P-9

事後分布とその実現手法

Numerical methods for obtaining posterior distributions

○池田宗馬 [1], 青柳美輝 [2]
Soma Ikeda, Miki Aoyagi

Abstract: In Bayesian statistics, it is necessary to obtain posterior distributions by calculating numerical approximations of multi-dimensional integrals. In this paper, we consider numerical methods for obtaining posterior distributions.

1. はじめに

ここでは事後分布とその実現手法について述べる．ベイズ推測を行う場合には事後分布を実現する必要があるが，一般に事後分布の積分を解析的に行うことはできないことが多く数値的な実現法が必要となる．通常，事後分布は簡単な確率モデルや共役事前分布などの特別な事前分布を用いない限りは解析的に計算することが困難なことが多い．

確率モデル $p(x|w)$，$(x \in \mathbb{R}^N, w \in W \subset \mathbb{R}^d)$ と事前分布 $\varphi(w)$ に対して，

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

とおく．

このとき事後分布は，$Z_n(\beta)$ を正規化定数として

$$p(w) = \frac{1}{Z_n(\beta)} \varphi(w) \prod_{i=1}^{n} p(X_i|w)^\beta = \frac{1}{Z_n(\beta)} \exp(-\beta H(w))$$

となる．ベイズ推測を行うためには，この確率分布を用いて平均値を計算することが必要となる．

以下，パラメータ $w \in \mathbb{R}^d$ に対し，関数 $H(w)$ が与えられたとき $\exp(-\beta H(w))$ に比例する確率分布に従うサンプル $\{w_k\}_{k=1}^K$ を生成するための方法を考察する．

これらの値は，

$$\int f(w)p(w)dw \sim \frac{1}{K} \sum_{k=1}^{K} f(w_k) \tag{1}$$

が成り立てば，高次元の平均を計算するために用いることができる．

数列 $\{w_1, w_2, w_3, \ldots\}$ がある条件付き確率 $p(w_{k+1}|w_k)$ に従って生成されているとする．このような確率過程のことをマルコフ過程とよぶ．マルコフ過程を用いて $\{w_1, w_2, w_3, \ldots\}$ を計算する方法をマルコフ連鎖モンテカルロ法（MCMC 法）という．

定理 1 (i)（詳細釣合い条件）任意の $w_a, w_b \in W$ について

$$p(w_b|w_a)p(w_a) = p(w_a|w_b)p(w_b)$$

が成り立つ．

(ii) 集合 W の任意の要素 w の近傍に到達する確率が 0 ではない．

これら 2 つの条件を満たすとき，$p(w_{k+1}|w_k)$ に従って生成される数列 $\{w_1, w_2, w_3, \ldots\}$ は，

$$\int f(w)p(w)dw = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} f(w_k)$$

を満たす．すなわち近似式 (**??**) が成立する．

2. メトロポリス法

メトロポリス法集合 $\{w(t) \in \mathbb{R}^d; t = 1, 2, 3, ...\}$ をつぎの手続きに従って得る．

1）初期値 $w(1)$ を決めて $t = 1$ とする．

2）得られている $w(t)$ から w' を条件確率 $r(w'|w(t))$ に従って生成する．ここで $r(w'|w(t))$ は，対称性「任意の w_1, w_2 について $r(w_1|w_2) = r(w_2|w_1)$」を満たすものであればよい．

3）$\Delta H \equiv H(w') - H(w(t))$ とおいて確率 $P = \min\{1, \exp(-\beta\Delta H)\}$ で $w(t+1) = w'$ とし，確率 $1-P$ で $w(t+1) = w(t)$ とする．

4）$t := t + 1$ とおいて 2) に戻る．これを繰り返す．

定理 2 $w(t)$ が与えられた時 $w(t+1)$ の条件付き確率を $p(w(t+1)|w(t))$ とおく．このとき，メトロポリス法は詳細釣合いの関係を満たしている．すなわち，任意の $w_a, w_b \in W$ について

$$p(w_b|w_a)p(w_a) = p(w_a|w_b)p(w_b)$$

が成り立つ．

3. ハイブリッド・モンテカルロ法

メトロポリス法の改良法としてハイブリッド・モンテカルロ法がある．新しい変数 $p \in \mathbb{R}^d$ を導入し，

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

と定義する．

1: 日大理工・院 (前)・数学　2: 日大理工・教員・数学

$\exp(-\mathcal{H}(w,p))$ に比例する確率密度関数のサンプリングを考える．この密度関数は w と p について独立であるから結局 $\exp(-\beta H(w))$ からのサンプリングを得ることができる．

1) $w(1)$ を初期化する．$t=1$ とする．

2) 平均 0 分散 1 の正規分布に従う d 個のサンプルを独立に発生して d 次元のベクトル p を構成する．

3) $(w(t),p)$ を初期値とするつぎの微分方程式を数値的に解いて，T 時刻後の (w',p') を得る．ここで t と T は無関係の値である．

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

4) $\Delta\mathcal{H}=\mathcal{H}(w',p')-\mathcal{H}(w(t),p)$ を求めて確率 $P=\min\{1,\exp(-\Delta\mathcal{H})\}$ で $w(t+1)=w'$ を採択し，$1-P$ で $w(t+1)=w(t)$ を採択する．

5)　2) に戻る．

このアルゴリズムは $\exp(-\mathcal{H}(w,p))$ に比例する確率分布に対する詳細釣合いを満たすので，得られた $\{(w(t),p(t))\}$ から $\{w(t)\}$ の部分のみを取り出せば，$\exp(-\beta H(w))$ に従うサンプリングができる．また 3) の微分法手式は常に $\frac{d\mathcal{H}}{d\tau}=0$ を満たすので，正確に解けるとすれば $\mathcal{H}$ を保存し，$\Delta\mathcal{H}=0$ である．すなわち，正確に解ければ $P=1$ である．微分方程式の数値解法によって誤差がで出ても P は比較的小さくなりにくく，T を大きめに設定できれば，確率 P 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．

微分方程式の解法は，次のリーブ・フロッグ法などが用いられている．

$$\frac{dw}{d\tau}=p,\quad \frac{dp}{d\tau}=f(w)$$

に対する繰り返しをつぎのように行う．

$$p(n+\frac{1}{2})=p(n)+\frac{\varepsilon}{2}f(w(n)),w(n+1)=w(n)+\varepsilon p(n+\frac{1}{2}),$$

$$p(n+1)=p(n+\frac{1}{2})+\frac{\varepsilon}{2}f(w(n+1))$$

ここで $\varepsilon>0$ は小さい定数である．

4. ギブス・サンプリング

パラメータ $w\in\mathbb{R}^d$ を 2 つの変数に分割して $w=(w_1,w_2)$ とする．目的とする確率分布を $p(w_1,w_2)$ として，この確率分布から定義される条件確率を $p(w_1|w_2)$ および $p(w_2|w_1)$ とする．$p(w_1|w_2)$ および $p(w_2|w_1)$ のそれぞれからのサンプリングが容易である場合に利用される．

集合 $\{w(t)=(w_1(t),w_2(t))\in\mathbb{R}^d; t=1,2,3,\ldots\}$ をつぎの手続きに従って得る．

1) 初期値 $w(1)$ を定めて，$t=1$ とする．

2) 確率 $\frac{1}{2}$ で $p(w_1'|w_2')p(w_2'|w_1'(t))$ に従って (w_1',w_2') を選出し，確率 $\frac{1}{2}$ で $p(w_2'|w_1')p(w_1'|w_2'(t))$ に従って (w_1',w_2') を選出する．

3) $w_{t+1}=(w_1',w_2')$ とおき，$t:=t+1$ とおいて 2) に戻る．

定理 3 ギブス・サンプリング法は詳細釣合いの関係を満たしている．

5. ランジュバン方程式を用いる方法

関数 $H(w)$ が与えられたとき $\exp(-\beta H(w))$ に比例する確率分布に従うサンプルを生成するための方法として，つぎのアルゴリズムを考える．$\varepsilon>0$ とする．

1) $w(t)$ を初期化する．$t=0$ とする．

2) $g(\varepsilon)$ を平均 0 分散 2ε の正規分布に従う確率変数とする．つぎの更新を行う．毎回独立な $g(\epsilon)$ を用いて

$$w(t+\varepsilon)=w(t)-\varepsilon\beta\nabla H(w(t))+g(\varepsilon)$$

とする．

3) $t=t+\varepsilon$ として 2) を繰り返す．

この方程式で $\varepsilon\to0$ として得られる確率微分方程式

$$\frac{dw}{dt}=-\beta\nabla H(w)+\frac{dB}{dt}$$

のことをランジュバン方程式という．ここで,$\frac{dB}{dt}$ は平均 0 分散 2ε の正規分布に従う確率変数を ε で割って，ε を零に近づけたものであり，通常の意味の確率変数にはならないが積分すると正規分布なるものである．これを白色雑音という．このアルゴリズムに従って生成される $w(t)$ が従う確率分布は次の補題のように，ある微分方程式を満たしている．

定理 4 上記のアルゴリズムに従って得られる確率変数 $W(t)$ の確率分布を $p(w,t)$ とする．$\varepsilon\to0$ とするとき

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

が成り立つ．これをフォッカー・ブランク方程式という．この定常解 $p(w,t)=\overline{p}(w)$ が存在するとすれば，$\overline{p}(w)$ は，$\exp(-\beta H(w))$ に比例する．すなわち目的の確率分布を得られる．

6. 参考文献

[1] 渡辺澄夫“ベイズ統計の理論と方法”, コロナ社, 2017

[2] S. Watanabe:“ Mathematical Theory of Bayesian Statistics ”, CRC Press, New York, USA, 2018.