本レジュメの巻末には、Appendix のページを付けている。必要に応じて参照する。

1　マルコフ連鎖モンテカルロ

前節では、関数の期待値を求めるために棄却サンプリング・重点サンプリングの手法を学んだ。これらの手法はサンプル $\boldsymbol{z}$ が高次元になると、計算負荷が指数的に上がってしまうという制約があった[*1]。そこで本節では、マルコフ連鎖モンテカルロ（以下 MCMC）と呼ばれる、上記の次元の問題を解決できて、かつ、より一般的な状況下で適用できる枠組みを取り扱っていく。

まず、サブセクション【マルコフ連鎖】においては、マルコフ連鎖とは何かということを理解する。そのあとで、遷移確率を導入し、求めたい分布からのサンプリングが可能であることを保証するのがポイントとなる。

実際のところ、遷移確率が求めたい分布から直接設計することは一般的にはできないことが多い（らしい）。そこで、棄却サンプリングや重点サンプリングと同様に、提案分布からのサンプリングをおこなうことを考える。

続いて、サブセクション【Metropolis-Hastings アルゴリズム】において、MCMC の基本的な手法である Metropolis-Hastings アルゴリズムを取り扱う。サンプル候補を提案分布から生成したとき、適切な規準[*2]に乗っ取ってサンプルを受理するかどうか考える。その規準の設定方法について、詳細に議論をおこなう[*3]。

1.1　マルコフ連鎖

まずマルコフ連鎖とは、何かについて紹介する。特に、本レジュメではマルコフ連鎖の中でも 1 次マルコフ連鎖と呼ばれるものを取り扱う。1 次マルコフ連鎖とは次のようなものである：確率変数の系列 $\boldsymbol{z}^{(m)}$ であって、$m \in \{1, \ldots, M-1\}$ であるとき、以下の条件付き独立性が成立する。

$$p(\boldsymbol{z}^{(m+1)} \mid \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}) = p(\boldsymbol{z}^{(m+1)} \mid \boldsymbol{z}^{(m)}) \tag{1}$$

これはグラフィカルモデルとしては、鎖の形をした有向グラフとして表すことができる[*4]。初期変数 $p(\boldsymbol{z}^{(0)})$ の確率分布と遷移確率 $T_m(\boldsymbol{z}^{(m)}, \boldsymbol{z}^{(m+1)}) \equiv p(\boldsymbol{z}^{(m+1)} \mid \boldsymbol{z}^{(m)})$ の形で後続変数の条件付き確率を与えることで、マルコフ連鎖を指定することができる。なお、すべての m について遷移確率が同じマルコフ連鎖のことを、均一マルコフ連鎖と呼ぶ。

ここで、均一マルコフ連鎖に対して不変性（定常性）を次のように定義する：均一マルコフ連鎖において

$$p^*(\boldsymbol{z}) = \sum_{\boldsymbol{z}'} T(\boldsymbol{z}', \boldsymbol{z}) p^*(\boldsymbol{z}') \tag{2}$$

が成り立つとき $p^*(\boldsymbol{z})$ は不変（定常）であるという。[*5]

いまサンプリングをおこなうための求めたい分布が不変となることを保証したい。保証するモチベとしては、サンプリングしていくなかで分布の形が変わってしまったら、つまり $p^*(\boldsymbol{z})$ でなくなってしまったら、我々が目的としていた $p^*(\boldsymbol{z})$ からサンプリングすることができなくなってしまう、それを防ぎたいということ

[*1] See p.246 上段
[*2] 基準と規準とは意味が異なることに注意せよ。前者はベンチマーク、後者はルールという意味を持っている。教科書の漢字は正しい。
[*3] Metropolis アルゴリズムについても、こちらのサブセクションで取り扱うことにする。
[*4] See p.104 上段
[*5] 遷移確率が恒等変換で与えられるなら任意の分布が不変性を満たす。

と理解した。合ってる？分布が不変であるための十分条件は $p^*(\boldsymbol{z})$ に対して

$$p^*(\boldsymbol{z})T(\boldsymbol{z}, \boldsymbol{z}') = p^*(\boldsymbol{z}')T(\boldsymbol{z}', \boldsymbol{z}) \tag{3}$$

で定義される詳細釣り合い条件が満たされるように遷移確率[*6]を選ぶことである。ある分布に詳細釣り合い条件を課すと、その分布が不変になることは簡単に示すことができる。これは $p^*(\boldsymbol{z})$ が不変の定義式 2 を満たすことを言えばいい。

$$\begin{aligned}\sum_{\boldsymbol{z}'} p^*(\boldsymbol{z}')T(\boldsymbol{z}', \boldsymbol{z}) &= \sum_{\boldsymbol{z}'} p^*(\boldsymbol{z})T(\boldsymbol{z}, \boldsymbol{z}') \ (\because \text{詳細釣り合い条件}) \\ &= p^*(\boldsymbol{z}) \sum_{\boldsymbol{z}'} T(\boldsymbol{z}, \boldsymbol{z}') \\ &= p^*(\boldsymbol{z}) \sum_{\boldsymbol{z}'} p(\boldsymbol{z}' \mid \boldsymbol{z}) \ (\because \text{遷移確率の定義}) \\ &= p^*(\boldsymbol{z})\end{aligned} \tag{4}$$

詳細釣り合い条件を満たすマルコフ連鎖は可逆であると言われる。

我々の目的は与えられた分布からのサンプリングをマルコフ連鎖を用いておこなうことである。そのためには、まず求めたい分布に対して詳細釣り合い条件を課すこと（分布の不変性を担保）。さらに、この詳細釣り合い条件に加えて、$m \to \infty$ のとき、初期分布 $p(\boldsymbol{z}^{(0)})$ の選択に関わらず、分布 $p(\boldsymbol{z}^{(m)})$ が求めたい不変分布 $p^*(\boldsymbol{z})$ に収束することも必要である。この収束に関する性質のことをエルゴード性と言い、また、このとき不変分布のことを平衡分布と呼ぶ。なおエルゴード性を構成する 3 つの条件については別資料を参照されたい[*7]。

合成遷移確率の話はスキップ。

1.2 Metropolis-Hastings アルゴリズム

「ベイズ深層学習」の p.86 を参照してほしい。最も基本的な MCMC の手法として Metropolis-Hastings 法を紹介する。サンプリングをおこなうための求めたい分布に関する条件については教科書を読み合わせる。MCMC のアルゴリズムを設計するには、遷移確率 T を設計する必要があるが、いま真の分布から直接設計することはできないため、提案分布 $q(\boldsymbol{z} \mid \boldsymbol{z}')$ を代用することを考える。

Metropolis-Hastings 法のアルゴリズムは p.86 の中段に示される通りである。遷移確率は提案分布を用いて

$$T(\boldsymbol{z}, \boldsymbol{z}') = q(\boldsymbol{z}' \mid \boldsymbol{z}) \min(1, \frac{\tilde{p}(\boldsymbol{z}')q(\boldsymbol{z} \mid \boldsymbol{z}')}{\tilde{p}(\boldsymbol{z})q(\boldsymbol{z}' \mid \boldsymbol{z})}) \tag{5}$$

と書ける[*8]。それでは、このように設計した遷移確率が求めたい分布に関して、詳細釣り合い条件を満たすことを確かめよう（See App）。

ここからは PRML に戻る。なお、提案分布が対称 $q(\boldsymbol{z}' \mid \boldsymbol{z}) = q(\boldsymbol{z} \mid \boldsymbol{z}')$ であるときは、式 5 右辺 min 内において、約分計算によってサンプルの受理率を表す式（11.44）が次のように書ける。

$$A_k(\boldsymbol{z}^*, \boldsymbol{z}^{(\tau)}) = \min(1, \frac{\tilde{p}(\boldsymbol{z}^*)}{\tilde{p}(\boldsymbol{z}^{\tau})}) \tag{6}$$

[*6] T の引数内は変数が入れ替わることで、T が別の量に変わることに注意せよ。これは同時分布の場合に、引数内の変数が入れ替わっても式としては等価であることとは対照的である。

[*7] https://mstour.hatenablog.com/entry/2020/07/21/073425 内の「定常分布に収束する条件」の項を見よ。

[*8] 教科書は $q(\boldsymbol{z}' \mid \boldsymbol{z})$ の引数内が逆になっている。これは間違いと思う。

これは Metropolis-Hastings 法における特殊な場合である Metropolis 法として知られている。Metropolis 法について手法の中身は読み合わせ（p.252）。

最後に提案分布の選択がアルゴリズムの性能に影響を与えることを見る。ここはよくわからなかったので一緒に解読したい。特に「状態空間での前進はゆっくりとしたランダムウォークの形を取り」のところが不明だった。

App 遷移確率が詳細釣り合い条件を満たすことの確認.

以下で与えられる遷移確率

$$T(z, z') = q(z'|z)\min\left(1, \frac{\tilde{p}(z')\, q(z|z')}{\tilde{p}(z)\, q(z'|z)}\right)$$

が求めたい分布 $p(z)$ に対して、次の詳細釣り合い条件の式

$$p(z)\, T(z, z') = p(z')\, T(z', z) \quad \cdots (*)$$

を満たすことを確かめる.

$$(\text{LHS of } (*)) = p(z)\, q(z'|z)\min\left(1, \frac{\tilde{p}(z')\, q(z|z')}{\tilde{p}(z)\, q(z'|z)}\right)$$

規格化定数を分子分母にかけ算

$$= p(z)\, q(z'|z)\min\left(1, \frac{p(z')\, q(z|z')}{p(z)\, q(z'|z)}\right)$$

$$= \min\left(p(z)\, q(z'|z),\ p(z')\, q(z|z')\right)$$

$$= \min\left(p(z')\, q(z|z'),\ p(z)\, q(z'|z)\right)$$

$$= p(z')\, q(z|z')\min\left(1, \frac{p(z)\, q(z'|z)}{p(z')\, q(z|z')}\right)$$

規格化定数で分子分母を割り算

$$= p(z')\, q(z|z')\min\left(1, \frac{\tilde{p}(z)\, q(z'|z)}{\tilde{p}(z')\, q(z|z')}\right)$$

$$= p(z')\, T(z', z) = (\text{RHS of } (*))$$

∎