Journal of Statistical Planning and Inference 143 (2013) 434–440

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Finite sample distributions of statistics in change point analysis

Reza Habibi

Department of Statistics, Central Bank of Iran, Iran

ARTICLE INFO

Article history:
Received 30 September 2009
Received in revised form
12 March 2012
Accepted 19 June 2012
Available online 6 July 2012

Keywords:
Beta distribution
Change point
Cusum
Multivariate normal
Non-central distribution

ABSTRACT

This paper is concerned with derivation of finite sampling distributions of some statistics which appear frequently in change point analysis. The exact distribution of cusum test statistic is approximated by two methods. Approximations are presented and their accuracies are measured. We first consider the change point in mean problem and we study the exact distribution of change point estimator. Finally, we consider the change point in variance case.

1. Introduction

It is very important for economic policy to identify change points in economic and financial series. For example, Hillebrand and Schnabl (2003) studied change point detection in volatility of Japanese foreign exchange intervention under GARCH modeling. During the last four decades, different methods are employed for detecting change points. For a comprehensive review, see Chen and Gupta (2000). This problem for a set of normally distributed observations may be described as follows.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent normal random variables with mean $\eta_i, i=1,2,\ldots,n$ and common variance σ^2 such that

$$\eta_i = \begin{cases} \theta_0, & i=1,2,\ldots,k_0, \\ \theta_0+\delta, & i=k_0+1,\ldots,n. \end{cases}$$

Without loss of generality, we assume that the size of change δ is negative. It is interested to test H_0 versus H_1 where

$$\begin{cases} H_0 : k_0 = n, \\ H_1 : 1 \le k_0 \le n-1. \end{cases}$$

The null hypothesis implies that there is no change point while under H_1, there is a shift in means of observations. The testing procedure depends on whether the nuisance parameter σ^2 is known or unknown. Let

$$S_k = \sum_{i=1}^{k} (X_i - \overline{X}), \quad k=1,\ldots,n-1,$$

E-mail address: rezahabibi2681@yahoo.com

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.06.016

where $\overline{X}=(1/n)\sum_{i=1}^{n}X_i$. Following Hinkley (1971), statistic M is the cusum procedure for testing H_0 against H_1, where

$$M=\max_{1\le k\le n-1}S_k.$$

Large values of M reject the null hypothesis. Under H_1, parameter k_0 is estimated by the maximizer of S_k (denote by $\widehat{k}$) which is defined by

$$S_{\widehat{k}}=M.$$

The finite sample distributions of M and $\widehat{k}$ are too complicated. Therefore, it is very important to approximate these distributions by methods of computational statistics. Conniffe and Spencer (2000) (hereafter CS) proposed a central chi-squared (beta) approximation for the null distribution of cusum statistic when σ^2 is known (unknown). However, they did not study the finite sample distribution of M under the alternative hypothesis. This distribution is necessary for power analysis. They also did not consider the finite distribution of $\widehat{k}$.

This paper is organized as follows. In Section 2, we approximate the alternative finite sample distribution of M using non-central chi-square (doubly non-central beta) distribution when σ^2 is known (unknown). Then, as alternative method, noticing that M is the maximum of a multivariate normal distribution, we apply Genz's (1992) numerical approach to solve the finite sample distribution problem. In Section 3, we derive the exact distribution of $\widehat{k}$, again using Genz's method. Finally, in Section 4, we consider the same problems for shift in variance case.

2. Distribution of test statistic

Here, we present two methods to approximate the finite sample distribution of M under the H_1. Both methods are given and their accuracies are measured. It is seen that Genz's approximation (the second method) is more accurate than the chi-square approximation (first method). Although, the difference is not too much. However, Genz's method might take long computing time for large n.

2.1. Chi-square approximation

First, let $\sigma^2=1$. Following CS, the chi-square approximation reduces to the fitting a non-central chi-square law $\chi^2_{d,\lambda}$ for

$$T=\frac{4}{n}M^2,$$

with non-integer degrees of freedom and a non-centrality parameter λ. The matching moment estimates of d and λ are

$$d=2\mu_T-\frac{\sigma_T^2}{2}\quad\text{and}\quad\lambda=\frac{\sigma_T^2}{2}-\mu_T,$$

where μ_T and σ_T^2 are the mean and variance of T under H_1. In practice, μ_T and σ_T^2 are unknown and they are estimated using a Monte Carlo study with $R=5000$ repetitions. As follows, we survey the accuracy of non-central chi-square approximation.

Denote the Monte Carlo quantiles of T by q_α, $\alpha=0.01,\ldots,0.99$. Let $F_{\widehat{d},\widehat{\lambda}}$ be the estimated cdf of non-central chi-square distribution. The approximated cdf may be evaluated by measuring some criteria such as $\max|e(\alpha)|$, $\min|e(\alpha)|$ and $\text{med}|e(\alpha)|$ (the median of absolute errors) where

$$e(\alpha)=F_{\widehat{d},\widehat{\lambda}}(q_\alpha)-\alpha.$$

Since the distribution of T does not depend on θ_0, we let $\theta_0=0$. We report parameter estimates and error analysis results in Table 1. Here, we let $n=10(10)50$ and $\delta=-1,-2$. The k_0 is selected at random. As it can be seen that the errors are negligible. We also propose a graphical comparison between the empirical density of T with density of distribution $F_{\widehat{d},\widehat{\lambda}}$ for various values of n,δ,k_0 (see Fig. 1).

Table 1
$\widehat{d},\widehat{\lambda}$, max, min, med of absolute errors; $\sigma^2=1$.

n	δ	k_0	$\widehat{d}$	$\widehat{\lambda}$	$\max\lvert e\rvert$	$\min\lvert e\rvert$	med $\lvert e\rvert$
10	−1	2	2.312	0.462	0.0265	0.000027	0.0034
10	−2	7	4.850	3.890	0.0091	0.000062	0.0033
20	−1	9	3.161	3.980	0.0089	0.000028	0.0029
20	−2	16	6.845	0.551	0.0143	0.000283	0.0064
30	−1	13	3.233	6.451	0.0064	0.000065	0.0025
30	−2	17	1.521	28.98	0.0045	0.000008	0.0015
40	−1	14	4.691	5.402	0.0122	0.000129	0.0031
40	−2	11	12.05	15.65	0.0077	0.000057	0.0030
50	−1	35	5.799	5.607	0.0142	0.000037	0.0041
50	−2	9	15.81	4.336	0.0151	0.000244	0.0063