## 概要

前回は一般的によく使われるサンプリング手法としてマルコフ連鎖モンテカルロ法を扱った。マルコフ連鎖モンテカルロ法は高次元空間からのサンプリングにおいてもよく機能する。サンプリングの方法は棄却サンプリングと同じく提案分布から得たサンプルを利用するが、直前のサンプルに依存して次のサンプルを得るという点が異なる。

さらに、マルコフ連鎖モンテカルロ法のうちの一つの手法として Metropolis-Hastings アルゴリズムを学んだ。Metropolis-Hastings アルゴリズムは、適切な規準を満たす受理確率を利用することで、得られるサンプルの不変性を保証するものである。

11.3 節では、 Metropolis-Hastings アルゴリズムのうち比較的単純で適用範囲の広いサンプリング手法である**ギブスサンプリング**を扱う。この手法は、複数の変数を同時にサンプリングする代わりに、ある変数以外を固定したもとでの一変数のサンプリングを繰り返すことで求めたいサンプル列を得る。サンプルを得る手続きにおいて受理確率が常に1である性質を持つ一方で、連続するサンプルの間に強い依存関係があるという性質がある。

11.4 節では、別のマルコフ連鎖モンテカルロ法である**スライスサンプリング**を扱う。これは Metropolis 法の難点であるステップサイズに敏感である点を改善する手法である。あるサンプルから次のサンプルを得るときには、真の分布の「スライス」を利用する。

## 11.3 ギブスサンプリング

ギブスサンプリングは単純で適用範囲の広いマルコフ連鎖モンテカルロアルゴリズムである。また、Metropolis-Hastings アルゴリズムの特別な場合とみなすこともできる。前節で学んだ Metropolis-Hastings 規準と照らし合わせながら手法を確認していく。

細かい手順は教科書を参照する。

サンプリングしたい確率分布 $p(\mathbf{z}) = p(z_1, \ldots, z_M)$ を考え，マルコフ連鎖のある初期状態を選択したと仮定しよう．ギブスサンプリングの各ステップでは，1 つの変数の値が置き換えられる．その際，残りの変数の値を固定した条件での，対象の変数の条件付き分布に従って抽出した値を用いる．すなわち，$z_i$ を分布 $p(z_i|\mathbf{z}_{\backslash i})$ から抽出された値で置き換える．ここで，$z_i$ は $\mathbf{z}$ の $i$ 番目の要素を表し，$\mathbf{z}_{\backslash i}$ は $z_1, \ldots, z_M$ から $z_i$ を除いたものを表す．この手続きは，各ステップで更新する変数をある決まった順序で循環するか，ある分布に従ってランダムに選択することで繰り返される．

真の条件付き分布

例えば，3 変数の分布 $p(z_1, z_2, z_3)$ を扱い，アルゴリズムのステップ $\tau$ で，値 $z_1^{(\tau)}$, $z_2^{(\tau)}, z_3^{(\tau)}$ を得ているとしよう．まず，$z_1^{(\tau)}$ を条件付き分布

$$p(z_1|z_2^{(\tau)}, z_3^{(\tau)}) \tag{11.46}$$

からサンプリングして得た新しい値 $z_1^{(\tau+1)}$ で置き換える．次に，新しい $z_1$ の値を以降のサンプリングのステップでそのまま用いて，$z_2^{(\tau)}$ を条件付き分布

$$p(z_2|z_1^{(\tau+1)}, z_3^{(\tau)}) \tag{11.47}$$

からサンプリングして得た値 $z_2^{(\tau+1)}$ で置き換える．そして，$z_3$ を

$$p(z_3|z_1^{(\tau+1)}, z_2^{(\tau+1)}) \tag{11.48}$$

から抽出したサンプル $z_3^{(\tau+1)}$ で更新する，といったことを 3 つの変数について順番に繰り返す．

✤ ギブスサンプリング ✤

1. $\{z_i : i = 1, \ldots, M\}$ を初期化する．
2. $\tau = 1, \ldots, T$ に対して以下を行う．
   - $z_1^{(\tau+1)} \sim p(z_1|z_2^{(\tau)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$ をサンプリングする．
   - $z_2^{(\tau+1)} \sim p(z_2|z_1^{(\tau+1)}, z_3^{(\tau)}, \ldots, z_M^{(\tau)})$ をサンプリングする．
   
     ⋮
   - $z_j^{(\tau+1)} \sim p(z_j|z_1^{(\tau+1)}, \ldots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \ldots, z_M^{(\tau)})$ をサンプリングする．
   
     ⋮
   - $z_M^{(\tau+1)} \sim p(z_M|z_1^{(\tau+1)}, z_2^{(\tau+1)}, \ldots, z_{M-1}^{(\tau+1)})$ をサンプリングする．

ギブスサンプリングでは、提案分布ではなく真の条件付き分布から直接サンプリングを行うため、実行が難しい場合がある。Metropolis-Hastings アルゴリズムでは「真の分布からサンプリングはできないが尤度は計算できる」という緩い条件であったのに対し、ギブスサンプリングの実行条件は少し厳しいものである。これについては節の後半で取り上げる。

次に、ギブスサンプリングが正しい分布 $p(\mathbf{z})$ からのサンプルを生成できることを確かめる。正しくサンプルを得るためには、求めたい分布 $p(\mathbf{z})$ がマルコフ連鎖における不変分布であることと、マルコフ連鎖のエルゴード性が成り立つことを示せば良い。ギブスサンプリングについては、不変性は成立するが、エルゴード性は一般的に成り立つとは限らない。

## 1. $p(\mathbf{z})$ が不変分布となることを示す。

詳細は教科書を参照する。方針としては以下の通りである。
前半：まず、ギブスサンプリングにおける1ステップ（一変数サンプリング）での不変さえ示せば、全体の不変が言える。（よくわからなかったので一緒に読んで理解したい）
後半：1ステップにおける不変性を示す。これは正しい条件付き分布を利用していることから言える。（下記に詳細釣り合い条件から示せるか試してみたので確認したい）

詳細釣り合い条件 $p(\mathbf{z})T(\mathbf{z},\mathbf{z}') = p(\mathbf{z}')T(\mathbf{z}',\mathbf{z})$ を示したい。
遷移確率である

$$T(\mathbf{z},\mathbf{z}') = p(z'_i \mid \mathbf{z}_{\setminus i})\,\mathbb{I}(\mathbf{z}'_{\setminus i} = \mathbf{z}_{\setminus i})$$

を利用すると。　∵ ギブスサンプリングでは、$\mathbf{z}_{\setminus i}$ が動く確率は 0 である

$$\begin{aligned}
p(\mathbf{z})T(\mathbf{z},\mathbf{z}') &= p(z_i \mid \mathbf{z}_{\setminus i})\,p(\mathbf{z}_{\setminus i})\,T(\mathbf{z},\mathbf{z}') \\
&= p(z_i \mid \mathbf{z}_{\setminus i})\,p(\mathbf{z}_{\setminus i})\,p(z'_i \mid \mathbf{z}_{\setminus i})\,\mathbb{I}(\mathbf{z}'_{\setminus i} = \mathbf{z}_{\setminus i}) \\
&= p(z'_i \mid \mathbf{z}_{\setminus i})\,p(\mathbf{z}_{\setminus i})\,p(z_i \mid \mathbf{z}_{\setminus i})\,\mathbb{I}(\mathbf{z}_{\setminus i} = \mathbf{z}'_{\setminus i}) \\
&= p(z'_i \mid \mathbf{z}'_{\setminus i})\,p(\mathbf{z}'_{\setminus i})\,p(z_i \mid \mathbf{z}'_{\setminus i})\,\mathbb{I}(\mathbf{z}_{\setminus i} = \mathbf{z}'_{\setminus i}) \quad (\because \mathbf{z}'_{\setminus i} = \mathbf{z}_{\setminus i} \text{ のみ残る}) \\
&= p(\mathbf{z}')T(\mathbf{z}',\mathbf{z})
\end{aligned}$$

## 2. エルゴード性を示す。

エルゴード性が成立する十分条件は、条件付き分布の確率が 0 となる場合がないことである。もしこれが成り立てば、要素変数のうち1つを更新するステップを有限回行うことで、z 空間の任意の点から任意の点へ到達可能である。もし成り立たない場合はエルゴード性の定義に立ち返って証明が必要となる。

ここまでギブスサンプリングから正しいサンプルが得られる保証することを示してきた。次に、ギブスサンプリングをより深く理解するために、 Metropolis-Hastings アルゴリズムの特別な場合として得ることを示す。 Metropolis-Hastings アルゴリズムにおける一つのステップで、変数 $z_k$ を扱い、残りの変数 $\mathbf{z}_{\backslash k}$ を固定し、 $\mathbf{z}$ から $\mathbf{z}^*$ への遷移確率が $q_k(\mathbf{z}^*|\mathbf{z}) = p(z_k^*|\mathbf{z}_{\backslash k})$ で与えられるものを考える。この時の受理確率は次のように 1 となり、必ず受理される。

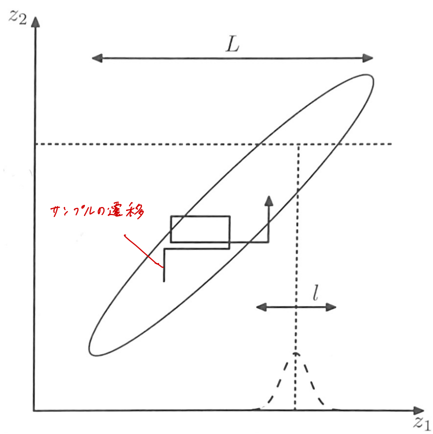
$$A(\mathbf{z}^*, \mathbf{z}) = \frac{p(\mathbf{z}^*)\, q_k(\mathbf{z}|\mathbf{z}^*)}{p(\mathbf{z})\, q_k(\mathbf{z}^*|\mathbf{z})}$$

$$= \frac{p(z_k^*|\mathbf{z}^*_{\backslash k})\, p(\mathbf{z}^*_{\backslash k})\, p(z_k|\mathbf{z}^*_{\backslash k})}{p(z_k|\mathbf{z}_{\backslash k})\, p(\mathbf{z}_{\backslash k})\, p(z_k^*|\mathbf{z}_{\backslash k})}$$

$$= \frac{p(z_k^*|\mathbf{z}_{\backslash k})\, p(\mathbf{z}_{\backslash k})\, p(z_k|\mathbf{z}_{\backslash k})}{p(z_k|\mathbf{z}_{\backslash k})\, p(\mathbf{z}_{\backslash k})\, p(z_k^*|\mathbf{z}_{\backslash k})} \qquad (\because\ \mathbf{z}^*_{\backslash k} = \mathbf{z}_{\backslash k})$$

$$= 1$$

次の図は二次元ガウス分布に対してギブスサンプリングを適用した例である。周辺分布の幅 L と条件付き分布の幅 $l$ に対して、独立なサンプルを得るためのステップ数が $(L/l)^2$ のオーダーとなるのが注目すべき点である。



この節の残りでは、ギブスサンプリングの弱点を補う拡張について取り上げる。ギブスサンプリングは変数を一つずつ順に更新するため、ステップ数が多く必要であるという弱点がある。これについては、**過剰緩和**により相関を持つ方向への移動を促進する効果を得ることができる。また、ギブスサンプリングを適用するには、条件付き分布からのサンプリングの容易さが必要である。これの判定のため、**マルコフブランケット**や**適応的棄却サンプリング法**を用いることができる。最後に、連続するサンプル間の依存関係を軽減するために、**ブロック化ギブスサンプリング**を紹介する。

## ■ 過剰緩和

ギブスサンプリングは一つずつ変数を更新するため、サンプルが広い範囲に分布するまでに多くのステップが必要になる。これは教科書の記述でいう「ランダムウォーク的な振る舞い」であり、これを低減する方法の一つが**過剰緩和**である。方法については教科書の読み合わせとする。

[ $z_i$ の平均と分散の計算]

$$
\begin{aligned}
E[z_i'] &= E[\mu_i + \alpha(z_i - \mu_i) + \sigma_i(1-\alpha^2)^{1/2}\nu] \\
&= \mu_i + \alpha(\mu_i - \mu_i) + \sigma_i(1-\alpha^2)^{1/2}\cdot 0 \\
&= \mu_i
\end{aligned}
$$

$$
\begin{aligned}
V[z_i'] &= V[\mu_i + \alpha(z_i - \mu_i) + \sigma_i(1-\alpha^2)^{1/2}\nu] \\
&= 0 + \alpha^2 V[z_i] + \sigma_i^2(1-\alpha^2)V[\nu] \\
&= \alpha^2\sigma_i^2 + \sigma_i^2(1-\alpha^2)\cdot 1 \\
&= \sigma_i^2
\end{aligned}
$$

(感想)

平均と分散が変化しないので、過剰緩和も求めたい分布を不変とする、というのはいささか納得ができなかった。ググってもそれっぽい答えが得られず。。

■ マルコフブランケットと適応的棄却サンプリングの利用

ギブスサンプリングが実際に適用可能かどうかは、条件付き分布 $P(z_k \mid \mathbf{z}_{\setminus k})$ からサンプルが容易に得られるかどうかに依存する。条件付き分布を計算する際には、すべての変数で条件付ける必要はなく、マルコフブランケットに含まれる変数についてのみ条件付けるだけでよい。マルコフブランケットで条件付けた分布は多くの場合対数凹になる。そのため、適応的棄却サンプリングを適用することができ、厳密な条件付き分布が必要なくサンプルを得られる。また、有向グラフ上の親子関係が共役性を保持する場合はより簡単にサンプルを得られる。この場合、条件付き分布はもとの分布と同じ関数形を保持するため、逆関数法などの標準的なサンプリング法を適用することができる。

（ICMの話は読み合わせ）

■ ブロック化ギブスサンプリング

こちらも読み合わせ

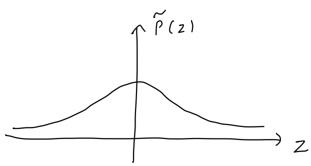
## 11.4 スライスサンプリング

Metropolis アルゴリズムの難点の一つはステップサイズに対して敏感であることを見てきた。もしステップサイズが小さすぎれば、ランダムウォーク的な振る舞いのために相関が消えるのが遅くなり、逆に大きすぎれば、高い棄却率のために非効率になる。**スライスサンプリング**と呼ばれる手法によって、分布の特徴に合わせてステップが自動的に調整される。（ギブスサンプリングでは真の条件付き分布を使うのでステップサイズの問題は起こらないはず。ギブスサンプリングと比較した時の優位点を最後に議論したい）

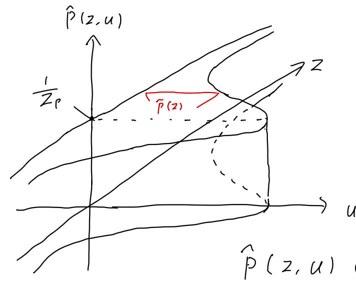
まず1変数の場合を考える。スライスサンプリングでは本来サンプリングしたい変数 $z$ に加え、付加的な変数 $u$ を同時にサンプリングする。ここでの目的は、次の分布に従うサンプルを得ることである。

$$\hat{p}(z, u) = \begin{cases} 1/Z_p & 0 \leq u \leq \tilde{p}(z) \text{ のとき} \\ 0 & \text{それ以外} \end{cases}$$

ここで、$\tilde{p}(z)$ は正規化されていない目的の分布であり、$Z_p$ はこれの正規化係数である。



$\tilde{p}(z)$ の図



$\hat{p}(z,u)$ の図

具体的なサンプリング方法については、図11.13 (a) を参照しながら確認する。

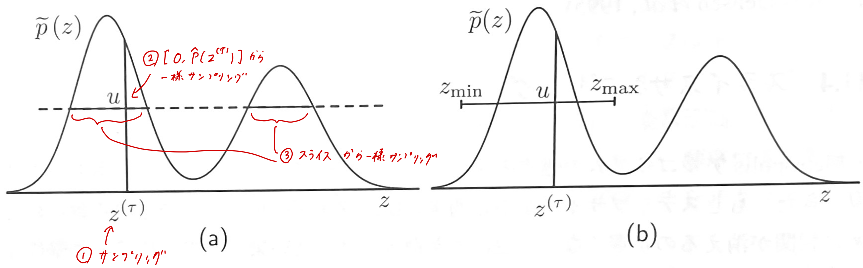



**図 11.13** スライスサンプリングの図解．(a) 与えられた値 $z^{(\tau)}$ に対して，$u$ の値は領域 $0 \leqslant u \leqslant \widetilde{p}(z^{(\tau)})$ から一様に選択され，これは水平の実線で示される，分布の「スライス」を定義する．(b) スライスから直接サンプリングすることは実行不可能なので，$z$ の新しいサンプルは，以前の値 $z^{(\tau)}$ を含む領域 $z_{\min} \leqslant z \leqslant z_{\max}$ から抽出される．

実際には、図中「③スライスから一様サンプリング」を実行するのは困難なことが多い。この場合には代わりに、スライスの領域にできるだけ一致するような領域を適応的に設定するアルゴリズムが利用される。図11.13 (b) のような $z_{min}$ $z_{max}$ を設定し、これらがぎりぎりスライスの外に出るまで領域を大きくしていくことで、スライスの領域を推定する。

また、ギブスサンプリングのように、各変数を順番にサンプリングすることで、スライスサンプリングを多変量の分布に拡張することができる。

（最後に）

ギブスサンプリングのように条件付き分布からサンプリングする方法と、スライスサンプリングのようにスライスから一様サンプリングする方法のどちらを使うべきなのか考察する。下記にまとめてみたところ、条件付き分布からサンプリングできる場合はギブスサンプリングを使うのがよく、そうでない場合はどちらもさほど変わらないように思えた。

■ ギブスサンプリング

条件付き分布からサンプリングできる必要がある。厳密に計算できない場合でも、条件付き分布が対数凹であれば適応的棄却サンプリング法を利用して実行できる。

■ スライスサンプリング

条件付き分布から直接サンプリングできる必要はない。ただし、スライスを推定するために、条件付き分布の尤度を計算できる必要がある。対数凹でない場合でも実行できるが、図 11.13 (b) を見たところ厳密なスライスを推定できていないので、実際には対数凹である必要がありそう。