

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

古作 創

2017年5月26日

1 概要

マルコフ連鎖モンテカルロ法 (MCMC; Markov Chain Monte Carlo method) とは、マルコフ連鎖の性質をもつ確率分布を用いつつ乱数でサンプリングを行うモンテカルロ法を組み合わせたサンプリング手法のことである。

2 マルコフ連鎖

マルコフ連鎖とは、現在の値 $z^{(\tau)}$ からのみ次の値 $z^{(\tau+1)}$ は決定することである。次の値 $z^{(\tau+1)}$ を決める際には現在の値 $z^{(\tau)}$ よりも前の値には依存しない。マルコフ連鎖の性質をもつ離散的なデータを持つ確率過程をマルコフ過程と呼ぶ。

マルコフ連鎖は現在の値が未来に影響するため独立ではない。そのため、現在発生させた乱数をもとに次にどのような乱数を発生させたら効率が良いかを調整する必要がある。その調整のためには、「詳細釣り合い条件」「エルゴード性」の2つの性質について考慮する必要がある。逆に、MCMC とはこの2つの性質をもたせたマルコフ連鎖を利用したモンテカルロ法による積分によって、未知の母数を効率良く求める手法のことである。

2.1 詳細釣り合い条件

確率分布 $p(\theta)$ とマルコフ連鎖 $z^{(1)}, z^{(2)}, \dots, z^{(\tau-2)}, z^{(\tau-1)}, z^{(\tau)}, z^{(\tau+1)}, \dots$ が存在し、以下の条件を満たしているとする。 $s(z^{(\tau)} \rightarrow z^{(\tau+1)})$ は状態 $z^{(\tau)}$ から状態 $z^{(\tau+1)}$ に移動する確率のことであり、遷移確率と呼ぶ。

任意の値 τ に対して

$$p(z^{(\tau)})s(z^{(\tau)} \rightarrow z^{(\tau+1)}) = p(z^{(\tau+1)})s(z^{(\tau+1)} \rightarrow z^{(\tau)})$$

詳細釣り合い条件のもとに状態が遷移するときには、以下のことがいえる。

- $p(z^{(\tau+1)}) > p(z^{(\tau)})$ であれば,
 $z^{(\tau)}$ よりも起こる可能性の少ない $z^{(\tau+1)}$ に移動する遷移確率 $s(z^{(\tau)} \rightarrow z^{(\tau+1)})$ は小さくなる
- $p(z^{(\tau+1)}) < p(z^{(\tau)})$ であれば,
 $z^{(\tau)}$ よりも起こる可能性の大きい $z^{(\tau+1)}$ に移動する遷移確率 $s(z^{(\tau)} \rightarrow z^{(\tau+1)})$ は大きくなる

2.2 エルゴード性

「任意の2つの状態 z と z' の間の遷移確率が有限個の0でない積で表すことができる」という性質のことをエルゴード性という。つまり、有限回のステップで任意の2つの状態間を行き来できるという性質のこと、どんなところでも必ずたどり着けることを意味する [?]。

3 ベイズの定理との関係

ベイズの定理を式 (??) に示す。式 (??) は式 (??) を表す。

$$P(\theta|Z) = \frac{P(Z|\theta)}{P(Z)} \cdot P(\theta) \quad (1)$$

$$\text{事後確率} = \frac{\text{尤度}}{\text{データが得られる確率}} \cdot \text{事前確率} \quad (2)$$

データが得られる確率 $P(Z)$ は、母数である θ を含んでいないため比例定数としてみることができる。これを考慮しないように変形すると以下の式が得られる。

$$P(\theta) \propto P(Z|\theta) \times P(\theta) \quad (3)$$

$$\text{事後確率} \propto \text{尤度} \times \text{事前確率} \quad (4)$$

連続変数であった場合は、式 (??) は式 (??) のようになる。

$$p(\theta|Z) = \frac{p(Z|\theta)}{p(Z)} \cdot p(\theta) \quad (5)$$

$$\text{事後分布} = \frac{\text{尤度}}{\text{データが得られる確率}} \cdot \text{事前分布} \quad (6)$$

ベイズの式は以下の式にも書き換えることができる。

$$p(\theta|Z) = \frac{p(Z|\theta)}{\int p(Z|\theta)p(\theta)d\theta} \cdot p(\theta) \quad (7)$$

式 (??) の分母の積分計算は、モデルが複雑な場合は解析的に解くことが困難になってしまう。そのため、積分計算の代わりに MCMC を用いてパラメータの推定を行う。

4 階層ベイズモデル

階層ベイズモデルとは、ベイズ法を用いて事後分布のパラメータを階層的に表した統計モデルのことである。現実の様々な事象は1つの確率分布で表すことは困難であるため、複数の確率分布を混同させることで複雑な事象を表現することができる。階層ベイズモデルを使用することで、個体差や環境差などを考慮して推定を行うことができる。

パラメータ θ に対するパラメータ ϕ を設定することで式 (??) のように書くことができる。

$$p(\theta|Z) \propto p(Z|\theta)p(\theta|\phi)p(\phi)$$

(8)

事後分布 \propto 尤度 \times 事前分布 \times 超事前分布

ここで使用される ϕ は超パラメータ (ハイパーパラメータ) と呼ばれ、 $p(\phi)$ のことを超事前分布と呼ぶ。これを用いることにより、パラメータ θ にもばらつきが発生し個体差や環境差を表現することができる。

5 サンプリング手法

5.1 Metropolis アルゴリズム

Metropolis アルゴリズムはMCMCの基本的なアルゴリズムで、Metropolis-Hastings アルゴリズムのもととなるアルゴリズムである。

Metropolis アルゴリズムを以下に示す [?].

Metropolis アルゴリズム

1. 初期値 $z^{(0)}$ をランダムに決定し、 $p(z^{(0)})$ を求める。
2. 現在の値を $z^{(\tau)}$ とし、ランダムに ε を決めて次の値の候補 z^* を $z^* = z^{(\tau)} + \varepsilon$ に決定し、 $p(z^*)$ を求める。
3. 遷移確率 $r = \frac{p(z^*)}{p(z^{(\tau)})}$ を決定する。
4. 条件により、値を更新する。
 - $r \geq 1$ ならば $z^{(\tau+1)} = z^*$ に更新する。
 - $r < 1$ ならば区間 0 から 1 の一様乱数 n を発生させる。
 - $r \geq n$ ならば $z^{(\tau+1)} = z^*$ に更新する。
 - $r < n$ でないなら $z^{(\tau+1)} = z^{(\tau)}$ に更新する。
5. $\tau = \tau + 1$ して 2 に戻る。

現在の値よりも確率が高くなれば必ず受理され、低いときは確率 r で棄却される。サンプリ

ングを行った後、サンプリング開始からのある程度の区間は不安定であるため burn-in 区間として削除する。

5.2 Metropolis-Hastings アルゴリズム

Metropolis-Hastings アルゴリズムは、Metropolis アルゴリズムの確率 r に関して提案分布 $q(\cdot)$ を導入したものであり、Metropolis アルゴリズムを一般化したものである [?]. 提案分布 $q(\cdot)$ とは、現在の値から次の値の候補を決めるために必要な確率分布である。

Metropolis-Hastings アルゴリズムを以下に示す。

Metropolis-Hastings アルゴリズム

1. 初期値 $z^{(0)}$ をランダムに決定し、 $p(z^{(0)})$ を求める。
2. 現在の値を $z^{(\tau)}$ とし、ランダムに ε を決めて次の値の候補 z^* を $z^* = z^{(\tau)} + \varepsilon$ に決定し、 $p(z^*)$ を求める。
3. 遷移確率 $r = \frac{p(z^*)q(z^{(\tau)}|z^*)}{p(z^{(\tau)})q(z^*|z^{(\tau)})}$ を決定する。
4. 条件により、値を更新する。
 - $r \geq 1$ ならば $z^{(\tau+1)} = z^*$ に更新する。
 - $r < 1$ ならば乱数 n を発生させる。
 - $r \geq n$ ならば $z^{(\tau+1)} = z^*$ に更新する。
 - $r < n$ でないなら $z^{(\tau+1)} = z^{(\tau)}$ に更新する。
5. $\tau = \tau + 1$ して 2 に戻る。

Metropolis-Hastings アルゴリズムも Metropolis アルゴリズム同様、サンプリング終了後に burn-in 区間を削除する。

5.3 Gibbs サンプリング

Gibbs サンプリングは、複数の確率分布のうち 1 つの確率分布以外を固定し条件付き分布としてサンプリングを行う。Gibbs サンプリングのアルゴリズムを以下に示す。

サンプリングしたい確率分布 $p(\mathbf{Z}) = p(z_1, \dots, z_M)$ について考える。

1. 確率変数 \mathbf{Z}_0 の初期値を設定する.
2. ステップ数 $\tau = 1, \dots, T$ に対して以下を行う.
 - $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \dots, z_M^{(\tau)})$ をサンプリングする.
 - $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \dots, z_M^{(\tau)})$ をサンプリングする.
 - \vdots
 - $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau)}, \dots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \dots, z_M^{(\tau)})$ をサンプリングする.
 - \vdots
 - $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \dots, z_{M-1}^{(\tau+1)})$ をサンプリングする.

ギブスサンプリングの初期値設定には、EM アルゴリズムを用いたりランダムに決定する。最終的に初期値からある程度の区間は burn-in 区間として切り捨てるため、あまり考えすぎなくて良い。

ギブスサンプリングは、メトロポリス・ヘイスティングス法の 1 つでもある。ギブスサンプリングの拡張として、Blocked Gibbs sampler, Collapsed Gibbs sampler がある。ギブスサンプリングの特徴として、事前分布，事後分布，尤度の確率分布が表??のように設定されている場合のみ効果を発揮する。

表 1: ギブスサンプリングに用いる確率分布

事前分布	尤度	事後分布
ベータ分布	二項分布	ベータ分布
正規分布	正規分布	正規分布
逆ガンマ分布	正規分布	逆ガンマ分布
ガンマ分布	ポアソン分布	ガンマ分布

参考文献

- [1] Logics of Blue, “ベイズと MCMC と統計モデルの関係”, <http://logics-of-blue.com/ベイズとmcmcと統計モデルの関係>.
- [2] 涌井良幸, “道具としてのベイズ統計”, 日本実業出版社, pp.140–175, 235, 2014.
- [3] C.M. ビショップ (著), 元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇 (監訳), “パターン認識と機械学習 下 ベイズ理論による統計的予測”, 丸善出版, pp.252–261, 2013.