Pure Mathematics 理论数学, 2022, 12(10), 1757-1764
Published Online October 2022 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2022.1210190

Hans汉斯

Lindley分布参数变点的贝叶斯估计

赵孟茹*，周菊玲#

新疆师范大学数学科学学院，新疆 乌鲁木齐

收稿日期：2022年9月19日；录用日期：2022年10月18日；发布日期：2022年10月26日

摘　要

利用贝叶斯方法研究了Lindley分布参数存在变点的参数估计问题，给出Lindley分布的变点模型，对参数选取无信息先验分布和伽玛分布两种情况，分别求出各参数的满条件分布，并通过R软件做随机模拟，得出各参数的MC误差都小于2%，且区间估计效果理想，表明通过贝叶斯估计研究各参数的估计值是有效的。

关键词

Lindley分布，变点，M-H抽样，贝叶斯估计

Bayesian Estimation of Parameter Change Points of Lindley Distribution

Mengru Zhao*, Juling Zhou#

School of Mathematical Sciences, Xinjiang Normal University, Urumqi Xinjiang

Received: Sep. 19th, 2022; accepted: Oct. 18th, 2022; published: Oct. 26th, 2022

Abstract

The parameter estimation problem of Lindley distribution with change points is studied by using Bayesian method. The change point model of Lindley distribution is given. The full conditional distribution of each parameter is calculated for the two cases of no information prior distribution and gamma distribution when the parameters are selected. The random simulation by R software shows that the MC error of each parameter is less than 2%, and the interval estimation effect is ideal. It shows that the estimation of each parameter by Bayesian estimation is effective.

*第一作者。
#通讯作者。

文章引用: 赵孟茹, 周菊玲. Lindley 分布参数变点的贝叶斯估计[J]. 理论数学, 2022, 12(10): 1757-1764.
DOI: 10.12677/pm.2022.1210190

Keywords

Lindley Distribution, Change Point, M-H Sampling, Bayes Estimation

1. 引言

Lindley 分布是可靠性研究中的一个重要分布，某些寿命数据可通过 Lindley 模型达到更好的拟合效果[1]。Krishna 等在逐步 II 型右删失数据下，采用极大似然方法和贝叶斯方法研究了 Lindley 分布的可靠性[2]。杨冬霞等分别在完全数据、逐步 I 型区间删失数据，逐步 II 型删失数据以及定数截尾样本下研究了 Lindley 分布的参数估计问题[3] [4] [5]；范梓淼等分别讨论了在 NA 随机样本序列和独立同分布样本下 Lindley 分布参数的经验贝叶斯检验函数问题[6] [7]；龙兵分析了 Lindley 分布参数的区间估计和假设检验问题[8]；近几年，变点问题也是统计方向研究的一个热点问题。何朝兵等在左截断右删失数据下对指数分布多变点模型进行了参数估计[9]。沙雪云等利用贝叶斯方法研究了 Lomax 分布形状参数变点的估计模型[10]；程静等用极大似然估计和贝叶斯估计讨论了两种分布的单变点问题[11] [12]。关于 Lindley 分布参数变点问题的研究较少，本文给出了 Lindley 分布的变点模型，在叙述了解决多变点模型问题的具体步骤后，主要研究 Lindley 分布参数的单变点模型，分别在无信息先验分布和伽玛分布为先验分布的条件下，利用贝叶斯估计研究参数和变点位置，并通过 R 软件进行随机模拟。结果显示：各参数的估计值和真实值之间的 MC 误差较小，表明其估计值的效果较为理想。

2. Lindley 分布变点模型

设随机变量X服从参数为θ的 Lindley 分布，则分布函数和密度函数如下：

$$F(x;\theta)=1-\left(1+\frac{\theta}{\theta+1}x\right)\exp(-\theta x),\ x>0.$$

$$f(x;\theta)=\frac{\theta^2}{\theta+1}(1+x)\exp(-\theta x),\ x>0.$$

其中参数$\theta>0$。

Lindley 分布多变点模型为：

$$X_i\sim\begin{cases}\text{Lindley}(\theta_1), i=1,2,\cdots,k_1\\ \text{Lindley}(\theta_2), i=k_1+1,\cdots,k_2\\ \vdots\\ \text{Lindley}(\theta_{m+1}), i=k_m+1,\cdots,n.\end{cases}$$

其中$\theta_1,\theta_2,\cdots,\theta_{m+1}$两两不等，$m$是变点个数，$k_1,k_2,\cdots,k_m$ (满足$1\le k_1<k_2<\cdots<k_m\le n-1$)是需要估计的变点位置。通过二分分段法来解决多变点的问题的具体步骤为：先确定 Lindley 分布的序列S中是否存在单变点，如果没有，则序列S中无变点；如果存在单变点，此变点将 Lindley 分布的序列S拆分成两个子序列，再次确定两个子序列中是否存在单变点，重复上述步骤，直至所有子序列中识别不到变点为止。

设随机变量 $X_i\left(i=1,2,\cdots,n\right)$ 相互独立且满足

$$X_i \sim \begin{cases} \text{Lindley}\left(\theta_1\right), & i=1,2,\cdots,k, \\ \text{Lindley}\left(\theta_2\right), & i=k+1,\cdots,n. \end{cases}$$

其中参数 $\theta_1,\theta_2>0, 1\le k\le n-1$ 且 $k\in N^+$， k,θ_1,θ_2 均未知，当 $\theta_1\ne\theta_2$ 时 k 就是要讨论的变点，此模型只含有一个变点，称其为 Lindley 分布的单变点模型。

下文确定各参数的贝叶斯估计，对k取无信息先验分布：$\pi\left(k\right)=\dfrac{1}{n-1}$，对参数 θ_1,θ_2 分别取无信息先验分布和伽玛分布后，再对变点 k 和参数 θ_1,θ_2 做贝叶斯估计。

3. Lindley 分布参数的贝叶斯估计

当 $\theta_1\ne\theta_2$ 时，设 k 是变点，故此变点问题的似然函数为

$$\begin{aligned} L\left(k,\theta_1,\theta_2\mid x\right) &= \frac{\theta_1^2}{\theta_1+1}\left(1+x_1\right)\exp\left(-\theta_1 x_1\right)\cdots\frac{\theta_1^2}{\theta_1+1}\left(1+x_k\right)\exp\left(-\theta_1 x_k\right) \\ &\quad\cdot\frac{\theta_2^2}{\theta_2+1}\left(1+x_{k+1}\right)\exp\left(-\theta_2 x_{k+1}\right)\cdots\frac{\theta_2^2}{\theta_2+1}\left(1+x_n\right)\exp\left(-\theta_2 x_n\right). \\ &= \frac{\theta_1^{2k}}{\left(\theta_1+1\right)^k}\prod_{i=1}^{k}\left(1+x_i\right)\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\frac{\theta_2^{2(n-k)}}{\left(\theta_2+1\right)^{n-k}} \\ &\quad\cdot\prod_{i=k+1}^{n}\left(1+x_i\right)\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right). \end{aligned}$$

1) 通过 Jeffreys 提出的用 Fisher 信息阵来确定 θ_1,θ_2 的无信息先验分布。

样本对数似然函数为：

$$\begin{aligned} l\left(\theta_1,\theta_2\mid x\right) &= 2k\ln\theta_1-k\ln\left(\theta_1+1\right)+\sum_{i=1}^{k}\ln\left(1+x_i\right)-\theta_1\sum_{i=1}^{k}x_i+2\left(n-k\right)\ln\theta_2 \\ &\quad-\left(n-k\right)\ln\left(\theta_2+1\right)+\sum_{i=k+1}^{n}\ln\left(1+x_i\right)-\theta_2\sum_{i=k+1}^{n}x_i. \end{aligned}$$

其中 $x=x_i\left(i=1,2,\cdots,n\right)$。

通过样本对数似然函数可以求得：

$$\frac{\partial l}{\partial\theta_1}=\frac{2k}{\theta_1}-\frac{k}{\theta_1+1}-\sum_{i=1}^{k}x_i,\ \frac{\partial^2 l}{\partial\theta_1^2}=-\frac{2k}{\theta_1^2}+\frac{k}{\left(\theta_1+1\right)^2},\ \frac{\partial^2 l}{\partial\theta_1\partial\theta_2}=0.$$

$$\frac{\partial l}{\partial\theta_2}=\frac{2\left(n-k\right)}{\theta_2}-\frac{n-k}{\theta_2+1}-\sum_{i=k+1}^{n}x_i,\ \frac{\partial^2 l}{\partial\theta_2^2}=-\frac{2\left(n-k\right)}{\theta_2^2}+\frac{n-k}{\left(\theta_2+1\right)^2},\ \frac{\partial^2 l}{\partial\theta_2\partial\theta_1}=0.$$

进而得到 θ_1,θ_2 的无信息先验矩阵为：

$$I\left(\theta_1,\theta_2\right)=E^{x\mid\theta_1,\theta_2}\left(-\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right)=\begin{pmatrix} \dfrac{2k}{\theta_1^2}-\dfrac{k}{\left(\theta_1+1\right)^2} & 0 \\ 0 & \dfrac{2\left(n-k\right)}{\theta_2^2}-\dfrac{n-k}{\left(\theta_2+1\right)^2} \end{pmatrix}.$$

其中 $i,j=1,2$。

故 θ_1,θ_2 的无信息先验分布为：

$$\pi(\theta_1,\theta_2)=\left[\det I(\theta_1,\theta_2)\right]^{\frac{1}{2}}=\left[\left(\frac{2k}{\theta_1^2}-\frac{k}{(\theta_1+1)^2}\right)\left(\frac{2(n-k)}{\theta_2^2}-\frac{n-k}{(\theta_2+1)^2}\right)\right]^{\frac{1}{2}}$$

$$=\left[\frac{k(n-k)\left[2(\theta_1+1)^2-\theta_1^2\right]\left[2(\theta_2+1)^2-\theta_2^2\right]}{\theta_1^2(\theta_1+1)^2\theta_2^2(\theta_2+1)^2}\right]^{\frac{1}{2}}.$$

由贝叶斯公式求得 k,θ_1,θ_2 的联合后验分布为：

$$\pi(k,\theta_1,\theta_2\mid x)\propto L(k,\theta_1,\theta_2\mid x)\pi(k)\pi(\theta_1,\theta_2)$$

$$=\frac{\theta_1^{2k}}{(\theta_1+1)^k}\prod_{i=1}^{k}(1+x_i)\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\cdot\frac{\theta_2^{2(n-k)}}{(\theta_2+1)^{n-k}}\cdot\prod_{i=k+1}^{n}(1+x_i)\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right)$$

$$\cdot\frac{1}{n-1}\cdot\left[\frac{k(n-k)\left[2(\theta_1+1)^2-\theta_1^2\right]\left[2(\theta_2+1)^2-\theta_2^2\right]}{\theta_1^2(\theta_1+1)^2\theta_2^2(\theta_2+1)^2}\right]^{\frac{1}{2}}.$$

各参数满条件分布为：

$$f(\theta_1\mid\theta_2,k,x_i)\propto\frac{\theta_1^{2k}}{(\theta_1+1)^k}\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\cdot\left[\frac{\left[2(\theta_1+1)^2-\theta_1^2\right]}{\theta_1^2(\theta_1+1)^2}\right]^{\frac{1}{2}};$$

$$f(\theta_2\mid\theta_1,k,x_i)\propto\frac{\theta_2^{2(n-k)}}{(\theta_2+1)^{n-k}}\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right)\cdot\left[\frac{\left[2(\theta_2+1)^2-\theta_2^2\right]}{\theta_2^2(\theta_2+1)^2}\right]^{\frac{1}{2}};$$

$$f(k\mid\theta_1,\theta_2,x_i)\propto\frac{\theta_1^{2k}}{(\theta_1+1)^k}\prod_{i=1}^{k}(1+x_i)\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\cdot\frac{\theta_2^{2(n-k)}}{(\theta_2+1)^{n-k}}$$

$$\prod_{i=k+1}^{n}(1+x_i)\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right)\cdot\left[k(n-k)\right]^{\frac{1}{2}}.$$

比较选取均匀分布作为先验分布来说，Jeffreys 提出的用 Fisher 信息阵来确定 θ_1,θ_2 的无信息先验分布在单调变换中具有不变性，能够保证不论采取什么样的参数化方法，它们的先验分布始终是互通的，从而后验分布也是互通的。

2) θ_1,θ_2 的先验分布为伽玛分布 $\left(Ga(b_i,c_i),i=1,2\right)$。

$$\begin{cases}\pi(\theta_1)=Ga(\theta_1\mid b_1,c_1)=\dfrac{c_1^{b_1}}{\Gamma(b_1)}\theta_1^{b_1-1}\mathrm{e}^{-c_1\theta_1},\theta_1>0,b_1>0,c_1>0.\\ \pi(\theta_2)=Ga(\theta_2\mid b_2,c_2)=\dfrac{c_2^{b_2}}{\Gamma(b_2)}\theta_2^{b_2-1}\mathrm{e}^{-c_2\theta_2},\theta_2>0,b_2>0,c_2>0.\end{cases}$$

且 k,θ_1,θ_2 相互独立，由贝叶斯公式得 k,θ_1,θ_2 的联合后验密度为：

$$\pi\left(k,\theta_1,\theta_2\mid x\right)\propto L\left(k,\theta_1,\theta_2\mid x\right)\pi\left(k\right)\pi\left(\theta_1\right)\pi\left(\theta_2\right)$$
$$=\frac{1}{n-1}\cdot\frac{\theta_1^{2k}}{\left(\theta_1+1\right)^k}\prod_{i=1}^{k}\left(1+x_i\right)\cdot\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\frac{\theta_2^{2(n-k)}}{\left(\theta_2+1\right)^{n-k}}\prod_{i=k+1}^{n}\left(1+x_i\right)$$
$$\cdot\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right)\frac{c_1^{b_1}}{\Gamma\left(b_1\right)}\theta_1^{b_1-1}\mathrm{e}^{-c_1\theta_1}\cdot\frac{c_2^{b_2}}{\Gamma\left(b_2\right)}\theta_2^{b_2-1}\mathrm{e}^{-c_2\theta_2}.$$

各参数满条件分布为：

$$f\left(\theta_1\mid\theta_2,k,x_i\right)\propto\frac{\theta_1^{2k}}{\left(\theta_1+1\right)^k}\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\cdot\theta_1^{b_1-1}\mathrm{e}^{-c_1\theta_1};$$

$$f\left(\theta_2\mid\theta_1,k,x_i\right)\propto\frac{\theta_2^{2(n-k)}}{\left(\theta_2+1\right)^{n-k}}\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right)\cdot\theta_2^{b_2-1}\mathrm{e}^{-c_2\theta_2};$$

$$f\left(k\mid\theta_1,\theta_2,x_i\right)\propto\frac{\theta_1^{2k}}{\left(\theta_1+1\right)^k}\prod_{i=1}^{k}\left(1+x_i\right)\exp\left(-\theta_1\sum_{i=1}^{k}x_i\right)\frac{\theta_2^{2(n-k)}}{\left(\theta_2+1\right)^{n-k}}\prod_{i=k+1}^{n}\left(1+x_i\right)\exp\left(-\theta_2\sum_{i=k+1}^{n}x_i\right).$$

4. 随机模拟

在随机模拟过程中，考虑到参数θ_1,θ_2,k的满条件分布比较复杂，因此选用 M-H 算法对各参数的满条件分布进行抽样。接下来介绍 *Markov Chain Monte Carlo* (*MCMC*)算法的几个具体步骤：

设初始点$\alpha^{(0)}=\left(k^{(0)},\theta_1^{(0)},\theta_2^{(0)}\right)$经过迭代后第$t-1$次迭代值为$\alpha^{(t-1)}=\left(k^{(t-1)},\theta_1^{(t-1)},\theta_2^{(t-1)}\right)$，则第 t 次迭代步骤如下：

1) $\theta_1^{(t)}\sim f\left(\theta_1\mid\theta_2,k,x_i\right)$，选取建议分布$q\left(\theta_1^{(t-1)},\theta_1'\right)$为均匀分布，并从中随机抽取$\theta_1'$，令$r\left(\theta_1^{(t-1)},\theta_1'\right)=\min\left\{\frac{\pi\left(\theta_1'\mid\cdot\right)}{\pi\left(\theta_1^{(t-1)}\mid\cdot\right)},1\right\}$，若随机数$u\le r\left(\theta_1^{(t-1)},\theta_1'\right)$，则$\theta_1^{(t)}=\theta_1'$，否则$\theta_1^{(t)}=\theta_1^{(t-1)}$；

2) $\theta_2^{(t)}\sim f\left(\theta_2\mid\theta_1,k,x_i\right)$，获取$\theta_2^{(t)}$与 1)类似；

3) $k^{(t)}\sim f\left(k\mid\theta_1,\theta_2,x_i\right)$，选取建议分布$q\left(k^{(t-1)},k'\right)$为取值$0,1,\cdots,n-1$的离散型均匀分布，并从中随机抽取$k'$，令$r\left(k^{(t-1)},k'\right)=\min\left\{\frac{\pi\left(k'\mid\cdot\right)}{\pi\left(k^{(t-1)}\mid\cdot\right)},1\right\}$，若随机数$u\le r\left(k^{(t-1)},k'\right)$，则$k^{(t)}=k'$，否则$k^{(t)}=k^{(t-1)}$。

设$\left(k^{(j)},\theta_1^{(j)},\theta_2^{(j)}\right),j=1,2,\cdots,B,\cdots,M$为迭代 M 次所得的 Gibbs 样本，若 B 次后迭代逐渐收敛，则将后$M-B$ $M-B$个迭代的均值作为参数k,θ_1,θ_2的估计值，

$$\hat{k}=\frac{1}{M-B}\sum_{t=B+1}^{M}k^{(t)},\ \hat{\theta}_i=\frac{1}{M-B}\sum_{t=B+1}^{M}\theta_i^{(t)},i=1,2.$$

取$n=200$个样本，参数$\left(k,\theta_1,\theta_2\right)$的真实值取$\left(100,3,8\right)$，此时 Lindley 分布的模型为：

$$X_i\sim\begin{cases}\frac{9}{4}\left(1+x\right)\exp\left(-3x\right),i=1,2,\cdots,100\\\frac{64}{9}\left(1+x\right)\exp\left(-8x\right),i=101,102,\cdots,200\end{cases}$$

利用各参数的满条件分布，运用 R 软件进行 MCMC 模拟。为确保参数的收敛性，先进行 10,000 次的预迭代，再进行 20,000 次迭代。结果如下所示：

1) 当θ_1,θ_2选取无信息先验分布时：

Table 1. Bayesian estimation of parameters k,θ_1,θ_2 under uninformative prior distribution
表 1. 无信息先验分布下参数k,θ_1,θ_2的贝叶斯估计

参数	真值	均值	标准差	MC 误差	2.5%分位数	中位数	97.5%分位数
θ_1	3	3.0229	0.2458	0.0164	2.5	3.1	3.4
θ_2	8	8.2129	0.6221	0.0079	6.8	8.3	9
k	100	104.0622	5.7847	0.0041	99.2	102	120

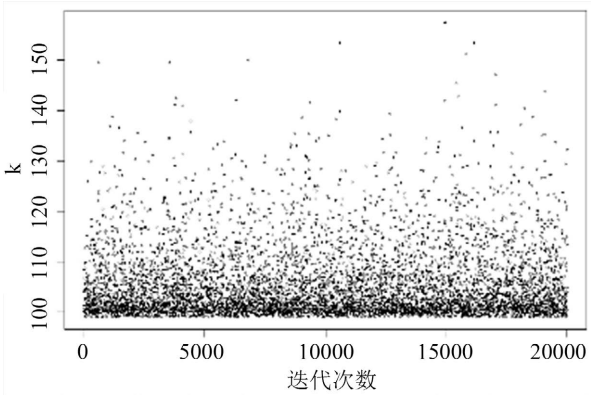

Figure 1. The iteration trajectory of parameter k
图 1. 参数 k 的迭代轨迹

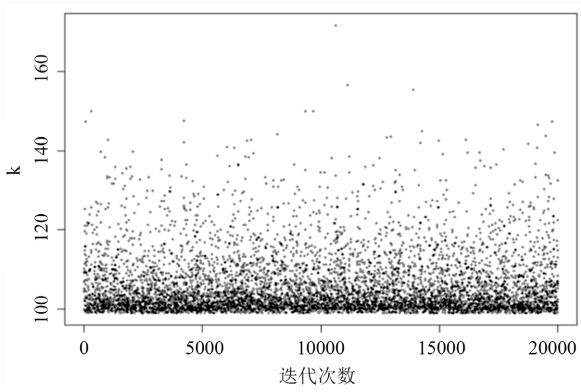

Figure 2. The iteration trajectory of parameter k
图 2. 参数 k 的迭代轨迹

2) 当θ_1,θ_2选取伽玛先验分布时：$\theta_1 \sim Ga(7,3)$，$\theta_2 \sim Ga(9,2)$。

Table 2. Bayesian estimation of parameters k,θ_1,θ_2 under conjugate prior distribution
表 2. 共轭先验分布下参数k,θ_1,θ_2的贝叶斯估计

参数	真值	均值	标准差	MC 误差	2.5%分位数	中位数	97.5%分位数
θ_1	3	3.0316	0.2445	0.0176	2.4	3.1	3.4
θ_2	8	8.1573	0.6327	0.0077	6.7	8.2	9
k	100	104.1521	5.9308	0.0039	99.1	102	120.6

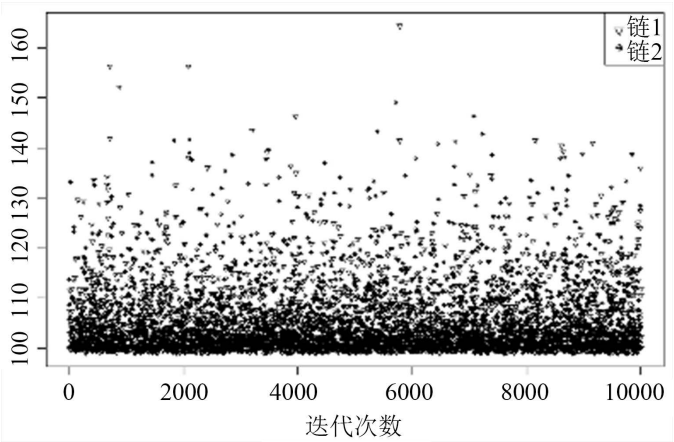

Figure 3. Two iteration trajectories of the parameter k
图 3. 参数 k 的两条迭代图

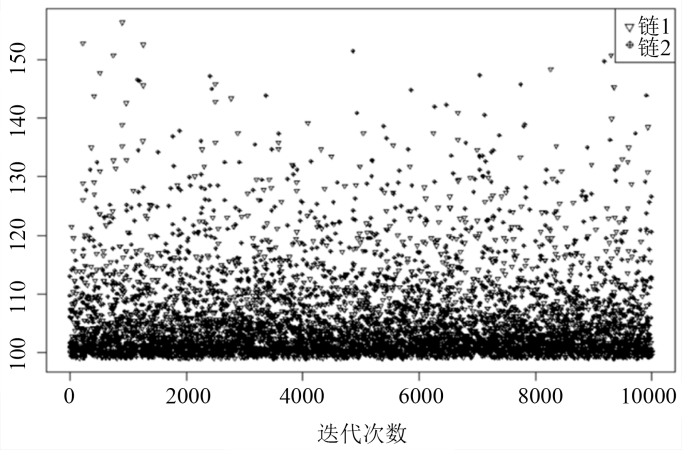

Figure 4. Two iteration trajectories of the parameter k
图 4. 参数 k 的两条迭代轨迹

结果分析：由表 1 和表 2 知，当参数选取不同的先验分布后再进行随机模拟，得到各参数估计值与真实值的 MC 误差均不超过 2%，因此各参数的估计值在较高水平上是有效的；各参数置信水平 0.95 的置信区间[2.5%分位数，97.5%分位数]较窄，说明区间估计效果良好；图 1，图 2 是变点 k 的抽样迭代轨迹，可以根据图上信息判断样本是否收敛。两张图上显示出抽样基本都在变点附近波动，具有一定的规律性；此外，由图 3，图 4 看出 k 的两条 Markov 链趋于重合，具有较好的收敛性。综上可得，Lindley 分布的参数和变点估计可由 MCMC 算法得到较为理想的效果，可用该方法解决 Lindley 分布的变点问题。

基金项目

国家自然科学基金项目(11801488)；新疆师范大学教学研究与改革项目(SDJG2020-30)；新疆师范大学科研发展专项项目(XJNUZX202001)。

参考文献

[1] 龙兵. II 型删失下 Lindley 分布的参数估计(英文) [J]. 湖南师范大学自然科学学报, 2017, 40(6): 71-75.

[2] Krishna, H. and Kumar, K. (2011) Reliability Estimation in Lindley Distribution with Progressively Type II Right Censored Sample. *Mathematics and Computers in Simulation*, **82**, 281-294.

https://doi.org/10.1016/j.matcom.2011.07.005

[3] 杨冬霞. Lindley 分布参数的贝叶斯估计[D]: [硕士学位论文]. 乌鲁木齐: 新疆师范大学, 2020.

[4] 代莹. Lindley 分布的统计分析[D]: [硕士学位论文]. 上海: 上海师范大学, 2018.

[5] 习长新, 刘华. 逐步 II 型删失下 Lindley 分布的参数估计[J]. 新余学院学报, 2017, 22(4): 24-26.

[6] 范梓淼, 周菊玲. NA 样本下 Lindley 分布参数的经验 Bayes 检验[J]. 贵州师范大学学报(自然科学版), 2016, 34(2): 68-70. https://doi.org/10.16614/j.cnki.issn1004-5570.2016.02.014

[7] 杜伟娟, 彭家龙, 李体政. Lindley 分布参数的经验 Bayes 检验的收敛速度[J]. 统计与决策, 2012(21): 23-26. https://doi.org/10.13546/j.cnki.tjyjc.2012.21.011

[8] 龙兵. Lindley 分布中参数的区间估计和假设检验[J]. 广西民族大学学报(自然科学版), 2014, 20(1): 59-62. https://doi.org/10.16177/j.cnki.gxmzzk.2014.01.003

[9] 何朝兵, 刘跃军, 刘华文. 左截断右删失数据下指数分布参数多变点的贝叶斯估计[J]. 西南师范大学学报(自然科学版), 2015, 40(1): 12-17. https://doi.org/10.13718/j.cnki.xsxb.2015.01.003

[10] 沙雪云, 周菊玲, 董翠玲. Lomax 分布形状参数变点的贝叶斯估计[J]. 淮阴师范学院学报(自然科学版), 2020, 19(4): 288-292. https://doi.org/10.16119/j.cnki.issn1671-6876.2020.04.002

[11] 程静, 周菊玲. Weibull 分布尺度参数变点的模型估计[J]. 河南科学, 2022, 40(3): 345-349.

[12] 程贝丽, 周菊玲. 对数伽玛分布变点模型的估计[J]. 淮阴师范学院学报(自然科学版), 2021, 20(2): 95-99. https://doi.org/10.16119/j.cnki.issn1671-6876.2021.02.001