサンプリング: p をランダムに生成するとすれば, w と p とは独立であるから, 分布 $p(w, p)$ からのサンプリングは分布 $p(w)$ からのサンプリングと等価.
- ★ 時間発展について: 次のハミルトン方程式

$$\frac{dw}{d\tau} = \frac{\partial \mathcal{H}}{\partial p} = p, \quad \frac{dp}{d\tau} = -\frac{\partial \mathcal{H}}{\partial w} = -\beta \nabla H(w) \quad (0 \leq \tau \leq T)$$

に従って時間発展させて, T 時間後の位置 (w', p') (候補点) を得る. 微分方程式の数値解法は時間反転と位相空間の体積保存を満たしている必要がある¹⁵¹⁶. 詳細つり合い!

¹⁵誤差を含んでもよい.

¹⁶HMC 法で利用されるリープ・フロッグ法はこれら 2 つを満たす.

ハイブリッドモンテカルロ法

ハイブリッドモンテカルロ法 (Hybrid Monte Carlo; HMC)

確率分布 $p(w) := \exp\{-\beta H(w)\}/Z$ ($w \in \mathbb{R}^d$) からのサンプリングを考える．新しい変数 $p \in \mathbb{R}^d$ を導入し， $\mathcal{H}(w, p) := \|p\|_2^2/2 + \beta H(w)$ と定義する．

- ▶ $w^{(1)}$ を初期化する．
- ▶ $k = 1, 2, \dots, K$ に対して以下を実行
 - ① $p \sim \mathcal{N}(0, I)$ を生成する．
 - ② $(w^{(k)}, p)$ を初期値とする次のハミルトン方程式を (数值的に) 解いて， T 時刻後の (w', p') を得る．

$$\frac{dw}{d\tau} = \frac{\partial \mathcal{H}}{\partial p} = p, \quad \frac{dp}{d\tau} = -\frac{\partial \mathcal{H}}{\partial w} = -\beta \nabla H(w) \quad (0 \leq \tau \leq T)$$

ただし，数値解法は時間反転と位相空間の体積保存を満たさなければならない¹⁷．

- ③ $\Delta \mathcal{H} := \mathcal{H}(w', p') - \mathcal{H}(w^{(k)}, p)$ を求める．確率 $P := \min\{1, \exp\{-\Delta \mathcal{H}\}\}$ で $w^{(t+1)} = w'$ を採択し，確率 $1 - P$ で $w^{(t+1)} = w^{(t)}$ を採択する．

¹⁷HMC 法で利用されるリープ・フロッグ法は条件を満たす．

ハイブリッドモンテカルロ法 (概要)

- ▶ HMC 法はメトロポリス法の改良版になっている。
 - ★ $\exp\{-\beta\mathcal{H}(w, p)\}$ に比例する確率分布に対してメトロポリス法を適用
 - ★ w と p は独立であること、ハミルトニアンが保存すること、がポイント。
- ▶ 受理確率について： $\Delta\mathcal{H} = 0$ より、ハミルトン方程式を正確に解ければ受理確率は 1 である¹⁸。
- ▶ ハミルトン方程式の数値解法について
 - ★ 数値解法は時間反転と位相空間の体積保存を満たしていなければならない。
 - ★ リープ・フロッグ法はこれらを満たす。

¹⁸数値解法によって誤差が出ても受理確率は小さくなりやすく、 T を大きめに設定できれば受理確率を小さくすることなく離れた場所に移動することができる。

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



メトロポリス・ヘイスティングス法 (概要)



simulated annealing; SA (焼きなまし法)

- ▶ SA とは： $\max_{x \in \mathcal{X}} f(x)$ の大域的最適解を見つけるための MCMC ベースな手法
- 局所的最適解から抜け出すために，熱的な揺らぎを導入
- ♣ アイデア：点列 $\{x_t\}_t$ ($x_t \sim f(x)^{1/T_t}$) を生成すること.
 - ★ $\{T_t\}_t$ ： アニーリングスケジュールと呼ばれる温度の点列． $T_t \geq T_{t+1}$, $\lim_{t \rightarrow \infty} T_t = 0$.
 - ★ 図
- ▶ 各 x_t が正確に $f(x)^{1/T_t}$ からサンプルされれば， $\{x_t\}_t$ は $T_t \rightarrow +0$ の極限で $f(x)$ の大域的最適解に収束する¹⁹.

¹⁹実際は，MCMC サンプリングはあくまでも“近似”であるため，大域的最適解への収束は保証されない。

simulated annealing

simulated annealing; SA (焼きなまし法)

- ▶ 初期状態 x_0 と初期温度 T_0 を決める. $t = 1$ とする.
- ▶ 各 $t = 1, 2, \dots$ に対して, 停止するまで以下の手順を繰り返す.
 - ① アニーリングスケジュールに従って, 温度 T_t をとる.
 - ② 近似的に $f(x)^{1/T_t}$ からのサンプル x_t を生成する.
 - ③ $t \leftarrow t + 1$ とする.
- ▶ SA の実装には様々なやり方がある.
 - ★ どの MCMC サンプリングアルゴリズムを使用するか.
 - ✓ 一般的な SA ではサンプラーとしてメトロポリス法が用いられる
 - ★ 各温度を更新するまでに生成するマルコフ連鎖の長さ
 - ★ アニーリングスケジュール
 - ✓ よく使用されるのは幾何冷却 $T_t = \beta^t T_0$ (β : cooling factor). 適切な T_0, β は問題に依存する (チューニングが必要²⁰).

simulated annealing

- ▶ 注意点
- ▶ そもそもサンプリング MCMC は近似であって，よって最適解への収束は自明ではない
- ▶ 温度を十分にゆっくり下げる，シミュレーション時間を十分に長くすると，必ず最小値が得られる．しかし，現実的にどのくらいの時間で最小値に行きつくかは自明ではない
- ▶ 得られた答が最小値かどうかどうやって判断するかも難しい
- ▶ 実用上問題ない極小値が得られた時点で終了することもある

SA の応用

- ▶ SA の典型的な応用例 → 最適化 $\min_{x \in \mathcal{X}} S(x)$ ($S(x)$ は既知の実数値関数)
 - ★ $S(x)$ を作用とするボルツマン分布 $f(x) \propto \exp\{-S(x)\}$, $x \in \mathcal{X}$ を定める.
 - ★ $T \rightarrow +0$ の極限では, $f(x)^{1/T} \propto \exp\{-S(x)/T\}$ の (最大値を与える) 大域的最適解は $S(x)$ の (最小値を与える) 大域的最適解になっている
- ボルツマン分布 $f(x)$ に対して SA を適用することで $\min_{x \in \mathcal{X}} S(x)$ を解ける ²¹.

²¹ 最大化問題も同様に処理できるが, 正規化できない場合には適用不可

SA with metropolis for minimization²²

- ▶ 最適化問題 $\min_{x \in \mathcal{X}} S(x)$ に対応するボルツマン分布 $f(x) \propto \exp\{-S(x)\}$, $x \in \mathcal{X}$ を定める.
- ▶ 初期状態 x_0 と初期温度 T_0 を決める. $t = 1$ とする.
- ▶ 停止条件が満たされるまで, 各 $t = 1, 2, \dots$ に対し以下の手順を繰り返す.
 - ① アニーリングスケジュールに従って, 温度 T_t をとる.
 - ② 候補点 y を対称な提案分布 $q(y|x_t) = q(x_t|y)$ から生成する.
 - ③ 受理確率を計算する.

$$a(x_t, y) = \min\left\{1, \frac{f(y)}{f(x_t)}\right\} = \min\left\{1, \exp\left\{-\frac{S(y) - S(x_t)}{T_t}\right\}\right\}$$

確率 $a(x_t, y)$ で $x_{t+1} = y$ を採択し, そうでない場合は棄却し $x_{t+1} = x_t$ とする.

- ④ $t \leftarrow t + 1$.

²²各温度に対して MCMC は 1 ステップのみ行われていることに注意.

レプリカ交換法

- ▶ レプリカ交換法とは：独立な複数の系を同時にシミュレーションする手法.
- ▶ レプリカ交換法は SA を改良した手法である ← 高温領域と低温領域とを行き来することができる
- ▶ 目的：次の確率分布 (目的分布) からのサンプリング

$$p(w) = \frac{1}{Z(\beta)} \exp\{-\beta H(w)\}, w \in \mathbb{R}^d$$

$Z(\beta)$ は分配関数, $H(w)$ はエネルギー

レプリカ交換法

♣ レプリカ交換法のアイデア 1: 複数の系を同時にシミュレーションする

- ▶ 変数 w が従う系のコピー (レプリカ) を K 個作成する²³.

$$(w_1, w_2, \dots, w_K) =: W.$$

- ▶ レプリカ $1, 2, \dots, K$ の逆温度を $\beta_1, \beta_2, \dots, \beta_K$ とする.

$$p^{(\beta_j)}(w_j) = \frac{1}{Z(\beta_j)} \exp\{-\beta_j H(w_j)\}$$

- ▶ K 個のレプリカを同時にシミュレーションする. つまり, W の確率分布

$$p(W) := p(w_1, w_2, \dots, w_K) = \prod_{j=1}^K p^{(\beta_j)}(w_j)$$

からのサンプリングを行う. これは各 w_j が独立であることから各温度で MCMC アルゴリズムを適用することで可能.

²³ただし, 各々の系は相互作用しない (独立である) とする. また, 添え字 k でコピーを区別する

レプリカ交換法

レプリカ間でサンプルの交換を行うつまり，各温度ごとに (各系ごとに) 通常の MCMC を行いながら，一定の間隔で ²⁴ 逆温度 β_j の系のパラメーター w_j と逆温度 β_{j+1} の系のパラメーター w_{j+1} とを交換する．

²⁴例えば，MCMC で 3 回更新するごとにレプリカを交換する等

レプリカ交換法

レプリカ交換法

a

レプリカ交換法

- ▶ レプリカ交換法において、逆温度 β_i, β_j ($\beta_i < \beta_j$) の間で交換が行われる確率は次で与えられる.

$$P(\beta_i, \beta_j) = 1 - \frac{1}{\sqrt{\pi}} \frac{\beta_j - \beta_i}{\beta_i} \frac{\Gamma(\lambda + 1/2)}{\Gamma(\lambda)}$$

ここで、 λ は $\hat{H}(w)$ から定まる実対数閾値 (real log canonical threshold; RLCT) である. 交換確率を各温度において一定にするためには、温度 $\{\beta_k\}$ を等比数列にすればよいことがわかる.

- ▶ レプリカ交換法は、逆温度が大きい領域 (低温領域) においてポテンシャル/エントロピー障壁のためにサンプリングが困難になる状況で、大局的なサンプリングが行えるという長所を有する²⁵.

²⁵低温領域と高温領域との間に頻繁に交換が行われるほど、効率の良いサンプリングが実現されると考えられる. ただ、交換の確率と近似効率の関係については不明な事項が多い.

メモ：マルコフ連鎖

マルコフ性

- ▶ 確率過程 (stochastic process; s.p.) : 確率変数の族 $\{X_\lambda\}_{\lambda \in \Lambda}$ を確率過程という.
 - ★ Λ は添え字集合
 - ★ $\Lambda = \mathbb{N} \cup \{0\}$ のとき離散時間 s.p., $\Lambda = [0, \infty)$ のとき連続時間 s.p. という.
- ▶ 推移確率密度 $p(a, b) : a, b \in S$ に対して $p(a, b) \in [0, 1]$ が対応しているとする.
 - ★ S が高々可算集合のとき, $p(i, j)$ ($i, j \in S$) が

$$\sum_{j \in S} p(i, j) = 1 \quad \forall i \in S$$

を満たすとき, $p(i, j)$ を推移確率 (密度) という.

- ★ S が非加算のとき, $p(x, y)$ ($x, y \in S$) が

$$\int_{y \in S} p(x, y) dy = 1 \quad \forall x \in S$$

を満たすとき, $p(x, y)$ を推移確率密度 (遷移核密度) という.

メモ：マルコフ連鎖

マルコフ性

- ▶ 推移確率関数 (遷移核) P : 推移確率 (密度) を $p(a, b)$ とする.
 - ★ S が高々可算集合のとき, 次で定まる関数 P を推移確率という.

$$P(i, A) := \sum_{j \in S} p(i, j)$$

- ★ S が非加算のとき, 次で定まる関数 P を推移確率という.

$$P(x, A) := \int_{y \in A} p(x, y) dy$$

- ▶ マルコフ性：確率過程 $\{X_\lambda\}_{\lambda \in \Lambda}$ が, 任意の $s \geq 0, t \in \Lambda$ に対して次の条件 (マルコフ性) を満たすとき²⁶, (推移確率 P をもつ) マルコフ過程であるという.

$$\mathbb{P}(X_{t+s} \in A \mid \mathcal{H}_{t-1}, X_t = x) = \mathbb{P}(X_{t+s} \in A \mid X_t = x) = P(x, A)$$

ここで, \mathcal{H}_{t-1} は時刻 $t-1$ までの $\{X_\lambda\}_{\lambda \in \Lambda}$ の履歴を表す.

²⁶簡単のため, ここでは推移確率が時刻 t に依存しない場合 (time-homogeneous; 斉次的) を扱う

メモ：マルコフ連鎖

マルコフ連鎖

- ▶ 特に，離散時間マルコフ過程のことをマルコフ連鎖という．
- ▶ 以降は，簡単な説明のために状態空間 S が高々可算集合であるとして話を進めることがある²⁷．

マルコフ連鎖（離散時間マルコフ過程）

確率空間 (Ω, \mathcal{F}, P) 上の離散時間確率過程 $\{X_n\}_{n=0}^{\infty}$ が，任意の $n \in \mathbb{N} \cup \{0\}$ と任意の $j \in S$ およびに $P(X_{t_0} = i_0, \dots, X_{t_{n-1}} = i_{n-1}, X_{t_n} = i) > 0$ なる任意の $i_0, \dots, i_{n-1}, i \in S$ に対して

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P(i, j)$$

を満たすとき，推移確率 P をもつマルコフ連鎖という．

²⁷MCMC では状態空間が非加算の場合もある．適宜 “いいかんじ” に読み替えたほうがよい

参考資料

このテンプレートを作るにあたり，以下を参考にした．

- ▶ "Beamer theme gallery". theme,color,font ごとに紹介している．

https://deic.uab.es/~iblanes/beamer_gallery/index.html

- ▶ 「Beamer V3.0 を使ってみる」. Beamer の簡単な説明．

<https://www-an.acs.i.kyoto-u.ac.jp/~fujiwara/tex/beamer.html>

- ▶ "Beamer - A LaTeX class for producing presentations"

<https://github.com/josephwright/beamer>

- ▶ 「LaTeX テンプレート: Beamer スライド, Beamer ポスター, 文書」

<https://qiita.com/birdwatcher/items/5dd8a5f453bdc0c6940e>

simple

A beamer theme

simple is a minimalist Beamer theme that features
テンプレート

- ▶ a **watermark** logo in the background
- ▶ slide **numbers**
- ▶ *emphasized* and **alerted** text

And of course...

blocks, columns, and all Beamer power