Detecting Small Shift on the Mean by Finite Moving Average*

PIERRE R. BERTRAND† and GÉRARD FLEURY‡

Abstract

We consider a sequence of independent Gaussian random variables whose mean remains constant until an unknown time \( r \) when it increases of \( \delta \), and then remains constant afterwards. We want to detect the abrupt change as soon as possible. We propose a new study of detection rules based on finite moving averages.

Finite moving averages (FMA) depend on two parameters: the window size \( A \) and a threshold \( h \). We study the asymptotic properties of finite moving averages detection rules for a fixed threshold \( h \) less than the assumed value of the shift \( \delta \), as the window size \( A \) is varying. For the stationary average delay time (SADT), the first order term of the asymptotic development with respect to the average run length (ARL) is the same (up to \( \epsilon \)) than for the Cusum, Shiryaev–Roberts or GLR detection rules, and the second term plays in favour of the finite moving average, as for exponentially weighted moving average (EWMA).

Numerical simulations confirm that finite moving average detection rules are better than GLR procedure for detecting small shifts and they show no real difference with the CUSUM procedure.

*Received: June 2, 2007; Accepted: October 29, 2007.

Key Words: quality control, finite moving average, average delay time, average run length, Monte–Carlo simulations.


†Mailing Address: Laboratoire de Mathématiques, UMR CNRS 6620 et Université de Clermont–Ferrand, 63177 Aubièere Cedex, France. Email: pierre.bertrand@math.univ-bpclermont.fr.

‡Mailing Address: Laboratoire de Mathématiques, UMR CNRS 6620 et Université de Clermont–Ferrand, 63177 Aubièere Cedex, France. Email: fleury@math.univ-bpclermont.fr
1 Introduction. Let \((X_i)_{i=1,...,n}\) be a sequence of independent Gaussian real random variables whose mean changes at an unknown time \(r\). We seek a family of stopping rules \(T_b\) which allows us to observe the r.v.’s \(X_i\) sequentially, detecting as soon as possible this shift on the mean and rarely stopping the process when there is no shift.

More formally, we consider \(X_i \overset{\mathcal{D}}{=} \mathcal{N}(\theta_0, \sigma_0)\) for \(i \leq r\), and \(X_i \overset{\mathcal{D}}{=} \mathcal{N}(\theta_1, \sigma_1)\) for \(i \geq r + 1\) with a shift on the mean \(\delta := \theta_1 - \theta_0 > 0\), \(\theta_0\) is assumed to be known, and for notational simplicity the variance is taken constant, i.e. \(\sigma_0 = \sigma_1 = \sigma\). A family of stopping times \(T_b\) for the filtration \(\mathcal{F}_k\) generated by \(\{X_1, \ldots, X_k\}\) is called a detection rule. Detection rules depend on parameter(s) \(b\) to be chosen for the application. We denote \(\mathbb{P}_0\) the distribution of the sequence \((X_i)_{i=1,...,n}\) when there is no change, \(\mathbb{P}_{r,\delta}\) the distribution when a change of size \(\delta\) occurs at time \(r\), \(\mathbb{E}_0\) and \(\mathbb{E}_{r,\delta}\) the associate expectations. We define the average run length (ARL) by \(\text{ARL}(b) = \mathbb{E}_0\{T_b\}\) and the average delay time (ADT) by \(\text{ADT}(b, r, \delta) = \mathbb{E}_{r,\delta}\{T_b - r | T_b \geq r + 1\}\). We require a detection rule for which \(\text{ARL}(b)\) is large, say, greater than a specified constant, and \(\text{ADT}(b, r, \delta)\) is as small as possible. Since \(\text{ADT}(b, r, \delta)\) depends on \(b, r, \delta\), different measures of smallness of \(\text{ADT}(b, r, \delta)\) are possible. When \(\delta\) is known, it is usual to consider the criterion \(\sup_{r \geq 0} \text{ADT}(b, r, \delta)\) or following \[18\] (Shiryaev, 1963) and \[19\] (Srivastava and Wu, 1993) \(S\text{ADT}(b, \delta) = \lim_{r \to +\infty} \text{ADT}(b, r, \delta)\), called stationary average delay time. In fact there could be many ways to compare maps \((r, \delta) \mapsto \text{ADT}(b, r, \delta)\). For instance, when \(r\) is known and \(\delta\) unknown, Roberts (1966) proposes the cost function \(\int_{\delta}^{3.5 \times \sigma} \text{ADT}(b, r, \delta) d\delta\), or a weighted integrated average delay time, with weights depending on the applications, see \[17, p.428\].

Several detection rules have been developed over the past decades, see for e.g. Montgomery (1985) or Hawkins and Olwell (1998). CUSUM procedure [Page (1954)] is the most popular and the most studied. This rule is based on the likelihood of a shift of known size \(\delta\), see for instance \[1, \text{chap. 2}\]. Another approach leads to quasi-Bayesian detection rules, called Shiryaev–Roberts [ Shiryaev (1963), Roberts (1966)]. Comparisons of the CUSUM procedure and the Shiryaev–Roberts procedure based on different criteria have been done by Pollack and Siegmund (1985), Moustakides (1986) and Srivastava and Wu (1993) and they conclude that the CUSUM procedure and the Shiryaev–Roberts procedure are equivalent to the first order and differ slightly to the second order, the Shiryaev–Roberts procedure is better with respect to SADT and the CUSUM procedure is better with respect to \(\sup_{r \geq 1} \text{ADT}(b, r, \delta)\), see \[19\].

The aim of this paper is to show a certain efficiency of finite moving averages (FMA)
for detecting small shift, for instance \( \delta/\sigma = 0.5 \). FMA depend on two parameters: window size \( A \) and a threshold \( h \). Their properties have been investigated for fixed small window sizes (\( A = 3,4,5 \) in [3] or 8,9 in [17]) and the asymptotic study of the scheme ARL vs. ADT done with respect to \( h \to +\infty \), see [7, 3]. We propose to choose a fixed threshold \( h < \delta \) and to study the asymptotic of ARL vs. ADT when \( A \to +\infty \). We find the same first order term as for Cusum procedure or Shiryaev–Roberts procedure multiplied by \( h/\delta \).

Our plan is the following. In Section 2, we study the finite moving averages with respect to the parameter \( A \) (size of window) and for a fixed threshold \( h \). In Section 3, we compare the finite moving average with the Cusum and GLR detection rules, from a theoretical point of view. In Section 4, we give numerical simulations, and present our conclusions on the interest of this new approach. The proofs of Section 2 are given in the appendix.

2 Some Properties of Finite Moving Averages. First, we precise the definition of the FMA detection rule. Following [2], we define

\[
D(A, k) = A^{-1} \left( \sum_{j=k-A+1}^{k} X_j \right) - \theta_0, \quad \text{for } k \geq A,
\]

and the stopping times

\[
T(A, h) := \inf \{ k \geq A \text{ such that } D(A, k) \geq h \}.
\]

Remark that the couple \( (A, h) \) appears as the parameter of the family of stopping times, thus, by replacing \( b \) by \( (A, h) \), we derive the definition of \( ARL(A, h) \) and \( ADT(A, h, r, \delta) \).

We recall the notations

\[
\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( -\frac{u^2}{2} \right) du \quad \text{and} \quad \Psi(x) = 1 - \Phi(x). \tag{2.1}
\]

The law of \( T(A, h) \) have exponential decay, moreover we have the following bounds on the average run length

**Proposition 2.1** (ARL). 1) For all \( A \in \mathbb{N}^* \) and \( h > 0 \), we have

\[
(A - 1) + \Psi^{-1} \left( \frac{h}{\sigma} \sqrt{A} \right) \leq ARL(A, h) \leq \Psi^{-1} \left( \frac{h}{\sigma} \sqrt{A} \right). \tag{2.2}
\]

2) We have another lower bound, \( \forall \epsilon > 0, \exists K_\epsilon > 0, \text{ such that } \forall A \in \mathbb{N}, \)

\[
ARL(A, h) \geq (K_\epsilon A) \times \Psi^{-1} \left( \frac{h}{\sigma} \sqrt{\frac{A}{1 + \epsilon}} \right). \tag{2.3}
\]
3) For every $k \geq 2A - 2$, we have

$$P_0 \{ T(A, h) > k + 1 \} = \beta \times P_0 \{ T(A, h) > k \} \quad (2.4)$$

for a constant $\beta \in (0, 1)$ given by (A.2).

**Proof.** Formula (2.2) follows from Lai [7, (8), p.140], the lower bound (2.3) and formula (2.4) are proved in Appendix. $\square$

The decision function $k \mapsto D(A, k)$ is the sum of a half hat function of size $\delta$ beginning at the change time $r$ and of an empirical process, i.e.

$$D(A, k) = \chi(r, k) \times \delta \times g_1 \left( \frac{k - r}{A} \right) + A^{-1} \left[ S_k - S_{k-A} \right] \quad (2.5)$$

with $\chi(r, k) = \begin{cases} 1 & \text{if } k \geq r, \\ 0 & \text{if } k < r, \end{cases}$ where by convention $r = \infty$ when there is no change point,

$$g_1(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } x \geq 1, \end{cases}$$

and $S_k = \sum_{i=1}^k (X_i - \bar{X}_i)$. We deduce that

$$\forall r \geq A, \quad ADT(A, h, r, \delta) = ADT(A, h, A, \delta) \quad (2.6)$$

and the following bounds.

**Proposition 2.2 (ADT).** 1) When $h \leq \delta$, we have

$$ADT(A, h, r, \delta) \leq A \times \zeta(A, h, \delta) \quad (2.7)$$

where $\zeta(A, h, \delta)$ does not depend on $r$ and

$$\zeta(A, h, \delta) \leq 1 + \Psi \left( \frac{\delta - h}{\sigma} \sqrt{A} \right) + 4 \Psi \left( \frac{\delta - h}{\sigma} \sqrt{A} \right)^2 \quad (2.8)$$

Moreover, if $h < \delta$, when $A \to +\infty$, we have

$$\zeta(A, h, \delta) = \frac{h}{\delta} + \mathcal{O} \left( A^{-1} \exp \left( -\frac{1}{2} A(h \wedge \delta - h)^2 \right) \right). \quad (2.9)$$

2) When $h \geq \delta$, we have

$$A + ARL(A, h - \delta) \leq ADT(A, h, r, \delta) \leq 4A + ARL(A, h - \delta). \quad (2.10)$$

**Proof.** See Appendix. $\square$
Asymptotic Properties. From the above propositions, we deduce the asymptotic of the parametric curve ADT versus ARL, when \( A \to \infty \).

**Corollary 2.1.** 1) When \( h < \delta \), we have
\[
\xi_1(A, h) \leq \text{ARL}(A, h) \leq \xi_2(A, h),
\]
and asymptotically when \( y \to +\infty \), with \( y = \text{ADT}(h, A, r, \delta) \)
\[
\ln(\xi_1) = \frac{1}{2} \left( \frac{h}{\delta} \right) \left( \frac{\delta}{\sigma} \right)^2 y + \frac{1}{2} \ln(y) + \ldots
\]
\[
\cdots + \frac{1}{2} \ln \left( \frac{h}{\sigma} \right) + \frac{1}{2} \ln \left( \frac{\delta}{\sigma} \right) + \frac{1}{2} \ln(2\pi) + O(1/y)
\]  
(2.11)

\[
\ln(\xi_2) = \frac{1}{2} \left( \frac{h}{\delta} \right) \left( \frac{\delta}{\sigma} \right)^2 y + \frac{3}{2} \ln(y) + \ldots
\]
\[
\cdots + \frac{1}{2} \ln \left( \frac{h}{\sigma} \right) + \frac{1}{2} \ln \left( \frac{\delta}{\sigma} \right) + \frac{1}{2} \ln(2\pi) + O(1/y)
\]  
(2.12)

2) When \( h > \delta \), as \( x = \text{ARL}(A, h) \to +\infty \), for \( y = \text{ADT}(A, h, r, \delta) \) we have
\[
\ln(y) = \ln(x) - \sqrt{2}\delta A^{1/2} \ln(x)^{1/2} + O(1)
\]  
(2.13)

**Proof.** 1) From (2.7) and (2.9), when \( A \to \infty \), we get
\[
y = A \left( \frac{h}{\delta} \right) + O \left( \exp \left( -\frac{1}{2} A(h \land \delta - h)^2 \right) \right).
\]
Since \( \Psi(x) \sim \frac{1}{\sqrt{2\pi}} \left( \frac{1}{x} \right) \exp \left( -\frac{x^2}{2} \right) \), when \( x \to +\infty \), using (2.2) we get
\[
\ln(\xi_1) = \frac{1}{2} \left( \frac{h}{\sigma} \right)^2 A + \frac{1}{2} \ln(A) + \frac{1}{2} \ln(2\pi) + \ln \left( \frac{h}{\sigma} \right) + O(1/A)
\]
which implies (2.11). Similarly, we obtain (2.12).

2) Since \( h - \delta > 0 \), when \( A \to +\infty \) we deduce from (2.2, 2.10)
\[
x = \text{ARL}(A, h) \sim \sqrt{2\pi} h \sqrt{A} \exp \left( \frac{1}{2} A h^2 \right),
\]
\[
y = \text{ADT}(A, h, r, \delta) \sim \sqrt{2\pi} (h - \delta) \sqrt{A} \exp \left( \frac{1}{2} A (h - \delta)^2 \right),
\]
which implies (2.13). □

We have shown a qualitative change on the asymptotic of the curve ARL vs. ADT as soon as \( h > \delta \).
Remark 2.1. From (2.2) and (2.10), we deduce the asymptotic for $A$ fixed when $h \to +\infty$

$$\ln(x) = \ln(y) + \sqrt{2} \delta A^{1/2} (\ln(y))^{1/2} - \frac{1}{2} A \delta^2 + o(1) \quad (2.14)$$

where $x = ARL(A, h)$ and $y = ADT(A, h, r, \delta)$. The first order term is not good compared to the asymptotic for $A \to +\infty$, see (2.11), or Figure 4.1.

3 Comparison with other Detection Rules.

3.1 Case $\delta$ known. Exact knowledge of $\delta$ is not very realistic for the applications, but simpler for a theoretical study. In this case we can choose the FMA detection rule with $h = \delta / \left(1 + \frac{\epsilon}{2}\right)$ for any $\epsilon > 0$ and we get

$$\ln(\xi_1) = \frac{1}{2 + \epsilon} \left(\frac{\delta}{\sigma}\right)^2 y + \frac{1}{2} \ln(y) + \ln\left(\frac{\delta}{\sigma}\right) + \frac{1}{2} \ln(2\pi) + \frac{\epsilon}{2} + O\left(1/y\right). \quad (3.1)$$

For the CUSUM detection rule, we have [19, Th. 6, p. 663]

$$\frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 y = \ln\left(\frac{1}{2} x (\delta/\sigma)\right) - 1.5 + O(\ln(x)/x) \quad (3.2)$$

where $y = SADT(h, \delta)$ and $x = ARL(h, \delta)$, or equivalently

$$\ln(x) = \frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 y - 2 \ln(\delta/\sigma) + \ln(2) + 1.5 + O\left(\frac{\ln(x)}{x}\right) \quad (3.3)$$

and for the Shiryaev–Roberts detection rule

$$\ln(x) = \frac{1}{2} \left(\frac{\delta}{\sigma}\right)^2 y - 2 \ln(\delta/\sigma) + \ln(2) + 1.5772 + O\left(\frac{\ln(x)}{x}\right). \quad (3.4)$$

If $h$ was equal to $\delta$, we should obtain the same first order term as for the CUSUM or Shiryaev–Roberts detection rules, the second term playing in favor of finite moving average for very large values of ARL. We find the same kind of results as for geometric moving average, most often called exponentially weighted moving average (EWMA) procedure, from [16, Cor. 1, p. 658] it results that for EWMA detection rule (with an optimal choice of the forgetting coefficient $\beta$), we have

$$\ln(x) = \frac{1}{2.4554} \left(\frac{\delta}{\sigma}\right)^2 y + \frac{1}{2} \ln(y) + O(1) \quad (3.5)$$
with \( x = ARL \) and \( y = SADT \). Therefore, we have the same second order term, but in the first order term \( 1/2.4554 \) is replaced by \( 1/(2 + \epsilon) \) for any \( \epsilon > 0 \).

**Remark 3.1.** For stationary average delay time, finite moving average appears as asymptotically better than CUSUM or Shiryaev–Roberts detection rules, but not for the criterion \( \sup_{r \geq 0} ADT(b, r, \delta) \). Indeed, considering the case \( h = \delta \), we have from (2.3) and (2.10)

\[
\eta = \sup_{r \geq 0} ADT(b, r, \delta) \geq A + ARL(A, 0) \geq A(1 + 2K_\epsilon)
\]

with \( K_\epsilon > 0 \), which leads to

\[
\ln(\xi_2) = \frac{1}{1 + 2K_\epsilon} \frac{1}{2} \left( \frac{h}{\delta} \right) \left( \frac{\delta}{\sigma} \right)^2 \eta + O(1).
\]

We have lost a multiplicative factor \( 1 + 2K_\epsilon \) in the first order term.

### 3.2 Case \( \delta \) unknown. Comparison with GLR.

We assume that \( \theta_0 \) is known (or estimated), but not \( \theta_1 \), therefore \( \delta \) is unknown. This is a case currently met in the applications. In this case CUSUM should be seen as a detection rule depending on the two parameters \((h, \nu)\), where \( \nu \) is the reference value of the shift. Another popular detection rule is generalized likelihood ratio (GLR), which is derived from the likelihood of existence of a positive shift \((\delta > 0)\), see [1]. The stopping time is defined by

\[
\tilde{T}_b = \inf \left\{ n \text{ such that } \max_{0 \leq k < n} \frac{(S_n - S_k)}{(n-k)^{1/2}} \geq b \right\}
\]

where \( S_n = \sum_{i=1}^{n} X_i \). This detection rule depends on one parameter, the threshold \( b \). The asymptotic of average run length is studied in Siegmund and Venkatraman (1995), see [16, Th. 1, p. 257 and (3.1), p.260]. They proved that, for \( \sigma = 1 \), when \( b \to \infty \),

\[
ARL(b) \sim Ctt \times \frac{\exp(b^2/2)}{b},
\]

\[
ADT(b, r, \delta) \leq ADT(b, 0, \delta) \sim \frac{b^2 - 1}{\delta^2} + g_0(\delta).
\]

where \( g_0(\delta) \) is an expression only depending on \( \delta \). We deduce the asymptotic of the curve \( ARL \) vs. \( ADT \) for GLR

\[
\ln(x) = \frac{1}{2} \left( \frac{\delta}{\sigma} \right)^2 \ln(y) - \frac{1}{2} \ln(y) + g_1 \left( \frac{\delta}{\sigma} \right) + o(1/y)
\]  \( (3.6) \)
Comparison with (2.12) gives a feeling of the relative efficiency of GLR and FMA. Indeed, when \( h = \delta \), considering asymptotic formulas (2.12) and (3.6) as exact, FMA is better than GLR iff for a given value of \( y = ADT \), \( x = ARL \) is larger, which is to say

\[
\frac{1}{2} \ln(y) + \ln\left(\frac{\delta}{\sigma}\right) + \frac{1}{2} \ln(2\pi) \geq - \frac{1}{2} \ln(y) + g_1\left(\frac{\delta}{\sigma}\right)
\]

\[\iff \ln(y) \geq g_2\left(\frac{\delta}{\sigma}\right).\] (3.7)

For any fixed value of \( \delta \), condition (3.7) is fulfilled for \( y \) large enough, and after for ARL large enough, more precisely when \( ARL(A, \delta) \geq \phi(\delta/\sigma) \) and \( r \geq A \). When \( \delta > h \), from (2.11) and (3.6) FMA is better than GLR iff

\[
\ln(y) \geq \frac{1}{2} \left(\frac{\delta (\delta - h)}{\sigma}\right) y + g_2\left(\frac{\delta}{\sigma}\right)
\] (3.8)

which implies that for each \( h \), when \( \delta > d_0(h) \), GLR is better than FMA, and when \( \delta \in (h, d_0(h)) \), FMA is better for a certain interval of values of \( y = ADT \).

**Remark 3.2.** We just propose an explanation of the observed facts. The previous calculations are not at all rigorous, but explain what the Monte–Carlo simulations plainly confirm, see next section.

4 Numerical Comparison.

4.1 Description of the Simulations. In this section, we compare CUSUM, GLR and FMA detection rules by Monte–Carlo simulations. We have fixed the change time at \( r = 500 \) and take the variance \( \sigma^2 = 1 \). Thus for each graph, we have simulated a sequence of 1000 independent Gaussian r.v.’s \( X_n \) with variance 1 and mean 0 for \( n \leq 500 \) and mean \( \delta \) for \( n \geq 501 \) by Box–Muller method using a Lehmer pseudo–random generator, see [5] or [8]. The real numbers are coded with 20 bits. Simulations done with the Lewis–Payne [9] pseudo–random generator lead to similar results. The three detection rules have been applied on the same 1000 sequences of r.v.’s.

Concerning the complexity, the FMA procedure is standard and can be implemented on EXCEL. We need an array of real of size \( A \), where \( A \) is the window size, indeed we have to stock the real numbers \( (X_i)_{k-A+1 \leq i \leq k} \) and \( S(A, k) = \sum_{i=A-k+1}^{k} X_i \) since the condition
\[ D(A, k) = \frac{1}{A} \sum_{i=A-k+1}^{k} X_i > h \text{ is equivalent to } S(A, k) > A \times h. \]

By a ring implementation, we just have to do two assignments at each step: replace \( X_{k-A+1} \) by \( X_{k+1} \), \( S(A, k+1) = S(A, k) - X_{k-A+1} + X_{k+1} \), and to shift the pointer on the ring. The space complexity is \( \mathcal{O}(A) \) and the time complexity is constant and very small at each step. The time to plot the parametric curves ARL vs. ADT is essentially the time to simulate the Gaussian random variables \((X_i)_{1 \leq i \leq n}\).

The GLR procedure needs to stock the data from the beginning, i.e. \((X_i)_{1 \leq i \leq k}\). The space complexity of GLR procedure is linearly increasing as the number of observations. This should induce to use mass memory, which would slow the computations. The CUSUM procedure only uses two memories of real number to stock \( S_n \) and \( \min_{k<n} (S_k - k \cdot \nu) \) since it is equivalent to test \( S_n - \frac{n \cdot \nu}{2} - \min_{k<n} (S_k - \frac{k \cdot \nu}{2}) > h \) with \( S_k = \sum_{i=1}^{k} X_i \).

### 4.2 Presentation of the Results.

Two kinds of representations are plotted. The curves represent the mean of 1000 simulated series.

1. When the shift \( \delta \) is known, we plot the parametric curves ARL vs. ADT, i.e. \( x = ARL(b), y = ADT(b, 500, \delta) \). The best detection rule corresponds to the lower curve, i.e. ADT smaller for a given ARL.

2. When the shift \( \delta \) is unknown, we fix a certain value of ARL, for instance \( ARL = 1000 \) which determine \( b_0 \) such that \( ARL(b_0) = 1000 \) for each detection rule, and we plot the parametric curve \( \delta, ADT(b_0, 500, \delta) \). Of course, \( ADT(b_0, 500, 0) = ARL(b_0) \). Once again, the better detection rule corresponds to the lower curve.

In the first series we compare FMA with fixed window size \( (A = 8) \) and a variable threshold \( h, \) FMA with a fixed threshold \( h = 0.9 \times \delta \) and the CUSUM procedure with \( \nu = \delta \). The value of the shift \( \delta \) is assumed to be known, in the two following pictures (Fig. 4.1 and Fig. 4.2), we have \( \delta = 0.25 \) and \( \delta = 0.5 \). Finite moving average appears as bad detection rules when window size \( A \) remains fixed but as good as the CUSUM detection rule when the threshold is fixed \( h < \delta \), for instance \( h = 0.9 \times \delta \) and window size is varying. Simulations suggest that when the r.v.’s \( X_i \) are dependent, then FMA detection rules could be better than the CUSUM detection rule.

In the second series, we compare the FMA, CUSUM and GLR detection rules when \( \delta \) is known. We have considered these detection rules for two-sided shift, i.e. for a positive shift
Detecting Small Shift on the Mean by Finite Moving Average

Figure 4.1: Case $\delta = 0.25$.

Figure 4.2: Case $\delta = 0.5$. 
δ or a negative shift −δ. In fact, numerical simulation yield the same kind of results as for one-sided shift. Since the value of the shift δ is known, we consider FMA with \( h = 0.9 \times \delta \) and CUSUM with \( \nu = \delta \), the two pictures Fig. 4.3 and Fig. 4.4 correspond to \( \delta = 0.5 \) and \( \delta = 1 \).

![Bilateral Tests, Delta=0.5](image)

Figure 4.3: Bilateral tests, case \( \delta = 0.5 \).

The GLR procedure appears as less efficient than FMA and CUSUM procedure when \( \delta \) is known, except for large value of the shift, i.e. \( \delta > 2 \sigma \) which have not been presented here.

In the last series (Fig. 4.5 and Fig. 4.6), we compare the FMA, CUSUM and GLR detection rules when the value of the shift δ is unknown. We take ARL = 1000, and consider FMA and CUSUM procedures corresponding to an assumed value of the shift \( \delta = 0.5 \), i.e. \( h = 0.45 \) and \( \nu = 0.5 \).

Finite moving average is slightly better than CUSUM when \( \delta < 0.4 \) and really better than GLR, for instance for \( \delta = 0.3 \), we have \( ADT_{FMA} = 80 \), \( ADT_{CUSUM} = 85 \) and \( ADT_{GLR} = 98 \), but FMA is the worst for \( \delta > 0.7 \). Fixing \( ARL = 500 \), we get the same result, the difference between FMA and CUSUM for small shift becoming very small, see Fig. 4.5 and Fig. 4.6.
Figure 4.4: Bilateral tests, case $\delta = 1$.

Figure 4.5: Bilateral tests, $\delta$ unknown, ARL=1000.
4.3 Conclusion. When window size varies instead of the threshold, finite moving averages appears as good detection rules. We have proved that when the shift $\delta$ is known for stationary average delay time the first order term is the same as for the CUSUM, Shiryaev–Roberts or GLR detection rules, and the second term plays in favor of the finite moving average, as for EWMA. Numerical simulations shows that the CUSUM procedure and Finite moving average detection rules are very closed when $\delta$ is known and $\delta < 2 \times \sigma$ or $\delta$ unknown. Both are better than GLR detection rule for detecting small shifts, for instance $\delta < 0.5 \times \sigma$.

More works should be done to obtain analytic formulas for FMA detection rules. Another question that should be studied is the case of abrupt change on the parameters for families of dependent random variables.


Formula (2.3). There is no change–point, using (2.5), we get after scaling

$$\sup_{A \leq k \leq N} D(A, k) \overset{(C)}{=} \sigma A^{-1} \sqrt{N} \rho \left( \frac{A}{N} \right)$$
with $\rho(x) \leq \sup_{u \in [0,1]} [W_{u+x} - W_u]$ where $W_t$ is a standard Brownian motion. Therefore

$$\mathbb{P}_0 \{ T(A, h) \leq N \} = \mathbb{P} \left\{ \frac{\rho(A/N)}{\sigma \sqrt{N}} \geq \frac{Ah}{\sqrt{N}} \right\}.$$  \hspace{1cm} (A.1)

From [4, Lemma 1.1.1, p. 24], for each $\epsilon > 0$ there exists $C_\epsilon > 0$ such that

$$\mathbb{P} \left\{ \rho(x) > v \sqrt{x} \right\} \leq C_\epsilon \frac{1}{x} \Psi \left( \frac{v}{\sqrt{2 + \epsilon/2}} \right).$$

Let $x = A/N$ and $v = h \sqrt{A}/\sigma$, we deduce

$$\mathbb{P}_0 \{ \tau_1^h(A) \leq N \} \leq C_\epsilon \frac{N}{A} \Psi \left( \frac{h \sqrt{A}}{\sqrt{2 + \epsilon/2}} \right).$$

Taking $N = \varphi(A)$ with $\varphi(A) = A/2 \cdot C_\epsilon^{-1}/\Psi \left( \frac{h}{\sigma} \sqrt{\frac{A}{2 + \epsilon/2}} \right)$, we get $\mathbb{P}_0 \{ \tau_1^h(A) < \varphi(A) \} \leq 1/2$.

We deduce

$$\mathbb{E}_0 \{ T(A, h) \} \geq \mathbb{E}_0 \{ T(A, h) \mathbf{1}_{[T(A, h) \geq \varphi(A)]} \} \geq \frac{1}{2} \varphi(A)$$

which induces formula (2.3) with $K_\epsilon = 1/(4C_\epsilon)$.

**Formula (2.4).** Denote $F(k) = \mathbb{P}_0 \{ T(A, h) > k \}$. For $k > A$, we get

$$F(k + 1) = \mathbb{P}_0 \{ D(A, k + 1) < h \text{ and } T(A, h) > k \}$$

$$= F(k) \times \mathbb{P}_0 \{ D(A, k + 1) < h \mid T(A, h) > k \}$$

$$= F(k) \times \mathbb{P}_0 \{ D(A, k + 1) < h \mid D(A, i) < h, \forall i \in [A, k] \}$$

$$= F(k) \times \mathbb{P}_0 \{ D(A, k + 1) < h \mid D(A, i) < h, \forall i \in [A, k] \cap [k + 2 - A, k] \}.$$

The last equality follows from the independence of the r.v.’s $D(A, k + 1)$ and $D(A, i)$ when $(k + 1) - A + 1 > i$. If moreover $k \geq 2A - 2$, then we have $[A, k] \cap [k + 2 - A, k] = [k + 2 - A, k]$ and the sequence of r.v.’s $D(A, i)$, $i = k + 2 - A, \ldots, k$ is identically distributed for every $k$. Therefore we deduce (2.4) with

$$\beta = \mathbb{P}_0 \{ D(A, k + 1) < h \mid D(A, i) < h, \forall i \in [k + 2 - A, k] \}.$$  \hspace{1cm} (A.2)

This finishes the proof of Proposition 2.1. □
Proof of Proposition 2.2. First, remark that for every \( h, \delta, A \), the map \( r \mapsto ADT(A, h, r, \delta) \) is decreasing. Combined with (2.6), we get

\[
\sup_{r \geq 0} ADT(A, h, r, \delta) = ADT(A, h, A, \delta) \geq ADT(A, h, 0, \delta) = A + ARL(A, h - \delta), \tag{A.3}
\]

the last equality being obvious on a picture. This gives us the first inequality of (2.10) and it suffices to prove the upper bound of (2.10) for \( r \geq A \). For notational convenience, let us omit conditioning by \( T_{h1}(A) \geq r \). For \( r \geq A \) we have

\[
ADT(A, h, r, \delta) = I_1(A, h, \delta) + I_2(A, h, \delta) + I_3(A, h, \delta),
\]

with

\[
I_1(A, h, \delta) = \mathbb{E}_{r, \delta} \left\{ [T(A, h) - r] 1_{\{T_{h1}^A(A) \in [r, r + A] \}} \right\},
\]

\[
I_2(A, h, \delta) = \mathbb{E}_{r, \delta} \left\{ [T(A, h) - r] 1_{\{T_{h1}^A(A) \in [r + A, r + 2A] \}} \right\},
\]

\[
I_3(A, h, \delta) = \mathbb{E}_{r, \delta} \left\{ [T(A, h) - r] 1_{\{T_{h1}^A(A) > r + 2A \}} \right\}.
\]

We bound \( I_1(A, h, \delta) \) by using the comparison lemma A.1 (see below). Let \( k \) be an integer \( k \leq A \), we get

\[
\mathbb{P}_{r, \delta} \{ T(h, A) \leq r + k \} = \mathbb{P}_{r, \delta} \left\{ \max_{r + 1 \leq i \leq r + k} D(A, i) \geq h \right\} \geq \mathbb{P}_{r, \delta} \{ D(A, r + k) \geq h \}.
\]

Since \( \mathbb{E}_{r, \delta} D(A, i) \) is a half hat-function, we have \( D(A, r + k) \overset{(c)}{=} \mathcal{N}(\frac{k \delta}{A}, \frac{\sigma}{\sqrt{A}}) \) which implies

\[
\mathbb{P}_{r, \delta} \{ T(h, A) \leq r + k \} \geq \Phi \left( \frac{\sqrt{A}}{\sigma} \left( \frac{k \delta}{A} - h \right) \right). \tag{A.4}
\]

Lemma A.1, gives us

\[
I_1(A) \leq \sum_{k=1}^{A} \Psi \left( \frac{\sqrt{A}}{\sigma} \left( \frac{k \delta}{A} - h \right) \right) \leq A. \tag{A.5}
\]

On the other hand, since \( \Psi(x) \) is a decreasing function, we have

\[
I_1(A, h, \delta) \leq I_1^*(A, h, \delta) \leq \int_0^A \Psi \left( \frac{\sqrt{A}}{\sigma} \left( \frac{t \delta}{A} - h \right) \right) dt = A \int_0^1 \Psi \left( \frac{\sqrt{A}}{\sigma} \left( \delta u - h \right) \right) du,
\]
which combined with the Lebesgue dominated convergence theorem induces

$$\lim_{A \to +\infty} A^{-1} I^*_1(A, h, \delta) = \begin{cases} \frac{h}{\delta} & \text{if } h \leq \delta, \\ 1 & \text{if } h \geq \delta. \end{cases}$$

(A.6)

Straight calculations lead to $A^{-1} I^*_1(A, h, \delta) = \frac{h}{\delta} + O \left( A^{-1} \exp \left( -\frac{1}{2} A (h \wedge \delta - h)^2 \right) \right)$. Exactly the same proof gives us

$$I_2(A, h, \delta) \leq A \Psi \left( \frac{\sqrt{A} (\delta - h)}{\sigma} \right).$$

(A.7)

It remains to bound $I_3(A, h, \delta)$. We have

$$I_3(A, h, \delta) = 2A \mathbb{P}_{r, \delta} \{ T(h, A) \geq r + 2A \} + \mathbb{E}_{r, \delta} \left\{ [T(h, A) - (r + 2A)] \mathbb{1}_{\{T(h, A) \geq r + 2A\}} \right\}$$

$$= \mathbb{P}_{r, \delta} \{ T(h, A) \geq r + 2A \} \times [2A + \text{ARL}(A, h - \delta)]$$

From one hand, when $h \leq \delta$, we have $\text{ARL}(A, h - \delta) \leq \text{ARL}(A, 0) \leq 2A$ by (2.2) which implies

$$I_3(A, h, \delta) \leq 4A \times \mathbb{P}_{r, \delta} \{ T(h, A) \geq r + 2A \}$$

(A.8)

But, since the r.v.'s $D_1(A, r + A + 1)$ et $D_1(A, r + 2A + 1)$ are independent, we get

$$\mathbb{P}_{r, \delta} \{ T(h, A) \geq r + 2A \} \leq \mathbb{P}_{r, \delta} \{ D(A, r + A) < h \text{ and } D(A, r + 2A) < h \}$$

$$= \Phi \left( \frac{h - \delta}{\sigma} \sqrt{A} \right)^2.$$ 

(A.9)

Combined with (A.5), (A.6), (A.7) and (A.8) this gives us (2.8) and (2.9). On the other hand, when $h \geq \delta$, we have

$$\text{ADT}(A, h, r, \delta) \leq A \times \left[ 1 + \Phi \left( \frac{\sqrt{A} (h - \delta)}{\sigma} \right) + 2 \Phi \left( \frac{\sqrt{A} (h - \delta)}{\sigma} \right)^2 \right]$$

$$+ \Phi \left( \frac{\sqrt{A} (h - \delta)}{\sigma} \right)^2 \times \text{ARL}(A, h - \delta)$$

which combined with (A.3) gives us (2.10). □

**Remark A.1.** When $\sigma_0 \neq \sigma_1$, we just have to replace $\sigma$ by $\sigma_1$ in the bounds of $I_2$, $I_3$, (A.7) and (A.9). To bound $I_1$, formula (A.4) should be adapted, we have to replace $\sigma$ by a function $s(k)$ see [2, Lemma B.1]. Anyway the calculations remain the same and we still obtain (2.8), (2.9) and (2.10).
Lemma A.1. Let $X$ and $Y$ be two r.v.’s with corresponding c.d.f’s $F(t) = \mathbb{P}\{X \leq t\}$ and $G(t) = \mathbb{P}\{Y \leq t\}$. Assume that for every $t > 0$, $F(t) \geq G(t)$, then for every $B > 0$ $\mathbb{E}(X 1_{X \in [0, B]}) \leq \int_0^B [1 - G(x)] \, dx$.

Proof. By integration by parts,

$$
\mathbb{E}(X 1_{X \in [0, B]}) = \int_0^B x \, dF(x) = \int_0^B [F(B) - F(x)] \, dx \leq \int_0^B [1 - G(x)] \, dx.
$$

\[\square\]

Acknowledgment: The authors are indebted to Professor Arcones, Editor–in–chief, to the Associate Editor and the referees for their comments and helpful suggestions. In particular, they suggest an improvement of the proof and the statement of Lemma A.1.

References


Detecting Small Shift on the Mean by Finite Moving Average


