

左截断右删失数据下几何分布参数多变点的贝叶斯估计*

何朝兵¹, 刘华文²

(1. 安阳师范学院 数学与统计学院, 河南 安阳 455000; 2. 山东大学 数学学院, 济南 250100)

摘要:首先通过添加数据得到了左截断右删失数据下几何分布的完全数据似然函数,然后研究了变点位置和其它参数的满条件分布,接着利用 Gibbs 抽样与 Metropolis-Hastings 算法相结合的 MCMC 方法得到了参数的 Gibbs 样本,把 Gibbs 样本的均值作为各参数的贝叶斯估计,最后进行了随机模拟,试验结果表明各参数贝叶斯估计的精度都较高。

关键词:完全数据似然函数;满条件分布;MCMC 方法;Gibbs 抽样;Metropolis-Hastings 算法

中图分类号:O213.2; O212.8

文献标志码:A

文章编号:1672-6693(2014)04-0100-06

变点分析研究始于 20 世纪 50 年代,自 20 世纪 70 年代初,这一问题引起了一些统计学家的重视,发展了一些变点分析方法^[1-3],如最小二乘法、极大似然法、贝叶斯法和非参数方法等。随着计算机技术的发展、贝叶斯方法的应用越来越广泛,特别是在水文统计^[4-5]和计量经济^[6]等领域。而贝叶斯计算方法中的 MCMC 方法,使变点分析的贝叶斯方法的实际操作变得非常方便。当进行可靠性寿命试验时,经常会出现左截断右删失数据,对左截断右删失模型的研究可参看文献[7-14]。几何分布是一种很重要的离散型寿命分布,它是指数分布的离散化,是特殊的离散型威布尔分布,它在可靠性理论和应用概率模型中有着很广泛的应用,例如,在排队论、植物种群物种的调查和库存需求分布等方面,文献[15-18]对几何分布的特征和参数估计进行了研究。而关于左截断右删失模型中几何分布的参数变点问题却很少有文献研究。下文首先通过添加数据得到了左截断右删失数据下几何分布的完全数据似然函数,然后研究了变点位置和其它参数的满条件分布,接着利用 Gibbs 抽样与 Metropolis-Hastings 算法相结合的 MCMC 方法得到了参数的 Gibbs 样本,把 Gibbs 样本的均值作为各参数的贝叶斯估计,最后进行了随机模拟,试验结果表明各参数贝叶斯估计的精度都较高。

1 离散型分布左截断右删失数据试验模型

设 (X, Y, T) 是一离散型随机变量,寿命变量 X 的分布函数为 $F(x, p) = P(X \leq x)$,分布律为 $f(x, p)$,这里 p 是未知参数; Y 是一右删失随机变量,分布函数为 $G(y)$,分布律为 $g(y)$; T 是一左截断随机变量,分布函数为 $H(t)$,分布律为 $h(t)$,且 Y, T 的分布与参数 p 无关。假定 X, Y, T 是相互独立取正整数的随机变量。对于 n 个受试样品(产品寿命),左截断右删失数据的试验模型是,仅在 $Z_i \geq T_i$ 时得到观察数据 (Z_i, T_i, δ_i) ,而在 $Z_i < T_i$ 下无法得到任何观察值,其中

$$Z_i = X_i \wedge Y_i = \min(X_i, Y_i), \delta_i = I(X_i \leq Y_i), i = 1, 2, \dots, n$$

下面求样本的似然函数。

$$P(Z_i = z_i, T_i = t_i, \delta_i = 1) = f(z_i, p) \bar{G}(z_i - 1) h(t_i), z_i \geq t_i$$

$$P(Z_i = z_i, T_i = t_i, \delta_i = 0) = g(z_i) \bar{F}(z_i, p) h(t_i), z_i \geq t_i$$

$$P(\text{无样本观察值}) = P(Z_i < T_i) = 1 - P(X_i \geq T_i, Y_i \geq T_i) = 1 - \sum_{k=1}^{\infty} h(k) \bar{G}(k-1) \bar{F}(k-1, p) \triangleq u(p)$$

* 收稿日期:2013-09-01 修回日期:2013-10-10 网络出版时间:2014-7-3 23:03

资助项目:国家自然科学基金(No. 61174099); 河南省教育厅自然科学基金(No. 2011B110001)

作者简介:何朝兵,男,讲师,研究方向为概率统计,E-mail:chaobing5@163.com

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20140703.2303.020.html

其中 $\bar{G}=1-G, \bar{F}=1-F$ 。为了研究方便,引入示性变量 $\nu_i=I(\min(X_i, Y_i) \geq T_i), i=1, 2, \dots, n_0$ 。

则基于数据 $\{(z_i, t_i, \delta_i): \nu_i=1, 1 \leq i \leq n\}$ 的似然函数为

$$L(p) = \prod_{i=1}^n \{ [f(z_i, p) \bar{G}(z_i - 1) h(t_i)]^{\nu_i \delta_i} [g(z_i) \bar{F}(z_i, p) h(t_i)]^{\nu_i (1-\delta_i)} [u(p)]^{1-\nu_i} \} =$$

$$A \prod_{i=1}^n \{ [f(z_i, p)]^{\nu_i \delta_i} [\bar{F}(z_i, p)]^{\nu_i (1-\delta_i)} \} [u(p)]^{n - \sum_{i=1}^n \nu_i}$$

其中 $A = [h(t_i)]^{\sum_{i=1}^n \nu_i} \prod_{i=1}^n \{ [\bar{G}(z_i - 1)]^{\nu_i \delta_i} [g(z_i)]^{\nu_i (1-\delta_i)} \}$, 且 A 与参数 p 无关。

2 左截断右删失数据下几何分布参数的完全数据似然函数

若 X 的分布律为 $P(X=k) = pq^{k-1}, k=1, 2, \dots, 0 < p < 1, q=1-p$, 则称 X 服从参数为 p 的几何分布, 记为 $X \sim Geo(p)$ 。左截断右删失数据的试验模型中, 假设产品寿命 $X \sim Geo(p)$ 。令 δ 表示 δ_i 组成的向量, z 表示 z_i 组成的向量, ν 表示 ν_i 组成的向量。几何分布 $Geo(p)$ 基于数据 $\{(z_i, t_i, \delta_i): \nu_i=1, 1 \leq i \leq n\}$ 的似然函数为

$$L(z, \delta, \nu | p) \propto \prod_{i=1}^n \{ (pq^{z_i-1})^{\nu_i \delta_i} (q^{z_i})^{\nu_i (1-\delta_i)} [u(p)]^{1-\nu_i} \} = p^{\sum_{i=1}^n \nu_i \delta_i} q^{\sum_{i=1}^n \nu_i \delta_i (z_i-1) + \sum_{i=1}^n \nu_i (1-\delta_i) z_i} [u(p)]^{n - \sum_{i=1}^n \nu_i}$$

设 p 的先验分布为 $\pi(p)$, 则

$$L(p | z, \delta, \nu) \propto \pi(p) p^{\sum_{i=1}^n \nu_i \delta_i} q^{\sum_{i=1}^n \nu_i \delta_i (z_i-1) + \sum_{i=1}^n \nu_i (1-\delta_i) z_i} [u(p)]^{n - \sum_{i=1}^n \nu_i} \quad (1)$$

从(1)式可以看出, 一般来说, p 的后验分布是比较复杂的。

下面添加缺损的 X_i 的值, 使(1)式转化为完全数据的似然函数。具体如下: 当 $\nu_i=1, \delta_i=0$ 时, 添加数据 $Z_{1i} = X_i = z_{1i}$ 。在 $X_i > z_i$ 的条件下, Z_{1i} 的条件分布律为

$$P(X_i = x | X_i > z_i) = pq^{x-1-z_i}, x = z_i + 1, z_i + 2, \dots$$

Z_{1i} 的条件分布是截断几何分布, 可以通过逆变换法随机产生 Z_{1i} 的值 z_{1i} 。当 $\nu_i=0$ 时, 添加数据 $Z_{2i} = X_i = z_{2i}$ 。在 $\nu_i=0$, 即 $\min(X_i, Y_i) < T_i$ 的条件下, Z_{2i} 的条件分布律为

$$P(X_i = x | \min(X_i, Y_i) < T_i) = [u(p)]^{-1} pq^{x-1} [\bar{G}(x) \bar{H}(x) + \sum_{k=1}^x g(k) \bar{H}(k)] \triangleq \phi(x, p)$$

则可由上面的分布随机抽取 1 个随机数作为 Z_{2i} 的值 z_{2i} 。

令 u_1 表示 z_{1i} 组成的向量, u_2 表示 z_{2i} 组成的向量, 则完全数据似然函数为

$$L(z, u_1, u_2, \nu, \delta | p) \propto \prod_{i=1}^n \{ (pq^{z_i-1})^{\nu_i \delta_i} (pq^{z_{1i}-1})^{\nu_i (1-\delta_i)} (pq^{z_{2i}-1})^{1-\nu_i} \} = (pq^{-1})^n q^{\sum_{i=1}^n \nu_i \delta_i z_i + \sum_{i=1}^n \nu_i (1-\delta_i) z_{1i} + \sum_{i=1}^n (1-\nu_i) z_{2i}}$$

注 如果 Y 服从 $Geo(p_y), T$ 服从 $Geo(p_t)$, 令 $q_y=1-p_y, q_t=1-p_t$, 则

$$u(p) = 1 - \sum_{k=1}^{\infty} (p_t q_t^{k-1}) q_y^{k-1} q^{k-1} = \frac{q_t(1-qq_y)}{1-qq_y q_t}$$

$$\phi(x, p) = [u(p)]^{-1} pq^{x-1} [\bar{G}(x) \bar{H}(x) + \sum_{k=1}^x g(k) \bar{H}(k)] = \left[\frac{q_t(1-qq_y)}{1-qq_y q_t} \right]^{-1} pq^{x-1} [q_y^x q_t^x + \sum_{k=1}^x p_y q_y^{k-1} q_t^k] =$$

$$\frac{1-qq_y q_t}{(1-q_y q_t)(1-qq_y)} pq^{x-1} [p_y + p_t q_y (q_y q_t)^{x-1}] = \alpha pq^{x-1} + \beta (1-qq_y q_t) (qq_y q_t)^{x-1}$$

其中 $\alpha = \frac{p_y(1-qq_y q_t)}{(1-q_y q_t)(1-qq_y)}, \beta = \frac{p_t p_y q_y}{(1-q_y q_t)(1-qq_y)}$, 且 $\alpha + \beta = 1$ 。

显然 pq^{x-1} 与 $(1-qq_y q_t) (qq_y q_t)^{x-1}$ 都是几何分布的分布律, 所以可以利用合成法来产生 Z_{2i} 的值。

3 左截断右删失数据下几何分布参数多变点的贝叶斯估计

几何分布参数多变点模型如下

$$X_i \sim \begin{cases} \text{Geo}(p_1), i=1, \dots, k_1 \\ \text{Geo}(p_2), i=k_1+1, \dots, k_2 \\ \text{Geo}(p_3), i=k_2+1, \dots, n \end{cases}$$

其中 p_1, p_2, p_3 两两不相等, $1 \leq k_1 < k_2 \leq n-1$ 。

下面对左截断右删失数据下几何分布的变点位置 k_1, k_2 及参数 p_1, p_2, p_3 进行估计。

令 $D_1 = \{1, 2, \dots, k_1\}, D_2 = \{k_1+1, \dots, k_2\}, D_3 = \{k_2+1, \dots, n\}$ 。

记 $\theta = (k_1, k_2, p_1, p_2, p_3)$, 则此变点问题的似然函数为

$$L(z, u_1, u_2, \nu, \delta | \theta) \propto [(p_1 q_1^{-1})^{k_1} q_1^{s_1}] [(p_2 q_2^{-1})^{k_2 - k_1} q_2^{s_2}] [(p_3 q_3^{-1})^{n - k_2} q_3^{s_3}]$$

其中 $s_m = \sum_{i \in D_m} [\nu_i \delta_i z_i + \nu_i (1 - \delta_i) z_{1i} + (1 - \nu_i) z_{2i}], m = 1, 2, 3$

下面确定各参数的先验分布。

1) 对于 (k_1, k_2) 取无信息先验分布: $\pi(k_1, k_2) = \frac{1}{C_{n-1}^2} = \frac{2}{(n-1)(n-2)}, 1 \leq k_1 < k_2 \leq n-1$

2) 对于 (p_1, p_2, p_3) 也取无信息先验分布: $\pi(p_1, p_2, p_3) = 1, 0 < p_1, p_2, p_3 < 1$

假设 (k_1, k_2) 与 (p_1, p_2, p_3) 独立, 则

$$L(\theta | z, u_1, u_2, \nu, \delta) \propto \pi(k_1, k_2) \pi(p_1, p_2, p_3) L(z, u_1, u_2, \nu, \delta | \theta) \propto L(z, u_1, u_2, \nu, \delta | \theta)$$

$$\text{当 } \nu_i = 1, \delta_i = 0 \text{ 时, } \pi(z_{1i} | \theta, z, z_{-1i}, u_2, \nu, \delta) \propto \varphi_1(z_{1i}; \theta, z_i) = \begin{cases} p_1 q_1^{z_{1i} - 1 - z_i}, i=1, 2, \dots, k_1 \\ p_2 q_2^{z_{1i} - 1 - z_i}, i=k_1+1, \dots, k_2 \\ p_3 q_3^{z_{1i} - 1 - z_i}, i=k_2+1, \dots, n \end{cases}$$

$$\text{当 } \nu_i = 0 \text{ 时, } \pi(z_{2i} | \theta, z, u_1, z_{-2i}, \nu, \delta) \propto \varphi_2(z_{2i}; \theta) = \begin{cases} \psi(z_{2i}, p_1), i=1, 2, \dots, k_1 \\ \psi(z_{2i}, p_2), i=k_1+1, \dots, k_2 \\ \psi(z_{2i}, p_3), i=k_2+1, \dots, n \end{cases}$$

其中 $z_{-1i} = \{z_{1j}; j \neq i\}, z_{-2i} = \{z_{2j}; j \neq i\}$ 。

为了书写方便, 把满条件分布中的“条件”用“ \cdot ”代替, 例如 $\pi(p_1 | k_1, k_2, p_2, p_3, z, u_1, u_2, \nu, \delta)$ 简记为 $\pi(p_1 | \cdot)$ 。各参数的满条件分布为 $\pi(p_1 | \cdot) \propto p_1^{k_1} q_1^{s_1 - k_1} \propto Be(k_1 + 1, s_1 - k_1 + 1)$,

$$\pi(p_2 | \cdot) \propto Be(k_2 - k_1 + 1, s_2 - (k_2 - k_1) + 1), \pi(p_3 | \cdot) \propto Be(n - k_2 + 1, s_3 - (n - k_2) + 1)$$

$$\pi(k_1 | \cdot) \propto [p_1 q_2 (p_2 q_1)^{-1}]^{k_1} q_1^{s_1} q_2^{s_2}, 1 \leq k_1 < k_2, \pi(k_2 | \cdot) \propto [p_2 q_3 (p_3 q_2)^{-1}]^{k_2} q_2^{s_2} q_3^{s_3}, k_1 < k_2 \leq n-1$$

由于得到了各参数的满条件分布, 下面利用 MCMC 方法获得各参数后验分布的平稳分布。参数 z_{1i} 的满条件分布是截断几何分布, p_1, p_2, p_3 的满条件分布是贝塔分布, z_{2i} 的满条件分布与截断删失变量的分布有关, 但一般不会太复杂, 所以这 5 个分布都可以采用 Gibbs 抽样; 但是 k_1, k_2 的满条件分布带有参数约束条件, 并且分布比较复杂, 进行 Gibbs 抽样比较困难, 可以利用 Metropolis-Hastings 算法进行抽样, 此时建议选取的分布为离散型均匀分布。

下面写出 MCMC 方法的具体步骤。

在给出起始点 $\theta^{(0)} = (k_1^{(0)}, k_2^{(0)}, p_1^{(0)}, p_2^{(0)}, p_3^{(0)})$ 后, 假定第 t 次迭代开始时的估计值为 $\theta^{(t-1)}$, 则第 t 次迭代分为如下几步:

1) 当 $\nu_i = 1, \delta_i = 0$ 时, 由分布 $\varphi_1(z_{1i}; \theta^{(t-1)}, z_i)$ 抽取 $z_{1i}^{(t)}$, 令 $u_1^{(t)}$ 表示 $z_{1i}^{(t)}$ 组成的向量;

2) 当 $\nu_i = 0$ 时, 由分布 $\varphi_2(z_{2i}; \theta^{(t-1)})$ 抽取 $z_{2i}^{(t)}$, 令 $u_2^{(t)}$ 表示 $z_{2i}^{(t)}$ 组成的向量;

3) 由 $\pi(p_1 | k_1^{(t-1)}, k_2^{(t-1)}, p_2^{(t-1)}, p_3^{(t-1)}, z, u_1^{(t)}, u_2^{(t)}, \nu, \delta)$ 抽取 $p_1^{(t)}$;

4) 由 $\pi(p_2 | k_1^{(t-1)}, k_2^{(t-1)}, p_1^{(t)}, p_3^{(t-1)}, z, u_1^{(t)}, u_2^{(t)}, \nu, \delta)$ 抽取 $p_2^{(t)}$;

5) 由 $\pi(p_3 | k_1^{(t-1)}, k_2^{(t-1)}, p_1^{(t)}, p_2^{(t)}, z, u_1^{(t)}, u_2^{(t)}, \nu, \delta)$ 抽取 $p_3^{(t)}$;

6) $k_1^{(t)} \sim \pi(k_1 | k_2^{(t-1)}, p_1^{(t)}, p_2^{(t)}, p_3^{(t)}, z, u_1^{(t)}, u_2^{(t)}, \nu, \delta) \triangleq \pi(k_1 | \cdot)$, 建议选择分布 $q(k_1^{(t-1)}, k_1')$ 为取值 $1, 2, \dots, k_2^{(t-1)} - 1$ 的离散型均匀分布, 即 $q(k_1^{(t-1)}, k_1') = 1 / (k_2^{(t-1)} - 1)$, 令

$$\alpha(k_1^{(t-1)}, k_1') = \min\{\pi(k_1' | \cdot) / \pi(k_1^{(t-1)} | \cdot), 1\}$$

从 $1, 2, \dots, k_2^{(t-1)} - 1$ 中任意抽取一个 k_1' , 然后产生一个随机数 u , 若 $u \leq \alpha(k_1^{(t-1)}, k_1')$, 则 $k_1^{(t)} = k_1'$, 否则 $k_1^{(t)} = k_1^{(t-1)}$ 。

7) $k_2^{(t)} \sim \pi(k_2 | k_1^{(t)}, p_1^{(t)}, p_2^{(t)}, p_3^{(t)}, z, u_1^{(t)}, u_2^{(t)}, \nu, \delta) \triangleq \pi(k_2 | \cdot)$, 建议选择分布 $q(k_2^{(t-1)}, k_2')$ 为取值 $k_1^{(t)} + 1, k_1^{(t)} + 2, \dots, n$ 的离散型均匀分布, 即 $q(k_2^{(t-1)}, k_2') = 1 / (n - k_1^{(t)})$, 令

$$\alpha(k_2^{(t-1)}, k_2') = \min\{\pi(k_2' | \cdot) / \pi(k_2^{(t-1)} | \cdot), 1\}$$

从 $k_1^{(t)} + 1, k_1^{(t)} + 2, \dots, n$ 中任意抽取一个 k_2' , 然后产生一个随机数 u , 若 $u \leq \alpha(k_2^{(t-1)}, k_2')$, 则 $k_2^{(t)} = k_2'$, 否则 $k_2^{(t)} = k_2^{(t-1)}$ 。这样就得到了 k_1, k_2, p_1, p_2, p_3 的一个联合样本 $(k_1^{(t)}, k_2^{(t)}, p_1^{(t)}, p_2^{(t)}, p_3^{(t)})$, 称之为 Gibbs 样本。重复这样的 t 步迭代 M 次, 便可得到 M 个独立同分布的 5 维随机样本。

设 $(k_1^{(j)}, k_2^{(j)}, p_1^{(j)}, p_2^{(j)}, p_3^{(j)})$, $j = 1, 2, \dots, M_0, \dots, M$ 为参数 k_1, k_2, p_1, p_2, p_3 的一个容量为 M 的 Gibbs 样本, 其中 M_0 为由于不稳定而舍弃的样本容量, 然后利用达到平稳状态的 $M - M_0$ 个独立样本的均值作为各参数的贝叶斯估计, 即

$$\hat{k}_m = \frac{1}{M - M_0} \sum_{j=M_0+1}^M k_m^{(j)}, m = 1, 2; \hat{p}_i = \frac{1}{M - M_0} \sum_{j=M_0+1}^M p_i^{(j)}, i = 1, 2, 3$$

4 随机模拟

基于上面的讨论, 下面进行随机模拟试验。取受试样品的个数 $n = 500, X_i \sim Geo(0.2), i = 1, \dots, 100; X_i \sim Geo(0.4), i = 101, \dots, 350; X_i \sim Geo(0.7), i = 351, \dots, 500$ 。右删失变量 $Y_i \sim Geo(0.6)$, 左截断变量 $T_i \sim Geo(0.9)$ 。则参数 $(k_1, k_2, p_1, p_2, p_3)$ 的真实值为 $(100, 350, 0.2, 0.4, 0.7)$ 。

表 1 参数 k_1, k_2, p_1, p_2, p_3 的贝叶斯估计

参数	真值	均值	相对误差	MC 误差	2.5%分位数	中位数	97.5%分位数	Gibbs 样本
k_1	100	90.854 9	0.091 5	0.077 2	79	94	103	10 000
k_2	350	342.220 5	0.022 2	0.155 4	332	339	363	10 000
p_1	0.2	0.174 7	0.126 5	0.000 2	0.128 5	0.174 2	0.224 1	10 000
p_2	0.4	0.446 4	0.116 0	0.000 3	0.398 9	0.445 9	0.497 6	10 000
p_3	0.7	0.670 2	0.042 6	0.000 3	0.603 7	0.669 7	0.737 9	10 000

利用前面推导出的各个参数的满条件分布使用 R 软件进行 MCMC 模拟, 重点估计分析位置参数 k_1, k_2 。在模拟运行过程中, 先进行 20 000 次 Gibbs 预迭代, 以确保参数的收敛性, 然后丢弃最初的预迭代, 再进行 10 000 次 Gibbs 迭代。迭代从第 20 001 次开始至第 30 000 次的 R 程序的运行结果见表 1。

位置参数的 Gibbs 抽样迭代过程见图 1 和图 2。

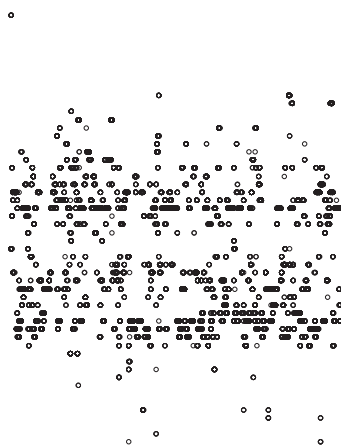


图 1 参数 k_1 的 Gibbs 抽样迭代过程

图 2 参数 k_2 的 Gibbs 抽样迭代过程

在模型的分析过程中, MCMC 收敛性诊断很重要, 模拟时不能简单地通过大量迭代作为预迭代, 而可以对参数进行多层链式迭代分析, 即输入多组初始值, 形成多层迭代链, 当参数收敛时, 迭代图形重合。在模拟过程中, 输入两组初始值分别进行 20 000 次迭代, 位置参数的多层迭代链轨迹见图 3 和图 4。

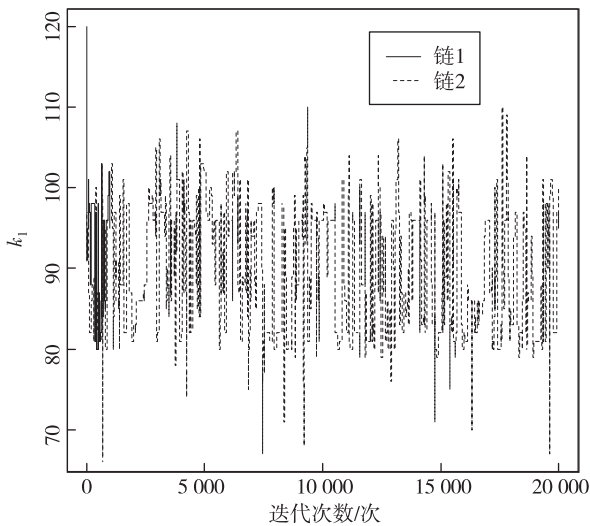


图 3 k_1 的多层迭代链轨迹

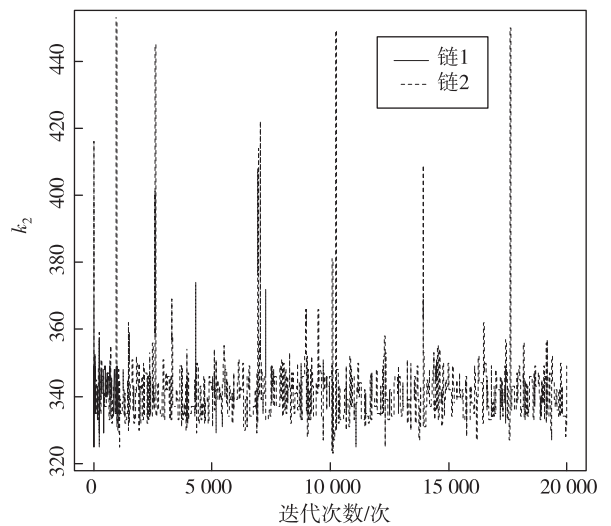


图 4 k_2 的多层迭代链轨迹

最后, 进行模拟结果分析。首先, 由表 1 可以看出位置参数的估计值与真值的相对误差不超过 10%, 其它参数的估计值与真值的相对误差不超过 13%。整体上估计的精度较高, 效果较好。其次, 要判断所产生的马尔科夫链是否收敛, 由图 3 和图 4 可以发现, 位置参数的两条迭代链都趋于重合, 收敛性较好。综上分析, 可以看出通过 MCMC 模拟所产生的效果较好。

注 编写 R 程序时用到的函数主要有 `rgeom()`, `rbeta()`, `min()`, `runif()`, `apply()`, `sum()`, `mean()`, `sd()`, `quantile()`, `plot()`, `lines()`, `legend()` 等。

参考文献:

- [1] Csörgő M, Horváth L. Limit Theorems in Change-Point Analysis[M]. New York: Wiley, 1997.
- [2] 陈希孺. 变点统计分析简介[J]. 数理统计与管理, 1991, 10(1): 55-59.
Chen X R. Introduction of change-point statistical analysis [J]. Journal of Applied Statistics and Management, 1991, 10(1): 55-59.
- [3] 项静恬, 史久恩. 非线性系统中数据处理的统计方法[M]. 北京: 科学出版社, 1997.
Xiang J T, Shi J E. The statistical method of data processing in nonlinear system[M]. Beijing: Science Press, 1997.
- [4] 熊立华, 周芬. 水文时间序列变点分析的贝叶斯方法[J]. 水电能源科学, 2003, 21(4): 39-41.
Xiong L H, Zhou F. Bayesian method for detecting change-points of hydrological time series[J]. Water Resources and Power, 2003, 21(4): 39-41.
- [5] Perreault L, Bernier J, Bobée B, et al. Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited[J]. Journal of Hydrology, 2000, 235(3): 221-241.
- [6] Kotz S, 吴喜之. 现代贝叶斯统计学[M]. 北京: 中国统计出版社, 2000.
Kotz S, Wu X Z. Modern bayesian statistics[M]. Beijing: China Statistics Press, 2000.
- [7] Balakrishnan N, Mitra D. Likelihood inference for lognormal data with left truncation and right censoring with an illustration[J]. Journal of Statistical Planning and Inference, 2011, 141(11): 3536-3553.
- [8] Cosslett S R. Efficient semiparametric estimation of censored and truncated regressions via a smoothed self-consistency equation[J]. Econometrica, 2004, 72(4): 1277-1293.
- [9] Gross S T, Lai T L. Nonparametric estimation and regression analysis with left-truncated and right-censored data [J]. Journal of the American Statistical Association, 1996,

91(435):1166-1180.

- [10] Molanes-lopez E M, Cao R, Keilegom I V A N. Smoothed empirical likelihood confidence intervals for the relative distribution with left-truncated and right-censored data [J]. Canadian Journal of Statistics, 2010, 38(3):453-473.
- [11] Gijbels I, Wang J L. Strong representations of the survival function estimator for truncated and censored data with applications[J]. Journal of Multivariate Analysis, 1993, 47(2):210-229.
- [12] Hudgens M G, Satten G A, Longini I M. Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation[J]. Biometrics, 2001, 57(1):74-80.
- [13] Hyde J. Testing survival under right censoring and left truncation[J]. Biometrika, 1977, 64(2):225-230.
- [14] Pan W, Chappell R. Computation of the NPMLE of distribution functions for interval censored and truncated data with applications to the Cox model[J]. Computational Statistics & Data Analysis, 1998, 28(1):33-50.
- [15] Ferguson T S. A characterization of the geometric distribution[J]. The American Mathematical Monthly, 1965, 72(3):256-260.
- [16] Barreto-Souza W, de Morais A L, Cordeiro G M. The Weibull-geometric distribution[J]. Journal of Statistical Computation and Simulation, 2011, 81(5):645-657.
- [17] Best D J, Rayner J C W. Tests of fit for the geometric distribution[J]. Communications in Statistics-Simulation and Computation, 2003, 32(4):1065-1078.
- [18] Lui K J. Estimation of the geometric survival distribution [J]. IEEE Transactions on Reliability, 1997, 46(1):134-141.

Bayes Estimation of Parameter of Geometric Distribution with Multiple Change Points for Truncated and Censored Data

HE Chaobing¹, LIU Huawen²

(1. School of Mathematics and Statistics, Anyang Normal University, Anyang Henan 455000;

2. School of Mathematics, Shandong University, Jinan 250100, China)

Abstract: This paper firstly obtains the complete-data likelihood function of geometric distribution for truncated and censored data, by adding data, then studies the full conditional distributions of change-point positions and other parameters, and gets Gibbs samples of the parameters by MCMC method of Gibbs sampling together with Metropolis-Hastings algorithm, and takes the means of Gibbs samples as Bayes estimations of the parameters. Finally random simulation tests are conducted, and the results show that Bayes estimations of the parameters are all fairly accurate.

Key words: complete-data likelihood function; full conditional distribution; MCMC method; Gibbs sampling; Metropolis-Hastings algorithm

(责任编辑 李若溪)