Conformal e-prediction for change detection

Vladimir Vovk

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 June 6, 2020.

Project web site:
http://alrw.net

Abstract

We adapt conformal e-prediction to change detection, defining analogues of the Shiryaev–Roberts and CUSUM procedures for detecting violations of the IID assumption. Asymptotically, the frequency of false alarms for these analogues does not exceed the usual bounds.

Contents

1 Introduction

We adapt conformal e-predictors, as defined in [7], to change detection. The standard approaches to change detection assume the independence of observations (given the change-point in the Bayesian approach) and known pre-change and post-change distributions (again given the change-point in the Bayesian approach). In this note we will just assume that before the change-point the observations are IID (the change-point may be already the first observation) and after the change-point the observations cease to be IID.

Since our problem has so little structure, we will be able to prove only validity results: before the change-point our procedures do not raise alarms too often. The efficiency (raising an alarm soon after the change) is a topic of further research, as we discuss in Section 5.

So far the only method of change detection with the general IID assumption (or the *assumption of randomness*) as null hypothesis has been conformal change detection (see, e.g., [8]). The approach of this note is also based on conformal prediction but is simpler. On the negative side, our validity results will be weaker. For further details, see Section 4.

We start the main part of this note, in Section 2, from another conformal version of the Shiryaev–Roberts procedure and a simple statement about its asymptotic validity. As a corollary, in Section 3 we obtain the asymptotic validity of an analogous conformal version of Page's CUSUM procedure.

Informally (and formally in the proof of Proposition 1), this note is based on the idea of reversing the time, which is standard in conformal prediction [9, Section 8.7]. This is how we obtain the procedures that we call Roberts–Shiryaev (reversing Shiryaev–Roberts) and MUSUC (reversing CUSUM). However, for simplicity, in Section 2 we first present the Roberts–Shiryaev procedure in its pure form, and only later, after stating the validity result, explain connections with its prototype in the standard theory of change detection.

2 Roberts–Shiryaev procedure

Let $\mathbf{Z}$ be the *observation space* (a measurable space), $(\Omega, \mathcal{A}, \mathbb{P})$ be an underlying probability space (with the expectation operator $\mathbb{E}$), and $Z_1, Z_2, \dots$ be an IID sequence of $\mathbf{Z}$-valued random elements. We are interested in a sequence $z_1, z_2, \dots$ of elements of $\mathbf{Z}$ and interpret $Z_1, Z_2, \dots$ as our observations and $z_1, z_2, \dots$ as their realized values.

We will use the notation $\lbag z_1, \dots, z_n \rbag$ for a bag (also known as multiset) consisting of elements $z_1, \dots, z_n$. It will be regarded as an equivalence class of sequences $(z_1, \dots, z_n)$, where two sequences are defined to be equivalent when they can be obtained from each other by permuting their elements.

A *conformal e-predictor* is a measurable function f that maps any finite sequence $(z_1, \dots, z_m)$, for any $m \in \{1, 2, \dots\}$, to a finite sequence $(\alpha_1, \dots, \alpha_m)$

of nonnegative numbers of the same length with average 1,

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

that satisfies the following property of equivariance: for any $m \in \{2,3,\dots\}$, any permutation π of $\{1,\dots,m\}$, any $(z_1,\dots,z_m) \in \mathbf{Z}^m$, and any $(\alpha_1,\dots,\alpha_m) \in [0,\infty)^m$,

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

In terms of betting [4], f is our bet and α_i shows how strange z_i looks in the bag $\lbag z_1,\dots,z_m \rbag$; for a large m and under the assumption of exchangeability of $z_1,\dots,z_m$, we do not expect α_i to be large for a significant proportion of z_i. It will be convenient to abuse the notation by setting

$$f(\lbag z_1,\dots,z_m \rbag, z) := \alpha, \tag{1}$$

where α is the last element of the sequence

$$(\alpha_1,\dots,\alpha_m,\alpha) := f(z_1,\dots,z_m,z).$$

(It is clear that the α in (1) does not depend on the ordering of the sequence $(z_1,\dots,z_m)$.)

With each conformal e-predictor f we can associate the sequence of nonnegative random variables $E_1, E_2, \dots$, where

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

Intuitively, large values of these random variables are evidence against $Z_1, Z_2, \dots$ being IID.

The *Roberts–Shiryaev procedure* for nonrandomness detection is the sequence of stopping times $\sigma_0 := 0$ and

$$\sigma_k := \min\left\{ n > \sigma_{k-1} \mid \sum_{i=\sigma_{k-1}+1}^{n} E_{\sigma_{k-1}+1}\dots E_i \ge c \right\}, \quad k = 1,2,\dots, \tag{3}$$

where $c > 1$ is the parameter of the procedure (usually c is a large number). The idea behind this definition is that we raise alarms at times $\sigma_1, \sigma_2, \dots$ warning the user that the IID assumption may have become violated. If the IID assumption is in fact never violated, we do not want to raise (false) alarms too often. The following proposition is a simple statement of validity. Remember that the sequence of observations $Z_1, Z_2, \dots$ is assumed to be IID, and so all alarms are false.

Proposition 1. *Let A_n be the number of alarms*

$$A_n := \max\{k \mid \sigma_k \le n\} \tag{4}$$

raised by the Roberts–Shiryaev procedure (3) *after seeing the first* n *observations* $Z_1, \dots, Z_n$. *Then*

$$\limsup_{n\to\infty} \frac{A_n}{n} \le \frac{1}{c} \quad \textit{in probability.} \tag{5}$$

The conclusion (5) can be spelled out as

$$\forall \epsilon > 0 \, \exists N_0 \, \forall N \ge N_0 : \mathbb{P}\left(\frac{A_N}{N} > \frac{1}{c} + \epsilon\right) \le \epsilon. \tag{6}$$

Let us see how the Roberts–Shiryaev procedure is obtained, informally, by reversing its standard counterpart. The *conformal e-pseudomartingale* corresponding to the random variables (2) is

$$S_n := E_1 \dots E_n, \quad n = 0, 1, 2, \dots, \tag{7}$$

where S_0 is understood to be 1. It is not a genuine martingale since we only have $\mathbb{E}\, E_n = 1$ for all n instead of $\mathbb{E}(E_n \mid E_1, \dots, E_{n-1}) = 1$. It is not clear what properties of validity the Shiryaev–Roberts procedure would retain if applied to S_n.

Instead, we can choose a large N and apply the Shiryaev–Roberts procedure ([5, 3]; we will use the description in [8, (13)]) to the martingale

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

(we will see that it is indeed a martingale in the proof of Proposition 1). The Shiryaev–Roberts procedure applied to the reverse martingale T_n divides the time $\{1, 2, \dots\}$ into intervals (a, b) such that, roughly,

$$\sum_{i=a+1}^{b} \frac{T_a}{T_i} \approx c.$$

By definition, this can be rewritten as

$$E_a + E_a E_{a+1} + \dots + E_a \dots E_{b-1} \approx c,$$

which motivates our definition (3).

Proof of Proposition 1. Let us fix $\epsilon > 0$ and find an N_0 satisfying (6). According to [8, Proposition 4.4], for the conformal Shiryaev–Roberts procedure the inequality in (5) holds almost surely; therefore, it holds in probability. Examination of the proof shows that (6) holds for the general Shiryaev–Roberts procedure (the underlying positive martingale does not have to be a conformal martingale) and, moreover, (6) holds uniformly in that N_0 depends only on ϵ and nothing else. Let us choose such an N_0. Fix any $N \ge N_0$.

Now we use the idea of reversing the time formally. Let $\mathcal{F}_n$ be the σ-algebra generated by the bag $\wr Z_1, \dots, Z_{n-1} \int$ and observations $Z_n, \dots, Z_N$. (Formally, $\mathcal{F}_n$ is the smallest σ-algebra containing the sets

$$\{(Z_1, \dots, Z_{n-1}) \in A\}, \ \{Z_n \in A_n\}, \dots, \ \{Z_N \in A_N\},$$

where $A \subseteq \mathbf{Z}^{n-1}$ is a symmetric measurable set of sequences of $n-1$ observations and $A_n, \ldots, A_N \subseteq Z$ are measurable sets of observations.) We also allow $n = N+1$, in which case $\mathcal{F}_{N+1}$ is the σ-algebra generated by the bag $\wr Z_1, \ldots, Z_N \int$. Then $(E_n, \mathcal{F}_n)$, $n = N, \ldots, 1$, is a stochastic sequence (meaning that each E_n is $\mathcal{F}_n$-measurable); moreover, it is a *martingale ratio*, in the sense

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

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

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

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 martingale $T_{N+1}, \ldots, T_1$. It gives us the stopping times $\tau_0 := N+1$ and

$$\tau_k := \max\left\{ n < \tau_{k-1} \mid \sum_{i=n}^{\tau_{k-1}-1} E_n \ldots E_i \geq c \right\}, \quad k = 1, 2, \ldots, \tag{8}$$

where $\max \emptyset := 0$. Let A'_N be the largest k such that $\tau_k > 0$; in words, A'_N is the total number of alarms raised by the Shiryaev–Roberts procedure.

Notice that each set $\{\sigma_k + 1, \ldots, \sigma_{k+1}\}$ with $\sigma_{k+1} \leq N$ contains at least one stopping time τ_j. This can be deduced from

$$\sum_{i=\sigma_k+1}^{\sigma_{k+1}} E_{\sigma_k+1} \ldots E_i \geq c. \tag{9}$$

Indeed, let j be the largest number satisfying $\tau_j > \sigma_{k+1}$ (our goal is to show that $\tau_{j+1} \geq \sigma_k + 1$). The inequality (9) implies

$$\sum_{i=\sigma_k+1}^{\tau_j-1} E_{\sigma_k+1} \ldots E_i \geq \sum_{i=\sigma_k+1}^{\sigma_{k+1}} E_{\sigma_k+1} \ldots E_i \geq c,$$

which in combination with (8) implies, in turn, that indeed $\tau_{j+1} \geq \sigma_k + 1$.

The argument of the previous paragraph shows that $A_N \leq A'_N$. Therefore, the outer inequality in (6) holds once it holds for A'_N in place of A_N (which we know to be true). □

3 MUSUC procedure

A procedure that is even more popular than the Shiryaev–Roberts procedure in change detection is Page's CUSUM procedure [2], which can be obtained from Shiryaev–Roberts by replacing $\sum$ with max [8, Section 4]. The *MUSUC procedure* is the sequence of stopping times defined by $\sigma_0 := 0$ and (3) with max in place of $\sum$. Notice that this definition can be simplified by dropping the max. We can say, equivalently, that the MUSUC procedure is the sequence of stopping times $\sigma_0 := 0$ and

$$\sigma_k := \min\left\{n > \sigma_{k-1} \mid E_{\sigma_{k-1}+1} \dots E_n \ge c\right\}, \quad k = 1, 2, \dots. \tag{10}$$

Notice that, if (7) were a genuine martingale, the conditional probability that the inequality in (10) holds for some n would not exceed $1/c$; it is a version of Ville's inequality [6, p. 100]. But since (7) is not necessarily a martingale, we only have the following weaker statement analogous to Proposition 1.

Proposition 2. *Let A_n be the number* (4) *of alarms raised by the MUSUC procedure* (3) *after seeing the first n observations $Z_1, \dots, Z_n$. Then* (5) *holds.*

Proof. The usual relation between the CUSUM and Shiryaev–Roberts procedures with the same parameter c is that the latter raises alarms more often than the former (see, e.g., [8], proof of Corollary 4.3). This relation still holds for the MUSUC and Roberts–Shiryaev procedures (although it becomes slightly less obvious), which, in combination with Proposition 1, implies Proposition 2.

Let us check this relation. Formally, the relation is that $\sigma_k \le \sigma'_k$ for all k, where σ_k (resp. σ'_k) is the time of the kth alarm raised by Roberts–Shiryaev (resp. MUSUC). Suppose it does not hold and let k be the smallest number such that $\sigma_k > \sigma'_k$. It is obvious that $k > 1$. By definition, $\sigma_{k-1} \le \sigma'_{k-1}$. It is obvious that, in this case, $\sigma_{k-1} < \sigma'_{k-1}$. It is only possible if

$$\prod_{i=\sigma_{k-1}+1}^{\sigma'_{k-1}} E_i < 1 \tag{11}$$

(otherwise we would have $\sigma_k \le \sigma'_k$). However, (11) contradicts the definition of MUSUC; in this case we would have $\sigma'_{k-1} \le \sigma_{k-1}$. □

4 Comparison with methods based on conformal martingales

The only existing approach to detecting nonrandomness online is based on conformal prediction; see [9, Section 7.1] and [1, 8]. The approach of this paper is based on conformal e-prediction. The two approaches are very different, and it is unlikely that either of them will be better in all interesting applications. These are some differences:

- Design of conformal 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, $n = 1, 2, \ldots$, conformal martingales may use the previous p-values $p_1, \ldots, p_{n-1}$. Such dependence on the past is not allowed for conformal e-pseudomartingales.
- Conformal martingales are randomized (without randomization we only obtain conformal supermartingales) whereas conformal e-pseudomartingales do not require randomization (it is optional and not used in this note).

5 Conclusion

As discussed in Section 1, this note only establishes simple validity results. The efficiency, in the sense of raising an alarm soon after the change, is an interesting topic of further research, theoretical or experimental (simulation or empirical studies).

Another interesting direction is to establish non-asymptotic validity results.

Acknowledgments

This research was partially supported by Amazon, Astra Zeneca, and Stena Line.

References

[1] Valentina Fedorova, Alex Gammerman, Ilia Nouretdinov, and Vladimir Vovk. Plug-in martingales for testing exchangeability on-line, On-line Compression Modelling project (New Series), `http://alrw.net`, Working Paper 4, April 2012. Conference version: ICML 2012.

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

[3] S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8:411–430, 1966.

[4] Glenn Shafer. The language of betting as a strategy for statistical and scientific communication. Technical Report arXiv:1903.06991 [math.ST], arXiv.org e-Print archive, March 2019. To appear as discussion paper in the *Journal of the Royal Statistical Society A*, and to be read in September 2020.

[5] Albert N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, 8:22–46, 1963.

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

[7] Vladimir Vovk. Cross-conformal e-prediction, On-line Compression Modelling project (New Series), `http://alrw.net`, Working Paper 26, January 2020.

[8] Vladimir Vovk. Testing randomness, On-line Compression Modelling project (New Series), `http://alrw.net`, Working Paper 24, June 2019.

[9] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, New York, 2005.

A Improved Roberts–Shiryaev and MUSUC procedures

One disadvantage of the definition (3) of the Roberts–Shiryaev procedure is that, when $E_{\sigma_{k-1}+1}$ is zero, the next alarm will never be raised, $\sigma_k = \infty$. And it is clear that the efficiency of the procedure may be severely affected when $E_{\sigma_{k-1}+1}$ is very small.

The *modified Roberts–Shiryaev procedure* is the sequence of stopping times $\sigma_0 := 0$ and

$$\sigma_k := \min \left\{ n > \sigma_{k-1} \mid \max_{m \in \{\sigma_{k-1}+1,\dots,n\}} \sum_{i=m}^{n} E_m \dots E_i \geq c \right\}, \quad k = 1, 2, \dots.$$

Essentially the same proof establishes the validity of the modified Roberts–Shiryaev procedure.

Proposition 3. *Proposition 1 continues to hold for the modified Roberts–Shiryaev procedure.*

Proof. Only the end of the proof of Proposition 1 has to be modified. Let us check that each set $\{\sigma_k + 1, \dots, \sigma_{k+1}\}$ with $\sigma_{k+1} \leq N$ contains at least one stopping time τ_j. Instead of (9) we now have

$$\sum_{i=m}^{\sigma_{k+1}} E_m \dots E_i \geq c$$

for some $m \in \{\sigma_k + 1, \sigma_{k+1}\}$. If j is the largest number satisfying $\tau_j > \sigma_{k+1}$, now we have

$$\sum_{i=m}^{\tau_j - 1} E_m \dots E_i \geq \sum_{i=m}^{\sigma_{k+1}} E_m \dots E_i \geq c,$$

which implies $\tau_{j+1} \geq m \geq \sigma_k + 1$. □

The analogous modification of the MUSUC procedure becomes very similar to the original CUSUM procedure. The *modified MUSUC procedure* is $\sigma_0 := 0$ and

$$\sigma_k := \min \left\{ n > \sigma_{k-1} \mid \max_{m \in \{\sigma_{k-1}+1,\dots,n\}} E_m \dots E_i \geq c \right\}, \quad k = 1, 2, \dots.$$

Proposition 4. *Proposition 2 continues to hold for the modified MUSUC procedure.*

Proof. We will again check that $\sigma_k \le \sigma'_k$ for all k, where σ_k (resp. σ'_k) is the time of the kth alarm raised by the modified Roberts–Shiryaev (resp. modified MUSUC) procedure. Suppose this does not hold and let k be the smallest number such that $\sigma_k > \sigma'_k$. It is obvious that $k > 1$. By definition, $\sigma_{k-1} \le \sigma'_{k-1}$. It is obvious that, in this case, $\sigma_{k-1} < \sigma'_{k-1}$. By the definition of the modified MUSUC procedure,

$$E_m \dots E_{\sigma'_k} > c$$

for some $m \in \{\sigma'_{k-1} + 1, \sigma'_k\}$. This implies, by the definition of the modified Robert–Shiryaev procedure, $\sigma_k \le \sigma'_k$, a contradiction. □