

## 可逆跳MCMC抽样在生物易感模式中的应用 \*

刘 瑞 银

(沈阳师范大学数学与系统科学学院, 沈阳, 110034)

### 摘 要

在一些生物实验中, 受试者对于一定处理的易感模式往往是不同的, 需要对不同的易感模式进行确定. 由于易感模式的个数和形式同时未知, 我们利用可逆跳MCMC方法, 对易感模式个数和参数联合起来建模, 基于他们的后验概率进行推断. 为了方便, 一般假定不同易感模式的分散程度是相同的. 本文对分散程度不同的情况作了分析. 数据分析表明可逆跳MCMC方法在易感模式处理中是非常显著的.

关键词: 混合模型, 可逆跳MCMC方法, 易感模式.

学科分类号: O213.9.

### §1. 引 言

在生物实验中, 实验室的受试者对于一定的处理其易感模式可能是不同的, 此时需要确定不同易感模式的具体形式. 在剂量反应模型中, Liu等(2011)利用RJMCMC抽样(参见Green (1995))的贝叶斯方法在有限混合正态模型下去估计参数并确定子总体(易感模式)的个数. 这个方法相比于以前的工作而言更加具有一般性和全面性. 一方面, 其考虑的是多个剂量水平而不是一个剂量水平; 另一方面, 在每个剂量水平下有多个子总体或者说有多个易感模式. 但是在Liu等(2011)中, 假定了不同易感模式的方差都是相同的. 实际中, 不同易感模式下方差不一定相同, 本文基于这个实际情况, 利用RJMCMC抽样(参见Green (1995))的贝叶斯方法在有限混合正态模型下去估计参数并确定子总体(易感模式)的个数.

### §2. 模 型

#### 2.1 问题

我们研究这样一个实际例子, 强化C5YSF1黑鼠服用羰基铁从而研究其体内葡萄糖水平. 此数据共有5个剂量, 在这个研究中, 高的葡萄糖含量被认为是不正常的, 极端高的葡萄糖水平认为和糖尿病有关. 此数据的来源参见Razzaghi和Kodell (2000)和Chen等(1996).

\*辽宁省教育厅科研基金(L2010514)资助.

本文2011年5月17日收到, 2012年9月3日收到修改稿.

羧基铁服用量和葡萄糖水平之间的关系见图1. 从图1可以看到, 在每个水平下葡萄糖水平都可以分成多个子总体. 在图1(a)中, 是Liu等(2011)中确定的三个易感模式函数, 在图1(b)中, 是本文最后确定的三个易感模式函数. 这里要研究的问题是, 在每个剂量下假定有 $K$ 个易感模式, 如何去确定易感模式的个数 $K$ 以及易感模式函数.

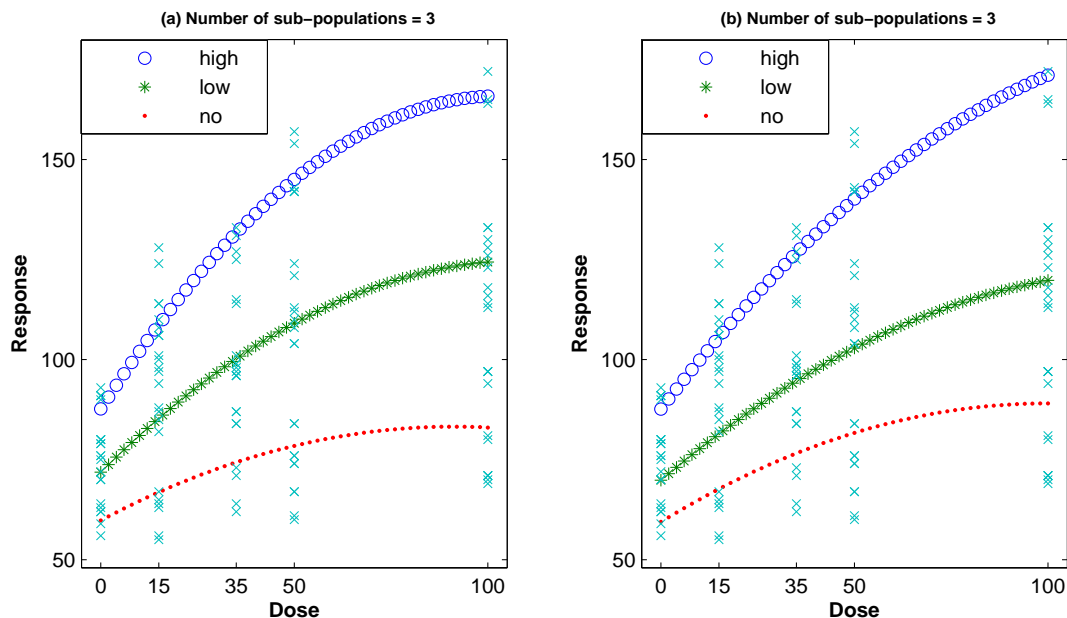


表1 羧基铁数据集的观测剂量反应模式,  $\times$ 代表C5Y5F1黑鼠在所有剂量下的观测葡萄糖水平, 曲线代表拟合的三个模型的剂量反应函数

## 2.2 混合贝叶斯模型

用 $d_i, i = 1, \dots, 5$ 表示第 $i$ 个剂量水平. 假设在 $d_i$ 剂量水平下有 $n_i$ 个受试者,  $X(d)$ 表示剂量水平 $d$ 下感兴趣的定量反应变量葡萄糖水平, 有限混合模型为

$$F(x) = P(X(d) \leq x) = \sum_{k=1}^K \omega_k \Phi \left\{ \frac{x - \mu_k(d)}{\sigma_k} \right\}.$$

$\Phi(\cdot)$ 表示标准正态累积分布函数,  $\mu_k$ 和 $\sigma_k, k = 1, 2, \dots, K$ 分别表示 $K$ 个子总体的均值反应函数和标准差. 另外, 混合系数 $\omega_k, k = 1, 2, \dots, K$ 满足 $\sum_{k=1}^K \omega_k = 1$ .

令 $x_j(d_i), j = 1, 2, \dots, n_i, i = 1, 2, \dots, 5$ 表示在剂量 $d_i$ 下的观测值. 对数似然可以写成

$$LL = \sum_{i=1}^5 \sum_{j=1}^{n_i} \log \left\{ \sum_{k=1}^K \omega_k \cdot \phi \left[ \frac{x_j(d_i) - \mu_k(d_i)}{\sigma_k} \right] \frac{1}{\sigma_k} \right\}.$$

$\phi(\cdot)$ 表示标准正态密度函数. 在生物实验中,  $\mu_k(d_i)$ 一般而言是 $d_i$ 的多项式 $\mu_k(d_i) = f(d_i, \theta'_k)$ ,  $\theta_k = (\theta_{k1}, \dots, \theta_{kS})', k = 1, \dots, K$ 表示参数向量. 本例中令均值函数为 $\mu_k(d) = a_k + b_k d +$

$c_k d^2$ , 则  $\theta_k = (a_k, b_k, c_k)$ . 分析的目的是做如下推断: (1)成分个数  $K$ ; (2)成分参数  $\theta_k, \sigma_k$ ; (3)成分权重  $\omega_k$ , 总和为1.

引进潜在变量  $z_{ij}$ , 表示第  $i$  个剂量下第  $j$  个观测值所来自的总体.  $z_{ij}$  彼此独立, 分布为  $P(Z_{ij} = k) = \omega_k, k = 1, 2, \dots, K$ .  $z_{ij}$  的值给定之后, 观测值的分布为

$$x_j(d_i) | z_{ij} \sim \phi \left\{ \frac{x_j(d_i) - \mu_{z_{ij}}(d_i)}{\sigma_k} \right\} \frac{1}{\sigma_k}.$$

未知参数  $K, \omega, \theta, \sigma$  是从一定的先验分布中抽取的, 所有变量的联合分布

$$p(K, \omega, z, \theta, \sigma, X) = p(K)p(\omega|K)p(z|\omega, K)p(\theta|z, \omega, K)p(\sigma|\theta, z, \omega, K)p(X|\theta, \sigma, z, \omega, K),$$

这里  $p(\cdot|\cdot)$  表示条件分布,  $\omega = (\omega_k)_{k=1}^K, z = (z_{ij})_{i=1, j=1}^{G, n_i}, \theta = (\theta_k)_{k=1}^K, X = (X_j(d_i))_{i=1, j=1}^{G, n_i}$ . 提出进一步的条件独立性, 使得

$$p(\theta|z, \omega, K) = p(\theta|K), \quad p(\sigma|\theta, z, \omega, K) = p(\sigma), \quad p(X|\theta, \sigma, z, \omega, K) = p(X|\theta, \sigma, z).$$

从而联合分布可以简化为

$$p(K, \omega, z, \theta, \sigma, X) = p(K)p(\omega|K)p(z|\omega, K)p(\theta|K)p(\sigma)p(X|\theta, \sigma, z).$$

为了更灵活, 再加一层先验, 使得  $K, \omega, \theta, \sigma$  的先验分别依赖于超参数  $\lambda, \delta, \eta, \zeta$ . 从而所有变量的联合分布为

$$\begin{aligned} & p(\lambda, \delta, \eta, \zeta, K, \omega, z, \theta, \sigma, X) \\ &= p(\lambda)p(\delta)p(\eta)p(\zeta)p(K|\lambda)p(\omega|K, \delta)p(z|\omega, K)p(\theta|K, \eta)p(\sigma|\zeta)p(X|\theta, \sigma, z). \end{aligned}$$

在模型中,  $\theta_k$  表示向量  $(a_k, b_k, c_k), k = 1, \dots, K$ , 从而

$$f(X|\theta_k, \sigma_k^2) = f(X(d)|a_k, b_k, c_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp \left\{ -\frac{(X(d) - \mu_k(d))^2}{2\sigma_k^2} \right\}.$$

$a_k, b_k, c_k, \sigma_k^{-2}$  的先验分布为:  $a_k \sim N(\xi_1, \kappa_1^{-1}), b_k \sim N(\xi_2, \kappa_2^{-1}), c_k \sim N(\xi_3, \kappa_3^{-1}), \sigma_k^{-2} \sim \Gamma(\alpha, \beta), \omega$  的先验取作狄力克莱分布  $\omega \sim D(\delta, \delta, \dots, \delta)$ . 为了方便,  $K$  的先验分布取作1到  $K_{\max}$  之间的均匀分布. 在  $\sigma_k^2$  的先验中, 令  $\beta$  服从Gamma分布  $\Gamma(g, h)$ , 从而引进了又一层次先验.  $\alpha, K_{\max}, \delta, g, h$  取作常数值. 先验分布形式的选取参照Richardson等(1997).

实例分析中, 先验参数的取值如下:  $\xi_1 = 76, \kappa_1^{-1} = 121, \xi_2 = 0.86, \kappa_2^{-1} = 4, \xi_3 = -0.004, \kappa_3^{-1} = 0.0001, \alpha = 2, g = 0.2, h = 0.001, \delta = 1, K_{\max} = 20$ . 这些值的选取是参照Tao等(2004)中模拟结果给出的.

### 2.3 正态混合的可逆跳移动

对于上述层次正态混合模型, 采取下面6种移动类型: (1)更新权重 $\omega$ ; (2)更新参数 $(\theta, \sigma)$ ; (3)更新定位参数 $z$ ; (4)更新超参数 $\beta$ ; (5)一个子成分分裂成两个, 或者把两个子成分合并成一个; (6)空成分的产生和消失.

推断的目标是参数 $\Psi = (\theta, \sigma, \beta, \omega, K)$ . 观测数据为 $x_j(d_i)$ ,  $i = 1, \dots, 5$ ,  $j = 1, \dots, n_i$ .  $(\Psi, z, x)$ 的联合密度为

$$\begin{aligned} f(\Psi, z, x) \\ = p(K)p(\omega|K)p(z|\omega, K)p(\theta|z, \omega, K)p(\sigma|\theta, z, \omega, K)p(\beta|\sigma, \theta, z, \omega, K)p(x|\beta, \sigma, \theta, z, \omega, K). \end{aligned}$$

由于共轭性, 权重参数 $\omega$ 的满条件分布仍然是狄力克莱分布 $\omega|\dots \sim D(\delta + m_1, \delta + m_2, \dots, \delta + m_K)$ . 其中 $m_k = \#\{z_{ij} = k\}$ ,  $i = 1, \dots, 5$ ,  $j = 1, \dots, n_i$ . 本文用 $|\dots$ 表示以所有其它变量为条件的条件分布. 因此 $\omega$ 可以用Gibbs抽样进行更新: 抽取独立的伽玛随机变量, 并通过变换使这些变量的总和为1.

参数 $\theta_k = (a_k, b_k, c_k)$ ,  $k = 1, \dots, K$ 的满条件分布分别为 $a_k|\dots \sim N(\mu_k^a, \sigma_k^a)$ ,  $b_k|\dots \sim N(\mu_k^b, \sigma_k^b)$ ,  $c_k|\dots \sim N(\mu_k^c, \sigma_k^c)$ , 其中均值为

$$\begin{aligned} \mu_k^a &= \left[ \frac{m_k}{\sigma_k^2} + \kappa_1 \right]^{-1} \left[ \frac{\sum_{z_{ij}=k} (x_j(d_i) - b_k d_i - c_k d_i^2)}{\sigma_k^2} + \kappa_1 \xi_1 \right], \\ \mu_k^b &= \left[ \kappa_2 + \frac{\sum_{i,j:z_{ij}=k} d_i^2}{\sigma_k^2} \right]^{-1} \left[ \xi_2 \kappa_2 + \frac{\sum_{i,j:z_{ij}=k} [x_j(d_i) - a_k d_i - c_k d_i^3]}{\sigma_k^2} \right], \\ \mu_k^c &= \left[ \kappa_3 + \frac{\sum_{i,j:z_{ij}=k} d_i^4}{\sigma_k^2} \right]^{-1} \left[ \xi_3 \kappa_3 + \frac{\sum_{i,j:z_{ij}=k} [x_j(d_i) d_i^2 - a_k d_i^2 - b_k d_i^3]}{\sigma_k^2} \right], \end{aligned}$$

方差分别为

$$\sigma_k^a = \left[ \frac{m_k}{\sigma_k^2} + \kappa_1 \right]^{-1}, \quad \sigma_k^b = \left[ \kappa_2 + \frac{\sum_{i,j:z_{ij}=k} d_i^2}{\sigma_k^2} \right]^{-1}, \quad \sigma_k^c = \left[ \kappa_3 + \frac{\sum_{i,j:z_{ij}=k} d_i^4}{\sigma_k^2} \right]^{-1}.$$

参数 $\sigma_k^{-2}$ 的满条件分布为

$$\sigma_k^{-2}|\dots \sim \Gamma \left\{ \alpha + \frac{m_k}{2}, \beta + \frac{1}{2} \sum_{i,j:Z_{ij}=k} (x_j(d_i) - \mu_{z_{ij}}(d_i))^2 \right\},$$

对于定位变量有

$$P(z_{ij} = k|\dots) \sim \frac{\omega_k}{\sigma_k} \exp \left\{ -\frac{(x_j(d_i) - \mu_k(d_i))^2}{2\sigma_k^2} \right\},$$

超参数 $\beta$ 为伽玛分布

$$\beta | \cdots \sim \Gamma\left(K\alpha + g, h + \sum_{k=1}^K \sigma_k^{-2}\right).$$

对于分裂合并移动(5), 需要用到可逆跳机制. 首先分别依概率 $p_k$ 和 $q_k = 1 - p_k$ 在分裂还是合并中随机选择一个. 其中 $q_1 = 0, p_{K_{\max}} = 0$ . 当 $k = 2, 3, \dots, K_{\max} - 1$ 时,  $p_k = q_k = 0.5$ . 当建议合并时, 随机选择一对相邻的子成分 $(k_1, k_2)$ , 即就参数 $\mu_k$ 而言,  $\mu_{k_1} < \mu_{k_2}$ , 在区间 $[\mu_{k_1}, \mu_{k_2}]$ 中没有别的 $\mu_k$ . 这两个成分合并之后, 所形成的新的子成分记作 $k^*$ , 这时需要把满足条件 $z_{ij} = k_1$ 和 $z_{ij} = k_2$ 的观察值 $x_j(d_i)$ 进行重新定位, 并产生新的参数值 $\omega_{k^*}, (a_{k^*}, b_{k^*}, c_{k^*}), \sigma_{k^*}^2$ . 重新定位只需要令 $z_{ij} = k^*$ 即可. 其它参数的更新通过匹配新成分和它所代替的两个旧成分的零阶矩和一阶矩来确定:

$$\left. \begin{aligned} \omega_{k^*} &= \omega_{k_1} + \omega_{k_2}, \\ a_{k^*} &= (\omega_{k_1} a_{k_1} + \omega_{k_2} a_{k_2}) / (\omega_{k_1} + \omega_{k_2}), \\ b_{k^*} &= (\omega_{k_1} b_{k_1} + \omega_{k_2} b_{k_2}) / (\omega_{k_1} + \omega_{k_2}), \\ c_{k^*} &= (\omega_{k_1} c_{k_1} + \omega_{k_2} c_{k_2}) / (\omega_{k_1} + \omega_{k_2}), \\ \sigma_{k^*}^2 &= (\omega_{k_1} \sigma_{k_1}^2 + \omega_{k_2} \sigma_{k_2}^2) / (\omega_{k_1} + \omega_{k_2}). \end{aligned} \right\} \quad (2.1)$$

分裂时, 随机选择成分 $k^*$ 并把它分成两个成分 $k_1$ 和 $k_2$ , 权重和参数根据方程组(2.1)来确定. 用beta分布如下产生随机数 $u_1 \sim \text{be}(2, 2), u_2 \sim \text{be}(2, 2), u_3 \sim \text{be}(2, 2), u_4 \sim \text{be}(2, 2), u_5 \sim \text{be}(2, 2)$ , 并且令

$$\left. \begin{aligned} \omega_{k_1} &= \omega_{k^*} u_1, & \omega_{k_2} &= \omega_{k^*} (1 - u_1), \\ a_{k_1} &= a_{k^*} - u_2 \sqrt{\omega_{k_2} / \omega_{k_1}}, & a_{k_2} &= a_{k^*} + u_2 \sqrt{\omega_{k_1} / \omega_{k_2}}, \\ b_{k_1} &= b_{k^*} - u_3 \sqrt{\omega_{k_2} / \omega_{k_1}}, & b_{k_2} &= b_{k^*} + u_3 \sqrt{\omega_{k_1} / \omega_{k_2}}, \\ c_{k_1} &= c_{k^*} + u_4 \sqrt{\omega_{k_2} / \omega_{k_1}}, & c_{k_2} &= c_{k^*} - u_4 \sqrt{\omega_{k_1} / \omega_{k_2}}, \\ \sigma_{k_1}^2 &= u_5 \sigma_{k^*}^2 \frac{\omega_{k^*}}{\omega_{k_1}}, & \sigma_{k_2}^2 &= (1 - u_5) \sigma_{k^*}^2 \frac{\omega_{k^*}}{\omega_{k_2}}. \end{aligned} \right\} \quad (2.2)$$

方程(2.2)可以给出所有需要的权重和参数, 并且满足方程(2.1). 解出方程之后需验证相邻条件是否满足. 如果不满足就停止这样的转移(因为这个移动即分裂合并对不可逆), 否则, 就需要对满足条件 $Z_{ij} = k^*$ 的观测值 $x_j(d_i)$ 在 $k_1$ 和 $k_2$ 之间进行重新定位. 只需要按照 $z_{ij}$ 在Gibbs定位的满条件分布的公式进行定位即可, 即

$$\begin{aligned} P(Z_{ij} = k_1 | \cdots) &\sim \frac{\omega_{k_1}}{\sigma_{k_1}} \exp \left\{ -\frac{1}{2\sigma_{k_1}^2} (x_j(d_i) - a_{k_1} - b_{k_1} d_i - c_{k_1} d_i^2)^2 \right\}, \\ P(Z_{ij} = k_2 | \cdots) &\sim \frac{\omega_{k_2}}{\sigma_{k_2}} \exp \left\{ -\frac{1}{2\sigma_{k_2}^2} (x_j(d_i) - a_{k_2} - b_{k_2} d_i - c_{k_2} d_i^2)^2 \right\}. \end{aligned}$$

分裂转移活动的接收概率为 $(1, A_s)$ , 其中

$$\begin{aligned} A_s &= \text{似然比} \times \text{先验比} \times \text{建议比} \times \text{雅可比} \\ &= \text{似然比} \times \frac{p(K+1)}{p(K)} (K+1)^3 \frac{\omega_{k_1}^{\delta-1+l_1} \omega_{k_2}^{\delta-1+l_2}}{\omega_{k^*}^{\delta-1+l_1+l_2} B(\delta, K\delta)} \\ &\quad \times \sqrt{\frac{\kappa_1}{2\pi}} \exp \left[ -\frac{1}{2} \kappa_1 \{ (a_{k_1} - \xi_1)^2 + (a_{k_2} - \xi_1)^2 - (a_{k^*} - \xi_1)^2 \} \right] \\ &\quad \times \sqrt{\frac{\kappa_2}{2\pi}} \exp \left[ -\frac{1}{2} \kappa_2 \{ (b_{k_1} - \xi_2)^2 + (b_{k_2} - \xi_2)^2 - (b_{k^*} - \xi_2)^2 \} \right] \\ &\quad \times \sqrt{\frac{\kappa_3}{2\pi}} \exp \left[ -\frac{1}{2} \kappa_3 \{ (c_{k_1} - \xi_3)^2 + (c_{k_2} - \xi_3)^2 - (c_{k^*} - \xi_3)^2 \} \right] \\ &\quad \times \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{\sigma_{k_1}^{-2} \sigma_{k_2}^{-2}}{\sigma_{k^*}^{-2}} \right)^{\alpha-1} \exp(-\beta(\sigma_{k_1}^{-2} + \sigma_{k_2}^{-2} - \sigma_{k^*}^{-2})) \\ &\quad \times \frac{q_{K+1}}{p_K P_{\text{alloc}}} \{g_{2,2}(u_1)g_{2,2}(u_2)g_{2,2}(u_3)g_{2,2}(u_4)g_{2,2}(u_5)\}^{-1} \frac{\omega_{k^*}^2 \sigma_{k^*}^2}{u_1^{3/2} (1-u_1)^{5/2}}. \end{aligned}$$

这里 $K$ 是分裂前的成分个数,  $l_1$ 和 $l_2$ 是重新定位后分配到 $k_1$ 和 $k_2$ 的观测值的个数,  $B(\cdot, \cdot)$ 是贝塔函数,  $P_{\text{alloc}}$ 是选择这种定位的概率,  $g_{p,q}$ 表示 $\text{beta}(p, q)$ 的密度函数, 似然比是 $f(x_j(d_i) | \theta_{Z_{ij}})$ 中新参数相对于旧参数的比值.

合并的接收概率为 $\min(1, A_c^{(-1)})$ ,  $A_c^{(-1)}$ 的形式与 $A_s$ 类似, 只需要做相应的替换即可.

生或死移动(6)相对简单. 首先随机选择是产生空成分还是使空成分消失, 用与上面分裂合并一样的概率 $p_k$ 和 $q_k$ . 对于一个新成分的产生, 其权重和参数用下式产生 $\omega_{k^*} \sim \text{be}(1, K)$ ,  $a_{k^*} \sim N(\xi_1, \kappa_1^{-1})$ ,  $b_{k^*} \sim N(\xi_2, \kappa_2^{-1})$ ,  $c_{k^*} \sim N(\xi_3, \kappa_3^{-1})$ ,  $\sigma_{k^*}^{-2} \sim \Gamma(\alpha, \beta)$ . 为了给新成分腾出空间, 当前的权重需要做一些调整,  $\omega'_k = \omega_k(1 - \omega_{k^*})$ , 从而使得生成新的空成分后所有的权重之和为1. 生成新的空成分的接收概率为 $\min(1, A_b)$ ,

$$A_b = \frac{p(K+1)}{p(K)} (K+1)^3 \frac{1}{B(K\delta, \delta)} \omega_{k^*}^{\delta-1} (1 - \omega_{k^*})^{N+K\delta-K} \times \frac{q_{K+1}}{p_k(k_0+1)} \frac{1}{g_{1,K}(\omega_{k^*})} (1 - \omega_{k^*})^K.$$

这里 $k_0$ 是当前的空成分数目,  $K$ 是当前的成分数目, 也即新的空成分产生之前的数目.

对于空成分的消失, 首先要在已有的空成分中随机选择一个, 然后剔除掉选中的空成分, 重新调整剩下的所有成分的权重,  $\omega'_k = \omega_k / (1 - \omega_{k^*})$ , 从而使得权重之和等于1. 空成分消失的接收概率为 $\min(1, A_d^{-1})$ ,  $A_d^{-1}$ 的形式与 $A_b$ 类似, 只需做相应替换.

### §3. 分析结果

在RJCMCMC分析中, 预热阶段为5000次迭代. 我们采用了预热后的5000次迭代的结果进行分析. 在链的实现过程当中, 子成分个数从未超过15, 因此 $K_{\max}$ 的选取是合理的.

为了检查链的稳定性, 我们给出了 $a_2$ 取不同初始值时累积后验均值的取值曲线, 这些值画在图2中, 从中可以看出, 在获得后验均值的稳定性方面5000次预热是充分的. 在图3中, 我们给出了10000次迭代的 $\hat{R}$ 的图形, 也即Gelman和Rubin (1992)给出的EPSR. 这个图形再次确定了这个链是接近于收敛的, 最后一次迭代中 $\hat{R}$ 的值为1.0105.

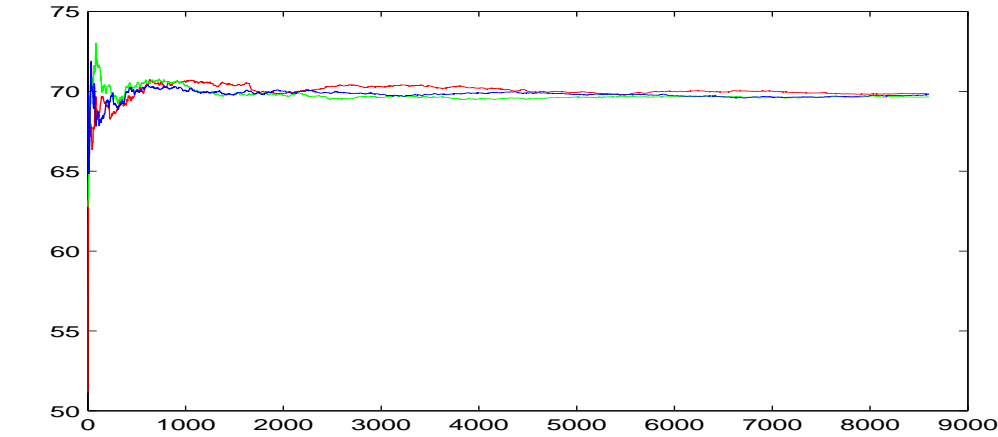


图2 三种不同的初始值下成分参数 $a_2$ 的后验均值轨迹, 表明接近收敛

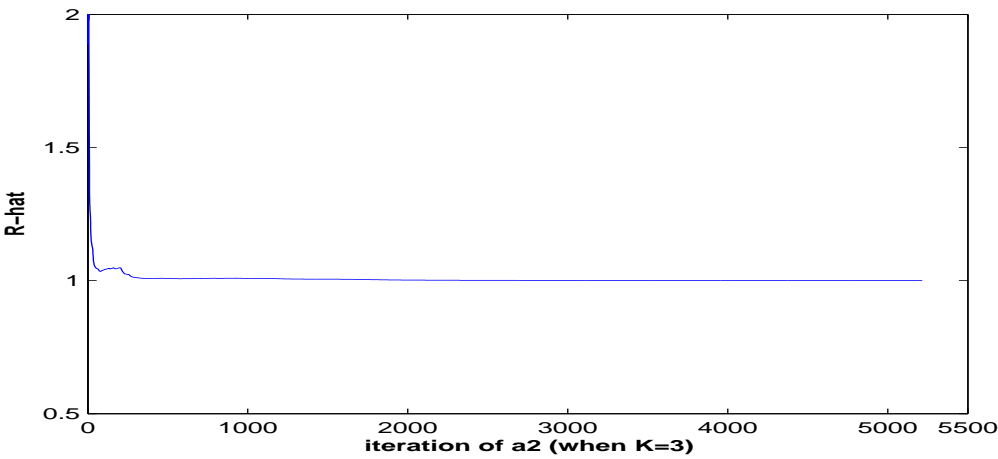


图3  $\hat{R}$ 的序列值, 非常接近1, 表明接近收敛.  $\hat{R}$ 的最后一次迭代的最终值为1.0105.

表1 当参数 $\beta$ 随机其它参数取默认值时 $K$ 的后验分布

$n$	$p(k x)$	移动被接受的比例(%)	
		分裂-合并	生-死
118	$p(1) = 0.02$	0.14	0.04
	$p(2) = 0.14$		
	$p(3) = 0.72$		
	$p(4) = 0.06$		
	$p(5) = 0.04$		
	$p(6) = 0.02$		



在表1中给出了 $K$ 的后验概率. 从中可以看到 $K$ 的后验支持3个子成分. 在表2中, 给出了相应参数的估计值. 这些估计值包括后验均值和置信区间. 置信区间是95%的等尾的区间, 它的区间端点分别是参数边际后验分布的2.5和97.5百分位数.

表2 每个剂量水平下有3个子总体时参数的估计值和0.95置信区间

参数	$a_1$	$a_2$	$a_3$	$b_1$
估计值	59.446	69.8540	87.6355	0.5940
置信区间	(38.7354,81.2149)	(47.9473,89.6234)	(70.2193,106.9325)	(0.1876,1.1567)
参数	$b_2$	$b_3$	$c_1$	$c_2$
估计值	0.8226	1.2668	-0.002980	-0.003232
置信区间	(0.2815,1.8857)	(0.1789,2.5121)	(-0.0067,0.0023)	(-0.0271,0.0043)
参数	$c_3$	$\omega_1$	$\omega_2$	$\omega_3$
估计值	-0.004325	0.2157	0.3427	0.4416
置信区间	(-0.0124,0.0089)	(0.1219,0.5734)	(0.1229,0.7064)	(0.2043,0.8264)
参数	$\sigma_1$	$\sigma_2$	$\sigma_3$	
估计值	42.3674	24.1454	28.7576	
置信区间	(34.1203,50.0718)	(17.1029,32.8967)	(20.9657,35.6623)	

§4. 结 论

在这篇文章中, 我们利用可以处理多个子总体的未知混合结构的RJMCMC方法联合考虑了参数的估计问题和子成分数目的确定问题. 在假设不同子总体(易感模式)的分散程度不同的情况下, 实验C5YSF1黑鼠数据的分析表明这个方法在实际中是非常有意义的. 从图1中可以看到, 易感模式的均值反应函数是单调的且是上凸的, 具备实际意义, 与Liu等(2011)中的结果很相似. 另外, 在本文中, 不同剂量处子总体的数目设为相同主要是为了方便和清晰. 事实上, 即使在不同剂量处子总体的数目不同也可以应用此方法.

参 考 文 献

[1] Liu, R.Y., Tao, J., Shi, N.Z. and He, X., Bayesian analysis of the patterns of biological susceptibility via reversible jump MCMC sampling, *Computational Statistics & Data Analysis*, **55(3)**(2011), 1498–1508.

[2] Green, P.J., Reversible jump Markov chain Monte carlo computation and Bayesian model determination, *Biometrika*, **82(4)**(1995), 711–732.

[3] Razzaghi, M. and Kodell, R.L., Risk assessment for quantitative responses using a mixture model, *Biometrics*, **56(2)**(2000), 519–527.



- [4] Chen, J.J., Kodell, R.L. and Gaylor, D.W., Risk assessment for non-quantal toxic effects, *Toxicology and Risk Assessment: Principles, Methods, and Applications*, Fan, A.M. and Chang, L.W. (Eds.), Marcel Dekker, New York, 1996, 503–513.
- [5] Richardson, S. and Green, P.J., On Bayesian analysis of mixtures with an unknown number of components (with discussion), *Journal of the Royal Statistical Society: Series B*, **59**(4)(1997), 731–792.
- [6] Tao, J., Shi, N.Z. and Lee, S.Y., Drug risk assessment with determining the number of sub-populations under finite mixture normal models, *Computational Statistics & Data Analysis*, **46**(4)(2004), 661–676.
- [7] Gelman, A. and Rubin, D.B., Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science*, **7**(4)(1992), 457–511.

## Application in Biological Susceptibility Patterns of the Reversible Jump Markov Chain Monte Carlo Sampling

LIU RUIYIN

(College of Mathematics and System Science, Shenyang Normal University, Shenyang, 110034)

In some biological experiments, it is quite common that laboratory subjects may be different in their patterns of susceptibility to a treatment. We need to determine the different patterns of susceptibility. In this paper we model the number of susceptibility's patterns and the parameters jointly, and base inference about these quantities on their posterior probabilities, making use of reversible jump Markov chain Monte Carlo methods that are capable of jumping between the parameter subspaces corresponding to different numbers of components in the mixture. For convenience, we always assume different patterns of susceptibility have common variances. The paper apply the methodology to the analysis of univariate normal mixtures with different variances. The practical significance of the proposed method is illustrated with a dose-response data set.

**Keywords:** Mixture models, reversible jump Markov chain Monte Carlo method, susceptibility patterns.

**AMS Subject Classification:** 62P10.