53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Detection Problem with Post-Change Drift Uncertainty

Heng Yang, Olympia Hadjiliadis

Abstract— **We consider the problem of detection of abrupt changes when there is uncertainty about the post-change distribution. In particular we examine this problem in the prototypical model of continuous time in which the drift of a Wiener process changes at an unknown time from zero to a random value. It is assumed that the change time is an unknown constant while the drift assumed after the change has a Bernoulli distribution with all values of the same sign independent of the process observed. We set up the problem as a stochastic optimization in which the objective is to minimize a measure of detection delay subject to a frequency of false alarm constraint. As a measure of detection delay we consider that of a worst detection delay weighed by the probabilities of the different possible drift values assumed after the change point to which we are able to compute a lower bound amongst the class of all stopping times. Our objective is to then construct low complexity, easy to implement decision rules, that achieve this lower bound exactly, while maintaining the same frequency of false alarms as the family of stopping times. In this effort, we consider a special class of decision rules that are delayed versions of CUSUM algorithm. In this enlarged collection, we are able to construct a family of computationally efficient decision rules that achieve the lower bound with equality, and then choose a best one whose performance is as close to the performance of a stopping time as possible.**

Keywords: change point, random drift, optimality, disorder problem, decision rule, min-max problem

I. INTRODUCTION

The disorder problem is concerned with detecting a change in the statistical behavior of sequential observations by balancing the trade-off between a small detection delay and a frequency of false alarms.

In this work we consider the problem of detecting a change in Wiener observations when there is uncertainty about the value of the post-change drift. In particular, we consider the case in which we only have noise before a signal arrives, which we represent by a zero drift Wiener model. The signal then arrives at an unknown constant in time otherwise known as the change point. We model the uncertainty about the post change drift by a Bernoulli distribution. In other words, we consider the case in which the signal can be a weak one represented by a small drift m_1 with a probability p, or a stronger signal represented by a larger drift m_2 with probability $1-p$. We assume that the uncertainty in the drift is independent of the observations.

This work was partially supported by the NSF-DMS grant 1222526.

Heng Yang is with the Department of Mathematics at the Graduate Center of City University of New York, hyang@gc.cuny.edu

Olympia Hadjiliadis is with the Department of Mathematics at Brooklyn College of the City University of New York, and with the Departments of Computer Science and Mathematics at the Graduate Center of City University of New York, ohadjiliadis@brooklyn.cuny.edu

Earlier studies have treated the case in which the post-change drift is a known constant after the unknown change time. In discrete time, Moustakides [11] has given the optimality of the cumulative sum (CUSUM) rule in Lorden's sense. And the optimality of the CUSUM also holds in continuous time Wiener processes as seen in Shiryaev [16], Beibel [2] and Moustakides [12]. As a result, if after the change time, the signal received has a signal strength equal to the linear combination of the weak and strong drifts, namely the drift $pm_1 + (1-p)m_2$, instead of a random assuming either the weak or the strong drift, then the optimal CUSUM stopping time is known, since the post-change drift is fixed.

In the Bayesian framework the change point is considered to be a random variable independent of the observations. In this framework, Beibel [3] and Beibel and Lerche [4] considered the case of uncertainty in the post-change drift in Wiener observations. More recently, Sezer [15] considered the case in which the post-change drift in a Wiener process is a known constant but the change time has a prior distribution which depends on the observations. The case of uncertainty in post change parameters has also been studied in Poisson observations within the Bayesian framework in Bayraktar, Dayanik and Karatzas [1].

In all of the above works the objective is to find optimal stopping times that balance a trade-off between an appropriately chosen measure of detection delay and a small probability of false alarms or a small frequency of false alarms. In other words the rules according the change point is decided are online rules in that the point in time at which they declare an alarm is also the point in time at which they estimate the location of the change point. This is in contract to many statistical works which provide frameworks for estimation techniques of one or multiple change points in an off-line fashion, that is by taking into consideration all of the observed data. In fact all such studies assume knowledge of the totality of the observation path on any time interval to provide an estimate of the change point. For a sample of such works please refer to [5], [9], [8] and [14].

In our work we initially consider the problem of online detection of the unknown change point in the presence of uncertainty in the drift and adopt a min-max approach of estimation of the change point. In this effect we consider a weighted average of a Lorden type measure of detection delay [10] with weights given by the probabilities of the Bernoulli distribution that captures the post-change drift uncertainty. The objective is to minimize this measure of detection delay subject to a constraint on the mean time to the first false alarm. We first compute a lower bound on the detection delay of all stopping times according to this measure, then enlarge the family of rules considered by allowing all decision rules which are a delayed version of stopping times multiplied by a positive constant which can take values less than unity. The idea is that following these decision rules the alarm is drawn according to a given stopping time but the estimation of the location of the change point is then given as the product of the constant and the time at which the stopping time alarm goes off. Clearly the closer the constant is to unity the more "online" is the estimation of the change point. Enlarging the class of rules considered beyond stopping times allows us to build low-complexity, computationally efficient schemes of estimation that, for the same frequency of false alarms as their stopping time counterparts, achieve exactly the lower bound of detection delay and are easy to implement. In this effort we find that a family of decision rules that use a λ parameter CUSUM statistic [6] [7] do achieve the lower bound with equality. It is then possible to select amongst them the decision rule with a constant factor as close to unity as possible that in fact often results in a very slight deviation from unity for a large number of parameter values.

In section II, we set up the problem mathematically by defining appropriate measures and filtrations and propose a new criterion to measure detection delay. We also derive a lower bound for the detection delay criterion for the family of all stopping times that satisfy the false alarm constraint. In section III, we consider an enlarged collection of decision rules, which contains not only the stopping times but also the delayed version of stopping times multiplied by positive constants less than unity. We show that there is a family of the decision rules in this class that achieve the lower bound of detection delay with equality and we choose the best decision rule in this family, namely the one whose performance is as close to the performance of a stopping time as possible. In section IV, we provide the examples and discuss the performance of the optimal rule we found. In section V, we give all the proofs of the theorems and lemmas.

II. THE OBSERVATIONS AND THE DELAY

Let $(\Omega, \mathcal{F})$ be a sample space. We observe the process $\{Z_t\}_{t\geq 0}$ on this space with initial value $Z_0 = 0$. Assume that there may be a change in the distribution of the observations process at the fixed but unknown time τ. When there is no change, we use P_∞ to denote the measure generated by $\{Z_t\}_{t\geq 0}$. It is the standard Wiener measure. When there is a change, assume the observation process changes from a standard Brownian motion to a Brownian motion with drift m; that is

$$dZ_t := \begin{cases} dW_t & t < \tau \\ mdt + dW_t & t \geq \tau. \end{cases} \tag{1}$$

We suppose that the post-change drift m can take the value m_1 with probability p and the value m_2 with probability $1-p$ respectively, for a parameter $p \in (0,1)$, that means m is a random variable with a Bernoulli distribution. And we suppose m is independent of the pre-change Brownian motion process $\{W_t\}_{t<\tau}$. Let P^m correspond to the measure generated by $\{m\}$. In our analysis, we assume that m_1 and m_2 are both positive. The case of negative drifts can be addressed by similar arguments.

To facilitate our analysis, let $\mathcal{G}_t$ correspond to the $\sigma-$algebra generated by the observation $\{Z_s\}_{s\leq t}$. Denote $\mathcal{G}_\infty = \bigcup_{t\geq 0} \mathcal{G}_t$, and so we have a filtration $\{\mathcal{G}_t\}_{t>0}$. We introduce the family of measures $P_\tau^{m_i}$ defined on this filtration, where $\tau \in [0, \infty)$ and $i = 1, 2$. $P_\tau^{m_i}$ is defined to be the measure generated by $\{Z_t\}_{t\geq 0}$, when $dZ_t = dW_t$ for $t < \tau$, and $dZ_t = m_i dt + dW_t$ for $t \geq \tau$. Thus under $P_0^{m_i}$, the observation process is a Brownian motion with drift: $dZ_t = m_i dt + dW_t$ for any $t \geq 0$.

Our objective is to find a $\mathcal{G}$-stopping rule that balances the trade off between the small detection delay and the lower bound of the time interval between two false alarms.

In this problem, we need to construct a new measure of detection delay that takes into account the distribution of both the random drift m and the observation path. For any $\mathcal{G}$-stopping time R, we define its detection delay given the post-change drift $m = m_i$ for $i = 1, 2$ as

$$J_i(R) := \sup_{\tau \geq 0} \operatorname{esssup} E_\tau^{m_i}[(R-\tau)^+ | \mathcal{G}_\tau]. \tag{2}$$

The detection delay of the $\mathcal{G}$-stopping time R is defined as

$$J(R) := pJ_1(R) + (1-p)J_2(R). \tag{3}$$

Here, we take the essential supremum over all path up to time τ and take supremum over all possible change time. Then, we take the average worst delay over all possible values of m which has a Bernoulli distribution.

We require a constraint on the mean time of the first false alarm, namely $E_\infty[R] \geq \gamma$. In fact, by using an argument similar to [11], we just need to consider the stopping times with $E_\infty[R] = \gamma$.

Let $\mathcal{S}$ be the collection of all $\mathcal{G}$-stopping times for which $E_\infty[R] = \gamma$. By the above setup, our objective becomes to minimize the detection delay $J(R)$ in (3) over all $\mathcal{G}$-stopping times in $\mathcal{S}$. For this purpose, we first find a lower bound of the detection delay for any $R \in \mathcal{S}$ in the following result:

Lemma 1: For any stopping time $R \in \mathcal{S}$, we have a lower bound of the delay as

$$J(R) \geq \frac{2p}{m_1^2} g\left(h^{-1}(\frac{m_1^2\gamma}{2})\right) + \frac{2(1-p)}{m_2^2} g\left(h^{-1}(\frac{m_2^2\gamma}{2})\right), \tag{4}$$

where $\quad h(x) := e^x - x - 1 \quad$ and $\quad g(x) := e^{-x} + x - 1 \quad (5)$

are increasing function as $x \geq 0$.

We denote the right hand side of (4) as LB, which is a constant depending on γ, m_1, m_2 and p. Unfortunately, we can not find a $\mathcal{G}$-stopping time that achieves LB with equality. So we enlarge the collection $\mathcal{S}$ of the $\mathcal{G}$-stopping times, by considering the delayed version of stopping times in the following section.

III. THE ENLARGED DETECTION PROBLEM AND THE OPTIMAL DECISION RULE

A. The Enlarged Detection Problem

Let $\mathcal{R}$ be the family of all rules of the form $R = C\hat{R}$, where $\hat{R}$ is a $\mathcal{G}$-stopping time, and $0 < C \leq 1$ is a constant positive coefficient, satisfying the false alarm $E_\infty[R] = \gamma$ and the delay $J(R) \geq LB$, which is the lower bound of the detection delay in Lemma 1. That is

$$\mathcal{R} := \left\{ \begin{array}{c} C\hat{R} : 0 < C \leq 1; \hat{R} \text{ is a } \mathcal{G}\text{-stopping time} \\ E_\infty[R] = \gamma; J(R) \geq LB \end{array} \right\}. \quad (6)$$

It is easy to see that $\mathcal{S} \subset \mathcal{R}$. The enlarged collection $\mathcal{R}$ contains not only the $\mathcal{G}$-stopping times in $\mathcal{S}$ but also some products of a $\mathcal{G}$-stopping time and a constant. A rule of the form $R = C\hat{R}$ is $\mathcal{G}$-measurable, so we can still use the detection delay in (3) to measure its performance.

Our enlarged problem is the minimization problem

$$\inf_{R \in \mathcal{R}} J(R). \quad \text{(P)}$$

By the definition of the enlarged class $\mathcal{R}$, the lower bound of the detection delay of any rule in $\mathcal{R}$ will be the same as that of the rule in $\mathcal{S}$ in (4), which is LB. In the following, we would like to find a family of rules in $\mathcal{R}$ to achieve LB, so that such rule has the least delay in $\mathcal{R}$.

As a remark, we call the rule in $\mathcal{R}$ to be a decision rule, since when the coefficient is smaller than 1, the weighted rule $R = C\hat{R}$ is no longer a $\mathcal{G}$-stopping time. In fact, if an alarm is drawn at time $\hat{R}$, the decision rule R claims the change position at $R = C\hat{R}$. If the coefficient is close to 1, the time at which we stop is not far from the position at which we claim the location of the change point. For example, when we use rule $0.98\hat{R}$ and the stopping time $\hat{R}$ draws an alarm at time 100, this decision rule declares the change at time $t = 98$.

B. The Decision Rule $T_{\lambda,C}$

We want to build the decision rules in $\mathcal{R}$, whose delays are equal to the lower bound LB in Lemma 1.

When the post-change drift is a known constant λ, which means $p = 1$ and $m_1 = \lambda$ in our setup, the optimal stopping rule is the CUSUM stopping time S_λ as in [6], that is

$$\begin{aligned} S_\lambda &:= \inf\{t > 0, V_t - \inf_{s \leq t} V_s \geq \nu\}, \\ \text{where} \quad V_t &:= \lambda Z_t - \frac{1}{2}\lambda^2 t \quad \text{and} \quad \nu > 0. \end{aligned} \quad (7)$$

Inspired by the optimality of the CUSUM stopping time when the post-change drift is known, we define a delayed version of CUSUM stopping time with tuning parameter λ

$$T_{\lambda,C} := CS_\lambda, \quad (8)$$

where $0 < C \leq 1$ is a constant parameter and S_λ is a CUSUM $\mathcal{G}$-stopping time with $\lambda > 0$ and $\nu > 0$ in (7). We can get its detection delay from the following lemma.

Lemma 2: For the decision rule $T_{\lambda,C}$ defined in (8), which satisfies $E_\infty[T_{\lambda,C}] = \gamma$, we have its detection delay

$$J(T_{\lambda,C}) = \frac{2pCg(\theta_1\nu)}{\lambda^2\theta_1^2} + \frac{2(1-p)Cg(\theta_2\nu)}{\lambda^2\theta_2^2}, \quad (9)$$

and ν can be represent as a function of C and λ in

$$\nu = h^{-1}(\frac{1}{2C}\lambda^2\gamma), \quad (10)$$

where

$$\theta_i := \frac{2m_i - \lambda}{\lambda} \quad (11)$$

is decreasing and $\theta_i > -1$ for $i = 1, 2$.

Our purpose is to find the parameters of the rule $T_{\lambda,C} \in \mathcal{R}$ to make the detection delay be equal to the lower bound in (4), which is given the following result.

Theorem 1: In the collection $\mathcal{R}$, there exists a family of the decision rules that solve the problem (P). More precisely, concerning the decision rule $T_{\lambda,C} = CS_\lambda$ in (8), for any parameter $\lambda > 0$, there exists a unique value C, namely C_λ, such that $J(T_{\lambda,C_\lambda}) = LB$, where LB is the lower bound of detection delay in (4).

We denote

$$T_\lambda := T_{\lambda,C_\lambda}. \quad (12)$$

Theorem 1 provides a family of rules $\{T_\lambda\}_{\lambda>0}$, whose delays reach the lower bound (4). We want a method to choose a best rule from this family.

Our original objective was to find a stopping time. So we want to choose one rule from the family $\{T_\lambda\}_{\lambda>0}$, whose behavior is as close to the behavior of a stopping time as possible. By running rule T_λ, we stop at S_λ and declare the estimatation of the change point to be $C_\lambda S_\lambda$. This suggests that the ideal choice of λ is the one to maximize C_λ.

Theorem 2: In the family of decision rules of the form $\{T_\lambda\}_{\lambda>0}$ in Theorem 1, there exists a rule $T_{\lambda^*} := C_{\lambda^*}S_{\lambda^*}$ whose behavior is closest to the behavior of a stopping time, for a $\lambda^* \in (m_1, m_2)$. More precisely, there exists a λ^* in (m_1, m_2) to maximize the coefficient C_λ.

Since the maximum of C_λ is located in (m_1, m_2), the number of values of λ to reach the maximum is finite. In case that there are more than one values of λ which give the maximum of C_λ in Theorem 2, we can choose any one of them, such as the smallest one, namely λ^*.

IV. EXAMPLES AND DISCUSSION

Given the values of m_1, m_2, p and γ, we would like to describe the method of choosing the parameters λ^* and C_{λ^*} used in the construction of the optimal rule T_{λ^*}. First, the equation (10) represents the threshold ν as a function of λ and C_λ. Then by equalizing expression (9) to the lower bound in (4), we obtain an equation involving two unknowns C_λ and λ. Now the objective becomes to identify the maximum C_λ by appropriately choosing λ, which can be achieved numerically always. Since $\lambda^* \in (m_1, m_2)$, in the iteration algorithm to find the maximum, we can always set the initial value of λ to be m_1. This produces the value of optimal choice λ^*, and thus leads to the rule T_{λ^*}.

Since T_{λ^*} is not always a stopping time, we are interested in the difference between the time that the alarm is drawn and the estimation of change time.

	$\gamma = 50$		$\gamma = 100$		$\gamma = 500$	
$p =$	alarm	change	alarm	change	alarm	change
0	2.79	2.79	3.36	3.36	4.75	4.75
0.1	3.01	2.99	3.66	3.63	5.26	5.19
0.2	3.23	3.19	3.95	3.9	5.74	5.62
0.3	3.44	3.4	4.24	4.17	6.19	6.06
0.4	3.65	3.6	4.51	4.44	6.63	6.49
0.5	3.85	3.81	4.78	4.71	7.06	6.93
0.6	4.05	4.01	5.04	4.98	7.48	7.36
0.7	4.25	4.22	5.3	5.24	7.89	7.8
0.8	4.45	4.42	5.55	5.51	8.3	8.23
0.9	4.64	4.62	5.8	5.78	8.7	8.67
1	4.83	4.83	6.05	6.05	9.1	9.1

TABLE I

THE CASE $m_1 = 1, m_2 = 1.5$

In Table I, given $m_1 = 1$, $m_2 = 1.5$ and $\gamma = 50, 100, 500$, we list the average time we stop under column "alarm", and the estimations of the change time under column "change", for different values of p from 0 to 1. We can see that when p is close to 0 or 1, the difference between the estimation of the change time and the time we stop is small. When p is far from 0 and 1, the difference is larger, since in such case it is harder to figure out the post-change drift value. In particular, the case $p = 0$ gives the CUSUM stopping time with tuning parameter m_2, and $p = 1$ gives the CUSUM stopping time with tuning parameter m_1. On the other hand, when γ increases, the threshold ν in the CUSUM stopping time increases. Then more time is necessary to declare the alarm, and thus the difference between the estimation of the change time and the time we stop gets larger.

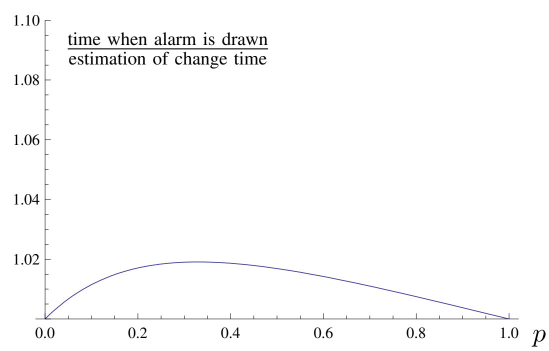

Fig. 1. The case $m_1 = 2$, $m_2 = 3$, $\gamma = 50$

In Figure 1, we consider the case $m_1 = 2, m_2 = 3, \gamma = 50$. The graph shows the ratio of the time at which the alarm is drawn and the estimation of the change time. We can see that the ratio is small when the post-change drift is more likely to be one specific value, and is large when it is hard to figure out the value of the post-change drift.

As a discussion, the idea of decision rule T_{λ^*} is to improve the performance of CUSUM stopping times. The CUSUM stopping times that satisfy the constraint on false alarms are in the collection $\mathcal{R}$. But their detection delays are always larger than the lower bound LB. Our strategy in T_{λ^*} is to modify the CUSUM stopping time to make the delay equal to the lower bound by weighting the stopping time, and then find a particular delayed CUSUM rule whose weight is closest to 1. Thus we can have an improved decision rule with a smaller delay and the same false alarm constraint as CUSUM stopping times.

V. PROOFS OF THEOREMS AND LEMMAS

Proof of Lemma 1:

The detection delay given the post-change drift $m = m_i$ in (2) is that of Lorden's criterion [10]. Since R is a $\mathcal{G}$-stopping time with $E_\infty[R] = \gamma$, from the optimality of the Cumulative Sum $\mathcal{G}$-stopping time in the case that the post change drift is known to be $m = m_i$ (see [16]), we know that

$$J_i(R) \geq E_0^{m_i}[S_{m_i}] = \frac{2}{m_i^2} g\left(h^{-1}(\frac{m_i^2}{2}\gamma)\right), \qquad (13)$$

where S_{m_i} is the CUSUM stopping time defined in (7). So we have the inequality (4). □

Proof of Lemma 2:

By simple computation (see [6] and [13]), for $i = 1, 2$, we have

$$E_\infty[S_\lambda] = \frac{2}{\lambda^2} h(\nu) \quad \text{and} \quad E_0^{m_i}[S_\lambda] = \frac{2g(\frac{2m_i - \lambda}{\lambda}\nu)}{(2m_i - \lambda)^2}. \qquad (14)$$

To compute the detection delay of decision rule $T_{\lambda,C} = CS_\lambda$, we use the fact that the worst detction delay over all possible paths will occur when the process $Y_t = V_t - \inf_{s \leq t} V_s$ is equal to 0 at time τ. That is, the worst detection delay takes place on those paths for which $\{Y_\tau = 0\}$, which is the same loaction for the Y_t process as the one that it takes at time 0 since $Y_0 = 0$. By Markov property, we have

$$\begin{aligned} \operatorname{esssup} E_\tau^{m_i}[(T_{\lambda,C} - \tau)^+ | \mathcal{G}_\tau] &= E_\tau^{m_i}[(CS_\lambda - \tau)^+ | Y_\tau = 0] \\ &= E_0^{m_i}[CS_\lambda]. \end{aligned} \qquad (15)$$

From (2), (14) and (15), for $i = 1, 2$ we obtain

$$J_i(T_{\lambda,C}) = \frac{2Cg(\theta_i \nu)}{\lambda^2 \theta_i^2}. \qquad (16)$$

From equations (3) and (16), we obtain (9).

Also, from $E_\infty[T_{\lambda,C}] = \gamma$ and (14), it follows that

$$\frac{2C}{\lambda^2} h(\nu) = \gamma. \qquad (17)$$

Since $h(x)$ is increasing on $[0, \infty)$, we obtain (10). □

Result 1: The function

$$r(x) := x \frac{e^x - 1}{e^x - x - 1} \qquad (18)$$

is positive and strictly increasing on $x \in (-\infty, \infty)$, with $r(0) = 2$, $r'(0) = 1/3$, $\lim_{x \to -\infty} r(x) = 1$ and $\lim_{x \to \infty} r(x) = \infty$.

Proof: The derivative of $r(x)$ is

$$r'(x) = \frac{(e^x-1)^2 - x^2 e^x}{(e^x-x-1)^2}. \quad (19)$$

Denote $r_1(x) = (e^x-1)^2 - x^2e^x$. We have $r_1(0) = 0$ and $r_1'(x) = 2e^x(e^x - 1 - x - \frac{1}{2}x^2)$. It is easy to see that $r_1'(x) > 0$ when $x > 0$ by Taylor expansion, and $r_1'(x) < 0$ when $x < 0$ by taking derivative of the term $e^x - 1 - x - \frac{1}{2}x^2$ twice. Then $r_1(x) > 0$ when $x \neq 0$, and so $r'(x) > 0$ when $x \neq 0$.

It is also easy to get $r(0) = 2$, $r'(0) = 1/3$, $r(-\infty) = 1$ and $r(\infty) = \infty$. Since $r'(0) = 1/3$, the function $r(x)$ is strictly increasing on $x \in (-\infty, \infty)$, and thus r(x) is positive on $x \in (-\infty, \infty)$. □

Result 2: The function

$$K(x) := \frac{e^x - x - 1}{x(e^x-1)} \quad (20)$$

is positive and strictly decreasing on $x \in (-\infty, \infty)$, with the values $K(0) = 1/2$, $K'(0) = -1/12$, $\lim_{x\to-\infty} K(x) = 1$ and $\lim_{x\to\infty} K(x) = 0$. Moreover, $K(x)$ is concave on $(-\infty, 0)$ and convex on $(0, \infty)$. And the graph of $K(x)$ is symmetric with respect to the point $(0, K(0))$.

Proof: We know $K(x)$ is strictly decreasing on $x \in (-\infty, \infty)$ with $K(0) = 1/2$ and $K'(0) = -1/12$ from Result 1. By computing

$$K(x) + K(-x) = \frac{-(e^x-1) - (e^{-x}-1)}{(e^x-1)(e^{-x}-1)} = 1, \quad (21)$$

we can get $1/2 - K(x) = K(-x) - 1/2$ for any x. Since $K(0) = 1/2$, the graph of $K(x)$ is symmetric with respect to the point $(0, K(0))$.

We have $K''(x) = \frac{2(e^x-2+e^{-x})^2 - x^3(e^x-e^{-x})}{x^3(e^x-2+e^{-x})^2}$. On $(0, \infty)$, the denominator is always positive. Denote the numerator as $K_1(x) = 2(e^x - 2 + e^{-x})^2 - x^3(e^x - e^{-x})$. Wh have $K_1(0) = 0$ and $K_1'(x) = 4(e^{2x} - e^{-2x}) - 8(e^x - e^{-x}) - 3x^2(e^x - e^{-x}) - x^3(e^x + e^{-x})$. To show $K_1'(x) > 0$ on $x > 0$, we use the Taylor expansion in each term to get

$$K_1'(x) = \sum_{n=1}^{\infty} \frac{8(2^{2n+1} - 2 - n - 3n^2 - 2n^3)}{(2n+1)!} x^{2n+1}. \quad (22)$$

Denote $s(n) = 2^{2n+1} - 2 - n - 3n^2 - 2n^3$. It is easy to see that $s(0) = s(1) = s(2) = 0$, $s(3) = 42$ and $s(n) > 0$ when $n \geq 3$. Then when $x > 0$, we have $K_1'(x) > 0$ and thus $K_1(x) > K_1(0) = 0$. And then $K''(x) > 0$ when $x > 0$, which leads to the result that $K(x)$ is convex when $x > 0$. By symmetry, $K(x)$ is concave when $x < 0$. □

Result 3: The function

$$L(x) = \frac{1 - e^{-x}}{x} \quad (23)$$

is positive, strictly decreasing and convex on $(-\infty, \infty)$.

Proof: It is easy to see that $L(x)$ is positive and its derivative is $L'(x) = -\frac{e^x - x - 1}{x^2 e^x} < 0$ with $L'(0) = -1/2$.

And we have $L''(x) = \frac{2}{x^3 e^x}(e^x - 1 - x - \frac{1}{2}x^2)$, with $L''(0) = 1/3$. When $x > 0$, $L''(x) > 0$ is given by Taylor expansion. When $x < 0$, we can see $e^x - 1 - x - \frac{1}{2}x^2 < 0$ by taking derivative twice, and thus $L''(x) > 0$. □

Proof of Theorem 1:

1) For any fixed $\lambda > 0$, to show that there exists a C to satisfy the equality $J(T_{\lambda,C}) = LB$, we first notice that the delay $J(T_{\lambda,C})$ in (9) is a continuous function of $C \in (0, 1]$.

At $C = 1$, we have $T_{\lambda,1} = S_\lambda$, which is a CUSUM $\mathcal{G}$-stopping time in $\mathcal{S}$, and so $J(T_{\lambda,1}) = J(S_\lambda) \geq LB$.

As $C \to 0^+$, from $e^\nu - \nu - 1 = \frac{\lambda^2\gamma}{2C}$, we have

$$\lim_{C\to0^+} \frac{\nu}{\ln C^{-1}} = 1. \quad (24)$$

When $\lambda < 2m_i$, we have $\theta_i > 0$ for $i = 1, 2$, and then

$$\lim_{C\to0^+} Cg(\theta_i\nu) = \lim_{C\to0^+} C\theta_i\nu = \lim_{C\to0^+} \theta_i C \ln C^{-1} = 0. \quad (25)$$

When $\lambda > 2m_i$, we have $-1 < \theta_i < 0$ for $i = 1, 2$, and then

$$\lim_{C\to0^+} Cg(\theta_i\nu) = \lim_{C\to0^+} Ce^{-\theta_i\nu} = \lim_{C\to0^+} C^{1+\theta_i} = 0. \quad (26)$$

When $\lambda = 2m_i$, we have $g(\theta_i\nu)/\theta_i^2 = \nu^2/2$, and then

$$\lim_{C\to0^+} \frac{C}{\theta_i^2\lambda^2} g(\theta_i\nu) = 0. \quad (27)$$

Thus from (24), (25), (26) and (27), we get $J(T_{\lambda,0^+}) = 0$.

Since $J(T_{\lambda,1}) \geq LB$ and $J(T_{\lambda,0^+}) = 0$, by continuity of the delay function $J(T_{\lambda,C})$ in (9), there exists a value of $C_\lambda \in (0, 1]$ such that $J(T_{\lambda,C_\lambda}) = LB$, for any $\lambda > 0$.

2) For uniqueness of C_λ, we take the derivative of the delay $J(T_{\lambda,C})$ with respect to C in (9).

From (16), for $i = 1, 2$, we can get

$$\frac{d}{dC} J_i(T_{\lambda,C}) = \frac{2\nu^2}{\lambda^2} L(\theta_i\nu) \left(K(-\theta_i\nu) - K(\nu)\right). \quad (28)$$

where $K(x)$ is defined in (20) and $L(x)$ is defined in (23). Since $\theta_i > -1$, we have $-\theta_i\nu < \nu$. Thus by Result 2 and Result 3, we have $\frac{d}{dC} J_i(T_\lambda) > 0$, for $i = 1, 2$. Then from (3), $J(T_{\lambda,C})$ is increasing in C.

From existence and uniqueness, there exists a unique $C_\lambda \in (0, 1]$ to satisfy $J(T_{\lambda,C_\lambda}) = LB$, for any $\lambda > 0$. Thus $T_{\lambda,C_\lambda} \in \mathcal{R}$ and it solves the problem (P). □

Proof of Theorem 2:

1) From Theorem 1, C_λ is a function of λ. By equations (9) and (10), the delay $J(T_\lambda)$ with parameters (λ, C_λ) is also a function of λ. Thus we can compute the derivatives of C_λ and $J(T_\lambda)$ with respect to λ.

By computation and equation (16), for $i = 1, 2$, we have

$$\frac{d}{d\lambda} J_i(T_\lambda) = \frac{2\nu^2}{\lambda^2} L(\theta_i\nu) \left(A(\nu, \theta_i)\frac{dC_\lambda}{d\lambda} - B(\nu, \theta_i)\frac{2C_\lambda}{\lambda}\right), \quad (29)$$

where $A(\nu, x) := K(-x\nu) - K(\nu)$, (30)

and $B(\nu, x) := \frac{1}{x}\left(K(x\nu) - \frac{1}{2} - x\left(K(\nu) - \frac{1}{2}\right)\right)$. (31)

From the constraint $J(T_\lambda) = LB$ in Theorem 1, we have $\frac{d}{d\lambda}J(T_\lambda) = 0$. Combining with equations (3) and (29), we can get the derivative of C_λ with respect to λ as

$$\frac{dC_\lambda}{d\lambda} = \frac{2C_\lambda}{\lambda} \frac{pL(\theta_1\nu)B(\nu,\theta_1) + (1-p)L(\theta_2\nu)B(\nu,\theta_2)}{pL(\theta_1\nu)A(\nu,\theta_1) + (1-p)L(\theta_2\nu)A(\nu,\theta_2)}. \tag{32}$$

To check the sign of $dC_\lambda/d\lambda$, we need to figure out the behavior of $A(\nu,x)$ and $B(\nu,x)$.

2) It is easy to check the behavior of the denominator in (32). Since $\theta_i > -1$, we have $-\theta_i\nu < \nu$. By Result 2, $K(x)$ is decreasing on $(-\infty,\infty)$, thus we have $A(\nu,\theta_i) > 0$ for $i = 1,2$ and $\lambda > 0$. From Result 3, $L(\theta_i\nu) > 0$ for $i = 1,2$ and thus the denominator in (32) is positive for $\lambda > 0$.

The behavior of $B(\nu,x)$ is related to the convexity of $K(x)$. By Result 2, $K(x) - 1/2$ is convex on $x > 0$ and concave on $x < 0$, and $K(x)-1/2$ is symmetric with respect to the point $(0, K(0) - 1/2) = (0,0)$.

When $x > 1$, we have $x\nu > \nu > 0$, and by convexity, $|K(x\nu) - 1/2|/|K(\nu) - 1/2| < x$. Since $K(x\nu) - 1/2 < K(\nu) - 1/2 < 0$, we get $K(x\nu) - 1/2 > x(K(\nu) - 1/2)$, which means $B(\nu,x) > 0$.

When $x = 1$, it is easy to see that $B(\nu,1) = 0$.

When $0 < x < 1$, we have $\nu > x\nu > 0$, and by convexity, $|K(x\nu)-1/2|/|K(\nu)-1/2| > x$ and $K(\nu)-1/2 < K(x\nu)-1/2 < 0$. So we can get $B(\nu,x) < 0$.

When $x = 0$, we have $B(\nu,0) = -\nu/12 - (K(\nu) - 1/2)$. We just need to take derivative with respect to ν to check $B(\nu,0)$ is decreasing in ν, and so $B(\nu,0) < 0$.

When $-1 < x < 0$, we have $K(x\nu) - 1/2 > 0 > K(\nu) - 1/2$. This gives $B(\nu,x) < 0$.

So we have

$$B(\nu,x) \begin{cases} > 0, & \text{when} \quad x > 1 \\ = 0, & \text{when} \quad x = 1 \\ < 0, & \text{when} \quad -1 < x < 1. \end{cases} \tag{33}$$

3) Now we can see the existence of the local maximum of C_λ as $\lambda \in (m_1, m_2)$.

When $0 < \lambda < m_1$, we have $\theta_1 > 1$ and $\theta_2 > 1$. Then from (33), we know $B(\nu,\theta_1) > 0$ and $B(\nu,\theta_2) > 0$. Thus in (32), we can see that $dC_\lambda/d\lambda > 0$.

When $\lambda = m_1$, we have $\theta_1 = 1$ and $\theta_2 > 1$. Then from (33), we know $B(\nu,\theta_1) = 0$ and $B(\nu,\theta_2) > 0$. Thus in (32), we can see that $dC_\lambda/d\lambda > 0$.

When $\lambda = m_2$, we have $-1 < \theta_1 < 1$ and $\theta_2 = 1$. Then from (33), we know $B(\nu,\theta_1) < 0$ and $B(\nu,\theta_2) = 0$. Thus in (32), we can see that $dC_\lambda/d\lambda < 0$.

When $\lambda > m_2$, we have $-1 < \theta_1 < 1$ and $-1 < \theta_2 < 1$. Then from (33), we know $B(\nu,\theta_1) < 0$ and $B(\nu,\theta_2) < 0$. Thus in (32), we can see that $dC_\lambda/d\lambda < 0$.

Since C_λ is increasing at $\lambda \le m_1$ and is decreasing at $\lambda \ge m_2$, there exists a maximum on (m_1, m_2). □

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we consider the problem of detection when the change time is an unknown constant. Our continuous sequential observations change from the standard Wiener process to Wiener process of drift m_1 with probability p, or to Wiener process of drift m_2 with probability $1-p$, where m_1 and m_2 are known constants which are both positive. Although we are unable to find stopping times to solve this problem, we demonstrate that it is possible to construct an easy to implement family of decision rules that achieve the lower bound of detection delay while in fact achieve a larger mean time to false alarm than their stopping time counterparts. These decision rules are delayed version of stopping times. Although, according to these decision rules, the change point is not declared when the alarm is drawn, the solution is still implementable online in that once the alarm is drawn an estimate of the change point is readily available. A problem of interest to consider in the future is that of detection problem when a general distribution on the random variable m is assumed. Another interesting problem is one in which we are uncertain about the value of p which would lead to the consideration of a family of different measures within the Bernoulli framework for the random post-change drift m.

VII. ACKNOWLEDGMENTS

We acknowledge the support of NSF-DMS grant 1222526 for this research.

REFERENCES

[1] E. Bayraktar, S. Dayanik and I. Karatzas, Adaptive Poisson Disorder problem, *The Annals of Applied Prob.* Vol.16, No.3, 1190-1261, 2006.

[2] M. Beibel, A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure, *Annals of Statistics*, Vol. 24, No. 2, pp. 1804-1812, 1996.

[3] M. Beibel, Sequential change-point detection in continuous time when the post-change drift is unknown, *Bernoulli* 3 457478, 1997.

[4] M. Beibel and H.R. Lerche, Sequential Bayes detection of trend changes, *Foundations of Statistical Inference(Y. Haitovsky, H.R. Lerche and Y. Ritov, eds.)* 117130, 2003.

[5] P.K.Bhattacharya, Some Aspects Of Change-point Analysis, *IMS Lecture Notes - Monograph Series*, Vol. 23, 1994

[6] O. Hadjiliadis, Optimality of the 2-CUSUM drift equalizer rules for detecting two-sided alternatives in the Brownian motion model, *Journal of Applied Probability*, Issue 4, Vol.42, pp. 1183-1193, 2005.

[7] O. Hadjiliadis and G. V. Moustakides, Optimal and Asymptotically Optimal CUSUM rules for change point detection in the Brownian Motion model with multiple alternatives, *Theory of Probability and its Applications* , issue 1, vol.50, pp. 131-144, 2006

[8] H.J.Kim and D. Siegmund, The Likelihood Ratio Test For a Change-Point in Simple Linear Regression, *Biometrika 76*, 409-423, 1989.

[9] H.J.Kim, Tests For a Change-Point in Linear Regression, *IMS Lecture Notes - Monograph Series*, Vol. 23, 1994.

[10] G. Lorden, Procedures for reacting to a change in distribution, *The Annals of Mathematical Statistics*, Vol. 42, No. 6, 1897-1908, 1971.

[11] G. V. Moustakides, Optimal stopping times for detecting changes in distributions, *Annals of Statistics*, Vol. 14, No. 4, pp. 1379-1387, 1986.

[12] G. V. Moustakides, Optimality of the CUSUM procedure in continuous time, *Annals of Statistics*, Vol. 32, No. 1, pp. 302-315, 2004.

[13] H.V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, Cambridge, UK, 2008.

[14] A.L. Rukhin, Asymptotic Minimaxity In The Change-Point Problem, *IMS Lecture Notes - Monograph Series*, Vol. 23, 1994

[15] S.O. Sezer, On the Wiener disorder problem, *The Annals of Aplied Prob.*, Vol. 20, No. 4, 1537-1566, 2010.

[16] A.N. Shiryaev, Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time, *Russian Mathem. Surv.* 51, 750-751, 1996.