

How to Capture a Stopping Time: The Independent Case

George V. Moustakides

Electrical and Computer Engineering, University of Patras, 26500 Rion, Greece
& Computer Science, Rutgers University, New Brunswick, NJ 08854-8019, USA
Email: moustaki@upatras.gr

Abstract—A stopping time τ is controlled by a process $\{X_t\}$ and we are interested in the detection of its onset using a sequential scheme. The sequential detector is an alternative stopping time T which is based on observations $\{\xi_t\}$ that are obtained sequentially. The occurrence of τ initiates a change in the statistical behavior of the pair processes $\{(X_t, \xi_t)\}$ for times after τ with τ modeling the change-imposing mechanism. We cast the detection problem as two possible constrained optimal stopping problems. We provide the optimum detector for each case when the pair process $\{(X_t, \xi_t)\}$ is i.i.d. and τ corresponds to the first entry time of X_t into some fixed and known set $\mathcal{A}$. The resulting optimum schemes accept an interesting form. In particular one of the two detectors is reduced to the well known Shiryaev test (with τ geometrically distributed) when the two processes $\{X_t\}$ and $\{\xi_t\}$ are independent from each other.

Index Terms—Sequential detection, Optimal stopping.

I. INTRODUCTION

A stopping time τ is adapted to the filtration $\{\mathcal{X}_t\}$ generated by some process $\{X_t\}$ with $\mathcal{X}_t = \sigma\{X_s, s \leq t\}$. The stopping rule which gives rise to τ is assumed known but $\{X_t\}$ is not observed. Instead, we sequentially observe an alternative process $\{\xi_t\}_{t \geq 0}$ which we like to use in order to detect the occurrence of τ . The onset of τ at time t , that is, $\tau = t$ creates a change in the statistical behavior of the observed process $\{\xi_t\}$ for times $t + 1, \dots$. The process $\{X_t\}$ can also experience a change but this is not necessary.

We assume that the pair process $\{(X_s, \xi_s)\}_{s=0}^t$ is i.i.d. before and after the change with known joint pdf $f_\infty(X, \xi), f_0(X, \xi)$. A change at $\tau = t$ induces a probability measure which we denote by $P_t(\cdot)$ while we reserve $E_t[\cdot]$ for the corresponding expectation. Clearly with this definition P_∞ is the measure where all data are under the nominal statistic while P_0 under the alternative. Finally, we use $P(\cdot)$ and $E[\cdot]$ for probability and expectation when we do not specify τ to take upon some particular deterministic value.

Any sequential detector can be seen as a stopping time T which, unlike τ , is adapted to the filtration $\{\mathcal{F}_t\}$ generated by the observations $\{\xi_t\}$, namely, $\mathcal{F}_t = \sigma\{\xi_s, 0 \leq s \leq t\}$ and $\mathcal{F}_0$ is a sigma-algebra which allows for randomizations at time 0. To quantify the detection capability of the stopping

time T we propose two possible performance measures

$$\mathcal{C}(T) = P(T = \tau^+ | T \geq \tau) \quad (1)$$

$$\text{or } \mathcal{D}(T) = E[T - \tau^+ | T \geq \tau]. \quad (2)$$

With (1) we are interested in the probability of immediately detecting (capturing) τ , while with (2) we quantify the expected detection delay. The first criterion should be applied to cases where detection must be performed immediately. This is for instance the case when the occurrence of τ signals the arrival of an imminent catastrophic event. Therefore its immediate detection, which is expressed through the event $\{T = \tau\}$, contributes towards maximizing the possibility of mitigating the upcoming disaster. The second criterion is more classical and considers the detection delay $T - \tau$ as a figure of merit. Both measures are conditioned on the event that we did not experience any false alarm before the occurrence of τ . By excluding false alarms from our criterion we quantify only successes. Failures, on the other hand, which are equally important, will be quantified through the false-alarm probability $P(T < \tau)$. An additional detail in both metrics is the fact that when $\tau < 0$ then the fastest we can hope in detecting τ is at $t = 0$. This is the reason why we consider the positive part τ^+ in our criterion¹.

Based on the performance metrics introduced in (1),(2) we are now in a position to propose suitable optimization problems whose solution will yield the optimum detector T for each case. Specifically we are interested in solving the following two constrained optimization problems

$$\sup_{T \geq 0} \mathcal{C}(T) = \sup_{T \geq 0} P(T = \tau^+ | T \geq \tau), \quad (3)$$

$$\text{or } \inf_{T \geq 0} \mathcal{D}(T) = \inf_{T \geq 0} E[T - \tau^+ | T \geq \tau], \quad (4)$$

$$\text{both subject to: } P(T < \tau) \leq \alpha. \quad (5)$$

In other words we would like to design our stopping time T so as to maximize the conditional capture probability or minimize the conditional average detection delay with simultaneous guarantee that the false-alarm probability will not exceed some prescribed level $\alpha \in (0, 1)$. In the next section we present a general analysis that will reduce (3) and (4) into classical optimal stopping problems. Before continuing with our analysis we need to make some useful remarks.

¹This also allows to summarize the past of τ with the help of the probability $P(\tau \leq 0)$.

Remark 1. We must mention that the problem of delayed detection of a stopping time also known as “tracking a stopping time” was first considered in [1] and [2] for special classes of continuous-time processes. Both articles provide asymptotically optimum procedures with the corresponding optimality being of first order. Here, we consider discrete-time signals and we target exact optimality.

Remark 2. The proposed analysis can be seen as a different formulation of the sequential change detection problem. In particular we model the change-imposing mechanism as a stopping time adapted to an information history which is not necessarily the same with the history generated by the observations (and used to detect the change). This idea was introduced in [3] where a worst-case analysis regarding τ was considered. In the current work we assume τ to be completely known as far as the stopping rule and its statistical description are concerned.

Remark 3. Our analysis, when applied to the i.i.d. case, will give rise to results that are more general than the well celebrated Shiryaev test [4]. Furthermore, our formulation will also be able to provide a realistic mechanism capable of delivering geometrically distributed change-times τ , which is the model adopted by Shiryaev in his Bayesian formulation.

II. BACKGROUND ANALYSIS

Problems where we need to optimize stopping times are usually solved by applying results from Optimal Stopping Theory. To write (3) and (4) under a form which is suitable for this theory we first need to make certain definitions. Regarding our stopping time T we equip it with a randomization probability $P(T = 0)$ which we apply at time 0. This is essential for addressing the case $\tau \leq 0$. Specifically at time 0, with probability $1 - P(T = 0)$ we decide to start sampling and with probability $P(T = 0)$ to stop at 0 without taking any samples.

Let us now define an $\{\mathcal{F}_t\}$ -adapted process $\{\pi_t\}$ with $\pi_0 = P(\tau \leq 0)$ and for $t > 0$ we have $\pi_t = P_\infty(\tau = t | \mathcal{F}_t)$. In other words π_t denotes the probability that $\tau = t$, given the observation history up to time t . Using $\{\pi_t\}$ we can rewrite the two performance metrics in a way that can be treated by Optimal Stopping Theory. We apply the following manipulations:

$$\begin{aligned} \mathcal{C}(T) &= \frac{P(T = \tau^+)}{P(T \geq \tau)} \\ &= \frac{E[\mathbb{1}_{\{T=0\}} \mathbb{1}_{\{\tau \leq 0\}} + \sum_{t=1}^{\infty} \mathbb{1}_{\{T=t\}} \mathbb{1}_{\{\tau=t\}}]}{E[\mathbb{1}_{\{\tau \leq 0\}} + \sum_{t=1}^{\infty} \mathbb{1}_{\{T \geq t\}} \mathbb{1}_{\{\tau=t\}}]} \\ &= \frac{E_\infty[\mathbb{1}_{\{T=0\}} P(\tau \leq 0) + \sum_{t=1}^{\infty} \mathbb{1}_{\{T=t\}} P_\infty(\tau = t | \mathcal{F}_t)]}{E_\infty[P(\tau \leq 0) + \sum_{t=1}^{\infty} \mathbb{1}_{\{T \geq t\}} P_\infty(\tau = t | \mathcal{F}_t)]} \\ &= \frac{E_\infty[\mathbb{1}_{\{T=0\}} \pi_0 + \sum_{t=1}^{\infty} \mathbb{1}_{\{T=t\}} \pi_t]}{E_\infty[\pi_0 + \sum_{t=1}^{\infty} \mathbb{1}_{\{T \geq t\}} \pi_t]} \\ &= \frac{E_\infty[\pi_T]}{E_\infty[\sum_{t=0}^T \pi_t]}, \quad (6) \end{aligned}$$

where $\frac{0}{0} = 1$ (to cover the case $\pi_0 = 0$ and $T = 0$). In (6), in the second equality, in order for the denominator to take the specific form, we used the fact that $T \geq 0$. Also the third equality is obtained by using the tower property of expectation and recalling that $\mathbb{1}_{\{T=t\}}$ is $\mathcal{F}_t$ -measurable, while $\mathbb{1}_{\{T \geq t\}}$ is $\mathcal{F}_{t-1}$ and therefore $\mathcal{F}_t$ -measurable. For the above manipulations to be valid we also need to assume that $P(\tau < \infty) = P(T < \infty) = 1$. Regarding the assumption that τ stops w.p.1, it is without loss of generality. Indeed if $0 < P(\tau < \infty) < 1$ we can simply rescale the probabilities $\{\pi_t\}$ using $P(\tau < \infty)$ and obtain an alternative version of the stopping time τ that stops w.p.1. This is the same as conditioning our whole analysis on the event $\{\tau < \infty\}$, provided the latter has nonzero probability. Note also that the event $\{\tau = \infty\}$ is not interesting since it results only in false alarms. Finally we note that all expectations are with respect to the nominal measure since by requiring immediate detection (capture) we do not experience any data under the alternative regime. Such data, however, will be required in the second metric.

Similarly we can analyze (2) and write

$$\begin{aligned} \mathcal{D}(T) &= \frac{E[(T - \tau^+)^+]}{P(T \geq \tau)} \\ &= \frac{E[T \mathbb{1}_{\{\tau \leq 0\}} + \sum_{t=1}^{\infty} (T - t)^+ \mathbb{1}_{\{\tau=t\}}]}{E[\mathbb{1}_{\{\tau \leq 0\}} + \sum_{t=1}^{\infty} \mathbb{1}_{\{T \geq t\}} \mathbb{1}_{\{\tau=t\}}]} \\ &= \frac{E[\sum_{s=0}^{\infty} \mathbb{1}_{\{T > s\}} \mathbb{1}_{\{\tau \leq 0\}} + \sum_{t=1}^{\infty} \sum_{s=t}^{\infty} \mathbb{1}_{\{T > s\}} \mathbb{1}_{\{\tau=t\}}]}{E[\mathbb{1}_{\{\tau \leq 0\}} + \sum_{t=1}^{\infty} \mathbb{1}_{\{T \geq t\}} \mathbb{1}_{\{\tau=t\}}]} \\ &= \frac{\sum_{s=0}^{\infty} E_0[\mathbb{1}_{\{T > s\}}] \pi_0 + \sum_{t=1}^{\infty} \sum_{s=t}^{\infty} E_t[\mathbb{1}_{\{T > s\}} \mathbb{1}_{\{\tau=t\}}]}{\pi_0 + \sum_{t=1}^{\infty} E_t[\mathbb{1}_{\{T \geq t\}} \mathbb{1}_{\{\tau=t\}}]}, \quad (7) \end{aligned}$$

We also note the following equality

$$E_0[\mathbb{1}_{\{T > s\}}] = E_\infty[\mathbb{1}_1^s \mathbb{1}_{\{T > s\}}] \quad (8)$$

and for $s \geq t > 0$

$$\begin{aligned} E_t[\mathbb{1}_{\{T > s\}} \mathbb{1}_{\{\tau=t\}}] &= E_t[\mathbb{1}_{\{T > s\}} E_t[\mathbb{1}_{\{\tau=t\}} | \mathcal{F}_s]] \\ &= E_t[\mathbb{1}_{\{T > s\}} E_\infty[\mathbb{1}_{\{\tau=t\}} | \mathcal{F}_t]] \quad (9) \\ &= E_\infty[\mathbb{1}_{t+1}^s \pi_t \mathbb{1}_{\{T > s\}}], \quad (10) \end{aligned}$$

where $\mathbb{1}_a^b$ denotes the likelihood ratio of the samples $\{\xi_a, \dots, \xi_b\}$ with $\mathbb{1}_a^b = 1$ whenever $b < a$. The equality in (9) is true due to our assumption that the pair process after $\tau = t$ is independent from the process up to t . All previous formulas are obtained by straightforward application of the tower property of expectation and change of measures. Finally, since $\{T \geq t\}$ is $\mathcal{F}_{t-1}$ -measurable it is also $\mathcal{F}_t$ -measurable, then on $\{\tau = t\}$ this event is under the nominal probability measure, consequently

$$\begin{aligned} E_t[\mathbb{1}_{\{T \geq t\}} \mathbb{1}_{\{\tau=t\}}] &= E_\infty[\mathbb{1}_{\{T \geq t\}} E_\infty[\mathbb{1}_{\{\tau=t\}} | \mathcal{F}_t]] \\ &= E_\infty[\pi_t \mathbb{1}_{\{T \geq t\}}] = E_\infty[\mathbb{1}_{t+1}^T \pi_t \mathbb{1}_{\{T \geq t\}}], \quad (11) \end{aligned}$$

with the last equality being true because $E_\infty[\mathbb{1}_{t+1}^T | \mathcal{F}_t] = 1$ due to Optional Sampling.

Substituting (8),(10),(11) in (7), interchanging the order of the two summations and swapping the roles of s and t in the numerator we obtain

$$\begin{aligned}\mathcal{D}(T) &= \frac{\mathbb{E}_\infty[\sum_{t=0}^{T-1} \sum_{s=t}^{T-1} \mathbf{L}_{t+1}^s \pi_t]}{\mathbb{E}_\infty[\sum_{t=0}^T \mathbf{L}_{t+1}^T \pi_t]} \\ &= \frac{\mathbb{E}_\infty[\sum_{t=0}^{T-1} \sum_{s=0}^t \mathbf{L}_{s+1}^t \pi_s]}{\mathbb{E}_\infty[\sum_{t=0}^T \mathbf{L}_{t+1}^T \pi_t]}.\end{aligned}\quad (12)$$

Finally, if we define the following statistic

$$R_t = \sum_{s=0}^t \mathbf{L}_{s+1}^t \pi_s \quad (13)$$

then our criterion takes its final form

$$\mathcal{D}(T) = \frac{\mathbb{E}_\infty[\sum_{t=0}^{T-1} R_t]}{\mathbb{E}_\infty[R_T]}, \quad (14)$$

where we define $\sum_a^b = 0$ when $b < a$. Regarding the false alarm constraint, from (6) and (14) we have that the denominator of our performance measures is equal to the complement of the false-alarm probability. Hence the constraint can be equivalently expressed as

$$\mathbb{E}_\infty[\sum_{t=0}^T \pi_t] = \mathbb{E}_\infty[R_T] \geq 1 - \alpha. \quad (15)$$

Summarizing, we distinguish the following two constrained optimization problems:

Problem 1:

$$\begin{aligned}\sup_{T \geq 0} \mathcal{C}(T) &= \sup_{T \geq 0} \frac{\mathbb{E}_\infty[\pi_T]}{\mathbb{E}_\infty[\sum_{t=0}^T \pi_t]} \\ \text{subject to: } &\mathbb{E}_\infty[\sum_{t=0}^T \pi_t] \geq 1 - \alpha,\end{aligned}\quad (16)$$

corresponding to the combination of (3) and (5); and

Problem 2:

$$\begin{aligned}\inf_{T \geq 0} \mathcal{D}(T) &= \inf_{T \geq 0} \frac{\mathbb{E}_\infty[\sum_{t=0}^{T-1} R_t]}{\mathbb{E}_\infty[R_T]} \\ \text{subject to: } &\mathbb{E}_\infty[R_T] \geq 1 - \alpha,\end{aligned}\quad (17)$$

corresponding to the combination of (4) and (5).

From this point on in order to simplify our presentation we will make the assumption that $\pi_0 = \mathbb{P}(\tau \leq 0) = 0$. In other words a change can occur after and including time 0. Under this condition we have the following interesting lemma.

Lemma 1. *The performance achieved by any stopping time T that satisfies the false-alarm constraint in the strict sense can be matched by an alternative stopping time that satisfies the constraint with equality.*

Proof: After distinguishing the cases $\mathbb{P}(T = 0)$ and $\mathbb{P}(T > 0)$ it is easy to verify that both metrics $\mathcal{C}(T)$ and $\mathcal{D}(T)$, when $\pi_0 = 0$, are independent from the randomization probability $\mathbb{P}(T = 0)$. Consequently, it is always possible to properly modify the randomization probability $\mathbb{P}(T = 0)$ in order to meet the constraint with equality without altering the corresponding performance metric. ■

Because of Lemma 1 we are allowed to limit ourselves to detectors that satisfy the constraint with equality suggesting

that the denominator in our metrics becomes constant and equal to $1 - \alpha$. Thus, we need only consider the optimization of the corresponding numerators. Consequently Problem 1 in (16) is equivalent to

$$\sup_{T \geq 0} \mathbb{E}_\infty[\pi_T], \quad \text{subject to: } \mathbb{E}_\infty[\sum_{t=0}^T \pi_t] = 1 - \alpha, \quad (18)$$

while Problem 2 in (17) becomes

$$\inf_{T \geq 0} \mathbb{E}_\infty[\sum_{t=0}^{T-1} R_t], \quad \text{subject to: } \mathbb{E}_\infty[R_T] = 1 - \alpha. \quad (19)$$

We are now ready to introduce the final form of the optimization problems we intend to solve.

Let λ be a Lagrange multiplier, then we define the following criterions:

$$\begin{aligned}\tilde{\mathcal{C}}(T) &= \mathbb{E}_\infty[\pi_T + \lambda \sum_{t=0}^T \pi_t] \\ \tilde{\mathcal{D}}(T) &= \mathbb{E}_\infty[\lambda R_T + \sum_{t=0}^{T-1} R_t],\end{aligned}$$

leading to the corresponding *unconstrained* optimization problems that replace (18) and (19)

$$\sup_{T \geq 0} \tilde{\mathcal{C}}(T) = \sup_{T \geq 0} \mathbb{E}_\infty[(1 + \lambda)\pi_T + \lambda \sum_{t=0}^{T-1} \pi_t] \quad (20)$$

$$\inf_{T \geq 0} \tilde{\mathcal{D}}(T) = \inf_{T \geq 0} \mathbb{E}_\infty[\lambda R_T + \sum_{t=0}^{T-1} R_t]. \quad (21)$$

Problems (20),(21) are under the standard form encountered in Optimal Stopping Theory [5] with $\{\pi_t\}, \{R_t\}$ being $\{\mathcal{F}_t\}$ -adapted. We can therefore directly apply the corresponding optimality results.

A. Optimal Stopping

We consider the sequences $\{U_t\}$ and $\{V_t\}$ of optimal gains

$$U_t = \sup_{T \geq t} \mathbb{E}_\infty[(1 + \lambda)\pi_T + \lambda \sum_{j=t}^{T-1} \pi_j | \mathcal{F}_t]$$

$$V_t = \inf_{T \geq t} \mathbb{E}_\infty[\lambda R_T + \sum_{j=t}^{T-1} R_j | \mathcal{F}_t]$$

where $\{U_t\}, \{V_t\}$ are $\{\mathcal{F}_t\}$ -adapted and for which we have the following backward updating formulas for $t = 0, 1, 2, \dots$, due to Optimal Stopping Theory:

$$U_t = \max\{(1 + \lambda)\pi_t, \lambda\pi_t + \mathbb{E}_\infty[U_{t+1} | \mathcal{F}_t]\} \quad (22)$$

$$V_t = \min\{\lambda R_t, R_t + \mathbb{E}_\infty[V_{t+1} | \mathcal{F}_t]\}. \quad (23)$$

Note that U_0, V_0 express the optimum gain for $T \geq 0$.

B. Time of First Entry into a Known Set

In the next section we will analyze (22),(23) for the special case where τ is the class of stopping times corresponding to the time of first entry of X_t into some known set $\mathcal{A}$, that is, $\tau = \inf\{t > 0 : X_t \in \mathcal{A}\}$. This suggests that for $t > 0$ we have

$$\begin{aligned}\pi_t &= \mathbb{P}_\infty(X_t \in \mathcal{A}, X_{t-1} \in \mathcal{A}^c, \dots, X_1 \in \mathcal{A}^c, \tau > 0 | \mathcal{F}_t) \\ &= \mathbb{P}_\infty(X_t \in \mathcal{A}, X_{t-1} \in \mathcal{A}^c, \dots, X_1 \in \mathcal{A}^c | \tau > 0, \mathcal{F}_t)(1 - \pi_0)\end{aligned}$$

where $\mathcal{A}^c$ denotes the complement of the set $\mathcal{A}$. The probability measure we use for the computation of π_t is the nominal since up to and including the change-time τ the processes follow the pre-change measure.

Under the assumption that the pair process $\{(X_t, \xi_t)\}$ is i.i.d. before and after the change we conclude that

$$\pi_t = \omega_t \prod_{j=0}^{t-1} (1 - \omega_j) \quad (24)$$

where $\omega_0 = \pi_0 = P(\tau \leq 0)$ and for $t > 0$ we have $\omega_t = P_\infty(X_t \in \mathcal{A} | \xi_t)$. Clearly $\{\omega_t\}_{t \geq 0}$, under P_∞ , is an i.i.d. sequence. Let us now apply this analysis for solving the two optimization problems of interest.

III. OPTIMUM DETECTORS

We first consider (16) and attempt to solve it by obtaining the solution of its unconstrained counterpart depicted in (20). We will then continue with the solution of (17) by treating (21).

A. Optimum Solution for Problem 1

Note that $P_\infty(\tau > t | \mathcal{F}_t) = \prod_{j=1}^t (1 - \omega_j)$ consequently, if we define the normalized gain $\mathcal{U}_t = U_t / P_\infty(\tau > t | \mathcal{F}_t)$ and also use (24) then (22) is equivalent to

$$\mathcal{U}_t = \lambda \frac{\omega_t}{1 - \omega_t} + \max \left\{ \frac{\omega_t}{1 - \omega_t}, E_\infty[(1 - \omega_{t+1})\mathcal{U}_{t+1} | \mathcal{F}_t] \right\}, \quad (25)$$

with the optimum terminal gain becoming $U_0 = \mathcal{U}_0 P_\infty(\tau > 0 | \mathcal{F}_0) = \mathcal{U}_0(1 - \pi_0) = \mathcal{U}_0$. Due to stationarity if we select $\mathcal{U}_t = \mathcal{U}(\omega_t)$ it is straightforward to show that it is a solution to the previous equation provided $\mathcal{U}(\omega)$ is defined as

$$\begin{aligned} \mathcal{U}(\omega) &= \lambda \frac{\omega}{1 - \omega} + \max \left\{ \frac{\omega}{1 - \omega}, C \right\} \\ &= \lambda \frac{\omega}{1 - \omega} + \frac{\omega}{1 - \omega} \mathbb{1}_{\{\omega \geq \frac{C}{1+C}\}} + C \mathbb{1}_{\{\omega < \frac{C}{1+C}\}}. \end{aligned}$$

Constant C , from (25), must satisfy $C = E_\infty[(1 - \omega_1)\mathcal{U}(\omega_1)]$. From Optimal Stopping Theory we also have that for $t > 0$ the stopping time

$$\begin{aligned} \mathcal{T}_\nu &= \inf\{t > 0 : \omega_t \geq \nu\} \\ &= \inf\{t > 0 : P_\infty(X_t \in \mathcal{A} | \xi_t) \geq \nu\} \end{aligned} \quad (26)$$

with $\nu = \frac{C}{1+C}$ can attain the optimum performance for $T > 0$. According to our definition however we need to include randomization at $T = 0$. The final optimum capture time T_o is given by the next theorem.

Theorem 1. *We distinguish two different forms of the optimum capture time T_o :*

- i) $1 - P_\infty(X_1 \in \mathcal{A}) \leq \alpha < 1$: Then T_o is a suitable randomization between 0 and 1.
- ii) $\alpha < 1 - P_\infty(X_1 \in \mathcal{A})$: Then $T_o = \mathcal{T}_\nu$ with properly selected threshold ν and no randomization at $t = 0$.

Proof: For Case i) consider $\mathcal{T}_\nu$ with $\nu = 0$ this means that $\mathcal{T}_0$ will necessarily stop at time 1. We can then see that stopping at 0 produces the same gain as stopping at time 1.

We are therefore allowed to use randomization between the two possibilities without altering the final optimum gain. To find the proper randomization probability we need to satisfy the false-alarm constraint with equality. As we can verify, this is possible if we select

$$P(T_o = 0) = 1 - \frac{(1 - \alpha)}{E_\infty[\omega_1]},$$

which, according to our assumption and after observing that $E_\infty[\omega_1] = P_\infty(X_1 \in \mathcal{A})$, is a legitimate probability with value in the interval $(0, 1)$.

For Case ii) we select the threshold $\nu \in (0, 1)$ consequently $C = \frac{\nu}{1-\nu} > 0$. This implies that stopping at 0 produces less gain than using S_ν suggesting that we must select $T_o = \mathcal{T}_\nu$ and never stop at 0. Again to find the proper threshold we need the false-alarm constraint to be satisfied with equality, that is

$$P_\infty(\omega_1 \geq \nu) + E_\infty[\omega_1 \mathbb{1}_{\{\omega_1 < \nu\}}] = \frac{E_\infty[\omega_1]}{1 - \alpha}.$$

A solution $\nu \in (0, 1)$ always exists since for $\nu = 0$ the left hand side is equal to 1 and thus strictly larger than the right, while for $\nu = 1$ the left becomes $E_\infty[\omega_1]$ which is smaller than the right. Evoking continuity arguments we prove existence of $\nu \in (0, 1)$.

In both cases if we solve the equation $C = E_\infty[(1 - \omega_1)\mathcal{U}(\omega_1)]$ we compute the Lagrange multiplier λ required in (25) to produce the specific optimum solution. ■

B. Optimum Solution for Problem 2

As in the previous problem in (23) we define the normalized gain $\mathcal{V}_t = V_t / P_\infty(\tau > t)$ and the normalized test statistic $\mathcal{R}_t = R_t / P_\infty(\tau > t)$, then (23) is equivalent to

$$\mathcal{V}_t = \min \{ \lambda \mathcal{R}_t, \mathcal{R}_t + E_\infty[(1 - \omega_{t+1})\mathcal{V}_{t+1} | \mathcal{F}_t] \}. \quad (27)$$

For the original optimum gain we have $V_0 = \mathcal{V}_0(1 - \pi_0) = \mathcal{V}_0$. Regarding $\{\mathcal{R}_t\}$ we can see that it satisfies the following time update for $t > 0$

$$\mathcal{R}_t = \frac{\mathcal{R}_{t-1}\ell_t + \omega_t}{1 - \omega_t}, \quad \mathcal{R}_0 = 0, \quad (28)$$

with $\ell_t = \frac{f_0(\xi_t)}{f_\infty(\xi_t)}$ the likelihood ratio of ξ_t . This implies that $\{\mathcal{R}_t\}_{t \geq 0}$ is first order Markov. Due to this Markovian nature, we can search for solutions in (27) that are of the form $\mathcal{V}_t = \mathcal{V}(\mathcal{R}_t)$, where $\mathcal{V}(\mathcal{R})$ satisfies the equation

$$\begin{aligned} \mathcal{V}(\mathcal{R}) &= \\ \min \left\{ \lambda \mathcal{R}, \mathcal{R} + E_\infty \left[(1 - \omega_1) \mathcal{V} \left(\frac{\mathcal{R}\ell_1 + \omega_1}{1 - \omega_1} \right) \right] \right\}, \end{aligned} \quad (29)$$

and, we recall, that $\omega_1 = P_\infty(X_1 \in \mathcal{A} | \xi_1)$. Our optimality result is given in the following theorem.

Theorem 2. *We distinguish two different forms of the optimum capture time T_o :*

- i) $1 - P_\infty(X_1 \in \mathcal{A}) \leq \alpha < 1$: Then T_o is a suitable randomization between 0 and 1.

ii) $\alpha < 1 - P_\infty(X_1 \in \mathcal{A})$: Then

$$T_o = \inf\{t > 0 : \mathcal{R}_t \geq \nu\}, \quad (30)$$

with the threshold $\nu > 0$ selected to satisfy the false-alarm constraint with equality.

Proof: Case i) is similar to Theorem 1. For Case ii) to prove that the optimum T_o is as in (30) we consider $\lambda < 0$ and we first show that the recursion

$$\mathcal{V}_n(\mathcal{R}) = \min \left\{ \lambda \mathcal{R}, \mathcal{R} + E_\infty \left[(1 - \omega_1) \mathcal{V}_{n-1} \left(\frac{\mathcal{R}\ell_1 + \omega_1}{1 - \omega_1} \right) \right] \right\}, \quad (31)$$

with $\mathcal{V}_0(\mathcal{R}) = 0$ converges to a function $\mathcal{V}(\mathcal{R})$. In particular we show that for all n we have the following properties: a) $\mathcal{V}_n(\mathcal{R}) \geq \mathcal{V}_{n+1}(\mathcal{R})$; b) $\mathcal{V}_n(\mathcal{R}) \geq \lambda(\mathcal{R} + 1)$ and c) $\mathcal{V}_n(\mathcal{R})$ is concave. All three properties are valid for $n = 0$. If we assume that they are true for $n = k$ then it is straightforward to show that they are also true for $n = k + 1$. Consequently they hold for all n .

To show existence of the limit $\mathcal{V}(\mathcal{R}) = \lim_{n \rightarrow \infty} \mathcal{V}_n(\mathcal{R})$ we note that for each fixed $\mathcal{R}$ we have from property a) that the sequence $\{\mathcal{V}_n(\mathcal{R})\}$ is decreasing in n and from property b) that it is lower bounded by $\lambda(\mathcal{R} + 1)$, consequently $\{\mathcal{V}_n(\mathcal{R})\}$ has a pointwise limit $\mathcal{V}(\mathcal{R})$ in n . Since the function $\mathcal{V}_n(\mathcal{R})$, for each n , is lower bounded by $\lambda(\mathcal{R} + 1)$ the same holds true for the limit $\mathcal{V}(\mathcal{R})$. Also from property c) we have $\mathcal{V}_n(\mathcal{R})$ to be concave for each n . Due to the “min” in recursion (31) we can show that this property is inherited by the limit, meaning that $\mathcal{V}(\mathcal{R})$ is concave and therefore continuous. Summarizing: there exists a function $\mathcal{V}(\mathcal{R})$ that satisfies (29) which is continuous, concave, and bounded from below by $\lambda(\mathcal{R} + 1)$.

The properties we just mentioned allow us to conclude that the function $\tilde{\mathcal{V}}(\mathcal{R}) = \mathcal{R} + E_\infty[(1 - \omega_1)\mathcal{V}(\frac{\mathcal{R}\ell_1 + \omega_1}{1 - \omega_1})]$ is concave (therefore continuous) and lower bounded by $(1 + \lambda)\mathcal{R} + \lambda$. Fig. 1 captures these facts and can help us understand why the two functions $\lambda\mathcal{R}$ and $\tilde{\mathcal{V}}(\mathcal{R})$ can intersect only at the single point $\mathcal{R} = \nu$. Indeed since the lower bound $(1 + \lambda)\mathcal{R} + \lambda$ intersects $\lambda\mathcal{R}$ and because $\tilde{\mathcal{V}}(0) = E_\infty[(1 - \omega_1)\mathcal{V}(\frac{\omega_1}{1 - \omega_1})] < 0$ due to $\mathcal{V}(\mathcal{R}) \leq 0$, we have that $\tilde{\mathcal{V}}(\mathcal{R})$ will also intersect $\lambda\mathcal{R}$ at some point. Function $\tilde{\mathcal{V}}(\mathcal{R})$ cannot intersect $\lambda\mathcal{R}$ in

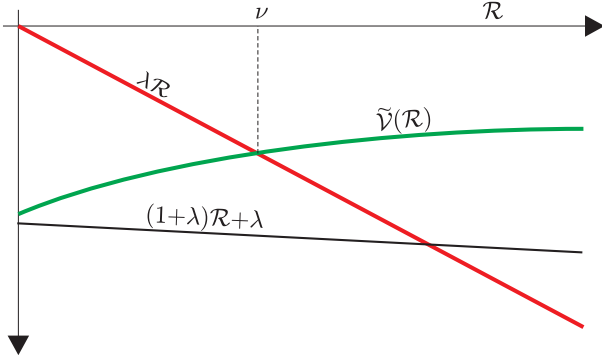


Fig. 1. Typical form of $\lambda\mathcal{R}$ (red), the function $\tilde{\mathcal{V}}(\mathcal{R}) = \mathcal{R} + E_\infty[(1 - \omega_1)\mathcal{V}(\frac{\mathcal{R}\ell_1 + \omega_1}{1 - \omega_1})]$ (green) and its lower bound $(1 + \lambda)\mathcal{R} + \lambda$ (black). The optimum (normalized) cost is $\mathcal{V}(\mathcal{R}) = \min\{\lambda\mathcal{R}, \tilde{\mathcal{V}}(\mathcal{R})\}$.

more than one points because due to its concavity this would require $\tilde{\mathcal{V}}(\mathcal{R})$ to have an asymptote for $\mathcal{R} \rightarrow \infty$ with slope that is steeper than λ . But if this were true, $\tilde{\mathcal{V}}(\mathcal{R})$ would have necessarily intersected the line $(1 + \lambda)\mathcal{R} + \lambda$ as well. This is a contradiction since this line is a lower bound to $\tilde{\mathcal{V}}(\mathcal{R})$.

From the previous discussion we conclude that there is only one point of intersection which we denote with ν . This means that for $\mathcal{R} < \nu$ we have $\lambda\mathcal{R} > \tilde{\mathcal{V}}(\mathcal{R})$ while for $\mathcal{R} \geq \nu$ the inequality changes direction. From Optimal Stopping Theory [5] we then know that the optimum stopping time is given by (30). Furthermore we can also show that there exists $\nu > 0$ so that T_o meets the false-alarm constraint with equality. ■

C. Case $\{X_t\}$ Independent from $\{\xi_t\}$

If we make the additional assumption that the two processes $\{X_t\}, \{\xi_t\}$ are independent from each other, then the detector can be simplified and reduced to the well known Shiryaev test [4]. Indeed, note that $\omega_t = P_\infty(X_t \in \mathcal{A} | \xi_t) = P_\infty(X_t \in \mathcal{A}) = \omega$, where $\omega \in (0, 1)$ is a constant. Therefore, we conclude that for $t > 0$

$$\pi_t = P(\tau = t | \mathcal{F}_t) = P(\tau = t) = (1 - \pi_0)\omega(1 - \omega)^{t-1}$$

and $P(\tau \leq 0) = \pi_0$, which is the (zero modified) geometric prior adopted by Shiryaev [4] for the changetime τ . Furthermore the test becomes

$$T_o = \inf\{t > 0 : \mathcal{R}_t \geq \nu\} = \inf\{t > 0 : \tilde{\mathcal{R}}_t \geq \tilde{\nu}\}$$

where $\tilde{\mathcal{R}}_t = (\omega^{-1} - 1)\mathcal{R}_t - 1$ and $\tilde{\nu} = (\omega^{-1} - 1)\nu - 1$. From (28) we can see that $\tilde{S}_t$ can be updated using

$$\tilde{\mathcal{R}}_t = (\tilde{\mathcal{R}}_{t-1} + 1) \frac{\ell_t}{1 - \omega}, \quad (32)$$

which is the well known updating formula for the statistic of the Shiryaev test.

Remark 4. From our previous analysis we conclude that Shiryaev’s formulation can be regarded as a particular setting of our model corresponding to the case where the two processes $\{X_t\}$ and $\{\xi_t\}$ are independent from each other and each process is i.i.d. before and after the change. Furthermore, we note that the first entry of the hidden process $\{X_t\}$ into the known set $\mathcal{A}$, can be regarded as a mechanism capable of generating the required (zero modified) geometric prior which is a key assumption in Shiryaev’s setup.

IV. ACKNOWLEDGEMENT

This work was supported by the US National Science Foundation under Grant CIF 1513373, through Rutgers University.

REFERENCES

- [1] U. Niesen and A. Tchamkerten, “Tracking stopping times through noisy observations,” *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 422–432, Jan. 2009.
- [2] M. V. Burnashev and A. Tchamkerten, “Tracking a Gaussian random walk first-passage time through noisy observations,” *Annals of Applied Probability*, Vol. 22, no. 5, pp. 1860–1879, 2012.
- [3] G. V. Moustakides, “Sequential change detection revisited,” *Annals of Statistics*, vol. 36, no. 2, pp. 787–807, Apr. 2008.
- [4] A. N. Shiryaev, “On optimal methods in quickest detection problems,” *Theory of Probability and Its Applications* vol. 8, pp. 22–46, 1963.
- [5] A. N. Shiryaev, *Optimal Stopping Rules*, Springer, New York, 1978.