

有限混合分布のベイズ推定とラベル・スイッチング

統計数理研究所 伊庭幸人

階層モデルへのギブス・サンプラーの応用について、具体例をひとつ論じよう。これは、最も簡単な例であるが、同時に、いろいろと興味深い疑問を生じさせる例でもある。

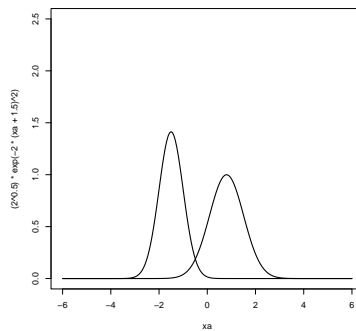


図 1: 有限混合分布。2つの正規分布が p_+ と p_- に対応する。両者の密度を重み付きで加えた分布が、有限混合分布の一番簡単な例である（図の場合は2つの山がある分布になる）。

確率モデルに基づいた教師なしの分類問題（クラスタリング）を考える。2つに分類する場合を考えて、サンプル $y = \{y_i\}$ の各々が密度 $p_+(y|\theta_+)$ か $p_-(y|\theta_-)$ のどちらか一方から来ているが、われわれはどちらか知らないとする。この仮定のもとで、個々のサンプルの起源の推定と密度の位置や形を決める未知パラメータ θ_+ , θ_- の推定を同時に行うのが課題である。図 1 では、 p_+ , p_- がともに 1 次元の正規分布の場合を描いてあるが、5 変量、10 変量の分布や正規分布以外の分布の場合も同様に考えられる。

y_i の起源を $(+1, -1)$ の 2 値を取るラベル x_i で示そう ($i = 1 \dots N$)。 $x_i = 1$ なら y_i は P_+ からのサンプル、 $x_i = -1$ なら P_- からのサンプルで

ある．以下，必要に応じて， $\mathbf{y} = (y_i)$ ， $\mathbf{x} = \{x_i\}$ とまとめて書く．

\mathbf{y} の分布は，未知のラベル $\mathbf{x} = \{x_i\}$ 及び分布 p_+ ， p_- のパラメータ θ_+ ， θ_- をパラメータとして，

$$p(\mathbf{y}|\mathbf{x}, \theta_+, \theta_-) = \prod_{i=1}^N \left\{ \left(\frac{1+x_i}{2} \right) p_+(y_i|\theta_+) + \left(\frac{1-x_i}{2} \right) p_-(y_i|\theta_-) \right\}$$

と書ける． $\frac{1+x_i}{2}$ ， $\frac{1-x_i}{2}$ が y_i がどちらの分布から来たかによって，0,1 の値を取るインジケータになっていることに注意．事前密度・事前分布を，それぞれ， $p(\theta_+)$ ， $p(\theta_-)$ ， $P(\mathbf{x}) = \prod_{i=1}^N P(x_i)$ とする．実際は， $P(x_i)$ もパラメータを含むべきだが，ここでは省く¹．

以上のもとで， \mathbf{x} ， θ_+ ， θ_- の事後密度は，

$$p(\mathbf{x}, \theta_+, \theta_- | \mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}, \theta_+, \theta_-) p(\theta_+) p(\theta_-) P(\mathbf{x})}{\sum_{\mathbf{x}} \int \int p(\mathbf{y}|\mathbf{x}, \theta_+, \theta_-) p(\theta_+) p(\theta_-) P(\mathbf{x}) d\theta_+ d\theta_-} \quad (1)$$

となる．この事後密度をサンプルするギブス・サンプラーを構成しよう．

まず， θ_+ ， θ_- を固定すると， \mathbf{x} の各成分 x_i は互いに独立になり， x_i の条件付き確率は

$$P(x_i = \pm 1 | \mathbf{y}, \theta_+, \theta_-) = \frac{p_{\pm}(y_i | \theta_{\pm}) P(x_i = \pm 1)}{p_+(y_i | \theta_+) P(x_i = +1) + p_-(y_i | \theta_-) P(x_i = -1)} \quad (2)$$

で与えられる（複号同順）．(2) は，2つの分布 P_+ ， P_- の中心や幅を与えるパラメータ θ_+ ， θ_- を固定したときに，分類の確率を与える式とみることが出来る．

一方， $\mathbf{x} = \{x_i\}$ を固定すると， θ_+ ， θ_- の条件付き密度は，それぞれ

$$p(\theta_{\pm} | \mathbf{y}, \mathbf{x}) = \frac{p(\theta_{\pm}) \prod_{i \in A_{\pm}(\mathbf{x})} p_{\pm}(y_i | \theta_{\pm})}{\int \left\{ p(\theta_{\pm}) \prod_{i \in A_{\pm}(\mathbf{x})} p_{\pm}(y_i | \theta_{\pm}) \right\} d\theta_{\pm}} \quad (3)$$

¹ サンプル y_i がそれぞれの分布に帰属される事前確率 $P(x_i = \pm 1)$ をデータから推定しないと，クラスターサイズが不均一な場合（正解が大体半々ずつに分類される場合以外）に効率が下がる．

となる(複号同順)。ここで、 $x_i = 1$ となる i の集合を $A_+(x)$ 、 $x_i = -1$ となる i の集合を $A_-(x)$ と書いた。(3)は、(2)と相補的に、分類を与えたときに、分布の位置や幅をきめるパラメータ θ_+ 、 θ_- の事後確率を決める式である。

(2)を用いた $\{x_i\}$ のサンプリングと(3)を用いた θ_+ 、 θ_- のサンプリングを交互に行うことで、(1)からのサンプルを生成するギブス・サンプラーが構成できる。(3)の右辺が簡単にサンプルできるかが問題であるが、たとえば「 p_1, p_0 が正規分布で θ_+ 、 θ_- がそれぞれの母平均をあらわすパラメータ」で「事前密度 $p(\theta_+)$ 、 $p(\theta_-)$ が正規分布」なら、(3)の右辺も正規分布になるので問題ない²。

ここで構成したギブス・サンプラーの動作は、パラメータ θ_+ 、 θ_- をきめて分類を行うことと、分類の結果を用いてパラメータ θ_+ 、 θ_- を選びなおすことを交互に繰り返して、自己組織的に分類を進めていると解釈できる。ただし、マルコフ連鎖モンテカルロ法の特徴として、 x も θ_+ 、 θ_- がどこか1点に収束するわけではない。 x も θ_+ 、 θ_- も、最初のうち系統的なトレンドを持って運動するが、そのあと、分布(1)からのサンプルを生成しながら、永遠に定常的に揺らぎ続けるのである。

趣旨は明快であるが、問題点はいろいろある。一番気になるのは次のことだろう。事前分布として、 $p(\theta_+)$ と $p(\theta_-)$ を同一の分布に取り、 $P(x_i) = P(-x_i)$ とすると、分布(1)は、2つの分布の形を決めるパラメータと所属をあらわすラベルを同時に入れ替える変換

$$\begin{aligned} \text{すべての } i \text{ について } ,x_i &\rightarrow -x_i \\ \theta_{\pm} &\rightarrow \theta_{\mp} \text{ (複号同順)} \end{aligned} \quad (4)$$

について確率の値が変わらないという性質を持っている。そこで、次のような疑問が起きる。

- (i) ギブス・サンプラーを実行したとき、変換(4)で対応する状態の対が、計算時間内に同じ頻度で列の中に出現するか。

² θ_{\pm} が分散や共分散も含む場合は $p(\theta_+)$ 、 $p(\theta_-)$ をそれらに対する「共役事前密度」に選ぶ必要がある。

(ii) 上記が満たされると、反転した状態ともとの状態の両方が同じ頻度で列の中に出てくるので、 $x_i = +1$ となる確率、 $x_i = -1$ となる確率はすべての i について 0.5 に収束する。これでは、周辺確率でみる限り「分類」したことにならないが、それでよいのか。

(i) についていえば、条件によって、計算時間内に両方出る場合も、一方に固着してしまう場合があると思われる。後者ではアルゴリズムが正しく働いていないことになるが、そこから、混合時間を短くする工夫をしたり「反転するまで待たなくても統計量が収束する場合があるか」等を考えるのは、アルゴリズムの研究として自然な展開である。

(ii) の方は、アルゴリズムの責任ではない。分布 (1) を書いた時点でそうなっているのである。それが問題になるのであれば、(1) に基づく定式化そのもの、また、それで「何がやりたいのか」を再考しなくてはならない³。アルゴリズムのせいにして「反転した状態が出ないほうがむしろよい」等と考え出すと、混乱してしまう。一般に、問題がアルゴリズムにあるのか、枠組みやモデルにあるのか、分けて考える必要がある。

マルコフ連鎖モンテカルロ法の導入によって生じたようにみえる問題が、実はもともとの枠組みやモデルの抱える問題の顕在化であることは、統計科学においてしばしばみられる。

³ 予測分布など変換 (4) で不変な量のみを考える、というのがひとつの立場であるが、それで本当に良いのかどうかは難しい問題である。