Conformal e-testing

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

On-line Compression Modelling Project (New Series)

Working Paper #29

First posted June 3, 2020. Last revised November 2, 2024.

Project web site:
http://alrw.net

Abstract

There is a useful counterpart of conformal prediction for e-values, called *conformal e-prediction*. Conformal prediction can serve as basis for testing the assumption of exchangeability, leading to *conformal testing*. Similarly, conformal e-prediction can also serve as basis for testing. The resulting *conformal e-testing* looks very different from but inherits some strengths of conformal testing; it even has some advantages over conformal testing. In this paper we discuss systematically both strengths and limitations of conformal e-testing.

Contents

1 Introduction

A useful application of conformal prediction is conformal testing, which is a technique for testing the assumption of exchangeability (or another online compression model). Conformal e-prediction is a modification of conformal prediction obtained by replacing the notion of a p-value by that of an e-value; it is reviewed in the sister article [15]. (It may be natural to refer to conformal testing as *conformal p-testing*, but we will never use this term.)

An important advantage of conformal prediction over conformal e-prediction is that its strong property of validity found in [13, Theorem 1] allows us to test the assumption of exchangeability. This strong property can be stated as the independence of the smoothed conformal p-values output at different steps, and to test exchangeability, we can bet against the conformal p-values being independent and uniformly distributed. This led to the introduction in 2003 [17] of conformal test martingales. The validity of conformal test martingales as means of testing exchangeability shows, e.g., in Ville's theorem [12, p. 100]: under exchangeability, the probability that a given conformal test martingale S ever exceeds a fixed threshold $c > 1$ is at most $1/c$. For example, we might feel justified in rejecting the hypothesis of exchangeability when S exceeds 100, since the probability of this event is at most 1%.

In general and informally, validity is the requirement that our testing methods should give false evidence against exchangeability (in the context of this paper) only with low probability. It will appear in various guises in this paper. Under the restriction of validity, we would also like our procedures to be efficient at discovering evidence against exchangeability. For a long time, nothing was known about the efficiency of conformal test martingales, and first results about their efficiency appeared in 2019 (see [14]); this is a major topic of [16, Part III].

In this paper we discuss conformal e-testing systematically and compare it with conformal testing. It turns out that, similarly to the case of conformal e-prediction [15], conformal e-testing can often emulate strengths of conformal testing. Moreover, conformal e-testing has some advantages of its own over conformal testing.

We start the main part of the paper in Sect. 2 by defining conformal e-prediction in a way adapted to the use in conformal e-testing. In the following Sect. 3 we introduce conformal e-testing and explore its validity. Instead of conformal test martingales, we obtain what we call "conformal e-pseudomartingales", and our main finding here is negative: conformal e-pseudomartingales can violate badly the property of validity expressed by Ville's inequality. While in Sect. 3 we use the online setting, which is standard in this area, in Sect. 4 we use a more limited batch setting, which allows us to establish results about both validity and efficiency of conformal e-testing.

In Sections 5 and 6 we discuss "multistage" ways of testing the exchangeability assumption, concentrating on the standard CUSUM procedure used on top of conformal e-prediction, which we call the "conformal CUSUM e-procedure". Section 5 is devoted to the validity of the conformal CUSUM e-procedure: we show that under exchangeability it raises false alarms with a frequency determined by its parameter (and it can be made as low as we want). This property is deduced from the analogous property of what we call the reverse Shiryaev–Roberts procedure. In this section we also give an example showing an undesirable property of the Shiryaev–Roberts procedure, standard or reverse. In Sect. 6 we discuss a property of efficiency of the conformal CUSUM e-procedure. While Sect. 5 shows that conformal e-testing inherits some strengths of conformal testing, Sect. 6 demonstrates an advantage of conformal e-testing.

Section 7 concludes and lists some directions of further research.

2 Conformal e-prediction

The task of conformal e-prediction, and predictive machine learning in general, is to predict the label of a test object x given a training set $z_1, \dots, z_n$ whose elements are labelled objects $z_i = (x_i, y_i)$. What distinguishes conformal (e-) prediction is that for each potential label y for the test object x it provides a non-negative number $f(z_1, \dots, z_n, (x, y))$ (we usually drop the internal parentheses) reflecting the plausibility of y being the true label of x.

The objects x_i are drawn from the *object space* $\mathbf{X}$ and the labels y_i from the *label space* $\mathbf{Y}$; both are required to be non-empty measurable spaces. The *observations* $z = (x, y)$ are drawn from the Cartesian product (the *observation space*) $\mathbf{Z} := \mathbf{X} \times \mathbf{Y}$. In this paper we will also be interested in the case where $\mathbf{Z}$ is unstructured (not a Cartesian product, which can be embedded into the structured case by setting $\mathbf{X}$ or $\mathbf{Y}$ to a one-element space).

We will use the notation $X^+ := \cup_{n=1}^{\infty} X^n$ for the set of all non-empty finite sequences of elements of X. If X is a measurable space, X^+ is also a measurable space.

A *nonconformity e-measure* is a measurable function $A : \mathbf{Z}^+ \to [0, \infty)^+$ that maps every finite sequence $(z_1, \dots, z_m)$, $m \in \{1, 2, \dots\}$, to a finite sequence $(\alpha_1, \dots, \alpha_m)$ of the same length such that

$$\frac{1}{m} \sum_{i=1}^{m} \alpha_i \le 1 \tag{1}$$

and that satisfies the property of *equivariance*: for any m and any permutation π of $\{1, \dots, m\}$,

$$(\alpha_1, \dots, \alpha_m) = A(z_1, \dots, z_m) \Longrightarrow (\alpha_{\pi(1)}, \dots, \alpha_{\pi(m)}) = A(z_{\pi(1)}, \dots, z_{\pi(m)}).$$

(We sometimes refer to the α_i as *nonconformity e-scores*.) The corresponding *conformal e-predictor* $f : \mathbf{Z}^+ \to [0, \infty)$ is defined as

$$f(z_1, \dots, z_n, x, y) := \alpha_{n+1}, \text{ where } (\alpha_1, \dots, \alpha_n, \alpha_{n+1}) := A(z_1, \dots, z_n, (x, y)).$$

For a training set $z_1, \dots, z_n$ and a test object x, the full prediction for x according to a conformal e-predictor f is given by the family of *potential conformal e-values*

$$(f(z_1, \dots, z_n, x, y) \mid y \in \mathbf{Y}).$$

We can make a confident prediction for x if the potential conformal e-values are large for all $y \in \mathbf{Y}$ except for one.

A nonconformity e-measure and the corresponding conformal e-predictor are *admissible* if we always have "=" in place of "≤" in (1) in its definition. Testing procedures based on nonconformity e-measures that are not admissible can be improved.

Now let us state what we regard as the main property of validity of conformal e-prediction (for more information, see [15, Sect. 3]). Let $Z_1, Z_2, \dots$ be the random observations, i.e., the random elements whose realizations are the observed $z_1, z_2, \dots$. In general, (X, Y) or Z are random observations, i.e., random elements taking values in the observation space $\mathbf{Z}$. Remember that a finite sequence of random elements is *exchangeable* if its joint distribution does not change if it is permuted (and an infinite sequence is exchangeable if its every finite beginning is exchangeable). For example, any IID sequence (i.e., a sequence of independent and identically distributed random elements) is exchangeable.

With each conformal e-predictor f we can associate the sequence of *conformal e-variables*

$$E_n := f(Z_1, \dots, Z_{n-1}, Z_n). \tag{2}$$

Intuitively, large values of the conformal e-variables are evidence against $Z_1, Z_2, \dots$ being exchangeable (in particular, IID).

The *exchangeable filtration* [7, Sect. 5.6] is $(\mathcal{F}_n)$, where $\mathcal{F}_n$ is the σ-algebra generated by the multiset $\wr Z_1, \dots, Z_{n-1} \int$ and the observations $Z_n, Z_{n+1}, \dots$. The following proposition is a stronger version of the property of validity of conformal e-prediction.

Proposition 1. *For any conformal e-predictor f and any n, if the sequence $Z_1, Z_2, \dots$ is exchangeable, then*

$$\mathbb{E}(E_n \mid \mathcal{F}_{n+1}) \le 1, \tag{3}$$

where $\mathcal{F}$ is the exchangeable filtration and E_n is the conformal e-variable (2) *(with "=" in place of "≤" in* (3) *if f is admissible).*

Proof. By the definition of conformal e-predictors we have

$$\mathbb{E}\left(f(Z_1, \dots, Z_n) \mid \wr Z_1, \dots, Z_n \int\right) \le 1,$$

and we can add $Z_{n+1}, Z_{n+2}, \dots$ to the condition since $Z_1, \dots, Z_n$ are exchangeable conditionally on $Z_{n+1}, Z_{n+2}, \dots$. This is equivalent to (3). □

We still have (3) if the sequence $Z_1, Z_2, \dots = Z_1, \dots, Z_N$ is finite provided its length N is at least n, $N \ge n$. In this case the exchangeable filtration is $(\mathcal{F}_n)_{n \le N+1}$, where $\mathcal{F}_{N+1}$ is the σ-algebra generated by the multiset $\wr Z_1, \dots, Z_N \int$.

A weaker version of (3) is $\mathbb{E}(E_n) \le 1$. This can be expressed as E_n being an e-variable, where an *e-variable* is defined to be a nonnegative random variable with expected value at most 1. The values taken by e-variables are referred to as *e-values*.

3 Conformal e-testing in the online setting

In this section we start our discussion of testing the assumption of exchangeability from the online setting. Namely, the testing process proceeds in time by processing a potentially infinite stream of observations $z_1, z_2, \ldots$ sequentially one by one, and at each moment we would like to have a measure of the amount of evidence that we have found against the assumption of their exchangeability. If such a measure exceeds a large threshold c, we might want to raise an alarm indicating that exchangeability is likely to have been violated. In conformal testing, such a measure is provided by conformal test martingales, and it satisfies a strong requirement of validity (an application of Ville's inequality): we will raise an alarm with probability at most $1/c$. In this section we will see that the natural counterpart of conformal test martingales in conformal e-testing, which we call conformal e-pseudomartingales, fails completely to satisfy this strong property of validity (Proposition 2 below).

The *conformal e-pseudomartingale* corresponding to the conformal e-variables (2) is

$$S_n := E_1 \ldots E_n, \quad n = 0, 1, 2, \ldots,$$

where S_0 is understood to be 1. It may not be a genuine martingale since by Proposition 1 we have $\mathbb{E}(E_n \mid \mathcal{F}_{n+1}) = 1$ for all n instead of $\mathbb{E}(E_n \mid E_1, \ldots, E_{n-1}) = 1$ required in the definition of martingales.

The definition of conformal e-pseudomartingales, nevertheless, is very similar to that of conformal test martingales, and the conformal e-variables E_n look analogous to the betting functions of conformal testing [16, Sect. 8.1.2]. It corresponds to the gambling picture in which we start from an initial capital of 1 and then compound the conformal e-values as usual by multiplying them. The crucial difference is that in conformal testing a betting function only depends on the past p-values, whereas in conformal e-testing it may also depend on the multiset of actual observations.

The following proposition uses the notation

$$S^*_\infty := \sup_{n=1,2,\ldots} S_n.$$

Ville's inequality can then be written as $P(S^*_\infty \geq c) \leq 1/c$ for any $c > 1$ and any test martingale S w.r. to P (i.e., any nonnegative martingale S satisfying $S_0 = 1$). To exclude the trivial case, let us assume that the σ-algebra on $\mathbf{Z}$ is different from $\{\emptyset, \mathbf{Z}\}$; in particular, $|\mathbf{Z}| > 1$.

Proposition 2. *For any $\epsilon > 0$ and $c > 1$, there exists an exchangeable probability measure P on $\mathbf{Z}^\infty$ and a conformal e-pseudomartingale such that $P(S^*_\infty \geq c) \geq 1 - \epsilon$.*

Proof. Assume, without loss of generality, that $\mathbf{Z} = \{0, 1\}$ and that c is an integer. Consider an e-predictor f that stakes everything on 1; in particular, for sequences of any length n,

$$f : (0, \ldots, 0, 1) \mapsto (0, \ldots, 0, n);$$

on the other hand, let

$$f : (0, \dots, 0) \mapsto (1, \dots, 1).$$

Let $A_c \subseteq \{0,1\}^\infty$ be the set of all sequences that have $0^{[c]}$ as their prefix apart from the sequence $0^{[\infty]}$ (where $0^{[a]}$ is the sequence consisting of a 0s). Then $A_c \subseteq \{S^*_\infty \geq c\}$; besides, $P(A_c) \geq 1 - \epsilon$ if, under P, the random observations $Z_1, Z_2, \dots$ are generated in the IID manner with the probability of 1 sufficiently small (but positive). □

We interpret the statement of Proposition 2 as a complete loss of validity (in the strong sense of Ville's inequality) for conformal e-pseudomartingales. In particular, Proposition 2 demonstrates that conformal e-pseudomartingales are not martingales in general. Despite conformal e-pseudomartingales violating Ville's inequality so badly, we will see in Sect. 5 that the CUSUM procedure based on conformal e-pseudomartingales still satisfies the standard property of validity (as in [16, Corollary 8.14]). Besides, the strong property of validity holds for a fixed time horizon, as Proposition 3 below will show.

4 Conformal e-testing in the batch setting

Having established loss of validity in the online setting, in this section we move on to the batch setting, which is standard in statistics (although we use terminology that is standard in conformal testing rather than statistics). Namely, we assume that the number N of observations is fixed and known in advance. At the end of step N we are required to make a decision whether to reject the hypothesis of exchangeability, or at least to present the amount of evidence that we have found against the null hypothesis of exchangeability.

Conformal e-testing in the batch mode is defined in two steps. First, a function $E : \mathbf{Z}^N \to [0, \infty)$ is called a *basic conformal e-test* if it has the form

$$E(z_1, \dots, z_N) = \prod_{n=1}^{N} f(z_1, \dots, z_n), \quad \forall (z_1, \dots, z_N) \in \mathbf{Z}^N,$$

for some conformal e-predictor f. In other words, if $E = S_N$ for some conformal e-pseudomartingale S. And second, a *conformal e-test* is defined as a convex combination of basic conformal e-tests: namely, the conformal e-tests are defined as convex mixtures $\lambda_1 E^{(1)} + \dots + \lambda_k E^{(k)}$ of basic conformal e-tests $E^{(1)}, \dots, E^{(k)}$, where $k \in \{1, 2, \dots\}$, $\lambda_i \in [0, 1]$, and $\lambda_1 + \dots + \lambda_k = 1$. We need the second step to achieve some efficiency later on, and the following proposition gives a property of validity for conformal e-pseudomartingales at a fixed time and, slightly more generally, for conformal e-tests.

Proposition 3. *Suppose $Z_1, \dots, Z_N$ are exchangeable. For any conformal e-test E, $E(Z_1, \dots, Z_N)$ is a bona fide e-variable. It satisfies $\mathbb{E}(E(Z_1, \dots, Z_N)) = 1$ if the underlying conformal e-predictor is admissible.*

Proof. Assume, without loss of generality, that E is a basic conformal e-test. We will use the notation (2), where f is the underlying conformal e-predictor, and the exchangeable filtration $(\mathcal{F}_n)$. We can show that $\mathbb{E}(E_1 \dots E_n \mid \mathcal{F}_{n+1}) \le 1$ a.s. by induction in n (for $n = 0$ this statement is vacuously true as equality):

$$\begin{aligned}\mathbb{E}(E_1 \dots E_n \mid \mathcal{F}_{n+1}) &= \mathbb{E}(\mathbb{E}(E_1 \dots E_n \mid \mathcal{F}_n) \mid \mathcal{F}_{n+1}) \\ &= \mathbb{E}(E_n \mathbb{E}(E_1 \dots E_{n-1} \mid \mathcal{F}_n) \mid \mathcal{F}_{n+1}) \le \mathbb{E}(E_n \mid \mathcal{F}_{n+1}) \le 1 \quad \text{a.s.},\end{aligned}$$

the last inequality following from Proposition 1. If the underlying conformal e-predictor is admissible, both "$\le$" become "$=$". □

If a conformal e-test E is chosen in advance and takes a very large value on the realized data sequence, we are justified in rejecting the assumption of exchangeability.

As the next step, we explore the efficiency of conformal e-testing along the lines of the treatment of the efficiency of conformal testing in [16, Sect. 9.1]. As in the case of conformal testing [16, Sect. 9.1], for our (rather weak) statement of efficiency we will simplify our task by only considering the binary case, $\mathbf{Z} = \{0, 1\}$.

First we define unrestricted testing of exchangeability. The *upper exchangeability probability* $\mathbb{P}^{\text{exch}}$ of measurable $A \subseteq \mathbf{Z}^N$ is defined as

$$\mathbb{P}^{\text{exch}}(A) := \sup_P P(A), \tag{4}$$

P ranging over the exchangeable probability measures on $\mathbf{Z}^N$. The intuition behind $\mathbb{P}^{\text{exch}}(A)$ is that, if it is very small, A can be used for testing the exchangeability of the data-generating distribution: we are entitled to reject exchangeability if $(z_1, \dots, z_N) \in A$ provided A is chosen in advance.

The *upper conformal e-probability* $\mathbb{P}^{\text{ce}}$ of $A \subseteq \mathbf{Z}^N$ is defined as

$$\mathbb{P}^{\text{ce}}(A) := \inf \{\alpha : \exists E \; \forall (z_1, \dots, z_N) \in A : E(z_1, \dots, z_N) \ge 1/\alpha\}, \tag{5}$$

E ranging over the conformal e-tests. The intuition is the same: if $\mathbb{P}^{\text{ce}}(A)$ is very small, we are entitled to reject exchangeability if we observe $(z_1, \dots, z_N) \in A$, again assuming that A is chosen in advance. The difference from $\mathbb{P}^{\text{exch}}(A)$ is that now the lack of exchangeability must be demonstrated via conformal e-testing.

The following proposition is an analogue of [16, Proposition 9.5]. Intuitively, its second statement says that conformal e-testing is universal in the batch mode: if lack of exchangeability can be demonstrated at all, it can be demonstrated (albeit less convincingly) using conformal e-testing. (And its first statement is another expression of validity.)

Proposition 4. *For any event $A \subseteq \mathbf{Z}^N$, $\mathbb{P}^{\text{exch}}(A) \le \mathbb{P}^{\text{ce}}(A)$. Assuming $\mathbf{Z} = \{0, 1\}$, $\mathbb{P}^{\text{ce}}(A) \le N \, \mathbb{P}^{\text{exch}}(A)$.*

Proof. The inequality $\mathbb{P}^{\text{exch}}(A) \le \mathbb{P}^{\text{ce}}(A)$ follows immediately from Markov's inequality applied to conformal e-tests E in combination with Proposition 3: if

$E(z_1, \dots, z_N) \ge 1/\alpha$ for all $(z_1, \dots, z_N) \in A$,

$$\mathbb{P}^{\text{exch}}(A) \le \mathbb{P}(E(Z_1, \dots, Z_N) \ge 1/\alpha) \le \frac{\mathbb{E}(E(Z_1, \dots, Z_N))}{1/\alpha} \le \alpha,$$

assuming $Z_1, \dots, Z_N$ are exchangeable.

Now assume $\mathbf{Z} = \{0, 1\}$; this part of the proof will be a modification of the proof of Proposition 9.5 in [16, Sect. 9.4.2]. As a first step, notice that it suffices to prove $\mathbb{P}^{\text{ce}}(A) \le \mathbb{P}^{\text{exch}}(A)$ for any nonempty $A \subseteq \mathbf{Z}^N$ such that each sequence in A has the same number of 1s. Let us fix such an A, and let $K \in \{0, \dots, N\}$ be the number of 1s in the elements of A. For each sequence $\zeta = (z_1, \dots, z_N) \in A$, consider the basic conformal e-test $E_\zeta = e_1 \dots e_N$, where the nth e-value e_n is

$$e_n := \begin{cases} n/k & \text{if } z_n = 1 \\ n/(n-k) & \text{if } z_n = 0, \end{cases}$$

k being the number of 1s among the first n elements of ζ. (This corresponds to a nonconformity e-measure satisfying

$$A\left(0^{[n-k]}, 1^{[k]}\right) := \begin{cases} \left(0^{[n-k]}, \left(\frac{n}{k}\right)^{[k]}\right) & \text{if } z_n = 1 \\ \left(\left(\frac{n}{n-k}\right)^{[n-k]}, 0^{[k]}\right) & \text{if } z_n = 0, \end{cases}$$

where $b^{[a]} = b, \dots, b$ (a times), as in the proof of Proposition 2. This was called "reckless gambling" in [16, end of Sect. 9.1.2].) The product $E_\zeta = e_1 \dots e_N$ will then be

$$\frac{N!}{K!(N-K)!} = \binom{N}{K}.$$

The arithmetic mean of E_ζ over $\zeta \in A$ witnesses that

$$\mathbb{P}^{\text{ce}}(A) \le |A| / \binom{N}{K} = \mathbb{P}^{\text{exch}}(A). \qquad \square$$

Proposition 2 can be restated in terms of $\mathbb{P}^{\text{exch}}$ and $\mathbb{P}^{\text{ce}}$ adapted to the online setting. Let $A \subseteq \mathbf{Z}^\infty$. Define $\mathbb{P}^{\text{exch}}(A)$ by (4), as before, with P ranging over the exchangeable probability measures on $\mathbf{Z}^\infty$. In the spirit of Ville's inequality, let us modify (5) as

$$\mathbb{P}^{\text{ce}}(A) := \inf \{\alpha : \exists S \; \forall (z_1, z_2, \dots) \in A : S^*_\infty(z_1, z_2, \dots) \ge 1/\alpha\},$$

where S ranges over the conformal e-pseudomartingales and we slightly abuse our notation by regarding S_n as functions of the observations. Then we can see that $\mathbb{P}^{\text{exch}}(A_c) = 1$ and $\mathbb{P}^{\text{ce}}(A_c) \le 1/c \to 0$ as $c \to \infty$ (where A_c are defined in the proof of Proposition 2). This can be expressed as $\mathbb{P}^{\text{exch}}$ and $\mathbb{P}^{\text{ce}}$ being entirely asymptotically separated (cf. [11, Sect. 3.10, especially (3)]).

5 The conformal CUSUM e-procedure: validity

In the previous section we discussed testing the assumption of exchangeability once, but in some important applications we would like to test it repeatedly over time in the online mode (see, e.g., [16, Sect. 8.3]): as soon as we suspect that exchangeability is violated, we raise an alarm, and we are allowed do so more than once. The topic of this and next sections is such multistage exchangeability testing. As usual, such procedures are required to satisfy properties of validity and efficiency. Here validity means that, under exchangeability, the probability or frequency of alarms (which in this case are *false alarms*) should be bounded by a prespecified constant. When applied to the moment when the first alarm is raised, such properties of validity are much weaker than the property of validity discussed in the previous sections: the first alarm is usually raised, sooner or later, with probability one. And efficiency means that, if exchangeability is violated at some point, an alarm should be raised as quickly as possible afterwards. In this section we concentrate on validity of multistage testing.

Suppose we observe a sequence of e-values $e_1, e_2, \dots$ output by a conformal e-predictor in the online protocol, as described in Sect. 3, under exchangeability. The *conformal CUSUM e-procedure* [5] raises the kth alarm, $k = 1, 2, \dots$, at the time

$$\tau_k := \min \left\{ n > \tau_{k-1} : \max_{i \in \{\tau_{k-1}+1, \dots, n\}} e_i \dots e_n \ge c \right\}, \tag{6}$$

where $\tau_0 := 0$ and $c > 1$ is the parameter of the procedure. It is usually applied in the situation where the observations z_i are generated independently first from a known probability measure $Q_0 \in \mathfrak{P}(\mathbf{Z})$ and then from another known probability measure $Q_1 \in \mathfrak{P}(\mathbf{Z})$ ($\mathfrak{P}(\mathbf{Z})$ being the family of all probability measures on $\mathbf{Z}$), and where e_i is the likelihood ratio of Q_1 to Q_0 evaluated at z_i (see, e.g., [6, Sect. 6.2] or (10) below). In our current context, where the e_i are conformal e-values, we may call it the *conformal CUSUM e-procedure.*

The following proposition gives an asymptotic property of validity for the conformal CUSUM e-procedure.

Proposition 5. *Let A_n be the number of alarms*

$$A_n := \max\{k \mid \tau_k \le n\} \tag{7}$$

raised by the conformal CUSUM e-procedure (6) *after processing the first n observations $Z_1, \dots, Z_n$. Then*

$$\limsup_{n \to \infty} \frac{A_n}{n} \le \frac{1}{c} \quad a.s. \tag{8}$$

provided the observations $Z_1, Z_2, \dots$ are exchangeable.

In practical applications, Proposition 5 can be applied when, after investigating each alarm, it is decided that the alarm was false, and so we can continue processing a single stream of observations. If the alarm was genuine, we need to reset the multistage procedure, and Proposition 5 is not applicable.

A popular modification of the CUSUM procedure is the *Shiryaev–Roberts procedure*, which replaces the max in (6) by the sum $\sum$ (see, e.g., [6, Sect. 6.5]). To state our next validity result (Proposition 6 below), we need the modification

$$\tau_k := \min\left\{ n > \tau_{k-1} : \max_{i \in \{\tau_{k-1}+1,\ldots,n\}} \sum_{j=i}^{n} e_i \ldots e_j \geq c \right\}, \qquad (9)$$

which we will call the *reverse Shiryaev–Roberts procedure* (or *conformal reverse Shiryaev–Roberts e-procedure* to emphasize e_i being produced by a conformal e-predictor).

Proposition 6. *Let A_n be defined by* (7) *for the conformal reverse Shiryaev–Roberts e-procedure* (9)*. Then we still have* (8) *provided the observations $Z_1, Z_2, \ldots$ are exchangeable.*

See Appendix A for the proofs. We will, of course, deduce Proposition 5 from Proposition 6: it is clear that the Shiryaev–Roberts procedure (regular or reverse) raises alarms at least as often as CUSUM does. However, the following example shows an advantage of CUSUM from an intuitive point of view.

Example 7. Consider the vacuous admissible conformal e-predictor identically equal to 1. The CUSUM procedure based on it will never raise alarms, while the Shiryaev–Roberts e-procedure will raise alarms every $\lceil c \rceil$th step, thereby fully exploiting (for an integer c) the leeway permitted by our target property of validity (8). This example shows one feature of the Shiryaev–Roberts procedure (shared by the reverse Shiryaev–Roberts procedure) that can be considered its disadvantage: while the CUSUM procedure raises an alarm when it has genuine evidence for disorder, the Shiryaev–Roberts procedure may raise an alarm simply because it is allowed to do so by a given constraint; we might not have any evidence for disorder.

Because of the feature of the Shiryaev–Roberts procedure illustrated in Example 7, in the next section we will concentrate on CUSUM-type procedures.

6 The conformal CUSUM e-procedure: efficiency

The standard version of the CUSUM procedure satisfies important properties of optimality [6, Chap. 6]: it is optimal in Lorden's [3] sense, as shown by Moustakides [4]; it is also optimal in Ritov's [8] very natural game-theoretic sense. However, as we mentioned in the previous section, the standard CUSUM procedure works under restrictive assumptions: we know the prechange distribution $Q_0 \in \mathfrak{P}(\mathbf{Z})$ and the postchange distribution $Q_1 \in \mathfrak{P}(\mathbf{Z})$, and the only unknown is the changepoint N_0; the observations $Z_1, \ldots, Z_{N_0}$ are generated from Q_0, and the observations $Z_{N_0+1}, Z_{N_0+2}, \ldots$ are generated from Q_1, all independently. Let f_0 and f_1 be probability densities of Q_0 and Q_1, respectively, w.r. to a σ-finite measure μ on $\mathbf{Z}$ (such as $\mu := Q_0 + Q_1$) under which both Q_0 and Q_1 are absolutely continuous. We will use $\mathbb{E}_0$ and $\mathbb{E}_1$ for expectations w.r. to Q_0 and Q_1, respectively.

The *standard CUSUM procedure* is based on the likelihood ratios

$$L_n := f_1(z_n)/f_0(z_n). \tag{10}$$

(For simplicity, the reader may assume that Q_0 and Q_1 are positive discrete distributions and replace $f_0(z)$ and $f_1(z)$ by $Q_0(\{z\})$ and $Q_1(\{z\})$, respectively.) As in (6), the kth alarm is raised at the time

$$\tau_k := \min\left\{n > \tau_{k-1} : \max_{i \in \{\tau_{k-1}+1,\dots,n\}} L_i \dots L_n \geq c\right\}, \tag{11}$$

with $\tau_0 := 0$. For comparison, we will also discuss experimental results for the *conformal CUSUM procedure*, which is defined in a similar way using conformal p-values, as explained in [16, Sect. 8.3.1].

This section implements a version of the "Burnaev–Wasserman programme" [16, Sect. 2.5] applied to the CUSUM procedure. Suppose we would like to detect deviations from exchangeability, but we have a prior model of the data generation mechanism in which the observations are first generated from a given $Q_0 \in \mathfrak{P}(\mathbf{Z})$ and then from another given probability measure, $Q_1 \in \mathfrak{P}(\mathbf{Z})$, independently of the previous observations. We however, do not trust our model and do not want the validity of our procedure to depend on it. Therefore, we "conformalize" the standard CUSUM procedure, using the likelihood ratios (10) (normalized to ensure (1)) as nonconformity e-scores. (Details will follow shortly.) The procedure of conformalization may be said to work well if the quality of the conformalized CUSUM is not significantly worse than the quality of the standard CUSUM even when the assumptions on which the standard CUSUM is based are fully satisfied.

Some experimental results for the case of Bernoulli observations are shown in Figure 1. In this experiment we generate $N_0 := 1000$ observations $Z_1, \dots, Z_{N_0}$ from the Bernoulli distribution with parameter 0.5 and another $N_1 := 1000$ observations $Z_{N_0+1}, \dots, Z_{N_0+N_1}$ from the Bernoulli distribution with parameter 0.6. The changepoint 1000 is shown as a dashed vertical line. The left panel of Figure 1 shows the paths of five stochastic processes:

- the likelihood ratio martingale $L_1 \dots L_n$, $n = 0, 1, \dots, N_0 + N_1$, in blue;
- the conformal e-pseudomartingale $E_1 \dots E_n$, $n = 0, 1, \dots, N_0 + N_1$ in orange; the conformal e-values are defined as the normalized L_n:

$$E_n := \frac{L_n}{\frac{1}{n}(L_1 + \dots + L_n)}; \tag{12}$$

- the green, red, and purple lines correspond to conformal testing; they will be discussed later and can be ignored for now.

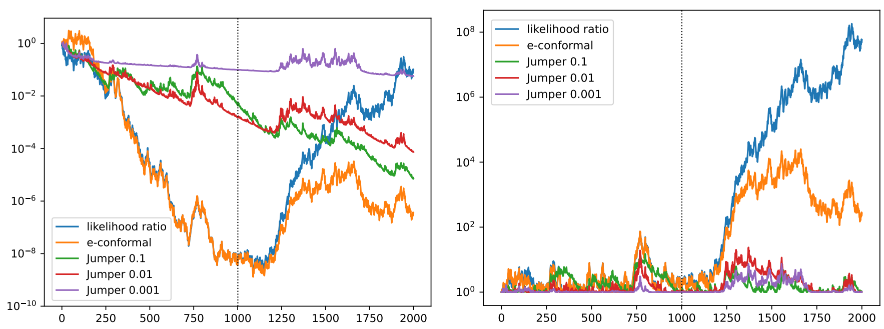

Figure 1: Five stochastic processes, as described in text, in the Bernoulli case (with the parameter 0.5 before the changepoint and 0.6 after the changepoint). Left panel: the paths of the processes. Right panel: the paths of the corresponding CUSUM statistics.

The right panel of Figure 1 shows the corresponding *CUSUM statistics*, where the CUSUM statistics of a process S_n, $n = 0, 1, 2, \ldots$ are defined as

$$S'_n := S_n / \min(S_0, \ldots, S_{n-1})$$

(notice that τ_1 in (6) and (11) is defined as the moment when the CUSUM statistic first reaches level c).

The five lines (CUSUM statistics) in the right panel of Figure 1 illustrate the efficiency of various versions of the CUSUM procedure in detecting the changepoint. The detection happens when the line first reaches the level c after the changepoint. (We ignore the rare case where the level c is reached between the changepoint and the last time before the changepoint when the CUSUM statistic is 0.) Without fixing c, we can judge how efficient the procedure is by the slope of the line after the changepoint. For example, we can see that the conformal CUSUM e-procedure (corresponding to the orange line) will raise an alarm not much later than the standard CUSUM procedure (corresponding to the blue line) for c up to about 10^3.

The blue and orange lines in the right panel are close to each other shortly after the changepoint, suggesting that conformal e-testing is efficient at first, but then they start diverging. This divergence illustrates the phenomenon of "decay" discussed in [16, Sect. 8.4.1]. In fact, decay is inevitable and by itself does not indicate lack of efficiency of conformal e-testing (unless it sets in too early). Remember that the likelihood ratio martingale and conformal e-pseudomartingale are testing very different null hypotheses: for the former the null hypothesis is Q_0, and for the latter it is exchangeability. If our null hypothesis is Q_0, we will be constantly surprised seeing observations from $Q_1 \neq Q_0$, whereas under exchangeability Q_1 will gradually become "the new normal", and we will stop being surprised. Therefore, decay will inevitably happen, and it is the moment when it sets in that is the hallmark of efficiency (the later the better).

In our experiments we always use the standard seed 42 for the numpy random number generator, but the results are qualitatively similar for other seeds as well (one unusual feature of Figure 1 is that the likelihood ratio martingale does not start its ascent right after the changepoint, which typically happens for other seeds).

The green, red, and purple lines in the left panel of Figure 1 show the paths of three conformal test martingales, as defined in [16, Sect. 8.1.2]. (This and next paragraphs depend on and use the terminology of [16, Sect. 8.1].) The nonconformity score of each observation Z_n (here and in Figure 2 below) is defined to be the likelihood ratio L_n (which is equivalent to using Z_n itself as nonconformity score in the current Bernoulli case). This produces a sequence of conformal p-values $p_1, p_2, \dots$. To turn conformal p-values into a conformal test martingale, we need to define betting functions. In Figure 1 we use the Simple Jumper betting functions [16, Sect. 8.1.2] omitting the adjective "Simple" in the caption. The resulting Simple Jumper martingale depends on a parameter called *jumping rate*, and we use three values for it as indicated in the legend.

It would be more in the spirit of the Burnaev–Wasserman programme to adapt the betting functions to the assumed Q_0 and Q_1. This is an interesting direction for further research, but in this paper we are only using Simple Jumper, a generic betting martingale that often gives satisfactory results for a wide range of datasets [16, Sect. 8.1.2, (8.8)]. (Using a generic betting martingale is analogous to our use of a generic way of betting, given by (12), in the conformal CUSUM e-procedure.)

To start theoretical analysis of the phenomena demonstrated by the blue and orange lines in Figure 1, let us see why the blue line changes its slope after the changepoint. Before the changepoint, $n \le N_0$, the expectation of the likelihood ratio is at most 1:

$$\mathbb{E}_0 L_n = \mathbb{E}_0 \frac{f_1(Z_n)}{f_0(Z_n)} = \int \frac{f_1}{f_0} f_0 \, \mathrm{d}\mu \le \int f_1 \, \mathrm{d}\mu = 1; \tag{13}$$

it is 1 if Q_1 is absolutely continuous w.r. to Q_0 (which is always the case in our experiments). After the changepoint, $n > N_0$, it becomes

$$\mathbb{E}_1 L_n = \mathbb{E}_1 \frac{f_1(Z_n)}{f_0(Z_n)} = \int \frac{f_1}{f_0} f_1 \, \mathrm{d}\mu = 1 + \int \frac{(f_1 - f_0)^2}{f_0} \, \mathrm{d}\mu = 1 + \chi^2(Q_0, Q_1), \tag{14}$$

where χ^2 stands for the χ^2 distance between probability measures (see, e.g., [1, Sect. 31, Definition 2]); the value (14) exceeds 1 unless $Q_0 = Q_1$. This suggests that the blue line is close to being horizontal before the changepoint while it starts increasing after the changepoint.

While the calculations (13)–(14) are useful and will be used in the proof of Proposition 8 below, it would be wrong to interpret them directly as indicators of the tendency of the likelihood ratio martingale to increase or decrease. For that, we should use

$$\mathbb{E}_0 \ln L_n = \int f_0 \ln \frac{f_1}{f_0} \,\mathrm{d}\mu = -\mathrm{KL}(Q_0, Q_1) < 0$$

before the changepoint and

$$\mathbb{E}_1 \ln L_n = \int f_1 \ln \frac{f_1}{f_0} \,\mathrm{d}\mu = \mathrm{KL}(Q_1, Q_0) > 0$$

after the changepoint, where KL stands for the Kullback–Leibler divergence. Therefore, the slope of the blue line is expected to be negative before the changepoint and positive after it. This agrees with what we see in Figure 1.

Let us now check informally that the conformal e-values E_n and the likelihood ratios L_n can be expected to be close to each other soon after the changepoint. Indeed, soon after the changepoint we will have

$$E_{N_0+n} = \frac{L_{N_0+n}}{\frac{1}{N_0+n}(L_1 + \dots + L_{N_0} + L_{N_0+1} + \dots + L_{N_0+n})} \tag{15}$$

$$\approx \frac{L_{N_0+n}}{\frac{N_0}{N_0+n} + \frac{1}{N_0+n}(L_{N_0+1} + \dots + L_{N_0+n})} \approx L_{N_0+n}. \tag{16}$$

The equality in (15) is just an application of the definition (12). The first approximate equality (which holds with high probability) in (16) follows from the law of large numbers assuming N_0 is large; we obtained it by replacing $L_1 + \dots + L_{N_0}$ by N_0 (see (13)). In the denominator in (16) we have a weighted average of 1 and the postchange likelihood ratios. The cumulative weight $\frac{n}{N_0+n}$ of the postchange likelihood ratios is small if $n \ll N$, and the second approximate equality in (16) assumes both $n \ll N$ and the postchange likelihood ratios being only moderately large. The presence of the postchange likelihood ratios in the denominator in (16) can be regarded as the usual conformal adjustment in this context. The closeness of E_n and L_n is manifested in the closeness of the blue and orange lines in Figure 1 soon after the changepoint.

The following proposition compares more formally the behaviour of the likelihood ratio martingale and the conformal e-pseudomartingale after the changepoint N_0. For simplicity we will assume that the likelihood ratios L_n are bounded; e.g., $L_n \in [0.8, 1.2]$ in the situation of Figure 1.

Proposition 8. *Suppose that Z_n, $n = 1, \dots, N_0$, are generated from $Q_0 \in \mathfrak{P}(\mathbf{Z})$, that Z_n, $n = N_0 + 1, \dots, N_0 + N_1$, are generated from $Q_1 \in \mathfrak{P}(\mathbf{Z})$, all independently, and that the likelihood ratios L_n are bounded: $0 \le a \le L_n \le b < \infty$. Suppose $\mathbb{E}_0 L_n = 1$ and $c := \mathbb{E}_1 L_n > 1$. For any $\epsilon \in (0,1)$, we have*

$$\forall n \in \{1, \dots, N_1\} : \ln \frac{L_{N_0+1} \dots L_{N_0+n}}{E_{N_0+1} \dots E_{N_0+n}} < \frac{n(n+1)}{N_0} \frac{c}{2} + \frac{n}{N_0^{1/2}} |b - a| \sqrt{\frac{1}{2} \ln \frac{2}{\epsilon}} + \frac{(N_1+1)^{3/2}}{N_0} |b - a| \sqrt{\frac{1}{6} \ln \frac{2}{\epsilon}} \tag{17}$$

with probability at least $1 - \epsilon$.

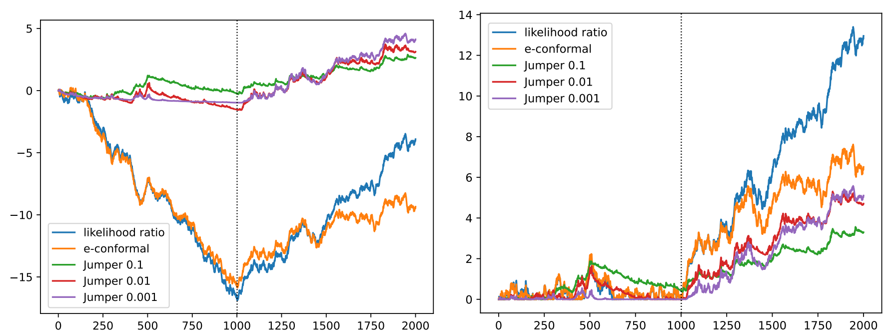

Figure 2: Five processes as in Figure 1 but for the Cauchy distributions with the location and scale parameters $(0,1)$ before the changepoint and $(0,0.7)$ after the changepoint. Left panel: the raw processes. Right panel: the corresponding CUSUM statistics.

The proof of Proposition 8 is given in Appendix B.

The inequality in (17) shows that, for a fixed ϵ, the conformal e-pseudomartingale grows after the changepoint almost as fast as the likelihood ratio martingale if $N_1 \ll \sqrt{N_0}$. In the situation of Figure 1, the proposition tells us that we can expect similar rates of growth for the conformal e-pseudomartingale and the likelihood ratio martingale for around $\sqrt{1000} \approx 30$ steps, whereas we observe similar rates of growth for about 300 steps.

Figure 2 is a counterpart of Figure 1 for continuous, namely Cauchy, distributions. Plots for Gaussian distributions, which are more standard, look similar, but we have chosen a more awkward case of a distribution without a mean, to make it less similar to the Bernoulli case. The conformal test martingales, despite the generic nature of their betting functions, now work much better than in Figure 1; to see that they are still much worse than the conformal e-pseudomartingale, the reader should pay attention to their behaviour soon after the changepoint at 1000, where their growth is relatively sluggish.

7 Conclusion

The only known approach to detecting exchangeability violations online before this paper was based on conformal prediction; see, e.g., [16, Part III]. The approach of this paper is based instead on conformal e-prediction. The two approaches are very different, and neither dominates the other in all interesting applications. These are some differences:

- Design of conformal test martingales involves two distinct steps: using a conformity measure to obtain p-values and then betting against those p-values. Conformal e-pseudomartingales do not involve such a rigid division and thus appear to be more flexible.

- On the other hand, when betting on the nth step against the nth p-value p_n, $n = 1, 2, \dots$, conformal test martingales may use the previous p-values $p_1, \dots, p_{n-1}$. Such dependence on the past is not allowed for conformal e-pseudomartingales.

- Conformal test martingales are randomized (without randomization we only obtain conformal test supermartingales) whereas conformal e-pseudomartingales do not require randomization (it is optional and not used in this paper).

We have discussed advantages and disadvantages of conformal testing and conformal e-testing. In summary:

1. In the online protocol, conformal testing relies on bona fide test martingales, which can then be used in one-off and multistage conformal testing. On the other hand, conformal e-pseudomartingales are not, in general, test martingales, and for them Ville's inequality can be violated badly (Sect. 3).

2. A weakened variants of an efficiency result for conformal testing adapted to conformal e-testing is discussed in Sect. 4.

3. In Sect. 5 we construct a conformal CUSUM e-procedure that is a natural modification of the standard CUSUM procedure, while construction of an efficient conformal CUSUM procedure is more difficult and dependent on the postulated Q_0 and Q_1 (whereas (12) is applicable universally).

4. A final advantage of conformal e-testing is that it requires no randomization.

The first strength, 1, is a clear advantage of conformal testing over conformal e-testing. For strength 2, the picture is more ambiguous, as the strength of conformal testing still partly survives for conformal e-testing. And we also have two advantages, 3 and 4, of conformal e-testing.

These are some interesting directions of further research:

- Developing betting functions that are better adapted to the assumed probability distributions Q_0 and Q_1 than the generic ones used in the Simple Jumper method (Sect. 6). (Perhaps in the spirit of the "Bayes–Kelly" approach of [16, Part III].)

- Is it possible to strengthen Proposition 8 to obtain performance guarantees for the conformal CUSUM e-procedure that are comparable with what we observe in the experimental results?

- One disadvantage of using the CUSUM procedure with a threshold c and the target asymptotic frequency of false alarms $1/c$ is that this method, while valid by Proposition 5, might be a conservative way of achieving this target: namely, Proposition 5 appears to be conservative, since even the

Shiryaev–Roberts procedure achieves the target. Is it possible to employ, e.g., adaptive conformal inference [2] or defensive forecasting [10, Chap. 12] to adapt online the threshold c to a target frequency of false alarms α?

Acknowledgments

This research has been partially supported by Astra Zeneca, Stena Line, and Mitie.

References

[1] Alexander A. Borovkov. *Mathematical Statistics*. Gordon and Breach, Amsterdam, 1998.

[2] Isaac Gibbs and Emmanuel J. Candès. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

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

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

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

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

[7] Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38:576–601, 2023.

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

[9] Sébastien Roch. *Modern Discrete Probability: An Essential Toolkit*. Cambridge University Press, Cambridge, 2024.

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

[11] Albert N. Shiryaev. *Probability-1*. Springer, New York, third edition, 2016.

[12] Jean Ville. *Etude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939.

[13] Vladimir Vovk. On-line Confidence Machines are well-calibrated. In *Proceedings of the Forty Third Annual Symposium on Foundations of Computer Science*, pages 187–196, Los Alamitos, CA, 2002. IEEE Computer Society.

[14] Vladimir Vovk. Testing randomness online. *Statistical Science*, 36:595–611, 2021. For earlier versions, see arXiv:1906.09256 [math.PR] (from 2019).

[15] Vladimir Vovk. Conformal e-prediction. *Pattern Recognition*, 0:0–0, 2024. Submitted for publication in the Special Issue on Conformal Prediction and Distribution-Free Uncertainty Quantification.

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

[17] Vladimir Vovk, Ilia Nouretdinov, and Alex Gammerman. Testing exchangeability on-line. In Tom Fawcett and Nina Mishra, editors, *Proceedings of the Twentieth International Conference on Machine Learning*, pages 768–775, Menlo Park, CA, 2003. AAAI Press.

A Proofs for Sect. 5

The idea of the proof of Proposition 6 consists, as usual, in reversing the direction of time. Let N be a sufficiently large natural number and $(\mathcal{F}_n)$ be the exchangeable filtration.

Remember that E_n is the nth conformal e-variable (2); it has e_n as it value. Then $(E_n, \mathcal{F}_n)$, $n = N, \dots, 1$, is an exact e-flow (in the terminology of [15]) over the finite time interval $N, \dots, 1$, in the sense

$$\mathbb{E}(E_n \mid \mathcal{F}_{n+1}) = 1, \quad n = N, \dots, 1.$$

The corresponding martingale is $(T_n, \mathcal{F}_n)$, $n = N+1, \dots, 1$, where

$$T_n := E_n \dots E_N, \quad n = N+1, N, \dots, 1,$$

with T_{N+1} understood to be 1. For simplicity, let us assume that all E_n are positive, so that T_n is a positive martingale.

Let us apply the Shiryaev–Roberts procedure to the e-flow $E_N, \dots, E_1, \dots$ continued by setting $E_0 := E_{-1} := \dots := 1$. It gives us the decreasing sequence of stopping times $\sigma_0 := N + 1$ and

$$\sigma_k := \max \left\{ n < \sigma_{k-1} \mid \sum_{i=n}^{\sigma_{k-1}-1} E_n \dots E_i \ge c \right\}, \quad k = 1, 2, \dots, \tag{18}$$

where n ranges over the integers. A useful property of the Shiryaev–Roberts procedure is $\mathbb{E}(\sigma_{k-1} - \sigma_k \mid \mathcal{F}_{\sigma_k}) \ge c$; see, e.g., [16, Proposition 8.13] (proved in [16, Sect. 8.5.2]; the current proof also uses some other ideas in that section).

To make the stopping times (18) more manageable, it will be convenient to force an alarm every L steps, where L is to be chosen later: $\sigma'_0 := N+1$ and

$$\sigma'_k := (\sigma'_{k-1} - L) \vee \max\left\{ n < \tau_{k-1} \mid \sum_{i=n}^{\sigma'_{k-1}-1} E_n \dots E_i \ge c \right\}, \quad k = 1, 2, \dots;$$

let us call this the *truncated Shiryaev–Roberts procedure*. As in the proof of [16, Corollary 8.4], induction in k shows that $\sigma'_k \ge \sigma_k$ for all k.

There is a useful connection between τ_k and σ'_k: namely, each set $\{\tau_{k-1}+1, \dots, \tau_k\}$ with $\tau_k \le N$ contains at least one stopping time σ'_l. (This is even true for σ_l, although we do not need it.) This can be deduced from

$$\sum_{j=i}^{\tau_k} E_i \dots E_j \ge c \text{ for some } i \ge \tau_{k-1} + 1. \tag{19}$$

Indeed, let l be the largest number satisfying $\sigma'_l > \tau_k$; arguing indirectly, let us suppose that $\sigma'_{l+1} < \tau_{k-1} + 1$. The inequality (19) implies

$$\sum_{j=i}^{\sigma'_l - 1} E_i \dots E_j \ge \sum_{j=i}^{\tau_k} E_i \dots E_j \ge c \text{ for some } i > \sigma'_{l+1},$$

which contradicts the definition of the stopping times σ'.

As in [16, Proof of Proposition 8.15 in Sect. 8.5.2], let us say that k is *slow* if

$$\mathbb{P}\left(\sigma'_{k-1} - \sigma'_k = L \mid \mathcal{F}_{\sigma'_{k-1}}\right) \ge c/L$$

and *fast* otherwise. We show there that

$$\mathbb{E}\left(\sigma'_{k-1} - \sigma'_k \mid \mathcal{F}_{\sigma'_{k-1}}\right) \ge c - c^2/L \tag{20}$$

if k is fast. On the other hand, if k is slow,

$$\mathbb{E}\left(\sigma'_{k-1} - \sigma'_k \mid \mathcal{F}_{\sigma'_{k-1}}\right) \ge L\mathbb{P}\left(\sigma'_{k-1} - \sigma'_k = L \mid \mathcal{F}_{\sigma'_{k-1}}\right) \ge c.$$

In both cases, we have (20).

Let A'_N be the largest k such that $\sigma'_k > 0$; we interpret A'_N as the number of alarms raised by the truncated Shiryaev–Roberts procedure. Remember that A_N, defined by (7), is the number of alarms raised by the conformal reverse Shiryaev–Roberts e-procedure. By Hoeffding's inequality [16, Sect. A.6.3], for arbitrarily small $\epsilon \in (0, c)$ and sufficiently large N,

$$\begin{aligned}\mathbb{P}\left(\frac{A_N}{N} \ge \frac{1}{c-\epsilon}\right) &\le \mathbb{P}\left(\frac{A'_N}{N} \ge \frac{1}{c-\epsilon}\right) = \mathbb{P}\left(A'_N \ge \left\lceil \frac{N}{c-\epsilon} \right\rceil\right) \\ &\le \mathbb{P}\left(\sum_{k=1}^{\lceil \frac{N}{c-\epsilon} \rceil} (\sigma'_{k-1} - \sigma'_k) \le N\right) \le \exp\left(-2\frac{\left(\lceil \frac{N}{c-\epsilon} \rceil \left(c - \frac{c^2}{L}\right) - N\right)^2}{L^2 \lceil \frac{N}{c-\epsilon} \rceil}\right)\end{aligned}$$

$$\leq \exp(-\epsilon' N),$$

where ϵ' is a positive constant (which requires a sufficiently large L). Since the series $\sum_N \exp(-\epsilon' N)$ converges,

$$\frac{A_N}{N} \geq \frac{1}{c-\epsilon}$$

happens only finitely often, which completes the proof of Proposition 6.

To deduce Proposition 5 notice that, by induction in k, $\tau_k^{\mathrm{SR}} \leq \tau_k^{\mathrm{CUSUM}}$, where τ_k^{CUSUM} are the τ_k defined by (6), and τ_k^{SR} are the τ_k defined by (9).

B Proof for Sect. 6

The proof will be based on the following "maximal" version of Hoeffding's inequality.

Proposition 9 (maximal Hoeffding inequality)**.** *Let $\mathcal{F}_0, \ldots, \mathcal{F}_N$ be a filtration. For any deterministic sequence $c_1, \ldots, c_N$ of positive numbers, any predictable sequence $a_1, \ldots, a_N$ w.r. to $(\mathcal{F}_i)$, any supermartingale difference $\xi_1, \ldots, \xi_N$ w.r. to $(\mathcal{F}_i)$ such that $\xi_i \in [a_i, a_i + c_i]$, $i = 1, \ldots, N$, and any $\beta > 0$,*

$$\mathbb{P}\left\{\max_{n=0,\ldots,N} \sum_{i=1}^{n} \xi_i \geq \beta\right\} \leq \exp\left(-\frac{2\beta^2}{\sum_{i=1}^{N} c_i^2}\right) \tag{21}$$

For a proof of Proposition 9, see, e.g., [9, Theorem 3.2.1] (or adapt a proof of Hoeffding's standard inequality, such as that given in [16, Sect. A.6.3]).

The interpretation of the three addends on the right-hand side of the inequality in (17) is: the first addend reflects the effect of the expectation c of the postchange likelihood ratios, the second addend reflects the volatility of the prechange likelihood ratios, and the third addend reflects the volatility of the postchange likelihood ratios. We start from the first addend and split the permitted probability ϵ of violating (17) into two equal parts, one controlling the prechange behaviour of the likelihood ratios and the other controlling their postchange behaviour.

By Hoeffding's inequality (21) we will have

$$\sum_{n=1}^{N_0} L_n < N + \beta \text{ with probability} \geq 1 - \frac{\epsilon}{2} \tag{22}$$

when

$$\exp\left(-\frac{2\beta^2}{N_0(b-a)^2}\right) = \frac{\epsilon}{2},$$

and so, solving this equation, we set

$$\beta := \sqrt{\frac{1}{2} N_0 (b-a)^2 \ln \frac{2}{\epsilon}}. \tag{23}$$

The log ratio in the left-hand side of the inequality in (17) can be bounded from above as follows:

$$\ln \frac{L_{N_0+1} \dots L_{N_0+n}}{E_{N_0+1} \dots E_{N_0+n}} \tag{24}$$
$$= \ln \left(\frac{L_1 + \dots + L_{N_0} + L_{N_0+1}}{N_0 + 1} \dots \frac{L_1 + \dots + L_{N_0} + L_{N_0+1} + \dots + L_{N_0+n}}{N_0 + n} \right)$$
$$< \ln \left(\frac{N_0 + \beta + L_{N_0+1}}{N_0} \dots \frac{N_0 + \beta + L_{N_0+1} + \dots + L_{N_0+n}}{N_0} \right)$$
$$\le \frac{n\beta + nL_{N_0+1} + \dots + L_{N_0+n}}{N_0}, \tag{25}$$

where the first inequality, which holds with probability at least $1 - \epsilon/2$, follows from the inequality in (22) (decreasing the denominators does not affect the validity of the first inequality), and the second inequality follows from the standard inequality $\ln(1 + x) \le x$. The expectation of the upper bound (25) on the log-ratio (24) is

$$\frac{n\beta + c(1 + \dots + n)}{N_0} = \frac{n\beta + cn(n + 1)/2}{N_0}. \tag{26}$$

By Hoeffding's maximal inequality (21) we will have

$$\forall n \in \{1, \dots, N_1\} : \frac{n(L_{N_0+1} - c) + \dots + (L_{N_0+n} - c)}{N_0} < B \tag{27}$$

with probability at least $1 - \epsilon/2$ when

$$\exp \left(-\frac{2B^2 N_0^2}{(b - a)^2 + \dots + N_1^2 (b - a)^2} \right) = \frac{\epsilon}{2}.$$

Solving the last equation and using the standard identity $1 + \dots + n^2 = n(n + 1)(2n + 1)/6$, we set

$$B := \frac{(N_1 + 1)^{3/2}}{N_0} \sqrt{\frac{1}{6}(b - a)^2 \ln \frac{2}{\epsilon}}$$
$$> \frac{1}{N_0} \sqrt{\frac{1}{12} N_1 (N_1 + 1)(2N_1 + 1)(b - a)^2 \ln \frac{2}{\epsilon}}. \tag{28}$$

Combining (23), (24)–(25), (26), (27), and (28), we obtain that (17) holds with probability at least $1 - \epsilon$, which completes the proof of Proposition 8.