

# Rで学ぶベイズ統計学入門

6.6 ベイズの計算での方策

6.7 グループ化データからの正規母集団を検討する

6.8 出力分析の事例

6.9 コーシー誤差によるデータのモデリング

6.10 スタンフォード心臓移植手術データの分析

茨城大学工学部情報工学科

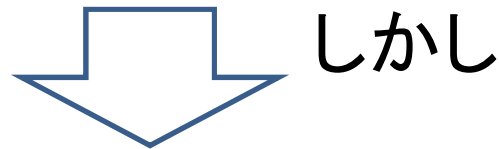
佐々木研究室

荒井悠有

## 6.6 ベイズの計算での方策

対数事後密度を要約するアプローチとして

- ・メトロポリス-ヘイスティングスのランダムウォークと独立連鎖
  - ・ギブスサンプリング・アルゴリズム
- がある。



パラメータのベクトル $\theta$ の位置と広がりについて初期値を指定する必要がある。

そこで、laplace関数()で事後密度を定める方法を例を用いて説明する。

# 6.7 グループ化データからの正規母集団を検討する

例．ランダムな標本が平均 $\mu$ で標準偏差 $\sigma$ の正規母集団から抽出される場合

→ある大学の男子学生の身長の平均と標準偏差を検討する

身長 <small>のビン</small> (インチ)	頻度
66未満	14
66－68	30
68－70	49
70－72	70
72－74	33
74以上	15

\* データはビン(区間幅)の頻度で記録されている。

## 正規パラメータの尤度

$$L(\mu, \sigma) \propto \Phi(66, \mu, \sigma)^{14} (\Phi(68, \mu, \sigma) - \Phi(66, \mu, \sigma))^{30} \\ \times (\Phi(70, \mu, \sigma) - \Phi(68, \mu, \sigma))^{49} (\Phi(72, \mu, \sigma) - \Phi(70, \mu, \sigma))^{70} \\ \times (\Phi(74, \mu, \sigma) - \Phi(72, \mu, \sigma))^{33} (1 - \Phi(74, \mu, \sigma))^{15}$$

$(\mu, \sigma)$ には $1/\sigma$ に比例する通常の無情報事前分布を割り当てる。  
すると、パラメータの事後密度は以下に比例する。

$$g(\mu, \sigma | data) \propto \frac{1}{\sigma} L(\mu, \sigma)$$

正の標準偏差を $\lambda = \log(\sigma)$ で変換すると、 $(\mu, \lambda)$ の事後密度は以下で与えられる。

$$g(\mu, \lambda | data) \propto L(\mu, \exp(\lambda))$$

## ▪ $(\mu, \log\sigma)$ の事後モードを計算する

- $(\mu, \lambda)$ の事後密度の対数を計算する(`groupeddatapost()`関数)

- $(\mu, \log\sigma)$ のモードを推定する

-`laplace()`関数を用いて $(\mu, \log\sigma)$ の事後モードを求める

```
$mode  
[1] 70.169880 0.973644
```

出力より、 $(\mu, \log\sigma)$ の事後モードは(70.17, 0.97)  
とわかる。

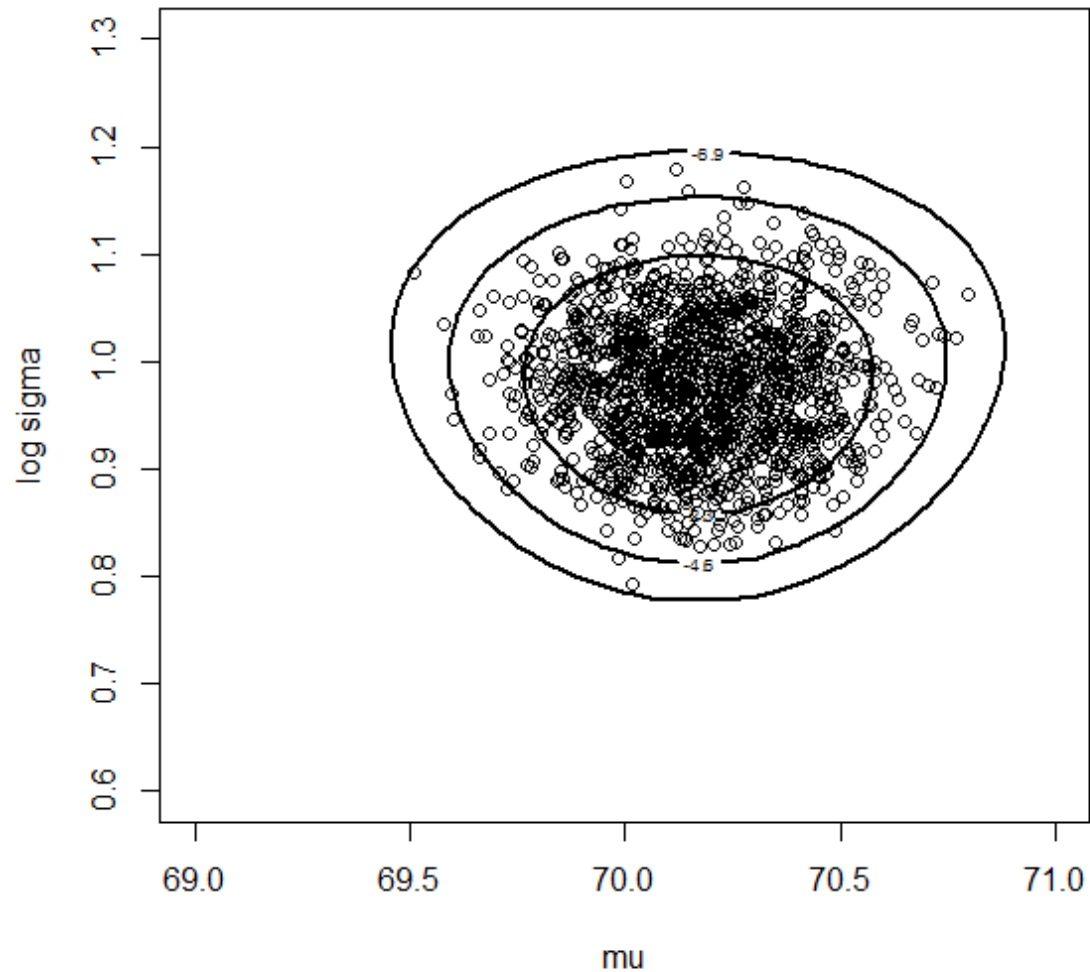
laplace()関数の平均と標準偏差をMCMCアルゴリズムによる値と比較する。

(初期値 $(\mu, \log\sigma)=(70, 1)$ , 尺度パラメータ=2)

		modal.sds	
mu	70.169880	0.18800834	←laplace()関数
log sigma	0.973644	0.05609341	
	post.means	post.sds	←メトロポリス・ランダム ウォーク・アルゴリズム
mu	70.1777610	0.1887366	
log sigma	0.9799327	0.0563362	

事後分布のモーメントの二つの集合はほぼ一致している。  
→事後分布に対するモードの近似は実用上正確である。

$\mu$ と $\log\sigma$ の事後密度の等高線図を確認すると、等高線が楕円形をしており、この例では正規近似が正確であることがわかる。



## 6.8 出力分析の事例

-メトロポリス・ランダムウォーク・アルゴリズムを、適切でない初期値と提案密度で実行してみる。



初期値 $(\mu, \log\sigma)=(65, 1)$  , 尺度パラメータ=0.2 に設定する。

	Mean	SD	Naïve SE	Batch SE
mu	70.1676	0.18474	0.0020654	0.012634
log sigma	0.9812	0.05789	0.0006472	0.004046

$(\mu, \log\sigma)=(65, 1)$   
尺度パラメータ=0.2

	Mean	SD	Naïve SE	Batch SE
mu	70.1679	0.19761	0.0022093	0.005650
log sigma	0.9832	0.05747	0.0006425	0.001754

$(\mu, \log\sigma)=(70, 1)$   
尺度パラメータ=2

適切な値を設定した方がMCMC連鎖のパフォーマンスが優れている。

## 6.9 コーシー誤差によるデータモデリング

-外れ値を含む可能性のあるデータのモデリングを取り上げる

例.  $y_1, \dots, y_n$  が位置パラメータ  $\mu$  で尺度パラメータ  $\sigma$  のコーシー密度からのランダム標本だとする

$$f(y | \mu, \sigma) = \frac{1}{\pi\sigma(1+z^2)}$$

$z=(x-\mu)/\sigma$ である。また $(\mu, \sigma)$ に通常は無情報事前分布を割り当てると仮定する。

$$g(\mu, \sigma) \propto \frac{1}{\sigma}$$

・ $\mu$ と $\sigma$ の事後密度

$$g(\mu, \sigma | data) \propto \frac{1}{\sigma} \prod_{i=1}^n f(y_i | \mu, \sigma)$$
$$= \frac{1}{\sigma} \prod_{i=1}^n \left[ \frac{1}{\sigma} \left( 1 + (y_i - \mu)^2 / \sigma^2 \right)^{-1} \right]$$

・ $(\mu, \lambda)$ の事後密度 ( $\lambda = \log \sigma$ )

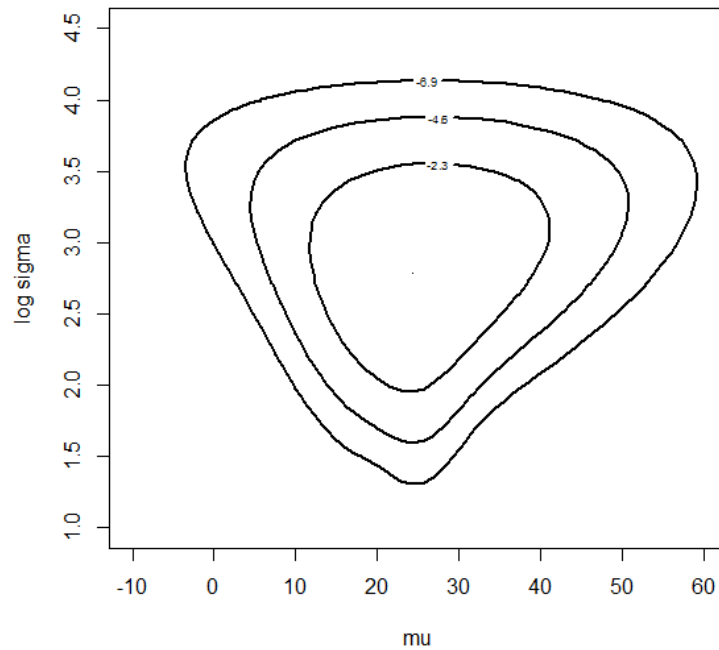
$$g(\mu, \lambda | data) \propto \prod_{i=1}^n \left[ \exp(-\lambda) (1 + \exp(-2\lambda)(y_i - \mu)^2)^{-1} \right]$$

・密度の対数

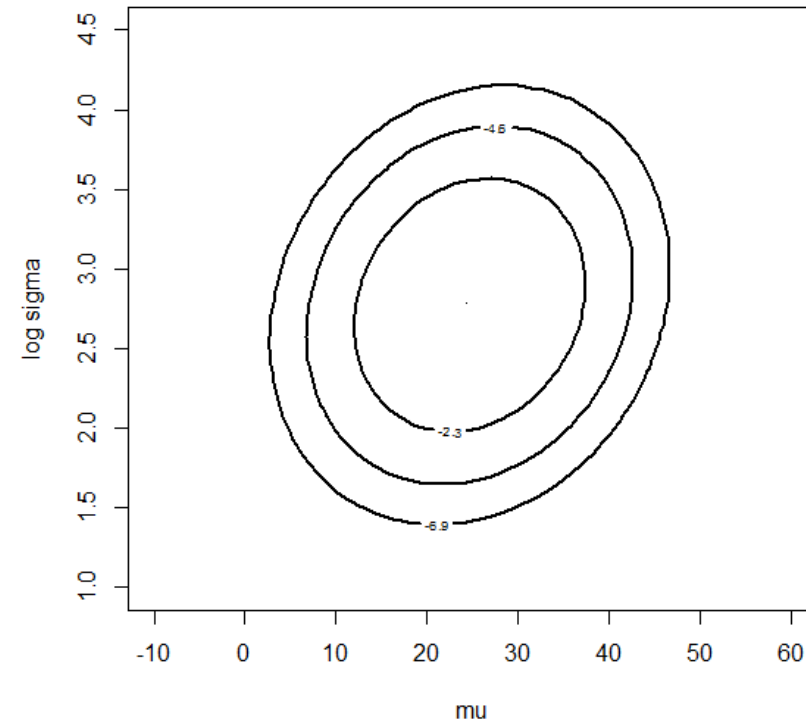
$$\log g(\lambda, \mu | data) = \sum_{i=1}^n \left[ -\lambda - \log(1 + \exp(-2\lambda)(y_i - \mu)^2) \right]$$



このモデルをダーウィンのデータセットにあてはめて、事後密度を考えていく。



事後分布の等高線図



事後分布の正規近似による等高線図

## 6.10 スタンフォード心臓移植手術データの分析

推論の目標: 心臓移植が患者の寿命を延ばすかどうかを判断すること

未知のパラメータである $(\tau, \lambda, p)$ を検討していく。



1. `laplace()`関数を実行して、事後分布の位置について初期値を与える。
2. メトロポリス・ランダムウォーク・アルゴリズム(`rwmetrop()`関数)をつかって事後密度からシミュレーションする。

## ・パラメータ( $\tau, \lambda, p$ )の要約

	[,1]	[,2]	[,3]
5%	0.4900237	12.65571	0.3061562
50%	0.9390019	27.90786	0.4680367
95%	2.0776187	66.86441	0.7563478



$\tau=1$ が事後分布の中心にあるので、このデータから  $\tau \neq 1$ と結論づけるには証拠不十分である。つまり、移植手術によって死亡のリスクが高くなる(あるいは低くなる)と結論づけるには証拠不十分である。

- 患者の生存曲線を推定する。

$$\text{生存関数 } S(t) = \frac{\lambda^p}{(\lambda + t)^p} \quad (t > 0)$$

