文章编号:0253-2778(2005)01-0051-08

至多一个变点的Γ分布的统计推断*

谭常春,缪柏其

(中国科学技术大学统计与金融系,合肥 230026)

摘要:对至多一个变点的Γ分布,即 $X_1,\cdots,X_n$ 为一列相互独立的随机变量序列,且 $X_1,\cdots,X_{k_0}$ $i.i.d \sim \Gamma(x;\nu_1,\lambda_1)$,$X_{k_0+1},\cdots,X_n$ $i.i.d \sim \Gamma(x;\nu_2,\lambda_2)$,其中 k_0 未知,称 k_0 为该序列的变点.借助 Gauss 过程理论和滑窗方法,利用第一型极值分布逼近本文提出的统计量的分布,给出了检测变点 k_0 的程序和变点的区间估计.最后对文中提出的统计量进行模拟并分析.

关键词:Γ分布;变点;区间估计;滑窗

中图分类号:O212.7 **文献标识码:**A

AMS Subject Classifications (2000): Primary 60F05;Secondary 62G05

0 引言

变点问题自上世纪70年代以来一直是统计中的一个热门话题,它广泛应用于工业质量控制、经济、金融等领域[1,2].关于独立随机变量序列中的变点问题,许多学者都进行了研究,如 Chernoff 和 Zacks[3]提出了检验正态分布均值变点的统计量,Haccon 和 Meelis[4]对独立指数随机变量序列中的变点提出了似然比检验方法,Daniel 等[5]用 Bayes 方法讨论了变点检验的问题,Siegmund 等[6]用似然比方法研究变点问题,文献[7~10]用非参数方法对变点进行检验和估计.陈希孺[11]研究了跳跃度存在变点的序列,Miao[12]研究了斜率存在变点的序列等.相对而言,关于分布中有多个参数的变点研究,大多是关于独立参数(如位置、刻度)的研究,如谭智平[13]讨论了跳跃度和斜率均存在变点的序列,缪柏其[14]同时讨论了变点个数和位置的检测.

Γ分布有两个参数 ν,λ,其均值和方差同时依赖于这两个参数.如何对这类分布变点进行检测是人们关心的一个问题.关于Γ分布参数变点的研究,已有一些结果,如 Hsu[15]给出了 ν 已知时 λ 的检验程序;Kimber[16]考虑了Γ样本中离群值上下界的检验;马超群[17]等在一定的先验分布下,用 Bayes 方法讨论了Γ分布参数多个转变点的问题.

设 $X \sim \Gamma(x;\nu,\lambda)$,其密度为

$$f(x)=\frac{\lambda^{\nu}}{\Gamma(\nu)}e^{-\lambda x}x^{\nu-1}\quad x>0$$

* **收稿日期:**2004-01-05

基金项目:国家自然科学基金资助项目(10471135).

作者简介:谭常春,男,1977年生.硕士生.研究方向:非参数统计推断.E-mail:tcczyl@mail.ustc.edu.cn

其中 ν 为形状参数，λ 为刻度参数. Γ 分布簇包含了其他一些分布簇，如指数分布簇 $E(\lambda)$ 是 $\nu=1$ 时的特例；$\chi^2(n)$ 是 $\nu=n,\lambda=1/2$ 的 Γ 分布. Erlang 分布就是当 $\nu=n$ 为正整数时的 Γ 分布，它常用于可靠性理论和排队论中，如一个复杂系统中从第一次故障到恰好再出现 n 次故障所需的时间. 因此研究了 Γ 分布簇参数变点的统计推断，也研究了 $\chi^2(n)$ 分布簇和指数分布簇 $E(\lambda)$ 的变点的统计推断. 若 $X_1,\cdots,X_n$，是一列独立的 Γ 分布随机变量，满足 $X_1,\cdots,X_{k_0}\ i.i.d \sim \Gamma(x;\nu_1,\lambda_1)$，$X_{k_0+1},\cdots,X_n\ i.i.d \sim \Gamma(x;\nu_2,\lambda_2)$ 其中 $\Gamma(x;\nu_1,\lambda_1)$，$\Gamma(x;\nu_2,\lambda_2)$ 参数分别为 $\nu_1,\lambda_1;\nu_2,\lambda_2$ 的两个 Γ 分布；称 k_0 为序列 $X_1,\cdots,X_n$ 的变点.

本文利用第一型极值分布理论，讨论 Γ 分布中两个参数 ν,λ 同时存在变点时变点位置的估计和检验. 为检验 Γ 序列 $X_1,\cdots,X_n$ 是否有变点，作原假设

$$H_0:\nu_1=\nu_2,\lambda_1=\lambda_2 \tag{1}$$

即不存在变点.

在本文中将讨论如下问题：第一，不论 λ 是否变化，ν 至多只有一个变点的检验；以及若变点存在，变点 k_0 的估计；第二，不论 ν 是否变化，λ 至多只有一个变点的检验；第三，模拟.

为方便起见，记

$$\overline{X}_{1,k}=\frac{1}{l}\sum_{k-l+1}^{k}X_i,\ \overline{X}_{2,k}=\frac{1}{l}\sum_{k+1}^{k+l}X_i,\ \hat{\sigma}_{1,k}^2=\frac{1}{l}\sum_{k-l+1}^{k}(X_i-\overline{X}_{1,k})^2,\ \hat{\sigma}_{2,k}^2=\frac{1}{l}\sum_{k+1}^{k+l}(X_i-\overline{X}_{2,k})^2,$$

$\mu=E(X_i)$，$\sigma^2=\mathrm{Var}(X_i)$.

1 ν 至多只有一个变点的检验

这一节只考虑

$$H_0:\nu_1=\nu_2\leftrightarrow H_1:\nu_1\neq\nu_2 \tag{2}$$

的检验，即 $H_0:\nu_1=\nu_2,\lambda_1=\lambda_2\leftrightarrow H_1:\nu_1\neq\nu_2,\lambda_1=\lambda_2$ 或 $H_0:\nu_1=\nu_2,\lambda_1\neq\lambda_2\leftrightarrow H_1:\nu_1\neq\nu_2,\lambda_1\neq\lambda_2$ 的检验.

因为 $E(X)=\dfrac{\nu}{\lambda}$，$\mathrm{Var}(X)=\dfrac{\nu}{\lambda^2}$，$\dfrac{E(X)}{\sqrt{\mathrm{Var}(X)}}=\sqrt{\nu}$，若 $X_1,\cdots,X_n$ 为 $i.i.d$ 序列，我们用 $\overline{X}_{2,k}$ 来估计 $E(X)$，用 $\hat{\sigma}_{2,k}^2$ 来估计 $\mathrm{Var}(X)$，则易证 $\dfrac{\overline{X}_{2,k}}{\hat{\sigma}_{2,k}}$ 为 $\sqrt{\nu}$ 的强相合估计，所以选用下面的 T_k 作为检验 ν 是否有变点的统计量. 又因为 $\dfrac{\overline{X}_{2,k}}{\hat{\sigma}_{2,k}}$ 与 λ 无关，且至多只有一个变点，因此次不论 λ 是否发生变化，T_k 仍可作为检验 ν 是否有变点的统计量（若 λ 发生变化，只会在 k_0 的两侧变化）.

定理1 设 $X_1,\cdots,X_n\ i.i.d$，$X_1\sim\Gamma(x;\nu,\lambda)$，$x>0$. $l=l_n$ 为正整数，满足条件

$$\lim_{n\to\infty}\frac{l}{n}=0;\quad \lim_{n\to\infty}\frac{(\log n)^2}{l}=0. \tag{3}$$

σ^2 已知，定义

$$T_k=\sqrt{\frac{l}{2}}\left(\frac{\overline{X}_{2,k}}{\hat{\sigma}_{2,k}}-\frac{\overline{X}_{1,k}}{\hat{\sigma}_{1,k}}\right) \tag{4}$$

$$\zeta_n=\max_{l\leqslant k\leqslant n-l}|T_k| \tag{5}$$

$$A_n(x)=\left[2\log\left(\frac{3n}{2l}-3\right)\right]^{-1/2}\left[x+2\log\left(\frac{3n}{2l}-3\right)+\frac{1}{2}\log\log\left(\frac{3n}{2l}-3\right)-\frac{1}{2}\log\pi\right] \quad (6)$$

则

$$\lim_{n\to\infty}P\{\zeta_n\leqslant A_n(x)\}=\exp(-2\mathrm{e}^{-x})\quad -\infty<x<\infty \quad (7)$$

由 T_k 的定义知

$$\begin{aligned}T_k=&\sqrt{\frac{l}{2}}\left(\frac{\overline{X}_{2,k}}{\hat{\sigma}_{2,k}}-\frac{\overline{X}_{1,k}}{\hat{\sigma}_{1,k}}\right)=\\&\frac{1}{\sqrt{2l}}\left(\frac{1}{\sigma}\sum_{k+1}^{k+l}X_i\left(\frac{\sigma}{\hat{\sigma}_{2,k}}-1+1\right)-\frac{1}{\sigma}\sum_{k-l+1}^{k}X_i\left(\frac{\sigma}{\hat{\sigma}_{1,k}}-1+1\right)\right)=\\&\frac{1}{\sqrt{2l}\sigma}(l\overline{X}_{2,k}-l\overline{X}_{1,k})+\frac{l\overline{X}_{2,k}}{\sqrt{2l}\sigma}\left(\frac{\sigma}{\hat{\sigma}_{2,k}}-1\right)+\frac{l\overline{X}_{1,k}}{\sqrt{2l}\sigma}\left(\frac{\sigma}{\hat{\sigma}_{1,k}}-1\right)\triangleq\\&T_{1k}+T_{2k}+T_{3k}\end{aligned} \quad (8)$$

引理 1　在定理 1 的条件下，有

$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}|T_{2k}|=0\quad a.s. \quad (9)$$

$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}|T_{3k}|=0\quad a.s. \quad (10)$$

证明　易知

$$\begin{aligned}T_{2k}=&\frac{l\overline{X}_{2,k}}{\sqrt{2l}\,\sigma}\left(\frac{\sigma}{\hat{\sigma}_{2,k}}-1\right)=\sqrt{\frac{l}{2}}\cdot\frac{1}{\sigma}\cdot\frac{\sigma^2-\hat{\sigma}_{2,k}^2}{\sigma(\sigma+\hat{\sigma}_{2,k})}\cdot\frac{\sigma}{\hat{\sigma}_{2,k}}\cdot\frac{\overline{X}_{2,k}}{\mu}\cdot\mu\\&\propto\sqrt{\frac{l}{2}}\cdot(\sigma^2-\hat{\sigma}_{2,k}^2)\quad a.s.\end{aligned}$$

所以
$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}|T_{2k}|=\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\sqrt{\frac{l}{2}}\left|\frac{1}{l}\sum_{k+1}^{k+l}(X_i-E(X_i)+E(X_i)-\overline{X}_{2,k})^2-\sigma^2\right|$$

$$\leqslant\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\sqrt{\frac{l}{2}}\left|\frac{1}{l}\sum_{k+1}^{k+l}[(X_i-E(X_i))^2-\sigma^2]\right|+\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\sqrt{\frac{l}{2}}(E(X_i)-\overline{X}_{2,k})^2\quad a.s.$$

由重对数律知

$$\begin{aligned}&\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\left(\frac{l}{2}\right)^{\frac{1}{4}}|(\overline{X}_{2,k}-E(X_i))|=\\&\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\left(\frac{l}{2}\right)^{\frac{1}{4}}\frac{\left|\sum_{i=1}^{l}X_i-lE(X_i)\right|}{\sqrt{2l\log\log l}}\frac{1}{l}\sqrt{2l\log\log l}=0\quad a.s.\end{aligned} \quad (11)$$

类似得
$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\sqrt{\frac{l}{2}}\left|\frac{1}{l}\sum_{k+1}^{k+l}[(X_i-E(X_i))^2-\sigma^2]\right|=0\quad a.s.$$

所以有
$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}|T_{2k}|=0\quad a.s.$$

同理可得
$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}|T_{3k}|=0\quad a.s.$$

引理 2　在定理 1 的条件下，有

$$\lim_{n\to\infty}|\hat{\sigma}_{2,k}-\sigma|\log n=0\quad \text{in}\quad P \quad (12)$$

证明　因为 $\lim_{n\to\infty}|\hat{\sigma}_{2,k}-\sigma|\log n\propto\lim_{n\to\infty}|\hat{\sigma}_{2,k}^2-\sigma^2|\log n\quad a.s.=$

$$\lim_{n\to\infty}\Big[\frac{1}{l}\sum_{k+1}^{k+l}(X_i-E(X_i)+E(X_i)-\overline{X}_{2,k})^2-\sigma^2\Big]\log n\leqslant$$

$$\lim_{n\to\infty}\Big[\frac{1}{l}\sum_{k+1}^{k+l}[(X_i-E(X_i))^2-\sigma^2]\Big]\log n+\lim_{n\to\infty}(E(X_i)-\overline{X}_{2,k})^2\log n\quad a.s.$$

由重对数律知,

$$\lim_{n\to\infty}|(E(X_i)-\overline{X}_{2,k})|\sqrt{\log n}=\lim_{n\to\infty}\frac{\left|\sum_{i=1}^{l}X_i-lE(X_i)\right|}{\sqrt{2l\log\log l}}\frac{1}{l}\sqrt{2l\log\log l}\sqrt{\log n}=0\quad a.s.\tag{13}$$

所以有 $\lim\limits_{n\to\infty}(E(X_i)-\overline{X}_{2,k})^2\log n=0\quad a.s.$

类似可得 $\lim\limits_{n\to\infty}\left|\frac{1}{l}\sum\limits_{k+1}^{k+l}[(X_i-E(X_i))^2-\sigma^2]\right|\log n=0\quad a.s$

所以有 $\lim\limits_{n\to\infty}|\hat{\sigma}_{2,k}-\sigma|\log n=0\quad \text{in}\quad P$

定理 1 的证明 记 $Z_i=\dfrac{X_i-E(X_i)}{\sigma}$, $S_k=\sum\limits_{i=1}^{k}Z_i$, σ^2 已知,因为 $X_1,\cdots,X_n\ i.i.d\sim\Gamma(x;\nu,\lambda)$,且 $E(e^{tX_i})<\infty$,由文献[18]知,存在一个标准布朗运动过程 $w(t)$,使得对任何 $c>0$ 有

$$\limsup_{n\to\infty}\Big\{\sup_{k\leqslant n}\frac{\left|S(k)-\frac{W(ck)}{\sqrt{c}}\right|}{\log n}\Big\}<\infty,a.s.\tag{14}$$

由 T_{1k} 的定义 $T_{1k}=\dfrac{1}{\sqrt{2l}}(S(k+l)-2S(k)+S(k-l))$

对 $k\leqslant n$,有

$$\lim_{n\to\infty}\max_{l\leqslant k\leqslant n-l}\left|T_{1k}-\frac{1}{\sqrt{2l}}\Big(\frac{1}{\sqrt{3}}W(3(k+l))-\frac{2}{\sqrt{3}}W(3(k))+\frac{1}{\sqrt{3}}W(3(k-l))\Big)\right|$$

$$\leqslant\lim_{n\to\infty}\frac{4}{\sqrt{2l}}\cdot\log n\cdot\sup_{l\leqslant k\leqslant n-l}\frac{\left|S(k)-\frac{1}{\sqrt{3}}W(3(k))\right|}{\log n}=0\quad a.s.\tag{15}$$

作一个 Gauss 过程 $Z(t)=\dfrac{1}{\sqrt{3}}\Big(W(t+\dfrac{3}{2})-2W(t)+W(t-\dfrac{3}{2})\Big)$,由式(3)及式(14)、(15)可知 $Z(t)$ 为平稳 Gauss 过程. 由文献[19]知

$$\lim_{n\to\infty}P\Big\{\max_{l\leqslant k\leqslant n-l}\left|Z\Big(\frac{3k}{2l}\Big)\right|\leqslant A_n(x)\Big\}=\exp(-2e^{-x})\qquad -\infty<x<\infty\tag{16}$$

又因为 $\lim\limits_{n\to\infty}P\{\zeta_n\leqslant A_n(x)\}=\lim\limits_{n\to\infty}P\{\max\limits_{l\leqslant k\leqslant n-l}|T_k|\leqslant A_n(x)\}=$

$$\lim_{n\to\infty}P\{\max_{l\leqslant k\leqslant n-l}|T_{1k}+T_{2k}+T_{3k}|\leqslant A_n(x)\}=$$

$$\lim_{n\to\infty}P\{\max_{l\leqslant k\leqslant n-l}|T_{1k}|\leqslant A_n(x)\}=\exp(-2e^{-x})$$

上面是在假设 $\mu=E(X_i)$, $\sigma^2=\mathrm{Var}(X_i)$ 已知情况下得到的结果,实际数据中,由于参数 μ, σ 未知,我们用 $\hat{\sigma}^2_{1,k}$, $\hat{\sigma}^2_{1,k}$ 来估计 σ^2,得

定理 2　设 $X_1, \cdots, X_n$　$i.i.d$, $X_1 \sim \Gamma(x;\nu,\lambda)$　$x > 0$. $l = l_n$ 为正整数，满足条件(3)，令

$$\tilde{\zeta}_n = \max_{l \leqslant k \leqslant n-l} |\tilde{T}_k| \tag{17}$$

$$\tilde{T}_k = \frac{1}{\sqrt{2l}}\left(\sum_{i=k+1}^{k+l} \frac{X_i}{\hat{\sigma}_{2,k}} - \sum_{i=k-l+1}^{k} \frac{X_i}{\hat{\sigma}_{1,k}}\right) \tag{18}$$

则有
$$\lim_{n\to\infty} P\{\tilde{\zeta}_n \leqslant A_n(x)\} = \exp(-2e^{-x}) \quad -\infty < x < \infty \tag{19}$$

证明由文献[10]中定理 2 和本文的引理 2 得证.

这样定理提供了一种检测原假设 $H_0: \nu_1 = \nu_2$ 的方法：选取水平 α，由方程 $\exp(-2e^{-x}) = 1-\alpha$ 得

$$x(\alpha) = -\log\left[-\frac{1}{2}\log(1-\alpha)\right] \tag{20}$$

则当且仅当

$$\tilde{\zeta}_n > A_n(x(\alpha)) \tag{21}$$

时，拒绝 H_0，即认为有变点，此时检验的渐近水平为 α.

变点位置 k_0 的估计

$$\hat{k} = \min_k \{k: |\tilde{T}_k| = \max_{l \leqslant j \leqslant n-l} |\tilde{T}_j|\} \tag{22}$$

k_0 的区间估计：$(\hat{k}-l, \hat{k}+l)$. k_0 的区间估计的长度为 $2l$，为提高估计的精度，应取较小 l_0，但它会带来两个问题：一是若 k_0 的存在要由上述检验决定时，l 过小会引起 $H_0: \nu_1 = \nu_2$ 不成立时，被接受的可能性增大；二是 l 太小时，置信系数 γ 会很低. 区间估计的置信系数 γ 有如下的估计：

$$\gamma = P(\hat{k}-l \leqslant k_0 \leqslant \hat{k}+l) \geqslant -\alpha + \Phi(\sqrt{l/2}(\sqrt{\nu_2 - \nu_1}) - A_n(x(\alpha))) \tag{23}$$

推导过程参见文献[15]，式(23)显示 l 较小时，γ 也就降低.

2　λ 至多只有一个变点的检验

这一节考虑 λ 至多只有一个变点的假设检验：

$$H_0: \lambda_1 = \lambda_2 \leftrightarrow H_1: \lambda_1 \neq \lambda_2 \tag{24}$$

同 ν 的检验讨论类似，在 $X_1, \cdots, X_n$ $i.i.d$ 的情况下，易证 $\dfrac{\overline{X}_{2,k}}{\hat{\sigma}^2_{2,k}}$，为 λ 的强相合估计，所以选用下面的 $\overline{T}_k$ 作为检验 λ 是否有变点的统计量.

定理 3　设 $X_1, \cdots, X_n$ $i.i.d$, $X_1 \sim \Gamma(x;\nu,\lambda)$, $x > 0$. $l = l_n$ 为正整数，满足条件(3)，$A_n(x)$ 定义同式(6)，令

$$\overline{T}_k = \sqrt{\frac{l}{2}}\left(\frac{\overline{X}_{2,k}}{\hat{\sigma}^2_{2,k}} - \frac{\overline{X}_{1,k}}{\hat{\sigma}^2_{1,k}}\right) \tag{25}$$

$$\overline{\zeta}_n = \max_{l \leqslant k \leqslant n-l} |\overline{T}_k| \tag{26}$$

则在 σ^2 已知时有

$$\lim_{n\to\infty} P\{\sigma \overline{\zeta}_n \leqslant A_n(x)\} = \exp(-2e^{-x}) \qquad -\infty < x < \infty \tag{27}$$

在 σ^2 未知时，若 $\hat{\sigma}$ 为 σ 的估计，且满足 $\lim\limits_{n\to\infty} |\hat{\sigma} - \sigma| \log n = 0$ in P，则仍有结论

$$\lim_{n\to\infty}P\{\tilde{\sigma}\zeta_n \leqslant A_n(x)\} = \exp(-2e^{-x}) \qquad -\infty < x < \infty \tag{28}$$

证明 同定理 1,2 完全类似.

变点检验程序、点估计和区间估计同 ν 的类似.

3 模拟

为了验证在第二第三节中提出的统计量用于检验变点位置的可行性,用 Matlab 对每组结果都进行了大量的模拟,在 $\alpha=0.05$ 的水平下,一些代表性的 ν 检验的结果见表 1,2,3. 其中 $\Gamma(x;\nu_1,\lambda_1)$,$\Gamma(x;\nu_2,\lambda_2)$ 表示变化前后的分布,k_0 表示变化前的随机产生的样本个数,即真实的变点位置;n 表示随机产生的总的样本个数,$\hat{k}$ 表示估计出来的变点位置,l 表示滑窗的长度.

从表 1 中可以看出在 λ 不变的情况下,ν_1,ν_2 之间的差距越大越容易检测出变点,即灵敏度越高;l 不能太小,随着 l 的增大,灵敏度越来越高,但是 l 也不能太大;同时 n 的取值也影响变点的检验.

表 1 λ 不变时,k_0 的估计

Tab. 1 Estimation of change point k_0 when λ is invariable

ν_1	λ_1	ν_2	λ_2	$A_n(x(\alpha))$	ζ_n	l	n	k_0	$\hat{k}$
1	1	5	1	4.298	5.689	30	2 000	1 000	1 020
				4.176	6.756	50	2 000	1 000	1 005
				4.132	9.048	60	2 000	1 000	1 021
				4.031	9.092	90	2 000	1 000	1 015
				4.273	7.575	50	3 000	1 000	1 008
				4.229	6.878	60	3 000	1 000	1 025
				4.193	9.994	70	3 000	1 000	1 003
				4.161	6.936	80	3 000	1 000	1 011
1	1	2	1	4.176	3.932	50	2 000	1 000	—
				4.112	4.987	65	2 000	1 000	1 086
				4.060	4.534	80	2 000	1 000	1 021
				4.325	3.181	40	3 000	1 000	—
				4.273	4.59	50	3 000	1 000	1 001
				4.229	4.881	60	3 000	1 000	1 053

从表 2 中看出:在 ν,λ 等比例变化时,即 Γ 分布均值不变时,用一般的均值方法无法检验出 ν 是否存在变点,由第一节提出的统计量 $\widetilde{T}_k$ 可以检验出 ν 是否存在变点,这是本文中提出用于检验变点的统计量的一大优点.

从表 3 中看出:上述方法对较小的 n 也适用,因而对实际数据的检验有很强的适用性.

另外还模拟了 $\lambda_1=0.1$,$\lambda_2=1$ 时 $\nu_1=1$,$\nu_2=1.5$ 的检验以及变点位置 k_0 较小和较大时的检验,在 λ_1 与 λ_2 不相同时,我们发现,λ_1 与 λ_2 是否相同对 ν 是否存在变点的检验几乎没有影响;变点位置 k_0 较小和较大时,检验的灵敏度相对要低一些. 对 λ 的模拟有类似的结论.

以上是对 ν,λ 至多一个变点的检验,由于我们用的是滑窗方法,所以同样适用于有多个

变点的 Γ- 分布序列的检验，只需先找到第一个，依次再找第二个，第三个…

变点 k_0 强相合的将另行讨论.

表 2 μ/λ 不变时，k_0 的估计

Tab. 2 Estimation of change point k_0 when μ/λ invariable

ν_1	λ_1	ν_2	λ_2	$A_n(x(\alpha))$	ζ_n	l	n	k_0	$\hat{k}$
1	0.5	2	1	4.176	5.025	50	2 000	1 000	1 010
				4.132	6.194	60	2 000	1 000	1 005
				4.061	6.971	80	2 000	1 000	1 009
				4.325	4.516	40	3000	1 000	1 004
				4.229	5.983	60	3 000	1 000	1 010
				4.161	6.755	80	3 000	1 000	1 012

表 3 n 较小时，k_0 的估计

Tab. 3 Estimation of change point k_0 when n is small

ν_1	λ_1	ν_2	λ_2	$A_n(x(\alpha))$	ζ_n	l	n	k_0	$\hat{k}$
1	0.1	5	1	4.003 6	5.409 2	10	200	100	109
				3.897 7	5.309 4	15	200	100	103
				3.820 0	8.403 5	20	200	100	102
				4.060 3	7.227 6	12	300	200	209
				3.971 3	5.139 1	17	300	200	204
				3.928 9	7.036 3	20	300	200	204

参考文献

[1] Braun J V, *et al*. Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation[J]. Biometrika, 2000, 87(2): 301-314.

[2] Christensen J, Rudemo M. Multiple change-point analysis of disease incidence rates[J]. Prve. Vet. Med. , 1996, 26: 53-76.

[3] Chernoff H, Zacks S. Estimating the cuurrent mean of normal distribution which is subjected tochange in time[J]. Ann. Math. Statistics. , 1964, 35: 999-1018.

[4] Haccou P, Meelis E. Asymptotic distribution of the likelihood ratio test for the changpoint problem for exponentially random variables [J]. Stochastic Processes and Appl, 1987, 27: 121-139.

[5] Daniel Barry D, Hartigan J A. A Bayesian analysis for change point problems(in theory and methods)[J]. JASA, 1993, 88: 309-319.

[6] Siegmund D, Venkatraman E S. Using the gneralized likelihood ratio statistic for sequential detection ofa change point[J]. Annals of Statistics, 1995, 23(1): 255-271.

[7] Krishnaiah P R, Miao B Q. Review about estimates of change-point[J]. Handbook of Statistics, 1988, 7: 375-402.

[8] CsörgöM, Horvath P. Nonparametric methods for change point problems[A]. Handbook of Statistics, Control and Reliabality [C]. New York: North-Holland, 1988, 7.

[9] 缪柏其. 关于只有一个变点模性的非参数推断[J]. 系统科学与数学, 1993, 13: 132-140.

[10] Ferger D. Nonparametric tests for nonstandard change-point problems (in Nonparametrics)[J]. Annals of Statistics, 1995, 23(5): 1848-1861.

[11] 陈希孺. 只有一个转变点的模型的假设检验和区间估计[J]. 中国科学 A 辑, 1988, 8: 817-

827 .

[12] Miao B Q. Inference in a model with at most one-slope changepoint [J]. J of Multivariate Analysis, 1988, 27: 375-391.

[13] 谭智平. 至多只有一个变点模性的统计推断[J]. 应用概率统计, 1996, 12: 43-54.

[14] 缪柏其, 赵林城, 谭智平. 关于变点个数及位置的检测和估计[J]. 应用数学学报, 2003, 26: 26-39.

[15] Hsu D A. Detecting shfits of parameter in gamma sequences with applications to stock price and Air Traffic Flow Analysis[J]. Journal of the American Statistical Association, 1979, 74: 31-40.

[16] Kimber A C. Testing upper and lower outlies pairs in gamma samples[J]. Commun. Statist. Sim and Comp, 1988, 17: 1055-1072.

[17] 马超群, 罗虎, 等. Grmma 分布参数多个转变点的 Bayes 推断[J]. 湖南大学学报, 1999, 26: 107-112.

[18] Csörgö M, Révész P. Strong approximations in probility and statistics[M]. New York: Academic press, 1981.

[19] Qualls C, Watanable H. Asymptotic properties of Gaussian processes[J]. Ann. Math. Statist, 1972, 43: 580-596.

[20] Pons O. Estimation in a Cox regression model with a change-point at an unknown time[J], Statistics, 2002, 36: 101-124.

[21] Pons O. Estimation in a Cox regression model with a change-point according to a threshold in a covariate[J]. Annals of Statistics, 2003, 31(2): 442-463.

Inference in Γ-distribution With at Most One Change-point

TAN Chang-chun, MIAO Bai-qi

(*Dept of Statistics and Finance*, *USTC*, *Hefei* 230026, *China*)

Abstract: Let $X_1, \cdots, X_{k_0}, X_{k_0+1}, \cdots, X_n$ be independent random variables such that $X_1, \cdots, X_{k_0}$ $i.i.d \sim \Gamma(\nu_1, \lambda_1)$ and $X_{k_0+1}, \cdots, X_n$ $i.i.d \sim \Gamma(\nu_2, \lambda_2)$. k_0 or k_0/n is called change-point. In this paper the change-point of Γ-distribution with two parameters is discussed. With the help of the theory of Gaussian process and the method of slipping window, the distribution of the statistics proposed can be approximated by the first type of extremal distribution. The detection procedures are also proposed by the statistics presented.

Key words: Γ-distribution; change-point; interval estimation; slipping window.