Journal of Statistical Planning and Inference 139 (2009) 1921 – 1936

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Truncated sequential change-point detection based on renewal counting processes II

Allan Gut[a,*], Josef Steinebach[b]

[a] *Department of Mathematics, Uppsala University, Box 480, SE-751 06 Uppsala, Sweden*
[b] *Mathematisches Institut, Universität zu Köln, Weyertal 86-90, D-50 931 Köln, Germany*

A R T I C L E I N F O

Article history:
Received 17 June 2008
Received in revised form
25 August 2008
Accepted 25 August 2008
Available online 6 September 2008

MSC:
primary 62G10
62G20
62G32
62L10
secondary 60F17
60G40
60G70
60J65
60K05

Keywords:
Change-point
Training period
Extreme value asymptotics
First passage time
Increments
Renewal counting process
Sequential test
Stopping time
Strong approximation
Wiener process

A B S T R A C T

The standard approach in change-point theory is to base the statistical analysis on a sample of fixed size. Alternatively, one observes some random phenomenon sequentially and takes action as soon as one observes some statistically significant deviation from the "normal" behaviour. The present paper is a continuation of Gut and Steinebach [2002. Truncated sequential change-point detection based on renewal counting processes. Scand. J. Statist. 29, 693–719] the main point being that here we look in more detail into the behaviour of the relevant stopping times, in particular the time it takes from the actual change-point until the change is detected, more precisely, we prove asymptotics for stopping times under alternatives.

1. Introduction

A typical situation in a series of observations is that if everything is in order, then the observations follow some kind of common pattern, whereas if something goes astray at some time point—known or unknown, then, from there on, the observations follow a different pattern. A simple and natural example is a possible change in the mean—known or unknown—caused by a failure in some mechanism in a (manufacturing) process. One obviously wishes to find out as soon as possible if something goes wrong

* Corresponding author. Tel.: +46 18 471 3182; fax: +46 18 471 3201.
E-mail addresses: allan.gut@math.uu.se (A. Gut), jost@math.uni-koeln.de (J. Steinebach).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.08.021

in order to take appropriate action, and, at the same time, minimize the probability of taking action if everything is in order. In this setting one talks about the case with *at most one change-point*, the so-called AMOC problem, which was introduced by Page (1954, 1955) in the context of quality control/control charts.

There are two basic approaches to tackle this kind of problems. One is to take a sample of fixed size and perform the analysis. The other one, which is our approach, is to observe the process sequentially and to take action as soon as one observes a significant deviation from the normal pattern. Most of the sequential analysis so far is based on (so-called) "open-end" procedures in which there is no pre-determined endpoint of the observation period. There is less work on truncated ("closed-end") versions which are considered here. The latter typically make use of certain extreme value asymptotics to determine their critical values. See, for example, the sequential testing procedures based on U-statistics or efficient score vectors, respectively, which have been constructed by Gombay (2000, 2003). For the corresponding background of our approach we refer to our predecessor, Gut and Steinebach (2002).

The present paper is a continuation of the previous one, the main point being that we look in more detail into the behaviour of the relevant stopping times, in particular the time it takes from the actual (unknown) change-point until one detects that a change actually has occurred, in other words, we prove asymptotics for stopping times under alternatives.

To the best of our knowledge, results of this type were initiated by Aue (2004) and Aue and Horváth (2004) who showed that, in the monitoring of a "change in the mean" model, certain stopping times based on cumulative sums of data (CUSUM's), which are observed after a training period of size m, have an asymptotically normal limiting distribution (as $m \to \infty$). The latter investigations have been extended in Aue et al. (2007a) to the monitoring of "early" shifts in the mean, for which extreme value type test statistics turn out to detect a change faster. Similar results have also been obtained for delay times in the monitoring of changes in linear regression models (cf., e.g., Kvesic, 2006; Aue et al., 2007b; Hušková et al., 2008). For corresponding results concerning sequential procedures based on moving sums (MOSUM's) instead of CUSUM's confer also Kühn (2008).

The basic setup is presented in Section 2. An essential technical tool is strong approximation, i.e., the fact that our counting processes can be appropriately approximated by Wiener processes, the details of which are presented in Section 3. In Section 4 we review some background results under the null hypothesis, taken from Gut and Steinebach (2002), before we take off into new results, which occupy the remainder of the paper.

2. Setup

Consider a *renewal counting process* $\{N(t), t \geqslant 0\}$ based on an i.i.d. sequence $X_1, X_2, \ldots$ up to (say) time k_n^*, after which it is based on another i.i.d. sequence $X_1^*, X_2^*, \ldots$, independent of the first one, up to time n. Set $EX_1 = \mu > 0$, $0 < \operatorname{Var} X_1 = \sigma^2 < \infty$, $EX_1^* = \mu^* = \mu - \Delta > 0$, and $0 < \operatorname{Var} X_1^* = \sigma^{*2} < \infty$. Our aim is to construct an (asymptotic) sequential test, either for the (one-sided) hypotheses

$$\mathrm{H}_0 : k_n^* = n \quad \text{vs.} \quad \mathrm{H}_1^+ : 1 \leqslant k_n^* < n, \ \ \Delta > 0, \tag{2.1}$$

or for the (two-sided) hypotheses

$$\mathrm{H}_0 : k_n^* = n \quad \text{vs.} \quad \mathrm{H}_1 : 1 \leqslant k_n^* < n, \ \ \Delta \neq 0, \tag{2.2}$$

based upon observations of the process $\{N(t), t \geqslant 0\}$ at the equidistant time points $k = 0, 1, \ldots, n$, where Δ thus denotes a possible change in the mean of the underlying inter-arrival times.

Remark 2.1. A typical situation in which a statistical analysis is based on counting data from a renewal process (instead of the corresponding inter-arrival times) is given in insurance mathematics. Here, for example, the number of claims are continuously observed for certain insurance portfolios and decisions (like, e.g., new calculations of premiums) have to be made based on these numbers.

For our analysis below we set $\theta = 1/\mu$ and $\eta^2 = \sigma^2/\mu^3$, and define

$$Y_k = Y_{k,n} = \frac{N(k) - N(k - h_n) - h_n\theta}{\eta\sqrt{h_n}}, \quad k = h_n, \ldots, n, \tag{2.3}$$

$$Z_k = Z_{k,n} = \frac{N(k) - k\theta}{\eta\sqrt{k}}, \quad k = 1, \ldots, n, \tag{2.4}$$

where, thus, h_n is the window size and the time interval $[0, h_n]$ corresponds to the abovementioned training period. Note that the window size should be chosen with some care. Namely, if it is "too small", then local randomness might cause false alarms, whereas a "too large" window size might unnecessarily delay action.