『社会科学のためのベイズ統計モデリング』第 4 章

Hiroshi Hamada
Tohoku University

May, 2020
at Tohoku University

4.1 ベルヌーイ試行の具体例

コインを 10 回投げて，表を 1 裏を 0 とするデータ $(0, 1, 1, 1, 1, 0, 1, 1, 0, 1)$ を得た．

モデリングの仮定

$$\begin{aligned} X_i &\sim \mathrm{Bernoulli}(q), \quad i = 1, \ldots, 10 \\ q &\sim \mathrm{Beta}(a, b) \end{aligned}$$

事後分布がベータ分布であることは既に分かっているが（3 章），あえて MCMC によって事後分布を導出する（4 章）．

4.2MCMC の導入

事後分布と同時確率分布 $p(x^n, \theta) = p(x^n|\theta)p(\theta)$ が比例する

$$p(\theta|x^n) = \frac{p(x^n|\theta)p(\theta)}{p(x^n)} \propto p(x^n|\theta)p(\theta)$$

という関係を利用（$\propto$ は比例関係を表す）.
分母部分の周辺尤度を無視して，$p(x^n|\theta)p(\theta)$ の情報から，事後分布の近似となる経験分布を得る

4.3 メトロポリス・アルゴリズム

1. 現在のパラメータの値 θ_0 を目標とする事後分布からの MCMC サンプルの 1 つの要素として採用
2. 次に移動する候補として，ランダムに別のパラメータの値 θ_1 をとる（一様分布からサンプリング）
3. θ_0 と θ_1 の事後確率の比（事後オッズ）r を計算

$$r = \frac{p(\theta_1|x^n)}{p(\theta_0|x^n)} = \frac{\frac{p(x^n|\theta_1)p(\theta_1)}{p(x^n)}}{\frac{p(x^n|\theta_0)p(\theta_0)}{p(x^n)}}$$

$$= \frac{p(x^n|\theta_1)}{p(x^n|\theta_0)}\frac{p(\theta_1)}{p(\theta_0)}$$

4. 現在のパラメータの値 θ_0 から候補パラメータ θ_1 へ移動するかどうかを決める．現在のパラメータの値 θ_0 から候補パラメータ θ_1 に移動する確率（採択確率）$\alpha(\theta_0, \theta_1)$ を
$$\alpha(\theta_0, \theta_1) = \min\{1, r\}$$
と定義．確率 $\alpha(\theta_0, \theta_1)$ で θ_1 に移動し，確率 $1 - \alpha(\theta_0, \theta_1)$ で θ_0 のまま．
5. 新しいパラメータの値を θ_0 として，(1) に戻る．

アルゴリズムの直感的意味

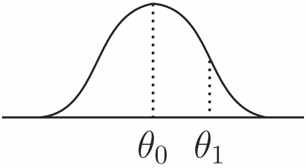
θ_0 から θ_1 へ確率 r で移動

θ_0 から θ_0 へ確率 $1-r$ で移動

$$p(\theta_0, \theta_1) = \alpha = \min\{1, r\} = r$$

$$r = \frac{p(\theta_1|x^n)}{p(\theta_0|x^n)}$$

$$p(\theta_0, \theta_0) = 1 - \alpha = 1 - r$$

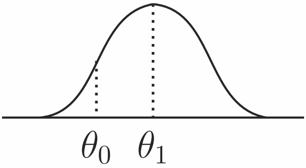


θ_0 から θ_1 へ確率 1 で移動

θ_0 から θ_0 へ確率 0 で移動

$$p(\theta_0, \theta_1) = \alpha = \min\{1, r\} = 1$$

$$p(\theta_0, \theta_0) = 1 - \alpha = 0$$



アルゴリズムの直感的意味

- 事後分布上は候補の確率が高い→確率 1 で移動
- 事後分布上は候補の確率が低い→低確率 $r < 1$ で移動
- 移動を繰り返すとサンプルの経験分布が事後分布に比例する
- ちゃんと事後分布からのサンプルに近似するか？ →マルコフ連鎖の性質を利用すれば可能

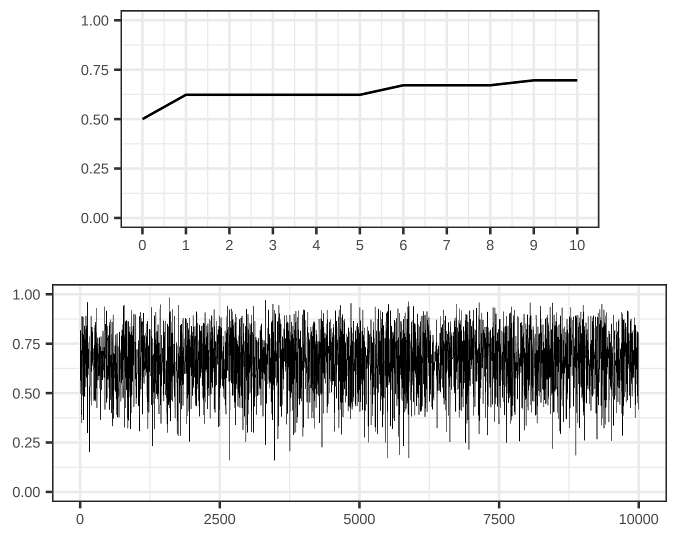


Figure: 10 試行までのトレースプロット (上)，1 万回のトレースプロット（下）

4.3.2 MCMC サンプリングの結果

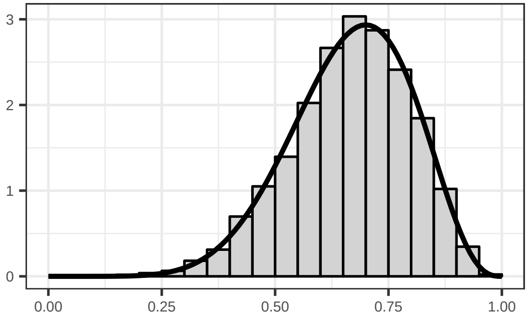


Figure: MCMC サンプルのヒストグラム．曲線は解析的に計算した事後分布（ベータ分布）

4.4 MCMC の一般的説明

$\{1, 2, \ldots, m\} = S$ を状態空間とする．すべての t について，状態 $i \in S$ にあるときに，$t+1$ 時点で状態 $j \in S$ にある確率が

$$
\begin{aligned}
&P(X_{t+1} = j | X_t = i) \\
&= P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0)
\end{aligned}
$$

となるとき，この確率過程をマルコフ連鎖 (Markov chain) を呼ぶ

推移確率

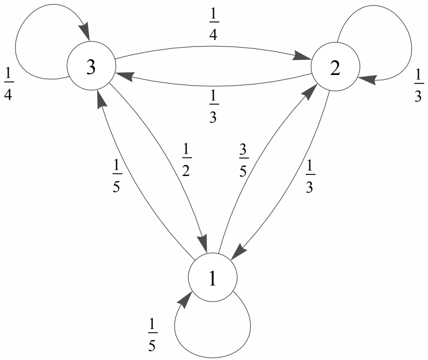
t 時における状態 $X_t = i \in S$ から状態 $X_{t+1} = j \in S$ に移る確率を推移確率とよんで，

$$p(i,j) = P(X_{t+1} = j | X_t = i)$$

で表す．このとき，

$$\sum_{j \in S} p(i,j) = 1.$$

つまり，状態 i から，（もとの状態 i を含む）いずれかの状態に至る確率は 1.



$$P = \begin{pmatrix} \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

Figure: 推移確率行列 P のグラフ

推移確率行列をかけても変化しない．つまり

$$\boldsymbol{\pi} = \boldsymbol{\pi} P \tag{1}$$

であるとき，$\boldsymbol{\pi}$ を定常分布という．
式 (1) を各状態 $i \in S$ についての式にすると，

$$\forall j \in S, \pi(j) = \sum_{i \in S} \pi(i) p(i, j) \tag{2}$$

4.4.3 詳細釣り合い条件

任意の $s, s' \in S$ について，

$$\forall s, s' \in S, \pi(s)p(s, s') = \pi(s')p(s', s) \tag{3}$$

が成り立つこと

詳細釣り合い条件 $\Longrightarrow \boldsymbol{\pi} = (\pi(1), \ldots, \pi(n))$ は定常分布

メトロポリス＝ヘイスティングス・アルゴリズム

1. （マルコフ連鎖の性質）任意の 2 つの状態間の推移確率がゼロではないという条件の下で，定常分布が存在するとき，どのような初期分布からでも，$t \to \infty$ のとき定常分布に収束する
2. MH アルゴリズムによって，目標分布 $\pi(s)$ が与えられたときに，詳細釣り合い条件を満たす推移確率を構成できる

事後分布（知りたい分布）がマルコフ連鎖の定常分布になるような推移確率を作る手続き＝MH アルゴリズム

MH アルゴリズム

1. 状態 s から始める
2. 状態 s から移る候補 s' を提案分布（現在の状態 s を条件とする条件付き確率分布）$q(\cdot|s)$ からのサンプリングで決める．
3. 以下の比 r を計算する．

$$r = \frac{\pi(s')q(s|s')}{\pi(s)q(s'|s)}$$

4. 現在の状態 s から移動候補 s' を採択する採択確率を

$$\alpha(s, s') = \min\{1, r\}$$

と定義．確率 $\alpha(s, s')$ で s' に移動し，確率 $1 - \alpha(s, s')$ で s にとどまる．

5. 新しい状態を s として，(1) に戻る．

アルゴリズムからの推移確率の構成

$$\underbrace{p(s,s')}_{\text{s から s' への推移確率}} = \underbrace{q(s'|s)}_{\text{候補 s' の提案確率}} \times \underbrace{\alpha(s,s')}_{\text{候補 s' の採択確率}}$$

$$\underbrace{p(s',s)}_{\text{s' から s への推移確率}} = \underbrace{q(s|s')}_{\text{候補 s の提案確率}} \times \underbrace{\alpha(s',s)}_{\text{候補 s の採択確率}}$$

提案分布は正規分布や一様分布を考えると
$q(s'|s) = q(s|s')$（対称）となる．正規分布の場合

$$q(s'|s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(s'-s)^2}{2\sigma^2}\right\} = q(s|s')$$

詳細釣り合い条件の導出

$$\begin{aligned}
\pi(s)p(s,s') &= \pi(s)q(s'|s)\alpha(s,s') \\
&= \pi(s)q(s'|s)\min\{1,r\} \\
&= \min\{\pi(s)q(s'|s),\pi(s)q(s'|s)r\} \\
&= \min\{\pi(s)q(s'|s),\pi(s')q(s|s')\} \\
&= \pi(s')q(s|s')\min\left\{\frac{\pi(s)q(s'|s)}{\pi(s')q(s|s')},1\right\} \\
&= \pi(s')q(s|s')\min\left\{r',1\right\} \\
&= \pi(s')q(s|s')\alpha(s',s) \\
&= \pi(s')p(s',s)
\end{aligned}$$

4.4.5 事後分布の MCMC

比 r を計算する際に，2 つのパラメータ値の事後分布の比をとることで，周辺尤度 $p(x^n)$ がキャンセルされる

$$
\begin{aligned}
r &= \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \\
&= \frac{p(\theta'|x^n)q(\theta|\theta')}{p(\theta|x^n)q(\theta'|\theta)} \\
&= \frac{p(x^n|\theta')\varphi(\theta')q(\theta|\theta')}{p(x^n|\theta)\varphi(\theta)q(\theta'|\theta)}
\end{aligned}
$$

対称な提案分布（一様分布や正規分布）を設定することで，対称分布もキャンセルされるので，比 r は

$$
\begin{aligned}
r &= \frac{\pi(\theta')}{\pi(\theta)} \\
&= \frac{p(\theta'|x^n)}{p(\theta|x^n)} \\
&= \frac{p(x^n|\theta')}{p(x^n|\theta)} \frac{\varphi(\theta')}{\varphi(\theta)}.
\end{aligned}
$$

これは 4.3 節で示したメトロポリス・アルゴリズム中の r と一致する．

4 章まとめ

- MCMC は共役事前分布が使えない場合に，事後分布を数値計算で近似する方法である．
- MCMC を用いることで，複雑なモデルでもベイズ推定が可能になる場合がある．