

Detecting Parameters Change Points in the Generalized Exponential Distribution

Aifang Qiao^{1*}, Xijuan Niu²

¹Department of Mathematics, Qinghai Normal University, Xining Qinghai

²College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

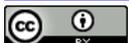
Email: ¹15202510232@163.com

Received: Dec. 3rd, 2015; accepted: Dec. 20th, 2015; published: Dec. 23rd, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper proposes a likelihood ratio method and a CUSUM method to detect change points of location parameters and scale parameters in the generalized exponential distribution. A Bootstrap method is introduced to approximate the critical values of the statistics for the scale parameters change points without explicit estimation leading to the critical values of statistic not easy to calculate. Simulations show that the likelihood ratio method is better than the CUSUM method when detecting change points of location parameters, however, the likelihood ratio statistic can't test change points of scale parameters. While the CUSUM method has the good performance testing two kinds of parameters change points. Combining the two methods can differentiate location parameters and scale parameters change points. Finally, the validity of proposed methods is demonstrated by analyzing a set of voltage data.

Keywords

Generalized Exponential Distribution, Change Point, Likelihood Ratio, CUSUM, Bootstrap

广义指数分布参数变点的检验

乔爱芳^{1*}, 牛玺娟²

¹青海师范大学数学系, 青海 西宁

²西北师范大学数学与统计学院, 甘肃 兰州

*通讯作者。

Email: 15202510232@163.com

收稿日期: 2015年12月3日; 录用日期: 2015年12月20日; 发布日期: 2015年12月23日

摘要

本文分别提出了检验广义指数分布位置参数及尺度参数变点的似然比方法和CUSUM方法, 并针对尺度参数无显式估计导致检验统计量的临界值不易计算的问题, 提出了用于近似统计量临界值的Bootstrap方法。模拟结果表明, 似然比方法在检验位置参数变点时优于CUSUM方法, 但无法检验出尺度参数变点, 而CUSUM方法对两类参数变点都有较好的检验效果, 结合两种方法能区分出两类参数变点。最后应用本文方法分析了一组电压数据, 说明所给方法的有效性和实用性。

关键词

广义指数分布, 变点, 似然比, CUSUM, Bootstrap

1. 引言

广义指数分布是由 Gupta 和 Kundu [1] 在 1999 年提出的, 其分布函数和密度函数分别为

$$\begin{aligned} F(x; \lambda, \beta) &= (1 - e^{-\lambda x})^\beta \quad (\lambda, \beta, x > 0), \\ f(x; \lambda, \beta) &= \beta \lambda (1 - e^{-\lambda x})^{\beta-1} e^{-\lambda x}, \end{aligned} \quad (1)$$

其中 λ 为尺度参数, β 为位置参数, 将该分布记为 $GE(\lambda, \beta)$ 。当 $\beta=1$ 时, GE 分布为指数分布。广义指数分布已被广泛应用到可靠性寿命分析、物理工程、临床医学、图像处理等领域。近年来对这一分布的研究已有了一些成果, 如文献[2]讨论了其次序统计量的矩; 文献[3]给出了其贝叶斯统计推断; 文献[4]在逐步增加 I 型截尾试验下, 给出了尺度参数 λ 的近似极大似然估计; 文献[5]给出了该分布的极大似然估计、矩估计、分位数估计、最小二乘估计和线性矩估计这五种估计, 并对各类估计进行了数值模拟, 得出极大似然估计均方误差最小; 文献[6]将该分布与 Gamma 分布和 Weibull 分布分别进行比较, 得出该分布在偏斜寿命数据中估计效果更好, 可以替代 Gamma 分布和 Weibull 分布。

指数分布、Weibull 分布、Gamma 分布、双参数指数分布、广义指数分布等都是描述产品可靠性的常用分布, 但由于诸多因素的影响, 这些寿命分布中的参数可能在某个时刻发生改变, 即产生变点。当数据中出现变点时, 如果继续按照原有分布描述数据就会产生错误, 因此检验这些寿命分布中的参数变点非常重要。为此, 黄志坚和张志华[7]研究了可靠性数据在变点前后服从不同参数的指数分布产品的寿命分布, 并且给出了参数的极大似然估计和贝叶斯估计。王黎明和王静龙[8]基于 U-统计量, 针对位置参数模型讨论了位置参数变点的检验问题, 并将结果应用到了双参数指数分布和 Weibull 分布尺度参数变点的检验问题中。谭常春等[9]针对 Gamma 分布利用累积和方法介绍了检测变点位置的程序, 并给出了变点的强相合估计和强收敛速度。王黎明[10]讨论了双参数指数分布变点的统计推断问题, 给出了检验的渐近分布和变点的估计。而关于广义指数分布参数变点的检验问题尚无研究成果。

本文介绍了检验广义指数分布序列中位置参数变点和尺度参数变点的似然比方法和 CUSUM 方法, 并针对尺度参数无显式估计导致检验统计量的临界值不易计算的问题, 提出了用于近似统计量临界值的 Bootstrap 方法。数值模拟结果表明, 似然比方法在检验位置参数变点时优于 CUSUM 方法, 但是无法检

验出尺度参数变点, 而 CUSUM 方法对两类参数变点都有较好的检验效果, 在检验时无法区分出位置参数变点和尺度参数变点, 所以结合这两种方法能更好地检测广义指数分布中的两类参数变点。此外, 和用渐近临界值做检验相比, 利用 Bootstrap 方法做检验也能够较好的控制经验水平, 且方便实际操作。最后应用本文方法分析了一组电压数据, 说明所给方法的有效性和实用性。

2. 方法介绍

2.1. 似然比方法

2.1.1. 位置参数 β 的似然比检验

设 x_1, x_2, \dots, x_n 是来自广义指数分布(1)的一组样本, 参数分别是 $(\lambda_1, \beta_1), (\lambda_2, \beta_2), \dots, (\lambda_n, \beta_n)$ 。假定 λ 已知, 不妨设 $\lambda = 1$, 则关于参数 β 的变点检验即是检验原假设 $H_0: \beta_1 = \beta_2 = \dots = \beta_n = \beta$ 和备择假设 $H_1: \beta_1 = \beta_2 = \dots = \beta_{k^*} \neq \beta_{k^*+1} = \dots = \beta_n$, 其中, k^* 是变点的位置, $1 < k^* < n$, n 为样本容量。在原假设 H_0 下的对数似然函数为

$$\ln L_0(\beta) = n \ln \beta + (\beta - 1) \sum_{i=1}^n \ln(1 - e^{-x_i}) - \sum_{i=1}^n x_i,$$

则 β 的极大似然估计为 $\hat{\beta} = -\frac{n}{\sum_{i=1}^n \ln(1 - e^{-x_i})}$, 由于 $\hat{\beta}$ 不是无偏的, 修正之后得到无偏估计

$\hat{\beta}_U = \frac{n-1}{n} \hat{\beta} = -\frac{n-1}{\sum_{i=1}^n \ln(1 - e^{-x_i})}$, 因此在 H_0 下的对数似然函数可化为

$$\ln L_0(\hat{\beta}_U) = n \ln \left[\frac{1-n}{\sum_{i=1}^n \ln(1 - e^{-x_i})} \right] + \left[\frac{1-n}{\sum_{i=1}^n \ln(1 - e^{-x_i})} - 1 \right] \sum_{i=1}^n \ln(1 - e^{-x_i}) - \sum_{i=1}^n x_i.$$

在备择假设 H_1 下的对数似然函数为

$$\begin{aligned} \ln L_1(\beta_1, \beta_n) &= k^* \ln \beta_1 - \sum_{i=1}^{k^*} x_i + (\beta_1 - 1) \sum_{i=1}^{k^*} \ln(1 - e^{-x_i}) + (n - k^*) \ln \beta_n \\ &\quad - \sum_{i=k^*+1}^n x_i + (\beta_n - 1) \sum_{i=k^*+1}^n \ln(1 - e^{-x_i}), \end{aligned}$$

则 β_1, β_n 的极大似然估计分别为 $\hat{\beta}_1 = \frac{-k^*}{\sum_{i=1}^{k^*} \ln(1 - e^{-x_i})}$, $\hat{\beta}_n = \frac{k^* - n}{\sum_{i=k^*+1}^n \ln(1 - e^{-x_i})}$,

修正 $\hat{\beta}_1, \hat{\beta}_n$ 得到其无偏估计分别为

$$\hat{\beta}_{1U} = \frac{k^* - 1}{k^*} \hat{\beta}_1 = \frac{1 - k^*}{\sum_{i=1}^{k^*} \ln(1 - e^{-x_i})}, \quad \hat{\beta}_{nU} = \frac{n - k^* - 1}{n - k^*} \hat{\beta}_n = \frac{1 + k^* - n}{\sum_{i=k^*+1}^n \ln(1 - e^{-x_i})}.$$

故而在 H_1 下的对数似然函数可化为

$$\begin{aligned} \ln L_1(\hat{\beta}_{1U}, \hat{\beta}_{nU}) &= k^* \ln \left[\frac{-k^*}{\sum_{i=1}^{k^*} \ln(1 - e^{-x_i})} \right] - \sum_{i=1}^{k^*} x_i + \left[\frac{-k^*}{\sum_{i=1}^{k^*} \ln(1 - e^{-x_i})} - 1 \right] \sum_{i=1}^{k^*} \ln(1 - e^{-x_i}) \\ &\quad + (n - k^*) \ln \left[\frac{1 + k^* - n}{\sum_{i=k^*+1}^n \ln(1 - e^{-x_i})} \right] - \sum_{i=k^*+1}^n x_i + \left[\frac{1 + k^* - n}{\sum_{i=k^*+1}^n \ln(1 - e^{-x_i})} - 1 \right] \sum_{i=k^*+1}^n \ln(1 - e^{-x_i}), \end{aligned}$$

, 因此, 检验 β 的似

然比统计量为

$$\Lambda_\beta = \max_{1 < k^* < n} \left\{ -2 \ln \frac{L_0(\hat{\beta}_U)}{L_1(\hat{\beta}_{1U}, \hat{\beta}_{nU})} \right\} = \max_{1 < k^* < n} f_n(x; k^*),$$

$$\text{其中, } f_n(x; k^*) = -2 \ln \left[\gamma^n(x) \theta^{-k^*}(x) \xi^{k^*-n}(x) \prod_{i=1}^{k^*} (1 - e^{-x_i})^{\gamma(x) - \theta(x)} \prod_{i=k^*+1}^n (1 - e^{-x_i})^{\gamma(x) - \xi(x)} \right],$$

$$\gamma(x) = \frac{1-n}{\sum_{i=1}^n \ln(1 - e^{-x_i})}, \quad \theta(x) = \frac{-k^*}{\sum_{i=1}^{k^*} \ln(1 - e^{-x_i})}, \quad \xi(x) = \frac{1+k^*-n}{\sum_{i=k^*+1}^n \ln(1 - e^{-x_i})}.$$

在给定显著性水平 α 下, 当 $\Lambda_\beta > c_\alpha$ (c_α 为临界值) 时, 拒绝原假设 H_0 , 认为序列中存在位置参数变点, 否则, 接受原假设 H_0 , 认为位置参数没有发生改变。

2.1.2. 尺度参数 λ 的似然比检验

与上节类似地, 关于参数 λ 的变点检验即是检验原假设 $H_0: \lambda_1 = \lambda_2 = \dots = \lambda_n = \lambda$ 和备择假设 $H_1: \lambda_1 = \lambda_2 = \dots = \lambda_{k^*} \neq \lambda_{k^*+1} = \dots = \lambda_n$, 其中, k^* 是变点的位置, $1 < k^* < n$, n 为样本容量。在原假设 H_0 下的对数似然函数为

$$\ln L_0(\lambda) = n \ln \beta + n \ln \lambda + (\beta - 1) \sum_{i=1}^n \ln(1 - e^{-\lambda x_i}) - \lambda \sum_{i=1}^n x_i,$$

$$\text{对 } \lambda \text{ 求偏导数: } \frac{\partial \ln L_0(\lambda)}{\partial \lambda} = \frac{n}{\lambda} + (\beta - 1) \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{1 - e^{-\lambda x_i}} - \sum_{i=1}^n x_i = 0, \text{ 得出}$$

$$v(\lambda) = \lambda, \quad (2)$$

$$\text{其中 } v(\lambda) = \left[\frac{1}{n} \sum_{i=1}^n \frac{x_i (1 - \beta e^{-\lambda x_i})}{1 - e^{-\lambda x_i}} \right]^{-1},$$

由文献[11]可知式(2)存在唯一解, 运用 Newton-Raphson 迭代法可求出其近似解 $\hat{\lambda}$ 。在备择假设 H_1 下的对数似然函数为

$$\begin{aligned} \ln L_1(\lambda_1, \lambda_n) &= n \ln \beta + k^* \ln \lambda_1 + (n - k^*) \ln \lambda_n - \lambda_1 \sum_{i=1}^{k^*} x_i - \lambda_n \sum_{i=k^*+1}^n x_i \\ &\quad + (\beta - 1) \sum_{i=1}^{k^*} \ln(1 - e^{-\lambda_1 x_i}) + (\beta - 1) \sum_{i=k^*+1}^n \ln(1 - e^{-\lambda_n x_i}) \end{aligned}$$

$$\text{对 } \lambda_1, \lambda_n \text{ 分别求偏导数: } \frac{\partial \ln L_1(\lambda_1, \lambda_n)}{\partial \lambda_1} = \frac{k^*}{\lambda_1} + (\beta - 1) \sum_{i=1}^{k^*} \frac{x_i e^{-\lambda_1 x_i}}{1 - e^{-\lambda_1 x_i}} - \sum_{i=1}^{k^*} x_i = 0,$$

$$\frac{\partial \ln L_1(\lambda_1, \lambda_n)}{\partial \lambda_n} = \frac{n - k^*}{\lambda_n} + (\beta - 1) \sum_{i=k^*+1}^n \frac{x_i e^{-\lambda_n x_i}}{1 - e^{-\lambda_n x_i}} - \sum_{i=k^*+1}^n x_i = 0, \text{ 得到}$$

$$u(\lambda_1) = \lambda_1, \quad (3)$$

$$h(\lambda_n) = \lambda_n, \quad (4)$$

$$\text{其中 } u(\lambda_1) = \left[\frac{1}{k^*} \sum_{i=1}^{k^*} \frac{x_i (1 - \beta e^{-\lambda_1 x_i})}{1 - e^{-\lambda_1 x_i}} \right]^{-1}, \quad h(\lambda_n) = \left[\frac{1}{n - k^*} \sum_{i=k^*+1}^n \frac{x_i (1 - \beta e^{-\lambda_n x_i})}{1 - e^{-\lambda_n x_i}} \right]^{-1}.$$

同理, 通过 Newton-Raphson 迭代法可以求出(3)、(4)的解 $\hat{\lambda}_1, \hat{\lambda}_n$ 。因此, 检验 λ 的似然比统计量为

$$\Lambda_\lambda = \max_{1 < k^* < n} \left\{ -2 \ln \frac{L_0(\hat{\lambda})}{L_1(\hat{\lambda}_1, \hat{\lambda}_n)} \right\}。$$

在给定显著性水平 α 下, 当 $\Lambda_\lambda > c_\alpha$ (c_α 为临界值)时, 拒绝原假设 H_0 , 认为序列中存在尺度参数变点, 否则接受原假设 H_0 , 认为尺度参数没有发生改变。虽然可以通过迭代的方法来计算 $\hat{\lambda}_1, \hat{\lambda}_n$, 但在实际应用中这种方法极为不便, 为此进一步提出如下 CUSUM 方法。

2.2. CUSUM 方法

设 x_1, x_2, \dots, x_n 是来自广义指数分布(1)的一组样本, 变点出现在 k^* 时刻, $1 < k^* < n$, n 为样本容量。

记 $\bar{x}_{k^*} = k^{*-1} \sum_{t=1}^{k^*} x_t$, $\bar{x}_{k^*}^* = (n - k^*)^{-1} \sum_{t=k^*+1}^n x_t$, $\bar{x} = n^{-1} \sum_{t=1}^n x_t$, $\sigma^2 = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2$, 构造统计量

$$\text{CUSUM}(k^*) = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{k^*} x_t - \frac{k^*}{n} \sum_{t=1}^n x_t \right) = \frac{k^*}{n} \left(1 - \frac{k^*}{n} \right) \left[\sqrt{n} (\bar{x}_{k^*} - \bar{x}_{k^*}^*) \right],$$

文献[12]给出了 $\text{CUSUM}(k^*)$ 的极限分布, 即

$$\frac{1}{\hat{\sigma}} \max_{1 < k^* < n} |\text{CUSUM}(k^*)| \xrightarrow{D} \sup_{0 \leq t \leq 1} |B(t)|,$$

其中, $\hat{\sigma}$ 是 σ 的一致估计, $B(t)$ 是布朗桥。

故而得到 CUSUM 检验统计量

$$U_{k^*} = \frac{1}{\hat{\sigma}} \max_{1 < k^* < n} |\text{CUSUM}(k^*)|。$$

该方法对检验位置参数 β 和尺度参数 λ 中的变点时都有效。在给定显著性水平 α 下, 当 $U_{k^*} > c_\alpha$ (c_α 为临界值)时, 拒绝原假设, 认为序列中存在变点, 否则接受原假设, 认为序列中的参数没有发生改变。

2.3. Bootstrap 近似法

由于似然比统计量的渐近分布不易得到, 故而构造 Bootstrap 抽样方法来确定统计量的经验临界值。Bootstrap 方法已广泛应用于抽样技术中, 陈占寿等[13]为确定监测统计量的临界值并避免估计厚尾指数, 提出了 Bootstrap 重抽样方法。赵春辉等[14]提出了一种基于均值函数的核估计检验统计量, 得到统计量在原假设和备择假设下的极限分布, 并构造了 Bootstrap 方法对非参数回归模型均值函数结构变点进行检验, 证明了检验和估计的一致性。

Bootstrap 方法的步骤如下:

第一步, 从 x_1, x_2, \dots, x_n 中任意抽取 n 个 Bootstrap 样本 $x_1^*, x_2^*, \dots, x_n^*$ 。

第二步, 将 $x_1^*, x_2^*, \dots, x_n^*$ 代入统计量 Λ_β , 计算统计量 Λ_β^* 的值。

第三步, 重复第一步和第二步 m 次, 用 Λ_β^* 的经验分位数作为 Λ_β 的临界值。

这里只给出了统计量 Λ_β 的 Bootstrap 近似方法, 实际上, 对于统计量 U_{k^*} 亦可采用类似方法计算临界值。

3. 数值模拟

本节通过数值模拟的方法, 来检验上节所提方法的有限样本的性质, 重点分析样本容量、参数值、变点位置等因素对上述两种检验统计量的影响。所有模拟程序均通过 matlab 软件生成(见附录), 首先利用渐近方法模拟临界值, 其方法是取样本容量为 5000, 令 $\beta = 1, \lambda = 1$, 通过 10,000 次循环计算出统计

量 Λ_β 和 U_{k^*} 的经验分位数作为其临界值。Bootstrap 方法的重抽样次数 $m = 1000$ 。

表 1 给出的是两种检验统计量在不同的样本容量条件下模拟所得的经验水平, 循环次数为 10,000 (下同), 检验水平 $\alpha = 0.1, 0.05, 0.01$ 。通过观察可以发现, 统计量 Λ_β 和 U_{k^*} 的经验水平都比较接近于检验水平, 经验水平扭曲的程度随样本容量的增大而减小。利用渐近临界值和 Bootstrap 方法做检验时, 两种检验统计量都能较好的控制经验水平, 但是实际中 λ 和 β 是未知的, 需提前估计, 且不同的 λ 和 β 对应不同的临界值, 这给实际应用带来极大的不便, 而 Bootstrap 方法不需要估计这些未知参数, 在实际应用中更加方便。

表 2 给出了位置参数 β 的取值在 k^* 处由 1 变为 2 和 1/2 时两种检验统计量的经验势, 变点位置 k^* 取 0.25, 0.5, 0.75, 检验水平为 0.05 (下同)。可以看出, 当样本容量逐渐增大时, 统计量 Λ_β 和 U_{k^*} 的经验势均在增高, 这说明两种统计量对位置参数 β 的检验是一致的。但是似然比统计量 Λ_β 明显比 CUSUM 统计量 U_{k^*} 的检验效率高, 说明似然比统计量在检验位置参数变点时优于 CUSUM 统计量。当变点位置 k^* 靠近序列中间时, 模拟所得的经验势比变点位置在两端时的经验势高, 说明两种检验统计量对在序列中间位置出现的位置参数变点的检验效果更好。此外, 利用渐近临界值和 Bootstrap 方法做检验所得的检验效率比较接近, 这说明在实际应用中可以用 Bootstrap 方法代替渐近方法模拟临界值。

表 3 给出了尺度参数 λ 的取值在 k^* 处由 1 变为 2 和 1/2 时两种检验统计量的经验势。可以看出, 似然比统计量的检验效率较低, 说明统计量 Λ_β 对尺度参数 λ 的变化不敏感, 无法检验出序列中的尺度参数变点。当样本容量逐渐增大时, 统计量 U_{k^*} 的经验势逐渐增高, 当样本容量 $n = 200$ 时, 其经验势已非常接近 1, 说明 CUSUM 统计量对尺度参数变点的检验效果较好。当变点位置 k^* 靠近序列中间时, CUSUM 统计量模拟所得的经验势比变点位置在两端时的经验势高, 这说明 CUSUM 统计量对在序列中间位置出现的尺度参数变点的检验能力较强。

综上所述, 在检验位置参数变点时, 两种检验统计量是一致性检验, 且似然比统计量 Λ_β 的检验效果优于 CUSUM 统计量, 若变点恰好发生在序列的中间位置时, 两种统计量的检验能力最强。在检验尺度参数变点时, 似然比统计量 Λ_β 对尺度参数 λ 的检验效果较差, 无法检验出尺度参数变点, 而 CUSUM 统计量对两类参数变点都比较敏感, 在检验时无法区分出位置参数变点和尺度参数变点, 所以结合这两种统计量能更好地检测广义指数分布序列中的两类参数变点。此外, 利用 Bootstrap 方法做检验时控制经验水平的能力和用渐近临界值做检验时是一致的, 且 Bootstrap 方法在实际操作时更为方便, 所以在实际应用中可以用 Bootstrap 方法来代替渐近方法。总之, 在实际应用时可以将似然比统计量和 CUSUM 统计量组合考虑来区分两类参数变点, 以便得出更准确的结果。

4. 实例分析

本节通过一组家庭电力消费的数据来说明本文方法的有效性。数据包括 2006 年 12 月 16 日 20:40~23:59 家庭电力消费中电压的 200 个观测值, 该数据集可以从 UCI 机器学习数据库 (<http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>) 中的 “Individual Household Electric Power Consumption Data Set” 中得到。图 1 是电压的原始数据图, 图 2 的 Q-Q 图表明电压数据服从广义指数分布, 因此, 我们所选用的数据符合前文中提出的假设。

下面对数据做变点检验, 按照 CUSUM 统计量 U_{k^*} 对图 1 数据进行计算, 得到 $U_{\max} = U_{97} = 5.9477 > 1.3865$ (临界值由 Bootstrap 方法模拟得到), 说明检验结果显著, 即第 97 个数据处存在变点。但无法确定是位置参数还是尺度参数发生了改变。按照似然比统计量 Λ_β 对图 1 数据进行计算, 检验结果不显著, 即没有发现变点。根据上节中两种检验统计量对两类参数变点的检验能力分析可知, 正是尺度参数 λ 的变化, 产生了变点。

Table 1. Empirical size (percent)

表 1. 经验水平(%)

n	统计量	渐近			Bootstrap		
		10%	5%	1%	10%	5%	1%
100	Λ_β	9.6	4.9	0.9	9.2	4.8	1.2
	U_{k^*}	10.1	5.9	1.0	10.1	4.7	1.4
200	Λ_β	9.9	5.1	1.0	9.9	5.0	1.1
	U_{k^*}	10.3	5.3	1.1	9.8	5.2	1.0

Table 2. Empirical power of β (percent)

表 2. 参数 β 的经验势(%)

n	β	渐近			Bootstrap	
		k^*	Λ_β	U_{k^*}	Λ_β	U_{k^*}
100	$1 \rightarrow \frac{1}{2}$	0.25	60.0	32.4	54.8	36.7
		0.5	79.7	51.0	70.9	52.5
		0.75	65.1	25.7	60.5	25.9
	$1 \rightarrow 2$	0.25	66.7	34.3	59.2	35.9
		0.5	78.1	60.4	73.3	65.9
		0.75	60.3	37.8	54.9	41.2
200	$1 \rightarrow \frac{1}{2}$	0.25	93.9	63.4	92.7	64.6
		0.5	98.8	87.3	97.9	88.0
		0.75	94.3	55.2	93.2	55.5
	$1 \rightarrow 2$	0.25	94.4	67.3	90.6	70.5
		0.5	98.5	92.3	98.1	94.6
		0.75	93.0	74.6	92.2	76.1

Table 3. Empirical power of λ (percent)

表 3. 参数 λ 的经验势(%)

n	λ	渐近			Bootstrap	
		k^*	Λ_β	U_{k^*}	Λ_β	U_{k^*}
100	$1 \rightarrow \frac{1}{2}$	0.25	50.6	45.6	49.7	44.1
		0.5	56.8	90.0	55.7	89.1
		0.75	45.1	77.7	44.5	77.4
	$1 \rightarrow 2$	0.25	16.5	78.2	15.3	77.8
		0.5	24.5	89.4	22.8	89.7
		0.75	18.8	44.5	16.6	44.9
200	$1 \rightarrow \frac{1}{2}$	0.25	74.9	89.2	75.3	90.2
		0.5	81.8	99.7	81.9	99.8
		0.75	69.3	98.1	70.1	99.2
	$1 \rightarrow 2$	0.25	40.9	98.8	38.9	98.5
		0.5	54.8	99.9	54.5	99.9
		0.75	46.5	88.3	47.2	90.5

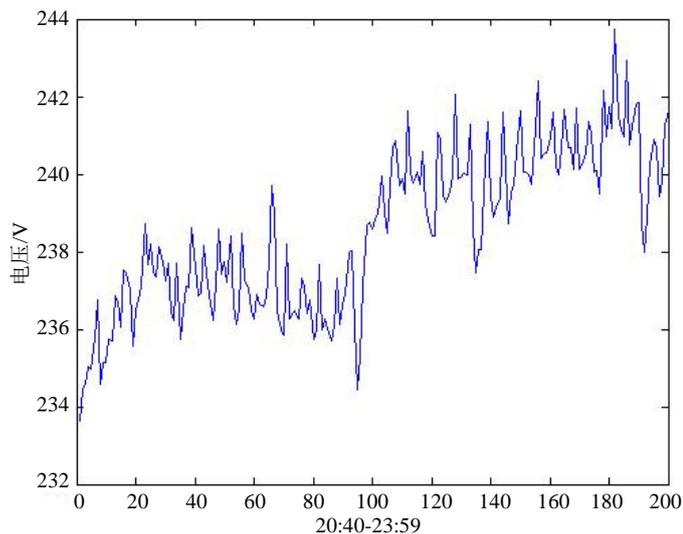


Figure 1. The original voltage data changing over time
图 1. 电压随时间变化的原始数据

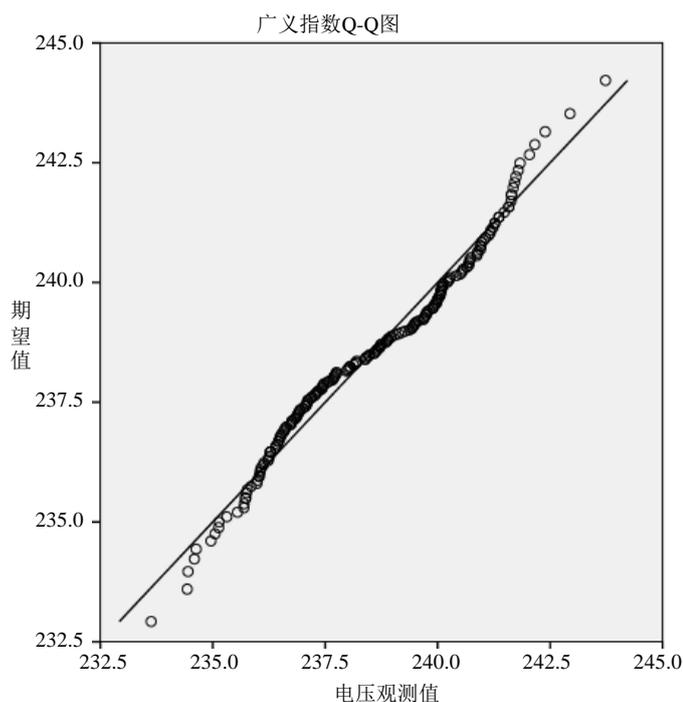


Figure 2. Q-Q plot of generalized exponential distribution
图 2. 广义指数分布的 Q-Q 图

综上所述, 该家庭在 2006 年 12 月 16 日 20:40~23:59 期间的电压有明显的突变, 变点出现在 22:16 时刻, 自 22:16~23:59 电压升高, 究其原因是因为这期间电器基本不运作, 用电量明显减少, 故而电压升高。

5. 结论

本文介绍了检验广义指数分布序列位置参数变点和尺度参数变点的似然比方法和 CUSUM 方法, 并

针对尺度参数无显式估计导致检验统计量的临界值不易计算的问题, 提出了用于近似统计量临界值的 Bootstrap 方法。数值模拟结果表明, 似然比方法在检验位置参数变点时优于 CUSUM 方法, 但是无法检验出尺度参数变点, 而 CUSUM 方法对两类参数变点都有较好的检验效果, 在检验时无法区分出位置参数变点和尺度参数变点, 所以结合这两种方法能更好地检测广义指数分布中的两类参数变点。此外, 和用渐近临界值做检验相比, 利用 Bootstrap 方法做检验也能够较好的控制经验水平, 且方便实际操作。最后应用本文方法分析了一组电压数据, 说明所给方法的有效性和实用性。

致 谢

在此, 我要感谢我的导师陈占寿老师, 感谢陈老师对我在生活上的关心和学业上的精心指导。他渊博的知识、严谨求实、精益求精的治学态度和对专业前沿问题敏锐的洞察力深深地影响了我, 使我不仅学到了本专业的理论知识, 而且培养了我发现问题、分析问题、解决问题的能力, 使我终身受益。我还要感谢国家自然科学基金、教育部人文社会科学基金、青海省自然科学基金的资助, 以及所引用文献的作者对科学做出的贡献。

基金项目

国家自然科学基金(11301291), 教育部人文社会科学基金(11XJA910001), 青海省自然科学基金(2015-ZJ-717)资助。

参考文献 (References)

- [1] Gupta, R.D. and Kundu, D. (1999) Generalized Exponential Distributions. *Australian and New Zealand Journal of Statistics*, **41**, 173-188. <http://dx.doi.org/10.1111/1467-842X.00072>
- [2] Raqab, M.Z. (2004) Generalized Exponential Distribution Moments of Order Statistics. *Statistics*, **38**, 29-41. <http://dx.doi.org/10.1080/0233188032000158781>
- [3] Raqab, M.Z. and Madi, M.T. (2005) Bayesian Inference for the Generalized Exponential Distribution. *Statistical Computation and Simulation*, **75**, 841-852. <http://dx.doi.org/10.1080/00949650412331299166>
- [4] Asgharzadeh, A. (2009) Approximate MLE for the Scaled Generalized Exponential Distribution under Progressive Type-II Censoring. *Journal of the Korean Statistical Society*, **3**, 223-229. <http://dx.doi.org/10.1016/j.jkss.2008.09.004>
- [5] Gupta, R.D. and Kundu, D. (2001) Generalized Exponential Distribution: Different Methods of Estimations. *Journal of Statistical Computation and Simulation*, **69**, 315-337. <http://dx.doi.org/10.1080/00949650108812098>
- [6] Gupta, R.D. and Kundu, D. (2001) Exponentiated Exponential Family: An Alternative to Gamma and Weibull Distributions. *Biometrical Journal*, **43**, 117-130. [http://dx.doi.org/10.1002/1521-4036\(200102\)43:1<117::AID-BIMJ117>3.0.CO;2-R](http://dx.doi.org/10.1002/1521-4036(200102)43:1<117::AID-BIMJ117>3.0.CO;2-R)
- [7] 黄志坚, 张志华. 基于指数分布数据的可靠性变点分析[J]. 武汉理工大学学报, 2008, 30(2): 157-160.
- [8] 王黎明, 王静龙. 位置参数变点的非参数检验及其渐近性质[J]. 数学年刊 A 辑(中文版), 2002(2): 229-234.
- [9] 谭常春, 缪柏其, 惠军. Γ 分布参数变点的非参数统计推断[J]. 中国科学技术大学学报, 2008, 38(2): 149-156.
- [10] 王黎明. 双参数指数分布参数变点的统计推断[J]. 系统工程, 2004, 22(3): 106-110.
- [11] 苏海军, 杨煜普, 王宇嘉. 微分进化算法的研究综述[J]. 系统工程与电子技术, 2008, 30(9): 1793-1797.
- [12] Robbins, M., Gallagher, C., Lund, R., et al. (2011) Mean Shift Testing in Correlated Data. *Journal of Time Series Analysis*, **32**, 498-511. <http://dx.doi.org/10.1111/j.1467-9892.2010.00707.x>
- [13] 陈占寿, 田铮. 一类厚尾随机信号平稳性的在线 bootstrap 监测[J]. 控制理论与应用, 2010, 27(7): 933-938.
- [14] 赵春辉, 田铮, 陈占寿. 非参数模型均值函数结构变点的 Bootstrap 检测[J]. 数理统计与管理, 2011, 30(4): 629-638.

附录

似然比统计量用渐近方法检验位置参数 β 的经验势的 matlab 模拟程序

```

clear
clc
j1=0;j2=0;j3=0;
c1=7.8948 ;
c2=9.3940 ;
c3=12.6240;
n=200;
M=10000;
af1=1;bt1=1;
af2=1;bt2=2;
kx=0.5;
KXT=kx*n;
for r=1:M
    e1=gexprnd(af1,bt1,1,KXT);
    e2=gexprnd(af2,bt2,1,n-KXT);
    x(1:KXT)=e1;
    x(KXT+1:n)=e2;
    for k=2:n-2
        gamma=(1-n)/sum(log(1-exp(-(x(1:n)))));
        beta=-k/sum(log(1-exp(-(x(1:k)))));
        kisi=(-n+k+1)/sum(log(1-exp(-(x(k+1:n)))));
        lamda(k)=-2*(n*log(gamma)-k*log(beta)-(n-k)*log(kisi)+k*(beta-gamma)/
            beta+(n-k-1)*(kisi-gamma)/kisi);
    end
    c=max(lamda);
    if c>c1
        j1=j1+1;
    end
    if c>c2
        j2=j2+1;
    end
    if c>c3
        j3=j3+1;
    end
end
E(1)=j1/M;
E(2)=j2/M;
E(3)=j3/M;
E

```