Validity and efficiency of the conformal CUSUM procedure

Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman

практические выводы
теории вероятностей
могут быть обоснованы
в качестве следствий
гипотез о *предельной*
при данных ограничениях
сложности изучаемых явлений

Abstract

In this paper we study the validity and efficiency of a conformal version of the CUSUM procedure for change detection both experimentally and theoretically.

Contents

1 Introduction

One of the most popular methods for change detection is Page's [9] CUSUM procedure; see [10, Chap. 6] for a modern exposition. The CUSUM procedure has been shown to be optimal in some important senses, including Lorden's [7, 8] worst-case definition and Ritov's [11] game-theoretic definition.

The basic setting in which the optimality of CUSUM has been shown is where the pre-change and post-change distributions are known, and the only unknown is the time when the change happens (the changepoint). The conformal CUSUM procedure, introduced in [14] and discussed in detail in [15, Chap. 8], only assumes that the pre-change data distribution is IID (the pre-change observations are independent and identically distributed).

In [15, Chap. 8] we only discussed the *validity* of the conformal CUSUM procedure, meaning the frequency of false alarms (i.e., alarms raised *before* the changepoint) being bounded (at least with a high probability) by a specified constant. Another desideratum for the conformal CUSUM procedure is its *efficiency*: we would like an alarm to be raised soon *after* the changepoint. We did not investigate the efficiency of our change detection procedures in [15] since the task of proving efficiency is in a sense too ambitious [15, Sect. 8.6.3]: conformal testing may be based on powerful algorithms, such as deep neural networks, involving elements of intelligence, and there is little hope of guaranteeing their efficiency.

In this paper we investigate the efficiency of conformal change detection in a very roundabout way, which is an instance of what we called the Burnaev–Wasserman programme in [15, Sect. 2.5]. Suppose we model our observations as coming in the IID fashion from a distribution Q_0 before the changepoint and from another distribution Q_1 after the changepoint. This is, however, a "soft model": we do not want the validity of our procedure to depend on it. The soft model provides us with a means of establishing the efficiency of our procedure: the efficiency is measured under this model.

We started implementing this programme in the easier case of conformal e-testing in [16]. In this paper we extend it to conformal testing, which has important advantages over conformal e-testing [16, Sect. 7].

We start the main part of this paper from a description of the conformal CUSUM procedure in Sect. 2. The conformal CUSUM procedure is determined by a threshold and an underlying conformal test martingale, and the latter is in turn determined by its underlying nonconformity measure and betting martingale. We define asymptotically optimal, in a natural sense, nonconformity measure and betting martingale in Sect. 3. The betting martingale consists in applying the same betting function at each step, and the betting function is explicitly computed in two popular special cases (both involving the Gaussian model) in Sect. 4. An experimental section, Sect. 5, illustrates the efficiency of the asymptotically optimal conformal CUSUM procedure that we derive in the previous sections, and Sect. 6 is an attempt of a theoretical analysis. Finally, we summarize and list some directions of further research in Sect. 7.

In this paper we will refer to the assumption that the observations are IID

as the assumption of randomness. By de Finetti's theorem this assumption is equivalent to that of exchangeability for observations taking values in a standard Borel space.

2 Conformal and standard CUSUM procedures

Let S be a conformal test martingale (CTM) as defined in [15, Sect. 8.1.2]; for simplicity we will assume that it is positive, $S > 0$. For a given threshold $c > 1$, the *conformal CUSUM procedure* raises the kth alarm, $k = 0, 1, \ldots$, at the time

$$\tau_k := \min \left\{ n > \tau_{k-1} : \max_{i = \tau_{k-1} \ldots, n-1} \frac{S_n}{S_i} \geq c \right\}, \tag{1}$$

where $\tau_0 := 0$ and $\min \emptyset := \infty$. A universally applicable but conservative property of validity of the conformal CUSUM procedure is that

$$\mathbb{E}(\tau_k - \tau_{k-1}) \geq c \tag{2}$$

for IID observations. This implies (albeit not trivially) that, under randomness,

$$\limsup_{n \to \infty} \frac{A_n}{n} \leq \frac{1}{c} \quad \text{a.s.,} \tag{3}$$

where $A_n := \max\{k : \tau_k \leq n\}$ is the number of alarms raised by the conformal CUSUM procedure by time n.

Next, with a view towards the Burnaev–Wasserman programme, we consider the standard CUSUM procedure. Let us use the notation of [16]: the observation space is $\mathbf{Z}$ (a measurable space), the pre-change distribution is Q_0, and the post-change distribution is Q_1. Let f_0 be the probability density of the pre-change distribution Q_0 and f_1 be the probability density of the post-change distribution Q_1 with respect to some σ-finite measure m (such as $m := Q_0 + Q_1$). The *likelihood ratio* is

$$L(z) := \frac{f_1(z)}{f_0(z)}, \quad z \in \mathbf{Z}; \tag{4}$$

for simplicity we will assume that both f_0 and f_1 are positive, so that $L \in (0, \infty)$. The corresponding *likelihood ratio martingale* (LRM) is

$$S_n := \prod_{i=1}^{n} L(z_i), \quad n = 0, 1, \ldots. \tag{5}$$

The *standard CUSUM procedure* corresponding to a threshold $c > 1$ and the pre-/post-change distributions (Q_0, Q_1) is defined by (1), where S is the LRM (5).

The standard CUSUM procedure satisfies the properties of validity, (2) and (3), that we mentioned earlier for the conformal CUSUM procedure. As we also mentioned earlier, these properties of validity are very conservative for CUSUM

2

(since, as shown in, e.g., [15, Chap. 4], they are also applicable to the Shiryaev–Roberts procedure, which raises many more alarms). A possible solution is to use computer simulations to find a suitable value of the threshold c for a given target value for the frequency of false alarms (as in [15, Sect. 8.4.3], where we also find confidence intervals for the frequency of false alarms). Many important pairs (Q_0, Q_1) of pre-/post-change distributions are even amenable to theoretic analysis, and it is possible for them to derive very accurate approximations to the value of c leading to a chosen "ARL2FA" ("average run length to false alarm"), a standard measure of validity for the standard CUSUM procedure. In particular, analysis of a discrete-time Gaussian case is performed in [3] (and described in [1, Sect. 5.2.2.1]; see also [13, Sect. 8.2.6.6]).

We would like to construct a CTM approaching in its validity and efficiency the LRM when both are plugged into the CUSUM scheme (1). In the next section we will see that in the case of validity this goal can achieved perfectly (see Proposition 1). In the case of efficiency, our results are much weaker (Sections 5 and 6).

3 Asymptotically optimal CTM

First we briefly give a general definition, which we then specialize to the more intuitive case of a finite observation space $\mathbf{Z}$. This specialization will allow us to justify our definition in a simple setting.

An *asymptotically optimal betting function* $f : [0, 1] \to [0, \infty)$ for the pair (Q_0, Q_1) of the pre-/post-change distributions is defined by the condition

$$Q_0(\{z : L(z) > f(p)\}) \leq p \leq Q_0(\{z : L(z) \geq f(p)\}),$$

required to hold for any $p \in [0, 1]$. In other words, it is defined to be an inverse function of the *survival function* $\bar{F}(t) := Q_0(\{z : L(z) > t\})$ of the likelihood ratio L under Q_0 [2, Definition A.18]. This function is always decreasing (not necessarily strictly decreasing).

The *canonical asymptotically optimal* (CAO) betting function f is defined to be the largest function in this class, i.e.,

$$f(p) := \inf\{t : \bar{F}(t) < p\} = \sup\{t : \bar{F}(t) \geq p\}.$$

In other words, the CAO betting function is the left-continuous inverse function of the survival function $\bar{F}$ (adapt [2, Lemma A.19] to the case of a decreasing function). Equivalently, the CAO betting function can be defined by $f(p) = F^{-1}(1 - p)$, where F^{-1} is the upper quantile function of the cumulative distribution function F of L under Q_0.

Let us now assume that the observation space $\mathbf{Z}$ is finite; we will drop the curly braces in the notation such as $Q_0(\{z\})$ for $z \in \mathbf{Z}$. To specialize the general definition to the case of a finite $\mathbf{Z}$, sort the likelihood ratios

$$L(z) = Q_1(z)/Q_0(z)$$

3

(cf. (4)) in the ascending order. Let the resulting unique values for $L(z)$ be $l_1 < \cdots < l_K$. Set $q_k := Q_0(\{z : L(z) = l_k\})$ for $k = 1, \ldots, K$. An asymptotically optimal betting function $f : [0, 1] \to [0, \infty)$ is defined by $f(p) := l_k$ for any $p \in [0, 1]$, where k is defined by the condition

$$q_{k+1} + \cdots + q_K \le p \le q_k \cdots + q_K \tag{6}$$

(this condition does not define k uniquely if p has the form $q_j + \cdots + q_K$ for some $j \in \{2, \ldots, K\}$, but the probability of a uniformly distributed $p \in [0, 1]$ having this form is zero).

Let us check that the term "asymptotically optimal" is indeed justified. Let N_0 be the number of pre-change observations. Consider a very large N_0 (so that we are talking about the asymptotics $N_0 \to \infty$) and a fixed n (or n bounded above by a fixed constant). Take the likelihood ratio $L(z)$ as the nonconformity score of an observation z. The conformal p-value at the $(N_0 + n)$th step is

$$p_{N_0+n} = \frac{|\{i : L_i > L_{N_0+n}\}| + \tau |\{i : L_i = L_{N_0+n}\}|}{N_0 + n} \approx q_{k+1} + \cdots + q_K + \tau q_k, \tag{7}$$

where i ranges over $\{1, \ldots, N_0 + n\}$, $\tau \in [0, 1]$ is a random number, and k is defined by the condition $l_k = L_{N_0+n}$. Since n is fixed and $N_0 \to \infty$, we can assume that i ranges over $\{1, \ldots, N_0\}$ and replace $N_0 + n$ by N_0 in the denominator of (7). The approximate equality "$\approx$" in (7) then follows from the law of large numbers (with a large probability). The asymptotically optimal betting function takes value l_k at this p-value by its definition (see (6)), as it should (if we regards the LRM as being optimal).

An *asymptotically optimal CTM* is a CTM based on an asymptotically optimal betting function f and on likelihood ratio L as nonconformity measure. It is the *CAO CTM* if f is the CAO betting function. The following proposition says that the conformal CUSUM procedure based on the CAO CTM, or any other asymptotically optimal CTM, is perfectly valid.

Proposition 1. *For any pair (Q_0, Q_1) of pre-/post-change distributions, the distribution of the likelihood ratio martingale under Q_0^∞ coincides with the distribution of any asymptotically optimal CTM under randomness (not necessarily under Q_0^∞).*

Proof. Let S be the LRM and $\bar{S}$ be an asymptotically optimal CTM based on (Q_0, Q_1). The relative increments S_n/S_{n-1}, $n = 1, 2, \ldots$, are independent because the observations are independent, and the relative increments $\bar{S}_n/\bar{S}_{n-1}$, $n = 1, 2, \ldots$, are independent because the conformal p-values are independent [15, Theorem 11.1]. Therefore, it suffices to prove that, for a fixed n, S_n/S_{n-1} and $\bar{S}_n/\bar{S}_{n-1}$ have the same distribution (the former under Q_0^∞ and the latter under randomness). The former's distribution is the pushforward of Q_0 by the likelihood ratio mapping (4). The latter's distribution is identical by definition and [2, Lemma A.23]. $\qquad\square$

4

According to Proposition 1, the conformal CUSUM procedure satisfies the same properties of validity as the standard CUSUM procedure but under radically weaker assumptions: instead of assuming Q^∞, we just assume randomness. In particular, we can use the whole arsenal of the existing results in the standard theory of change detection to find a suitable threshold c for the CUSUM procedure. (See the mini-review at the end of Sect. 2.) Alternatively, we can use computer simulations, as in [15, Sect. 8.4.3].

Proposition 1 demonstrates a significant advantage of the conformal CUSUM procedure over the conformal CUSUM e-procedure; it is unlikely that anything similar to Proposition 1 can be proved for the conformal CUSUM e-procedure, since we know the distribution of the conformal p-values (it is uniform in $[0, 1]^\infty$) but not of conformal e-values.

4 Computing the CAO betting function in popular special cases

Now we take the observation space to be the real line $\mathbb{R}$, $\mathbf{Z} = \mathbb{R}$, and take m to be Lebesgue measure; we still assume that the pre-change and post-change distributions are absolutely continuous w.r. to m with positive probability densities. As before, f_0 is the probability density of the pre-change distribution Q_0 and f_1 is the probability density for the post-change distribution Q_1. The likelihood ratio is defined by (4), where now $z \in \mathbb{R}$.

The special cases for which we will compute the CAO betting function in this section will all involve the Gaussian statistical model. First let us assume that the pre-change distribution is $Q_0 := N(0, 1)$ and the post-change distribution is $Q_1 := N(\mu, 1)$. For concreteness, let us assume $\mu > 0$ (this does not really restrict generality because of symmetry). The likelihood ratio is

$$L(z) = \exp(\mu z - \mu^2/2).$$

Solving the inequality $L(z) \geq l$ for $l \in (0, \infty)$, we obtain the optimal Neyman–Pearson region

$$z \geq \frac{\ln l}{\mu} + \frac{\mu}{2}. \tag{8}$$

Its probability under Q_0 (i.e., the *type I error probability* of the corresponding statistical test) is

$$p = \mathbb{P}\left(Z \geq \frac{\ln l}{\mu} + \frac{\mu}{2}\right) = 1 - \Phi\left(\frac{\ln l}{\mu} + \frac{\mu}{2}\right),$$

where $Z \sim N(0, 1)$ and Φ stands for the distribution function of $N(0, 1)$. Expressing l in terms of p, we obtain the betting function

$$f(p) = l = \exp\left(\mu \Phi^{-1}(1 - p) - \mu^2/2\right), \tag{9}$$

5

Φ^{-1} being the inverse function of Φ (i.e., Φ^{-1} is the quantile function of $N(0, 1)$). If $\mu < 0$, we should replace the "$\geq$" in (8) by "$\leq$", which leads to

$$f(p) = l = \exp\left(\mu\Phi^{-1}(p) - \mu^2/2\right) \tag{10}$$

in place of (9). (We can also see that (9) implies (10) by the symmetry between p and $1 - p$.)

Next we consider the case where the pre-change distribution is $N(0, 1)$ and the post-change distribution is $N(0, \sigma^2)$. The likelihood ratio is

$$L(z) = \frac{1}{\sigma} \exp\left((1 - 1/\sigma^2)z^2/2\right).$$

Let us first assume $\sigma > 1$ (the observations are becoming more spread out after the changepoint). Then $L(z)$ ranges over $[1/\sigma, \infty)$. Solving $L(z) \geq l$ for $l \geq 1/\sigma$ gives the optimal Neyman–Pearson region

$$\left(-\infty, -\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}\right] \cup \left[\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}, \infty\right). \tag{11}$$

The type I error probability is

$$p = \mathbb{P}\left(Z \in \left(-\infty, -\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}\right] \cup \left[\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}, \infty\right)\right)$$

$$= 2N\left(-\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}\right). \tag{12}$$

For the change $N(0, 1) \to N(0, \sigma^2)$ for $\sigma > 1$, (12) gives

$$f(p) = l = \frac{1}{\sigma} \exp\left(\frac{1 - 1/\sigma^2}{2}\left(\Phi^{-1}(p/2)\right)^2\right) \in [1/\sigma, \infty). \tag{13}$$

Similarly, we get

$$f(p) = l = \frac{1}{\sigma} \exp\left(\frac{1 - 1/\sigma^2}{2}\left(\Phi^{-1}((1 - p)/2)\right)^2\right) \in (0, 1/\sigma] \tag{14}$$

for $\sigma < 1$; this is obtained by replacing the optimal Neyman-Pearson region (11) by its complement.

There is a convenient equivalent definition of the CAO betting function that sometimes leads to nicer and more intuitive mathematical expressions. For each threshold $l \in (0, \infty)$, find, alongside the type I error probability

$$p_0(l) := \int_{z:L(z) \geq l} f_0(z)\mathrm{d}z \in [0, 1], \tag{15}$$

the probability

$$p_1(l) := \int_{z:L(z) \geq l} f_1(z)\mathrm{d}z \in [0, 1] \tag{16}$$

6

(this is the *power*, or 1 minus the *type II error probability*). Let us assume, for simplicity, that the functions p_0 and p_1 defined by (15) and (16) are continuous and strictly increasing over some interval. (These conditions are satisfied in the examples considered in this section.) The *Neyman–Pearson ROC curve R* for the pair (Q_0, Q_1) (with Q_0 interpreted as null hypothesis and Q_1 as alternative) is defined to be p_1 as function of p_0:

$$R(p) := \sup\{p_1(l) : p_0(l) = p\}.$$

This definition is motivated, of course, by the Neyman–Pearson lemma [6, Sect. 3.2].

Proposition 2. *The Neyman–Pearson ROC curve R is differentiable, and the CAO betting function f is equal to its derivative, $f = R'$.*

Another way to express Proposition 2 is: the CAO betting function f is equal to the minus derivative of the type II error probability $1 - p_1$ expressed as function of the type I error probability p_0.

Proof of Proposition 2. Under our assumptions, as $\Delta l \to 0$,

$$\frac{\mathrm{d}p_1}{\mathrm{d}p_0}(l) \sim \frac{p_1(l + \Delta l) - p_1(l)}{p_0(l + \Delta l) - p_0(l)}$$

$$= \int_{z:L(z)\in[l,l+\Delta l]} L(z)f_0(z)\,\mathrm{d}z \bigg/ \int_{z:L(z)\in[l,l+\Delta l]} f_0(z)\,\mathrm{d}z \sim l. \qquad \square$$

Proposition 2 gives more intuitive expressions for the CAO betting functions discussed so far. Start, again, from the change $N(0,1) \to N(\mu,1)$ for $\mu > 0$. The probability of (8) under Q_1 (i.e., the power) is

$$p^* = \mathbb{P}\left(Z + \mu \geq \frac{\ln l}{\mu} + \frac{\mu}{2}\right) = 1 - \Phi\left(\frac{\ln l}{\mu} - \frac{\mu}{2}\right).$$

(So that now we are writing p for p_0 and p^* for p_1.) Expressing p^* via p, we obtain the Neyman–Pearson ROC curve

$$R(p) = p^* = 1 - \Phi(\Phi^{-1}(1 - p) - \mu).$$

Differentiating this function gives the optimal betting function

$$f(p) = R'(p) = \frac{\phi(\Phi^{-1}(1 - p) - \mu)}{\phi(\Phi^{-1}(1 - p))}, \tag{17}$$

where ϕ is the probability density of $N(0,1)$. It is easy to check that the expression (17) is equivalent to our previous expression (9). If $\mu < 0$, replacing the "$\geq$" in (8) by "$\leq$" gives

$$f(p) = R'(p) = \frac{\phi(\Phi^{-1}(p) - \mu)}{\phi(\Phi^{-1}(p))} \tag{18}$$

in place of (17); the last expression is again equivalent to (10).

We can interpret the betting functions (17) and (18) as likelihood ratios. For example, (17) is the likelihood ratio of the alternative hypothesis Q_1 to the null hypothesis Q_0 after transforming the observed p-value p to the Gaussian scale $\Phi^{-1}(1-p)$.

In the case $N(0,1) \to N(0,\sigma^2)$ with $\sigma > 1$, the power is

$$p^* = \mathbb{P}\left(\sigma Z \in \left(-\infty, -\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}\right] \cup \left[\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}, \infty\right)\right)$$

$$= 2N\left(-\frac{1}{\sigma}\sqrt{\frac{2\ln(l\sigma)}{1 - 1/\sigma^2}}\right).$$

Therefore, the Neyman–Pearson ROC curve is

$$R(p) = p^* = 2N\left(\frac{1}{\sigma}\Phi^{-1}(p/2)\right),$$

and differentiation gives the betting function

$$f(p) = R'(p) = \frac{\phi\left(\frac{1}{\sigma}\Phi^{-1}(p/2)\right)}{\sigma\phi(\Phi^{-1}(p/2))}. \tag{19}$$

The formula for the betting function for the case $\sigma < 1$ is

$$f(p) = R'(p) = \frac{\phi\left(\frac{1}{\sigma}\Phi^{-1}((1-p)/2)\right)}{\sigma\phi(\Phi^{-1}((1-p)/2))}; \tag{20}$$

as usual, it is obtained by replacing (11) by its complement. Of course, (19) and (20) are equivalent to (13) and (14), but the former look more natural.

Remark 3. The conformal test martingales based on the CAO betting functions developed in this section can be integrated over a probability measure on the parameter (the post-change mean μ or the post-change standard deviation σ). This allows us to achieve good performance for wider ranges of post-change distributions.

Remark 4. We can't easily combine the two Gaussian cases that we have considered, $N(0,1) \to N(\mu,1)$ and $N(0,1) \to N(0,\sigma^2)$, into the case of the pre-change distribution being $N(0,1)$ and the post-change distribution being $N(\mu,\sigma^2)$. The general case of a change in both location and scale parameters does not appear to admit a simple closed-form solution.

5 Efficiency and validity: Experimental results

In this section we will report some experimental results for conformal CUSUM procedures and, for comparison, conformal CUSUM e-procedures considered in
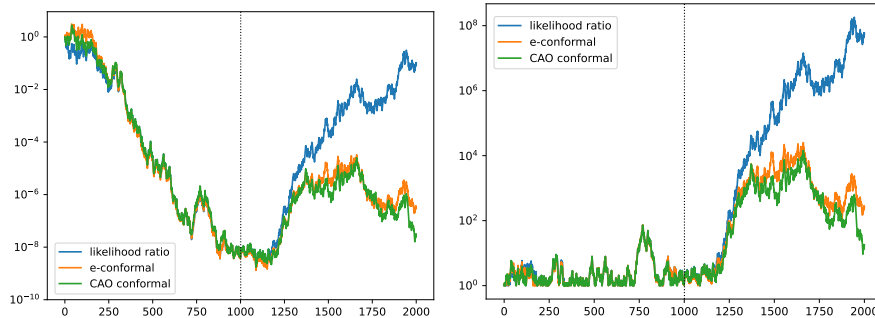
Figure 1: Left panel: the likelihood ratio martingale, conformal e-pseudomartingale, and conformal test martingale, as described in text, in the Bernoulli case (with the parameter equal to 0.5 pre-change and 0.6 post-change). Right panel: the corresponding CUSUM statistics.

[16]. The latter are based on universal e-values

$$E_n := \frac{L_n}{\frac{1}{n}(L_1 + \cdots + L_n)},\qquad(21)$$

where $L_n := L(z_n)$, and no further tuning to Q_0 and Q_1 (like choosing the betting functions in the case of conformal testing) is needed. We omit the details referring the reader to [16].

Experimental results for Bernoulli observations are given in Fig. 1. The first $N_0 := 1000$ observations $z_1, \ldots, z_{N_0}$ are generated from the Bernoulli distribution with parameter 0.5 and another $N_1 := 1000$ observations $z_{N_0+1}, \ldots, z_{N_0+N_1}$ from the Bernoulli distribution with parameter 0.6. The changepoint 1000 is shown as a dashed vertical line. The left panel of Fig. 1 shows three paths:

- the likelihood ratio martingale $L_1 \ldots L_n$, $n = 0, 1, \ldots, N_0 + N_1$, in blue;

- the conformal e-pseudomartingale $E_1 \ldots E_n$, $n = 0, 1, \ldots, N_0 + N_1$, in orange, with the conformal e-values defined as the normalized L_n: see (21);

- the CAO conformal test martingale in green; as the green line was drawn after the orange line and the two lines are very close, the green line often obscures the orange one.

The right panel shows the corresponding *CUSUM statistics* S_n^* of these processes defined as

$$S_n^* := S_n / \min(S_0, \ldots, S_{n-1}),$$

S_n being the original process. The CAO betting function is, according to the definition above,

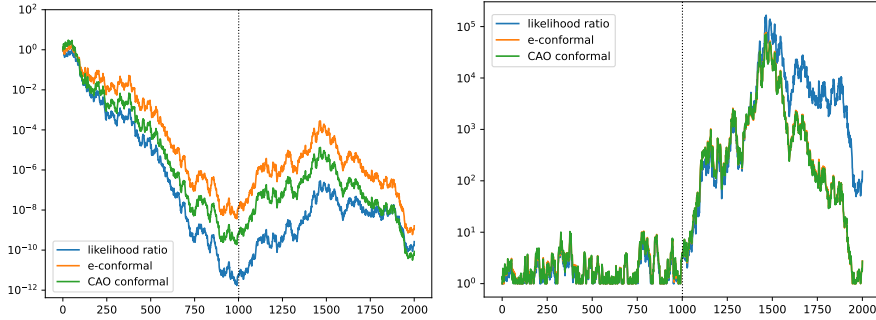$$f(p) := \begin{cases} 1.2 & \text{if } p \le 0.5 \\ 0.8 & \text{otherwise.} \end{cases}$$

9

Figure 2: Left panel: the three processes as in Fig. 1, but with a change from $N(0, 1)$ to $N(0.2, 1)$. Right panel: the corresponding CUSUM statistics.
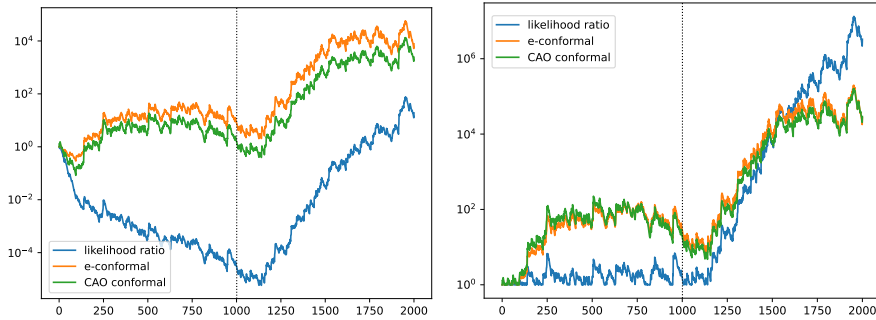

Figure 3: Left panel: the three processes as in Fig. 1, but with a change from $N(0, 1)$ to $N(0, 1.1)$. Right panel: the corresponding CUSUM statistics.

In Fig. 1 we can clearly see a phenomenon that we called *decay* in [15, Sect. 8.4]: when testing the assumption of randomness (the orange and green lines), the new distribution Q_1 will gradually become the "new normal" after the changepoint, and the growth of the CTM will slow down or stop. Of course, there is no decay for the blue line. Decay is inevitable when testing randomness, and by itself it does not indicate any inefficiency.

Figure 2 is the analogue of Fig. 1 for a change in mean for a Gaussian distribution. Now the first $N_0 := 1000$ observations are generated from $N(0, 1)$ and the other $N_1 := 1000$ from $N(0.2, 1)$. The orange and green lines are remarkably close to each other in the right panel (and this remains true even if we vary the seed of the random number generator; for other seeds, of course, the blue line does not show any signs of decay, which is an artefact of our default choice of "42" as seed).

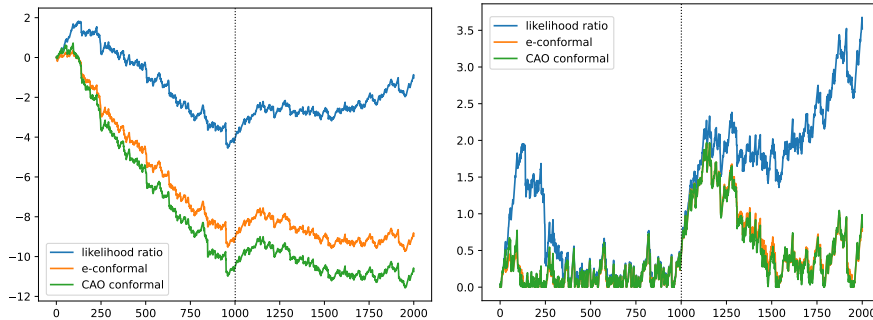Figure 3 is the analogue of Figures 1 and 2 for a change from $N(0, 1)$ to

10

Figure 4: Left panel: the three processes as in Fig. 1, but with a change from $N(0, 1)$ to $N(0, 0.9)$. Right panel: the corresponding CUSUM statistics.

$N(0, 1.1)$. The orange and green lines are still fairly close in the right panel. And finally, Fig. 4 is the analogue for a change from $N(0, 1)$ to $N(0, 0.9)$. We use the expressions (9), (13), and (14) for the CAO betting functions (but of course using (17), (19), and (20) would have led to the identical results).

We will finish this section by exploring the validity of the conformal CUSUM procedure. It might appear that there is a contradiction between Proposition 1 and the left panels of Figures 3 and 4: those panels seem to suggest that the stochastic behaviour of the LRM and the CAO CTM is very different before the changepoint. In the rest of this section we will see that, despite the frequent remarkably different behaviour of the LRM and the CAO CTM before the changepoint on the same dataset, their distributions are very close to each other (and we know from Proposition 1 that they are even identical), and both are typically close to the distribution of the conformal e-pseudomartingale (for which we do not have a theoretical explanation at this time).

Figure 5 demonstrates the closeness in two cases shown in its two panels. We perform 10,000 simulations of 1000 observations from a pre-change distribution computing the final values of three processes. The boxplots in the figure show the median (as the orange middle line with notches showing its default Matplotlib confidence limits), the first and third quartiles (as the lower and upper sides of the boxes), and the 5% and 95% quantiles (as the whiskers) of the decimal logarithms of the three processes. The left panel is for the Bernoulli case with the same parameters as before (0.5 pre-change and 0.6 post-change), and the right panel is for the Gaussian case with a change in mean with the same parameters ($N(0, 1)$ pre-change and $N(0.2, 1)$ post-change). There is no changepoint, and we only generate $N_0 := 1000$ observations from the pre-change distributions. The boxplots in each panel are given in this order: the LRM, the CAO CTM, and the conformal e-pseudomartingale (CeP). All three boxplots are very close in each of the two panels.

Figure 6 is the analogue of Fig. 5 for the upward and downward change in the variance of the standard Gaussian distribution generating the data, again
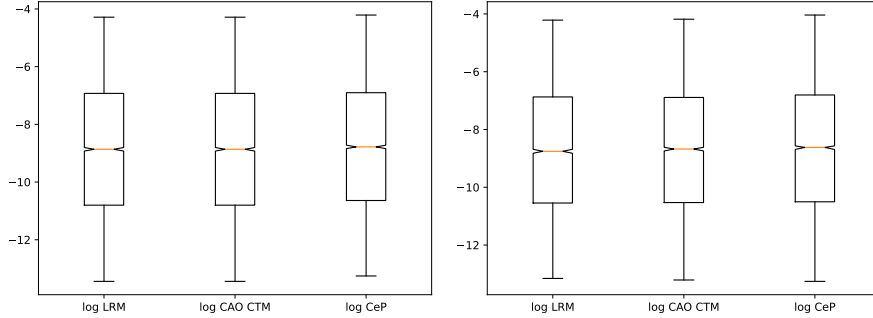
Figure 5: Left panel: the validity for the Bernoulli case. Right panel: the validity for the Gaussian case with a change in mean. As described in text, with 10,000 simulations.

with the same parameters as before. While the first two boxplots are still very close to each other (as they must be by Proposition 1), the boxplot for conformal e-pseudomartingale looks visibly different. Figure 7 gives a more extreme example of broken validity. In it, the pre-change distribution is Bernoulli with parameter 0.1 while the post-change distribution is Bernoulli with parameter 0.9; the number of observations is $N_0 := 100$.

6   Efficiency: Theoretical analysis

In this section we report a simple preliminary theoretical result about the efficiency of conformal testing in the context of change detection. We only consider the Bernoulli case.

Theorem 5. *Consider the Bernoulli case with the pre-change parameter θ_0 and post-change parameter θ_1. Let B be given by*

$$B := \left| \ln \frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right|$$

(this is a kind of distance between θ_0 and θ_1). Let p_n be the nth conformal p-value based on the likelihood ratio nonconformity measure and f be the CAO betting function. Then, for any $\epsilon > 0$, the log ratio of the likelihood ratio martingale to the CAO CTM after the changepoint is bounded above as

$$\forall n \in \{1, \dots, N_1\} : \ln \frac{L_{N_0+1} \dots L_{N_0+n}}{f(p_{N_0+1}) \dots f(p_{N_0+n})}$$
$$\leq B \left(\ln \frac{2}{\epsilon} + 5N_1 \sqrt{\frac{\ln \frac{4}{\epsilon}}{2N_0}} + \frac{5}{2} \frac{N_1(N_1 + 1)}{N_0} \right) \quad (22)$$
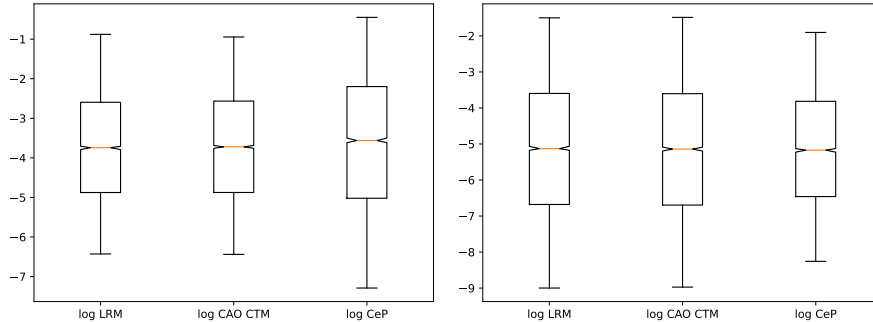
12

Figure 6: Left panel: the validity for the Gaussian case with an upward change in variance. Right panel: the validity for the Gaussian case with a downward change in variance. The number of simulations is 10,000 in both panels; see the description in text.

with probability at least $1 - \epsilon$.

Assuming that $\ln \frac{1}{\epsilon}$ is only moderately large, the right-hand side of (22) will be moderately large if N_1 has the same order of magnitude as $\sqrt{N_0}$ (this is similar to our conclusion about [16, Proposition 8]).

For a proof of Theorem 5 see Appendix A.

7   Conclusion

In this paper we started implementing the Burnaev–Wasserman programme in the case of change detection. We showed that the conformal CUSUM procedure has perfect validity and investigated its efficiency both experimentally and, in a very basic way, theoretically.

Our secondary goal was to compare conformal testing and conformal e-testing in the context of change detection. These are some advantages of conformal testing:

- The conformal CUSUM procedure is perfectly valid, as we saw in Proposition 1. This allows us to find a suitable value of the parameter c for a given target value of the frequency of false alarms using computer simulations or existing theoretical results. Proposition 1 also demonstrates the perfect validity of the conformal version of the Shiryaev–Roberts procedure, which is an important competitor of CUSUM. Validity can be violated for the conformal CUSUM e-procedure, as we saw in Sect. 5 (Fig. 7).

- Conformal testing produces conformal test martingales, which are genuine martingales, and merging genuine martingales is easy: we can just average them. This does not work for conformal e-pseudomartingales, while it is
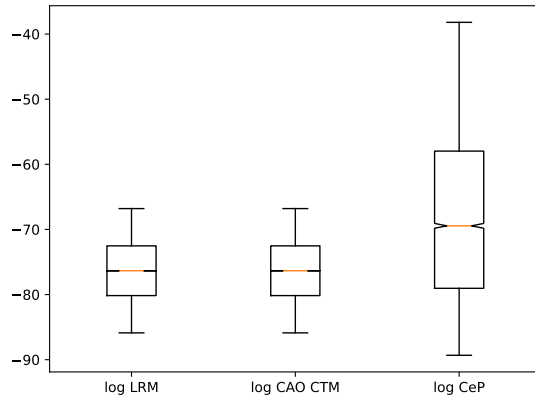
13

Figure 7: Validity and lack thereof for a Bernoulli case, as described in text (with 10,000 simulations).

important in change detection (e.g., when we do not have a single post-change distribution even as our soft model).

This short list complements the analogous list in [16, Sect. 7]. One item on that list is the need for randomization in conformal testing, which may be regarded as its disadvantage. Randomization is really harmful if it significantly affects the outcome of testing; it becomes much more tolerable if (as is often the case) it is performed repeatedly and the law of large numbers becomes applicable.

These are some open problems and directions of further research along these lines:

- We saw in Sect. 5 that Proposition 1 does not hold for conformal e-testing (Fig. 7) but in typical cases (Figures 5–6) the distributions of the final values for likelihood ratio martingales and conformal e-pseudomartingales still look similar before the changepoint. Can we prove partial validity results for conformal e-testing?

- The experimental results in Sect. 5 and Theorem 5 show the closeness of the CUSUM statistics for the LRM and CAO CTM, but Theorem 5 covers only the Bernoulli case. It would be nice to have more general and stronger efficiency results.

- Similarly, [16] demonstrated the closeness of the CUSUM statistics for the LRM and the conformal e-pseudomartingale (21). Those efficiency results can also be generalized and strengthened.

- In Sect. 5 we observed extreme closeness of the CUSUM statistics for conformal e-pseudomartingales and conformal test martingales. Is this a manifestation of a general efficiency result?

References

[1] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes: Theory and Application.* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[2] Hans Föllmer and Alexander Schied. *Stochastic Finance: An Introduction in Discrete Time.* De Gruyter, Berlin, fourth edition, 2016.

[3] Amrit L. Goel and S. M. Wu. Determination of A. R. L. and a contour nomogram for Cusum charts to control normal mean. *Technometrics*, 13:221–230, 1971.

[4] Michel Goemans. Chernoff bounds, and some applications: Lecture notes. Principles of Discrete Applied Mathematics, Massachusetts Institute of Technology, February 2015.

[5] William Kuszmaul and Qi Qi. The multiplicative version of Azuma's inequality, with an application to contention analysis. Technical Report arXiv:2102.05077 [cs.DS], arXiv.org e-Print archive, February 2021.

[6] Erich L. Lehmann and Joseph P. Romano. *Testing Statistical Hypotheses.* Springer, Cham, fourth edition, 2022.

[7] Gary Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42:1897–1908, 1971.

[8] George V. Moustakides. Optimal stopping times for detecting a change in distribution. *Annals of Statistics*, 14:1379–1388, 1986.

[9] Ewan S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[10] H. Vincent Poor and Olympia Hadjiliadis. *Quickest Detection.* Cambridge University Press, Cambridge, 2009.

[11] Ya'acov Ritov. Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics*, 18:1464–1469, 1990.

[12] Glenn Shafer and Vladimir Vovk. *Game-Theoretic Foundations for Probability and Finance.* Wiley, Hoboken, NJ, 2019.

[13] Alexander Tartakovsky, Igor Nikiforov, and Michèle Basseville. *Sequential Analysis: Hypothesis Testing and Changepoint Detection.* CRC Press, Boca Raton, FL, 2015.

[14] Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov, Alex Gammerman, and Vladimir Vovk. Inductive conformal martingales for change-point detection. *Proceedings of Machine Learning Research*, 60:132–153, 2017. COPA 2017.

[15] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World.* Springer, Cham, second edition, 2022.

[16] Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman. Conformal e-testing. Technical Report arXiv:2006.02329v2 [math.ST], arXiv.org e-Print archive, November 2024.

A   Proof of Theorem 5

To compare conformal test martingales with likelihood ratio martingales, we will use the martingale multiplicative version of the Chernoff bound to be proved in Appendix B (Theorem 6).

The betting function in Theorem 5 is

$$f(p) := \begin{cases} \frac{\theta_1}{\theta_0} & \text{if } p \le \theta_0 \\ \frac{1-\theta_1}{1-\theta_0} & \text{otherwise.} \end{cases}$$

We split ϵ into two equal parts of $\epsilon/2$. By Hoeffding's inequality (see, e.g., [15, Sect. A.6.3]),

$$\mathbb{P}\left(\left|\frac{K_0}{N_0} - \theta_0\right| \ge \delta\right) \le 2\exp(-2\delta^2 N_0),$$

where K_0 is the number of 1s in the first N_0 steps. Solving the equation

$$2\exp(-2\delta^2 N_0) = \frac{\epsilon}{2}$$

gives

$$\delta = \sqrt{\frac{\ln\frac{4}{\epsilon}}{2N_0}}. \tag{23}$$

Let us fix the first N_0 observations such that

$$\left|\frac{K_0}{N_0} - \theta_0\right| < \delta.$$

We say that step n after the changepoint is *anomalous* if one of the events $p_n \le \theta_0$ and $z_n = 1$ happens while the other does not happen. Only if step n after the changepoint is anomalous can we have $L_{N_0+n} \ne f(p_{N_0+n})$, and even if that step is anomalous we still have

$$\ln L_{N_0+n} - \ln f(p_{N_0+n}) \le B.$$

Therefore, our proof strategy will consist in bounding the number of anomalous steps.

The expectation of the number of anomalous steps up to (and including) step N_1 after the changepoint is at most

$$\sum_{n=1}^{N_1}\left(\delta + \frac{n}{N_0}\right) = N_1\delta + \frac{N_1(N_1+1)}{2N_0}.$$

By the martingale multiplicative Chernoff bound (Theorem 6 below),

$$\mathbb{P}\left(A \geq (1+\Delta)\left(N_1\delta + \frac{N_1(N_1+1)}{2N_0}\right)\right)$$

$$\leq \exp\left(-\frac{\Delta^2}{2+\Delta}\left(N_1\delta + \frac{N_1(N_1+1)}{2N_0}\right)\right), \quad (24)$$

where A is the number of anomalous post-change steps and Δ is a positive constant. Solving the equation

$$\exp\left(-\frac{\Delta^2}{2+\Delta}\left(N_1\delta + \frac{N_1(N_1+1)}{2N_0}\right)\right) = \frac{\epsilon}{2}$$

in Δ is equivalent to solving

$$\frac{\Delta^2}{2+\Delta} = c, \quad (25)$$

where

$$c := \frac{\ln\frac{2}{\epsilon}}{N_1\delta + \frac{N_1(N_1+1)}{2N_0}}. \quad (26)$$

Solving the quadratic equation (25) gives

$$\Delta = \frac{c + \sqrt{c^2 + 8c}}{2} \leq \sqrt{c^2 + 8c} \leq c + 4.$$

We set, conservatively, $\Delta := c + 4$. According to (24), we have

$$A < (1+\Delta)\left(N_1\delta + \frac{N_1(N_1+1)}{2N_0}\right) = (c+5)\left(N_1\delta + \frac{N_1(N_1+1)}{2N_0}\right)$$

with probability at least $1 - \frac{\epsilon}{2}$. It remains to plug (26) and then (23) into the last inequality.

B   Martingale multiplicative Chernoff bound

This result was stated and proved in [5, Corollary 6]; we have used it in Appendix A.

Theorem 6. *Fix a filtration $(\mathcal{F}_n)$. Suppose (ξ_n) is a (super)martingale difference such that ξ_n take values in $\{0,1\}$, and set $\theta_n := \mathbb{P}(\xi_n = 1 \mid \mathcal{F}_{n-1})$. Then, for any constant $s > 0$,*

$$S_n := \prod_{i=1}^{n} \exp\left(s\xi_i - \theta_i(e^s - 1)\right) \quad (27)$$

is a test supermartingale. For any $\delta > 0$, any N, and any constant μ such that $\sum_{i=1}^{N} \theta_i \leq \mu$ a.s.,

$$\mathbb{P}\left(\sum_{i=1}^{N} \xi_i \geq (1+\delta)\mu\right) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right). \quad (28)$$

Theorem 6 can be easily stated in terms of game-theoretic probability [12], but in this paper we prefer the more familiar measure-theoretic statement. Notice that the left-hand side of (28) can be replaced by

$$\mathbb{P}\left(\exists n \in \{1,\ldots,N\} : \sum_{i=1}^{n} \xi_i \geq (1+\delta)\mu\right)$$

as $\xi_i \geq 0$. For completeness, we will give a simple proof of Theorem 6.

Proof of Theorem 6. The proof is standard (see, e.g., [4] for a nice exposition). To check that (S_n) is a supermartingale it suffices to notice that (assuming (ξ_n) is a martingale)

$$\mathbb{E}\left(\exp\left(s\xi_n - \theta_n(e^s - 1)\right) \mid \mathcal{F}_{n-1}\right) = (\theta_n e^s + 1 - \theta_n)\exp\left(-\theta_n(e^s - 1)\right)$$
$$\leq \exp\left(\theta_n(e^s - 1)\right)\exp\left(-\theta_n(e^s - 1)\right) = 1,$$

where the inequality "$\leq$" follows from $1 + x \leq e^x$.

This is a derivation of (28):

$$\mathbb{P}\left(\sum_{i=1}^{N}\xi_i \geq (1+\delta)\mu\right) = \mathbb{P}\left(\exp\left(s\sum_{i=1}^{N}\xi_i\right) \geq \exp\left(s(1+\delta)\mu\right)\right)$$

$$\leq \mathbb{P}\left(\exp\left(s\sum_{i=1}^{N}\xi_i - (e^s - 1)\sum_{i=1}^{N}\theta_i\right) \geq \exp\left(s(1+\delta)\mu - (e^s - 1)\mu\right)\right) \quad (29)$$

$$\leq \exp\left(-s(1+\delta)\mu + (e^s - 1)\mu\right) = \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \leq \exp\left(-\frac{\delta^2}{2+\delta}\mu\right), \quad (30)$$

where $s := \ln(1+\delta) > 0$. The inequality "$\leq$" in (29) follows from the condition $\sum_{i=1}^{N}\theta_i \leq \mu$ a.s. The first inequality "$\leq$" in (30) follows from (27) being a test supermartingale. The equality in (30) follows from $s := \ln(1+\delta)$ (which we obtain by minimizing the expression to the left of the "=" in s). The second inequality "$\leq$" in (30) follows from

$$\ln(1+\delta) \geq \frac{\delta}{1 + \frac{\delta}{2}}, \quad \delta > 0, \quad (31)$$

applied to the logarithm of the expression on its left-hand side ((31) can be proved by, e.g., differentiating both sides). □

18