ギブスサンプリングによる多変量回帰

羅洪帥 (57165029)

2017 年 7 月 17 日

目次

---

1 照井本 89 ページ例題を New MCMCpack での再実行

ベイジアン正規線形回帰モデルの実例として照井本の 89 ページの例を実行した。結果は'1.3.6' バージョンの MCMCpack は照井本と同じ結果が出力されることを確認した。R コード及び出力は添付した HW_Rcode.R と HW_output.pdf に記載される。

2　MCMCregress の事前分布とパラメータ設定

R_MCMCpack.pdf の 110 ページの記載によると、R の MCMCregress の事前分布はセミ共役事前分布で、$\boldsymbol{\beta}$ は

$$\boldsymbol{\beta} \sim \mathcal{N}(b_0, B_0^{-1})$$

であり、σ^2 は

$$(\sigma^2)^{-1} \sim Gamma(c_0/2, d_0/2)$$

もしくは、

$$\sigma^2 \sim IGamma(c_0/2, d_0/2)$$

である。

b0

　$\boldsymbol{\beta}$ の期待値ベクター。スカラーの値を定義する場合は、全部の β にそのスカラーの値を期待値と利用する。

B0

　$\boldsymbol{\beta}$ の精度の行列を定義する。B0 は 0.2 と設定する場合に、K=3 の $\boldsymbol{\beta}$ の分散共分散行列は対角要素は B0 の逆数になるもの。

$$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

c0, d0

　$c_0/2, d_0/2$ は σ^2 の分布を決めるパラメータ。

sigma.mu,sigma.var

　sigma.mu,sigma.var は σ^2 の期待値と分散から分布を決めるパラメータ。例えば、sigma.mu=5 と sigma.var=25 を設定する場合は、逆ガンマ分布のパラメータと以下の関係で対応できる。

$$\frac{c_0/2}{d_0/2 - 1} = \text{sigma.mu} = 5$$

$$\frac{(c_0/2)^2}{(d_0/2 - 1)^2(d_0/2 - 2)} = \text{sigma.var} = 25$$

　連立方程式を解けば、$c_0 = 20, d_0 = 6$ だと分かる。

beta.start

　$\boldsymbol{\beta}$ の初期値設定。デフォルトでは最小 2 乗推定を用いる。

3　K=3 での回帰モデル

本節では、K=3 のモデルでギブスサンプリングによる回帰モデルを記述する。設定したモデルから乱数を用いて発生させたデータを観測データとする。説明変数は 2 つの属性があり、それぞれ x_{1i} と x_{2i} で記録する。切片を加えて、K=3 のモデルとなる。

3.1　事前の設定

切片付きのモデルを下記の用に設定する。

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2) \quad i = 1, ..., n$$

興味あるパラメータは係数 $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ と誤差項の分散 σ^2 である。観測データは $\mathrm{y} = (y_1, ..., y_n)'$ と X である。X は下記のようである。ただし、$x_{0i} = 1$ である。

$$\begin{bmatrix} x_{01} & x_{11} & x_{21} \\ x_{02} & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ x_{0n} & x_{1n} & x_{2n} \end{bmatrix}$$

3.2　最小 2 乗推定値

本節では、以降の数式変形の準備として、係数の最小 2 乗推定値 $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ を求める。モデルの予測値と観測値の差 e_i である残差の 2 乗和は

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (x_{0i}\hat{\beta}_0 + x_{1i}\hat{\beta}_1 + x_{2i}\hat{\beta}_2))$$

である。最小化するには、下記の連立方程式が得られる。

$$\begin{aligned} \frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_j} &= -2 \sum_{i=1}^{n} x_{ji}(y_i - (x_{0i}\hat{\beta}_0 + x_{1i}\hat{\beta}_1 + x_{2i}\hat{\beta}_2)) \\ &= -2 \sum_{i=1}^{n} x_{ji} e_i \\ &= 0 \qquad j = 0, 1, 2 \end{aligned}$$

$\mathrm{e} = (e_1, ..., e_n)' = \mathrm{y} - \mathrm{X}\hat{\beta}$ として、連立方程式方程式から、$\mathrm{X}'\mathrm{e} = 0$ が分かる。

3.3　尤度関数

尤度関数 $p(\mathrm{y}|\mathrm{X}, \boldsymbol{\beta}, \sigma^2)$ を求めたいですが、現在は σ^2 の分布しか分からないである。$J_{\epsilon \to y}$ を利用して、変数変換する必要はある。つまり、

$$p(\mathrm{y}|\mathrm{X}, \boldsymbol{\beta}, \sigma^2) = p(\epsilon|\mathrm{X}, \boldsymbol{\beta}, \sigma^2)|J_{\epsilon \to y}|$$

$n \times 1$ の誤差項 $\epsilon = (\epsilon_1, ..., \epsilon_n)'$ は正規分布で、かつ独立のため、同時確率は下記のようである。

$$p(\epsilon|\mathrm{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\epsilon_i - 0)^2}{2\sigma^2}\right)$$

変数変換のヤコビアンは下記である。

$$J = \begin{bmatrix} \frac{\partial \epsilon_1}{\partial y_1} & \cdots & \frac{\partial \epsilon_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \epsilon_n}{\partial y_1} & \cdots & \frac{\partial \epsilon_n}{\partial y_n} \end{bmatrix}$$

$\epsilon_i = y_i - (\beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2)$ にたいして、y_j で偏微分する。

$$\frac{\partial \epsilon_i}{\partial y_j} = \frac{\partial y_i}{\partial y_j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

したがって、上記ヤコビアンは単位行列になる。

$$J = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

したがって、観測データの同時確率は

$$p(\mathrm{y}|\mathrm{X}, \boldsymbol{\beta}, \sigma^2) = p(\epsilon|\mathrm{X}, \boldsymbol{\beta}, \sigma^2)|J_{\epsilon \to y}| = p(\epsilon|\mathrm{X}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\epsilon_i - 0)^2}{2\sigma^2}\right)$$

$\epsilon_i = y_i - (\beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2)$ を利用して、行列で書き換えて、比例関係ある項だけを残る。

$$\begin{aligned} p(\mathrm{y}|\mathrm{X}, \boldsymbol{\beta}, \sigma^2) &= (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})\right) \\ &\propto (\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})\right) \end{aligned}$$

以降の式変形のため、$(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})$ をさらに変形する。前述した最小 2 乗推定では $\hat{\mathrm{y}} = \mathrm{X}\hat{\beta}$ と $\mathrm{e} = \mathrm{y} - \mathrm{X}\hat{\beta} = \mathrm{y} - \hat{\mathrm{y}}$ および $\mathrm{X}'\mathrm{e} = 0$ を利用して、下記の式変形をする。

$$\begin{aligned} (\mathrm{y} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{y} - \mathrm{X}\boldsymbol{\beta}) &= (\mathrm{y} - \hat{\mathrm{y}} + \hat{\mathrm{y}} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{y} - \hat{\mathrm{y}} + \hat{\mathrm{y}} - \mathrm{X}\boldsymbol{\beta}) \\ &= (\mathrm{e} + \hat{\mathrm{y}} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{e} + \hat{\mathrm{y}} - \mathrm{X}\boldsymbol{\beta}) \\ &= [\mathrm{e} + \mathrm{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]'[\mathrm{e} + \mathrm{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ &= \mathrm{e}'\mathrm{e} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathrm{X}'\mathrm{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathrm{X}'\mathrm{e} \\ &= \mathrm{e}'\mathrm{e} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\mathrm{X}'\mathrm{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \end{aligned}$$

3.4　事前分布

セミ共役事前分布で、$\boldsymbol{\beta}$ と σ^2 の事前確率は独立として、その同時事前分布は

$$p(\boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{\beta})p(\sigma^2)$$

である。ただし、事前分布は $\boldsymbol{\beta}$ は

$$\boldsymbol{\beta} \sim \mathcal{N}_3(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$$

であり、σ^2 は

$$\sigma^2 \sim IGamma(v_0/2, s_0/2)$$

と定義される。多変量正規分布の確率密度関数は照井本 4.3.1 多変量正規分布と尤度関数に記載されるように、

$$p(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0) = (2\pi)^{-1/2}|\boldsymbol{\Sigma}_0|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)$$

となる。逆ガンマ分布の確率密度関数のカーネルは照井本 4.2.2(34 ページ) で示したように、

$$p(\sigma^2|v_0, s_0) \propto (\sigma^2)^{-(v_0/2+1)} \exp(-\frac{s_0}{2\sigma^2})$$

事前分布の同時確率は多変量正規分布と逆ガンマ分布の積となって、下記のようにまとめられる。

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2) &= p(\boldsymbol{\beta})p(\sigma^2) \\ &\propto |\boldsymbol{\Sigma}_0|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)(\sigma^2)^{-(v_0/2+1)} \exp(-\frac{s_0}{2\sigma^2}) \end{aligned}$$

3.5　事後分布

尤度関数と事前分布の積により、事後分布は下記のように求める。

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2|\mathrm{X}, \mathrm{y}) &= p(\boldsymbol{\beta}, \sigma^2)p(\mathrm{y}|\mathrm{X}, \boldsymbol{\beta}, \sigma^2) \\ &\propto (\sigma^2)^{-n/2} \exp\left(\frac{-1}{2\sigma^2}(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})'(\mathrm{y} - \mathrm{X}\boldsymbol{\beta})\right) \\ &\quad \times |\boldsymbol{\Sigma}_0|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)(\sigma^2)^{-(v_0/2+1)} \exp(-\frac{s_0}{2\sigma^2}) \end{aligned}$$

3.6　条件付事後分布

ギッブスサンプリングでは条件付事後分布でパラメータを生成するため、条件付事後分布を求める必要がある。

$p(\sigma^2|\boldsymbol{\beta},\mathrm{X},\mathrm{y})$ を求めるには、同時事後確率から σ^2 を含める項目だけを残す。

$$\begin{aligned}
p(\sigma^2|\boldsymbol{\beta},\mathrm{X},\mathrm{y}) &\propto (\sigma^2)^{-n/2}\exp\left(\frac{-1}{2\sigma^2}(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})'(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})\right)(\sigma^2)^{-(v_0/2+1)}\exp(-\frac{s_0}{2\sigma^2})\\
&= (\sigma^2)^{-(\frac{v_0+n}{2}+1)}\exp\left(\frac{-1}{2\sigma^2}\left(s_0+(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})'(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})\right)\right)\\
&= (\sigma^2)^{-(\frac{v_1}{2}+1)}\exp\left(\frac{s_1}{2\sigma^2}\right)
\end{aligned}$$

ここの $v_1=v_0+n, s_1=s_0+(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})'(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})$ であるため、σ^2 の条件付事後分布は $IG(v_1/2, s_1/2)$ である。

$p(\boldsymbol{\beta}|\sigma^2,\mathrm{X},\mathrm{y})$ を求めるには、同時事後確率から $\boldsymbol{\beta}$ を含める項目だけを残す。前提した下記の式変形を同時事後確率に代入して、

$$(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})'(\mathrm{y}-\mathrm{X}\boldsymbol{\beta}) = \mathrm{e}'\mathrm{e}+(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathrm{X}'\mathrm{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})$$

$\boldsymbol{\beta}$ を含む項だけを残す。

$$\begin{aligned}
p(\boldsymbol{\beta}|\sigma^2,\mathrm{X},\mathrm{y}) &\propto \exp\left(\frac{-1}{2\sigma^2}(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})'(\mathrm{y}-\mathrm{X}\boldsymbol{\beta})\right)\exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right)\\
&\propto \exp\left(\frac{-1}{2\sigma^2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathrm{X}'\mathrm{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right)\exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right)\\
&= \exp\left[-\frac{1}{2}\left\{\sigma^{-2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathrm{X}'\mathrm{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})+(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\}\right]
\end{aligned}$$

$\boldsymbol{\beta}$ を含む項をさらに分解するため、上記の式の指数部分の式を、照井本 43 ページの公式 B で適用する。

$$\begin{aligned}
&\sigma^{-2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathrm{X}'\mathrm{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})+(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
&= (\boldsymbol{\beta}-\boldsymbol{\beta}^*)'(\sigma^{-2}\mathrm{X}'\mathrm{X}+\boldsymbol{\Sigma}_0^{-1})(\boldsymbol{\beta}-\boldsymbol{\beta}^*)+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\left((\sigma^{-2}\mathrm{X}'\mathrm{X})^{-1}+(\boldsymbol{\Sigma}_0^{-1})^{-1}\right)^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)
\end{aligned}$$

上式の右第 2 項はすでに $\boldsymbol{\beta}$ が含まない。この中の β^* の定義及び $\mathrm{X}'\mathrm{e}=0$ を利用した式変形は

$$\begin{aligned}
\beta^* &= (\sigma^{-2}\mathrm{X}'\mathrm{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\sigma^{-2}\mathrm{X}'\mathrm{X}\hat{\boldsymbol{\beta}}+\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0)\\
&= (\sigma^{-2}\mathrm{X}'\mathrm{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\sigma^{-2}\mathrm{X}'\hat{\boldsymbol{y}}+\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0)\\
&= (\sigma^{-2}\mathrm{X}'\mathrm{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}\left(\sigma^{-2}\mathrm{X}'(\boldsymbol{y}-\boldsymbol{e})+\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0\right)\\
&= (\sigma^{-2}\mathrm{X}'\mathrm{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\sigma^{-2}\mathrm{X}'\boldsymbol{y}+\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0)
\end{aligned}$$

すると、条件付き確率式に代入し、$\boldsymbol{\beta}$ を含む項だけを残す。

$$p(\boldsymbol{\beta}|\sigma^2,\mathrm{X},\mathrm{y}) \propto \exp\left(\frac{-1}{2}\left((\boldsymbol{\beta}-\boldsymbol{\beta}^*)'\boldsymbol{\Sigma}_*^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^*)\right)\right)$$

よって、$\boldsymbol{\beta}$ は $N_3(\boldsymbol{\beta}^*, \boldsymbol{\Sigma}_*)$3 次元の多変量正規分布。この中の $\boldsymbol{\Sigma}_* = (\sigma^{-2}\mathrm{X}'\mathrm{X} + \boldsymbol{\Sigma}_0^{-1})^{-1}$ である。従って、$\boldsymbol{\beta}^* = \boldsymbol{\Sigma}_*(\sigma^{-2}\mathrm{X}'\boldsymbol{y} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0)$ と書き換える。

(疑問: 今回では β を全体としてサンプリングするが、$\beta_0, \beta_1, \beta_2$ 個別個別の順でサンプリングするべきでしょうか? また､その差異はあるでしょうか? これを今後の課題としたいと思う。)

3.7 R コードへの反映

1. 初期化。パラメータの初期値 beta.start を決める。シミュレーション回数 mcmc，バーンイン期間 burnin を設定する。
2. サンプリングの繰り返し。毎回では､逆ガンマ分布の σ^2 と 多変数正規分布の $\boldsymbol{\beta}$ をサンプリング値を生成する。
3. 終了。mcmc+burnin 回を実行しましたら､終了する。

3.8 自作ギブスサンプリングと MCMCregress の結果を比較

	True Value	MCMCregress	Original Gibbs
β_0	1	1.0936	1.0938
β_1	0.5	0.4616	0.4616
β_2	-0.5	-0.4745	-0.4745
σ^2	0.25	0.2937	0.2937

4 実証分析:ポートランドの不動産価格

オレゴン州のポートランド市の 47 件不動産のデータを使って､不動産価格と部件の面積と寝室数の関係を求める。 LiveInPorland.txt には各部件の平方フィート単位での面積、寝室の数、ドル単位での価格を記録されている。

この実証分析は以下の懸念点があると思う。

- 面積と寝室の数の間は相関があり、面積大きいほうに、寝室が多い。
- 寝室の数は 2,3,4 に集中する。
- 価格の数は数十万程度の数字に対し、面積の数は数千ぐらいである。

データそのままで使用は不適切だと思う。一旦すべてのデータを正規化した。変換後のデータでは被説明変数を価格にして y で表記、説明変数を面積と寝室数にして、それぞれ x_1 と x_2 で表記する。

自作ギブスサンプリングと MCMCregress の結果を下記に記述する。

	MCMCregress	SD	Original Gibbs	SD
β_0	0.0002086	0.1200	0.0000527	0.0369982
β_1	0.8816722	0.1456	0.8845845	0.0452540
β_2	-0.0515708	0.1452	-0.0530577	0.0452108
σ^2	0.6729265	0.1399	0.0644245	0.0040801