



**Subject Areas:**

Sequential analysis, Optimal stopping.

**Keywords:**

Optimality, Hypothesis testing, Quickest detection, Change-point, False alarm, Average detection delay.

**Author for correspondence:**

Peter Johnson

e-mail:

[peter.johnson-3@manchester.ac.uk](mailto:peter.johnson-3@manchester.ac.uk)

# Detecting Changes in Real-Time Data: A User's Guide to Optimal Detection

P. Johnson<sup>1</sup>, J. Moriarty<sup>2</sup> and G. Peskir<sup>1</sup>

<sup>1</sup>School of Mathematics, University of Manchester, UK

<sup>2</sup>School of Mathematical Sciences, Queen Mary

University of London, UK

The real-time detection of changes in a noisily observed signal is an important problem in applied science and engineering. The study of parametric optimal detection theory began in the 1930s, motivated by applications in production and defence. Today this theory, which aims to minimise a given measure of detection delay under accuracy constraints, finds applications in domains including radar, sonar, seismic activity, global positioning, psychological testing, quality control, communications and power systems engineering.

This paper reviews developments in optimal detection theory and sequential analysis, including sequential hypothesis testing and change-point detection, in both Bayesian and classical (non-Bayesian) settings. For clarity of exposition we work in discrete time and provide a brief discussion of the continuous time setting, including recent developments using stochastic calculus. Different measures of detection delay are presented, together with the corresponding optimal solutions. We emphasise the important role of the signal-to-noise ratio and discuss both the underlying assumptions and some typical applications for each formulation.

## 1. Introduction

Methods for detecting changes as quickly as possible in a noisily observed signal have a wide array of applications including finance [1], cyber security [2], epidemiology [3], metrology [4], statistical diagnosis [5] and, recently, the detection of instability in power systems [6]. These problems have been studied since the 1930s. Motivated by applications in production and defence, they were cast as parametric constrained optimisation problems. To date a wide variety of formulations have been posed and solved, including both Bayesian and classical (non-Bayesian) settings, and in both continuous and discrete time. This theory now finds applications in domains including radar, sonar, seismic activity, global positioning, psychological testing, quality control, communications and power systems engineering.

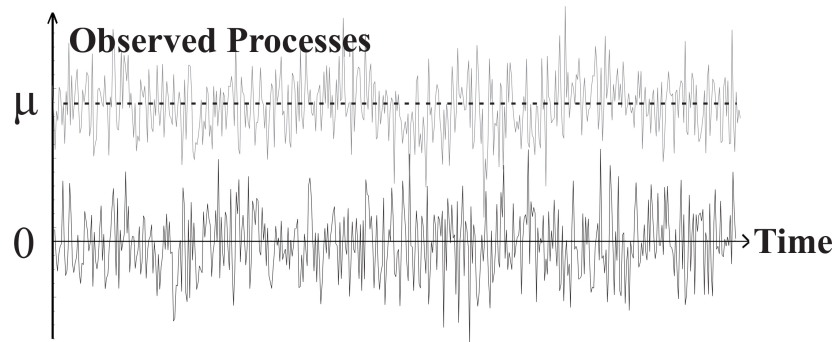
The most basic detection problem is that of *hypothesis testing*. In the simplest problem, a fixed dataset is given and the goal is to detect which of two distinct hypotheses has better statistical support. An optimal solution was provided in 1933 by Neyman and Pearson [7]. In the 1940s the field of sequential analysis emerged, in which the data is not fixed but is assumed to be observed sequentially. The goal is to decide between the two hypotheses as quickly as possible, while respecting constraints on detection accuracy. Wald and co-authors proposed the *sequential probability ratio test* in 1940s [8]. In this approach a test statistic is calculated at each step in the sequence. Depending on the value of this statistic it is decided to either accept one hypothesis or the other, or to seek more data by continuing to the next observation in the sequence. The optimality of this approach was proved in 1948 [9].

A natural extension is the *sequential change-point detection problem*, in which the process generating the data is assumed to change its probabilistic characteristics at an unknown *change-point* in the sequence. In this problem the challenge is to detect the change-point with minimal average delay under constraints on premature detection or *false alarms*. Such problems were considered in a classical setting by Shewhart in 1931 [10] then later by Girshick and Rubin in 1952 [11] and Page in 1954 [12]. In the *classical* approach the change is assumed to occur at a deterministic unknown time. This contrasts with the *Bayesian* approach, in which a prior probabilistic distribution is assigned to the change-point and then used to inform the constrained optimisation procedure. In 1963 Shiryaev studied this Bayesian formulation of the change-point detection problem [13]. The optimality of Page's approach was later established by Moustakides [14] (see also [15]) in the context of a formulation due to Lorden [16] in 1971.

In this paper we review techniques which can be shown to be optimal in the above problems, in the sense that they solve the constrained optimisation problem exactly. We also wish to emphasise the important role that the signal-to-noise ratio plays throughout these problems. For a clearer exposition we formulate the problems in discrete time and provide a separate discussion of the continuous time setting. While technically more demanding, the latter setting has the advantage that solutions may be obtained using stochastic calculus.

In line with the canonical work of Neyman and Pearson and the first attempts at change-point detection in 1931 by Shewhart [10], our running example is that of normally distributed data whose mean takes one of two possible values. We use this example to connect the problems and their solutions. Since both classical and Bayesian settings are presented, random variables will be denoted using bold font throughout this review. We do not assume prior knowledge of the material.

The paper is organised as follows. Section 2 considers the hypothesis testing problem and the sequential probability ratio test. Section 3 presents the change-point detection problem in various formulations, including possible formulations of average delay and constraints on false alarms. Discussion of the continuous time setting is provided in Section 4.



**Figure 1.** Observed processes from two possible hypotheses, one with a mean of zero ( $H_0$ ) and the other with a mean  $\mu$  different from zero ( $H_1$ ). In this scenario the underlying hypotheses are easily recognised in the data sets, however as  $\mu$  decreases towards zero (or the variance increases) the underlying hypothesis becomes more difficult to determine and so sequential analysis is required.

## 2. Hypothesis testing

A main distinction in detection theory is between *online* and *offline* testing. In offline testing the dataset is fixed and detection accuracy is the only concern. In online testing the data arrives in sequence to form a running (i.e. increasing) dataset and the average detection delay is also of primary importance.

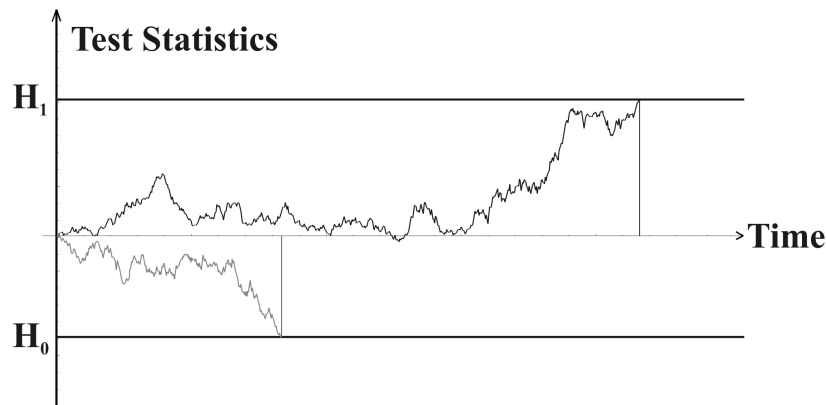
The observed data is assumed to be generated by a particular stochastic process. The probabilistic properties of this process are known, conditional on the value of an unknown parameter. In both the offline and online formulations we must decide which of two or more hypotheses about this parameter are best supported by the data. Figure 1 shows examples of observed stochastic processes generated with two different parameter values. They are simulations from a Gaussian white noise process with differing mean values, and this model will be used in a running example below.

The general methodological approach is to create a *sufficient statistic* (a test statistic) from the data which contains all information necessary to make a decision optimally. In online testing we therefore seek a *process* of sufficient statistics, updated with each new observation. This process of sufficient statistics is a stochastic process itself (see Fig. 2).

In the canonical setting we consider two hypotheses, the null hypothesis  $H_0$  and the alternative hypothesis  $H_1$ . It is apparent that there are two errors to be avoided in this setting. A *type 1 error* is to accept  $H_1$  when  $H_0$  is true. This is also known as a *false alarm* and in the following we shall denote the probability of false alarm by  $P_{FA}$ . Conversely a *type 2 error* occurs upon accepting  $H_0$  when  $H_1$  is true. This is a *missed detection* or a *false serenity*, the probability of which will be denoted  $P_{FS}$ . Hence the *probability of detection* is given by  $P_D = 1 - P_{FS}$ .

The offline case was solved by Neyman and Pearson in 1933, resulting in the classical Neyman-Pearson (NP) test [7]. For a fixed data set that has a known distribution under each hypothesis, the NP test was proven to be optimal in terms of maximising the probability of detection for a given probability of false alarm. When the distribution of the fixed or sequential data is known under each hypothesis this is known as a *simple hypothesis test*.

The NP test has been adapted to the online setting in the form of a sliding NP test. However the former is fundamentally designed to maximise the detection probability rather than to address detection delay. Indeed the sliding NP test has not been shown to be optimal in terms of minimising any useful notion of detection delay, such as those discussed in Section 3, under fixed false alarm constraints. We therefore begin consideration of online detection with the pioneering work of Wald on sequential analysis and the sequential probability ratio test (SPRT) in the 1940s.



**Figure 2.** An example of possible test statistics which can be compared to defined boundaries in order to detect which hypothesis the data is deemed to be derived from. In this case hitting the upper boundary would provide the decision that the data is derived from the alternative hypothesis  $H_1$ , while on hitting the lower boundary we accept  $H_0$ .

### (a) Sequential probability ratio testing

Sequential probability ratio testing (SPRT) originated in Abraham Wald's canonical work on sequential analysis in the 1940s. Building on the NP test, he derived a *sequential hypothesis test* which is designed to gather data until a desired level of confidence has been reached in one of the hypotheses.

The following example will be used throughout this paper. Consider a probability space  $(\Omega, \mathcal{F}, P)$  on which a stochastic process

$$\mathbf{x}[i] = \sigma \mathbf{w}[i] + \theta \mu, \quad i = 1, 2, \dots \quad (2.1)$$

is defined where it is assumed that  $\sigma > 0$  and  $\mu \neq 0$ . The discrete-time Gaussian white noise  $\mathbf{w}[i]$  (for  $i = 1, 2, \dots$ ) are defined to be independent and identically distributed (i.i.d.)  $N(0, 1)$  random variables. We set the task of choosing between the hypotheses

$$H_0: \quad \theta = 0 \quad (2.2)$$

$$H_1: \quad \theta = 1 \quad (2.3)$$

so that the generating processes under the two hypotheses are

$$H_0: \quad \mathbf{x}[i] = \sigma \mathbf{w}[i] \quad (2.4)$$

$$H_1: \quad \mathbf{x}[i] = \sigma \mathbf{w}[i] + \mu \quad (2.5)$$

for  $i \geq 1$ . These are the two processes depicted in Figure 1. In this example we are therefore aiming to detect the presence of a constant *signal*  $\mu$  with additional random *noise*  $\mathbf{w}$ , where under  $H_0$  only the white Gaussian noise is observed while under  $H_1$  the signal is also present. At each time  $i$ , Wald's SPRT statistic makes one of three decisions: to stop and decide that  $\theta = 0$ ; stop and decide that  $\theta = 1$ ; or to continue the observations.

It is clear from (2.4)-(2.5) that the hypothesis determines the probability distribution of the observations  $\mathbf{x}[i]$ , and hence their associated probability measure. In the following we will write  $P_0(\cdot) = P(\cdot|H_0)$  and  $P_1(\cdot) = P(\cdot|H_1)$ . Hence  $P_0(A)$  indicates the probability of an event  $A$  under the null hypothesis  $H_0$ , and  $P_1(A)$  indicates its probability under the alternative hypothesis  $H_1$ . Similarly we define the associated expectation operators  $E_0(\cdot) = E(\cdot|H_0)$  and  $E_1(\cdot) = E(\cdot|H_1)$ . Letting  $\mathbf{X}_n$  denote the random vector of the first  $n$  observations so that  $\mathbf{X}_n = (\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[n])$ , its (joint) probability density function under the two hypotheses will be denoted by  $f_0^n(\mathbf{X}_n)$  and

$f_1^n(\mathbf{X}_n)$  under  $H_0$  and  $H_1$  respectively. For simplicity throughout the paper the notation will be simplified by writing  $f_i(\mathbf{X}_n)$  instead of  $f_i^n(\mathbf{X}_n)$  under  $H_i$ , for  $i = 0, 1$ . By the assumption of independence we have

$$f_i(\mathbf{X}_n) = \prod_{k=1}^n f_i(\mathbf{x}[k]) \quad (2.6)$$

for  $i = 0, 1$ .

Before stating the main result of this section we introduce the *relative entropy* or *Kullback-Leibler distance* between the densities  $f_0$  and  $f_1$  by setting

$$D(f_0||f_1) = \int_{-\infty}^{\infty} f_0(x) \log \left( \frac{f_0(x)}{f_1(x)} \right) dx \quad (2.7)$$

and similarly

$$D(f_1||f_0) = \int_{-\infty}^{\infty} f_1(x) \log \left( \frac{f_1(x)}{f_0(x)} \right) dx. \quad (2.8)$$

**Theorem 2.1** (Sequential probability ratio test, cf. [17, Ch. 2]). *Let  $\mathbf{X}_n = (x[1], x[2], \dots, x[n])$  where  $x[i]$  are i.i.d. random variables with density either  $f_1$  or  $f_0$ . The likelihood ratio is then defined as*

$$L(\mathbf{X}_n) = \frac{f_1(\mathbf{X}_n)}{f_0(\mathbf{X}_n)}. \quad (2.9)$$

*Then the first entry time  $\tau$  to the stopping set  $D = (0, \gamma_0] \cup [\gamma_1, \infty)$  given by*

$$\tau = \inf\{n \geq 1 | L(\mathbf{X}_n) \in D\} \quad (2.10)$$

*with*

$$\gamma_0 = \frac{1 - P_D}{1 - P_{FA}} \quad \text{and} \quad \gamma_1 = \frac{P_D}{P_{FA}} \quad (2.11)$$

*minimises the expected test length  $E_i[\tau]$  for both  $i = 0$  and  $i = 1$ , with false alarm probability  $P_{FA}$  and detection probability  $P_D$ . The expected test lengths are approximately given by*

$$E_0[\tau] \approx D(f_0||f_1)^{-1} \left( (1 - P_{FA}) \log(\gamma_0^{-1}) - P_{FA} \log(\gamma_1) \right) \quad (2.12)$$

$$E_1[\tau] \approx D(f_1||f_0)^{-1} \left( P_D \log(\gamma_1) - (1 - P_D) \log(\gamma_0^{-1}) \right). \quad (2.13)$$

2(a)(i). *Signal-to-noise ratio.* In the case where the observed data is normally distributed as in (2.4) and (2.5) we have that  $D(f_0||f_1) = D(f_1||f_0) = \mu^2/2\sigma^2$ .

The ratio  $\mu^2/\sigma^2$  is often referred to as the *signal-to-noise ratio* (SNR) and it follows from (2.12)-(2.13) that in the Gaussian case, increasing the SNR decreases the average test length under both hypotheses. A practical consequence of this fact is that if the SNR can be increased (e.g. using signal processing techniques - see for example [18]) while preserving the Gaussian data output, then this will increase the average speed of detection. This ratio is also important in continuous-time sequential analysis due to the Gaussian nature of the Brownian motion (see Section 4).

More generally it can be seen that the relative entropies, or Kullback-Leibler distances, between the densities  $f_1$  and  $f_0$  are inversely proportional to the expected detection delays. This corresponds to the intuitive idea that two distributions may be distinguished more quickly when they are more distinct.

2(a)(ii). *Example.* In our running example of detecting a constant amplitude signal among Gaussian white noise, the likelihood ratio derived from the first  $n$  observations  $\mathbf{X}_n =$

$(\mathbf{x}[1], \mathbf{x}[2], \dots, \mathbf{x}[n])$  is

$$L(\mathbf{X}_n) = \prod_{i=1}^n \frac{f_1(\mathbf{x}[i])}{f_0(\mathbf{x}[i])} = \exp \left[ -\frac{1}{2\sigma^2} \left( -2\mu \sum_{i=1}^n \mathbf{x}[i] + n\mu^2 \right) \right]. \quad (2.14)$$

By Theorem 2.1, for each  $n \geq 1$  it is optimal to compare the value of  $L(\mathbf{X}_n)$  with the boundaries given by (2.11). Equivalently we may define

$$T(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{x}[i] \quad (2.15)$$

$$\tilde{\gamma}_i = \frac{\sigma^2}{n\mu} \log \gamma_i + \frac{\mu}{2} \quad (2.16)$$

for  $i = 0, 1$  and compare the *running mean* test statistic  $T(\mathbf{X}_n)_{n \geq 1}$  with the thresholds  $\tilde{\gamma}_0$  and  $\tilde{\gamma}_1$  as follows. At each time step ( $n = 1, 2, \dots$ ) one of the following decisions is made:

- i) if  $T(\mathbf{X}_n) \leq \tilde{\gamma}_0$ , stop observing and accept the hypothesis  $H_0$ ;
- ii) if  $T(\mathbf{X}_n) \geq \tilde{\gamma}_1$ , stop observing and accept the hypothesis  $H_1$ ;
- iii) if  $\tilde{\gamma}_0 < T(\mathbf{X}_n) < \tilde{\gamma}_1$ , continue to time step  $n + 1$ .

## (b) Bayesian formulation

In the Bayesian formulation of the hypothesis testing problem, the unknown value  $\theta$  is assumed to be a random variable taking the value 0 or 1 with a known distribution (and independent of the noise process  $(w[i])_{i \geq 1}$ ). In this setting we work with the prior probability measure  $P_\pi$ , which is derived from the probability measures  $P_i$  above for  $i = 0, 1$ . Writing  $\pi$  for the probability that  $\theta$  takes the value 1, and setting

$$P_0(\theta = 1) = 0 \quad (2.17)$$

$$P_1(\theta = 1) = 1 \quad (2.18)$$

we define the prior probability measure  $P_\pi(\cdot)$  as

$$P_\pi(\cdot) = \pi P_1(\cdot) + (1 - \pi) P_0(\cdot). \quad (2.19)$$

Under this new measure we have

$$\begin{aligned} P_\pi(\theta = 1) &= \pi P_1(\theta = 1) + (1 - \pi) P_0(\theta = 1) \\ &= \pi \end{aligned} \quad (2.20)$$

and similarly  $P_\pi(\theta = 0) = 1 - \pi$ .

Then we define the *posterior process* to be the probability of  $H_1$  (equivalently  $\theta = 1$ ) given the data we have observed

$$\mathbf{II}_n = P_\pi(H_1 | \mathbf{X}_n). \quad (2.21)$$

This posterior process can be related to the likelihood ratio at time  $n$ . For simplicity we indicate the derivation in the case that the observations are drawn from a discrete distribution; the derivation for continuous random variables is analogous. Using Bayes' formula  $P(A|B) = P(A)P(B|A)/P(B)$ , and the law of total probability  $P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$ , we have that

$$\mathbf{II}_n = \frac{P_\pi(H_1)P_\pi(\mathbf{X}_n|H_1)}{P_\pi(H_0)P_\pi(\mathbf{X}_n|H_0) + P_\pi(H_1)P_\pi(\mathbf{X}_n|H_1)} \quad (2.22)$$

$$= \frac{\pi P_\pi(\mathbf{X}_n|H_1)}{(1 - \pi)P_\pi(\mathbf{X}_n|H_0) + \pi P_\pi(\mathbf{X}_n|H_1)} \quad (2.23)$$

$$= \frac{\frac{\pi}{1 - \pi} L(\mathbf{X}_n)}{1 + \frac{\pi}{1 - \pi} L(\mathbf{X}_n)} \quad (2.24)$$

where the likelihood ratio at time  $n$  is defined as

$$L(\mathbf{X}_n) = \frac{P_\pi(\mathbf{X}_n|H_1)}{P_\pi(\mathbf{X}_n|H_0)}. \quad (2.25)$$

**Theorem 2.2** (Bayesian sequential testing, cf. [19, Ch. 4]). *Let  $\mathbf{X}_n = (x[1], x[2], \dots, x[n])$  where  $x[i]$  are i.i.d. random variables having density  $f_1$  with probability  $\pi$  and density  $f_0$  with probability  $1 - \pi$ . Let*

$$\mathbf{II}_n = \frac{\frac{\pi}{1-\pi}L(\mathbf{X}_n)}{1 + \frac{\pi}{1-\pi}L(\mathbf{X}_n)}, \quad \mathbf{II}_0 = \pi \quad (2.26)$$

with

$$L(\mathbf{X}_n) = \frac{f_1(\mathbf{X}_n)}{f_0(\mathbf{X}_n)}. \quad (2.27)$$

Then the first entry time  $\tau$  of  $\mathbf{II}_n$  to the stopping set  $D = (0, A] \cup [B, 1)$  given by

$$\tau = \inf\{n \geq 0 | \mathbf{II}_n \in D\} \quad (2.28)$$

minimises the expected test length  $E_i[\tau]$  for both  $i = 0$  and  $i = 1$ , where the constants  $0 < A < B < 1$  are chosen such that the required false alarm probability  $P_{FA}$  and detection probability  $P_D$  satisfy

$$P_0(\mathbf{II}_\tau = B) = P_{FA} \quad (2.29)$$

$$P_1(\mathbf{II}_\tau = A) = 1 - P_D. \quad (2.30)$$

For further details see [19, pp 172-180].

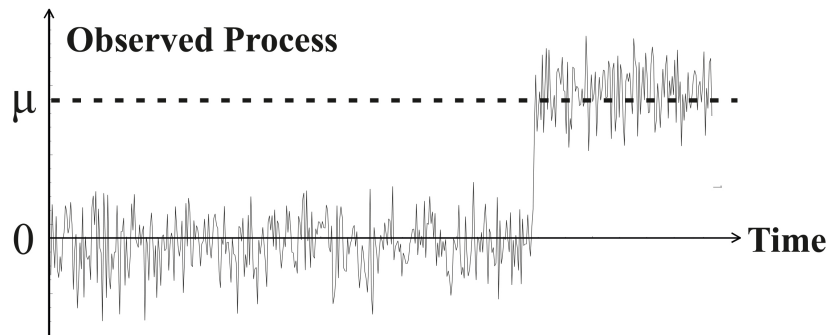
This Bayesian formulation generalises the *classical* formulation (2.14) by taking account of the prior probability measure  $P_\pi$ . In particular, if prior information about the value of  $\theta$  is available then the Bayesian formulation offers the advantage of including it in the analysis. It also allows the observer to monitor a running probability  $\mathbf{II}_n$  that the data is derived from the alternate hypothesis  $H_1$ . It may be checked that the classical formulation is recovered as a special case upon taking the *uninformed prior*  $\pi = 1/2$ , which favours both hypotheses equally.

Recent work has also used neural network techniques to estimate this posterior process, and hence test hypotheses for observed data with complicated dynamics [6].

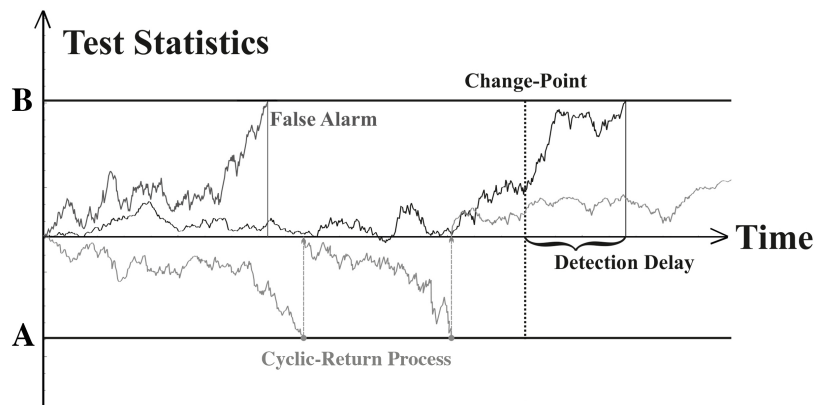
### 3 Change-point detection

In this paper we have real-time engineering applications in mind. Therefore, although detection problems may be formulated either online or offline (as for hypothesis testing), we consider only the online setting. In this setting we wish to find a *stopping time*  $\tau$  at which point we declare that there is sufficient evidence that a change-point has occurred. An *optimal stopping time* is then one which minimises a given measure of delay under certain false alarm constraints. This means that, given this performance criterion and the modelling assumptions, the method cannot be outperformed by any other method. In this section we will consider which procedures provide optimal stopping times for different measures of the delay and false alarm constraints.

In the change-point detection problem the observed data  $(x[1], x[2], \dots, x[i], \dots)$  is generated by the following variation on the process (2.1). There is now an unobservable *change-point* at time  $\nu$  such that for times  $i = 1, 2, \dots, \nu$  we have  $\theta = 0$  while for times  $i = \nu + 1, \nu + 2, \dots$  we have  $\theta = 1$ . This construction is illustrated in Figure 3. It is interesting to note that in the sequential hypothesis testing problem of the previous section there are just two hypotheses in total ( $H_0$  and  $H_1$ ). However in change-point detection, while only two hypotheses are considered at each time step  $i$ , there are effectively as many hypotheses as there are possible change-points: an interesting explanation of this is provided in [17, Sect. 2.6]. Detection algorithms then convert all the required information from the observed process into a test statistic. When this test statistic hits a corresponding boundary  $B$  then it is declared that a change has occurred. If this hitting time  $\tau$  occurs before a change has occurred then this is called a *false alarm* while if this happens after



**Figure 3.** An observed process with a change in mean, from zero to  $\mu$ , at an easily observable change-point. However as  $\mu$  decreases towards zero (or the variance increases) the change-point becomes more difficult to determine and so sequential analysis is required.



**Figure 4.** Three possible trajectories for test statistics in change-point detection problems. For the change-point shown, the black test statistic results in a detection with the indicated detection delay; the dark grey test statistic results in a false alarm; and the light grey test statistic (a cyclic-return process, see 3(a)) results in no detection over this time period.

the change point then the distance between these times is called the *detection delay* (see Fig. 4). In the classical approach  $\nu$  is assumed deterministic and unknown, while the Bayesian approach assumes  $\nu$  to be a random variable which is assigned a prior distribution.

We define the following probability measures. Let  $P^\nu(A)$  denote the probability of the event  $A$  given that the change-point takes the value  $\nu$ , with  $E^\nu$  the associated expectation and  $f^{\nu,n}$  the (joint) probability density function of  $\mathbf{X}_n = (x[1], x[2], \dots, x[n])$  under  $P^\nu$ . Hence  $P^0$  is the measure under which the change occurs at time zero (effectively before the first observation) and under  $P^\infty$  the change never occurs. Similarly to the previous sequential testing problems, we will simplify the notation of the joint densities  $f^{0,n}$ ,  $f^{\infty,n}$  and  $f^{\nu,n}$  which in the following will be denoted  $f^0$ ,  $f^\infty$  and  $f^\nu$  respectively. It can be noted that the sequential testing problem may be interpreted as a special case of this set-up, with just two possible change-points: namely,  $\nu = 0$  and  $\nu = \infty$ . As such it is possible to see that  $f_0 = f^\infty$  and  $f_1 = f^0$  and the likelihood ratio process from the sequential testing problem may be written

$$L(\mathbf{X}_n) = \frac{f^0(\mathbf{X}_n)}{f^\infty(\mathbf{X}_n)}. \quad (3.1)$$



In change-point detection, the problem is to minimise the average delay in detecting the change given defined constraints on the false alarms. We now make these terms precise.

### (i) Constraints on false alarms

In the Bayesian setting we may simply restrict the *probability of false alarm* ( $P_{FA}$ ) to lie below a prescribed level  $\alpha$ . Then we consider only those stopping times  $\tau$  that satisfy

$$P(\tau < \nu) \leq \alpha. \quad (3.2)$$

In the classical setting this false alarm probability depends on the deterministic unknown value of  $\nu$ , and so an alternate formulation must be used. Lorden [16] suggested bounding the *frequency of false alarms* by considering only those stopping times  $\tau$  which are on average greater than a parameter  $T$  if the change never occurs, i.e. requiring

$$E^\infty[\tau] \geq T. \quad (3.3)$$

There has also been some work which limits a certain conditional probability of false alarm during time intervals of a given length  $l$  (see [20], [21]), specifying for any  $k > 0$  that

$$P^\infty(k < \tau < k + l \mid \tau > k) \leq \alpha. \quad (3.4)$$

However this is a much more complicated problem.

### (ii) Measures of delay

Having specified the set of admissible stopping times  $\tau$  in the previous section, it remains to specify the concept of optimality. In general, if a false alarm does not occur then we define the detection delay as  $\tau - \nu$  and the goal is to minimise the *average detection delay*, which we will abbreviate ADD. That is, we seek to find a stopping time  $\tau^*$  such that  $\mathcal{J}(\tau^*) = \inf_{\tau} \mathcal{J}(\tau)$  given the relevant constraint from Section 3(i), where  $\mathcal{J}(\tau)$  is one of the ADD formulations defined below.

(1). In the Bayesian formulation the ADD may be defined simply as the average of  $\tau - \nu$  given that the change-point has occurred

$$\mathcal{J}_\pi(\tau) = E_\pi[\tau - \nu \mid \tau > \nu] \quad (3.5)$$

$$= E_\pi[(\tau - \nu)^+] / P_\pi(\tau \geq \nu) \quad (3.6)$$

where the second equality follows from Bayes' formula. A version of this constrained minimisation problem was solved in [13] when the prior distribution of the change-point  $\nu$  is geometric with the possible addition of a weight at zero (see Section 3(b)). The key points about the Bayesian formulation are:

- i) Provides a means to include useful information known about the change-point;
- ii) Provides a running probability of a change-point having occurred;
- iii) Can constrain the probability of false alarm as described in (3.2).

There are also a number of proposed classical (non-Bayesian) measures of delay where the change-point is seen as a deterministic unknown. In this classical setting it is not possible to minimise the probability of false alarm constraint (3.2) so often the frequency of false alarms are bounded as in (3.3).

(2). One classical ADD proposed is the *relative integral average detection delay* (RIADD)

$$\mathcal{J}_{RI}(\tau) = \sum_{\nu=0}^{\infty} \mathbb{E}^{\nu}[(\tau - \nu)^+] / \mathbb{E}^{\infty}[\tau]. \quad (3.7)$$

The RIADD formulation may also be obtained theoretically as a certain limit of the formulation (3.6), as the prior distribution of the change-point tends to the (improper) uniform distribution between zero and infinity. This means that in most cases a false alarm will be expected before the change-point occurs and so it may be preferable to allow a string of false alarms before the successful detection (see [22]). This makes the RIADD very suitable to situations where the cost of false alarms is low but long delays are costly. When the data is i.i.d. the RIADD is also equivalent to the stationary average detection delay (STADD) defined in [22, Thm. 2]. The key points about the RIADD formulation are:

- i) Good delay measure if the first change-point is as likely in the very distant future as it is now;
- ii) Assumes any change-point is equally likely;
- iii) Practically this assumption means that the change-point is likely to occur after one or more false alarms.

(3). In the classical formulation, Pollak and Siegmund (see [23], [24]) proposed to minimise the ADD for the *worst possible* change-point by setting

$$\mathcal{J}_{PS}(\tau) = \sup_{\nu} \mathbb{E}^{\nu}[\tau - \nu | \tau \geq \nu]. \quad (3.8)$$

Since this approach minimises the delay for the change-point at which the delay is the greatest on average (maximum), this is termed a *minimax formulation*. The key points about the Pollack-Siegmund formulation are:

- i) Assumes that the pre-change data/noise has no effect on the change-point (i.e. these are independent);
- ii) Takes into account the worst possible change-point;
- iii) Currently no known optimal procedure for any given constraint on the frequency of false alarms.

(4). At a similar time, Lorden proposed another minimax formulation [16] which additionally considered the most unfavourable pre-change data  $\mathbf{X}_{\nu}$  so that the ADD is formulated as

$$\mathcal{J}_L(\tau) = \sup_{\nu} (\text{ess sup } \mathbb{E}^{\nu}[\tau - \nu | \mathbf{X}_{\nu}]). \quad (3.9)$$

This means it is a good measure of delay to use if the pre-change data can affect the time the change-point appears. The key points about Lorden's formulation are:

- i) Good delay measure to consider if the pre-change data/noise affects the change-point (i.e. these are not independent);
- ii) Takes into account the worst possible change-point and pre-change data;
- iii) If pre-change data and change-point are independent this may be too pessimistic;

The formulations (3.5), (3.8) and (3.9) are related by the inequalities  $\mathcal{J}_{\pi}(\tau) \leq \mathcal{J}_{PS}(\tau) \leq \mathcal{J}_L(\tau)$  (see [25]), however this does not mean that one delay measure outperforms the others. Rather the Bayesian measure assumes that (a) the probability distribution of the change-point is known and (b) the change-point is independent from the pre-change data; in contrast the Pollak-Siegmund measure drops (a) while still retaining (b); finally Lorden's measure drops both (a) and (b). For further discussion see [25].

While (3.8) is more natural if the change-point is not affected by the pre-change data, currently there is no known optimal solution in the latter formulation under the constraint that  $E^\infty[\tau] \geq T$ . In contrast, it is known that Page's CUSUM procedure minimises  $\mathcal{J}_L(\tau)$  given the additional constraint  $E^\infty[\tau] \geq T$  (see [14]).

### (a) CUSUM algorithm

The CUSUM (CUmulative SUM) procedure was formulated by Page [12] shortly after Wald's work with the SPRT. In contrast with the SPRT there is a single boundary for the test statistic.

**Theorem 3.1** (CUSUM procedure, cf. [14]). *Let  $(\mathbf{x}[1], \dots, \mathbf{x}[\nu], \mathbf{x}[\nu+1], \dots)$  be a sequence of observed data where the  $\mathbf{x}[i]$  are independent random variables with probability density  $f^\infty$  for  $i \leq \nu$  and  $f^0$  for  $i > \nu$  but the change-point  $\nu$  is a deterministic unknown.*

*Define the sufficient statistic at time  $n$  by the recursive relation*

$$\mathbf{S}_n = \max(\mathbf{S}_{n-1}, 1) \frac{f^0(\mathbf{x}[n])}{f^\infty(\mathbf{x}[n])}, \quad \mathbf{S}_0 = 0 \quad (3.10)$$

*and the stopping time*

$$\tau = \inf\{n \geq 0 | \mathbf{S}_n \geq B\} \quad (3.11)$$

*where  $B$  is chosen as small as possible subject to the constraint*

$$E^\infty[\tau] \geq T. \quad (3.12)$$

*Then the first entry time  $\tau$  of  $\mathbf{S}_n$  above the boundary  $B$  is the optimal stopping time which minimises Lorden's delay measure (3.9) subject to the bounded frequency of false alarm constraint (3.12).*

It should be noted here that the optimal stopping time is not an estimate of the change-point but the point at which there is enough evidence that a change-point has already occurred. Offline change-point detection can then be used to estimate the exact change-point in the data. However the CUSUM procedure can also provide an estimate of the change-point, which is

$$\nu \approx \sup\{n < \tau | \mathbf{S}_n = 1\}. \quad (3.13)$$

The CUSUM procedure can be thought of in a few different ways. Firstly, it can be thought of as a sequential probability ratio test which restarts if the sufficient statistic  $L(\mathbf{X}_n)$  given in (2.9) falls below the level 1. This is a type of *cyclic return process*, a process that is returned to a given level upon hitting/crossing a boundary, which in this case is the level  $A = 1$  (see Fig. 4).

Secondly, it is also sometimes viewed as

$$\mathbf{S}_n = \frac{L(\mathbf{X}_n)}{\min_{0 \leq i \leq n} L(\mathbf{X}_i)} \quad (3.14)$$

whereby  $\mathbf{S}_n$  is the divergence of the SPRT sufficient statistic (3.1) from its current minimum. Finally, it can also be seen as a sliding SPRT or a reverse time SPRT starting from the latest data point. For further discussion on the CUSUM procedure see [17, Sect. 2.6].

3(a)(i). *Example.* In our Gaussian example, we see that by letting  $\mathbf{L}_n = (2\sigma^2/\mu) \log(\mathbf{S}_n)$  for  $n > 0$  we have

$$\mathbf{L}_n = \frac{2\sigma^2}{\mu} \max(\log \mathbf{S}_{n-1}, \log 1) + \frac{2\sigma^2}{\mu} \log \frac{f^0(\mathbf{x}[n])}{f^\infty(\mathbf{x}[n])} \quad (3.15)$$

$$= \max(\mathbf{L}_{n-1}, 0) + \mathbf{x}[n] - \mu/2 \quad (3.16)$$

$$= \max(\mathbf{L}_{n-1} + \mathbf{x}[n] - \mu/2, \mathbf{x}[n] - \mu/2). \quad (3.17)$$

This can be seen to be equivalent to the formulation often used

$$\mathbf{L}_n = \max(\mathbf{L}_{n-1} + \mathbf{x}[n] - \frac{\mu}{2}, 0), \quad \mathbf{L}_0 = 0. \quad (3.18)$$

### (b) Bayesian quickest detection

In Section 3(ii) we referred to the work [13] on the Bayesian quickest detection problem, in which the prior distribution on  $\nu$  is the geometric distribution with a weight  $\pi$  at zero. This choice is typical when one has knowledge of the average value of the change-point, but no further information. This follows since in the set of all distributions on the positive integers with a given mean, the geometric distribution has the greatest entropy and hence is the most uncertain. The probability measure associated to the quickest detection problem is then

$$\begin{aligned} P_\pi(\mathbf{X}_n \in A) & \quad (3.19) \\ &= \pi P^0(\mathbf{X}_n \in A) + (1 - \pi) \sum_{s=1}^{n-1} p(1-p)^{s-1} P^s(\mathbf{X}_n \in A) + (1 - \pi)(1-p)^{n-1} P^\infty(\mathbf{X}_n \in A). \end{aligned}$$

This expansion can be seen as the summation of probabilities of three scenarios: i) the change occurs at time zero, ii) the change occurs at a geometrically distributed time  $1 \leq s \leq n-1$ , and iii) the geometrically distributed change-point has not yet happened in the observed data  $\mathbf{X}_n$ .

**Theorem 3.2** (Bayesian quickest detection, cf. [19]). *Let  $(x[1], \dots, x[\nu], x[\nu+1], \dots)$  be a sequence of observed data where the  $x[i]$  are independent random variables with probability density  $f^\infty$  for  $i \leq \nu$  and  $f^0$  for  $i > \nu$ . The change-point  $\nu$  is a weighted geometrically distributed random variable with parameter  $p$  and a weighting of  $\pi$  at time zero. Define the sufficient statistic at time  $n$  by the recursive relation*

$$\mathbf{\Pi}_n = \frac{\mathbf{\Pi}_{n-1} f^0(\mathbf{x}[n]) + (1 - \mathbf{\Pi}_{n-1}) p f^0(\mathbf{x}[n])}{\mathbf{\Pi}_{n-1} f^0(\mathbf{x}[n]) + (1 - \mathbf{\Pi}_{n-1}) p f^0(\mathbf{x}[n]) + (1 - \mathbf{\Pi}_{n-1}) (1-p) f^\infty(\mathbf{x}[n])}, \quad \mathbf{\Pi}_0 = \pi \quad (3.20)$$

and the stopping time

$$\tau = \inf\{n \geq 0 : \mathbf{\Pi}_n \geq B\} \quad (3.21)$$

where

$$B = 1 - \alpha. \quad (3.22)$$

Then the first entry time  $\tau$  of  $\mathbf{\Pi}_n$  above the boundary  $B$  is the optimal stopping time which minimises the Bayesian (weighted-geometric) delay measure (3.5) subject to the probability of false alarm constraint

$$P_{FA} = \alpha. \quad (3.23)$$

In Theorem 3.2 the process  $(\mathbf{\Pi}_n)_{n \geq 1}$  is the running probability that the change-point has occurred given the information observed until time  $n$ , i.e.

$$\mathbf{\Pi}_n = P_\pi(\nu < n | \mathbf{X}_n) \quad (3.24)$$

which is also known as the *posterior probability process*. This is linked to the likelihood ratio  $L(\mathbf{X}_n)$  from (3.1), which has been present throughout these problems, via

$$\mathbf{\Pi}_n = \frac{\varphi_n}{1 + \varphi_n} \quad (3.25)$$

where

$$\varphi_n = \frac{L(\mathbf{X}_n)}{(1-p)^{n-1}} \left( \frac{\pi}{1-\pi} + p \sum_{k=1}^{n-1} \frac{(1-p)^{k-1}}{L(\mathbf{X}_k)} \right) \quad (3.26)$$

and  $\varphi_0 = \pi/(1-\pi)$  and the summation is defined to be equal to zero at  $n=1$ . It can be seen that

$$\varphi_n = \frac{1}{(1-p)} \frac{f^0(\mathbf{x}[n])}{f^\infty(\mathbf{x}[n])} (\varphi_{n-1} + p), \quad \varphi_0 = \frac{\pi}{1-\pi}. \quad (3.27)$$

The recursive formula (3.20) is then given by combining (3.25) and (3.27).

### (c) Shiryaev-Roberts procedure

When  $\pi \rightarrow 0$  and  $p \rightarrow 0$ , the weighted-geometric distribution used in the previous section may be interpreted as approaching an improper (i.e. non-normalised) uniform distribution on the positive integers. In the limit we obtain a *generalised Bayesian approach* in which the change is equally likely to occur at any time between 1 and  $+\infty$ . The solution in this case is provided by the *Shiryaev-Roberts* (SR) procedure.

**Theorem 3.3** (Shiryaev-Roberts procedure, cf. [22]). *Let  $(x[1], \dots, x[\nu], x[\nu + 1], \dots)$  be a sequence of observed data where the  $x[i]$  are independent random variables with density  $f^\infty$  for  $i \leq \nu$  and density  $f^0$  for  $i > \nu$  where the change-point  $\nu$  is a deterministic unknown value.*

*Define the sufficient statistic at time  $n$  by the recursive relation*

$$\mathbf{R}_n = (1 + \mathbf{R}_{n-1}) \frac{f^0(\mathbf{x}[n])}{f^\infty(\mathbf{x}[n])}, \quad \mathbf{R}_0 = 0 \quad (3.28)$$

*and the stopping time*

$$\tau = \inf\{n \geq 0 : \mathbf{R}_n \geq B\} \quad (3.29)$$

*where  $B$  is chosen as small as possible subject to the constraint*

$$\mathbf{E}^\infty[\tau] \geq T. \quad (3.30)$$

*Then the first entry time  $\tau$  of  $\mathbf{R}_n$  above the boundary  $B$  is the optimal stopping time which minimises the RIADD delay measure (3.7) subject to the frequency of false alarm constraint (3.30).*

In our Gaussian running example, it is known that the optimal boundary is equal to the lower bound on the frequency of false alarm constraint (see [26]), i.e.

$$B = T. \quad (3.31)$$

It can also be noted that the sufficient statistic for the SR procedure can be defined in terms of the likelihood ratio function  $L(\mathbf{X}_n)$  as follows

$$\mathbf{R}_n = \sum_{k=1}^n \frac{L(\mathbf{X}_n)}{L(\mathbf{X}_k)}. \quad (3.32)$$

3(c)(i). *Relationship with Bayesian setting.* It can be seen that starting with the Bayesian delay measure and sufficient statistic from (3.26) one has

$$\frac{\varphi_n}{p} \rightarrow \mathbf{R}_n \quad (3.33)$$

as  $p \rightarrow 0$  and  $\pi \rightarrow 0$ . It is also known that

$$\frac{\mathbf{E}_\pi[(\tau - \nu)^+]}{p} \rightarrow \sum_{\nu=0}^{\infty} \mathbf{E}^\nu[(\tau - \nu)^+] \quad (3.34)$$

$$\frac{\mathbf{P}_\pi(\tau \geq \nu)}{p} \rightarrow \mathbf{E}^\infty[\tau] \quad (3.35)$$

and hence the delay measure

$$\mathcal{J}_\pi(\tau) \rightarrow \mathcal{J}_{RI}(\tau) \quad (3.36)$$

as  $p \rightarrow 0$  and  $\pi \rightarrow 0$ .

3(c)(ii). *The SR-r procedure.* The Shiryaev-Roberts procedure with a starting point  $R_0 = r$  such that

$$\mathbf{R}_n = (1 + \mathbf{R}_{n-1}) \frac{f^0(\mathbf{x}[n])}{f^\infty(\mathbf{x}[n])}, \quad \mathbf{R}_0 = r \quad (3.37)$$

is also referred to as the *SR-r procedure*. In [24] it is shown that by letting  $\mathbf{r}$  be a defined random variable the SR-r procedure is asymptotically optimal in minimising the Pollak-Siegmund delay metric as the frequency of false alarm constraint  $T \rightarrow \infty$ . However, a counter example in [27] showed that this procedure is not strictly optimal for any given lower bound on the frequency of false alarms.

## 4 Continuous time processes

The above problems have natural analogues in the continuous time setting, when time is indexed by the non-negative real numbers. In our white noise example the process (2.1) corresponds naturally to Brownian motion  $(\mathbf{B}_t)_{t \geq 0}$ , the fundamental continuous time stochastic process, and in the hypothesis testing problem we then have:

$$H_0: \mathbf{X}_t = \sigma \mathbf{B}_t \quad (4.1)$$

$$H_1: \mathbf{X}_t = \sigma \mathbf{B}_t + \mu t. \quad (4.2)$$

The hypotheses therefore differ through the presence or absence of the *drift* term  $\mu t$ . In the analogous change-point problem we have

$$\mathbf{X}_t = \mathbf{B}_t + \mu(t - \nu)I(t \geq \nu). \quad (4.3)$$

The optimal procedures for the Bayesian hypothesis testing problem and the constrained quickest detection problems have also been solved in the setting of a Brownian motion with a possible (constant) drift. The Bayesian hypothesis testing problem was solved in [28], the Bayesian quickest detection problem in [29], Lorden's constrained problem in [30] and [31], and the relative integral average delay formulation in [32].

The continuous time setting can serve as an approximation to the discrete time setting (or *vice versa*), but has the advantage that methods of continuous-time stochastic calculus may be applied. This has led to further progress on problems with more general observed processes in continuous time, which we now mention briefly.

4(i). *General diffusion processes.* Consider the continuous-time observed process

$$d\mathbf{X}_t = \sigma(\mathbf{X}_t)d\mathbf{B}_t + (\mu_0(\mathbf{X}_t) + \theta(\mu_1(\mathbf{X}_t) - \mu_0(\mathbf{X}_t))) dt, \quad \mathbf{X}_0 = x \quad (4.4)$$

for the hypothesis testing problem, or

$$d\mathbf{X}_t = \sigma(\mathbf{X}_t)d\mathbf{B}_t + (\mu_0(\mathbf{X}_t) + I(t \geq \nu)(\mu_1(\mathbf{X}_t) - \mu_0(\mathbf{X}_t))) dt, \quad \mathbf{X}_0 = x \quad (4.5)$$

in the case of change-point detection.

These observation processes have been considered in the Bayesian framework (see [33] and [34]). In this setting the SNR is also important and is similarly defined as  $(\mu_1(x) - \mu_0(x))^2 / \sigma(x)^2$  (e.g. see [33, Eq. 2.16]). When the SNR is constant the one-dimensional theory of *optimal stopping* may be applied. Non-constant SNR processes give rise to 2-dimensional sufficient statistics, which depend on both the posterior process and the current position of the observed process, making their analysis more challenging. However recent breakthroughs have been made for certain processes with non-constant SNR (see [35] and [36]). For a general discussion of this problem and the causes of increased dimensionality see [35, Sect. 2] and [36, Sect. 3].

4(ii). *Further results.* The above problems have also been considered with a finite time horizon (see [37] and [38]), exponential time-costs (see [39], [34] and [40]), and in the context of multiple hypothesis testing (see [41] and [42]). Discontinuous processes in continuous time have been

studied, see for example [43], [44] and [45]. The first two papers deal with a continuous time counting process (a Poisson process), which is a natural model in the problem of detecting nuclear material against noise from background radiation. Further extensions include problems where the change-point depends on the observed trajectory (termed *self-exciting* processes, see [46]) and problems in which the change occurs when the observed process hits an unobservable random level in space rather than in time (see [47] and [48]). Further details may also be found in [49], [26] and [50] and, specifically for quickest detection problems, in [51].

**Authors' Contributions.** This review is a collection of the experience of all three authors, who together drafted and revised the manuscript and approved the final text.

**Funding.** This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grants EP/K00557X/2 and EP/M507969/1.

**Acknowledgements.** The authors would like to thank their industrial collaborators for valuable discussions over many years, and also the referees for insightful comments which improved the presentation of the paper.

## References

1. Shiryaev AN.  
Quickest detection problems in the technical analysis of the financial data.  
In: Math. Fin. - Bachelier Cong. 2000. Springer; 2002. p. 487–521.
2. Polunchenko AS, Tartakovsky AG, Mukhopadhyay N.  
Nearly optimal change-point detection with an application to cybersecurity.  
Sequential Anal. 2012;31(3):409–435.
3. Yu X, Baron M, Choudhary PK.  
Change-point detection in binomial thinning processes, with applications in epidemiology.  
Sequential Anal. 2013;32(3):350–367.
4. Zucca C, Tavella P, Peskir G.  
Detecting atomic clock frequency trends using an optimal stopping method.  
Metrologia. 2016;53(3):S89.
5. Brodsky E, Darkhovsky BS.  
Non-Parametric Statistical Diagnosis: Problems and Methods.  
Springer Science & Business Media; 2000.
6. Gonzalez J, Kitapbayev Y, Guo T, Milanovic JV, Peskir G, Moriarty J.  
Application of sequential testing problem to online detection of transient stability status for power systems.  
In: Decis. Control (CDC), 2016 IEEE 55th Conf.; 2016. p. 1536–1541.
7. Neyman J, Pearson ES.  
On the problem of the most efficient tests of statistical hypotheses.  
Philos Trans R Soc A. 1933;231(694-706):289–337.
8. Wald A.  
Sequential tests of statistical hypotheses.  
Ann Math Stat. 1945;16(2):117–186.
9. Wald A, Wolfowitz J.  
Optimum character of the sequential probability ratio test.  
Ann Math Stat. 1948;19(3):326–339.
10. Shewhart WA.  
Economic Control of Quality of Manufactured Product.  
D. Van Nostrand Company, Inc., New York; 1931.
11. Girshick MA, Rubin H.  
A Bayes approach to a quality control model.  
Ann Math Stat. 1952;23(1):114–125.
12. Page ES.  
Continuous inspection schemes.  
Biometrika. 1954;41(1):100–115.
13. Shiryaev A.  
On optimum methods in quickest detection problems.  
Theory Probab Appl. 1963;8(1):22–46.

14. Moustakides GV.  
Optimal stopping times for detecting changes in distributions.  
Ann Stat. 1986;14(4):1379–1387.
15. Ritov Y.  
Decision theoretic optimality of the CUSUM procedure.  
Ann Stat. 1990;18(3):1464–1469.
16. Lorden G.  
Procedures for reacting to a change in distribution.  
Ann Math Stat. 1971;42(6):1897–1908.
17. Siegmund D.  
Sequential analysis: tests and confidence intervals.  
Springer Science & Business Media; 2013.
18. Kay SM.  
Fundamentals of Statistical Signal Processing: Detection Theory. vol. 2.  
Prentice-Hall; 1998.
19. Shiryaev A.  
Optimal Stopping Rules.  
Springer; 1978.
20. Tartakovsky AG.  
Asymptotic performance of a multichart CUSUM test under false alarm probability constraint.  
In: Proc. 44th IEEE Conf. Decis. Control. Eur. Control Conf.; 2005. p. 320–325.
21. Tartakovsky AG.  
Discussion on "Is average run length to false alarm always an informative criterion?" by Yajun Mei.  
Sequential Anal. 2008;27(4):396–405.
22. Pollak M, Tartakovsky AG.  
On optimality properties of the Shiryaev-Roberts procedure.  
Stat Sin. 2009;19(4):1729–1739.
23. Pollak M, Siegmund D.  
Approximations to the expected sample size of certain sequential tests.  
Ann Stat. 1975;3(6):1267–1282.
24. Pollak M.  
Optimal detection of a change in distribution.  
Ann Stat. 1985;13(1):206–227.
25. Moustakides GV.  
Sequential change detection revisited.  
Ann Stat. 2008;36(2):787–807.
26. Shiryaev A.  
Quickest detection problems: Fifty years later.  
Sequential Anal. 2010;29(4):345–385.
27. Polunchenko AS, Tartakovsky AG.  
On optimality of the Shiryaev-Roberts procedure for detecting a change in distribution.  
Ann Stat. 2010;38(6):3445–3457.
28. Mikhalevich VS.  
A Bayes test of two hypotheses concerning the mean of a normal process.  
Visnik Kiiv Univ. 1958;1(101-104).
29. Shiryaev A.  
Two problems of sequential analysis.  
Cybernetics. 1967;3(2):63–69.
30. Shiryaev A.  
Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time.  
Russ Math Surv. 1996;51(4):750–751.
31. Beibel M.  
A note on Ritov's Bayes approach to the minimax property of the cusum procedure.  
Ann Stat. 1996;24(4):1804–1812.
32. Feinberg EA, Shiryaev A.



- Quickest detection of drift change for Brownian motion in generalized Bayesian and minimax settings.  
Stat Decis. 2006;24(4):445–470.
33. Gapeev P, Shiryaev A.  
On the sequential testing problem for some diffusion processes.  
Stochastics. 2011;83(4-6):519–535.
  34. Gapeev P, Shiryaev A.  
Bayesian quickest detection problems for some diffusion processes.  
Adv Appl Probab. 2013;45(1):164–185.
  35. Johnson P, Peskir G.  
Sequential testing problems for Bessel processes.  
Trans Amer Math Soc. 2016;(to appear).
  36. Johnson P, Peskir G.  
Quickest detection problems for Bessel processes.  
Ann Appl Probab. 2016;(to appear).
  37. Gapeev P, Peskir G.  
The Wiener sequential testing problem with finite horizon.  
Stoch and Stoch Rep. 2004;76(1):59–75.
  38. Gapeev P, Peskir G.  
The Wiener disorder problem with finite horizon.  
Stoch Process Appl. 2006;116(12):1770–1791.
  39. Poor HV.  
Quickest detection with exponential penalty for delay.  
Ann Stat. 1998;26(6):2179–2205.
  40. Buonaguidi B, Muliere P.  
A note on some sequential problems for the equilibrium value of a Vasicek process.  
Pioneer J Theor Appl Stat. 2012;4(2):101–116.
  41. Novikov A.  
Optimal sequential multiple hypothesis tests.  
Kybernetika. 2009;45(2):309–330.
  42. Zhitlukhin MV, Shiryaev A.  
A Bayesian sequential testing problem of three hypotheses for Brownian motion.  
Stat Risk Model. 2011;28(3):227–249.
  43. Peskir G, Shiryaev A.  
Sequential testing problems for Poisson processes.  
Ann Stat. 2000;28(3):837–859.
  44. Peskir G, Shiryaev A.  
Solving the Poisson disorder problem.  
In: Adv. Financ. Stochastics. Springer; 2002. p. 295–312.
  45. Buonaguidi B, Muliere P.  
Sequential testing problems for Lévy processes.  
Sequential Anal. 2013;32(1):47–70.
  46. Aliev A.  
Disorder problem for self-exciting process.  
Theory Probab Appl. 2013;57(3):512–520.
  47. Peskir G.  
Optimal detection of a hidden target: The median rule.  
Stoch Process Appl. 2012;122(5):2249–2263.
  48. Peskir G.  
Quickest detection of a hidden target and extremal surfaces.  
Ann Appl Probab. 2014;24(6):2340–2370.
  49. Peskir G, Shiryaev A.  
Optimal Stopping and Free-Boundary Problems.  
Lectures in Mathematics, ETH Zürich, Birkhäuser; 2006.
  50. Polunchenko AS, Tartakovsky AG.  
State-of-the-art in sequential change-point detection.  
Methodol Comput Appl Probab. 2012;14(3):649–684.
  51. Poor HV, Hadjiliadis O.

Quickest Detection.  
Cambridge University Press; 2009.