Advances in Applied Mathematics 应用数学进展, 2016, 5(3), 321-326
Published Online August 2016 in Hans. http://www.hanspub.org/journal/aam
http://dx.doi.org/10.12677/aam.2016.53040

Hans汉斯

A Method of Detecting Multiple Change Point for Normal Distribution Process

Huihui Shen[1,2]

[1]Hubei University of Economics, Wuhan Hubei
[2]China University of Geosciences, Wuhan Hubei
Email: sophy0209@126.com

Received: Jul. 20th, 2016; accepted: Aug. 12th, 2016; published: Aug. 15th, 2016

Abstract

The problem of structure model occurs multiple change points in the economic system of mathematical models. In this paper, we give the detection method for change point problems about the variance changes. We combine the Bayesian method with the maximum likelihood method on the detection about the variance multiple change points in the same mean. The elimination extra parameters can make use of Bayesian method; the maximum likelihood method can avoid the unknown problems of the prior distribution information of the change points number. It is a practical method.

Keywords

Change Point, Bayesian Method, Maximum Likelihood Method, Prior Distribution, Likelihood Density Function

正态分布下多个方差转变点的检测与方法探讨

沈卉卉[1,2]

[1]湖北经济学院，湖北 武汉
[2]中国地质大学，湖北 武汉
Email: sophy0209@126.com

收稿日期：2016年7月20日；录用日期：2016年8月12日；发布日期：2016年8月15日

文章引用: 沈卉卉. 正态分布下多个方差转变点的检测与方法探讨[J]. 应用数学进展, 2016, 5(3): 321-326.
http://dx.doi.org/10.12677/aam.2016.53040

摘　要

经济系统数学模型中含有多个转变点的结构模型问题，本文对在均值相同情况下，方差多个转变点的检测方法采取的是将贝叶斯方法和极大似然方法结合起来，利用贝叶斯方法消去多余参数，极大似然方法可以回避转变点个数的先验分布信息未知的问题，给出有效的检测方法。

关键词

转变点，贝叶斯方法，极大似然方法，先验分布，似然密度函数

1. 引言

转变点问题一直是统计学前沿研究的热点问题，具有重要的应用价值，尤其在经济学和金融学方面有广泛的应用[1]。在经济飞速发展的现代社会，需要处理经济、金融的问题也越来越多。经济周期中的变点分析一直都是经济学家和统计学家关心的问题，因为研究经济周期中的变点问题对我们及时调整经济政策，使经济稳定发展有很大帮助。

转变点的定义：按统计学的定义，对于某一随机变量序列，如果存在一个时间点，在这个点以前的序列服从一种概率分布而在这个时间点以后的序列则服从另一种概率分布(或同一种概率分布而参数不同)，那么在这个序列中存在一个转变点。因此，一个转变点问题就包含以下两方面的内容[1]：① 确定是否存在变化，② 估计未知转变点的个数及位置。

一般地，估计转变点的方法有很多，常见的方法有 Schwarz 信息准则方法、二分分段法、似然比检验、(加权)最小二乘法、非参数方法、累加和方法[2]、贝叶斯方法[3]-[5]、极大似然方法等。Schwarz (1978)提出了信息准则的方法用于检测转变点的存在性时，无需导出其复杂的分布函数，来估计变点的个数和位置[6]。Vostrikova (1981)提出的二分分段法能够同时检测出转变点的数量和它们的位置，并能够节约大量运算时间[7]。文献[8]结合前人方法，将二分分段法和信息准则方法结合来检测均值转变点情况。Stergios Fotopoulos, Venkata Jandhyala (2001) [9]针对独立同分布于指数分布的一系列的随机变量，用极大似然估计的方法估计分布参数的转变点问题，但其方法是寻找转变点位置的一个近似结果。Inclan 和 Tao (1994) [10]用累积平方和的方法解决多个方差转变点的问题，在某种程度上类似于二分分段方法，该方法是节省了计算量，但是检测转变点必须整个时间序列分割，难以保证转变点是全局意义上的。如今 Yuehjen E. Shao 和 Ke-Shan Lin (2015) [11]针对随机变量分布在未知的情况下，提出了一种人工神经网络(ANN)的新方法。

然而总体的参数会是随时间而变的，关于分布参数的转变点[12]，文献[12] [13]用极大似然的方法来估计参数的转变点。一般转变点问题，常常考虑总体均值和方差的变化情况，且是在假设检验的框架下进行的，基于模型的总体分布是正态[14]，这样一来变点问题的推断等价于均值或者方差变化的检测。

2. 方差转变点的检测

Shao 和 Hou (2006) [15]运用 S—控制图和极大似然方法相结合来估计服从 gamma 分布的随机变量的转变点问题。Wenzhi Zhao, Zheng Tian, Zhiming Xia (2010) [16]针对有长期记忆的线性过程用比值判别法来检测方差转变点，但是此方法需要一定的限制假设条件。文献[17]结合了 X 控制图与贝叶斯估计方法，文献[18]的应用实证中也运用贝叶斯方法来定位找到转变点的位置，通过一系列模拟。结果表明，贝

叶斯估计量与之前的信息更准确和更精确。

鉴于以上研究情况，对于正态随机变量的方差多个转变点问题，我们采取将 Bayes 方法和极大似然方法相结合，先利用经验贝叶斯方法消去多余参数，然后利用极大似然方法寻找转变点位置。

设随机正态变量时间序列$\{x_t\}$，$(t=1,2,\cdots,T)$令$x_1,x_2,\cdots,x_T$，$x_t\sim N\left(0,\sigma_t^2\right)$，且

$$\sigma_t^2=\tau_0^2,\quad 1\le t\le k_1$$

$$\sigma_t^2=\tau_1^2,\quad k_1<t\le k_2$$

$$\sigma_t^2=\tau_2^2,\quad k_2<t\le k_3$$

$$\cdots$$

$$\sigma_t^2=\tau_m^2,\quad k_m<t\le T$$

其中$1\le k_1<k_2<k_3<\cdots<k_m<T$表示的是方差转变点的下标，假定有$m$个转变点，转变点下标$k=\left(k_1,k_2,\cdots,k_m\right)'$，$m=1,2,\cdots,T-1$。为表达方便，令$k_0=0,k_{m+1}=T$。转变点将时间序列分割成$m+1$个时间段，每个时间段里$x_t$分布相同，$x_t\sim N\left(0,\tau_j^2\right)$，$k_j<t\le k_{j+1}$，$j=0,1,\cdots,m$。在第$j$个时间段中，$x_t$的均值是 0，方差为$\tau_j^2$，$m$的观测数是$d_j=k_{j+1}-k_j$，$j=0,1,\cdots,m$。现在的主要问题是如何根据样本$x=\left(x_1,x_2,\cdots,x_T\right)'$去估计转变点的位置$k=\left(k_1,k_2,\cdots,k_m\right)'$以及个数，转变点个数实际上是转变点位置向量$k$的维数。

我们检测的是方差转变点，而方差$DX=\int_{-\infty}^{+\infty}\left(x-EX\right)^2 f\left(x\right)\mathrm{d}x$是样本分布的密度函数，所以将方差转变点的检测和分析问题最终可看作是样本密度函数的估计问题，然后利用极大似然方法估计其转变点的位置。

因为$x_t\sim N\left(0,\sigma_t^2\right)$，$(t=1,2,\cdots,T)$所以$x=\left(x_1,x_2,\cdots,x_T\right)'$的联合分布密度$f\left(x\,|\,\sigma,k,m\right)$，以及似然函数$L\left(x_1,x_2,\cdots,x_n;\sigma\right)$为：

$$L\left(x_1,x_2,\cdots,x_n;\sigma\right)=f\left(x_1;\sigma\right)f\left(x_2;\sigma\right)\cdots f\left(x_n;\sigma\right)=L\left(x;\sigma,k,m\right)=f\left(x\,|\,\sigma,k,m\right)$$

$$\begin{aligned}L\left(x;\sigma,k,m\right)&=f\left(x\,|\,\sigma,k,m\right)=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{x_1^2}{2\sigma_1^2}}\times\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{x_2^2}{2\sigma_2^2}}\times\cdots\times\frac{1}{\sqrt{2\pi}\sigma_T}\mathrm{e}^{-\frac{x_T^2}{2\sigma_T^2}}\\&=\left(\frac{1}{\sqrt{2\pi}}\right)^T\prod_{t=1}^{T}\frac{1}{\sigma_t}\cdot\mathrm{e}^{-\frac{1}{2}\frac{x_t^2}{\sigma_t^2}}=\frac{1}{\left(\sqrt{2\pi}\right)^{\mathrm{T}}\sigma_1\sigma_2\cdots\sigma_T}\mathrm{e}^{\sum_{t=1}^{T}\left(-\frac{x_t^2}{2\sigma_t^2}\right)}\\&=\frac{1}{\left(\sqrt{2\pi}\right)^T\sigma_1\sigma_2\cdots\sigma_T}\mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T}\frac{x_t^2}{\sigma_t^2}}=\left(\frac{1}{\sqrt{2\pi}}\right)^{\mathrm{T}}\left(\sigma_1\sigma_2\cdots\sigma_T\right)^{-1}\mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T}\frac{x_t^2}{\sigma_t^2}}\end{aligned}$$

其中$\sigma=\left(\sigma_1,\sigma_2,\cdots,\sigma_T\right)'$，$k=\left(k_1,k_2,\cdots,k_m\right)'$，$m$为变点的个数。

或者由 Inclan (1993)给出的如下联合分布密度函数[2]：

$$f\left(x\,|\,\tau,k,m\right)=\left(\frac{1}{\sqrt{2\pi}}\right)^T\prod_{j=0}^{m}\frac{1}{\tau_j^{d_j}}\cdot\mathrm{e}^{-\frac{1}{2\tau_j^2}\sum_{t=k_j+1}^{k_{j+1}}x_t^2}=L\left(x;\tau,k,m\right)\tag{1}$$

其中$\tau=\left(\tau_0,\tau_1,\cdots,\tau_m\right)'$，$k=\left(k_1,k_2,\cdots,k_m\right)'$，$m$为变点的个数。$\tau_j^2=\sigma_t^2$，$d_j=k_{j+1}-k_j$，

$(t=1,2,\cdots,T;\ j=0,1,\cdots,m)$，$k_j<t\le k_{j+1}$，令 $k_0=0$，$k_{m+1}=T$。

我们要解决问题的关键是参数 k，m，也就是转变点位置和转变点个数，而 τ 相对不重要，可看作是多余的参数。如果直接用极大似然方法处理多余参数，存在很多困难，因为 σ 和 τ 都不知道，而贝叶斯方法就可以处理这种问题。运用贝叶斯公式可得到参数的后验分布密度函数 $f(k,m,\tau\mid x)$，然后通过后验分布密度 $f(k,m,\tau\mid x)$ 对 τ 积分可以得到 k，m 的边缘分布密度。

2.1. 贝叶斯检测

要运用贝叶斯方法消去多余参数 τ，必须先构造 τ 的先验分布。在没有 τ 的较好先验信息的情况下，经验贝叶斯方法可以解决此问题。经验贝叶斯的思想是从样本信息中构造 τ 的先验分布 $f(\tau\mid k,m)$。

因为 $x_t\sim N(0,\sigma_t^2)$，$\tau_j^2=\sigma_t^2$，$k_j<t\le k_{j+1}$，$d_j=k_{j+1}-k_j$，在 k，m 给定的条件下有：

$$\sum_{t=k_j+1}^{k_{j+1}}\frac{x_t^2}{\tau_j^2}\sim\chi^2(d_j) \tag{2}$$

这样可以根据式(2)导出 τ_j 的分布，并把它作为 τ_j 的先验分布。于是就得到 τ 的先验分布 $f(\tau\mid k,m)$。这个先验分布用到了样本的信息，因此我们用的是贝叶斯方法。结合式(1)得：

令 $y_j=\sum_{t=k_j+1}^{k_{j+1}}\frac{x_t^2}{\tau_j^2}\sim\chi^2(d_j)$，$y=(y_0,y_1,\cdots,y_m)'$，$j=0,1,\cdots,m$，则

$$\tau_j=\left(\frac{1}{y_j}\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{\frac{1}{2}},\quad f(y_j)=\frac{1}{2^{\frac{d_j}{2}}\Gamma\left(\frac{d_j}{2}\right)}y_j^{\frac{d_j}{2}-1}\mathrm{e}^{-\frac{y_j}{2}},\ (y_j>0)$$

$f(x,\tau\mid k,m)=f(x\mid\tau,k,m)f(\tau\mid k,m)$，等式两边同时对 τ 积分得：

$$\int f(x,\tau\mid k,m)\mathrm{d}\tau=f(x\mid k,m)=\int f(x\mid\tau,k,m)f(\tau\mid k,m)\mathrm{d}\tau=E_{\tau\mid k,m}\left[f(x\mid\tau,k,m)\right]$$

由式(1)知：$f(x\mid\tau,k,m)=\left(\frac{1}{\sqrt{2\pi}}\right)^T\prod_{j=0}^{m}\frac{1}{\tau_j^{d_j}}\cdot\mathrm{e}^{-\frac{1}{2\tau_j^2}\sum_{t=k_j+1}^{k_{j+1}}x_t^2}$，则似然函数为：

$$L(x;k,m)=f(x\mid k,m)=E_y\left[f(x\mid\tau,k,m)\right]=\int_0^{+\infty}f(x\mid\tau,k,m)f(y_j)\mathrm{d}y_j\quad\left(\tau_j=\left(\frac{1}{y_j}\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{\frac{1}{2}}\right)$$

$$=\int_0^{+\infty}\left(\frac{1}{\sqrt{2\pi}}\right)^T\prod_{j=0}^{m}\frac{1}{\tau_j^{d_j}}\cdot\mathrm{e}^{-\frac{1}{2\tau_j^2}\sum_{t=k_j+1}^{k_{j+1}}x_t^2}\times\frac{1}{2^{\frac{d_j}{2}}\Gamma\left(\frac{d_j}{2}\right)}\cdot y_j^{\frac{d_j}{2}-1}\cdot\mathrm{e}^{-\frac{y_j}{2}}\mathrm{d}y_j$$

$$=\int_0^{+\infty}\left(\frac{1}{\sqrt{2\pi}}\right)^T\prod_{j=0}^{m}\left(\frac{1}{y_j}\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\cdot\mathrm{e}^{-\frac{y_j}{2}}\frac{1}{2^{\frac{d_j}{2}}\Gamma\left(\frac{d_j}{2}\right)}\cdot y_j^{\frac{d_j}{2}-1}\cdot\mathrm{e}^{-\frac{y_j}{2}}\mathrm{d}y_j$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^T\prod_{j=0}^{m}\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\int_0^{+\infty}y_j^{\frac{d_j}{2}}\cdot\mathrm{e}^{-\frac{y_j}{2}}\frac{1}{2^{\frac{d_j}{2}}\Gamma\left(\frac{d_j}{2}\right)}\cdot y_j^{\frac{d_j}{2}-1}\cdot\mathrm{e}^{-\frac{y_j}{2}}\mathrm{d}y_j$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^{T}\prod_{j=0}^{m}\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\int_0^{+\infty}\cdot\frac{1}{2^{\frac{d_j}{2}}\Gamma\left(\frac{d_j}{2}\right)}\cdot y_j^{d_j-1}\cdot e^{-y_j}dy_j$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^{T}\prod_{j=0}^{m}\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\frac{1}{2^{\frac{d_j}{2}}}\cdot\int_0^{+\infty}\frac{1}{\Gamma\left(\frac{d_j}{2}\right)}\cdot y_j^{d_j-1}\cdot e^{-y_j}dy_j$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^{T}\prod_{j=0}^{m}\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\cdot 2^{-\frac{d_j}{2}}\cdot\frac{\Gamma\left(d_j\right)}{\Gamma\left(\frac{d_j}{2}\right)}=\left(\frac{1}{\sqrt{2\pi}}\right)^{T}\prod_{j=0}^{m}2^{-\frac{d_j}{2}}\cdot\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\cdot\frac{\Gamma\left(d_j\right)}{\Gamma\left(\frac{d_j}{2}\right)}$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^{T}\prod_{j=0}^{m}2^{-\frac{k_{m+1}-k_0}{2}}\cdot\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\cdot\frac{\Gamma\left(d_j\right)}{\Gamma\left(\frac{d_j}{2}\right)}=\left(\frac{1}{\sqrt{2\pi}}\right)^{T}\cdot 2^{-\frac{T}{2}}\prod_{j=0}^{m}\left(\sum_{t=k_j+1}^{k_{j+1}}x_t^2\right)^{-\frac{d_j}{2}}\cdot\frac{\Gamma\left(d_j\right)}{\Gamma\left(\frac{d_j}{2}\right)}$$

$$\left(\Gamma\left(d_j\right)=\int_0^{+\infty}y_j^{d_j-1}\cdot e^{-y_j}dy_j\ ,\ \ d_j=k_{j+1}-k_j\ ,\ \ k_0=0\ ,\ \ k_{m+1}=T\right)$$

2.2. 极大似然方法检测

由此可知，要运用贝叶斯方法，还需要知道 m 的先验分布，由于 m 先验信息未知，且 m 的先验分布对问题求解起到至关重要的作用。因此，我们对上面求出的联合密度函数 $f\left(x\mid k,m\right)$ 也即是似然函数 $L\left(x;k,m\right)$，用极大似然方法求出 k， m，这样就避开了 m 的先验分布未知的问题。先给定 m，然后通过将 $f\left(x\mid k,m\right)$ 表达式最大化来确定 $\hat{k}\mid m$ 作为转变点个数为 m 时的转变点位置的估计。

以上过程中我们并不是只用贝叶斯一种方法，而是将贝叶斯方法和极大似然方法结合起来，先用贝叶斯方法消去多余参数 τ，然后再用极大似然方法寻找转变点位置 k。这样既消去了多余参数；又回避了转变点在先验分布上未知的问题；而且极大似然方法的好处是只需要在解空间中找出似然密度函数最大的解，从而使方差多个转变点的检测问题得以解决。

在实际的经济模型应用中，我们经常会碰到既有多余参数又没有什么强的先验信息的情况。单纯利用贝叶斯方法或极大似然方法效果都不理想，这样把贝叶斯方法和极大似然方法结合起来利用各自方法的优点来解决各种问题，是一个较好的选择。可以先用贝叶斯方法消去多余参数，然后再利用极大似然方法估计我们感兴趣的参数。

基于模型总体分布是正态分布下的均值和方差多变点的情况，在前期的研究基础上讨论了同方差下的均值转变点检测问题[8]，本文讨论同均值下的方差转变点估计问题，还有不同方差下的均值转变点，以及均值和方差同时有转变点的估计问题[12]。以及对于经济模型中参数分布为指数分布、Poisson 分布和 t 分布等情况的转变点结构模型问题，都有待我们去进一步研究和探讨。由于转变点问题常常涉及到非独立随机变量的分布问题，这是非常难以处理的，这个问题的研究在理论上难度很大，富有挑战性。

基金项目

转变点在经济领域中的数学建模与应用探讨(Q20121902)，湖北省教育厅科学技术研究计划优秀中青年人才项目。

参考文献 (References)

[1] Chen, J. and Gupta, A.K. (2000) Parametric Statistical Change Point Analysis. Birkhauser, Boston. http://dx.doi.org/10.1007/978-1-4757-3131-6

[2] Inclan, C. (1993) Detection of Multiple Changes of Variance Using Posterior Odds. *Journal of Business & Economic Statistics*, **11**, 289-300.

[3] 孙军, 姜诗意, 李宏纲. 经济序列变点的 Bayes 分析[J]. 统计研究, 2001(8): 27-30.

[4] Son, Y.S. and Kim, S.K. (2005) Bayesian Single Change Point Detection in a Sequence of Multivariate Normal Observations. *Statistics*, **39**, 373-387. http://dx.doi.org/10.1080/02331880500315339

[5] Chen, J., Yigiter, A. and Chang, K.C. (2011) A Bayesian Approach to Inference about a Change Point Model with Application to DNA Copy Number Experimental Data. *Journal of Applied Statistics*, **38**, 1899-1913. http://dx.doi.org/10.1080/02664763.2010.529886

[6] Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464. http://dx.doi.org/10.1214/aos/1176344136

[7] Vostrikova, L.J.U. (1981) Detecting "Disorder" in Multidimensional Random Processes. *Soviet Mathematics Doklady*, **24**, 55-59.

[8] 沈卉卉. 转变点判定及其在我国居民消费中的应用分析[J]. 商业时代, 2010 (3): 9-10.

[9] Jandhyala, V.K., Fotopoulos, S.B. and Hawkins, D.M. (2002) Detection and Estimation of Abrupt Changes in the Variability of a Process. *Computational Statistics and Data Analysis*, **40**, 1-19. http://dx.doi.org/10.1016/S0167-9473(01)00108-6

[10] Inclan, C. and Tiao, G.C. (1994) Use of Cumulative Sum of Squares for Retrospective Detection of Changes of Variances. *Journal of American Statistcal Association*, **89**, 913-923.

[11] Shao, Y.E. and Lin, K.-S. (2015) Change Point Determination for an Attribute Process Using an Artificial Neural Network-Based Approach. *Discrete Dynamics in Nature and Society*, **2015**, Article ID: 892740.

[12] Fotopoulos, S. and Jandhyala, V. (2001) Maximum Likelihood Estimation of a Change-Point for Exponentially Distributed Random Variables. *Statistics & Probability Letters*, **2001**, 423-429. http://dx.doi.org/10.1016/S0167-7152(00)00185-1

[13] Niaki, S.T.A. and Khedmati, M. (2014) Monotonic Change-Point Estimation of Multivariate Poisson Processes Using a Multi-Attribute Control Chart and MLE. *International Journal of Production Research*, **52**, 2954-2982. http://dx.doi.org/10.1080/00207543.2013.857797

[14] Pignatiello, J.J. and Samuel, T.R. (2001) Estimation of the Change Point of a Normal Process Mean in SPC Applications. *Journal of Quality Technology*, **33**, 82-95.

[15] Shao, Y.E., Hou, C.D. and Wang, H.J. (2006) Estimation of the Change Point of a Gamma Process by Using the S Control Chart and MLE. *Journal of the Chinese Institute of Industrial Engineers*, **23**, 207-214. http://dx.doi.org/10.1080/10170660609509010

[16] Zhao, W.Z., Tian, Z. and Xia, Z.M. (2010) Ratio Test for Variance Change Point in Linear Process with Long Memory. *Stat Papers*, **51**, 397-407. http://dx.doi.org/10.1007/s00362-009-0202-3

[17] Monfared, M.E.D. and Lak, F. (2013) Bayesian Estimation of the Change Point Using $\bar{X}$ Control Chart. *Communications in Statistics—Theory and Methods*, **42**, 1572-1582. http://dx.doi.org/10.1080/03610926.2011.594536

[18] Plummer, P.J. and Chen, J. (2014) A Bayesian Approach for Locating Change Points in a Compound Poisson Process with Application to Detecting DNA Copy Number Variations. *Journal of Applied Statistics*, **41**, 423-438. http://dx.doi.org/10.1080/02664763.2013.840272