Change-point detection of two-sided alternatives in the Brownian motion model and its connection to the gambler's ruin problem with relative wealth perception

Olympia Hadjiliadis

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2005

©2005

Olympia Hadjiliadis

All Rights Reserved

Change-point detection of two-sided alternatives in the Brownian motion model and its connection to the gambler's ruin problem with relative wealth perception

Olympia Hadjiliadis

This thesis addresses the problem of change-point detection in the Brownian motion model with multiple alternatives. Attention is drawn to the 2-CUSUM stopping time and its properties as a means of detecting a two-sided change. It is shown that the 2-CUSUM stopping rule is second-order asymptotically optimal as the frequency of false alarms tends to infinity. The above problem can be related to the gambler's ruin problem in which gamblers make their decisions to quit the game based on the relative change in their wealth. Probabilities of exiting after a significant upward rally in the gambler's wealth (or a significant downward fall) are worked out both in the discrete time framework and in the continuous time framework.

Contents

List of Figures **iii**

Acknowledgments **iv**

Chapter 1 Introduction **1**

Chapter 2 CUSUM rules for detecting a regime change in the Brownian motion model with multiple alternatives **13**

2.1 Mathematical formulation of the problem. 15

2.2 The one-sided CUSUM stopping time 17

2.3 Different drift signs and the 2-CUSUM stopping time 22

2.3.1 The harmonic mean 2-CUSUM rules 23

2.3.2 2-CUSUM drift equalizer rules 25

2.4 Asymptotic optimality in opposite sign drifts 28

2.4.1 The case of equal in absolute value drifts 29

2.4.2 The case of different in absolute value drifts 31

2.5 Conclusions & Future Work 33

Chapter 3 Optimality of the 2-CUSUM drift equalizer rules among

the harmonic mean 2-CUSUM rule class for detecting two-sided alternatives in the Brownian motion model **36**

3.1 The 2-CUSUM rules & the harmonic mean rule (revisited) 37

3.2 Equalizer rules are best 40

3.3 Conclusions 45

Appendix A **46**

A.1 46

A.2 47

A.3 48

A.4 49

Chapter 4 The gambler's ruin problem with relative wealth perception **51**

4.1 Gambler's ruin problem with relative wealth perception 52

4.1.1 The discrete time framework 52

4.1.2 The continuous time framework 65

4.2 Conclusions & Future work 74

Appendix B **76**

List of Figures

2.1 Typical form of the upper and lower bounds of the performance of the optimum stopping rule for the case $\mu_1 = -\mu_2 = 1$. 31

2.2 Typical form of the upper and lower bounds of the performance of the optimal stopping rule for the case $\mu_2 < 0 < \mu_1$, with $\mu_1 = 1$ and $\mu_2 = -1.05, -1.15, -1.3$. 32

Acknowledgments

I would like first of all to thank my advisor Dr. Večeř for his constant support in the difficult and in the good times. Dr. Večeř has always been available to address all of the issues that have come up through this time as well as offer his valuable help. He is a good listener and a patient man. He is also able to help me prioritize matters in order to accomplish my goals. Jan is able to channel positive energy in the way that proves most effective. I am very grateful to Dr. Moustakides who is the person that not only introduced me to the problem of change-point detection but also guided me through the writing of my very first paper. The results that appear in the first Chapter were done in collaboration with Dr. Moustakides. In particular two of the results, that appear in Sections 2.2 and 2.4.1 and Theorem 2.2 are his contribution. Lemma 2.4 was also inspired by his observation. Dr. Moustakides was also the one that suggested the problem whose solution appears in the second Chapter of this thesis, pointed out the existence of the third case in it, and helped with the calculation that appears in Appendix A.1. Dr. Moustakides is an honest man with very good intentions towards his students.

I would like to thank Dr. Rabinowitz for challenging me with very interesting problems and offering me his insight in the change-point problems. I also appreciate

all of the time he spent advising me on the career decisions that could be beneficial to me in later life. Dr. Rabinowitz has also always been available to answer any questions I have. I am grateful to him for offering me the invaluable opportunity to teach that is something that I literary enjoy. Moreover, I would like to thank Dr. Heyde for his interest in my work, his offer to serve as a member of my Committee as well as his advice on the submission of my second paper. Dr. Heyde examined my thesis very thoroughly and offered very useful suggestions for improvement. This is a process that requires both time and effort and I would like to say that I am very grateful to him for this. I would like to also thank Dr. Mishra for his interest in my work, his invitation to the Courant Institute and his offer to serve as a member of the Committee. I am grateful to Dr. Mark Brown for offering to serve as a member of the Committee and would also like to thank him for his interest in my work.

I would like to thank Dr. Shiryaev for all our valuable conversations on the subject of change-point detection, for him taking the time to read and evaluate my work and for his suggestion to submit my first paper to the Theory of Probability and its Applications. I would also like to thank Dr. de la Peña for never losing his faith in me and always offering his moral support. I would also like to thank Dr. Lo and Dr. Ying for the opportunity to teach and for delivering a good course in statistical inference during my first year. Moreover, I would like to thank Dr. Karatzas whose constant dedication to the field has been a constant source of inspiration for me.

I am grateful to my colleague Gerardo Hernandez for his valuable suggestions. I would also like to thank my friends Jeffrey Said and Daqing Zhang for the constant source of support. Furthermore, I would like to thank Dood Kalicharan

our departmental administrator for her patience and support.

Moreover, I would like to thank my parents Dr. Nick Hadjiliadis and Marietta Lamera as well as my brother Dr. Denis Hadjiliadis for their constant support. Finally, I would like to say that there are no words to express my gratitude to my husband Dr. Ioannis Stamos that helped me through this difficult ride.

To my husband, Dr. Ioannis Stamos

Chapter 1

Introduction

This thesis is a collection of two related works. The first work falls under the broader topic of statistical surveillance. The second work is a study of the gambler's ruin problem as examined from a different point of view on the investor's behavior.

The need for statistical surveillance has been noted in many different areas. Applications include:

- Statistical quality control

  Historically the need for quality control was noted in Shewhart (see [33]). The topic continued being of great interest in the 50's and 60's when we see the more systematic construction of online detection schemes (see [25, 30, 34]). Statistical quality control consists of the generation of alarms for the attention of the operator after which the technological process has to be stopped, checked and repaired, if necessary. The purpose is the on-line detection of changes in the parameters of the model used to describe the raw input or the production material (see for example [47, 3]).

- Epidemiology (see for example [46, 39, 29, 19, 37]).

  An example is when the number of cases of a disease is recorded sequentially with the aim of detecting an increased incidence. In this case, a decision concerning whether the incident has increased or not must be made on the basis of the data collected so far.

- Medicine (see for example [11, 13]).

  An example is the monitoring of adverse drug reactions (ADR) after its approval for distribution in the market. The post-marketing surveillance system functions as a crucial medium for providing additional safety information that cannot realistically be obtained before the approval of each drug.

- Biomedical signal processing (see for example [3, 7])

  The need for online detection algorithms in this area, has been motivated by the automatic processing of biomedical signals such as Electroencephalogram (EEC) and Electrocardiogram (ECG).

- Finance (see for example [1])

  An example includes detecting change points in business cycles such as a peak or a trough in the economy through the means of prospective analysis base on data of a leading economical indicator.

- Fault detection in navigational systems (see for example [3])

  The purpose is to extract the useful signal (geodesic coordinates, velocity) and isolate the faulty sensors as soon as possible.

- Seismology (see [3] for details)

Traditionally, Shewhart's (see [33]) $\bar{x}$ charts with various modifications dating back to 1931 have been very popular for this problem. The above chart is based on the fact that it is expected that the average of the observations will dramatically change as soon as the process goes out of control. This procedure, however, was found to be somewhat inefficient in detecting small shifts. To overcome this shortcoming several stopping rules have been introduced during the past few decades. One of the most popular techniques has been the CUSUM stopping rule first introduced by Page in 1954 (see [25]). The CUSUM rule is defined to be the difference of the log of the Radon Nikodym derivative of the change from its running minimum. It is characterized by two parameters, namely the drift parameter of the change and the threshold parameter. Its properties have been thoroughly studied by Van Dobben de Bruyn in 1968 (see [44]). The problem of determining its distribution has also been studied by Zacks in 1981 (see [52]). These results were later generalized by Woodall in 1983 (see [48, 49]) and Yaschin in 1985 (see [51]) where the distribution of the two-sided CUSUM (2-CUSUM) was studied, all in the discrete time setting. In the continuous time setting one can find the exact computation of the Laplace transform of the one-sided CUSUM in Taylor (1975, see [43]) and Lehoczky (1977, see [18]). Lorden in 1971 (see [20]) proved that the one-sided CUSUM stopping rule is first-order asymptotically optimal for a specific min-max performance measure with the assumption that τ is an unknown parameter. Roberts in 1959 (see [30]) proposed the EWMA rule. Later, Shiryaev in 1963 (see [34]) and Roberts in 1966 (see [31]) independently proposed what became known as the Shiryaev-Roberts rule. This rule has been employed when τ is assumed to be a random variable with a given prior distribution.

For the Bayesian setting, in which the change point τ is assumed to have a prior distribution that is assumed to be exponential (in the continuous time model) and geometric (in the discrete time model) and the magnitude of the change is known a priori, it is shown in [34] that the Shiryaev-Roberts procedure is the optimal stopping rule in minimizing the Bayes risk. This measure penalizes the probability of false alarms and the detection delay of the change point by a given constant c per time unit. In other words, the objective is to find a stopping rule that minimizes the sum of the probability that $\{T < \tau\}$ and c times the expected value of $T - \tau$. The solution consists of computing the posterior density of the change point τ and stopping the first time it exceeds a given threshold that depends on the constant c. The proof is subsequently simplified by Beibel in 1996 (see [6]), a paper which also demonstrates the relationship between the posterior density of the change point τ and the CUSUM stopping rule. The later work of Karatzas in 2002 (see [14]) should also be mentioned, where the performance measure is replaced by an expected miss criterion which is equal to the expected absolute deviation of the stopping rule T from the change point τ. In this setting, it is shown that the optimal stopping rule is the first time that the conditional odds-ratio exceeds a threshold h that depends on the exponential parameter of the change point τ. Later, Karatzas in 2003 addressed the adaptive problem of also estimating the magnitude of the change with a more general performance measure. In this paper (see [15]), it is shown that a closed form solution involving a two-dimensional sufficient statistic for the change point τ exists only when the distribution of the random variable representing the magnitude of the change is a symmetric Bernoulli. A first adaptive problem of estimation of the mean parameter of a normal distribution which is

subjected to changes in time and where observations are taken in discrete time, appears in Chernoff & Zacks (1964, see [8]).

A comparison of the CUSUM and the Shiryaev-Roberts rules for detecting one-sided alternatives in the Brownian motion was conducted by Pollak & Siegmund in 1985 (see [27]). This comparison was based on the conditional average delay time in detecting the change point, given no false alarm was made, and is also based on the extreme assumptions that the change point is either equal to 0 or ∞. The results were that the Shiryaev-Roberts rule is as powerful as the CUSUM rule. Later, Srivastava & Wu in 1993 (see [40]) compared the Shiryaev-Roberts and CUSUM rules with the EWMA rule. The comparison was based on a performance measure called SADT (Stationary Average Delay Time) first advocated by Shiryaev in [34]. SADT is the limiting value of ADT, that is defined as the expected value of the time until the first actual detection of a change point. The SADT is also shown to be equal to a weighted average of the CADT (Conditional Average Delay Time) also first advocated by Shiryaev in [34]. Shiryaev in [34] suggested that SADT is preferable as a performance measure when the change point rarely occurs or when the cost of false alarms is relatively small compared to the loss due to delay in detection. The asymptotic properties of the three were examined as the in-control ARL(or the mean time between false alarms) tends to infinity in the above mentioned work of Srivastava & Wu. The results were that the EWMA is less efficient than the other two. The mean time between false alarms is nothing but the expected value of T when the observed process does not change. This is referred to in the literature as the "in-control Average Run Length (ARL)" (see [9, 50, 35]) , since the process is considered to be in-control when the change has not

occurred.

In the work that appears in Chapters 2 and 3, the change point τ is assumed to be an unknown parameter and multiple alternatives exist after the change. Today it is known that the CUSUM rule (see [35], [6]), with Lorden's criterion (see [20]), in the single alternative case where the change is a known constant, is optimal. This criterion considers the worst detection delay over all possible paths and all possible change-points as a performance measure. This result was extended by Tartakovsky in 1995 (see [41]), where the drift assumed after the change is a deterministic function of time. The result was further extended by Moustakides in 2004 (see [24]), where the drift assumed after the change is a measurable function of the observations, and the performance measure used is an alternative to Lorden's criterion, namely the Kullback-Leibler divergence. In discrete time it is known that for a single alternative and with an independence assumption before and after the change, the CUSUM is also optimal (see [23]) even when an exponential penalty for delay is used as a performance measure(see [28]).

The first one to suggest the cumulative sum tests for two-sided alternatives was Barnard in 1959 (see [2]). Later the problem of multiple alternatives in the discrete time exponential family model was examined by Lorden in 1971 (see [20]). He proposed that the generalized CUSUM be used in the case where the magnitude of the assumed drift after the change is unknown, and proved that for two-sided alternatives, as the in-control ARL tends to infinity, the procedure is first-order asymptotically optimal. Dragalin in 1994 (see [10]) improved on this result and showed that the generalized CUSUM stopping rule, for a specific choice of threshold, as the in-control ARL tends to infinity, is second-order asymptotically optimal. The

problem of multiple alternatives was subsequently addressed by Tartakovsky in 1994 (see [42]). His objective was to find a rule that would not only detect the change but that would also specifically point it out. With $\sup_\tau E_\tau^i [T - \tau \mid T > \tau] \quad \forall\, i$ as a performance measure, he found that the N-CUSUM stopping rule is first-order asymptotically optimal as the in-control ARL tends to infinity. The N-CUSUM consists of running N one-sided CUSUM schemes in parallel, each designed to detect the respective changes. The 2-CUSUM stopping rule was proposed as an alternative to the generalized CUSUM by Dragalin in 1997 (see [9]). Although he only considered one-sided alternatives in the discrete time exponential family model (in which the magnitude of the drift assumed after the change is unknown) he used a min-max type of criterion for the performance measure subject to the usual constraint on the in-control ARL.

The first two chapters are concerned with the problem of change-point detection in the Brownian motion model with multiple alternatives. In other words, both chapters employ the Brownian motion model in which observations are taken sequentially. The objective is to detect a change in the constant drift by means of a stopping rule when there are multiple but known possibilities for such a change. As a performance measure an extended Lorden criterion is proposed. In other words, the worst detection delay over all paths, over all change-points and over all possible changes, is considered. The goal is to minimize the worst case detection delay, subject to a constraint in the frequency of false alarms. First in Chapter 2, it is shown that, when the drifts have the same sign, the CUSUM rule designed to detect the smallest in absolute value drift, is the optimal stopping rule. If the drifts have opposite signs of known magnitude the rule traditionally suggested in the litera-

ture (see for example [36, 42, 2, 9]) is the 2-CUSUM rule that consists of running two one-sided CUSUM schemes in parallel, each designed to detect the respective changes. In this case, a specific 2-CUSUM rule is shown to be asymptotically optimal as the frequency of false alarms tends to infinity. In particular it is shown that, when the drifts are equal in absolute value, the difference in performance between the unknown optimal rule and the proposed scheme remains uniformly bounded although both quantities tend to infinity. For unequal in absolute value drifts the asymptotic optimality is even stronger since the corresponding difference tends to zero. Note that this is a clear improvement of what exists in the literature (see for example [42]) where it is only shown that the ratio of the above performances tends to a constant. The work that appears in Chapter 3 is a closer examination of the proposed scheme in the case of two-sided alternatives. More specifically, attention is drawn to a class of 2-CUSUM stopping rules that exhibit a property which allows for the exact computation of their expectations. These 2-CUSUM rules are called the harmonic mean 2-CUSUM rules. The proposed scheme is drawn from a special class of 2-CUSUM stopping rules amongst this category, called drift equalizer rules. Drift equalizer 2-CUSUM rules exhibit the exact detection delay under both the positive and the negative change and are shown to have strictly better performance than non-equalizer 2-CUSUM harmonic mean rules for the proposed performance measure in Chapter 3. In other words, by allowing an extra degree of freedom on the choice of the drift parameters of the 2-CUSUM we can get a strictly better performance than for any of the 2-CUSUM rules that have been proposed in the literature.

In the last Chapter, the gambler's ruin problem is revisited. The gambler's

ruin problem is one of the well known problems in probability theory. In the traditional setup a gambler quits once his or her wealth reaches some upper or lower level for the first time. The evolution of the gambler's wealth is assumed to be a biased random walk in the discrete time model, and a Brownian motion with non-positive drift in the continuous time model. In this setup, one can explicitly compute the probability of stopping the game at the upper level in contrast to stopping the game at the lower level. Reaching the upper bound can be viewed as winning in the betting game, while reaching the lower bound as losing in the game. Computing these probabilities is an easy consequence of the Optional Sampling Theorem and we review this result in Appendix B.

However, people often make decisions based on relative change in contrast to absolute change of their wealth. As a consequence, some gamblers (and investors in general) may have a tendency to stop after their wealth makes a significant positive or negative movement. In the last Chapter we consider this situation, i.e., the case when the gambler decides to stop either when his or her current wealth is above a certain level in comparison to the historical minimum of his or her wealth (upward rally), or when his or her current wealth is below a certain level in comparison to the historical maximum of his or her wealth (downward fall). The gambler would stop as soon as either the upward rally or the downward fall reach some pre-specified values. In other words, at each point in time, the gambler considers the following two quantities:

- The difference between his or her current wealth and the running minimum of it since he joined the game.

- The difference between the running maximum and the current value of his or

her wealth.

The gambler stops the first time either of these quantities exceeds a given threshold. Quitting on the downward fall can be perceived as losing in the game, while quitting on the upward rally that can be perceived as winning in the game. The problem is to determine the probability of quitting on the upward rally (or equivalently on the downward fall). There is obviously a clear connection between this problem and the 2-CUSUM stopping rule because of the properties of both of the quantities introduced above, based upon which gamblers make their decisions. More specifically, the 2-CUSUM is the minimum of its one-sided CUSUM branches each of which declares a stop when the difference between the value of the Radon-Nikodym derivative and its running minimum exceeds a given threshold. The gambler, equivalently, decides to quit the game by comparing the value of his or her wealth to its running minimum or running maximum and stopping the first time either of these quantities exceeds a given threshold.

In the setting described above we compute the probabilities of quitting the game on upward rally (or downward fall) are computed both in the discrete and in the continuous time framework. The probabilities are computed by means of the distribution function of the random variables $Y^+_{T_1(a)}$ and $Y^-_{T_2(b)}$, where $Y^+_{T_1(a)}$ represents the value of the upward rally when the downward fall reaches the level a for the first time, and $Y^-_{T_2(b)}$ represents the value of the downward fall when the upward rally reaches the level b for the first time. Moreover, the expected value of the minimum of the time it takes the downward fall of the wealth and the upward rally of the wealth to reach their respective thresholds is computed.

In the discrete time framework, it is shown that the distribution of each of

$Y^+_{T_1(a)}$ and $Y^-_{T_2(b)}$ is geometric with a parameter that is related to the gambler's ruin probability in the traditional setting, but with an additional mass at 0. The mass at 0 is computed in terms of the expected values of the time it takes the downward fall or the upward rally to reach their respective thresholds. This is achieved using the method described by Siegmund (see [36]) in the computation of the expected value of the one-sided CUSUM stopping time. As a side result one can also compute the expected value of the minimum of the time it takes the downward fall of the wealth and the upward rally of the wealth to reach their respective thresholds is also computed. This stopping rule is a version of the 2-CUSUM stopping rule and the explicit computation of its first moment in the case of different thresholds in its one-sided CUSUM branches is a clear improvement of the existing result that appears in [51]. Hence, it is worth noting that the usefulness of this result is also seen in two-sided alternative change-point detection.

In the continuous time framework, the computation of the probabilities is achieved using the distributional properties of $y^+_{T^c_1(a)}$ and $y^-_{T^c_2(b)}$ – the continuous time counterparts of the above mentioned random variables. Using results of Taylor [43] and Lehoczky [18] for the distribution of a stopped drifted Brownian motion at the first time of the downfall of level a, we are able to show that the probability density function of $y^+_{T^c_1(a)}$ and $y^-_{T^c_2(b)}$ is exponential, but with an additional mass at 0. The mass at 0 can be computed in a similar fashion as in the discrete case. As a side result, we also get the expected value of the minimum of the above times, which is related to the expectation of the 2-CUSUM stopping times with equal drift parameters but unequal thresholds in their respective one-sided CUSUM branches and is not known in the current literature in the continuous time framework.

It is worth mentioning that the probability densities of the random variables $y^{+}_{T_1^c(a)}$ and $y^{-}_{T_2^c(b)}$ are the first stepping stone to the computation of the joint density of the random variables maximal downward fall and maximal upward rally of a Brownian motion. This is an extension of the result that exists in the current literature where one can find the density of the maximal downward fall of a Brownian motion (see [21]). It is also worth mentioning that the Laplace transform of the maximum of a random walk appears in Kemperman 1961 (see [17])

Chapter 2

CUSUM rules for detecting a regime change in the Brownian motion model with multiple alternatives

In this work, we examine the simplest continuous model (a model appropriate when the observation process is assumed to be continuous). This is the Brownian motion model.

The setting of the change-point detection problem in the Brownian motion model is as follows: The stochastic process $\{\xi_t\}_{t\geq 0}$ is observed and is assumed to be a Brownian Motion which during the interval $[0, \tau]$ has zero drift. During the interval (τ, ∞) it assumes a known (simple alternative case) or unknown drift which is equivalent to several known drifts (multiple alternative case). We seek a

stopping rule T which 'detects' the change point τ as soon as possible while at the same time controls the mean time between false alarms. In other words, at each decision time point, t, we want to discriminate between two states of the process, the state, $\{T > s\}$ and the state $\{T \leq s\}$ as explained by Sonesson in [38]. Sonesson in [38] clearly indicates that this is achieved by an alarm statistic, i.e a process that is a function of the observations, and an alarm limit. As soon as the alarm limit is exceeded by the above statistic, an alarm is drawn. More specifically, the stopping rule T balances the trade-off between controlling the mean time between false alarms while minimizing the detection delay of the change.

As our problem involves multiple alternatives after the change, we extend Lorden's criterion in a min-max way as described in Section 2.1. Properties of the one-sided CUSUM are presented in Section 2.2 along with an optimality result applicable when all alternatives are of the same sign. In Section 2.3 the 2-CUSUM rule is introduced for detecting a two-sided alternative. A special class of 2-CUSUM stopping rules with the harmonic mean rule property is presented. The property permits the explicit computation of the first moment of the 2-CUSUM rule. Among that class, the smaller class of drift equalizer rules is subsequently presented . The asymptotically best equalizer rule is found in the case of unequal-in-absolute-value drifts. In Section 2.4, it is shown that in the latter case, the difference in the performance between the optimal rule and the asymptotically best equalizer rule tends to 0 as the in control ARL tends to infinity. In the symmetric case it is shown that the difference between the 2-CUSUM stopping rule with drift parameter equal to the absolute value of the change and the optimal stopping rule tends to a constant as the in-control ARL tends to infinity. In Section 2.5, we close with concluding

remarks and suggestions for future work. In Chapter 3 we revisit the class of 2-CUSUM drift equalizer harmonic mean stopping rules, whereby we demonstrate their optimality amongst the class of 2-CUSUM harmonic mean stopping rules.

2.1 Mathematical formulation of the problem.

We begin by considering the observation process $\{\xi_t\}_{t>0}$ with the following dynamics:

$$d\xi_t = \begin{cases} dw_t & t \leq \tau \\ \mu_i\, dt + dw_t & t > \tau, \quad i = 1, 2. \end{cases}$$

where τ, the time of change, is assumed to be an unknown constant; μ_i, the possible drifts the process can change to, are assumed known, but the specific drift the process is changing to is assumed to be determined by nature and is thus unknown. Our goal is to detect the change and not to infer which of the changes occurred.

The probability triple is

$$\big(C[0,\infty], \mathcal{F}, \{\mathcal{F}_t\}, \{P^i_\tau\}\big) \quad \forall\ i = 1, 2 \text{ and } \tau \ \in\ [0, \infty),$$

functions, $\mathcal{F}_t = \sigma\{\xi_t, t \geq 0\}$, $\mathcal{F} = \mathcal{F}_\infty = \cup_{t>0}\mathcal{F}_t$, and $\{P^i_\tau\}$ is the family of probability measures generated by the observation process $\{\xi_t\}$ when the change is $i = 1, 2$ and the change-point is τ. Notice that P_∞ is the Wiener measure.

The objective is to detect the change as soon as possible, which is achieved through the means of a stopping rule T adapted to the filtration $\mathcal{F}_t$. This means that at each instant t it is decided whether to stop or continue sampling based only on the information that is available up to that instant. If P_τ is the true distribution, then in the event that $\{T \geq \tau\}$ it is desired that the conditional expectation of

$T - \tau$ should be small. Notice that $\{T \geq \tau\} \in \mathcal{F}_\tau$. But of course, $\forall\ t > \tau$, $\{T = t\} \in \mathcal{F}_t \supset \mathcal{F}_\tau$. One of the possible performance measures of the detection delay, suggested by Lorden in [20], considers the worst detection delay over all paths before the change and all possible change points τ. It is

$$J(T) = \sup_\tau \operatorname{ess\,sup} E_\tau \left[(T-\tau)^+ | \mathcal{F}_\tau\right], \tag{2.1}$$

giving rise to the following constrained stochastic optimization problem:

$$\begin{aligned} &\inf_T J(T) \\ &E_\infty [T] \geq \gamma. \end{aligned} \tag{2.2}$$

One can immediately notice that the small detection delay requirement is offset by the requirement that the frequency of "false reactions" be controlled. In other words, the meaning of the requirement that $E_\infty [T] \geq \gamma$ is that, the mean time between alarms under the Wiener measure (i.e. the measure corresponding to there not being any change) is at least as big as γ. One can also write

$$E_\infty [T] = \int_0^\infty P_\infty (T > t) dt$$

and notice that the above requirement is equivalent to the requirement of a small $P_\infty [T < t]$, which is the probability of a false alarm (type I error).

In order to incorporate the different possibilities for the μ_i, we extend Lorden's performance measure inspired by the idea of the worst detection delay regardless of the change (along the lines of [9]). It is

$$J_L(T) = \max_i \sup_\tau \operatorname{ess\,sup} E_\tau^i \left[(T-\tau)^+ | \mathcal{F}_\tau\right], \tag{2.3}$$

which results in a corresponding optimization problem of the form:

$$\begin{aligned} &\inf_T J_L(T) \\ &E_\infty [T] \geq \gamma. \end{aligned} \tag{2.4}$$

It is easily seen that, in seeking solutions to the above problem, we can restrict our attention to stopping times that satisfy the false alarm constraint with equality. This is because, if $E_\infty[T] > \gamma$, we can produce a stopping time that achieves the constraint with equality without increasing the detection delay, simply by randomizing between T and the stopping time that is identically 0. This was first seen by Moustakides in the discrete case [see [23]]. To this effect, we introduce the following definition:

Definition 2.1 *Define $\mathcal{K}$ to be the set of all stopping rules T that are adapted to $\mathcal{F}_t$ and that satisfy $E_\infty[T] = \gamma$.*

2.2 The one-sided CUSUM stopping time

The CUSUM statistic process and the corresponding one-sided CUSUM stopping time are defined as follows:

Definition 2.2 *Let $\lambda \in \mathcal{R}$ and $\nu \in \mathcal{R}_+$. Define the following processes:*

1. $u_t(\lambda) = \lambda\xi_t - \frac{1}{2}\lambda^2 t$; $m_t(\lambda) = \inf_{0\le s\le t} u_s(\lambda)$.
2. $y_t(\lambda) = u_t(\lambda) - m_t(\lambda) \ge 0$, *which is the CUSUM statistic process.*
3. $T_c(\lambda,\nu) = \inf\{t \ge 0; y_t(\lambda) \ge \nu\}$, *which is the CUSUM stopping time.*

We are now in a position to examine two very important properties of the one-sided CUSUM stopping time. The first is a characteristic specifically inherent in the CUSUM statistic, and is summarized in the following lemma:

Lemma 2.1 *Fix $\tau \in [0, \infty)$. Let $t \geq \tau$, and consider the process*

$$y_{t,\tau} = u_t - u_\tau - \inf_{\tau \leq s \leq t} (u_t - u_\tau).$$

This is the CUSUM process when starting at time τ. We have that $y_t \geq y_{t,\tau}$ with equality if $y_\tau = 0$.

Proof: Note that

$$y_t = y_{t,\tau} + \left(\inf_{\tau \leq s \leq t} (u_s - u_\tau) + y_\tau \right)^+ \geq y_{t,\tau} \tag{2.5}$$

and that $\inf_{\tau \leq s \leq t}(u_s - u_\tau) \leq 0$. $\diamond$

By its definition it is clear that $y_{t,\tau}$ depends only on information received *after* time τ. Let us remind ourselves that the CUSUM stopping rule is a function of our CUSUM statistic process only. Thus, we conclude that all contribution of the observation process $\{\xi_t\}$ before time τ to our CUSUM stopping rule, is summarized in y_τ. Relation (2.5), therefore, suggests that the worst detection delay *before* τ occurs whenever $y_\tau = 0$. In other words,

$$\operatorname{ess\,sup} E_\tau \left[(T_c(\lambda, \nu) - \tau)^+ | \mathcal{F}_\tau \right] = E_\tau \left[(T_c(\lambda, \nu) - \tau)^+ | y_\tau = 0 \right] = E_0 \left[T_c(\lambda, \nu) \right]. \tag{2.6}$$

Equation (2.6) states that the CUSUM stopping time is an equalizer rule over τ, in the sense that its performance does not depend on the value of this parameter.

The second property of the one-sided CUSUM comes as a result of noticing that m_t is nonincreasing and that when it changes (decreases) we necessarily have $m_t = u_t$. In other words, when m_t changes, y_t attains its smallest value, that is 0. When this happens we will say that the CUSUM statistic process *restarts*. This important observation combined with standard results appearing in [16] allow for the computation of the CUSUM delay function.

Lemma 2.2 *Suppose a CUSUM stopping rule is based on the CUSUM statistic with drift parameter $\lambda \in \mathcal{R}$ and has threshold $\nu \in \mathcal{R}_+$. Then the detection delay when the observation process ξ_t has drift $\mu \in \mathcal{R}$ is given by $E\left[T_c(\lambda,\nu)\right] = (2/\lambda^2)g(\nu,\rho)$, where*

$$g(\nu,\rho) = \frac{e^{-\rho\nu} + \rho\nu - 1}{\rho^2} \quad \text{and} \quad \rho = 2\frac{\mu}{\lambda} - 1.$$

Proof: Consider the function $f(y) = \frac{2}{\lambda^2}[g(\nu,\rho) - g(y,\rho)]$. Then f is a twice continuously differentiable function of y satisfying

$$\rho f'(y) + f''(y) = -1, \text{ with } f'(0) = f(\nu) = 0.$$

Using standard Itô calculus on the process $f(y_t)$ and the results appearing in [16, Pages 149, 210] it is easy to show that for any stopping time T with $E[T] < \infty$, we have

$$E[f(y_T)] - f(y_0) = -E[T].$$

The desired formula follows by noticing that $y_0 = 0$ and for the CUSUM stopping time we have $y_{T_c} = \nu$ (for more details see also [24]). $\diamond$

Notice that for $\alpha \neq 0$ we have $\frac{1}{\alpha^2}g(\nu,\rho) = g(\frac{\nu}{|\alpha|}, \rho|\alpha|)$. This suggests the following alternative expression for the delay function

$$E[T_c(\lambda,\nu)] = 2g\left(\frac{\nu}{|\lambda|}, \text{sign}(\lambda)(2\mu - \lambda)\right). \tag{2.7}$$

In [6] and [35] it is shown that when there is only one possible alternative for the drift μ, the CUSUM stopping rule $T_c(\mu,\nu)$, with ν satisfying $\frac{2}{\mu^2}g(\nu,-1) = \gamma$, solves the optimization problem defined in (2.2).

When the sign of the alternative drifts is the same, with the help of the following lemma we can show that the one-sided CUSUM stopping rule that detects the smallest in absolute value drift is the optimal solution of the problem in (2.4).

Lemma 2.3 *For every path of the Brownian motion w_t, the process $y_t(\lambda)$ is an increasing (decreasing) function of the drift of the observation process ξ_t when $\lambda > 0$ ($\lambda < 0$).*

Proof: Consider two possible drift values μ_1, μ_2 with $\mu_1 < \mu_2$. We define two observation processes $\xi_t(\mu_i) = \mu_i(t-\tau)^+ + w_t,\ i = 1, 2$, that lead to the corresponding CUSUM processes

$$\begin{aligned} u_t(\lambda, \mu_i) &= \lambda \xi_t(\mu_i) - \frac{1}{2}\lambda^2 t = \lambda\{w_t + \mu_i(t-\tau)^+\} - \frac{1}{2}\lambda^2 t \\ m_t(\lambda, \mu_i) &= \inf_{0 \le s \le t} u_s(\lambda, \mu_i) \\ y_t(\lambda, \mu_i) &= u_t(\lambda, \mu_i) - m_t(\lambda, \mu_i). \end{aligned}$$

Consider the difference $y_t(\lambda, \mu_2) - y_t(\lambda, \mu_1) = \delta(t-\tau)^+ - m_t(\lambda, \mu_2) + m_t(\lambda, \mu_1)$ where $\delta = \lambda(\mu_2 - \mu_1)$. Notice now that $\lambda > 0$ implies $\delta > 0$ and we can write

$$u_s(\lambda, \mu_2) = u_s(\lambda, \mu_1) + \delta(s-\tau)^+ \le u_s(\lambda, \mu_1) + \delta(t-\tau)^+.$$

Taking the infimum over $0 \le s \le t$ we get $m_t(\lambda, \mu_2) \le m_t(\lambda, \mu_1) + \delta(t-\tau)^+$ from which, by rearranging terms, we get that $y_t(\lambda, \mu_2) \ge y_t(\lambda, \mu_1)$. The case $\lambda < 0$ can be shown similarly. $\diamond$

From Lemma 2.3 it also follows that $\mu_1 \le \mu_2$ implies $E^1[T_c(\lambda, \nu)] \ge E^2[T_c(\lambda, \nu)]$ when $\lambda > 0$ and the opposite when $\lambda < 0$. As a direct consequence of this fact comes our first optimality result concerning drifts with the same sign.

Theorem 2.1 *Let $0 < \mu_1 \le \mu_2$ or $\mu_2 \le \mu_1 < 0$, then the one-sided CUSUM stopping time $T_c(\mu_1, \nu_1)$ with ν_1 satisfying $\frac{2}{\mu_1^2} g(\nu_1, -1) = \gamma$ solves the optimization problem defined in (2.4).*

Proof: The proof is straightforward. Since ν_1 was selected so that $T_c(\mu_1, \nu_1)$ satisfies the false alarm constraint, we have $T_c(\mu_1, \nu_1) \in \mathcal{K}$. Then, $\forall \quad T \in \mathcal{K}$ we have

$$\begin{aligned} J_L(T) &= \max_i \sup_\tau \operatorname{ess\,sup} E_\tau^i \left[(T-\tau)^+ | \mathcal{F}_\tau\right] \\ &\geq \sup_\tau \operatorname{ess\,sup} E_\tau^1 \left[(T-\tau)^+ | \mathcal{F}_\tau\right] \\ &\geq E_0^1[T_c(\mu_1, \nu_1)] = \max_i E_0^i[T_c(\mu_1, \nu_1)] = J_L(T_c(\mu_1, \nu_1)) = \frac{2}{\mu_1^2} g(\nu_1, 1). \end{aligned}$$

The last inequality comes from the optimality of the one-sided CUSUM stopping rule and the last three equalities are due to Lemma 2.3, the definition of the performance measure $J_L(T)$ in (2.3) and Lemma 2.2. $\diamond$

It is worth pointing out that if we had n alternative drifts (instead of two) of the form $0 < \mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ or $0 > \mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$ and we used the extended Lorden criterion in (2.3), the optimality of $T_c(\mu_1, \nu_1)$, presented in Theorem 1, would still be valid. Our result should be compared to [9] (which refers to discrete time and the exponential family), where for the same type of changes only asymptotically optimum schemes are offered.

We also have the following corollary of Lemma 3:

Corollary 2.1 *Let* $0 < |\mu_1| \leq |\mu_2|$ *and define* $\eta_i, \ i = 1, 2,$ *so that* $\frac{2}{\mu_i^2} g(\eta_i, -1) = \gamma > 0$. *Then we have*

$$\frac{1}{\mu_1^2} g(\eta_1, 1) \geq \frac{1}{\mu_2^2} g(\eta_2, 1). \tag{2.8}$$

Proof: Since the result is independent of the sign of the two drifts, without loss of generality we may assume $0 < \mu_1 \leq \mu_2$. Consider the two CUSUM rules $T_c(\mu_i, \eta_i), \ i = 1, 2$. Because the two thresholds η_i were selected to satisfy the false

alarm constraint, using Lemma 1, Lemma 3 and the optimality of the one-sided CUSUM stopping time, the following inequalities hold $\forall\, T \in \mathcal{K}$:

$$\begin{aligned}
\frac{2}{\mu_1^2} g(\eta_1, 1) = E_0^1[T_c(\mu_1, \eta_1)] &\geq E_0^2[T_c(\mu_1, \eta_1)] \\
&= \sup_{\tau} \operatorname{ess\,sup} E_\tau^2 \left[(T_c(\mu_1, \eta_1) - \tau)^+ | \mathcal{F}_\tau \right] \\
&\geq \inf_{T} \sup_{\tau} \operatorname{ess\,sup} E_\tau^2 \left[(T - \tau)^+ | \mathcal{F}_\tau \right] \\
&= E_0^2[T_c(\mu_2, \eta_2)] = \frac{2}{\mu_2^2} g(\eta_2, 1). \qquad \diamond
\end{aligned}$$

2.3 Different drift signs and the 2-CUSUM stopping time

Let us now consider the case $\mu_2 < 0 < \mu_1$. The very interesting problem of knowing the amplitude of the drift but not the sign falls into this setting. What has traditionally been done in the literature, dating as far back as Barnard in [2], is to use the minimum of the stopping rules $T_c(\mu_1, \nu_1)$ and $T_c(\mu_2, \nu_2)$ each tuned to detect the respective changes μ_1 and μ_2. To this effect, we introduce the following 2-CUSUM stopping rule:

Definition 2.3 *Let $\lambda_2 < 0 < \lambda_1$. The 2-CUSUM stopping time $T_{2c}(\lambda_1, \lambda_2, \nu_1, \nu_2)$ is defined by: $T_{2c}(\lambda_1, \lambda_2, \nu_1, \nu_2) = T_c(\lambda_1, \nu_1) \wedge T_c(\lambda_2, \nu_2)$.*

We will, from now on, denote all 2-CUSUM rules by T_{2c} unless it is necessary to give emphasis to their four parameters. By the definition of the 2-CUSUM stopping rule it is apparent that it consists of running the two CUSUM statistic processes $y_t(\lambda_1)$ and $y_t(\lambda_2)$ in parallel, and stopping whenever one of the two hits its corresponding

threshold for the first time. From Lemma 2.1 we can conclude that

$$\operatorname{ess\,sup} E_\tau^i \left[(T_{2c}-\tau)^+|\mathcal{F}_\tau\right] = E_\tau^i \left[(T_{2c}-\tau)^+|y_\tau(\lambda_1)=y_\tau(\lambda_2)=0\right] = E_0^i\left[T_{2c}\right], \quad (2.9)$$

from which we get

$$J_L(T_{2c}) = \max_i \sup_\tau \operatorname{ess\,sup} E_\tau^i \left[(T_{2c}-\tau)^+|\mathcal{F}_\tau\right] = \max_i E_0^i\left[T_{2c}\right].$$

As we have seen the 2-CUSUM stopping rule is characterized by the four parameters, $\lambda_1, \lambda_2, \nu_1$ and ν_2. Since our intention is to propose a specific rule as the "preferable" one, we need to come up with a specific selection of these parameters. For this purpose, up to this point, we only have one equation available, namely, the false alarm constraint $E_\infty[T_{2c}] = \gamma$. Hence, we will gradually impose additional constraints on our 2-CUSUM structure in order to arrive to a unique stopping rule. Once our rule is specified we will support its selection by demonstrating that it enjoys a strong asymptotic optimality property.

2.3.1 The harmonic mean 2-CUSUM rules

First we draw our attention to a specific class of 2-CUSUM stopping rules that allow for the exact computation of their performance.

Definition 2.4 *Define*

$$\mathcal{G} = \{T_{2c}(\lambda_1,\lambda_2,\nu_1,\nu_2); \nu_1 = |\lambda_1|\nu \text{ and } \nu_2 = |\lambda_2|\nu\}.$$

For $T_{2c} \in \mathcal{G}$ we have the following characteristic property:

Lemma 2.4 *Let $T_{2c} \in \mathcal{G}$ then, when T_{2c} stops, one of its CUSUM statistic processes hits its corresponding threshold while the other necessarily restarts.*

Proof: Although the proof given in [36, Page 28], for discrete time and the exponential family, applies here as well (without major changes), we will give an alternative proof. Consider the process:

$$Y_t = \frac{y_t(\lambda_1)}{|\lambda_1|} + \frac{y_t(\lambda_2)}{|\lambda_2|} \quad = \quad -\frac{1}{2}(|\lambda_1| + |\lambda_2|)t - \frac{m_t(\lambda_1)}{|\lambda_1|} - \frac{m_t(\lambda_2)}{|\lambda_2|}.$$

Since $y_t(\lambda_i) \geq 0$ we clearly have $Y_t \geq 0$. We can now distinguish the three cases:

1. if neither of $\frac{y_t(\lambda_1)}{|\lambda_1|}$ or $\frac{y_t(\lambda_1)}{|\lambda_1|}$ are equal to 0, then Y_t decreases in time (since in this case $\frac{m_t(\lambda_1)}{|\lambda_1|}$ and $\frac{m_t(\lambda_2)}{|\lambda_2|}$ remain constant in time).

2. if both of $\frac{y_t(\lambda_1)}{|\lambda_1|}$ and $\frac{y_t(\lambda_1)}{|\lambda_1|}$ are equal to 0, then Y_t equals 0.

3. if one of $\frac{y_t(\lambda_1)}{|\lambda_1|}$ or $\frac{y_t(\lambda_1)}{|\lambda_1|}$ restarts (becomes equal to 0), then Y_t increases in time.

Therefore Y_t increases only when either of the processes $\frac{y_t(\lambda_1)}{|\lambda_1|}$ or $\frac{y_t(\lambda_2)}{|\lambda_2|}$ restarts. In other words, when $\frac{y_t(\lambda_1)}{|\lambda_1|}$ reaches its threshold ν for the first time $\frac{y_t(\lambda_2)}{|\lambda_2|}$ restarts and the other way around. $\diamond$

The following lemma uses the above property to derive a formula for the expected delay of the 2-CUSUM rule.

Lemma 2.5 *Let $T_{2c} = T_1 \wedge T_2$ with $T_{2c} \in \mathcal{G}$ and T_1, T_2 the corresponding one-sided CUSUM branches. Then the expected delay of the 2-CUSUM stopping time T_{2c} is related to the corresponding delays of its one-sided CUSUM branches through the formula*

$$(E[T_{2c}])^{-1} = (E[T_1])^{-1} + (E[T_2])^{-1}. \tag{2.10}$$

Proof: By using Itô calculus on the function $g(x,\rho)$ as it appears in Lemma 2.2 we get:

$$\begin{aligned} E\left[g(y^+_{T_{2c}},\rho_1)\right] &= \frac{\lambda_1^2}{2}E\left[T_{2c}\right], && (2.11)\\ \text{where } \rho_1 &= 2\frac{\mu}{\lambda_1}-1 \text{ and} \\ E\left[g(y^-_{T_{2c}},\rho_2)\right] &= \frac{\lambda_2^2}{2}E\left[T_{2c}\right], && (2.12)\\ \text{where } \rho_2 &= 2\frac{\mu}{\lambda_2}-1. \end{aligned}$$

As a consequence of Lemma 2.4 and the fact that $T_{2c}\in\mathcal{G}$, it follows that

$$P(y^-_{T_{2c}}=0)+P(y^+_{T_{2c}}=0) = 1, \qquad (2.13)$$

and therefore the RHS of equations (2.11) and (2.12) become:

$$\begin{aligned} g(\nu,\rho_1)\cdot P(y^-_{T_{2c}}=0) &= \frac{\lambda_1^2}{2}E\left[T_{2c}\right], && (2.14)\\ g(\nu,\rho_2)\cdot P(y^+_{T_{2c}}=0) &= \frac{\lambda_2^2}{2}E\left[T_{2c}\right]. && (2.15) \end{aligned}$$

The result now follows from equations (2.13), (2.14), (2.15), and Lemma 2.2. $\diamond$

2.3.2 2-CUSUM drift equalizer rules

It is well known that min-max problems, such as (2.4), are solved by *equalizer rules*. In other words, by stopping rules that demonstrate the same performance under the two changes. This is shown, in particular, in Chapter 3. Thus, we further restrict ourselves among the class of equalizer rules.

Definition 2.5 *Define*

$$\mathcal{D}=\{T_{2c}\in\mathcal{G};E_0^1\left[T_{2c}\right]=E_0^2\left[T_{2c}\right]\}.$$

By the definition of the class of equalizer rules it follows that $\mathcal{D} \subset \mathcal{G}$. Let us now find a simple condition that guarantees this property.

By using equations. (2.7), (2.10) we get

$$E_0^i[T_{2c}] = \left(\frac{1}{2g(\nu, \mathrm{sgn}(\lambda_1)(2\mu_i - \lambda_1))} + \frac{1}{2g(\nu, \mathrm{sgn}(\lambda_2)(2\mu_i - \lambda_2))} \right)^{-1}, \quad i = 1, 2. \tag{2.16}$$

From (2.16) we can see that in order to have $T_{2c} \in \mathcal{D}$ we need

$$\mathrm{sgn}(\lambda_1)(2\mu_1 - \lambda_1) = \mathrm{sgn}(\lambda_2)(2\mu_2 - \lambda_2) \tag{2.17}$$

$$\mathrm{sgn}(\lambda_2)(2\mu_1 - \lambda_2) = \mathrm{sgn}(\lambda_1)(2\mu_2 - \lambda_1). \tag{2.18}$$

One can now easily verify that both of the above equations (2.17) and (2.18) are satisfied whenever

$$\lambda_1 + \lambda_2 = 2(\mu_1 + \mu_2). \tag{2.19}$$

In other words, if we select λ_1, λ_2 to satisfy (2.19) then the corresponding 2-CUSUM stopping rule has the same performance under both drifts μ_1, μ_2. Note that a more elegant proof of this appears in Chapter 3.

By limiting ourselves to the class $\mathcal{D}$ (i.e. selecting $\nu_1 = |\lambda_1|\nu$, $\nu_2 = |\lambda_2|\nu$ and using (2.19)), apart from the false alarm constraint, we impose two additional constraints on our four parameters. In order for the 2-CUSUM rule to be completely specified we need one final condition. Our intention is to select the parameter λ_1 so that the corresponding detection delay is asymptotically minimized (as $\gamma \to \infty$).

Theorem 2.2 *Let $\mu_2 < 0 < \mu_1$ with $|\mu_1| \leq |\mu_2|$. Consider all 2-CUSUM stopping times $T_{2c} \in \mathcal{K} \cap \mathcal{D}$. Then among all such stopping rules the one with $\lambda_1 = \mu_1$, $\lambda_2 = 2\mu_2 + \mu_1$ is asymptotically optimal as $\gamma \to \infty$.*

Proof: Since $\mu_1 + \mu_2 \leq 0$, for any $\lambda_1 > 0$, from equation (2.19), we get $|\lambda_1| \leq |\lambda_2|$. Let us first consider the false alarm constraint. Using equations. (2.7), (2.10) with $\mu = 0$ and $\nu_1 = |\lambda_1|\nu$, $\nu_2 = |\lambda_2|\nu$, we get

$$E_\infty[T_{2c}] = \left(\frac{1}{2g(\nu, -|\lambda_1|)} + \frac{1}{2g(\nu, -|\lambda_2|)}\right)^{-1} = \gamma. \tag{2.20}$$

By carefully examining the exponential rates of the two terms in (2.20) we conclude that the leading term is the one containing λ_1. Hence, we get

$$\lambda_1 \nu = \log \gamma (1 + o(1)). \tag{2.21}$$

For the common detection delay, using equation (2.16) and substituting $\lambda_2 = 2(\mu_1 + \mu_2) - \lambda_1$ we have the estimates:

$$\begin{aligned} E_0^i[T_{2c}] &= \left(\frac{1}{2g(\nu, 2\mu_1 - \lambda_1)} + \frac{1}{2g(\nu, 2\mu_2 - \lambda_1)}\right)^{-1} \\ &= \begin{cases} \frac{2\nu}{2\mu_1 - \lambda_1}(1 + o(1)) & \text{for } 2\mu_1 > \lambda_1 \geq 0 \\ \nu^2(1 + o(1)) & \text{for } 2\mu_1 = \lambda_1 \\ \frac{2e^{\nu|2\mu_1 - \lambda_1|}}{(2\mu_1 - \lambda_1)^2}(1 + o(1)) & \text{for } 2\mu_1 < \lambda_1. \end{cases} \end{aligned} \tag{2.22}$$

The objective is to minimize the detection delay with respect to λ_1 in order to find the best selection for this parameter. From (2.22) it is clear that it is sufficient to limit ourselves to the case $0 \leq \lambda_1 < 2\mu_1$, since for $\lambda_1 \geq 2\mu_1$ the detection delay increases significantly faster as ν increases. For $0 \leq \lambda_1 < 2\mu_1$, the detection delay, after substituting ν from (2.21), can be written as

$$\frac{2 \log \gamma}{\lambda_1(2\mu_1 - \lambda_1)}(1 + o(1)),$$

which is clearly minimized, asymptotically, for $\lambda_1 = \mu_1$. Using equation (2.19), we also get $\lambda_2 = 2\mu_2 + \mu_1$. $\diamond$

Let us now summarize our results. We propose the following 2-CUSUM rule for the case $\mu_2 < 0 < \mu_1$: when $|\mu_1| \leq |\mu_2|$ select $\lambda_1 = \mu_1$, $\lambda_2 = 2\mu_2 + \mu_1$, $\nu_1 = |\mu_1|\nu$, $\nu_2 = |2\mu_2 + \mu_1|\nu$. If $|\mu_1| \geq |\mu_2|$ then $\lambda_1 = 2\mu_1 + \mu_2$, $\lambda_2 = \mu_2$, $\nu_1 = |2\mu_1 + \mu_2|\nu$, $\nu_2 = |\mu_2|\nu$. Finally, the parameter ν is selected so as to satisfy the false alarm constraint.

2.4 Asymptotic optimality in opposite sign drifts

For the specific 2-CUSUM rule introduced at the end of the previous Section, we are going to demonstrate two asymptotic optimality results. By means of an upper and a lower bound on the performance of the unknown optimal stopping rule, we will show that in the case of equal in absolute value drifts the *difference* in performance between the unknown optimum rule and the proposed 2-CUSUM rule tends to a constant as $\gamma \to \infty$. In the case of different in absolute value drifts we have a stronger asymptotic result. In particular, we will demonstrate that the *difference* in performance between the unknown optimal rule and the proposed 2-CUSUM rule tends to 0 as $\gamma \to \infty$. This should be compared to most existing asymptotic optimality results (see for example [42]) where it is shown that the *ratio* between the performance of the optimum and the proposed scheme tends to unity (first order optimality). Our form of asymptotic optimality is clearly stronger since it implies first order optimality, while the opposite is not necessarily true.

Let T_{2c} denote the specific 2-CUSUM rule proposed in the previous Section with the threshold ν selected so that the false alarm constraint is satisfied with equality. Since T_{2c} constitutes a possible choice in the class $\mathcal{K}$, equation (2.9) and

Lemma 2.2 imply that $\forall\, T \in \mathcal{K}$

$$E_0^1[T_{2c}] = E_0^2[T_{2c}] = J_L(T_{2c}) \geq \inf_T J_L(T). \tag{2.23}$$

To find a lower bound, we observe that $\forall\, T \in \mathcal{K}$ we can write

$$\begin{aligned} \inf_T J_L(T) &= \inf_T \max_i \sup_\tau \operatorname{ess\,sup} E_\tau^i[(T-\tau)^+|\mathcal{F}_\tau] \\ &\geq \max_i \left(\inf_T \sup_\tau \operatorname{ess\,sup} E_\tau^i[(T-\tau)^+|\mathcal{F}_\tau] \right) \\ &= \max_i \frac{2}{\mu_i^2} g(\eta_i, 1), \end{aligned} \tag{2.24}$$

where for the last equality we used the optimality of the one-sided CUSUM stopping rule and the expression for its worst detection delay from Lemma 2. The two thresholds η_i, $i = 1, 2$, are selected to satisfy the false alarm constraint $\frac{2}{\mu_i^2} g(\eta_i, -1) = \gamma$. The asymptotic results that follow examine the way the two bounds approach each other. Since the performance of the optimal stopping rule is between the two bounds, this will also determine the rate with which the 2-CUSUM approaches the optimal solution.

2.4.1 The case of equal in absolute value drifts

We first consider the special case $\mu_1 = -\mu_2 = \mu$. Here our parameter selection takes the form $\lambda_1 = \mu_1 = \mu$ and $\lambda_2 = 2\mu_2 + \mu_1 = \mu_2 = -\mu$ which coincides with the 2-CUSUM scheme proposed in the literature. Let us now examine the two bounds. The upper bound, from (2.16), with this specific parameter selection becomes

$$J_L(T_{2c}) = E_0^i[T_{2c}] = \left(\frac{1}{2g(\nu, \mu)} + \frac{1}{2g(\nu, -3\mu)} \right)^{-1}, \quad i = 1, 2, \tag{2.25}$$

with the threshold ν computed from the false alarm constraint (2.20) that takes the form

$$E_\infty[T_{2c}] = \left(\frac{1}{2g(\nu,-\mu)} + \frac{1}{2g(\nu,-\mu)}\right)^{-1} = g(\nu,-\mu) = \gamma. \tag{2.26}$$

Similarly, the lower bound becomes $\frac{2}{\mu^2}g(\eta,1)$ with the threshold η satisfying $\frac{2}{\mu^2}g(\eta,-1) = \gamma$.

Theorem 2.3 *The difference in the performance between the proposed 2-CUSUM stopping rule and the optimal stopping rule, is bounded above by a quantity that tends to the constant* $\frac{2\log 2}{\mu^2}$*, as the false alarm constraint* $\gamma \to \infty$*.*

Proof: Solving for ν from (2.26) we obtain $\mu\nu = \log\gamma + \log\frac{\mu^2}{2} + \log 2 + o(1)$. On the other hand, we can write (2.25) as $J_L(T_{2c}) = \frac{2}{\mu^2}\{\mu\nu + e^{-\mu\nu} - 1\}\{1 + O(\mu\nu e^{-3\mu\nu})\}$. Substituting the estimate for ν we get

$$J_L(T_{2c}) = \frac{2}{\mu^2}\left\{\log\gamma + \log\frac{\mu^2}{2} - 1 + \log 2 + o(1)\right\}.$$

Similarly, for the lower bound we have that the threshold η as a function of γ becomes $\eta = \log\gamma + \log\frac{\mu^2}{2} + o(1)$. Therefore, the lower bound is of the form $\frac{2}{\mu^2}\{\log\gamma + \log\frac{\mu^2}{2} - 1 + o(1)\}$. Since the difference between the upper and the lower bound, bounds the difference $J_L(T_{2c}) - \inf_T J_L(T)$, we conclude that

$$0 \le J_L(T_{2c}) - \inf_T J_L(T) \le \frac{2}{\mu^2}\{\log 2 + o(1)\},$$

from which the result follows by letting $\gamma \to \infty$. $\diamond$

Figure 2.1 depicts the upper and lower bound as a function of the false alarm constraint γ for the case $\mu_1 = -\mu_2 = 1$. Since, as we can see, the difference of the two bounds is increasing with γ, the constant proposed by Theorem 2.3 corresponds to a *worst case* performance attained only in the limit as $\gamma \to \infty$.

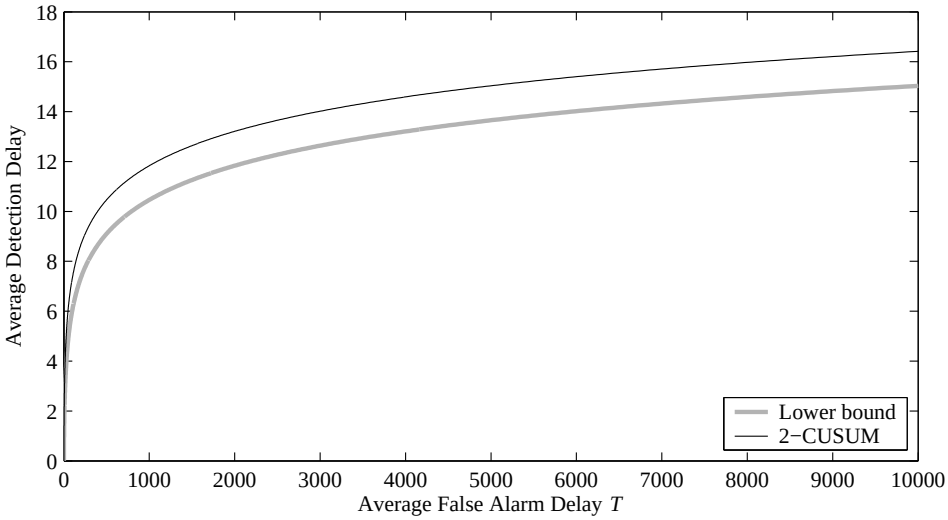

Figure 2.1: Typical form of the upper and lower bounds of the performance of the optimum stopping rule for the case $\mu_1 = -\mu_2 = 1$.

2.4.2 The case of different in absolute value drifts

Theorem 2.4 *The difference in the performance between the proposed 2-CUSUM stopping rule and the optimal stopping rule is bounded above by a quantity that tends to* 0*, as the false alarm constraint* $\gamma \to \infty$.

Proof: We will only examine the case $|\mu_1| < |\mu_2|$. From Corollary 1 and equation (2.8) it follows that the maximum in the lower bound in (2.24) is achieved for μ_1. Hence, as in Theorem 2.3, we get $\frac{2}{\mu_1^2}\{\log\gamma + \log\frac{\mu_1^2}{2} - 1 + o(1)\}$ for the lower bound.

The upper bound is the detection delay of the proposed 2-CUSUM stopping time T_{2c}. From (2.16), with $\lambda_1 = \mu_1$, $\lambda_2 = 2\mu_2 + \mu_1$, we have

$$\begin{aligned} J_L(T_{2c}) = E_0^i[T_{2c}] &= \left(\frac{1}{2g(\nu,\mu_1)} + \frac{1}{2g(\nu, 2\mu_2 - \mu_1)}\right)^{-1} \\ &= \frac{2}{\mu_1^2}\{e^{-\mu_1\nu} + \mu_1\nu - 1\}\{1 + O(\mu_1\nu e^{(2\mu_2-\mu_1)\nu})\}, \quad (2.27) \end{aligned}$$

where ν is selected to satisfy the false alarm constraint, which from (2.20) takes the form

$$E_\infty[T_{2c}] = \left(\frac{1}{2g(\nu, -\mu_1)} + \frac{1}{2g(\nu, 2\mu_2 + \mu_1)} \right)^{-1} = \gamma. \tag{2.28}$$

From (2.28) we get the estimate $\mu_1 \nu = \log \gamma + \log \frac{\mu_1^2}{2} + o(1)$. This, when substituted in (2.27), produces:

$$J_L(T_{2c}) = E_0^i[T_{2c}] = \frac{2}{\mu_1^2} \left\{ \log \gamma + \log \frac{\mu_1^2}{2} - 1 + o(1) \right\}. \tag{2.29}$$

Subtracting now the lower bound expression from the upper bound expression in (2.29) we obtain

$$0 \leq J_L(T_{2c}) - \inf_T J_L(T) \leq o(1),$$

which tends to 0 as $\gamma \to \infty$. $\diamond$

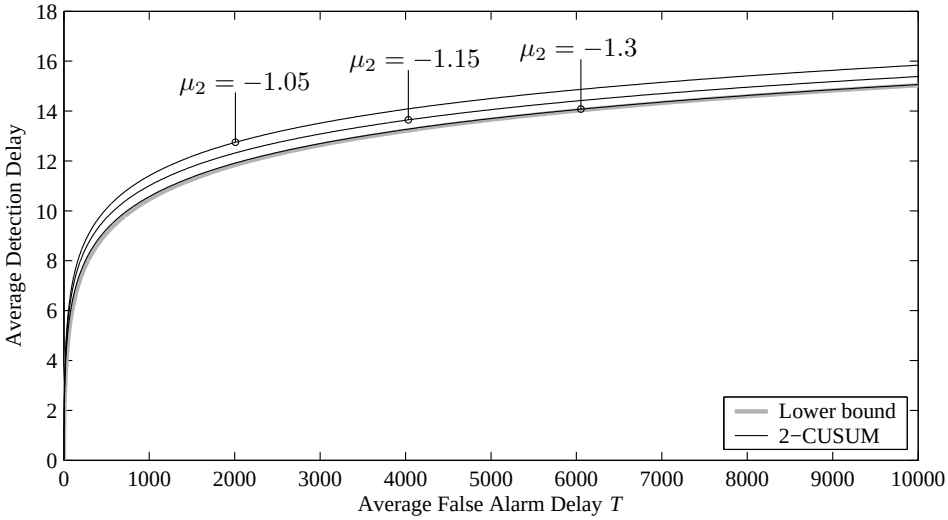

Figure 2.2: Typical form of the upper and lower bounds of the performance of the optimal stopping rule for the case $\mu_2 < 0 < \mu_1$, with $\mu_1 = 1$ and $\mu_2 = -1.05, -1.15, -1.3$.

In Figure 2.2 we present the two bounds for $\mu_1 = 1$ and $\mu_2 = -1.05, -1.15, -1.3$. We recall that the upper bound is the detection delay of the 2-CUSUM rule

$T_{2c} \in \mathcal{G} \cap \mathcal{K}$ with parameters $\lambda_1 = \mu_1$ and $\lambda_2 = 2\mu_2 + \mu_1$. We can see that the difference between the two curves is tending to zero as the false alarm tends to infinity, thus corroborating Theorem 2.4. What is more interesting, however, is the fact that the two curves rapidly approach each other, *uniformly* over γ, as the ratio $|\mu_2|/|\mu_1|$ of the two drifts increases. As we can see, in the case $\mu_1 = 1,\ \mu_2 = -1.3$ the two bounds become almost indistinguishable. This suggests that the proposed 2-CUSUM rule can be (extremely) close to the unknown optimal rule, not only asymptotically, as proposed by Theorem 2.4, but also uniformly over all false alarm values.

It is also worth noting that the difference in the performance of the optimal rule and any 2-CUSUM rule in $\mathcal{G}$ with parameters $\lambda_1 = \mu_1$ and $\lambda_2 \in (-\mu_1\ 2\mu_2 + \mu_1]$ (one such possibility is the selection proposed in the literature $\lambda_1 = \mu_1,\ \lambda_2 = \mu_2$) also tends to 0 as $\gamma \to \infty$. Therefore, asymptotically optimal solutions allow for many different choices. It is, however, our selection that leads to an equalizer rule.

2.5 Conclusions & Future Work

In this Chapter we identify the harmonic mean drift equalizer rule with the best asymptotic performance in the case of unequal in absolute value drifts (as the frequency of false alarms tends to infinity) and are able to prove a stronger asymptotic optimality result than is known in the literature. In particular, in this case, the difference in the detection delay of the optimal unknown scheme and the 2-CUSUM stopping rule (with the choice of parameters $\lambda_1 = \mu_1$ and $\lambda_2 = 2\mu_2 + 2\mu_1 - \lambda_1$) tends to 0, even though both of the detection delay quantities are unbounded as the frequency of false alarms tends to infinity. In fact, this difference tends faster

to 0 as the difference between μ_2 and μ_1 increases. It is interesting to notice however, that this strong asymptotic optimality result holds even when the choice of drift parameters is the one that coincides with the 2-CUSUM rules traditionally proposed in the literature. Nevertheless, the drift equalizer 2-CUSUM rule choice is preferable for all values of the frequency of false alarms as is demonstrated in the next Chapter. Moreover, in the symmetric case the difference in the detection delay of the optimum scheme and the specific 2-CUSUM rule with drift parameter equal to the absolute value of the change tends to the constant $\frac{2\log 2}{\mu^2}$ as the frequency of false alarms tends to infinity, where μ is the absolute value of the two-sided possible changes. Notice that in this case too, both detection delays become unbounded as the frequency of false alarms tends to infinity.

Yet, the choice of stopping rules within the class $\mathcal{G}$ was made only due to the fact that we can readily compute their expected values. Remark 4.2 in the last Chapter provides the first stepping stone to the explicit computation of the expected value of 2-CUSUM stopping rules that are not members of $\mathcal{G}$. It would be of great interest to contrast 2-CUSUM rules that are members of $\mathcal{G}$ with those that are not in order to identify optimal behavior.

Another problem of great interest is the problem of identifying the best 2-CUSUM rule among the family of 2-CUSUM rules generated by the set of all pairs of possible drift parameter values $\mathcal{M} = \{\mu_1, \mu_2, \ldots, \mu_n\}$ that contains both positive and negative values. This knowledge would enable us to select the best 2-CUSUM rule in the case that we know that the change is two-sided but only know a possible range of values that the two drifts of different signs can take. Obviously in the case that all μ_i's have the same sign the one-sided CUSUM with the smallest in absolute

value drift will be the optimal stopping rule as shown by Moustakides in Section 2.2.

Moreover, it is worth pointing out that we can only hope to find strict optimality in the two-sided alternative case if we restrict ourselves to a class of stopping rules that satisfy a symmetry condition. This should be true due to the fact that in two-sided-alternative hypotheses testing there is no uniformly most powerful test, but there does exist a uniformly most powerful unbiased test that satisfies a symmetry condition.

Finally, an area of interest would also be to try to generalize the results that appear in this Chapter to general Lévy processes. For the simplest case of jump processes, that is in the case of the Poisson disorder problem, results of interest appear in [45, 12, 32, 26, 4, 5, 22].

Chapter 3

Optimality of the 2-CUSUM drift equalizer rules among the harmonic mean 2-CUSUM rule class for detecting two-sided alternatives in the Brownian motion model

In the previous Chapter we confined our attention to 2-CUSUM stopping rules in the class $\mathcal{G}$ that satisfy the harmonic mean rule which enables us to compute their first moment exactly. We further restricted our attention to drift equalizer rules. This Chapter shows strictly optimal performance of drift equalizer harmonic

mean rules. We begin by redefining the 2-CUSUM stopping time by using only positive drift parameters and proceed to prove optimality of the 2-CUSUM drift equalizer rules amongst the class of harmonic mean 2-CUSUM rules. Drift equalizer 2-CUSUM rules constitute a clear improvement over what has been proposed and traditionally used in the literature for detecting two-sided alternatives. This is because a strictly better performance is achieved by means of a careful selection of their drift parameters.

This Chapter is structured as follows. In Section 3.1 we redefine the 2-CUSUM stopping rules with only positive parameters and revisit the harmonic mean rule. In Section 3.2 optimality of drift equalizer 2-CUSUM harmonic mean rules is proven. The proof is made up of results that appear in Appendix A.

3.1 The 2-CUSUM rules & the harmonic mean rule (revisited)

We sequentially observe a process $\{\xi_t\}$ with the dynamics:

$$d\xi_t = \begin{cases} dw_t & t \leq \tau \\ \\ \mu_1 dt + dw_t & \\ \quad \text{or} & t \geq \tau \\ -\mu_2 dt + dw_t & \end{cases}$$

where τ, the time of change, is assumed deterministic but unknown; μ_i, the possible drifts the process can change to, are assumed known, but the specific drift the process is changing to is assumed to be unknown. Both μ_1, μ_2 are assumed to be

positive. Without loss of generality we can assume that $\mu_2 \geq \mu_1$. Our goal is to detect the change and not to infer which of the two changes occurred.

The probabilistic setting of this problem is identical to the one that appears in Chapter 2 and our objective is identical to the one in Chapter 2, namely to solve the stochastic optinizaton problem that appears in 2.4.

Let us redefine the CUSUM statistics with only positive parameters.

Definition 3.1 *The normalized CUSUM statistics with drift parameters $\lambda_1 > 0$ and $\lambda_2 > 0$, tuned to detect the positive and negative changes in the drift of the Brownian motion are defined respectively as follows:*

1. $\frac{y_t^+(\lambda_1)}{\lambda_1} = \xi_t - \frac{1}{2}\lambda_1 t - \inf_{s \leq t}(\xi_s - \frac{1}{2}\lambda_1 s)$,
2. $\frac{y_t^-(\lambda_2)}{\lambda_2} = -\xi_t - \frac{1}{2}\lambda_2 t - \inf_{s \leq t}(-\xi_s - \frac{1}{2}\lambda_2 s)$.

We now proceed to define the 2-CUSUM stopping rules.

Definition 3.2 *The 2-CUSUM stopping rule with drift parameters $\lambda_1 > 0$, $\lambda_2 > 0$ and threshold parameters $\nu_1 > 0$, $\nu_2 > 0$ is defined as follows:*

$$T(\lambda_1, \lambda_2, \nu_1, \nu_2) = T^1 \wedge T^2,$$

where

1. $T^1 = \inf\{t > 0; \frac{y_t^+(\lambda_1)}{\lambda_1} > \nu_1\}$,
2. $T^2 = \inf\{t > 0; \frac{y_t^-(\lambda_2)}{\lambda_2} > \nu_2\}$.

We proceed to consider the smaller class of 2-CUSUM rules that satisfy the property of the harmonic mean rule. In particular, we will consider all 2-CUSUM rules whose both two CUSUM stopping-time branches T^1 and T^2 have the same threshold. The

harmonic mean rule enables us to explicitly compute the expected value of the 2-CUSUM stopping rule in terms of the expected values of its corresponding one sided CUSUM stopping times, as seen in Lemma 2.5. To this effect, we introduce the harmonic mean 2-CUSUM class of stopping rules with positive parameters only:

Definition 3.3 *With* $\lambda_1,\ \lambda_2,\ \nu_1,\ \nu_2 \in \mathcal{R}_+$, *define* $\mathcal{G} = \{T(\lambda_1, \lambda_2, \nu_1, \nu_2); \nu_1 = \nu_2\}$.

From now on we only consider 2-CUSUM rules in $\mathcal{G}$ and denote them by $T(\lambda_1, \lambda_2, \nu)$. In the previous Chapter, we showed that under any of the measures P_0^1, P_0^2, P_∞ we have:

$$\frac{1}{E\left[T(\lambda_1, \lambda_2, \nu)\right]} = \frac{1}{E\left[T^1\right]} + \frac{1}{E\left[T^2\right]}. \tag{3.1}$$

At this point, it is worth noting that for any T, CUSUM stopping rule the worst detection delay over all paths is the one that occurs when y_τ^+ and y_τ^- are 0. This is essentially a consequence of the non-negativity of the CUSUM statistic processes and can more formally be seen as a result of Lemma 2.1. It appears in equation (2.9) and we recall it here.

$$\begin{aligned} J_L(T) &= \max_i \sup_\tau \ \text{essup}\ E_\tau^i \left[(T-\tau)^+ | \mathcal{F}_\tau\right] \\ &= \max\{E_0^1\left[T\right], E_0^2\left[T\right]\}. \end{aligned} \tag{3.2}$$

As shown in Lemma 2.2, by applying Itô's rule and using existing results in

stochastic analysis we get:

$$\frac{1}{2}E_\infty(T^1) = \frac{h(\lambda_1\nu)}{\lambda_1^2}, \tag{3.3}$$

$$\frac{1}{2}E_\infty(T^2) = \frac{h(\lambda_2\nu)}{\lambda_2^2}, \tag{3.4}$$

$$\frac{1}{2}E_0^1(T^1) = \frac{h\Big((\lambda_1 - 2\mu_1)\nu\Big)}{(\lambda_1 - 2\mu_1)^2}, \tag{3.5}$$

$$\frac{1}{2}E_0^1(T^2) = \frac{h\Big((\lambda_2 + 2\mu_1)\nu\Big)}{(\lambda_2 + 2\mu_1)^2}, \tag{3.6}$$

$$\frac{1}{2}E_0^2(T^1) = \frac{h\Big((\lambda_1 + 2\mu_2)\nu\Big)}{(\lambda_1 + 2\mu_2)^2}, \tag{3.7}$$

$$\frac{1}{2}E_0^2(T^2) = \frac{h\Big((\lambda_2 - 2\mu_2)\nu\Big)}{(\lambda_2 - 2\mu_2)^2}, \tag{3.8}$$

where $h(x) = e^x - x - 1$.

3.2 Equalizer rules are best

We now proceed to inspect the dynamics of the CUSUM statistic processes when the change is μ_1 and when the change is $-\mu_2$:

	change is μ_1
$\frac{y_t^+}{\lambda_1}$	$w_t + (\mu_1 - \frac{1}{2}\lambda_1)t - \inf_{s\le t}(w_s + (\mu_1 - \frac{1}{2}\lambda_1)s)$
$\frac{y_t^-}{\lambda_2}$	$-w_t - (\mu_1 + \frac{1}{2}\lambda_2)t - \inf_{s\le t}(-w_s - (\mu_1 + \frac{1}{2}\lambda_2)s)$

Table 3.1: The dynamics of the two CUSUMs when the change is μ_1

We notice that if equation

$$\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1 \tag{3.9}$$

	change is μ_2
$\frac{y_t^+}{\lambda_1}$	$w_t - (\mu_2 + \frac{1}{2}\lambda_1)t - \inf_{s\le t}(w_s - (\mu_2 + \frac{1}{2}\lambda_1)s)$
$\frac{y_t^-}{\lambda_2}$	$-w_t + (\mu_2 - \frac{1}{2}\lambda_2)t - \inf_{s\le t}(-w_s + (\mu_2 - \frac{1}{2}\lambda_2)s)$

Table 3.2: The dynamics of the two CUSUMs when the change is μ_2

holds, then $\frac{y_t^+}{\lambda_1}$ when the change is μ_1 has the same law as $\frac{y_t^-}{\lambda_2}$ when the change is μ_2 and that $\frac{y_t^+}{\lambda_1}$ when the change is μ_2 has the same law as $\frac{y_t^-}{\lambda_2}$ when the change is μ_1. In particular, this means that $T(\lambda_1, \lambda_2, \nu) = T^1 \wedge T^2$ has the same distribution under both measures P_0^1 and P_0^2. Therefore, when equation (3.9) holds, $E_0^1\big(T(\lambda_1, \lambda_2, \nu)\big) = E_0^2\big(T(\lambda_1, \lambda_2, \nu)\big)$. This allows us to distinguish among all 2-CUSUM harmonic mean rules, the equalizer rules whose performance is the same under both measures P_0^1 and P_0^2.

Definition 3.4 *We define the class of all equalizer rules as follows:*

$$\mathcal{E} \;=\; \{T(\lambda_1, \lambda_2, \nu); \lambda_2 - \lambda_1 = 2(\mu_2 - \mu_1), \nu > 0\}.$$

In the sequel we will use S for any stopping rule that belongs to the class $\mathcal{E}$ and by T any stopping rule that does not belong to $\mathcal{E}$. Notice that if $\mu_2 = \mu_1$ (the symmetric case) any choice of $\lambda \in \mathcal{R}_+$ will result in an equalizer rule for $\lambda_2 = \lambda_1 = \lambda$.

Our focus thus, is on the case that $\mu_2 > \mu_1$. The objective is that for any arbitrary rule T we would like to be able to find an equalizer rule S that achieves the same frequency of false alarms while lowering the detection delay. In other words, for any arbitrary rule T we want to always be able to find a rule $S \in \mathcal{E}$ that has better performance.

To this effect let us define two classes of non-equalizer rules:

Definition 3.5 *We define the following two classes of non-equalizer rules*

1. $\mathcal{D}_g = \{T(\lambda_1, \lambda_2, \nu); \lambda_2 - \lambda_1 > 2\mu_2 - 2\mu_1\}$,

2. $\mathcal{D}_s = \{T(\lambda_1, \lambda_2, \nu); \lambda_2 - \lambda_1 < 2\mu_2 - 2\mu_1\}$.

Notice that $\mathcal{E}^c = \mathcal{D}_g \cup \mathcal{D}_s$.

Theorem 3.1 $\forall$ *2-CUSUM rules* $T \in \mathcal{D}_g \cup \mathcal{D}_s$, $\exists\ S \in \mathcal{E}$ *such that*

$$E_\infty(T) = E_\infty(S), \tag{3.10}$$

while

$$\max\{E_0^1(T), E_0^2(T)\} > E_0^1(S) = E_0^2(S).$$

Proof: We can distinguish the following three cases

1. $T(\lambda_1', \lambda_2', \nu) \in \mathcal{D}_g$. $\exists\ \lambda_2$, $\lambda_1 > 0$ with $\lambda_2 < \lambda_2'$ and $\lambda_1 > \lambda_1'$ such that $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$ for which (3.10) holds.

2. $T(\lambda_1', \lambda_2', \nu) \in \mathcal{D}_s$ and $\lambda_2' > \lambda_1'$. (The justification for the additional assumption $\lambda_2' > \lambda_1'$ is given in Appendix A.3 and should be read after the sequel). $\exists\ \lambda_2$, $\lambda_1 > 0$ with $\lambda_2 > \lambda_2'$ and $\lambda_1' > \lambda_1$ such that $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$ and the same frequency of false alarms can be achieved by an equalizer rule for the same threshold ν. More specifically, $\exists\ S(\lambda_1, \lambda_2, \nu)$ for which (3.10) holds.

3. There does not exist $\lambda_1 > 0$ such that with $\lambda_2 > \lambda_2'$ and $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$ the same frequency of false alarms can be achieved by an equalizer rule with the same threshold.

To prove the result in the first case it suffices to show the following two inequalities:

$$\frac{1}{E_0^2(S^1)} - \frac{1}{E_0^2(T^1)} > \frac{1}{E_\infty(S^1)} - \frac{1}{E_\infty(T^1)}, \tag{3.11}$$

$$\frac{1}{E_\infty(T^2)} - \frac{1}{E_\infty(S^2)} > \frac{1}{E_0^2(T^2)} - \frac{1}{E_0^2(S^2)}. \tag{3.12}$$

This is because the RHS of inequality (3.11) is equal to the LHS of inequality (3.12) as can be seen by using equations (3.1) and (3.10). It then follows that the LHS of the former inequality is greater than the RHS of the latter and using equation (3.1) again we get that $E_0^2(T) > E_0^2(S)$. Using equations (3.7), (3.8), (3.3), (3.4) we can rewrite inequality (3.11) in the following way:

$$\frac{1}{\frac{h\left((\lambda_1+2\mu_2)\nu\right)}{(\lambda_1+2\mu_2)^2}} - \frac{1}{\frac{h\left((\lambda_1'+2\mu_2)\nu\right)}{(\lambda_1'+2\mu_2)^2}} > \frac{1}{\frac{h(\lambda_1\nu)}{\lambda_1^2}} - \frac{1}{\frac{h(\lambda_1\nu)}{\lambda_1'^2}}. \tag{3.13}$$

Multiplying both sides of the equation by ν^2 and using the convexity of the function $g(x) = \frac{x^2}{h(x)}$ (see Appendix A.1) the result follows. Similarly we can show (3.12).

In cases 2 and 3 the result follows from inequalities

$$\frac{1}{E_0^1(S^2)} - \frac{1}{E_0^1(T^2)} > \frac{1}{E_\infty(S^2)} - \frac{1}{E_\infty(T^2)}, \tag{3.14}$$

$$\frac{1}{E_\infty(T^1)} - \frac{1}{E_\infty(S^1)} > \frac{1}{E_0^1(T^1)} - \frac{1}{E_0^1(S^1)}. \tag{3.15}$$

Notice that in both cases 2 and 3 we have $E_0^1(T) > E_0^1(S)$. In case 2 the two inequalities are a direct consequence of the convexity of the function $g(x) = \frac{x^2}{h(x)}$ after using equations (3.5), (3.6), (3.3), and (3.4).

In case 3 the situation is slightly more involved, since in order to achieve the same frequency of false alarms for an equalizer rule, we need to lower the threshold to $\nu' < \nu$. In other words, we can still find an $S(\lambda_1, \lambda_2, \nu') \in \mathcal{E}$ for which (3.10)

holds by selecting $\lambda_1 = \lambda_1'$, $\lambda_2 > \lambda_2'$ such that $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$ and consequently a threshold $\nu' < \nu$. We can now rewrite inequalities (3.14) and (3.15) with the above choice of parameters using equations (3.5), (3.6), (3.3), and (3.4) as follows:

$$\frac{1}{\frac{h\Big((\lambda_2+2\mu_1)\nu'\Big)}{(\lambda_2+2\mu_1)^2}} - \frac{1}{\frac{h\Big((\lambda_2'+2\mu_1)\nu\Big)}{(\lambda_2'+2\mu_1)^2}} > \frac{1}{\frac{h(\lambda_2\nu')}{\lambda_2^2}} - \frac{1}{\frac{h(\lambda_2'\nu)}{\lambda_2'^2}}, \tag{3.16}$$

$$\frac{1}{\frac{h(\lambda_1\nu)}{\lambda_1^2}} - \frac{1}{\frac{h(\lambda_1\nu')}{\lambda_1^2}} > \frac{1}{\frac{h\Big((\lambda_1-2\mu_1)\nu\Big)}{(\lambda_1-2\mu_1)^2}} - \frac{1}{\frac{h\Big((\lambda_1-2\mu_1)\nu'\Big)}{(\lambda_1-2\mu_1)^2}}. \tag{3.17}$$

For a proof of inequality (3.17) please refer to Appendix A.4. Notice that the parameters are chosen in such a way that (3.10) holds and therefore the RHS of (3.16) is equal to the LHS of (3.17) and they are both negative. Thus, we have $\frac{h(\lambda_2'\nu)}{\lambda_2'^2} < \frac{h(\lambda_2\nu')}{\lambda_2^2}$. If $\frac{h\Big((\lambda_2'+2\mu_1)\nu\Big)}{(\lambda_2'+2\mu_1)^2} \geq \frac{h\Big((\lambda_2+2\mu_1)\nu'\Big)}{(\lambda_2+2\mu_1)^2}$, then (3.16) trivially holds. We will now proceed to examine the contrary case. We make two selections $\lambda_2'' \in [\lambda_2', \lambda_2]$ and $\lambda_2^x \in [\lambda_2', \lambda_2]$ such that

$$\frac{h(\lambda_2\nu')}{\lambda_2^2} = \frac{h(\lambda_2''\nu)}{\lambda_2''^2}, \tag{3.18}$$

$$\frac{h\Big((\lambda_2+2\mu_1)\nu'\Big)}{(\lambda_2+2\mu_1)^2} = \frac{h\Big((\lambda_2^x+2\mu_1)\nu\Big)}{(\lambda_2^x+2\mu_1)^2}. \tag{3.19}$$

But from Appendix A.2, it follows that $\lambda_2^x < \lambda_2''$ and since the function $\frac{h(x)}{x^2}$ is strictly increasing $\forall\ x \in \mathcal{R}_+$ we have:

$$\frac{1}{\frac{h\Big((\lambda_2^x+2\mu_1)\nu\Big)}{(\lambda_2^x+2\mu_1)^2}} - \frac{1}{\frac{h\Big((\lambda_2'+2\mu_1)\nu\Big)}{(\lambda_2'+2\mu_1)^2}} > \frac{1}{\frac{h\Big((\lambda_2''+2\mu_1)\nu\Big)}{(\lambda_2''+2\mu_1)^2}} - \frac{1}{\frac{h\Big((\lambda_2'+2\mu_1)\nu\Big)}{(\lambda_2'+2\mu_1)^2}}. \tag{3.20}$$

Inequality (3.16) now readily follows from (3.20), (3.18), (3.19), the convexity of the function $g(x) = \frac{x^2}{h(x)}$ (see Appendix A.1) and the fact that $\lambda_2'' > \lambda_2'$. This completes the proof. $\diamond$

3.3 Conclusions

After restricting ourselves to the class of harmonic mean 2-CUSUM rules that impose a first constraint on the thresholds of each one-sided branch we proceed to examine the best selection of drift parameters in the case that the two opposite sign drifts are known. The result presented in this Chapter allows us to select the drifts λ_1, λ_2 in such a way that we can construct harmonic mean 2-CUSUM rules with a strictly better performance for all frequencies of false alarms especially in the case that the absolute values of the possible opposite sign drifts assumed after the change are not equal. This result clearly enhances the results in the literature whereby the suggested 2-CUSUM rules that have been used were selected amongst the class of harmonic mean 2-CUSUM rules with drift parameters exactly equal to the drifts assumed after the change.

Appendix A

A.1

Lemma A.1 *The function*

$$g(x) = \frac{1}{\frac{h(x)}{x^2}}$$

where $h(x) = e^x - x - 1$*, is strictly convex.*

Proof: It suffices to show that $g''(x) > 0$. We have

$$g''(x) = \frac{[2(e^x - x - 1) - x^2 e^x]\,(e^x - x - 1) - 2(e^x - 1)\,[2x(e^x - x - 1)]}{(e^x - x - 1)^3}$$

In order to show that $g''(x) > 0$ it suffices to show that the function $f(x) = [(e^x - x - 1)^3]\,g''(x)$ is positive $\forall\ x \neq 0$. Notice that $f(0) = 0$. It suffices to show that $f'(x)$ has the same sign as x. But

$$f'(x) = xe^x \left[2xe^x - 6e^x + x^2 + 4x + 6\right]$$

Let $k(x) = 2xe^x - 6e^x + x^2 + 4x + 6$. We have $k(0) = 0$ and $k''(x) = 2e^x\,[e^{-x} + x - 1] > 0\ \forall\ x \neq 0$. Therefore $k(x) > 0\ \forall\ x \neq 0$ and $k(0) = 0$. Hence $f'(x)$ has the same sign as x and $f(0) = 0$ from which it follows that $f(x) \geq 0$ with equality only at $x = 0$. This completes the proof. $\diamond$

A.2

Lemma A.2 *Suppose we pick x_1, x_2, ν' and ν all positive, with $\nu > \nu'$ and $x_1 < x_2$ so that*

$$\frac{h(x_1\nu)}{x_1^2} = \frac{h(x_2\nu')}{x_2^2}. \tag{A.1}$$

Then $\forall\ a \in \mathcal{R}_+$, we have:

$$\frac{h\Big((x_1+a)\nu\Big)}{(x_1+a)^2} > \frac{h\Big((x_2+a)\nu'\Big)}{(x_2+a)^2}, \tag{A.2}$$

where $h(x) = e^x - x - 1$.

Proof: Notice that, since $x_1 < x_2$, for equality (A.1) to hold we need $x_1\nu < x_2\nu'$. Therefore, $\frac{1}{x_1\nu} + \frac{1}{(x_1\nu)^2} > \frac{1}{x_2\nu'} + \frac{1}{(x_2\nu')^2}$. Using this and the fact that $\nu > \nu'$, we get that:

$$\left[\frac{x_1\nu+1}{(x_1\nu)^2}\right]\nu^2 - \left[\frac{x_2\nu'+1}{(x_2\nu')^2}\right]\nu'^2 > 0. \tag{A.3}$$

From equation (A.1) and inequality (A.3), it follows that:

$$\frac{e^{x_1\nu}}{x_1^2} > \frac{e^{x_2\nu'}}{x_2^2} \Leftrightarrow \frac{x_2^2}{x_1^2} > \frac{e^{x_2\nu'}}{e^{x_1\nu}}. \tag{A.4}$$

We now have:

$$\begin{aligned}\frac{e^{x_2\nu'}}{e^{x_1\nu}} &> \frac{e^{x_2\nu'} - x_1\nu e^{-a\nu}}{e^{x_1\nu} - x_1\nu e^{-a\nu}} > \frac{e^{x_2\nu'} - x_2\nu' e^{-a\nu'}}{e^{x_1\nu} - x_1\nu e^{-a\nu}} \\ &> \frac{e^{x_2\nu'} - x_2\nu' e^{-a\nu'} - a\nu' e^{-a\nu'} - e^{-a\nu'}}{e^{x_1\nu} - x_1\nu e^{-a\nu} - a\nu e^{-a\nu} - e^{-a\nu}},\end{aligned} \tag{A.5}$$

where the first inequality follows from the fact that $x_1\nu e^{-a\nu} > 0$, the second inequality from the fact that $x_1\nu < x_2\nu'$ and $e^{-a\nu'} > e^{-a\nu}$, and the last inequality by

noting that the function $(x+1)e^{-x}$ is decreasing $\forall\ x > 0$. Using inequalities (A.4), (A.5) as well as the fact that $a > 0$ and the fact that $\frac{e^{a\nu'}}{e^{a\nu}} < 1$, we get:

$$\begin{aligned} \frac{(x_2+a)^2}{(x_1+a)^2} &> \frac{x_2^2}{x_1^2} > \frac{e^{x_2\nu'} - (x_2+a)\nu' e^{-a\nu'} - e^{-a\nu'}}{e^{x_1\nu} - (x_1+a)\nu e^{-a\nu} - e^{-a\nu}} \\ &> \frac{e^{a\nu'}}{e^{a\nu}} \frac{e^{x_2\nu'} - (x_2+a)\nu' e^{-a\nu'} - e^{-a\nu'}}{e^{x_1\nu} - (x_1+a)\nu e^{-a\nu} - e^{-a\nu}}. \end{aligned} \tag{A.6}$$

The result follows from the final inequality and a rearrangement of terms. $\diamond$

A.3

What remains for the proof of Theorem 3.1 is to justify that, whenever $E_0^1[T] > E_0^1[S]$ (cases 2 and 3), it is sufficient to consider 2-CUSUM rules T for which the second drift parameter is greater than the first. To this effect, let us define the following two classes of stopping rules:

Definition A.1 *Define*

1. $\mathcal{C}_1 = \{T(\lambda_1, \lambda_2, \nu); \lambda_2 > \lambda_1 > 0, \lambda_2 - \lambda_1 < 2\mu_2 - 2\mu_1, \nu > 0\}$.

2. $\mathcal{C}_2 = \{T(\lambda_1, \lambda_2, \nu); 0 < \lambda_2 < \lambda_1, \lambda_2 - \lambda_1 < 2\mu_2 - 2\mu_1, \nu > 0\}$.

Notice that $\mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{D}_s$. The following lemma is sufficient to justify our adherence to rules that belong to the class $\mathcal{C}_1$, whenever $E_0^1[T] > E_0^1[S]$.

Lemma A.3 $\forall\ T_{c_2} \in \mathcal{C}_2\ \ \exists\, T_{c_1} \in \mathcal{C}_1$ *such that* $E_\infty[T_{c_1}] = E_\infty[T_{c_2}]$ *while* $E_0^1[T_{c_2}] > E_0^1[T_{c_1}]$.

Proof: Let $\lambda_1' > \lambda_2' > 0$. Then $T(\lambda_1', \lambda_2', \nu) \in \mathcal{C}_2$. From equations (3.3), (3.4) and (3.1), we get that

$$\frac{1}{E_\infty[T(\lambda_1', \lambda_2', \nu)]} = \frac{1}{\frac{h\left((\lambda_1')\nu\right)}{(\lambda_1')^2}} + \frac{1}{\frac{h\left((\lambda_2')\nu\right)}{(\lambda_2')^2}}. \tag{A.7}$$

Now let $\lambda_1'' = \lambda_2'$ and $\lambda_2'' = \lambda_1'$. Then $T(\lambda_1'', \lambda_2'', \nu) \in \mathcal{C}_1$, while

$$\begin{aligned} \frac{1}{E_\infty\left[T(\lambda_1'', \lambda_2'', \nu)\right]} &= \frac{1}{\frac{h\left((\lambda_1'')\nu\right)}{(\lambda_1'')^2}} + \frac{1}{\frac{h\left((\lambda_2'')\nu\right)}{(\lambda_2'')^2}} \\ &= \frac{1}{\frac{h\left((\lambda_1')\nu\right)}{(\lambda_1')^2}} + \frac{1}{\frac{h\left((\lambda_2')\nu\right)}{(\lambda_2')^2}}. \end{aligned} \tag{A.8}$$

Therefore both of the above rules have the same frequency of false alarms.

The desirable result comes as a direct consequence of the following two inequalities:

$$\frac{1}{\frac{h\left((\lambda_1'-2\mu_1)\nu\right)}{(\lambda_1'-2\mu_1)^2}} - \frac{1}{\frac{h\left((\lambda_1''-2\mu_1)\nu\right)}{(\lambda_1''-2\mu_1)^2}} < \frac{1}{\frac{h\left((\lambda_1')\nu\right)}{(\lambda_1')^2}} - \frac{1}{\frac{h\left((\lambda_1'')\nu\right)}{(\lambda_1'')^2}} \tag{A.9}$$

and

$$\frac{1}{\frac{h\left((\lambda_2'')\nu\right)}{(\lambda_2'')^2}} - \frac{1}{\frac{h\left((\lambda_2')\nu\right)}{(\lambda_2')^2}} < \frac{1}{\frac{h\left((\lambda_2''+2\mu_1)\nu\right)}{(\lambda_2''+2\mu_1)^2}} - \frac{1}{\frac{h\left((\lambda_2'+2\mu_1)\nu\right)}{(\lambda_2'+2\mu_1)^2}}. \tag{A.10}$$

Notice that from (A.8) and (A.7), it follows that the RHS of (A.9) is equal to the LHS of (A.10). Therefore, *the LHS of (A.9) is greater than the RHS of (A.10).* From the result just mentioned, a rearrangement of terms and equations (3.6), (3.5), (3.3), (3.4) as well as (3.1), we get $\frac{1}{E_0^1\left(T(\lambda_1'',\lambda_2'',\nu)\right)} > \frac{1}{E_0^1\left(T(\lambda_1',\lambda_2',\nu)\right)}$, which is the desired inequality.

Inequalities (A.9) and (A.10) follow by multiplying both sides by ν^2 and using the convexity of the function $\frac{x^2}{h(x)}$ (see Appendix A.1) along with the fact that $\lambda_1' > \lambda_1''$ and $\lambda_2'' > \lambda_2'$ respectively. This completes the proof of the lemma. $\diamond$

A.4

Definition A.2 *Let $\mu \in [0, \infty)$ and $\nu > \nu' > 0$. Define the following functions:*

1. $C_\nu(\mu) = \frac{h\Big((\lambda_1-2\mu)\nu\Big)}{(\lambda_1-2\mu)^2}$,

2. $C_{\nu'}(\mu) = \frac{h\Big((\lambda_1-2\mu)\nu'\Big)}{(\lambda_1-2\mu)^2}$,

3. $f(\mu) = \frac{1}{C_\nu(\mu)} - \frac{1}{C_{\nu'}(\mu)}$,

where $h(x) = e^x - x - 1$.

After introducing the above definition we can rewrite (3.17) in the following way:

$$f(0) > f(\mu_1). \tag{A.11}$$

To prove inequality (A.11) it suffices to show that $f(\mu)$ is strictly decreasing.

Lemma A.4 *The function $f(\mu)$ is strictly decreasing.*

Proof: By differentiating $f(\mu)$ w.r.t μ we get

$$f'(\mu) \quad = \quad \frac{-C'_\nu(\mu)}{[C_\nu(\mu)]^2} + \frac{C'_{\nu'}(\mu)}{[C_{\nu'}(\mu)]^2}. \tag{A.12}$$

Hence it suffices to show that

$$-C'_\nu(\mu)\,[C_{\nu'}(\mu)]^2 \quad < \quad -C'_{\nu'}(\mu)\,[C_\nu(\mu)]^2\,. \tag{A.13}$$

Using $\nu > \nu'$ and doing a term-by-term comparison the result follows. $\diamond$

Chapter 4

The gambler's ruin problem with relative wealth perception

This Chapter is structured as follows. In Section 4.1.1 the explicit probabilities of exiting on the upward rally (or the downward fall) are computed along with the probability mass functions of the random variables $Y^{+}_{T_1(a)}$ and $Y^{-}_{T_2(b)}$ in the discrete time framework. Moreover, the expected value of the minimum of the two stopping times described above is given in a remark. All computations are also given in the special case of an unbiased random walk. In Section 4.1.2 the explicit probabilities of winning (or losing) as well as the probability density functions of the random variables $y^{+}_{T^c_1(a)}$ and $y^{-}_{T^c_2(b)}$ are computed in the continuous time framework. Furthermore, the expected value of the minimum of the two stopping times described above is given in a remark. All computations are also given in the special case of a standard Brownian motion model. Concluding remarks and suggestions for future work appear in Section 4.2. A short review of the traditional gambler's ruin problem in both the discrete and in the continuous time framework

appears in Appendix B.

4.1 Gambler's ruin problem with relative wealth perception

4.1.1 The discrete time framework

Assume that the evolution of the gambler's wealth S_n follows biased random walk, i.e., at time n

$$S_n = \sum_{i=1}^{n} Z_i,$$

where

$$Z_i = \begin{cases} 1, & \text{with probability } \ p, \\ -1, & \text{with probability } \ q, \end{cases}$$

with $p + q = 1$ and $p < q$. The quantity

$$S_n - \min_{k \in [0,n] \cap \mathcal{N}} S_k$$

measures the size of the upward rally comparing the present value of the wealth to its historical minimum, while the quantity

$$\max_{k \in [0,n] \cap \mathcal{N}} S_k - S_n$$

measures the size of the downward fall comparing the present value of the wealth to its historical maximum.

The aim of this Section is to determine the probability that the gambler would quit the game on the upward rally in contrast to quitting the game on the

downward fall. To this effect, we introduce the stopping times:

$$T_1(a) = \inf\{n \in \mathcal{N} : \max_{k\in[0,n]\cap\mathcal{N}} S_k - S_n = a, \quad a \in \mathcal{N}\},$$

and

$$T_2(b) = \inf\{n \in \mathcal{N} : S_n - \min_{k\in[0,n]\cap\mathcal{N}} S_k = b, \quad b \in \mathcal{N}\}.$$

The gambler stops at $T(a,b) = T_1(a) \wedge T_2(b)$. The stopping times $T_1(a)$ and $T_2(b)$ indicate the first time of reaching the critical level of the downward fall $T_1(a)$, or the first time of reaching the critical level of the upward rally $T_2(b)$. In this Section, we compute probabilities of the events $\{T(a,b) = T_1(a)\}$, which represents stopping the game on the downward fall, and $\{T(a,b) = T_2(b)\}$, which represents stopping the game on the upward rally.

In order to simplify notation we introduce the following processes:

$$\begin{aligned}
M_n^+ &:= \min_{k\in[0,n]\cap\mathcal{N}} S_k, \\
M_n^- &:= \min_{k\in[0,n]\cap\mathcal{N}} (-S_k) = -\sup_{k\in[0,n]\cap\mathcal{N}} S_k, \\
Y_n^+ &:= S_n - M_n^+, \\
Y_n^- &:= -S_n - M_n^-.
\end{aligned}$$

Therefore we can re-express $T_1(a)$ and $T_2(b)$ as:

$$\begin{aligned}
T_1(a) &= \min\{n \in \mathcal{N} : Y_n^- = a, \quad a \in \mathcal{N}\}, \\
T_2(b) &= \min\{n \in \mathcal{N} : Y_n^+ = b, \quad b \in \mathcal{N}\}.
\end{aligned}$$

Theorem 4.1 *Let $S_n = \sum_{i=1}^n Z_i$ be the evolution of the wealth of the gambler and let $T(a,b)$, $T_1(a)$ and $T_2(b)$ be stopping times defined as above. We distinguish the following three cases:*

1. $b \geq a+1 > 1$

   The probabilities of stopping the game on downward fall or upward rally are given by

$$P\Big(T(a,b) = T_1(a)\Big) = m_A + (1-m_A)\cdot(1-R_A^{b-a}), \tag{4.1}$$

$$P\Big(T(a,b) = T_2(b)\Big) = (1-m_A)\cdot R_A^{b-a}, \tag{4.2}$$

   respectively, where

$$m_A = \frac{\left(\frac{q}{p}\right)^{a+1} - (a+1)\left(\frac{q}{p}\right) + a}{\left[1-\left(\frac{q}{p}\right)^{-a}\right]\cdot\left[\left(\frac{q}{p}\right)^{a+1} - 1\right]}, \tag{4.3}$$

   and

$$R_A = \frac{1-\left(\frac{q}{p}\right)^{a}}{1-\left(\frac{q}{p}\right)^{a+1}}. \tag{4.4}$$

2. $a \geq b+1 > 1$

   The probabilities of stopping the game on downward fall or upward rally are given by

$$P\Big(T(a,b) = T_1(a)\Big) = (1-m_B)\cdot R_B^{a-b}, \tag{4.5}$$

$$P\Big(T(a,b) = T_2(b)\Big) = m_B + (1-m_B)\cdot(1-R_B^{a-b}), \tag{4.6}$$

   respectively, where

$$m_B = \frac{\left(\frac{q}{p}\right)^{-(b+1)} - (b+1)\left(\frac{q}{p}\right)^{-1} + b}{\left[1-\left(\frac{q}{p}\right)^{-b}\right]\cdot\left[\left(\frac{q}{p}\right)^{b+1} - 1\right]}, \tag{4.7}$$

   and

$$R_B = \left(\frac{q}{p}\right)\cdot\frac{1-\left(\frac{q}{p}\right)^{b}}{1-\left(\frac{q}{p}\right)^{b+1}}. \tag{4.8}$$

3. $a = b$

The probabilities of stopping the game on downward fall or upward rally are given by

$$P\Big(T(a,a) = T_1(a)\Big) = \frac{\left(\frac{q}{p}\right)^{a+1} - (a+1)\left(\frac{q}{p}\right) + a}{\left[1 - \left(\frac{q}{p}\right)^{-a}\right] \cdot \left[\left(\frac{q}{p}\right)^{a+1} - 1\right]}, \tag{4.9}$$

$$P\Big(T(a,a) = T_2(a)\Big) = \frac{\left(\frac{q}{p}\right)^{-(a+1)} - (a+1)\left(\frac{q}{p}\right)^{-1} + a}{\left[1 - \left(\frac{q}{p}\right)^{-a}\right] \cdot \left[\left(\frac{q}{p}\right)^{a+1} - 1\right]} \tag{4.10}$$

respectively.

The proof of the above theorem uses the following proposition:

Proposition 4.1 *The probability distribution functions of the random variables $Y^+_{T_1(a)}$ and $Y^-_{T_2(b)}$ are given by the following:*

1.

$$p_0^A = P(Y^+_{T_1(a)} = 0) = m_A + (1 - m_A) \cdot (1 - R_A), \tag{4.11}$$
$$p_k^A = P(Y^+_{T_1(a)} = k) = (1 - m_A) \cdot (1 - R_A) \cdot R_A^k, \quad \forall\ k \in \mathcal{N}^* \ ,$$

where m_A and R_A are given by equations (4.3) and (4.4) respectively.

2.

$$p_0^B = P(Y^-_{T_2(b)} = 0) = m_B + (1 - m_B) \cdot (1 - R_B), \tag{4.12}$$
$$p_k^B = P(Y^-_{T_2(b)} = k) = (1 - m_B) \cdot (1 - R_B) \cdot R_B^k, \quad \forall\ k \in \mathcal{N}^* \ ,$$

where m_B and R_B are given by equations (4.7) and (4.8) respectively.

In order to prove Proposition 4.1 and Theorem 4.1, we will need the following two lemmas.

Lemma 4.1 *For $a,\ b \in \mathcal{N}$, we have:*

$$E\left[T_1(a)\right] = \frac{1}{p} \cdot \frac{\left(\frac{q}{p}\right)^{-(a+1)} - (a+1)\left(\frac{q}{p}\right)^{-1} + a}{\left[\left(\frac{p}{q}\right)-1\right]\cdot\left[1-\left(\frac{q}{p}\right)\right]}, \tag{4.13}$$

$$E\left[T_2(b)\right] = \frac{1}{q} \cdot \frac{\left(\frac{q}{p}\right)^{b+1} - (b+1)\left(\frac{q}{p}\right) + b}{\left[\left(\frac{q}{p}\right)-1\right]\cdot\left[1-\left(\frac{p}{q}\right)\right]}. \tag{4.14}$$

Proof. The proof is similar to the procedure that appears in Siegmund (1985) (see [36]) for the purpose of computing the expectation of the CUSUM stopping time. With $S_n = \sum_{i=1}^{n} Z_i$, define the sequence of stopping times $\{N_k\}$ in the following way:

$$N_1 = \inf\{n \geq 1; S_n \notin (-1, b)\}.$$

If $S_{N_1} = b$, then $T_2 = N_1$, otherwise

$$S_{N_1} = \min_{k \in [0, N_1] \cap \mathcal{N}} S_k,$$

and

$$N_2 = \inf\{n \geq 1; S_{N_1+n} - S_{N_1} \notin (-1, b)\}.$$

If $S_{N_1+N_2} = b$, then $T_2 = N_1 + N_2$, else

$$S_{N_1+N_2} = \min_{k \in [0, N_1] \cap \mathcal{N}} S_k.$$

In general we have:

$$N_k = \inf\{n \geq 1; S_{N_1+\ldots+N_{k-1}+n} - S_{N_1+\ldots+N_{k-1}} \notin (-1, b)\},$$

and $T_2(b) = \sum_{i=1}^{M} N_i$, where

$$M = \inf\{k; S_{N_1+\ldots+N_k} - S_{N_1+\ldots+N_{k-1}} = b\}.$$

Since the $Z_i's$ and the $N_i's$ are independent, from Wald's identity it follows that

$$E\left[S_{N_1}\right] = E\left[Z_1\right] E\left[N_1\right] = (p-q) \cdot E\left[N_1\right], \tag{4.15}$$

$$E\left[T_2(b)\right] = E\left[N_1\right] E\left[M\right] = \frac{E\left[N_1\right]}{P(S_{N_1} = b)}, \tag{4.16}$$

since $M \sim \text{Geometric}\Big(P(S_{N_1} = b)\Big)$. From Theorem B.1 mentioned in the Appendix, we can write

$$P(S_{N_1} = b) = P\left(U(b,1) = U_1(b)\right) = \frac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^{b+1}}, \tag{4.17}$$

$$P(S_{N_1} \leq -1) = P\left(U(b,1) = U_2(1)\right) = \left(\frac{q}{p}\right) \frac{1-\left(\frac{q}{p}\right)^{b}}{1-\left(\frac{q}{p}\right)^{b+1}}. \tag{4.18}$$

Finally, we have

$$E\left[N_1\right] = b \cdot P(S_{N_1} = b) + 1 \cdot P(S_{N_1} = -1). \tag{4.19}$$

Using equations (4.19), (4.18), (4.17), (4.16), and (4.15), we get (4.14).

Equation (4.13) follows similarly by noticing that

$$Y_n^- = \sum_{i=1}^{n} R_i - \inf_{k \in [0,n] \cap \mathcal{N}} \sum_{i=1}^{k} R_i,$$

where

$$R_i = \begin{cases} 1, & \text{with probability } q, \\ -1, & \text{with probability } p. \end{cases}$$

Equation (4.15) becomes:

$$E\left[S_{N_1}\right] = E\left[Z_1\right] E\left[N_1\right] = (q-p) \cdot E\left[N_1\right]. \tag{4.20}$$

The result follows by using equations (4.19), (4.18), (4.17), (4.16), where we substitute p in place of q and q in place of p, and (4.20). This concludes the proof of the lemma. $\diamond$

Lemma 4.2 *We have*

$$Y_k^+ + Y_k^- = \max_{i \in [0,k] \cap \mathcal{N}} \{Y_i^+, Y_i^-\}.$$

Proof. Observe that

$$Y_k^+ + Y_k^- = -M_k^+ - M_k^-. \tag{4.21}$$

We notice that the process $Y_k^+ + Y_k^-$ can only increase when either $S_k = M_k^+$ or $-S_k = M_k^-$, both of which cannot happen since that would imply that (4.21) is 0. Therefore, $Y_k^+ + Y_k^-$ is constant in time unless either $Y_k^+ = 0$ or $Y_k^- = 0$, at which instant

$$\max\{Y_k^+, Y_k^-\} = \max_{i \in [0,k] \cap \mathcal{N}} \Big\{ \max\{Y_i^+, Y_i^-\} \Big\}.$$

This completes the proof of the lemma. $\diamond$

An important consequence of this lemma is that

$$Y_{T_1(a)}^+ = (\max_{n \le T_1(a)} Y_n^+ - a) \vee 0, \tag{4.22}$$

$$Y_{T_2(b)}^- = (\max_{n \le T_2(b)} Y_n^- - b) \vee 0. \tag{4.23}$$

We can now proceed to the proof of Proposition 4.1 and then to the proof of Theorem 4.1.

Proof of Proposition 4.1. Let us compute the probability distribution function of the random variable $Y_{T_1(a)}^+$, since the computation of the probability mass function of the random variable $Y_{T_2(b)}^-$ is done in a similar way. From equation (4.22), it

follows that

$$P\Big(Y^+_{T_1(a)} = 0\Big) = P\Big(\max_{n\le T_1(a)} Y^+_n < a\Big) + \\ +P\Big(\max_{n\le T_1(a)} Y^+_n \ge a\Big)\cdot P\Big(Y^+_{T_1(a)} = 0 \mid \max_{n\le T_1(a)} Y^+_n \ge a\Big), \tag{4.24}$$

while

$$P\Big(Y^+_{T_1(a)} = k\Big) = P\Big(\max_{n\le T_1(a)} Y^+_n \ge a\Big)\cdot P\Big(Y^+_{T_1(a)} = k \mid \max_{n\le T_1(a)} Y^+_n \ge a\Big). \tag{4.25}$$

We prove this proposition in three basic steps:

In the first step we compute the distribution of the random variable

$$\max_{n\le T_1(a)} S_n.$$

In the second step we show that

$$P\Big(Y^+_{T_1(a)} = k \mid \max_{n\le T_1(a)} Y^+_n \ge a\Big) \;=\; P\Big(\max_{n\le T_1(a)} S_n = k\Big), \quad k \in \mathcal{N}. \tag{4.26}$$

In the last step we compute $P(\max_{n\le T_1(a)} Y^+_n < a)$.

Beginning with the distribution of

$$\max_{n\le T_1(a)} S_n,$$

we notice that $\max_{n\le T_1(a)} S_n = k$ is the same event as k times going up by 1 before going down by a, and then going down by a before going up by 1. Thus we have

$$P\Big(\max_{n\le T_1(a)} S_n = k\Big) = P\Big(U_2(1) < U_1(a)\Big)^k \cdot P\Big(U_1(a) < U_2(1)\Big),$$

where the last equality follows from the definition of $U_1(a)$ and $U_2(b)$ as it appears in the Appendix. Therefore, using the result of Theorem B.1, we get that

$$\max_{n\le T_1(a)} S_n \sim \text{ Geometric}(\pi), \tag{4.27}$$

where $\pi = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^{a+1}}{1-\left(\frac{q}{p}\right)^{a+1}}$.

Let us proceed to the second step where we demonstrate

$$\mathcal{L}\Big(Y^+_{T_1(a)} \mid \max_{n\le T_1(a)} Y^+_n \ge a\Big) \;=\; \mathcal{L}\Big(\max_{n\le T_1(a)} S_n\Big). \tag{4.28}$$

To see this, let

$$R_1 = \sup\{n \le T_1(a); Y^+_n = 0\}. \tag{4.29}$$

Fix $k \in \mathcal{N}$. Then

$$\begin{aligned}
P\Big(Y^+_{T_1(a)} = k \mid \max_{n\le T_1(a)} Y^+_n \ge a\Big) &= \frac{P\Big(S_{T_1(a)} - \inf_{n\le T_1(a)} S_n = k\Big)}{P\Big(\max_{n\le T_1(a)} Y^+_n \ge a\Big)} \\
&= \frac{P\Big(S_{T_1(a)} - S_{R_1} + S_{R_1} - \inf_{n\le T_1(a)} S_n = k \mid R_1 < T_1(a)\Big)\cdot P\Big(R_1 < T_1(a)\Big)}{P\Big(\max_{n\le T_1(a)}(S_n - S_{R_1} + S_{R_1} - \inf_{k\le n} S_k) \ge a \mid R_1 < T_1(a)\Big)\cdot P\Big(R_1 < T_1(a)\Big)} \\
&= \frac{P\Big(S_{T_1(a)} - S_{R_1} + S_{R_1} - \inf_{n\le R_1} S_n = k \mid R_1 < T_1(a)\Big)}{P\Big(\max_{n\le T_1(a)}(S_n - S_{R_1} + S_{R_1} - \inf_{k\le R_1} S_k) \ge a \mid R_1 < T_1(a)\Big)} \\
= \frac{P\Big(S_{T_1(a)} - S_{R_1} = k \mid R_1 < T_1(a)\Big)}{P\Big(\max_{n\le T_1(a)} S_n - S_{R_1} \ge a \mid R_1 < T_1(a)\Big)} &= \frac{P\Big(S_{T_1(a)} - S_{R_1} = k \mid R_1 < T_1(a)\Big)}{P\Big(\max_{R_1\le n\le T_1(a)} S_n - S_{R_1} \ge a \mid R_1 < T_1(a)\Big)} \\
&= \frac{P(S_{T_1(a)} = k)}{P(\max_{n\le T_1(a)} S_n \ge a)} = \frac{(1-\pi)^{k+a}\pi}{\sum_{k=a}^{\infty}(1-\pi)^k\pi} = \frac{(1-\pi)^{k+a}\pi}{(1-\pi)^a} = P(\max_{n\le T_1(a)} S_n = k),
\end{aligned}$$

where $\pi = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^{a+1}}{1-\left(\frac{q}{p}\right)^{a+1}}$. Therefore we get

$$P\Big(Y^+_{T_1(a)} = k \mid \max_{t\le T_1(a)} Y^+_t \ge a\Big) \sim \text{Geometric}(\pi), \quad k \in \mathcal{N}. \tag{4.30}$$

What remains to be computed is $P(\max_{n\le T_1(a)} Y^+_n < a)$. From equation (4.22), it follows that

$$P(\max_{n\le T_1(a)} Y^+_n < a) \;=\; P\Big(T_1(a) < T_2(a)\Big). \tag{4.31}$$

To compute $P\Big(T_1(a) < T_2(a)\Big)$, we first notice that

$$T_1(a) = T(a,b) + \Big(T_1(a) - T(a,b)\Big)\mathbf{1}_{\{T(a,b)=T_2(b)\}}, \tag{4.32}$$

$$T_2(b) = T(a,b) + \Big(T_2(b) - T(a,b)\Big)\mathbf{1}_{\{T(a,b)=T_1(a)\}}. \tag{4.33}$$

Taking expectations we get

$$E\left[T_1(a)\right] = E\left[T(a,b)\right] + E\left[\Big(T_1(a) - T(a,b)\Big)\mathbf{1}_{\{T(a,b)=T_2(b)\}}\right], \tag{4.34}$$

$$E\left[T_2(b)\right] = E\left[T(a,b)\right] + E\left[\Big(T_2(b) - T(a,b)\Big)\mathbf{1}_{\{T(a,b)=T_1(a)\}}\right]. \tag{4.35}$$

With $a = b$ and equation (4.22), it follows that

$$E\left[T_1(a)\right] = E\left[T(a,a)\right] + E\left[T_1(a)\right] \cdot P\Big(T_2(a) < T_1(a)\Big), \tag{4.36}$$

$$E\left[T_2(a)\right] = E\left[T(a,a)\right] + E\left[T_2(a)\right] \cdot P\Big(T_1(a) < T_2(a)\Big). \tag{4.37}$$

Using

$$P\Big(T_1(a) < T_2(a)\Big) + P\Big(T_2(a) < T_1(a)\Big) = 1$$

and equations (4.36) and (4.37), we conclude that

$$P\Big(T_1(a) < T_2(a)\Big) = \frac{E\left[T_2(a)\right]}{E\left[T_2(a)\right] + E\left[T_1(a)\right]}. \tag{4.38}$$

The result now follows by substituting (4.30) and (4.38) into (4.24) and (4.25), using Lemma 4.1 and the fact that

$$P(\max_{n \le T_1(a)} Y_n^+ < a) + P(\max_{n \le T_1(a)} Y_n^+ \ge a) = 1.$$

This concludes the proof of the proposition. $\diamond$

We can now proceed to the proof of Theorem 4.1.

Proof of Theorem 4.1. We will prove the result in the case $b \ge a + 1 > 1$ since the result is proven similarly in the case when $a \ge b + 1 > 1$.

From Lemma 4.2 and equation (4.22), it follows that on the event $\{T_1(a) < T_2(b)\}$, we have

$$Y^+_{T_1(a)} = \begin{cases} 0 & \text{if } \max_{n \le T_1(a)} Y^+_n < a, \\ \max_{n \le T_1(a)} Y^+_n - a & \text{if } a \le \max_{n \le T_1(a)} Y^+_n < b. \end{cases} \tag{4.39}$$

From equation (4.39) it becomes obvious that on the event $\{T_1(a) < T_2(b)\}$, $Y^+_{T_1(a)}$ cannot exceed the level $b - a$, or cannot be exactly equal to this level. Therefore

$$P\Big(T_1(a) < T_2(b)\Big) \;=\; \sum_{k=0}^{b-a-1} P\Big(Y^+_{T_1(a)} = k\Big). \tag{4.40}$$

Using Proposition 4.1 the result follows. This completes the proof of the Theorem 4.1. $\diamond$

It is worth noting that we can readily get the expectation of $T(a,b) = T_1(a) \wedge T_2(b)$.

Remark 4.1 *We can distinguish the following three cases for the expectation of $T(a,b) = T_1(a) \wedge T_2(b)$ in terms of the expectations of $T_1(a)$ and $T_2(b)$ (as they appear in Lemma 4.1):*

1. $b \ge a + 1 > 1$

$$E\left[T(a,b)\right] \;=\; E\left[T_1(a)\right] \cdot \left[1 - (1 - m_A) \cdot R_A^{b-a}\right], \tag{4.41}$$

where R_A and m_A as they appear in equations (4.4) and (4.3) respectively.

2. $a \ge b + 1 > 1$

$$E\left[T(a,b)\right] \;=\; E\left[T_2(b)\right] \cdot \left[1 - (1 - m_B) \cdot R_B^{a-b}\right], \tag{4.42}$$

where R_B and m_B as they appear in equations (4.8) and (4.7) respectively.

3. $b = a$

$$E\left[T(a,a)\right] = \frac{E\left[T_2(a)\right] \cdot E\left[T_1(a)\right]}{E\left[T_2(a)\right] + E\left[T_1(a)\right]}. \tag{4.43}$$

Proof: The proof, for case 1, is a mere consequence of the following equation

$$\begin{aligned} E\left[T_1(a)\right] &= E\left[T(a,a)\right] + E\left[T_1(a) - T_2(b) \mid T_2(b) < T_1(a)\right] \cdot P\Big(T_2(b) < T_1(a)\Big) \\ &= E\left[T(a,b)\right] + E\left[T_1(a)\right] \cdot P\Big(T_2(b) < T_1(a)\Big), \end{aligned}$$

and Theorem 4.1. $\diamond$

It is interesting to see the probabilities of stopping on downward fall or upward rally for an unbiased random walk.

Corollary 4.1 *Let* $S_n = \sum_{i=1}^{n} Z_i$ *be the evolution of the wealth of the gambler in a game of equal odds* ($p = q = \frac{1}{2}$)*, and let* $T(a,b)$*,* $T_1(a)$ *and* $T_2(b)$ *be stopping times defined as above. We distinguish the following three cases:*

1. $b \geq a + 1 > 1$

The probabilities of stopping the game on downward fall or upward rally are given by

$$P\Big(T(a,b) = T_1(a)\Big) = 1 - \frac{1}{2} \cdot \left(\frac{a}{a+1}\right)^{b-a}, \tag{4.44}$$

$$P\Big(T(a,b) = T_2(b)\Big) = \frac{1}{2} \cdot \left(\frac{a}{a+1}\right)^{b-a}. \tag{4.45}$$

2. $a \geq b + 1 > 1$

The probabilities of stopping the game on downward fall or upward rally are given by

$$P\Big(T(a,b) = T_1(a)\Big) = \frac{1}{2} \cdot \Big(\frac{b}{b+1}\Big)^{a-b}, \tag{4.46}$$

$$P\Big(T(a,b) = T_2(b)\Big) = 1 - \frac{1}{2} \cdot \Big(\frac{b}{b+1}\Big)^{a-b}. \tag{4.47}$$

3. $a = b$

The probabilities of stopping the game on downward fall or upward rally are given by

$$P\Big(T(a,a) = T_1(a)\Big) = P\Big(T(a,a) = T_2(a)\Big) = \frac{1}{2}. \tag{4.48}$$

Proof. All of the above results are a simple consequence of taking the limit as $p \to \frac{1}{2}$ in Theorem 4.1. $\diamond$

Corollary 4.2 *Let $S_n = \sum_{i=1}^{n} Z_i$ be the evolution of the wealth of the gambler in a game of equal odds ($p = q = \frac{1}{2}$). The probability distribution functions of the random variables $Y^+_{T_1(a)}$ and $Y^-_{T_2(b)}$ are given by the following:*

1.

$$p_0^A = P(Y^+_{T_1(a)} = 0) = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{a+1}, \tag{4.49}$$

$$p_k^A = P(Y^+_{T_1(a)} = k) = \frac{1}{2} \cdot \frac{1}{a+1} \cdot \Big(\frac{a}{a+1}\Big)^k, \quad \forall\ k \in \mathcal{N}^*.$$

2.

$$p_0^B = P(Y^-_{T_2(b)} = 0) = \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{b+1}, \tag{4.50}$$

$$p_k^B = P(Y^-_{T_2(b)} = k) = \frac{1}{2} \cdot \frac{1}{b+1} \cdot \Big(\frac{b}{b+1}\Big)^k, \quad \forall\ k \in \mathcal{N}^*.$$

Proof. This corollary is a simple consequence of Proposition 4.1 by taking the limit as $p \to \frac{1}{2}$. $\diamond$

4.1.2 The continuous time framework

In the continuous time framework, the wealth of the gambler at each time point t is assumed to follow

$$X_t = W_t - \mu t,$$

where $\mu > 0$ and W_t is a Brownian motion.

The quantity

$$X_t - \inf_{s \in [0,t]} X_s$$

measures the size of the upward rally comparing the present value of the wealth to its historical minimum, while the quantity

$$\sup_{s \in [0,t]} X_s - X_t$$

measures the size of the downward fall comparing the present value of the wealth to its historical maximum.

The aim of this Section is to determine the probability that the gambler would quit the game on the upward rally in contrast to quitting the game on the downward fall. We introduce the stopping times:

$$T_1^c(a) = \inf\{t \geq 0 : \sup_{s \in [0,t]} X_s - X_t = a, \quad a \in \mathcal{R}_+\},$$

and

$$T_2^c(b) = \inf\{t \geq 0 : X_t - \inf_{s \in [0,t]} X_s = b, \quad b \in \mathcal{R}_+\}.$$

The gambler stops at $T^c(a,b) = T_1^c(a) \wedge T_2^c(b)$. The stopping times $T_1^c(a)$ and $T_2^c(b)$ indicate the first time of reaching the critical level of the downward fall $T_1^c(a)$, or the first time of reaching critical level of the upward rally $T_2^c(b)$. In this Section we compute the probabilities of the events $\{T^c(a,b) = T_1^c(a)\}$, which represents stopping the game on downward fall, and $\{T^c(a,b) = T_2^c(b)\}$, which represents stopping the game on upward rally.

In order to simplify notation we introduce the following processes:

$$\begin{aligned} m_t^+ &:= \inf_{s\in[0,t]} X_s, \\ m_t^- &:= \inf_{s\in[0,t]} (-X_s) = -\sup_{s\in[0,t]} X_s, \\ y_t^+ &:= X_t - m_t^+, \\ y_t^- &:= -X_t - m_t^-. \end{aligned}$$

Using the above notation, the stopping times $T_1^c(a)$ and $T_2^c(b)$ become

$$T_1^c(a) = \inf\{t \geq 0 : y_t^- = a, \quad a \in \mathcal{R}_+\},$$

$$T_2^c(b) = \inf\{t \geq 0 : y_t^+ = b, \quad b \in \mathcal{R}_+\}.$$

Theorem 4.2 *Let $X_t = W_t - \mu t$ be the evolution of the wealth of the gambler and let T^c, T_1^c and T_2^c be stopping times defined as above and $\mu > 0$. We distinguish the following two cases:*

1. $b \geq a > 0$

   The probabilities of stopping at downward fall or upward rally are given by

$$P\Big(T^c(a,b) = T_1^c(a)\Big) = m_A^c + (1 - m_A^c) \cdot \left[1 - \exp\left(-\tfrac{2\mu}{1-e^{-2\mu a}} \cdot (b-a)\right)\right], \tag{4.51}$$

$$P\Big(T^c(a,b) = T_2^c(b)\Big) = (1 - m_A^c) \cdot \exp\left(-\tfrac{2\mu}{1-e^{-2\mu a}} \cdot (b-a)\right), \tag{4.52}$$

where

$$m_A^c = \frac{e^{2\mu a} - 2\mu a - 1}{e^{2\mu a} + e^{-2\mu a} - 2}. \tag{4.53}$$

2. $a \geq b > 0$

The probabilities of stopping at downward fall or upward rally are given by

$$P\Big(T^c(a,b) = T_1^c(a)\Big) = (1 - m_B^c) \cdot \exp\left(-\tfrac{2\mu}{e^{2\mu b}-1} \cdot (a-b)\right), \tag{4.54}$$

$$P\Big(T^c(a,b) = T_2^c(b)\Big) = m_B^c + (1 - m_B^c) \cdot \left[1 - \exp\left(-\tfrac{2\mu}{e^{2\mu b}-1} \cdot (a-b)\right)\right], \tag{4.55}$$

where

$$m_B^c = \frac{e^{-2\mu b} + 2\mu b - 1}{e^{2\mu b} + e^{-2\mu b} - 2}. \tag{4.56}$$

The proof of the theorem uses the next proposition:

Proposition 4.2 *The probability distribution functions of the random variables* $y^+_{T_1^c(a)}$ *and* $y^-_{T_2^c(b)}$ *are given by:*

1.

$$P(y^+_{T_1^c(a)} = 0) = m_A^c \tag{4.57}$$

$$P(y^+_{T_1^c(a)} \in dr) = (1 - m_A^c) \cdot \left[\tfrac{2\mu}{1-e^{-2\mu a}} \cdot \exp\left(-\tfrac{2\mu}{1-e^{-2\mu a}} \cdot r\right)\right] dr \ , \ r > 0, \tag{4.58}$$

where m_A^c *is given by equation (4.53).*

2.

$$P(y^-_{T_2^c(b)} = 0) = m_B^c \tag{4.59}$$

$$P(y^-_{T_2^c(b)} \in dr) = (1 - m_B^c) \cdot \left[\frac{2\mu}{e^{2\mu b}-1} \cdot \exp\left(-\frac{2\mu}{e^{2\mu b}-1} \cdot r\right)\right] dr \ , \ r > 0, \tag{4.60}$$

where m_B^c is given by equation (4.56).

In order to prove Proposition 4.2 and Theorem 4.2, we will need the following two lemmas.

Lemma 4.3 *For $a,\ b \in \mathcal{R}_+$, we have:*

$$E\left[T_1^c(a)\right] = \frac{e^{-2\mu a} + 2\mu a - 1}{2\mu^2}, \tag{4.61}$$

$$E\left[T_2^c(b)\right] = \frac{e^{2\mu b} - 2\mu b - 1}{2\mu^2}. \tag{4.62}$$

Proof. Let $g_2(x) = e^{2\mu x} - 2\mu x - 1$. By applying Itô's rule to the processes $g_2(y_t^+)$ we get

$$dg_2(y_t^+) = g_2'(y_t^+)dW_t - \mu g_2'(y_t^+)dt - g_2'(y_t^+)dm_t^+ + \frac{1}{2}g_2''(y_t^+)dt. \tag{4.63}$$

We notice that the third term in the right hand side of the above equality disappears because $dm_t^+ \neq 0$ only when $y_t^+ = 0$ and $g_2'(0) = 0$. We also notice that the function g_2 satisfies the second order differential equation

$$-\mu g_2'(x) + \frac{1}{2}g_2''(x) = 2\mu^2. \tag{4.64}$$

By integrating from 0 to $T_2^c(b)$, we have

$$g_2(y^+_{T_2^c(b)}) - g_2(0) = \int_0^{T_2^c(b)} g_2'(y_t^+)dW_t + \int_0^{T_2^c(b)} \left(-\mu g_2'(y_t^+) + \frac{1}{2}g_2''(y_t^+)\right) dt.$$

Using equation (4.64), $y^+_{T_2^c(b)} = b$, $g_2(0) = 0$ and taking expectations we get

$$g_2(b) \;=\; 2\mu^2 E\left[T_2^c(b)\right]. \tag{4.65}$$

Consequently,

$$E\left[T_2^c(b)\right] \;=\; \frac{g_2(b)}{2\mu^2}. \tag{4.66}$$

Similarly, by applying Itô's rule to $g_1(y_t^-)$, where $g_1(x) = e^{-2\mu x} + 2\mu x - 1$, we have

$$g_1(y^-_{T_1^c(a)}) - g_1(0) \;=\; -\int_0^{T_1^c(a)} g_1'(y_t^-)dW_t + \int_0^{T_1^c(a)} \left(\mu g_1'(y_t^-) + \frac{1}{2}g_1''(y_t^-)\right) dt$$

from which it follows that

$$E\left[T_1^c(a)\right] \;=\; \frac{g_1(a)}{2\mu^2}. \tag{4.67}$$

This concludes the proof of the lemma. $\diamond$

Lemma 4.4 *We have*

$$y_t^+ + y_t^- = \max_{s\le t}\{y_s^+, y_s^-\}.$$

Proof. Observe that

$$y_t^+ + y_t^- = -m_t^+ - m_t^-. \tag{4.68}$$

We notice that the process $y_t^+ + y_t^-$ can only increase when either $X_t = m_t^+$ or $-X_t = m_t^-$, both of which cannot happen at the same time since that would imply that $y_t^+ + y_t^-$ is 0. Therefore, $y_t^+ + y_t^-$ is a constant as a function of time unless either $y_t^+ = 0$ or $y_t^- = 0$, at which instant

$$\max\{y_t^+, y_t^-\} = \sup_{s\in[0,t]}\Big\{\max\{y_s^+, y_s^-\}\Big\}.$$

This completes the proof of the lemma. $\diamond$

As a consequence of this lemma we have

$$y^+_{T_1^c(a)} = (\max_{t \le T_1^c(a)} y_t^+ - a) \vee 0, \tag{4.69}$$

$$y^-_{T_2^c(b)} = (\max_{t \le T_2^c(b)} y_t^- - b) \vee 0. \tag{4.70}$$

Finally, in order to proceed to the proof of Proposition 4.2 and Theorem 4.2, we will use the results of Taylor in [43] and Lehoczky in [18]. Taylor computes the bivariate Laplace transform of $X_{T_1^c(a)}$ and $T_1^c(a)$, where T_1^c is defined as above. Lehoczky pointed out that the random variable $X_{T_1^c(a)} + a = \sup_{t \le T_1^c(a)} X_t$ has the exponential distribution:

$$X_{T_1^c(a)} + a \sim \text{Exp}\left(\tfrac{2\mu}{1-e^{-2\mu a}}\right). \tag{4.71}$$

Note that the exponential parameter becomes equal to $\frac{1}{a}$ in the case when $\mu = 0$. Now we can proceed to the proof of Proposition 4.2 and then to the proof of Theorem 4.2.

Proof of Proposition 4.2. We will only compute the probability density function of the random variable $y^+_{T_1^c(a)}$ since the computation of the probability density function of the random variable $y^-_{T_2^c(b)}$ is done in a similar way. From equation (4.69), it follows that

$$P\Big(y^+_{T_1^c(a)} = 0\Big) = P\Big(\max_{t \le T_1^c(a)} y_t^+ < a\Big), \tag{4.72}$$

while

$$\begin{aligned} P\Big(y^+_{T_1^c(a)} \in dr\Big) &= P\Big(\max_{t \le T_1^c(a)} y_t^+ \ge a\Big) \cdot P\Big(y^+_{T_1^c(a)} \in dr \mid \max_{t \le T_1^c(a)} y_t^+ \ge a\Big) \\ &= P\Big(y^+_{T_1^c(a)} > 0\Big) \cdot P\Big(y^+_{T_1^c(a)} \in dr \mid y^+_{T_1^c(a)} > 0\Big), \ r > 0. \end{aligned} \tag{4.73}$$

In the next discussion we first demonstrate

$$\mathcal{L}\Big(y^+_{T_1^c(a)} \mid y^+_{T_1^c(a)} > 0\Big) \;=\; \mathcal{L}\Big(X_{T_1^c(a)} + a\Big). \tag{4.74}$$

To see this, let

$$R_1^c = \sup\{t \le T_1^c(a); y_t^+ = 0\}. \tag{4.75}$$

Fix $r > 0$. Then

$$\begin{aligned}
P\Big(y^+_{T_1^c(a)} \in dr \mid y^+_{T_1^c(a)} > 0\Big) &= \frac{P\Big(X_{T_1^c(a)} - \inf_{s \le T_1^c(a)} X_s \in dr\Big)}{P\Big(\max_{t \le T_1^c(a)} y_t^+ \ge a\Big)} \\
&= \frac{P\Big(X_{T_1^c(a)} - X_{R_1^c} + X_{R_1^c} - \inf_{s \le T_1^c(a)} X_s \in dr \mid R_1^c < T_1^c(a)\Big) \cdot P\Big(R_1^c < T_1^c(a)\Big)}{P\Big(\max_{t \le T_1^c(a)} (X_t - X_{R_1^c} + X_{R_1^c} - \inf_{s \le t} X_s) \ge a \mid R_1^c < T_1^c(a)\Big) \cdot P\Big(R_1^c < T_1^c(a)\Big)} \\
= \frac{P\Big(X_{T_1^c(a)} - X_{R_1^c} + X_{R_1^c} - \inf_{s \le R_1^c} X_s \in dr \mid R_1^c < T_1^c(a)\Big)}{P\Big(\max_{t \le T_1^c(a)} (X_t - X_{R_1^c} + X_{R_1^c} - \inf_{s \le R_1^c} X_s) \ge a \mid R_1^c < T_1^c(a)\Big)} &= \frac{P\Big(X_{T_1^c(a)} - X_{R_1^c} \in dr \mid R_1^c < T_1^c(a)\Big)}{P\Big(\max_{t \le T_1^c(a)} X_t - X_{R_1^c} \ge a \mid R_1^c < T_1^c(a)\Big)} \\
&= \frac{P\Big(X_{T_1^c(a)} - X_{R_1^c} \in dr \mid R_1^c < T_1^c(a)\Big)}{P\Big(\max_{R_1^c \le t \le T_1^c(a)} X_t - X_{R_1^c} \ge a \mid R_1^c < T_1^c(a)\Big)} = \frac{P(X_{T_1^c(a)} \in dr)}{P(\max_{t \le T_1^c(a)} X_t \ge a)} \\
&= \frac{\lambda e^{-\lambda r} e^{-\lambda a} dr}{e^{-\lambda a}} = \lambda e^{-\lambda r} dr = P(X_{T_1^c(a)} + a \in dr),
\end{aligned}$$

where $\lambda = \frac{2\mu}{1 - e^{-2\mu a}}$. Therefore we get

$$P\Big(y^+_{T_1^c(a)} \in dr \mid y^+_{T_1^c(a)} > 0\Big) \sim \mathrm{Exp}\Big(\frac{2\mu}{1 - e^{-2\mu a}}\Big),\ r > 0. \tag{4.76}$$

From equation (4.69), it follows that

$$P(y^+_{T_1^c(a)} = 0) \;=\; P\Big(T_1^c(a) < T_2^c(a)\Big). \tag{4.77}$$

With T_1^c, T_2^c in place of T_1 and T_2 respectively in equations (4.36) and (4.37), we get

$$E\left[T_1^c(a)\right] \;=\; E\left[T^c(a,a)\right] + E\left[T_1^c(a)\right] \cdot P\Big(T_2^c(a) < T_1^c(a)\Big), \tag{4.78}$$

$$E\left[T_2^c(a)\right] \;=\; E\left[T^c(a,a)\right] + E\left[T_2^c(a)\right] \cdot P\Big(T_1^c(a) < T_2^c(a)\Big). \tag{4.79}$$

Using

$$P\Big(T_1^c(a) < T_2^c(a)\Big) + P\Big(T_2^c(a) < T_1^c(a)\Big) = 1$$

and equations (4.78) and (4.79), we conclude that

$$P\Big(T_1^c(a) < T_2^c(a)\Big) \quad = \quad \frac{E\left[T_2^c(a)\right]}{E\left[T_2^c(a)\right] + E\left[T_1^c(a)\right]}. \tag{4.80}$$

The result now follows by substituting (4.76), (4.77), (4.80) into equation (4.73) using Lemma 4.3. This completes the proof of Proposition 4.2. $\diamond$

Proof of Theorem 4.2. We will prove the theorem in the case that $b \geq a$ since the proof is similar in the case $a \geq b$. Suppose that $b \geq a$.

From Lemma 4.4 and equation (4.69), it follows that on the event $\{T_1^c(a) < T_2^c(b)\}$ we have

$$y^+_{T_1^c(a)} = \begin{cases} 0 & \text{if } \max_{s \leq T_1^c(a)} y_s^+ < a, \\ \max_{s \leq T_1^c(a)} y_s^+ - a & \text{if } a \leq \max_{s \leq T_1^c(a)} y_s^+ < b. \end{cases} \tag{4.81}$$

Therefore,

$$P\Big(T_1^c(a) < T_2^c(b)\Big) \quad = \quad P(y^+_{T_1^c(a)} = 0) + \int_{0^+}^{b-a} P(y^+_{T_1^c(a)} \in dr), \tag{4.82}$$

and the result is obtained from Proposition 4.2. This completes the proof of the Theorem 4.2. $\diamond$

Remark 4.2 *We can distinguish the following three cases for the expectation of $T^c(a,b) = T_1^c(a) \wedge T_2^c(b)$ in terms of the expectations of $T_1^c(a)$ and $T_2^c(b)$ (as they appear in Lemma 4.3):*

1. $b \geq a > 0$

$$E\left[T^c(a,b)\right] \quad = \quad E\left[T_1^c(a)\right] \cdot \left[1 - (1 - m_A^c) \cdot e^{-\lambda_{neg}(b-a)}\right], \tag{4.83}$$

where $\lambda_{neg} = \frac{2\mu}{1-e^{-2\mu a}}$ and m_A^c as it appears in equation (4.53).

2. $a \geq b > 0$

$$E\left[T^c(a,b)\right] = E\left[T_2^c(b)\right] \cdot \left[1 - (1 - m_B^c) \cdot e^{-\lambda_{pos}(a-b)}\right], \tag{4.84}$$

where $\lambda_{pos} = \frac{2\mu}{e^{2\mu b}-1}$ *and* m_B^c *as it appears in equation (4.56).*

Proof: The proof for $b > a$ is very similar to the proof of Lemma 4.1. The case $a > b$ is done in a similar way too. $\diamond$

Corollary 4.3 *Let* $X_t = W_t$ *be the evolution of the wealth of the gambler and let* T^c, T_1^c *and* T_2^c *be stopping times defined as above in a game of equal chances. We distinguish the following two cases:*

1. $b \geq a > 0$

The probabilities of stopping at downward fall or upward rally are given by

$$P\Big(T^c(a,b) = T_1^c(a)\Big) = \frac{1}{2} + \frac{1}{2} \cdot \left[1 - e^{-\frac{1}{a}(b-a)}\right], \tag{4.85}$$

$$P\Big(T^c(a,b) = T_2^c(b)\Big) = \frac{1}{2} \cdot e^{-\frac{1}{a}(b-a)}. \tag{4.86}$$

2. $a \geq b > 0$

The probabilities of stopping at downward fall or upward rally are given by

$$P\Big(T^c(a,b) = T_1^c(a)\Big) = \frac{1}{2} \cdot e^{-\frac{1}{b}(a-b)}, \tag{4.87}$$

$$P\Big(T^c(a,b) = T_2^c(b)\Big) = \frac{1}{2} + \frac{1}{2} \cdot \left[1 - e^{-\frac{1}{b}(a-b)}\right]. \tag{4.88}$$

Proof. It is a simple consequence of Theorem 4.2 by taking the limit as $\mu \to 0$.

$\diamond$

Corollary 4.4 *Let $X_t = W_t$ be the evolution of the wealth of the gambler. The probability distribution function of the random variables $y^+_{T_1^c(a)}$ and $y^-_{T_2^c(b)}$ are given by*

1.

$$P(y^+_{T_1^c(a)} = 0) = \frac{1}{2} \tag{4.89}$$

$$P(y^+_{T_1^c(a)} \in dr) = \frac{1}{2} \cdot \left[\frac{1}{a} e^{-\frac{1}{a}r}\right] dr \ , \ r > 0. \tag{4.90}$$

2.

$$P(y^-_{T_2^c(b)} = 0) = \frac{1}{2} \tag{4.91}$$

$$P(y^-_{T_2^c(b)} \in dr) = \frac{1}{2} \cdot \left[\frac{1}{b} e^{-\frac{1}{b}r}\right] dr \ , \ r > 0. \tag{4.92}$$

Proof. The above corollary is a consequence of Proposition 4.2 by letting $\mu \to 0$.

◇

4.2 Conclusions & Future work

In this Chapter we are able to compute explicitly the probabilities of winning or losing in a game of chance based on quitting the game after a significant upward rally or downward fall both in the continuous and in the discrete time framework. In doing so, we have also managed to compute the distributions of the random variables $Y^+_{T_1}$, $Y^-_{T_2}$ in discrete time with their continuous counterparts $y^+_{T_1^c}$ and $y^-_{T_2^c}$ respectively. These results are the first step to getting the joint distribution of the random variables $\max_{s \le t} y^+_s$ and $\max_{s \le t} y^-_s$ in both the continuous and the discrete time setting. This is a clear improvement of the already existing that appears in

[21], whereby the marginal distribution of $\max_{s \leq t} y_s^-$ is computed in the continuous time framework.

Another very important result that follows directly from the above probabilities appears in the computation of the expected value of the minimum stopping rule, namely in Remarks 4.2 and 4.1. The importance of this result is its connection to the 2-CUSUM stopping rule, since $T(b, a)$ can be seen as the two-sided CUSUM stopping time with $T_1(b)$ and $T_2(a)$ as its one sided CUSUM branches. The 2-CUSUM stopping rule, as seen in the first two chapters has been widely used in the literature for the purpose of detecting two-sided changes. In Yashchin (1985) (see [51]), one can find the Laplace transform of the 2-CUSUM stopping time in the discrete time framework when $b = a$. Although an expression for the Laplace transform is also given for $a \neq b$, only upper and lower bounds for the expected value of $T(a, b)$ are achieved, and as a result, only upper and lower bounds are given for the $P\Big(T_1(a) < T_2(b)\Big)$. His work only deals with the discrete time model. Our result provides the exact computation of the expected value of the 2-CUSUM stopping rule with equal drift parameters in each of its one-sided CUSUM stopping branches. This is a result that can help us find the best 2-CUSUM rule (in the sense of the first Chapter) among the family of 2-CUSUM rules with different thresholds and equal drift parameters in their respective one-sided CUSUM branches. This result could then potentially be extended to identifying the best 2-CUSUM rule amongst all 2-CUSUM rules that are members of the class $\mathcal{G}$ and which therefore have different drift and threshold parameters in their respective one-sided CUSUM branches.

Appendix B

Review of the Gambler's Ruin Problem in the Traditional Setup

This Section reviews the very well known result of the gambler's ruin problem. We distinguish between the discrete time and the continuous time framework.

The discrete time framework

Let Z_i, $i \in \mathcal{N}$ be a sequence of independent identically distributed random variables with the following distribution

$$Z_i = \begin{cases} 1, & \text{with probability } \ p, \\ -1, & \text{with probability } \ q, \end{cases}$$

where $p + q = 1$ and $p < q$. Each Z_i represents a win or loss of the gambler on the i-th bet. The wealth (or cumulative winnings) of the gambler after n bets is given by

$$S_n = \sum_{i=1}^{n} Z_i.$$

The gambler stops as soon as his or her wealth reaches some upper level b or some lower level $-a$, where $a, b \in \mathcal{N}$. This event occurs at the stopping time

$$U(a, b) = \inf\{n \in \mathcal{N} : S_n = -a \text{ or } S_n = b\}.$$

Let us introduce the stopping times

$$U_1(a) = \inf\{n \in \mathcal{N} : S_n = -a\},$$

and

$$U_2(b) = \inf\{n \in \mathcal{N} : S_n = b\}.$$

In other words, $U_1(a)$ is the time when gambler's wealth reaches the level $-a$, and $U_2(b)$ is the time at which his or her wealth reaches the level b. We are interested in computing the probabilities of the events $\{U(a,b) = U_1(a)\}$, i.e., exiting the game on a loss, and $\{U(a,b) = U_2(b)\}$, i.e., exiting the game on a win. We have the following result which determines these probabilities:

Theorem B.1 *Let $S_n = \sum_{i=1}^n Z_i$ be the evolution of the wealth of the gambler and let $U(a,b)$, $U_1(a)$ and $U_2(b)$ be stopping times defined as above, with $a, b \in \mathcal{N}$. Then*

$$P\Big(U(a,b) = U_1(a)\Big) = P\Big(U_1(a) < U_2(b)\Big) = \frac{\left(\frac{q}{p}\right)^a - \left(\frac{q}{p}\right)^{a+b}}{1 - \left(\frac{q}{p}\right)^{a+b}}, \tag{B.1}$$

and

$$P\Big(U(a,b) = U_2(b)\Big) = P\Big(U_2(b) < U_1(a)\Big) = \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^{a+b}}. \tag{B.2}$$

Proof. The result is a simple consequence of the Optional Sampling Theorem applied to the discrete time martingale $M_n = \left(\frac{q}{p}\right)^{S_n}$. In particular,

$$1 = E\left[M_{U(a,b)}\right] \quad = \quad P\Big(U(a,b) = U_1(a)\Big) \cdot \left(\frac{q}{p}\right)^{-a} + P\Big(U(a,b) = U_2(b)\Big) \cdot \left(\frac{q}{p}\right)^{b}.$$

The fact that $P\Big(U(a,b) = U_1(a)\Big) + P\Big(U(a,b) = U_2(b)\Big) = 1$ concludes the proof.

$\diamond$

Remark B.1 For the case of equal odds ($p = q = \frac{1}{2}$), we can pass to the limit in the previously computed probabilities to conclude

$$P\Big(U(a,b) = U_1(a)\Big) = P\Big(U_1(a) < U_2(b)\Big) = \frac{b}{a+b}, \tag{B.3}$$

and

$$P\Big(U(a,b) = U_2(b)\Big) = P\Big(U_2(b) < U_1(a)\Big) = \frac{a}{a+b}. \tag{B.4}$$

The continuous time framework

In the continuous time framework, the wealth X_t of the gambler follows a drifted Brownian motion

$$X_t = W_t - \mu t, \tag{B.5}$$

for $\mu > 0$, where W_t is a standard Brownian motion. The analogous stopping times introduced above now become

$$U^c(a,b) = \inf\{t \geq 0 : X_t = -a \text{ or } X_t = b\},$$

$$U_1^c(a) = \inf\{t \geq 0 : X_t = -a\},$$

and

$$U_2^c(b) = \inf\{t \geq 0 : X_t = b\},$$

with $a, b \in \mathcal{R}_+$. The following theorem determines probabilities of events $\{U^c(a,b) = U_1^c(a)\}$ and $\{U^c(a,b) = U_2^c(b)\}$.

Theorem B.2 *Let $X_t = W_t - \mu t$ be the evolution of the wealth of the gambler and let $U^c(a,b)$, $U_1^c(a)$ and $U_2^c(b)$ be the stopping times defined above, with $a, b \in \mathcal{R}_+$, $\mu > 0$. Then*

$$P\Big(U^c(a,b) = U_1^c(a)\Big) = P\Big(U_1^c(a) < U_2^c(b)\Big) = \frac{e^{2\mu b} - 1}{e^{2\mu b} - e^{-2\mu a}}, \tag{B.6}$$

and

$$P\Big(U^c(a,b) = U_2^c(b)\Big) = P\Big(U_2^c(b) < U_1^c(a)\Big) = \frac{1 - e^{-2\mu a}}{e^{2\mu b} - e^{-2\mu a}}. \tag{B.7}$$

Proof. Consider the martingale $M_t = e^{2\mu X_t}$. Then, according to the Optional Sampling Theorem,

$$1 = E\left[M_{U^c(a,b)}\right] = P\Big(U^c(a,b) = U_1^c(a)\Big) \cdot e^{-2\mu a} + P\Big(U^c(a,b) = U_2^c(b)\Big) \cdot e^{2\mu b}.$$

Since

$$P\Big(U^c(a,b) = U_1^c(a)\Big) + P\Big(U^c(a,b) = U_2^c(b)\Big) = 1,$$

simple algebra concludes the proof. $\diamond$

Remark B.2 When $\mu = 0$, we can take the limit as $\mu \to 0$ in the previously computed probabilities to conclude

$$P\Big(U^c(a,b) = U_1^c(a)\Big) = P\Big(U_1^c(a) < U_2^c(b)\Big) = \frac{b}{a+b}, \tag{B.8}$$

$$P\Big(U^c(a,b) = U_2^c(b)\Big) = P\Big(U_2^c(b) < U_1^c(a)\Big) = \frac{a}{a+b}. \tag{B.9}$$

Bibliography

[1] E. Andersson. Monitoring cyclical processes: A non-parametric approach. *Journal of Applied Statistics*, 29:973–990, 2002.

[2] G.A. Barnard. Control charts and stochastic processes. *Journal of the Royal Statistical Society, Series B*, 11:239–271, 1959.

[3] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes: Theory and Applications.* Prentice Hall, Englewood Cliffs, NJ, 1993.

[4] E. Bayraktar and S. Dayanik. Poisson disorder problem with exponential penalty for delay. Princeton University, 2003.

[5] E. Bayraktar, S. Dayanik, and I. Karatzas. The standard Poisson disorder revisited. Princeton University, 2004.

[6] M. Beibel. A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure. *Annals of Statistics*, 24:1804–1812, 1996.

[7] B.E. Brodsky and B.S. Darkhovsky. *Non-parametric statistical diagnosis.* Kluwer, The Netherlands, 2000.

[8] H. Chernoff and S. Zacks. Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, 35:999–1018, 1964.

[9] V. Dragalin. The design and analysis of 2-CUSUM procedure. *Communications in Statistics - Simulations*, 26:67–81, 1997.

[10] V.P. Dragalin. Optimality of the generalized CUSUM procedure in quickest detection problem. *Statistics and Control of Random Processes: Proceedings of the Steklov Institute of Mathematics*, 202:107–120, 1994. American Mathematical Society, Providence, Rhode Island.

[11] M. Frisen. Evaluations of methods for statistical surveillance. *Statistics in Medicine*, 11:1489–1502, 1992.

[12] L. J. Galchuk and B. L. Rozovsky. The "disorder" problem for a Poisson process. *Theory of Probability and its Applications*, 16:712–717, 1971.

[13] E.M. Hillson, J.H. Reeves, and C.A. McMillan. A statistical signalling model for use in surveillance of adverse drug reaction data. *Journal of Applied Statistics*, 25:23–40, 1998.

[14] I. Karatzas. A note on Bayesian detection of change-points with an expected miss criterion. *Statistics and Decisions*, 21:3–14, 2002.

[15] I. Karatzas. Bayesian adaptive detection of a change-point. submitted to the Journal of Statistics and Decisions, 2003.

[16] I. Karatzas and E.S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 2nd edition, 1991.

[17] J. H. B. Kemperman. *The passage problem for a stationary Markov chain*, volume 1 of *Statistical research monographs*. The University of Chicago press, 1961.

[18] J.P. Lehoczky. Formulas for stopped diffusion processes with stopping times based on the maximum. *Annals of Probability*, 5:601–607, 1977.

[19] B. Levin and J. Kline. The CUSUM test of homogeneity with an application in spontaneous abortion epidemiology. *Statistics in Medicine*, 4:469–488, 1985.

[20] G. Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42:1897–1908, 1971.

[21] M. Magdon-Ismail, A.F. Atiya, A. Pratap, and Y.S. Abu-Mostafa. On the Maximum drawdown of a Brownian motion. *Journal of Applied Probability*, 41:147–161, 2004.

[22] G. V. Moustakides. Closed form expressions for the performance of cusum schemes. University of Thessaly, Greece, 2004.

[23] G.V. Moustakides. Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14:1379–1387, 1986.

[24] G.V. Moustakides. Optimality of the CUSUM procedure in continuous time. *Annals of Statistics*, 32:302–315, 2004.

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

[26] G. Peskir and A. N. Shiryaev. Solving the Poisson disorder problem. *Advances in finance and stochastics*, pages 295–312, 2002.

[27] M. Pollak and D. Siegmund. A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, 72:267–280, 1985.

[28] V.H. Poor. Quickest detection with exponential penalty for delay. *Annals of Statistics*, 26:2179–2205, 1998.

[29] G. Radaelli. Using the cuscore technique in the surveillance of rare health events. *Journal of Applied Statistics*, 19:75–81, 1992.

[30] S.W. Roberts. Control chart tests based on geometric moving average. *Technometrics*, 1:239–250, 1959.

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

[32] A. Segall, M. H. A. Davis, and T. Kailath. Nonlinear filtering with counting observations. *IEEE Transactions on Information Theory*, IT-21:143–149, 1975.

[33] W. Shewhart. *Economic Control of Quality of Manufactured Product.* Van Nostrand, Princeton, 1931.

[34] A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and its Applications*, 13:22–46, 1963.

[35] A. N. Shiryaev. Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time. *Russian Mathematical Surveys*, 51:750–751, 1996.

[36] D. Siegmund. *Sequential Analysis.* Springer-Verlag, 1st edition, 1985.

[37] D. Siegmund. The Wald memorial lectures. Boundary crossing probabilities and statistical applications. *Annals of Statistics*, 14:261–404, 1986.

[38] C. Sonesson. Evaluations of some exponentially weighted moving average methods. *Journal of Applied Statistics*, 30:1115–1133, 2003.

[39] C. Sonesson and D. Bock. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society, Series A*, 166:5–21, 2003.

[40] M.S. Srivastava and Y. Wu. Comparison of EWMA, CUSUM and Shirayaev-Roberts procedures for detecting a shift in the mean. *Annals of Statistics*, 21:645–670, 1993.

[41] A. Tartakovsky. Asymptotic properties of CUSUM and Shiryaev's procedures for detecting a change in non-homogeneous gaussian processes. *Mathematical Methods of Statistics*, 4:389–404, 1995.

[42] A.G. Tartakovsky. Asymptotically minimax multi-alternative sequential rule for disorder detection. *Statistics and Control of Random Processes: Proceedings of the Steklov Institute of Mathematics*, 202:229–236, 1994. American Mathematical Society, Providence, Rhode Island.

[43] H.M. Taylor. A stopped Brownian motion formula. *Annals of Probability*, 3:234–246, 1975.

[44] C.S. Van Dobben de Bruyn. *Cumulative Sum Tests*. Griffin, London, 1968.

[45] C. B. Wan and M. H. A. Davis. The general point process disorder problem. *IEEE Transactions on Information Theory*, pages 538–540, 1977.

[46] J.A.C. Weatherall and J.C. Haskey. Surveillance of malformations. *British Medical Bulletin*, 32:39–44, 1976.

[47] D.W. Wilson, K. Griffiths, K.W. Kemp, A.B.J. Nix, and R.J. Rowlands. Intrenal quality control of radioimmunoassays. *Journal of Endocrinology*, 80:365–372, 1979.

[48] W.H. Woodall. The distribution of the run length of one-sided CUSUM procedures for continuous random variables. *Technometrics*, 25:295–300, 1983.

[49] W.H. Woodall. On the Markov chain approach to the two-sided CUSUM. *Technometrics*, 26:41–46, 1984.

[50] B. Yakir. A lower bound on the ARL to detection of a change with a probability constraint of false alarm. *Annals of Statistics*, 24:431–435, 1996.

[51] E. Yashchin. On a unified approach to the analysis of two-sided cumulative sum control schemes with headstarts. *Advances in Applied Probability*, 17:562–593, 1985.

[52] S. Zacks. The probability distribution and the expected value of a stopping variable associated with one-sided CUSUM procedures for non-negative integer valued random variables. *Communications in Statistics A*, 10:2245–2258, 1981.