NON-PARAMETRIC VERTICAL BOX CONTROL CHART FOR MONITORING THE MEAN

EWARYST RAFAJŁOWICZ
TECHNICAL UNIVERSITY WROCŁAW, POLAND,

MIROSŁAW PAWLAK
UNIVERSITY OF MANITOBA, CANADA,

ANSGAR STELAND
RUHR-UNIVERSITY BOCHUM, GERMANY

Abstract. A new class of non-parametric control charts for detecting the change in the process mean is examined. The method, called a Vertical Box Control Chart (V-Box Chart), offers a simple and quick detection of the mean change in an observed process. No parametric assumption on the distribution function of the process is required. Furthermore, the V-Box Chart outperforms the classical Shewhart control chart by lowering the probability of detection of the out-of-control situation with the zero delay. Theoretical bounds on in-control and out-of-control behaviors of the V-Box Chart are worked out. The developed theory is supported by simulation examples.

The research of the E. Rafajłowicz was supported by the Council for Scientific Research of Poland under grant ranging from 2002 to 2006. A. Steland acknowledges the support of the Deutsche Forschungsgemeinschaft (SFB 475, Reduction of Complexity in Multivariate Data Structures).

1. Introduction

Classical control charts, such as CUSUM, EWMA and the Shewhart charts, are the most widely used techniques for detecting changes in parameters of time series models. Due to their simplicity and relative good efficiency they provide standards for many problems in quality control, financial time series and signal processing. They have been extensively examined and extended into many directions, see [18] and the references cited therein. These basic methods posses a common feature based on the utilization of a certain averaging scheme of past observations. The resulting differences of the generalized averages form the test statistic for the parameter change. Such a strategy allows to tune the in-control average run length and to accumulate results of small changes of the controlled process. On the other hand, averaging across the change point yields an substantial delay of its detection. Furthermore there is an additional reduction in the performance due to the usage of incorrect parameter estimates in control charts, see [1] for a recent discussion of this serious issue.

In order to alleviate such shortcomings of the aforementioned classical methods we propose a new approach for constructing control charts which does not average past observations and do not require parametric knowledge of the distribution of the process. In fact, the method counts how many past data points fell into a rectangular box which has a properly controlled width and height. The box moves together with the most recently obtained observation which defines the right-hand side of the rectangular. We call such a scheme as the Vertical Box Control Chart (V-Box Chart). Moreover, the front position of the most recent observation allows to detect jumps of a moderate size without virtually any delay. Furthermore, unlike in the classical theory, the V-Box Chart does not require the parametric knowledge of the underlying probability distributions. We refer, however, to [5] and [3] for extensions of control charts to nonparametric cases, being understood in the sense that no assumptions on the underlying distribution of the error terms are made.

In this paper we are concerned with the change detection in a function

which cannot be parametrized. Nevertheless, in our theoretical considerations we shall focus on the change of the step form. The proposed chart implicitly uses the general concept of vertically weighted regression (see [10], [11], [16], [12], [17]), but we do not need a general theory of this notion. We also compare our method to the Shewhart control chart not only due to its popularity but because it is also a technique relying on the most recent observations. We refer to [2] for an extensive discussion of the classical control charts. Furthermore in [8], [9], [13] , [14], [15], [6] various aspects and extensions of the CUSUM and Shewhart control charts are examined.
The paper is organized as follows. In Section 2 we introduce our vertically weighted control chart and give a detailed description of its usage. Section 3 examines an important issue of selecting a parameter which controls the accuracy of the method, i.e., the parameter which can reduce the probability of false alarm. In Section 4 we establish exponential bounds on the probability of the false alarm and the probability of not detecting the change. Finally Section 5 reports some simulation results for the proposed control chart and in particular its performance relative to the Shewhart chart.

2. Definition of V-Box Control chart

2.1. **Model of observations.** Let Y_i, $i = 1, 2, \ldots, n$ be a sequence of observations such that

$$Y_i = m_i \mathbf{1}_q(i) + \epsilon_i, \tag{2.1}$$

where ϵ_i's are unobserved random errors, $q > 0$ is an unknown change point, i.e., the discrete time point at which Y_i's change their distribution function. By definition

$$\mathbf{1}_q(i) = \begin{cases} 0 & i < q \\ 1 & i \geq q \end{cases},$$

m_i, $i = 1, 2, \ldots n$ is unknown, but it is either nondecreasing or nonincreasing sequence of numbers. One may interpret the m_i's as equidistant samples, $m_i = m(t_i)$, for a hypothetical (unknown) quality characteristic. This interpretation by no means being convenient is not necessary for validity of the results presented in this paper.

Concerning random errors, we assume that ϵ_i are i.i.d. random variables with a distribution function which is symmetric with respect to zero. Note that we do not assume the existence of any moments of ϵ_i's. For simplicity of exposition we assume the existence of a p.d.f. of ϵ_i's, denoted further by f_ϵ, but this assumption can be relaxed.

2.2. **Definition of V-Box Control Chart.** The proposed method of detecting change point q is as follows.

Preparations: Assume that a certain number of observations has been collected. Denote by n the index of the current observation.

- Choose $L > 1$, the number of past observations, which are taken into account when deciding whether an out-of-control state is reached or not. Define a box of the form

$$B(L, H, n, Y_n) \stackrel{def}{=} [n - L, n] \times [Y_n - H, Y_n + H], \tag{2.2}$$

  where $H > 0$ is the height of the box.
- Select $H > 0$ and $0 < \theta < 1$ in such a way that θL is the fraction of observations Y_{n-L}, ..., Y_{n-1}, Y_n which are "typically" in the box $[n - L, L] \times [Y_n - H, Y_n + H]$, if the process is in-control state (choice of L, H and θ is discussed below in more details).
- Collect observations $Y_1, Y_2, \ldots, Y_L$ and set $n = L + 1$.

Step 1: Calculate the number $b_{LH}(n)$ of observations among Y_{n-1}, ..., Y_{n-L} which fell to box $B(L, H, n, Y_n)$. Thus,

$$b_{LH}(n) = Card\{Y_{n-j} \in B(L, H, n, Y_n),\ j = 1, 2, \ldots L\},$$

where $Card$ denotes the cardinality of a set. *Note that Y_n is always in $B(L, H, n, Y_n)$ but it is not included in $b_{LH}(n)$.*

Step 2: If $b_{LH}(n) > \theta L$, then increase n by one and go to Step 1.

Step 3: If $b_{LH}(n) \leq \theta L$, then stop and signal out-of-control state.

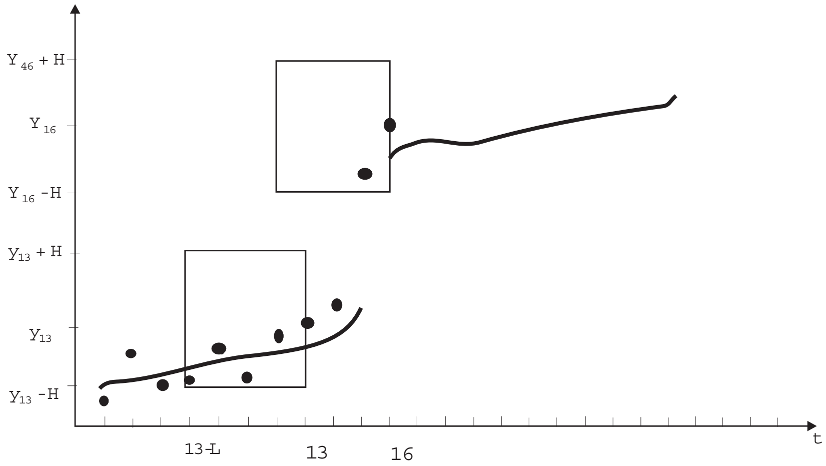

FIGURE 1. A sketch of performance of V-Box Chart for an unobserved (hypothetical) quality characteristic of the underlying process (fat line) and given observations (fat dots) for parameter values $L = 4$, $\theta = 0.5$. Left box: in-control behavior, right box: out-of-control state is signaled.

Thus, the stopping time has the form

$$N(L, H, \theta) = \inf_n \{n : b_{LH}(n) \leq \theta L\}, \tag{2.3}$$

which is the first time index when a signal is occurring.

The performance of the above proposed chart is schematically shown in Fig. 1. Intuitive explanation of how the V-Box Chart works is the following.

(1) If the underlying process is in-control state, then most of the observations Y_j with time indexes $j \in \{n, n-1, \dots n-L\}$ are contained in box $B(L, H, n, Y_n)$ and an alarm is not signaled, since condition $b_{LH}(n) > \theta(L+1)$ holds for any reasonable chosen $0 < \theta < 1$ and $H > 0$. A false alarm can be signaled, if either Y_n is an outlier observation and θ is too small or random errors have relatively large variance and H is chosen too small.
(2) If the process runs out-of-control at time q and the jump $m(t_q) - m(t_{q-1})$ is relatively large, then the box $B(L, H, q, Y_q)$ contains relatively small number of observations with time indexes $j \in$

$\{q, q-1, \dots q-L\}$. Thus condition $b_{LH}(n) \leq \theta(L+1)$ holds and out-of-control state is signaled. In the most favorite case, box $B(L, H, q, Y_q)$ contains only one observation, namely, Y_q.

(3) If it happens that the jump is not detected in time q (due to unfavorable random pattern of errors ϵ_i's), then the chart can detect it in later time instants, $q+1, q+2, \dots$. However, if after q the out-of-control state is constant ($m(t)$ is a step function), then the chart can detect it not later than $q + \lfloor \theta L \rfloor$, since after that time the chart treats the state after the jump as "normal". In other words, by increasing L we increase the probability of detecting jump, if it was not detected immediately.

(4) The long memory L of the chart stabilizes its in-control behavior, but for practical reasons L can not be too large, since the first L observations are lost for detecting out-of-control states.

A practical way of overcoming the last difficulty is to use smaller L at the beginning of the observation process and then enlarge it gradually with time. Nevertheless, for theoretical purposes we will consider the case $L \to \infty$.

It is worth noting that our V-Box Chart resembles charts for p-charts for detecting changes in the frequency (see [19] for a recent bibliography on this subject). Nevertheless, our chart detects changes in the mean by counting events in a vertically and horizontally wandering box.

3. Selecting θ

Assume that errors have a finite support, $\epsilon_i \in [-Z, Z]$ a.s. for a certain $Z > 0$. Assume also that for certain constants $0 < C_0 \leq C_1 < \infty$ the p.d.f. f_ϵ of the error terms fulfils

$$C_0 \;\leq\; f_\epsilon(x) \;\leq\; C_1, \;\; x \in [-Z, Z]. \tag{3.1}$$

Clearly, $C_0 \leq 1/(2Z)$ and $C_1 \geq 1/(2Z)$.

Denote by $P\{Y_{n-j} \in B(L, H, n, Y_n) \mid InC\}$ the probability that the $(n-j)-th$ observation is in the box $B(L, H, n, Y_n)$, $j = 1, 2, \dots L$, assuming that the process was in the in-control state in time instants n, $n-1$, ... ,$n-L$, what is denoted by the symbol InC .

The following two lemmas are important for our future considerations.

Lemma 3.1. *Assume* $0 < H \leq 2\,Z$. *Assuming that for* f_ϵ *conditions* (3.1) *hold, define* $p(H,Z) = C_0^2\,(4HZ - H^2)$. *Then, we have for* $j = 1, 2, \ldots, L$

$$P\{Y_{n-j} \in B(L,H,n,Y_n) \mid InC\} \;\geq\; p(H,Z). \tag{3.2}$$

Furthermore,

$$0 \;<\; p(H,Z) \;\leq\; \frac{H}{Z} - \frac{H^2}{4Z^2} \;<\; 1. \tag{3.3}$$

The equality in (3.2) *is attained for the distribution uniform in* $[-Z, Z]$ *and then*

$$P\{Y_{n-j} \in B(L,H,n,Y_n) \mid InC\} \;=\; \frac{H}{Z} - \frac{H^2}{4Z^2}. \tag{3.4}$$

PROOF. Denote by $P\{Y_{n-j} \in B(L,H,n,Y_n) \mid InC, Y_n = y\}$ the probability of finding Y_{n-j} in $B(L,H,n,Y_n)$, providing that $Y_n = y$. Then,

$$P\{Y_{n-j} \in B(L,H,n,Y_n) \mid InC, Y_n = y\} \;=\; \int_{y-H}^{y+H} f_\epsilon(t)dt \;\geq \tag{3.5}$$

$$C_0\,[-\,((H-y+Z)\;\mathbf{1}(-H+y-Z)) - (H+y-Z)\;\mathbf{1}(H+y-Z) +$$
$$+\,(H-y-Z)\;\mathbf{1}(-H+y+Z) + (H+y+Z)\;\mathbf{1}(H+y+Z)]$$

Denote by $F(y,Z,H)$ the r.h.s. of (3.5). Then,

$$P\{Y_{n-j} \in B(L,H,n,Y_n) \mid InC\} \;\geq\; \int_{-Z}^{Z} F(y,Z,H) f_\epsilon(y)dy \geq \tag{3.6}$$

$$C_0 \int_{-Z}^{Z} F(y,Z,H)dy \;=\; C_0^2\left(4HZ - H^2\right). \quad \bullet$$

The case considered in Lemma 3.1 is depicted in Fig. 2 – left box. The right box in this figure corresponds to the next lemma.
Denote by $P\{Y_{q-j} \in B(L,H,q,Y_q) \mid OutC\}$ the probability that $(q-j)-th$ observation is in box $B(L,H,q,Y_q)$, $j = 1, 2, \ldots L$, assuming that the process was in the in-control state in time instants $q-1, \ldots$,$q-L$, and in q it changes to the out-of-control state, what is marked by $OutC$. The scenario of switching between these states is assumed to be $a \cdot \mathbf{1}_q(i)$.

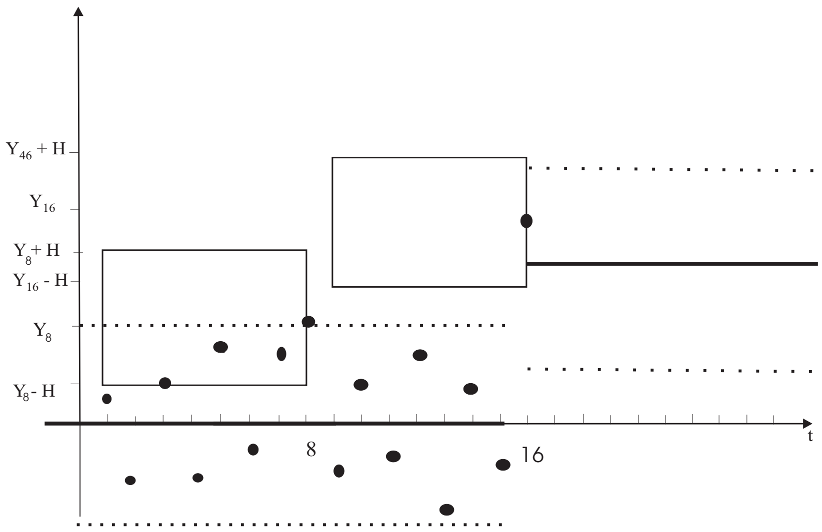

FIGURE 2. A sketch of performance of V-Box chart for step like change from in-control to out-of-control state. Dotted lines are boundaries of errors, assuming that their distribution has a finite support.

Lemma 3.2. *Assume* $0 < H \leq Z$, $H < a$ *and* (3.1). *Depending on the jump height* a *we have for* $j = 1, 2, \dots, L$ *the following bounds.*
A) If $|a| > 2\,Z + H$, *then*

$$P\{Y_{q-j} \in B(L, H, q, Y_q) \,|\, OutC\} \;=\; 0. \tag{3.7}$$

B) If $2\,Z - H < |a| \leq 2\,Z + H$, *then*

$$P\{Y_{q-j} \in B(L, H, q, Y_q) \,|\, OutC\} \;\leq\; C_1^2\, \frac{(-|a| + H + 2\,Z)^2}{2}. \tag{3.8}$$

C) If $|a| \leq 2\,Z - H$, *then*

$$P\{Y_{q-j} \in B(L, H, q, Y_q) \,|\, OutC\} \;\leq\; 2\,C_1^2\, H\,(2\,Z - |a|). \tag{3.9}$$

Thus, for every $a \in R$ *we have*

$$P\{Y_{q-j} \in B(L, H, q, Y_q) \,|\, OutC\} \;\leq\; C_1^2 \frac{(-|a| + H + 2\,Z)^2}{2}. \tag{3.10}$$

The equalities in (3.8) *and* (3.9) *are attained for the distribution uniform in* $[-Z, Z]$ *if we set* $C_1 = 1/(2\,Z)$.

Proof. For the probability $P\{Y_{q-j} \in B(L,H,q,Y_q)\,|\,OutC, Y_q = y\}$, given $Y_q = y$, we have similar expressions as (3.5), since for $j = q-1, q-2, \ldots$ the process is still in-control state. Thus,

$$(3.11) \quad P\{Y_{q-j} \in B(L,H,q,Y_q)\,|\,OutC, Y_q = y\} = \int_{y-H}^{y+H} f_\epsilon(t)dt \le$$

$$\le \frac{C_1}{C_0} F(y,Z,H)\,,$$

where F is the same as in the proof of Lemma 3.1 and the unconditional probability is bounded by

$$(3.12) \quad P\{Y_{q-j} \in B(L,H,q,Y_q)\,|\,OutC\} \le \frac{C_1^2}{C_0} \int_{a-Z}^{a+Z} F(y,Z,H)\,dy =$$

$$\frac{C_1^2}{2}\left[-4\,H\,(a-2\,Z) - (-a+H+2\,Z)^2\,\mathbf{1}(a-H-2\,Z) \right.$$
$$\left. + (a+H-2\,Z)^2\,\mathbf{1}(a+H-2\,Z)\right].$$

¿From (3.12) we obtain:
– $P\{Y_{q-j} \in B(L,H,q,Y_q)\,|\,OutC\} = 0$ for a large jump $|a| > 2Z + H$,
– inequality (3.8) for a moderate jump $2\,Z - H < |a| \le 2\,Z + H$,
– inequality (3.9) for a small jump $|a| \le 2\,Z - H$.
This completes the proof, since (3.10) follows by comparing cases A), B) and C). Also the last statement can be verified by direct computation.
•

Remark 3.3. The above can be carried for showing the following lower bound

$$(3.13) \quad P\{Y_{q-j} \in B(L,H,q,Y_q)\,|\,OutC\} \ge C_0^2 \frac{(-|a|+H+2\,Z)^2}{2},$$

where C_0 is the constant defined in (3.1).

Define the following function

$$(3.14) \quad r(H,Z,a) = \begin{cases} 0 & \text{if} \quad |a| > 2Z + H \\ 2\,C_1^2\,H\,(2\,Z - |a|) & \text{if} \quad |a| \le 2\,Z - H \\ \frac{C_1^2}{2}\,(-|a|+H+2\,Z)^2 & \text{in} \quad \text{other cases.} \end{cases}$$

Then, the statement of Lemma 3.2 can be rewritten as follows

$$(3.15) \quad P\{Y_{q-j} \in B(L,H,q,Y_q)|OutC\} \le r(H,Z,a), \quad j = 1,2,\ldots L\,.$$

¿From (3.14) it follows that selecting $H > 0$ sufficiently small one can force $r(H, Z, a)$ to be close to zero. On the other hand, selecting H sufficiently close to $2\,Z$ we can force $p(H, Z)$ to be close to 1. Additionally, one can verify directly that for uniformly distributed ϵ_i's we have $r(H, Z, a) < p(H, Z)$.

The above considerations justify the following choice of θ

$$r(H, Z, a) \; < \; \theta \; < \; p(H, Z)\,. \tag{3.16}$$

These inequalities are sufficient for theoretical results that are developed in the next section.

For practical applications it is expedient to give more precise indications concerning the choice of θ. This is possible if we know the type of errors distribution. Below, we give such indications for the errors uniformly distributed in $[-Z, Z]$ for which (3.16) is fulfilled if we take θ such that

$$\frac{(-|a| + H + 2\,Z)^2}{8\,Z^2} \; < \; \theta \; < \; \frac{H}{Z} - \frac{H^2}{4Z^2}\,. \tag{3.17}$$

Introducing the normalized variables $h \stackrel{def}{=} H/Z$ and $J \stackrel{def}{=} |a|\,/Z$, one can write (3.17) equivalently as

$$\frac{(2 + h - J)^2}{8} \; < \; \theta \; < \; h - \frac{h^2}{4}. \tag{3.18}$$

The area in (h, θ) plane, which are admissible in the sense (3.18) is plotted in Fig. 3 (left panel) for $J = 1$.

As one can notice, the admissible area still provides a freedom in selecting (h, θ). Let us note that the left and the right hand sides of the inequalities in (3.18) are bounds for p and r, respectively, i.e., the bounds for the probability that an observation is in $B(L, H, n, Y_n)$ for in-control and out-of-control states. In order to ensure better distinguishability of these states it is desirable to have a large difference between these bounds, since it pushes away also probabilities p and r. In the r.h.s. panel of Fig. 3 the difference $p - r$ versus h is plotted for different normalized jump heights J. Existence of h for which this difference is maximized and its dependence on J are clearly visible.

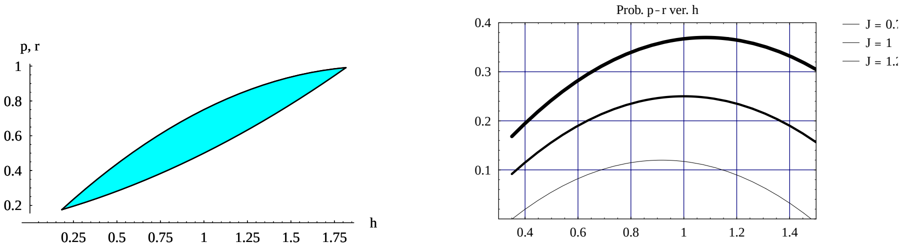

FIGURE 3. Left panel: An example of the area for which inequalities (3.17) for θ hold for $J = 1$.
Right panel: Difference between bounds $p(H, Z) - r(H, Z, a)$ as a function of h parametrized by J.

4. BOUNDS ON ERRORS OF V-BOX CHART

Our aim in this section is to derive exponential bounds on the probabilities:

- of signaling out-of-control state when the process is in the in-control situation,
- of not signaling immediately that the process is out-of-control.

We need the following Chernoff's bounds (see [4] for the proof), which, in slightly different notation than in [4], have the form.

Lemma 4.1. *A) Let b be a binomial random variable with parameters: $L > 1$ – the number of trials and $0 < \varrho < 1$ – the success probability in one trial. Then, for $1 > \theta > \varrho > 0$,*

$$(4.1)\qquad P\{b \, > \, L\theta\} \, \le \, \exp\left\{-L\left[\theta \log\frac{\theta}{\varrho} + (1-\theta)\, \log\frac{1-\theta}{1-\varrho}\right]\right\} \, \le$$

$$\le \, \exp\left\{-L\left[\varrho - \theta + \theta\log(\theta/\varrho)\right]\right\} \, .$$

B) Let β be a binomial random variable with parameters: $L > 1$ – the number of trials and $0 < \varpi < 1$ – the success probability in one trial. Then, for $1 > \varpi > \theta > 0$,

$$(4.2)\qquad P\{\beta \, < \, L\theta\} \, \le \, \exp\left\{-L\left[\theta \log\frac{\theta}{\varpi} + (1-\theta)\, \log\frac{1-\theta}{1-\varpi}\right]\right\} \, \le$$

$$\le \, \exp\left\{-L\left[\varpi - \theta + \theta\log(\theta/\varpi)\right]\right\} \, .$$

Note that apparent similarity between (4.2), (4.1) follows from the fact that if β is a binomial r.v. with parameters L and ϖ, then $L-\beta$ is a binomial r.v. with parameters L and $1-\varpi$.

4.1. **In-control behavior.** Consider the probability of false alarm in n-th step, providing $n > L$. According to Step 3 of V-Box Chart, this event occurs if $b_{LH}(n) \leq \theta\, L$, provided that the process is in-control state. Note that $b_{LH}(n)$, being the number of observations captured in box $B(L, H, n, Y_n)$, is a binomial r.v. with the probability of success in one trial equal to

$$(4.3) \qquad \varpi(L,H) \stackrel{def}{=} P\{Y_{n-j} \in B(L,H,n,Y_n) \mid InC\}$$

and L being the number of trials. In (4.3) we have dropped time index n, since random variables $b_{LH}(n)$, $n = L+1, L+2, \ldots$ form a stationary sequence, provided the the process is in-control state. Thus, for the probability of false alarm we have

$$(4.4) \qquad P\{b_{LH}(n) \leq \theta\, L \,|\, InC\} \;\leq\; P\{b_{LH}(n) < \theta\,(L+1) \,|\, InC\} \;\leq$$

$$\exp\left\{-(L+1)\left[\varpi(L,H) - \theta + \theta \log(\theta/\varpi(L,H))\right]\right\} ,$$

where the last inequality follows from part B) of Lemma 4.1. Note that the assumption $1 > \varpi(L,H) > \theta > 0$ of this lemma holds, since by (4.3) and (3.16) we have

$$(4.5) \qquad \varpi(L,H) \geq p(H,Z) > \theta\,.$$

Theorem 4.2. *Let assumptions of Lemma 3.1 hold and let θ be selected according to* (3.16). *Then, for $n > L$ the probability of false alarm in n-th time instant does not depend on n. This probability is further denoted by $\mathcal{P}(\theta, L, H)$ and we have*
(4.6)

$$\mathcal{P}(\theta,L,H) \;\leq\; \exp\left\{-(L+1)\left[\varpi(L,H) - \theta + \theta \log(\theta/\varpi(L,H))\right]\right\} .$$

Furthermore, $\mathcal{P}(\theta, L, H)$ can be made arbitrarily close to zero by selecting L sufficiently large.

Proof. Inequality (4.6) was proved above. To justify the second statement it suffices to prove that the function in the square brackets in

(4.6) is positive. This follows immediately, since this function is equal to

$$\varpi(L,H)-\theta\left(1+\log\left(\frac{\varpi(L,H)}{\theta})\right)\right) > \varpi(L,H)+\theta\left(-\frac{\varpi(L,H)}{\theta}\right) = 0,$$

where we have used the elementary inequality $-(1+\log(x)) > -x$ valid for $x > 0$ and $x \neq 1$, substituting in it $x = \frac{\varpi(L,H)}{\theta} > 1$. •

Corollary 4.3. *Under assumptions of Thm. 4.2 the in-control averaged run length of V-Box Chart, calculated for time instants starting from* $n = L + 1$, *is not smaller than*

$$\exp\left\{(L+1)\left[\varpi(L,H)-\theta+\theta\log(\theta/\varpi(L,H))\right]\right\} \tag{4.7}$$

and it can be made arbitrarily large by selecting L *large enough.*

PROOF. Let us shift the origin of the time scale to the point $n = L$ and let new time index k equals zero for $n = L$. Denote by $\mathcal{Q}_k$ the probability that the false alarm is signaled exactly at time $k > 1$ and it was not signaled at time instants $k-1, k-2, \ldots, 1$. Let $Q(j, j-1)$ be the probability that the alarm was not signaled exactly at time j conditioned on the event that it also was not signaled at time $(j-1)$. Then, $\mathcal{Q}_k$ depends on the probability of false alarm appearing at time k and on the probability that it was not signaled at times $k-1, k-2, \ldots, 1$, which can be calculated by subsequent conditioning. These yield

$$\mathcal{Q}_k = \mathcal{P}(\theta,L,H)\prod_{j=1}^{k-1} Q(j,j-1), \quad k = 2,3,\ldots \tag{4.8}$$

Now, $Q(j, j-1)$, being the conditional probability, is not smaller than the unconditional probability that the alarm is not signaled at time j, which equals to $1 - \mathcal{P}(\theta, L, H)$ (see Thm. 4.2). This fact and (4.8) imply

$$\mathcal{Q}_k \geq \mathcal{P}(\theta,L,H)\ (1-\mathcal{P}(\theta,L,H))^{k-1}. \tag{4.9}$$

The r.h.s. of this inequality is formally the geometric distribution with parameter $\mathcal{P}(\theta, L, H)$. The expectation of a random variable with this distribution equals $1/\mathcal{P}(\theta, L, H)$, what finishes the proof by invoking (4.6). •

We refer the reader to [7] for the discussion on relationships between a single alarm probability and the average run length.

4.2. **Out-of-control behavior.** In the same vain as above one can analyze the probability of not detecting a jump of the process quality exactly at time instant q when it occurred. We shall denote this probability by $\mathcal{R}(H, Z, a)$.

Assume that the process runs out-of-control at time q, where $q > L$. Then, according to Step 2 of V-Box Chart the alarm is not signaled immediately if $b_{LH}(q) > \theta L$. Thus,

$$\mathcal{R}(H, Z, a) = P\{b_{LH}(q) > \theta L | OutC\} . \tag{4.10}$$

b_{LH} is a binomial r.v. with L as the number of trials and

$$\varrho(H, Z, a) \stackrel{def}{=} P\{Y_{q-j} \in B(L, H, q, Y_q) | OutC\} , \quad j = 1, 2, \ldots L \tag{4.11}$$

as the probability of "success" in one trial, where we have used the fact that $P\{Y_{q-j} \in B(L, H, q, Y_q) \,|\, OutC\} , \quad j = 1, 2, \ldots L$ does not depend on j. According to (3.15) and (3.16) we have

$$\varrho(H, Z, a) < r(H, Z, a) < \theta , \tag{4.12}$$

what allows to apply part A) of Lemma 4.1 for evaluating $P\{b_{LH}(q) > \theta L | OutC\}$ and we obtain.

Theorem 4.4. *Let assumptions of Lemma 3.2 hold and let θ be selected according to* (3.16). *Then, for $q > L$*

$$\mathcal{R}(H, Z, a) \leq \exp\{-L [\varrho(H, Z, a) - \theta + \theta \log(\theta/\varrho(H, Z, a))]\} . \tag{4.13}$$

Furthermore, $\mathcal{R}(H, Z, a)$ can be made arbitrarily small by selecting L large enough.

Proof. We need only to prove that for $\varrho(H, Z, a) < \theta$ we have

$$\varrho(H, Z, a) - \theta + \theta \log(\theta/\varrho(H, Z, a)) > 0. \tag{4.14}$$

This follows from the following elementary considerations in which the arguments of ϱ are dropped

$$\varrho - \theta + \theta \log(\theta/\varrho) = \theta \left[\frac{\varrho}{\theta} - \left(\log \left(\frac{\varrho}{\theta} \right) + 1 \right) \right] > \theta \left[\frac{\varrho}{\theta} - \frac{\varrho}{\theta} \right] = 0, \tag{4.15}$$

where we have used the elementary inequality $-(\log(x)+1) > -x$, valid for $x \neq 1$ and substituting $x = \frac{\varrho}{\theta} < 1$. •

Remark 4.5. Let us note that the exponent on the right hand side of (4.13) is a monotonically increasing function of $\varrho(H, Z, a)$. Therefore, we can replace $\varrho(H, Z, a)$ by the expression given in Remark 3.3, yielding a simple upper bound for $\mathcal{R}(H, Z, a)$.

Remark 4.6. $\mathcal{R}(H, Z, a)$ is the probability of the signal has not appeared immediately after the change point. If one wants to evaluate the joint probability of signaling in a sequence of later time instances, then one should proceed as in the proof of Thm. 4.2, using the conditioning technique.

Remark 4.7. The main results of this paper have been obtained under the assumption that the support of errors is compact. This assumption can be relaxed to the unbounded support case. In fact, if $f_\epsilon(t)$ is symmetric around $t = 0$ and monotonic for $t > 0$, then Lemma 3.1 takes the following form

$$P\{Y_{n-j} \in B(L, H, n, Y_n) \geq 4\,H \int_0^\infty f_\epsilon(y+H) \cdot f_epsilon(y)\,dy.$$

In the case of $f_\epsilon = N(0, \sigma^2)$ we have

$$P\{Y_{n-j} \in B(L, H, n, Y_n) \geq (2/\sqrt{\pi}) \cdot h \exp(-h^2/4)[1 - Erf(h/2)],$$

where $h = H/\sigma$, while $Erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^(-t^2)\,dt$.

4.3. **Further remarks on ARL.** In contrast to the Shewhart chart we do not have simple expressions for the ARL of the V-Box chart. Whereas the Shewhart chart follows a geometric distribution, the V-Box chart uses the most recent L data points at each time point. Thus, the time windows overlap and the sequence of signals is no longer a sequence of independent Bernoulli trials. In the following we propose a simple modification of the V-Box design for which simple formulas for the ARL can be derived.

The basic idea is to apply a modified version of the V-Box chart only at each L-th data point. The modified version tries to guess the decision of the V-Box chart but uses only the data located in the current time window. Notice that in most instances when the V-Box chart gives a signal, the observation corresponding to the position of the signal is the maximum of the most recent L observations located in the box. This means, in this case the most recent L data points before a maximum are checked whether they are located in the box. We mimic that situation by checking only data points prior to the maximum. The r-th application of the modified chart occurs at the time point $n = rL$, r an integer. Define $X_i = Y_{n-L+i-1}$, $i = 1, \dots, n$, and let $X_{(1)} \leq \dots \leq X_{(n)}$ denote the order statistic. Let m^* be the position of the maximum, i.e., $X_{m^*} = X_{(n)}$ and consider

$$b^*_{LH}(n) = Card\{X_{(i)} \in B(L, H, m^*, X_{m^*}), i = 1, \dots, m^* - 1\}.$$

We now compare $b^*_{LH}(n)$ at time points $n = rL + 1$, $r \in N$, with m^*L, and give a signal, if the proportion of values in the box is less than θ, i.e.,

$$R^* = \inf\{r - 1 : \ r \in N, b^*_{LH}(rL + 1) \leq \theta \cdot m^*\}.$$

Then, $E(R^*) = 1/\varpi^*(L, H) - 1$, where

$$\varpi^*(L, H) \stackrel{def}{=} P\{b^*_{LH}(L + 1) \leq \theta \cdot m^* | InC\}\,.$$

Note that R^* gives a signal if the proportion of data points before the maximum is less or equal L, and therefore our guess of the run length of V-Box chart is

$$N^* = R^*L + 1.$$

Clearly, we have the in-control ARL

$$E(N^*) = L(\varpi^*(L, H)^{-1} - 1) + 1.$$

5. Simulations

The Vertical Box Control Chart was compared to the Shewhart Chart, in which the alarm state is signaled if $|Y_n| > C_S$, where $C_S > 0$ is a constant selected in such a way that an in-control average run length (ARL) is not smaller than selected by the statistician.

As a model of in-control behavior we take $N(0, \sigma)$, i.e. zero mean gaussian errors with dispersion $\sigma = 0.25$. The change from in-control

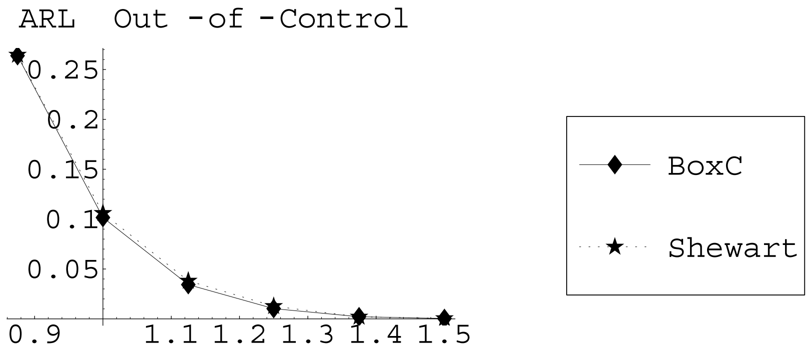

FIGURE 4. Out-of-control ARL of V-Box chart and the Shewhart chart as a function of the jump height a.

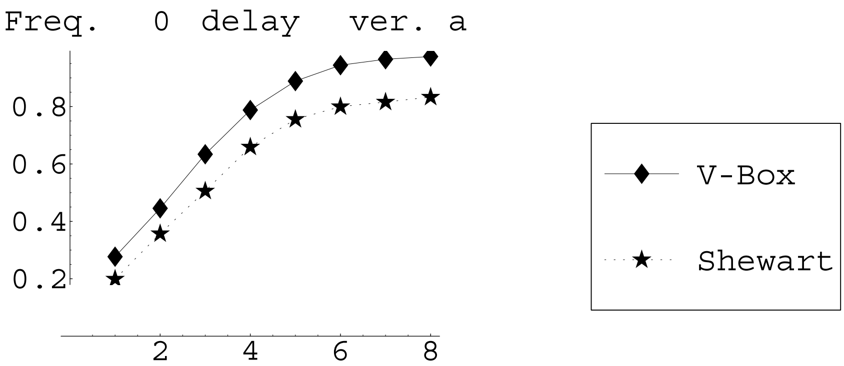

FIGURE 5. Empirical frequency of detecting the jump with zero delay versus its height a for V-Box and the Shewhart charts.

to out-of-control scenario was modeled as a simple step function with height a, which is corrupted by $N(0, \sigma)$ errors.

Both charts were tuned to in-control average run length (ARL) equal to 100. This resulted in $C_S = 0.675$ for the Shewhart chart and in $H = 0.675$, $\theta = 0.6$, $L = 25$ for V-Box chart. The charts were tuned by averaging 10^4 runs.

Then, both charts were tested for detecting the step jump of height $a > 0$ hidden in $N(0, \sigma)$ noise. The resulting out-of-control ARL are plotted in Fig. 4. As one can notice, out-of-control ARL of the both charts is, with the accuracy of simulations, the same for a wide range of jumps heights. On the other hand, Fig. 5 indicates that the frequency of detecting the jump exactly at the time of its occurrence is visible larger for V-box chart than for the Shewhart chart.

6. Concluding remarks

In the above, V-box chart was investigated as a chart for detecting step like changes in the process mean. It is however apparent from the construction of the chart that it can also be used for monitoring more general changes of non-parametric nature. One can also expect that V-box chart is robust against outliers of a moderate size. The discussion of these aspects is outside the scope of this paper.

Our approach can easily be generalized to smooth counting function, by replacing statistics $b_{LH}(n)$ with

$$\sum_{j=1}^{L} V\left(\frac{Y_{n-j} - Y_n}{H}\right)$$

for a general class of window functions $V(t)$. The choice $V(t) = \mathbf{1}_{[-1,1]}(t)$ gives the method considered in this paper.

References

[1] Albers, W. and Kallenberg, C.M. (2004). Estimation in Shewhart control charts: effects and corrections. Metrika, 59, 207-234.

[2] Brodsky B. E. , Darkhovsky B. S., Nonparametric Methods in Change-Point Problems. Kluwer, Dordrecht, 1993.

[3] Chakraborti S., Van der Laan P. and Bakir S. T., Nonparametric control charts: an overview and some results. J. Quality Techn., vol. 33, No 3, 2001, pp 303-315.

[4] Gyorfi L., Kohler M., Krzyzak A., Walk H. A Distribution-Free Theory of Nonparametric Regression. Springer-Verlag, New York, 2003.

[5] Lee S,. Ha, J. and Na, O. (2003), The Cusum test for parameter change in time series. Scand. J. of Statistics, 30, pp 781-796.

[6] Lu, C.W. and Reynolds, M.R. (1999). EMWA control charts for monitoring the mean of autocorrelated process. J. Quality Technology, 31 ,166-188.

[7] Margavio T.M. et al. (1995), Alarm rates for quality control charts. Stat. and Probab. Letters, 24, pp 219-224.

[8] Page, E.S. (1954). Continuous inspection schemes. Biometrika, 1, 100-115.

[9] Page, E.S. (1955). A test for a change in a parameter occurring at an unknown point. Biometrika, 42, 523-526.

[10] Rafajłowicz, E. (1996). Model free control charts for continuous time production processes. Politechnika Wrocław, Technical Report No. 56/96.

[11] Pawlak, M. and Rafajłowicz, E. (2000). Vertically weighted regression - a tool for non- linear data analysis and constructing control charts. Journal of the German Statistical Association, 84, 367-388.

[12] Pawlak, M. and Rafajłowicz, E., and Steland, A. (2004). On detecting jumps in time series - Nonparametric setting. Journal of Nonparametric Statistics, 16, 329-347.

[13] Ritov, Y. (1990). Decision theoretic optimality of the CUSUM procedure. Annals of Statistics, 18, 1464-1469.

[14] Schmid W., On the run length of a Shewhart chart for correlated data., Statistical Papers, vol. 36, pp. 111-130, 1995.

[15] Schmid, W. and Schöne, A. (1997). Some properties of the EWMA control chart in the presence of autocorrelation. Annals of Statistics, 25, 3, 1277-1283.

[16] Steland, A. (2002). Nonparametric monitoring of financial time series by jump-preserving estimators. Statistical Papers, 43, 361-377.

[17] Steland, A. (2004). Jump-preserving monitoring of dependent processes using pilot estimators, Statistics and Decision, 21, 343-366.

[18] Woodall, W.H. and Montgomery (1999). Research issues and ideas in statistical process control. J. Quality Technology, 31, 376-386.

[19] Woodall, W.H (1997) Control charts based on attribute data. J. Quality Technology, 29, 172-183.