「ベイズ統計の理論と方法」
渡辺 澄夫 著

5 事後分布の実現

5.1 マルコフ連鎖モンテカルロ法

令和2年1月18日

文責 @animegazer

自己紹介

- Twitter ID:
  - @animegazer（オタク活動用のアカウント、名前の通り）
  - @HironakaKenichi（真面目なアカウント、最近作った）
- 理論生物学者（cf. 数理生物学、システム生物学）
- 九大（学振DC）→理研＆阪大（学振PD）→東大（特任助教）
- 九大のときは巌佐研（緑本の久保先生の出身ラボ）
- 卒論「カオスニューラルネットワークの動的な想起過程における情報論的解析」
- 博論「キイロショウジョウバエ翅原基においてDpp依存的形態形成を支配する遺伝子制御ネットワークの理論的研究」

↑一昨年書いた記事、結構読まれて嬉しかった

↑去年出した論文（https://www.bdr.riken.jp/）

本書における５章の位置付け（個人的見解）

１～２章
- ベイズ推測の枠組みを定義

３～４章
- ベイズ推測の有効性を理論的に評価

５章
- ベイズ推測の実用で必要になる数値計算手法を紹介

６章
- ベイズ推測にまつわる様々な話題

７～８章
- 統計学の基礎

5章前半（5.1）のアウトライン

- 5.1. マルコフ連鎖モンテカルロ
  - マルコフ過程
  - 詳細釣り合い条件
- 5.1.1 メトロポリス法
  - ハイブリッド・モンテカルロ法
- 5.1.2 ギブズ・サンプリング
- 5.1.3 ランジュバン方程式を用いる方法
  - フォッカー・プランク方程式
- 5.1.4 自由エネルギーの近似
  - レプリカ交換法

5 章で考える問題：事後分布の実現

ベイズ推測とは：

　ある確率変数$\boldsymbol{x} \in \mathbb{R}^N$が従う真の確率分布 $q(\boldsymbol{x})$が、おおよそ予測分布

$$p^*(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{w})p(\boldsymbol{w}|\boldsymbol{X}^n)d\boldsymbol{w}$$

の形をしているであろうと推測すること。

$\boldsymbol{w} \in W \subset \mathbb{R}^d$：パラメータ

人間に与えられるもの：

サンプル

$\boldsymbol{X}^n = (\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n)$

人間が与えるもの：

確率モデル	事前分布	逆温度
$p(\boldsymbol{x}\|\boldsymbol{w})$	$\varphi(\boldsymbol{w})$	β

ベイズ推測を行うために、パラメータの**事後分布**

$$p(\boldsymbol{w}|\boldsymbol{X}^n) = \frac{1}{Z_n(\beta)}\varphi(\boldsymbol{w})\prod_{i=1}^{n} p(\boldsymbol{X}_i|\boldsymbol{w})^\beta$$

を求める必要があるが、これは一般的に解析的に求めることができない（注意51）。

よって、数値的に事後分布（あるいはそれによる積分）を求める必要がある。

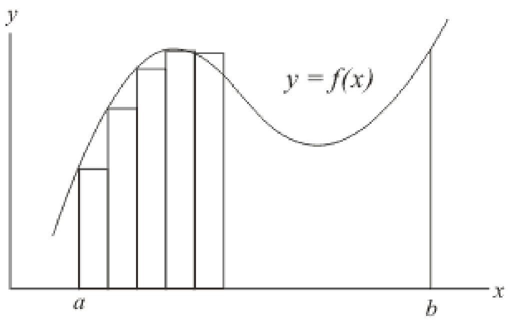
事後分布を求める数値アルゴリズムとして、
　5.1では **マルコフ連鎖モンテカルロ法（MCMC）** を
　5.2では **平均場近似法（変分ベイズ法）** を
紹介する。

(注意51) 事後分布の実現にまつわる困難

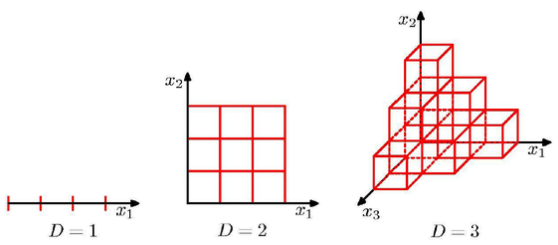
(1) 次元の呪い

　もし事後分布が陽に与えられるなら、それにまつわる平均は愚直な数値積分によって求めることができる。しかし、パラメータの次元が大きくなると格子点が爆発的に増え、計算の効率が非常に悪くなる。

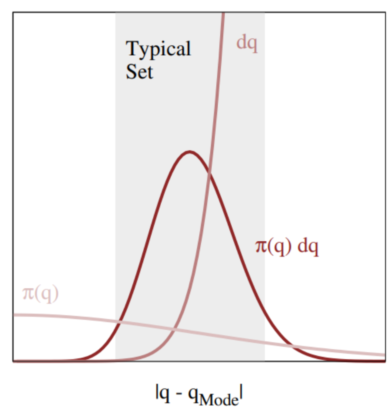
$$\int_0^1 f(w)p(w)dw \approx \frac{1}{K}\sum_{k=1}^{K} f\left(\frac{k}{K}\right)p\left(\frac{k}{K}\right)$$



http://www.isc.meiji.ac.jp/~re00108/ch03/index.html



https://slideplayer.com/slide/3526023/



https://arxiv.org/abs/1701.02434

(注意51) 事後分布の実現にまつわる困難

(2) **分配関数**

実際はそもそも事後分布を陽に与えることができないのが一般的である。というのは、規格化定数である**分配関数** $Z_n(\beta)$ が未知だからである。

普通、我々が知りうるのは事後分布が比例している関数（**ボルツマン因子**）だけである。

事後分布

$$p(\boldsymbol{w}) = \frac{1}{Z_n(\beta)} \varphi(\boldsymbol{w}) \prod_{i=1}^{n} p(\boldsymbol{X}_i|\boldsymbol{w})^{\beta}$$

$$= \frac{1}{Z_n(\beta)} \underbrace{\exp(-\beta H(\boldsymbol{w}))}_{\text{ボルツマン因子}}$$

ハミルトニアン

$$H(\boldsymbol{w}) = \widehat{H}(\boldsymbol{w}) - \frac{1}{\beta} \log \varphi(\boldsymbol{w})$$

$$= -\sum_{i=1}^{n} \log p(\boldsymbol{X}_i|\boldsymbol{w}) - \frac{1}{\beta} \log \varphi(\boldsymbol{w})$$

（尤度）+（事前分布）

・統計力学のカノニカル分布のアナロジーと思われるので、同じ用語を用いた
・ハミルトニアンや事後分布はサンプル$\boldsymbol{X}^n$に依存するので、$H(\boldsymbol{w}|\boldsymbol{X}^n)$や$p(\boldsymbol{w}|\boldsymbol{X}^n)$と書けるが、５章では大体$(\cdot\,|\boldsymbol{X}^n)$は略する（サンプルが変わることを考えない）
・後述するように、何を「ハミルトニアン」と呼ぶかは文脈に依存する（$H, \widehat{H}, \mathcal{H}$）

カノニカル分布の気持ち（1）
最適化問題の結果としてのカノニカル分布

熱力学では、U：内部エネルギー、T：温度、S：エントロピー として、等温環境で自由エネルギー $F \coloneqq U - TS$ が最小の状態が平衡状態になることが知られている。

統計力学では、物理的な系がとりうる微視的状態を確率変数 $\boldsymbol{w}$ として、人間が日常的に知覚できるレベルの巨視的状態は確率分布 $p(\boldsymbol{w})$から導かれる統計量だと考える。たとえば、$H(\boldsymbol{w})$を微視的状態$\boldsymbol{w}$のハミルトニアン、 kをボルツマン定数として、内部エネルギー $U \coloneqq E[H(\boldsymbol{w})]$、エントロピー $S \coloneqq E[-k \log p(\boldsymbol{w})]$など。

そこで、確率分布 $p(\boldsymbol{w})$の**自由エネルギー最小化問題**を考えてみると、その答えは**カノニカル分布** $p(\boldsymbol{w}) \propto \exp\bigl(-\beta H(\boldsymbol{w})\bigr)$ になる。ここで、$\beta = 1/kT$は逆温度。

（証明）

ラグランジュの未定乗数法を使う。 $L[p] \coloneqq F[p] + \lambda\bigl(\int p(\boldsymbol{w})\, d\boldsymbol{w} - 1\bigr)$ として、

$$\delta L = \int (H(\boldsymbol{w}) + kT \log p(\boldsymbol{w}) + \lambda)\delta p(\boldsymbol{w})\, d\boldsymbol{w} = 0$$

$$\therefore p(\boldsymbol{w}) \propto \exp\left(-\frac{1}{kT} H(\boldsymbol{w})\right)$$

カノニカル分布の気持ち（2）最適化問題の結果としてのカノニカル分布

熱力学では、U：内部エネルギー、T：温度、S：エントロピー として、等温環境で自由エネルギー $F \coloneqq U - TS$ が最小の状態が平衡状態になることが知られている。

統計力学では、物理的な系がとりうる微視的状態を確率変数 $\boldsymbol{w}$ として、人間が日常的に知覚できるレベルの巨視的状態は確率分布 $p(\boldsymbol{w})$から導かれる統計量だと考える。たとえば、$H(\boldsymbol{w})$を微視的状態$\boldsymbol{w}$のハミルトニアン、kをボルツマン定数として、内部エネルギー $U \coloneqq E[H(\boldsymbol{w})]$、エントロピー $S \coloneqq E[-k \log p(\boldsymbol{w})]$など。

そこで、確率分布 $p(\boldsymbol{w})$の**自由エネルギー最小化問題**を考えてみると、その答えは**カノニカル分布** $p(\boldsymbol{w}) \propto \exp\bigl(-\beta H(\boldsymbol{w})\bigr)$ になる。ここで、$\beta = 1/kT$は逆温度。

この熱力学drivenな最適化問題＝「自由エネルギーFを下げたい気持ち」の中には「内部エネルギーUを下げたい気持ち」と「エントロピーSを上げたい気持ち」が混ざっており、温度 T はそのバランスを決めるパラメータになっている。

温度 T が**ゼロ** → **内部エネルギー**最小化 → **基底状態**の実現（δ関数）

温度 T が**無限大** → **エントロピー**最大化 → **一様分布**の実現

カノニカル分布の気持ち（３）
最適化問題の結果としてのカノニカル分布

統計力学では自由エネルギーを次のように置いた：

$$F[p] = U - TS = \int p(\boldsymbol{w}) H(\boldsymbol{w}) d\boldsymbol{w} + \frac{1}{\beta} \int p(\boldsymbol{w}) \log p(\boldsymbol{w})\, d\boldsymbol{w}$$

ここで、**ベイズ推測**で定義した「ハミルトニアン」

$$H(\boldsymbol{w}) = \widehat{H}(\boldsymbol{w}) - \frac{1}{\beta} \log \varphi(\boldsymbol{w}) = -\sum_{i=1}^{n} \log p(\boldsymbol{X}_i | \boldsymbol{w}) - \frac{1}{\beta} \log \varphi(\boldsymbol{w})$$

を上の自由エネルギーに代入してやると、次のようにまとめられる：

$$F[p] = \int p(\boldsymbol{w}) \widehat{H}(\boldsymbol{w}) d\boldsymbol{w} + \frac{1}{\beta} \int p(\boldsymbol{w}) \log \frac{p(\boldsymbol{w})}{\varphi(\boldsymbol{w})}\, d\boldsymbol{w}$$

このFの最小化問題の解として、本書における**事後分布**の形が与えられる。

このベイズ推測の最適化問題＝「自由エネルギーFを下げたい気持ち」の中には
「尤度を上げたい気持ち」と「事前分布の形に近付けたい気持ち」が混ざっており、
逆温度 β はそのバランスを決めるパラメータになっている。
逆温度βが**無限大** **→** **尤度**最大化 **→** **最尤推定**の実現（δ関数）
逆温度βが**ゼロ** **→** **KLダイバージェンス**最小化 **→** **事前分布**の実現

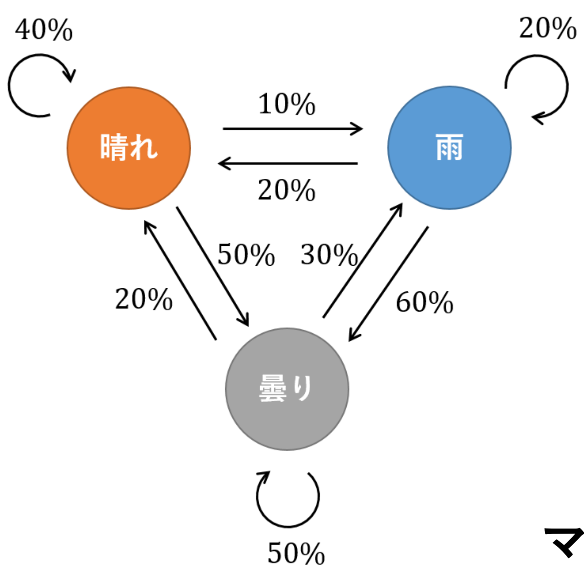
5.1 マルコフ連鎖モンテカルロ法
Markov chain Monte Carlo methods (MCMC)

- **マルコフ連鎖**＝次の状態が現在の状態だけで決まる確率過程
- **モンテカルロ法**＝乱数を使う手法
- **MCMC**＝マルコフ連鎖によって目的の分布に従う乱数を生成する手法
  - $p(\boldsymbol{w})$に従う乱数列$\{\boldsymbol{w}_k\}_{k=1}^{K}$が手に入れば、次の式によって期待値操作を行うことができる：

$$\int f(\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w} \approx \frac{1}{K}\sum_{k=1}^{K} f(\boldsymbol{w}_k)$$

0 日目	1 日目	2 日目	3 日目	4 日目	5 日目	…
晴れ	曇り	曇り	雨	曇り	雨	…

マルコフ連鎖による
乱数列の出力例
（晴れ＝1, 曇り＝2, 雨＝3）



今日の天気

$$\begin{array}{cc} & \begin{array}{ccc} \text{晴れ} & \text{曇り} & \text{雨} \end{array} \\ \text{明日の天気} \begin{array}{c} \text{晴れ} \\ \text{曇り} \\ \text{雨} \end{array} & \begin{pmatrix} 0.4 & 0.2 & 0.2 \\ 0.5 & 0.5 & 0.6 \\ 0.1 & 0.3 & 0.2 \end{pmatrix} \end{array}$$

推移確率行列

マルコフ連鎖は推移確率行列
によって特徴づけられる
（一対一に対応する）

マルコフ連鎖の例（気象連鎖）

マルコフ連鎖に目的の分布を生成させるには？

1. 確率分布が**定常分布**に収束する必要がある
   - 定常分布 $\boldsymbol{\pi}$ は推移確率行列（積分核） Pの固有ベクトル（関数）
     - つまり固有方程式 $P\boldsymbol{\pi} = \boldsymbol{\pi}$ を満たす
   - **エルゴード性**：ユニークな定常分布に収束する条件
     1. **既約**：すべての状態を行き来できる
     2. **正再帰**：必ず元の状態に戻れる
        - 状態が有限で既約の場合は必ず正再帰になる
     3. **非周期**：同じ状態に留まれる
        - これは不正確な表現で、実際はもっと複雑で非直観的な定義であるが省略…
2. 定常分布が**目的の分布**と一致する必要がある
   - 上記の固有方程式＝大域的釣り合い条件……必要十分条件
     - 与えられた $\boldsymbol{\pi}$ に対して P をどう設計する？
   - MCMCでは下記の**詳細釣り合い条件**を用いる……十分条件

$$p(\boldsymbol{w}_b|\boldsymbol{w}_a)p(\boldsymbol{w}_a) = p(\boldsymbol{w}_a|\boldsymbol{w}_b)p(\boldsymbol{w}_b)$$

$$\Rightarrow \int d\boldsymbol{w}_a p(\boldsymbol{w}_b|\boldsymbol{w}_a)p(\boldsymbol{w}_a) = \int d\boldsymbol{w}_a p(\boldsymbol{w}_a|\boldsymbol{w}_b)p(\boldsymbol{w}_b)$$

$$\Rightarrow \int d\boldsymbol{w}_a p(\boldsymbol{w}_b|\boldsymbol{w}_a)p(\boldsymbol{w}_a) = p(\boldsymbol{w}_b)$$

$$(\Rightarrow P\boldsymbol{\pi} = \boldsymbol{\pi})$$

5.1.1 メトロポリス法

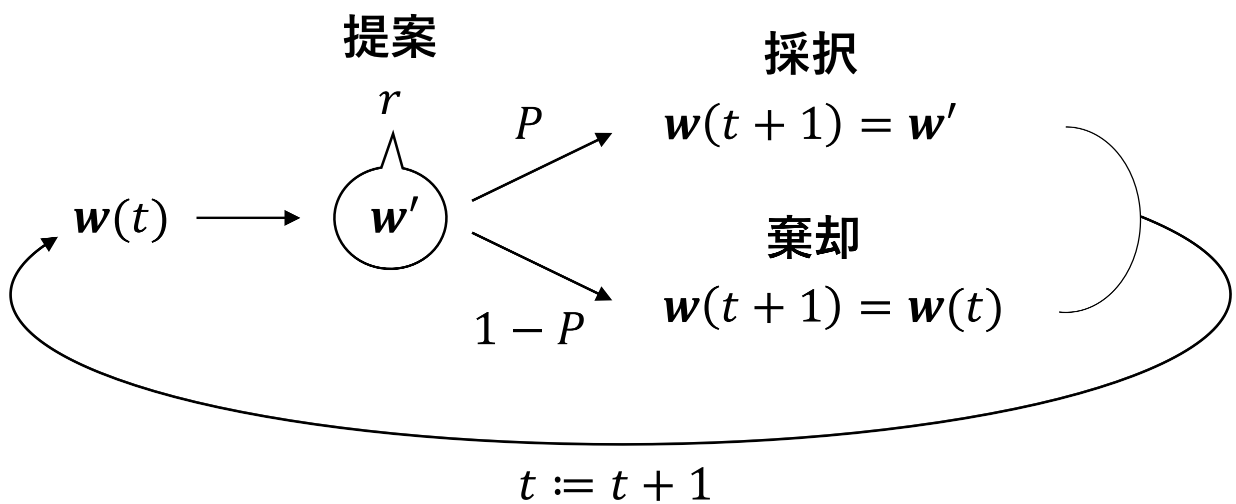
集合$\{\boldsymbol{w}(t) \in \mathbb{R}^d; t = 1,2,3 \ldots\}$を次の手続きに従って得る。

1) 初期値$\boldsymbol{w}(1)$を決めて$t = 1$とする。

2) 得られている$\boldsymbol{w}(t)$から$\boldsymbol{w}'$を**提案分布** $r(\boldsymbol{w}'|\boldsymbol{w}(t))$に従って生成する。ここで$r(\boldsymbol{w}'|\boldsymbol{w}(t))$は、対称性「任意の$\boldsymbol{w}_1, \boldsymbol{w}_2$について$r(\boldsymbol{w}_1|\boldsymbol{w}_2) = r(\boldsymbol{w}_2|\boldsymbol{w}_1)$」を満たすものであればよい。

3) $\Delta H \equiv H(\boldsymbol{w}') - H\big(\boldsymbol{w}(t)\big)$とおいて**採択率** $P = \min\{1, \exp(-\beta \Delta H)\}$ で$\boldsymbol{w}(t+1) = \boldsymbol{w}'$ とし、確率 $1 - P$ で $\boldsymbol{w}(t+1) = \boldsymbol{w}(t)$ とする。

4) $t \coloneqq t + 1$ とおいて2)に戻る。これを繰り返す。



定理18 メトロポリス法は詳細釣り合いを満たす

（証明1/2）

カノニカル分布において分配関数は共通なので、ボルツマン因子だけに注目した詳細釣り合い条件

$$p(\boldsymbol{w}_a|\boldsymbol{w}_b)\exp(-\beta H(\boldsymbol{w}_b)) = p(\boldsymbol{w}_b|\boldsymbol{w}_a)\exp(-\beta H(\boldsymbol{w}_a))$$

を示せば良い。

まず、1ステップの推移確率$p(\boldsymbol{w}_a|\boldsymbol{w}_b)$を陽に書くことを目指す。

「$\boldsymbol{w}(t)$ から$\boldsymbol{w}'$ が提案され、かつ採択される確率」は

$$r(\boldsymbol{w}'|\boldsymbol{w}(t))P(\Delta H(\boldsymbol{w}', \boldsymbol{w}(t)))$$

なので、これを$\boldsymbol{w}'$ について積分すれば、「提案される状態が何であれ**採択**される確率」Qが得られる：

$$Q\big(\boldsymbol{w}(t)\big) = \int r(\boldsymbol{w}'|\boldsymbol{w}(t))P(\Delta H(\boldsymbol{w}', \boldsymbol{w}(t)))d\boldsymbol{w}'$$

逆に、「提案される状態が何であれ**棄却**される確率」は$1-Q$。

よって、1ステップの推移確率は「**採択**によって**動く**場合」と「**棄却**によって**動かない**場合」の足し合わせとして、

$$p(\boldsymbol{w}_a|\boldsymbol{w}_b) = r(\boldsymbol{w}_a|\boldsymbol{w}_b)P(\Delta H(\boldsymbol{w}_a, \boldsymbol{w}_b)) + \delta(\boldsymbol{w}_a - \boldsymbol{w}_b)\big(1 - Q(\boldsymbol{w}_b)\big)$$

と書ける。

定理18 メトロポリス法は詳細釣り合いを満たす

（証明2/2）

推移確率とボルツマン因子の掛け算（詳細釣り合いの左辺）は、

$$
\begin{aligned}
&p(\boldsymbol{w}_a|\boldsymbol{w}_b)\exp(-\beta H(\boldsymbol{w}_b)) \\
&= r(\boldsymbol{w}_a|\boldsymbol{w}_b)P\big(\Delta H(\boldsymbol{w}_a,\boldsymbol{w}_b)\big)\exp\big(-\beta H(\boldsymbol{w}_b)\big) \\
&\quad +\delta(\boldsymbol{w}_a-\boldsymbol{w}_b)\big(1-Q(\boldsymbol{w}_b)\big)\exp(-\beta H(\boldsymbol{w}_b))
\end{aligned}
$$

と書ける。ここで、３つの対称性

$$
\begin{gathered}
r(\boldsymbol{w}_a|\boldsymbol{w}_b) = r(\boldsymbol{w}_b|\boldsymbol{w}_a) \\
P\big(\Delta H(\boldsymbol{w}_a,\boldsymbol{w}_b)\big)\exp\big(-\beta H(\boldsymbol{w}_b)\big) = P\big(\Delta H(\boldsymbol{w}_b,\boldsymbol{w}_a)\big)\exp\big(-\beta H(\boldsymbol{w}_a)\big) \\
\delta(\boldsymbol{w}_a-\boldsymbol{w}_b)f(\boldsymbol{w}_b) = \delta(\boldsymbol{w}_b-\boldsymbol{w}_a)f(\boldsymbol{w}_a)
\end{gathered}
$$

を使えば、

$$
p(\boldsymbol{w}_a|\boldsymbol{w}_b)\exp(-\beta H(\boldsymbol{w}_b)) = p(\boldsymbol{w}_b|\boldsymbol{w}_a)\exp(-\beta H(\boldsymbol{w}_a))
$$

が示される。（証明終了）

ボルツマン因子$\exp\big(-\beta H(\boldsymbol{w}_b)\big) > 0$をminの中に入れて$\exp\big(-\beta H(\boldsymbol{w}_a)\big)$をminの外に出す

$\{(\boldsymbol{w}_a,\boldsymbol{w}_b);\boldsymbol{w}_a=\boldsymbol{w}_b\}\subset W\times W$という台を持つ関数として両辺は同じものを指す

注意52 メトロポリス・ヘイスティングス法

提案分布rとして対称性を持たないものを用意した場合でも、メトロポリス方の手続き3)を次のものに置き換えることで、釣り合い条件を満たすことができる。

3’) 確率

$$P = \min\left\{1, \frac{r(\boldsymbol{w}(t)|\boldsymbol{w}')\exp(-\beta H(\boldsymbol{w}'))}{r(\boldsymbol{w}'|\boldsymbol{w}(t))\exp(-\beta H(\boldsymbol{w}(t)))}\right\}$$

で$\boldsymbol{w}(t+1) = \boldsymbol{w}'$ を選ぶ

（証明）

１ページ前の証明における対称性のうち１番目と２番目を

$$r(\boldsymbol{w}_a|\boldsymbol{w}_b)P\big(\Delta H(\boldsymbol{w}_a, \boldsymbol{w}_b)\big)\exp\big(-\beta H(\boldsymbol{w}_b)\big)$$
$$= r(\boldsymbol{w}_b|\boldsymbol{w}_a)P\big(\Delta H(\boldsymbol{w}_b, \boldsymbol{w}_a)\big)\exp\big(-\beta H(\boldsymbol{w}_a)\big)$$

で置き換えるだけ。

例20 ハイブリッド・モンテカルロ法

　メトロポリス法はステップ間の相関が強く、また棄却も起きうるためサンプリング効率が悪い。そこで古典力学に基づいた$\boldsymbol{w}$の「運動」を考えることで、「１ステップで大きく動く」かつ「ほとんど棄却が起きない」提案分布を作りたい。

　まず、$\boldsymbol{w} \in \mathbb{R}^d$に付随する「運動量（速度）」$\boldsymbol{p} \in \mathbb{R}^d$を用意し、ハミルトニアンを次のように修正する：

$$\mathcal{H}(\boldsymbol{w}, \boldsymbol{p}) = \frac{1}{2}\|\boldsymbol{p}\|^2 + \beta H(\boldsymbol{w})$$

　つまり、これまでのハミルトニアン$H(\boldsymbol{w})$はポテンシャルエネルギーとして振る舞う。この場合のカノニカル分布

$$\begin{aligned} p(\boldsymbol{w}, \boldsymbol{p}) &= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\bigl(-\mathcal{H}(\boldsymbol{w}, \boldsymbol{p})\bigr) \\ &= Z_{\boldsymbol{w}}^{-1} \exp\bigl(-\beta H(\boldsymbol{w})\bigr) \cdot Z_{\boldsymbol{p}}^{-1} \exp\left(-\frac{1}{2}\|\boldsymbol{p}\|^2\right) \\ &= p(\boldsymbol{w}) \cdot p(\boldsymbol{p}) \end{aligned}$$

から$(\boldsymbol{w}, \boldsymbol{p})$の組をサンプリングできれば、$\boldsymbol{w}$と$\boldsymbol{p}$は上式のように独立であるため、$\boldsymbol{w}$を$\exp\bigl(-\beta H(\boldsymbol{w})\bigr)$からサンプリングしたのと同じことになる。ここで、各分布$p(\boldsymbol{w}), p(\boldsymbol{p}), p(\boldsymbol{w}, \boldsymbol{p})$の分配関数（規格化定数）を$Z_{\boldsymbol{w}}, Z_{\boldsymbol{p}}, Z_{(\boldsymbol{w},\boldsymbol{p})} = Z_{\boldsymbol{w}} Z_{\boldsymbol{p}}$と置いた。

　具体的なサンプリング法を次ページに記す。

例20 ハイブリッド・モンテカルロ法

集合$\{\boldsymbol{w}(t) \in \mathbb{R}^d; t = 1,2,3 \dots\}$を次の手続きに従って得る。

1) 初期値$\boldsymbol{w}(1)$を決めて$t = 1$とする。

2) 正規分布$p(\boldsymbol{p}) = Z_{\boldsymbol{p}}^{-1} \exp\left(-\frac{1}{2}\|\boldsymbol{p}\|^2\right)$に従う$d$次元のベクトル$\boldsymbol{p}$を構成する。

3) $(\boldsymbol{w}(t), \boldsymbol{p})$を初期値とする次の微分方程式を数値的に解いて、$T$ 時刻後の$(\boldsymbol{w}', \boldsymbol{p}')$ を得る。ここでtとTは関係のない値である。

$$\frac{d\boldsymbol{w}}{d\tau} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}} = \boldsymbol{p},\ \frac{d\boldsymbol{p}}{d\tau} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{w}} = -\beta \boldsymbol{\nabla} H(\boldsymbol{w}) \qquad (0 \leq \tau \leq T).$$

この手続き$g: (\boldsymbol{w}, \boldsymbol{p}) \mapsto (\boldsymbol{w}', \boldsymbol{p}')$は誤差を含んでいてもよいが、

・**時間反転対称性：** $g^{-1}(\boldsymbol{w}', \boldsymbol{p}') = (\boldsymbol{w}, \boldsymbol{p}) = S \cdot g \cdot S(\boldsymbol{w}', \boldsymbol{p}')$

（ここで$S: (\boldsymbol{w}, \boldsymbol{p}) \mapsto (\boldsymbol{w}, -\boldsymbol{p})$は運動量の符号反転操作）

・**位相空間の体積保存性**（シンプレクティック性）：$\left|\frac{\partial(\boldsymbol{w}', \boldsymbol{p}')}{\partial(\boldsymbol{w}, \boldsymbol{p})}\right| = 1$

が満たされている必要がある。

4) $\Delta\mathcal{H} \equiv \mathcal{H}(\boldsymbol{w}', \boldsymbol{p}') - H(\boldsymbol{w}(t), \boldsymbol{p})$を求めて採択率 $P = \min\{1, \exp(-\Delta\mathcal{H})\}$ で$\boldsymbol{w}(t+1) = \boldsymbol{w}'$ とし、確率 $1-P$ で $\boldsymbol{w}(t+1) = \boldsymbol{w}(t)$ とする。

5) $t \coloneqq t+1$ とおいて2)に戻る。これを繰り返す。

（予備知識）変数変換には２通りの方法がある

写像 $\boldsymbol{f}: \mathbb{R}^m \supset X \ni \boldsymbol{x} \mapsto \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{y} \in Y \subset \mathbb{R}^n$ と、 Y上の試験関数 $g: \mathbb{R}^n \to \mathbb{R}$を考える。

（**積分変数**の変数変換の場合）

・ $m = n$で、写像$\boldsymbol{f}$が全単射で、ヤコビ行列が正則であるならば、

$$\int_Y g(\boldsymbol{y}) d\boldsymbol{y} = \int_X g(\boldsymbol{f}(\boldsymbol{x})) \left| \frac{d\boldsymbol{f}}{d\boldsymbol{x}} \right| d\boldsymbol{x}$$

・関数の名前$\boldsymbol{f}$と変数の名前$\boldsymbol{y}$を同一にすれば、微小体積について形式的に次のように書ける：

$$d\boldsymbol{y} = \left| \frac{d\boldsymbol{y}}{d\boldsymbol{x}} \right| d\boldsymbol{x}$$

（**確率変数**の変数変換の場合）

・確率変数$\boldsymbol{y}$が$\boldsymbol{f}$を通して確率変数$\boldsymbol{x}$に従属するとき、確率密度関数$p(\boldsymbol{x}), p(\boldsymbol{y})$の関係は次の式で与えられる：

$$p(\boldsymbol{y}) = \int_X p(\boldsymbol{x}) \delta(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}))\, d\boldsymbol{x}$$

・ $m = n$でも、写像$\boldsymbol{f}$が全単射でも、微分可能でもある必要はなく、「無意識な統計学者の法則」が成り立つ：

$$\int_Y g(\boldsymbol{y}) p(\boldsymbol{y}) d\boldsymbol{y} = \int_Y \int_X g(\boldsymbol{y}) p(\boldsymbol{x}) \delta(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}))\, d\boldsymbol{x}\, d\boldsymbol{y} = \int_X g(\boldsymbol{f}(\boldsymbol{x}))\, p(\boldsymbol{x}) d\boldsymbol{x}$$

・微小体積について形式的に次のように書ける：

$$p(\boldsymbol{y}) d\boldsymbol{y} = p(\boldsymbol{x}) d\boldsymbol{x}$$

・もし$m = n$で、写像$\boldsymbol{f}$が全単射で、ヤコビ行列が正則なら、次のようにも書ける：

$$p(\boldsymbol{y}) \left| \frac{d\boldsymbol{y}}{d\boldsymbol{x}} \right| = p(\boldsymbol{x})$$

HMC法は詳細釣り合い条件を満たす

（証明1/2）

$$p(\boldsymbol{w}'|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}d\boldsymbol{w}' = p(\boldsymbol{w}|\boldsymbol{w}')p(\boldsymbol{w}')d\boldsymbol{w}'d\boldsymbol{w}$$

を示す（変数変換を介するため、明示的に微小体積を考えたい）。

まず左辺において$p(\boldsymbol{w})$を陽に書くと、

$$p(\boldsymbol{w}'|\boldsymbol{w})p(\boldsymbol{w})d\boldsymbol{w}d\boldsymbol{w}' = Z_{\boldsymbol{w}}^{-1}p(\boldsymbol{w}'|\boldsymbol{w})\exp\left(-\beta H(\boldsymbol{w})\right)d\boldsymbol{w}d\boldsymbol{w}'$$

ここで、

$$p(\boldsymbol{w}'|\boldsymbol{w})d\boldsymbol{w}' = Z_{\boldsymbol{p}}^{-1}\exp\left(-\frac{1}{2}\|\boldsymbol{p}\|^2\right)P\left((\boldsymbol{w},\boldsymbol{p}) \to g(\boldsymbol{w},\boldsymbol{p})\right)d\boldsymbol{p} + \delta(\boldsymbol{w}-\boldsymbol{w}')\left(1-Q(\boldsymbol{w}')\right)d\boldsymbol{w}'$$

を代入して（対称性が自明な第二項は以後省略）、

$$= Z_{\boldsymbol{w}}^{-1}Z_{\boldsymbol{p}}^{-1}\exp\left(-\frac{1}{2}\|\boldsymbol{p}\|^2\right)P\left((\boldsymbol{w},\boldsymbol{p}) \to g(\boldsymbol{w},\boldsymbol{p})\right)\exp\left(-\beta H(\boldsymbol{w})\right)d\boldsymbol{w}d\boldsymbol{p} + \cdots$$

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1}\exp\left(-\mathcal{H}(\boldsymbol{w},\boldsymbol{p})\right)P\left((\boldsymbol{w},\boldsymbol{p}) \to g(\boldsymbol{w},\boldsymbol{p})\right)d\boldsymbol{w}d\boldsymbol{p} + \cdots$$

ここで、

$$\exp\left(-\mathcal{H}(\boldsymbol{w},\boldsymbol{p})\right)P\left((\boldsymbol{w},\boldsymbol{p}) \to g(\boldsymbol{w},\boldsymbol{p})\right) = \exp\left(-\mathcal{H}(\boldsymbol{w},\boldsymbol{p})\right)\min\left\{1,\exp\left(-\left(\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)-\mathcal{H}(\boldsymbol{w},\boldsymbol{p})\right)\right)\right\}$$

$$= \min\left\{\exp\left(-\mathcal{H}(\boldsymbol{w},\boldsymbol{p})\right),\exp\left(-\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)\right)\right\}$$

$$= \exp\left(-\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)\right)\min\left\{\exp\left(-\left(\mathcal{H}(\boldsymbol{w},\boldsymbol{p})-\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)\right)\right),1\right\}$$

$$= \exp\left(-\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)\right)P\left(g(\boldsymbol{w},\boldsymbol{p}) \to (\boldsymbol{w},\boldsymbol{p})\right)$$

より、

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1}\exp\left(-\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)\right)P\left(g(\boldsymbol{w},\boldsymbol{p}) \to (\boldsymbol{w},\boldsymbol{p})\right)d\boldsymbol{w}d\boldsymbol{p} + \cdots$$

HMC法は詳細釣り合い条件を満たす

（証明2/2）

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\left(-\mathcal{H}\left(g(\boldsymbol{w},\boldsymbol{p})\right)\right) P\big(g(\boldsymbol{w},\boldsymbol{p}) \to (\boldsymbol{w},\boldsymbol{p})\big) d\boldsymbol{w} d\boldsymbol{p} + \cdots$$

$(\boldsymbol{w},\boldsymbol{p}) \to (\boldsymbol{w}',\boldsymbol{p}') \coloneqq g(\boldsymbol{w},\boldsymbol{p})$と変数変換すると、

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\big(-\mathcal{H}(\boldsymbol{w}',\boldsymbol{p}')\big) P\big((\boldsymbol{w}',\boldsymbol{p}') \to g^{-1}(\boldsymbol{w}',\boldsymbol{p}')\big) \left|\frac{\partial(\boldsymbol{w},\boldsymbol{p})}{\partial(\boldsymbol{w}',\boldsymbol{p}')}\right| d\boldsymbol{w}' d\boldsymbol{p}' + \cdots$$

gの時間反転対称性とシンプレクティック性から、

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\big(-\mathcal{H}(\boldsymbol{w}',\boldsymbol{p}')\big) P\big((\boldsymbol{w}',\boldsymbol{p}') \to S \cdot g \cdot S(\boldsymbol{w}',\boldsymbol{p}')\big) d\boldsymbol{w}' d\boldsymbol{p}' + \cdots$$

運動エネルギーにおける$\boldsymbol{p}$の符号反転対称性から、

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\big(-\mathcal{H}(\boldsymbol{w}',-\boldsymbol{p}')\big) P\big((\boldsymbol{w}',-\boldsymbol{p}') \to g(\boldsymbol{w}',-\boldsymbol{p}')\big) d\boldsymbol{w}' d\boldsymbol{p}' + \cdots$$

$\boldsymbol{p}' \to \boldsymbol{p}'' \coloneqq -\boldsymbol{p}'$と変数変換して、

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\big(-\mathcal{H}(\boldsymbol{w}',\boldsymbol{p}'')\big) P\big((\boldsymbol{w}',\boldsymbol{p}'') \to g(\boldsymbol{w}',\boldsymbol{p}'')\big) \left|\frac{\partial \boldsymbol{p}'}{\partial \boldsymbol{p}''}\right| d\boldsymbol{w}' d\boldsymbol{p}'' + \cdots$$

$$= Z_{(\boldsymbol{w},\boldsymbol{p})}^{-1} \exp\left(-\frac{1}{2}\|\boldsymbol{p}''\|^2\right) P\big((\boldsymbol{w}',\boldsymbol{p}'') \to g(\boldsymbol{w}',\boldsymbol{p}'')\big) \exp\big(-\beta H(\boldsymbol{w}')\big) d\boldsymbol{w}' d\boldsymbol{p}'' + \cdots$$

$$= p(\boldsymbol{w}|\boldsymbol{w}')p(\boldsymbol{w}')d\boldsymbol{w}'d\boldsymbol{w}$$

となり、ハイブリッド・モンテカルロ法が詳細釣り合い条件を満たすことが確かめられた。（証明終了）

シンプレクティック数値積分法$\boldsymbol{g}$の例：リープ・フロッグ法（速度ベルレ法）

$$\frac{d\boldsymbol{w}}{d\tau} = \boldsymbol{p}, \frac{d\boldsymbol{p}}{d\tau} = \boldsymbol{f}(\boldsymbol{w})$$

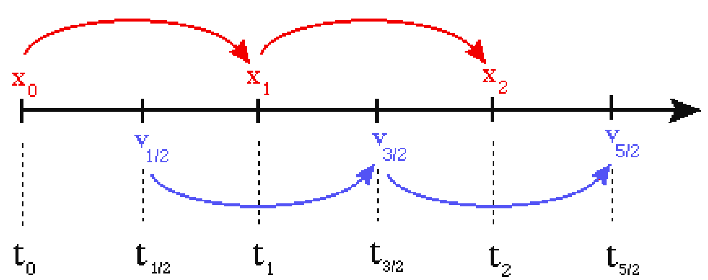
に対する繰り返しを次のようにする。

$$\boldsymbol{p}\left(n + \frac{1}{2}\right) = \boldsymbol{p}(n) + \frac{\epsilon}{2}\boldsymbol{f}\big(\boldsymbol{w}(n)\big),$$

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \epsilon\boldsymbol{p}\left(n + \frac{1}{2}\right),$$

$$\boldsymbol{p}(n+1) = \boldsymbol{p}\left(n + \frac{1}{2}\right) + \frac{\epsilon}{2}\boldsymbol{f}\big(\boldsymbol{w}(n+1)\big).$$

ここで$\epsilon > 0$は小さい定数である。



https://www.physics.drexel.edu/~valliere/PHYS305/Diff_Eq_Integrators/leapfrog/

各ステップを線形化すると：

$$\begin{pmatrix} \boldsymbol{p}_{n+\frac{1}{2}} \\ \boldsymbol{w}_n \end{pmatrix} = \begin{pmatrix} I & \dfrac{\epsilon}{2}\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}} \\ O & I \end{pmatrix} \begin{pmatrix} \boldsymbol{p}_n \\ \boldsymbol{w}_n \end{pmatrix}$$

$$\begin{pmatrix} \boldsymbol{p}_{n+\frac{1}{2}} \\ \boldsymbol{w}_{n+1} \end{pmatrix} = \begin{pmatrix} I & O \\ \epsilon I & I \end{pmatrix} \begin{pmatrix} \boldsymbol{p}_{n+\frac{1}{2}} \\ \boldsymbol{w}_n \end{pmatrix}$$

$$\begin{pmatrix} \boldsymbol{p}_{n+1} \\ \boldsymbol{w}_{n+1} \end{pmatrix} = \begin{pmatrix} I & \dfrac{\epsilon}{2}\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{w}} \\ O & I \end{pmatrix} \begin{pmatrix} \boldsymbol{p}_{n+\frac{1}{2}} \\ \boldsymbol{w}_{n+1} \end{pmatrix}$$

となり、各ステップにおいてヤコビアンが1に等しいので、シンプレクティック性を持つ。オイラー法やRK法のようにすべての変数を同時更新するとこうはならない。

※ Aを正則行列、Dを正方行列としたとき、ブロック行列 $T = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ に対して、

$$|T| = |A||D - CA^{-1}B|$$

また、手続きを逆に辿って$(\boldsymbol{p}_{n+1}, \boldsymbol{w}_{n+1})$から$(\boldsymbol{p}_n, \boldsymbol{w}_n)$を得ようとすると、順手続きにおいて$\boldsymbol{p}$ の符号を反転させたときと同じ形になるので、時間反転対称性を持つ。

5.1.2 ギブス・サンプリング

パラメータ$\boldsymbol{w} \in \mathbb{R}^d$を二つの変数に分割して$\boldsymbol{w} = (\boldsymbol{w}_1, \boldsymbol{w}_2)$とする。目的とする確率分布を$p(\boldsymbol{w}_1, \boldsymbol{w}_2)$として、この確率分布から定義される条件付き確率を$p(\boldsymbol{w}_1 | \boldsymbol{w}_2)$および$p(\boldsymbol{w}_2 | \boldsymbol{w}_1)$とする。

集合$\{\boldsymbol{w}(t) = (\boldsymbol{w}_1(t), \boldsymbol{w}_2(t)) \in \mathbb{R}^d; t = 1,2,3 \ldots\}$を次の手続きに従って得る。

1) 初期値$\boldsymbol{w}(1)$を決めて$t = 1$とする。

2) 確率1/2で$p(\boldsymbol{w}_1' | \boldsymbol{w}_2') p(\boldsymbol{w}_2' | \boldsymbol{w}_1(t))$に従って$(\boldsymbol{w}_1', \boldsymbol{w}_2')$を選出し、確率1/2で$p(\boldsymbol{w}_2' | \boldsymbol{w}_1') p(\boldsymbol{w}_1' | \boldsymbol{w}_2(t))$に従って$(\boldsymbol{w}_1', \boldsymbol{w}_2')$を選出する。

3) $\boldsymbol{w}(t+1) = (\boldsymbol{w}_1', \boldsymbol{w}_2')$とおき、$t \coloneqq t + 1$ とおいて2)に戻る。

$$
\begin{tikzcd}
 & \begin{array}{c}\boldsymbol{w}_2'\text{生成}\\ \sim p(\boldsymbol{w}_2' | \boldsymbol{w}_1)\end{array} \arrow[r] & \begin{array}{c}\boldsymbol{w}_1'\text{生成}\\ \sim p(\boldsymbol{w}_1' | \boldsymbol{w}_2')\end{array} \arrow[dr] & \\
\boldsymbol{w}(t) = (\boldsymbol{w}_1(t), \boldsymbol{w}_2(t)) \arrow[ur, "1/2"] \arrow[dr, "1/2"'] & & & \boldsymbol{w}(t+1) = (\boldsymbol{w}_1', \boldsymbol{w}_2') \arrow[lll, bend left=40, "t \coloneqq t + 1"] \\
 & \begin{array}{c}\boldsymbol{w}_1'\text{生成}\\ \sim p(\boldsymbol{w}_1' | \boldsymbol{w}_2)\end{array} \arrow[r] & \begin{array}{c}\boldsymbol{w}_2'\text{生成}\\ \sim p(\boldsymbol{w}_2' | \boldsymbol{w}_1')\end{array} \arrow[ur] &
\end{tikzcd}
$$

補題26 ギブス・サンプリングにおける詳細釣り合い

（証明）

$$p(\boldsymbol{w}_1',\boldsymbol{w}_2'|\boldsymbol{w}_1,\boldsymbol{w}_2)p(\boldsymbol{w}_1,\boldsymbol{w}_2)=p(\boldsymbol{w}_1,\boldsymbol{w}_2|\boldsymbol{w}_1',\boldsymbol{w}_2')p(\boldsymbol{w}_1',\boldsymbol{w}_2')$$

を示す。まず、推移確率を２通りについて分けて書くと、

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

この第１項のみについて詳細釣り合いの左辺を開くと、

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

同様にして、第２項については、

$$p(\boldsymbol{w}_1'|\boldsymbol{w}_2')p(\boldsymbol{w}_2'|\boldsymbol{w}_1)p(\boldsymbol{w}_1,\boldsymbol{w}_2)=p(\boldsymbol{w}_2|\boldsymbol{w}_1)p(\boldsymbol{w}_1|\boldsymbol{w}_2')p(\boldsymbol{w}_1',\boldsymbol{w}_2')$$

まとめると、

$$\begin{aligned}&p(\boldsymbol{w}_1',\boldsymbol{w}_2'|\boldsymbol{w}_1,\boldsymbol{w}_2)p(\boldsymbol{w}_1,\boldsymbol{w}_2)\\&=\frac{1}{2}p(\boldsymbol{w}_1|\boldsymbol{w}_2)p(\boldsymbol{w}_2|\boldsymbol{w}_1')p(\boldsymbol{w}_1',\boldsymbol{w}_2')+\frac{1}{2}p(\boldsymbol{w}_2|\boldsymbol{w}_1)p(\boldsymbol{w}_1|\boldsymbol{w}_2')p(\boldsymbol{w}_1',\boldsymbol{w}_2')\\&=p(\boldsymbol{w}_1,\boldsymbol{w}_2|\boldsymbol{w}_1',\boldsymbol{w}_2')p(\boldsymbol{w}_1',\boldsymbol{w}_2')\end{aligned}$$

（証明終了）

注意55 ギブス・サンプリング

１ステップで$\boldsymbol{w_1}$か$\boldsymbol{w_2}$の一方のみを確率1/2で動かす場合もギブス・サンプリングと呼ぶ。この場合の推移確率は

$$p(\boldsymbol{w}_1', \boldsymbol{w}_2' | \boldsymbol{w}_1, \boldsymbol{w}_2) = \frac{1}{2}\{p(\boldsymbol{w}_1' | \boldsymbol{w}_2)\delta(\boldsymbol{w}_2' - \boldsymbol{w}_2) + p(\boldsymbol{w}_2' | \boldsymbol{w}_1)\delta(\boldsymbol{w}_1' - \boldsymbol{w}_1)\}$$

となるが、やはり詳細釣り合いを満たす。

（証明）

前ページと同様に、詳細釣り合いの左辺を開いて第１項を書くと、

$$p(\boldsymbol{w}_1' | \boldsymbol{w}_2)\delta(\boldsymbol{w}_2' - \boldsymbol{w}_2)p(\boldsymbol{w}_1, \boldsymbol{w}_2) = \frac{p(\boldsymbol{w}_1', \boldsymbol{w}_2)}{p(\boldsymbol{w}_2)}\delta(\boldsymbol{w}_2' - \boldsymbol{w}_2)p(\boldsymbol{w}_1, \boldsymbol{w}_2)$$
$$= p(\boldsymbol{w}_1', \boldsymbol{w}_2)\delta(\boldsymbol{w}_2' - \boldsymbol{w}_2)p(\boldsymbol{w}_1 | \boldsymbol{w}_2)$$

デルタ関数の性質から、

$$= p(\boldsymbol{w}_1', \boldsymbol{w}_2')\delta(\boldsymbol{w}_2 - \boldsymbol{w}_2')p(\boldsymbol{w}_1 | \boldsymbol{w}_2')$$
$$= p(\boldsymbol{w}_1 | \boldsymbol{w}_2')\delta(\boldsymbol{w}_2 - \boldsymbol{w}_2')p(\boldsymbol{w}_1', \boldsymbol{w}_2')$$

となる。第２項についても同様。項ごとに$\boldsymbol{w}$と$\boldsymbol{w}'$の入れ替えについて対称性があるため、詳細釣り合いが満たされることがわかる。また、確率は1/2 ずつでなくてもいい。

（証明終了）

章末問題【1】混合指数型分布におけるギブス・サンプリング

- 潜在変数$\boldsymbol{y}^n$を導入した形(5.18)で、これをパラメータとみなし、$\boldsymbol{y}^n$と$\boldsymbol{w}=(\boldsymbol{a},\boldsymbol{b})=\{(a_k,\boldsymbol{b}_k);k=1,2,\ldots,K\}$を分けてサンプリングするのが容易。
- $\boldsymbol{y}^n$が確定した下では$\boldsymbol{a}$と$\boldsymbol{b}$は独立な分布とみなせる（条件付き独立）。
- $\boldsymbol{y}^n$ は独立なn個のカテゴリカル分布からサンプリングできる：

$$p(\boldsymbol{y}^n|\boldsymbol{w},\boldsymbol{x}^n)=\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{A_{ik}}{\sum_l^K A_{il}}\right)^{y_{ik}}$$

$$\text{where } A_{ik}\coloneqq a_k v(\boldsymbol{x}_i)\exp(\boldsymbol{f}(\boldsymbol{b}_k)\cdot\boldsymbol{g}(\boldsymbol{x}_i))$$

- $\boldsymbol{a}$ は1個のディリクレ分布からサンプリングできる：

$$p(\boldsymbol{a}|\boldsymbol{x}^n,\boldsymbol{y}^n)\propto\prod_{k=1}^{K}(a_k)^{\sum_i^n y_{ik}+\phi_k-1}$$

- $\boldsymbol{b}$ は独立なK個の指数型分布からサンプリングできる：

$$p(\boldsymbol{b}|\boldsymbol{x}^n,\boldsymbol{y}^n)\propto\prod_{k=1}^{K}\exp\left(\boldsymbol{f}(\boldsymbol{b}_k)\cdot\left(\boldsymbol{\eta}_k+\sum_i^n y_{ik}\boldsymbol{g}(\boldsymbol{x}_i)\right)\right)$$

5.1.3 ランジュバン方程式を用いる方法

　ノイズ項のある最急降下法（勾配降下法）を微分方程式で表すと：

$$\frac{d\boldsymbol{w}}{dt} = -\beta\nabla H(\boldsymbol{w}) + \frac{d\boldsymbol{B}}{dt},\ \ \frac{d\boldsymbol{B}}{dt}\text{: ノイズ項}$$

これは物理では**過減衰ランジュバン方程式** (overdamped Langevin equation)と呼ばれ、粘性流体中の小さな質量の物体などの運動方程式に対応する。過減衰ランジュバン方程式を時間刻み幅 $\epsilon > 0$ を用いて離散化して、MCMCとして利用する方法を**ランジュバン・モンテカルロ法**（**LMC**）と呼ぶ。つまり：

　集合$\{\boldsymbol{w}(t) \in \mathbb{R}^d; t = 0,1,2 \ldots\}$を次の手続きに従って得る。

1) $\boldsymbol{w}(t)$の初期値を決める。$t = 0$とする。

2) $\boldsymbol{g}(\epsilon)$を平均0分散確率2ϵの正規分布$q(\boldsymbol{g})$に従う確率変数とする。次の更新を行う。毎回独立な$\boldsymbol{g}(\epsilon)$を用いて

$$\boldsymbol{w}(t+\epsilon) = \boldsymbol{w}(t) - \epsilon\beta\nabla H\big(\boldsymbol{w}(t)\big) + \boldsymbol{g}(\epsilon)$$

とする

3) $t \coloneqq t + \epsilon$ とおいて2)に戻る。

　なお、2)のランジュバン方程式の時間発展を必ず採択するのではなく、あくまでもメトロポリス・ヘイスティングス法の**提案分布**として2)を用いて、その後に棄却する可能性も考慮する場合は、**メトロポリス補正ランジュバン法**（**MALA**）と呼ぶ。MALAは理論的にはハイブリッド・モンテカルロ法を１ステップだけ進めることと同じである (Girolami and Calderhead 2011)。

補題27 ランジュバン方程式に従う変数の確率分布はフォッカー・プランク方程式を満たす

　前ページのアルゴリズムに従う確率変数（実現値$\boldsymbol{w}$）の確率分布を$\boldsymbol{p(w,t)}$とする。$\epsilon \to 0$とするとき、フォッカー・プランク方程式

$$\frac{\partial}{\partial t} p(\boldsymbol{w},t) - \beta \boldsymbol{\nabla} \cdot \Big(\big(\boldsymbol{\nabla} H(\boldsymbol{w}) \big) p(\boldsymbol{w},t) \Big) = \Delta p(\boldsymbol{w},t)$$

が成り立つ。ここで$\Delta \coloneqq \boldsymbol{\nabla}^2 = \boldsymbol{\nabla} \cdot \boldsymbol{\nabla}$はラプラシアン。

（証明1/2）

　手続き2)の更新式

$$\boldsymbol{w}_{t+\epsilon} = \boldsymbol{w}_t - \epsilon \beta \boldsymbol{\nabla} H(\boldsymbol{w}_t) + \boldsymbol{g}(\epsilon)$$

を確率分布で表すと、各ステップで生成される$\boldsymbol{g}$は$\boldsymbol{w_t}$と独立なので、

$$p(\boldsymbol{w}_{t+\epsilon}, t+\epsilon) d\boldsymbol{w}_{t+\epsilon} = p(\boldsymbol{w}_t, t) q(\boldsymbol{g}(\epsilon)) d\boldsymbol{w}_t d\boldsymbol{g}$$

と書ける。この更新後の確率分布をフーリエ変換$(\boldsymbol{w} \to \boldsymbol{k})$していく。変換後の周波数表示の関数の記号は、元の関数の記号にハットをつけたものと約束しておこう。たとえば$p(\boldsymbol{w},t)$のフーリエ変換は：

$$\hat{p}(\boldsymbol{k},t) = \mathcal{F}(p)(\boldsymbol{k},t) = \int p(\boldsymbol{w},t) \exp(i\boldsymbol{k} \cdot \boldsymbol{w}) d\boldsymbol{w}$$

補題27 ランジュバン方程式に従う変数の確率分布はフォッカー・プランク方程式を満たす

（**証明2/2**）

正規分布qのフーリエ変換が、分散の逆数を取った正規分布：

$$\hat{q}(\boldsymbol{k}) = \int \frac{1}{\sqrt{4\pi\epsilon}} \exp\left(-\frac{\|\boldsymbol{w}\|^2}{4\epsilon}\right) \exp(i\boldsymbol{k}\cdot\boldsymbol{w}) d\boldsymbol{w} = \exp(-\epsilon\|\boldsymbol{k}\|^2)$$

となることを踏まえれば、 $p(\boldsymbol{w}_{t+\epsilon}, t+\epsilon)$のフーリエ変換は

$$\hat{p}(\boldsymbol{k}, t+\epsilon) = \int p(\boldsymbol{w}_{t+\epsilon}, t+\epsilon) \exp(i\boldsymbol{k}\cdot\boldsymbol{w}_{t+\epsilon}) d\boldsymbol{w}_{t+\epsilon}$$

$$= \int p(\boldsymbol{w}_t, t) q(\boldsymbol{g}) \exp\big(i\boldsymbol{k}\cdot(\boldsymbol{w}_t - \epsilon\beta\boldsymbol{\nabla}H(\boldsymbol{w}_t) + \boldsymbol{g})\big) d\boldsymbol{w}_t d\boldsymbol{g}$$

$$= \exp(-\epsilon\|\boldsymbol{k}\|^2) \int p(\boldsymbol{w}_t, t) \exp\left(-i\epsilon\beta\boldsymbol{k}\cdot\boldsymbol{\nabla}H(\boldsymbol{w}_t)\right) \exp(i\boldsymbol{k}\cdot\boldsymbol{w}_t) d\boldsymbol{w}_t$$

フーリエ基底以外の指数関数を１次近似して、

$$= (1-\epsilon\|\boldsymbol{k}\|^2) \int p(\boldsymbol{w}_t, t)\big(1 - i\epsilon\beta\boldsymbol{k}\cdot\boldsymbol{\nabla}H(\boldsymbol{w}_t)\big) \exp(i\boldsymbol{k}\cdot\boldsymbol{w}_t) d\boldsymbol{w}_t + \mathcal{O}(\epsilon^2)$$

展開して出てくる微小量ϵ^2の項を後ろにまとめて、

$$= \hat{p}(\boldsymbol{k}, t) - \epsilon\left\{i\beta\boldsymbol{k}\cdot\mathcal{F}\left(\big(\boldsymbol{\nabla}H(\boldsymbol{w}_t)\big)p\right)(\boldsymbol{k}, t) + \|\boldsymbol{k}\|^2\hat{p}(\boldsymbol{k}, t)\right\} + \mathcal{O}(\epsilon^2)$$

$\mathcal{F}(\boldsymbol{\nabla}p) = -i\boldsymbol{k}\hat{p}$であることに注意して、両辺を逆フーリエ変換すれば、

$$p(\boldsymbol{w}, t+\epsilon) = p(\boldsymbol{w}, t) - \epsilon\left\{-\beta\boldsymbol{\nabla}\cdot\left(\big(\boldsymbol{\nabla}H(\boldsymbol{w})\big)p(\boldsymbol{w}, t)\right) - \Delta p(\boldsymbol{w}, t)\right\} + \mathcal{O}(\epsilon^2)$$

$$\frac{p(\boldsymbol{w}, t+\epsilon) - p(\boldsymbol{w}, t)}{\epsilon} = \beta\boldsymbol{\nabla}\cdot\left(\big(\boldsymbol{\nabla}H(\boldsymbol{w})\big)p(\boldsymbol{w}, t)\right) + \Delta p(\boldsymbol{w}, t) + \mathcal{O}(\epsilon^2)$$

ここで$\epsilon \to 0$とすれば、フォッカー・プランク方程式が得られる（証明終了）。

フォッカープランク方程式の定常解はカノニカル分布をなす

フォッカープランク方程式

$$\frac{\partial}{\partial t}p(\boldsymbol{w},t) - \beta\boldsymbol{\nabla}\cdot\Big(\big(\boldsymbol{\nabla}H(\boldsymbol{w})\big)p(\boldsymbol{w},t)\Big) = \Delta p(\boldsymbol{w},t)$$

が時間に依存しない解、つまり定常分布 $p(\boldsymbol{w},t) = \bar{p}\,(\boldsymbol{w})$ を持つとすると、その解はカノニカル分布をなす。

（証明）

定常分布 $\bar{p}$ が満たすべき方程式は、フォッカープランク方程式から時間微分$\partial_t p$を取り除いた

$$-\boldsymbol{\nabla}\cdot\Big(\big(\beta\boldsymbol{\nabla}H(\boldsymbol{w})\big)\bar{p}(\boldsymbol{w})\Big) = \Delta\bar{p}(\boldsymbol{w})$$

である。この一般解は、$\boldsymbol{C_1}$を定数ベクトルとして、

$$-\big(\beta\boldsymbol{\nabla}H(\boldsymbol{w})\big)\bar{p}(\boldsymbol{w}) = \boldsymbol{\nabla}\bar{p}(\boldsymbol{w}) + \boldsymbol{C_1}$$

である。ここで、パラメータが発散しない、つまり$\bar{p}$ が無限遠で0となる自然な制約

$$\lim_{\|\boldsymbol{w}\|\to\infty}\bar{p}(\boldsymbol{w}) = 0$$

を入れると、 $\boldsymbol{C}_1 = \boldsymbol{0}$となるので、

$$-\big(\beta\boldsymbol{\nabla}H(\boldsymbol{w})\big)\bar{p}(\boldsymbol{w}) = \boldsymbol{\nabla}\bar{p}(w)$$

$$-\beta\boldsymbol{\nabla}H(\boldsymbol{w}) = \boldsymbol{\nabla}\log\bar{p}(\boldsymbol{w})$$

この一般解は、C_2を定数スカラーとして、

$$-\beta H(\boldsymbol{w}) = \log\bar{p}(\boldsymbol{w}) + C_2$$

$$\therefore\ \bar{p}\,(\boldsymbol{w}) \propto \exp\big(-\beta H(\boldsymbol{w})\big)$$

（証明終了）

5.1.4 自由エネルギーの近似

　ここまでは事後分布の規格化定数である分配関数を知ることなしに事後分布から乱数を生成する方法を紹介してきたが、モデル選択などの枠組みにおいては$\beta = 1$のときの分配関数そのもの（ひいては自由エネルギー、モデルエビデンス）を知りたいことも多い。実は、**異なる複数の逆温度下におけるサンプリング**を実行することによって、$\beta = 1$のときの**分配関数**や**自由エネルギー**を求めることができる。

　まず、負の対数尤度（経験対数損失関数$L(\boldsymbol{w})$のn倍）を新たに、

$$\hat{H}(\boldsymbol{w}) = -\sum_{i=1}^{n} \log p(\boldsymbol{X}_i|\boldsymbol{w}) > 0$$

とおくと、逆温度βにおける分配関数、ならびに逆温度βにおける期待値はそれぞれ、

$$Z_n(\beta) = \int e^{-\beta\hat{H}(\boldsymbol{w})}\varphi(\boldsymbol{w})d\boldsymbol{w}$$

$$\mathbb{E}_{\boldsymbol{w}}^{(\beta)}[\,] = \frac{1}{Z_n(\beta)}\int (\,)e^{-\beta\hat{H}(\boldsymbol{w})}\varphi(\boldsymbol{w})d\boldsymbol{w}$$

と書ける。次に、逆温度の集合$\{\beta_k; k = 0,1,\dots,J\}$を

$$0 = \beta_0 < \beta_1 < \cdots < \beta_J = 1$$

となるように取れば、$Z_n(0) = 1$から、$\beta = 1$のときの分配関数$Z_n(1)$および自由エネルギー$F_n(1)$は：

$$Z_n(1) = \prod_{k=0}^{J-1}\frac{Z_n(\beta_{k+1})}{Z_n(\beta_k)} = \prod_{k=0}^{J-1}\mathbb{E}_{\boldsymbol{w}}^{(\beta_k)}\left[e^{-(\beta_{k+1}-\beta_k)\hat{H}(\boldsymbol{w})}\right]$$

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

注意56

(1) イェンセンの不等式より

$$F_n(1) = -\sum_{k=0}^{J-1} \log \mathbb{E}_{\boldsymbol{w}}^{(\beta_k)}\left[e^{-(\beta_{k+1}-\beta_k)\widehat{H}(\boldsymbol{w})}\right] \leq \sum_{k=0}^{J-1} \mathbb{E}_{\boldsymbol{w}}^{(\beta_k)}\left[(\beta_{k+1}-\beta_k)\widehat{H}(\boldsymbol{w})\right]$$

である。一方、$f(\beta) \coloneqq \beta F_n(\beta) = -\log Z_n(\beta)$とおくと、

$$F_n(1) = f(1) = \int_0^1 \frac{df}{d\beta}(\beta) d\beta = \int_0^1 \mathbb{E}_{\boldsymbol{w}}^{(\beta)}\left[\widehat{H}(\boldsymbol{w})\right] d\beta$$

となる。前ページの手続きはこの積分を離散化したものと考えることができる。

(2) 各温度におけるパラメータの集合をまとめて$W = \left(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{J-1}\right)$とする。全パラメータの同時分布

$$P(W) = \prod_{j=1}^{J-1} p^{(\beta_j)}(\boldsymbol{w}_j),\ p^{(\beta_j)}\left(\boldsymbol{w}_j\right) \coloneqq \frac{1}{Z(\beta_j)} \exp(-\beta_j \widehat{H} - \log \varphi(\boldsymbol{w}))$$

から効率的にサンプリングできる手法として、**レプリカ交換法**がある。これは各温度で通常のモンテカルロ法を行いつつ、一定間隔で次の確率に従って温度β_jと温度β_{j+1}の間で状態$\boldsymbol{w}_j$と$\boldsymbol{w}_{j+1}$を交換する。

$$\min\left\{1, \exp\left\{-\left(\beta_{j+1}-\beta_j\right)(\widehat{H}(\boldsymbol{w}_j) - \widehat{H}\left(\boldsymbol{w}_{j+1}\right)\right)\right\}\right\}$$

この交換は詳細釣り合いを満たし、$P(W)$が定常分布であることを保証する。レプリカ交換法は高温環境の存在が低温環境に対して焼きなまし法（SA）のように振る舞い、局所解に陥ることを回避できる。

(3) レプリカ交換法による逆温度$\beta_1, \beta_2 (\beta_2 > \beta_1)$の間で交換が行われる確率は

$$P(\beta_1, \beta_2) = 1 - \frac{1}{\sqrt{\pi}} \frac{\beta_2 - \beta_1}{\beta_1} \frac{\Gamma\left(\lambda + \frac{1}{2}\right)}{\Gamma(\lambda)}$$

となる (Nagata and Watanabe 2008)。よって、温度$\{\beta_j\}$を等比数列にすることで交換確率は一定になる。

注意57

$$p(\boldsymbol{w}) = \frac{1}{Z_n(\beta)} \varphi(\boldsymbol{w}) \exp\left(-\beta \widehat{H}(\boldsymbol{w})\right)$$

であることから、

$$\mathbb{E}_{\boldsymbol{w}}^{\beta}[\exp(\beta \widehat{H}(\boldsymbol{w}))] = \int p(\boldsymbol{w}) \exp(\beta \widehat{H}(\boldsymbol{w}))\, d\boldsymbol{w} = \frac{1}{Z_n(\beta)} \int \varphi(\boldsymbol{w}) d\boldsymbol{w} = \frac{1}{Z_n(\beta)}$$

となるはずである。しかし、この期待値は「確率的に出にくい点ほど寄与が大きい」という性質を持つため、MCMCによる正確な計算は困難である（つまり、ボルツマン因子$\exp\left(-\beta \widehat{H}(\boldsymbol{w})\right)$が小さい$\boldsymbol{w}$ほど、その逆数$\exp(\beta \widehat{H}(\boldsymbol{w}))$が大きい）。