Detection of a Stopping Time Through Noisy Observations

Urs Niesen[1] and Aslan Tchamkerten[2]

[1] Massachusetts Institute of Technology,
Cambridge, USA
uniesen@mit.edu

[2] Telecom ParisTech,
Paris, France
aslan.tchamkerten@telecom-paristech.fr

Abstract. A novel quickest detection setting is proposed, generalizing the Bayesian change-point detection model. Suppose that $\{(X_i, Y_i)\}_{i\geq 1}$ is a sequence of pairs of random variables, and that S is a stopping time with respect to $\{X_i\}_{i\geq 1}$. The problem is to find a stopping time T with respect to $\{Y_i\}_{i\geq 1}$ that optimally tracks S, in the sense that T minimizes the expected *reaction delay* $\mathbb{E}(T-S)^+$, while keeping the *false-alarm probability* $\mathbb{P}(T<S)$ below a given threshold $\alpha \in [0,1]$. This problem formulation applies in several areas, such as in communication, detection, forecasting, and quality control.

Our results apply to the situation where the X_i's and Y_i's take values in finite alphabets and where S is bounded by some positive integer κ. By using elementary methods based on the analysis of the tree structure of stopping times, we exhibit an algorithm that computes the optimal average reaction delays for all $\alpha \in [0,1]$, and constructs the associated optimal stopping times T. Under certain conditions on $\{(X_i, Y_i)\}_{i\geq 1}$ and S, the algorithm running time is shown to be polynomial in κ.

1 Problem Statement

The tracking stopping time (TST) problem is defined as follows. Let $\{(X_i, Y_i)\}_{i\geq 1}$ be a sequence of pairs of random variables. Alice observes $X_1, X_2, \ldots$ and chooses a stopping time (s.t.) S with respect to that sequence. Knowing the distribution of $\{(X_i, Y_i)\}_{i\geq 1}$ and the stopping rule S, but having access only to the Y_i's, Bob wishes to find a s.t. that gets as close as possible to Alice's. Specifically, Bob aims to find a s.t. T with respect to $\{Y_i\}_{i\geq 1}$ minimizing the expected reaction delay $\mathbb{E}(T-S)^+ \triangleq \mathbb{E}\max\{0, T-S\}$, while keeping the false-alarm probability $\mathbb{P}(T<S)$ below a certain threshold $\alpha \in [0,1]$.

Example 1. **Forecasting**

A large manufacturing machine breaks down as soon as its cumulative fatigue hits a certain threshold. Knowing that a machine replacement takes, say, ten days, the objective is to order a new machine so that it is operational at the time the old machine breaks down. This prevents losses due to an interrupted manufacturing process as well as storage costs caused by an unused backup machine.

The problem of determining the operating start date of the new machine can be formulated as follows. Let X_n be the cumulative fatigue of the current machine up to day n, and let S denote the first day n that X_n crosses the critical fatigue threshold. Since the replacement period is ten days, the first day T a new machine is operational can be scheduled only on the basis of a (possibly randomized) function of $\{X_i\}_{i=1}^{T-10}$. By defining Y_i to be equal to X_{i-10} if $i > 10$ and else equal to zero, the day T is now a s.t. with respect to $\{Y_i\}_{i\geq 1}$, and we can formulate the requirement on T as aiming to minimize $\mathbb{E}(T-S)^+$ while keeping $\mathbb{P}(T<S)$ below a certain threshold.

The TST setting generalizes the Bayesian version of the change-point detection problem, a long studied problem dating back to the 1940's with applications to industrial quality control (Anscombe et al., 1947; Shiryaev, 1963; Yakir, 1994; Lai, 1998; Moustakides, 2008; Tartakovsky and Veeravalli, 2005). The Bayesian change-point problem is formulated as follows. Let θ be a random variable taking values in the positive integers. Let $\{Y_i\}_{i\geq 1}$ be a sequence of random variables such that, given the value of θ, the conditional probability of Y_n given $Y^{n-1} \triangleq \{Y_i\}_{i=1}^{n-1}$ is $P_0(\cdot|Y^{n-1})$ for $n < \theta$ and is $P_1(\cdot|Y^{n-1})$ for $n \geq \theta$. We are interested in a s.t. T with respect to the Y_i's minimizing the change-point

This work was supported in part by NSF under Grant No. CCF-0515122, by DoD MURI Grant No. N00014-07-1-0738, and by a University IR&D Grant from Draper Laboratory. The material presented here is part of the recently published journal article Niesen and Tchamkerten (2009).

reaction delay $\mathbb{E}(T-\theta)^+$, while keeping the false-alarm probability $\mathbb{P}(T<\theta)$ below a certain threshold $\alpha \in [0,1]$.

To see that the Bayesian change-point problem can be formulated as a TST problem, it suffices to define the sequence of binary random variables $\{X_i\}_{i\geq 1}$ such that $X_i = 0$ if $i < \theta$ and $X_i = 1$ if $i \geq \theta$, and to let $S \triangleq \inf\{i : X_i = 1\}$ (i.e., $S = \theta$). The change-point problem defined by θ and $\{Y_i\}_{i\geq 1}$ becomes the TST problem defined by S and $\{(X_i, Y_i)\}_{i\geq 1}$. However, it can be shown that the TST problem cannot, in general, be formulated as a Bayesian change-point problem.

2 Main results

As is argued in the last section, the TST problem is a generalization of the Bayesian change-point problem, which itself is analytically tractable only in special cases. This makes an analytical treatment of the general TST problem difficult. Instead, we present an algorithmic solution to this problem. Let $\{(X_i, Y_i)\}_{i\geq 1}$ be a discrete-time process where the X_i's and Y_i's take value in some finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let S be a s.t. with respect to $\{X_i\}_{i\geq 1}$ such that $S \leq \kappa$ almost surely for some constant $\kappa \geq 1$. We aim to find for any $\alpha \in [0,1]$

$$d(\alpha) \triangleq \min_{\substack{T:\mathbb{P}(T<S)\leq\alpha \\ T\leq\kappa}} \mathbb{E}(T-S)^+$$

where the minimization is over all (possibly randomized) s.t.'s T with respect to $\{Y_i\}_{i\geq 1}$. The restriction $T \leq \kappa$ induces no loss of generality.

It can be shown that the function $d(\alpha)$ is convex and piecewise linear, with break-points $\{\alpha_m, d_m\}_{m=1}^M$ achieved by non-randomized s.t.'s. Its typical shape is depicted in Figure 1. Defining

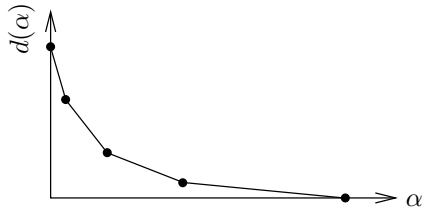

Fig. 1. Typical shape of the expected delay $d(\alpha)$ as a function of false-alarm probability α. The break-points are achieved by non-randomized stopping times.

for $\lambda \geq 0$ the Lagrangian

$$J_\lambda(T) \triangleq \mathbb{E}(T-S)^+ + \lambda\mathbb{P}(T<S),$$

we have by duality

$$d(\alpha) = \sup_{\lambda\geq 0} \min_{T\leq\kappa} \left(J_\lambda(T) - \lambda\alpha\right), \tag{1}$$

where the minimization here can be restricted to non-randomized s.t.'s.

Now, to any non-randomized s.t. T, we associate a unique $|\mathcal{Y}|$-ary tree $\mathcal{T}$ (i.e., all the nodes of $\mathcal{T}$ have either zero or exactly $|\mathcal{Y}|$ children) having each node specified by some y^n. Similarly, to each such tree $\mathcal{T}$, we can associate a unique non-randomized s.t. $T(\mathcal{T})$. To solve the TST problem, we propose a tree pruning algorithm similar to the CART algorithm for the construction of classification and regression trees (see Breiman et al., 1984). More precisely, for a given s.t. S, the algorithm constructs a sequence of s.t.'s $\{T(\mathcal{T}^m)\}_{m=0}^M$ and Lagrange multipliers $\{\lambda_m\}_{m=0}^M$ such that the s.t's $\{T(\mathcal{T}^m)\}_{m=1}^M$ achieve the break points of $d(\alpha)$, and the $\{\lambda_m\}_{m=0}^M$ are the slopes at the corresponding points.

We show that the worst case running time of the proposed algorithm is $\exp(O(\kappa))$. This is to be compared, for instance, with exhaustive search that has a $\Omega(\exp\exp(\kappa))$ running time. Moreover, we show that for a certain class of s.t. (essentially exchangeable s.t.'s) and i.i.d. processes, the running time of the algorithm is only *polynomial* in κ, and hence the optimal s.t. T can be found efficiently.

Example 2. Let $\{(X_i, Y_i)\}_{i\geq 1}$ be i.i.d. with the X_i's taking values in $\{0,1\}$. Consider the s.t. $S \triangleq \inf\{i : X_i = 1\}$. This example satisfies the aforementioned conditions, and hence the proposed algorithm has only polynomial running time in κ.

The problem considered in the last example is actually a Bayesian change-point problem. Here the change-point $\Theta \triangleq S$ has distribution $\mathbb{P}(\Theta = n) = p(1-p)^{n-1}$, where $p \triangleq \mathbb{P}(X = 1)$. The conditional distribution of Y_i given Θ is

$$\mathbb{P}(Y_i = y_i|\Theta = n) = \begin{cases} \mathbb{P}(Y_i = y_i|X_i = 0) & \text{if } i < n, \\ \mathbb{P}(Y_i = y_i|X_i = 1) & \text{if } i = n, \\ \mathbb{P}(Y_i = y_i) & \text{if } i > n. \end{cases}$$

Note that, unlike the case considered by Shiryaev, the distribution of the process at the change-point differs from the ones before and after it. We now give an example that cannot be formulated as a change-point problem.

Example 3. Let $\{(X_i, Y_i)\}_{i\geq 1}$ be i.i.d. where the X_i's and Y_i's take values in $\{0,1\}$, and let $S \triangleq \inf\{i \geq 1 : \sum_{j=1}^{i} X_j = 2\}$. In this case, it can again be shown that the proposed algorithm has only polynomial running time in κ.

3 An algorithm for computing $d(\alpha)$

We first establish a few preliminary results later used to evaluate $\min_T J_\lambda(T)$. Emphasis is put on the finite tree representation of bounded s.t.'s with respect to finite alphabet processes. We then provide an algorithm that computes the entire curve $d(\alpha)$. Due to space constraints, we present results without proofs. These can be found in the full version of the paper (Niesen and Tchamkerten, 2009).

We introduce a few notational conventions. The set $\mathcal{Y}^*$ represents all finite sequences over $\mathcal{Y}$. An element in $\mathcal{Y}^*$ is denoted either by y^n or by $\boldsymbol{y}$, depending on whether or not we want to emphasize its length. To any non-randomized s.t. T, we associate a unique $|\mathcal{Y}|$-ary tree $\mathcal{T}$ having each node specified by some $\boldsymbol{y} \in \mathcal{Y}^*$, where $\rho\boldsymbol{y}$ represents the vertex path from the root ρ to the node $\boldsymbol{y}$. A node $y^n \in \mathcal{T}$ is a leaf if $\mathbb{P}(T = n|Y^n = y^n) = 1$. We denote by $\mathcal{L}(\mathcal{T})$ the leaves of $\mathcal{T}$ and by $\mathcal{I}(\mathcal{T})$ the intermediate (or non-terminal) nodes of $\mathcal{T}$. The notation $T(\mathcal{T})$ is used to denote the (non-randomized) s.t. T induced by the tree $\mathcal{T}$. Given a node $\boldsymbol{y}$ in $\mathcal{T}$, let $\mathcal{T}_{\boldsymbol{y}}$ be the subtree of $\mathcal{T}$ rooted in $\boldsymbol{y}$. Finally, let $\mathcal{D}(\mathcal{T}_{\boldsymbol{y}})$ denote the descendants of $\boldsymbol{y}$ in $\mathcal{T}$. The next example illustrates these notational conventions.

Example 4. Let $\mathcal{Y} = \{0,1\}$ and $\kappa = 2$. The tree $\mathcal{T}$ depicted in Figure 2 corresponds to the non-randomized s.t. T taking value one if $Y_1 = 1$ and value 2 if $Y_1 = 0$. The sets $\mathcal{L}(\mathcal{T})$ and $\mathcal{I}(\mathcal{T})$ are given by $\{00, 01, 1\}$ and $\{\rho, 0\}$, respectively. The subtree $\mathcal{T}_0$ of $\mathcal{T}$ consists of the nodes $\{0, 00, 01\}$, and its descendants $\mathcal{D}(\mathcal{T}_0)$ are $\{00, 01\}$. The subtree $\mathcal{T}_\rho$ is the same as $\mathcal{T}$, and its descendants $\mathcal{D}(\mathcal{T}_\rho)$ are $\{0, 1, 00, 01\}$.

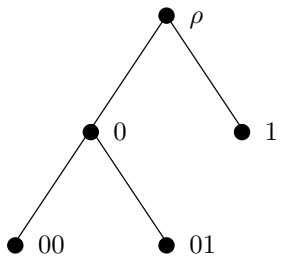

Fig. 2. Tree corresponding to the s.t. T defined by $T = 1$ if $Y_1 = 1$, and $T = 2$ else.

Below, we describe an algorithm that, for a given s.t. S, constructs a sequence of s.t.'s $\{T(\mathcal{T}^m)\}_{m=0}^M$ and Lagrange multipliers $\{\lambda_m\}_{m=0}^M$ with the following two properties. First, the $\mathcal{T}^m$'s and λ_m's are ordered in the sense that $\mathcal{T}^M \subset \mathcal{T}^{M-1} \subset \ldots \subset \mathcal{T}^0$ and $0 = \lambda_M \leq \lambda_{M-1} \leq \ldots \leq \lambda_1 \leq \lambda_0 = \infty$. (Here the symbol $\subset$ denotes inclusion, not necessarily strict.) Second, for any $m \in \{0, \ldots, M\}$ and $\lambda \in (\lambda_m, \lambda_{m-1}]$ the tree $\mathcal{T}^{m-1}$ minimizes $J_\lambda(\mathcal{T}) \triangleq J_\lambda(T(\mathcal{T}))$ among all non-randomized s.t.'s.

Before we state the algorithm, we need to introduce a few quantities. Given a non-randomized s.t. T represented by its $|\mathcal{Y}|$-ary tree $\mathcal{T}$, we write the Lagrangian $J_\lambda(\mathcal{T})$ as

$$\begin{aligned} J_\lambda(\mathcal{T}) &= \mathbb{E}(T-S)^+ + \lambda \mathbb{P}(T < S) \\ &= \sum_{\boldsymbol{y} \in \mathcal{L}(\mathcal{T})} \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}) \Big(\mathbb{E}\big((l(\boldsymbol{y}) - S)^+ | \boldsymbol{Y} = \boldsymbol{y}\big) \\ &\quad + \lambda \mathbb{P}\big(S > l(\boldsymbol{y}) | \boldsymbol{Y} = \boldsymbol{y}\big) \Big) \\ &= \sum_{\boldsymbol{y} \in \mathcal{L}(\mathcal{T})} b(\boldsymbol{y}) + \lambda a(\boldsymbol{y}) \\ &= \sum_{\boldsymbol{y} \in \mathcal{L}(\mathcal{T})} J_\lambda(\boldsymbol{y}), \end{aligned}$$

where

$$\begin{aligned} a(\boldsymbol{y}) &\triangleq \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}) \mathbb{P}(S > l(\boldsymbol{y}) | \boldsymbol{Y} = \boldsymbol{y}), \\ b(\boldsymbol{y}) &\triangleq \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y}) \mathbb{E}\big((l(\boldsymbol{y}) - S)^+ | \boldsymbol{Y} = \boldsymbol{y}\big), \\ J_\lambda(\boldsymbol{y}) &\triangleq b(\boldsymbol{y}) + \lambda a(\boldsymbol{y}) \ . \end{aligned}$$

We extend the definition of $J_\lambda(\cdot)$ to subtrees of $\mathcal{T}$ by setting

$$J_\lambda(\mathcal{T}_{\boldsymbol{y}}) \triangleq \sum_{\boldsymbol{\gamma} \in \mathcal{L}(\mathcal{T}_{\boldsymbol{y}})} J_\lambda(\boldsymbol{\gamma}).$$

With this definition

$$J_\lambda(\mathcal{T}_{\boldsymbol{y}}) = \begin{cases} J_\lambda(\boldsymbol{y}) & \text{if } \boldsymbol{y} \in \mathcal{L}(\mathcal{T}), \\ \sum_{\gamma \in \mathcal{Y}} J_\lambda(\mathcal{T}_{\boldsymbol{y}\gamma}) & \text{if } \boldsymbol{y} \in \mathcal{I}(\mathcal{T}). \end{cases}$$

Similarly, we define

$$\begin{aligned} a(\mathcal{T}_{\boldsymbol{y}}) &\triangleq \sum_{\boldsymbol{\gamma} \in \mathcal{L}(\mathcal{T}_{\boldsymbol{y}})} a(\boldsymbol{\gamma}), \\ b(\mathcal{T}_{\boldsymbol{y}}) &\triangleq \sum_{\boldsymbol{\gamma} \in \mathcal{L}(\mathcal{T}_{\boldsymbol{y}})} b(\boldsymbol{\gamma}). \end{aligned}$$

For a given $\lambda \geq 0$ and $\mathcal{T}$, define $\mathcal{T}(\lambda) \subset \mathcal{T}$ to be the subtree of $\mathcal{T}$ having the same root, and such that $J_\lambda(\mathcal{T}(\lambda)) \leq J_\lambda(\mathcal{T}')$ for all subtrees (with same root) $\mathcal{T}' \subset \mathcal{T}$, and $\mathcal{T}(\lambda) \subset \mathcal{T}'$ for all subtrees (with same root) $\mathcal{T}' \subset \mathcal{T}$ satisfying $J_\lambda(\mathcal{T}(\lambda)) = J_\lambda(\mathcal{T}')$. In words, among all subtrees of $\mathcal{T}$ yielding a minimal cost for a given λ, the tree $\mathcal{T}(\lambda)$ is the smallest. It can be shown that such a smallest subtree always exists, and hence $\mathcal{T}(\lambda)$ is well defined.

Remark 1. Note that $\mathcal{T}_{\boldsymbol{y}}(\lambda)$ is different from $(\mathcal{T}(\lambda))_{\boldsymbol{y}}$. Indeed, $\mathcal{T}_{\boldsymbol{y}}(\lambda)$ refers to the optimal subtree of $\mathcal{T}_{\boldsymbol{y}}$ with respect to λ, whereas $(\mathcal{T}(\lambda))_{\boldsymbol{y}}$ refers to subtree rooted in $\boldsymbol{y}$ of the optimal tree $\mathcal{T}(\lambda)$.

Example 5. Consider again the tree $\mathcal{T}$ in Figure 2. Assume $J_\lambda(\rho) = 4, J_\lambda(0) = 2, J_\lambda(1) = J_\lambda(00) = J_\lambda(01) = 1$. Then

$$\begin{aligned} J_\lambda(\mathcal{T}) &= J_\lambda(1) + J_\lambda(00) + J_\lambda(01) = 3, \\ J_\lambda(\mathcal{T}_0) &= J_\lambda(00) + J_\lambda(01) = 2. \end{aligned}$$

The smallest optimal subtree of $\mathcal{T}$ having the same root is $\mathcal{T}(\lambda) = \{\rho, 0, 1\}$ and

$$J_\lambda(\mathcal{T}(\lambda)) = J_\lambda(0) + J_\lambda(1) = 3.$$

The smallest optimal subtree of $\mathcal{T}_0$ having the same root is $\mathcal{T}_0(\lambda) = \{0\}$ and

$$J_\lambda(\mathcal{T}_0(\lambda)) = J_\lambda(0) = 2.$$

For a tree $\mathcal{T}$ such that $\mathcal{I}(\mathcal{T}) \neq \emptyset$, define for any $\boldsymbol{y} \in \mathcal{I}(\mathcal{T})$

$$g(\boldsymbol{y}, \mathcal{T}) \triangleq \frac{b(\mathcal{T}_{\boldsymbol{y}}) - b(\boldsymbol{y})}{a(\boldsymbol{y}) - a(\mathcal{T}_{\boldsymbol{y}})},$$

where we set $0/0 \triangleq 0$. The quantity $g(\boldsymbol{y}, \mathcal{T})$ captures the tradeoff between the reduction in delay $b(\mathcal{T}_{\boldsymbol{y}}) - b(\boldsymbol{y})$ and the increase in probability of false-alarm $a(\boldsymbol{y}) - a(\mathcal{T}_{\boldsymbol{y}})$ if we stop at the intermediate node $\boldsymbol{y}$ instead of stopping at the leaves $\mathcal{L}(\mathcal{T}_{\boldsymbol{y}})$ of $\mathcal{T}$.

Let $\mathcal{T}^0$ denote the complete tree of depth κ. Starting with $\lambda_0 = \infty$, for $m = \{1, \ldots, M\}$ recursively define

$$\begin{aligned}\lambda_m &\triangleq \inf\{\lambda \leq \lambda_{m-1} : \mathcal{T}^{m-1}(\lambda) = \mathcal{T}^{m-1}\},\\ \mathcal{T}^m &\triangleq \mathcal{T}^{m-1}(\lambda_m),\end{aligned}$$

where M is the smallest integer such that $\lambda_{M+1} = 0$, and with $\lambda_1 \triangleq \infty$ if the set over which the infimum is taken is empty. It can be shown that for two consecutive transition points λ_m and λ_{m+1}, we have $\mathcal{T}^0(\lambda) = \mathcal{T}^0(\lambda_m)$ for all $\lambda \in (\lambda_{m+1}, \lambda_m]$ as shown in Figure 3.

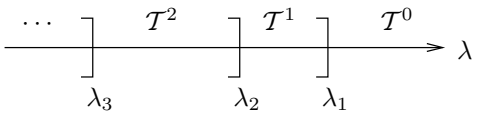

Fig. 3. For all $m \in \{0, 1, \ldots, M-1\}$ the tree $\mathcal{T}^m$ is the smallest tree minimizing the cost $J_\lambda(\cdot)$ for any $\lambda \in (\lambda_{m+1}, \lambda_m]$.

We are now ready to state the algorithm that fully characterizes $d(\alpha)$ by computing its set of break-points $\{(\alpha_m, d_m)\}_{m=1}^M$.

Algorithm: Compute the break-points $\{(\alpha_m, d_m)\}_{m=1}^M$ of $d(\alpha)$

$m \Leftarrow 0$
$\lambda_0 \Leftarrow \infty$
$\mathcal{T}^0 \Leftarrow$ complete tree of depth κ
repeat
  $m \Leftarrow m + 1$
  $\lambda_m \Leftarrow \max_{\boldsymbol{y} \in \mathcal{I}(\mathcal{T}^{m-1})} g(\boldsymbol{y}, \mathcal{T}^{m-1})$
  $\mathcal{T}^m \Leftarrow \mathcal{T}^{m-1} \setminus \bigcup_{\substack{\boldsymbol{y} \in \mathcal{I}(\mathcal{T}^{m-1}):\\ g(\boldsymbol{y}, \mathcal{T}^{m-1}) = \lambda_m}} \mathcal{D}(\mathcal{T}^{m-1}_{\boldsymbol{y}})$
  $\alpha_m \Leftarrow \mathbb{P}(T(\mathcal{T}^m) < S)$
  $d_m \Leftarrow \mathbb{E}(T(\mathcal{T}^m) - S)^+$
until $\lambda_m = 0$
$M \Leftarrow m - 1$

As a $|\mathcal{Y}|$-ary tree has less than $|\mathcal{Y}|^\kappa$ non-terminal nodes, the algorithm terminates after at most that many iterations. Further, one may check that each iteration has a running time that is $\exp(O(\kappa))$. Therefore, the worst case running time of the algorithm is $\exp(O(\kappa))$. This is to be compared, for instance, with exhaustive search that has a $\Omega(\exp\exp(\kappa))$ running time (because all break-points of $d(\alpha)$ are achieved by non-randomized s.t.'s and there are already $2^{|\mathcal{Y}|^{\kappa-1}}$ $|\mathcal{Y}|$-ary trees having leaves at either depth κ or $\kappa-1$). Moreover, it can be shown that under certain conditions on $\{(X_i, Y_i)\}_{i\geq 1}$ and S, the running time of the algorithm is only *polynomial* in κ (see Niesen and Tchamkerten, 2009, for the details).

4 Conclusions

We introduced the tracking stopping time problem, a novel quickest detection problem that generalizes the Bayesian change-point setting. By exploiting the finite tree structure of bounded stopping times defined over finite alphabet processes, we derived an algorithm solving the tracking stopping time problem. This algorithm computes the minimum reaction delays for tracking a stopping time through noisy observations for any fixed probability of false-alarm, and finds the corresponding optimal tracking stopping time. The running time of this algorithm is exponential in the bound of the stopping time we want to track and, in certain cases, even polynomial. In comparison, an exhaustive search has a running time that is doubly exponential.

References

Anscombe, F. J., Godwin, H. J., and Plackett, R. L. (1947). Methods of deferred sentencing in testing the fraction defective of a continuous output. *Supplement to the Journal of the Royal Statistical Society*, **9**(2):198–217.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Chapman & Hall.

Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, **44**(7):2917–2929.

Moustakides, G. (2008). Sequential change detection revisited. *the Annals of Statistics*, **36**(2):787–807.

Niesen, U. and Tchamkerten, A. (2009). Tracking Stopping Times Through Noisy Observations. *IEEE Transactions on Information Theory*, **55**(1):422–432.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Th. Prob. and its App.*, **8**(1):22–46.

Tartakovsky, A. G. and Veeravalli, V. V. (2005). General asymptotic Bayesian theory of qickest change detection. *Th. Prob. Appl.*, **49**(3):458–497.

Yakir, B. (1994). Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space. In *Change-point problems*, vol. 23. Institute of Mathematical Statistics, Lecture Notes, Monograph Series, pages 346–358.