Some Fluctuation Results Related to Draw-down Times for Spectrally Negative Lévy Processes

And

On Estimation of Entropy and Residual Entropy for Nonnegative Random Variables

Nhat Linh Vu

A Thesis
in the Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

August 2020

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Nhat Linh Vu

Entitled:

Some fluctuation results related to draw-down times for spectrally negative Lévy processes
And
On estimation of entropy and residual entropy for nonnegative random variables

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Mathematics and Statistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Supervisor
Chaubey, Yogendra

______________________________ Co-Supervisor
Zhou, Xiaowen

______________________________ External Examiner
Ding, Deng

______________________________ External to Program Examiner
Nebebe, Fassil

______________________________ Internal Examiner
Sen, Arusharka

______________________________ Internal Examiner
Sun, Wei

Approved by ____
Galia Dafni, Graduate Program Director

11 August, 2020 ____
Pascale Sicotte, Dean of Faculty

ABSTRACT

Nhat Linh Vu, Ph.D.
Concordia University, 2020

Part I: Some Fluctuation Results on Draw-down Times for Spectrally Negative Lévy Processes

In this thesis, we first introduce and review some fluctuation theory of Lévy processes, especially for general spectrally negative Lévy processes and for spectrally negative Lévy taxed processes. Then we consider a more realistic model by introducing draw-down time, which is the first time a process falls below a predetermined draw-down level which is a function of the running maximum. Particularly, we present the expressions for the classical two-sided exit problems for these processes with draw-down times in terms of scale functions. We also find the expressions for the discounted present values of tax payments with draw-down time in place of ruin time. Finally, we obtain the expression of the occupation times for the general spectrally negative Lévy processes to spend in draw-down interval killed on either exiting a fix upper level or a draw-down lower level.

Part II: On Estimation of Entropy and Residual Entropy for Nonnegative Random Variables

Entropy has become more and more essential in statistics and machine learning. A large number of its applications can be found in data transmission, cryptography, signal processing, network theory, bio-informatics, and so on. Therefore, the question of entropy estimation comes naturally. Generally, if we consider the entropy of a random variable knowing that it has survived up to time t, then it is defined as the residual entropy. In this thesis we focus on entropy and residual entropy estimation for nonnegative random variable. We first present a quick review on properties of popular existing estimators. Then we propose some candidates for entropy and residual entropy estimator along with simulation study and comparison among estimators.

Contribution from Authors

- **Part I of this thesis is under supervision of Dr. Xiaowen Zhou**
  The following two papers had been published based on Section 3 of Part 1 of this thesis.

  1. Florin Avram, Nhat Linh Vu and Xiaowen Zhou (2017). On taxed spectrally negative Lévy processes with draw-down stopping. *Insurance: Mathematics and Economics*, Volume 76, Issue C, page 69–74.

     All authors have equal contributions in this paper.

  2. Bo Li, Nhat Linh Vu and Xiaowen Zhou (2019). Exit problems for general draw-down times of spectrally negative Lévy processes. *Journal of Applied Probability*, Volume 56, Issue 2, page 441–457.

     The first and the third authors have more contributions in this paper.

- **Part II of this thesis is under supervision of Dr. Yogen Chaubey**
  The following two papers, which have been submitted, are both based on Section 3 and 4 of Part 2 of this thesis.

  1. Yogen Chaubey and Nhat Linh Vu (submitted in 2020). A numerical study of entropy and residual entropy estimators based on smooth density estimators for nonnegative random variables. *Journal of Statistical Research*.

     Both authors have equal contributions in this paper.

  2. Yogen Chaubey and Nhat Linh Vu (submitted in 2020). On an entropy estimator based on a nonparametric density estimator for nonnegative data. *IEEE Transactions on Information Theory*.

     Both authors have equal contributions in this paper.

Contents

List of Figures

List of Tables

PART I: Some fluctuation results on draw-down times for spectrally negative Lévy processes

1 Introduction

Lévy processes have been applied widely in a lot of fields of science. Some examples can be listed here such as for the study of turbulence, laser cooling and quantum field theory in physics; for the study of networks, queues and dams in engineering; for continuous time-series models and risk models in economics, for risk theory in actuarial mathematics, and of course, in mathematical finance, for the stock price in the market and calculations of insurance and re-insurance risk. Readers, who are interested in a deep comprehensive overview of Lévy process applications, can find answers in Prabhu (1998), in Barndorff *et al.* (2001), in Pistorius (2003), in Kyprianou *et al.* (2005), and in Kyprianou (2006).

Lévy processes are stochastic processes with independent and stationary increments. The best known and most important examples are Poisson processes, Brownian motion, Cauchy processes, and more general the stable processes. They are prototypes of Markov processes (actually they form the class of space-time homogeneous Markov processes) and of semi-martingales. Historically, the first researches go back to the late 20's with the study of infinitely divisible distributions, and their general structure had been gradually discovered by de Finetti, Kolmogorov, Lévy, Khintchine and Itô. After the pioneer contribution of Hunt in the mid-50's, the developments of the theory of Markov processes and their connection with abstract potential theories have had a considerable impact on Lévy processes. Many important properties of sample paths of Lévy processes have been noted by Getoor (1961), Rogozin (1972), and others. Further developments in this setting are made quite recently by Bertoin (1996), Barndorff *et al.* (2001), Doney and Kyprianou (2006), Sato (2013) and others.

In mathematical finance, back to the very beginning, people used Brownian motion to model and describe the observed reality of financial markets. However, In the real world, Brownian motion is not a good candidate for financial modeling because the actual asset price processes have jumps or spikes, and risk managers have to take them into consideration. As a result, risk managers seek for models that accurately fit return distributions for the estimation of profit and loss distributions. Another similar situation in the risk-neutral world, traders realized that the model of Black and Scholes (1973) cannot model the implied volatilities which can be constant neither across strikes nor across maturities. Therefore, in order to handle the risk of trades, traders need models that can capture the behavior of the implied volatility smiles more accurately. Consequently, Lévy processes are becoming more and more fashionable and one of the best choices in mathematical finance because Lévy processes can provide the appropriate tools to adequately and consistently describe all these observations, both in the real and risk-neutral world.

One of the most obvious and fundamental problems that can be stated for Lévy processes, particularly in relation to their role as modeling tools, is the distributional characterization of the time at which a Lévy process first exits an interval together with its overshoot and undershoot beyond the boundary of that interval. With the solution of one-sided and two-sided exit problems on hand, researchers can develop lots of relative properties of Lévy processes. The theory of Lévy processes forms the cornerstone of an enormous volume of mathematical literature which supports a wide variety of applied and theoretical stochastic models. As a family of stochastic processes, Lévy processes are now well understood and the exit problems have been solved by many different approaches dating back to the 1960s. Namely, Getoor (1961) and Rogozin (1972) are the foundation for other researchers to study more on the exit problems of Lévy processes.

Nonetheless, the theory of Lévy processes is still a large field to study and it is very challenging to characterize their properties without restricting ourselves into their sub-classes and then explore them separately. Consequently, Lévy processes are categorized into different classes such as stable processes, jump-diffusion processes, processes with one-sided jump and so on. The latter processes have attracted many researchers, and many of their fluctuation identities have been established explicitly or semi-explicitly due to their essential and obvious characteristics. That is, they are allowed to have only one-sided jump, either positive or negative jump but not both (for more details see Bertoin (1996), Avram *et al*. (2004), Chiu and Yin (2005), Doney and Kyprianou (2006), and Baurdoux (2009)). Among the class one-sided jump processes, spectrally negative Lévy processes (SNLPs) have been noticed recently because of their special applications in risk theory for insurance (an introduction of spectrally negative Lévy processes can be found in Kyprianou (2006)).

For the sake of expressing the solution of fluctuation identities associated with the one-sided and two-sided exit problems for spectrally negative Lévy processes in a closed, nice and simple form, researchers have literally introduced the so-called q-scale functions. Despite of the convenient use of q-scale functions, their explicit form is not available for most of Lévy processes as a result of the complexity of the Laplace exponent. So the numerical estimation must be employed in these cases. However, for some spectrally negative Lévy processes, such as the Brownian motion with or without drift, specially one-sided compound Poisson processes, spectrally one-sided α-stable processes with $\alpha \in (0, 2)$, and jump-diffusion processes, one can obtain the explicit form of the q-scale functions. Readers can refer to Kuznetsov *et al*. (2013) for the detail on evaluating scale functions of spectrally negative Lévy processes.

Beside the solution to the exit problems, occupation time for stochastic processes is also an important quantity that is used in many fields such as mathematical finance and risk theory. In the former, the distribution of occupation times is the key for the pricing of a certain class of average options (so-called α-quantile options), while in the latter, the Laplace transform of the occupation time is associated with the bankruptcy probability. The idea was first introduced in Gerber (1990) as follow. Some companies can have enough funds available or ask for external funds to support short periods in which the surplus of the company falls below zero, in the hope that it will recover soon in the future. Therefore, there is distinction between ruin (negative surplus) and bankruptcy (going out of business). That is, the Omega risk model assumes that the business still continue until bankruptcy occurs. The question arises here is the duration that the recovery will take in order to decide whether or not to continue the business. This is, indeed, related to the occupation time for the surplus process.

The study of occupation times has attracted researchers since the paper of Lévy (1939) in which he derived the density of occupation time for standard Brownian motion. In the recent decade, the investigation on occupation times for Lévy processes has grown widely, and many interesting results of occupation times have been derived. However, most of the existing papers can be classified into two categories depending on the the assumptions of the underlying process. The first group works with the occupation times for those Lévy processes whose two-sided jumps follow exponential or hyper-exponential distribution only (see Cai *et al*. (2010) and Wu and Zhou (2016)), whereas the second group focuses on the occupation times for SNLPs whose the jumps are only one-sided but no restriction on their distribution (see Laudriault *et al*. (2011) and Loeffen *et al*. (2014)).

Back to the very beginning of the theory of insurance risk, the classical Cramér-Lundberg surplus process, introduced in Lundberg (1903), were used to model the insurance risk. And then it was soon replaced by SNLPs which could capture the fluctuation of insurance risk better. However, to make it more practical to investigate the influences of tax on quantitative and qualitative behavior of the infinite time ruin probability, the model was modified to spectrally negative Lévy taxed processes (SNLTPs). It is assumed that the tax is paid at a fixed rate γ of the policy holders income (premium) whenever their risk process is at running maximum (profitable time). For the past ten years, the SNLTPs have been used to defined the so called risk process with tax in actuarial mathematics. The fluctuation identities for SNLTPs can be found in Albrecher and Hipp (2007) and Albrecher *at al*. (2008).

More recently, the models have been modified to be more flexible by replacing the ruin time by a varying draw-down time, which is a function of the running maximum. The draw-down can be interpreted as the investor's sustained loss between a peak (new maximum) and subsequent valley (points in between two maxima). It has recently become more and more considerably interesting in various areas of applied probability such as in queuing theory, risk theory and mathematical finance. For example, in the fund management industry, draw-down is used as the quoted indices; in mathematical finance, it is an indicator of risk in performance measure like the Calmar ratio, the Sterling ratio, and the Burke ratio. For further literature review on draw-down, readers are referred to Landriault *et al*. (2017). Therefore, associating draw-down times to SNLPs, SNLTPs, and occupation times is our main goal in this thesis.

Among different methodologies in dealing with stochastic processes, excursion theory has been introduced and successfully used to derive many well-known results. Especially, it comes to handy when we work on spectrally negative Lévy processes with draw-down times because many explicit calculations can be carried out using the fundamental property of excursion process, which is a Poisson point process. In particular, with the help of excursion theory, we were able to obtain the expression of the classical two-sided exit problems in terms of scale functions for SNLPs and SNLTPs with draw-down times. Also, we found the expressions for the discounted present values of tax payments, and the solution to the occupation times of a SNLP in a given draw-down interval.

This thesis is organized as follows. The introduction to Lévy processes along with their well-known properties are given in the first part of Section 2. The second part of Section 2 is devoted to a brief introduction of excursion theory which will play a central role in our main results. The definition of spectrally negative Lévy taxed processes and some existing results regarding to this process are given in the subsequent section. And the rest of the Section 2 is reserved for occupation times along with their previous results. Section 3 contains our main works and results involving drawn-down times regarding to exit problem of the SNLPs (in the first part), the SNLTPs (in the second part), and the occupation times (in the last part). Finally in Section 4, we present some results that related to draw-down times together with our interests and future works in joint distribution of a SNLP and the occupation times in a given draw-down intervals.

This part of thesis is based on Avram *et al.* (2017) and Li *et al.* (2019).

2 Spectrally negative Lévy process and its fluctuation theory

2.1 Lévy processes

2.1.1 Definitions and examples

In this section, we introduce some basic concepts of Lévy processes.

Definition 2.1. *(Lévy process) A process $X = \{X_t : t \geq 0\}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, is said to be one-dimensional Lévy process taking real value if it possesses the following properties:*
(i) The paths of X are $\mathbb{P}$-almost surely right-continuous with left limit.
(ii) $\mathbb{P}(X_0 = 0) = 1$.
(iii) For $0 \leq s \leq t, X_t - X_s$ follows the same distribution as X_{t-s}.
(iv) For $0 \leq s \leq t, X_t - X_s$ is independent of $\{X_u : u \leq s\}$.

From the definition above, it is difficult to see how rich the class of Lévy processes is. De Finetti (1929) introduced the notion of infinitely divisible distributions and showed that they have an intimate relationship with Lévy processes.

Definition 2.2. *We say that a real-valued random variable, Θ, has an infinitely divisible distribution if, for each $n = 1, 2, ...$, there exists a sequence of i.i.d. random variables $\Theta_{1,n}, ..., \Theta_{n,n}$ such that*

$$\Theta \stackrel{d}{=} \Theta_{1,n} + ... + \Theta_{n,n},$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Alternatively, we can express this relation in terms of probability laws. That is to say, the law η of a real-valued random variable is infinitely divisible if, for each $n = 1, 2, ...$, there exists another law η_n of a real-valued random variable such that $\eta = \eta_n^{*n}$. (Here η_n^{*n} denotes the n-fold convolution of η_n.) So, one way to establish whether a given random variable has an infinitely divisible distribution is via its characteristic exponent. Suppose that Θ has characteristic exponent $\Psi(u) := -\log \mathbb{E}(e^{iu\Theta})$, defined for all $u \in \mathbb{R}$. Then Θ has an infinitely divisible distribution if, for all $n \geq 1$, there exists a characteristic exponent of a probability distribution, say Ψ_n, such that $\Psi(u) = n\Psi_n(u)$, for all $u \in \mathbb{R}$. The full extension to which we may characterize infinitely divisible distributions is described by the characteristic exponent Ψ and an expression known as the Lévy Khintchine formula, which can be found in Kyprianou (2006).

Theorem 2.1. *(Lévy-Khintchine formula) A probability law, η, of a real-valued random variable is infinitely divisible with characteristic exponent (Lévy exponent) Ψ*

$$\int_{\mathbb{R}} e^{i\theta x}\eta(dx) = e^{-\Psi(\theta)}, \quad \text{or} \quad \Psi(\theta) := -\log \mathbb{E}(e^{i\theta X}) \quad \text{for } \theta \in \mathbb{R},$$

if and only if there exists a triple (μ, σ, Π), where $\mu, \sigma \in \mathbb{R}$, and Π is a measure concentrated on $\mathbb{R}\backslash\{0\}$ satisfying $\int_{\mathbb{R}}(1 \wedge x^2)\Pi(dx) < \infty$, such that

$$\Psi(\theta) = i\mu\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}}(1 - e^{i\theta x} + i\theta x \mathbf{1}_{|x|<1})\Pi(dx), \tag{2.1}$$

for every $\theta \in \mathbb{R}$. Moreover, the triple (μ, σ^2, Π) is unique.

Note that the measure Π is called the Lévy (characteristic) measure. This Lévy measure describes the size and the rate of jumps of the Lévy process. The condition $\int_{\mathbb{R}}(1\wedge x^2)\Pi(dx) < \infty$ in the theorem above ensures that the integral in the Lévy-Khintchine formula converges. Roughly speaking, in a small period of time dt, a jump of size x will occur with probability $\Pi(dx)dt + o(dt)$. In fact, the smaller the jump size results in the greater the intensity, and so the discontinuities in the path of the Lévy process is predominantly made up of arbitrarily small jumps. The converse of Theorem 2.1 which defines a Lévy process is given in the following theorem.

Theorem 2.2. *(Lévy-Khintchine formula for Lévy processes) Suppose that $\mu, \sigma \in \mathbb{R}$, and Π is a measure concentrated on $\mathbb{R}\backslash\{0\}$ such that $\int_{\mathbb{R}}(1 \wedge x^2)\Pi(dx) < \infty$. From this triple, define for each $\theta \in \mathbb{R}$,*

$$\Psi(\theta) = i\mu\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}}(1 - e^{i\theta x} + i\theta x \mathbf{1}_{|x|<1})\Pi(dx).$$

Then there exists a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, on which a Lévy process is defined having the characteristic exponent Ψ.

To clarify the concept of Lévy processes, we present here some examples of Lévy processes.

Compound Poisson processes

The first example of Lévy processes concerns processes whose paths are of bounded variation over finite time horizons. The necessary and sufficient conditions for a SNLP $X = (X_t)_{t\geq 0}$ to have paths of bounded variation are

$$\int_{(-1,0)} |x|\Pi(dx) < 0 \qquad \text{and} \qquad \sigma = 0.$$

In this case, X can be rewritten as

$$X_t = \mu t + S_t, \qquad t \geq 0,$$

where $\{S_t : t \geq 0\}$ is a pure jump subordinator which we will define later. Let $\{N_t : t \geq 0\}$ be a Poisson process such that, for each $t > 0$, N_t is a Poisson distribution with parameter λt. Then a compound Poisson process X_t is defined as

$$X_t = \sum_{i=1}^{N_t} Y_i,$$

where $\{Y_i\}_{i \geq 0}$ is a sequence of independent identical random variables with common law F. Also, it is well-known that

$$\mathbb{E}\big[e^{i\theta X_1}\big] = \exp\bigg\{ -\lambda \int_{\mathbb{R}} (e^{i\theta x} - 1) F(dx) \bigg\} = \bigg[\exp\bigg\{ -\frac{\lambda}{n} \int_{\mathbb{R}} (e^{i\theta x} - 1) F(dx) \bigg\} \bigg]^n.$$

From the above expression, we see that the distribution of X_t is infinite divisible. So it is a Lévy process. Also its characteristic exponent is given by $\Psi(\theta) = \lambda \int_{\mathbb{R}} (e^{i\theta x} - 1) F(dx)$, which implies that $\mu = -\lambda \int_{0<|x|<1} x F(dx), \sigma = 0$ and $\Pi(dx) = \lambda F(dx)$.

Linear Brownian Motion

Our second example of Lévy processes is processes with unbounded variation over finite time horizons. A linear Brownian Motion is defined as

$$X_t := \mu t + \sigma B_t, \qquad t \geq 0, \sigma > 0,$$

where $B = \{B_t : t \geq 0\}$ is a standard Brownian motion. With some algebra, one can show that

$$\mathbb{E}\big[e^{i\theta X_1}\big] = e^{-\frac{1}{2}\sigma^2\theta^2 + i\theta\mu} = \bigg[e^{-\frac{1}{2}\left(\frac{\sigma}{\sqrt{n}}\right)^2 \theta^2 + i\theta \frac{\mu}{n}} \bigg]^n,$$

which is of the form of an infinitely divisible distribution. So, it is a Lévy process with characteristic exponent $\Psi(\theta) = \frac{1}{2}\sigma^2\theta^2 - i\theta\mu$ with $\mu = -\mu, \sigma = \sigma$ and $\Pi = 0$.

Jump-diffusion processes

A jump-diffusion process X_t is just a sum of the compound Poisson process and an independent linear Brownian motion. That is

$$X_t = \mu t + \sigma B_t + \sum_{i=1}^{N_t} Y_i.$$

A well-known application of this process is modeling the stock price introduced in Merton (1976). That is, the stock price can be defined as $S_t = S_0 e^{X_t}$, where X_t is the jump-diffusion process and Y_i follows a Gaussian distribution. The process X_t has the characteristic exponent

$$\Psi(\theta) = -i\theta\mu + \frac{\sigma^2\theta^2}{2} - \lambda \int_{\mathbb{R}} (e^{i\theta x} - 1) F(dx).$$

We observe that this is very close to (1). Indeed, by the Lévy-Itô Decomposition Theorem, in general any Lévy process can be written as

$$X_t = \mu t + \sigma B_t + Z_t,$$

where μt is interpreted as the drift component, σB_t is the diffusion component, and Z_t is the jump process with possibly infinitely many jumps over any finite time interval.

Stable processes

Stable processes are those processes with characteristic exponent of the form of stable distributions which is another example of infinitely divisible distributions. Y is called a stable distribution if, for all $n \geq 1$, it can be decomposed into

$$n^{1/\alpha} Y + b_n \stackrel{d}{=} Y_1 + \cdots + Y_n,$$

where $Y_1, ..., Y_n$ are independent copies of Y and $b_n \in \mathbb{R}$ and $\alpha \in (0, 2]$ is known as the stability index. Note that for the case $\alpha = 2$, it turns out to be a zero mean Gaussian distribution. The Stable process X has a characteristic exponent of the form

$$\Psi(\theta) = \begin{cases} c|\theta|^\alpha \big(1 - i\beta \tan \frac{\pi\alpha}{2} \text{sgn } \theta\big) + i\theta\eta & \text{for } \alpha \in (0,1) \cup (1,2), \\ c|\theta| \big(1 + i\beta \frac{2}{\pi} \text{sgn } \theta \log|\theta|\big) + i\theta\eta & \text{for } \alpha = 1, \end{cases}$$

where $\beta \in [-1, 1]$, $\eta \in \mathbb{R}$, $c > 0$ and $\text{sgn } \theta = \mathbf{1}_{(\theta>0)} - \mathbf{1}_{(\theta<0)}$. This results in $\sigma = 0$,

$$\Pi(dx) = \begin{cases} c_1 x^{-1-\alpha} dx & \text{for } x \in (0, \infty), \\ c_2 |x|^{-1-\alpha} dx & \text{for } x \in (-\infty, 0), \end{cases}$$

where $c_1, c_2 \geq 0$, and the choice of $a \in \mathbb{R}$ is implicit.

Lévy processes with one-sided jumps

In general, it is very difficult to study the properties of the whole class of Lévy process due to the complexity jump part. It could be either a positive jump or a negative jump and it could happen infinite many times in a short interval. However, if we narrow our research to only one-sided jump Lévy processes, then we can explore a lot of interesting properties of this sub-class of Lévy processes.

Definition 2.3. *Suppose that* $\Pi(-\infty, 0) = 0$*, which implies that the corresponding Lévy processes have no negative jumps. Also, suppose further that* $\int_{(0,\infty)} (1 \wedge x)\Pi(dx) < \infty$*,* $\sigma = 0$*, and positive drift* $\mu > 0$*, then the process is called subordinator.*

A process X is called a spectrally positive Lévy process if $\Pi(-\infty, 0) = 0$, X does not have monotone paths, and it is not a pure negative linear drift. Lastly, if $-X$ is spectrally positive, then X is called a spectrally negative Lévy process. In this thesis, we only focus on the spectrally negative Lévy processes (SNLPs).

2.1.2 Some basic properties and facts about SNLPs

In order to facilitate the expression of results, we write

$$\mathbb{P}_x = \mathbb{P}(\cdot|X_0 = x) \quad \text{and} \quad \mathbb{E}_x = \mathbb{E}(\cdot|X_0 = x).$$

And for the case of $x = 0$, we write $\mathbb{P} = \mathbb{P}_0$ and $\mathbb{E} = \mathbb{E}_0$, respectively.

Laplace exponent

Since a SNLP does not have positive jumps, its Laplace exponent exists and is defined as

$$\mathbb{E}[e^{\theta X_t}] = e^{t\psi(\theta)},$$

so

$$\psi(\theta) = \mu\theta + \frac{1}{2}\sigma^2\theta^2 - \int_{(-\infty,0)} (1 - e^{\theta x} + \theta x \mathbf{1}_{(x>-1)})\Pi(dx), \tag{2.2}$$

given the triple (μ, σ, Π) for all $\theta \geq 0$. The function $\psi : [0, \infty) \to \mathbb{R}$ satisfies
(i) $\psi(0) = 0$.
(ii) $\lim_{x\to\infty} \psi(x) = \infty$.
(iii) ψ is infinitely differentiable and strictly convex on $(0, \infty)$.
(iv) $\psi'(0+) = \mathbb{E}[X_1] \in [-\infty, \infty)$.
For each $q \geq 0$, the right inverse of ψ is defined as

$$\Phi(q) = \sup\{\lambda \geq 0 : \psi(\lambda) = q\}. \tag{2.3}$$

Note that if the overall drift $\psi'(0+) \geq 0$, then ψ is strictly increasing. So $\lambda = 0$ is the unique solution to $\psi(\lambda) = 0$. If $\psi'(0+) < 0$, then the equation $\psi(\lambda) = 0$ has two solutions. One of them is zero, and the other is greater than 0.

Creeping upwards

Given a fixed level $a > 0$, the first passage time above this level a is defined as $\tau_a^+ := \inf\{t \geq 0 : X_t > a\}$ with the convention that $\inf \emptyset := \infty$. Also, we define the first passage time below level a as $\tau_a^- := \inf\{t \geq 0 : X_t < a\}$. Due to the fact that the SNLP has no positive jumps, it is shown by Corollary 3.13 in Kyprianou (2006) that

$$\mathbb{P}\big[X_{\tau_a^+} = a|\tau_a^+ < \infty\big] = 1. \tag{2.4}$$

That is, SNLPs necessarily creep upwards. But if $\sigma > 0$, then the process can creep downwards.

Drifting and oscillating

Since $\psi'(0+) = \mathbb{E}[X_1]$ is the overall drift of the process, the SNLP
(i) drifts to ∞ if and only if $\psi'(0+) > 0$,

(ii) oscillates if and only if $\psi'(0+) = 0$,
(iii) drifts to $-\infty$ if and only if $\psi'(0+) < 0$.

The Wiener-Hopf factorisation

For $t > 0$, define

$$\bar{X}_t = \sup_{s \le t} X_s \quad \text{and} \quad \underline{X}_t = \inf_{s \le t} X_s.$$

By the duality Lemma in Kyprianou (2006), the pairs $(\bar{X}_t, \bar{X}_t - X_t)$ and $(X_t - \underline{X}_t, -\underline{X}_t)$ have the same distribution in $\mathbb{P}$. Then we have the following Wiener-Hopf factorization theorem which plays an essential role in developing fluctuation identities of Lévy processes. For $\beta \ge 0$,

$$\mathbb{E}\big[e^{-\beta \bar{X}_{e_p}}\big] = \frac{\Phi(p)}{\Phi(p) + \beta} \text{ and } \mathbb{E}\big[e^{\beta \underline{X}_{e_p}}\big] = \frac{p}{\Phi(p)} \times \frac{\Phi(p) - \beta}{p - \psi(\beta)}, \tag{2.5}$$

where e_p is an independent exponential distribution with parameter p. The first expression implies that $\bar{X}_{e_p}$ follows an exponential distribution with parameter $\Phi(p)$.

q-Scale functions

It is surprising and interesting that most of the properties of SNLPs can be expressed in terms of the so-called q-scale functions. For $q \ge 0$, the q-scale function $W^{(q)}$ of a process X is defined on $[0, \infty)$ as a continuous function with Laplace transform of the form

$$\mathcal{L}[W^{(q)}](\lambda) := \int_0^\infty e^{-\lambda y} W^{(q)}(y) dy = \frac{1}{\psi(\lambda) - q}, \quad \text{for } \lambda > \Phi(q). \tag{2.6}$$

The function $W^{(q)}$ is unique, positive and strictly increasing for $x \ge 0$. To extend the domain of $W^{(q)}$ to the whole real line, we set $W^{(q)}(x) = 0$ for $x < 0$. For simplicity, we write $W = W^{(0)}$ whenever $q = 0$. Furthermore, we define another scale function $Z^{(q)}$ as

$$Z^{(q)}(x) := 1 + q \int_0^x W^{(q)}(y) dy. \tag{2.7}$$

It is shown that for all SNLPs, q-scale functions exist for all $q \ge 0$ (see Kuznetsov *et al.* (2013)). This is a fundamental result because from here we can express the fluctuation identities of general SNLPs in terms of scale functions. Moreover, the q-scale function $W^{(q)}$ is continuous and almost everywhere differentiable. Indeed, for each $q \ge 0$, the scale function $W^{(q)}$ belongs to $\mathcal{C}^1(0, \infty)$ if and only if at least one of the following three criteria holds

(i) $\sigma = 0$.

(ii) $\int_{(-1,0)} |x| \Pi(dx) = \infty$.

(iii) $\overline{\Pi}(dx) := \Pi(-\infty, -x)$ is continuous.

Let $d := \mu - \int_{-1}^{0} z\Pi(dz)$, then the initial values of $W^{(q)}$ and $W'^{(q)}$ are

$$W^{(q)}(0+) = \begin{cases} 1/d & \text{if } \sigma = 0 \text{ and } \int_{-1}^{0} z\Pi(dz) < \infty, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$W'^{(q)}(0+) = \begin{cases} 2/\sigma^2 & \text{if } \sigma > 0, \\ (\Pi(-\infty, 0) + q)/d^2 & \text{if } \sigma = 0 \text{ and } \int_{-1}^{0} z\Pi(dz) < \infty, \\ \infty & \text{otherwise.} \end{cases}$$

Here we derive some explicit expressions of q-scale function taken from Biffis and Kyprianou (2010) and Kuznetsov *et al.* (2013) (for a deep study on scale functions of SNLPs, readers are referred to Chan *et al.* (2009) and Hubalek *et al.* (2010)).

- The first example is the linear Brownian motion with drift X which is of the form

  $$X_t := \mu t + \sigma B_t, \qquad \text{where} \qquad \sigma > 0, \mu \in \mathbb{R}.$$

  Its Laplace exponent can be obtained directly from its corresponding characteristic exponent.

  $$\psi(\theta) = -\Psi(-i\theta) = \mu\theta + \frac{\sigma^2\theta^2}{2} \qquad \text{for } \theta \in \mathbb{R}.$$

  Then the Laplace transform of the scale function $W^{(q)}$ for the linear Brownian motion with drift X can be expressed as

  $$\mathcal{L}[W^{(q)}](s) = \frac{1}{\frac{\sigma^2 s^2}{2} + \mu s - q} = \frac{2}{\sigma^2(s - s_1)(s - s_2)},$$

  where $s_1 = \frac{-\mu - \sqrt{\mu^2 + 2\sigma^2 q}}{\sigma^2}$ and $s_2 = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2 q}}{\sigma^2}$. So, by partial fraction decomposition, we can re-write the Laplace transform as

  $$\mathcal{L}[W^{(q)}](s) = \frac{1}{\sqrt{2q\sigma^2 + \mu^2}} \left[\frac{1}{s - s_2} - \frac{1}{s - s_1} \right].$$

  Therefore, by inverse Laplace transform, we obtain the explicit expression of the q-scale function $W^{(q)}$:

  $$W^{(q)}(x) = \frac{1}{\sqrt{2q\sigma^2 + \mu^2}} \left[e^{(\sqrt{2q\sigma^2 + \mu^2} - \mu)\frac{x}{\sigma^2}} - e^{-(\sqrt{2q\sigma^2 + \mu^2} + \mu)\frac{x}{\sigma^2}} \right].$$

- The second example is the compound Poisson processes with a positive drift and with exponential jumps

  $$X_t = \mu t - \sum_{i=1}^{N_t} Y_i,$$

  where $\mu > 0$ and Y follows exponential distribution with parameter ρ and N_t is an independent Poisson process with intensity $\lambda > 0$. The Laplace exponent of X can be derived directly as

$$\begin{aligned}\psi(\theta) &= -\Psi(-i\theta)\\ &= \mu\theta + \lambda\int_{(-\infty,0)}(e^{\theta x}-1)F(dx)\\ &= \mu\theta + \lambda\int_0^\infty (e^{-\theta x}-1)\rho e^{-\rho x}dx\\ &= \mu\theta - \frac{\lambda\theta}{\rho+\theta}.\end{aligned}$$

So after some algebra, the Laplace transform of $\mathcal{L}[W^{(q)}](s)$ can be expressed as

$$\frac{\rho+s}{\mu s^2-(\lambda+q+\rho\mu)s-\rho q} = \frac{\rho+s}{\mu(s-s_1)(s-s_2)} = \frac{1}{\mu(s_2-s_1)}\left(\frac{\rho+s_2}{s-s_2}-\frac{\rho+s_1}{s-s_1}\right),$$

where

$$\begin{aligned}s_1 &= \frac{1}{2\mu}\left((\lambda+q+\rho\mu)-\sqrt{(\lambda+q+\rho\mu)^2+4\mu\rho q}\right),\\ s_2 &= \frac{1}{2\mu}\left((\lambda+q+\rho\mu)+\sqrt{(\lambda+q+\rho\mu)^2+4\mu\rho q}\right).\end{aligned}$$

As a result, by inverse Laplace transform, we obtain the explicit expression of q-scale function for compound Poisson processes with a positive drift and exponential jumps

$$W^{(q)}(x) = \frac{1}{\mu(s_2-s_1)}\left((\rho+s_2)e^{s_2x}-(\rho+s_1)e^{s_1x}\right).$$

Note that, for the special case where $q=0$, the expression of 0-scale function is simplified into the form

$$W(x) = \frac{\rho}{\lambda-\rho\mu}\left(\frac{\lambda}{\rho\mu}e^{(\frac{\lambda}{\rho}-\mu)\frac{\rho}{\mu}x}-1\right).$$

- The third example is the α-stable process X, $\alpha \in (1,2)$, which is defined in a way that for each $t>0$, X_t is equal in distribution to $t^{1/\alpha}X_1$. It is interesting that the 0-scale function for this process has a very simple form. That is

$$W(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}.$$

If the α-stable process plus a drift μt is considered, then its 0-scale function has a form

$$W(x) = \frac{1}{\mu}(1-E_{\alpha-1,1}(-cx^{\alpha-1})),$$

where

$$E_{\alpha-1,1}(z) = \sum_{k\geq 0} z^k/\Gamma(1+(\alpha-1)/k).$$

Exponential change of measure

Thanks to the fact that, for each $c \geq 0$, $\{e^{cX_t-\psi(c)t} : t \geq 0\}$ is a positive martingale, we can define the change of measure for each $c \geq 0$,

$$\left.\frac{d\mathbb{P}^{(c)}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = e^{cX_t-\psi(c)t}, \tag{2.8}$$

where $(\mathcal{F}_t)_{t\geq 0}$ is the filtration generated by X. The change of measure plays an important role in deriving some properties of SNLPs in such a way that it changes a SNLP X with non-positive drift $\psi'(0+) \leq 0$ into another SNLP X with a positive drift $\psi_c'(0+) > 0$. Particularly with an appropriate choice of c, given $(X, \mathbb{P})$ is a SNLP, then by the change of measure from $\mathbb{P}$ to $\mathbb{P}^{(c)}$, $(X, \mathbb{P}^{(c)})$ is also a spectrally negative Lévy process with Laplace exponent given by

$$\begin{aligned}
\psi_c(\theta) &= \psi(\theta+c) - \psi(c) \\
&= \theta\Big(\sigma^2 c - \mu + \int_{(-\infty,0)} x(e^{cx}-1)\mathbf{1}_{(x>-1)}\Pi(dx)\Big) + \frac{1}{2}\sigma^2\theta^2 \\
&\qquad + \int_{(-\infty,0)} (e^{\theta x} - 1 - \theta x\mathbf{1}_{(x>-1)})e^{cx}\Pi(dx),
\end{aligned}$$

for $\theta \geq -c$ given the triple (μ, σ, Π) of X. By inspecting the above expression, we see that the triple of the process X under the new measure becomes

$$\begin{aligned}
\mu_c &= -\Big(\sigma^2 c - \mu + \int_{(-\infty,0)} x(e^{cx}-1)\mathbf{1}_{(x>-1)}\Pi(dx)\Big), \\
\sigma_c &= \sigma, \\
\Pi_c(dx) &= e^{cx}\Pi(dx).
\end{aligned}$$

Moreover, the q-scale function $W_c^{(q)}$ and $Z_c^{(q)}$ of X under $\mathbb{P}^{(c)}$ can be expressed in terms of the q-scale function $W^{(q)}$ and $Z^{(q)}$ of X under $\mathbb{P}$ respectively as

$$W_c^{(q)}(x) = e^{-cx}W^{(q+\psi(c))}(x) \qquad \text{and} \qquad Z_c^{(q)}(x) = 1 + q\int_0^x W_c^{(q)}(y)dy. \tag{2.9}$$

The one-sided and two-sided exit problems

The solutions for the one-sided and two-sided exit problems are stated in the following theorem (Kyprianou 2006).

Theorem 2.3. *For any $x \in \mathbb{R}$ and $q \geq 0$,*

$$\mathbb{E}\big[e^{-q\tau_a^+}\mathbf{1}_{(\tau_a^+<\infty)}\big] \quad = \quad e^{-\Phi(q)a}, \tag{2.10}$$

$$\mathbb{E}_x\big[e^{-q\tau_0^-}\mathbf{1}_{(\tau_0^-<\infty)}\big] \quad = \quad Z^{(q)}(x) - \frac{q}{\Phi(q)}W^{(q)}(x), \tag{2.11}$$

where we understand $q/\Phi(q)$ in the limiting sense for $q = 0$, so that

$$\mathbb{P}_x\big[\tau_0^- < \infty\big] = \begin{cases} 1 - \psi'(0+)W(x) & \text{if } \psi'(0+) \geq 0, \\ 1 & \text{if } \psi'(0+) < 0. \end{cases} \tag{2.12}$$

Also, for any $x \leq a$ and $q \geq 0$,

$$\mathbb{E}_x\big[e^{-q\tau_a^+}\mathbf{1}_{(\tau_0^->\tau_a^+)}\big] = \frac{W^{(q)}(x)}{W^{(q)}(a)}, \tag{2.13}$$

$$\mathbb{E}_x\big[e^{-q\tau_0^-}\mathbf{1}_{(\tau_0^-<\tau_a^+)}\big] = Z^{(q)}(x) - Z^{(q)}(a)\frac{W^{(q)}(x)}{W^{(q)}(a)}, \tag{2.14}$$

and for $u, v \geq 0$ the joint Laplace transform of τ_0^- and $X_{\tau_0^-}$ is

$$\mathbb{E}_x\big[e^{-u\tau_0^- + vX_{\tau_0^-}}\big] = e^{vx}\Big(Z_v^{(p)}(x) - W_v^{(p)}(x)p/\Phi_v(p)\Big), \tag{2.15}$$

where $p = u - \psi(v)$, $\Phi_v(p)$ is the largest root of $\psi_v(\theta) = p$, and $W_v^{(p)}, Z_v^{(p)}$ are the q-scale functions under the new measure $\mathbb{P}^v$.

Resolvent measures

The q-potential measure, or known as the resolvent measure, is defined as

$$U^{(q)}(a, x, dy) := \int_0^\infty e^{-qt}\mathbb{P}_x(X_t \in dy, \tau > t)dt, \tag{2.16}$$

where $\tau := \tau_a^+ \wedge \tau_0^-$. Also we denote $U := U^{(0)}$. If for each $x \in [0, a]$, a density of $U^{(q)}(a, x, dy)$ exists with respect to Lebesgue measure, then $u^{(q)}(a, x, dy)$ is called the potential density. Potential measure have played an important role in derivation of fluctuation identities of Lévy processes especially in SNLPs. Moreover, It is interesting for the case of SNLPs, in which the potential density always exists and can be expressed in terms of scale functions. The following theorem is taken from Kyprianou (2006).

Theorem 2.4. *For $q \geq 0$ and $x, y \in [0, a]$, the density $u^{(q)}(a, x, y)$ of q-potential measure of a SNLP killed on exiting $[0, a]$ is given by*

$$u^{(q)} = \frac{W^{(q)}(x)W^{(q)}(a-y)}{W^{(q)}(a)} - W^{(q)}(x-y). \tag{2.17}$$

Reflected SNLPs

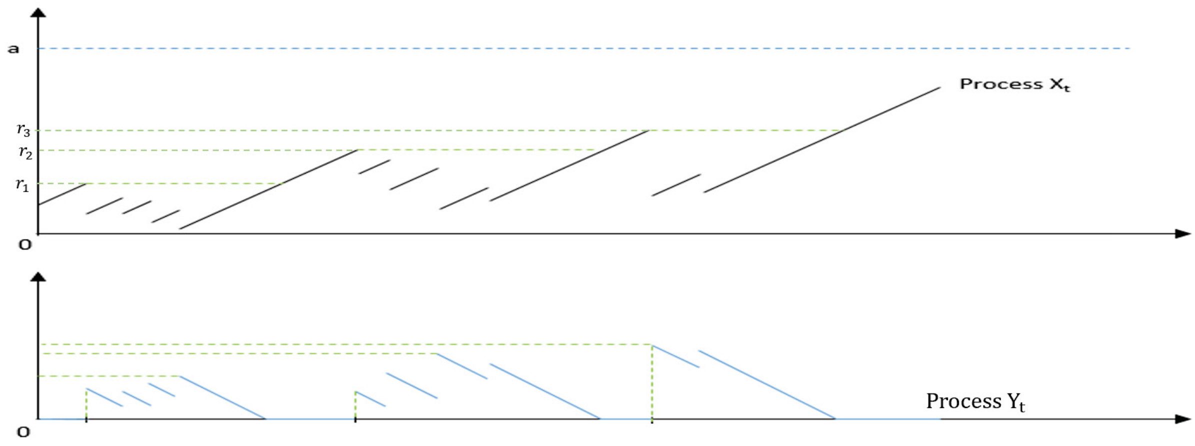

Figure 1: An example of sample path of SNLP X_t along with the reflected SNLP Y_t.

We want to close this sub-section by a brief introduction to reflected SNLPs which take essential part in excursion theory. For $x \geq 0$, the process

$$Y_t := (x \vee \bar{X}_t) - X_t, \qquad t \geq 0 \tag{2.18}$$

is called the process reflected at its supremum. The following two theorems from Kyprianou (2006) present the exit problem and the potential measure of the process Y respectively.

Theorem 2.5. *Suppose for a fixed $a > 0$, we define the first passage time of the process Y as $\sigma_a^+ := \inf\{t > 0 : Y_t > a\}$. Then, for $x \in [0, a]$, $\theta \in \mathbb{R}$ such that $\psi(\theta) < \infty$, we have*

$$\mathbb{E}_x[e^{-q\sigma_a^+ - \theta Y_{\sigma_a^+}}] = e^{-\theta x}\left(Z_\theta^{(p)}(a-x) - W_\theta^{(p)}(a-x)\frac{pW_\theta^{(p)}(a) + \theta Z_\theta^{(p)}(a)}{W_\theta^{(p)\prime}(a) + \theta W_\theta^{(p)}(a)}\right), \tag{2.19}$$

where $p := q - \psi(\theta)$ and $W^{(q)\prime}(a)$ is the right derivative of $W^{(q)}$ at a.

Theorem 2.6. *For $a > 0, q \geq 0$, and $x, y \in [0, a]$ the potential measure of the reflected process Y, denoted by $\bar{U}^{(q)}$, can be expressed as*

$$\bar{U}^{(q)}(a, x, dy) = \left(W^{(q)}(a-x)\frac{W^{(q)}(0)}{W^{(q)\prime}(a)}\right)\delta_0(dy) + \left(W^{(q)}(a-x)\frac{W^{(q)\prime}(y)}{W^{(q)\prime}(a)} - W^{(q)}(y-x)\right)dy.$$

2.2 Excursion theory for SNLPs

Since most of our main results mainly rely on the excursion theory, we want to dedicate this sub-section to briefly introduce this concept. In order to introduce the theory of excursion, we need the notion of the Poisson random measure which plays a central role in the theory of Lévy processes.

2.2.1 Poisson random measure

First, we present the definition of random measure.

Definition 2.4 (Random measure)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ be a measurable space. Then $M : \Omega \times \mathcal{E} \to \mathbb{R}$ is a random measure if*

(i) for every $\omega \in \Omega$, $M(\omega, \cdot)$ is a measure on $\mathcal{E}$,

(ii) for every $A \in \mathcal{E}$, $M(\cdot, A)$ is measurable.

Then the Poisson random measure is defined as follow.

Definition 2.5 (Poisson random measure)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{E}, \eta)$ be a measurable space. Then $N : \Omega \times \mathcal{E} \to \mathbb{R}$ is a Poisson random measure with intensity η measure on $\mathcal{E}$ if*

(i) for every $A \in \mathcal{E}$ with $\eta(A) < \infty$, $N(\cdot, A)$ follows the Poisson distribution with parameter $\eta(A)$,

(ii) for any disjoint sets $A_1, ..., A_n \in \mathcal{E}$, $N(\cdot, A_1), ..., N(\cdot, A_n)$ are independent,

(iii) for every $\omega \in \Omega$, $N(\omega, \cdot)$ is a measure on $\mathcal{E}$.

The existence of the Poisson random measure for SNLPs is guaranteed by Theorem 2.4 in Kyprianou (2006). Although there are many important properties regarding to the Poisson measure, we present here the two most useful results which can be found in Kyprianou (2006). The first one is the Exponential Formula.

Theorem 2.7. *Let N be a Poisson random measure on $(E, \mathcal{E}, \eta)$ with intensity η, $B \in \mathcal{E}$ and let f be a measurable function with $\int_B |e^{f(x)} - 1|\eta(dx) < \infty$. Then*

$$\mathbb{E}\left[e^{\int_B f(x)N(dx)}\right] = \exp\left\{ \int_B (e^{f(x)} - 1)\eta(dx) \right\}.$$

The second one is the Compensation Formula under Poisson random measure.

Theorem 2.8. *Suppose $\phi : [0, \infty) \times \mathbb{R} \times \Omega \to [0, \infty)$ is a random time-space function such that*

(i) function $\phi = \phi(t, x)(\omega)$ is measurable,

(ii) for each $t \geq 0$, $\phi(t, x)[\omega]$ is $\mathcal{B}(\mathbb{R} \times \mathcal{F}_t)$-measurable,

(iii) for each $x \in \mathbb{R}$, with probability one, $\{\phi(t, x) : t \geq 0\}$ is a left continuous process.

Then, for all $t \geq 0$,

$$\mathbb{E}\left[\int_{[0,t]} \int_{\mathbb{R}} \phi(s, x)N(ds \times dx) \right] = \mathbb{E}\left[\int_0^t \int_{\mathbb{R}} \phi(s, x)\eta(dx)ds \right],$$

with the understanding that the right-hand side is infinite if and only if the left-hand side is.

2.2.2 Local time and excursion processes

To illustrate how an excursion process is constructed, we consider a very simple case of SNLP of bounded variation but not a compound Poisson process. With the assumption that the SNLP X drifts to infinity, we are able to decompose the path of X in the following way. First we initiate at $T_0 = 0$ and $H_0 = 0$. Then we define $T_1 := \sigma_0^+ = \inf\{t > 0 : Y_t > 0\}$ and $H_1 = Y_{T_1}$ if $T_1 < \infty$, otherwise $H_1 := \infty$, where $Y = \bar{X} - X$ is the reflected process. Now we construct the sequence of pairs of variables (T_n, H_n) as

$$T_n := \begin{cases} \inf\{t > T_{n-1} : Y_t > H_{n-1}\} & \text{if } T_{n-1} < \infty, \\ \infty & \text{if } T_{n-1} = \infty, \end{cases} \qquad H_n := \begin{cases} Y_{T_n} & \text{if } T_n < \infty, \\ \infty & \text{if } T_n = \infty. \end{cases}$$

We can see that the sequence of pair (T_n, H_n) are simply the jump times and the consecutive heights of the new maxima of Y (see Figure 6). By the strong Markov property and stationary independent increments, an excursion of X from its maximum, defined as

$$\epsilon_n := \{Y_t - Y_{T_{n-1}} : T_{n-1} < t \leq T_n\}, \tag{2.20}$$

is independent of $\mathcal{F}_{T_{n-1}}$ and has the same law as $\{Y_t : 0 < t \leq \sigma_0^+\}$. Unfortunately, it is known that a general Lévy process may have an infinite number of new maxima, which results in infinite number of excursions, over any given finite interval of time. This requires a quantity that can monitor and index these excursions, so the notion of local time at the maximum is introduced. Here we assume neither X nor $-X$ is a subordinator, and we refer to the reflected process $\bar{X} - X$, which is a strong Markov process by Kyprianou (2006).

Definition 2.6. *A continuous, non-decreasing, nonnegative, $\mathbb{F}_t$-adapted process, $L = \{L_t : t \geq 0\}$, is called a local time at the maximum for X if the followings hold:*

(i) The support of the Stieltjes measure dL is the closure of the (random) set of times $\{t \geq 0 : \bar{X}_t = X_t\}$.

(ii) For every $\mathbb{F}$-stopping time T such that $\bar{X}_T = X_T$ on $\{T < \infty\}$ almost surely, the shifted process $\{L_{T+t} - L_T : t \geq 0\}$ is independent of $\mathbb{F}_T$ on $\{T < \infty\}$ and has the same law as L under $\mathbb{P}$.

Then the inverse local time process, $L^{-1} := \{L_t^{-1} : t \leq 0\}$, is

$$L_t^{-1} := \begin{cases} \inf\{s > 0 : L_s > t\} & \text{if } t < L_\infty, \\ \infty & \text{otherwise.} \end{cases}$$

Particularly, the local time in the case of SNLPs has a special form.

Theorem 2.9. *The point 0 is regular for $(0, \infty)$ and the continuous increasing process $\bar{X}_t = \sup\{X_s : 0 \leq s \leq t\}$ is a local time at 0 for the reflected process $\bar{X} - X$.*

For each moment of local time $t > 0$, the general decomposition of the path of Lévy process is given in terms of its excursions and inverse local time from the maximum as the following.

$$\epsilon_t = \begin{cases} \{X_{L_{t-}^{-1}+s} - X_{L_{t-}^{-1}} : 0 < s \leq L_t^{-1} - L_{t-}^{-1}\} & \text{if } L_{t-}^{-1} < L_t^{-1}, \\ \partial & \text{if } L_{t-}^{-1} = L_t^{-1}, \end{cases}$$

where ∂ is some "dummy" state. Before stating the fundamental result of excursion theory, we need the following definition from Kyprianou (2006).

Definition 2.7. *Let $\mathcal{E}$ be the space of excursions of X from its running maximum, that is, the space of mappings which are right-continuous with left limits satisfying*

$$\begin{aligned} \varepsilon : (0, \zeta) \to [0, \infty) \qquad & \textit{for some } \zeta \in (0, \infty], \\ \varepsilon : \{\zeta\} \to [0, \infty) \qquad & \textit{if } \zeta < \infty, \end{aligned}$$

where $\zeta = \zeta(\varepsilon)$ is the excursion length or excursion lifetime. Write $h = h(\varepsilon)$ for the terminal value of the excursion, so that $h(\varepsilon) = \varepsilon(\zeta)$. Finally, let $\bar{\varepsilon} = -\inf_{s\in(0,\zeta)} \varepsilon(s)$ for the excursion height.

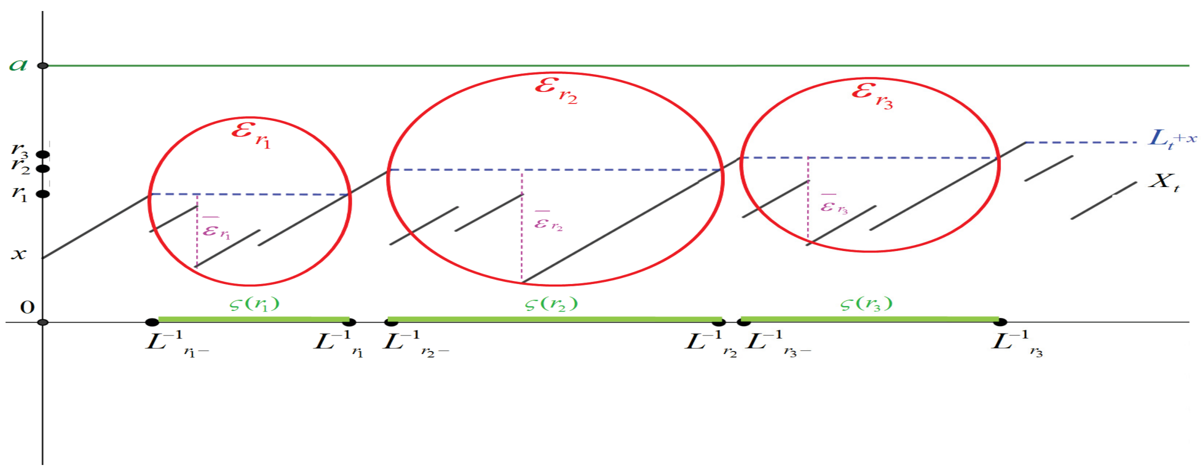

Figure 2: An example of sample path of X_t along with the excursion processes $\epsilon_r(t)$. Here, L_t is the local time of the reflected process $Y_t = \bar{X}_t - X_t$ at zero.

It turns out interestingly that ε is a Poisson point process, and $\bar{\varepsilon}$ is also the Poisson point process with characteristic measure (also called the excursion measure) ν. The theorem below is taken from Kyprianou (2006).

Theorem 2.10. *There exists a sigma-algebra Σ and σ-finite measure n such that $(\mathcal{E}, \Sigma, n)$ is a measure space and Σ is rich enough to contain sets of the form*

$$\{\varepsilon \in \mathcal{E} : \zeta(\varepsilon) \in A,\ \bar{\varepsilon} \in B,\ h(\varepsilon) \in C\},$$

where, for a given $\varepsilon \in \mathcal{E}$, A, B, and C are Borel sets in $[0, \infty]$.

(i) If $\mathbb{P}(\limsup_{t\uparrow\infty} X_t = \infty) = 1$, then $\{(t, \epsilon_t) : t \geq 0$ and $\epsilon_t \neq \partial\}$ is a Poisson point process on $\big([0, \infty] \times \mathcal{E}, \mathcal{B}[0, \infty) \times \Sigma, dt \times dn\big)$ with characteristic measure n.

(ii) If $\mathbb{P}(\limsup_{t\uparrow\infty} X_t < \infty) = 1$, then $\{(t,\epsilon_t) : t \leq L_\infty \text{ and } \epsilon_t \neq \partial\}$ is a Poisson point process on $\big([0,\infty]\times\mathcal{E}, \mathcal{B}[0,\infty)\times\Sigma, dt\times dn\big)$ with characteristic measure ν stopped at the first arrival of an excursion in $\mathcal{E}_\infty := \{\epsilon\in\mathcal{E} : \zeta(\epsilon)=\infty\}$.

Moreover, by Bertoin (1996), the excursion height $\bar{\epsilon}$ is a Poisson point process with characteristic measure ν. In order to study this characteristic measure ν, we recall the well-known two-sided exit problem. For $x, y > 0$, we have

$$\mathbb{P}[\tau_y^+ < \tau_{-x}^-] = \frac{W(x)}{W(x+y)}.$$

On the other hand, by excursion theory, the event $\{\tau_y^+ < \tau_{-x}^-\}$ coincides to the event $\{\bar{\epsilon}_t \leq x+t \text{ for all } t\in[0,y]\}$. And since $\bar{\epsilon}$ is a Poisson point process with charateristic measure ν, the event $\{\bar{\epsilon}_t \leq x+t \text{ for all } t\in[0,y]\}$ is equivalent to $\{N=0\}$, where N is he Poisson random variable counting the number of Poisson points

$$\big\{(t,\epsilon_t)\in\mathbb{R}^2 \big|\ t\in[0,y],\ \epsilon_t > x+t\big\}.$$

Thus,

$$\begin{aligned}
\frac{W(x)}{W(x+y)} &= \mathbb{P}[\tau_y^+ < \tau_{-x}^-] \\
&= \mathbb{P}[N=0] \\
&= \exp\Big\{-\int_0^y \nu(x+t,\infty)dt\Big\} \\
&= \exp\Big\{-\int_x^\infty \nu(t,\infty)dt + \int_{x+y}^\infty \nu(t,\infty)dt\Big\}
\end{aligned}$$

which implies that

$$W(x) = \exp\Big\{-\int_x^\infty \nu(t,\infty)dt\Big\}.$$

Finally, we want to close this sub-section with one of the most important results for the excursion theory, the compensation formula in the framework of excursion theory from Bertoin (1996). Consider a real value Markov process $M=(M_t, t\geq 0)$ having right-continuous sample paths and with $\mathbb{P}(M_0=0)=1$. We denote $\mathcal{L}=\{t : M_t=0\}$ the zero set of M and its closure as $\overline{\mathcal{L}}$. We call (g,d) as an excursion interval if $M_t\neq 0$ for all $t\in(g,d)$, $g\in\overline{\mathcal{L}}$ and $d\in\overline{\mathcal{L}}\cup\{\infty\}$. Intuitively, g, d, and $l=d-g$ are the left-end point, right-end point, and the length of the excursion interval. For every left-end point $g<\infty$ of an excursion interval, we denote $\mathfrak{e}=\{M_{g+t}, 0\leq t<d-g\}$ the excursion starting at time g. Then, the compensation formula is stated in the following theorem.

Theorem 2.11. *Consider a measurable function $F : \mathbb{R}^+\times\Omega\times\mathcal{E}\to[0,\infty)$ such that for every $\mathfrak{e}\in\mathcal{E}$, the process $F_t(\mathfrak{e}) = F(t,\omega,\mathfrak{e})$ is left-continuous and adapted, then we have*

$$\mathbb{E}\Big[\sum_g F_g(\mathfrak{e}_g)\Big] = \mathbb{E}\Big[\int_0^\infty dL(s)\Big(\int_{\mathcal{E}} F_s(\mathfrak{e})n(d\mathfrak{e})\Big)\Big], \tag{2.21}$$

where the sum in the left-hand side is taken over all the left-end points of excursion intervals.

2.3 Spectrally negative Lévy taxed processes (SNLTP)

In this sub-section, we will take a quick review on SNLTPs. The idea of investigating how tax influences the behavior of the fluctuation theory was first applied on the Lundberge's risk process by Albrecher and Hipp (2007). Recall that for an insurance conpany, the surplus process is modeled by a classical Lundberg's risk process which has the form

$$R_0(t) = s + ct - \sum_{i=1}^{N_t} Y_i,$$

where s is the initial fund, $\{Y_i\}_{i\geq 1}$ is the claim sizes which followed independent and identical distribution F of mean μ, N_t is an independent Poisson process with intensity λ, and the premium intensity $c > 0$ has a positive safety loading: $c > \lambda\mu$. This process was later replaced by a SNLP X which can capture the functuation of surplus processes better.

To distinguish from the general SNLPs X, we will denote the SNLTP by U throughout this thesis. As mentioned in the introduction that it is more realistic for policy holders paying tax whenever they make profits. For instance, the policy holders will pay an amount of tax $c\gamma(\bar{X})$, where the tax rate $\gamma(\bar{X})$ is a function of $\bar{X}$ which is the situation the company is making profit, and c is the premium rate at profitable time; but do not pay anything otherwise. Therefore, the premium income is reduced from c to $(1-\gamma(\bar{X}))c$ at the time of profit. We first review some results with a fixed tax rate.

2.3.1 Spectrally negative Lévy processes with fixed tax rate

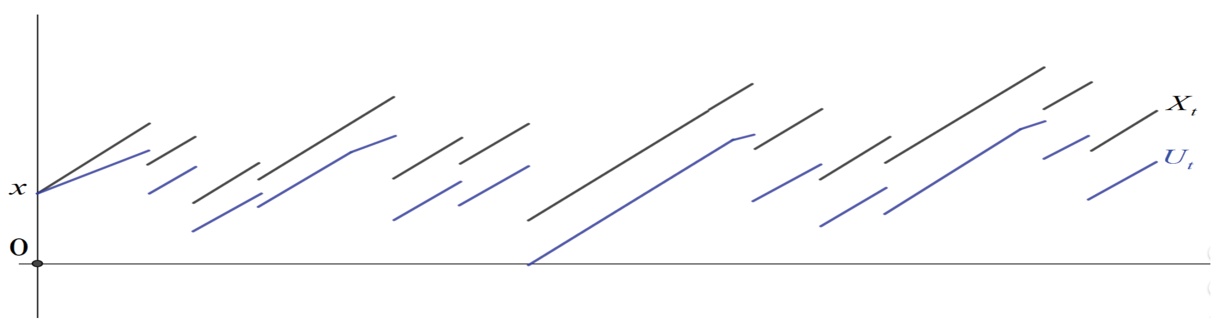

Figure 3: An example of sample path of SNLP X_t along with the SNLTP U_t.

Definition 2.8. *Let X be the underlying SNLP with differentiable scale functions. For a fixed tax rate $0 < \gamma < 1$, define a process $U_\gamma = (U_\gamma(t))_{t\geq 0}$ by*

$$U_\gamma(t) := X(t) - \gamma(\bar{X}(t) - X(0)). \tag{2.22}$$

In terms of finance, U_γ can be interpreted as surplus processes of an insurance company. Given a fixed tax rate γ, the company will pay out an amount $\gamma(\bar{X}_t - X(0))$ of taxes whenever they are in a profitable situation. That is the surpluses are at the running maximum. Particularly, for the case of $\gamma = 1$, the amount the company pays out will be the dividends for any capital above its initial value.

Albrecher *et al.* (2008) generalized the exit problems for SNLTPs as in the following theorem.

Theorem 2.12. *For any $0 < u < a$ and $q \geq 0$, let $\tau^+_{a,\gamma} = \inf\{t > 0 : U_\gamma(t) > a\}$ be the first passage time and $\tau^-_{0,\gamma} = \inf\{t > 0 : U_\gamma(t) < 0\}$ be the ruin time. If $\gamma < 1$, then*

$$\mathbb{E}_u\left[e^{-q\tau^+_{a,\gamma}}; \tau^+_{a,\gamma} < \tau^-_{0,\gamma}\right] = \left(\frac{W^{(q)}(u)}{W^{(q)}(a)}\right)^{1/(1-\gamma)}. \tag{2.23}$$

One can easily verify that the result agrees with Theorem 2.3 above by setting $\gamma = 0$ to get back to the usual SNLPs. Another interested quantity that we want to present here is the discounted tax payments which is given in the following definition.

Definition 2.9. *Given the force of interest $\delta \geq 0$, let $D(t) = \bar{X}(t) - X(0)$ and denote*

$$T_{\gamma,\delta} = \gamma \int_0^{\tau^-_{0,\gamma}} e^{-\delta t} dD(t) \tag{2.24}$$

be the present value of all tax payments until the ruin time $\tau^-_{0,\gamma}$. Then the discounted tax payments $v_{\gamma,\delta}(u)$ of a process is the discounted expectation of the present value of all dividends paid until ruin time when a horizontal barrier is at level u. That is

$$v_{\gamma,\delta}(u) = \mathbb{E}_u(T_{\gamma,\delta}) = \gamma \mathbb{E}_u \int_0^{\tau^-_{0,\gamma}} e^{-\delta t} dD(t).$$

The following theorem gives the expression for the discounted tax payments until ruin time by Albrecher *et al.* (2008).

Theorem 2.13. *If $\gamma < 1$ and $\delta > 0$, then the expected discounted sum of tax payments until ruin time is given by*

$$v_{\gamma,\delta}(u) = \frac{\gamma}{1-\gamma} \int_u^\infty \left(\frac{W^{(q)}(u)}{W^{(q)}(s)}\right)^{1/(1-\gamma)} ds. \tag{2.25}$$

2.3.2 Spectrally negative Lévy processes with general tax rate

In this sub-section, we consider a more general case where the tax rate is allowed to vary in a way that it is a function of running maximum $\bar{X}$. This idea was proposed by Kyprianou and Zhou (2009). The aggregate surplus process $U_{\gamma(\bar{X})}$ is defined as

$$U_{\gamma(\bar{X})}(t) := X(t) - \int_0^t \gamma(\bar{X}(s)) d\bar{X}(s), \tag{2.26}$$

where $\gamma : [0, \infty) \to [0, 1)$ is a measurable function that satisfies $\int_0^\infty (1 - \gamma(s))ds = \infty$. We observe that when γ is a constant in $(0, 1)$, we get back the Definition 2.8. And in the case $\gamma = 0$, we are back to a regular SNLP. Kyprianou and Zhou (2009) were able to obtain some important results. The first one concerns about the two-sided exit problem.

Theorem 2.14. *For any $0 < u < a$ and $q \geq 0$, let $\tau^+_{a,\gamma(\bar{X})} = \inf\{t > 0 : U_{\gamma(\bar{X})}(t) > a\}$ be the first passage time and $\tau^-_{0,\gamma(\bar{X})} = \inf\{t > 0 : U_{\gamma(\bar{X})}(t) < 0\}$ be the ruin time of the surplus process $U_{\gamma(\bar{X})}$. Then we have*

$$\mathbb{E}_u\big[e^{-q\tau^+_{a,\gamma(\bar{X})}}; \tau^+_{a,\gamma(\bar{X})} < \tau^-_{0,\gamma(\bar{X})}\big] = \exp\bigg\{ - \int_u^a \frac{W^{(q)\prime}(y)}{W^{(q)}(y)(1 - \gamma(\bar{\gamma}^{-1}(y)))} du \bigg\}, \tag{2.27}$$

where $\bar{\gamma}(s) := s - \int_u^s \gamma(y)dy$ for $s \geq u$, and $\bar{\gamma}^{-1}$ is its inverse.

And the second result gives the expression of the net present value of tax paid until ruin.

Theorem 2.15. *For any $0 < u < a$, we have*

$$\mathbb{E}_u\bigg[\int_0^{\tau^-_{0,\gamma(\bar{X})}} e^{-qs}\gamma(\bar{X}(s))ds \bigg] = \int_u^\infty \exp\bigg\{ - \int_u^t \frac{W^{(q)\prime}(\bar{\gamma}(s))}{W^{(q)}(\bar{\gamma}(s))} ds \bigg\}\gamma(t)dt. \tag{2.28}$$

2.4 Occupation times of spectrally negative Lévy processes

For the general SNLPs, the Laplace transforms and the joint Laplace transforms of occupation times have been developed with different degrees of complexity and by different approaches. For instance,

- By under-estimating and over-estimating the occupation time by a union of disjoint random time periods which can be solved directly with the help of solutions to the exit problems, then taking an appropriate limit using Lebesgue's dominated convergence theorem twice, Laudriault *et al.* (2011) showed that the two bounds converge to the desired occupation time.

- Later on, Loeffen *et al.* (2014) proposed a different method to obtain the joint Laplace transform of the first passage time and the occupation time using that fact that any SNLP X can be approximated by a sequence $\{X^n\}_{n\geq 1}$ of SNLPs of bounded variation. The result was first derived for the case of a SNLP of bounded variation and expressed in a new extension form of scale functions which will be given in the sub-section below. The case of unbounded variation then was generalized by the dominated convergence theorem.

- Meanwhile, motivated by a new approach concerning the values of the Lévy process observed at independent Poisson arrival times, Li and Zhou (2014) presented the Laplace transforms of pre-exit joint occupation times. Here, the authors studied the joint occupation times over interval $(0, a)$ and (a, b) killed on exiting the interval $[0, b]$. Each sub-interval can be identified as the arrival times of independent Poisson processes with different rate avoiding the time duration when the SNLP occupies that sub-interval. Then, by relying on the key identities initially obtain in Lemma 2.2 of Loeffen *et al.* (2014), Li and Zhou (2014) were able to find the recursive relationship between occupation times at different position at time zero. As a result, the occupation time is obtained by solving the recursive equations.

We begin with an introduction to new form of scale functions.

2.4.1 Extension of scale functions

In order to express the solutions to occupation times, we present some extensions of scale functions. By taking Laplace transforms on both sides of (2.6), we obtain

$$\begin{aligned}(q-p)\int_0^a W^{(p)}(a-y)W^{(q)}(y)dy &= W^{(q)}(a)-W^{(p)}(a),\\ (q-p)\int_0^a W^{(p)}(a-y)Z^{(q)}(y)dy &= Z^{(q)}(a)-Z^{(p)}(a).\end{aligned}$$

Then, using the properties above we can define these functions below in terms of q-scale function $W^{(q)}$ and $Z^{(q)}$ as

$$\begin{aligned}\mathcal{W}_a^{(p,q)}(x) &:= W^{(p+q)}(x)-q\int_0^a W^{(p+q)}(x-y)W^{(p)}(y)dy\\ &= W^{(p)}(x)+q\int_a^x W^{(p+q)}(x-y)W^{(p)}(y)dy, \qquad (2.29)\end{aligned}$$

$$\begin{aligned}\mathcal{Z}_a^{(p,q)}(x) &:= Z^{(p+q)}(x)-q\int_0^a W^{(p+q)}(x-y)Z^{(p)}(y)dy\\ &= Z^{(p)}(x)+q\int_a^x W^{(p+q)}(x-y)Z^{(p)}(y)dy. \qquad (2.30)\end{aligned}$$

These two functions above were used to express the joint Laplace transforms of occupation time for SNLPs in Loeffen *et al.* (2014) and Li and Zhou (2014).

Finally, the idea of scale functions is generalized to ω-scale functions suggested by Li and Palmowski (2018) as follows. First, we let $\omega : \mathbb{R} \to \mathbb{R}_+$ be a locally bounded nonnegative measurable function, then the two families of functions: $\{\mathcal{W}^{(\omega)}(x, y), x \in \mathbb{R}, y \in \mathbb{R}\}$ and

$\{\mathcal{Z}^{(\omega)}(x,y), x\in\mathbb{R}, y\in\mathbb{R}\}$, are defined uniquely as the solutions to the following equations:

$$\mathcal{W}^{(\omega)}(x,y) = W(x-y) + \int_y^x W(x-z)\omega(z)\mathcal{W}^{(\omega)}(z,y)dz, \tag{2.31}$$

$$\mathcal{Z}^{(\omega)}(x,y) = 1 + \int_y^x W(x-z)\omega(z)\mathcal{Z}^{(\omega)}(z,y)dz, \tag{2.32}$$

respectively, where $W(x)$ is a classical 0-scale function. For special case of a constant function $\omega(x) = q$, it was shown in Li and Palmowski (2018) that $(\mathcal{W}^{(\omega)}(x,0), \mathcal{Z}^{(\omega)}(x,0)) = (W^{(q)}(x), Z^{(q)}(x))$.

Moreover, for $p, q \geq 0$ by taking

$$\omega(x) = p + q\mathbf{1}_{(a,b)}(x),$$

the two ω-scale functions have a form

$$\mathcal{W}^{(\omega)}(x) = \mathcal{W}_a^{(p,q)}(x) - q\int_b^x W^{(p)}(x-z)\mathcal{W}_a^{(p,q)}(z)dz, \tag{2.33}$$

$$\mathcal{Z}^{(\omega)}(x) = \mathcal{Z}_a^{(p,q)}(x) - q\int_b^x W^{(p)}(x-z)\mathcal{Z}_a^{(p,q)}(z)dz. \tag{2.34}$$

We are now ready to state some interesting results concerning about occupation times of a general SNLP.

2.4.2 Previous results on occupation times of SNLPs

The very first result is the total occupation time of the negative half-line $(-\infty, 0)$ found by Laudriault *et al.* (2011).

Theorem 2.16. *If $\psi'(0+) > 0$, then for $q \geq 0$*

$$\mathbb{E}\left[e^{-q\int_0^\infty \mathbf{1}_{\{X_t\leq 0\}}dt}\right] = \psi'(0+)\frac{\phi(q)}{q}, \tag{2.35}$$

where $\phi(q)/q$ is to be understood in the limiting sense when $q = 0$. Or more generally for $x \geq 0$

$$\mathbb{E}_x\left[e^{-q\int_0^\infty \mathbf{1}_{\{X_t\leq 0\}}dt}\right] = \psi'(0+)\phi(q)\int_0^\infty e^{-\phi(q)z}W(x+z)dz. \tag{2.36}$$

Also, they found the occupation time of $(-\infty, 0)$ until a negative level $-b$, $b > 0$, is crossed for the first time.

Theorem 2.17. *If $\psi'(0+) \geq 0$, then for $q \geq 0$*

$$\mathbb{E}\left[e^{-q\int_0^{\tau_{-b}^-} \mathbf{1}_{\{X_t\leq 0\}}dt}\right] = \frac{\psi'(0+) + \frac{\sigma^2}{2}\frac{A_1^{(q)}(b)}{W^{(q)}(b)} + \int_{-\infty}^{0-} A_2^{(q)}(x)\int_0^\infty \Pi(dx-y)dy}{\psi'(0+) + \frac{\sigma^2}{2}\frac{W'^{(q)}(b)}{W^{(q)}(b)} + \int_{-\infty}^{0-} A_3^{(q)}(x)\int_0^\infty \Pi(dx-y)dy}, \tag{2.37}$$

where

$$\begin{aligned} A_1^{(q)}(b) &= Z^{(q)}(b)W'^{(q)}(b) - q(W^{(q)}(b))^2, \\ A_2^{(q)}(x) &= Z^{(q)}(x+b) - Z^{(q)}(b)\frac{W^{(q)}(x+b)}{W^{(q)}(b)}, \\ A_3^{(q)}(x) &= 1 - \frac{W^{(q)}(x+b)}{W^{(q)}(b)}. \end{aligned}$$

In their paper, the main idea in the proof of the two theorems above is based on under-estimating and over-estimating the occupation time. Later on in Loeffen *et al.* (2014), they extended the result to the joint Laplace transforms of occupation time given a closed interval $[a, b] \subset [0, \infty)$ with the ruin time τ_0^-.

Theorem 2.18. *For* $0 \le a \le b \le c$, $p, q \ge 0$ *and* $0 \le x \le c$,

$$\mathbb{E}_x\left[e^{-p\tau_0^- - q\int_0^{\tau_0^-} \mathbf{1}_{(a,b)}(X_t)dt}; \tau_0^- < \tau_c^+\right] = \mathcal{Z}_a^{(p,q)}(x) - q\int_b^x W^{(p)}(x-z)\mathcal{Z}_a^{(p,q)}(z)dz$$

$$-\left(\mathcal{W}_a^{(p,q)}(x) - q\int_b^x W^{(p)}(x-z)\mathcal{W}_a^{(p,q)}(z)dz\right)\frac{\mathcal{Z}_a^{(p,q)}(c) - q\int_b^c W^{(p)}(c-z)\mathcal{Z}_a^{(p,q)}(z)dz}{\mathcal{W}_a^{(p,q)}(c) - q\int_b^c W^{(p)}(c-z)\mathcal{W}_a^{(p,q)}(z)dz}, \quad (2.38)$$

and

$$\mathbb{E}_x\left[e^{-p\tau_c^+ - q\int_0^{\tau_c^+} \mathbf{1}_{(a,b)}(X_t)dt}; \tau_c^+ < \tau_0^-\right] = \frac{\mathcal{W}_a^{(p,q)}(x) - q\int_b^x W^{(p)}(x-z)\mathcal{W}_a^{(p,q)}(z)dz}{\mathcal{W}_a^{(p,q)}(c) - q\int_b^c W^{(p)}(c-z)\mathcal{W}_a^{(p,q)}(z)dz}. \quad (2.39)$$

Meanwhile, Li and Zhou (2014) considered the Laplace transforms of joint occupation times over disjoint intervals $(0, a)$ and (a, b) before it first exits interval $(0, b)$ for $0 < a < b$. By identifying this joint Laplace transform with the probability that two independent sequences of Poisson arrival times, they derived the results in the theorem below.

Theorem 2.19. *For any* $0 < a < b$, $0 \le x \le b$, *and* $p, q \ge 0$, *we have*

$$\begin{aligned} \mathbb{E}_x\left[e^{-p\int_0^{\tau_b^+} \mathbf{1}_{(0,a)}(X_t)dt - q\int_0^{\tau_b^+} \mathbf{1}_{(a,b)}(X_t)dt}; \tau_b^+ < \tau_0^-\right] &= \frac{\mathcal{W}_a^{(p,q)}(x)}{\mathcal{W}_a^{(p,q)}(b)}, \\ \mathbb{E}_x\left[e^{-p\int_0^{\tau_0^-} \mathbf{1}_{(0,a)}(X_t)dt - q\int_0^{\tau_0^-} \mathbf{1}_{(a,b)}(X_t)dt}; \tau_0^- < \tau_b^+\right] &= \mathcal{Z}_a^{(p,q)}(x) - \mathcal{Z}_a^{(p,q)}(b)\frac{\mathcal{W}_a^{(p,q)}(x)}{\mathcal{W}_a^{(p,q)}(b)}. \end{aligned}$$

We end this sub-section by some results in Li and Palmowski (2018) concerning the ω-type resolvent. The following theorem gives the expression of ω-type resolvent for SNLPs.

Theorem 2.20. *For $x, y \in [0, c]$, we have*

$$U^{(\omega)}(x, dy) := \int_0^\infty \mathbb{E}_x\left[\exp\left(-\int_0^t \omega(X_s)ds\right); t < \tau_0^- \wedge \tau_c^+, X_t \in dy\right]dt$$
$$= \left(\frac{\mathcal{W}^{(\omega)}(x,0)}{\mathcal{W}^{(\omega)}(c,0)}\mathcal{W}^{(\omega)}(c,y) - \mathcal{W}^{(\omega)}(x,y)\right)dy. \quad (2.40)$$

The power of this type of ω-scale function is that, with an appropriate choice of ω, we can obtain the various expressions involving the occupation time. For instance, by letting $\omega(x) = p + q\mathbf{1}_{(a,b)}(x)$, and $(a, b) \subset [0, c]$, we get

$$U^{(\omega)}(x, dy) = \int_0^\infty \mathbb{E}_x\left[\exp\left(-\int_0^t (p + q\mathbf{1}_{(a,b)}(X_s))ds\right); t < \tau_0^- \wedge \tau_c^+, X_t \in dy\right]dt$$
$$= \int_0^\infty e^{-pt}\mathbb{E}_x\left[\exp\left(-q\int_0^t (\mathbf{1}_{(a,b)}(X_s))ds\right); t < \tau_0^- \wedge \tau_c^+, X_t \in dy\right]dt. \quad (2.41)$$
$$= \left(\frac{\mathcal{W}^{(\omega)}(x,0)}{\mathcal{W}^{(\omega)}(c,0)}\mathcal{W}^{(\omega)}(c,y) - \mathcal{W}^{(\omega)}(x,y)\right)dy.$$

where $\mathcal{W}^{(\omega)}(x, y)$ is the unique solution to the equations:

$$\mathcal{W}^{(\omega)}(x,y) = W(x-y) + \int_y^x W(x-z)\omega(z)\mathcal{W}^{(\omega)}(z,y)dz.$$

The above expression is the resolvent of the occupation time on the interval (a, b) killed on exiting the external interval $[0, c]$. Moreover, Li and Palmowski (2018) obtained the similar result for the reflected process Y.

Theorem 2.21. *For $x, y \in [0, c]$, define $T_c = \inf\{t \geq 0 : Y_t > c\}$, then we have*

$$L^{(\omega)}(x, dy) := \int_0^\infty \mathbb{E}_x\left[\exp\left(-\int_0^t \omega(Y_s)ds\right); t < T_c, Y_t \in dy\right]dt$$
$$= \left(\frac{\mathcal{Z}^{(\omega)}(x,0)}{\mathcal{Z}^{(\omega)}(c,0)}\mathcal{W}^{(\omega)}(c,y) - \mathcal{W}^{(\omega)}(x,y)\right)dy. \quad (2.42)$$

where $\mathcal{Z}^{(\omega)}(x, y)$ is the unique solution to the equations:

$$\mathcal{Z}^{(\omega)}(x,y) = 1 + \int_y^x W(x-z)\omega(z)\mathcal{Z}^{(\omega)}(z,y)dz,$$

3 Fluctuation results related to draw-down times

As mention in the Introduction Section, draw-down time has recently become more and more considerably interesting because of its practical application. For instance, in risk theory, suppose the price process of an underlying security is modeled by a SNLP X, and we are consider the decision to sell that stock. Obviously, it is ideal to sell a stock at the highest price or just right before it starts to decline. However, it is not easy to sell a stock at this ideal time, and it totally depends on luck. Instead, it would be reasonable to sell a stock either when it hits a price target, says, thirty percents drop from the previous maximum. That means, the stock will be sold at a draw-down time. As a result, studying the Laplace transform of the draw-down time will help us to understand the fluctuation of the time to sell stocks. In the above example, the draw-down function is $\xi(x) = 0.7x$ and it is a measurable non-stochastic function in $\mathbb{R}$. One may consider the case of stochastic draw-down function. However, since it is not very interesting, we only consider the case of no-stichastic draw-down function.

Our goal in this section is to generalize some results in the Section 2 with the association of draw-down times. More specifically, we will solve the two-sided exit problems for the general SNLP X and SNLTP U with draw-down times; the expression of discounted tax payment of SNLTPs accumulated up to a draw-down time will be given; and lastly we will generalize the occupation times of a general SNLP in a given draw-down interval.

Firstly, we start by introducing the notion of draw-down times of a general SNLP. Let ξ be a measurable function on $\mathbb{R}$, then we define the draw-down time with respect to the draw-down function ξ as

$$\tau_\xi := \inf\{t \geq 0 : X(t) < \xi(\bar{X}_t)\} = \inf\{t > 0 : Y_t > \bar{\xi}(\bar{X}_t)\},$$

where $Y_t := \bar{X}_t - X_t$ is the reflected process of X at its running maximum, $\bar{\xi}(z) := z - \xi(z) > 0$ and $\{\xi(\bar{X}_t), t \geq 0\}$ is the associated draw-down level process. Indeed, draw-down times can be interpreted as the perceptual accumulated loss due to a sequence of drops in prices of an investment.

3.1 General SNLPs with draw-down times

In this sub-section, we will work out the two-sided exit problems for the general SNLPs with the draw-down time τ_ξ in place of the ruin time τ_0^-. In order to archive this, we rely mainly on the excursion theory for Markov processes in which the compensation formula and the exponential formula for Poisson point processes are employed. Before stating our main results, we want to present some notation regarding to the local time.

It is well-known that the reflected process Y is a Markov process with 0 being instantaneous whenever $W(0) = 0$, so we can define a local time process L of Y at 0 that is unique up to a multiplicative factor ν. That is, for all $t \geq 0$

$$\int_0^t \mathbf{1}_{\{s \in \overline{\mathcal{L}}\}} ds = \nu L(t), \tag{3.1}$$

where $\mathcal{L} := \{t > 0, Y_t = 0\}$ is the zero set of Y and $\overline{\mathcal{L}}$ is its closure. The inverse local time is defined as

$$L_t^{-1} := \inf\{s > 0 : L(s) > t\}, \qquad t \geq 0.$$

Under the new time scale, the excursion process of Y away from zero, associated with L and denoted by $\epsilon = \{\epsilon_r, r \geq 0\}$ takes values in the excursion space $\mathcal{E}$ with an additional isolated point ∂. Where,

$$\epsilon_r := \begin{cases} \{Y_t, 0 \leq t \leq L_r^{-1} - L_{r-}^{-1}\} & \text{if } L_{r-}^{-1} < L_r^{-1}, \\ \partial & \text{otherwise.} \end{cases}$$

Since $\mathcal{L}$ differs from $\overline{\mathcal{L}}$ by at most countable points, we can rewrite (3.1) as

$$\int_0^{L_r^{-1}} \mathbf{1}_{\{s \in \overline{\mathcal{L}}\}} ds = \nu r. \tag{3.2}$$

We first consider the case for an SNLP X under $\mathbb{P}$, then by shifting argument, we can generalize our results to the case under $\mathbb{P}_x$. It is worthy to notice that under $\mathbb{P}$, the running maximum process $\bar{X}$ meets all the requirements to be a local time, so it is the best choice to let $\bar{X}$ be the local time of Y at 0. The benefits of this setup is that we get $L_s^{-1} = \tau_s^+$ as a subordinator with Laplace exponent Φ, and

$$\nu = \lim_{s\to\infty} \frac{\Phi(s)}{s} = \lim_{s\to\infty} \frac{s}{\psi(s)} = W(0).$$

We are now ready to state our first results concerning the two-sided exit problems of SNLP X associated with the draw-down time τ_ξ.

Theorem 3.1 (Part 1 of Proposition 3.1 in Li *et al.* (2019))**.** *For any $q > 0$ and $x < b$, we have*

$$\mathbb{E}_x\left[e^{-q\tau_b^+}; \tau_b^+ < \tau_\xi\right] = \exp\left\{-\int_x^b \frac{W^{(q)\prime}(\bar{\xi}(y))}{W^{(q)}(\bar{\xi}(y))} dy\right\}. \tag{3.3}$$

Theorem 3.2 (Part 2 of Proposition 3.1 in Li *et al.* (2019))**.** *For any $q > 0$ and $x < b$, we have*

$$\mathbb{E}_x\left[e^{-q\tau_\xi}; \tau_\xi < \tau_b^+\right] = \int_x^b e^{-\int_x^z \frac{W^{(q)\prime}(\bar{\xi}(y))}{W^{(q)}(\bar{\xi}(y))}dy} \left(\frac{W^{(q)\prime}(\bar{\xi}(z))}{W^{(q)}(\bar{\xi}(z))} Z^{(q)}(\bar{\xi}(z)) - qW^{(q)}(\bar{\xi}(z))\right) dz. \tag{3.4}$$

Remark 3.1. *If we fix the draw-down function $\xi(z) = c$ for all $c \leq x \leq b$ in both of the* *theorems above, then we recover the classical two-sided exit problems. Indeed*

$$\begin{aligned}
\mathbb{E}_x\Big[e^{-q\tau_b^+};\tau_b^+<\tau_c^-\Big] &= \exp\Big\{-\int_x^b \frac{W^{(q)\prime}(y-c)}{W^{(q)}(y-c)}dy\Big\}\\
&= \exp\Big\{-\int_{x-c}^{b-c} \frac{W^{(q)\prime}(y)}{W^{(q)}(y)}dy\Big\}\\
&= \exp\Big\{-\int_{x-c}^{b-c} \frac{d}{dy}\ln(W^{(q)}(y))dy\Big\}\\
&= \exp\Big\{\ln(W^{(q)}(x-c))-\ln(W^{(q)}(b-c))\Big\}\\
&= \frac{W^{(q)}(x-c)}{W^{(q)}(b-c)}.
\end{aligned}$$

And since $Z^{(q)\prime}(x)=qW^{(q)}(x)$, we have

$$\begin{aligned}
\mathbb{E}_x\Big[e^{-q\tau_c^-};\tau_c^-<\tau_b^+\Big] &= \int_x^b e^{-\int_x^z \frac{W^{(q)\prime}(y-c)}{W^{(q)}(y-c)}dy}\Big(\frac{W^{(q)\prime}(z-c)}{W^{(q)}(z-c)}Z^{(q)}(z-c)-qW^{(q)}(z-c)\Big)dz\\
&= \int_x^b \frac{W^{(q)}(x-c)}{W^{(q)}(z-c)}\Big(\frac{W^{(q)\prime}(z-c)}{W^{(q)}(z-c)}Z^{(q)}(z-c)-Z^{(q)\prime}(z-c)\Big)dz\\
&= W^{(q)}(x-c)\int_{x-c}^{b-c}\frac{W^{(q)\prime}(z)Z^{(q)}(z)-Z^{(q)\prime}(z)W^{(q)}(z)}{(W^{(q)}(z))^2}dz\\
&= W^{(q)}(x-c)\int_{x-c}^{b-c}\frac{d}{dz}\Big(-\frac{Z^{(q)}(z)}{W^{(q)}(z)}\Big)dz\\
&= Z^{(q)}(x-c)-\frac{W^{(q)}(x-c)}{W^{(q)}(b-c)}Z^{(q)}(b-c).
\end{aligned}$$

Remark 3.2. *If we let $q\to 0$ in the two theorems above, we obtain an important probability concerning of the occurrence of first passage time and draw-down time.*

$$\mathbb{P}_x\big[\tau_b^+<\tau_\xi\big] = \exp\Big\{-\int_x^b \frac{W'(\bar{\xi}(y))}{W(\bar{\xi}(y))}dy\Big\}. \tag{3.5}$$

Moreover, by applying the change of measure on Theorem 3.2, we can obtain the joint Laplace transform with additional SNLP X at the draw-down time. Thus, we have the following corollary.

Corollary 3.1. *For any $p,q>0$ and $x<b$, by defining $u:=q-\psi(p)$ we have*

$$\mathbb{E}_x\Big[e^{-q\tau_\xi+pX(\tau_\xi)};\tau_\xi<\tau_b^+\Big]=\int_x^b e^{px-\int_x^z\frac{W_p^{(u)\prime}(\bar{\xi}(y))}{W_p^{(u)}(\bar{\xi}(y))}dy}\Big(\frac{W_p^{(u)\prime}(\bar{\xi}(z))}{W_p^{(u)}(\bar{\xi}(z))}Z_p^{(u)}(\bar{\xi}(z))-uW_p^{(u)}(\bar{\xi}(z))\Big)dz. \tag{3.6}$$

where $W_p^{(u)}(\cdot)$ and $Z_p^{(u)}(\cdot)$ are scale functions under the new measure $\mathbb{P}^{(p)}$.

Example of the linear draw-down function

Suppose we consider the linear draw-down function which defined as

$$\xi(t) = \xi t - d$$

for some constant and $\xi, d > 0$. Then the draw-down time of SNLP X has the form

$$\tau_{d,\xi} = \inf\{t \geq 0 : X(t) + d < \xi \bar{X}(t)\}.$$

As a result, we can work out the explicit expression of (3.3) and (3.4). For example

$$\begin{aligned}
\mathbb{E}_x\left[e^{-q\tau_b^+}; \tau_b^+ < \tau_{d,\xi}\right] &= \exp\left\{-\int_x^b \frac{W^{(q)\prime}(\bar{\xi}(y))}{W^{(q)}(\bar{\xi}(y))}dy\right\} \\
&= \exp\left\{-\int_x^b \frac{W^{(q)\prime}((1-\xi)y+d)}{W^{(q)}((1-\xi)y+d)}dy\right\} \\
&= \exp\left\{-\frac{1}{1-\xi}\int_{(1-\xi)x+d}^{(1-\xi)b+d} \frac{W^{(q)\prime}(y)}{W^{(q)}(y)}dy\right\} \\
&= \exp\left\{-\frac{1}{1-\xi}\int_{(1-\xi)x+d}^{(1-\xi)b+d} \frac{d}{dy}\ln(W^{(q)}(y))dy\right\} \\
&= \exp\left\{-\frac{1}{1-\xi}\Big[\ln(W^{(q)}(y))\Big]_{(1-\xi)x+d}^{(1-\xi)b+d}\right\} \\
&= \left(\frac{W^{(q)}((1-\xi)x+d)}{W^{(q)}((1-\xi)b+d)}\right)^{\frac{1}{1-\xi}}.
\end{aligned}$$

The rest of this sub-section will be devoted to the proof of the two theorems and the corollary.

Proof of Theorem 3.1.

Recall that for a generic excursion $\varepsilon \in \mathcal{E}$, we denote its lifetime by $\zeta = \zeta(\varepsilon)$ and its excursion height by $\bar{\varepsilon}$. Also, we define its first passage time as

$$\rho_c^+ = \rho_c^+(\varepsilon) := \inf\{s \in (0, \zeta) : \varepsilon(s) > c\}.$$

If $\tau_b^+ < \infty$ a.s., then under $\mathbb{P}$ we can rewrite the first passage time τ_b^+ as

$$\tau_b^+ = L_b^{-1} = \int_0^{L_b^{-1}} \mathbf{1}_{\{t \in \overline{\mathcal{L}}\}}dt + \int_0^{L_b^{-1}} \mathbf{1}_{\{t \notin \overline{\mathcal{L}}\}}dt = \nu b + \sum_{r \in [0,b]} \zeta(\epsilon_r). \tag{3.7}$$

Where the first term after the last equation follows by (3.2) and the second term is the total lifetime of all excursions occurring up to the first passage time τ_b^+. Similarly, the draw-down time can be decomposed into the first passage time of the first excursion whose height surpasses the draw-down level, plus all the previous time before this excursion. That is the draw-down time τ_ξ can be rewritten as

$$\tau_\xi = L^{-1}_{r-} + \rho^{+}_{\bar{\xi}(r)}(\epsilon_r) : r \geq 0 \qquad \text{where} \;\; \bar{X}(\tau_\xi) = L(\tau_\xi) = r. \tag{3.8}$$

Now, considering the event $\{\tau_b^+ < \tau_\xi\}$, which is the event that the SNLP X reaches the level b before the draw-down level, we see that it is equivalent to the event that the excursion height $\bar{\epsilon}_r$ is at most the draw-down depth $\bar{\xi}(r)$ for all local time $r \in [0, b]$. That is

$$\{\tau_b^+ < \tau_\xi\} = \{\bar{\epsilon}_r \leq \bar{\xi}(r) \;\text{ for all } r \in [0,b]\} = \bigcap_{r\in[0,b]} \{\bar{\epsilon}_r \leq \bar{\xi}(r)\}. \tag{3.9}$$

Therefore, from (3.7), (3.9), and with the help of the Exponential Formula, we have

$$\begin{aligned}
\mathbb{E}\Big[e^{-q\tau_b^+}; \tau_b^+ < \tau_\xi\Big] &= \mathbb{E}\Big[e^{-q\nu b - q\sum_{r\in[0,b]}\zeta(\epsilon_r)}; \tau_b^+ < \tau_\xi\Big] \\
&= e^{-q\nu b}\mathbb{E}\Big[e^{-q\sum_{r\in[0,b]}\zeta(\epsilon_r)} \prod_{r\in[0,b]} \mathbf{1}_{\{\bar{\epsilon}_r \leq \bar{\xi}(r)\}}\Big] \\
&= e^{-q\nu b}\mathbb{E}\Big[\exp\Big\{-q \sum_{r\in[0,b]} \big(\zeta(\epsilon_r) + \infty \cdot \mathbf{1}_{\{\bar{\epsilon}_r \geq \bar{\xi}(r)\}}\big)\Big\}\Big] \\
&= \exp\Big\{-\int_0^b \Big(q\nu + \int_{\mathcal{E}} \big(1 - e^{-q\zeta}\mathbf{1}_{\{\bar{\varepsilon} \leq \bar{\xi}(r)\}}\big) n(d\varepsilon)\Big) dr\Big\},
\end{aligned}$$

with the understanding that $e^{-\infty} = 0$ and $\infty \cdot 0 = 0$. Clearly, it is not possible to directly integrate the integral with respect to excursion measure. However, if we fix the draw-down function to a constant $\xi \equiv c < 0$, then we get back to the classical two-sided exit problems whose solution is of the form

$$\frac{W^{(q)}(-c)}{W^{(q)}(b-c)} = \mathbb{E}_x\Big[e^{-q\tau_b^+}; \tau_b^+ < \tau_c^-\Big] = \exp\Big\{-\int_0^b \Big(q\nu + \int_{\mathcal{E}} \big(1 - e^{-q\zeta}\mathbf{1}_{\{\bar{\varepsilon} \leq r-c\}}\big) n(d\varepsilon)\Big) dr\Big\}.$$

Then, by taking log and differentiating with respect to b on both sides of the equation above, we obtain

$$q\nu + \int_{\mathcal{E}} \big(1 - e^{-q\zeta}\mathbf{1}_{\{\bar{\varepsilon} \leq b-c\}}\big) n(d\varepsilon) = \frac{W^{(q)\prime}(b-c)}{W^{(q)}(b-c)}.$$

This implies that, for $z > 0$

$$q\nu + \int_{\mathcal{E}} \big(1 - e^{-q\zeta}\mathbf{1}_{\{\bar{\varepsilon} \leq z\}}\big) n(d\varepsilon) = \frac{W^{(q)\prime}(z)}{W^{(q)}(z))}.$$

Therefore,

$$\mathbb{E}\Big[e^{-q\tau_b^+}; \tau_b^+ < \tau_\xi\Big] = \exp\Big\{-\int_0^b \frac{W^{(q)\prime}(\bar{\xi}(y))}{W^{(q)}(\bar{\xi}(y))} dy\Big\}.$$

To generalize this result to the case $\mathbb{P}_x$ with $x < b$, we define a function $\varsigma(y) := \xi(y + x) - x$.

Then

$$\bar{\varsigma}(y) = y + x - \xi(y+x) = \bar{\xi}(y+x).$$

Because X is spatially homogeneous, we have $(X, \bar{X}, \tau_\xi)\big|_{\mathbb{P}_x} = (x + X, x + \bar{X}, \tau_\varsigma)\big|_{\mathbb{P}}$. Hence,

$$\begin{aligned}
\mathbb{E}_x\Big[e^{-q\tau_b^+}; \tau_b^+ < \tau_\xi\Big] &= \mathbb{E}\Big[e^{-q\tau_{b-x}^+}; \tau_{b-x}^+ < \tau_\varsigma\Big] \\
&= \exp\Big\{ - \int_0^{b-x} \frac{W^{(q)\prime}(\bar{\varsigma}(y))}{W^{(q)}(\bar{\varsigma}(y))} dy \Big\} \\
&= \exp\Big\{ - \int_x^{b} \frac{W^{(q)\prime}(\bar{\xi}(y))}{W^{(q)}(\bar{\xi}(y))} dy \Big\}.
\end{aligned}$$

□

In summary, the idea in the proof of Theorem 3.1 above is as following. First, we try to express the expectation in terms of excursion processes, excursion highs, and the excursion lifetimes, then using the exponential formula (or the conpensation formula in later proofs) we can associate the excursion measure into the expression. With the help of the previous result on the fixed constant draw-down function, we are able to obtain the explicit form of the associated excursion measure. Therefore, the results follow by plugging back into the previous expression. Indeed, the key factor in the proof is to obtain the expression of the associated excursion measure, and we will carry this approach throughout most of the proofs in this section.

Proof of Theorem 3.2. We begin by considering the event $\{\tau_\xi < \tau_b^+\}$ which can be expressed as

$$\{\tau_\xi < \tau_b^+\} \equiv \Big\{ \bigcup_{r\in[0,b]} \Big(\Big[\bigcap_{s<r} \{\bar{\epsilon}_s \le \bar{\xi}(s)\} \Big] \{\bar{\epsilon}_r > \bar{\xi}(r)\} \Big) \Big\}.$$

Then, we have

$$\begin{aligned}
\mathbb{E}\Big[e^{-q\tau_\xi}; \tau_\xi < \tau_b^+\Big] &= \mathbb{E}\Big[\sum_{r\in[0,b]} e^{-q\left(L_{r-}^{-1} + \rho_{\bar{\xi}(r)}^+(\epsilon_r)\right)} \Big(\prod_{s<r} \mathbf{1}_{\{\bar{\epsilon}_s \le \bar{\xi}(s)\}} \Big) \mathbf{1}_{\{\bar{\epsilon}_r > \bar{\xi}(r)\}} \Big] \\
&= \mathbb{E}\Big[\sum_{r\in[0,b]} \Big(e^{-qL_{r-}^{-1}} \prod_{s<r} \mathbf{1}_{\{\bar{\epsilon}_s \le \bar{\xi}(s)\}} \Big) e^{-q\rho_{\bar{\xi}(r)}^+(\epsilon_r)} \mathbf{1}_{\{\bar{\epsilon}_r > \bar{\xi}(r)\}} \Big] \\
&= \mathbb{E}\Big[\int_0^\infty \Big(e^{-qt} \prod_{s<L(t)} \mathbf{1}_{\{\bar{\epsilon}_s \le \bar{\xi}(s)\}} \Big) \Big(\int_{\mathcal{E}} e^{-q\rho_{\bar{\xi}(L(t))}^+(\varepsilon)} \mathbf{1}_{\{\bar{\varepsilon} > \bar{\xi}(L(t))\}} n(d\varepsilon) \Big) dL(t) \Big] \\
&= \mathbb{E}\Big[\int_0^b \Big(e^{-qL_{r-}^{-1}} \prod_{s<r} \mathbf{1}_{\{\bar{\epsilon}_s \le \bar{\xi}(s)\}} \Big) \Big(\int_{\mathcal{E}} e^{-q\rho_{\bar{\xi}(r)}^+} \mathbf{1}_{\{\bar{\varepsilon} > \bar{\xi}(r)\}} n(d\varepsilon) \Big) dr \Big] \\
&= \int_0^b \mathbb{E}\Big[e^{-qL_{r-}^{-1}} \prod_{s<r} \mathbf{1}_{\{\bar{\epsilon}_s \le \bar{\xi}(s)\}} \Big] \times n\big(e^{-q\rho_{\bar{\xi}(r)}^+}; \bar{\varepsilon} > \bar{\xi}(r)\big) dr
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^b \mathbb{E}\Big[e^{-qL_r^{-1}}; L_r^{-1} < \tau_\xi\Big] \times n\big(e^{-q\rho^+_{\bar\xi(r)}}; \bar\varepsilon > \bar\xi(r)\big)dr \\
&= \int_0^b \mathbb{E}\Big[e^{-q\tau_r^+}; \tau_r^+ < \tau_\xi\Big] \times n\big(e^{-q\rho^+_{\bar\xi(r)}}; \bar\varepsilon > \bar\xi(r)\big)dr \\
&= \int_0^b e^{-\int_0^r \frac{W^{(q)\prime}(\bar\xi(y))}{W^{(q)}(\bar\xi(y))}dy} \times n\big(e^{-q\rho^+_{\bar\xi(r)}}; \bar\varepsilon > \bar\xi(r)\big)dr.
\end{aligned}
$$

Where we applied the Compensation Formula in the fourth equation; the change of variable $r = L(t)$ is used in the fifth equation, and the seventh equation follows by the fact that L_r^{-1} differs from L_{r-}^{-1} by at most countable points. Again, the only unknown quantity in the equation above is the excursion measure $n\big(e^{-q\rho^+_{\bar\xi(r)}} : \bar\varepsilon > \bar\xi(r)\big)$ which can be derived by the same approach in the proof of Theorem 3.1. Particularly, considering the case where $\xi \equiv c < 0$, we have

$$
\begin{aligned}
Z^{(q)}(-c) - \frac{W^{(q)}(-c)}{W^{(q)}(b-c)} Z^{(q)}(b-c) &= \mathbb{E}\Big[e^{-q\tau_c^-}; \tau_c^- < \tau_b^+\Big] \\
&= \int_0^b \frac{W^{(q)}(-c)}{W^{(q)}(r-c)} \times n\big(e^{-q\rho^+_{r-c}} : \bar\varepsilon > r-c\big)dr.
\end{aligned}
$$

By differentiating with respect to b on both sides of the equation above, we get

$$
n\big(e^{-q\rho_z^+} : \bar\varepsilon > z\big) = \frac{W^{(q)\prime}(z)}{W^{(q)}(z)} Z^{(q)}(z) - qW^{(q)}(z), \qquad \text{for } z > 0. \tag{3.10}
$$

As a result,

$$
\mathbb{E}\Big[e^{-q\tau_\xi}; \tau_\xi < \tau_b^+\Big] = \int_0^b e^{-\int_0^z \frac{W^{(q)\prime}(\bar\xi(y))}{W^{(q)}(\bar\xi(y))}dy} \left(\frac{W^{(q)\prime}(\bar\xi(z))}{W^{(q)}(\bar\xi(z))} Z^{(q)}(\bar\xi(z)) - qW^{(q)}(\bar\xi(z))\right)dz.
$$

And by shifting argument, we again obtain

$$
\mathbb{E}_x\Big[e^{-q\tau_\xi}; \tau_\xi < \tau_b^+\Big] = \int_x^b e^{-\int_x^z \frac{W^{(q)\prime}(\bar\xi(y))}{W^{(q)}(\bar\xi(y))}dy} \left(\frac{W^{(q)\prime}(\bar\xi(z))}{W^{(q)}(\bar\xi(z))} Z^{(q)}(\bar\xi(z)) - qW^{(q)}(\bar\xi(z))\right)dz.
$$

□

Proof of Corollary 3.1.

Recall that, for each $c \geq 0$, we have the following exponential change of measure

$$
\left.\frac{d\mathbb{P}^{(c)}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = e^{cX_t - \psi(c)t}.
$$

Let $u := q - \psi(p) > 0$, taking this change of measure on the expectation for the case $X(0) = 0$ we get

$$\mathbb{E}\left[e^{-q\tau_\xi+pX(\tau_\xi)};\tau_\xi<\tau_b^+\right]=\mathbb{E}^{(p)}\left[e^{-u\tau_\xi};\tau_\xi<\tau_b^+\right]$$
$$=\int_0^b e^{-\int_0^z \frac{W_p^{(u)\prime}(\bar{\xi}(y))}{W_p^{(u)}(\bar{\xi}(y))}dy}\left(\frac{W_p^{(u)\prime}(\bar{\xi}(z))}{W_p^{(u)}(\bar{\xi}(z))}Z_p^{(u)}(\bar{\xi}(z))-uW_p^{(u)}(\bar{\xi}(z))\right)dz.$$

To generalize this result to the case $\mathbb{P}_x$ with $x<b$, similar to the proof of Theorem (3.1), we define a function $\varsigma(y):=\xi(y+x)-x$. Then

$$\bar{\varsigma}(y)=y+x-\xi(y+x)=\bar{\xi}(y+x).$$

Because X is spatially homogeneous, we have $(X,\bar{X},\tau_\xi)\big|_{\mathbb{P}_x}=(x+X,x+\bar{X},\tau_\varsigma)\big|_{\mathbb{P}}$. Hence,

$$\mathbb{E}_x\left[e^{-q\tau_\xi+pX(\tau_\xi)};\tau_\xi<\tau_b^+\right]=e^{px}\mathbb{E}\left[e^{-q\tau_\varsigma+pX(\tau_\varsigma)};\tau_\varsigma<\tau_{b-x}^+\right]$$
$$=\int_0^{b-x} e^{px-\int_0^z \frac{W_p^{(u)\prime}(\bar{\varsigma}(y))}{W_p^{(u)}(\bar{\varsigma}(y))}dy}\left(\frac{W_p^{(u)\prime}(\bar{\varsigma}(z))}{W_p^{(u)}(\bar{\varsigma}(z))}Z_p^{(u)}(\bar{\varsigma}(z))-uW_p^{(u)}(\bar{\varsigma}(z))\right)dz.$$
$$=\int_x^{b} e^{px-\int_x^z \frac{W_p^{(u)\prime}(\bar{\xi}(y))}{W_p^{(u)}(\bar{\xi}(y))}dy}\left(\frac{W_p^{(u)\prime}(\bar{\xi}(z))}{W_p^{(u)}(\bar{\xi}(z))}Z_p^{(u)}(\bar{\xi}(z))-uW_p^{(u)}(\bar{\xi}(z))\right)dz.$$

□

3.2 SNLTPs with draw-down times

In this section, we extend the results (2.23) in Theorem 2.12 and (2.25) in Theorem 2.13 to a more flexible and realistic situation where the ruin times is now replaced by the draw-down times. For simplicity, we only consider ξ as a linear draw-down function $\xi(t)=\xi t-d$ for some constants ξ and d satisfying $d>0$ and $(1-\xi)a+d>0$, where $a>0$ is an upper bound of the process. Thus, the draw-down time of SNLTPs has the form

$$\tau_{\xi,\gamma}=\tau_{d,\xi,\gamma}:=\inf\left\{t\geq 0:U_\gamma(t)+d<\xi\bar{U}_\gamma(t)\right\}. \tag{3.11}$$

The following theorem generalizes the exit problem for SNLTPs with draw-down times which can be found in Avram *et al.* (2017).

Theorem 3.3 (Theorem 1.1 in Avram *et al.* (2017))**.** *For any $q\geq 0,\gamma\in[0,1),\xi<1$ and $0<u\leq a$, we have*

$$\mathbb{E}_u\left[e^{-q\tau_{a,\gamma}^+};\tau_{a,\gamma}^+<\tau_{\xi,\gamma}\right]=\left(\frac{W^{(q)}((1-\xi)u+d)}{W^{(q)}((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}. \tag{3.12}$$

Remark 3.3. *By letting $\xi = d = 0$, we recover the solution (2.23) of the two-sided exit problem for SNLTPs in Albrecher et al. (2008). And the well-known result (2.13) of the two-sided exit problem for the general SNLPs can be recovered by letting $\gamma = \xi = d = 0$.*

Remark 3.4. *Similar to the power identities in Albrecher et al. (2007) and Albrecher et al. (2014), we have a power relation between the results with and without taxes. Furthermore, for the special case $\xi = 1$ we can obtain a generalization of Landriault et al. (2017). That is*

$$\mathbb{E}_u\big[e^{-q\tau_{a,\gamma}^+};\tau_{a,\gamma}^+ < \tau_{1,\gamma}\big] = \exp\bigg\{-\frac{(a-u)W'^{(q)}(d)}{(1-\gamma)W^{(q)}(d)}\bigg\}. \tag{3.13}$$

And by letting $d \to \infty$, we get

$$\mathbb{E}_u\big[e^{-q\tau_{a,\gamma}^+}\big] = \exp\bigg\{-\frac{(a-u)}{(1-\gamma)}\Phi(q)\bigg\}.$$

Remark 3.5. *For $\psi'(0+) > 0$, letting $q = 0$ and $a \to \infty$ we get*

$$\mathbb{P}_u[\tau_{\xi,\gamma} = \infty] = \Big(\psi'(0+)W((1-\xi)u + d)\Big)^{\frac{1}{(1-\xi)(1-\gamma)}}.$$

With the solution of two-sided exit problem for SNLTPs with draw-down times on hand, we are ready to derive the expression of its discounted taxed payments. Recall that the expected present value of all tax payments is defined as

$$v_{\xi,\gamma,\delta}(u) = v_{d,\xi,\gamma\delta}(u) := \gamma\mathbb{E}_u(T_{\xi,\gamma,\delta}) = \gamma\mathbb{E}_u\int_0^{\tau_{a,\gamma}^+\wedge\tau_{\xi,\gamma}} e^{-\delta t}d(\bar{X}(t) - u).$$

Theorem 3.4 (Theorem 1.2 in Avram *et al.* (2017))**.** *For any $\delta \geq 0, \gamma \in [0,1), \xi < 1$ and $0 < u < a \leq \infty$, we have*

$$\mathbb{E}_u\int_0^{\tau_{a,\gamma}^+\wedge\tau_{\xi,\gamma}} e^{-\delta t}d(\bar{X}(t) - u) = \frac{1}{1-\gamma}\int_u^a\left(\frac{W^{(\delta)}((1-\xi)u + d)}{W^{(\delta)}((1-\xi)s + d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}} ds, \tag{3.14}$$

and

$$\mathbb{E}_u\left[\int_0^{\tau_{a,\gamma}^+} e^{-\delta t}d(\bar{X}(t) - u);\tau_{a,\gamma}^+ < \tau_{\xi,\gamma}\right] =$$

$$\frac{1}{1-\gamma}\int_u^a\left(\frac{W^{(\delta)}((1-\xi)u + d)W((1-\xi)s + d)}{W^{(\delta)}((1-\xi)s + d)W((1-\xi)a + d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}} ds. \tag{3.15}$$

Remark 3.6. *To obtain the expression (2.25) of Theorem 3.2 in Albrecher et al. (2008), we let $a \to \infty$ and $d = \xi = 0$.*

Remark 3.7. *For the special case $\xi = 1$, one can follow the proof of Theorem 3.4 given below and the Remark 3.4 to get*

$$\begin{aligned}\mathbb{E}_u \int_0^{\tau_{a,\gamma}^+ \wedge \tau_{1,\gamma}} e^{-\delta t} d(\bar{X}(t) - u) &= \frac{1}{1-\gamma} \int_u^a \exp\left\{ - \frac{(s-u)W'^{(\delta)}(d)}{(1-\gamma)W^{(\delta)}(d)} \right\} ds \\ &= \frac{W^{(\delta)}(d)}{W'^{(\delta)}(d)} \left(1 - \exp\left\{ - \frac{(s-u)W'^{(\delta)}(d)}{(1-\gamma)W^{(\delta)}(d)} \right\} \right).\end{aligned}$$

And

$$\mathbb{E}_u \left[\int_0^{\tau_{a,\gamma}^+ \wedge \tau_{1,\gamma}} e^{-\delta t} d(\bar{X}(t) - u); \tau_{a,\gamma}^+ < \tau_{1,\gamma} \right] =$$

$$\frac{1}{1-\gamma} \int_u^a \exp\left\{ - \frac{(s-u)W'^{(\delta)}(d)}{(1-\gamma)W^{(\delta)}(d)} - \frac{(a-s)W'(d)}{(1-\gamma)W(d)} \right\} ds.$$

Remark 3.8. *In addition, when $\gamma = 1$, we also have*

$$\mathbb{E}_u \int_0^{\tau_{\xi,1}} e^{-\delta t} d(\bar{X}(t) - u) = \frac{W^{(\delta)}((1-\xi)u + d)}{W'^{(\delta)}((1-\xi)u + d)}.$$

Proof of Theorem 3.3.

We can mimic an argument from Chapter VII.2 of Bertoin (1996). We first assume that $X_t \to \infty$ as $t \to \infty$ and $q = 0$. Then $\bar{X} - X$ is a Markov process with local time $\bar{X} - X(0)$ at 0 by Theorem 2.9 above. Let ϵ be the excursion process of $\bar{X} - X$ away from 0, and h for the excursion height process. Using the excursion theory, we can rewrite the event $\{\tau_{a,\gamma}^+ < \tau_{\xi,\gamma}\}$ in terms of the excursion height process as

$$\left\{ h_s \leq (1-\xi)(u + (1-\gamma)s) + d, \quad \forall s \in \left[0, \frac{a-u}{1-\gamma}\right] \right\}.$$

This is due to the fact that if the process X starts the excursion at the level $u + s$ for $0 \leq s \leq \frac{a-u}{1-\gamma}$, which is the level $u + (1-\gamma)s$ for the taxed process U_γ, then to guarantee the time $\tau_{a,\gamma}^+$ occurring before the time $\tau_{\xi,\gamma}$, the excursion height h_s cannot exceed the sum

$$\{u + (1-\gamma)s\} + \{d - \xi(u + (1-\gamma)s)\} = (1-\xi)(u + (1-\gamma)s),$$

for $0 \leq s \leq \frac{a-u}{1-\gamma}$ (see figure 4 for the illustration). Thus, with the assumption X drifts to ∞ then by Theorem 2.10, h is a Poisson point process with characteristic measure ν such that

$$W((1-\xi)(u + (1-\gamma)s) + d) = e^{-\int_{(1-\xi)(u+(1-\gamma)s)+d}^{\infty} \nu(t,\infty)\mathrm{d}t},$$

which implies that

$$\nu((1-\xi)(u+(1-\gamma)s)+d,\infty)=\frac{W'((1-\xi)(u+(1-\gamma)s)+d)}{W((1-\xi)(u+(1-\gamma)s)+d)}.$$

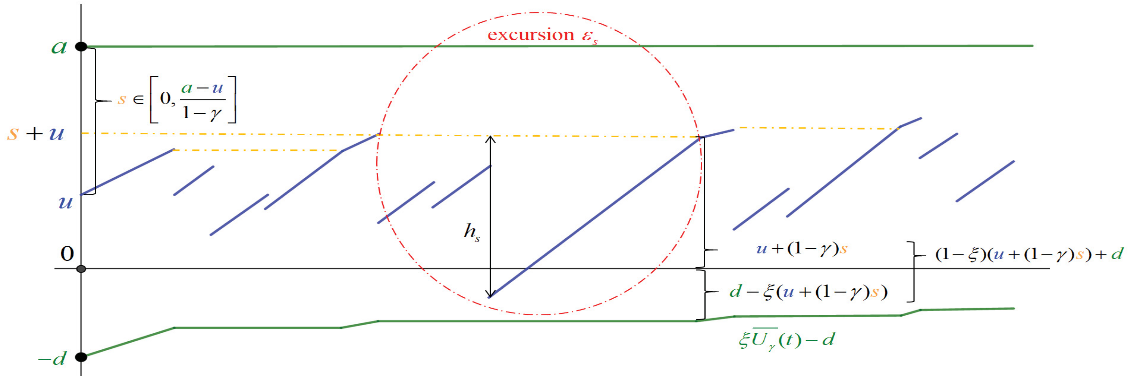

Figure 4: Assume one of simple paths of the tax SNLP $U(t)$ is given in the figure. At the beginning of an excursion at level $u+s$ for process $X(t)$ where $0 \le s \le \frac{a-u}{1-\gamma}$ (this corresponds to the level $u+(1-\gamma)s$ for the taxed process $U(t)$), the process $U(t)$ starts at u within the two barriers: the fixed level a above and the varying level $\xi\bar{U}_\gamma - d$ below. The event $\{\tau^+_{a,\gamma} < \tau_{\xi,\gamma}\}$ is equivalent to the event $\{0 \le h_s \le (1-\xi)(u+(1-\gamma)s)+d\}$.

Now, let N be the Poisson random variable counting the number of Poisson points (s, h_s)

$$\#\left\{(s,h_s)\in\mathbb{R}^2 \,\middle|\, 0\le s\le \frac{a-u}{1-\gamma},\quad h_s \ge (1-\xi)(u+(1-\gamma)s)+d\right\}.$$

Then its parameter is of the form

$$\int_0^{\frac{a-u}{1-\gamma}} \nu((1-\xi)(u+(1-\gamma)t)+d,\infty)dt=\int_0^{\frac{a-u}{1-\gamma}}\frac{W'((1-\xi)(u+(1-\gamma)t)+d)}{W((1-\xi)(u+(1-\gamma)t)+d)}\mathrm{d}t.$$

Therefore, for $0<u<a$,

$$\begin{aligned}
\mathbb{P}_u\{\tau^+_{a,\gamma}<\tau_{\xi,\gamma}\} &= \mathbb{P}\left\{h_s \le (1-\xi)(u+(1-\gamma)s)+d,\quad \forall s\in\left[0,\frac{a-u}{1-\gamma}\right]\right\}\\
&= \mathbb{P}\{N=0\}\\
&= \exp\left\{-\int_0^{\frac{a-u}{1-\gamma}}\frac{W'((1-\xi)(u+(1-\gamma)t)+d)}{W((1-\xi)(u+(1-\gamma)t)+d)}\mathrm{d}t\right\}\\
&= \exp\left\{-\frac{1}{(1-\xi)(1-\gamma)}\int_{(1-\xi)u+d}^{(1-\xi)a+d}\frac{W'(t)}{W(t)}\mathrm{d}t\right\}\\
&= \exp\left\{-\frac{1}{(1-\xi)(1-\gamma)}\int_{(1-\xi)u+d}^{(1-\xi)a+d}\frac{d}{dt}\ln(W(t))\mathrm{d}t\right\}\\
&= \left(\frac{W((1-\xi)u+d)}{W((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}.
\end{aligned}$$

In general, for $q>0$ we consider a new measure $\mathbb{P}^{\Phi(q)}$ with Radon-Nikodým derivative martingale $e^{\Phi(q)(X_t-X_0)-qt}$. X drifts to ∞ under $\mathbb{P}^{\Phi(q)}$ and $W_{\Phi(q)}(u)=e^{-\Phi(q)u}W^{(q)}(u)$. Also, we observe that if process U_t reaches a new maximum, then X_t and $\bar{X}_t$ also reach the same new maximum at the same time. That is, at the time $\tau^+_{a,\gamma}$, $X(\tau^+_{a,\gamma})=\bar{X}(\tau^+_{a,\gamma})$ for $\tau^+_{a,\gamma}<\infty$. Then

$$a=U_\gamma(\tau^+_{a,\gamma})=X(\tau^+_{a,\gamma})-\gamma\big(\bar{X}(\tau^+_{a,\gamma})-u\big),$$

which implies that

$$X(\tau^+_{a,\gamma})\mathbf{1}_{(\tau^+_{a,\gamma}<\infty)}=\frac{a-\gamma u}{1-\gamma}\mathbf{1}_{(\tau^+_{a,\gamma}<\infty)}.$$

Therefore, using the result above for the process X under the new measure we obtain

$$\begin{aligned}\mathbb{P}^{\Phi(q)}_u\{\tau^+_{a,\gamma}<\tau_{\xi,\gamma}\}&=\left(\frac{W_{\Phi(q)}((1-\xi)u+d)}{W_{\Phi(q)}((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}\\&=e^{\Phi(q)(a-u)}\left(\frac{W^{(q)}((1-\xi)u+d)}{W^{(q)}((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}.\end{aligned}$$

Moreover, using the change of measure, we can rewrite $\mathbb{P}^{\Phi(q)}_u\{\tau^+_{a,\gamma}<\tau_{\xi,\gamma}\}$ as

$$\begin{aligned}\mathbb{P}^{\Phi(q)}_u\{\tau^+_{a,\gamma}<\tau_{\xi,\gamma}\}&=\mathbb{E}_u[e^{\Phi(q)(X(\tau^+_{a,\gamma})-u)-q\tau^+_{a,\gamma}};\tau^+_{a,\gamma}<\tau_{\xi,\gamma}]\\&=e^{\Phi(q)(a-u)}\mathbb{E}_u[e^{-q\tau^+_{a,\gamma}};\tau^+_{a,\gamma}<\tau_{\xi,\gamma}].\end{aligned}$$

Hence

$$\mathbb{E}_u(e^{-q\tau^+_{a,\gamma}};\tau^+_{a,\gamma}<\tau_{\xi,\gamma})=\left(\frac{W^{(q)}((1-\xi)u+d)}{W^{(q)}((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}.$$

The result follows for the case $q=0$ by letting $q\to 0^+$. □

Proof of Theorem 3.4.

The idea in this proof is to divide the interval $[u,a]$ into sub-intervals of length $h>0$, then we condition on each iteration the arrival time of $\tau^+_{x+h,\gamma}$ and $\tau_{\xi,\gamma}$. The detail is as following. Assume $U_\gamma(0)=x>0$ and for $h>0$, we have

$$\begin{aligned}x+h&=U_\gamma(\tau^+_{x+h,\gamma})\\&=X(\tau^+_{x+h,\gamma})-\gamma(\bar{X}(\tau^+_{x+h,\gamma})-x)\\&=\bar{X}(\tau^+_{x+h,\gamma})-\gamma(\bar{X}(\tau^+_{x+h,\gamma})-x)\\&=(1-\gamma)\bar{X}(\tau^+_{x+h,\gamma})+\gamma x,\end{aligned}$$

which gives

$$\bar{X}(\tau^+_{x+h,\gamma})-x=\frac{h}{1-\gamma}.$$

By conditioning on the arrival time between $\tau^+_{x+h,\gamma}$ and $\tau_{\xi,\gamma}$ we get

$$\mathbb{E}_x[T_{\xi,\gamma,\delta}] = \mathbb{E}_x[T_{\xi,\gamma,\delta}; \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}] + \mathbb{E}_x[T_{\xi,\gamma,\delta}; \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}]. \tag{3.16}$$

For the first expectation in (3.16), we can bound it from above by

$$\begin{aligned}
& \mathbb{E}_x[T_{\xi,\gamma,\delta}; \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}] \\
= \ & \mathbb{E}_x\left[\int_0^{\tau^+_{x+h,\gamma}\wedge\tau_{\xi,\gamma}} e^{-\delta t} d(\bar{X}(t)-x);\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
\leq \ & \mathbb{E}_x\left[\int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t} d(\bar{X}(t)-x);\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
= \ & \mathbb{E}_x\left[e^{-\delta t}(\bar{X}(t)-x)\Big|_0^{\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
& + \mathbb{E}_x\left[\delta\int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t}(\bar{X}(t)-x)dt;\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
\leq \ & \mathbb{E}_x\left[e^{-\delta\tau^+_{x+h,\gamma}}\frac{h}{1-\gamma};\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
& + \mathbb{E}_x\left[\frac{\delta h}{1-\gamma}\int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t}dt;\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
= \ & \frac{h}{1-\gamma}\mathbb{E}_x\left[e^{-\delta\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
& + \frac{h}{1-\gamma}\mathbb{E}_x\left[\left(1-e^{-\delta\tau^+_{x+h,\gamma}}\right);\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
= \ & \frac{h}{1-\gamma}\mathbb{E}_x\left[e^{-\delta\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
& + \frac{h}{1-\gamma}\mathbb{P}_x\left[\tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] - \frac{h}{1-\gamma}\mathbb{E}_x\left[e^{-\delta\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}\right] \\
= \ & \frac{h}{1-\gamma}\left(1-\left(\frac{W((1-\xi)x+d)}{W((1-\xi)(x+h)+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}\right) \\
= \ & \frac{h}{1-\gamma} - \frac{h}{1-\gamma}\left(\frac{W((1-\xi)x+d)}{W((1-\xi)(x+h)+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}} \\
= \ & \frac{h}{1-\gamma} + o(h),
\end{aligned}$$

where the second inequality follows by the fact that the integrand is positive and the condition $\tau^+_{x+h,\gamma} > \tau_{\xi,\gamma}$. In the third equality we apply the integral by parts. In the fourth inequality, it follows by the maximal value of the integrand. And in the last equality we apply Theorem

3.3. For the second expectation in (3.16), we can again bound it from above by

$$
\begin{aligned}
& \mathbb{E}_x[T_{\xi,\gamma,\delta}; \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}] \\
= \ & \mathbb{E}_x\left[\int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t} d(\bar{X}(t) - x);\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
= \ & \mathbb{E}_x\left[e^{-\delta t}(\bar{X}(t) - x)\Big|_0^{\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
& + \mathbb{E}_x\left[\delta \int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t}(\bar{X}(t) - x)dt;\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
\leq \ & \frac{h}{1-\gamma}\mathbb{E}_x\left[e^{-\delta\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
& + \mathbb{E}_x\left[\frac{\delta h}{1-\gamma}\int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t}dt;\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
= \ & \frac{h}{1-\gamma}\mathbb{E}_x\left[e^{-\delta\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
& + \frac{h}{1-\gamma}\mathbb{E}_x\left[1 - e^{-\delta\tau^+_{x+h,\gamma}};\ \tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
= \ & \frac{h}{1-\gamma}\mathbb{P}_x\left[\tau^+_{x+h,\gamma} < \tau_{\xi,\gamma}\right] \\
= \ & \frac{h}{1-\gamma}\left(\frac{W((1-\xi)x + d)}{W((1-\xi)(x+h) + d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}} \\
= \ & o(h).
\end{aligned}
$$

So, it follows that

$$
\begin{aligned}
& \mathbb{E}_u\left[\int_{\tau^+_{x,\gamma}\wedge\tau_{\xi,\gamma}}^{\tau^+_{x+h,\gamma}\wedge\tau_{\xi,\gamma}} e^{-\delta t} d(\bar{X}_t - u)\right] \\
= \ & \mathbb{E}_u\left[e^{-\delta\tau^+_{x,\gamma}};\ \tau^+_{x,\gamma} < \tau_{\xi,\gamma}\right]\mathbb{E}_x\left[\int_0^{\tau^+_{x+h,\gamma}\wedge\tau\xi,\gamma} e^{-\delta t} d(\bar{X}_t - u)\right] \\
= \ & \left(\frac{W^{(\delta)}((1-\xi)u + d)}{W^{(\delta)}((1-\xi)x + d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}} \mathbb{E}_x[T_{\xi,\gamma,\delta}] \\
= \ & \left(\frac{W^{(\delta)}((1-\xi)u + d)}{W^{(\delta)}((1-\xi)x + d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}} \left(\frac{h}{1-\gamma} + o(h)\right).
\end{aligned}
$$

Therefore, by Riemann sum we obtain

$$\begin{aligned}
\mathbb{E}_u[T_{\xi,\gamma,\delta}] &= \lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}_u\Big[\int_{\tau^+_{u+(i-1)(a-u)/n,\gamma}\wedge\tau_{\xi,\gamma}}^{\tau^+_{u+i(a-u)/n,\gamma}\wedge\tau_{\xi,\gamma}} e^{-\delta t}d(\bar{X}_t-u)\Big]\\
&= \lim_{n\to\infty}\sum_{i=1}^{n}\left(\frac{W^{(\delta)}((1-\xi)u+d)}{W^{(\delta)}((1-\xi)(u+(i-1)(a-u)/n)+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}\left(\frac{a-u}{(1-\gamma)n}+o(\frac{1}{n})\right)\\
&= \frac{1}{1-\gamma}\int_u^a\left(\frac{W^{(\delta)}((1-\xi)u+d)}{W^{(\delta)}((1-\xi)s+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}ds.
\end{aligned}$$

Similarly, for the second part of the theorem, we have

$$\begin{aligned}
&\mathbb{E}_u\Big[\int_{\tau^+_{x,\gamma}}^{\tau^+_{x+h,\gamma}} e^{-\delta t}d(\bar{X}(t)-u);\ \tau^+_{a,\gamma}<\tau_{\xi,\gamma}\Big]\\
=\ &\mathbb{E}_u\Big[e^{-\delta\tau^+_{x,\gamma}};\tau^+_{x,\gamma}<\tau_{\xi,\gamma}\Big]\mathbb{E}_x\Big[\int_0^{\tau^+_{x+h,\gamma}} e^{-\delta t}d(\bar{X}(t)-x);\ \tau^+_{x+h,\gamma}<\tau_{\xi,\gamma}\Big]\mathbb{P}_{x+h}\{\tau^+_{a,\gamma}<\tau_{\xi,\gamma}\}\\
=\ &\left(\frac{W^{(\delta)}((1-\xi)u+d)}{W^{(\delta)}((1-\xi)x+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}\left(\frac{h}{1-\gamma}+o(h)\right)\left(\frac{W((1-\xi)(x+h)+d)}{W((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_u\Big[\int_0^{\tau^+_{a,\gamma}} e^{-\delta t}d(\bar{X}(t)-u);\ \tau^+_{a,\gamma}<\tau_{\xi,\gamma}\Big]\\
=\ &\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}_u\Big[\int_{\tau^+_{u+(i-1)(a-u)/n,\gamma}}^{\tau^+_{u+i(a-u)/n,\gamma}} e^{-\delta t}d(\bar{X}_t-u);\ \tau^+_{a,\gamma}<\tau_{\xi,\gamma}\Big]\\
=\ &\frac{1}{1-\gamma}\int_u^a\left(\frac{W^{(\delta)}((1-\xi)u+d)W((1-\xi)s+d)}{W^{(\delta)}((1-\xi)s+d)W((1-\xi)a+d)}\right)^{\frac{1}{(1-\xi)(1-\gamma)}}ds.
\end{aligned}$$

□

3.3 Occupation times of general SNLPs in a draw-down interval.

In this sub-section, we will study the Laplace transform of the occupation time, which is the total amount of time a SNLP X spends in an interval killed on exiting a pre-determined interval. However, instead of considering the classical occupation times on a fixed interval, we allow the interval varying over time. Particularly, we want to understand the law of occupation time in a draw-down interval, killed on either exiting the upper barrier b or another lower draw-down level (see Figure 5 for an illustration).

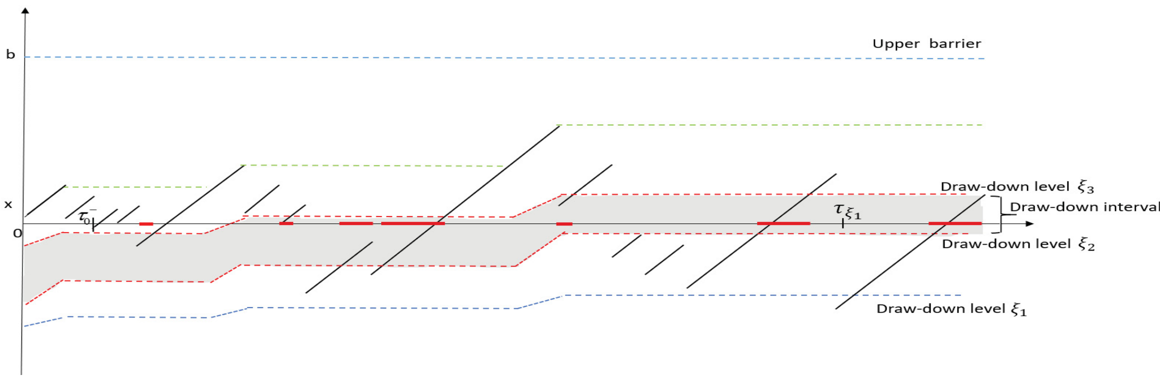

Figure 5: The occupation time of the SNLP X in the draw-down interval (shaded area): $(\xi_2\bar{X}(t), \xi_3\bar{X}(t))$. The sum of all red intervals represents the occupation time without killing.

That is, suppose there are three ordered drawdown functions $\xi_1(\cdot), \xi_2(\cdot)$ and $\xi_3(\cdot)$ such that $\xi_1(x) \le \xi_2(x) \le \xi_3(x) \le x$ for all $x \le b$, and also $\bar{\xi}_i(x) := x - \xi_i(x) > 0$. We are interested in the joint Laplace transform of the forms

$$\mathbb{E}_x\left[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_b^+ < \tau_{\xi_1}\right], \tag{3.17}$$

and

$$\mathbb{E}_x\left[e^{-p\tau_{\xi_1} - q\int_0^{\tau_{\xi_1}} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_{\xi_1} < \tau_b^+\right]. \tag{3.18}$$

Observe that with the condition $\xi_i(x) < x$ for all x and $i = 1, 2, 3$, the occupation time of the process X under a drawdown level $\xi_i(\bar{X}_t)$ only happens during excursions at certain time. Particularly, for each excursion ϵ_r with lifetime $\zeta(\epsilon_r)$ where $r = \bar{X}_t$, the indicator for the occupation time under the drawdown level $\xi_i(r)$ is $\mathbf{1}_{\{\epsilon_r(t) > \bar{\xi}_i(r)\}}$ for $0 \le t \le \zeta(\epsilon_r)$. Therefore, under $\mathbb{P}$, we can use excursion process to express the occupation time in (3.17) as

$$\int_0^{\tau_b^+} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt = \sum_{r\in[0,b]} \int_0^{\zeta(\epsilon_r)} \mathbf{1}_{\{\bar{\xi}_3(r) < \epsilon_r(t) < \bar{\xi}_2(r)\}}dt. \tag{3.19}$$

On the other hand, to express the occupation time in (3.18), we recall the first passage time of a genetic excursion process (introduced in Section (3.1)).

$$\rho_c^+ = \rho_c^+(\varepsilon) := \inf\{s \in (0, \zeta) : \varepsilon(s) > c\}.$$

Then, we can rewrite τ_{ξ_1} as

$$\tau_{\xi_1} = L^{-1}_{r-} + \rho^{+}_{\bar{\xi}_1(r)}(\epsilon_r),$$

where $\bar{X}(\tau_\xi) = L(\tau_\xi) = r$. As a result, the occupation time in (3.18) can be expressed as

$$\int_0^{\tau_{\xi_1}} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt = \sum_{s\in[0,r)} \int_0^{\zeta(\epsilon_s)} \mathbf{1}_{\{\bar{\xi}_3(s)<\epsilon_s(t)<\bar{\xi}_2(s)\}}dt + \int_0^{\rho^{+}_{\bar{\xi}_1(r)}(\epsilon_r)} \mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt. \tag{3.20}$$

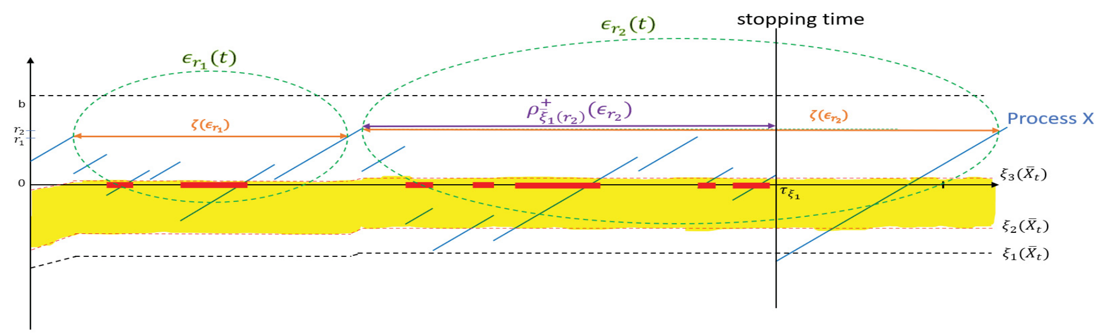

Figure 6: The occupation time of the SNLP X is decomposed into excursion components.

Before stating our main results on the occupation times, we define two new scale functions that will help putting our expressions in a compact and nested form

$$\mathbb{W}^{(p,q)}_c(x,a) \;:=\; \mathcal{W}^{(p,q)}_c(x) - q\int_0^a \mathcal{W}^{(p,q)}_c(x-z)W^{(p)}(z)dz, \tag{3.21}$$

and

$$\mathbb{Z}^{(p,q)}_c(x,a) \;:=\; \mathcal{Z}^{(p,q)}_c(x) - q\int_0^a \mathcal{Z}^{(p,q)}_c(x-z)W^{(p)}(z)dz, \tag{3.22}$$

where

$$\begin{aligned}\mathcal{W}^{(p,q)}_a(x) &:= W^{(p+q)}(x) - q\int_0^a W^{(p+q)}(x-y)W^{(p)}(y)dy\\ &= W^{(p)}(x) + q\int_a^x W^{(p+q)}(x-y)W^{(p)}(y)dy,\end{aligned}$$

and

$$\begin{aligned}\mathcal{Z}^{(p,q)}_a(x) &:= Z^{(p+q)}(x) - q\int_0^a W^{(p+q)}(x-y)Z^{(p)}(y)dy\\ &= Z^{(p)}(x) + q\int_a^x W^{(p+q)}(x-y)Z^{(p)}(y)dy\end{aligned}$$

are given in Section 2.3. Now we are ready for the following two theorems.

Theorem 3.5. *For all $x \le b$ and $p, q \ge 0$, we have*

$$\mathbb{E}_x\left[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_b^+ < \tau_{\xi_1}\right] = \exp\left\{ -\int_x^b \frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))} dy\right\}. \tag{3.23}$$

Theorem 3.6. *For all $x \le b$ and $p, q \ge 0$, we have*

$$\begin{aligned}\mathbb{E}_x\left[e^{-p\tau_{\xi_1} - q\int_0^{\tau_{\xi_1}} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_{\xi_1} < \tau_b^+\right] &= \int_x^b \exp\left\{ -\int_x^z \frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))} dy\right\} \\ \times \left(\frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z))}{\mathbb{W}^{(p,q)}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z))}\mathbb{Z}^{(p,q)}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z)) - \mathbb{Z}^{(p,q)\prime}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z))\right)dz.\end{aligned} \tag{3.24}$$

Remark 3.9. *By letting $\xi_1 \equiv 0 \le \xi_2 \equiv c \le \xi_3 \equiv a \le b$, we can recover the results in Loeffen et al. (2014). We have*

$$\begin{aligned}\mathbb{E}_x\left[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(c,a)}(X_t)dt}, \tau_b^+ < \tau_0^-\right] &= \exp\left\{ -\int_x^b \frac{\mathbb{W}_c^{(p,q)\prime}(y, y-a)}{\mathbb{W}_c^{(p,q)}(y, y-a)} dy\right\} \\ &= \exp\left\{ -\int_x^b \frac{d}{dy} \ln\left\{\mathbb{W}_c^{(p,q)}(y, y-a)\right\} dy\right\} \\ &= \frac{\mathbb{W}_c^{(p,q)}(x, x-a)}{\mathbb{W}_c^{(p,q)}(b, b-a)},\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}_x\left[e^{-p\tau_0^- - q\int_0^{\tau_0^-} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_0^- < \tau_b^+\right] \\ &= \int_x^b \exp\left\{ -\int_x^z \frac{\mathbb{W}_c^{(p,q)\prime}(y, y-a)}{\mathbb{W}_c^{(p,q)}(y, y-a)} dy\right\}\left(\frac{\mathbb{W}_c^{(p,q)\prime}(z, z-a)}{\mathbb{W}_c^{(p,q)}(z, z-a)}\mathbb{Z}_c^{(p,q)}(z, z-a) - \mathbb{Z}_c^{(p,q)\prime}(z, z-a)\right)dz \\ &= \mathbb{W}_c^{(p,q)}(x, x-a)\int_x^b \left(\frac{\mathbb{W}_c^{(p,q)\prime}(z, z-a)\mathbb{Z}_c^{(p,q)}(z, z-a) - \mathbb{Z}_c^{(p,q)\prime}(z, z-a)\mathbb{W}_c^{(p,q)}(z, z-a)}{\left(\mathbb{W}_c^{(p,q)}(z, z-a)\right)^2}\right)dz \\ &= \mathbb{W}_c^{(p,q)}(x, x-a)\int_x^b \frac{d}{dz}\left(-\frac{\mathbb{Z}_c^{(p,q)}(z, z-a)}{\mathbb{W}_c^{(p,q)}(z, z-a)}\right)dz \\ &= \mathbb{Z}_c^{(p,q)}(x, x-a) - \frac{\mathbb{W}_c^{(p,q)}(x, x-a)}{\mathbb{W}_c^{(p,q)}(b, b-a)}\mathbb{Z}_c^{(p,q)}(b, b-a).\end{aligned}$$

Remark 3.10. *By letting $q = 0$ in Theorem 3.5 and Theorem 3.6 , we recover the solutions of two-sided exit problems of Theorem 3.1 and 3.2. Particularly,*

$$\begin{aligned}\mathbb{E}_x\left[e^{-p\tau_b^+}, \tau_b^+ < \tau_{\xi_1}\right] &= \exp\left\{-\int_x^b \frac{\mathbb{W}^{(p,0)\prime}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))}{\mathbb{W}^{(p,0)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))} dy\right\} \\ &= \exp\left\{-\int_x^b \frac{\mathcal{W}^{(p,0)\prime}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y))}{\mathcal{W}^{(p,0)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y))} dy\right\} \\ &= \exp\left\{-\int_x^b \frac{W^{(p)\prime}(\bar{\xi}_1(y))}{W^{(p)}(\bar{\xi}_1(y))} dy\right\},\end{aligned}$$

where the second and third equation are followed by the definition of $\mathbb{W}^{(p.q)}_c(\cdot,\cdot)$ and $W^{(p,q)}_a(\cdot)$ respectively. Moreover,

$$\begin{aligned}&\mathbb{E}_x\left[e^{-p\tau_{\xi_1}}, \tau_{\xi_1} < \tau_b^+\right] \\ = &\int_x^b \exp\left\{-\int_x^z \frac{\mathbb{W}^{(p,0)\prime}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))}{\mathbb{W}^{(p,0)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))} dy\right\} \\ &\times\left(\frac{\mathbb{W}^{(p,0)\prime}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z))}{\mathbb{W}^{(p,0)}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z))}\mathbb{Z}^{(p,0)}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z)) - \mathbb{Z}^{(p,0)\prime}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z), \bar{\xi}_3(z))\right) dz \\ = &\int_x^b \exp\left\{-\int_x^z \frac{\mathcal{W}^{(p,0)\prime}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))}{\mathcal{W}^{(p,0)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y), \bar{\xi}_3(y))} dy\right\} \\ &\times\left(\frac{\mathcal{W}^{(p,0)\prime}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z))}{\mathcal{W}^{(p,0)}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z))}\mathcal{Z}^{(p,0)}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z)) - \mathcal{Z}^{(p,0)\prime}_{\xi_2(z)-\xi_1(z)}(\bar{\xi}_1(z))\right) dz \\ = &\int_x^b \exp\left\{-\int_x^z \frac{W^{(p)\prime}(\bar{\xi}_1(y))}{W^{(p)}(\bar{\xi}_1(y))} dy\right\}\left(\frac{W^{(p)\prime}(\bar{\xi}_1(z))}{W^{(p)}(\bar{\xi}_1(z))} Z^{(p)}(\bar{\xi}_1(z)) - pW^{(p)}(\bar{\xi}_1(z))\right) dz.\end{aligned}$$

Proof of Theorem 3.5.

We first prove it under $\mathbb{P}$. As mentioned at the beginning of Section 3.1 that under $\mathbb{P}$, the running maximum process $\bar{X}$ is the local time of the reflected process Y at 0. Then its inverse $L_s^{-1} = \tau_s^+$ is a subordinator with Laplace exponent Φ, and

$$\nu = \lim_{s\to\infty} \frac{\Phi(s)}{s} = \lim_{s\to\infty} \frac{s}{\psi(s)} = W(0).$$

Also, the first passage time can be expressed as

$$\tau_b^+ = L_b^{-1} = \nu b + \sum_{r\in[0,b]} \zeta(\epsilon_r),$$

where $\zeta = \zeta(\varepsilon)$ is the lifetime of a generic excursion process. By the expression (3.19) and with the help of exponential compensation formula, we have

$$\begin{aligned}
&\mathbb{E}\Big[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_b^+ < \tau_{\xi_1}\Big] \\
&= \mathbb{E}\Big[\exp\Big\{ -p\big(\nu b + \sum_{r\in[0,b]} \zeta(\epsilon_r)\big) - q\sum_{r\in[0,b]}\int_0^{\zeta(\epsilon_r)} \mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt\Big\} \prod_{r\in[0,b]} \mathbf{1}_{\{\bar{\epsilon}_r\le\bar{\xi}_1(r)\}}\Big] \\
&= e^{-p\nu b}\mathbb{E}\Big[\exp\Big\{ -\sum_{r\in[0,b]}\Big(p\zeta(\epsilon_r) + q\int_0^{\zeta(\epsilon_r)} \mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt\Big)\Big\} \prod_{r\in[0,b]} \mathbf{1}_{\{\bar{\epsilon}_r\le\bar{\xi}_1(r)\}}\Big] \\
&= e^{-p\nu b}\mathbb{E}\Big[\exp\Big\{ -\sum_{r\in[0,b]}\Big(p\zeta(\epsilon_r) + q\int_0^{\zeta(\epsilon_r)} \mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt + \infty\cdot\mathbf{1}_{\{\bar{\epsilon}_r>\bar{\xi}_1(r)\}}\Big)\Big\}\Big] \\
&= \exp\Big\{ -\int_0^b \Big(p\nu + \int_{\mathcal{E}}\Big(1-\exp\Big\{-p\zeta - q\int_0^{\zeta}\mathbf{1}_{\{\bar{\xi}_3(r)<\varepsilon(t)<\bar{\xi}_2(r)\}}dt\Big\}\mathbf{1}_{\{\bar{\varepsilon}\le\bar{\xi}_1(r)\}}\Big)n(d\varepsilon)\Big)dr\Big\}.
\end{aligned}$$

Consider the case where $\xi_1 \equiv d \le \xi_2 \equiv c \le \xi_3 \equiv a \le 0 \le b$. Recall that the results in Loeffen *et al.* (2014) can be expressed in a nice form using $\mathbb{W}_c^{(p,q)}(x,a)$ and $\mathbb{Z}_c^{(p,q)}(x,a)$ as

$$\mathbb{E}_x\Big[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_b^+ < \tau_0^-\Big] = \frac{\mathbb{W}_c^{(p,q)}(x, x-a)}{\mathbb{W}_c^{(p,q)}(b, b-a)},$$

which means

$$\mathbb{E}\Big[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_b^+ < \tau_d^-\Big] = \frac{\mathbb{W}_{c-d}^{(p,q)}(-d, -a)}{\mathbb{W}_{c-d}^{(p,q)}(b-d, b-a)}.$$

Therefore, putting everything together we get

$$\begin{aligned}
&\frac{\mathbb{W}_{c-d}^{(p,q)}(-d, -a)}{\mathbb{W}_{c-d}^{(p,q)}(b-d, b-a)} = \mathbb{E}\Big[e^{-p\tau_b^+ - q\int_0^{\tau_b^+} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_b^+ < \tau_d^-\Big] \\
&= \exp\Big\{ -\int_0^b \Big(p\nu + \int_{\mathcal{E}}\Big(1-\exp\Big\{-p\zeta - q\int_0^{\zeta}\mathbf{1}_{\{r-a<\varepsilon(t)<r-c\}}dt\Big\}\mathbf{1}_{\{\bar{\varepsilon}\le r-d\}}\Big)n(d\varepsilon)\Big)dr\Big\}.
\end{aligned}$$

Taking log on both sides of the above equation we get

$$\ln\left[\frac{\mathbb{W}^{(p,q)}_{c-d}(b-d,b-a)}{\mathbb{W}^{(p,q)}_{c-d}(-d,-a)}\right]$$
$$=\int_0^b\left(p\nu+\int_{\mathcal{E}}\left(1-\exp\left\{-p\zeta-q\int_0^\zeta\mathbf{1}_{\{r-a<\varepsilon(t)<r-c\}}dt\right\}\mathbf{1}_{\{\bar{\varepsilon}\le r-d\}}\right)n(d\varepsilon)\right)dr.$$

Differentiating on both side of the above equation with respect to b we obtain

$$p\nu+\int_{\mathcal{E}}\left(1-\exp\left\{-p\zeta-q\int_0^\zeta\mathbf{1}_{\{b-a<\varepsilon(t)<b-c\}}dt\right\}\mathbf{1}_{\{\bar{\varepsilon}\le b-d\}}\right)n(d\varepsilon)=\frac{\mathbb{W}'^{(p,q)}_{c-d}(b-d,b-a)}{\mathbb{W}^{(p,q)}_{c-d}(b-d,b-a)}.$$

Or in general, for $y\ge 0$

$$p\nu+\int_{\mathcal{E}}\left(1-\exp\left\{-p\zeta-q\int_0^\zeta\mathbf{1}_{\{\bar{\xi}_3(y)<\varepsilon(t)<\bar{\xi}_2(y)\}}dt\right\}\mathbf{1}_{\{\bar{\varepsilon}\le\bar{\xi}_1(y)\}}\right)n(d\varepsilon)=\frac{\mathbb{W}'^{(p,q)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y),\bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y),\bar{\xi}_3(y))}.$$

Hence, we obtain our desired result of Theorem 3.5 under $\mathbb{P}$.

For the general case under $\mathbb{P}_x$ where $x<b$, we consider functions $\varsigma_i(y):=\xi_i(y+x)-x$, $i=1,2,3$. Then $\bar{\varsigma}_i(y)=y+x-\xi_i(y+x)=\bar{\xi}_i(y+x)$. Because X is spatially homogeneous, by a shifting argument, we have

$$\mathbb{E}_x\left[e^{-p\tau_b^+-q\int_0^{\tau_b^+}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt},\tau_b^+<\tau_{\xi_1}\right]$$
$$=\mathbb{E}\left[e^{-p\tau_{b-x}^+-q\int_0^{\tau_{b-x}^+}\mathbf{1}_{(\varsigma_2(\bar{X}_s),\varsigma_3(\bar{X}_s))}(X_s)ds},\tau_{b-x}^+<\tau_{\varsigma_1}\right]$$
$$=\exp\left\{-\int_0^{b-x}\frac{\mathbb{W}'^{(p,q)}_{\varsigma_2(y)-\varsigma_1(y)}(\bar{\varsigma}_1(y),\bar{\varsigma}_3(y))}{\mathbb{W}^{(p,q)}_{\varsigma_2(y)-\varsigma_1(y)}(\bar{\varsigma}_1(y),\bar{\varsigma}_3(y))}dy\right\}$$
$$=\exp\left\{-\int_x^{b}\frac{\mathbb{W}'^{(p,q)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y),\bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\xi_2(y)-\xi_1(y)}(\bar{\xi}_1(y),\bar{\xi}_3(y))}dy\right\}.$$

□

Proof of Theorem 3.6.

Recall that the draw-down time τ_ξ can be expressed as

$$\tau_\xi=L^{-1}_{r-}+\rho^+_{\bar{\xi}(r)}(\epsilon_r),$$

where $\bar{X}(\tau_\xi)=L(\tau_\xi)=r$, and $\rho^+_c=\rho^+_c(\varepsilon):=\inf\{s\in(0,\zeta):\varepsilon(s)>c\}$ is the first passage time of a genetic excursion process. Also, the occupation time has the form

$$\int_0^{\tau_{\xi_1}} \mathbf{1}_{(\xi_2(\bar{X}_s),\xi_3(\bar{X}_s))}(X_s)ds = \sum_{s\in[0,r)} \int_0^{\zeta(\epsilon_s)} \mathbf{1}_{\{\bar{\xi}_3(s)<\epsilon_s(t)<\bar{\xi}_2(s)\}}dt + \int_0^{\rho^+_{\bar{\xi}_1(r)}(\epsilon_r)} \mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt.$$

Using the same approach as in the proof of Theorem 3.5, under $\mathbb{P}$ we have

$$\begin{aligned}
&\mathbb{E}\Big[e^{-p\tau_{\xi_1}-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_{\xi_1}<\tau_b^+\Big]\\
&=\mathbb{E}\Big[\sum_{r\in[0,b]}\Big(\exp\Big\{-p(L^{-1}_{r-}+\rho^+_{\bar{\xi}_1(r)}(\epsilon_r))-q\sum_{s\in[0,r)}\int_0^{\zeta(\epsilon_s)}\mathbf{1}_{\{\bar{\xi}_3(s)<\epsilon_s(t)<\bar{\xi}_2(s)\}}dt\\
&\qquad -q\int_0^{\rho^+_{\bar{\xi}_1(r)}(\epsilon_r)}\mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt\Big\}\prod_{s<r}\mathbf{1}_{\{\bar{\epsilon}_s\le\bar{\xi}_1(s)\}}\mathbf{1}_{\{\bar{\epsilon}_r>\bar{\xi}_1(r)\}}\Big)\Big]\\
&=\mathbb{E}\Big[\sum_{r\in[0,b]}\Big(\Big(\exp\Big\{-pL^{-1}_{r-}-q\sum_{s\in[0,r)}\int_0^{\zeta(\epsilon_s)}\mathbf{1}_{\{\bar{\xi}_3(s)<\epsilon_s(t)<\bar{\xi}_2(s)\}}dt\Big\}\prod_{s<r}\mathbf{1}_{\{\bar{\epsilon}_s\le\bar{\xi}_1(s)\}}\Big)\\
&\times\Big(\exp\Big\{-p\rho^+_{\bar{\xi}_1(r)}(\epsilon_r)-q\int_0^{\rho^+_{\bar{\xi}_1(r)}(\epsilon_r)}\mathbf{1}_{\{\bar{\xi}_3(r)<\epsilon_r(t)<\bar{\xi}_2(r)\}}dt\Big\}\mathbf{1}_{\{\bar{\epsilon}_r>\bar{\xi}_1(r)\}}\Big)\Big)\Big]\\
&=\mathbb{E}\Big[\int_0^{L_b^{-1}}\Big(\exp\Big\{-pt-q\sum_{s\in[0,L(t))}\int_0^{\zeta(\epsilon_s)}\mathbf{1}_{\{\bar{\xi}_3(s)<\epsilon_s(t)<\bar{\xi}_2(s)\}}dt\Big\}\prod_{s<L(t)}\mathbf{1}_{\{\bar{\epsilon}_s\le\bar{\xi}_1(s)\}}\\
&\times\int_{\mathcal{E}}\exp\Big\{-p\rho^+_{\bar{\xi}_1(L(t))}-q\int_0^{\rho^+_{\bar{\xi}_1(L(t))}}\mathbf{1}_{\{\bar{\xi}_3(L(t))<\varepsilon(t)<\bar{\xi}_2(L(t))\}}dt\Big\}\mathbf{1}_{\{\bar{\varepsilon}>\bar{\xi}_1(L(t))\}}n(d\varepsilon)\Big)dL(t)\Big]\\
&=\mathbb{E}\Big[\int_0^{b}\Big(\exp\Big\{-pL^{-1}_{r-}-q\sum_{s\in[0,r)}\int_0^{\zeta(\epsilon_s)}\mathbf{1}_{\{\bar{\xi}_3(s)<\epsilon_s(t)<\bar{\xi}_2(s)\}}dt\Big\}\prod_{s<r}\mathbf{1}_{\{\bar{\epsilon}_s\le\bar{\xi}_1(s)\}}\\
&\times\int_{\mathcal{E}}\exp\Big\{-p\rho^+_{\bar{\xi}_1(r)}-q\int_0^{\rho^+_{\bar{\xi}_1(r)}}\mathbf{1}_{\{\bar{\xi}_3(r)<\varepsilon(t)<\bar{\xi}_2(r)\}}dt\Big\}\mathbf{1}_{\{\bar{\varepsilon}>\bar{\xi}_1(r)\}}n(d\varepsilon)\Big)dr\Big]\\
&=\int_0^b\mathbb{E}\Big[e^{-pL_r^{-1}-q\int_0^{L_r^{-1}}\mathbf{1}_{(\xi_2(\overline{X}_t),\xi_3(\overline{X}_t))}(X_t)dt}, L_r^{-1}<\tau_{\xi_1}\Big]\\
&\qquad\times n\Big(e^{-p\rho^+_{\bar{\xi}_1(r)}-q\int_0^{\rho^+_{\bar{\xi}_1(r)}}\mathbf{1}_{\{\bar{\xi}_3(r)<\varepsilon(t)<\bar{\xi}_2(r)\}}dt}\mathbf{1}_{\{\bar{\varepsilon}>\bar{\xi}_1(r)\}}\Big)dr\\
&=\int_0^b\mathbb{E}\Big[e^{-p\tau_r^+-q\int_0^{\tau_r^+}\mathbf{1}_{(\xi_2(\overline{X}_t),\xi_3(\overline{X}_t))}(X_t)dt}, \tau_r^+<\tau_{\xi_1}\Big]\\
&\qquad\times n\Big(e^{-p\rho^+_{\bar{\xi}_1(r)}-q\int_0^{\rho^+_{\bar{\xi}_1(r)}}\mathbf{1}_{\{\bar{\xi}_3(r)<\varepsilon(t)<\bar{\xi}_2(r)\}}dt}\mathbf{1}_{\{\bar{\varepsilon}>\bar{\xi}_1(r)\}}\Big)dr\\
&=\int_0^b e^{-\int_0^r\frac{\mathbb{W}'^{(p,q)}_{\bar{\xi}_2-\bar{\xi}_1}(\bar{\xi}_1(y),\bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\bar{\xi}_2-\bar{\xi}_1}(\bar{\xi}_1(y),\bar{\xi}_3(y))}dy}\, n\Big(e^{-p\rho^+_{\bar{\xi}_1(r)}-q\int_0^{\rho^+_{\bar{\xi}_1(r)}}\mathbf{1}_{\{\bar{\xi}_3(r)<\varepsilon(t)<\bar{\xi}_2(r)\}}dt}\mathbf{1}_{\{\bar{\varepsilon}>\bar{\xi}_1(r)\}}\Big)dr.
\end{aligned}$$

In the third equation, we applied the compensation formula in excursion theory; the fourth equation was obtained by the change of variable $r = L(t)$; and we again use the fact the $L_r^{-1} \neq L_{r-}^{-1}$ for at most countably many values of r in the fifth equation.

Now consider the case where $\xi_1 \equiv d \leq \xi_2 \equiv c \leq \xi_3 \equiv a \leq 0 \leq b$. Recall that by Loeffen *et al.* (2014) we have

$$\mathbb{E}_x\left[e^{-p\tau_0^- - q\int_0^{\tau_0^-} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_0^- < \tau_b^+\right] = \mathbb{Z}_c^{(p,q)}(x, x-a) - \frac{\mathbb{W}_c^{(p,q)}(x, x-a)}{\mathbb{W}_c^{(p,q)}(b, b-a)}\mathbb{Z}_c^{(p,q)}(b, b-a),$$

Or it can be modified as

$$\mathbb{E}\left[e^{-p\tau_d^- - q\int_0^{\tau_d^-} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_d^- < \tau_b^+\right] = \mathbb{Z}_{c-d}^{(p,q)}(-d, -a) - \frac{\mathbb{W}_{c-d}^{(p,q)}(-d, -a)}{\mathbb{W}_{c-d}^{(p,q)}(b-d, b-a)}\mathbb{Z}_{c-d}^{(p,q)}(b-d, b-a).$$

Thus

$$\begin{aligned}
&\mathbb{Z}_{c-d}^{(p,q)}(x-d, x-a) - \frac{\mathbb{W}_{c-d}^{(p,q)}(x-d, x-a)}{\mathbb{W}_{c-d}^{(p,q)}(b-d, b-a)}\mathbb{Z}_{c-d}^{(p,q)}(b-d, b-a) \\
&= \mathbb{E}_x\left[e^{-p\tau_d^- - q\int_0^{\tau_d^-} \mathbf{1}_{(c,a)}(X_t)dt}; \tau_d^- < \tau_b^+\right] \\
&= \int_0^b e^{-\int_0^r \frac{\mathbb{W}_{c-d}^{(p,q)\prime}(y-d,y-a)}{\mathbb{W}_{c-d}^{(p,q)}(y-d,y-a)}dy} n\left(e^{-p\rho_{r-d}^+ - q\int_0^{\rho_{\bar{\xi}_1(r)}^+} \mathbf{1}_{\{r-a<\varepsilon(t)<r-c\}}dt}\mathbf{1}_{\{\bar{\varepsilon}>r-d\}}\right)dr.
\end{aligned}$$

Differentiating with respect to b on both sides of the equation above will result in

$$\begin{aligned}
&n\left(e^{-p\rho_{b-d}^+ - q\int_0^{\rho_{\bar{\xi}_1(b)}^+} \mathbf{1}_{\{b-a<\varepsilon(t)<b-c\}}dt}\mathbf{1}_{\{\bar{\varepsilon}>b-d\}}\right) \\
&= \left(\frac{\mathbb{W}_{c-d}^{(p,q)\prime}(b-d, b-a)}{\mathbb{W}_{c-d}^{(p,q)}(b-d, b-a)}\mathbb{Z}_{c-d}^{(p,q)}(b-d, b-a) - \mathbb{Z}_{c-d}^{(p,q)\prime}(b-d, b-a)\right),
\end{aligned}$$

which implies that, for any $z \geq 0$ we have

$$\begin{aligned}
&n\left(e^{-p\rho_{\bar{\xi}_1(z)}^+ - q\int_0^{\rho_{\bar{\xi}_1(b)}^+} \mathbf{1}_{\{\bar{\xi}_3(z)<\varepsilon(t)<\bar{\xi}_2(z)\}}dt}\mathbf{1}_{\{\bar{\varepsilon}>\bar{\xi}_1(z)\}}\right) \\
&= \left(\frac{\mathbb{W}_{\xi_2(z)-\xi_1(z)}^{(p,q)\prime}(\bar{\xi}_1(z), \bar{\xi}_3(z))}{\mathbb{W}_{\xi_2(z)-\xi_1(z)}^{(p,q)}(\bar{\xi}_1(z), \bar{\xi}_3(z))}\mathbb{Z}_{\xi_2(z)-\xi_1(z)}^{(p,q)}(\bar{\xi}_1(z), \bar{\xi}_3(z)) - \mathbb{Z}_{\xi_2(z)-\xi_1(z)}^{(p,q)\prime}(\bar{\xi}_1(z), \bar{\xi}_3(z))\right).
\end{aligned}$$

As a result, Theorem 3.6 is proved for the case under $\mathbb{P}$. For the case of $X(0) = x$, it follows by the shifting argument. □

Indeed, we can add more terms into Theorem 3.6. That is, we can obtain the expression of the joint Laplace transform

$$\mathbb{E}_x\left[e^{-u\tau_{\xi_1}+\nu X(\tau_{\xi_1})-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_{\xi_1}<\tau_b^+\right].$$

But, first we need to introduce a new notation $\mathcal{W}_{a,c}^{(p,q)}(x)$, which is just the (p,q)-scale function under a new measure $\mathbb{P}^c$. That is, for $c,p,q\geq 0, p+q+\psi(c)\geq 0$, we have

$$\begin{aligned}
\mathbb{W}_{a,c}^{(p,q)}(x) &= W_c^{(p+q)}(x)-q\int_0^a W_c^{(p+q)}(x-y)W_c^{(p)}(y)dy\\
&= e^{-cx}W^{(p+q+\psi(c))}(x)-q\int_0^a e^{-c(x-y)}W^{(p+q+\psi(c))}(x-y)e^{-cy}W^{(p+\psi(c))}(y)dy\\
&= e^{-cx}W^{(p+q+\psi(c))}(x)-qe^{-cx}\int_0^a W^{(p+q+\psi(c))}(x-y)W^{(p+\psi(c))}(y)dy\\
&= e^{-cx}\left(W^{(p+q+\psi(c))}(x)-q\int_0^a W^{(p+q+\psi(c))}(x-y)W^{(p+\psi(c))}(y)dy\right)\\
&= e^{-cx}\mathcal{W}_a^{(p+\psi(c),q)}(x).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{Z}_{a,c}^{(p,q)}(x) &= Z_c^{(p+q)}(x)-q\int_0^a W_c^{(p+q)}(x-y)Z_c^{(p)}(y)dy\\
&= Z_c^{(p)}(x)+q\int_a^x W_c^{(p+q)}(x-y)Z_c^{(p)}(y)dy.
\end{aligned}$$

Now, we are ready for the following corollary.

Corollary 3.2. *For any $u,\nu,q\geq 0$ and $p=u-\psi(\nu)$, we have*

$$\begin{aligned}
&\mathbb{E}_x\left[e^{-u\tau_{\xi_1}+\nu X(\tau_{\xi_1})-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt}, \tau_{\xi_1}<\tau_b^+\right]\\
&=\int_x^b \exp\left\{\nu x-\int_x^z \frac{\mathbb{W}_{\xi_2(y)-\xi_1(y),\nu}^{(p,q)\prime}(\bar{\xi}_1(y),\bar{\xi}_3(y))}{\mathbb{W}_{\xi_2(y)-\xi_1(y),\nu}^{(p,q)}(\bar{\xi}_1(y),\bar{\xi}_3(y))}dy\right\}\\
&\quad\times\left(\frac{\mathbb{W}_{\xi_2(z)-\xi_1(z),\nu}^{(p,q)\prime}(\bar{\xi}_1(z),\bar{\xi}_3(z))}{\mathbb{W}_{\xi_2(z)-\xi_1(z),\nu}^{(p,q)}(\bar{\xi}_1(z),\bar{\xi}_3(z))}\mathbb{Z}_{\xi_2(z)-\xi_1(z),\nu}^{(p,q)}(\bar{\xi}_1(z),\bar{\xi}_3(z))-\mathbb{Z}_{\xi_2(z)-\xi_1(z),\nu}^{(p,q)\prime}(\bar{\xi}_1(z),\bar{\xi}_3(z))\right)dz.
\end{aligned}\tag{3.25}$$

Proof. First we assume $X(0)=0$, then by the change of measure, we have

$$\begin{aligned}
&\mathbb{E}\left[e^{-u\tau_{\xi_1}+\nu X(\tau_{\xi_1})-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt},\tau_{\xi_1}<\tau_b^+\right]\\
&=\mathbb{E}^{(\nu)}\left[e^{-u\tau_{\xi_1}+\psi(\nu)\tau_{\xi_1}-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt},\tau_{\xi_1}<\tau_b^+\right]\\
&=\mathbb{E}^{(\nu)}\left[e^{-p\tau_{\xi_1}-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt},\tau_{\xi_1}<\tau_b^+\right]\\
&=\int_0^b\exp\left\{-\int_0^z\frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(y)-\xi_1(y),\nu}(\bar{\xi}_1(y),\bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\xi_2(y)-\xi_1(y),\nu}(\bar{\xi}_1(y),\bar{\xi}_3(y))}dy\right\}\\
&\quad\times\left(\frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))}{\mathbb{W}^{(p,q)}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))}\mathbb{Z}^{(p,q)}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))-\mathbb{Z}^{(p,q)\prime}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))\right)dz.
\end{aligned}$$

To generalize this result to the case $\mathbb{P}_x$ with $x<b$, similar to the proof of Theorem (3.1), we define a function $\varsigma(y):=\xi(y+x)-x$. Then

$$\bar{\varsigma}(y)=y+x-\xi(y+x)=\bar{\xi}(y+x).$$

Because X is spatially homogeneous, we have $(X,\bar{X},\tau_\xi)\big|_{\mathbb{P}_x}=(x+X,x+\bar{X},\tau_\varsigma)\big|_{\mathbb{P}}$. Hence,

$$\begin{aligned}
&\mathbb{E}_x\left[e^{-u\tau_{\xi_1}+\nu X(\tau_{\xi_1})-q\int_0^{\tau_{\xi_1}}\mathbf{1}_{(\xi_2(\bar{X}_t),\xi_3(\bar{X}_t))}(X_t)dt},\tau_{\xi_1}<\tau_b^+\right]\\
&=e^{\nu x}\mathbb{E}\left[e^{-u\tau_{\varsigma_1}+\nu X(\tau_{\varsigma_1})-q\int_0^{\tau_{\varsigma_1}}\mathbf{1}_{(\varsigma_2(\bar{X}_t),\varsigma_3(\bar{X}_t))}(X_t)dt},\tau_{\varsigma_1}<\tau_{b-x}^+\right]\\
&=\int_0^{b-x}\exp\left\{\nu x-\int_0^z\frac{\mathbb{W}^{(p,q)\prime}_{\varsigma_2(y)-\varsigma_1(y),\nu}(\bar{\varsigma}_1(y),\bar{\varsigma}_3(y))}{\mathbb{W}^{(p,q)}_{\varsigma_2(y)-\varsigma_1(y),\nu}(\bar{\varsigma}_1(y),\bar{\varsigma}_3(y))}dy\right\}\\
&\quad\times\left(\frac{\mathbb{W}^{(p,q)\prime}_{\varsigma_2(z)-\varsigma_1(z),\nu}(\bar{\varsigma}_1(z),\bar{\varsigma}_3(z))}{\mathbb{W}^{(p,q)}_{\varsigma_2(z)-\varsigma_1(z),\nu}(\bar{\varsigma}_1(z),\bar{\varsigma}_3(z))}\mathbb{Z}^{(p,q)}_{\varsigma_2(z)-\varsigma_1(z),\nu}(\bar{\varsigma}_1(z),\bar{\varsigma}_3(z))-\mathbb{Z}^{(p,q)\prime}_{\varsigma_2(z)-\varsigma_1(z),\nu}(\bar{\varsigma}_1(z),\bar{\varsigma}_3(z))\right)dz\\
&=\int_x^b\exp\left\{\nu x-\int_x^z\frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(y)-\xi_1(y),\nu}(\bar{\xi}_1(y),\bar{\xi}_3(y))}{\mathbb{W}^{(p,q)}_{\xi_2(y)-\xi_1(y),\nu}(\bar{\xi}_1(y),\bar{\xi}_3(y))}dy\right\}\\
&\quad\times\left(\frac{\mathbb{W}^{(p,q)\prime}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))}{\mathbb{W}^{(p,q)}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))}\mathbb{Z}^{(p,q)}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))-\mathbb{Z}^{(p,q)\prime}_{\xi_2(z)-\xi_1(z),\nu}(\bar{\xi}_1(z),\bar{\xi}_3(z))\right)dz.
\end{aligned}$$

This completes the proof of Corollary 3.2. □

4 Other results related to draw-down time for SNLPs and future works

4.1 Other results related to draw-down time

4.1.1 Draw-down Parisian ruin for SNLPs

In risk theory, we are interested in the first time the process has stays below zero for a consecutive period of time that is greater than a pre-determined duration of length r, and we call this a Parisian ruin, denoted by κ_r. This concept was first initiated for option pricing in mathematical finance, then it was soon studied under the framework of Brownian motion and the Lundberg risk processes. Consequently, the Parisian ruin problem has been brought to SNLPs. Recently, with the new development of draw-down time, it is practical to consider the draw-down Parisian ruin for SNLPs. Wang and Zhou (2020) worked on this and found the solution to the two-sided exit problems via excursion theory as well as the potential measure for the process killed at the draw-down Parisian time.

For $r > 0$, the Parisian ruin time is defined by

$$\kappa_r := \inf\{t > r : t - g_t > r\} \qquad \text{where} \qquad g_t := \sup\{0 < s < t : X(s) \geq 0\}. \tag{4.1}$$

Given the draw-down function ξ, the draw-down Parisian ruin time is defined by

$$\kappa_r^{\xi} := \inf\{t > r : t - g_t^{\xi} > r\} \qquad \text{where} \qquad g_t^{\xi} := \sup\{0 < s < t : X(s) \geq \xi(\bar{X}(s))\}. \tag{4.2}$$

The following theorem in Wang and Zhou (2020) concerns about the two-sided exit problem with the draw-down Parisian ruin time.

Theorem 4.1. *For $x \in (-\infty, a)$ and $q \geq 0$, we have*

$$\mathbb{E}_x\left[e^{-q\tau_a^+}, \tau_a^+ < \kappa_r^{\xi}\right] = \exp\left\{-\int_x^a \frac{l_r^{(q)\prime}(\bar{\xi}(z))}{l_r^{(q)}(\bar{\xi}(z))}dz\right\}, \tag{4.3}$$

where

$$l_r^{(q)}(x) := \int_0^{\infty} W^{(q)}(x+z)\frac{z}{r}\mathbb{P}(X(r) \in dz) \tag{4.4}$$

can be viewed as a scale function associated to the Parisian ruin. We denote $l_r := l_r^{(0)}$ for simplicity.

The potential measure for the process killed at the draw-down Parisian time is also derived in Wang and Zhou (2020).

Theorem 4.2. *For any $q, \lambda \geq 0, r > 0, a \geq x$ and bounded differentiable function f, we have*

$$\int_0^\infty e^{-q(t-r)}\mathbb{E}_x\big(f(X(t)); t < \kappa_r^\xi \wedge \tau_a^+\big)dt \tag{4.5}$$
$$= \int_x^a \exp\Big\{-\int_x^w \frac{l_r^{(q)\prime}(\bar{\xi}(z))}{l_r^{(q)}(\bar{\xi}(z))}dz\Big\}\Big[\frac{l_r^{(q)\prime}(\bar{\xi}(w))}{l_r^{(q)}(\bar{\xi}(z))}\Big(\int_0^r e^{q(r-s)}\mathbb{E}\big(f(w+X(s))\big)ds$$
$$-\int_0^{\bar{\xi}(w)} W^{(q)}(\bar{\xi}(w)-z)\mathbb{E}\big(f(z+\xi(w)+X(r))\big)dz - \int_0^r \mathbb{E}\big(f(\xi(w)+X(r-s))\big)l_s^{(q)}(\bar{\xi}(w))ds\Big)$$
$$-\int_0^r e^{q(r-s)}\mathbb{E}\big(f'(w+X(s))\big)ds + \int_0^{\bar{\xi}(w)} W^{(q)\prime}(\bar{\xi}(w)-z)\mathbb{E}\big(f(z+\xi(w)+X(r))\big)dz$$
$$+W^{(q)}(0+)\mathbb{E}\big(f(w+X(r))\big) + \int_0^r \mathbb{E}\big(f(\xi(w)+X(r-s))\big)l_s^{(q)\prime}(\bar{\xi}(w))ds\Big]dw.$$

4.1.2 Draw-down reflected SNLPs

The reflected spectrally negative Lévy process has been well-known due to its wide range of applications in queuing theory, the optimal stopping problems, the optimal control problem and so on. To generalize these results, Wang and Zhou (2019) recently studied a new process R that is obtained by reflecting the SNLP from consecutive draw-down levels. Intuitively, for all the time before the first draw-down time of R, the process R agrees with the SNLP X. Right after that time, it evolves according to the process X reflected at the draw-down level until it reaches its previous height. The process keeps repeating the previous behavior for the updated draw-down times and draw-down levels. Wang and Zhou (2019) have worked out the Laplace transform of the upper exiting time for this process R.

Theorem 4.3. *For $q > 0$ and $x \leq b$, define the first passage time of $b \in \mathbb{R}$ for the process R as $\kappa_b^+ := \inf\{t \geq 0 : R(t) > b\}$, then we have*

$$\mathbb{E}_x[e^{-q\kappa_b^+}] = \exp\Big(-\int_x^b \frac{qW^{(q)}(\bar{\xi}(z))}{Z^{(q)}(\bar{\xi}(z))}dz\Big), \tag{4.6}$$

where the draw-down function $\xi(\cdot)$ and $\bar{\xi}(\cdot)$ are same in the previous sections.

Wang and Zhou (2019) also obtained an expression of the resolvent density for the process R.

Theorem 4.4. *For $q > 0$ and $x, r \leq b$, the resolvent measure of R is absolutely continuous with respect to the Lebesgue measure with density given by*

$$\int_0^\infty e^{-qt}\mathbb{P}_x[R(t) \in dr, t < \kappa_b^+]dt$$
$$= \int_x^b \exp\Big(-\int_x^y \frac{qW^{(q)}(\bar{\xi}(z))}{Z^{(q)}(\bar{\xi}(z))}dz\Big)\Big(W^{(q)\prime}(y-r) - \frac{qW^{(q)}(\bar{\xi}(y))}{Z^{(q)}(\bar{\xi}(y))}W^{(q)}(y-r)\Big)\mathbf{1}_{(\xi(y),y)}(r)dydr$$
$$+W^{(q)}(0)\exp\Big(-\int_x^r \frac{qW^{(q)}(\bar{\xi}(z))}{Z^{(q)}(\bar{\xi}(z))}dz\Big)\mathbf{1}_{(x,b)}(r)dr.$$

The expression of the expectation of the total discounted capital injections and the Laplace transform of the accumulated capital injections until time κ_b^+ are also given in Wang and Zhou (2019).

4.2 Future work

We are interested in the following quantity

$$\int_0^\infty e^{-pt}\mathbb{E}_x\left[e^{-q\int_0^t \mathbf{1}_{(\xi_2(\bar{X}_s),\xi_3(\bar{X}_s))}(X_s)ds}; X_t \in dy, t < \tau_b^+ \wedge \tau_{\xi_1}\right]dt, \tag{4.7}$$

where $\xi_1(x) \le \xi_2(x) \le \xi_3(x) \le x \le b$ for all $x \in \mathbb{R}$. This quantity is just a general version of (2.41) in Li and Palmowski (2018). That is

$$\begin{aligned} U^{(\omega)}(x, dy) &= \int_0^\infty e^{-pt}\mathbb{E}_x\left[\exp\left(-q\int_0^t (\mathbf{1}_{(a,b)}(X_s))ds\right); t < \tau_0^- \wedge \tau_c^+, X_t \in dy\right]dt. \\ &= \left(\frac{\mathcal{W}^{(\omega)}(x,0)}{\mathcal{W}^{(\omega)}(c,0)}\mathcal{W}^{(\omega)}(c,y) - \mathcal{W}^{(\omega)}(x,y)\right)dy. \end{aligned}$$

In order to work out the expression (4.7) using the excursion theory, our idea is first try to rewrite (4.7) in terms of excursion processes, then using the expression of $U^{(\omega)}(x, dy)$ above, we can obtain the expression of associated excursion measure. Therefore, we will be able to use the same technique that we have used in the proofs in this thesis to obtain the expression (4.7).

References

[1] Albrecher, H. and Hipp, C. (2007). Lundberge's risk process with tax. *Blatter der DVGFM*, **28(1)**, 13–28.

[2] Albrecher, H., Renaud, J.F. and Zhou, X. (2008). A Lévy insurance risk process with tax. *Journal of Applied Probability*, **45(2)**, 363–375.

[3] Albrecher, H. and Ivanovs, J. (2014). Power identities for Lévy risk models under taxation and capital injections. *Stochastic Systems*, **4(1)**, 157–172.

[4] Avram, F., Kyprianou, A.E. and Pistorius, M. (2004). Exits problems for spectrally negative Lévy processes and applications to Russian, American and Canadized options. *Annals of Applied Probability*, **14(1)**, 215–238.

[5] Avram, F., Vu, N.L. and Zhou, X. (2017). On taxed spectrally negative Lévy processes with draw-down stopping. *Insurance Mathematics and Economics*, **76**, 69–74.

[6] Barndorff, N., Mikosch, O.E. and Resnick, S. (2001). Lévy processes: theory and applications. *Birkhauser*.

[7] Baurdoux, E.J. (2009). Some excursion calculations for reflected Lévy Processes. *Latin American Journal of Probability and Mathematical Statistics*, **6**, 149–162.

[8] Bertoin, J. (1996). Lévy processes. *Cambridge University Press*.

[9] Biffis, E. and Kyprianou, A.E. (2010). A note on scale functions and the time value of ruin for Lévy insurance risk processes. *Insurance Mathematics and Economics*, **46(1)**, 85–91.

[10] Black, F. and Scholes, M. (1976). The pricing of options and corporate liabilities. *The Journal of Political Economy*, **81(3)**, 637–654.

[11] Cai, N., Chen, N., and Wan, X. (2010). Occupation times of jump-diffusion processes with double exponential jumps and the pricing options. *Mathematics of Operations Research*, **35**, 412–437.

[12] Chan, T., Kyprianou, A.E., and Savov, M. (2009). Smoothness of scale function for spectrally negative Lévy processes. *Probability Theory and Related Fields*, (in press).

[13] Chiu, S.N. and Yin, C. (2005). Passage times for a spectrally negative Lévy process with applications to risk theory. *Bernoulli*, **11(3)**, 511–522.

[14] De Finetti, B. (1929). Sulle funzioni ad incremento aleatorio. *Rendiconti della R. Accademia Nazionnale dei Lincei (Ser VI)*, **1**, 65–70.

[15] Doney, R.A. and Kyprianou, A.E. (2006). Overshoots and undershoots of Lévy processes. *Annals of Applied Probability*, **16(1)**, 91–106.

[16] Gerber, H.U. (1990). When does the surplus reach a given target? *Insurance: Mathematics and Economics*, **7**, 15–23.

[17] Getoor, R.K. (1961). First passage times for symmetric stable processes in space. *Transactions of the American Mathematical Society*, **101(1)**, 75–90.

[18] Hubalek, F., Kyprianou, A.E. (2010). Old and new examples of scale functions for spectrally negative Lévy processes. *Sixth Seminar on Stochastic Analysis, Random Fields and Applications, Progress in Probability, Birkhäuser*.

[19] Kuznetsov, A., Kyprianou, A.E. and Rivero, V. (2013). The theory of scale functions for spectrally negative Lévy processes. *Lévy Matters, Springer Lecture Notes in Mathematics*, **2061**, 97–186.

[20] Kyprianou, A.E., Schoutens, W. and Wilmott, P. (2005). Exotic option pricing and advanced Lévy models. *Wiley*.

[21] Kyprianou, A.E. (2006). Fluctuations of Lévy processes with applications. *Springer*.

[22] Kyprianou, A.E. and Zhou, X. (2009). General tax structures and the Lévy insurance risk model. *Journal of Applied Probability*, **46**, 1146–1156.

[23] Landriault, D., Renaud, J.F., and Zhou, X. (2011). Occupation times of spectrally negative Lévy processes with applications. *Stochastic Process and Their Applications*, **121**, 2629–2641.

[24] Landriault, D., Li, B. and Zhang, H. (2017). On magnitude, asymptotic and duration of draw-downs for Lévy models. *Bernoulli*, **23(1)**, 432–458.

[25] Lévy, P. (1939). Sur certains processus stochatiques homogènes. *Compositio Mathematica*, **7**, 283–339.

[26] Li, B. and Palmowski, Z. (2018). Fluctuations of Omega-killed spectrally negative Lévy processes. *Stochastic Processes and Their Applications*, **128(10)**, 3273–3299.

[27] Li, B., Vu, N.L. and Zhou, X. (2019). Exit problems for general draw-down times of spectrally negative Lévy processes. *Journal of Applied Probability*, **56(2)**, 441–457.

[28] Li, Y. and Zhou, X. (2014). On pre-exit joint occupation times for spectrally negative Lévy processes. *Statistics and Probability Letters*, **94**, 48–55.

[29] Loeffen, R.L., Renaud, J.F., and Zhou, X. (2014). Occupation times of intervals until first passage times for spectrally negative Lévy processes. *Stochastic Process and Their Applications*, **124**, 1408–1435.

[30] Lundberg, F. (1903). Approximerad framstallning av sannolikhetsfunktionen. *Aterforsakring av kollektivrisker,Almqvist and Wiksell, Uppsala*.

[31] Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, **3(1-2)**, 125–144.

[32] Pistorius, M.R. (2003). Exit problems of Lévy processes with applications in finance. *Proefschrift*.

[33] Prabhu, N.U. (1998). Stochastic storage processes. (2nd ed.). *Springer*.

[34] Rogozin, B.A. (1972). The distribution of the first hit for stable and asymptotically stable walks on an interval. *Theory of Probability and Its Applications*, **17(2)**, 332–338.

[35] Sato, K.I. (2013). Lévy processes and infinitely divisible distributions. *Cambridge*.

[36] Wang, W. and Zhou, X. (2019). A draw-down reflected spectrally negative Lévy process. *In press. Journal of Theoretical Probability*.

[37] Wang, W. and Zhou, X. (2020) Draw-down Parisian ruin for spectrally negative Lévy processes. *To appear in: Advances in Applied Probability*.

[38] Wu, L., and Zhou, J. (2016). Occupation times of hyper-exponential jump diffusion processes with application to price step options. *Journal of Computational and Applied Mathematics*, **294**, 251–274.

PART II: On estimation of entropy and residual entropy for nonnegative random variables

1 Introduction

Originally, the concept of entropy comes from a principle in a branch of physics called thermodynamics which deals with energy. A German physicist, Rudolph Clausius, used the word "entropy" to describe the measurement of disorder. However, entropy theory started to attract applied statisticians since a publication of a paper by Shannon in 1949, in which entropy is used to measure the unpredictability of the state (or of its information content). Suppose X is a discrete random variable with alphabet size K and associated cell probabilities $p_1, p_2, ..., p_K$ where $p_i > 0$ and $\sum_i p_i = 1$. An entropy H of this discrete random variable X is defined as

$$H = -\sum_{i=1}^{K} p_i \log p_i. \tag{1.1}$$

If X is a continuous random variable with probability density function (pdf) $f(x)$, then the entropy is defined as

$$H_f = -\int_{-\infty}^{\infty} f(x) \log f(x) dx. \tag{1.2}$$

After the introduction to entropy by Shannon, many researchers have been studying this quantity and it becomes a central place in statistical theory and applications (see Cover and Thomas, 1991 Kapur, 1993 and Kapur and Kesavan, 1992). One of the well known applications of entropy in statistics is the test of normality for a random variable because of the characterizing property that the normal distribution attains the maximum entropy among all continuous distributions with a given variance [see Vacicek, 1976 and an adaptation to testing exponentiality by Chaubey, Mudholkar and Smethurst, 1993].

Ebrahimi (1996) provided an interpretation of H_f as a measure of uncertainty associated with f. In the context of life testing, X being the lifetime of a unit, knowing that the unit has survived up to time t, a more appropriate measure for uncertainty in f is defined as

$$H_f(t) = -\int_{t}^{\infty} \left(\frac{f(x)}{R(t)}\right) \left(\log \frac{f(x)}{R(t)}\right) dx, \tag{1.3}$$

where $R(t) = \mathbb{P}(X > t)$ is the reliability or the survival function. $H_f(t)$ is called *residual entropy* of X given the event $\{X > t\}$. In this thesis, we focus on estimating H_f and more generally $H_f(t)$ when X is continuous and more specifically, when X is a nonnegative random variable.

However, in practice the underlying pdf (or pmf for the discrete case) of a random variable is unknown, so the question of entropy estimation comes naturally. In recent decades, there exists a huge number of researchers working on entropy estimation, and their proposals have formed a collection of estimators which can be classified as follows.

(i) A naive approach is to discretize the support of the underlying density function into m bins, then for each bin we compute the empirical measures, corresponding to $p_i = \int_{A_i} f(x)dx$. That is if $X_1, X_2, ..., X_n$ are independently and identically distributed (i.i.d.) continuous random variables, then $\hat{p}_i = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{A_i}(X_j)$, where $\{A_i\}_{i=1}^{m}$ are mutually disjoint subsets of $\mathbb{R}$ such that $\cup_{i=1}^{m} A_i = \text{support}(X)$. Clearly, $\hat{p}_i$ is the maximum likelihood estimator (MLE) for p_i. With this estimator $\hat{p}_i$, a "naive" or "plug-in" MLE estimator of entropy is given as

$$\hat{H}_f^{MLE} = -\sum_{i=1}^{m} \hat{p}_i \log \hat{p}_i. \tag{1.4}$$

It has been shown that this MLE estimator of entropy results in heavy bias. Consequently, the problem of bias correction and the choice of m have drawn the attention of many researchers; see Paninski (2003) for a detailed account.

(ii) One straightforward approach for entropy estimation is to estimate the underlying density function $f(x)$ by some well-known density estimator $\hat{f}_n(x)$, then plug it into (1.2) to obtain the entropy estimator

$$\hat{H}_f^{Plugin} = -\int_0^{\infty} \hat{f}_n(x) \log \hat{f}_n(x) dx. \tag{1.5}$$

The *fixed symmetric kernel density estimator* $\hat{f}_n^{Fixed}(x)$, which is already a well-known and popular approach for estimating the pdf with an unbounded support, is an example of density estimator which is defined as

$$\hat{f}_n^{Fixed}(x) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right), \tag{1.6}$$

where K is a symmetric density function with mean zero and variance one, and $b := b(n)$ is the smoothing parameter, called the bandwidth. It is obvious that the performance of this plugin entropy estimator totally depends on the density estimator $\hat{f}_n$. However, when dealing with non-negative random variables, this $\hat{f}_n^{Fixed}(\cdot)$ is shown to produce a heavy bias near the boundary. Consequently, it would result in a better estimator if we replace $\hat{f}_n^{Fixed}$ by some density estimators that is free of boundary effect.

(iii) Motivated by the representation of entropy as an expected value

$$H_f = -\int_0^{\infty} f(x) \log f(x) dx = -\mathbb{E}[\log f(X)].$$

By the strong law of large number we have $-\frac{1}{n}\sum_{i=1}^{n}\log f(X_i) \overset{a.s.}{\to} H_f$. Thus we obtain a new entropy estimator if we replace $f(\cdot)$ by an appropriate density estimator $\hat{f}_n(\cdot)$, as given by

$$\hat{H}_f^{Meanlog} = -\frac{1}{n}\sum_{i=1}^{n}\log \hat{f}_n(X_i). \tag{1.7}$$

(iv) The entropy can be estimated by another approach, called "*spacing*", which is initiated by Vasicek (1975). By the change of variable $p = F(x)$, the entropy (1.2) can be expressed in the form

$$H_f = -\int_{-\infty}^{\infty} f(x)\log f(x)dx = \int_0^1 \log\left\{\frac{d}{dp}F^{-1}(p)\right\}dp.$$

To estimate H_f, the distribution F is replaced by the empirical distribution F_n, and the differential operator is replaced by the difference operator. As a result, the derivative of $F^{-1}(p)$ is estimated by $\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})$ for $(i-1)/n < p \le i/n$, $i = m+1, m+2, ..., n-m$, where $X_{(i)}$'s are the order statistics and m is a positive integer smaller than $n/2$. When $p \le m/n$ or $p > (n-m)/n$, one-sided differences are used. That is, $(X_{(i+m)} - X_{(1)})$ and $(X_{(n)} - X_{(i-m)})$ are in place of $(X_{(i+m)} - X_{(i-m)})$ respectively. All together this leads to the following estimator of entropy

$$\hat{H}_f^{Vasicek} = \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}. \tag{1.8}$$

Motivated by the idea of spacing, researchers have followed this direction and proposed other versions of entropy estimator, which are claimed to have a better performance. We will list some of them in the next section. One of the greatest weakness of this spacing estimator is the choice of spacing parameter m, which does not have the optimal form.

(v) Lastly, different from all of the above estimators, Bouzebda *et al.* (2013) presented a potentially estimator of entropy based on smooth estimator of quantile density function. Their idea again starts with the expression of the entropy

$$H_f = \int_0^1 \log\left\{\frac{d}{dp}F^{-1}(p)\right\}dp = \int_0^1 \log\left\{\frac{d}{dp}Q(p)\right\}dp = \int_0^1 \log q(p)dp,$$

where $Q(p) := \inf\{t : F(t) \ge p\}$ for $0 \le p \le 1$ is the quantile function and $q(p) := dQ(p)/dp = 1/f(Q(p))$ is the quantile density function. Then a new entropy estimator can be obtained by substituting $q(\cdot)$ by its appropriate estimator $\hat{q}_n(\cdot)$. That is

$$\hat{H}_f^{Quantile} = \int_0^1 \log \hat{q}_n(p)dp. \tag{1.9}$$

Bouzebda *et al.* (2013) were motivated by the work of Cheng and Parzen (1997), which introduced a kernel type estimator $\hat{q}_n(\cdot)$ that has good asymptotic properties.

On the other hand, for the residual entropy estimation, Belzunce *et al.* (2001) suggested that we can rewrite (1.3) as

$$H_f(t) = \log(R(t)) - \frac{1}{R(t)} \int_t^\infty f(x) \log f(x) dx. \tag{1.10}$$

Then the residual entropy can be estimated if we replace $R(t)$ by an empirical or kernel estimator $\hat{R}(t)$, and an estimator $\hat{g}_n(x) = \hat{f}_n(x) \log \hat{f}_n(x)$ in place of the functional $g(x) = f(x) \log f(x)$.

The organization of this thesis is as follow. Section 2 will explore more detailed properties and behaviors of existing estimators introduced above. Our main results are in the Section 3, in which we show our proposed estimators of entropy and residual entropy. Also, we obtain some asymptotic properties of our estimators. The simulation study to show and compare the performance between our proposals and the existing ones will be carried out in the Section 4.

2 Properties of entropy and residual entropy estimators

2.1 Estimation by direct use of density estimators

Obviously, this is one of the most straightforward approach in enstimating entropy. With the density estimator on hand, the entropy and residual entropy can be estimated by

$$\begin{aligned}
\hat{H}_f^{Plugin} &= -\int_0^\infty \hat{f}_n(x) \log \hat{f}_n(x) dx, \\
\hat{H}_f^{Meanlog} &= -\frac{1}{n} \sum_{i=1}^n \log \hat{f}_n(X_i), \\
\hat{H}_f^{Plugin} &= \log \hat{R}(t) - \frac{1}{\hat{R}(t)} \int_t^\infty \hat{f}_n(x) \log \hat{f}_n(x) dx, \\
\text{where } \hat{R}(t) &= \int_t^\infty \hat{f}_n(x) dx.
\end{aligned}$$

When dealing with pdfs with bounded support, the fixed symmetric kernel density estimator has been shown to have poor performance especially for those estimations near the boundaries, which is so-called the "boundary effect", "boundary bias", or the "spill-over effect". To overcome these issues, in literature, there are a huge number of papers concerning the estimation of density function with bounded support (see, e.g., Müller (1991), Marron and Ruppert (1994), Jones (1993), Jones and Foster (1996)). Namely, some of approaches are: fixed kernel density estimator with boundary bias correction, varying asymmetric kernel density estimator, transformation kernel density estimator, and so on. Roughly speaking, although there exists many kernel density estimators for pdf with bounded support, so far none of them seems to dominate the others both in terms of precision (smallest mean integrated squared error (MISE)) and performance (the rate of convergence).

2.1.1 Fixed symmetric kernel density estimators with bias correction

The first naive idea of fixing the boundary effect is to modify the standard kernel density estimator in such a way that the resulting estimator will have less bias while keeping the order of convergence. However, even though this method seems to work well in certain cases, it does not produce an estimator with good properties. For instance, the new density estimator may be negative, may not have the same support as the true underlying pdf, or may not integrate to one. Some examples of this type are

- The reflection density estimator (Schuster, 1958)

$$\hat{f}_n^{Ref}(x) = \frac{1}{nb}\sum_{i=1}^{n}\left[K\left(\frac{x-X_i}{b}\right) + K\left(\frac{x+X_i}{b}\right)\right], \tag{2.1}$$

  where K is any symmetric density function and b is the bandwidth. Note that the bandwidth is a function of n but for simplification, we use b for $b(n)$.

- The cut-and-normalized density estimator (Gasser and Müller ,1979)

$$\hat{f}_n^{CN}(x) = \frac{\frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x-X_i}{b}\right)}{K_0(p)}, \tag{2.2}$$

  where $(-S_K, S_K)$ is the support of kernel K, $p = x/b$, and $K_j(p) = \int_{-S_K}^{\min\{p,S_K\}} u^j K(u)du$.

- The bounded density estimator (Jones, 1993)

$$\hat{f}_n^{Bound}(x) = \alpha_x \hat{f}_n^{CN}(x, K) + \beta_x \hat{f}_n^{CN}(x, L), \tag{2.3}$$

  where $\hat{f}_n^{CN}(x, K)$ and $\hat{f}_n^{CN}(x, L)$ are the cut-and-normalized density estimator with 2 different kernel K and kernel L, and

$$\alpha_x = \frac{L_1(p)K_0(p)}{L_1(p)K_0(p) - K_1(p)L_0(p)} \qquad \beta_x = \frac{-K_1(p)L_0(p)}{L_1(p)K_0(p) - K_1(p)L_0(p)}.$$

- The Jones and Foster estimator (Jones and Foster, 1996)

$$\hat{f}_n^{J\&F}(x) = \hat{f}_n^{CN}(x)\exp\left\{\frac{\hat{f}_n^{Bound}(x)}{\hat{f}_n^{CN}(x)} - 1\right\}. \tag{2.4}$$

2.1.2 Varying Asymmetric kernel density estimators

Another simple idea in fixing the boundary effect is to replace the symmetric kernel K by an asymmetric kernel density with the shape parameter as a function of the point of estimation x. Some properties of varying asymmetric kernel density estimators, which make them one of the best candidates for density estimation, are nonnegative, same support as the true underlying density function, and integrating to unity. The latter is an obvious property because the asymmetric kernel itself is a density function. Below we give some examples of potentially good asymmetric kernels along with their bias, variance, and mean squared error (where available).

- The inverse Gaussian kernel estimator (Bouezmarni and Scaillet, 2005)

$$\hat{f}_n^{IG}(x) = \frac{1}{n}\sum_{i=1}^{n} K^{IG}(x, b, X_i), \tag{2.5}$$

where $K^{IG}(x, b, t)$ is the inverse Gaussian density with the mean x and the shape parameter $1/b$. That is

$$K^{IG}(x, b, t) = \frac{1}{\sqrt{2\pi b t^3}} \exp\left(-\frac{1}{2bx}\left(\frac{t}{x} - 2 + \frac{x}{t}\right)\right).$$

- The reciprocal inverse Gaussian kernel estimator (Bouezmarni and Scaillet, 2005)

$$\hat{f}_n^{RIG}(x) = \frac{1}{n}\sum_{i=1}^{n} K^{RIG}(x, b, X_i), \tag{2.6}$$

where $K^{RIG}(x, b, t)$ is the reciprocal inverse Gaussian density

$$K^{RIG}(x, b, t) = \frac{1}{\sqrt{2\pi b t}} \exp\left(-\frac{x-b}{2b}\left(\frac{t}{x-b} - 2 + \frac{x-b}{t}\right)\right).$$

- The modified Gamma kernel estimator (Chen, 2000)

$$\hat{f}_n^{Gam}(x) = \frac{1}{n}\sum_{i=1}^{n} K^{Gam}(x, b, X_i), \tag{2.7}$$

where $K^{Gam}(x, b, t)$ is the pdf of the Gamma distribution with the scale parameter b and the shape parameter $x/b + 1$.

$$K^{Gam}(x, b, t) = \frac{t^{x/b}\exp(-t/b)}{b^{x/b+1}\Gamma(x/b+1)}.$$

If $b \to 0$ and $nb \to \infty$ as $n \to \infty$, its asymptotic bias and variance have the form

$$\begin{aligned} Bias\big[\hat{f}^{Gam}(x)\big] &= bf'(x) + \frac{1}{2}bxf''(x) + o(b), \\ Var\big[\hat{f}^{Gam}(x)\big] &\approx \begin{cases} \frac{f(x)}{nb}\mathcal{C}_b(x) & \text{if } x/b \to \kappa, \\ \frac{f(x)}{2n\sqrt{\pi b x}} & \text{if } x/b \to \infty, \end{cases} \end{aligned}$$

where κ is a nonnegative constant, and $\mathcal{C}_b(x) = \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}$.

The following theorem from Bouezmarni and Scaillet (2005) shows the uniformly weak consistency of these varying asymmetric kernel estimators mentioned above.

Theorem 2.1. *Let f be a continuous and bounded probability density function on $[0, \infty)$ and $\hat{f}_b$ be one of the three varying asymmetric kernel density estimators mentioned above. For any compact set I in $[0, \infty)$, if $\lim_{n\to\infty} b = 0$ and $\lim_{n\to\infty} nb^{2a} = \infty$, where $a = 1$ for the Gamma kernel and $a = \frac{5}{2}$ for the IG and RIG kernels. Then,*

$$\sup_{x \in I} |\hat{f}_b(x) - f(x)| \xrightarrow{\mathbb{P}} 0 \qquad \text{as } n \to \infty.$$

For the case of unbounded density function at $x = 0$, in order to achieve the weak convergence result, the kernel density must satisfy the condition that it will assign almost all weight to the boundary points as $b \to 0$. Bouezmarni and Scaillet (2005) also provided the convergence of these estimators in this case.
Later on, Bouezmarni *et al.* (2011) obtained the asymptotic normality of $\hat{f}_n^{Gam}(x)$ which is given in the following theorem.

Theorem 2.2. *Under assumptions that $f(x)$ is twice continuous differentiable, $\log^2 n/(nb^{3/2}) \to 0$, and $nb^{5/2} = o(1)$, we have*

$$n^{1/2}b^{1/4}\left(\frac{\hat{f}_n^{Gam}(x) - f(x)}{\sqrt{Var[\hat{f}_n^{Gam}(x)]}}\right) \xrightarrow{D} \mathcal{N}(0, 1).$$

2.1.3 Data-transformed kernel density estimators

Beside the above ideas, one can easily find a solution for boundary effect problem by applying the usual symmetric kernel density estimator on the transformed data. Let $g_\lambda(\cdot)$ be a monotonic increasing transformation function with the transformation parameter λ in such a way that the transformed data are defined as $Y_i = g_\lambda(X_i)$ for $i = 1, ..., n$. Then the transformed kernel density estimator is given by

$$\hat{f}_\lambda(x) = \frac{1}{nb}\sum_{i=1}^{n} g'_\lambda(x) K\left(\frac{g_\lambda(x) - g_\lambda(X_i)}{b}\right), \tag{2.8}$$

where $K(\cdot)$ is a symmetric kernel density with mean zero. It is pointed in Koekemoer and Swanepoel (2008) that the choice of $K(\cdot)$ is not crucial. Thus the standard normal density is chosen for the computation simplification reason. Here we only present the special case of the function $g_\lambda(\cdot)$ as the logarithmic function.

$$
\begin{aligned}
\hat{f}_n^{LogTrans}(x) &= \frac{1}{nb}\sum_{i=1}^{n}\phi\biggl(\frac{\log x-\log X_i}{b}\biggr)\frac{1}{x}\\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{xb\sqrt{2\pi}}\exp\biggl\{-\frac{1}{2}\biggl(\frac{\log x-\log X_i}{h}\biggr)^2\biggr\}\\
&= \frac{1}{n}\sum_{i=1}^{n}LN(x,\log X_i,b),
\end{aligned}
$$

where $\phi(\cdot)$ is the standard Gaussian density and $LN(x,\log X_i,b)$ is the log-normal density with medians X_i and variance $(e^{b^2}-1)e^{b^2}X_i^2$. That is the kernel density estimation based on a log-transformed of data is identical to the log-normal kernel density estimation in the original data. Moreover, its bias and variance are shown in Charpentier and Flachaire (2015) to have the form

$$
\begin{aligned}
Bias[\hat{f}_n^{LogTrans}(x)] &\approx \frac{b^2}{2}\biggl(f(x)+3xf'(x)+x^2f''(x)\biggr),\\
Var[\hat{f}_n^{LogTrans}(x)] &\approx \frac{1}{nbx}f(x)\int_{\infty}^{\infty}K^2(u)du.
\end{aligned}
$$

2.1.4 Other approaches for density estimation

To end this sub-section, we introduce a different approach proposed by Chaubey *et al.* (2012). The idea is based on smoothing the empirical distribution function by the generalization of Hille's lemma.

$$
\hat{f}_{v,\epsilon}(x)=\frac{1}{n(x+\epsilon)^2}\sum_{i=1}^{n}X_iq_v\biggl(\frac{X_i}{x+\epsilon}\biggr), \tag{2.9}
$$

where $q_v(\cdot)$ is any density with mean $\mu_n\to x$ and variance $v_n^2\to 0$. If we take $g_{x,v}(t):=\frac{1}{x}q_v(t/x)$, where $q_v(\cdot)$ is the gamma density with shape $\alpha=1/v^2$ and scale $\beta=1/\alpha$, then the density estimator $\hat{f}_{v,\epsilon}(x)$ can be rewritten as

$$
\hat{f}_{v,\epsilon}(x)=\frac{1}{n}\sum_{i=1}^{n}K^{Gam}_{1/v^2,(x+\epsilon)v^2}(X_i)=\frac{1}{n}\sum_{i=1}^{n}\frac{X_i^{1/v^2-1}\exp(-X_i/((x+\epsilon)v^2))}{b^{1/v^2}\Gamma(1/v^2)}.
$$

With this form, we can compare it with other varying asymmetric kernel density estimators introduced above. By direct computation, one can obtain the bias and variance of this estimator shown below

$$
\begin{aligned}
Bias[\hat{f}_{v,\epsilon}(x)] &= (xv^2+\epsilon)f'(x)+o(v^2+\epsilon),\\
Var[\hat{f}_{v,\epsilon}(x)] &= \frac{I_2(q)f(x)}{nv(x+\epsilon)}+o((vn)^{-1}),
\end{aligned}
$$

where $I_2(q) = \lim_{v\to 0} v \int_0^\infty (q_v(t))^2 dt$ and v, ϵ are such that $v^2 \to 0, nv \to \infty$ and $\epsilon \to 0$, as $n \to \infty$. The uniformly strong consistency and asymptotic normality of the density estimator $\hat{f}_{v,\epsilon}(x)$ hold under certain conditions on the density $q_v(\cdot)$ and the true underlying pdf $f(x)$, which are provided in Chaubey *et al.* (2012).

Theorem 2.3 (Uniformly strong consistency)**.** *If*

(i) $\sup_{x\geq 0} \int_0^\infty \left|\frac{d}{dx}(q_v(t/(x+\epsilon)))\right| dt = o\left(\frac{n^{\frac{1}{2}}}{\log\log n}\right)$,

(ii) $\sup_{u,v>0} uq_v(u) < \infty$,

(iii) $f(\cdot)$ *is Lipschitz continuous on* $[0, \infty)$,

then as $n \to \infty$, *we have* $\sup_{x\geq 0} |\hat{f}_{v,\epsilon}(x) - f(x)| \xrightarrow{a.s.} 0$.

Theorem 2.4 (Asymptotic normality)**.** *Define* $q^*_m, v(t) := \frac{(q_v(t))^m}{\int_0^\infty (q_v(u))^m du}$, $\mu_{m,v} := \int_0^\infty t q^*_{m,v}(t)dt$, *and* $\sigma^2_{m,v} := \int_0^\infty (t - \mu_{m,v})^2 q^*_{m,v}(t)dt$, *if*

(i) $f(\cdot)$ *is continuously differentiable on* $[0, \infty)$,

(ii) $\int_0^\infty (q_v(t))^m dt = O(v^{-(m-1)})$ *as* $v \to 0$, *for* $1 \leq m \leq 3$, *and* $I_2(q)$ *exists,*

(iii) $\mu_{m,v} = 1 + O(v)$; $\quad \sigma^2_{m,v} = O(v^2)$; $\quad \sup_{0<v<\eta} \int_0^\infty t^{4+\delta} q^*_{m,v}(t)dt < \infty$ *for some* $\delta, \eta > 0$.

(iv) $nv \to \infty, \quad nv\epsilon \to \infty, \quad nv^5 \to 0, \quad nv\epsilon^2 \to 0$ *as* $n \to \infty$,

then

$$\begin{aligned} \sqrt{nv}\left(\hat{f}_{v,\epsilon}(x) - f(x)\right) &\xrightarrow{D} \mathcal{N}\left(0, I_2(q)\frac{f(x)}{x}\right) \text{ for } x > 0 \\ \sqrt{nv}\left(\hat{f}_{v,\epsilon}(0) - f(0)\right) &\xrightarrow{D} \mathcal{N}\left(0, I_2(q)f(0)\right). \end{aligned}$$

2.2 Spacing entropy estimators

As mentioned in the Introduction section, entropy can be estimated by a well-known non-parametric approach called ”spacing”. However, this approach has not been generalized to the residual entropy estimation. In this sub-section, we introduce some common estimators of entropy using spacing approach.

- Vasicek's estimator (Vasicek, 1976)

$$\hat{H}_f^{Vasicek} = \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}. \tag{2.10}$$

It will be easy to study the properties of $\hat{H}_f^{Vasicek}$ if we rewrite it as

$$\hat{H}_f^{Vasecek} = -\frac{1}{n}\sum_{i=1}^{n}\log f(X_i) + V_m + U_m,$$

where

$$\begin{aligned} V_m &= \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{F(X_{(i+m)}) - F(X_{(i-m)})}{f(X_{(i)})(X_{(i+m)} - X_{(i-m)})}\right], \\ U_m &= \frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{n}{2m}(F(X_{(i+m)}) - F(X_{(i-m)})\right]. \end{aligned}$$

The first term of the above expression is the minimum variance unbiased estimate of H_f given the variance of $-\log f(X_i)$ is finite, and the values $f(X_1)$, $f(X_2)$,...,$f(X_n)$. V_m is the error estimation due to the estimation of derivative by finite differences, and U_m is the error estimation due to the estimation of increments of F by F_n. The behavior of both V_m and U_m are in opposite direction, so simultaneously minimizing them requires $m \to \infty$ and $m/n \to 0$. Vasicek showed that U_m converges to zero in probability as $n, m \to \infty$, so does $\hat{H}_f^{Vasicek}$. Furthermore, it can be shown that under appropriate conditions on m and n, the Vasicek's estimator is also asymptotically normal.

- Ebrahimi's estimator (Ebrahimi *et al.*, 1992)

$$\hat{H}_f^{Ebrahimi} = \frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{n}{c_i m}(X_{(i+m)} - X_{(i-m)})\right\}, \tag{2.11}$$

where

$$c_i = \begin{cases} 1 + \frac{i-1}{m} & 1 \le i \le m, \\ 2 & m+1 \le i \le n-m, \\ 1 + \frac{n-i}{m} & n-m+1 \le i \le n. \end{cases}$$

It was shown that $\hat{H}_f^{Ebrahimi} \xrightarrow{\mathbb{P}} H_f$ as $n, m \to \infty$, and $m/n \to 0$. Also, its bias is smaller than that of Vasicek's.

- Van Es estimator (Van Es, 1992)

$$\hat{H}_f^{VanEs} = -\frac{1}{n-m}\sum_{i=1}^{n-m}\log\left\{\frac{n+1}{m}(X_{(i+m)} - X_{(i)})\right\} + \sum_{k=m}^{n}\frac{1}{k} + \log(m) - \log(n+1). \tag{2.12}$$

Van established, under certain conditions, the strong consistency and asymptotic normality of $\hat{H}_f^{VanEs}$.

- <u>Correa's estimator (Correa, 1995)</u>

Motivation from the Vasicek's estimator of entropy, Correa studied the behavior of this estimator and suggested the estimation of entropy using local linear model based on $2m+1$ points: $F(X_{(j)}) = a + bX_{(j)}$ $j = m-i, ..., m+i$.

$$\hat{H}_f^{Correa} = -\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})(j-i)}{n\sum_{j=i-m}^{i+m}(X_{(j)} - \bar{X}_{(i)})^2}\right\}, \tag{2.13}$$

where $\bar{X}_{(i)} := \frac{1}{2m+1}\sum_{j=i-m}^{i+m} X_j$. The resulting estimator of entropy is shown to attain a smaller mean squared error compared to the Vasicek's estimator.

- <u>WG's estimator (Wieczorkowski and Grzegorzewski, 1999)</u>

By modifying the Vasicek's estimator, Wieczorkowski and Grzegorzewski obtained a new estimator of entropy whose bias is shown to be smaller than that of Vasicek's.

$$\hat{H}_f^{WG} = \hat{H}_f^{Vasicek} - \log n + \log(2m) - \left(1 - \frac{2m}{n}\right)\Psi(2m) \tag{2.14}$$

$$+\Psi(n+1) - \frac{2}{n}\sum_{i=1}^{m}\Psi(i+m-1),$$

where $\Psi(k) = \sum_{i=1}^{k-1}\frac{1}{i} - \gamma$ is the di-Gamma function defined on integer set, and $\gamma = 0.57721566...$ is the Euler's constant.

- <u>Noughabi's estimator (Noughabi, 2010)</u>

By combining both spacing and kernel methodology, Noughabi presented a new potential estimator of entropy. He followed exactly the initial steps in Vasicek (1975), but he approximated the difference $(F(X_{(i+m)}) - F(X_{(i-m)})$ by the first order of Taylor expansion to get back to estimator involving the density function $f(x)$. Then by putting the fixed kernel estimator $\hat{f}_n^{Fixed}(x)$ in place of $f(x)$, he obtained a new estimator for entropy as below.

$$\hat{H}_f^{Noughabi} = -\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{\hat{f}_n^{Fixed}(X_{(i+m)}) + \hat{f}_n^{Fixed}(X_{(i-m)})}{2}\right\}, \tag{2.15}$$

where the bandwidth is fixed to $b = 1.06sn^{-1/5}$ and s is the sample standard deviation.

Similar to $\hat{H}_f^{Vasicek}$, Noughabi showed that $\hat{H}_f^{Noughabi}$ weakly converges to H_f using the same proof. Also, he showed that the scale of the random variable X has no effect on the accuracy of $\hat{H}_f^{Nougahbi}$.

For the choice of spacing parameter m, Dudewicz and van de Meulen (1981) discussed about the optimal choice of m, but in the the context of spacing entroy estimator for testing uniformly. For a class of alternatives, one can consider the following choice of m:

$$m = \lfloor n/2 - 1 \rfloor, \tag{2.16}$$

$$m = \lfloor \sqrt{n} + 0.5 \rfloor. \tag{2.17}$$

These suggested choice of m above may be a working rule in providing the best power. However, whether it will be the best in still unknown.

2.3 Estimation by direct use of quantile density estimators

Recall that, the entropy can be expressed in terms of quantile density function $q(p)$ as

$$H_f = \int_0^1 \log q(p) dp.$$

Therefore, one straightforward way to obtain the entropy and residual estimator is to plugin directly the estimator of quantile density function. That is

$$\hat{H}_f^{Quantile} = \int_0^1 \log \hat{q}_n(p) dp.$$

This approach was proposed by Bouzebda *et al.* (2013), in which the kernel-type quantile density estimator $\hat{q}_n^{CP}(p)$ suggested by Cheng and Parzen (1997) is in place. However, since $\hat{q}_n^{CP}(p)$ performs very poor at the boundaries, Bouzebda and co-authors had estimated $H_f(\epsilon)$ instead, where

$$H_{\epsilon,f} = \epsilon \log q(\epsilon) + \epsilon \log(q(1-\epsilon)) + \int_\epsilon^{1-\epsilon} \log q(p) dp, \qquad \epsilon \in (0, 0.5).$$

We observe that $|H_f - H_{\epsilon,f}| = o(\eta(\epsilon))$ if $\eta(\epsilon) \to 0$ as $\epsilon \to 0$. Therefore, the estimator of entropy given by Bouzebda *et al.* (2013) has the following form

$$\hat{H}_{\epsilon,f}^{Quantile} = \epsilon \log \hat{q}_n^{CP}(\epsilon) + \epsilon \log(\hat{q}_n^{CP}(1-\epsilon)) + \int_\epsilon^{1-\epsilon} \log \hat{q}_n^{CP}(p) dp, \tag{2.18}$$

$$\hat{q}_n^{CP}(p) := \frac{d}{dp} \int_0^1 \hat{Q}_n(x) K_n(p, x) d\mu_n(x), \tag{2.19}$$

where $\hat{Q}_n(\cdot)$ is the empirical quantile function, $K_n(p, x)$ is the sequence density kernel functions defined on $(0, 1) \times [0, 1]$, and $\mu_n(x)$ is a sequence of σ-finite measure on $[0, 1]$. The asymptotic properties of $\hat{H}_{\epsilon,f}^{Quantile}$ had been studied thoroughly by Bouzebda *et al.* (2013) which are based on the following assumptions of the true quantile density function:

Q1 The quantile density function $q(\cdot)$ is twice differentiable in $(0,1)$.

Q2 There exists a constant $\gamma > 0$ such that $\sup_{t\in(0,1)}\{t(1-t)|\frac{d}{dt}\log q(t)|\} \leq \gamma$.

Q3 Either $q(0) < \infty$ or $q(t)$ is non-increasing in some interval $(0, t_*)$, and either $q(1) < \infty$ or $q(t)$ is non decreasing in some interval $(t^*, 1)$, where $0 < t_* < t^* < 1$.

Q4 $\mathbb{E}[\log^2(q(F(X)))] < \infty$.

Also, for $U(\epsilon) = [\epsilon, 1-\epsilon]$, the sequence of kernel $K_n(\cdot,\cdot)$ needs to satisfy:

K1 $\int_0^1 K_n(p,x)d\mu_n(x) = 1$ for each $n \geq 1$.

K2 There is a sequence $\delta = \delta_n \to 0$ such that $\sup_{p\in U(\epsilon)}\left[1 - \int_{p-\delta}^{p+\delta} K_n(p,x)d\mu_n(x)\right] \to 0$.

K3 For any function $g(\cdot)$ that is at least three times differentiable in $(0,1)$, $\int_0^1 g(x)K_n(p,x)d\mu_n(x)$ is differentiable in p on $U(\epsilon)$, and

$$\sup_{p\in U(\epsilon)}\left|g(p) - \int_0^1 g(x)K_n(p,x)d\mu_n(x)\right| = O(n^{-\alpha}), \quad \alpha > 0;$$
$$\sup_{p\in U(\epsilon)}\left|g'(p) - \frac{1}{dp}\int_0^1 g(x)K_n(p,x)d\mu_n(x)\right| = O(n^{-\beta}), \quad \beta > 0.$$

The following theorem in Bouzebda *et al.* (2013) shows the asymptotic consistency of $\hat{H}^{Quantile}_{\epsilon,f}$.

Theorem 2.5. *If the sequence of random variable $X_1, ..., X_n$ with a quantile density function $q(\cdot)$ fulfilling conditions Q1-Q3 and the kernel $K_n(\cdot,\cdot)$ satisfies K1-K3, then we have*

$$|\hat{H}^{Quantile}_{\epsilon,f} - H_f| = O_{\mathbb{P}}\big(n^{-1/2}M(q) + n^{-\beta}\big),$$

where

$$M(q) = M_q\hat{M}(1)\sqrt{\delta\log\delta^{-1}} + M_{q'} + \sqrt{\hat{M}(q^2)R'(1)} + n^{-1/2}A_\gamma(n)M_q\hat{M}(1),$$

$$\hat{M}(g) = \sup_{p\in U(\epsilon)}\int_{i=0}^m |g(x)K(p,x)|d\mu(x); \quad R'(g) = \sup_{p\in U(\epsilon)}\int_{[0,1]\setminus U(\epsilon+\delta)}|g(x)K(p,x)|d\mu(x),$$

$$M_g = \sup_{p\in U(\epsilon)}|g(p)|; \quad A_\gamma(n) = \begin{cases}\log n, & \max\{q(0),q(1)\} < \infty \ \text{or}\ \gamma\leq 2\\ (\log\log n)^\gamma(\log n)^{(1+\epsilon)(\gamma-1)}, & \gamma > 2\end{cases},$$

Also, the asymptotic normality of $\hat{H}^{Quantile}_{\epsilon,f}$ is derived in Bouzebda *et al.* (2013).

Theorem 2.6. *If the conditions Q1-Q4 are satisfied and the kernel $K_n(\cdot,\cdot)$ satisfies K1-K3 with $\alpha > 0.5$ and $\beta > 0.5$, then we have*

$$\sqrt{n}\big(\hat{H}^{Quantile}_{\epsilon,f} - H_{\epsilon,f}\big) - \psi_n(\epsilon) = O_{\mathbb{P}}(\{2\epsilon\log\epsilon^{-1}\}^{1/2}) + o_{\mathbb{P}}(1),$$

where

$$\psi_n(\epsilon) := \int_{U(\epsilon)} (q'(x)/q(x)) B_n(x) dx$$

is a centered Gaussian random variable with variance

$$Var[\psi_n(\epsilon)] = Var[\log(q(F(X)))] + o(\eta(\epsilon) + \Phi(\epsilon)),$$

.

$$\Phi(\epsilon) = \mathbb{E}\Big[\log^2(q(F(X)))\Big] - \Big\{\epsilon \log^2 q(\epsilon) + \epsilon \log^2(q(1-\epsilon)) + \int_{\epsilon}^{1-\epsilon} \log^2 q(p) dp\Big\}.$$

3 New entropy and residual entropy estimators and their asymptotic properties

3.1 Entropy estimators

3.1.1 Entropy estimators based on smooth Poisson histogram density estimation

In this subsection, we will propose entropy estimators based on the idea of (1.5), (1.7) and (1.9). Also, we will discuss briefly about their asymptotic properties. Recall that the entropy of a random variable can be estimated by the direct plugin approach

$$\hat{H}_f^{Plugin} = -\int_0^{\infty} \hat{f}_n(x) \log \hat{f}_n(x) dx,$$

or by the sample mean of $\log f(X)$

$$\hat{H}_f^{Meanlog} = -\sum_{i=1}^{n} \log \hat{f}_n(X_i).$$

We observe that the performance of $\hat{H}_f^{Plugin}$ and $\hat{H}_f^{Meanlog}$ mainly depend on the density estimator, so we would expect to obtain a good entropy estimator if the chosen density estimator is well-performed, consistent, and well-behaved. In literature, regarding to non-negative random variable, there are so many candidates for the density estimators with nice asymptotic properties. Among these, we especially focus on the Poisson smooth histogram density estimator. This density estimation is motivated by Hill's Lemma in Feller (1966), which states that

Lemma 3.1. *If $u(x)$ is a bounded, continuous function and $p_{k,i}$ is a family of lattice distributions, then as $k \to \infty$*

$$\sum_{i \geq 0} u(i/k) p_{k,i}(x) \to u(x)$$

uniformly in any finite interval.

One can obtain a smooth histogram estimator of distribution function $F(\cdot)$ by substituting of the empirical distribution $F_n(\cdot)$ in place of $u(\cdot)$.

$$\tilde{F}_n(x) = \sum_{i\geq 0} F_n(i/k)p_{k,i}(x).$$

Consequently, taking derivative of $\tilde{F}_n(\cdot)$ provides a vaid estimator of density $f(\cdot)$.

$$\tilde{f}_n(x) = \sum_{i\geq 0} F_n(i/k)\frac{d}{dx}(p_{k,i}(x))$$

Among these choice of $p_{k,i}(\cdot)$, we especially focus on the Poisson distribution which was developed by Gawronski and Stadtmüller (1980, 1981) and studied more detail in Chaubey and Sen (1996) and Chaubey *et al.* (2010) which first considered the estimator of survival function

$$\tilde{R}_n(x) = \sum_{i=0}^{\infty} R_n\left(\frac{i}{k}\right)e^{-kx}\frac{(kx)^i}{i!}, \tag{3.1}$$

where $R_n(x) = 1-F_n(x)$ is the empirical survival function and $k = k(n)$ can be viewed as the smoothing parameter. It was shown in Chaubey and Sen (1996) that under the conditions $k \to \infty$, $n^{-1}k \to 0$, and $f(x)$ is absolutely continuous with a bounded derivative $f'(\cdot)$ a.e. on $\mathbb{R}^+$, then

$$||\tilde{R}_n - R_n|| = \sup_{x\in\mathbb{R}^+}\{\tilde{R}_n(x) - R_n(x)\} = O(n^{-3/4}(\log n)^{1+\delta}) \;\; a.s. \;\text{ as } n \to \infty, \tag{3.2}$$

where $\delta > 0$ is arbitrary. The estimator of density is obtained by taking the derivative of $\tilde{R}_n(x)$.

$$\hat{f}_n^{Pois}(x) = -\frac{d}{dx}\tilde{R}_n(x) = k\sum_{i=0}^{\infty}\left[F_n\left(\frac{i+1}{k}\right) - F_n\left(\frac{i}{k}\right)\right]e^{-kx}\frac{(kx)^i}{i!}. \tag{3.3}$$

The estimator $\hat{f}_n^{Pois}(x)$ can be interpreted as a random weighted sum of Poisson mass functions. We observe that the summation in (3.3) is up to a certain value of i because the differences $F_n((i+1)/k) - F_n(i/k)$ will vanish for all $i \geq kX_{(n)} + 1$.

The asymptotic properties of $\hat{f}_n^{Pois}(x)$ have been studied and its weak convergence was proven by Bouezmarni and Scaillet (2005) under the assumption of $\lim_{n\to\infty} k = \infty$ and $\lim_{n\to\infty} nk^{-2} = \infty$, we have

$$\sup_{x\in I}|\hat{f}_n^{Pois}(x) - f(x)| \xrightarrow{\mathbb{P}} 0 \qquad \text{as } n \to \infty$$

uniformly on a compact set I in $\mathbb{R}^+$. Not only that, they also obtained the weak convergence for the case of unbounded pdf f at $x = 0$. Later on, Chaubey *et al.* (2010) filled the gap in the asymptotic theory of $\hat{f}_n^{Pois}$ by proposing the asymptotic bias, asymptotic variance, strong consistency and the asymptotic normality of the estimator. Particularly, under the assumptions that $k = cn^h$ for some constant c and $0 < h < 1$, and $f'(x)$ satisfies the Lipschitz condition of order $\alpha > 0$, i.e. there exists a finite positive K such that

$$|f'(s) - f'(t)| \leq K|s-t|^\alpha \qquad \forall s, t \in \mathbb{R}^+,$$

then asymptotic bias and variance of $\hat{f}^{Pois}(\cdot)$ are given by

$$Bias\big[\hat{f}_n^{Pois}(x)\big] \approx \frac{f'(x)}{2cn^h}, \tag{3.4}$$

$$Var\big[\hat{f}_n^{Pois}(x)\big] \approx \frac{\mathbb{E}[X]}{2}\sqrt{\frac{c}{2\pi x^3}} f(x) n^{h/2-1}. \tag{3.5}$$

For the strong consistency of $\hat{f}_n^{Pois}(x)$, if $f(x)$ and $f'(x)$ is bounded and $k_n = O(n^h)$ for some $h > 0$, then

$$||\hat{f}_n^{Pois} - f|| = \sup_{x \in \mathbb{R}}\{\hat{f}_n^{Pois}(x) - f(x)\} \stackrel{a.s.}{\rightarrow} 0. \tag{3.6}$$

For the asymptotic normality of $\hat{f}_n^{Pois}(x)$, if $k_n = O(n^{2/5})$, and $f'(x)$ satisfies the Lipschitz order α condition, then for x in a compact set $I \subset \mathbb{R}^+$,

$$n^{2/5}(\hat{f}_n^{Pois}(x) - f(x)) - \frac{1}{2\delta^2} f'(x) \stackrel{D}{\rightarrow} \mathcal{G}, \tag{3.7}$$

where $\mathcal{G}$ is the Gaussian process with covariance function $\gamma_x^2 \delta_x s$, $\gamma_x^2 = \frac{\mathbb{E}[X]}{2}(2\pi x^3)^{-1/2} f(x)\delta$, $\delta_x s = 0$ for $x \neq s$, $\delta_x s = 1$ for $x = s$, and $\delta = \lim_{n\to\infty}(n^{-1/5}k_n^{1/2})$.
Similar to the problem of bandwidth selection in kernel density estimators, the choice of the smoothing parameter k strongly affects the performance of the resulting density estimators. In general, small value of k corresponds to over-smoothing, whereas large value of k corresponds to under-smoothing. Here we list some suggestions for the choice of k (see Chaubey and Sen 2009).

- A stochastic choice of k proposed by Chaubey and Sen (1996):

$$k_{(1)} = \frac{n}{\max(X_1, ..., X_n)}. \tag{3.8}$$

  This choice sastifies the condition of $k \to \infty$ if $\mathbb{E}[X] < \infty$, but for the conpact support, this choice will not satisfy $n^{-1}k \to 0$.

- Another choice suggested by Chaubey and Sen (1998):

$$k_{(2)} = \frac{n}{X_{(n-r_n+1)} \log\log r_n}, \tag{3.9}$$

  where $X_{(i)}$ is the i^{th} order statistic of the random sample $(X_1, ..., X_n)$ and $r_n = o(\log\log n)$.

- Deterministic choice of k given bey Gowronski and Stadtmüller (1981):

$$k_{(3)} = \lfloor n^{2/5} \rfloor + 1. \tag{3.10}$$

This choice is due to the strong convergence of the density estimator $\hat{f}_n^{Pois}(x)$.

- Two choices of cross validation methods, one is based on the likelihood and the other is based on mean integrated squared error (MISE).

With these nice asymptotic properties of $\hat{f}_n^{Pois}(\cdot)$, we propose here two entropy estimators of the form

$$\hat{H}_f^{Plugin-Pois} = -\int_0^\infty \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x)dx, \tag{3.11}$$

$$\hat{H}_f^{Meanlog-Pois} = -\sum_{i=1}^{n} \log \hat{f}_n^{Pois}(X_i). \tag{3.12}$$

The following theorem concerns of the asymptotic properties of $\hat{H}_f^{Meanlog-Pois}$.

Theorem 3.1. *Assume the following conditions hold:*

- $\mathbb{E}[(\log f(X))^2] < \infty$,
- $f'(x)$ *is bounded with* $\int_0^\infty f'(x)dx < \infty$, *and satisfies Lipschitz order of* α *condition,*
- $f(x)$ *is twice differentiable and* $\int_0^\infty \frac{f''(x)}{f(x)}dx < \infty$,
- $k_n = o(n^h)$ *for some* $h \in (0, 3/4)$,

then

$$\hat{H}_f^{Meanlog-Pois} = -\frac{1}{n}\sum_{i=1}^{n} \log f(X_i) + o(n^{-1/2}) \ a.s. \text{ as } n \to \infty. \tag{3.13}$$

Consequently, we get

$$\left| \hat{H}_f^{Meanlog-Pois} - H_f \right| \overset{a.s.}{\to} 0, \tag{3.14}$$

and

$$\sqrt{n}\left(\hat{H}_f^{Meanlog-Pois} - H_f \right) \overset{D}{\to} \mathcal{N}(0, Var[\log f(X)]). \tag{3.15}$$

Proof. Writing $\hat{H}_f^{Meanlog-Pois}$ as an integral with respect to the empirical distribution function $F_n(x)$, we have

$$\begin{aligned} -\frac{1}{n}\sum_{i=1}^{n} \log \hat{f}_n^{Pois}(X_i) &= -\int_0^\infty \log \hat{f}_n^{Pois}(x)dF_n(x) \\ &= -\int_0^\infty \log f(x)dF_n(x) - \int_0^\infty \Big(\log \hat{f}_n^{Pois}(x) - \log f(x) \Big) dF_n(x) \\ &= -\frac{1}{n}\sum_{i=1}^{n} \log f(X_i) - I_n, \end{aligned}$$

where

$$I_n := \int_0^\infty \bigg(\log \hat{f}_n^{Pois}(x) - \log f(x) \bigg) dF_n(x).$$

On the other hand, we know that $-\frac{1}{n}\sum_{i=1}^n \log f(X_i)$ is an unbiased, strongly consistent estimator for H_f by the strong law of large numbers, and is root-n asymptotic normal by the *central limit theorem* if $\mathbb{E}[\log f(X)^2] < \infty$. That is

$$\left| -\frac{1}{n}\sum_{i=1}^n \log f(X_i) - H_f \right| \stackrel{a.s.}{\to} 0,$$

and

$$\sqrt{n}\bigg(-\frac{1}{n}\sum_{i=1}^n \log f(X_i) - H_f \bigg) \stackrel{D}{\to} \mathcal{N}(0, \mathrm{Var}[\log f(X)]),$$

as $n \to \infty$. Therefore, it is sufficient to prove that

$$I_n = o(n^{-1/2}) \text{ a.s. as } n \to \infty.$$

In order to study the asymptotic behavior of I_n, we decompose it into two parts as,

$$\begin{aligned} I_n &= \int_0^\infty \bigg(\log \hat{f}_n^{Pois}(x) - \log f(x) \bigg) d(F_n(x) - F(x)) + \int_0^\infty \bigg(\log \hat{f}_n^{Pois}(x) - \log f(x) \bigg) dF(x) \\ &= I_{n,1} + I_{n,2}, \text{ say.} \end{aligned}$$

- Analysis of $I_{n,2}$:

  Since the function $\log z$ is continuous and differentiable for all $z > 0$, we can apply the Taylor expansion centering at a to get

  $$\log z = \log a + \frac{z-a}{tz + (1-t)a},$$

  where $t \in (0,1)$. By letting $z = \hat{f}_n^{Pois}(x)$ and $a = f(x)$, we obtain

  $$\log \hat{f}_n^{Pois}(x) - \log f(x) = \frac{\hat{f}_n^{Pois}(x) - f(x)}{t\hat{f}_n^{Pois}(x) + (1-t)f(x)}.$$

  As $||\hat{f}_n^{Pois}(x) - f(x)|| \stackrel{a.s.}{\to} 0$ uniformly, it implies $t\hat{f}_n^{Pois}(x) + (1-t)f(x) \stackrel{a.s.}{\to} f(x)$. Thus,

  $$\log \hat{f}_n^{Pois}(x) - \log f(x) \stackrel{a.s.}{\to} \frac{\hat{f}_n^{Pois}(x) - f(x)}{f(x)}.$$

As a result, $I_{n,2}$ can be expressed as

$$\begin{aligned}
I_{n,2} &\stackrel{a.s.}{\to} \int_0^\infty \left(\frac{\hat{f}_n^{Pois}(x) - f(x)}{f(x)}\right) f(x)dx \\
&= \int_0^\infty \left(\hat{f}_n^{Pois}(x) - f(x)\right)dx \\
&= \int_0^\infty \hat{f}_n^{Pois}(x)dx - \int_0^\infty f(x)dx \\
&= 0,
\end{aligned}$$

that follows, since the Poisson smooth density estimator integrates to unity.

- Analysis of $I_{n,1}$:

Using integration by part we have

$$\begin{aligned}
I_{n,1} &= \big(F_n(x) - F(x)\big) \log\left(\frac{\hat{f}_n^{Pois}(x)}{f(x)}\right)\Bigg|_0^\infty \\
&\quad - \int_0^\infty \left(\frac{\hat{f}_n^{\prime Pois}(x)}{\hat{f}_n^{Pois}(x)} - \frac{f'(x)}{f(x)}\right)\big(F_n(x) - F(x)\big)dx.
\end{aligned}$$

It is well-known that by the law of the iterated logarithm, we have

$$||F_n - F||_\infty = O\big(n^{-1/2}(\log\log n)^{1/2}\big) \text{ a.s.}$$

Meanwhile, Recall from Chaubey *et al.* (2010) that if $f'(x)$ satisfies the Lipschitz of order $\alpha > 0$, i.e. there exists a finite positive K such that

$$|f'(s) - f'(t)| \leq K|s - t|^\alpha \qquad \forall s, t \in \mathbb{R}^+,$$

then for fixed $x \in \mathbb{R}$ we have

$$\hat{f}_n^{Pois}(x) - f(x) = \frac{1}{2k} f'(x) + O(k^{-1-\alpha}), \tag{3.16}$$

which implies that

$$\frac{\hat{f}_n^{Pois}(x)}{f(x)} = 1 + \frac{1}{2k}\frac{f'(x)}{f(x)} + O(k^{-k-\alpha}) = 1 + O(k^{-1}). \tag{3.17}$$

If we restrict $k = o(n^h)$ for $0 < h < 3/4$, then for fixed $x \in \mathbb{R}$ we get

$$\sqrt{n}\Big(F_n(x) - F(x)\Big) \log\left(\frac{\hat{f}_n^{Pois}(x)}{f(x)}\right) = O\big((\log\log n)^{1/2}\big) \log(1 + o(n^{-h})).$$

By L'Hôpital's rule, we get

$$\begin{aligned}
\lim_{n\to 0}(\log\log n)^{1/2}\log(1+n^{-h}) &= \lim_{n\to 0} 2h\frac{\big(\log\log n\big)^{3/2}\log n}{n^h} \\
&= \lim_{n\to 0}\frac{2\big(\log\log n\big)^{3/2}+3\big(\log\log n\big)^{1/2}}{n^h} \\
&= \lim_{n\to 0}\frac{3\big(\log\log n\big)^{1/2}+\frac{3}{2}\big(\log\log n\big)^{-1/2}}{hn^h\log n} \\
&= 0.
\end{aligned}$$

Thus, we obtain

$$\sqrt{n}\left[\big(F_n(x)-F(x)\big)\log\left(\frac{\hat{f}_n^{Pois}(x)}{f(x)}\right)\Bigg|_0^\infty\right] \overset{a.s.}{=} o(1),$$

which means that

$$\left[\big(F_n(x)-F(x)\big)\log\left(\frac{\hat{f}_n^{Pois}(x)}{f(x)}\right)\Bigg|_0^\infty\right] \overset{a.s.}{=} o(n^{-1/2}).$$

On the other hand, since $||\hat{f}_n^{Pois}(x)-f(x)|| \overset{a.s.}{\to} 0$ uniformly, we can bound the second term of $I_{n,1}$ as

$$\begin{aligned}
&\left|\int_0^\infty\left(\frac{\hat{f}_n'^{Pois}(x)}{\hat{f}_n^{Pois}(x)}-\frac{f'(x)}{f(x)}\right)\big(F_n(x)-F(x)\big)dx\right| \\
\overset{a.s.}{\to}\ &\left|\int_0^\infty\left(\frac{\hat{f}_n'^{Pois}(x)}{f(x)}-\frac{f'(x)}{f(x)}\right)\big(F_n(x)-F(x)\big)dx\right| \\
\leq\ & ||F_n-F||_\infty\int_0^\infty\frac{\left|\hat{f}_n'^{Pois}(x)-f'(x)\right|}{f(x)}dx.
\end{aligned}$$

By differentiating on both sides of (3.16) and dividing by $f(x)$, we get

$$\frac{\hat{f}_n'^{Pois}(x)-f'(x)}{f(x)} = \frac{1}{2k}\frac{f''(x)}{f(x)}+O(k^{-1-\alpha}).$$

So, if $k=o(n^h)$ where $1/2<h<1$ and $\int_0^\infty\frac{f''(x)}{f(x)}dx<\infty$, then

$$\begin{aligned}
\left|\int_0^\infty\left(\frac{\hat{f}_n'^{Pois}(x)}{\hat{f}_n^{Pois}(x)}-\frac{f'(x)}{f(x)}\right)\big(F_n(x)-F(x)\big)dx\right| &\leq O\big(n^{-1/2}(\ln\ln n)^{1/2}\big)o(n^{-1/2} \\
&= o(n^{-1/2}).
\end{aligned}$$

Therefore, $I_{n,1} = o(n^{-1/2})$ a.s. Putting everything together, we get

$$\hat{H}_f^{Meanlog-Pois} = -\frac{1}{n}\sum_{i=1}^{n} \log f(X_i) + o(n^{-1/2}) \text{ almost surely.}$$

This completes the proof of the theorem. □

Next we establish asymptotic properties of the plug-in estimator $\hat{H}_f^{Plugin-Pois}$.

Theorem 3.2. *Assume the following conditions hold:*

- $\mathbb{E}[(\log f(X))^2] < \infty$,
- $f(x) > 0$ *for all* $x \in (0, \infty)$,
- $f'(x)$ *is bounded with* $\int_0^\infty f'(x)dx < \infty$*, and satisfies Lipschitz order of* α *condition,*
- $\int_0^\infty \frac{f''(x)}{f(x)}dx < \infty$,
- $k = O(n^h)$ *for* $1/2 < h < 3/4$,

then

$$\hat{H}_f^{Plugin-Pois} = -\frac{1}{n}\sum_{i=1}^{n} \log f(X_i) + o(n^{-1/2}) \ a.s. \text{ as } n \to \infty. \tag{3.18}$$

Consequently, we get

$$\left|\hat{H}_f^{Plugin-Pois} - H_f\right| \xrightarrow{a.s.} 0, \tag{3.19}$$

and

$$\sqrt{n}\left(\hat{H}_f^{Plugin-Pois} - H_f\right) \xrightarrow{D} \mathcal{N}(0, Var[\log f(X)]). \tag{3.20}$$

Proof. We have

$$\begin{aligned}
\hat{H}_f^{Plugin-Pois} &= -\int_0^\infty \left(\hat{f}_n^{Pois}(x) - f(x) + f(x)\right)\left(\log \hat{f}_n^{Pois} - \log f(x) + \log f(x)\right)dx \\
&= -\int_0^\infty \left(\hat{f}_n^{Pois}(x) - f(x)\right)\left(\log \hat{f}_n^{Pois} - \log f(x)\right)dx \\
&\quad - \int_0^\infty \left(\hat{f}_n^{Pois}(x) - f(x)\right)\log f(x)dx \\
&\quad - \int_0^\infty f(x)\left(\log \hat{f}_n^{Pois} - \log f(x)\right)dx - \int_0^\infty f(x)\log f(x)dx \\
&= U_{n,1} - U_{n,2} - I_{n,2} + H_f,
\end{aligned}$$

where

$$U_{n,1} := -\int_0^\infty \Big(\hat{f}_n^{Pois}(x) - f(x)\Big)\Big(\log \hat{f}_n^{Pois} - \log f(x)\Big)dx, \tag{3.21}$$

$$U_{n,2} := -\int_0^\infty \Big(\hat{f}_n^{Pois}(x) - f(x)\Big)\log f(x)dx, \tag{3.22}$$

$$I_{n,2} := -\int_0^\infty f(x)\Big(\log \hat{f}_n^{Pois} - \log f(x)\Big)dx. \tag{3.23}$$

From the proof of the asymptotic properties of the meanlog entropy estimator $\hat{H}_f^{Meanlog-Pois}$, we already have

$$I_{n,2} \overset{a.s.}{\to} 0.$$

- Analysis of $U_{n,1}$:

  Since under the assumptions in the theorem, $||\hat{f}_n^{Pois}(x) - f(x)|| \overset{a.s.}{\to} 0$ uniformly, we get $\hat{f}_n^{Pois}(x)/f(x) \overset{a.s.}{\to} 1$ uniformly. Thus, by the dominant convergence theorem (DCT), $U_{n,1}$ becomes

$$\begin{aligned} U_{n,1} &= -\int_0^\infty \big(\hat{f}_n^{Pois}(x) - f(x)\big)\log\left(\frac{\hat{f}_n^{Pois}(x)}{f(x)}\right)dx \\ &\overset{a.s.}{\to} 0. \end{aligned}$$

- Analysis of $U_{n,2}$:

  We can decompose $U_{n,2}$ as following

$$\begin{aligned} U_{n,2} &= -\int_0^\infty \log f(x)d\big(\tilde{F}_n(x) - F(x)\big) \\ &= -\int_0^\infty \log f(x)d\big(R(x) - \tilde{R}_n(x)\big) \\ &= -\int_0^\infty \log f(x)d\big(R(x) - R_n(x)\big) - \int_0^\infty \log f(x)d\big(R_n(x) - \tilde{R}_n(x)\big) \\ &= -\int_0^\infty \log f(x)d\big(F_n(x) - F(x)\big) - \int_0^\infty \log f(x)d\big(R_n(x) - \tilde{R}_n(x)\big) \\ &= -\int_0^\infty \log f(x)dF_n(x) + \int_0^\infty \log f(x)dF(x) - \int_0^\infty \log f(x)d\big(R_n(x) - \tilde{R}_n(x)\big) \\ &= -\frac{1}{n}\sum_{i=1}^n \log f(X_i) - H_f - \int_0^\infty \log f(x)d\big(R_n(x) - \tilde{R}_n(x)\big) \end{aligned}$$

  Recall that under the conditions $k \to \infty$, $n^{-1}k \to 0$, and $f(x)$ is absolutely continuous with a bounded derivative $f'(\cdot)$ a.e. on $\mathbb{R}^+$, Chaubey and Sen (1996 see their Theorem 3.2) showed that

$$||\tilde{R}_n - R_n|| = \sup_{x \in \mathbb{R}^+} \{\tilde{R}_n(x) - R_n(x)\} = O(n^{-3/4}(\log n)^{1+\delta}) \ \ a.s. \ \text{ as } n \to \infty, \quad (3.24)$$

where $\delta > 0$ is arbitrary. Therefore, for the second term in the last equation above, under the assumptions in the theorem, we can use integration by part to obtain

$$\begin{aligned}
&\left|\int_0^\infty \log f(x) d\big(R_n(x) - \tilde{R}_n(x)\big)\right| \\
&= \left|\Big[\log f(x)\big(R_n(x) - \tilde{R}_n(x)\big)\Big]_0^\infty - \int_0^\infty \frac{f'(x)}{f(x)}\big(R_n(x) - \tilde{R}_n(x)\big)dx\right| \\
&\leq \lim_{a \to \infty}\left|\log f(a)\big(R_n(a) - \tilde{R}_n(a)\big)\right| + \lim_{b \to \infty}\left|\log f(b)\big(R_n(b) - \tilde{R}_n(b)\big)\right| \\
&\qquad\qquad + \left|\sup_{x \in \mathbb{R}^+}\{\tilde{R}_n(x) - R_n(x)\}\int_0^\infty \frac{f'(x)}{f(x)}dx\right| \\
&\stackrel{a.s.}{\to} 0,
\end{aligned}$$

that follows from Equation (3.24). Therefore, putting back everything, we obtain

$$\hat{H}_f^{Plugin-Pois} = -\frac{1}{n}\sum_{i=1}^n \log f(X_i) + o(n^{-1/2}) \ a.s. \text{ as } n \to \infty,$$

which completes the proof of Theorem 3.2. □

Remark 2.1. The asymptotic results established here using the Poisson weight smoothing estimator of the density function follows very closely to those established in Hall and Morton (1993), though under some stringent smoothness conditions. Such results may also be established using alternative asymmetric kernel density estimators such as those proposed in Chaubey *et al.* (2012), Chen (2000), Cheng and Parzen (1997) and others. However, Poisson weights based entropy estimator, especially $\hat{H}_f^{Meanlog-Pois}$ may be computationally preferable over others. This is important as there may not exist an uniformly best estimator, as demonstrated through numerical studies in the next section.

3.1.2 Entropy estimator based on quantile density estimator

As mentioned in the introduction that the entropy can be expressed in terms of quantile density function as the following.

$$\hat{H}_f^{Quantile} = \int_0^1 \hat{q}_n(p)dp.$$

Consequently, we consider a particular estimator of quantile density using an approach called Bernstein polynomial approximation. First, we start with the Bernstein polynomial estimator of a smooth quantile function proposed by Cheng (1995).

$$\tilde{Q}_n(p) = \sum_{i=0}^{m} \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p) \qquad p \in [0,1],$$

where $\hat{Q}_n(\cdot)$ is the empirical quantile function, $b(i,m,p) = \mathbb{P}[Y = i]$ where Y follows the binomial(m,p), and m is a function of n such that $m \to \infty$ as $n \to \infty$. Then the estimator of $q(\cdot)$ can be obtained by differentiating $\tilde{Q}_n(\cdot)$.

$$\begin{aligned}
\tilde{q}_n(p) &= \frac{d\tilde{Q}_n(p)}{dp} \\
&= \frac{d}{dp}\sum_{i=0}^{m} \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p) \\
&= \sum_{i=0}^{m} \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p)\left[\frac{i-mp}{p(1-p)}\right]. \qquad (3.25)
\end{aligned}$$

Therefore, the entropy can be estimated by

$$\hat{H}_f^{Quantile} = \int_0^1 \tilde{q}_n(p)dp. \qquad (3.26)$$

The advantage of using this quantile density estimator $\tilde{q}_n(\cdot)$ not only results in a simple and efficient estimator in terms of computation, but also resolves the boundary problem of the general case of $\tilde{q}_n^{CP}(\cdot)$. It is easy to see that this estimator $\tilde{q}_n(\cdot)$ is free of the boundary problem. For instance,

$$\begin{aligned}
\tilde{q}_n(0) &= \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p)\left[\frac{i-mp}{p(1-p)}\right]\mathbf{1}[i=0] + \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p)\left[\frac{i-mp}{p(1-p)}\right]\mathbf{1}[i=1] \\
&= m\big(\hat{Q}_n(1/m) - \hat{Q}_n(0)\big) \\
&\to \left.\frac{d\hat{Q}_n(p)}{dp}\right|_{p=0} = \hat{q}_n(0) \overset{a.s.}{\to} q(0) \qquad \text{as } m \to \infty.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\tilde{q}_n(1) &= \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p)\left[\frac{i-mp}{p(1-p)}\right]\mathbf{1}[i=m-1] + \hat{Q}_n\left(\frac{i}{m}\right) b(i,m,p)\left[\frac{i-mp}{p(1-p)}\right]\mathbf{1}[i=m] \\
&= m\big(\hat{Q}_n(1) - \hat{Q}_n((m-1)/m)\big) \\
&\to \left.\frac{d\hat{Q}_n(p)}{dp}\right|_{p=1} = \hat{q}_n(1) \overset{a.s.}{\to} q(1) \qquad \text{as } m \to \infty.
\end{aligned}$$

Furthermore, recall that the kernel-type quantile density estimator $\hat{q}_n^{CP}(\cdot)$, proposed by Cheng and Parzen (1997), has the form

$$\hat{q}_n^{CP}(p) = \frac{d}{dp}\int_0^1 \hat{Q}_n(x)K_n(p,x)d\mu_n(x),$$

where $\hat{Q}_n(\cdot)$ is the empirical quantile function, $K_n(p,x)$ is the sequence density kernel functions defined on $(0,1)\times[0,1]$, and $\mu_n(x)$ is a sequence of σ-finite measure on $[0,1]$. Then, we see that $\tilde{q}_n(\cdot)$ is the special case of the kernel-type quantile density estimator $\hat{q}_n^{CP}(\cdot)$ in the way that

$$\begin{aligned} K_m(p,x) &= \frac{\Gamma(m+1)}{\Gamma(mx+1)\Gamma(m-mx+1)}p^{mx}(1-p)^{m-mx}, \\ d\mu_m(x) &= \begin{cases} 1 & x=\frac{i}{m} \quad i=1,...,m, \\ 0 & \text{otherwise}, \end{cases} \end{aligned}$$

where $m\to\infty$ as $n\to\infty$. As a result, $\hat{H}_f^{Quantile}$ inherits all the asymptotic properties from $\hat{H}_{\epsilon,f}^{Quantile}$ introduced in the Section 2.3. That is, under certain conditions on the quantile density functions,

$$\begin{aligned} |\hat{H}_f^{Quantile} - H_f| &\xrightarrow{\mathbb{P}} 0, \\ \sqrt{n}\big(\hat{H}_f^{Quantile} - H_f\big) &\xrightarrow{D} \mathcal{N}\big(0, \log(q(F(X)))\big). \end{aligned}$$

3.2 Residual entropy estimator

In this sub-section, we consider the residual entropy estimators, and we propose here three candidates along with their asymptotic properties.
Motivated by the well-behavior of $\hat{f}_n^{Pois}(\cdot)$ and $\tilde{q}_n(\cdot)$, we suggest a direct plugin residual entropy estimator. That is, our proposed residual entropy estimators are of the form

$$\hat{H}_f^{Plugin-Pois}(t) = \log(\hat{R}^{Pois}(t)) - \frac{1}{\hat{R}^{Pois}(t)}\int_t^\infty \hat{f}_n^{Pois}(x)\log\hat{f}_n^{Pois}(x)dx, \qquad (3.27)$$

$$\hat{H}_f^{Quantile}(t) = \log(\hat{R}^{Pois}(t)) + \frac{1}{\hat{R}^{Pois}(t)}\int_0^1 \log\tilde{q}_n(p)dp, \qquad (3.28)$$

where $\hat{R}^{Pois}(t) = \int_t^\infty \hat{f}_n^{Pois}(x)dx$. The asymptotic consistency of $\hat{H}_f^{Plugin-Pois}(t)$ is stated in the following theorem.

Theorem 3.3. *Assume the following conditions:*

- *$f'(x)$ is bounded, satisfies Lipschitz order of α condition,*
- *$\int_0^\infty f'(x)\log f(x)dx < \infty$,*

- $k_n = o(n^h)$ *for some* $0 < h < 3/4$,

then,

$$\left|\hat{H}_f^{Plugin-Pois}(t) - H_f(t)\right| \overset{a.s.}{\to} 0. \tag{3.29}$$

Proof. We have

$$\begin{aligned}
& \left|\hat{H}_f^{Plugin-Pois}(t) - H_f(t)\right| \\
\leq \; & \left|\log \hat{R}_n(t) - \log R(t)\right| + \left|\frac{1}{\hat{R}_n(t)}\int_t^\infty \hat{f}_n(x)\log \hat{f}_n(x)dx - \frac{1}{R(t)}\int_t^\infty f(x)\log f(x)dx\right| \\
= \; & T_n^{(1)}(t) + T_n^{(2)}(t).
\end{aligned}$$

- <u>Analysis of $T_n^{(1)}(t)$</u>

  We first observe that

  $$|\hat{R}^{Pois}(t) - R(t)| \leq \int_t^\infty \left|\hat{f}_n^{Pois}(x) - f(x)\right| dx.$$

  Since $||\hat{f}_n^{Pois}(x) - f(x)|| \overset{a.s.}{\to} 0$ uniformly, by DCT we obtain

  $$|\hat{R}^{Pois}(t) - R(t)| \overset{a.s.}{\to} 0.$$

  Then, for any $\alpha \in (0,1)$ by Taylor expansion, we have

  $$\left|\log \hat{R}^{Pois}(t) - \log R(t)\right| = \left|\frac{\hat{R}^{Pois}(t) - R(t)}{\alpha\hat{R}^{Pois}(t) + (1-\alpha)R(t)}\right|.$$

  Since $\alpha\hat{R}^{Pois}(t) + (1-\alpha)R(t) \overset{a.s.}{\to} R(t)$, we get $T_n^{(1)}(t) \overset{a.s.}{\to} 0$.

- <u>Analysis of $T_n^{(2)}(t)$</u>

  Since $|\hat{R}(t) - R(t)| \overset{a.s.}{\to} 0$, we deduce that for sufficiently large n, $T_n^{(2)}(t)$ has the same limit as

  $$\begin{aligned}
  T_n^{(2)}(t) &= \left|\frac{1}{R(t)}\int_t^\infty \left[\hat{f}_n^{Pois}(x)\log \hat{f}_n^{Pois}(x) - f(x)\log f(x)\right]dx\right| \\
  &\leq \frac{1}{R(t)}\int_t^\infty \left|\hat{f}_n^{Pois}(x)\log \hat{f}_n^{Pois}(x) - f(x)\log f(x)\right| dx.
  \end{aligned}$$

We have

$$
\begin{aligned}
&\int_t^\infty \left| \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x) - f(x)\log f(x) \right| dx \\
&\le \int_t^\infty \left| \big(\hat{f}_n^{Pois}(x) - f(x)\big)\big(\log \hat{f}_n^{Pois}(x) - \log f(x)\big) \right| dx \\
&\quad + \int_t^\infty \left| \big(\hat{f}_n^{Pois}(x) - f(x)\big) \log f(x) \right| dx + \int_t^\infty \left| f(x)\big(\log \hat{f}_n^{Pois}(x) - \log f(x)\big) \right| dx \\
&= T_{n,1}^{(2)}(t) + T_{n,2}^{(2)}(t) + T_{n,3}^{(2)}(t).
\end{aligned}
$$

- For $T_{n,1}^{(2)}(t)$, Since under the assumptions in the theorem, $||\hat{f}_n^{Pois}(x) - f(x)|| \overset{a.s.}{\to} 0$ uniformly, we get $\hat{f}_n^{Pois}(x)/f(x) \overset{a.s.}{\to} 1$ uniformly. Thus, by the dominant convergence theorem (DCT), $T_{n,1}^{(2)}(t)$ becomes

$$
\begin{aligned}
& \int_t^\infty \left| \big(\hat{f}_n^{Pois}(x) - f(x)\big)\big(\log \hat{f}_n^{Pois}(x) - \log f(x)\big) \right| dx \\
= & \int_t^\infty \left| \big(\hat{f}_n^{Pois}(x) - f(x)\big) \log \left(\frac{\hat{f}_n^{Pois}(x)}{f(x)} \right) \right| dx \\
\overset{a.s.}{\to} & \ 0.
\end{aligned}
$$

- For $T_{n,2}^{(2)}(t)$, recall that if $f'(x)$ satisfies the Lipschitz of order $\alpha > 0$, then for fixed $x \in \mathbb{R}$ we have

$$
\hat{f}_n^{Pois}(x) - f(x) = \frac{1}{2k} f'(x) + O(k^{-1-\alpha}),
$$

  Thus, we get

$$
\int_t^\infty \left| \big(\hat{f}_n^{Pois}(x) - f(x)\big) \log f(x) \right| dx = \frac{1}{2k} \int_0^\infty |f'(x) \log f(x)|\, dx + O(k^{-1-\alpha}) = o(1),
$$

  given that $k = o(n^h)$ and $\int_0^\infty f'(x) \log f(x) dx < \infty$. Therefore $T_{n,2}^{(2)}(t) \overset{a.s.}{\to} 0$.

- For $T_{n,3}^{(2)}(t)$, again, since $||\hat{f}_n^{Pois}(x) - f(x)|| \overset{a.s.}{\to} 0$ uniformly and by Taylor expansion and DCT, for $t \in (0,1)$ we get

$$
\begin{aligned}
\int_t^\infty \left| f(x)\big(\log \hat{f}_n^{Pois}(x) - \log f(x)\big) \right| dx &= \int_t^\infty \left| \left(\frac{\hat{f}_n^{Pois}(x) - f(x)}{t\hat{f}_n^{Pois}(x) - (1-t) f(x)} \right) f(x) \right| dx \\
&\overset{a.s.}{\to} \int_t^\infty \left| \left(\frac{\hat{f}_n^{Pois}(x) - f(x)}{f(x)} \right) f(x) \right| dx \\
&= \int_t^\infty \left| \hat{f}_n^{Pois}(x) - f(x) \right| dx \\
&\overset{a.s.}{\to} 0.
\end{aligned}
$$

Consequently, $T_n^{(2)}(t) \stackrel{a.s.}{\to} 0$. This completes the proof of the theorem.

□

4 Simulation study

4.1 Simulation study on entropy estimators

To study the performance of entropy and residual entropy estimators, we run simulations for a wide range of densities which consists fifteen non-negative densities below. They are categorized into three groups: monotone density, uni-modal density, and bimodal density.

1. Monotone density
   - Standard Exponential(1) with true $H_f = 1$.
   - Exponential(10) with true $H_f = -\log 10 - 1 \approx -1.3026$.
   - Pareto(2,1) with true $H_f \approx 0.8069$.
   - Log-Normal(0,2) with true $H_f \approx 2.1121$.
   - Weibull(0.5,0.5) with true $H_f \approx 0.4228$.
2. Uni-modal density
   - Gamma(2,2) with true $H_f \approx 0.8841$.
   - Gamma(7.5,1) with true $H_f \approx 2.3804$.
   - Log-Normal(0,0.5) with true $H_f \approx 0.7258$.
   - Maxwell(1) with true $H_f \approx 0.9962$.
   - Maxwell(20) with true $H_f \approx -0.5017$.
   - Weibull(2,2) with true $H_f \approx 1.2886$.
3. Bimodal density
   - Mix Gamma: (1/2)Gamma(0.5,0.5)+(1/2)Gamma(2,2) with true $H_f \approx 2.2757$.
   - Mix Lognorm: (1/2)Lognorm(0,0.5)+(1/2)Lognorm(0,2) with true $H_f \approx 1.6724$.
   - Mix Maxwell: (1/2)Maxwell(1)+(1/2)Maxwell(20) with true $H_f \approx 0.8014$.
   - Mix Weibull: (1/2)Weibull(0.5,0.5)+(1/2)Weibull(2,2) with true $H_f \approx 1.1330$.

The graphs of densities for each group along with the corresponding function $f(x) \log f(x)$ are given in the Figure 7, 8, 9, and 10. The densities in the monotone group are chosen with different rates of decay, from the slowest standard Exp(1) to the fastest Exp(10). The Pareto distribution is included here to show the effect of different support $[1, \infty)$ on entropy estimation. On the contrary to the first group, the second group consists of well-behaved densities from highly concentrated Maxwell(20) to widely spread Gamma(7.5,1). Lastly, the simulations are extended to the bimodal densities where each density is a mixture of the same family density but with different parameters. Note that for some densities with high rate of decay and small variance like Exp(10) or Maxwell(20), the integrand $f(x)\log f(x)$ produces non-finite value for large value of x in R software due to the round-up. However, since the contribution of the right tail of the integrand to the entropy is insignificant for sufficiently large x, we obtain the approximately true value of entropy by cutting of the negligible right tail of the integration. That is, instead of integrating over the entire support of $f(x)\log f(x)$, we integrate up to a certain value at which the right tail is negligible.

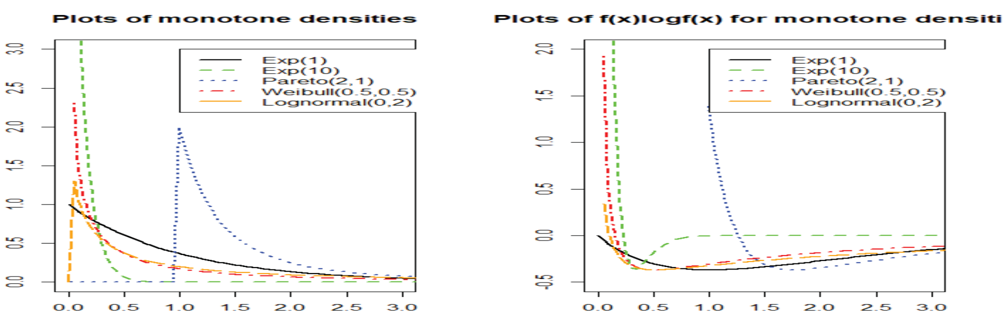

Figure 7: Plots of density in the first group along with their $f(x)\log f(x)$ plots. The exp(1)-black solid line, exp(10)-green dashed line, Pareto(2,1)-blue dotted line, Weibull(0.5,0.5)-red dotted dashed line, and logNormral(0,2)-orange long dashed line.

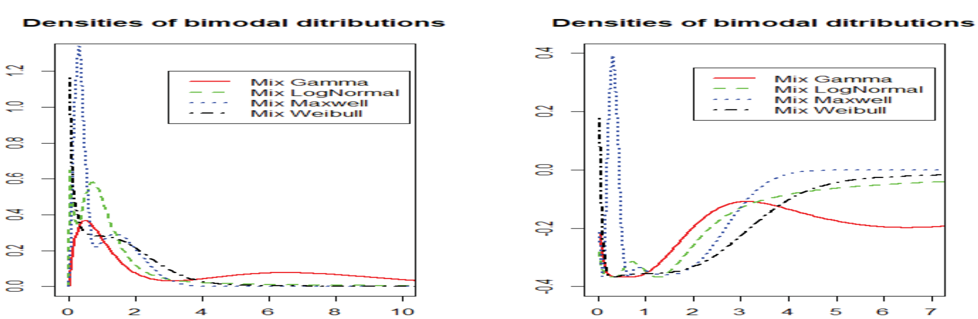

Figure 8: Plots of density in the third group along with their $f(x)\log f(x)$. The mixture Gamma-red solid line, mixture logNormal-green dashed line, mixture Maxwell-blue dotted line, and mixture Weibull-black dotted dashed line.

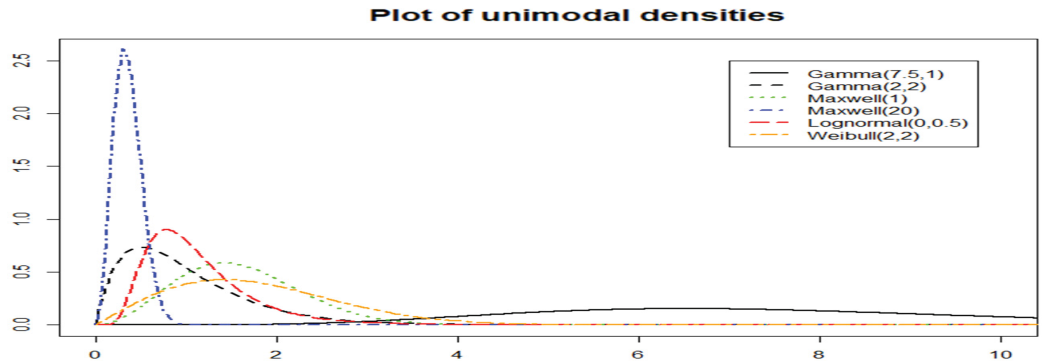

Figure 9: Plots of density in the second group. Gamma(7.5,1)-black solid line, Gamma(2,2)-black dashed line, Maxwell(1)-green dotted line, Maxwell(20) blue dotted dashed line, logNormal(0,0.5)-red long dashed line, and Weibull(2,2)-orange mixed dashed line.

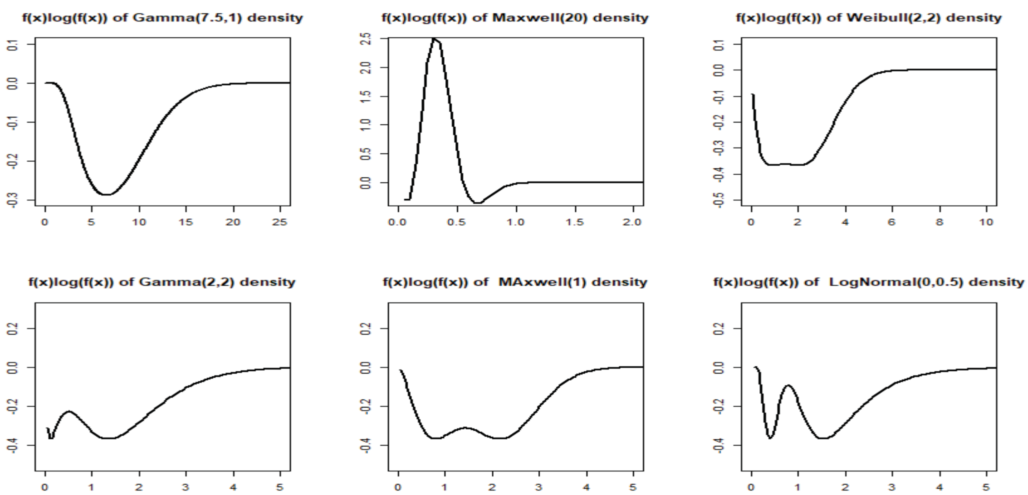

Figure 10: Plots of $f(x)\log f(x)$ for density in the second group.

We run simulation experiments on different estimators which are classified into two groups, namely, 'spacing estimators' and 'non-spacing estimators'.

Spacing estimators

1. Vasicek's estimator (Vasicek, 1976)

$$\hat{H}_1 = \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{n}{2m}(X_{(i+m)} - X_{(i-m)})\right\}. \tag{4.1}$$

2. Ebrahimi's estimator (Ebrahimi *et al.*, 1992)

$$\hat{H}_2 = \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{n}{c_i m}(X_{(i+m)} - X_{(i-m)})\right\}, \tag{4.2}$$

where

$$c_i = \begin{cases} 1 + \frac{i-1}{m} & 1 \leq i \leq m, \\ 2 & m+1 \leq i \leq n-m, \\ 1 + \frac{n-i}{m} & n-m+1 \leq i \leq n. \end{cases}$$

3. Van Es estimator (Van Es, 1992)

$$\hat{H}_3 = -\frac{1}{n-m} \sum_{i=1}^{n-m} \log \left\{ \frac{n+1}{m} (X_{(i+m)} - X_{(i)}) \right\} + \sum_{k=m}^{n} \frac{1}{k} + \log \left(\frac{m}{n+1} \right). \tag{4.3}$$

4. Correa's estimator (Correa, 1995)

$$\hat{H}_4 = -\frac{1}{n} \sum_{i=1}^{n-m} \log \left\{ \frac{\sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})(j-i)}{n \sum_{j=i-m}^{i+m} (X_{(j)} - \bar{X}_{(i)})} \right\}, \tag{4.4}$$

where $\bar{X}_{(i)} := \frac{1}{2m+1} \sum_{j=i-m}^{i+m} X_{(j)}$.

5. WG's estimator (Wieczorkowski and Grzegorzewski, 1999)

$$\hat{H}_5 = \hat{H}_1 + \log \left(\frac{2m}{n} \right) - \frac{n-2m}{n} \Psi(2m) + \Psi(n+1) - \frac{2}{n} \sum_{i=1}^{m} \Psi(i+m-1). \tag{4.5}$$

where $\Psi(k) = \sum_{i=1}^{k-1} \frac{1}{i} - \gamma$ is the di-Gamma function defined on integer set, and $\gamma = 0.57721566...$ is the Euler's constant.

6. Noughabi's estimator (Noughabi, 2010)

$$\hat{H}_6 = -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{\hat{f}_n^{Fixed}(X_{(i+m)}) + \hat{f}_n^{Fixed}(X_{(i-m)})}{2} \right\}. \tag{4.6}$$

where the bandwidth in $\hat{f}_n^{Fixed}(x)$ is fixed to $b = 1.06 s n^{-1/5}$ and s is the sample standard deviation.

7. Gama-spacing estimator:
Motivated by Noughabi's estimator, we also want to test the performance of an entropy estimator obtained by replacing $\hat{f}_n^{Fixed}$ by $\hat{f}_n^{Gam}$

$$\hat{H}_6 = -\frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{\hat{f}_n^{Fixed}(X_{(i+m)}) + \hat{f}_n^{Fixed}(X_{(i-m)})}{2} \right\}. \tag{4.7}$$

Non-spacing estimators

To see how our entropy estimators perform, we also run simulation on other direct plugin and mean of log entropy estimators that use the fixed symmetric kernel density estimator $\hat{f}_n^{Fixed}$

and the log transformed kernel density estimator $\hat{f}_n^{Logtrans}$. Thus, the non-spacing estimators consist of

$$\text{Plugin-Fixed} \quad : \quad \hat{H}_8 = -\int \hat{f}_n^{Fixed}(x) \log \hat{f}_n^{Fixed}(x) dx. \tag{4.8}$$

$$\text{Meanlog-Fixed} \quad : \quad \hat{H}_9 = -\frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_n^{Fixed}(X_i). \tag{4.9}$$

$$\text{Plugin-Gam} \quad : \quad \hat{H}_{10} = -\int \hat{f}_n^{Gam}(x) \log \hat{f}_n^{Gam}(x) dx. \tag{4.10}$$

$$\text{Meanlog-Gam} \quad : \quad \hat{H}_{11} = -\frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_n^{Gam}(X_i). \tag{4.11}$$

$$\text{Plugin-Logtrans} \quad : \quad \hat{H}_{12} = -\int \hat{f}_n^{LogTrans}(x) \log \hat{f}_n^{LogTrans}(x) dx. \tag{4.12}$$

$$\text{Meanlog-Logtrans} \quad : \quad \hat{H}_{13} = -\frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_n^{LogTrans}(X_i). \tag{4.13}$$

$$\text{Plugin-Pois} \quad : \quad \hat{H}_{14} = -\int \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x) dx. \tag{4.14}$$

$$\text{Meanlog-Pois} \quad : \quad \hat{H}_{15} = -\frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_n^{Pois}(X_i). \tag{4.15}$$

$$\text{Plugin-Quantile} \quad : \quad \hat{H}_{16} = \int_0^1 \log \tilde{q}_n(p) dp. \tag{4.16}$$

The simulation study is organized as follows. For each density, three sample sizes: $n = 10$, $n = 50$, and $n = 100$ are tested. While 500 replication data are generated for sample size 10 and 50, only 100 copies are used to obtain estimators for sample size 100 due to calculation expense. Since the optimal choice for m in spacing estimators is still an opening problem, we use the following heuristically deterministic formula for $m = \lfloor \sqrt{n} + 0.5 \rfloor$. However, to verify the effect of the choice of m on the spacing estimator performance, we also run simulation with different choices of m starting from 2 up to $\lfloor n/2 \rfloor$ for the sample size $n = 50$ (see Figure 11 and 12). While the optimal bandwidth is chosen by biased cross validation in $\hat{H}_8$, the rest kernel and smoothed histogram estimators are implemented with the unbiased cross validation for the bandwidth selection. Lastly, the choice of m in the quantile density entropy estimator $\hat{H}_{16}$ is fixed to $m = n/\log n$.

The simulation results for the entropy estimators performance comparison with different sample size are given in the Table 1, 2, and 3 (for $n = 10$), Table 4, 5, and 6 (for $n = 50$), and Table 7, 8, and 9 (for $n = 100$). To compare the performance between estimators, for each density, we compute the point estimate and its mean squared error shown in the parentheses. Each column in the table corresponds to one density with the associated true entropy right below, and the bold symbol in that column indicates the best estimator with the smallest mean squared error for the associated density. Moreover, the last column of each table presents the average time consumption over the fifteen densities.

		Monotone densities					
		Exp(1) H=1	Exp(10) -H=1.3026	Pareto(2,1) H=0.8069	Lnorm(0,2) H=2.1121	Weibull(0.5,0.5) H=0.4228	time (second)
Spacing estimators	$\hat{H}_1$	0.5838(0.3018)	1.7188(0.3018)	0.4311(0.4433)	2.0332(0.7174)	0.3468(0.4641)	<1
	$\hat{H}_2$	0.8400(0.1542)	1.4626(0.1542)	0.6873(0.3164)	2.2894(0.7426)	0.6030(0.4908)	<1
	$\hat{H}_3$	0.9180(0.1407)	1.3846(0.1407)	0.6566(0.2885)	2.1096(0.5984)	0.5333(0.5014)	<1
	$\hat{H}_4$	0.7770(0.1808)	1.5256(0.1808)	0.6448(0.3456)	2.2908(0.7822)	0.5911(0.4899)	<1
	$\hat{H}_5$	1.0052(0.1286)	1.2973(0.1286)	0.8526(0.3042)	2.4547(0.8285)	0.7683(0.5777)	<1
	$\hat{H}_6$	1.1257(0.1459	1.1769(0.1459)	1.0851(0.5262)	3.0921(2.3025)	1.2261(1.1944)	1-2
	$\hat{H}_7$	0.8838(0.2347)	1.4222(0.2375)	0.4827(0.4214)	1.2974(1.2827)	-0.4488(1.9748)	4-8
Non-spacing estimators	$\hat{H}_8$	**1.0882(0.0763)**	0.8664(0.3279)	1.1821(0.1755)	0.8069(1.8228)	1.3560(0.9734)	2-3
	$\hat{H}_9$	1.1503(0.1382)	1.1800(0.1611)	1.2108(0.4928)	3.0693(2.1123)	1.6624(1.8076)	2-3
	$\hat{H}_{10}$	1.0032(0.0936)	0.9930(0.2956)	0.7367(0.1587)	1.7854(0.3306)	0.4866(0.1848)	5-10
	$\hat{H}_{11}$	0.9719(0.1988)	1.3343(0.2027)	0.6199(0.2874)	1.5046(1.2888)	-0.0069(1.0753)	4-8
	$\hat{H}_{12}$	0.8024(0.1004)	1.2433(0.1147)	1.2516(0.3127)	1.7247(0.3886)	Divergent	6-11
	$\hat{H}_{13}$	0.9144(0.1746)	1.4077(0.1942)	0.6424(0.3232)	1.8315(0.6200)	Divergent	5-10
	$\hat{H}_{14}$	1.1036(0.1635)	1.0567(0.1262)	0.9933(0.1460)	1.9366(0.6304)	0.5667(0.2419)	15-30
	$\hat{H}_{15}$	0.8668(0.1138)	**1.3361(0.0827)**	**0.7643(0.1364)**	**2.0139(0.3078)**	**0.3003(0.2977)**	15-25
	$\hat{H}_{16}$	0.7272(0.2140)	1.5754(0.2140)	0.6924(0.3503)	2.2866(0.8415)	0.6751(0.5048)	2-3

Table 1: Simulation results for $n = 10$ and $m = 3$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Unimodal densities						
		Gam(2,2) H=0.8841	Gam(7.5,1) H=2.3804	Lnorm(0,0.5) H=0.7258	Maxwell(1) H=0.9962	Maxwell(20) -H=0.5017	Weibull(2,2) H=1.2886	time (second)
Spacing estimators	$\hat{H}_1$	0.4019(0.3240)	1.8622(0.3400)	0.2194(0.3711)	0.4766(0.3377)	1.0213(0.3377)	0.7422(0.3590)	<1
	$\hat{H}_2$	0.6581(0.1426)	2.1184(0.1401)	0.4756(0.1773)	0.7328(0.1371)	0.7651(0.1371)	0.9984(0.1447)	<1
	$\hat{H}_3$	0.7290(0.1368)	2.1948(0.1178)	0.5398(0.1527)	0.8144(0.1199)	0.6834(0.1199)	1.0847(0.1144)	<1
	$\hat{H}_4$	0.5877(0.1799)	2.0431(0.1868)	0.4049(0.2220)	0.6547(0.1843)	0.8432(0.1843)	0.9172(0.1997))	<1
	$\hat{H}_5$	0.8234(0.0952)	2.2837(0.0808)	0.6409(0.1219)	0.8980(0.0774)	0.5998(0.0774)	1.1637(0.0761	<1
	$\hat{H}_6$	0.8922(0.0943)	2.3308(0.0728)	0.6992(0.1187)	0.9491(0.0646)	0.5488(0.0646)	1.2082(0.0646)	1-2
	$\hat{H}_7$	0.7568(0.1853)	2.2605(0.1230)	0.5533(0.1854)	0.9132(0.1131)	0.5840(0.1140)	1.2220(0.1287)	4-8
Non-spacing estimators	$\hat{H}_8$	**0.9154(0.0651)**	**2.4492(0.0549)**	0.8567(0.0850)	1.0506(0.0437)	0.4169(0.0563)	**1.3072(0.0429)**	2-3
	$\hat{H}_9$	0.8654(0.1049)	2.2705(0.0754)	0.7067(0.0924)	0.9040(0.0668)	0.5939(0.0668)	1.2224(0.0723)	2-3
	$\hat{H}_{10}$	0.9262(0.0808)	2.4770(0.0836)	0.8139(0.0872)	1.1052(0.0649)	0.3363(0.0846)	1.3544(0.0701)	5-10
	$\hat{H}_{11}$	0.7993(0.1677)	2.2555(0.1073)	0.5746(0.1756)	0.9027(0.1109)	0.5439(0.0695)	1.2287(0.1203)	4-8
	$\hat{H}_{12}$	1.0383(0.1004)	2.5829(0.0927)	0.9169(0.1271)	1.2317(0.1145)	0.3534(0.0574)	1.4629(0.0755)	6-11
	$\hat{H}_{13}$	0.7502(0.1830)	2.2617(0.1185)	0.5944(0.1594)	0.8980(0.1373)	0.5996(0.1356)	1.2118(0.1190)	5-10
	$\hat{H}_{14}$	1.0000(0.0837)	2.4187(0.0715)	0.8937(0.0821)	1.1755(0.0739)	0.2505(0.0801)	1.3507(0.0632)	15-30
	$\hat{H}_{15}$	0.7714(0.0801)	2.2060(0.1060)	**0.6519(0.0691)**	**0.9998(0.0325)**	**0.5043(0.0266)**	1.1502(0.0832)	15-25
	$\hat{H}_{16}$	0.5040(0.2332)	1.9697(0.2387)	0.3450(0.2533)	0.5998(0.2212)	0.8980(0.2212)	0.9114(0.2112)	2-3

Table 2: Simulation results for $n = 10$ and $m = 3$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Bimodal densities				
		MixGam H=2.2757	MixLnorm H=1.6724	MixMaxwell H=0.8014	MixWeibull H=1.1330	time (second)
Spacing estimators	$\hat{H}_1$	1.9857(0.1193)	1.5080(0.5933)	0.4343(0.1810)	0.9092(0.1874)	<1
	$\hat{H}_2$	2.2419(0.0363)	1.7642(0.5746)	0.6905(0.0585)	1.1654(0.1384)	<1
	$\hat{H}_3$	2.4405(0.0630)	1.5013(0.3675)	0.8391(0.0595)	1.2905(0.1311)	<1
	$\hat{H}_4$	2.1881(0.0419)	1.7520(0.6218)	0.6283(0.0747)	1.1150(0.1457)	<1
	$\hat{H}_5$	2.4071(0.0525)	1.9294(0.6323)	0.8558(0.0492)	1.3307(0.1764)	<1
	$\hat{H}_6$	2.6566(0.1843)	2.3774(1.9139)	1.0576(0.1191)	1.4830(0.2522)	1-2
	$\hat{H}_7$	2.0863(0.2145)	1.2056(0.4836)	0.6552(0.2129)	0.8225(0.9557)	4-8
Non-spacing estimators	$\hat{H}_8$	**2.2745(0.0144)**	2.0122(0.2035)	1.0400(0.0756)	1.3273(0.0966)	2-3
	$\hat{H}_9$	2.6655(0.1881)	2.3348(1.7550)	1.0453(0.1101)	1.5408(0.3009)	2-3
	$\hat{H}_{10}$	2.4110(0.0506)	1.4978(0.1739)	0.9117(0.0726)	**1.2040(0.0831)**	5-10
	$\hat{H}_{11}$	2.4676(0.1307)	1.4059(0.3730)	0.7468(0.1808)	1.0405(0.5687)	4-8
	$\hat{H}_{12}$	2.5385(0.0907)	**1.5993(0.1518)**	1.0180(0.0915)	Divergent	6-11
	$\hat{H}_{13}$	2.4821(0.1217)	1.4474(0.3560)	0.7744(0.1132)	Divergent	5-10
	$\hat{H}_{14}$	1.8600(0.2022)	1.4875(0.2488)	0.9962(0.0540)	1.3002(0.3714)	15-30
	$\hat{H}_{15}$	2.0699(0.1140)	1.3372(0.2563)	**0.7778(0.0491)**	1.0404(0.1865)	15-25
	$\hat{H}_{16}$	2.1290(0.0563)	1.7298(0.8283)	0.5777(0.0961)	1.0935(0.1563)	2-3

Table 3: Simulation results for $n = 10$ and $m = 3$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Monotone densities					
		Exp(1) H=1	Exp(10) -H=1.3026	Pareto(2,1) H=0.8069	Lnorm(0,2) H=2.1121	Weibull(0.5,0.5) H=0.4228	time (second)
Spacing estimators	$\hat{H}_1$	0.8697(0.0393)	1.4329(0.0393)	0.7608(0.0665)	2.1770(0.1357)	0.4912(0.1032)	<1
	$\hat{H}_2$	0.9697(0.0232)	1.3329(0.0232)	0.8608(0.0673)	2.2771(0.1587)	0.5912(0.1269)	<1
	$\hat{H}_3$	0.8933(0.0327)	1.4093(0.0327)	0.6793(0.0681)	2.0010(0.1192)	0.4593(0.1022)	<1
	$\hat{H}_4$	0.9661(0.0242)	1.3365(0.0242)	0.8725(0.0727)	2.3044(0.1737)	0.6126(0.1361)	<1
	$\hat{H}_5$	1.0246(0.0228)	1.2820(0.0228)	0.9117(0.0753)	2.3279(0.1780)	0.6421(0.1466)	<1
	$\hat{H}_6$	1.1041(0.0338)	1.1985(0.0338)	1.2363(0.3931)	3.4108(2.2796)	1.3615(1.0130)	3-5
	$\hat{H}_7$	0.9347(0.0359)	1.3697(0.0380)	0.5094(0.1523)	1.6531(0.3173)	Divergent	6-12
Non-spaicng estimators	$\hat{H}_8$	1.0947(0.0281)	0.9171(0.1816)	1.2517(0.1105)	**2.1595(0.0668)**	1.2787(0.7648)	5-10
	$\hat{H}_9$	1.1503(0.1382)	1.1615(0.0522)	1.2681(0.2444)	3.5506(2.6429)	1.5301(1.3145)	4-8
	$\hat{H}_{10}$	**1.0407(0.0214)**	1.1201(0.0832)	0.6520(0.0517)	1.8444(0.1333)	Divergent	10-18
	$\hat{H}_{11}$	1.0238(0.0330)	1.2788(0.0329)	0.6614(0.0657)	1.9008(0.1549)	Divergent	10-15
	$\hat{H}_{12}$	1.0881(0.0275)	1.1446(0.0531)	Divergent	1.7022(0.4329)	Divergent	12-25
	$\hat{H}_{13}$	0.9935(0.0271)	0.4187(1.1567)	Divergent	1.8315(0.6200)	Divergent	10-18
	$\hat{H}_{14}$	1.1021(0.0681)	0.8268(0.0738)	1.0397(0.0749)	1.3202(0.6532)	0.6322(0.0896)	20-50
	$\hat{H}_{15}$	0.9624(0.0253)	**1.2167(0.0168)**	**0.9008(0.0386)**	1.9096(0.2498)	**0.4715(0.0681)**	20-40
	$\hat{H}_{16}$	1.0215(0.025)	1.2811(0.025)	0.9867(0.1284)	2.5532(0.3923)	0.751(0.2102)	3-5

Table 4: Simulation results for $n = 50$ and $m = 7$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Unimodal densities						
		Gam(2,2) H=0.8841	Gam(7.5,1) H=2.3804	Lnorm(0,0.5) H=0.7258	Maxwell(1) H=0.9962	Maxwell(20) -H=0.5017	Weibull(2,2) H=1.2886	time (second)
Spacing estimators	$\hat{H}_1$	0.7312(0.0383)	2.2093(0.0422)	0.5819(0.0384)	0.8331(0.0369)	0.6648(0.0369)	1.1213(0.0383)	<1
	$\hat{H}_2$	0.8313(0.0177)	2.3093(0.0180)	0.6819(0.0196)	0.9331(0.0143)	0.5648(0.0143)	1.2213(0.0149)	<1
	$\hat{H}_3$	0.7371(0.0378)	2.2041(0.0442)	0.5550(0.0484)	0.8390(0.0363)	0.6588(0.0363)	1.1426(0.0333)	<1
	$\hat{H}_4$	0.8244(0.0189)	2.3014(0.0196)	0.6762(0.0204)	0.9246(0.0155)	0.5733(0.0155)	1.2118(0.0166)	<1
	$\hat{H}_5$	0.8821(0.0149)	2.3602(0.0133)	0.7328(0.0178)	0.9840(0.0105)	0.5139(0.0105)	1.2722(0.0106)	<1
	$\hat{H}_6$	0.8878(0.0166)	2.3380(0.0144)	0.7046(0.0193)	0.9744(0.0109)	0.5234(0.0109)	1.2744(0.0114)	3-5
	$\hat{H}_7$	0.8277(0.0271)	2.3259(0.0220)	0.6396(0.0297)	0.9672(0.0155)	0.5309(0.0156)	1.2550(0.0171)	6-12
Non-spaicng estimators	$\hat{H}_8$	0.9401(0.0157)	2.4845(0.0206)	0.8495(0.0300)	1.0703(0.0132)	0.4158(0.0160)	**1.3289(0.0084)**	5-10
	$\hat{H}_9$	0.9130(0.0241)	**2.3623(0.0113)**	0.7320(0.0172)	0.9736(0.0103)	0.5243(0.0103)	1.2805(0.0099)	4-8
	$\hat{H}_{10}$	0.9283(0.0161)	2.4610(0.0202)	0.7937(0.0204)	1.0724(0.0160)	0.4138(0.0175)	1.3420(0.0142)	10-18
	$\hat{H}_{11}$	0.8570(0.0225)	2.3392(0.0151)	0.6751(0.0240)	0.9665(0.0131)	0.5307(0.0105)	1.2493(0.0148)	10-15
	$\hat{H}_{12}$	1.0107(0.0340)	2.5282(0.0318)	0.8451(0.0319)	1.1377(0.0301)	0.3603(0.0300)	1.4200(0.0293)	12-25
	$\hat{H}_{13}$	0.8598(0.0191)	2.3366(0.0181)	0.6898(0.0226)	0.1530(0.8343)	0.0898(0.2137)	1.2544(0.0157)	10-18
	$\hat{H}_{14}$	0.9949(0.0244)	2.4523(0.0200)	0.9028(0.0409)	1.1460(0.0281)	0.2306(0.0770)	1.3504(0.0129)	20-50
	$\hat{H}_{15}$	**0.8667(0.0126)**	2.3436(0.0132)	**0.7344(0.0125)**	**0.9939(0.0075)**	**0.4466(0.0088)**	1.2444(0.0108)	20-40
	$\hat{H}_{16}$	0.8705(0.0169)	2.3480(0.0151)	0.7267(0.0259)	0.9621(0.0105)	0.5358(0.0105)	1.2434(0.0138)	3-5

Table 5: Simulation results for $n = 50$ and $m = 7$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Bimodal densities				
		MixGam H=2.2757	MixLnorm H=1.6724	MixMaxwell H=0.8014	MixWeibull H=1.1330	time (second)
Spacing estimators	$\hat{H}_1$	2.1708(0.0192)	1.7774(0.1005)	0.6618(0.0272)	1.1667(0.0481)	<1
	$\hat{H}_2$	2.2708(0.0083)	1.8774(0.1315)	0.7619(0.0093)	1.2667(0.0648)	<1
	$\hat{H}_3$	2.2500(0.0094)	1.4962(0.0939)	0.7127(0.0162)	**1.2032(0.0308)**	<1
	$\hat{H}_4$	2.2720(0.0083)	1.9080(0.1495)	0.7563(0.0103)	1.2847(0.0749)	<1
	$\hat{H}_5$	2.3217(0.0104)	1.9283(0.1550)	0.8127(0.0079)	1.3176(0.0810)	<1
	$\hat{H}_6$	2.6009(0.1132)	2.7151(1.7748)	0.9912(0.0450)	1.5292(0.1911)	3-5
	$\hat{H}_7$	2.1663(0.0412)	1.2977(0.1947)	0.6940(0.0303)	Divergent	6-12
Non-spaicng estimators	$\hat{H}_8$	**2.3096(0.0049)**	1.9924(0.1466)	1.0265(0.0556)	1.3418(0.0556)	5-10
	$\hat{H}_9$	2.6187(0.1268)	2.8481(2.3296)	1.0257(0.0595)	1.5318(0.1848)	4-8
	$\hat{H}_{10}$	2.4627(0.0097)	**1.5492(0.0544)**	0.8414(0.0096)	Divergent	10-18
	$\hat{H}_{11}$	2.4743(0.0585)	1.5405(0.0819)	0.7259(00213)	Divergent	10-15
	$\hat{H}_{12}$	2.5912(0.1065)	1.0476(0.3977)	0.9233(0.0246)	Divergent	12-25
	$\hat{H}_{13}$	2.4904(0.0590)	1.6021(0.0795)	0.7714(0.0167)	Divergent	10-18
	$\hat{H}_{14}$	1.9261(0.1279)	1.8067(0.1110)	0.9689(0.0112)	1.2632(0.2374)	20-50
	$\hat{H}_{15}$	2.2249(0.0126)	1.7353(0.0983)	**0.8331(0.0073)**	1.1285(0.1352)	20-40
	$\hat{H}_{16}$	2.3299(0.0121)	2.1617(0.4259)	0.8168(0.0079)	1.3226(0.0996)	3-5

Table 6: Simulation results for $n = 50$ and $m = 7$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Monotone densities					
		Exp(1) H=1	Exp(10) -H=1.3026	Pareto(2,1) H=0.8069	Lnorm(0,2) H=2.1121	Weibull(0.5,0.5) H=0.4228	time (second)
Spacing estimators	$\hat{H}_1$	0.9192(0.0183)	1.3834(0.0183)	0.7491(0.0280)	**2.1815(0.0466)**	0.4963(0.0448)	1
	$\hat{H}_2$	0.9876(0.0119)	1.3150(0.0119)	**0.8114(0.0247)**	2.2499(0.0608)	0.5647(0.0595)	1
	$\hat{H}_3$	0.9079(0.0196)	1.3947(0.0196)	0.6613(0.0452)	1.9895(0.0473)	**0.4467(0.0368)**	1
	$\hat{H}_4$	0.9878(0.0125)	1.3148(0.0125)	0.8290(0.0260)	2.2713(0.0694)	0.5806(0.0648)	1
	$\hat{H}_5$	1.0200(0.0122)	1.2826(0.0122)	0.8467(0.0262)	2.2823(0.0708)	0.5972(0.0698)	1
	$\hat{H}_6$	1.1058(0.0223)	1.1968(0.0223)	1.1725(0.1764)	3.5184(2.4516)	1.3713(0.9682)	5-10
	$\hat{H}_7$	0.9477(0.0189)	1.3533(0.0186)	0.5269(0.1136)	1.7883(0.1695)	Divergent	15-30
Non-spaicng estimators	$\hat{H}_8$	1.0766(0.0155)	0.9866(0.1170)	1.5534(0.5829)	2.2007(0.0472)	1.2544(0.7132)	10-20
	$\hat{H}_9$	1.1154(0.0272)	1.1872(0.0272)	1.1505(0.2464)	3.6329(2.6706)	1.4436(1.1107)	8-15
	$\hat{H}_{10}$	1.0479(0.0115)	1.1095(0.0474)	Divergent	1.8300(0.1246)	Divergent	20-50
	$\hat{H}_{11}$	1.0375(0.0128)	1.2651(0.0128)	0.6637(0.0502)	1.9686(0.0903)	Divergent	15-40
	$\hat{H}_{12}$	0.9321(0.0112)	1.1766(0.0274)	Divergent	1.8476(0.0994)	Divergent	30-60
	$\hat{H}_{13}$	0.9928(0.0123)	1.3098(0.0124)	0.7202(0.0391)	2.0834(0.0670)	Divergent	20-50
	$\hat{H}_{14}$	1.0504(0.0105)	1.1435(0.0335)	1.1832(0.1536)	1.5465(0.3589)	0.6636(0.0861)	100-200
	$\hat{H}_{15}$	**0.9984(0.0100)**	**1.2920(0.0095)**	1.2978(0.5435)	2.1144(0.0604)	0.6181(0.0859)	80-150
	$\hat{H}_{16}$	1.040(0.0122)	1.2626(0.0122)	0.9361(0.0599)	2.4328(0.1961)	0.6599(0.1108)	5-7

Table 7: Simulation results for $n = 100$ and $m = 10$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Unimodal densities						
		Gam(2,2) H=0.8841	Gam(7.5,1) H=2.3804	Lnorm(0,0.5) H=0.7258	Maxwell(1) H=0.9962	Maxwell(20) -H=0.5017	Weibull(2,2) H=1.2886	time (second)
Spacing estimators	$\hat{H}_1$	0.8138(0.0106)	2.2987(0.0130)	0.6546(0.0130)	09112(0.0122)	0.5867(0.0122)	1.1917(0.0139)	1
	$\hat{H}_2$	**0.8822(0.0056)**	2.3662(0.0064)	0.7230(0.0079)	0.9796(0.0052)	0.5183(0.0052)	1.2600(0.0054)	1
	$\hat{H}_3$	0.7785(0.0177)	2.2469(0.0236)	0.5767(0.0291)	0.8722(0.0218)	0.6257(0.0218)	1.1649(0.0197)	1
	$\hat{H}_4$	0.8798(0.0058)	2.3644(0.0066)	0.7241(0.0084)	0.9772(0.0054)	0.5207(0.0054)	1.2561(0.0057)	1
	$\hat{H}_5$	0.9147(0.0066)	2.3986(0.0065)	0.7554(0.0088)	1.0120(0.0052)	0.4858(0.0052)	1.2925(0.0046)	1
	$\hat{H}_6$	0.9060(0.0075)	**2.3619(0.0056)**	0.7054(0.0084)	0.9862(0.0064)	0.5117(0.0064)	**1.2797(0.0044)**	5-10
	$\hat{H}_7$	0.8400(0.0124)	2.3426(0.0120)	0.6613(0.0174)	0.9673(0.0088)	0.5304(0.0088)	1.2480(0.0099)	15-30
Non-spaicng estimators	$\hat{H}_8$	0.9302(0.0092)	2.4680(0.0149)	0.8441(0.0206)	1.0659(0.0088)	0.4233(0.0105)	1.3197(0.0056)	10-20
	$\hat{H}_9$	0.9060(0.0101)	2.3700(0.0078)	0.7242(0.0109)	0.9897(0.0050)	0.5081(0.0050)	1.2792(0.0057)	8-15
	$\hat{H}_{10}$	0.9122(0.0090)	2.4490(0.0136)	0.7812(0.0142)	1.0588(0.0111)	0.4386(0.0111)	1.3280(0.0083)	20-50
	$\hat{H}_{11}$	0.8628(0.0100)	2.3522(0.0103)	0.6953(0.0137)	0.9700(0.0073)	0.5278(0.0073)	1.2546(0.0079)	15-40
	$\hat{H}_{12}$	0.9831(0.0204)	2.5158(0.0232)	0.8252(0.0197)	1.1162(0.0189)	0.3890(0.0174)	1.3873(0.0168)	30-60
	$\hat{H}_{13}$	0.8704(0.0093)	2.0834(0.0670)	0.7077(0.0126)	0.9716(0.0070)	0.5263(0.0070)	1.2572(0.0078)	20-50
	$\hat{H}_{14}$	0.9883(0.0171)	2.4238(0.0193)	0.9055(0.0334)	1.1451(0.0252)	0.2192(0.0819)	1.3754(0.0117)	100-200
	$\hat{H}_{15}$	0.8810(0.0072)	2.3825(0.0074)	**0.7603(0.0027)**	**1.0078(0.0041)**	0.4330(0.0079)	1.2713(0.0050)	80-150
	$\hat{H}_{16}$	0.8917(0.0083)	2.3830(0.0086)	0.7542(0.0134)	0.9879(0.0042)	**0.5099(0.0042)**	1.2623(0.0072)	5-7

Table 8: Simulation results for $n = 100$ and $m = 10$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

		Bimodal densities				
		MixGam H=2.2757	MixLnorm H=1.6724	MixMaxwell H=0.8014	MixWeibull H=1.1330	time (second)
Spacing estimators	$\hat{H}_1$	2.2124(0.0076)	1.7776(0.0453)	0.7302(0.0091)	1.1999(0.0271)	1
	$\hat{H}_2$	**2.2807(0.0036)**	1.8460(0.0644)	0.7986(0.0041)	1.2683(0.0409)	1
	$\hat{H}_3$	2.2304(0.0057)	1.5117(0.0482)	0.7284(0.0097)	**1.1804(0.0134)**	1
	$\hat{H}_4$	2.2827(0.0039)	1.8698(0.0759)	0.7983(0.0042)	1.2864(0.0481)	1
	$\hat{H}_5$	2.3132(0.0050)	1.8785(0.0767)	0.8310(0.0050)	1.3007(0.0507)	1
	$\hat{H}_6$	2.5558(0.0812)	2.8557(2.0431)	0.9835(0.0371)	1.5245(0.1723)	5-10
	$\hat{H}_7$	2.1820(0.0217)	1.3675(0.1308)	0.7286(0.0115)	Divergent	15-30
Non-spaicng estimators	$\hat{H}_8$	2.4148(0.0211)	2.0156(0.1463)	1.0140(0.0484)	1.3455(0.0529)	10-20
	$\hat{H}_9$	2.5659(0.0879)	2.9796(2.3270)	0.8951(0.0261)	1.5313(0.1746)	8-15
	$\hat{H}_{10}$	2.4707(0.0410)	1.5518(0.0368)	0.8300(0.0045)	Divergent	20-50
	$\hat{H}_{11}$	2.4534(0.0393)	1.5709(0.0510)	0.7508(0.0076)	Divergent	15-40
	$\hat{H}_{12}$	2.5749(0.0941)	**1.5994(0.0271)**	0.8870(0.0120)	Divergent	30-60
	$\hat{H}_{13}$	2.4709(0.0439)	1.6266(0.0324)	0.7785(0.0055)	Divergent	20-50
	$\hat{H}_{14}$	2.5162(0.0629)	1.5034(0.0629)	0.9706(0.0305)	1.2541(0.0263)	100-200
	$\hat{H}_{15}$	2.4422(0.0314)	1.6866(0.0368)	0.8428(0.0037)	1.2897(0.1050)	80-150
	$\hat{H}_{16}$	2.3291(0.0059)	2.0210(0.2035)	**0.8277(0.0035)**	1.2993(0.0616)	5-7

Table 9: Simulation results for $n = 100$ and $m = 10$. True entropy is shown below density name. For each density, the point estimate and mean squared error (in parentheses) are computed for each estimator. The bold symbol is the best estimator with the smallest mean squared error.

By observing the simulation results above, we have some important remarks.

Remark 4.1. *There does not exist the uniquely best estimator in all cases.*

When comparing entropy estimators, not only the smallest mean squared error is considered, but the computation expense is also an important key to determine the best estimator. In general, if both criteria are considered, there exists no unique estimator that beats out all other estimators.

Remark 4.2. *Spacing entropy estimators always win the other group of entropy estimators in terms of computation expense.*

When speed is the only one in the consideration, the estimators in the spacing group always dominate the other groups of estimators to stand in the first rank. This result does not come to surprise because there is no parameter optimization in their definitions. Definitely, this could be the main reason why most of papers in literature only focus on spacing estimators but a few in the other group. And if the density estimation or any density related approaches are employed, researchers try to simplify their estimator by using a simple bandwidth selection. Noughabi's estimator $\hat{H}_6$ is an example of such a case, where the fixed bandwidth selection is used, and it stands on the second rank in the speed test. However, the computation expense is improved significantly when the fixed kernel density is used due to its simple and symmetric form. Therefore, our second and third entropy estimator candidates ($\hat{H}_8$ and $\hat{H}_9$) do a good job when time consumption is taking into account to achieve the third rank. Among all, the Poisson smoothed histogram entropy estimator $\hat{H}_{15}$ appears to be the lowest one in the race of speed.

Remark 4.3. *For spacing entropy estimator, the optimal m varies case by case, and there exists no expression for the optimal m.*

In order to study more the behavior of spacing estimators with different choices of m, we provide the simulation results for the same density set but with various values of m in the Figure 11 and 12 below. From the figures, we see that the effect of the choice of m on the estimator's performance is quite noticeable. Especially for those values of m that bigger than $\lfloor\sqrt{n}+0.5\rfloor$, the MSEs start to disperse. Although the heuristically deterministic choice of $m = \lfloor\sqrt{n}+0.5\rfloor = 7$ does not attain the optimum, it could be used as an upper bound for the selection of m. Nonetheless, the choice of m is still the biggest weakness in spacing entropy estimators, and the issue seems to exist insolvably because there exists no relating loss function of m like mean squared error to optimize. On the other hand, by assuming that someone can define a loss function of m, then by solving the optimization problem to find the optimal m, the spacing estimators will loose the strength of fast computation.

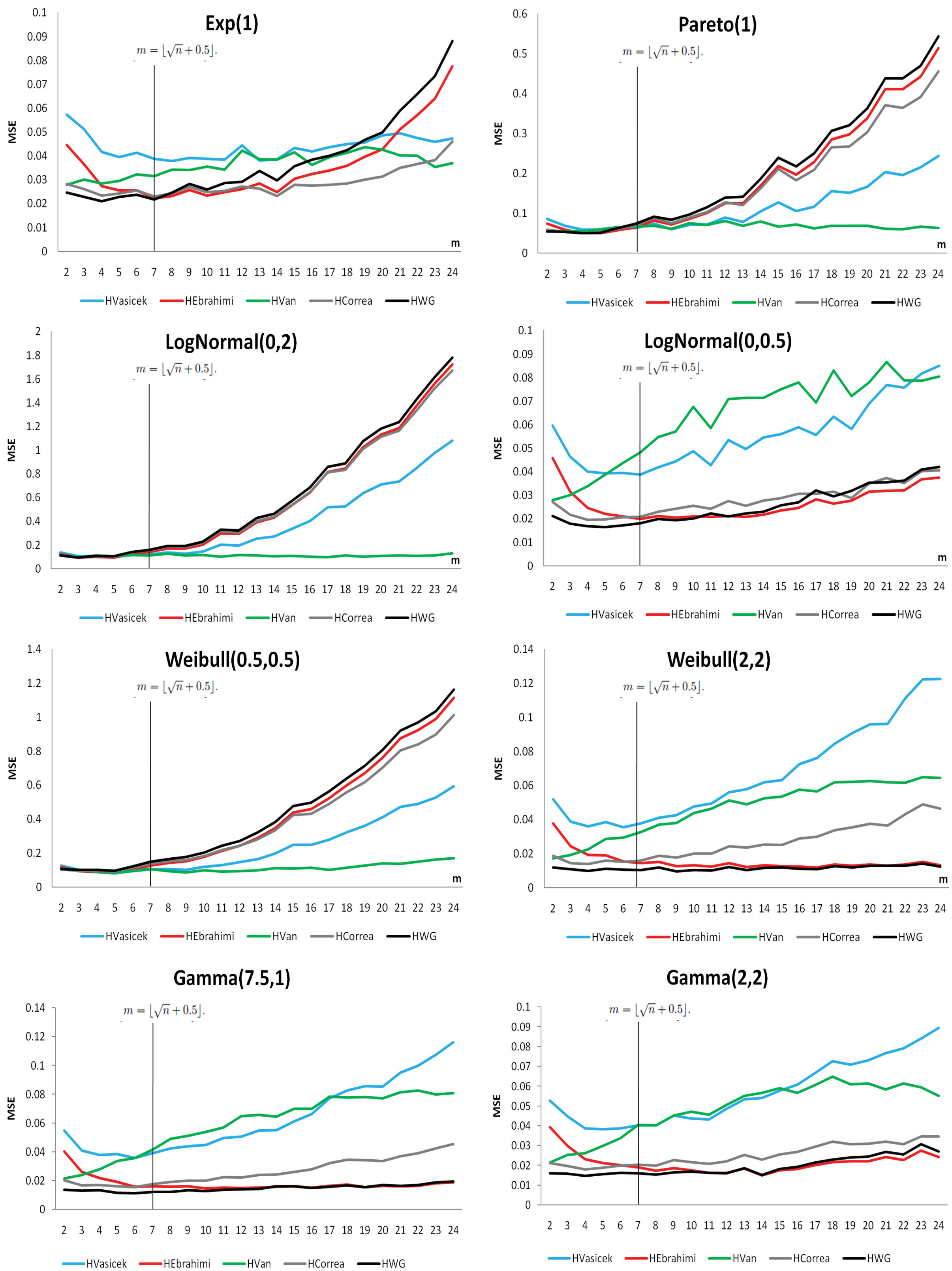

Figure 11: Plots of Simulation result for $n = 50$ and different m within the spacing estimator group. The vertical line $m = \lfloor \sqrt{n} + 0.5 \rfloor$ is set to be the upper bound for the choice of m.

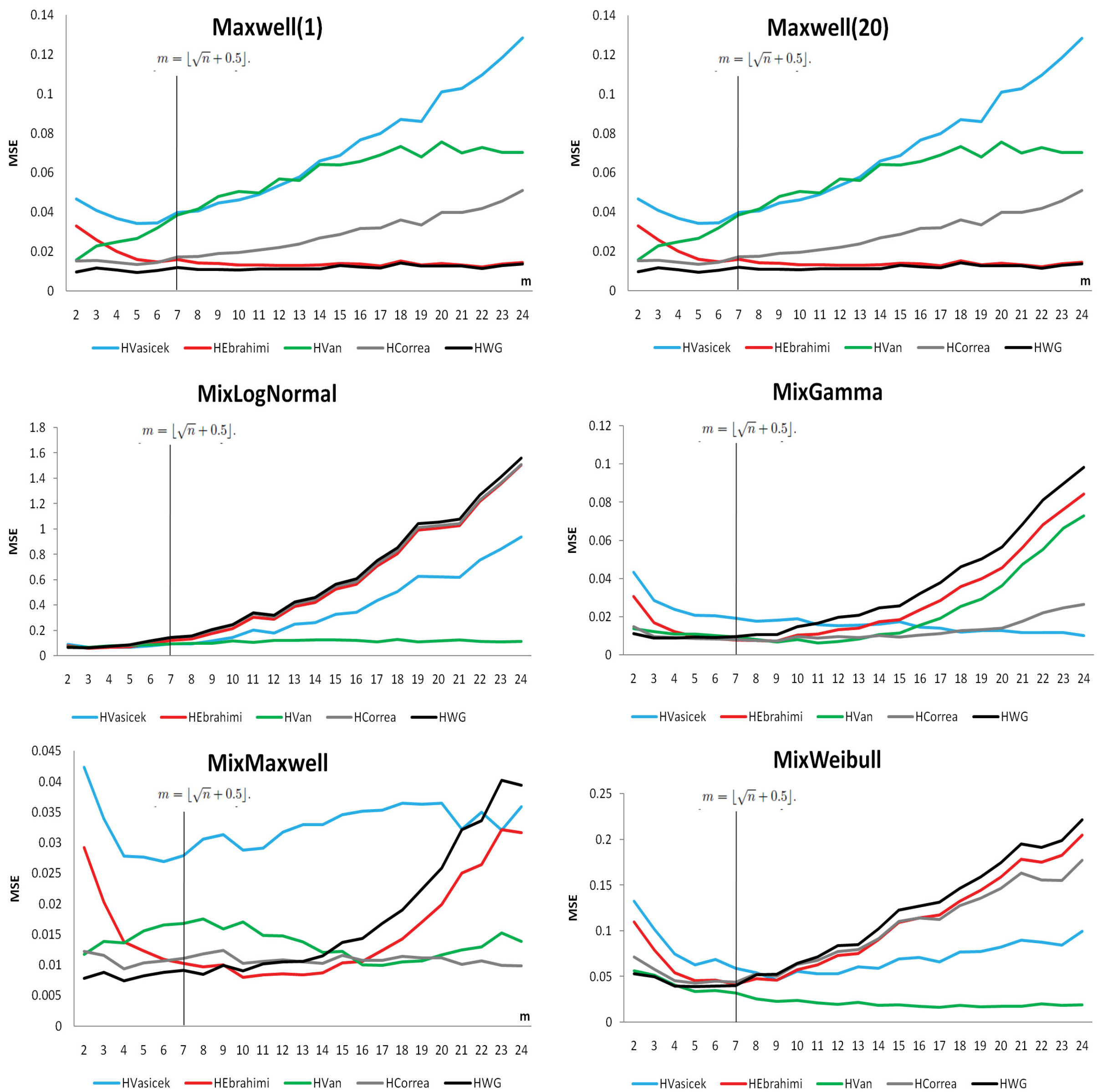

Figure 12: Plots of Simulation result for $n = 50$ and different m within the spacing estimator group. The vertical line $m = \lfloor \sqrt{n} + 0.5 \rfloor$ is set to be the upper bound for the choice of m.

Remark 4.4. *Among the estimators in the spacing group, The WG's entropy estimator $\hat{H}_5$ seems to outperform the other members in most cases.*

We see that in most cases, if we only consider those values m smaller than $\lfloor \sqrt{n} + 0.5 \rfloor$, the WG's estimator seems to produces the smallest mean squared error comparing to the others, and it is quite robust to the choice of m. Even though in some cases where it is not the winner such as logNormal(0,2), Weibull(0.5,0.5), and the mixture densities, the differences in mean squared error between the WG's estimator and the winner's are very small and insignificant. Therefore, it would be the best choice when an spacing entropy estimator is in concern.

Remark 4.5. *Among our candidates for entropy estimators by means of kernel density estimator, the "plug-in" Fixed kernel entropy estimator $\hat{H}_8$ surprisingly performs better than* *other members in kernel density estimator group.*

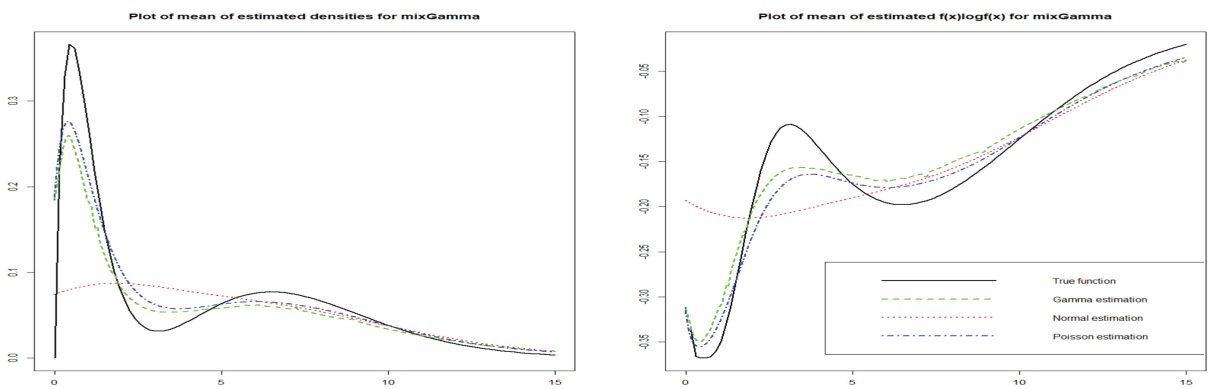

Figure 13: Density estimators comparison for mixture Gamma.

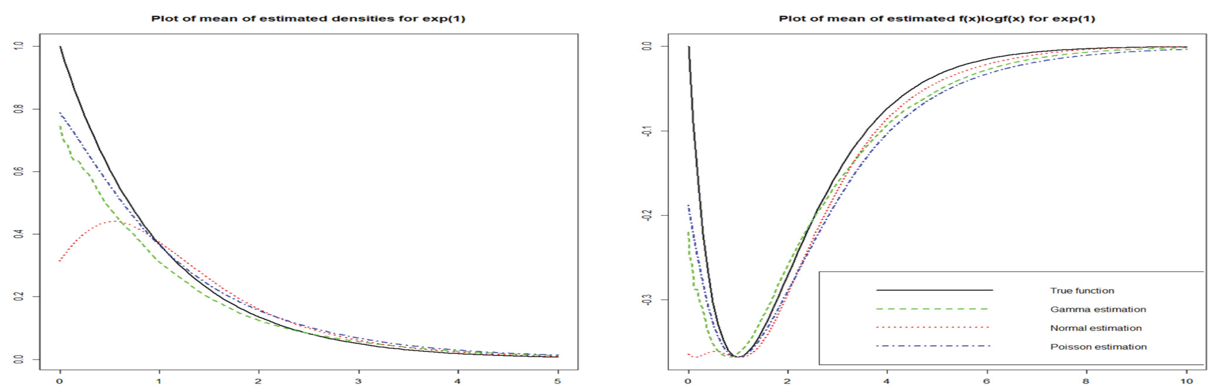

Figure 14: Density estimators comparison for Exp(1).

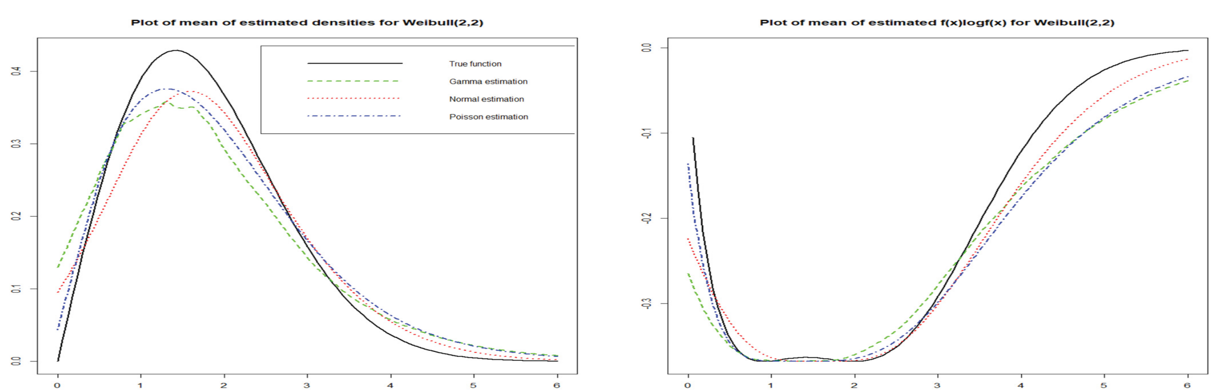

Figure 15: Density estimators comparison for Weibull(2,2).

As shown in Table 1-9 above, in most of cases, the estimator $\hat{H}_8$ ("plug-in" fixed kernel density entropy estimator) and $\hat{H}_{15}$ (mean of log of Poisson density points) achieve the best result among other candidates. This is an unanticipated result for $\hat{H}_8$ because it is very well-known that the fixed kernel density estimator performs very poorly for nonnegative random variable due to "boundary effect" problem. This suggests that we need to have a deeper inspection in these cases where $\hat{H}_8$ outperforms the others for instance the Exponential(1), Weibull(2,2), and the mixture Gamma. Consequently, we present the graphs of the estimator of the function $g(x) = f(x) \log f(x)$. These graphs below are produced for the case $n = 10$ using the mean of 500 samples of fixed kernel density entropy estimator, Gamma kernel density entropy estimator, and Poisson smoothed histogram density entropy estimator. The left panels are the density estimations along with the true density, while the right panels are the estimations of the $f(x) \log f(x)$ function. These graphs show a very poor estimation of the fixed kernel density estimator in the case of Exponential(1) and mixture Gamma, still its entropy estimation produces the smallest mean squared error. This can be explained as by accidentally the area under the estimated curve of $f(x) \log f(x)$ is very close to the true one. On the other hand, by observing the estimation graphs for Weibull(2,2), we see that the fixed kernel density estimator does a good job. This could be simply because the true density of Weibull(2,2) is very similar to Normal($\mu = 1.5, \sigma = 1$). Therefore, we could conclude that the fixed kernel entropy estimator $\hat{H}_8$ and $\hat{H}_9$ are not a good choice for entropy estimator for nonnegative random variable.

Remark 4.6. *The mean of log of Poisson density point $\hat{H}_{15}$ seems to be a potential candidate for entropy estimator when the sample size is small.*

To have an insight of the performance of $\hat{H}_{15}$ for small sample size, we run the simulation for more choices of sample size $n = 10, 20, 30, 40$, and 50. These graphs below show the performance comparison between the mean of log of Poisson density point $\hat{H}_{15}$, the "plug-in" Poisson smoothed histogram density estimator $\hat{H}_{14}$, the "plug-in" fixed kernel density estimator $\hat{H}_8$, the "plug-in" Gamma kernel density estimator $\hat{H}_{10}$, and the WG's estimator $\hat{H}_5$ with the optimal m (here we run all the possible choices of m from 2 to $\lfloor \sqrt{n} + 0.5 \rfloor$, then obtain the optimal m). We see that most of the times, the MSE of $\hat{H}_{15}$ is the smallest comparing to that of others estimators. Even the best spacing entropy estimator WG with the optimal m cannot beat $\hat{H}_{15}$, especially for the case of very small sample size like $n = 10$. The only weakness of the estimator $\hat{H}_{15}$ is the computation expense. It would be ideal if we could improve the speed of computation while maintain its performance. It is worth notice that those computed optimal bandwidths by unbiased cross validation in our proposal entropy estimator are only local but not global because there is no function in R can guarantee the result is a global optimal. As a result, we could improve the estimator $\hat{H}_{15}$ by using a deterministic bandwidth.

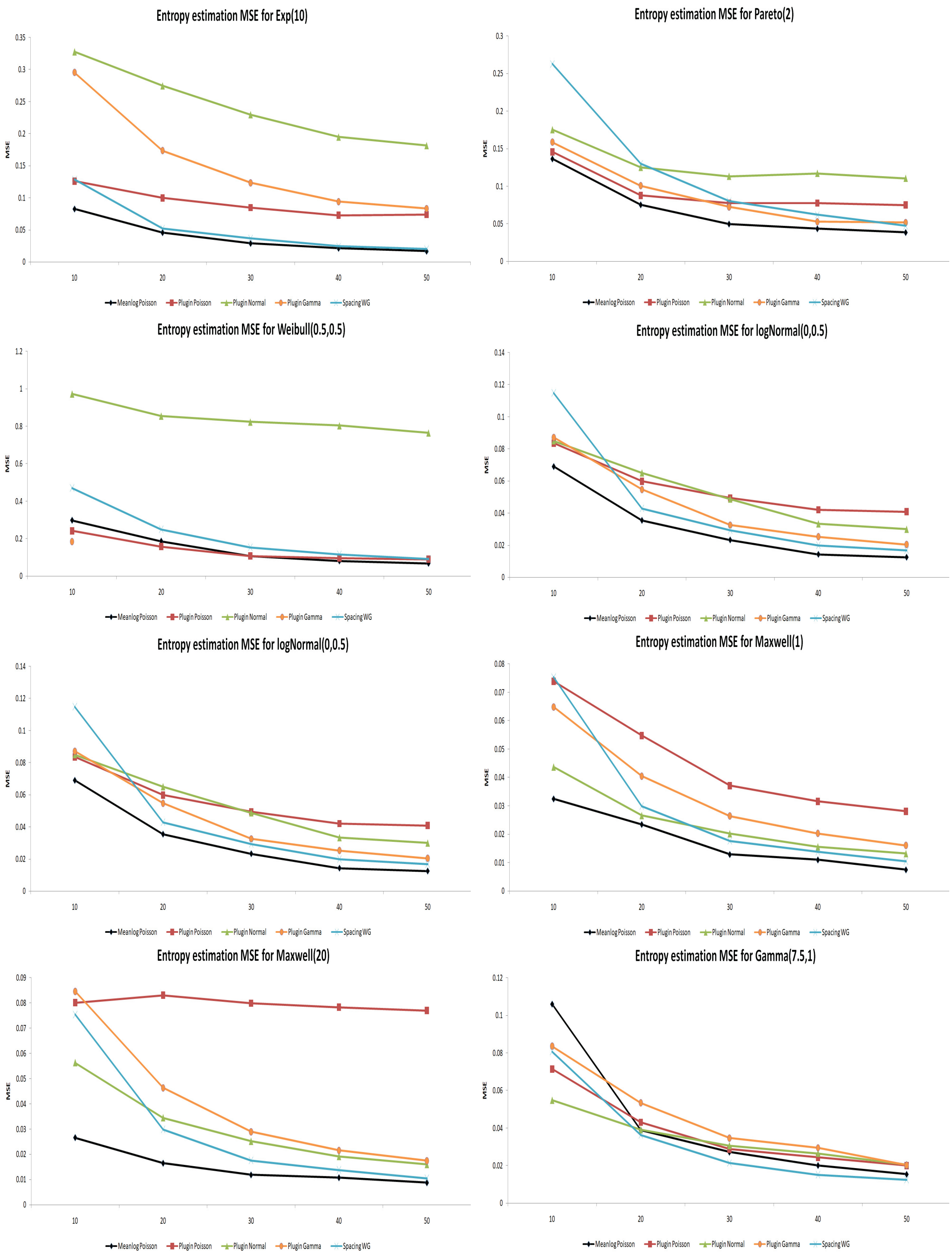

Figure 16: Entropy estimation MSE comparison with different sample size $n = 10, 20, 30, 40$, and 50. Where diamond-black is $\hat{H}_{15}$, square-red is $\hat{H}_{14}$, triangle-green is $\hat{H}_8$, circle-orange is $\hat{H}_{10}$, and asterisk-blue is $\hat{H}_5$.

Remark 4.7. *As the sample size increases, the quantile density entropy estimator $\hat{H}_{16}$ seems to converge faster than all other candidates. As a result, for a large sample size, the quantile density entropy estimator and the WG's estimator could be the best choice for entropy estimation.*

We notice that for a small sample size like $n = 10$ (Table 1-3), the performance of the quantile density entropy estimator $\hat{H}_{16}$ is quite poor among other estimators. That is the differences in MSE between it and the best one are very significant. However, its rate of convergence picks up very fast as the sample size increases like in $n = 100$ (Table 7-9). For instance, it becomes the best estimators in some distributions such as Maxwell(20), mix-Maxwell, or very close to the best one in most cases. Besides, it takes a small amount of time to compute the estimator $\hat{H}_{16}$, which makes it become more favorable than other non-spacing estimators in large sample cases. Similarly, by studying the graphs of entropy estimation (Figure 16), we see that although the WG's estimator with optimal m has a large MSE when the sample size is very small ($n = 10$), its rate of convergence increases faster than all that of other estimators when the sample size starts to increase ($n = 20$). Moreover, its performance becomes better than most of the others when $n = 50$. Consequently, one would pick the quantile density entropy estimator or the WG's estimator for entropy estimation when the sample size is sufficient large.

In conclusion, there is always a trade-off between the speed and the performance for each estimator. It is obvious that any spacing estimator would do a good job and is the best choice when dealing with large sample sizes (though the answer to the question what value of m would be considered for a large sample size is still unknown) due to its fast rate of convergence and fast computation. However, when sample size is small, the choice of estimator for entropy becomes more difficult and requires careful studies. This is why the density entropy estimators come to handy not only because they give a more precise result, but also the computation expense is significantly improved for small sample size. Among our candidates for entropy estimation presented above, only the the mean of log of Poisson density points $\hat{H}_{15}$ is outstanding for a good potential entropy estimator. Particularly, if we are able to improve the time spending on the bandwidth selection while keeping the same performance, we will obtain a better entropy estimator among the existing ones. Indeed, we have been testing some pre-deterministic bandwidths and it turns out that our "best" optimal bandwidth must be of the order $o(n^{2/5})$ as proposed by Gawronski and Stadtmüler (1981). We have tried many bandwidths as a linear function of $n^{2/5}$, some of them produce a huge improvement of entropy estimation on certain distributions but become unstable on other distributions. After running a number of bandwidth choices, we notice that the bandwidth of the form $k^* = n^{2/5} + 1$ is the best one because it consistently results in an improvement for entropy estimation in most of distributions we use in this thesis. Figure 17 and 18 below demonstrate this situation for $n = 10$ and $n = 50$, where we compare the performance of the mean of log of Poisson entropy estimators using cross validation bandwidth $\hat{H}_{15}$ versus the fixed bandwidth k^* above. We see that in most cases, the estimator with fixed bandwidth k^* performs better than that of the cross validation bandwidth one, except for the case of Exp(10), Maxwell(20), and LogNormal(0,2). Furthermore, with this pre-deterministic k^*, it results in a significant improvement of calculation expense. In particular, it takes smaller amount of time to compute the estimator than computing the standard kernel base entropy estimator $\hat{H}_8$.

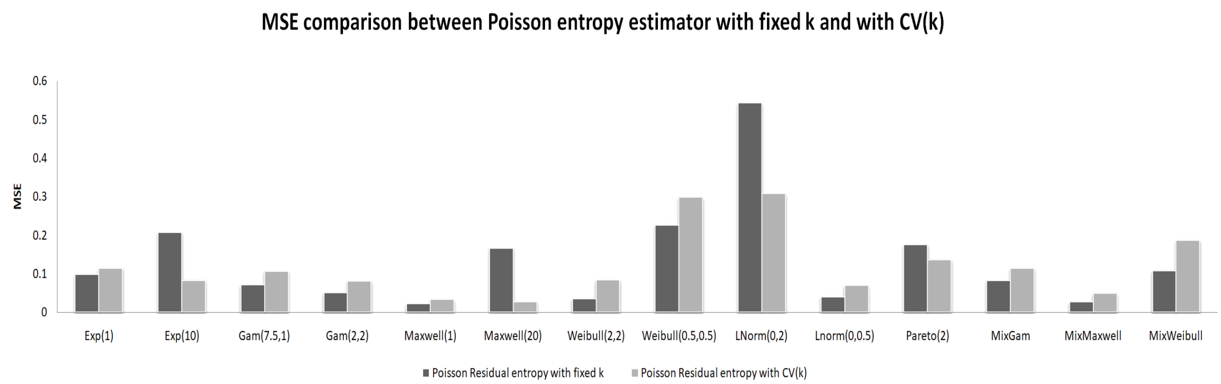

Figure 17: Entropy estimation MSE comparison of mean of log of Poisson density with different bandwidths. Here sample size is $n = 10$.

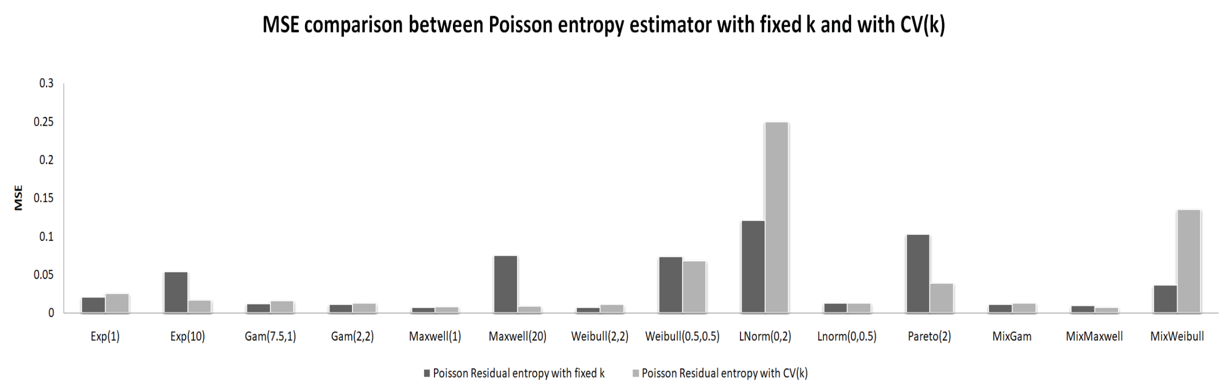

Figure 18: Entropy estimation MSE comparison of mean of log of Poisson density with different bandwidth. Here sample size is $n = 50$

4.2 Simulation study on residual entropy estimators

In this section, we want to see the performance of our proposal estimators for residual entropy function as well as comparing them with the ones suggested by Belzunce *et al.* (2001). Particularly, we still use the same testing set of distributions in the previous section, and the set of estimators are

$$\hat{H}_f^{Belzunce1}(t) = \log \hat{R}(t) - \frac{1}{\hat{R}(t)} \int_t^{\infty} \hat{f}_n^{Fixed}(x) \log \hat{f}_n^{Fixed}(x) dx \tag{4.17}$$

$$\hat{H}_f^{Belzunce2}(t) = \log \hat{R}(t) - \frac{1}{\hat{R}(t)} \sum_{i=1}^{n} R_K\left(\frac{t - X_i}{b}\right) \log \hat{f}_i^{Fixed}(X_i) \tag{4.18}$$

$$\hat{H}_f^{Plugin-Pois}(t) = \log \hat{R}^{Pois}(t) - \frac{1}{\hat{R}^{Pois}(t)} \int_t^{\infty} \hat{f}_n^{Pois}(x) \log \hat{f}_n^{Pois}(x) dx \tag{4.19}$$

$$\hat{H}_f^{Quantile}(t) = \log(\hat{R}^{Pois}(t)) - \frac{1}{\hat{R}^{Pois(t)}} \int_{\hat{F}(t)}^{1} \log \tilde{q}_n(p) dp, \tag{4.20}$$

where R_K is the survival function associated to the kernel function K (here we choose the normal density for K); the estimated survival function $\hat{R}(t)$ is computed based on the integration of $\hat{f}_n^{Fixed}(\cdot)$; the smoothing parameter k in $\hat{H}_f^{Plugin-Pois}(t)$ is set equal to $n^{2/5}+1$; and the parameter m in estimation of $\tilde{q}_n(\cdot)$ is set equal to $n/\log n$.
The setup of simulation study is as following. We present here both performance comparison by graphics and by mean integrated squared error (MISE). In order to produce the plots of residual entropy estimators, for each distribution, we run 500 replications of the sample size $n = 50$, then we compute the sample mean residual entropy estimator functions of these 500 replications. As mentioned in Belzunce et al (2001) that the behavior of residual entropy function is smooth up to certain bound because $R(t) \to 0$ very fast when $t \to \infty$, it is recommended to stop the estimation at the time $t^* = \inf\{t : R(t) \leq 0.01\}$. Therefore, the MISEs are just computed on the interval $[0, t^*]$ (for Pareto(2,1) is $[1, t^*]$). Similarly, the MISEs comparison between estimators are done based on 500 replications for sample size $n = 50$ and $n = 100$ for each distribution, but only 100 replications for sample size $n = 500$. The plots and tables below show the performance of our estimators and the ones in Belzunce *et al.* (2001).

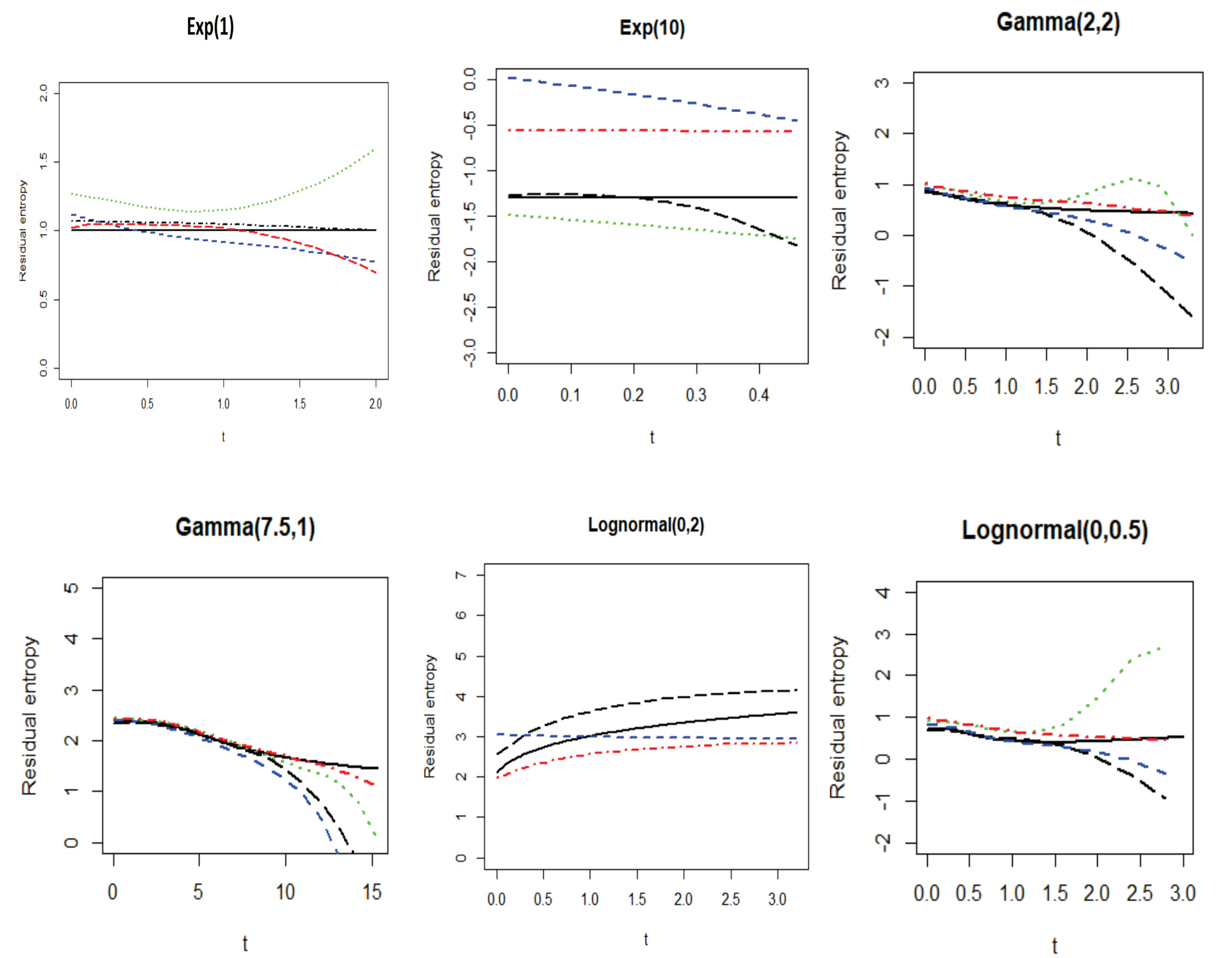

Figure 19: Plots of residual entropy estimators. The true function is in black solid line, $\hat{H}_f^{Plugin-Pois}(t)$ in red dotted-dashed line, $\hat{H}_f^{Quantile}(t)$ in black long dashed line, $\hat{H}^{Belzunce1}(t)$ in blue dashed line, and $\hat{H}_f^{Belzunce2}(t)$ in green dotted line. There are some plots without the green dotted line because the function is not in the range.

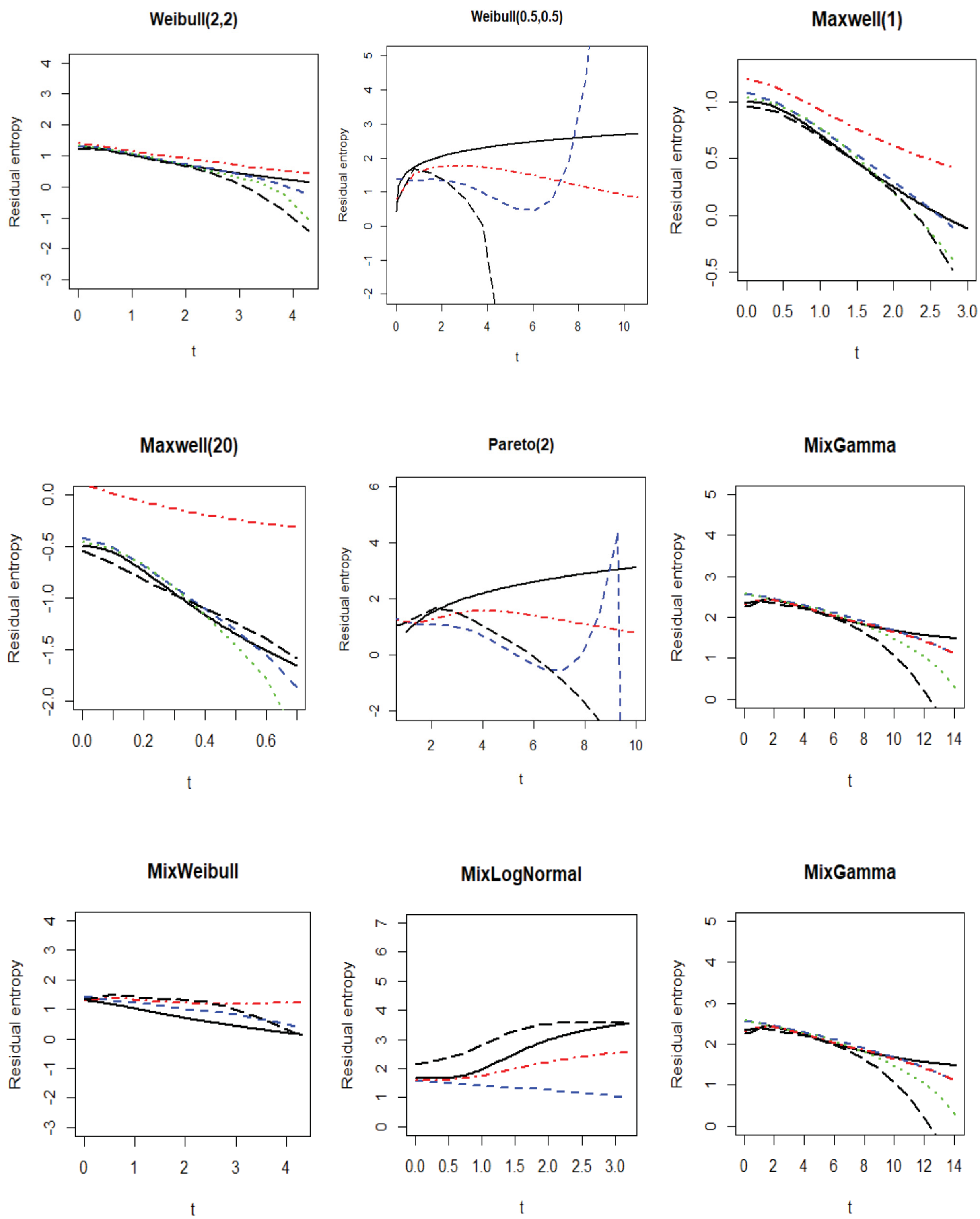

Figure 20: Plots of residual entropy estimators. The true function is in black solid line, $\hat{H}_f^{Plugin-Pois}(t)$ in red dotted-dashed line, $\hat{H}_f^{Quantile}(t)$ in black long dashed line, $\hat{H}^{Belzunce1}(t)$ in blue dashed line, and $\hat{H}_f^{Belzunce2}(t)$ in green dotted line. There are some plots without the green dotted line because the function is not in the range.

	Interval	Intergrated Mean Squared Error (IMSE) $\hat{H}^{Pois}(X,t)$	$\hat{H}_1^{Belzunce}(X,t)$	$\hat{H}^{Quantile}(X,t)$
Exp(1)	(0,4.605)	**0.608**	2.5816	9.8387
Exp(10)	(0,0.461)	0.2535	1.3814	**0.0719**
Gamma(7.5,1)	(0,15.289)	**0.9315**	1.5832	14.4769
Gamma(2,2)	(0,3.319)	**0.1993**	0.9969	3.1009
LogNormal(0,2)	(0,3.200)	2.2236	**0.9033**	2.9778
LogNormal(0,0.5)	(0,3.200)	**0.2631**	2.188	3.1081
Weibull(0.5,0.5)	(0,10.604)	**15.6378**	>10000	>10000
Weibull(2,2)	(0,4.292)	**0.2945**	0.3276	1.797
Maxwell(1)	(0,3.368)	0.4408	**0.1853**	0.6615
Maxwell(20)	(0,0.753)	0.7173	0.0379	**0.0367**
Pareto(2)	(1,10.024)	**19.6957**	6509.037	181.6876
MixGamma	(0,14.130)	1.1988	**0.8582**	17.3737
MixLogNormal	(0,3.200)	**2.5622**	3.2082	3.6732
MixWeibull	(0,4.292)	2.3112	**0.7388**	3.9595
MixMaxwell	(0,3.136)	0.365	**0.1949**	0.9866
Average time consumption (mins)		10-15	7-10	**3-5**

Table 10: MISEs of three estimators on 500 replications with sample size $n = 50$. The MISEs are computed on the indicated interval for each distribution.

	Interval	Intergrated Mean Squared Error (IMSE) $\hat{H}^{Pois}(X,t)$	$\hat{H}_1^{Belzunce}(X,t)$	$\hat{H}^{Quantile}(X,t)$
Exp(1)	(0,4.605)	**0.4124**	2.4588	4.5337
Exp(10)	(0,0.461)	0.1674	0.2396	**0.0362**
Gamma(7.5,1)	(0,15.289)	**0.6**	0.944	5.9302
Gamma(2,2)	(0,3.319)	**0.1505**	0.6284	1.5077
LogNormal(0,2)	(0,3.200)	**0.7871**	0.9182	1.7078
LogNormal(0,0.5)	(0,3.200)	**0.1826**	0.8065	1.4114
Weibull(0.5,0.5)	(0,10.604)	**9.5602**	40.3328	62.3692
Weibull(2,2)	(0,4.292)	0.2113	**0.2012**	0.7826
Maxwell(1)	(0,3.368)	0.3158	**0.1346**	0.3129
Maxwell(20)	(0,0.753)	0.5532	**0.03**	0.031
Pareto(2)	(1,10.024)	**13.5616**	26773.74	66.4542
MixGamma	(0,14.130)	**0.7175**	0.8547	6.9898
MixLogNormal	(0,3.200)	**1.0251**	2.8335	2.8198
MixWeibull	(0,4.292)	2.5913	**0.9241**	4.6132
MixMaxwell	(0,3.136)	**0.258**	0.2852	0.4339
Average time consumption (mins)		15-25	15-20	**5-10**

Table 11: MISEs of three estimators on 500 replications with sample size $n = 100$. The MISEs are computed on the indicated interval for each distribution.

	Interval	Intergrated Mean Squared Error (IMSE) $\hat{H}^{Pois}(X,t)$	$\hat{H}_1^{Belzunce}(X,t)$	$\hat{H}^{Quantile}(X,t)$
Exp(1)	(0,4.605)	**0.1013**	0.7366	0.4834
Exp(10)	(0,0.461)	0.0583	0.06	**0.0099**
Gamma(7.5,1)	(0,15.289)	**0.1677**	0.3434	0.5875
Gamma(2,2)	(0,3.319)	**0.0537**	0.1211	0.1674
LogNormal(0,2)	(0,3.200)	0.3641	0.7894	**0.2377**
LogNormal(0,0.5)	(0,3.200)	**0.0722**	0.1486	0.2194
Weibull(0.5,0.5)	(0,10.604)	**1.8837**	28.6434	3.6078
Weibull(2,2)	(0,4.292)	0.0843	**0.0597**	0.124
Maxwell(1)	(0,3.368)	0.1273	**0.0356**	0.0548
Maxwell(20)	(0,0.753)	0.2719	**0.0066**	0.0181
Pareto(2)	(1,10.024)	**3.3839**	>1000	3.4142
MixGamma	(0,14.130)	**0.1736**	0.8307	0.5792
MixLogNormal	(0,3.200)	**0.4343**	1.9141	0.5238
MixWeibull	(0,4.292)	3.02	**0.8817**	6.2564
MixMaxwell	(0,3.136)	0.103	0.1164	**0.0786**
Average time consumption (mins)		20-35	20-30	**10-15**

Table 12: MISEs of three estimators on 100 replications with sample size $n = 500$. The MISEs are computed on the indicated interval for each distribution.

To illustrate the rate of convergence of each estimators as the sample size increases, we also show the following graphs which are MISEs as a function of sample size for each distribution.

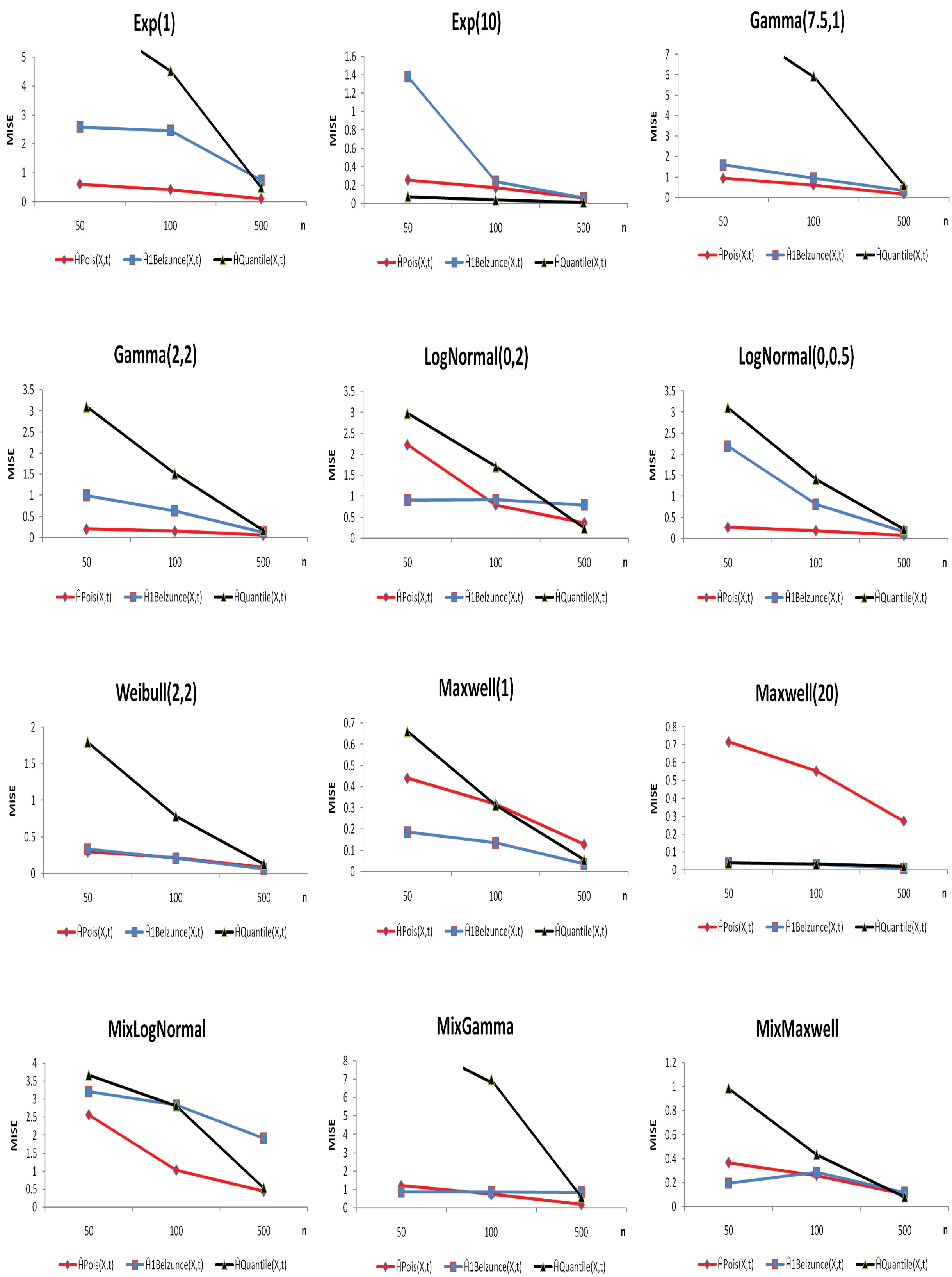

Figure 21: Residual entropy estimation MISE comparison with different sample size $n = 50, 100$ and 500. Where diamond-red is $\hat{H}_f^{Plugin-Pois}(t)$, square-blue is $\hat{H}_f^{Belzunce1}(t)$, and triangle-black is $\hat{H}_f^{Quantile}(t)$.

Remark 4.8. *$\hat{H}_f^{Belzunce2}(t)$ performs very poor in most cases.*

By the plot of residual entropy estimations from Figure 19-20 , we see that $\hat{H}_f^{Belzunce2}(t)$ is not as good as the others among all estimators. It is even out of the range in some distributions. Therefore, we can remove it out of our consideration.

Remark 4.9. *All estimators tend to diverge away from the true residual entropy over the time in most cases.*

We notice that most of estimators have a quite good startup as t is small. However, as the time t increase, in which the true survival function tends to zero rapidly, the estimators begin to diverge away from the true residual entropy. How large of the time t does this phenomenon happens depends on the true distribution.

Remark 4.10. *None of estimators outperforms the others in all cases. However, the $\hat{H}_f^{Plugin-Pois}(t)$ estimator seems to achieve a better estimation in most cases.*

If we just focus on the performance comparison between estimators, we see that in most cases, estimator $\hat{H}_f^{Plugin-Pois}(t)$ captures the trend of the true function throughout the plots and achieves better precision in terms of MISE shown in the Table 10-12. It is defeated by $\hat{H}_f^{Belzunce1}(t)$ only in Maxwell distributions which have the bell-normal shape. Although the quantile density entropy estimator $\hat{H}_f^{Quantile}(t)$ is the fastest one in terms of computation expense, the time consumption differences between estimators are quite small. Therefore, the $\hat{H}_f^{Plugin-Pois}(t)$ is recommended in a small sample size.

Remark 4.11. *The rate of convergence of the $\hat{H}_f^{Quantile}(t)$ estimator seems to be faster than other competitors.*

Similar to the entropy estimation, in this residual entropy estimation, the $\hat{H}_f^{Quantile}(t)$ estimator still has the fastest rate of convergence in most cases as the sample size increases. We admit that for a small sample size, $\hat{H}_f^{Quantile}(t)$ does not show up to be a good estimator, and it is not recommended. However, if the sample size is sufficiently large, it becomes one of the best potential choices for residual entropy estimator not only because of its fast rate of convergence, but also its speed in computation.

References

[1] Belzunce, F., Guillamon, A., Navarro, J., and Ruiz, J.M. (2001). Kernel estimation of residual entropy. *Communications in Statistics-Theory and Methods*, **30**, 1243–1255.

[2] Bouzebda, S., Elhattab, I., Keziou, A., and Lounis, T. (2013). New entropy estimator with an application to test of normality. *Communication in Statistics-Theory and Methods*, **42**, 2245–2270.

[3] Bouezmarni, T. and Scaillet, O. (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory*, **21**, 390–412.

[4] Bouezmarni, T. Ghouch, A.E, Mesfioui, M (2011). Gamma kernel estimators for density and hazard rate of right-censored data. *Journal of Probability and statistics*, **2011**, 1–16.

[5] Charpentier, A. and Flachaire, E. (2015). Log-transform kernel density estimation of income distribution. *L'Actualite Economique, Revue d'Analyse Economique*, **91**.

[6] Chaubey, Y. P., Mudholkar, G. S. and Smethurst, P. A. (1993). On entropy-based goodness-of-fit tests: a practical strategy. In *Probability and Statistics,* S. K. Basu and B. K. Sinha, eds., Narosa Publishing House, New Delhi, 116–120.

[7] Chaubey, Y.P. and Sen, P.K. (1996). On smooth estimation of survival and density functions. *Statistical Decisions*, **14**, 1–22.

[8] Chaubey, Y.P. and Sen, P.K. (1998). On smooth estimation of hazard and cumulative hazard functions. *Frontiers of Probability and Statistics,*, S.P. Mukherjee *et al.* (eds.) Narosa: New Delhi, 91–99.

[9] Chaubey, Y. P. and Sen, P. K. (2009). On the selection of the smoothing parameter in Poisson smoothing of histogram estimator. *Computational Aspects. Pakistan Journal of Statistics*, **25(4)**, 385–401.

[10] Chaubey, Y.P, Li, J., and Sen, P. (2010). Smooth density estimation for length-biased data. *Journal of The Indian Society of Agricultural Statistics*, **64(2)**, 145–155.

[11] Chaubey, Y.P., Li, J., Sen, A., and Sen, P.K. (2012). A New smooth density estimator for nonnegative random variables. *Journal of the Indian Statistical Association*, **50**, 83–104.

[12] Chen, S. X. (2000). Probability density functions estimation using Gamma kernels. *Annals of the Institute of Statistical Mathematics*, **52**, 471–480.

[13] Cheng, C. (1995). The Bernstein polynomial estimator of a smooth quantile function. *Statistics and Probability Letters*, **24(4)**, 321–330.

[14] Cheng, C. and Parzen, E. (1997). Unified estimators of smooth quantile and quantile density functions. *Journal of Statistical Planning and Inference*, **59**, 291–307.

[15] Correa, J.C. (1995). A new estimator of entropy. *Communications in Statistics: Theory and Methods*, **24**, 2439–2449.

[16] Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*, Wiley, New York.

[17] Dudewicz, E. J. and van der Meulen, E. C. (1981). Entropy-Based Tests of Uniformity. *Journal of the American Statistical Association*, **76(376)**, 967–974

[18] Ebrahimi, N., Habibullah, M., and Soofi, E. (1992). Testing exponentiality based on KullbackLeibler information. *Journal of the Royal Statistical Society: Series B*, **54**, 739–748.

[19] Ebrahimi, N. (1996). How to measure uncertainty in the life time distributions. *The Indian Journal of Statistics*. Series A **58**:48–56.

[20] Feller, W. (1966). Generalized moment generating functions of random variables and their probability density functions. *An Introduction to Probability Theory and Application.* Volume II, second edition, John Wiley and Sons, Inc. New York.

[21] Gasser, T. and Müller, H. (1979). Kernel estimation of regression functions. *Lecture Notes in Mathematics 757 Springer*, 23–68.

[22] Gawronski, W. and Stadmüller, U. (1980). On density estimation by means of Poisson's distribution. *Scandinavian Journal of Statistics*, **7(2)**, 90–94.

[23] Gawronski, W. and Stadmüller, U. (1981). Smoothing of histogram by means of lattice and continuous distributions. *Metrika*, **28**, 155–164.

[24] Hall, P. and Morton, S.C. (1993). On the estimation of entropy. *Annals of the Institute of Statistatical Mathematics*, **45**, 69–88.

[25] Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, **3**, 135–146.

[26] Jones, M.C. and Foster, P.J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, **6**, 1005–1013.

[27] Kapur, J. N. (1993). *Maximum Entropy Models in Science and Engineering.* John Wiley & Sons, Inc., New York, USA.

[28] Kapur, J. N., Kesavan, H. K. (1992). *Entropy optimization principles with applications.* Academic press, Inc., New York.

[29] Koekemoer, G. and Swanepoel, J.W.H. (September 2008). Transformation kernel density estimation with applications. *Journal of Computational and Graphical Statistics*, **17**, 750–769.

[30] Noughabi, H.A. (2010). A new estimator of entropy and its application in testing normality. *Journal of Statistical Computation And Simulation*, **80**, 1151–1162.

[31] Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Computation*, **15**, 1191–1253.

[32] Schuster, E. (1958). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics, Part A - Theory and Methods*, **14**, 1123–1136.

[33] Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, **27**, 379-423. Reprinted in C. E . Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

[34] Van Es, B. (1992) Estimating functional related to a density by a class of statistics based on spacings. *Scandinavian Journal of Statistics*, **19**, 61–72.

[35] Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society*, **38**, 54–59.

[36] Wieczorkowski, R. and Grzegorzewski, P. (1999). Entropy estimators improvements and comparisons. *Communications in Statistics-Simulation and Computation*, **28(2)**, 541–567.