

## 重尾序列均值变点的经验似然比检验\*

王丹\* 皮林

(西北大学数学学院, 西安, 710127)

**摘要:** 本文通过经验似然思想建立假设检验的方法, 研究了重尾序列均值变点的检测问题. 首先, 基于重尾模型, 在原假设和备择假设下得到经验似然函数. 其次, 基于经验似然函数构造似然比检验统计量, 并给出在原假设成立时该似然比统计量的渐近分布. 最后, 进行 Monte Carlo 数值模拟验证该方法的有效性, 模拟结果表明本方法对重尾序列均值变点的检测具有良好效果.

**关键词:** 重尾序列; 均值变点; 经验似然比检验

**中图分类号:** O212.1

**英文引用格式:** WANG D, PI L. Empirical likelihood ratio test for mean change-point in heavy-tailed sequence [J]. Chinese J Appl Probab Statist, 2021, 37(2): 111-122. (in Chinese)

### §1. 引言

统计学中, 变点问题一直是一个热点课题, 一般认为变点问题的研究始于 Page<sup>[1]</sup> 在 *Biometrika* 上发表的一篇关于连续抽样检验的文章. 后来陈希孺院士对变点的一般性定义进行了总结, 陈希孺院士<sup>[2]</sup>指出, 变点即“模型中具体某个或某些参数发生突然变化的点”. 变点检验问题, 就是应用恰当的方法对模型中是否存在变点进行分析.

变点问题自 1954 年由 Page 在工业应用上提出以来, 许多专家学者对其进行了深入的研究. 近些年来, 关于变点问题的理论研究, 其广度和深度都得到了加强和发展, 积累了很多的文献. Wang 和 Wang<sup>[3]</sup> 对长记忆性的无穷阶移动平均模型 (MA) 的方差变点进行了研究. Madurkayová<sup>[4]</sup> 利用 RCUSUM 函数的比率构造 Ratio 统计量的方法研究了渐变序列的变点检测问题. Steinebach 和 Timmermann<sup>[5]</sup> 对随机误差项满足弱不变原理的渐变随机过程的变点进行了研究. Ciuperca<sup>[6]</sup> 通过构造 S-估计量的方法研究了随机误差项为长相依序列的两阶段线形回归模型的变点问题. Dehling 等<sup>[7]</sup> 通过基于两样本的 Wilcoxon 检验方法对长相依序列数据中变点问题进行了讨论. Zou 等<sup>[8]</sup> 针对多变点问题应用非参数最大似然方法进行了研究. Jirak<sup>[9]</sup> 对高维数据的变点检测进行了分析. Aminikhanghahi 和 Cook<sup>[10]</sup> 对时间序列的变点检测进行了研究.

\*国家自然科学基金面上项目 (批准号: 11771353)、陕西省科技计划项目 (自然科学基金基础研究计划一般项目) (批准号: 2018JQ1075) 和陕西省教育厅专项科研计划项目 (批准号: 15JK1737) 资助.

\*通讯作者, E-mail: wangdan@nwu.edu.cn.

本文 2019 年 9 月 2 日收到, 2020 年 4 月 17 日收到修改稿.

在现实生活中,许多金融资产收益率的分布都具有不能由正态分布来准确描述的高尖峰、新息重尾等特性<sup>[11]</sup>.近年来,随着诸如计量经济学等新兴学科的发展,具有尾概率为  $P(|Y_t| > x) \approx Cx^{-\kappa}$  的重尾序列成为专家学者的研究热点<sup>[12]</sup>.比如对短期利率、通货膨胀率、外汇率、股票数据等重尾序列展开研究.由于受政策波动、金融危机、经济泡沫等突发事件的影响,而在某个未知的时刻序列发生突变,从而造成金融资产可能的损失,所以,针对股票数据等重尾序列均值变点的检测就显得尤为重要.然而在实际应用中,重尾指数  $\kappa$  是难以估计的,应用参数方法对重尾序列变点进行检测就不再适用,因为非参数方法对总体的假定较少,因而应用非参数方法对重尾序列是否存在均值变点进行检测就成为有力工具之一.

本文针对重尾时间序列问题,应用经验似然思想对均值变点进行研究.经验似然方法在变点研究方面自从被 Owen<sup>[13]</sup> 提出后,已经广泛被研究者们应用到统计学各种模型中解决实际问题.Chen<sup>[14,15]</sup> 在线性回归模型中应用了此方法.Chen<sup>[16]</sup> 应用此方法研究了非参数统计问题.Einmahl 和 McKeague<sup>[17]</sup> 提出在有且仅有一个变点的情况下并基于经验似然的综合假设检验方法检验独立性和指数性.Guan<sup>[18]</sup> 在半参数问题上应用了经验似然方法.Zhu 和 Xue<sup>[19]</sup> 在部分线性模型中应用了经验似然方法.本文研究的重尾指数  $\kappa$  满足  $\kappa \in (1, 2)$ ,此时重尾序列  $\{Y_t\}$  均值存在,方差不存在.Kokoszka 和 Wolf<sup>[20]</sup> 假设序列  $\{Y_t\}$  为零均值的重尾随机变量过程,并基于此研究了重尾序列中不含变点的均值估计.本文假定重尾序列中含有一个均值变点,然后通过经验似然思想建立假设检验,并在原假设和备择假设下得到经验似然函数,基于经验似然函数构造了经验似然比统计量,并证明经验似然比统计量在原假设成立时的渐近分布.最后,进行 Monte Carlo 数值模拟验证了该方法的有效性,模拟结果表明本方法对重尾序列均值变点的检测具有良好效果.

## §2. 模型与假设

考虑如下模型:

$$X_t = u(t) + Y_t, \quad t = 1, 2, \dots, n,$$

其中  $u(t)$  是非随机函数,序列  $\{Y_t\}$  是一个重尾序列,且满足如下假设.

**假设 1** 随机变量序列  $\{Y_t\}$  是严平稳序列,有一维对称边缘分布,且满足

$$n \times P(Y_1/a_n \in \cdot) \implies u(\cdot),$$

其中  $a_n$  定义为  $n \times P(|e_1| > a_n) \rightarrow 1$ ,符号“ $\implies$ ”表示弱收敛,测度  $u(\cdot)$  定义为

$$2u(dx) = \kappa|x|^{-\kappa-1}I\{x < 0\}dx + \kappa x^{-\kappa-1}I\{x > 0\}dx,$$

其中  $1 < \kappa < 2$ .

**注记 2** 假设是一维边缘分布属于特征指数为  $\kappa$  的稳定分布的吸收域所满足的条件, 在一定条件下, 正则变化分布与 Pareto 分布满足该假设, 而且选取适当的参数, GARCH(1,1) 过程满足该假设, GARCH(1,1) 过程是典型的重尾分布模型.

若假设 1 成立, 则对于给定的数据  $X_t$ , 我们考虑如下假设检验问题:

$$H_0: X_t = u_1 + Y_t, \quad t = 1, 2, \dots, n.$$

$$H_1: X_t = \begin{cases} u_1 + Y_t, & t = 1, 2, \dots, k^*; \\ u_2 + Y_t, & t = k^* + 1, k^* + 2, \dots, n, \end{cases}$$

其中  $u_1, u_2$  为常数且  $u_1 \neq u_2$ ,  $n$  是样本容量. 在原假设下, 重尾序列  $\{X_t\}$  没有均值变点, 即  $X_1, X_2, \dots, X_n$  有共同的连续分布函数  $F$ , 在备择假设下, 重尾序列  $\{X_t\}$  在  $k^*$  处有一个均值变点, 即  $X_1, X_2, \dots, X_{k^*}$  有一个共同的连续分布函数  $F$ ,  $X_{k^*+1}, X_{k^*+2}, \dots, X_n$  有共同的连续分布函数为  $G$ , 分布函数  $F$  和  $G$  未知, 而且  $F \neq G$ .

### §3. 主要结果

假设  $X_1, X_2, \dots, X_n$  是独立的, 令  $\tilde{F}(t)$  和  $\tilde{G}(t)$  分别是样本  $X_1, X_2, \dots, X_{k^*}$  和  $X_{k^*+1}, X_{k^*+2}, \dots, X_n$  的经验分布函数, 则  $k^*\tilde{F}(t) \sim B(k^*, F(t))$ ,  $(n - k^*)\tilde{G}(t) \sim B(n - k^*, G(t))$ . 则在  $H_0$  下的非参数最大似然函数为

$$\tilde{L}_{H_0}(t) = [\tilde{F}_n(t)]^{n\tilde{F}_n(t)} \cdot [1 - \tilde{F}_n(t)]^{n[1-\tilde{F}_n(t)]},$$

同理,  $H_1$  下的非参数最大似然函数为

$$\tilde{L}_{H_1}(t) = [F(t)]^{k^*\tilde{F}(t)} \cdot [1 - \tilde{F}(t)]^{k^*[1-\tilde{F}(t)]} \cdot [\tilde{G}(t)]^{(n-k^*)\tilde{G}(t)} \cdot [1 - \tilde{G}(t)]^{(n-k^*)[1-\tilde{G}(t)]},$$

其中  $\tilde{F}_n(t)$  是  $X_1, X_2, \dots, X_n$  的经验分布函数,  $\tilde{F}_n(t) = [k^*\tilde{F}(t) + (n - k^*)\tilde{G}(t)]/n$ .

定义经验似然比为

$$\tilde{R}(t) = \frac{\tilde{L}_{H_0}(t)}{\tilde{L}_{H_1}(t)}, \quad (1)$$

对 (1) 式两边同时取对数有

$$\ln \tilde{R}(t) = \ln \tilde{L}_{H_0}(t) - \ln \tilde{L}_{H_1}(t), \quad (2)$$

其中

$$\ln \tilde{L}_{H_0}(t) = n\{\tilde{F}_n(t) \ln \tilde{F}_n(t) + [1 - \tilde{F}_n(t)] \ln [1 - \tilde{F}_n(t)]\}, \quad (3)$$

$$\begin{aligned} \ln \tilde{L}_{H_1}(t) &= k^*\{\tilde{F}(t) \ln \tilde{F}(t) + [1 - \tilde{F}(t)] \ln [1 - \tilde{F}(t)]\} \\ &\quad + (n - k^*)\{\tilde{G}(t) \ln \tilde{G}(t) + [1 - \tilde{G}(t)] \ln [1 - \tilde{G}(t)]\}, \end{aligned} \quad (4)$$

把  $\tilde{F}_n(t) = [k^*\tilde{F}(t) + (n - k^*)\tilde{G}(t)]/n$  带入 (3) 式和 (4) 式并结合 (2) 式得到对数经验似然比为

$$\begin{aligned} \ln \tilde{R}(t) = & k^* \left\{ \tilde{F}(t) \ln \frac{\tilde{F}_n(t)}{\tilde{F}(t)} + [1 - \tilde{F}(t)] \ln \frac{1 - \tilde{F}_n(t)}{1 - \tilde{F}(t)} \right\} \\ & + (n - k^*) \left\{ \tilde{G}(t) \ln \frac{\tilde{F}_n(t)}{\tilde{G}(t)} + [1 - \tilde{G}(t)] \ln \frac{1 - \tilde{F}_n(t)}{1 - \tilde{G}(t)} \right\}, \end{aligned} \quad (5)$$

则构造经验似然比检验统计量为

$$\tilde{T}_n = -2 \int_{-\infty}^{\infty} \ln \tilde{R}(t) d\tilde{F}_n(t) = -\frac{2}{n} \sum_{i=1}^n \ln \tilde{R}(X_i),$$

则提出如下定理.

**定理 3** 在原假设  $H_0$  下, 令  $F$  是连续的函数, 且假设当  $n \rightarrow \infty$  时,  $k, n - k \rightarrow \infty$ , 则

$$\tilde{T}_n \xrightarrow{D} \int_0^1 \frac{B^2(y)}{y(1-y)} dy,$$

其中符号 “ $\xrightarrow{D}$ ” 表示依分布收敛,  $B$  是一个标准布朗桥.

**注记 4** 定理 3 是一个一般性结论, 本文提出的是非参数的处理均值变点问题的方法, 所以除了本文的重尾序列外, 对一些类似真实分布未知的复杂数据均值变点问题, 也可用此检验统计量进行处理.

欲证明定理 3, 需要给出如下两个引理.

**引理 5** (Helly-Bray 定理) 令  $F$  和  $F_1, F_2, \dots$  是实数域上的累积分布函数, 若  $F_n$  弱收敛于  $F$ , 则当  $n \rightarrow \infty$  时, 对每一个连续函数  $g$  都有

$$\int_{\mathcal{R}} g(x) dF_n(x) \rightarrow \int_{\mathcal{R}} g(x) dF(x).$$

以下引理 6 来自文献 [21] 中第 86 页的第 18 个表达式.

**引理 6** [21] 令  $Y_n(y) = \sqrt{n}[\Psi_n(y) - y]$ , 则

$$\sqrt{n}[\Psi_n(y) - y] \xrightarrow{D} B(y), \quad y \in [0, 1],$$

其中  $\Psi_n$  是  $U_i, i = 1, 2, \dots, n$  的经验分布函数,  $U_i = F(X_i)$ , 符号 “ $\xrightarrow{D}$ ” 表示依分布收敛,  $B$  是一个标准布朗桥.

**定理 3 的证明:** 令  $U_i = F(X_i)$ ,  $\Psi_n, \Psi_{1n}$  和  $\Psi_{2n}$  分别是  $U_i, i = 1, 2, \dots, n; U_i, i = 1, 2, \dots, k^*$  和  $U_i, i = k^* + 1, k^* + 2, \dots, n$  的经验分布函数. 令  $0 < \varepsilon < 1/2$ , 要使定理 3 成立, 我们只需证当  $n \rightarrow \infty$  时, 有

$$\tilde{T}_{1n} = -2 \int_{\varepsilon}^{1-\varepsilon} \ln \tilde{R}(y) d\Psi_n(y) \xrightarrow{D} \int_{\varepsilon}^{1-\varepsilon} \frac{B^2(y)}{y(1-y)} dy \quad (6)$$

和

$$\tilde{T}_{2n} = \tilde{T}_n - \tilde{T}_{1n} = O_p(\sqrt{\varepsilon}) \quad (7)$$

在  $\varepsilon$  上一致成立.

首先考虑  $\tilde{T}_{1n}$ , 即证 (6) 式, 在  $\varepsilon < y < 1 - \varepsilon$  时, 通过泰勒展开, 有

$$-2 \ln \tilde{R}(y) = \frac{k^* [\Psi_n(y) - \Psi_{1n}(y)]^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} + \frac{(n - k^*) [\Psi_n(y) - \Psi_{2n}(y)]^2}{\Psi_{2n}(y)[1 - \Psi_{2n}(y)]} + o_p(1),$$

在原假设  $H_0$  成立的条件下, 有  $\Psi_{1n}(y) = \Psi_{2n}(y)$ , 故

$$-2 \ln \tilde{R}(y) = \frac{n [\Psi_n(y) - \Psi_{1n}(y)]^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} + o_p(1),$$

所以对 (6) 式有

$$\begin{aligned} \tilde{T}_{1n} &= -2 \int_{\varepsilon}^{1-\varepsilon} \ln \tilde{R}(y) d\Psi_n(y) \\ &= \int_{\varepsilon}^{1-\varepsilon} \left\{ \frac{n [\Psi_n(y) - \Psi_{1n}(y)]^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} + o_p(1) \right\} d\Psi_n(y) \\ &= \int_{\varepsilon}^{1-\varepsilon} \frac{n [\Psi_n(y) - \Psi_{1n}(y)]^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} d\Psi_n(y) \\ &= \int_{\varepsilon}^{1-\varepsilon} \frac{\{\sqrt{n} [\Psi_n(y) - \Psi_{1n}(y)]\}^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} d\Psi_n(y), \end{aligned}$$

由引理 6 有  $\Psi_{1n}(y) \xrightarrow{D} y$ ,  $\Psi_{2n}(y) \xrightarrow{D} y$ . 结合引理 5 和引理 6,

$$\tilde{T}_{1n} = \int_{\varepsilon}^{1-\varepsilon} \frac{\{\sqrt{n} [\Psi_n(y) - \Psi_{1n}(y)]\}^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} d\Psi_n(y) \xrightarrow{D} \int_{\varepsilon}^{1-\varepsilon} \frac{B^2(y)}{y(1-y)} dy,$$

即 (6) 式得证.

现在证 (7) 式, 欲证明 (7) 式, 需要给出如下两个引理, 以下引理 7 来自文献 [21] 中第 415-416 页的表达式.

**引理 7** [21] 令  $0 < b_n \leq 1/2$ , 当  $n \rightarrow \infty$  时,  $nb_n \rightarrow \infty$ , 则对任意的  $\delta > 0$ , 都有

$$P \left\{ \left[ \sup_{b_n \leq y \leq 1} \frac{\Psi_{1n}(y)}{y} \right] \vee \left[ \sup_{b_n \leq y \leq 1} \frac{y}{\Psi_{1n}(y)} \right] > 1 + \delta \right\} \rightarrow 0.$$

由文献 [22] 中定理 3 的简单推导可得如下引理 8.

**引理 8** [22] 在引理 7 的条件下, 存在一个标准的布朗桥序列  $\{B_n\}$  使得

$$\sup_{b_n \leq y \leq 1-b_n} \frac{|Y_n(y) - B_n(y)|}{[y(1-y)]^{1/4}} \xrightarrow{P} 0,$$

其中符号 “ $\xrightarrow{P}$ ” 表示依概率收敛.

现在证明 (7) 式. 只考虑在  $0 < y \leq 1/2$  的区域. 我们将区域  $0 < y \leq 1/2$  分成几个子区域, 首先考虑区域  $1/n^{3/8} \leq y \leq 1/2$ , 由泰勒展开和引理 7 可知, 在此区域内对于很大的  $n$  一致有

$$\begin{aligned} |\ln \tilde{R}(y)| &\leq k^* \frac{[\Psi_n(y) - \Psi_{1n}(y)]^2}{\Psi_{1n}(y)[1 - \Psi_{1n}(y)]} + (n - k^*) \frac{[\Psi_n(y) - \Psi_{2n}(y)]^2}{\Psi_{2n}(y)[1 - \Psi_{2n}(y)]} \\ &= \frac{k^*(n - k^*)^2 [\Psi_{1n}(y) - \Psi_{2n}(y)]^2}{n^2 \Psi_{1n}(y)[1 - \Psi_{1n}(y)]} + \frac{k^{*2}(n - k^*) [\Psi_{1n}(y) - \Psi_{2n}(y)]^2}{n^2 \Psi_{2n}(y)[1 - \Psi_{2n}(y)]}. \end{aligned}$$

我们只考虑求和的第一项, 第二项可类似处理. 由引理 7 可知, 第一项以很高的概率在区域上一致有上界,

$$\frac{2Y_n^2(y)}{y} \frac{y}{\Psi_{1n}(y)} \leq \frac{3Y_n^2(y)}{y} = 3 \left[ \frac{Y_n(y)}{y^{1/4}} \right]^2 \frac{1}{y^{1/2}}.$$

由引理 8, 有

$$\begin{aligned} \int_{1/n^{3/8}}^{1/2} \left[ \frac{Y_n^2(y)}{y^{1/4}} \right]^2 \frac{1}{y^{1/2}} d\Psi_n(y) &\stackrel{D}{=} \int_{1/n^{3/8}}^{1/2} \left[ \frac{B(y)}{y^{1/4}} \right]^2 \frac{1}{y^{1/2}} d\Psi_n(y) + o_p(1) \\ &= O_p(1) \int_0^\varepsilon \frac{1}{y^{1/2}} dy + o_p(1) \\ &= O_p(\sqrt{\varepsilon}), \end{aligned}$$

同理, 可知求和第二项的阶为  $O_p(\sqrt{\varepsilon})$ . 对于其他区域类似可证, 即 (7) 式得证.  $\square$

## §4. 数值模拟

### 1) 检验的临界值

本节我们给出模拟结果, 通过数值模拟来说明重尾序列均值变点的似然比检验. 由文献 [23] 模拟可得到经验似然比检验统计量  $\tilde{T}_n$  的渐近分布  $\int_0^1 \{B^2(y)/[y(1-y)]\} dy$  在不同样本容量  $n$  下各分位数的对应值. 如表 1 所示, 各分位数的对应值即为原假设下的临界值  $T_n$ , 若  $\tilde{T}_n > T_n$ , 则拒绝原假设.

表 1  $H_0$  下似然比统计量渐近分布的各分位数对应值

样本量	90%	95%
$n = 500$	1.958	2.630
$n = 1000$	1.893	2.499
$n = 1500$	2.039	2.609
$n = 2000$	1.982	2.520

## 2) 检验的经验水平和功效函数

考虑如下数据生成过程

$$X_t = \begin{cases} u_1 + Y_t, & t = 1, 2, \dots, k^*; \\ u_2 + Y_t, & t = k^* + 1, k^* + 2, \dots, n, \end{cases}$$

假设检验问题:

$$H_0: u_1 = u_2; \quad H_1: u_1 \neq u_2,$$

其中  $u_1 = 0$ ,  $u_2 = 0, 0.2, 0.35, 0.5, 0.8$ , 若  $u_2 = 0$ , 则序列不存在变点, 若  $u_2 = 0.2, 0.35, 0.5, 0.8$ , 则序列存在结构变点, 令  $k^* = [n\tau]$ ,  $\tau$  分别取 0.2, 0.5, 0.8, 其中显著性水平  $\alpha = 0.05, 0.10$ , 样本容量  $n = 500, 1000, 1500, 2000$ . 误差序列  $\{Y_t\}$  是满足假设 1 的重尾序列. 重尾序列由 Nolan 教授程序生成, 具体可见 Nolan 教授个人主页.

对样本容量为  $n$  的数据分别重复实验 2000 次, 检验的经验水平和功效函数值用实验 2000 次中拒绝原假设的频率来近似. 模拟结果分别见表 2、表 3、表 4 和表 5.

表 2  $H_0$  下的经验水平值

$n$	$\kappa = 1.46$		$\kappa = 1.97$	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
500	0.0425	0.0990	0.0505	0.1060
1000	0.0570	0.1185	0.0510	0.1045
1500	0.0435	0.0860	0.0430	0.0890
2000	0.0475	0.0945	0.0540	0.0910

表 3  $H_1$  下的功效函数值,  $\alpha = 0.05, \tau = 0.2$

$n$	$\kappa = 1.46$				$\kappa = 1.97$			
	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0.5$	$u_2 = 0.8$	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0.5$	$u_2 = 0.8$
500	0.1850	0.4550	0.7720	0.9940	0.2080	0.5360	0.8355	0.9990
1000	0.3440	0.7960	0.9740	1.0000	0.4155	0.8620	0.9905	1.0000
1500	0.4605	0.9195	0.9980	1.0000	0.5480	0.9505	0.9995	1.0000
2000	0.5925	0.9685	0.9995	1.0000	0.6860	0.9920	1.0000	1.0000

通过模拟可知的主要结论:

(a) 从表 2 可以看出, 在没有变点的情况下, 经验水平值是接近于显著性水平  $\alpha$  的. 且样本容量  $n$  越大, 经验水平值越接近于显著性水平  $\alpha$ , 说明水平失真越小, 检验有效.

(b) 从表 3、表 4 和表 5 了解到, 对于一个固定的数  $n$ , 随着  $u_2$  的不断增大, 功效函数值也相应地不断增大, 对于一个固定的  $u_2$ , 功效函数值随着样本容量  $n$  的增大而增大, 随着样本容量  $n$  和  $u_2$  的增加, 功效函数值逼近 1.

表 4  $H_1$  下的功效函数值,  $\alpha = 0.05, \tau = 0.5$ 

$n$	$\kappa = 1.46$				$\kappa = 1.97$			
	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0.5$	$u_2 = 0.8$	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0.5$	$u_2 = 0.8$
500	0.2550	0.6600	0.9260	0.9990	0.3085	0.7460	0.9655	1.0000
1000	0.4935	0.9350	0.9970	1.0000	0.5880	0.9625	1.0000	1.0000
1500	0.6270	0.9820	0.9995	1.0000	0.7385	0.9965	1.0000	1.0000
2000	0.7715	0.9955	1.0000	1.0000	0.8680	1.0000	1.0000	1.0000

表 5  $H_1$  下的功效函数值,  $\alpha = 0.05, \tau = 0.8$ 

$n$	$\kappa = 1.46$				$\kappa = 1.97$			
	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0.5$	$u_2 = 0.8$	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0.5$	$u_2 = 0.8$
500	0.1950	0.4685	0.7650	0.9865	0.2130	0.5225	0.8400	0.9985
1000	0.3405	0.7690	0.9720	1.0000	0.4130	0.8555	0.9885	1.0000
1500	0.4455	0.9040	0.9980	1.0000	0.5175	0.9510	1.0000	1.0000
2000	0.5900	0.9670	1.0000	1.0000	0.6770	0.9865	1.0000	1.0000

(c) 结合表 2、表 3、表 4 和表 5,  $\kappa = 1.97$  的结果优于  $\kappa = 1.46$ , 这是因为特征指数  $\kappa$  刻画的是重尾序列的特性: 指数  $\kappa$  越小, 取到“奇异点”的概率越大。

### 3) 经验似然比方法与极大似然比方法比较

本节我们给出经验似然比方法与极大似然比方法分别对正态分布数据与重尾分布数据的均值变点检验效果比较。

极大似然比方法是经典的均值变点检验方法, 该方法是参数方法, 极大似然比统计量如下:

$$\Lambda_k = \frac{\sup_{(u, \delta^2) \in \Theta_0} \prod_{i=1}^n f(x_i; u, \delta^2)}{\sup_{(u_1, u_2, \delta^2) \in \Theta} \prod_{i=1}^k f(x_i; u_1, \delta^2) \prod_{i=k+1}^n f(x_i; u_2, \delta^2)},$$

数据服从正态分布, 其中  $u_1, u_2$  是正态分布的均值,  $\delta^2$  是正态分布的方差,  $n$  为样本容量, 该方法采取如下的检验统计量:

$$Z_n = \max_{1 \leq k \leq n-1} (-2 \ln \Lambda_k),$$

当  $Z_n$  较大时, 则拒绝原假设, 认为序列存在结构变点, 否则, 接受原假设。

由文献 [24] 可知, 该检验统计量的上侧  $\alpha$  分位数为

$$c_n(\alpha) = \frac{D_d(\ln n) - \ln[-\ln(1 - \alpha)] + \ln 2}{2 \ln \ln n},$$



其中  $D_d(x) = 2 \ln x + (d/2) \ln \ln x - \ln \Gamma(d/2)$ ,  $d$  是未知参数的维数, 本文中  $d = 1$ .

该上侧  $\alpha$  分位数即为原假设下检验的临界值,  $n = 2000$ ,  $\alpha = 0.05$ , 可计算出临界值为 3.1352. 在本小节中, 数据生成过程与前文类似, 其中  $u_1 = 0$ , 重尾数据的尾指数  $\kappa = 1.46$ , 正态数据方差为 1.

类似地, 我们对正态分布数据与重尾分布数据分别重复检验 2000 次, 可得经验水平值和功效函数值. 下表中,  $u_2 = 0$ , 在不存在变点的情形下, 比较了两种方法的经验水平值.  $u_2 = 0.1, 0.2, 0.35$ ,  $\tau = 0.5$ , 在有变点的情形下, 比较了两种方法的功效函数值.

表 6 两种方法的经验水平值与功效函数值,  $\alpha = 0.05$ ,  $\tau = 0.5$

	极大似然比方法				经验似然比方法			
	$u_2 = 0$	$u_2 = 0.1$	$u_2 = 0.2$	$u_2 = 0.35$	$u_2 = 0$	$u_2 = 0.1$	$u_2 = 0.2$	$u_2 = 0.35$
重尾数据	0.0670	0.1050	0.2090	0.4280	0.0475	0.2700	0.7715	0.9955
正态数据	0.0695	0.6795	0.9975	1.0000	0.0415	0.5665	0.9940	1.0000

通过模拟可知的主要结论:

(a) 从表 6 可以看出, 在没有变点的情况下, 相较于极大似然比方法, 经验似然比方法的经验水平值更接近于显著性水平值  $\alpha$ , 说明在没有变点的情形下, 经验似然比方法的检验效果较极大似然比方法的检验效果有微弱优势.

(b) 从表 6 可以了解到, 在有变点的情况下, 当数据为重尾分布数据时, 经验似然比方法模拟出的功效函数值比极大似然比方法模拟出的功效函数值都要大且更接近于 1; 反之, 当数据服从正态分布时, 极大似然比方法模拟出的功效函数值比经验似然比方法模拟出的功效函数值更接近于 1, 说明对于变点问题, 当模型是参数模型时, 极大似然比方法的检验效果更有效, 当模型是非参数模型时, 本文的经验似然比方法的检验效果更有效.

#### 4) 实例分析

在现实生活中, 许多金融资产都具有重尾特性, 本节用一组股票数据对重尾序列均值变点的经验似然比检验进行实例分析, 此数据是鲁北化工自 2004 年 7 月 9 日至 2005 年 12 月 16 日共计 351 个股票收盘价 (数据取自 <http://finance.sina.com.cn/realstock/company/sh600727/nc.shtml>). 韩四儿<sup>[25]</sup>已经验证这组数据是重尾指数  $\kappa = 1.81$  的重尾分布.

根据图 1, 我们可以看到股票价格在  $k = 100$  附近有一个明显的变化, 因此我们猜测在这个位置附近可能存在一个结构变点. 首先通过经验似然比检验统计量  $\tilde{T}_n$  的渐近分布  $\int_0^1 \{B^2(y)/[y(1-y)]\} dy$  在样本容量下的经验分位数确定临界值, 其次计算原统计量并和临界值作比较, 最后我们发现原统计量的值大于临界值, 因此拒绝原假设, 即该组股票数据存在一个均值变点 ( $\hat{k} = 120$ , 参见文献 [25]). 现在我们将这 351 个股票价格的观测值分为前后两个部分, 分别把样本前 120 个观测值和后 231 个观测值称为鲁北化工股票 1 和鲁北化工股票 2, 图 1、图 2 和图 3 分别表明了这些股票的实际价格.



图 1 鲁北化工股票价格

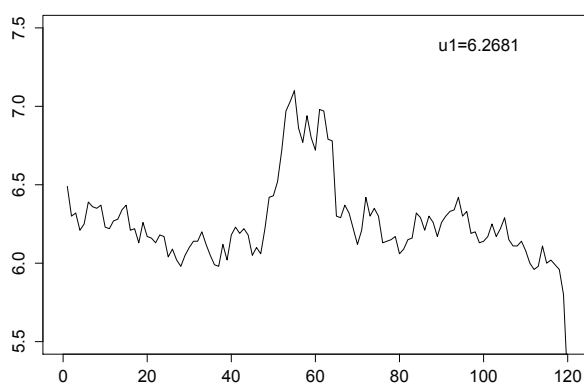


图 2 鲁北化工股票 1 价格

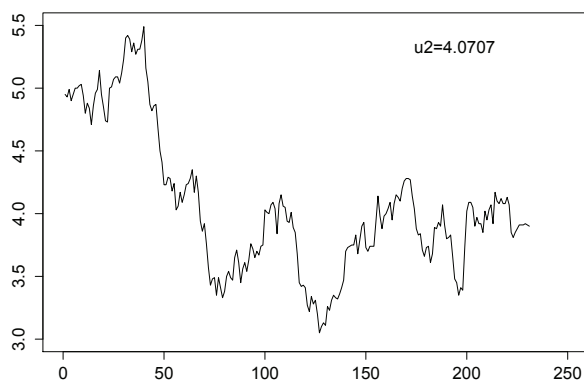


图 3 鲁北化工股票 2 价格

事实上,  $\hat{k} = 120$  所对应的实际时间是 2004 年 12 月 31 日, 这一天该公司董事会发表了这样一条声明: 决定对公司的不规范之处进行整改. 这可能是导致股票价格下跌的一个因素, 在股价下跌一段时间之后, 最终股票价格在 4 元上下波动, 实例分析的结果与我们的理论结果相吻合.

## §5. 结 论

本文通过经验似然思想建立假设检验的方法, 研究了重尾序列均值变点问题. 在现实生活中, 由于许多金融资产收益率的分布都具有高尖峰、新息重尾等特性, 然而重尾指数  $\kappa$  在实际操作中是难以估计的, 所以应用参数方法对重尾序列的变点进行检验就不再适用, 故本文应用非参数方法对重尾序列是否存在均值变点进行检验. 首先基于重尾数据, 在原假设和备择假设下得到经验似然函数. 其次, 基于经验似然函数构造似然比检验统计量, 并给出在原假设成立时该似然比统计量的渐近分布. 最后, 进行 Monte Carlo 数值模拟验证该方法的有效性. 模拟结果表明, 在没有变点的情况下, 经验水平值是接近于显著性水平  $\alpha$  的, 且样本容量  $n$  越大, 经验水平值越接近于显著性水平  $\alpha$ , 在有变点的情况下, 功效函数值随着样本容量和变点跃度的增加而不断逼近于 1, 说明本方法对重尾序列均值变点的检测具有良好效果; 本文的最后我们把经验似然比方法与极大似然比方法分别对正态分布数据与重尾分布数据均值变点的检验效果进行了比较. 当数据服从正态分布时, 参数检验方法, 即极大似然比方法检验效果更好, 但当数据服从重尾分布时, 本文的非参数检验方法, 即经验似然比方法检验效果更好.

## 参 考 文 献

- [1] PAGE E S. Continuous inspection schemes[J]. *Biometrika*, 1954, **41(1-2)**: 100–115.
- [2] 陈希儒. 变点统计分析简介[J]. *数理统计与管理*, 1991, **10(2)**: 52–59.
- [3] WANG L H, WANG J D. Change-of-variance problem for linear processes with long memory[J]. *Statist Papers*, 2006, **47(2)**: 279–298.
- [4] MADURKAYOVÁ B. Ratio tests for gradual changes[C]// ŠAFRÁNKOVÁ J, PAVLŮ J. (eds.) *WDS'07 Proceedings of Contributed Papers: Part I - Mathematics and Computer Sciences*, Prague: Matfyzpress, 2007: 175–180.
- [5] STEINEBACH J G, TIMMERMANN H. Sequential testing of gradual changes in the drift of a stochastic process[J]. *J Statist Plann Inference*, 2011, **141(8)**: 2682–2699.
- [6] CIUPERCA G. The S-estimator in the change-point random model with long memory[J]. *Statistics*, 2012, **46(6)**: 697–718.
- [7] DEHLING H, ROOCH A, TAQQU M S. Non-parametric change-point tests for long-range dependent data[J]. *Scand J Stat*, 2013, **40(1)**: 153–173.
- [8] ZOU C L, YIN G S, FENG L, et al. Nonparametric maximum likelihood approach to multiple change-point problems[J]. *Ann Statist*, 2014, **42(3)**: 970–1002.
- [9] JIRAK M. Uniform change point tests in high dimension[J]. *Ann Statist*, 2015, **43(6)**: 2451–2483.
- [10] AMINIKHANGHAHI S, COOK D J. A survey of methods for time series change point detection[J]. *Knowl Inf Syst*, 2017, **51(2)**: 339–367.
- [11] MANDELNBROT B. The variation of certain speculative prices[J]. *J Business*, 1963, **36(4)**: 394–419.

- [12] MEERSCHAERT M M, SCHEFFLER H P. Chapter 15 - Portfolio modeling with heavy tailed random vectors [M] // RACHEV S T. (ed.) *Handbook of Heavy Tailed Distributions in Finance: Volume 1 in Handbooks in Finance*. Amsterdam, Netherlands: Elsevier Science B V, 2003: 595–640.
- [13] OWEN A B. Empirical likelihood ratio confidence intervals for a single functional [J]. *Biometrika*, 1988, **75(2)**: 237–249.
- [14] CHEN S X. On the accuracy of empirical likelihood confidence regions for linear regression model [J]. *Ann Inst Statist Math*, 1993, **45(4)**: 621–637.
- [15] CHEN S X. Empirical likelihood confidence intervals for linear regression coefficients [J]. *J Multivariate Anal*, 1994, **49(1)**: 24–40.
- [16] CHEN S X. Empirical likelihood confidence intervals for nonparametric density estimation [J]. *Biometrika*, 1996, **83(2)**: 329–341.
- [17] EINMAHL J H J, MCKEAGUE I W. Empirical likelihood based hypothesis testing [J]. *Bernoulli*, 2003, **9(2)**: 267–290.
- [18] GUAN Z. A semiparametric changepoint model [J]. *Biometrika*, 2004, **91(4)**: 849–862.
- [19] ZHU L X, XUE L G. Empirical likelihood confidence regions in a partially linear single-index model [J]. *J R Stat Soc Ser B Stat Methodol*, 2006, **68(3)**: 549–570.
- [20] KOKOSZKA P, WOLF M. Subsampling the mean of heavy-tailed dependent observations [J]. *J Time Series Anal*, 2004, **25(2)**: 217–234.
- [21] SHORACK G R, WELLNER J A. *Empirical Processes with Applications to Statistics* [M]. New York: Wiley, 1986.
- [22] KOMLÓS J, MAJOR P, TUSNÁDY G. An approximation of partial sums of independent RV's, and the sample DF. I [J]. *Z Wahrscheinlichkeitstheorie und Verw Gebiete*, 1975, **32(1-2)**: 111–131.
- [23] 钱敏平, 龚光鲁. 随机过程论 [M]. 北京: 北京大学出版社, 1997.
- [24] CSÖRGŐ M, HORVÁTH L. *Limit Theorems in Change-Point Analysis* [M]. Chichester: Wiley, 1997.
- [25] 韩四儿. 两类厚尾相依序列的变点分析 [D]. 西安: 西北工业大学, 2007.

## Empirical Likelihood Ratio Test for Mean Change-Point in Heavy-Tailed Sequence

WANG Dan      PI Lin

(School of Mathematics, Northwest University, Xi'an, 710127, China)

**Abstract:** This paper establishes a empirical likelihood method to detect change-point in the mean of heavy-tailed sequence. Firstly, under the null and the alternative hypothesis, the empirical likelihood functions are obtained in the heavy-tailed observations. Secondly, the empirical likelihood ratio statistics is constructed based on empirical likelihood functions. And under the null hypothesis, the asymptotic distribution of statistics is given. Finally, Monte Carlo simulation is carried out to verify the correctness of the method. The simulation results show that the performance of our method is well to detect mean change in heavy-tailed sequence.

**Keywords:** heavy-tailed sequence; change-point in the mean; empirical likelihood ratio test

**2020 Mathematics Subject Classification:** 62G10; 62G05