階層ベイズモデル

伴　正隆

ベイズモデルの推定では，尤度とパラメータの事前分布の積によって表されるパラメータの事後分布を評価することが目的となる．階層ベイズモデル（hierarchical bayes model）は，ベイズモデルにおけるパラメータの事前分布について，さらに多段階の事前分布を追加的に設定したモデルである．階層構造を持つモデルを構築することが自然であるような状況や問題は数多く存在し，さまざまな分野で応用されている．

1. 階層ベイズモデルとは

まずベイズモデルの推定では，尤度とパラメータの事前分布の積によって表されるパラメータの事後分布を評価することが目的となる．階層ベイズモデル（hierarchical bayes model）は，ベイズモデルにおけるパラメータの事前分布について，さらに多段階の事前分布を追加的に設定したモデルである[2]．

例えば，ある地域で病院ごとに心臓疾患患者の生存率を推測する場合，その病院内のサンプルだけを用いて推測することも考えられるが，各病院の生存率の背後にその地域の病院間で共通の生存率が存在し，各病院の生存率はその背景からのサンプルであると考えることもできる．そのような階層性が想定される問題に対して階層構造を持つモデルを構築することは自然な設定であり，さまざまな分野で応用されている[1]．

以下では，より一般的な表現で階層ベイズモデルについて説明したのち，具体的な説明として，階層ベイズ回帰モデルを紹介する．

2. モデルの概要

データを y とし，パラメータを β とすると，非階層のベイズ推測では β に関して事前分布 $p(\beta)$ を設定し，尤度 $p(y|\beta)$ と事前分布を掛け合わせたパラメータの事後分布

$$p(\beta|y) \propto p(y|\beta)p(\beta) \tag{1}$$

によって β を推測する．

しかしいま，パラメータ β がさらに別のパラメータ θ に依存している場合，β の事前分布は $p(\beta|\theta)$ になり，ベイズ推定を行うためにはさらに θ の事前分布 $p(\theta)$ が必要になる．すると(1)式で定義したパラメータ β の事後分布は，事前分布が入れ替わり，

$$p(\beta|y) \propto p(y|\beta)p(\beta|\theta)p(\theta) \tag{2}$$

となる．ここでは2段階までの例にとどめるが，階層ベイズモデルではパラメータの依存関係がなくなるまで多段階の階層構造が構築される．

ここで，どのパラメータにも依存しないパラメータ θ の事前分布を規定するパラメータはハイパーパラメータ（hyper parameter）と呼ばれる．全体に影響を与えない事前分布になるような値を置くなど，既知とすることが多い．階層ベイズモデルでは階層化の段階が増えることで最下層に存在するハイパーパラメータの影響は小さくなり，事前分布を設定することの恣意性が弱められる[3]．

階層が増えるほど多数のパラメータを扱うことになる階層ベイズモデルでは，パラメータの同時事後分布が複雑なものになることが多く，事後分布を解析的に積分評価することは困難である．そこでモデルの推定には目的のパラメータ以外をすべて条件付きとした完全条件付き事後分布（full conditional posterior distribution）から逐次乱数を発生させることで同時分布からの乱数を得る，マルコフ連鎖モンテカルロ法（Markov chain Monte Carlo 法：MCMC 法）が使われるのが一般的である．

3. 階層ベイズ線形回帰モデル

ここでは例として，回帰係数について階層化した線形回帰モデルを紹介する．ただし，紙面の制約と説明の単純化のために単回帰を想定し，階層モデルについても定数項だけの簡素なものを考える．

ばん　まさたか
目白大学 経営学部
〒161-8539 新宿区中落合 4-31-1

主体を $i(i=1,\cdots,I)$, $t(t=1,\cdots,T)$ を各データに付与された番号とし，主体ごとに T 個のデータが観測されている状況を考える．例えばマーケティングであれば主体を消費者と読み替え，消費者ごとに T 回の購買時点データがある状況，あるいは教育学関連の分野であればクラス i ごとに学生 T 人分の試験の成績がデータとして存在しているような状況である．

いま，y を目的変数，x を説明変数として，両者の間につぎの線形回帰モデルを仮定する．

$$y_{it}=\alpha_i+x_{it}\beta_i+\varepsilon_{it} \quad ; \varepsilon_{it}\sim N(0,\sigma^2) \tag{3}$$

ここで α, β は回帰係数である．$X_{it}=(1,x_{it})$, $\theta_i=(\alpha_i,\beta_i)'$ として(3)式をベクトル表記すると，$y_{it}=X_{it}\theta_i+\varepsilon_{it}$ となる．これは主体別のパラメータから構成されており，主体ごとの特徴を表現しているので，主体内（within-subject）モデルと呼ばれる．

また，パラメータ θ_i の階層モデルを

$$\theta_i=\bar{\theta}+\delta_i \quad ; \delta_i\sim N(0,\Delta) \tag{4}$$

とする．ここで $\bar{\theta}$ は主体共通のパラメータである．このモデルは主体間の共通性を意味していることから，主体間（between-subject）モデルと呼ばれる．

二つのモデルのパラメータは主体内モデルでは $\{\theta_i, \sigma^2\}$，主体間モデルでは $\{\bar{\theta},\Delta\}$ である．主体共通のパラメータそれぞれについて不確実性を許容し，以下のように事前分布を設定する．

$$\begin{cases}\sigma^2\sim \text{Inverted Gamma}\,(q_0/2,\ r_0/2)\\ \bar{\theta}\sim N(s_0,t_0)\\ \Delta\sim \text{Inverted Wishart}\,(u_0,v_0)\end{cases} \tag{5}$$

$(q_0,r_0,s_0,t_0,u_0,v_0)$ はこのモデルにおけるハイパーパラメータである．

(2)式に対応させるとパラメータの同時事後分布は，ハイパーパラメータを $\phi_0=(q_0,r_0\,;s_0,t_0\,;u_0,v_0)$ とおくと，

$$\begin{aligned}&p(\{\theta_i\},\sigma^2,\bar{\theta},\Delta|y)\\ &\quad\propto p(y|\{\theta_i\},\sigma^2)p(\{\theta_i\}|\bar{\theta},\Delta)\\ &\qquad p(\sigma^2|q_0,r_0)p(\bar{\theta}|s_0,t_0)p(\Delta|u_0,v_0)P(\phi_0)\end{aligned} \tag{6}$$

となる．

また，MCMC法でパラメータを推定する際には，各パラメータの完全条件付き事後分布から逐次乱数を発生させる．初期値の影響を受けない十分な回数を繰り返したあとのサンプルを用いて，目的のパラメータの事後平均や事後分散を求める．特に θ_i の事後分布は，$X_i=(X_{i1},\cdots,X_{iT})'$, $Y_i=(y_{i1},\cdots,y_{iT})'$ とすると，

$$p(\theta_i|\sigma^2,\bar{\theta},\Delta,y)\sim N(k_iV_i,V_i),$$
$$\begin{cases}k_i=\sigma^{-2}X_i'Y_i+\Delta^{-1}\bar{\theta},\\ V_i=(\sigma^{-2}X_i'X_i+\Delta^{-1})^{-1}\end{cases} \tag{7}$$

となる．事後分布の平均の分子部分 k_i では，主体間の共通部分である $\bar{\theta}$ が，主体間の異質性の程度を示す Δ との積の形で構成要素となっており，Δ が尤度と事前分布の融合度を決定していることが分かる．

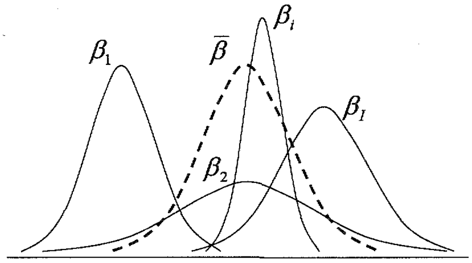

図1 パラメータ事後分布の例

MCMC によるパラメータの推定では θ_i の事後分布が主体ごとに導出されるとともに，主体間での共通性 $\bar{\theta}=(\bar{\alpha},\bar{\beta})'$ も推定され，特に β_i と $\bar{\beta}$ の事後分布を描くと図1のようになる．主体ごとに平均も分散も異なる分布を実線で描き，主体の共通部分を示す事後分布を破線で描いている．先ほどの例に戻ると，マーケティングでいえば，消費者の購買時点での価格に対する反応の大きさが消費者ごとに分かるとともに，消費者共通の反応の大きさを推定しているような状況である．また，クラスごとに学生の出席回数が成績に与える影響を測定しながら，その背景にあると考えられる学年全体で共通の成績への影響度を測定している状況も例として挙げられる．

参考文献

[1] Gelman, A., et al., Bayesian Data Analysis, Second Edition, Chapman & Hall/CRC (2004).
[2] Greenberg, E., Introduction to Bayesian Econometrics, Cambridge University Press (2007).
[3] 照井伸彦, R によるベイズ統計分析（シリーズ〈統計科学のプラクティス〉2), 朝倉書店 (2010).