COMPOUND POISSON DISORDER PROBLEM WITH UNIFORMLY DISTRIBUTED DISORDER TIME

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

INDUSTRIAL ENGINEERING

By

Çağın Ürü

July 2019

Compound Poisson Disorder Problem with Uniformly Distributed Disorder Time
By Çağın Ürü
July 2019

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Savaş Dayanık(Advisor)

Semih Onur Sezer

Arnab Basu

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

COMPOUND POISSON DISORDER PROBLEM WITH UNIFORMLY DISTRIBUTED DISORDER TIME

Çağın Ürü
M.S. in Industrial Engineering
Advisor: Savaş Dayanık
July 2019

Suppose that arrival rate and jump distribution of a compound Poisson process change suddenly at an unknown and unobservable time. The problem of detecting the change (disorder) as soon as it occurs is known as compound Poisson disorder. In practice, an unfavorable regime shift may require immediate action, and a quickest detection rule can allow the decision maker to react to the change and take the necessary countermeasures in a timely manner.

Dayanık and Sezer [Compound Poisson disorder problem, Math. Oper. Res., vol. 31, no. 4, pp. 649-672, 2006] completely solve the compound Poisson disorder problem assuming a change-point with an exponential prior distribution. Although the exponential prior is convenient when solving the problem, it has flaws when expressing reality due to the memoryless property. Besides, as an informative prior, it fails to represent the case when the decision maker has no prior information on the change-point. Considering these defects, we assume a uniformly distributed change-point instead in our study. Unlike the exponential prior, the uniform prior has a memory and can be used when the decision maker does not have a strong belief on the change-point. We reformulate the quickest detection problem as a finite-horizon optimal stopping problem for a piecewise-deterministic and Markovian sufficient statistic. With Monte Carlo simulation and Chebyshev interpolation, we calculate the value function numerically via successive approximations. Studying the sample-paths of the sufficient statistic, we describe an explicit quickest detection rule and provide numerical examples for our solution method.

Keywords: Poisson disorder problem; compound Poisson process; quickest detection; optimal stopping.

ÖZET

TEKDÜZE DAĞILAN DÜZENSİZLİK ZAMANLI BİLEŞİK POISSON DÜZENSİZLİK PROBLEMİ

Çağın Ürü
Endüstri Mühendisliği, Yüksek Lisans
Tez Danışmanı: Savaş Dayanık
Temmuz 2019

Bir bileşik Poisson sürecinin hızının ve atlama dağılımının bilinmeyen ve gözlemlenemeyen bir zamanda aniden değiştiğini düşünelim. Bu değişimi (düzensizliği) meydana geldikten sonra en kısa sürede tespit etme problemine bileşik Poisson düzensizlik problemi denir. Birçok uygulamada, olumsuz bir rejim değişikliği acil eylem gerektirebilir ve en hızlı tespit kuralı karar vericinin değişime zamanında tepki vermesine ve gerekli önlemleri almasına izin verebilir.

Dayanık ve Sezer [Compound Poisson disorder problem, Math. Oper. Res., vol. 31, no. 4, pp. 649-672, 2006] bileşik Poisson düzensizlik problemlemini üssel ön dağılımlı bir değişim noktası varsayarak tamamen çözmüştür. Üssel dağılım, problemin çözümü için uygun olsa dahi hafızasız olma özelliği sebebiyle gerçekliği yansıtırken kusurludur. Ayrıca, bilgilendirici bir ön dağılım olduğu için karar vericinin değişim noktası hakkında bir ön bilgisinin olmadığı durumu temsil edemez. Bu eksiklikleri göz önünde bulundurarak, çalışmamızda üssel yerine tekdüze dağılan bir değişim noktası varsayıyoruz. Üssel dağılımdan farklı olarak, tekdüze dağılım bir hafızaya sahiptir ve karar vericinin değişim noktası hakkında güçlü bir öngörüsünün bulunmadığı durumda kullanılabilir. En hızlı tespit problemini parçalı-deterministik ve Markoviyen bir yeterli istatistik için sonlu bir en iyi durma problemine dönüştürüyoruz. Ardışık yaklaşımlar yoluyla, değer fonksiyonunu Monte Carlo simülasyonu ve Chebyshev enterpolasyonuyla sayısal olarak hesaplıyoruz. Yeterli istatistiğin örnek yollarını inceleyip en hızlı tespit kuralını tarif ediyoruz ve çözüm yöntemimiz için sayısal örnekler sunuyoruz.

Anahtar sözcükler: Poisson düzensizlik problemi; bileşik Poisson süreci; en hızlı tespit; en iyi durma.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my advisor, Prof. Savaş Dayanık, for his continuous guidance, patience and understanding. Without his supervision this thesis would not be possible.

I am indebted to Assoc. Prof. Semih Onur Sezer for guiding and supporting me at every step of the way during my undergraduate and graduate studies.

I am grateful to Prof. Hans Frenk for his infectious enthusiasm which nurtured my passion to continue my academic journey.

I would like to thank to Assoc. Prof. Arnab Basu for accepting to read and review this thesis.

I cannot fully express my gratitude to Nazlıcan Arslan, my parents Seval and Hasan Ürü, and my brother Umut Tanın Ürü for their endless love and support.

During my time at Bilkent and Sabancı, I have been privileged to have teachers and mentors who constantly encouraged and inspired me. It is a great honor to be one of their students. I would like to offer my thanks to these role models. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

Chapter 1

Preliminaries

Definition 1.1 (Sigma-algebra)**.** *Let Ω be a given set. A collection $\mathcal{F}$ of subsets of Ω is called a σ-algebra on Ω if it satisfies the following properties.*

(i) $\Omega \in \mathcal{F}$.

(ii) $\mathcal{F}$ *is closed under complements: if* $A \in \mathcal{F}$, *then* $A^c = \Omega \setminus A \in \mathcal{F}$.

(iii) $\mathcal{F}$ *is closed under countable unions: if* $A_1, A_2, \ldots \in \mathcal{F}$, *then* $\cup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

Definition 1.2 (Probability measure)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space. A set function $\mathbb{P} : \mathcal{F} \mapsto \mathbb{R}_+$ is called a probability measure on $(\Omega, \mathcal{F})$ if it satisfies the following properties.*

(i) $\mathbb{P}(\Omega) = 1$.

(ii) $\mathbb{P}$ *is countably additive: if* $A_1, A_2, \ldots \in \mathcal{F}$ *are disjoint sets, then*

$$\mathbb{P}(\cup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

Definition 1.3 (Probability space)**.** *A probability space is the triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.*

Definition 1.4 (Random variable)**.** *Let* $(\Omega, \mathcal{F})$ *and* $(E, \mathcal{E})$ *be measurable spaces. A mapping* $X : \Omega \mapsto E$ *is called a random variable taking values in* $(E, \mathcal{E})$ *if it is measurable with respect to* $\mathcal{F}$ *and* $\mathcal{E}$*, that is, if the pre-image*

$$X^{-1}A = \{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\}$$

is in $\mathcal{F}$ *for every* A *in* $\mathcal{E}$*.*

Definition 1.5 (Stochastic process)**.** *Let* $(E, \mathcal{E})$ *be a measurable space and* T *be an arbitrary set. For each* $t \in T$*, let* X_t *be a random variable taking values in* $(E, \mathcal{E})$*. Then the collection of random variables* $X = \{X_t; t \in T\}$ *is called a stochastic process with state space* $(E, \mathcal{E})$ *and index set* T*.*

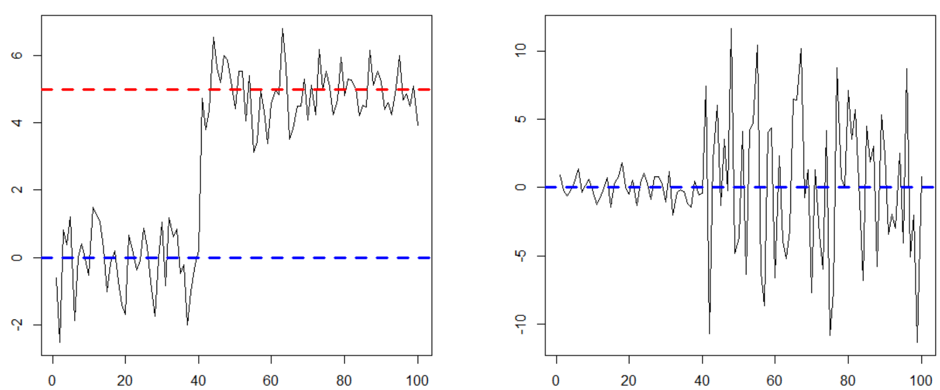
Definition 1.6 (Filtration)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* T *be a subset of* $\mathbb{R}_+$*. The family of* σ*-algebras* $\mathcal{G} = \{\mathcal{G}_t; t \in T\}$ *is called a filtration on* T *if* $\mathcal{G}_t \subset \mathcal{F}$ *for every* $t \in T$ *and* $\mathcal{G}_s \subset \mathcal{G}_t$ *for* $s < t \in T$*. In other words, a filtration is an increasing family of sub-*σ*-algebras.*

Definition 1.7 (Stopping time)**.** *Let* $\mathcal{G}$ *be a filtration on* T*. A random time* $\tau : \Omega \mapsto T \cup \{+\infty\}$ *is called a stopping time of* $\mathcal{G}$ *if* $\{\tau \leq t\} \in \mathcal{G}_t$ *for each* $t \in T$*.*

Chapter 2

Introduction and literature review

Illustrated with two examples in Figure 2.1, a sudden change in the probability distribution of a stochastic process or the statistical behavior of a time series has drawn considerable interest among decision makers for many years. The classical problem of detecting this change is known as statistical change-detection, change-point detection, or disorder detection.



(a) Change in mean (b) Change in variance

Figure 2.1: Change in the statistical behavior of a sequence of data points

With various applications in numerous contexts, the change-detection problems have been widely studied in the statistical analysis literature. Dating back to 1930s, the first studies in the change-detection literature are on the quality control applications in manufacturing processes; see Basseville and Nikiforov [1] and Poor and Hadjiliadis [2]. Since then, change-detection problems have been addressed in a variety of fields, and the application areas grew immensely. Today, those application areas include finance, healthcare, network and security systems, and many others.

Many applications in areas such as text or image analysis depend on the off-line methods to detect a change. Employing methods such as maximum likelihood estimation, the off-line approaches aim to identify the occurrence time of a change-point during a past time frame using the recorded observations. Hence, they cannot react to a change in real-time, and cannot be used to analyze streaming data. However, many real-world applications require the detection of a change-point in real-time. While addressing this problem known as the quickest detection, the on-line detection techniques seek to identify an alarm time as close as possible to the disorder time using the revealed information obtained from the gathered observations. While the off-line methods deal with "when has the change happened?" the on-line methods consider "when is the change going to happen?"

A sudden change in the statistical behavior of a system can be catastrophic, and the identification of a quickest detection rule may help alleviating some of the detrimental effects of the change by allowing decision makers to react and take the necessary countermeasures in a timely manner. For instance, monitoring the arrival process of seismic waves and declaring an emergency when there is a change in their behavior may help with the efforts to foresee an emerging earthquake, and could potentially save lives and minimize the damage. Likewise, observing the arrival process of patients to medical centers can contribute to the disease outbreak investigations greatly. Therefore, detecting the change-point as soon as it occurs is a crucial task in many applications.

2.1 Literature review

The Bayesian formulation of the quickest detection problem dates back to the works of Kolmogorov and Shiryaev in the early 1960s; see Poor and Hadjiliadis [2]. Assuming some prior knowledge on the probability law governing the change-point, the Bayesian formulation aims to minimize a cost function, known as the Bayes risk, consisting of the probability of falsely detecting a change and the average detection delay cost. The change-point is assumed to have a geometric and exponential prior distribution in discrete-time and continuous-time models, respectively. In those Bayesian models, the primary objects of study are the sequences of random variables and their continuous time analogs. The solution methods to those problems depend on the reformulation to an optimal stopping problem and reduction to a free-boundary problem. When solving the reformulated optimal stopping problem, both martingale and Markovian approaches are studied; see Peskir and Shiryaev [3].

We refer the reader to the monographs of Poor and Hadjiliadis [2], Basseville and Nikiforov [1], and Peskir and Shiryaev [3] for a comprehensive review of the quickest detection problems.

2.1.1 Poisson disorder problem

A quickest detection problem for a Poisson process was formulated by Galchuk and Rozovskii [4] for the first time. In their formulation, the intensity of a Poisson process changes suddenly from a known constant λ_0 to another known constant λ_1 at some unknown time θ. The prior distribution of θ is assumed to be conditionally exponential with parameter λ given that $\theta > 0$. That is,

$$\mathbb{P}\{\theta = 0\} = \pi \quad \text{and} \quad \mathbb{P}\{\theta > t\} = (1-\pi)e^{-\lambda t}, \quad t \geq 0,\ \pi \in [0,1),\ \lambda > 0.$$

In this simple Poisson disorder problem, the objective is to declare an alarm as close as possible to the change-time θ while observing the Poisson process. Hence, authors looked for a stopping time τ of the observed point process that minimizes the Bayes risk

$$R_\tau^1(\pi) = \mathbb{P}\{\tau < \theta\} + c\mathbb{E}[(\tau - \theta)^+],$$

consisting of the false alarm probability $\mathbb{P}\{\tau < \theta\}$, the expected detection delay $\mathbb{E}[(\tau - \theta)^+]$ and the positive cost parameter c. Assuming $\lambda + c \geq \lambda_1 > \lambda_0$, they provided a partial solution to the problem.

Later, Davis [5] extended those results to a more general case when $\lambda + c \geq \lambda_1 - \lambda_0 \geq 0$. As another notable contribution, Davis [5] showed that the choice of the Bayes risk does not affect the optimal quickest detection rule. In fact, he proposed the different measures of the Bayes risk

$$\begin{aligned} R_\tau^2(\pi) &= \mathbb{P}\{\tau < \theta - \varepsilon\} + c\mathbb{E}[(\tau - \theta)^+], \quad \varepsilon > 0, \\ R_\tau^3(\pi) &= \mathbb{E}[(\theta - \tau)^+] + c\mathbb{E}[(\tau - \theta)^+], \end{aligned}$$

are special cases of a "standard" cost function

$$R_\tau^S(\pi) = a + b\int_0^\tau (\Pi_s - k)ds, \qquad a \in \mathbb{R}, \quad b \in \mathbb{R}_+, \quad k \in [0, 1],$$

where $\Pi_t \triangleq \mathbb{P}\{\theta \leq t | \mathcal{F}_t\}$ is the posterior probability process and $\mathcal{F} = \{\mathcal{F}_t; t \geq 0\}$ is the filtration generated by the Poisson process. Observe that when $\varepsilon = 0$, we simply have the Bayes risk of Galchuk and Rozovskii [4].

Three decades after its first formulation, the simple Poisson disorder problem was completely solved by Peskir and Shiryaev [6]. In their solution, authors reformulated the problem as an optimal stopping problem with the value function

$$V(\pi) = \inf_\tau \left(\mathbb{P}\{\tau < \theta\} + c\mathbb{E}(\tau - \theta)^+\right)$$

and reduced the optimal stopping problem to a free-boundary problem.

Considering the applications in finance, Bayraktar and Dayanik [7] solved the simple Poisson disorder problem with the Bayes risk

$$R_\tau^4(\pi) = \mathbb{P}\{\tau < \theta\} + c\mathbb{E}[e^{\alpha(\tau - \theta)^+} - 1], \qquad \alpha > 0,$$

by assuming an exponential penalty for the detection delay. Later, Bayraktar et al. [8] revisited the problem, and showed that the Bayes risk $R_\tau^4(\pi)$ is also a "standard" Bayes risk. They solved the problem using the odds-ratio process $\Phi_t \triangleq \Pi_t/(1 - \Pi_t)$ defined by the posterior probability process Π.

2.1.2 Compound Poisson disorder problem

While the earlier studies in the literature have worked on the simple Poisson disorder problem, the compound Poisson disorder problem remained to be addressed due to its difficulty. With the first attempt to solve the problem of detecting a change in the probability law of a compound Poisson process, Gapeev [9] gave a partial solution to the compound Poisson disorder problem under a very specific scenario where the marks were exponentially distributed and their expectations were equal to the arrival intensities before and after the disorder.

Dayanik and Sezer [10] completely solved the compound Poisson disorder problem by adapting a method of Gugerli [11] and Davis [12], and using the optimal stopping theory for the piecewise-deterministic Markov processes. Unlike the earlier analytical methods, their probabilistic solution involves studying the sample-paths of the odds-ratio process Φ defined previously. This approach also forms the basis of our solution technique in this study.

2.1.3 Prior distribution of the disorder time

Consider a sequence of independent experiments at the deterministic times $\sigma, 2\sigma, 3\sigma, \ldots$, for $\sigma > 0$. Suppose each experiment is a success with probability $\lambda\sigma$. At the time of the first success, a change occurs. Hence, the change-point has a geometric prior distribution. This was the setup of the first disorder detection problem formulated by Kolmogorov and Shiryaev in 1960s.

As $\sigma \to 0^+$, we have $\mathbb{P}\{\text{time of the first success} > t\} \to e^{-\lambda t}$, and we reach to the asymptotic approximation of exponential distribution with geometric distribution. Therefore, when addressing the continuous-time equivalent of the disorder detection problem, the later studies replaced the geometric prior with the exponential one. In fact, the exponential prior soon became the common assumption for the Bayesian change-point in the change-detection literature and the Poisson disorder problem in particular.

Exponential distribution was a reasonable choice for the prior of the change-point when modeling and solving a quickest detection problem. For example, the lifetime of machines or parts are often modeled to have an exponential distribution in reliability theory. One can imagine that the end of life of those components would be a potential change-point in the system. Thus, exponentially distributed change-point makes sense when modeling a quickest detection problem. More importantly, having an exponential prior proves useful when obtaining explicit results with several solution methods. For instance, the sufficient statistics such as the conditional probability process or the odds-ratio process mentioned above are shown to be one-dimensional when the change-point is exponentially distributed; see Sezer [13]. Due to those analytical advantages, exponential distribution remained as the common assumption for the change-point's prior in the literature.

Even though the exponential prior seems convenient, it has a serious flaw, the memorylessness. This well-known property of the exponential distribution can be summarized as the independence of the probability distribution from history. Let us describe why this constitutes a problem with a concrete example. In the previous paragraph, we mentioned how the end of lifetime of some component would be a potential change-point in a system. Suppose the lifetime of that component is exponentially distributed. Now, imagine you have already observed the system for some time and the component is yet to fail. At this point in time, the memoryless property sets the clock back to time 0 regardless of the elapsed time and takes that aged component as if it is brand new. Although an aged component is often more likely to fail, the memoryless property overlooks this fact. Therefore, the exponential prior has defects when representing the reality.

Considering the limitations, some studies replaced the exponential prior at the expense of losing analytical tractability. Most notably, Bayraktar and Sezer [14] took the prior distribution of the disorder time to be a phase-type distribution, which is the distribution of the absorption time of a continuous time Markov chain with a finite state space. However, the sufficient statistics are no longer one-dimensional in that case. Although explicit results are difficult to obtain, asymptotically optimal solutions can be reached. We refer the reader to Baron and Tartakovsky [15] for an overview of the works on asymptotical detection.

2.2 Introduction

In this study we revisit the compound Poisson disorder problem in which arrival rate and jump distribution of a compound Poisson process change at an unknown and unobservable time θ. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space hosting a such compound Poisson process

$$X_t = X_0 + \sum_{k=1}^{N_t} Y_k, \quad t \geq 0. \tag{2.1}$$

where $N = \{N_t;\, t \geq 0\}$ is standard Poisson process with rate $\lambda_0 > 0$, and $Y_1, Y_2, \ldots$ are independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random variables with a common distribution $\nu_0(\cdot)$ independent of the process N.

Suppose at time θ, the arrival rate λ_0 and the jump distribution ν_0 shift to λ_1 and ν_1, respectively. Assuming those arrival rates and jump distributions are known, we would like to identify the change-point θ as soon as it occurs.

In practice, decision makers often neither have a rational assessment nor hold any strong belief on the change-point due to lack of any prior knowledge. Imagine a day-trader. She is buying and selling financial assets during a trading day. Suppose a change is expected to happen at some random time the next day, and she absolutely has no idea about its timing. This might be because she does not have any experience in the job or any financial data to extract information from. Alternatively, perhaps the market fluctuates a lot; so, any outcome is equally likely to occur. With no available prior information on the change-point, what would be a reasonable approach to model it?

The absence of information in some applications suggests the use of an uninformative prior for the distribution of the disorder time. As opposed to an informative prior such as the exponential distribution, an uninformative prior does not carry any definite information about the variable of interest as its name implies. Moreover, an uninformative prior is invariant to parametrization. That is, it expresses no information even when the variable is parametrized in another way.

When modeling the scenario where the decision maker has no prior information on the change-point, uniform distribution is the first choice that comes to mind for the change-point's prior. It is reasonable to assign uniform distribution to a random variable when there is no information but the range of its possible values. Uniform distribution is simple and puts equal weights to every possible outcome. Thus, it can successfully express the notion of having no information and would be a good fit for the the change-point's prior. Furthermore, it has a memory unlike the exponential prior. Hence, it fits better to some real cases. Note that the uniform prior is not invariant under reparametrization. Therefore, it is useful for modeling an uninformative change-time, but has its own defects.

For those reasons, while the earlier work in the literature assume an exponentially distributed change-point, we consider a uniformly distributed one. That is, we assume that the disorder time θ is a uniformly distributed random variable taking values in the interval $[0, T]$ with the following zero-modified prior distribution.

$$\mathbb{P}\{\theta = 0\} = \pi \quad \text{and} \quad \mathbb{P}\{\theta > t\} = (1-\pi)\left(\frac{T-t}{T}\right), \tag{2.2}$$

where $0 \leq t \leq T$ and $\pi \in [0, 1)$ for some known fixed $T \in \mathbb{R}_+$. Interested in a quickest detection rule, we would like to find a $[0, T]$-valued stopping time τ adapted to the history $\mathcal{F}$ of the compound Poisson process X whose Bayes risk

$$R_\tau(T, \pi) \triangleq \mathbb{P}\{\tau < \theta\} + c\mathbb{E}(\tau - \theta)^+, \quad \pi \in [0, 1),\ \tau \in \mathcal{F}, \tag{2.3}$$

which consists of the false-alarm frequency $\mathbb{P}\{\tau < \theta\}$ and the expected detection delay cost $c\mathbb{E}(\tau - \theta)^+$, is the smallest. If an $\mathcal{F}$-stopping time τ attains the minimum Bayes risk

$$V(T, \pi) \triangleq \inf_{\tau \in \mathcal{S}_{[0,T]}} R_\tau(T, \pi), \quad \pi \in [0, 1), \tag{2.4}$$

where $\mathcal{S}_{[0,T]}$ is the collection of all $\mathcal{F}$-adapted $[0, T]$-valued stopping times, then it is called a Bayes-optimal alarm time.

In this study, we adopt the solution method presented in Dayanik and Sezer [10]. We study the sample-path behavior of a one-dimensional sufficient statistic, which is piece-wise deterministic and Markovian, approximate the Bayes risk successively, and implement a numerical algorithm for the solution of the problem.

The remainder of this thesis is organized as follows. In Chapter 3, we present the model containing the random elements of our problem, describe the quickest detection problem and reformulate it to an optimal stopping problem for a suitable Markovian sufficient statistic. In Chapter 4, we approach the optimal stopping problem with the method of variational inequalities and partition the time-space domain of the sufficient statistic into continuation and stopping regions. Due to the free-boundary separating those regions and the jumping sufficient statistic, solving the optimal stopping problem analytically turns out to be difficult. Therefore, in Chapter 5, we introduce the successive approximations of the value function of the optimal stopping problem and derive some useful results for our alternative solution method. In Chapter 6, we suggest a schema to evaluate the value function numerically with Monte Carlo simulation and Chebyshev interpolation. We characterize an explicit quickest detection rule and outline its on-line implementation. In Chapter 7, we present numerical examples and analyze the effects of different problem parameters on the distribution of the alarm time. Chapter 8 mentions some interesting findings, provides future research directions and concludes the thesis. Lengthy derivations and long proofs are moved to Appendix.

Chapter 3

Model and problem description

In this chapter, we construct a model containing the random elements of our problem. Then we describe our quickest detection problem and reformulate it as an optimal stopping problem for a piece-wise deterministic and Markovian sufficient statistic Q defined in (3.9).

3.1 Model

Let $(\Omega, \mathcal{F}, \mathbb{P}_0)$ be a probability space hosting the following independent stochastic elements:

(i) a standard Poisson process $N = \{N_t;\, t \geq 0\}$ with the arrival rate λ_0,

(ii) i.i.d. $\mathbb{R}^d$-valued random variables $Y_1, Y_2, \ldots$ with a common distribution $\nu_0(B) \triangleq \mathbb{P}_0\{Y_1 \in B\}$ for every Borel set B in the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ and $\nu_0(\{0\}) = 0$,

(iii) a random variable θ with the distribution

$$\mathbb{P}_0\{\theta = 0\} = \pi \quad \text{and} \quad \mathbb{P}_0\{\theta > t | \theta > 0\} = 1 - \frac{t}{T} \tag{3.1}$$

for $0 \leq t \leq T,\ \pi \in [0, 1),\ T \in \mathbb{R}_+$.

Let $X = \{X_t;\, t \geq 0\}$ defined by (2.1) be a compound Poisson process with the arrival rate λ_0, jump distribution $\nu_0(\cdot)$, and the jump times

$$\sigma_n \triangleq \inf\{t > \sigma_{n-1} : X_t \neq X_{t-}\}, \quad n \geq 1 \quad (\sigma_0 \equiv 0). \tag{3.2}$$

Let us denote the augmentation of its natural filtration $\sigma(X_s,\, s \leq t)$, $t \geq 0$ with $\mathbb{P}_0$-null sets by $\mathcal{F} = \{\mathcal{F}_t; t \geq 0\}$. We will describe the enlargement of this filtration $\mathcal{F}$ with the sigma-algebra $\sigma(\theta)$ generated by θ with $\mathcal{G} = \{\mathcal{G}_t; t \geq 0\}$. That is, $\mathcal{G}_t \triangleq \mathcal{F}_t \vee \sigma(\theta)$ for every $t \geq 0$.

Let $\lambda_1 > 0$ be a constant, and $\nu_1(\cdot)$ be a probability measure on the measurable space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ absolutely continuous with respect to the distribution $\nu_0(\cdot)$. Radon-Nikodym derivative

$$f(y) \triangleq \frac{d\nu_1}{d\nu_0}\Big|_{\mathcal{B}(\mathbb{R}^d)}(y), \quad y \in \mathbb{R}^d \tag{3.3}$$

of $\nu_1(\cdot)$ with respect to $\nu_0(\cdot)$ exists and is a $\nu_0 - a.e.$ nonnegative Borel function.

We describe a new probability measure $\mathbb{P}$ on the measurable space $(\Omega, \vee_{s\geq 0}\mathcal{G}_s)$ by a change of measure with Radon-Nikodym derivative Z_t of $\mathbb{P}$ with respect to $\mathbb{P}_0$ as follows.

$$\frac{d\mathbb{P}}{d\mathbb{P}_0}\Big|_{\mathcal{G}_t} = Z_t \triangleq 1_{\{t<\theta\}} + 1_{\{t\geq\theta\}} e^{-(\lambda_1-\lambda_0)(t-\theta)} \prod_{k=N_{\theta-}+1}^{N_t} \left[\frac{\lambda_1}{\lambda_0} f(Y_k)\right], \quad t \geq 0, \tag{3.4}$$

where $N_{\theta-}$ is the number of arrivals in the time interval $[0, \theta)$. Note that Z_t can also be written as

$$Z_t \triangleq 1_{\{t<\theta\}} + 1_{\{t\geq\theta\}} \frac{R_t}{R_\theta}, \quad t \geq 0 \tag{3.5}$$

in terms of the likelihood ratio process $R = \{R_t; t \geq 0\}$ given by

$$R_t \triangleq e^{-(\lambda_1-\lambda_0)t} \prod_{k=1}^{N_t} \left[\frac{\lambda_1}{\lambda_0} f(Y_k)\right], \quad t \geq 0. \tag{3.6}$$

Probability measures $\mathbb{P}_0$ and $\mathbb{P}$ are called the reference and physical probability measures, respectively. Observe that since $Z_0 = 1$ $\mathbb{P}_0$-a.s. and the probability measures $\mathbb{P}$ and $\mathbb{P}_0$ agree on $\mathcal{G}_0 = \sigma(\theta)$, the disorder time θ and the compound Poisson process X have the same properties under both probability measures. Calculations which are difficult under $\mathbb{P}$ become easier under $\mathbb{P}_0$ due to the independence of the previously defined stochastic elements.

3.2 Problem description

In the compound Poisson disorder problem, arrival rate and jump distribution of a compound Poisson process X change at an unknown and unobservable disorder time θ. The problem is to find an $\mathcal{F}$-stopping time τ that minimizes the Bayes risk in (2.3) and detect the disorder time θ as soon as possible while observing the process X. In this section, we will reformulate this quickest detection problem as an optimal stopping problem for a suitable Markov process.

As shown in Appendix A.1, we can rewrite the Bayes risk in (2.3) as

$$R_\tau(T,\pi) = 1-\pi+\frac{c(1-\pi)}{T}\,\mathbb{E}_0\left[\int_0^\tau \left((T-t)\Phi_t-\frac{1}{c}\right)dt\right],$$

where $\pi \in [0,1)$ and $\tau \in \mathcal{F}$, in terms of the odds-ratio process Φ defined by

$$\Phi_t = \frac{\Pi_t}{1-\Pi_t} = \frac{\mathbb{P}\{\theta \le t|\mathcal{F}_t\}}{\mathbb{P}\{\theta > t|\mathcal{F}_t\}}, \quad t \ge 0 \tag{3.7}$$

in terms of the posterior probability process Π.

Let us define the deterministic process $D = \{D_t; t \ge 0\}$ by

$$D_t \triangleq \frac{1}{T-t}. \tag{3.8}$$

Observe that the random variable D_t is the multiplicative inverse of the remaining time until time T at time t and may be viewed as a discounting factor for the process Q. When we replace $T-t$ with $\frac{1}{D_t}$, the Bayes risk becomes

$$R_\tau(T,\pi) = 1-\pi+\frac{c(1-\pi)}{T}\,\mathbb{E}_0\left[\int_0^\tau \left(\frac{\Phi_t}{D_t}-\frac{1}{c}\right)dt\right].$$

Let us define

$$Q_t \triangleq \frac{\Phi_t}{D_t} \tag{3.9}$$

in terms of the odd-ratio process Φ and the deterministic process D. In Appendix A.2, we present the dynamics of the process Q. As it turns out, the process Q belongs to the family of piece-wise deterministic Markov processes.

If we define

$$x(q,t) \triangleq \begin{cases} \frac{1}{\lambda_1-\lambda_0} + \left(q - \frac{1}{\lambda_1-\lambda_0}\right) e^{-(\lambda_1-\lambda_0)t}, & \lambda_1 \neq \lambda_0, \\ q+t, & \lambda_1 = \lambda_0, \end{cases} \tag{3.10}$$

for $t \in \mathbb{R}_+$ and $q \in \mathbb{R}_+$, then for $n \geq 0$ we have

$$Q_t = \begin{cases} x(Q_{\sigma_n}, t-\sigma_n), & \sigma_n \leq t < \sigma_{n+1}, \\ x(Q_{\sigma_n}, \sigma_{n+1}-\sigma_n)\frac{\lambda_1}{\lambda_0} f(Y_{n+1}), & t = \sigma_{n+1}. \end{cases} \tag{3.11}$$

This means that the process Q follows the deterministic curves $t \mapsto x(q,t)$ in (3.10) between two consecutive jumps of the process X and jumps instantly at the jump times of the process X as described in (3.11).

Finally, substituting the process Q in the Bayes risk gives

$$R_\tau(T,\pi) = 1-\pi+\frac{c(1-\pi)}{T}\,\mathbb{E}_0\left[\int_0^\tau \left(Q_t - \frac{1}{c}\right) dt\right]. \tag{3.12}$$

Hence, the minimum Bayes risk in (2.4) is given by

$$V(T,\pi) = 1-\pi+\frac{c(1-\pi)}{T}\,U\left(T, T\frac{\pi}{1-\pi}\right), \quad T \in \mathbb{R}_+,\ \pi \in [0,1)$$

in terms of the value function

$$U(T,q) \triangleq \inf_{\tau \in \mathcal{S}_{[0,T]}} \mathbb{E}_{0,q}\left[\int_0^\tau g(Q_t)dt\right], \quad T \in \mathbb{R}_+,\ q = Q_0 \in \mathbb{R}_+ \tag{3.13}$$

of an optimal stopping problem with the running cost

$$g(q) \triangleq q - \frac{1}{c}, \quad q \in \mathbb{R}_+. \tag{3.14}$$

Hence, the quickest detection problem for a compound Poisson process can be reformulated as the finite-horizon optimal stopping problem of the process Q. In the next chapter, we will approach this optimal stopping problem with the method of variational inequalities.

Chapter 4

Variational inequalities

In this chapter, we derive and present the variational inequalities associated with the optimal quickest detection rule and look for a solution to our optimal stopping problem with the value function U in (3.13).

Note that if we have not stopped and raised an alarm yet, it is optimal either to stop immediately or wait for infinitesimally short h units of time and act optimally afterwards. Both decisions suggest a variational inequality that our value function U must satisfy. In Appendix B, we derive those variational inequalities in detail. The results can be summarized as follows.

1) Immediately stop:

$$-U(T, q) \geq 0,$$

2) Wait h units of time and act optimally afterwards:

$$g(q) + [1 - (\lambda_1 - \lambda_0)q]U_2(T, q) - U_1(T, q) + (DU)(T, q) \geq 0,$$

where g is the cost function in (3.14), $U_1(\cdot, \cdot)$ and $U_2(\cdot, \cdot)$ are the partial derivatives of $U(\cdot, \cdot)$ with respect to first and second arguments, and $(DU)(T, q)$ is the jump difference operator defined by

$$(DU)(T, q) \triangleq \int_{\mathbb{R}^d} \left[U\left(T, q\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T, q) \right] \lambda_0 \nu_0(dy).$$

As one of the actions must be optimal, the following must hold

$$\min\{-U(T,q),\, g(q) + [1-(\lambda_1-\lambda_0)q]U_2(T,q) - U_1(T,q) + (DU)(T,q)\} = 0. \tag{4.1}$$

Time-space domain, namely $\mathbb{R}_+ \times \mathbb{R}_+$, can be partitioned into a continuation region

$$C \triangleq \{(T,q) \in \mathbb{R}_+ \times \mathbb{R}_+ : U(T,q) < 0\}$$

and a stopping region

$$\Gamma \triangleq \{(T,q) \in \mathbb{R}_+ \times \mathbb{R}_+ : U(T,q) = 0\}.$$

Value function U is a solution of the integro-partial-differential equations

$$\left\{\begin{aligned} U(T,q) < 0 \\ g(q) + [1-(\lambda_1-\lambda_0)q]U_2(T,q) - U_1(T,q) + (DU)(T,q) = 0 \end{aligned}\right\},\ (T,q) \in C,$$
$$\left\{\begin{aligned} U(T,q) = 0 \\ g(q) + [1-(\lambda_1-\lambda_0)q]U_2(T,q) - U_1(T,q) + (DU)(T,q) > 0 \end{aligned}\right\},\ (T,q) \in \Gamma.$$

We can describe the quickest detection rule with the minimum Bayes risk in (2.4) among all of the stopping times of $\mathcal{S}_{[0,T]}$ as the first crossing time $\inf\{t \geq 0 : Q_t \in \Gamma\}$ of Q to the stopping region Γ. Then the solution to our optimal stopping problem requires simultaneously determining the value function U, the continuation region C, and the stopping region Γ. However, determining those analytically is a rather difficult task due to the jumping sufficient statistic Q and the free-boundary that separates the mentioned regions. Therefore, we will calculate the value function numerically with successive approximations and provide an approximate solution to the optimal stopping problem in the following chapters.

Chapter 5

Successive approximations

Let us define the family of optimal stopping problems

$$U_n(T,q) \triangleq \inf_{\tau \in \mathcal{S}_{[0,T]}} \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_n}\left(Q_t - \frac{1}{c}\right)dt\right], \quad T \in \mathbb{R}_+,\ q \in \mathbb{R}_+,\ n \in \mathbb{N}, \quad (5.1)$$

obtained from stopping the process Q at the nth jump time σ_n of the observation process X. Since the integrand in (5.1) is bounded from below by $-\frac{1}{c}$, the expectation in (5.1) is well defined for every stopping time $\tau \in \mathcal{S}_{[0,T]}$.

Note that stopping the process Q at time 0, i.e. taking $\tau = 0$, results in $U_n(T,q) = 0$. Hence, 0 is an upper bound for the $(U_n(T,q))_{n\geq 1}$ sequence. Since $Q_t \geq 0$ for all $t \in \mathbb{R}_+$ and $\tau \leq T$ almost surely,

$$\begin{aligned}\int_0^{\tau\wedge\sigma_n}\left(Q_t - \frac{1}{c}\right)dt &\geq -\int_0^{\tau\wedge\sigma_n}\frac{1}{c}dt\\ &\geq -\int_0^{\tau}\frac{1}{c}dt\\ &\geq -\int_0^{T}\frac{1}{c}dt\\ &= -\frac{T}{c}.\end{aligned}$$

Taking the expectation and infimum on both sides over $\tau \in \mathcal{S}_{[0,T]}$ shows $-\frac{T}{c}$ is a lower bound for the $(U_n(T,q))_{n\geq 1}$ sequence. Hence, $-\frac{T}{c} \leq U_n(T,q) \leq 0$ for all $n \in \mathbb{N}$.

The sequence $(U_n(T,q))_{n\geq 1}$ is decreasing because the sequence $(\sigma_n)_{n\geq 1}$ of jump times of the process X is increasing almost surely. Therefore, $\lim_{n\to\infty} U_n(T,q)$ exists everywhere. It can be seen that $U_n(T,q) \geq U(T,q)$ for all $n \in \mathbb{N}$.

Proposition 5.1. *For every $T_{max} \in \mathbb{R}_+$, as $n \to \infty$ the sequence $U_n(T,q)$ converges to $U(T,q)$ uniformly in $T \in [0, T_{max}]$, $q \in \mathbb{R}_+$. In fact, for every $n \in \mathbb{N}$, $T \in [0, T_{max}]$ and $q \in \mathbb{R}_+$, we have*

$$-\frac{T}{c}\left(1 - e^{-\lambda_0 T}\right)^n \leq U(T,q) - U_n(T,q) \leq 0. \tag{5.2}$$

Before proceeding with the proof let us define

$$\|f\|_{T_{\max},\infty} \triangleq \sup_{(T,q)\in[0,T_{\max}]\times\mathbb{R}_+} |f(T,q)|, \quad \forall T_{\max} \in \mathbb{R}_+,$$

for a function $f : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}_-$.

Proof. Because $Q_t \geq 0$ for all $t \in \mathbb{R}_+$ and $\tau \leq T$ almost surely, we have

$$\begin{aligned}
&\mathbb{E}_{0,q}\left[\int_0^\tau \left(Q_t - \frac{1}{c}\right) dt\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_n} \left(Q_t - \frac{1}{c}\right) dt + 1_{\{\tau>\sigma_n\}} \int_{\sigma_n}^\tau \left(Q_t - \frac{1}{c}\right) dt\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_n} \left(Q_t - \frac{1}{c}\right) dt\right] + \mathbb{E}_{0,q}\left[1_{\{\tau>\sigma_n\}} \int_{\sigma_n}^\tau \left(Q_t - \frac{1}{c}\right) dt\right] \\
&\geq U_n(T,q) + \mathbb{E}_{0,q}\left[1_{\{\tau>\sigma_n\}} \int_{\sigma_n}^\tau \left(Q_t - \frac{1}{c}\right) dt\right] \\
&\geq U_n(T,q) - \frac{1}{c}\mathbb{E}_0\left[1_{\{\tau>\sigma_n\}}(\tau - \sigma_n)\right] \\
&\geq U_n(T,q) - \frac{1}{c}\mathbb{E}_0\left[1_{\{T>\sigma_n\}}(T - \sigma_n)\right] \\
&\geq U_n(T,q) - \frac{T}{c}\mathbb{P}_0\{\sigma_n \leq T\}.
\end{aligned}$$

Since the nth jump time σ_n of the process X is the sum of n i.i.d. interarrival times, denoted by the sequence $(\widetilde{\sigma_n})_{n\geq 1}$, having exponential distribution with common parameter λ_0 under $\mathbb{P}_0$, we have

$$\begin{aligned}\mathbb{P}_0\{\sigma_n \leq T\} &= \mathbb{P}_0\{\widetilde{\sigma_1} + .. + \widetilde{\sigma_n} \leq T\} \\ &\leq \mathbb{P}_0\{\widetilde{\sigma_1} \leq T, .., \widetilde{\sigma_n} \leq T\} \\ &= (\mathbb{P}_0\{\widetilde{\sigma_1} \leq T\})^n \\ &= \left(1 - e^{-\lambda_0 T}\right)^n .\end{aligned}$$

Hence,

$$\mathbb{E}_{0,q}\left[\int_0^\tau \left(Q_t - \frac{1}{c}\right) dt\right] \geq U_n(T,q) - \frac{T}{c}\left(1 - e^{-\lambda_0 T}\right)^n .$$

Taking the infimum of both sides over $\tau \in \mathcal{S}_{[0,T]}$ gives the first inequality in (5.2).

Thus,

$$U_n(T,q) \geq U(T,q) \geq U_n(T,q) - \frac{T}{c}\left(1 - e^{-\lambda_0 T}\right)^n \tag{5.3}$$

holds for $n \in \mathbb{N}$. As a result, we have

$$\lim_{n\to\infty} \|U_n(T,q) - U(T,q)\|_{T_{\max},\infty} = 0, \quad \forall\, T_{\max} \in \mathbb{R}_+,$$

and the sequence $(U_n(T,q))_{n\geq 1}$ converges to $U(T,q)$ uniformly in $T \in [0, T_{\max}]$, $q \in \mathbb{R}_+$. □

Let us define the following operators

$$(J_t w)(T,q) \triangleq \mathbb{E}_{0,q}\left[\int_0^{t\wedge\sigma_1} \left(Q_s - \frac{1}{c}\right) ds + 1_{\{t\geq\sigma_1\}} w(T - \sigma_1, Q_{\sigma_1})\right], \quad t \in [0,T] \tag{5.4}$$

$$(Jw)(T,q) \triangleq \inf_{t\in[0,T]} (J_t w)(T,q) \tag{5.5}$$

acting on bounded Borel functions $w : [0,T] \times \mathbb{R}_+ \mapsto \mathbb{R}_-$ for $T \in \mathbb{R}_+,\ q \in \mathbb{R}_+$.

Note that σ_1 is exponentially distributed with parameter λ_0 under $\mathbb{P}_0$. Hence, using the Fubini theorem, for $t \in [0,T]$ we can write

$$\begin{aligned}(J_t w)(T,q) &= \mathbb{E}_{0,q}\left[\int_0^t 1_{\{s<\sigma_1\}} \left(Q_s - \frac{1}{c}\right) ds + 1_{\{t\geq\sigma_1\}} w(T - \sigma_1, Q_{\sigma_1})\right] \\ &= \mathbb{E}_{0,q}\left[\int_0^t 1_{\{s<\sigma_1\}} \left(x(q,s) - \frac{1}{c}\right) ds + 1_{\{t\geq\sigma_1\}} w\left(T - \sigma_1, x(q,\sigma_1)\frac{\lambda_1}{\lambda_0} f(Y_1)\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= \int_0^t \mathbb{E}_{0,q}\left[1_{\{s<\sigma_1\}}\left(x(q,s)-\frac{1}{c}\right)\right]ds \\
&\quad + \mathbb{E}_{0,q}\left[1_{\{t\geq\sigma_1\}} w\left(T-\sigma_1, x(q,\sigma_1)\frac{\lambda_1}{\lambda_0}f(Y_1)\right)\right] \\
&= \int_0^t \mathbb{P}_0\{s<\sigma_1\}\mathbb{E}_{0,q}\left[x(q,s)-\frac{1}{c}\middle| s<\sigma_1\right]ds \\
&\quad + \mathbb{E}_{0,q}\left[1_{\{t\geq\sigma_1\}} w\left(T-\sigma_1, x(q,\sigma_1)\frac{\lambda_1}{\lambda_0}f(Y_1)\right)\right] \\
&= \int_0^t e^{-\lambda_0 s}\left(x(q,s)-\frac{1}{c}\right)ds \\
&\quad + \int_0^t \lambda_0 e^{-\lambda_0 s}\left[\int_{y\in\mathbb{R}^d} w\left(T-s, x(q,s)\frac{\lambda_1}{\lambda_0}f(y)\right)\nu_0(dy)\right]ds \\
&= \int_0^t e^{-\lambda_0 s}\Big(g(x(q,s)) + \lambda_0\cdot(Sw)(T-s,x(q,s))\Big)ds,
\end{aligned} \tag{5.6}
$$

where S is the linear operator

$$
(Sw)(t,x) \triangleq \int_{y\in\mathbb{R}^d} w\left(t, x\frac{\lambda_1}{\lambda_0}f(y)\right)\nu_0(dy), \quad t\in[0,T],\ x\in\mathbb{R}_+,
$$

defined on the collection of bounded Borel functions $w:[0,T]\times\mathbb{R}_+\mapsto\mathbb{R}_-$.

Remark 5.2. *For every $T<\infty$ and bounded Borel function $w:[0,T]\times\mathbb{R}_+\mapsto\mathbb{R}_-$, the integrand of* (5.6) *is also bounded. Therefore, the mapping $t\mapsto(J_t w)(T,q):[0,T]\mapsto\mathbb{R}_-$ is continuous and bounded. Hence, the infimum $(Jw)(T,q)$ in* (5.5) *is attained.*

The next lemma shows that the operator J preserves boundedness and monotonicity.

Lemma 5.3. *For every bounded Borel function $w:[0,T]\times\mathbb{R}_+\mapsto\mathbb{R}_-$ taking values in $[-\frac{T}{c},0]$, we have*

$$
-\frac{T}{c}\leq(Jw)(T,q)\leq 0, \quad T\in\mathbb{R}_+,\ q\in\mathbb{R}_+. \tag{5.7}
$$

If $w_1(\cdot,\cdot)$ and $w_2(\cdot,\cdot)$ are bounded Borel functions defined on $[0,T]\times\mathbb{R}_+$ with $w_1(\cdot,\cdot)\leq w_2(\cdot,\cdot)$, then $(Jw_1)(\cdot,\cdot)\leq(Jw_2)(\cdot,\cdot)$.

Proof. Since $w(\cdot,\cdot)$ takes values in $[-\frac{T}{c}, 0]$, $(Sw)(T-s, x(q,s)) \geq -\frac{T-s}{c}$ holds for $s \in [0,T]$, $T \in \mathbb{R}_+$ and $q \in \mathbb{R}_+$. Then as $x(\cdot,\cdot)$ is positive and $t < T$,

$$\begin{aligned}
(J_t w)(T,q) &\geq \int_0^t e^{-\lambda_0 s}\left(x(q,s) - \frac{1}{c}\right) ds - \int_0^t \lambda_0 e^{-\lambda_0 s}\left(\frac{T-s}{c}\right) ds \\
&\geq -\frac{1}{c\lambda_0}\int_0^t \lambda_0 e^{-\lambda_0 s} ds - \frac{T}{c}\int_0^t \lambda_0 e^{-\lambda_0 s} ds + \frac{1}{c}\int_0^t s\lambda_0 e^{-\lambda_0 s} ds \\
&= -\frac{1}{c\lambda_0}(1 - e^{-\lambda_0 t}) - \frac{T}{c}(1 - e^{-\lambda_0 t}) + \frac{1 - e^{-\lambda_0 t}(\lambda_0 t + 1)}{c\lambda_0} \\
&= -\frac{T}{c}(1 - e^{-\lambda_0 t}) - \frac{t e^{-\lambda_0 t}}{c} \\
&= -\frac{T}{c} + (T-t)\frac{e^{-\lambda_0 t}}{c} \\
&\geq -\frac{T}{c}.
\end{aligned}$$

Taking infimum over $t \in [0,T]$ gives the first inequality. The second inequality follows when $t = 0$ which gives $(Jw)(T,q) = 0$.

For $t \in [0,T]$ and $q \in \mathbb{R}_+$, $w_1(t,q) \leq w_2(t,q)$ implies $(Sw_1)(t,q) \leq (Sw_2)(t,q)$ due to the linearity of the S operator. As the constant λ_0 is positive, $(J_t w_1)(T,q) \leq (J_t w_2)(T,q)$ for all $t \in [0,T]$. Taking the infimum over $t \in [0,T]$, we have $(Jw_1)(T,q) \leq (Jw_2)(T,q)$. □

The next lemma demonstrates when we apply the operator J to an increasing and concave function w, (Jw) will still be increasing and concave.

Lemma 5.4. *For every fixed $T \in \mathbb{R}_+$, if the mapping $q \mapsto w(T,q)$ is increasing and concave, then so is $q \mapsto (Jw)(T,q)$.*

Proof. Refer to Appendix C.1. □

This will be useful when we calculate the value function $U(\cdot,\cdot)$ numerically as described in the next chapter. For a fixed T, as the mapping $q \mapsto (Jw)(T,q)$ is increasing and bounded above by 0, it is enough to compute $(Jw)(T,q)$ until some q^* where $(Jw)(T,q^*) = 0$. For $q \geq q^*$, $(Jw)(T,q)$ is simply equal to 0. Hence, there is no need for extra computation.

This result also gives us a clue on the shape of the continuation and stopping regions introduced in the previous chapter. Recall that the stopping region Γ corresponds to points (T, q) in time-space domain $\mathbb{R}_+ \times \mathbb{R}_+$ where $U(T, q) = 0$. Hence, for a fixed time point T, the stopping region is formed by the points $\{(T, q) : q > q^*\}$ as opposed to a set of scattered points.

Let us define the successive approximations $\overline{u}_n : [0, T] \times \mathbb{R}_+ \mapsto \mathbb{R}_-$ and $\underline{u}_n : [0, T] \times \mathbb{R}_+ \mapsto \mathbb{R}_-$, $n \in \mathbb{N}_0$, sequentially by

$$\overline{u}_0 \equiv 0, \quad \overline{u}_n \triangleq J\overline{u}_{n-1}, \quad \text{and} \quad \underline{u}_0 \equiv -\frac{T}{c}, \quad \underline{u}_n \triangleq J\underline{u}_{n-1}. \tag{5.8}$$

Proposition 5.5. *For every $n \in \mathbb{N}_0$, $T \in [0, T_{max}]$ and $q \in \mathbb{R}_+$, we have*

$$-\frac{T}{c} \equiv \underline{u}_0 \leq \underline{u}_1 \leq \underline{u}_2 \leq \ldots \leq \ldots \leq \overline{u}_2 \leq \overline{u}_1 \leq \overline{u}_0 \equiv 0, \quad \forall T_{max} \in \mathbb{R}_+. \tag{5.9}$$

Proof. Observe that as $\overline{u}_0 \equiv 0 \in [-\frac{T}{c}, 0]$, we have $-\frac{T}{c} \leq \overline{u}_1 \equiv J\overline{u}_0 \leq \overline{u}_0 = 0$ as the operator J preserves boundedness by Lemma 5.3. Hence, $\overline{u}_1 \leq \overline{u}_0$ holds. Applying the operator J to $\overline{u}_1$ and $\overline{u}_0$, we have $-\frac{T}{c} \leq \overline{u}_2 \equiv J\overline{u}_1 \leq J\overline{u}_0 \equiv \overline{u}_1 \leq 0$ as the operator J preserves monotonicity and boundedness by Lemma 5.3. Successively applying the operator J to $\overline{u}_n$'s leads us to the decreasing sequence

$$-\frac{T}{c} \leq \ldots \leq \overline{u}_2 \leq \overline{u}_1 \leq \overline{u}_0 \equiv 0.$$

Similarly, as the operator J preserves monotonicity and boundedness by Lemma 5.3, continuously applying the operator J to $\underline{u}_n$'s gives the increasing sequence

$$-\frac{T}{c} \equiv \underline{u}_0 \leq \underline{u}_1 \leq \underline{u}_2 \leq \ldots \leq 0.$$

Let us show $\underline{u}_n \leq \overline{u}_n$ for $n \in \mathbb{N}_0$ by an induction argument. Note that $-\frac{T}{c} \equiv \underline{u}_0 \leq \overline{u}_0 \equiv 0$ by construction. Next, let us assume $\underline{u}_k \leq \overline{u}_k$ holds for some $k \in \mathbb{N}$. Since the operator J preserves monotonicity by Lemma 5.3, when we apply the operator J to $\underline{u}_k$ and $\overline{u}_k$ we have $\underline{u}_{k+1} \equiv J\underline{u}_k \leq J\overline{u}_k \equiv \overline{u}_{k+1}$ which implies $\underline{u}_{k+1} \leq \overline{u}_{k+1}$. This shows $\underline{u}_n \leq \overline{u}_n$ for $n \in \mathbb{N}_0$ and completes the proof. □

With the next proposition we will show that the limit of the bounded increasing sequence $(\underline{u}_n)_{n\geq 0}$ and the bounded decreasing sequence $(\overline{u}_n)_{n\geq 0}$ coincide.

Proposition 5.6. *For every $T_{max} \in \mathbb{R}_+$, as $n \to \infty$ the sequences of successive approximations $(\underline{u}_n)_{n\geq 0}$ and $(\overline{u}_n)_{n\geq 0}$ defined in (5.8) converge to the same limit*

$$u \triangleq \lim_{n\to\infty} \underline{u}_n = \lim_{n\to\infty} \overline{u}_n \tag{5.10}$$

uniformly in $T \in [0, T_{max}]$, $q \in \mathbb{R}_+$. In fact, for every $n \in \mathbb{N}_0$, $T \in [0, T_{max}]$ and $q \in \mathbb{R}_+$, we have

$$0 \leq \overline{u}_n(T,q) - \underline{u}_n(T,q) \leq \frac{T}{c}(1 - e^{-\lambda_0 T})^n. \tag{5.11}$$

Proof. In Proposition 5.5, we have showed that the first inequality in (5.11) holds for all $n \in \mathbb{N}_0$. Now, let us prove the second inequality.

By Remark 5.2, we know that

$$\overline{u}_n(T,q) = J\overline{u}_{n-1}(T,q) = J_{\overline{t}_n}\overline{u}_{n-1}(T,q),$$

$$\underline{u}_n(T,q) = J\underline{u}_{n-1}(T,q) = J_{\underline{t}_n}\underline{u}_{n-1}(T,q)$$

for some $\overline{t}_n$ and $\underline{t}_n$ in the interval $[0,T]$. Then by the linearity of the operator S

$$\begin{aligned}
\overline{u}_n - \underline{u}_n &= J_{\overline{t}_n}\overline{u}_{n-1}(T,q) - J_{\underline{t}_n}\underline{u}_{n-1}(T,q)\\
&\leq J_{\underline{t}_n}\overline{u}_{n-1}(T,q) - J_{\underline{t}_n}\underline{u}_{n-1}(T,q)\\
&= \int_0^{\underline{t}_n} e^{-\lambda_0 s}\lambda_0 (S(\overline{u}_{n-1} - \underline{u}_{n-1}))(T-s, x(q,s))ds\\
&\leq \|\overline{u}_{n-1} - \underline{u}_{n-1}\|_{T_{\max},\infty} \int_0^{\underline{t}_n} e^{-\lambda_0 s}\lambda_0 ds\\
&\leq \|\overline{u}_{n-1} - \underline{u}_{n-1}\|_{T_{\max},\infty} \int_0^{T_{\max}} e^{-\lambda_0 s}\lambda_0 ds\\
&= \|\overline{u}_{n-1} - \underline{u}_{n-1}\|_{T_{\max},\infty}(1 - e^{-\lambda_0 T_{\max}})
\end{aligned}$$

holds for all $n \in N_0$. Taking the supremum of $\overline{u}_n - \underline{u}_n$ over $(T,q) \in [0, T_{\max}] \times \mathbb{R}_+$ gives

$$\|\overline{u}_n - \underline{u}_n\|_{T_{\max},\infty} \leq \|\overline{u}_{n-1} - \underline{u}_{n-1}\|_{T_{\max},\infty}(1 - e^{-\lambda_0 T_{\max}}).$$

This shows that the gap between the terms of the sequences $(\underline{u}_n)_{n\geq 0}$ and $(\overline{u}_n)_{n\geq 0}$

contracts by the factor $(1-e^{-\lambda_0 T_{\max}})$ with each new term. Hence,

$$\begin{aligned}
\|\overline{u}_n - \underline{u}_n\|_{T_{\max},\infty} &\leq \|\overline{u}_{n-1} - \underline{u}_{n-1}\|_{T_{\max},\infty}(1-e^{-\lambda_0 T_{\max}}) \\
&\leq \|\overline{u}_0 - \underline{u}_0\|_{T_{\max},\infty}(1-e^{-\lambda_0 T_{\max}})^n \\
&= \frac{T}{c}(1-e^{-\lambda_0 T_{\max}})^n.
\end{aligned}$$

Therefore, (5.11) holds for all $n \in \mathbb{N}_0$. As a result, as $n \to \infty$ the sequences $(\underline{u}_n)_{n\geq 0}$ and $(\overline{u}_n)_{n\geq 0}$ converge to the common limit u in (5.10). $\square$

This means that whether we start with $\underline{u}_0 \equiv -\frac{T}{c}$ or $\overline{u}_0 \equiv 0$, we will end up eventually at the fixed limit u as $n \to \infty$. In fact, we can reach to the this limit u starting with any bounded function $f : [0,T] \times \mathbb{R}_+ \mapsto \mathbb{R}_-$ taking values in $[-\frac{T}{c}, 0]$. In the next proposition, we will prove this claim.

Before stating the proposition, let us introduce the f-successive approximation sequence $u_n(f),\ n = 0, 1, \ldots$ of the function $f : [0,T] \times \mathbb{R}_+ \mapsto [-\frac{T}{c}, 0]$ defined by

$$u_0(f)(T,q) \triangleq f(T,q), \tag{5.12}$$

$$u_n(f)(T,q) \triangleq J(u_{n-1}(f))(T,q) \tag{5.13}$$

for $T \in [0, T_{\max}],\ q \in \mathbb{R}_+$.

Proposition 5.7. *For every $f : [0,T] \times \mathbb{R}_+ \mapsto [-\frac{T}{c}, 0]$ and bounded Borel functions $\overline{u}_n$ and $\underline{u}_n$, $n \in \mathbb{N}_0$ we have*

$$\underline{u}_n \equiv u_n\left(-\frac{T}{c}\right) \leq u_n(f) \leq u_n(0) \equiv \overline{u}_n, \quad n \in \mathbb{N}_0.$$

Moreover,

$$0 \leq u_n(f) - \underline{u}_n \leq \overline{u}_n - \underline{u}_n \leq \frac{T}{c}(1-e^{-\lambda_0 T})^n, \quad n \in \mathbb{N}_0.$$

Therefore,

$$u = \lim_{n\to\infty} \underline{u}_n = \lim_{n\to\infty} \overline{u}_n = \lim_{n\to\infty} u_n(f), \tag{5.14}$$

exists.

Proof. The conclusions follow from repeated application of the operator J and Propositions 5.3 and 5.6. $\square$

Proposition 5.8. *For every $n \in \mathbb{N}$, the functions $\overline{u}_n$ of* (5.8) *and U_n of* (5.1) *coincide. For every $\varepsilon \geq 0$, let*

$$r_n^{\varepsilon}(T,q) \triangleq \inf\{s \in (0,T] : J_s\overline{u}_n(T,q) \leq J\overline{u}_n(T,q) + \varepsilon\},$$

for $n \in \mathbb{N}_0,\ T \in \mathbb{R}_+,\ q \in R_+,$ and

$$S_1^{\varepsilon} \triangleq r_0^{\varepsilon}(T,q) \wedge \sigma_1$$

$$S_{n+1}^{\varepsilon} \triangleq \begin{cases} r_n^{\varepsilon/2}(T,q), & \text{if } \sigma_1 > r_n^{\varepsilon/2}(T,q) \\ \sigma_1 + S_n^{\varepsilon/2} \circ \theta_{\sigma_1}, & \text{if } \sigma_1 \leq r_n^{\varepsilon/2}(T,q) \end{cases}, \quad n \in \mathbb{N},$$

where θ is the shift operator. Then

$$\mathbb{E}_{0,q}\left[\int_0^{S_n^{\varepsilon}} g(Q_t)dt\right] \leq \overline{u}_n(T,q) + \varepsilon, \quad \forall n \in N,\ \forall \varepsilon \geq 0. \tag{5.15}$$

Proof. Refer to Appendix C.2. □

Propositions 5.1 and 5.8 suggest that we can approximate the value function U of the optimal stopping problem uniformly with the functions $\overline{u}_n$ defined sequentially in (5.5). The sequence $(\overline{u}_n)_{n\geq 0}$ converges to U at an exponential rate, and the explicit bounds in Proposition 5.1 determine the number of iterations to achieve any desired accuracy. In the next chapter, we will describe how to compute the successive approximations $\overline{u}_n$, $n \in \mathbb{N}$ and obtain the value function U numerically.

Chapter 6

Solution

Note that between two consecutive arrivals of the compound Poisson process X, no new information is observed. Thus, our problem evolves deterministically and the all stopping rules at or before the next arrival are constant. Hence, our original problem turns out to be nothing but a series of deterministic subproblems connected by the random jumps of X. Therefore, we can solve those subproblems iteratively to obtain a solution to the optimal stopping problem by relying on the dynamic programming principle.

In the previous chapter, we have showed how to approximate the value function U in (3.13) of the optimal stopping problem with the successive approximations by using a dynamic programming approach by Propositions 5.1 and 5.8. The dynamic programming operator $(J_t w)$ given in (5.4) can be interpreted as the value of waiting until the constant $t \in [0, T]$, and then acting according to the function w. Minimizing this value over $t \in [0, T]$ is a deterministic optimization problem. The minimum objective value of this problem, the cost of waiting until the optimal $t \in [0, T]$ and then acting according to w, is stored in the dynamic programming operator (Jw) defined in (5.5). Thus, by repeatedly applying the operator J to the successive approximations and solving the deterministic subproblems, we can calculate U and provide a solution to our original optimal stopping problem.

In the following sections, we will explain how to calculate the value function U numerically, describe the quickest detection rule explicitly, and present a clear solution algorithm.

6.1 Calculation of U

We would like to approximate the value function $U(T, q)$ for $0 \leq T \leq T_0$ and $0 \leq q \leq Q_0$ for some large T_0 and Q_0. One way to achieve this is to put a grid on $[0, T_0] \times [0, Q_0]$, evaluate the value function on the grid nodes and interpolate the remaining points in between.

In the following subsections, we will describe how to evaluate (Jw) numerically with the Monte Carlo methods, approximate the value function U with an iterative algorithm and later interpolate it with Chebyshev interpolation.

6.1.1 Numeric evaluation of (Jw)

The calculation of the value function U of the optimal stopping problem requires the calculation of the dynamic programming operator (Jw) in (5.5) when applied to the successive approximations. Hence, for a given continuous and bounded function $w : [0, T] \times \mathbb{R}_+ \mapsto [-\frac{T}{c}, 0]$, we would like to calculate $(Jw)(T, q)$ for $0 \leq T \leq T_0$ and $0 \leq q \leq Q_0$ for some large T_0 and Q_0.

Let us start with separating $(J_t w)(T, q)$ into two parts as

$$\begin{aligned}(J_t w)(T, q) &= \int_0^t e^{-\lambda_0 s} \left(x(q, s) - \frac{1}{c} + \lambda_0 (Sw)\left(T - s, x(q, s)\right) \right) ds \\ &= \underbrace{\int_0^t e^{-\lambda_0 s} \left(x(q, s) - \frac{1}{c} \right) ds}_{\equiv I(t,q)} + \int_0^t \lambda_0 e^{-\lambda_0 s} (Sw)\left(T - s, x(q, s)\right) ds,\end{aligned}$$

and explicitly calculate the first part $I(t,q)$ as

$$\begin{aligned}
I(t,q) &\equiv \int_0^t e^{-\lambda_0 s}\left(x(q,s)-\frac{1}{c}\right)ds \\
&= \left\{\begin{array}{ll} \dfrac{1-e^{-\lambda_0 t}}{\lambda_0(\lambda_1-\lambda_0)} + [q(\lambda_1-\lambda_0)-1]\dfrac{1-e^{-\lambda_1 t}}{\lambda_1(\lambda_1-\lambda_0)}, & \lambda_1 \neq \lambda_0 \\ \dfrac{1}{\lambda_0^2}\left[\lambda_0 q + 1 - e^{-\lambda_0 t}\big(\lambda_0(q+t)+1\big)\right], & \lambda_1 = \lambda_0 \end{array}\right\} - \frac{1-e^{-\lambda_0 t}}{c\lambda_0}.
\end{aligned}$$

The remaining second part

$$\begin{aligned}
(J_t w)(T,q) - I(q,t) &= \int_0^t \lambda_0 e^{-\lambda_0 s}(Sw)\left(T-s, x(q,s)\right)ds \\
&= \int_0^t \lambda_0 e^{-\lambda_0 s}\left[\int_{y\in\mathbb{R}^d} w\left(T-s, x(q,s)\frac{\lambda_1}{\lambda_0}f(y)\right)\nu_0(dy)\right]ds
\end{aligned} \tag{6.1}$$

can be calculated with the Monte Carlo methods. By sampling random time points in the interval $[0,t]$ and random marks from the distribution ν_0, we can get a good estimate of the double integral in (6.1).

In fact, a careful look at the integrand suggests that sampling exponential time points that are clustered around 0, rather than points scattered uniformly in the interval $[0,t]$ would improve our approximation of the outer integral. With this importance sampling technique, we can get a better estimate of the outer integral while sampling fewer points. Indeed, the expression for the remaining part can be rewritten as

$$\begin{aligned}
(J_t w)(T,q) - I(q,t) &= \int_0^t \lambda_0 e^{-\lambda_0 s}(Sw)\left(T-s, x(q,s)\right)ds \\
&= \int_0^\infty \lambda_0 e^{-\lambda_0 s}1_{\{s\leq t\}}(Sw)\big(T-s, x(q,s)\big)ds \\
&= \mathbb{E}\left[1_{\{E\leq t\}}(Sw)\big(T-E, x(q,E)\big)\right] \\
&= \mathbb{E}\left[(Sw)\big(T-E, x(q,E)\big)\Big| E\leq t\right]\mathbb{P}\{E\leq t\} \\
&= \mathbb{E}\left[(Sw)\big(T-E_t, x(q,E_t)\big)\right](1-e^{-\lambda_0 t})
\end{aligned}$$

in terms of a truncated exponential random variable E_t with the cumulative distribution function

$$
\begin{aligned}
F_t(s) &:= \mathbb{P}\{E_t \le s\} = \mathbb{P}\{E \le s \mid E \le t\} \\
&= \begin{cases} 0, & \text{if} \quad s < 0, \\ \dfrac{\mathbb{P}\{T \le s\}}{\mathbb{P}\{T \le t\}} = \dfrac{1 - e^{-\lambda_0 s}}{1 - e^{-\lambda_0 t}}, & \text{if} \quad 0 \le s \le t, \\ 1, & \text{if} \quad s \ge t. \end{cases}
\end{aligned}
$$

Hence, combining two parts we have

$$
(J_t w)(T, q) = I(q, t) + \mathbb{E}\left[(Sw)\big(T - E_t, x(q, E_t)\big)\right](1 - e^{-\lambda_0 t}).
$$

Then we can generate a series of truncated exponential random variates by the inverse transform method to calculate the expectation in the expression above. Observe that for every $U \sim \text{Uniform}(0,1)$, $F_t^{-1}(U)$ and E_t have the same distribution, where

$$
F_t^{-1}(u) = -\frac{1}{\lambda_0} \ln\left(1 - u + ue^{-\lambda_0 t}\right), \quad 0 \le u \le 1.
$$

Thus, we can sample the exponential random variates $E_1, \ldots, E_n$ by drawing n i.i.d. Uniform (0,1) random variates $U_1, \ldots, U_n$ and setting $E_{t,i} := F_t^{-1}(U_i)$, $i = 1, \ldots, n$. Hence, we can approximate $(J_t w)(T, q)$ by

$$
(J_t w)(T, q) \approx I(q, t) + (1 - e^{-\lambda_0 t}) \frac{1}{n} \sum_{i=1}^{n} (Sw)\big(T - E_{t,i}, x(q, E_{t,i})\big)
$$

for $0 \le t \le T \le T_0$ and $0 \le q \le Q_0$.

Similarly, if we draw m i.i.d. random variates $Y_1, \ldots, Y_m$ from distribution ν_0, then we can estimate the operator (Sw) by

$$
(Sw)(t, x) \approx \frac{1}{m} \sum_{j=1}^{m} w\left(t, x \frac{\lambda_1}{\lambda_0} f(Y_j)\right)
$$

for $0 \le t \le T_0$ and $0 \le q \le Q_0$.

Finally, when we combine the last two expressions, we have

$$
(J_t w)(T, q) \approx I(q, t) + (1 - e^{-\lambda_0 t}) \frac{1}{n\, m} \sum_{i=1}^{n} \sum_{j=1}^{m} w\left(T - E_{t,i}, x(q, E_{t,i}) \frac{\lambda_1}{\lambda_0} f(Y_j)\right)
$$

for $0 \le t \le T \le T_0$ and $0 \le q \le Q_0$.

In an alternative formulation, we can work with the $E_{T_0,i}$ random variates instead of the $E_{t,i}$ random variates, which may allow us to sample once and calculate $(J_t w)(T,q)$ for every $0 \le t \le T \le T_0$ and $0 \le q \le Q_0$ at once. This can be done by noting that we can also write (6.1) as

$$\begin{aligned}
(J_t w)(T,q) - I(q,t) &= \int_0^\infty \lambda_0 e^{-\lambda_0 s} 1_{\{s \le t\}} (Sw)\big(T-s, x(q,s)\big) ds \\
&= \mathbb{E}\left[1_{\{E \le T_0\}} 1_{\{E \le t\}} (Sw)\big(T-E, x(q,E)\big)\right] \\
&= \mathbb{E}\left[1_{\{E \le t\}} (Sw)\big(T-E, x(q,E)\big) \Big| E \le T_0\right] \mathbb{P}\{E \le T_0\} \\
&= \mathbb{E}\left[1_{\{E_{T_0} \le t\}} (Sw)\big(T-E_{T_0}, x(q,E_{T_0})\big)\right] (1 - e^{-\lambda_0 T_0}).
\end{aligned}$$

Hence, $(J_t w)(T,q)$ is given by

$$(J_t w)(T,q) = I(q,t) + \mathbb{E}\left[1_{\{E_{T_0} \le t\}} (Sw)\big(T-E_{T_0}, x(q,E_{T_0})\big)\right] (1 - e^{-\lambda_0 T_0}).$$

Let $E_{T_0,1}, \ldots, E_{T_0,n}$ be i.i.d. random variates drawn from $F_{T_0}(\cdot)$ distribution. Then $(J_t w)(T,q)$ can be approximated by

$$\begin{aligned}
(J_t w)(T,q) &\approx I(q,t) + (1 - e^{-\lambda_0 T_0}) \frac{1}{n} \sum_{i=1}^{n} 1_{\{E_{T_0,i} \le t\}} (Sw)\big(T - E_{T_0,i}, x(q, E_{T_0,i})\big) \\
&\approx I(q,t) + (1 - e^{-\lambda_0 T_0}) \frac{1}{n\,m} \sum_{i=1}^{n} 1_{\{E_{T_0,i} \le t\}} \sum_{j=1}^{m} w\left(T - E_{T_0,i}, x(q, E_{T_0,i}) \frac{\lambda_1}{\lambda_0} f(Y_j)\right)
\end{aligned}$$

for every $0 \le t \le T \le T_0$ and $0 \le q \le Q_0$.

As a result, taking the infimum over $t \in [0,T]$ simply gives $(Jw)(T,q)$ as

$$\begin{aligned}
(Jw)(T,q) = \inf_{t \in [0,T]} (J_t w)(T,q) \approx \inf_{t \in [0,T]} \Bigg\{ I(q,t) \\
+ (1 - e^{-\lambda_0 T_0}) \frac{1}{n\,m} \sum_{i=1}^{n} 1_{\{E_{T_0,i} \le t\}} \sum_{j=1}^{m} w\left(T - E_{T_0,i}, x(q, E_{T_0,i}) \frac{\lambda_1}{\lambda_0} f(Y_j)\right) \Bigg\}.
\end{aligned}$$

In order to take this infimum numerically, we calculate $(J_t w)(T,q)$ at the T_0-truncated exponential random variates in the interval $[0,T]$, and take their minimum to assign to $(Jw)(T,q)$. Although this does not guarantee optimality, it gives satisfactory results.

6.1.2 Grid calculation of U

Thus far we have explained how to evaluate the dynamic programming operator J when it is applied to the successive approximations. Now, we will describe how to calculate the value function U on a dynamic grid with an iterative algorithm.

Before we start with the algorithm, we must first decide on where to put the grid points in $[0, T_0] \times [0, Q_0]$. This decision will be important when we interpolate the value function U. Note that the selection of the grid points is itself a research problem worth considering. Although there are many alternatives, we select a mesh consisting of the Chebyshev grid points. The reason for this choice and further details of those special grid points will be described in the next subsection.

The iterative algorithm that calculates the value function U revolves around two types of iterations: the Chebyshev and the value function iterations. At each Chebyshev iteration, we observe a series of value function iterations. Those value function iterations correspond to the sequence of successive approximations $\overline{u}_n$, $n \in \mathbb{N}_0$ in (5.5). In Proposition 5.8, we showed that we can approximate U with those successive approximations. In fact, we proved that we can reach to U starting from any function f taking values in the interval $[-\frac{T}{c}, 0]$ by Proposition 5.7. When the value function iterations converge, the last successive approximation gives a good approximate of the value function U. Then we update the grid and move to the next Chebyshev iteration. In this new Chebyshev iteration, we again calculate the successive approximations until they converge. We continue in this manner until the Chebyshev iterations converge.

The algorithm begins with a fixed number of Chebyshev grid points in each dimension, and assigns 0 to the value of each grid point as $\overline{u}_0 \equiv 0$. Then we repeatedly apply the dynamic programming operator J to each grid point as described in the previous subsection and obtain the subsequent successive approximations. At each value function iteration, we calculate the maximum relative distance between the old and the new successive approximation. When this distance becomes negligible, the value function iterations converge. This concludes the first Chebyshev iteration.

For the next Chebyshev iteration, we add new grid points in each dimension while keeping the old ones. In this new Chebyshev iteration, we again compute the successive approximations iteratively until the stopping criteria is met. Thus, at each Chebyshev iteration, we update our dynamic grid and calculate the value function iteratively. To check the convergence of the Chebyshev iterations, we compute the maximum relative distance between the value function approximations obtained at the end of two consecutive Chebyshev iterations on the shared grid points of the preceding Chebyshev iteration. When this distance goes below a certain threshold level, Chebyshev iterations converge and the algorithm stops.

We considered the following ways to speed up the algorithm and save computation time.

1. At the first Chebyshev iteration, we start the value function iterations with $\overline{u}_0 \equiv 0$. At each subsequent Chebyshev iteration, we take the last successive approximation of the preceding Chebyshev iteration as the first successive approximation. This way we save the information we have learned about the value function; thus, the successive approximations converge with fewer iterations.

2. In Lemma 5.4, we showed that if the mapping $q \mapsto w(T, q)$ is increasing and concave, then so is $q \mapsto (Jw)(T, q)$. By this proposition, we only need to calculate the successive approximations on a subset of the grid points since the successive approximations are bounded above by 0. At each value function iteration, for a fixed time grid point T, we calculate the next successive approximation until some Q grid point q^* where the value of the next successive approximation at the point (T, q^*) is 0. For every Q grid point $q \geq q^*$, we simply equate the value of the next successive approximation at the points (T, q) to 0. This reduces the amount of computations at each value function iteration.

3. At each value function iteration, since the calculations of the successive approximations at different grid points are independent, we do them in parallel. Hence, we reduce the computation time with the help of multiple cores.

6.1.3 Interpolation of U

We have discussed how to compute the value function U on the Chebyshev grid points scanning $[0, T_0] \times [0, Q_0]$ in the previous subsection. Now, we will describe how to interpolate the remaining points in between.

Interpolation methods range from simple methods such as piecewise constant and linear interpolation to more involved methods such as polynomial and spline interpolation, including the use of basis functions. When simple methods provide results that are good enough, it is reasonable to use them to avoid lengthy computations and extra computation time. However, in some cases, we need more advanced interpolation techniques to have a better fit with the minimum error and the maximum accuracy.

Properties of U, shown in the previous chapter, suggest that polynomial interpolation would be suitable for our task. Therefore, in this study we use polynomial interpolation with Chebyshev points as described on Section 8.5 in Greenbaum and Chartier [16]. The idea behind the Chebyshev interpolation is to use a distorted grid with the interpolation points clustered near the endpoints of the interpolation interval instead of an equally spaced grid to avoid the difficulties associated with high-degree polynomial interpolation; see Runge's phenomenon. In fact, Theorem 8.5.1 on Section 8.5 in Greenbaum and Chartier [16] states that if f is a continuous function on [-1, 1], p_n is its degree n polynomial interpolant at the Chebyshev points, and p_n^* is its best approximation among all nth-degree polynomials on the interval [-1, 1] in the ∞-norm $\|\cdot\|_\infty$, then

$$\|f - p_n\|_\infty \leq \left(2 + \frac{2}{\pi}\log n\right) \|f - p_n^*\|_\infty.$$

That means the polynomial interpolant at the Chebyshev points is close to the best polynomial approximation in the ∞-norm.

Chebyshev interpolation works as follows. Suppose we want to approximate a function $f : [0, Q_0] \mapsto \mathbb{R}$ by some n-degree Lagrange polynomial $p_n(\cdot)$. These are the steps to be followed.

1. Scale the domain from $[0, Q_0]$ to $[-1, 1]$ with $q \mapsto x := \varphi(q) = -1 + 2q/Q_0$, and define $F : [-1, 1] \mapsto \mathbb{R}$ by

$$F(x) := f(q) = f\left(\frac{x+1}{2}\right).$$

2. Calculate F at the Chebyshev points

$$x_j = \cos\left(\frac{\pi j}{n}\right), \quad j = 0, \dots, n.$$

   Let $F_j = F(x_j)$, $j = 0, \dots, n$.

3. Calculate the Lagrange polynomial passing through (x_j, y_j), $j = 0, \dots, n$.

$$p_n(x) = \sum_{j=0}^{n} F_j \prod_{i \neq j} \frac{x - x_i}{x_j - x_i} = \sum_{j=0}^{n} F_j w_j \frac{\phi(x)}{x - x_j} = \phi(x) \sum_{j=0}^{n} \frac{w_j}{x - x_j} F_j,$$

   where

$$w_j := \prod_{i \neq j} \frac{1}{x_j - x_i} = \frac{1}{\prod_{i \neq j}(x_i - x_j)}, \quad \phi(x) := \prod_{j=0}^{n} (x - x_j).$$

   Suppose each $F_j = 1$. Then $\widetilde{p}_n$ is another n-degree interpolating polynomial. Because p_n and $\widetilde{p}$ agree at $n+1$ locations $(x_0, \dots, x_n)$, those polynomials must coincide everywhere:

$$1 \equiv \widetilde{p}(x) = p_n(x) = \phi(x) \sum_{j=0}^{n} \frac{w_j}{x - x_j} \quad \text{for every } x \in [-1, 1].$$

   Hence, we have

$$\phi(x) = \frac{1}{\sum_{j=0}^{n} \frac{w_j}{x - x_j}},$$

   and the interpolating Lagrange polynomial becomes

$$p_n(x) = \begin{Bmatrix} \displaystyle\sum_{j=0}^{n} \frac{w_j}{x - x_j} F_j \Bigg/ \sum_{k=0}^{n} \frac{w_k}{x - x_k}, & x \neq x_j, \ j = 0, \dots, n \\ F_j, & x = x_j, \ j = 0, \dots, n \end{Bmatrix},$$

In the meantime

$$
\begin{aligned}
w_j &= 1 \Big/ \prod_{i \neq j} (x_i - x_j) \\
&= 1 \Big/ \prod_{i \neq j} \left[\cos\left(\frac{\pi i}{n}\right) - \cos\left(\frac{\pi j}{n}\right) \right] \\
&= \frac{2^{n-1}}{n} \begin{cases} \dfrac{(-1)^j}{2}, & j = 0 \text{ or } j = n, \\ (-1)^j, & \text{otherwise.} \end{cases}
\end{aligned}
$$

In the interpolation, we can drop the common constant $2^{n-1}/n$, because they appear in both numerator and denominator, and cancel each other.

Hence, the interpolation over the original domain becomes $f_n(q) = p_n(x) = p_n(2q/Q_0 - 1)$.

As a result, when we are interpolating the value function with Chebyshev interpolation in the Q dimension we follow the recipe below.

1. Set the degree of a polynomial to a large integer n.
2. Calculate $n + 1$ Chebyshev points x_j in $[-1, 1]$.
3. Calculate the weights w_j.
4. Transform Chebyshev points linearly into $[0, Q_0]$, $x_j \mapsto q_j = 0.5(x_j + 1)Q_0$.
5. Calculate the function at Chebyshev points by means of dynamic programming equation, $F_j = F(q_j)$.
6. Interpolate the value function over $[0, Q_0]$ with the Lagrange polynomial as described and implemented above.

Likewise, when we are interpolating the value function with Chebyshev interpolation in the T dimension, we follow the steps described above while replacing Q_0 with T_0.

6.2 Quickest detection rule

In this section, we will briefly describe the quickest detection rule, state the optimal time to stop and raise an alarm depending on the sample-path of the sufficient statistic Q, and explain how to solve the optimal stopping problem with a clear and concise solution algorithm.

Recall that in the variational inequalities chapter, we have partitioned the state space of the process Q and its time index set into a continuation region C and a stopping region Γ which correspond to the points (T, q) in the time-space domain where $U(T, q) < 0$ and $U(T, q) = 0$, respectively. We described the quickest detection rule as the first crossing time $\inf\{t \geq 0 : Q_t \in \Gamma\}$ of Q to the stopping region Γ.

However, since deriving analytical results on the free-boundary was difficult, we had no idea about the structure of the continuation and the stopping regions. In fact, the time-space domain could have been partitioned in any way imaginable. Yet, Lemma 5.4 revealed some information and suggested that for a fixed time point the stopping region consisted of a set points above a certain threshold level. Although the threshold levels for different time points were unknown, we learned that the stopping region lies above the continuation region at every fixed time point.

In line with this observation, we can develop the following insight. The sufficient statistic Q can intuitively be viewed as information, and the random value Q_t may represent our belief on the occurrence of the change-point at or before time t. Hence, a low value of the process Q implies that we do not have enough information to make the stopping decision, and it is likely that the change has not happened yet. When this is the case, it is better to wait, continue with our observations, and raise the alarm at a later moment in time. On the other hand, a high value of Q indicates that we have enough information and strongly believe that the change has happened already. Therefore, when the value of Q reaches to a certain level the best action is to raise an alarm and stop immediately.

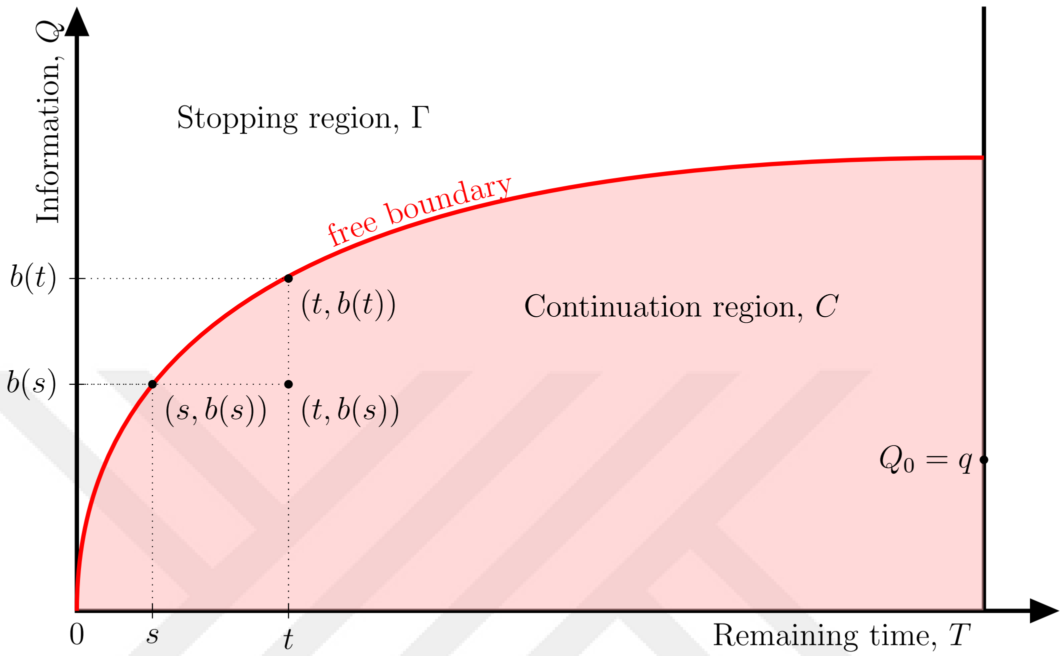


Figure 6.1: Stopping and continuation regions in time-space domain.

Observe that since we expect the change to occur until time T, we must stop at T at the latest. This suggests that the threshold level the process Q has to pass in order to raise the alarm depends on the remaining time period until time T. While this level is expected to be higher when the remaining time period is long, it should decrease and fall to zero as we get closer to, and hit time T, respectively. With this intuition, we expect the free-boundary to be a monotone increasing function of the remaining time and the time-space domain to resemble Figure 6.1.

In the previous sections, we have explained in detail how to obtain the value function $U(T, q)$ for $0 \leq T \leq T_0$ and $0 \leq q \leq Q_0$ for some large T_0 and Q_0. Thus, we can calculate U and separate those two regions numerically. With our numerical study, we verify that the time-space domain indeed has the expected shape and the free-boundary is monotone. We provide various examples with different parameters; see Figures 7.2 and 7.3. Now, let us try to see why that is the case. Firstly, observe that the mapping $T \mapsto U(T, q)$ is decreasing for a fixed $q \in \mathbb{R}_+$ as taking the infimum in (3.13) over a larger set results in a decreased objective value. Suppose the free-boundary is represented with the function $b(t)$ of the remaining time. Then for points $s \leq t$ in the interval $[0, T]$ with $U(s, b(s)) = U(t, b(t)) = 0$, we want to show that $b(s) \leq b(t)$. Note that we have $U(t, b(s)) \leq U(s, b(s)) = U(t, b(t)) = 0$ as the mapping $T \mapsto U(T, q)$ is decreasing for a fixed $q \in \mathbb{R}_+$. That implies $U(t, b(s)) \leq U(t, b(t))$. Hence, we can conclude $b(s) \leq b(t)$ as the mapping $q \mapsto U(t, q)$ is increasing for a fixed $t \in [0, T]$ by Lemma 5.4. This shows that the free-boundary is a monotone increasing function of the remaining time.

Let us describe the on-line implementation of the quickest detection rule and present the solution algorithm. We start with setting $Q_0 = T\left(\frac{\pi}{1-\pi}\right)$ at time 0. If the point (T, Q_0) falls into the stopping region Γ, i.e. $U(T, Q_0) = 0$, then we stop and raise an alarm. If not, we update the remaining time T and sufficient statistic Q according to (3.11) by observing the compound Poisson process X in real-time. When Q enters Γ, we immediately stop and raise an alarm. Note that Q can enter Γ during a deterministic motion or by a jump.

Between two consecutive jumps, Q follows a deterministic path. As a result, the time that it would enter the stopping region Γ if it were to follow this deterministic path is also deterministic. If it were to jump, then it would move to another deterministic path. Again, the time that it would enter Γ if it were to follow this new deterministic path is also deterministic. In other words, we have a deterministic alarm time from the start. If Q jumps before this alarm time, then we update it; otherwise, we stop. Hence, the on-line implementation of the quickest detection rule can be summarized by the following solution algorithm.

Step 0 Choose the best deterministic $t^* = t^*(T, Q_0)$ to set an alarm before the next information arrival.

Step 1 If no information arrives before time t^*, i.e. $t^* < \sigma_1$, stop and raise an alarm at time t^*.

Step 2 If information arrives before time t^*, i.e. $t^* \geq \sigma_1$, then

- update your belief about the change, i.e. $Q_0 \to Q_{\sigma_1}$,
- go to Step 0, calculate and act according to the new $t^* \leftarrow t^*(T - \sigma_1, Q_{\sigma_1})$.

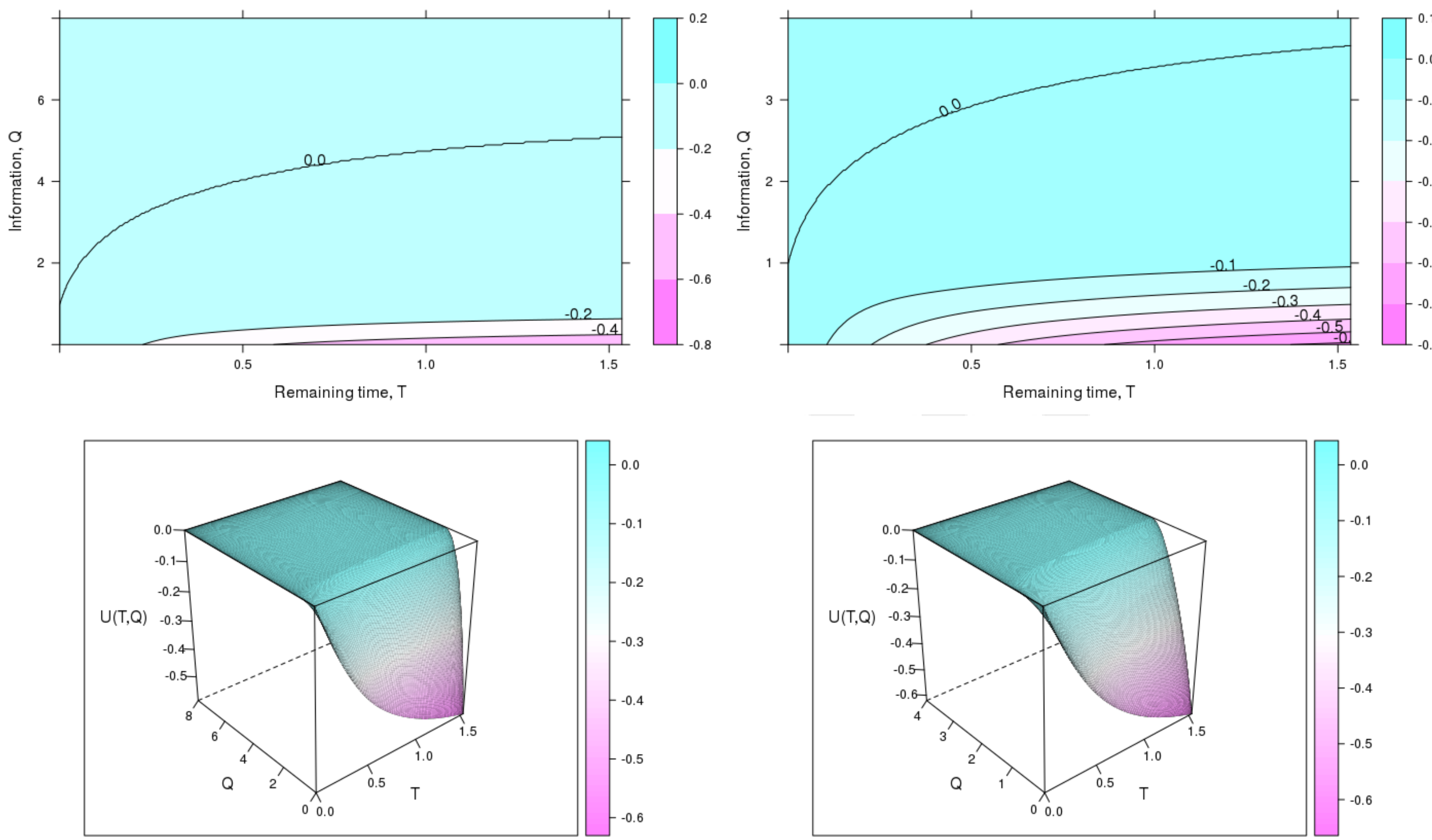
Chapter 7

Numerical examples

In the previous chapter, we discussed how to calculate the value function U and determine the free-boundary in the time-space domain numerically. Then we described the quickest detection rule and presented our solution algorithm. In this chapter, we illustrate our solution method with numerical examples.

In all numerical examples we present, marks before and after the change are assumed to be normally distributed. That is, both ν_0 and ν_1 are assumed to be normal. The mark distribution before the change is taken as standard normal, i.e. normal with mean of 0 and standard deviation of 1. After the change-point, the mean changes from $\mu_0 \equiv 0$ to μ_1 while the standard deviation remains constant.

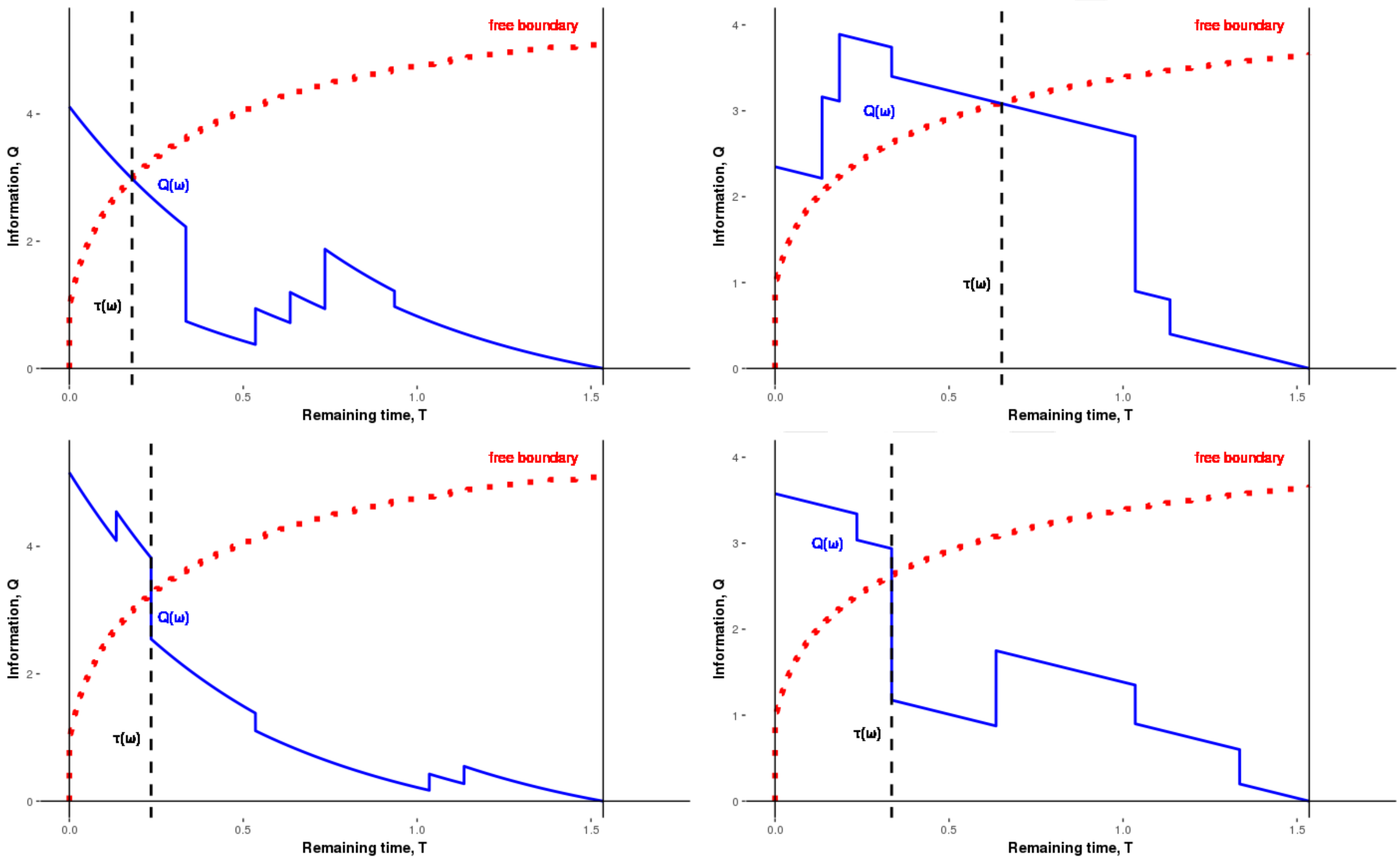
Contour plots and wireframe plots in Figure 7.1 present some examples of value functions while Figures 7.2 and 7.3 portray some sample-paths realizations of the sufficient statistic Q. In those figures, the x and y axes display the remaining time and the value of the process Q, respectively. In Figures 7.2 and 7.3 the time horizon T is shown with a vertical line. Note that the remaining time decreases as time goes on. Hence, the sample-path realizations, represented with the blue solid lines, start at time T and move in the direction of the origin. When they hit the red dotted line, i.e. the free-boundary, we stop and raise an alarm. The realizations of the stopping time τ is depicted with a dashed line.



(a) $\lambda_0 > \lambda_1 = 1.5$

(b) $\lambda_0 = \lambda_1 = 3$

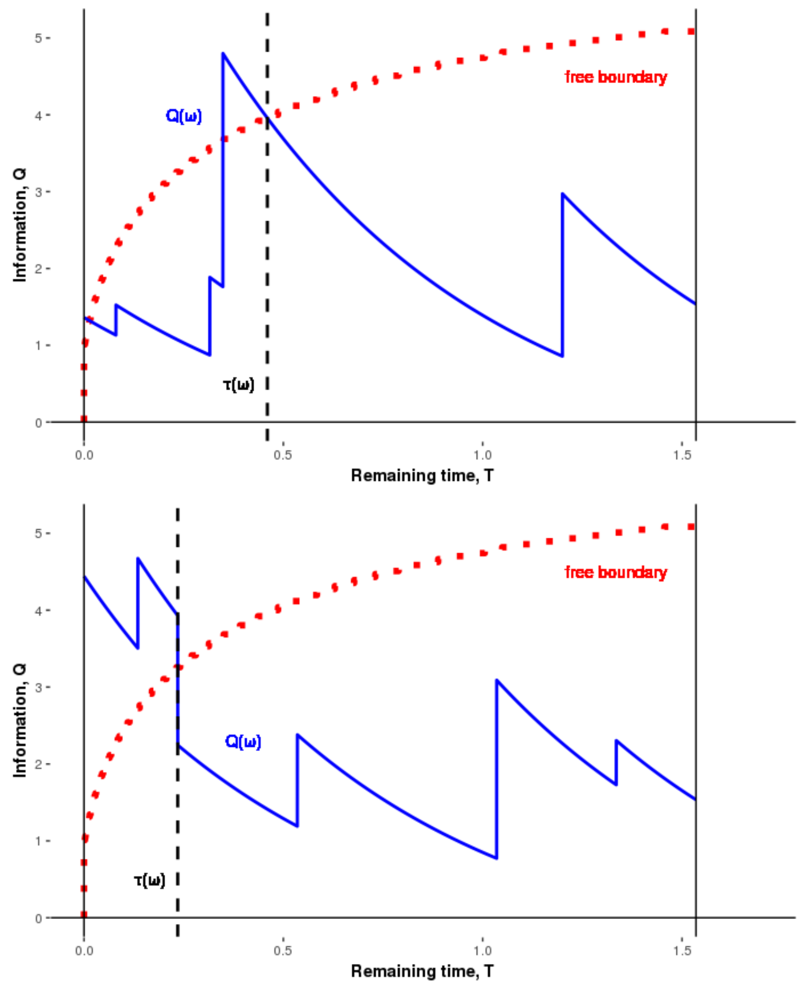
Figure 7.1: Contour plots and wireframe plots of the value function U ($T = 1.5351,\ \lambda_0 = 3,\ c = 1,\ \mu_1 = 1$)



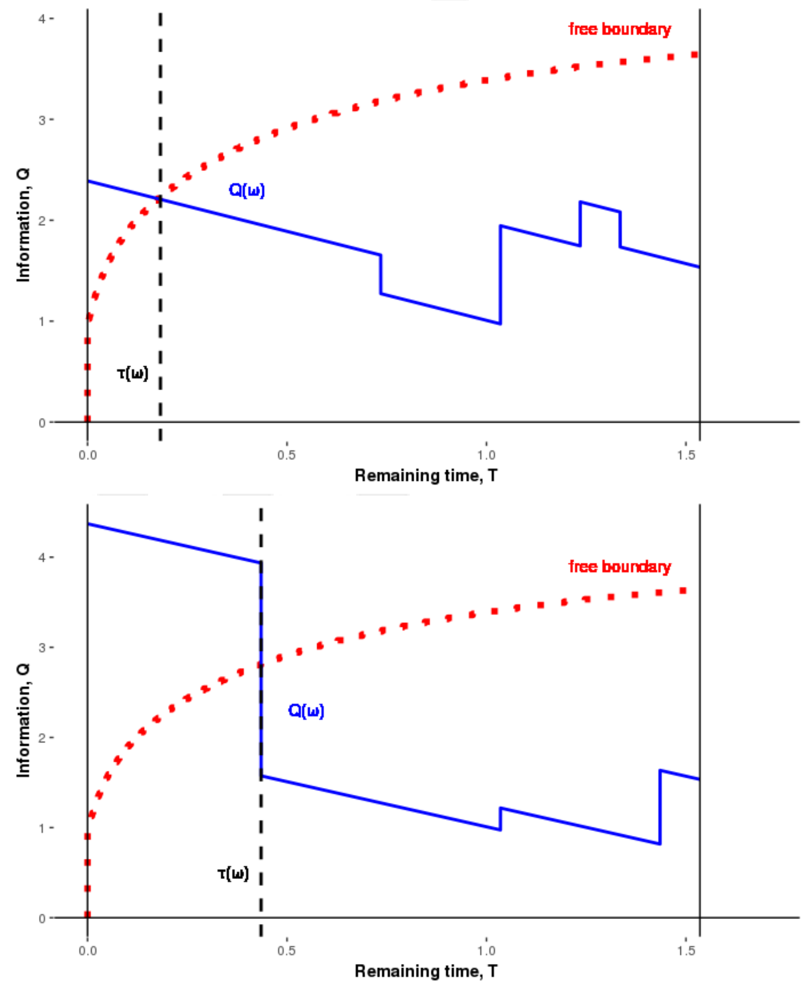
(a) $\lambda_0 > \lambda_1 = 1.5$

(b) $\lambda_0 = \lambda_1 = 3$

Figure 7.2: Sample-paths of the process Q.
$(T = 1.5351,\ \lambda_0 = 3,\ c = 1,\ \mu_1 = 1)$



(a) $\lambda_0 > \lambda_1 = 1.5$



(b) $\lambda_0 = \lambda_1 = 3$

Figure 7.3: Sample-paths of the process Q.
$(T = 1.5351,\ \lambda_0 = 3,\ c = 1,\ \mu_1 = 1)$

In Figures 7.2 and 7.3, we take $\pi_0 = 0$ and $\pi_0 = 0.5$, respectively. As a result, the sample-path realizations start at points $(T, 0)$ and (T, T) with different initial Q values. In sub-figures on the left side of the panels, the arrival rates λ_0 and λ_1 are not equal. Consequently, the process Q's motion forms curves. However, in sub-figures on the right side of the panels, the arrival rates λ_0 and λ_1 coincide and the process Q's motion forms straight lines. In the upper sub-figures of the panels, the process Q enters the stopping region during a deterministic motion, whereas in the lower sub-figures of the panels, the process Q enters the stopping region at a jump time.

A careful look at the process Q's trajectory in the sample-path realizations in Figures 7.2 and 7.3 reveals an interesting observation. Recall that Q moves in deterministic paths described in (3.10) between the jumps of the compound Poisson process X. At the jump times of X, the process Q also jumps and continues its motion on another deterministic path. When $\lambda_1 > \lambda_0$, we have $x(q, t) \to \frac{1}{\lambda_1 - \lambda_0}$ as $t \to \infty$. Thus, $\frac{1}{\lambda_1 - \lambda_0}$ becomes the mean-reversion level for Q. Although it may move away from this level with a jump depending on the mark observed, the process Q ultimately reverts to this mean-reversion level while following the integral curves as demonstrated in Figure 7.4. On the other hand, when $\lambda_1 \leq \lambda_0$, we have $x(q, t) \to \infty$ as $t \to \infty$. Hence, the process Q increases unboundedly between the jumps as depicted in Figure 7.5.

Interested in the effects of different problem parameters on the alarm time τ, we simulated the compound Poisson process X and recorded the mean and standard deviation values of the alarm time observations; i.e., the realizations of the random variable τ. The Tables 7.1, 7.2, and 7.3 summarize the results. Note that, we selected the time horizon T as the 25^{th}, 50^{th}, and 75^{th} quantiles of the exponential distribution with rate λ_0 in Tables 7.1, 7.2, and 7.3, respectively. Figures 7.6 and 7.6 provide a graphical representation of the results with box-plots. In the upper box-plots on the panels, the arrival rate $\lambda_0 = 1$, while in the lower box-plots on the panels $\lambda_0 = 10$. In each box-plot, the data are grouped according to the mean value μ_1. Moreover, for every μ_1 level, data are further grouped according to the arrival rate λ_1: in the box-plot on the left $\lambda_1 = \lambda_0/2$, in the box-plot in the middle $\lambda_1 = \lambda_0$ and in the box-plot on the right $\lambda_1 = 2\lambda_0$.

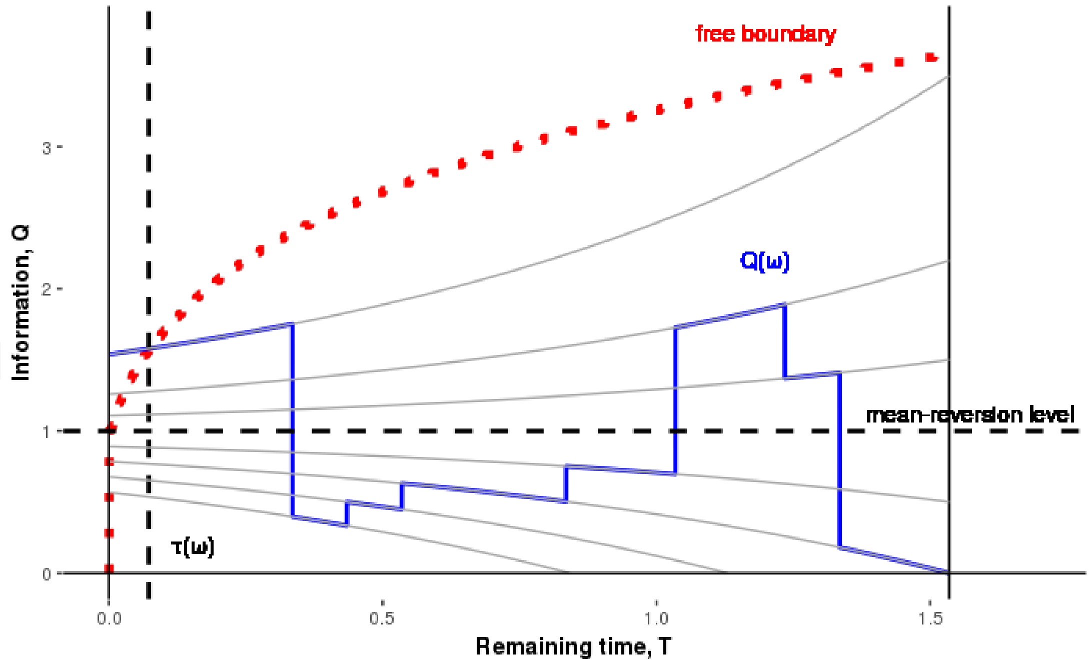


Figure 7.4: Process Q reverts to the mean-reversion level ($T = 1.5351,\ 3 = \lambda_0 < \lambda_1 = 4,\ c = 1,\ \mu_1 = 1$).

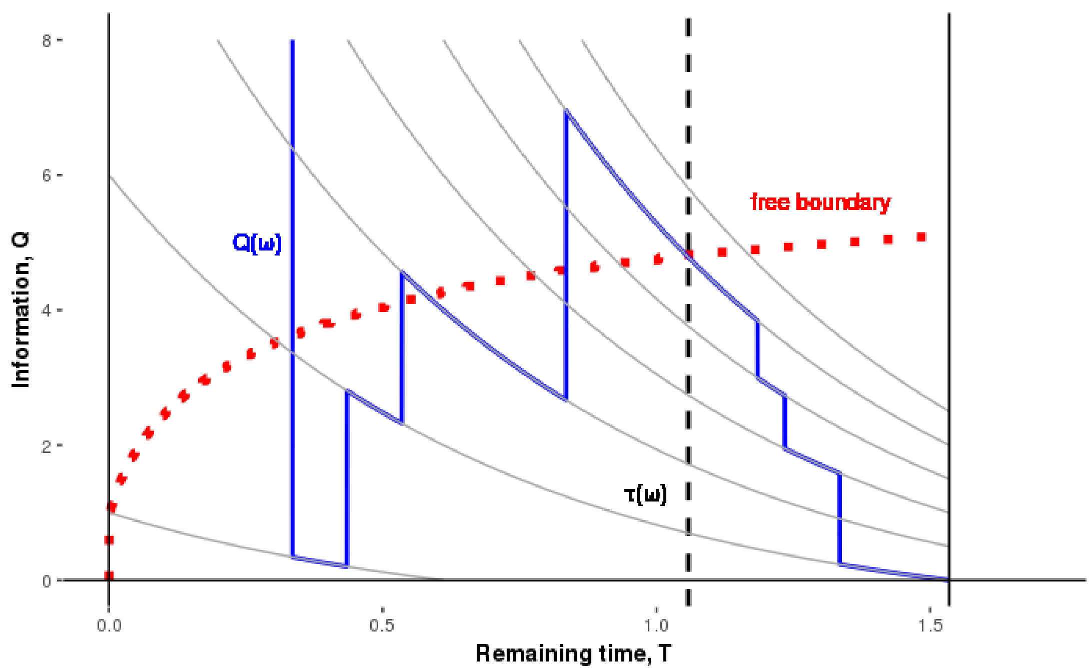


Figure 7.5: Process Q increases unboundedly ($T = 1.5351,\ 3 = \lambda_0 > \lambda_1 = 1.5,\ c = 1,\ \mu_1 = 1$).

Simulation results clearly suggest that when the cost parameter c increases, the expected alarm time decreases as waiting becomes more costly. However, the effects of the other problem parameters on the distribution of the alarm time are inconclusive and need further analysis.

#	T	c	λ_0	λ_1	μ_1	τ-mean	τ-sd
1	2.8768	1	1	0.5	0.2	1.17	0.36
2	2.8768	1	1	0.5	1	1.41	0.50
3	2.8768	1	1	0.5	5	2.14	0.60
4	2.8768	1	1	1	0.2	1.02	0.10
5	2.8768	1	1	1	1	2.61	0.39
6	2.8768	1	1	1	5	2.61	0.37
7	2.8768	1	1	2	0.2	1.55	0.75
8	2.8768	1	1	2	1	2.22	0.79
9	2.8768	1	1	2	5	2.87	0.09
10	0.2877	1	10	5	0.2	0.29	0.00
11	0.2877	1	10	5	1	0.29	0.01
12	0.2877	1	10	5	5	0.29	0.01
13	0.2877	1	10	10	0.2	0.29	0.00
14	0.2877	1	10	10	1	0.29	0.01
15	0.2877	1	10	10	5	0.29	0.00
16	0.2877	1	10	20	0.2	0.29	0.01
17	0.2877	1	10	20	1	0.28	0.02
18	0.2877	1	10	20	5	0.29	0.01
19	2.8768	10	1	0.5	0.2	0.11	0.01
20	2.8768	10	1	0.5	1	0.10	0.01
21	2.8768	10	1	0.5	5	0.11	0.02
22	2.8768	10	1	1	0.2	0.10	0.00
23	2.8768	10	1	1	1	0.10	0.01
24	2.8768	10	1	1	5	0.13	0.02
25	2.8768	10	1	2	0.2	0.10	0.01
26	2.8768	10	1	2	1	0.11	0.01
27	2.8768	10	1	2	5	0.14	0.03
28	0.2877	10	10	5	0.2	0.13	0.04
29	0.2877	10	10	5	1	0.14	0.05
30	0.2877	10	10	5	5	0.21	0.06
31	0.2877	10	10	10	0.2	0.10	0.01
32	0.2877	10	10	10	1	0.14	0.05
33	0.2877	10	10	10	5	0.27	0.03
34	0.2877	10	10	20	0.2	0.15	0.07
35	0.2877	10	10	20	1	0.21	0.08
36	0.2877	10	10	20	5	0.29	0.01

Table 7.1: Results on the alarm time

#	T	c	λ_0	λ_1	μ_1	τ-mean	τ-sd
37	6.9315	1	1	0.5	0.2	1.16	0.32
38	6.9315	1	1	0.5	1	1.78	0.77
39	6.9315	1	1	0.5	5	2.67	1.48
40	6.9315	1	1	1	0.2	1.06	0.11
41	6.9315	1	1	1	1	1.48	0.55
42	6.9315	1	1	1	5	6.03	1.20
43	6.9315	1	1	2	0.2	1.61	0.99
44	6.9315	1	1	2	1	3.59	2.04
45	6.9315	1	1	2	5	6.90	0.36
46	0.6931	1	10	5	0.2	0.68	0.04
47	0.6931	1	10	5	1	0.68	0.04
48	0.6931	1	10	5	5	0.69	0.01
49	0.6931	1	10	10	0.2	0.69	0.02
50	0.6931	1	10	10	1	0.65	0.10
51	0.6931	1	10	10	5	0.69	0.00
52	0.6931	1	10	20	0.2	0.66	0.09
53	0.6931	1	10	20	1	0.67	0.08
54	0.6931	1	10	20	5	0.69	0.01
55	6.9315	10	1	0.5	0.2	0.11	0.01
56	6.9315	10	1	0.5	1	0.11	0.01
57	6.9315	10	1	0.5	5	0.13	0.03
58	6.9315	10	1	1	0.2	0.10	0.00
59	6.9315	10	1	1	1	0.11	0.01
60	6.9315	10	1	1	5	0.19	0.05
61	6.9315	10	1	2	0.2	0.11	0.01
62	6.9315	10	1	2	1	0.11	0.01
63	6.9315	10	1	2	5	0.12	0.02
64	0.6931	10	10	5	0.2	0.13	0.04
65	0.6931	10	10	5	1	0.13	0.05
66	0.6931	10	10	5	5	0.29	0.16
67	0.6931	10	10	10	0.2	0.10	0.01
68	0.6931	10	10	10	1	0.15	0.06
69	0.6931	10	10	10	5	0.63	0.09
70	0.6931	10	10	20	0.2	0.16	0.10
71	0.6931	10	10	20	1	0.26	0.18
72	0.6931	10	10	20	5	0.69	0.03

Table 7.2: Results on the alarm time

#	T	c	λ_0	λ_1	μ_1	τ-mean	τ-sd
73	13.8629	1	1	0.5	0.2	1.55	0.48
74	13.8629	1	1	0.5	1	2.26	1.13
75	13.8629	1	1	0.5	5	3.89	2.53
76	13.8629	1	1	1	0.2	1.03	0.11
77	13.8629	1	1	1	1	1.29	0.47
78	13.8629	1	1	1	5	12.90	1.91
79	13.8629	1	1	2	0.2	1.63	1.00
80	13.8629	1	1	2	1	3.54	2.85
81	13.8629	1	1	2	5	13.81	0.60
82	1.3863	1	10	5	0.2	1.20	0.25
83	1.3863	1	10	5	1	1.24	0.25
84	1.3863	1	10	5	5	1.38	0.05
85	1.3863	1	10	10	0.2	1.08	0.24
86	1.3863	1	10	10	1	1.30	0.21
87	1.3863	1	10	10	5	1.38	0.06
88	1.3863	1	10	20	0.2	1.26	0.25
89	1.3863	1	10	20	1	1.28	0.25
90	1.3863	1	10	20	5	1.39	0.01
91	13.8629	10	1	0.5	0.2	0.11	0.01
92	13.8629	10	1	0.5	1	0.11	0.02
93	13.8629	10	1	0.5	5	0.12	0.02
94	13.8629	10	1	1	0.2	0.11	0.00
95	13.8629	10	1	1	1	0.11	0.01
96	13.8629	10	1	1	5	0.12	0.02
97	13.8629	10	1	2	0.2	0.11	0.01
98	13.8629	10	1	2	1	0.11	0.01
99	13.8629	10	1	2	5	0.13	0.02
100	1.3863	10	10	5	0.2	0.11	0.03
101	1.3863	10	10	5	1	0.13	0.05
102	1.3863	10	10	5	5	0.74	0.39
103	1.3863	10	10	10	0.2	0.10	0.01
104	1.3863	10	10	10	1	0.13	0.04
105	1.3863	10	10	10	5	0.97	0.36
106	1.3863	10	10	20	0.2	0.16	0.10
107	1.3863	10	10	20	1	0.31	0.24
108	1.3863	10	10	20	5	1.37	0.10

Table 7.3: Results on the alarm time

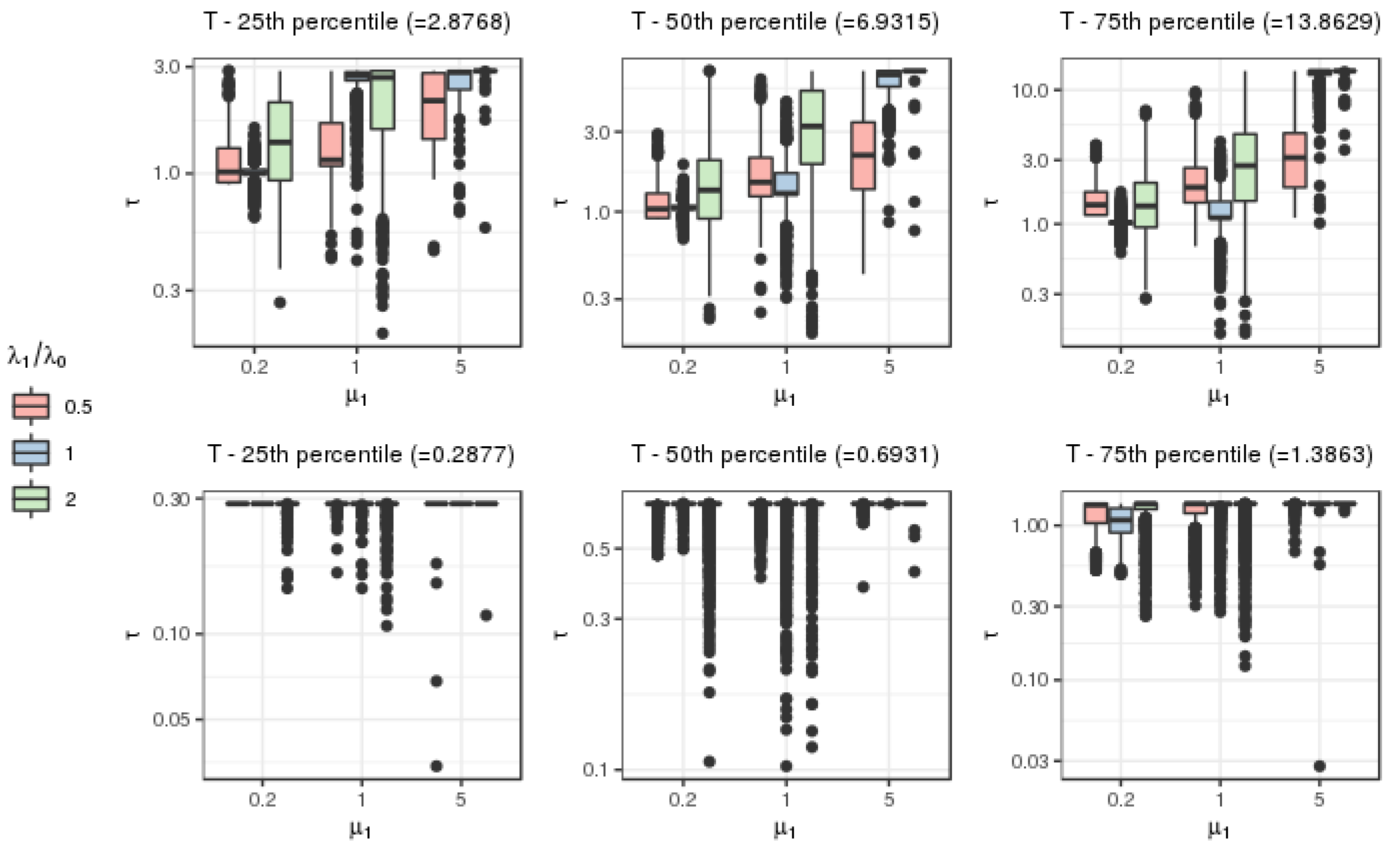


Figure 7.6: Box-plots on the alarm time (c = 1)
$\lambda_0 = 1$ on the first row, $\lambda_0 = 10$ on the second row.

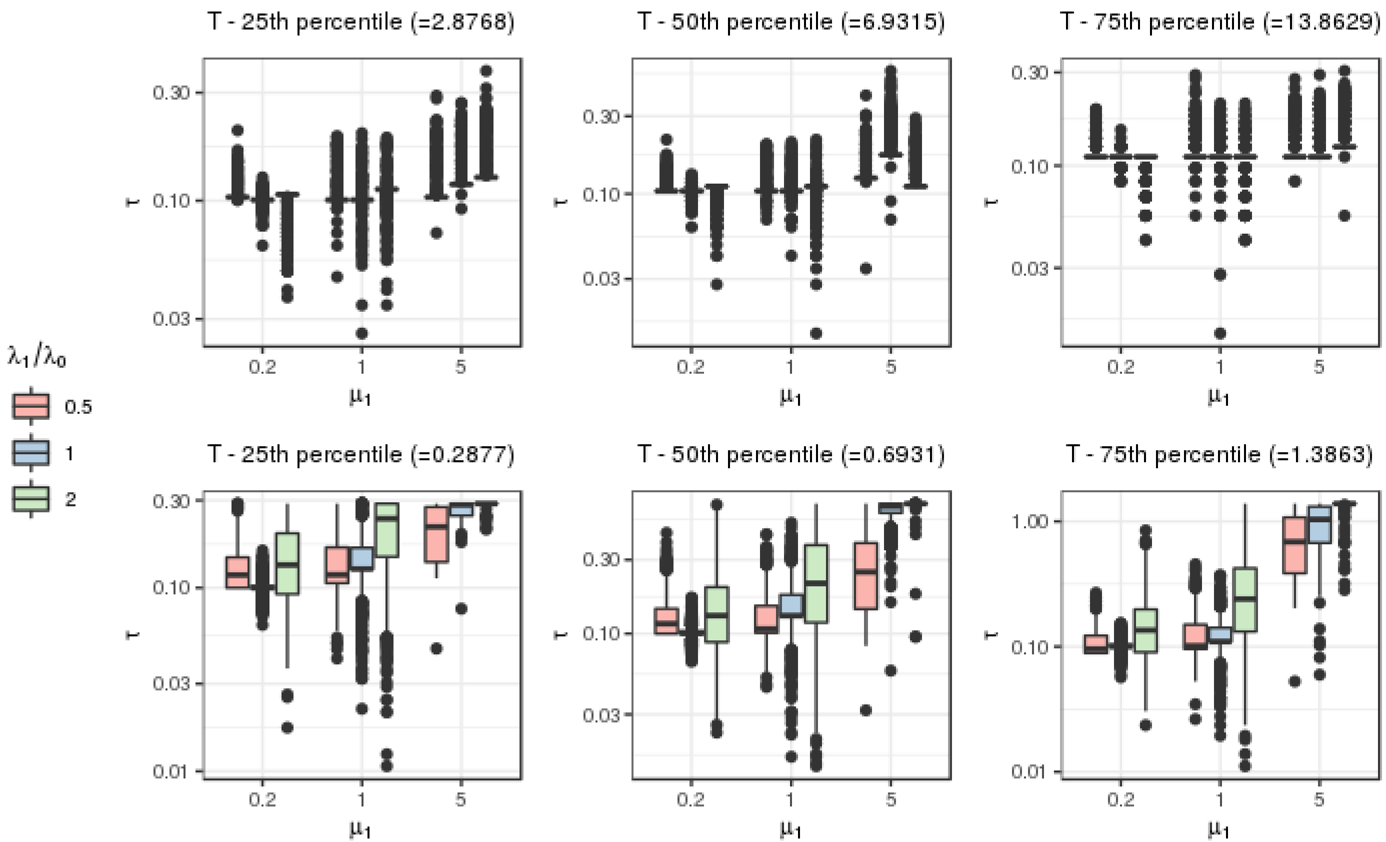


Figure 7.7: Box-plots on the alarm time (c = 10)
$\lambda_0 = 1$ on the first row, $\lambda_0 = 10$ on the second row.

Chapter 8

Conclusion

In this thesis, we revisit the compound Poisson disorder problem in which decision maker aims to detect a change in the probability law of a compound Poisson process. In practice, the identification of a quickest detection rule may help with mitigating the detrimental effects of an unfavorable regime change. Hence, detecting the change-point as soon as it occurs is an important task.

Dayanik and Sezer [10] completely solve the compound Poisson disorder problem assuming a change-point with an exponential prior distribution. Although having an exponentially distributed change-point facilitates the analysis, the memoryless property of the exponential distribution prevents an accurate representation of reality. Moreover, although the decision makers often have no prior knowledge of the change-point, the exponential prior conveys information about the change-point as an informative prior. Those limitations suggest the use of uninformative priors which also have a history. Therefore, we consider a uniformly distributed change-point in our study.

We reformulate the quickest detection problem as a finite-horizon optimal stopping problem for a piecewise-deterministic Markov process. Firstly, we approach the optimal stopping problem with the method of variational inequalities. Although solving the problem is possible with this method, it is difficult because of the jumping sufficient statistic and the free-boundary in time-space domain. Since the sufficient statistic has piecewise-deterministic sample-paths, we are able solve the optimal stopping problem with the successive approximations of the value function.

It is interesting to see that the compound Poisson disorder problem with uniformly distributed disorder time admits a one-dimensional Markovian sufficient statistic. Because when the change-point is not exponentially distributed, this is not the case. As another notable finding, the dynamics of sufficient statistic happened to be completely free of the time horizon T.

For future work, we can obtain accurate threshold levels and bounds on the quickest detection rule. The numerical calculation of the value function can be improved in various ways. Instead of calculating the value function on grid points and interpolating the points in between, we can use linear regression and basis functions to obtain it. Furthermore, we may extract several features from the problem structure as we did in Lemma 5.4.

Future research directions may include studying the compound Poisson disorder problem under other uninformative priors, such as the Jeffrey's prior. Note that the uniform prior is not the only way to describe the absence of any prior knowledge of the change-point. In fact, there are other priors that can be used for modeling an uninformative change-time which are also invariant to reparametrization unlike the uniform prior. Moreover, we assume that the arrival rate and the jump distribution of the compound Poisson process before and after the change-point to be known in advance. However, one can work on both the detection of the change-point and the identification of the new regime after the change. Perhaps, one can begin with introducing a simple Bernoulli distribution for the mark distribution after the change-time. Hence, making the jump distribution after the change random could lead to another possible research path. In our study, we assume a linear detection delay cost, but one can also consider an exponential penalty for the detection delay. It would be interesting to see how our problem changes in that case.

Bibliography

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

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

[3] G. Peskir and A. Shiryaev, *Optimal Stopping and Free-Boundary Problems*. Birkhäuser Basel, 01 2006.

[4] L. Gal'chuk and B. Rozovskii, "The 'disorder' problem for a Poisson process," *Theory of Probability & Its Applications*, vol. 16, no. 4, pp. 712–716, 1971.

[5] M. Davis, "A note on the Poisson disorder problem," *Banach Center Publications*, vol. 1, no. 1, pp. 65–72, 1976.

[6] G. Peskir and A. N. Shiryaev, *Solving the Poisson disorder problem*, pp. 295–312. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[7] E. Bayraktar and S. Dayanik, "Poisson disorder problem with exponential penalty for delay," *Mathematics of Operations Research*, vol. 31, no. 2, pp. 217–233, 2006.

[8] E. Bayraktar, S. Dayanik, and I. Karatzas, "The standard Poisson disorder problem revisited," *Stochastic Processes and their Applications*, vol. 115, no. 9, pp. 1437 – 1450, 2005.

[9] P. V. Gapeev, "The disorder problem for compound Poisson processes with exponential jumps," *Ann. Appl. Probab.*, vol. 15, pp. 487–499, 02 2005.

[10] S. Dayanik and S. O. Sezer, “Compound Poisson disorder problem,” *Mathematics of Operations Research*, vol. 31, no. 4, pp. 649–672, 2006.

[11] U. Gugerli, “Optimal stopping of a piecewise-deterministic Markov process,” *Stochastics*, vol. 19, no. 4, pp. 221–236, 1986.

[12] M. Davis, “Markov Models and Optimization,” *Monographs on Statistics and Applied Probability*, 1993.

[13] S. O. Sezer, “On the Wiener disorder problem,” *The Annals of Applied Probability*, vol. 20, no. 4, pp. 1537–1566, 2010.

[14] E. Bayraktar and S. Sezer, “Online change detection for a Poisson process with a phase-type change-time prior distribution,” *Sequential Analysis*, vol. 28, no. 2, pp. 218–250, 2009.

[15] M. Baron and A. G. Tartakovsky, “Asymptotic optimality of change-point detection schemes in general continuous-time models,” *Sequential Analysis*, vol. 25, no. 3, pp. 257–296, 2006.

[16] A. Greenbaum and T. P. Chartier, *Numerical Methods: Design, Analysis, and Computer Implementation of Algorithms.* Princeton University Press, 2012.

[17] A. N. Shiryaev, *Probability.* Springer-Verlag, 1996.

[18] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*, vol. 50. Springer, 1981.

Appendix A

Calculations

A.1 Reformulation of the Bayes risk

Recall that the Bayes risk consists of two parts: the probability of a false alarm and the expected detection delay cost. Let us work on those parts separately. Note that the expectation $\mathbb{E}_0$ is with respect to the reference probability measure $\mathbb{P}_0$.

False-alarm frequency

$$
\begin{aligned}
\mathbb{P}\{\tau < \theta\} &= \mathbb{E}[1_{\{\tau<\theta\}}] = \mathbb{E}_0[Z_{\tau\wedge\theta}1_{\{\tau<\theta\}}] = \mathbb{E}_0[Z_\tau 1_{\{\tau<\theta\}}] \\
&= \mathbb{P}_0\{\tau < \theta\} = 1 - \mathbb{P}_0\{\tau \geq \theta\} \\
&= 1 - \pi - \frac{(1-\pi)}{T}\mathbb{E}_0[\tau].
\end{aligned}
$$

Expected detection delay

Before calculating the expected detection delay, we will present and prove the following proposition; see generalized Bayes theorem (Shiryev [17], pp. 230-231).

Proposition A.1. *For a $\mathcal{G}_t$-measurable integrable random variable U, we have*

$$\mathbb{E}[U|\mathcal{F}_t] = \frac{\mathbb{E}_0[Z_t U|\mathcal{F}_t]}{\mathbb{E}_0[Z_t|\mathcal{F}_t]} \quad \mathbb{P}\text{-}a.s.$$

Proof. For every integrable $V \in \mathcal{F}_t$, we have

$$\begin{aligned}
\mathbb{E}\big[\mathbb{E}[U \mid \mathcal{F}_t]V\big] &= \mathbb{E}[UV] = \mathbb{E}_0[Z_t UV] = \mathbb{E}_0\big[\mathbb{E}_0[Z_t U \mid \mathcal{F}_t]V\big] \\
&= \mathbb{E}_0\left[\mathbb{E}_0[Z_t \mid \mathcal{F}_t]\frac{\mathbb{E}_0[Z_t U \mid \mathcal{F}_t]}{\mathbb{E}_0[Z_t \mid \mathcal{F}_t]}V\right] \\
&= \mathbb{E}_0\left[Z_t\frac{\mathbb{E}_0[Z_t U \mid \mathcal{F}_t]}{\mathbb{E}_0[Z_t \mid \mathcal{F}_t]}V\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}_0[Z_t U \mid \mathcal{F}_t]}{\mathbb{E}_0[Z_t \mid \mathcal{F}_t]}V\right].
\end{aligned}$$

□

Proposition (A.1) implies that

$$\Phi_t = \frac{\mathbb{E}[1_{\{\theta \le t\}}|\mathcal{F}_t]}{\mathbb{E}[1_{\{\theta > t\}}|\mathcal{F}_t]} = \frac{\mathbb{E}_0[Z_t 1_{\{\theta \le t\}}|\mathcal{F}_t]}{\mathbb{E}_0[Z_t|\mathcal{F}_t]}\left(\frac{\mathbb{E}_0[Z_t 1_{\{\theta > t\}}|\mathcal{F}_t]}{\mathbb{E}_0[Z_t|\mathcal{F}_t]}\right)^{-1}.$$

Since τ and θ are $\mathcal{F}_t$-measurable and $\mathcal{G}_t$-measurable, respectively

$$\begin{aligned}
\mathbb{E}[(\tau-\theta)^+] &= \mathbb{E}\left[1_{\{\tau\ge\theta\}}\int_\theta^\tau dt\right] = \mathbb{E}\left[\int_0^T 1_{\{\tau>t\}}1_{\{\theta\le t\}}dt\right] = \int_0^T \mathbb{E}\left[1_{\{\tau>t\}}1_{\{\theta\le t\}}\right]dt \\
&= \int_0^T \mathbb{E}_0\left[Z_t 1_{\{\tau>t\}}1_{\{\theta\le t\}}\right]dt = \int_0^T \mathbb{E}_0\left[1_{\{\tau>t\}}\mathbb{E}_0[Z_t 1_{\{\theta\le t\}}|\mathcal{F}_t]\right]dt \\
&= \int_0^T \mathbb{E}_0\left[1_{\{\tau>t\}}\Phi_t\mathbb{E}_0[Z_t 1_{\{\theta> t\}}|\mathcal{F}_t]\right]dt = \int_0^T \mathbb{E}_0\left[1_{\{\tau>t\}}\Phi_t\mathbb{P}_0\{\theta>t\}\right]dt \\
&= (1-\pi)\mathbb{E}_0\left[\int_0^{T\wedge\tau}\left(\frac{T-t}{T}\right)\Phi_t dt\right] = (1-\pi)\mathbb{E}_0\left[\int_0^{\tau}\left(\frac{T-t}{T}\right)\Phi_t dt\right].
\end{aligned}$$

Hence, when we combine the two components the Bayes risk is given by

$$\begin{aligned}
R_\tau(\pi) &= \underbrace{1-\pi-\frac{(1-\pi)}{T}\mathbb{E}_0[\tau]}_{\text{false alarm frequency}} + \underbrace{c(1-\pi)\mathbb{E}_0\left[\int_0^\tau\left(\frac{T-t}{T}\right)\Phi_t dt\right]}_{\text{expected detection delay cost}} \\
&= 1-\pi+\frac{c(1-\pi)}{T}\mathbb{E}_0\left[\int_0^\tau\left((T-t)\Phi_t-\frac{1}{c}\right)dt\right].
\end{aligned}$$

A.2 Dynamics of the sufficient statistic

A.2.1 Dynamics of Φ and D

Firstly, let us show the dynamics of the likelihood ratio process R given in (3.6). Observe that, at every time t between any two consecutive jumps σ_n and σ_{n+1} of the process X, we have

$$\frac{R_t}{R_{\sigma_n}} = \frac{e^{-(\lambda_1-\lambda_0)t}\prod_{k=1}^{N_t}\left[\frac{\lambda_1}{\lambda_0}f(Y_k)\right]}{e^{-(\lambda_1-\lambda_0)\sigma_n}\prod_{k=1}^{N_{\sigma_n}}\left[\frac{\lambda_1}{\lambda_0}f(Y_k)\right]} = \frac{e^{-(\lambda_1-\lambda_0)t}}{e^{-(\lambda_1-\lambda_0)\sigma_n}}, \quad \sigma_n \le t < \sigma_{n+1},$$

because $N_t = N_{\sigma_n} = n$ which implies

$$R_t = R_{\sigma_n}e^{-(\lambda_1-\lambda_0)(t-\sigma_n)},$$

and thus

$$dR_t = -(\lambda_1-\lambda_0)R_t dt, \quad \sigma_n \le t < \sigma_{n+1}.$$

At the jump time σ_{n+1} of the process X,

$$dR_{\sigma_{n+1}} = R_{\sigma_{n+1}} - R_{\sigma_{(n+1)-}} = R_{\sigma_{(n+1)-}}\left(\frac{\lambda_1}{\lambda_0}f(Y_{n+1}) - 1\right).$$

Therefore, the dynamics of the process R between the jumps and at the jump times are given by

$$dR_t = \begin{cases} -(\lambda_1-\lambda_0)R_t dt, & t \in [\sigma_n,\, \sigma_{n+1}), \\ R_{t-}\left(\frac{\lambda_1}{\lambda_0}f(Y_{n+1}) - 1\right), & t = \sigma_{n+1}. \end{cases} \tag{A.1}$$

Since the random variable θ is independent of the process X and has the uniform distribution in (2.2) under $\mathbb{P}_0$, we can write Φ_t as

$$\begin{aligned}\Phi_t &= \frac{\mathbb{E}_0[Z_t 1_{\{\theta\le t\}}|\mathcal{F}_t]}{\mathbb{E}_0[Z_t|\mathcal{F}_t]}\left(\frac{\mathbb{E}_0[Z_t 1_{\{\theta> t\}}|\mathcal{F}_t]}{\mathbb{E}_0[Z_t|\mathcal{F}_t]}\right)^{-1} \\ &= \frac{\mathbb{E}_0[Z_t 1_{\{\theta\le t\}}|\mathcal{F}_t]}{\mathbb{E}_0[Z_t 1_{\{\theta> t\}}|\mathcal{F}_t]} \\ &= \frac{\mathbb{E}_0[1_{\{\theta\le t\}}\frac{R_t}{R_\theta}|\mathcal{F}_t]}{\mathbb{P}_0\{\theta > t|\mathcal{F}_t\}}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\pi R_t + (1-\pi)\int_0^t \frac{R_t}{R_s}\frac{1}{T}ds}{(1-\pi)\left(\frac{T-t}{T}\right)} \\
&= \frac{R_t}{T-t}\left(T\frac{\pi}{1-\pi} + \int_0^t \frac{1}{R_s}ds\right) \\
&\equiv W_t V_t
\end{aligned}
$$

in terms of

$$
W_t = \frac{R_t}{T-t} \quad \text{and} \quad V_t = T\frac{\pi}{1-\pi} + \int_0^t \frac{1}{R_s}ds.
$$

The product rule for the finite-variation process is

$$
d\Phi_t = dW_t V_t + W_t dV_t.
$$

Therefore, between jumps σ_n and σ_{n+1} of the process X we have

$$
\begin{aligned}
d\Phi_t &= \frac{dR_t(T-t) + R_t dt}{(T-t)^2}V_t + W_t\frac{1}{R_t}dt \\
&= -(\lambda_1 - \lambda_0)\Phi_t dt + \frac{1}{T-t}\Phi_t dt + \frac{1}{T-t}dt \\
&= \left[\frac{1}{T-t} + \left(\frac{1}{T-t} - \lambda_1 + \lambda_0\right)\Phi_t\right]dt, \quad \sigma_n \le t < \sigma_{n+1},
\end{aligned}
$$

and at times $t = \sigma_n$ where the process X makes a jump we have

$$
\begin{aligned}
d\Phi_t &= \Phi_t - \Phi_{t-} \\
&= \underbrace{(R_t - R_{t-})}_{dR_t}\frac{1}{T-t}V_t \\
&= R_{t-}\left(\frac{\lambda_1}{\lambda_0}f(Y_{n+1}) - 1\right)\frac{1}{T-t}V_t \\
&= \Phi_{t-}\left(\frac{\lambda_1}{\lambda_0}f(Y_{n+1}) - 1\right).
\end{aligned}
$$

Replacing $\frac{1}{T-t}$ with D_t, we have the dynamics of the odds-ratio process Φ as

$$
d\Phi_t = \begin{cases} [D_t + (D_t - \lambda_1 + \lambda_0)\Phi_t]dt, & t \in [\sigma_n, \sigma_{n+1}), \\ \Phi_{t-}\left(\frac{\lambda_1}{\lambda_0}f(Y_{n+1}) - 1\right), & t = \sigma_{n+1}. \end{cases} \tag{A.2}
$$

The dynamics of the deterministic process D is simply given by the differential equation

$$
dD_t = \frac{dt}{(T-t)^2} = D_t^2 dt. \tag{A.3}
$$

A.2.2 Dynamics of Q

The dynamics of the sufficient statistic Q are given as follows. For $t \in [\sigma_n, \sigma_{n+1})$,

$$\begin{aligned}
dQ_t &= \frac{d\Phi_t}{D_t} - \frac{\Phi_t}{D_t^2} dD_t \\
&= \frac{1}{D_t}[D_t + (D_t - \lambda_1 + \lambda_0)\Phi_t]dt - \frac{\Phi_t}{D_t^2} D_t^2 dt \\
&= \left[1 + (D_t - \lambda_1 + \lambda_0)\frac{\Phi_t}{D_t}\right] dt - \Phi_t dt \\
&= \left[1 + \Phi_t - (\lambda_1 - \lambda_0)\frac{\Phi_t}{D_t} - \Phi_t\right] dt \\
&= \left[1 - (\lambda_1 - \lambda_0)\frac{\Phi_t}{D_t}\right] dt \\
&= [1 - (\lambda_1 - \lambda_0)Q_t]dt,
\end{aligned}$$

and for $t = \sigma_{n+1}$,

$$\begin{aligned}
dQ_t = \Delta Q_t &= Q_t - Q_{t-} \\
&= \frac{\Phi_t}{D_t} - \frac{\Phi_{t-}}{D_t} \\
&= \frac{1}{D_t}(\Phi_t - \Phi_{t-}) = \frac{d\Phi_t}{D_t} \\
&= \frac{\Phi_{t-}}{D_{t-}}\left[\frac{\lambda_1}{\lambda_0} f(Y_{n+1}) - 1\right] \\
&= Q_{t-}\left[\frac{\lambda_1}{\lambda_0} f(Y_{n+1}) - 1\right].
\end{aligned}$$

In a more compact form, the dynamics of the process Q are given by

$$dQ_t = \begin{cases} [1 - (\lambda_1 - \lambda_0)Q_t]dt, & t \in [\sigma_n, \sigma_{n+1}), \\ Q_{t-}\left(\frac{\lambda_1}{\lambda_0} f(Y_{n+1}) - 1\right), & t = \sigma_{n+1}. \end{cases} \tag{A.4}$$

Note that the process Q is the solution of the differential equation

$$dQ_t = [1 - (\lambda_1 - \lambda_0)Q_t]dt + Q_{t-}\int_{\mathbb{R}^d}\left(\frac{\lambda_1}{\lambda_0} f(y) - 1\right) p_0(dt, dy). \tag{A.5}$$

This suggests that the process Q is an autonomous process: it is a piecewise-deterministic Markov process driven by the point process $\{p_0((0,t] \times A),\ t \geq 0\}$, which has $\mathbb{P}_0$-mean measure $\lambda_0 t \nu_0(A)$. Between the jumps of the process X, the process Q follows the integral curves of the differential equation in (A.5). At every jump of X, the process Q is updated instantaneously.

By (A.4), for $\sigma_n \le t < \sigma_{n+1}$ we have

$$\frac{dQ_t}{dt} = 1 - (\lambda_1 - \lambda_0) Q_t,$$

which implies

$$Q_t' + (\lambda_1 - \lambda_0) Q_t = 1.$$

Solving this differential equation, we have

$$d\left(Q_t e^{(\lambda_1 - \lambda_0)t}\right) = e^{(\lambda_1 - \lambda_0)t} dt.$$

Integrating both sides on $[\sigma_n, t]$ gives

$$Q_t e^{(\lambda_1-\lambda_0)t} - Q_{\sigma_n} e^{(\lambda_1-\lambda_0)\sigma_n} = \begin{cases} \frac{1}{\lambda_1-\lambda_0}\left(e^{(\lambda_1-\lambda_0)t} - e^{(\lambda_1-\lambda_0)\sigma_n}\right), & \lambda_1 \neq \lambda_0, \\ t - \sigma_n, & \lambda_1 = \lambda_0. \end{cases}$$

Then, Q_t is given by

$$Q_t = \begin{cases} \frac{1}{\lambda_1-\lambda_0} + \left(Q_{\sigma_n} - \frac{1}{\lambda_1-\lambda_0}\right) e^{-(\lambda_1-\lambda_0)(t-\sigma_n)}, & \lambda_1 \neq \lambda_0, \\ Q_{\sigma_n} + t - \sigma_n, & \lambda_1 = \lambda_0. \end{cases} \tag{A.6}$$

Likewise, for $t = \sigma_{n+1}$ we have

$$Q_t - Q_{t-} \equiv dQ_t = Q_{t-}\left(\frac{\lambda_1}{\lambda_0} f(Y_{n+1}) - 1\right),$$

which implies

$$Q_t = Q_{t-} \frac{\lambda_1}{\lambda_0} f(Y_{n+1}).$$

Hence, $Q_{\sigma_{n+1}}$ is given by

$$\begin{aligned} Q_{\sigma_{n+1}} &= Q_{\sigma_{(n+1)-}} \frac{\lambda_1}{\lambda_0} f(Y_{n+1}) \\ &= \begin{cases} \frac{\lambda_1}{\lambda_0} f(Y_{n+1}) \left[\frac{1}{\lambda_1-\lambda_0} + \left(Q_{\sigma_n} - \frac{1}{\lambda_1-\lambda_0}\right) e^{-(\lambda_1-\lambda_0)(\sigma_{n+1}-\sigma_n)}\right], & \lambda_1 \neq \lambda_0, \\ \frac{\lambda_1}{\lambda_0} f(Y_{n+1}) \left[Q_{\sigma_n} + \sigma_{n+1} - \sigma_n\right], & \lambda_1 = \lambda_0. \end{cases} \end{aligned} \tag{A.7}$$

Appendix B

Variational inequalities

Firstly, note that Q_t can be written as

$$Q_t = Q_{t_0} + \int_{t_0}^{t} [1 - (\lambda_1 - \lambda_0) Q_s] ds + \int_{t_0}^{t} \int_{\mathbb{R}^d} Q_{s-} \left(\frac{\lambda_1}{\lambda_0} f(y) - 1 \right) p_0(ds, dy), \quad \text{(B.1)}$$

for $0 \leq t_0 < t$ by (A.5).

If we have not stopped already, we can either stop immediately or wait for infinitesimally short h units of time and act optimally afterwards. In the first case, taking $\tau = 0$ results in

$$\mathbb{E}_{0,q} \left[\int_0^{\tau=0} g(Q_t) dt \right] = 0.$$

Taking the infimum shows $U(T, q) \leq 0$ must hold. For the latter case, we split the time interval $[0, \tau]$ into two and obtain

$$\begin{aligned} U(T, q) &\leq \mathbb{E}_{0,q} \left[\int_0^{\tau} g(Q_t) dt \right] \\ &= \mathbb{E}_{0,q} \left[\int_0^{h} g(Q_t) dt \right] + \mathbb{E}_{0,q} \left[\int_h^{h+\tau^* \circ \theta_h} g(Q_t) dt \right], \end{aligned} \quad \text{(B.2)}$$

where θ is the shift-operator on $\Omega : X_t \circ \theta_h = X_{h+t}$, and $\tau^* \in \mathcal{S}_{[h,T]}$ is an optimal stopping time in $T - h$ unit time starting at Q_h.

In the time interval $(0, h]$, there is a single arrival with probability $\lambda_0 h$ and no arrival with probability $1 - \lambda_0 h$. Hence, we can write the first expectation in

(B.2) as

$$\mathbb{E}_{0,q}\left[\int_0^h g(Q_t)dt\right] = \left[\underbrace{(1-\lambda_0 h)g(Q_0)}_{\text{no jump in}(0,h]} + \underbrace{(\lambda_0 h)\int_{\mathbb{R}^d} g\left(Q_0\frac{\lambda_1}{\lambda_0}f(y)\right)\nu_0(dy)}_{\text{one jump in }(0,h]}\right] h$$
$$+ o(h)$$
$$= g(Q_0)h + o(h),$$

and the second expectation in (B.2) as

$$\begin{aligned}
\mathbb{E}_{0,q}\left[\int_h^{h+\tau^*\circ\theta_h} g(Q_t)dt\right] &= \mathbb{E}_{0,q}\left[\int_0^{\tau^*\circ\theta_h}\left(Q_{t+h}-\frac{1}{c}\right)dt\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{\tau^*\circ\theta_h}\left(Q_t\circ\theta_h-\frac{1}{c}\right)dt\right] \\
&= \mathbb{E}_{0,q}\left[\left\{\int_0^{\tau^*}\left(Q_t-\frac{1}{c}\right)dt\right\}\circ\theta_h\right] \\
&= \mathbb{E}_{0,q}\left(\mathbb{E}_{0,q}\left[\left\{\int_0^{\tau^*}\left(Q_t-\frac{1}{c}\right)dt\right\}\circ\theta_h\middle|\mathcal{F}_h\right]\right) \\
&= \mathbb{E}_{0,q}\left(\mathbb{E}_{0,q}\left[\int_0^{\tau^*}\left(Q_t-\frac{1}{c}\right)dt\right]\Bigg|_{q=Q_h}\right) \\
&= \mathbb{E}_{0,q}[U(T-h,Q_h)],
\end{aligned}$$

as a result of the Markov property of Q at time h. Combining those two expectations, we have

$$U(T,q) \le g(Q_0)h + \mathbb{E}_{0,q}[U(T-h,Q_h)] + o(h).$$

Now, let us work on the expectation $\mathbb{E}_{0,q}[U(T-h,Q_h)]$. Assuming that U is continuously differentiable, for $t \in [\sigma_n, \sigma_{n+1})$

$$\begin{aligned}
dU(T-t,Q_t) &= [-U_1(T-t,Q_t)dt + U_2(T-t,Q_t)dQ_t] \\
&= \Big\{[1-(\lambda_1-\lambda_0)Q_t]U_2(T-t,Q_t) - U_1(T-t,Q_t)\Big\}dt,
\end{aligned}$$

where $U_1(\cdot,\cdot)$ and $U_2(\cdot,\cdot)$ are the partial derivatives of $U(\cdot,\cdot)$ with respect to first and second arguments, respectively. For $t = \sigma_{n+1}$

$$
\begin{aligned}
&dU(T-t, Q_t) = U(T-t, Q_t) - U(T-t, Q_{t-}) \\
&= \int_{\mathbb{R}^d} U\left(T-t, Q_{t-} + Q_{t-}\left[\frac{\lambda_1}{\lambda_0} f(y) - 1\right]\right) p_0(dt, dy) - U(T-t, Q_{t-}) \\
&= \int_{\mathbb{R}^d} U\left(T-t, Q_{t-}\frac{\lambda_1}{\lambda_0} f(y)\right) p_0(dt, dy) - U(T-t, Q_{t-}) \\
&= \int_{\mathbb{R}^d} \left[U\left(T-t, Q_{t-}\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T-t, Q_{t-})\right] p_0(dt, dy).
\end{aligned}
$$

Therefore, for $0 \leq t \leq T$

$$
\begin{aligned}
dU(T-t, Q_t) = &\Big\{[1 - (\lambda_1 - \lambda_0)Q_t]U_2(T-t, Q_t) - U_1(T-t, Q_t)\Big\}dt \\
&+ \int_{\mathbb{R}^d} \left[U\left(T-t, Q_{t-}\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T-t, Q_{t-})\right] p_0(dt, dy).
\end{aligned}
$$

So, for $0 \leq t_0 < t$

$$
\begin{aligned}
U(T-t, Q_t) = &\, U(T-t_0, Q_{t_0}) \\
&+ \int_{t_0}^{t} \Big\{[1 - (\lambda_1 - \lambda_0)Q_s]U_2(T-s, Q_s) - U_1(T-s, Q_s)\Big\}ds \\
&+ \int_{t_0}^{t} \int_{\mathbb{R}^d} \left[U\left(T-s, Q_{s-}\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T-s, Q_{s-})\right] p_0(ds, dy).
\end{aligned}
$$

When $t_0 = 0 < t = h$ and $q = Q_0$

$$
\begin{aligned}
&U(T-h, Q_h) = U(T, q) \\
&+ \int_0^h \Big\{[1 - (\lambda_1 - \lambda_0)Q_t]U_2(T-t, Q_t) - U_1(T-t, Q_t)\Big\}dt \\
&+ \int_0^h \int_{\mathbb{R}^d} \left[U\left(T-t, Q_{t-}\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T-t, Q_{t-})\right] \\
&\times \Big[\underbrace{p_0(dt, dy) - \lambda_0 dt \nu_0(dy)}_{\equiv \widetilde{p_0}(dt, dy)} + \lambda_0 dt \nu_0(dy)\Big] \\
&= U(T, q) + \int_0^h \Bigg\{[1 - (\lambda_1 - \lambda_0)Q_t]U_2(T-t, Q_t) - U_1(T-t, Q_t) \\
&+ \int_{\mathbb{R}^d} \left[U\left(T-t, Q_{t-}\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T-t, Q_{t-})\right] \lambda_0 \nu_0(dy)\Bigg\}dt \\
&+ \underbrace{\int_0^h \int_{\mathbb{R}^d} \left[U\left(T-t, Q_{t-}\frac{\lambda_1}{\lambda_0} f(y)\right) - U(T-t, Q_{t-})\right] \widetilde{p_0}(dt, dy)}_{\equiv M_h^u}.
\end{aligned}
$$

Observe that M_h^u is a martingale. Let us define the jump difference operator,

$$(DU)(T-t,Q_{t_-}) \triangleq \int_{\mathbb{R}^d} \left[U\left(T-t,Q_{t-}\frac{\lambda_1}{\lambda_0}f(y)\right) - U(T-t,Q_{t-})\right]\lambda_0\nu_0(dy). \tag{B.3}$$

Then

$$\begin{aligned} U(T-h,Q_h) = U(T,q) + \int_0^h \Big\{&[1-(\lambda_1-\lambda_0)Q_t]U_2(T-t,Q_t) \\ &- U_1(T-t,Q_t) + (DU)(T-t,Q_{t_-})\Big\}dt + M_h^u. \end{aligned}$$

When we take the expectation, the martingale term M_h^u disappears and we get

$$\begin{aligned} \mathbb{E}_{0,q}[U(T-h,Q_h)] = U(T,q) + \mathbb{E}_{0,q}\Bigg[\int_0^h \Big\{&[1-(\lambda_1-\lambda_0)Q_t]U_2(T-t,Q_t) \\ &- U_1(T-t,Q_t) + (DU)(T-t,Q_{t_-})\Big\}dt\Bigg]. \end{aligned}$$

Thus, when we go back to the second variational inequality in (B.2)

$$\begin{aligned} U(T,q) \le{}& g(q)h + U(T,q) + \mathbb{E}_{0,q}\Bigg[\int_0^h \Big\{[1-(\lambda_1-\lambda_0)Q_t]U_2(T-t,Q_t) \\ &- U_1(T-t,Q_t) + (DU)(T-t,Q_{t_-})\Big\}dt\Bigg] + o(h) \\ ={}& g(q)h + U(T,q) + \mathbb{E}_{0,q}\Bigg[\Big\{[1-(\lambda_1-\lambda_0)q]U_2(T,q) \\ &- U_1(T,q) + (DU)(T,q)\Big\}h + o(h)\Bigg] + o(h) \\ ={}& g(q)h + U(T,q) + \Big\{[1-(\lambda_1-\lambda_0)q]U_2(T,q) \\ &- U_1(T,q) + (DU)(T,q)\Big\}h + o(h) \\ 0 \le{}& \Bigg\{g(q) + [1-(\lambda_1-\lambda_0)q]U_2(T,q) - U_1(T,q) + (DU)(T,q)\Bigg\}h + o(h) \end{aligned}$$

As we divide both sides by h and let $h \downarrow 0$, $o(h)$ term disappears as $\frac{o(h)}{h} \to 0$. Hence, we obtain

$$g(q) + [1-(\lambda_1-\lambda_0)q]U_2(T,q) - U_1(T,q) + (DU)(T,q) \ge 0$$

as the second variational inequality.

Appendix C

Long proofs

C.1 Proof of Lemma 5.4

Proof. Observe that for a fixed $t \in \mathbb{R}_+$, the mappings $q \mapsto x(q,t)$ and $q \mapsto g(x(q,t))$ are increasing. For every fixed $T \in \mathbb{R}_+$, if we assume $q \mapsto w(T,q)$ to be increasing, then $q \mapsto (Sw)(T,q)$ is also increasing. As the coefficients $e^{-\lambda_0 t}$ and λ_0 are positive, $q \mapsto (J_t w)(T,q)$ is increasing for all $t \in [0,T]$. Since $q \mapsto (J_t w)(T,q)$ is increasing for all $t \in [0,T]$, the mapping $q \mapsto (Jw)(T,q)$ is increasing as well.

Note that the mapping $q \mapsto x(q,t)$ is (affine) concave. As concavity is preserved when $x(q,t)$ is shifted by $-\frac{1}{c}$, $q \mapsto g(x(q,t))$ is concave as well. Assuming $q \mapsto w(T,q)$ to be concave and using the linearity of the S operator, for $\alpha \in [0,1]$, $T \in \mathbb{R}_+$ and $q_1, q_2 \in \mathbb{R}_+$ we have

$$
\begin{aligned}
&(J_t w)(T, \alpha q_1 + (1-\alpha) q_2) \\
&= \int_0^t e^{-\lambda_0 s} \Big[g(x(\alpha q_1 + (1-\alpha) q_2, s)) \Big] ds \\
&+ \int_0^t \lambda_0 e^{-\lambda_0 s} \left[\int_{y \in \mathbb{R}^d} w\left(T - s, x(\alpha q_1 + (1-\alpha) q_2, s) \frac{\lambda_1}{\lambda_0} f(y) \right) \nu_0(dy) \right] ds \\
&\geq \int_0^t e^{-\lambda_0 s} \left[g(\alpha x(q_1, s) + (1-\alpha) x(q_2, s)) \right] ds
\end{aligned}
$$

$$
\begin{aligned}
&+\int_0^t \lambda_0 e^{-\lambda_0 s}\Bigg[\int_{y\in\mathbb{R}^d}\Bigg[w\left(T-s,\big(\alpha x(q_1,s)+(1-\alpha)x(q_2,s)\big)\frac{\lambda_1}{\lambda_0}f(y)\right)\Bigg]\nu_0(dy)\Bigg]ds\\
&\geq \int_0^t e^{-\lambda_0 s}\Big[\alpha g(x(q_1,s))+(1-\alpha)g(x(q_2,s))\Big]ds\\
&+\int_0^t \lambda_0 e^{-\lambda_0 s}\Bigg[\int_{y\in\mathbb{R}^d}\Bigg[\alpha w\left(T-s,x(q_1,s)\frac{\lambda_1}{\lambda_0}f(y)\right)\\
&+(1-\alpha)w\left(T-s,x(q_2,s)\frac{\lambda_1}{\lambda_0}f(y)\right)\Bigg]\nu_0(dy)\Bigg]ds\\
&=\alpha\int_0^t e^{-\lambda_0 s}g(x(q_1,s))ds+(1-\alpha)\int_0^t e^{-\lambda_0 s}g(x(q_2,s))ds\\
&+\alpha\int_0^t \lambda_0 e^{-\lambda_0 s}(Sw)(T-s,x(q_1,s))ds\\
&+(1-\alpha)\int_0^t \lambda_0 e^{-\lambda_0 s}(Sw)(T-s,x(q_2,s))ds\\
&=\alpha\left[\int_0^t e^{-\lambda_0 s}\Big(g(x(q_1,s))+\lambda_0(Sw)(T-s,x(q_1,s))\Big)ds\right]\\
&+(1-\alpha)\left[\int_0^t e^{-\lambda_0 s}\Big(g(x(q_2,s))+\lambda_0(Sw)(T-s,x(q_2,s))\Big)ds\right]\\
&=\alpha(J_t w)(T,q_1)+(1-\alpha)(J_t w)(T,q_2).
\end{aligned}
$$

where we used the fact that g and w are increasing and concave in the first and second inequality, respectively. When we take the infimum, we have

$$
\begin{aligned}
(Jw)(T,\alpha q_1+(1-\alpha)q_2) &= \inf_{t\in[0,T]}(J_t w)(T,\alpha q_1+(1-\alpha)q_2)\\
&\geq \inf_{t\in[0,T]}\left[\alpha(J_t w)(T,q_1)+(1-\alpha)(J_t w)(T,q_2)\right]\\
&\geq \inf_{t\in[0,T]}\alpha(J_t w)(T,q_1)+\inf_{t\in[0,T]}(1-\alpha)(J_t w)(T,q_2)\\
&=\alpha\inf_{t\in[0,T]}(J_t w)(T,q_1)+(1-\alpha)\inf_{t\in[0,T]}(J_t w)(T,q_2)\\
&=\alpha(Jw)(T,q_1)+(1-\alpha)(Jw)(T,q_2)
\end{aligned}
$$

Hence, the mapping $q\mapsto(Jw)(T,q)$ is concave. $\square$

C.2 Proof of Proposition 5.8

For the next proof, we will use the following result on the characterization of $\mathcal{F}$-stopping times; see Brémaud [18] and Davis [12].

Lemma C.1. *For every $\mathcal{F}$-stopping time τ and every $n \in \mathbb{N}_0$, there is an $\mathcal{F}_{\sigma_n}$-measurable random variable $R_n : \Omega \mapsto [0, \infty]$ such that $\tau \wedge \sigma_{n+1} = (\sigma_n + R_n) \wedge \sigma_{n+1}$ $\mathbb{P}_0$-a.s. on the event $\{\tau \geq \sigma_n\}$.*

Proof. Firstly, let us show that

$$\mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_n} g(Q_t)dt\right] \geq \overline{u}_n(T, q) \tag{C.1}$$

for every $n \in \mathbb{N}_0$, by proving inductively on $k = 1, ..., n+1$ that

$$\begin{aligned}
&\mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_n} g(Q_t)dt\right] \\
&\geq \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_{n-k+1}} g(Q_t)dt + 1_{\{\tau\geq\sigma_{n-k+1}\}}\overline{u}_{k-1}(T-\sigma_{n-k+1}, Q_{\sigma_{n-k+1}})\right] \\
&\equiv RHS_{k-1},
\end{aligned} \tag{C.2}$$

Observe that (C.1) is verified when we set $k = n+1$ in (C.2).

When $k = 1$, the inequality in (C.2) holds as an equality due to $\overline{u}_0 \equiv 0$. By an induction argument, we will show that (C.2) holds for $k+1$ assuming it holds for some $1 \leq k < n+1$. Let us rewrite (C.2) as

$$\begin{aligned}
\mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_n} g(Q_t)dt\right] \geq RHS_{k-1} &= RHS^{(1)}_{k-1} + RHS^{(2)}_{k-1} + RHS^{(3)}_{k-1} \\
&\triangleq \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_{n-k}} g(Q_t)dt\right] \\
&+ \mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}\left(\int_{\sigma_{n-k}}^{\tau\wedge\sigma_{n-k+1}} g(Q_t)dt\right)\right] \\
&+ \mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}1_{\{\tau\geq\sigma_{n-k+1}\}}\overline{u}_{k-1}(T-\sigma_{n-k+1}, Q_{\sigma_{n-k+1}})\right],
\end{aligned}$$

using

$$\int_0^{\tau\wedge\sigma_{n-k+1}} = \int_0^{\tau\wedge\sigma_{n-k}} + 1_{\{\tau\geq\sigma_{n-k}\}}\int_{\tau\wedge\sigma_{n-k}}^{\tau\wedge\sigma_{n-k+1}},$$

and

$$1_{\{\tau\geq\sigma_{n-k+1}\}} = 1_{\{\tau\geq\sigma_{n-k}\}}1_{\{\tau\geq\sigma_{n-k+1}\}}.$$

By the characterization of the $\mathcal{F}$-stopping times in Lemma (C.1), there exists a $\mathcal{F}_{\sigma_{n-k}}$-measurable random variable R_{n-k} such that

$$\tau\wedge\sigma_{n-k+1} = (\sigma_{n-k} + R_{n-k})\wedge\sigma_{n-k+1},$$

$\mathbb{P}_0$-almost surely on the event $\{\tau\geq\sigma_{n-k}\}$. Then, by the strong Markov property of the process X, the second expectation denoted by $RHS^{(2)}_{k-1}$ becomes

$$\begin{aligned}
&\mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}\int_{\sigma_{n-k}}^{\tau\wedge\sigma_{n-k+1}} g(Q_t)dt\right] \\
&= \mathbb{E}_{0,q}\left[\mathbb{E}_{0,q}\left(1_{\{\tau\geq\sigma_{n-k}\}}\int_0^{R_{n-k}\wedge\sigma_1\circ\theta_{\sigma_{n-k}}} g(Q_t)\circ\theta_{\sigma_{n-k}}dt\middle|\mathcal{F}_{\sigma_{n-k}}\right)\right] \\
&= \mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}\mathbb{E}_{0,q}\left(\left\{\int_0^{r\wedge\sigma_1} g(Q_t)dt\right\}\circ\theta_{\sigma_{n-k}}\middle|\mathcal{F}_{\sigma_{n-k}}\right)\Bigg|_{r=R_{n-k}}\right] \\
&= \mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}\mathbb{E}_{0,q}\left(\int_0^{r\wedge\sigma_1} g(Q_t)dt\right)\Bigg|_{r=R_{n-k},\, q=Q_{\sigma_{n-k}}}\right],
\end{aligned}$$

and the third expectation denoted by $RHS^{(3)}_{k-1}$ becomes

$$\begin{aligned}
&\mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}1_{\{\tau\geq\sigma_{n-k+1}\}}\overline{u}_{k-1}(T-\sigma_{n-k+1}, Q_{\sigma_{n-k+1}})\right] \\
&= \mathbb{E}_{0,q}\left[\mathbb{E}_{0,q}\left(1_{\{\tau\geq\sigma_{n-k}\}}1_{\{R_{n-k}\geq\sigma_1\circ\theta_{\sigma_{n-k}}\}}\{\overline{u}_{k-1}(T-\sigma_1, Q_{\sigma_1})\}\circ\theta_{\sigma_{n-k}}\middle|\mathcal{F}_{\sigma_{n-k}}\right)\right] \\
&= \mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}\mathbb{E}_{0,q}\left(\{1_{\{r\geq\sigma_1\}}\overline{u}_{k-1}(T-\sigma_1, Q_{\sigma_1})\}\circ\theta_{\sigma_{n-k}}\middle|\mathcal{F}_{\sigma_{n-k}}\right)\Big|_{r=R_{n-k}}\right] \\
&= \mathbb{E}_{0,q}\left[1_{\{\tau\geq\sigma_{n-k}\}}\mathbb{E}_{0,q}\left(1_{\{r\geq\sigma_1\}}\overline{u}_{k-1}(s-\sigma_1, Q_{\sigma_1})\right)\Big|_{r=R_{n-k},\, q=Q_{\sigma_{n-k}},s=T-\sigma_{n-k}}\right]
\end{aligned}$$

where θ is the shift operator. Hence, we have

$$
\begin{aligned}
RHS_{k-1} &= \mathbb{E}_{0,q}\Bigg[\int_0^{\tau\wedge\sigma_{n-k}} g(Q_t)dt + 1_{\{\tau\geq\sigma_{n-k}\}}\Bigg(\mathbb{E}_{0,q}\Big(\int_0^{r\wedge\sigma_1} g(Q_t)dt \\
&+ 1_{\{r\geq\sigma_1\}}\overline{u}_{k-1}(s-\sigma_1, Q_{\sigma_1})\Big)\Big|_{r=R_{n-k},\, q=Q_{\sigma_{n-k}},\, s=T-\sigma_{n-k}}\Bigg)\Bigg] \\
&= \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_{n-k}} g(Q_t)dt + 1_{\{\tau\geq\sigma_{n-k}\}}(J_r\overline{u}_{k-1})(s,q)\Big|_{r=R_{n-k},\, q=Q_{\sigma_{n-k}},\, s=T-\sigma_{n-k}}\right] \\
&\geq \mathbb{E}_{0,q}\left[\int_0^{\tau\wedge\sigma_{n-k}} g(Q_t)dt + 1_{\{\tau\geq\sigma_{n-k}\}}\overline{u}_k(T-\sigma_{n-k}, Q_{\sigma_{n-k}})\right] \\
&= RHS_k,
\end{aligned}
$$

as $(J_r\overline{u}_{k-1})(s,q) \geq (J\overline{u}_{k-1})(s,q) \equiv \overline{u}_k(s,q)$. This completes the proof of (C.2) by induction on k, and (C.1) follows by setting $k = n+1$ in (C.2). Taking the infimum of both sides in (C.1), we obtain $U_n \geq \overline{u}_n, n \in \mathbb{N}$.

Observe that to prove the opposite inequality $U_n \leq \overline{u}_n, n \in \mathbb{N}$, it is enough to show (5.15) holds because every $\mathcal{F}$-stopping time S_n^ε is less than or equal to σ_n $\mathbb{P}_0$-a.s. by construction. By induction on $n \in \mathbb{N}$, let us show (5.15) holds. For $n = 1$,

$$
\mathbb{E}_{0,q}\left[\int_0^{S_1^\varepsilon} g(Q_t)dt\right] = \mathbb{E}_{0,q}\left[\int_0^{r_0^\varepsilon(T,q)\wedge\sigma_1} g(Q_t)dt\right] = (J_{r_0^\varepsilon(T,q)}\overline{u}_0)(T,q).
$$

Since $(J_{r_0^\varepsilon(T,q)}\overline{u}_0)(T,q) \leq (J\overline{u}_0)(T,q) + \varepsilon$ by Remark (5.2) and $(J\overline{u}_0)(T,q) \equiv \overline{u}_1(T,q)$, we have $\mathbb{E}_{0,q}\left[\int_0^{S_1^\varepsilon} g(Q_t)dt\right] \leq \overline{u}_1(T,q) + \varepsilon$, and (5.15) holds for $n = 1$.

Suppose that (5.15) holds for every $\varepsilon > 0$ for some $n \in \mathbb{N}$. Let us show it also holds when n is replaced with $n+1$. Because $S_{n+1}^\varepsilon \wedge \sigma_1 = r_n^{\varepsilon/2}(T,q) \wedge \sigma_1$ $\mathbb{P}_0$-a.s., we have

$$
\begin{aligned}
&\mathbb{E}_{0,q}\left[\int_0^{S_{n+1}^\varepsilon} g(Q_t)dt\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{S_{n+1}^\varepsilon\wedge\sigma_1} g(Q_t)dt + 1_{\{S_{n+1}^\varepsilon\geq\sigma_1\}}\int_{\sigma_1}^{S_{n+1}^\varepsilon} g(Q_t)dt\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{r_n^{\varepsilon/2}(T,q)\wedge\sigma_1} g(Q_t)dt\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \mathbb{E}_{0,q}\left[1_{\{r_n^{\varepsilon/2}(T,q)\geq\sigma_1\}}\mathbb{E}_{0,q}\left(\int_0^{S_n^{\varepsilon/2}\circ\theta_{\sigma_1}} g(Q_{t+\sigma_1})dt\middle|\mathcal{F}_{\sigma_1}\right)\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{r_n^{\varepsilon/2}(T,q)\wedge\sigma_1} g(Q_t)dt\right] \\
&+ \mathbb{E}_{0,q}\left[1_{\{r_n^{\varepsilon/2}(T,q)\geq\sigma_1\}}\mathbb{E}_{0,q}\left(\left\{\int_0^{S_n^{\varepsilon/2}} g(Q_t)dt\right\}\circ\theta_{\sigma_1}\middle|\mathcal{F}_{\sigma_1}\right)\right] \\
&= \mathbb{E}_{0,q}\left[\int_0^{r_n^{\varepsilon/2}(T,q)\wedge\sigma_1} g(Q_t)dt\right] + \mathbb{E}_{0,q}\left[1_{\{r_n^{\varepsilon/2}(T,q)\geq\sigma_1\}}\mathbb{E}_{0,q}\left[\int_0^{S_n^{\varepsilon/2}} g(Q_t)dt\right]\Bigg|_{q=Q_{\sigma_1}}\right]
\end{aligned}
$$

by the strong Markov property of the process Q.

By induction hypothesis, $\mathbb{E}_{0,q}\left[\int_0^{S_n^{\varepsilon/2}} g(Q_t)dt\right]\Bigg|_{q=Q_{\sigma_1}} \leq \overline{u}_n(T-\sigma_1, Q_{\sigma_1}) + \varepsilon/2$. Therefore,

$$
\begin{aligned}
&\mathbb{E}_{0,q}\left[\int_0^{S_{n+1}^{\varepsilon}} g(Q_t)dt\right] \\
&\leq \mathbb{E}_{0,q}\left[\int_0^{r_n^{\varepsilon/2}(T,q)\wedge\sigma_1} g(Q_t)dt + 1_{\{r_n^{\varepsilon/2}(T,q)\geq\sigma_1\}}\overline{u}_n(T-\sigma_1, Q_{\sigma_1})\right] + \varepsilon/2 \\
&= (J_{r_n^{\varepsilon/2}(T,q)}\overline{u}_n)(T,q) + \varepsilon/2.
\end{aligned}
$$

Because $(J_{r_n^{\varepsilon/2}(T,q)}\overline{u}_n)(T,q) \leq (J\overline{u}_n)(T,q) + \varepsilon/2$ by the definition of $r_n^{\varepsilon/2}(T,q)$, and $(J\overline{u}_n)(T,q) \equiv \overline{u}_{n+1}(T,q)$, we have $\mathbb{E}_{0,q}\left[\int_0^{S_{n+1}^{\varepsilon}} g(Q_t)dt\right] \leq \overline{u}_{n+1}(T,q) + \varepsilon$, and (5.15) holds for $n+1$. This proves (5.15) by induction. □