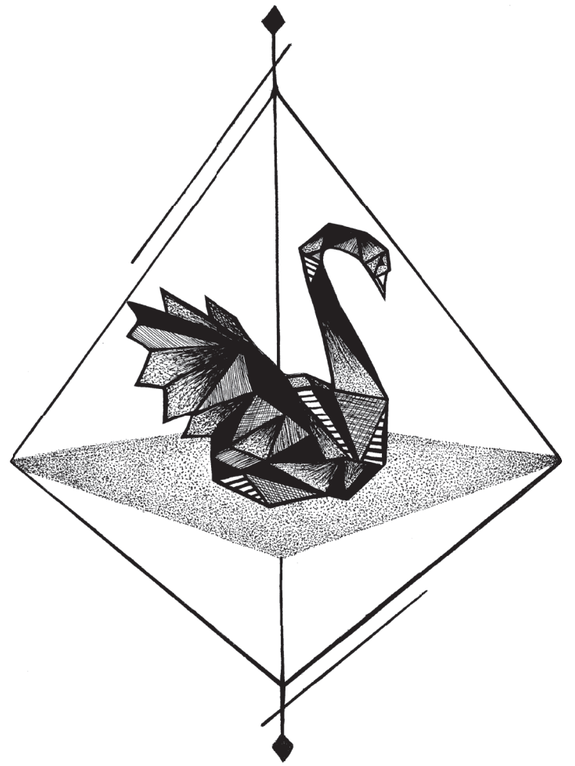
◆ K. Bisewski ◆ RARE EVENT SIMULATION AND TIME DISCRETIZATION ◆

RARE EVENT SIMULATION AND TIME DISCRETIZATION

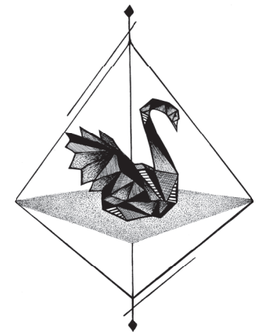


Krzysztof Bisewski

INVITATION

You are cordially invited to
the public defense of
my doctoral thesis

RARE EVENT SIMULATION
AND TIME DISCRETIZATION



On Tuesday
15 October 2019 at 12:00

In the Agnietenkapel
Oudezijds Voorburgwal 229-231
1012 EZ Amsterdam

Krzysztof Bisewski
kbisewski@gmail.com

Rare Event Simulation and Time Discretization

by

Krzysztof Bisewski

This work is part of the 'Mathematics of Planet Earth' program, funded by the Netherlands Organization for Scientific Research (NWO). The research work was carried out in the Scientific Computing group of the Centrum Wiskunde & Informatica (CWI), the Dutch national research institute for mathematics and computer science.

Rare Event Simulation and Time Discretization

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof.dr.ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde

commissie, in het openbaar te verdedigen in de Agnietenkapel

op dinsdag 15 oktober 2019, te 12:00 uur door

Krzysztof Łukasz Bisewski

geboren te Gdynia

Promotiecommissie

Promotores:

prof.dr. D.T. Crommelin	Universiteit van Amsterdam
prof.dr. M.R.H. Mandjes	Universiteit van Amsterdam

Copromotor:

prof.dr. J.H. van Zanten	Universiteit van Amsterdam

Overige leden:

prof.dr. P.G. Bolhuis	Universiteit van Amsterdam
dr. J.L. Dorsman	Universiteit van Amsterdam
prof.dr.ir. J.E. Frank	Universiteit Utrecht
prof.dr. R. Núñez Queija	Universiteit van Amsterdam
dr. A.A.N. Ridder	Vrije Universiteit Amsterdam
prof.dr. A.P. Zwart	Technische Universiteit Eindhoven

Faculteit:

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

To my parents,

Jolanta and Robert

Acknowledgements

To begin with, I would like to thank my supervisors Daan Crommelin and Michel Mandjes for guiding me through the entire PhD. You have given me a lot of freedom to pursue topics interesting to me while keeping me organized. You certainly taught me a whole lot about doing research and I feel very lucky for having you as my supervisors. Furthermore, I thank Harry van Zanten for his co-supervision in the early stages of my PhD.

I would like to acknowledge Jevgenijs Ivanovs for collaboration which lead to Chapter 3 of this thesis. I really enjoyed working together with you Jev and I hope we will have a chance to collaborate more in the future!

Special thanks go to Nick Verheul for translating the summary of this thesis to Dutch on a very short notice — you're awesome. It was great having you around in the office talking about games and feelings.

Jolusia, thank you designing and drawing the cover of this thesis. Even more importantly, thank you for taking care of me and your patience during these final, nervous stages of my PhD.

Dziękuję całej mojej rodzinie za wsparcie. W szczególności dziękuję rodzicom za zaufanie — od najmłodszych lat daliście mi wolność do podróżowania własną ścieżką. Gdyby nie wasza wiara w moje możliwości, myślę, że nie doszłoby w ogóle do tego doktoratu.

Dziękuję ekipie moich przyjaciół z Rumi: Adam, Grzesiu, Maciek (Hamer), Marta, Monika, Piotrek (Sąsiad), Szymon (Shimi), Witek i Żaneta. Doceniam to, że zawsze potrafiliście znaleźć dla mnie czas, gdy wracalem do domu. Bardzo ważne jest dla mnie abyśmy nie urwali kontaktu. Dziękuję również górskiej ekipie: Andrzejowi, Filipowi (Młodemu), Grześkowi i Tomkowi za czas spędzony razem na szczytach (i przy okazji wspólne robienie zadanek) — mam nadzieję, że przed nami czeka jeszcze wiele wspólnych wyjazdów. Wiktor, dziękuję za twoje wizyty w Diemen — twoja pozytywność jest zaraźliwa!

Ki Wai, Prashant and Sangeetika, thank you for all the time we spent together — for the times we went out and for the times we stayed late at work eating takeaway food and drinking whisky. Thanks to the Squash Team: Alvaro, Paco, Prashant and Richard for keeping me in quasi-shape and for all the fun *Zatte* evenings we had. Prashant, we spent so much time together during the last two years; it was great having you around.

Furthermore, I would like to thank all my colleagues for creating a friendly atmosphere at work: Alessandro, Anastasia, Andrea, Anne, Ashutosh, Barry, Bart, Beatriz, Behnaz, Benjamin, Bohan, Brendan, Chang-Han, Dmitrii, Evelien, Fredrik, Gabriele, Hema, Jurriaan, Kees, Kristoffer, Laurent, Luis, Mandar, Marco, Mariska, Masha, Michael, Mihail, Nicos, Pablo, Qian, Rakesh, Sirshendu, Svetlana, Tommaso, Valerio, Willem Jan, Wouter, Yous and Zhichao. Also, thanks to Bikkie, Duda, Evelien and Nada for always being helpful and teaching me Dutch (Bikkie).

My dear office mate Debarati, you took care of me since day one. It's mostly because of you that I enjoyed working at CWI so much (and maybe also a little bit because of math). I honestly can't imagine how these four years would have been without you — well, certainly much more dull. Thank you for being there with me.

Bence, Kuba and Sandra, thank you for being my friends. Sandra, your friendship means a lot to me. Just knowing that you're somewhere out there was enough to make me feel better during some lonely moments. Kuba, our e-mail correspondence helped me go through some rough patches during my stay in the Netherlands. Bence (and also you, Ildikó!), I could always count on you to listen to me and feed me whenever I felt down.

Guido and Teresa, you are special to me in the sense that you're both my dear friends and math-related academics at the same time. Having more experience you could always relate to me — you had a fair share of shaping my view on academia and life in general. Teresa, your willingness to listen is indispensable, thank you for sharing your insight with me. Guido, thank you for all the mantime with had together, I really needed that. I'm lucky to have you both as my friends.

Summary

Rare Event Simulation and Time Discretization

Rare events are of great importance in many areas of application. Particularly, when they have a potentially catastrophic impact, there is a clear need to accurately estimate their likelihood. Examples include earthquakes, extreme weather conditions, bankruptcy of an insurance company, simultaneous failure of multiple components of a machine, etc.

The probability of such an event usually can not be expressed with a mathematical formula due to the complex nature of the underlying system and is typically estimated via simulation. In this thesis we consider possible errors arising while estimating rare event probabilities.

The first, and perhaps, most widely known problem is the presence of

(i) *sampling error*.

The sampling error can arise due to a finite number of samples, potentially resulting in a large variance of an estimator. Due to the very nature of the rare event, one has to simulate the system many times before observing even one relevant occurrence (and even more times to estimate its probability accurately). When simulation of the system is computationally expensive, this task becomes practically impossible.

When the underlying stochastic system is time-continuous, there is an evident need for time discretization, as a whole continuous trajectory can not be simulated and stored on a finite computer. This can give rise to two different sources of error:

(ii) *discretization error*; and

(iii) *detection error*.

Discretization error arises due to non-exact (approximated) sampling of the process on the grid. For example, processes driven by Stochastic Differential

Equations, need to be simulated via approximated numerical schemes (such as Euler or Runge-Kutta methods) because their exact solution is typically unknown.

While certain time-continuous stochastic processes such as Brownian Motion, Gaussian or Lévy processes can be simulated exactly (with no discretization error), they still need to be simulated over a finite grid. This loss of information between observation times leads to the detection error. The detection error can arise, for example, in first passage time problems, where one might fail to observe the passage while looking only on the discrete grid.

In Chapter 1 of this thesis, we introduce the errors (i)-(iii) above in greater detail. We also lay out preliminaries for the remaining four chapters.

In Chapter 2 we study the first hitting time of a large threshold for a standard Brownian Motion. As it turns out, the combination of sampling and detection error leads to an extremely inefficient Monte Carlo estimator. Not only do we need more and more independent samples to countereffect the growing sampling error, as the threshold grows large, but also a finer and finer discretization grid in order to deal with the detection error. We mitigate this effect by a particular (but explicit) placement of the observation points in combination with a rare event simulation algorithm. This concept can be generalized to a broad class of stochastic processes and is empirically shown to outperform the uniform discretization grid.

In Chapter 3 we consider the same setting as in Chapter 2, but for a general Lévy process. Here, we derive the exact asymptotics of the detection error, as the number of observation points on the uniform grid tends to infinity. This limit result can be used in many areas of application to provide correction terms, for example, in finance, it can be used to price barrier options more accurately.

In Chapter 4 we study the discretization error for the estimation of the tail of the stationary distribution of a Stochastic Differential Equation. Interestingly, we show, both theoretically and numerically, that various discretization schemes (e.g. Euler, Milstein) can lead to completely different tail behavior as compared with the real, continuous-time process — fat tails may become light and vice versa. We develop a theoretical tool that, roughly speaking, determines the 'fatness' of the tail of the stationary distribution.

In Chapter 5 we focus exclusively on the sampling error and devise an appli-

cable algorithm for the estimation of tail probabilities for the stationary distribution of a Stochastic Differential Equation. The algorithm is straightforward to implement and does not require detailed knowledge of the system — it can be even applied to so called black-box models. The method relies on the Markov chain's underlying recurrent structure (a concept akin to regeneration) in combination with the Multilevel Splitting method. Under certain assumptions, the algorithm is shown to be asymptotically optimal. We successfully apply this algorithm to a four-dimensional non-linear SDE, which has features of a climate model.

Samenvatting

Rare Event Simulatie en Tijd Discretisatie

Zeldzame gebeurtenissen zijn van groot belang in allerlei toepassingsgebieden. Er is een duidelijke noodzaak om hun waarschijnlijkheid nauwkeurig te schatten, vooral wanneer ze een mogelijk catastrofale impact kunnen hebben. Voorbeelden zijn aardbevingen, extreme weersomstandigheden, faillissement van een verzekeringsmaatschappij, gelijktijdig falen van meerdere componenten van een machine, enzovoort.

De kansen op dergelijke gebeurtenissen kunnen meestal niet uitgedrukt worden in een wiskundige formule vanwege de complexiteit van het onderliggende systeem. Ze worden doorgaans geschat met behulp van simulaties. In dit proefschrift beschouwen we de mogelijke fouten die voorkomen bij het schatten van kansen van zeldzame gebeurtenissen.

Het eerste, en mogelijk bekendste, probleem is het optreden van

(i) *sampling fouten.*

Sampling fouten doen zich voor bij een eindig aantal samples, en kunnen resulteren in een hoge variantie van de schatter. Omdat het gaat om gebeurtenissen die zeldzaam zijn moet men het systeem veelvuldig simuleren voordat zo'n gebeurtenis zich voordoet (en nog veel vaker om de kans nauwkeurig te kunnen schatten). Dit wordt vrijwel onmogelijk als simulaties van het systeem computationeel duur zijn.

Wanneer het onderliggende stochastische systeem continu in de tijd is, is het noodzakelijk om de tijd te discretiseren, omdat een volledig continue pad niet gesimuleerd en opgeslagen kan worden op een eindige computer. Dit kan twee soorten fouten veroorzaken:

(ii) *discretisatiefouten*, en

(iii) *detectiefouten.*

Discretisatiefouten ontstaan door niet-exacte (benaderde) sampling van het proces op een grid. Bijvoorbeeld, processen aangedreven door stochastische differentiaalvergelijkingen moeten gesimuleerd worden via numerieke benaderingen (zoals Euler of Runge-Kutta) omdat hun exacte oplossing over het algemeen niet bekend is.

Alhoewel sommige stochastische processen die continu zijn in de tijd, zoals Brownse beweging en Gaussische of Lévy processen, exact gesimuleerd kunnen worden (zonder discretisatiefout), moeten ze nog steeds gesimuleerd worden over een eindig grid. Het verlies van informatie tussen de observatiemomenten leidt tot detectiefouten. Detectiefouten kunnen bijvoorbeeld optreden bij zogeheten *first passage time* problemen, waar het moment van eerste overschrijding (*first passage*) mogelijk niet waargenomen wordt op het discrete grid.

In hoofdstuk 1 van dit proefschrift introduceren we bovenstaande fouten (i)-(iii) in meer detail. We presenteren hier ook basisbegrippen voor de resterende vier hoofdstukken.

In hoofdstuk 2 bestuderen we de eerste raaktijd van een hoge drempelwaarde voor standaard Brownse beweging. Het blijkt dat de combinatie van sampling fout en detectiefout leidt tot een uiterst inefficiente Monte Carlo schatter. Als de drempelwaarde groter wordt hebben we een toenemend aantal onafhankelijke samples nodig om de sampling fout niet te laten groeien. Daarbovenop hebben we ook een hogere resolutie van het discretisatie grid nodig om de detectiefout niet groter te laten worden. We doen deze effecten deels teniet met een specifieke (maar expliciete) keuze van de observatiepunten in combinatie met een rare event simulatie algoritme. Dit concept kan gegeneraliseerd worden tot een brede klasse van stochastische processen. We tonen empirisch aan dat het beter presteert dan een uniform discretisatiegrid.

In Hoofdstuk 3 beschouwen we dezelfde configuratie als in hoofdstuk2, maar dan voor een algemeen Lévy proces. Hier leiden we het exacte asymptotische gedrag af van de detectiefout onder de conditie dat het aantal observatiepunten op het uniforme grid naar oneindig gaat. Dit limietresultaat kan gebruikt worden voor het vinden van correctietermen in allerlei toepassingsgebieden. Het kan bijvoorbeeld in de financiële wereld gebruikt worden om de prijs van barrieropties nauwkeuriger te bepalen.

In hoofdstuk 4 bestuderen we de discretisatiefout voor het schatten van

de staart van de stationaire verdeling van een stochastische differentiaalvergelijking. We laten zien, zowel theoretisch als numeriek, dat verschillende discretisatieschema's (bijvoorbeeld Euler, Milstein) kunnen leiden tot totaal verschillend staartgedrag vergeleken met het echte continue-tijds proces: dikke staarten kunnen licht worden en vice versa. We ontwikkelen hier een theoretisch hulpmiddel dat, kort gezegd, bepaalt wat de 'dikte' is van de staart van de stationaire verdeling.

In hoofdstuk 5 richten we ons geheel op de samplingfout en ontwikkelen we een algoritme voor het schatten van kansen in de staart van de stationaire verdeling van een stochastische differentiaalvergelijking. Het algoritme is eenvoudig te implementeren en vereist geen gedetailleerde kennis van het systeem — het kan zelfs toegepast worden op zogenaamde black-box modellen. De methode maakt gebruik van de onderliggende recurrente structuur van de Markov-keten (een concept vergelijkbaar met regeneratieve processen) in combinatie met de multi-level splitting methode. We laten zien dat dit algoritme onder bepaalde voorwaarden asymptotisch optimaal is. We passen het algoritme met succes toe op een vierdimensionaal niet-lineaire stochastische differentiaalvergelijking die eigenschappen deelt met een klimaatmodel.

Contents

Chapter 1

Introduction

This thesis treats the problem of efficient and accurate estimation of *rare event* probabilities. Such events, by definition, occur very infrequently; depending on the application this could mean that they have a probability of order 10^{-3} (e.g. in climate science) or even 10^{-7} and less (e.g. in telecommunication networks). These extreme events usually represent some undesirable or even catastrophic phenomena. In climate science examples include floods and extreme weather conditions such as heatwaves or and high precipitation. Other examples include earthquakes, bankruptcy of an insurance company, electric grid power overload or fuel starvation in air traffic control. Given the possible aftermath of such events, there is a clear motivation to accurately estimate their likelihood.

Formally, we are interested in a general problem of estimating

$$\gamma := \mathbb{P}(X \in S), \tag{1.1}$$

where X is the stochastic process of interest and S the rare event. Mathematical models for X used in practice are often extremely complex (such as models used for climate predictions) and in that case finding a formula for γ is out of reach.

Here is when simulation becomes an indispensable tool, as it makes it possible to statistically estimate values of interest, often without the need of theoretical study of the mathematical properties of the system. Nevertheless, *rare event simulation*, that is simulation methods designed for estimation of rare event probabilities comes with errors and these errors are the topic of this thesis. We distinguish three sources of error — sampling, discretization and detection error, which are described in the following three sections of this chapter.

Organization of this thesis. For the remainder of this introductory chapter we introduce possible sources of error in rare event simulation. Each of the four following chapters treats at least one of these errors. Individual chapters can be read stand-alone and no uniform notation is pursued.

1 Sampling error

Sampling error is an error that arises due to intrinsic randomness in statistical estimation. In this thesis it will be synonymous with squared relative error defined in (1.4) below.

When simulation of the stochastic system X is available there is a straightforward way to estimate probability γ: one could draw many independent samples of X and take the average number of times, the event S occurred. This so-called naïve *Monte Carlo* estimator is formally defined as

$$\widehat{\gamma}^{\mathrm{MC}} := \frac{1}{N} \sum_{i=1}^{N} 1_{\{X_i \in S\}}, \tag{1.2}$$

where $X_1, X_2, \ldots$ are N independent statistical copies of X. The Law of Large Numbers guarantees that $\widehat{\gamma}^{\mathrm{MC}}$ converges to γ as N grows large; but how large should N be in order to obtain a reasonably accurate estimator? This question can be answered with help of the Central Limit Theorem: when N is large, $\widehat{\gamma}^{\mathrm{MC}}$ is approximately normally distributed,

$$\frac{\widehat{\gamma}^{\mathrm{MC}} - \gamma}{\gamma} \sim \mathcal{N}\Big(0, \mathrm{RE}^2\big(\widehat{\gamma}^{\mathrm{MC}}\big)\Big), \tag{1.3}$$

where $\mathrm{RE}^2(\cdot)$ is the *squared relative error* defined as

$$\mathrm{RE}^2(\widehat{\gamma}) := \frac{\mathbb{V}\mathrm{ar}(\widehat{\gamma})}{(\mathbb{E}\widehat{\gamma})^2} \tag{1.4}$$

for any estimator $\widehat{\gamma}$. In Eq. (1.3) we focus on the *relative* difference between $\widehat{\gamma}^{\mathrm{MC}}$ and γ, because in the rare event context γ is very small and so the absolute error is expected to be close to 0 making it not a meaningful measure of accuracy.

Now say we want our Monte Carlo estimator to be within 10% distance from the ground truth with confidence at least 95%; formally we require

$$\mathbb{P}\Big(\tfrac{|\widehat{\gamma}^{\mathrm{MC}}-\gamma|}{\gamma} < 10\%\Big) \geq 95\%. \tag{1.5}$$

Using the normal approximation in Eq. (1.3) we see that we need an estimator with relative error satisfying $\mathrm{RE}(\widehat{\gamma}) \cdot z_{0.025} = 10\%$, where z_α is the α-quantile of the normal distribution. Using the fact that $\mathrm{RE}^2\big(\widehat{\gamma}^{\mathrm{MC}}\big) = \frac{1-\gamma}{\gamma N}$ and consulting a standard normal table, we finally see that approximately $N = 400\gamma^{-1}$ independent samples are needed.

This result underlines the extreme inefficiency of the Monte Carlo method in rare event setting: a reliable MC estimator must be based on a sample size at least two orders of magnitude larger than γ^{-1}. When γ is small, this becomes a virtually impossible task, especially so, when the simulation of X is computationally costly. This gives a clear motivation for devising algorithms more efficient than Monte Carlo.

While the relative error is an important indicator of the *accuracy* of the estimator $\widehat{\gamma}$ (as it determines the size of confidence intervals), it doesn't say anything about its *efficiency* of the algorithm that produced $\widehat{\gamma}$. Let $W(\widehat{\gamma})$ denote the computational time needed to obtain an estimate $\widehat{\gamma}$. We define the *work-normalized* squared relative error (or relative time variance product)

$$\mathrm{RTV}(\widehat{\gamma}) = \mathrm{RE}^2(\widehat{\gamma})W(\widehat{\gamma}). \tag{1.6}$$

Now notice that the value of $\mathrm{RTV}(\widehat{\gamma}^{\mathrm{MC}})$ is independent of sample size N. Among two estimators we will prefer the one with lower RTV, as it takes less computational time to reach the same level of accuracy.

The field of rare event simulation aims at devising algorithms to decrease

the sampling error of the naïve Monte Carlo method; they also belong to a larger class of algorithms called *variance reduction* techniques. Preferably, they should have small RTV which doesn't grow too fast, as $\gamma \to 0$. We distinguish two notions of asymptotic efficiency of rare event simulation algorithms. We say that an estimator is (i) *strongly efficient* if its work-normalized squared relative error is bounded as $\gamma \to 0$, i.e. $\mathrm{RTV}(\widehat{\gamma}) = O(1)$ and (ii) *logarithmically efficient* when it grows slower than any negative power of γ, i.e. $\mathrm{RTV}(\widehat{\gamma}) = o(\gamma^{-\varepsilon})$ for any $\varepsilon > 0$. In Chapter 2 and Chapter 5 we introduce algorithms which satisfy these notions of efficiency. For more details and other notions of asymptotic efficiency in rare event simulation the reader is referred to a textbook on the topic [70].

We can see that RTV of the naïve Monte Carlo estimator in Eq. (1.2) is proportional to γ^{-1}, which makes it neither strongly, nor logarithmically efficient. The growth in efficiency of using some logarithmically efficient algorithm in place of Monte Carlo is difficult to exaggerate. For instance, in Example 1.1 below we will see a thousandfold growth in efficiency. There are many techniques, which can be proven to satisfy the above notions of efficiency, to name a few: importance Sampling, using control variates, stratified sampling, and particle methods such as multilevel splitting, RESTART, genealogical particle analysis. There are several monographs on this topic, e.g. [8, 70, 85].

Rare event simulation methods have been successfully implemented in nuclear physics, to derive the probability that a particle is transmitted through a shield (pioneering work in the field by Kahn and Harris [65]), in telecommunication networks to derive a Packet Loss Ratio [51], in air traffic control to find the chance of fuel starvation, in molecular dynamics and computational chemistry to find a chance of rare transitions. In electrical power grids studies it was applied to find the chance of connection overload [97], and in climate science to find the probability of an extreme heatwave [81]. This is still a very active field of research both on theoretical and practical frontiers.

In this thesis we pay special attention to *Multilevel Splitting* (MLS), as it is a topic of Chapter 5. This method can be used in the context of stochastic processes evolving in time, in particular to estimate probabilities of hitting times:

$$\mathbb{P}(\tau_B < T), \tag{1.7}$$

where $\tau_B := \inf\{t > 0 : X_t \in B\}$ is the first time process X enters set B and T is an arbitrary stopping time (e.g. T might be deterministic or $T = \tau_A$ for some set A). Due the intrinsic randomness in Monte Carlo simulation, a lot of computational effort is wasted on simulating irrelevant paths (only a small fraction of paths actually reaches the set B). The idea behind the Multilevel Splitting method is to reinforce the 'good behavior' of the path, which is accomplished by *splitting* it in a number of copies whenever it approaches B. This way, we have more control over the simulation, by forcing the process into interesting regions. An illustration of the algorithm is included in Figure 1.1. MLS is, under certain regularity assumptions, proven to be logarithmically efficient. The method is formally introduced in Chapter 5, where also the choice of intermediate sets is discussed in more detail. The interested reader is referred to the following works [51, 74].

To give the reader some idea about the sampling error we introduce the following illustrative example.

Example 1.1. Let $X = (X_t)_{t\geq 0}$ be a standard Ornstein-Uhlenbeck (OU) process, i.e. a solution to the Stochastic Differential Equation:

$$\mathrm{d}X_t = \theta(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{1.8}$$

with parameters $\mu \in \mathbb{R}$, and $\theta, \sigma > 0$, and initial condition $X_0 = x_0$, where $(W_t)_{t\geq 0}$ is a standard Wiener process. We are interested in the probability of *upcrossing* a large threshold u before *downcrossing* l, where $l < x_0 < u$, that is

$$\gamma = \mathbb{P}(\tau_u^{\uparrow} < \tau_l^{\downarrow}), \tag{1.9}$$

where $\tau_u^{\uparrow} := \inf\{t > 0 : X_t > u\}$ and $\tau_l^{\downarrow} := \inf\{t > 0 : X_t < l\}$. In our example we put $\mu = x_0 = 0$, $\theta = 1$, $\sigma = 1$ and thresholds $u = 4$, $l = -1$. Now, we want to estimate γ using Monte Carlo method described in (1.2) and we choose N to satisfy (1.5). The exact mathematical expression for γ is known, $\gamma \approx 1.27 \cdot 10^{-6}$, and we will use it as reference.

This experiment took over 16 minutes to on my standard personal computer. For reference, an MLS estimator for with the same accuracy executed in half of a second making it over 2000 times (sic) faster. This tremendous growth in efficiency is perhaps surprising until one fully realizes how extremely inefficient

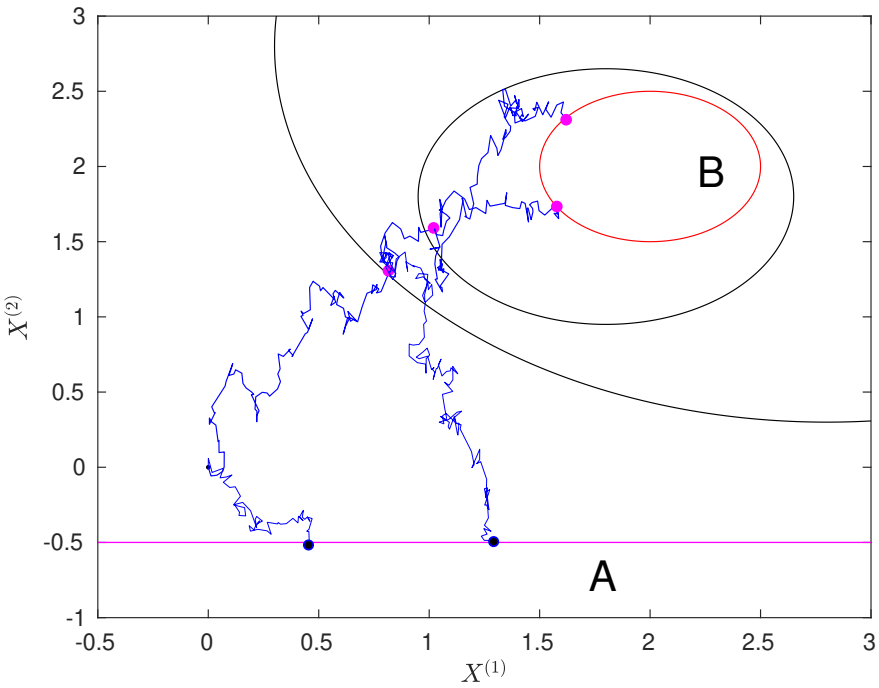


Figure 1.1: An illustration of the mechanics behind the Multilevel Splitting algorithm. The rare event of interest is entering set B before A. The path starts at $X_0 = \mathbf{0}$ and it is split in two independent copies upon hitting one of the intermediate sets (with black boundary).

Monte Carlo method is.

Chapters 2 and 5 of this thesis directly relate to the sampling error theme. In Chapter 2 we introduce a strongly efficient algorithm for the estimation of the tail of the supremum of Brownian Motion over a finite time interval. It relies on the simulation of Brownian Motion on a particular grid (to decrease the detection error discussed below in Section 3) and application of the algorithm developed in [3]. In Chapter 5 we devise an algorithm for the estimation of tail probabilities of the stationary distribution of a multidimensional Stochastic Differential Equation. A new framework is developed, which makes it possible to apply rare event simulation algorithms for this problem — note that it is not clear how to apply such methods, as the probability in question is not of the form (1.7). The algorithm is straightforward to implement and does not require

detailed knowledge of the system — it can be even applied to so called black-box models. We are not aware of any other such widely applicable algorithm in this context.

Throughout this section we tacitly assumed that the estimator $\widehat{\gamma}$ is unbiased (that is $\mathbb{E}\widehat{\gamma} = \gamma$). This is often not the case and in the following two sections we concentrate on two different sources of bias, which can arise when the underlying stochastic system is time-continuous. We will see that in the rare event simulation setting they are often as troublesome as sampling error.

2 Discretization error

Discretization error can arise in simulation of time-continuous stochastic processes. It is evidently impossible to simulate and store the entire continuous path of a process on the computer (due to the infinite time and memory requirement) and thus one must sample the path of the process on a discrete time grid. One typically chooses a certain *time step* $h > 0$ and simulates and stores the system on the grid $0, h, 2h, \ldots$. Some stochastic systems such as Gaussian, Lévy processes or finite state space Markov chains can be simulated *exactly* on a given finite grid. This means that the vector $(X_0, X_h, \ldots, X_{Nh})$ can be sampled, without an error, on a computer. Unfortunately, exact simulation is often unavailable. A large class of stochastic processes driven by Stochastic Differential Equations (SDEs) does not enjoy that property. In that case, one is forced to sample from a distribution that only approximates $(X_0, X_h, \ldots, X_{Nh})$ inducing a *discretization error* for the estimation of γ.

In this thesis we consider discretization error arising from numerical approximation of SDEs. Let $X = (X_t)_{t\geq 0}$ be driven by the following d-dimensional SDE:

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t \tag{1.10}$$

with some initial condition X_0. The functions $f, g : \mathbb{R}^d \to \mathbb{R}^d$ satisfy certain regularity assumptions so that (1.10) has a unique solution and W is a d-dimensional Wiener process. In order to simulate a single (approximate) trajectory of X, we choose a time step $h > 0$ and define the following Euler discretization scheme: let $X^h = (X^h_t)_{t\geq 0}$ be defined on the grid $0, h, 2h, \ldots$ through

the following recursion:

$$X^h_{h(n+1)} = X^h_{hn} + f(X^h_{hn})h + g(X^h_{hn})\Delta W^h_n, \tag{1.11}$$

where $\Delta W^h_n := W_{h(n+1)} - W_{hn}$ is an increment of a Wiener process and for $t \geq 0$ not lying on the grid $0, h, 2h, \ldots$ we put $X^h_t := X^h_{h[t/h]}$ so that X^h is piecewise constant and right-continuous. This is arguably the most straightforward discretization method; in practice one could use more sophisticated but also more computationally demanding methods. For a standard texbook on discretization methods for SDEs we refer to [69]. Higham [62] gives a very good non-technical introduction to numerical simulation of SDEs and lays out main issues with the discretization error.

Example 1.1 (continued)**.** In the previous section, we did not discuss how a single path of an Ornestein-Uhlenbeck was generated. We can do so using the Euler method given by the recursion in (1.11) and obtain

$$X^h_{(n+1)h} := X^h_{nh} + \theta(\mu - X^h_{nh})h + \sqrt{h}\sigma Z_n, \tag{1.12}$$

where $Z_1, Z_2, \ldots$ is an iid sequence of standard normal random variables. However, as was already mentioned, (1.12) is only an approximation of the real path driven by (1.8). The Ornstein-Uhlenbeck process is one of few diffusions whose closed form solutions are known; its single trajectory satisfies the following recursion

$$X_{(n+1)h} := X_{nh} + (\mu - X_{nh})(1 - e^{-\theta h}) + \frac{\sigma\sqrt{1 - e^{-2\theta h}}}{\sqrt{2\theta}} \cdot Z_n, \tag{1.13}$$

where $Z_1, Z_2, \ldots$ are as in (1.12). It is clear that (1.12) is not equivalent to (1.13) however, as $h \to 0$, it is getting closer and closer to the ground truth.

Now, let's go back to the problem of estimation of γ defined in (1.9) with the same parameters, as in the first part of this example on page 5. The numerical results are shown in Tab. 1.1. As expected, the discretization error decreases, as $h \to 0$. The Euler approximation (1.12) always overestimates the probability γ, as compared to the exact solution (1.13) for all values of h considered. This is most likely due to a slightly larger variance in each step of the Euler scheme making it 'easier' for system (1.12) to reach high values. Referring back to

h	10^{-1}	$5 \cdot 10^{-2}$	10^{-2}	$5 \cdot 10^{-3}$	10^{-3}
$\gamma^h_{\text{euler}}/\gamma^h_{\text{exact}}$	209%	144%	108%	104%	100%

Table 1.1: Size of the discretization error depending on the value of time step h for the problem described in the second part of Example 1.1 on page 8. The simulation results are accurate up to $\pm 1\%$ with 95% confidence.

Section 1, it is worth noting that the numbers in Tab. 1.1 were obtained via Multilevel Splitting algorithm. Achieving the same accuracy via Monte Carlo method would have taken over 3 months.

The Euler approximation to (1.10) is expected to be more and more accurate, as $h \to 0$. There is a large body of literature studying the discretization error for SDEs. The focus is on the rate of convergence of X^h_T to the 'true' X_T, as $h \to 0$ for a fixed $T > 0$. Here, authors typically distinguish two notions of convergence: weak and strong. The former concerns the convergence in distribution $X^h_T \rightsquigarrow X_T$ and the latter the convergence of the expected absolute error $\mathbb{E}|X^h_T - X_T|$, as $h \to 0$. For more notions of convergence we again we refer to [69]. From the recursion (1.11) we see that the absolute error $|X^h_T - X_T|$ will typically accumulate with time so the path-wise approximation is expected to get worse, as T grows and h is fixed.

We say that X has an invariant (stationary, steady-state) distribution μ if the law of X_t converges weakly to μ. When a stationary measure exists, an interesting problem is to study the weak convergence of $X^h_\infty \rightsquigarrow X_\infty$, where we put $X^h_\infty \sim \mu_h$ and $X_\infty \sim \mu$. The problem of studying this weak convergence can be more complicated, as μ_h might not even exist [84]. Some work has been done for the study of the rate of convergence of μ_h to μ for various (also higher-order) schemes, see [1, 92].

Hardly any work has been done in context of extremes. Note that strong and weak convergence focus on the 'bulk' of the distribution. This means in particular that they don't capture the size of the error for tail probabilities $|\mathbb{P}(X^h_\infty > u) - \mathbb{P}(X_\infty > u)|$ well. According to our findings, the error can be significant. In Chapter 4 we show that various discretization schemes can lead to completely different tail behavior of the stationary distribution μ_h as compared with the real, continuous-time one μ. Strikingly — fat tails may become light and vice versa.

While in this thesis we focus mainly on quantifying the discretization error in rare event setting, it is worth mentioning that there are methods aimed at decreasing it such as Multilevel Monte Carlo (MLMC) [53] or even eliminating it [83]. The idea is to boost the efficiency without the loss of accuracy by allocating more computational power to sampling lower resolution solutions. These methods require however some knowledge about the rate of convergence of the discretization error. It could be interesting to explore the performance of MLMC in rare event setting but this lies beyond the scope of this thesis. Finally, we note that a certain subclass of SDEs can be simulated exactly using acceptance-rejection method [14] thus eliminating the discretization error completely (possibly at the cost of much longer simulation time). We also mention that there exists an exact simulation scheme for general diffusions [20] but at its current stage of developments it is perhaps more of a theoretical value as it has infinite expected simulation time.

3 Detection error

In the previous section we argued that it is not possible to simulate a full time-continuous path of the whole process X on a computer and thus one is forced to observe the process only on the chosen grid $0, h, 2h, \ldots$. As a result, even when exact simulation of the process is possible (in a sense described in Section 2), one still has to work with a piece-wise linear process $X^h_t := X_{h[t/h]}$ instead of the real X. The *detection error*, is then the error one makes due to the lost information between the observation times, even when the observations at times $0, h, 2h, \ldots$ carry no error. It should not be confused with the *discretization* error, as the latter is due to *approximate* sampling on the grid points. Depending on the application at hand, these errors can arise separately or simultaneously.

Example 1.1 (continued)**.** In the previous section we introduced the discretization error, which can arise from using approximate sampling of SDEs (for example, when using Euler scheme). However, even when simulating the path with the exact recursion in Eq. (1.13), we still loose the information about the path between the observation points $0, h, 2h, \ldots$. Because of that, we might fail to detect whether the event

$$\{\tau_u^{\uparrow} < \tau_l^{\downarrow}\}$$

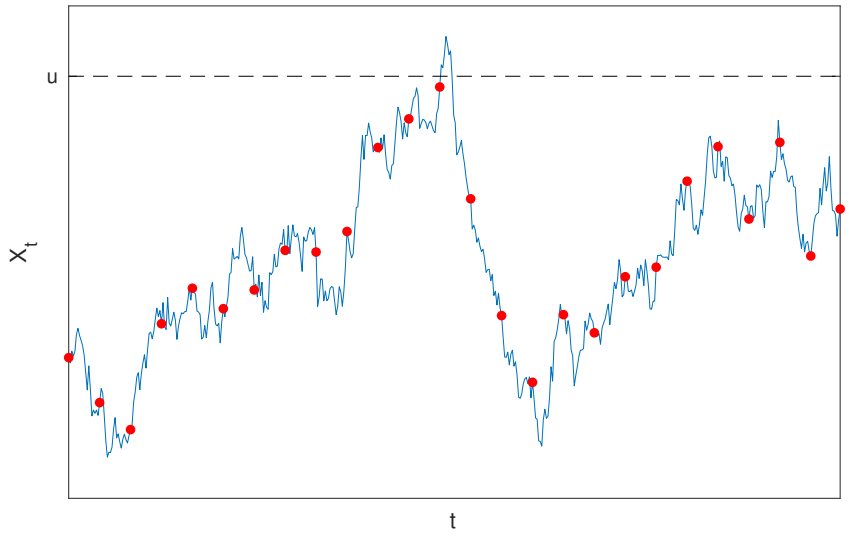


Figure 1.2: An illustration of detection error. Here, the continuous path (in blue) has crossed the threshold u but we have failed to detect that on the discrete grid (red dots).

h	10^{-2}	$5 \cdot 10^{-3}$	10^{-3}	$5 \cdot 10^{-4}$	10^{-4}
$\gamma^h_{\text{exact}}/\gamma$	72%	79%	91%	92%	96%

Table 1.2: Size of the detection error depending on the value of time step h for the problem described in the third part of Example 1.1 on page 10. The simulation results are accurate up to $\pm 1\%$ with 95% confidence.

took place. In Fig. 1.2 we show an example of a path, which hits $[u, \infty)$ in continuous time, but this is not detected on the discrete grid. Similarly, one might fail to detect downcrossing below l. In Tab. 1.2 we present values of detection error for various time steps h. We see that γ is underestimated in all cases and that the error is surprisingly large, of order 10% even for h as small as 10^{-3}. This hints that the OU process spends very little time above u making the upcrossing hard to detect on a finite grid. The detection error is expected to be even more significant, as $u \to \infty$. A similar situation occurs in the first passage problem for Brownian Motion, which we describe in more detail in Chapter 2.

There are many instances in which detection errors can arise. In this thesis

we focus on the detection error for the supremum on a finite time interval $[0,1]$, with $h = 1/n$,

$$\Delta_n := \sup_{t\in[0,1]} X - \sup_{t\in[0,1]} X^{1/n} \tag{1.14}$$

and the related probability of *failure in detection* of threshold exceedance

$$\Delta_n(u) := \mathbb{P}\big(\sup_{t\in[0,1]} X > u\big) - \mathbb{P}\big(\sup_{t\in[0,1]} X^{1/n} > u\big). \tag{1.15}$$

In the rare event simulation setting, that is, as $u \to \infty$, the failure in detection will eventually vanish so in Chapter 2 we work with the *relative* detection error $\Delta_n(u)/\mathbb{P}(\sup_{t\in[0,1]} X > u)$.

In the context of Lévy processes, several authors are interested in the asymptotic behavior of Δ_n, as n grows large. In [7], the rate of convergence of Δ_n and $\mathbb{E}\Delta_n$ and exact asymptotics were established for Brownian Motion. Chen [30] established upper and lower bounds on the rate of convergence of $\mathbb{E}\Delta_n$ under certain regularity assumptions. More recently, weak convergence of Δ_n (under a proper scaling) was derived by Ivanovs [63]. The failure in detection was studied by Broadie et al. [22], who found the exact asymptotics for $\Delta_n(u)$ for Brownian Motion and applied it in the context of more accurate pricing of barrier options.

Discretization and detection errors are often (but not always) closely related and just like in the previous section, algorithms introduced by Giles [53] and Rhee and Glynn [83] can be used to decrease or even completely eliminate the detection error. The study of the asymptotic behavior of the detection error is helpful for establishing efficient MLMC algorithms, see also [54].

In Chapter 2 we study the behavior of $\Delta_n(u)$ for a standard Brownian Motion in an asymptotic regime, when both n and u tend to infinity. We propose to deal with the detection error by using a *threshold-dependent* discretization grid meaning that we change the observation points depending on the value of the threshold u. This approach is proven to significantly decrease the detection error and can be applied to various stochastic processes (Gaussian, Lévy). In Chapter 3 we build on the result of Ivanovs [63] and derive the exact asymptotics of $\mathbb{E}\Delta_n^p$ and also $\Delta_n(u)$, as $n \to \infty$ under mild regularity assumptions. This limit result can be used in many areas of application to provide correction terms; e.g. in finance, to price barrier options more accurately.

Chapter 2

Controlling the time discretization bias for the supremum of Brownian Motion

In this chapter we consider the detection error (also called discretization bias) arising from time discretization when estimating the threshold crossing probability $w(b) := \mathbb{P}(\sup_{t\in[0,1]} B_t > b)$, with $(B_t)_{t\in[0,1]}$ a standard Brownian Motion. We prove that if the discretization is equidistant, then to reach a given target value of the relative detection error, the number of grid points has to grow quadratically in b, as b grows. When considering non-equidistant discretizations (with threshold-dependent grid points), we can substantially improve on this: we show that for such grids the required number of grid points is independent of b, and in addition we point out how they can be used to construct a strongly efficient algorithm for the estimation of $w(b)$. Finally, we show how to apply the resulting algorithm for a broad class of stochastic processes; it is empirically shown that the threshold-dependent grid significantly outperforms

its equidistant counterpart.

1 Introduction

Extreme values of random processes play a prominent role in a broad range of practical problems. It is often of interest to find the tail of the distribution of the supremum of a continuous-time stochastic process $(X_t)_{t\geq 0}$ over a finite time interval. In this chapter the focus is on the level crossing probability

$$w(b) := \mathbb{P}\left(\sup_{t\in[0,1]} X_t > b\right).$$

For many classes of processes, such as the Gaussian processes [2], typically no explicit expressions for $w(b)$ are available, with Brownian Motion and the Ornstein-Uhlenbeck process being notable exceptions. When an explicit expression for $w(b)$ is unavailable one usually resorts to using high-dimensional numerical integration and simulation-based methods, see e.g. [52] for further reading.

For most of the available numerical methods, the underlying continuous-time process needs to be discretized in time. One chooses a certain *finite grid* $T \subset [0,1]$ and then approximates $w(b)$ with $w_T(b) := \mathbb{P}\big(\sup_{t\in T} X_t > b\big)$. We note that this always leads to an underestimation, i.e., $w_T(b) \leq w(b)$. We quantify this underestimation by $\beta_T(b) := (w(b) - w_T(b))/w(b)$, the relative *discretization bias* (as $b \to \infty$, both $w_T(b)$ and $w(b)$ tend to 0, so that the *absolute* bias is not a meaningful accuracy measure). Typically T is chosen to be an *equidistant grid* $T = \{\frac{1}{n}, \frac{2}{n}, \ldots, 1\}$ and in that case, $\beta_T(b)$ can be reduced only by changing the *grid size* n. The finer the grid, the smaller the bias, but also, the larger the computational effort to estimate $w_T(b)$. The main drawback of using equidistant grids is that typically, to reach a given target value of the discretization bias, the grid size n has to grow with the threshold b. In that case, for large b, the appropriate grid size can become so large that the computation is not feasible. Two central questions arise from these observations: How fast does n have to grow in b? Furthermore, can we identify a more efficient family of grids?

In this chapter we address these issues for standard Brownian Motion. Although in this case $w(b)$ can be computed explicitly, there are no available expressions for $\beta_T(b)$. We conduct a thorough study of the influence of the choice of the grid on the corresponding relative bias. Furthermore, we argue that exploring the case of standard Brownian Motion is a first step towards finding efficient grids for a more general class of processes. We demonstrate numerically how our analysis of efficient grids for Brownian Motion leads to a useful procedure to determine efficient grids for a broad range of other processes.

The contributions of this chapter are the following. (i) The first finding can be seen as a negative result: we show that to *uniformly control* the relative bias, the size n of the equidistant grid must grow at least quadratically in b; see Theorem 2.1 in Section 3. Here, *uniform control* means that for a fixed $\varepsilon > 0$, we have that $\beta_T(b) < \varepsilon$ for all $b > 0$; the grid T can change in b. (ii) The second finding is that we can do much better by using a *threshold-dependent* family of grids, meaning that grid points change their location with b (but the number of points does not increase). The discretization bias induced by this particular family of grids is uniformly controlled without having to increase the number of grid points; see Theorem 2.2 in Section 4. According to the best of the authors' knowledge, this is the first result which shows that a careful choice of the grid can drastically increase the accuracy of the discrete estimator of $w(b)$. Using threshold-dependent grids makes it feasible to estimate $w(b)$ with moderate grid sizes even for very high thresholds b, which would be impossible to estimate using equidistant grids. In particular, in Section 5 we present a strongly efficient algorithm for the estimation of $w(b)$ that relies on threshold-dependent grids. (iii) In the third place, we point out how the ideas underlying our threshold-dependent grid can be used for a broad class of stochastic processes (including Gaussian processes, such as fractional Brownian Motion, and Lévy processes); it is empirically shown that the threshold-dependent grid significantly outperforms its equidistant counterpart.

An efficient grid (both small in size and inducing a small discretization bias) is particularly relevant for situations with large b. In this respect, the work presented here connects to the rare event simulation literature. As b approaches infinity, $w(b)$ decays exponentially to 0 and standard simulation-based methods like Crude Monte Carlo to estimate $w(b)$ become extremely time consuming.

We emphasize that rare event simulation methods commonly aim to control the sampling error, not the bias due to the discretization. Adler et al. [3] develop an algorithm that is strongly efficient (with bounded relative sampling error) for estimation of $w_T(b)$ (rather than $w(b)$). We will show that combining their algorithm with the use of threshold-dependent grids provides a strongly efficient algorithm for estimation of $w(b)$.

A topic closely related to ours concerns the quantification of the difference between the supremum of the stochastic process taken over $[0,1]$ and the supremum taken over a finite grid $T \subset [0,1]$, i.e.

$$\Delta(T) = \sup_{t\in[0,1]} X_t - \sup_{t\in T} X_t.$$

There are several results in the literature that study the behavior of $\Delta(T)$ for standard Brownian Motion. Asmussen et al. [7] showed that for the equidistant grids $T_n^{\text{eq}} = \{\frac{1}{n}, \ldots, \frac{n}{n}\}$, $\sqrt{n}\,\Delta(T_n^{\text{eq}})$ has a tight, non-degenerate weak limit, as $n \to \infty$ and [64] derived an expansion for $\mathbb{E}\Delta(T_n^{\text{eq}})$. For *random grids* $T_n^{\text{rnd}} = \{U_1, \ldots, U_n\}$, where $U_1, \ldots, U_n$ are i.i.d. uniform samples on $(0,1)$, independent of the Brownian Motion $(X_t)_{t\in[0,1]}$, [25] establish the weak limit of $\sqrt{n}\,\Delta(T_n^{\text{rnd}})$. Finally, [24] proposed a class of *adaptive* grids, meaning that the consecutive grid-points t_{k+1} are chosen based on $((t_1, B_{t_1}), \ldots, (t_k, B_{t_k}))$; given any $\delta > 0$, an adaptive grid $T_n^\delta = \{t_1^\delta, \ldots, t_n^\delta\}$ is provided such that $n^{1-\delta/2}\Delta(T_n^\delta)$ has a weak limit.

In our study we do not focus on the difference $\Delta(T)$ between the values of the maxima of the discrete and continuous-time Brownian Motion, but rather on the $\beta_T(b)$, i.e., the relative difference between the probabilities that these maxima lie above a certain fixed threshold.

There are several approaches to tackle the discretization bias available in the literature. Arguably, the most widely applicable method is *Multilevel Monte Carlo* (MLMC) [53]. It can be applied together with any numerical method that relies on discretization. The idea is to use several different *levels of discretization* and spend less computational effort (draw less samples) at the finest levels of discretization. MLMC effectively reduces the computational effort, and the time saved can be used to produce even finer levels of discretization. It could be

interesting to explore the combination of MLMC method together with the idea of threshold-dependent grids but further exploiting this procedure lies beyond the scope of this article.

One of the methods that aims to directly decrease the bias induced by equidistant grids is *continuity correction*. Since the discrete-time approximation $w_T(b)$ is always smaller than $w(b)$, one could *slightly* lower the threshold b to compensate for the underestimation. Broadie et al. [22], using the machinery developed in [89], proposed a way of lowering the threshold which improves the rate of convergence of the relative bias from $O(n^{-1/2})$, cf. Proposition 2.1, to $O(n^{-1})$, as the number of grid points n grows large. However, in the non-Brownian case, it remains a non-trivial problem how much b should be decreased. In fact, there is no direct way of making sure whether lowering b decreases the absolute relative bias, as lowering b by *too much* leads to overcompensation and thus to an estimate that is *larger* than $w(b)$. By contrast, it is straightforward to compare the bias induced by two different grids — the larger the discrete estimator $w_T(b)$, the smaller the relative bias.

There are also several simulation-based algorithms that do not rely on pre-discretization. Li and Liu [75] propose a strongly efficient algorithm for estimation of $w(b)$ for a large class of Gaussian processes (most prominently, processes with constant variance function). However, when the underlying process has a unique point of maximal variance (such as Brownian Motion), the algorithm requires the simulation of a random time $\tau \in [0, 1]$ from a density $f(t) \propto \mathbb{P}(X_t > b)$, which becomes a rare event simulation problem when b is large. While for an arbitrary process, the random discretization proposed in the algorithm requires a computational effort cubic in the number of grid points (in order to simulate a discrete Gaussian path), pre-discretization requires only quadratic effort; see the discussion in Section 5.

This chapter is organized as follows. Section 2 provides definitions, preliminaries, and develops a general intuition. In Section 3 we introduce useful upper and lower bounds for the discretization bias (see Lemma 2.1) and show that the number of points on the equidistant grid has to grow quadratically in the threshold b in order to uniformly control the discretization bias. In Section 4, as an alternative to equidistant grids, we study threshold-dependent grids,

which control the relative bias with a constant grid size, independently of b. The proofs of all lemmas and a proposition are postponed to Section 8. In Section 5 we present an algorithm by [3], that we use throughout the chapter for producing the numerical results; combining this algorithm with the use of threshold-dependent grids yields a strongly efficient algorithm for estimation of $w(b)$, see Corollary 2.1. In Section 6 we apply threshold-dependent grids developed in previous section to stochastic processes other than Brownian Motion: Brownian Motion with jumps, Ornstein-Uhlenbeck process and fractional Brownian Motion. Lastly, in Section 7 we present concluding remarks and discuss some ideas for future research of *optimal grids*. In the appendices we collect various technical results used throughout the chapter.

2 Preliminary results

Let $(B_t)_{t\in[0,1]}$ be a standard Brownian Motion on the time interval $[0,1]$ with $B_0 = 0$. We consider the probability of crossing a positive threshold b, that is

$$w(b) := \mathbb{P}\Big(\sup_{t\in[0,1]} B_t > b\Big). \tag{2.1}$$

For a standard Brownian Motion, an explicit formula for the threshold-crossing probability (2.1) is known, namely $w(b) = 2\,\mathbb{P}(B_1 > b)$, which follows directly using the *reflection principle* (see e.g. [78]). Given a *finite grid* T we define a discrete-time approximation of $w(b)$:

$$w_T(b) := \mathbb{P}\Big(\sup_{t\in T} B_t > b\Big), \tag{2.2}$$

where $T = \{t_1, \ldots, t_n\}$ is a finite subset of the interval $[0,1]$, ordered such that $t_1 < \ldots < t_n$. As we are mostly interested in choosing the grid T efficiently, we define the following performance measure.

Definition 2.1. *Let T be a finite grid on $[0,1]$, then*

$$\beta_T(b) := \frac{w(b) - w_T(b)}{w(b)} = \mathbb{P}\Big(\sup_{t\in T} B_t < b \;\Big|\; \sup_{t\in[0,1]} B_t > b\Big)$$

is called the relative bias induced by the grid T.

The second representation of relative bias in Definition 2.1 is especially intuitive. It means that the relative bias is *the probability that B_t stays below b on the grid T, given that its supremum over $[0,1]$ is greater than b.* Notice that any grid which includes the endpoint $t = 1$ will induce a relative bias no greater than $\frac{1}{2}$. Indeed, if $1 \in T$, then $w_T(b) = \mathbb{P}(\sup_{t\in T} B_t > b) \geq \mathbb{P}(B_1 > b)$ and thus

$$\beta_T(b) = 1 - \frac{w_T(b)}{w(b)} \leq 1 - \frac{\mathbb{P}\big(B_1 > b\big)}{2\,\mathbb{P}\big(B_1 > b\big)} = \frac{1}{2}.$$

Our objective is to accurately estimate $w(b)$ using discrete approximations $w_T(b)$, in a computationally efficient manner. Brownian Motion has continuous paths and thus it is always possible for a given b to find a fine enough grid to bound the bias up to a desired accuracy. However, the computational cost of estimating $w_T(b)$ grows in the grid size and thus it might be infeasible to numerically compute $w_T(b)$ for large grids.
At this point, we emphasize that we are not as much interested in the behaviour of $\beta_T(b)$ for a fixed b or a fixed n but rather in asymptotic regimes in which b and/or n approach infinity. For every b we allow to use a different grid so it seems natural to treat the grid as a function of threshold. For every b we define a collection of grids of all possible sizes $\{T_1(b), T_2(b), \ldots\}$, where $T_n(b)$ has n elements, and we denote $\beta_n(b) := \beta_{T_n(b)}(b)$. For a given *family of grids* we are interested in behavior of $\beta_n(b)$ as n or b tend to infinity. The most straightforward choice for the family of grids is the following.

Definition 2.2. *The family $\{T_n\}_{n\in\mathbb{N}}$, where $T_n := \{t_1^n, \ldots, t_n^n\}$ with $t_k^n := \frac{k}{n}$ is called the equidistant family of grids.*

Notice that the location of grid points on the equidistant grid is independent of b. Since the distance between neighboring points is equal to $\frac{1}{n}$, and since Brownian paths are continuous, it follows that $\beta_n(b) \to 0$, as $n \to \infty$ for any fixed b. It has been established in [7] that for T_n, equidistant grid, the difference between the continuous-time and discrete-time supremum $\varepsilon_n = \sup_{t\in[0,1]} B_t - \sup_{t\in T_n} B_t$ is of order $n^{-1/2}$. More precisely, the sequence $(\sqrt{n}\varepsilon_n)_{n\in\mathbb{N}}$ has a tight and non-degenerate weak limit.

Proposition 2.1. *Let $(B_t)_{t\in[0,1]}$ denote standard Brownian Motion and $\{T_n\}_{n\in\mathbb{N}}$ be the equidistant family of grids from Definition 2.2 with $\beta_n(b) := \beta_{T_n}(b)$. For*

any threshold $b > 0$ there exist positive constants C_1, C_2 such that

$$C_1\, n^{-1/2} \leq \beta_n(b) \leq C_2\, n^{-1/2}.$$

The proof of the Proposition 2.1 is given in Appendix C. The proof we give strongly resembles the proof of Theorem 2.1 below in Section 3, but we remark that it is also possible to derive it using the tools developed in [22].

Proposition 2.1 states that $\beta_n(b)$ decays like $n^{-1/2}$, when n grows large *for a fixed* b but it does not describe the behavior of the relative bias when b varies. In Theorem 2.1 in the following section, we derive an upper bound for $\beta_n(b)$ for n and b simultaneously.

Figure 2.1 shows the evolution of the relative bias for four different thresholds $b = 5, 6, 7, 8$ against the size of the equidistant grid. Even though all four graphs show the $n^{-1/2}$ decay, the graphs rise up with growing threshold. In particular, for thresholds $b = 5$ and 8 respectively $n = 700$ and 1700 points are needed to arrive at around 10% relative bias. It indicates that, as b grows *increasingly many grid-points are needed to arrive at the target relative bias.* Using the threshold-dependent grid that we develop in Section 4 one can arrive at 10% relative bias using approximately $n = 100$ grid-points, independently of the value of the threshold. This amounts to a substantial improvement of the computational efficiency.

In some cases, the equidistant family of grids is the best possible choice, in the sense that other grid families require at least equally fast asymptotic growth of n as b increases, in order to control the relative bias. Adler et al. [3] prove that for *centered, homogeneous and twice continuously differentiable (in a mean squared sense) Gaussian processes*, n has to grow linearly in b to uniformly control the relative bias. Moreover, if n grows sublinearly in b, then the relative bias of any family of grids (not necessarily equidistant) tends to its maximal value, as b approaches infinity. It is noted, however, that Brownian Motion does not belong to the family of Gaussian processes for which the result of [3] applies.

In the following two sections we analyze the asymptotic behavior of the relative bias $\beta_n(b)$ for two families of grids. We prove that the equidistant grid requires quadratic growth of n in b (see Theorem 2.1 in Section 3). As an

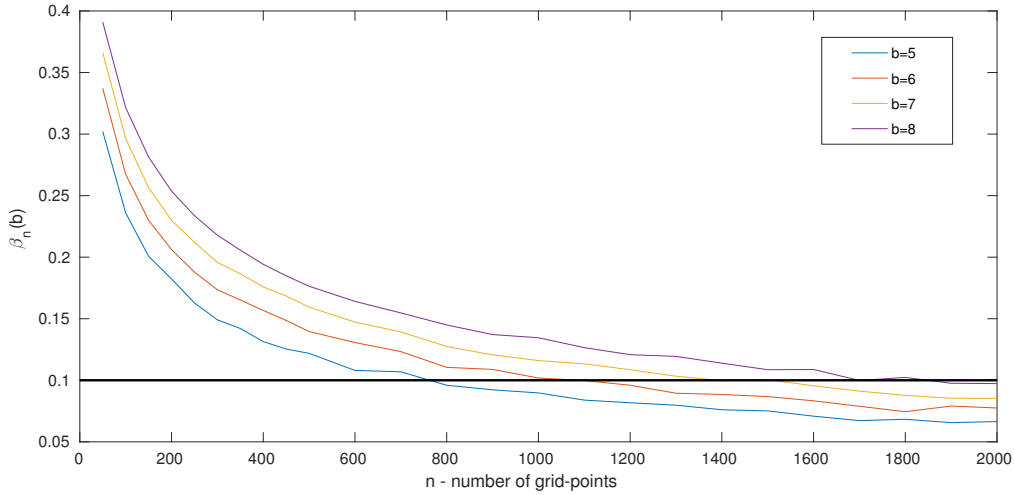


Figure 2.1: Plots of the relative bias $\beta_n(b)$ against the grid size n for the equidistant family of grids for four different thresholds. The numerical results are computed using an algorithm described in Section 5.

alternative, we develop the threshold-dependent family of grids, for which we prove that the relative bias can be made arbitrarily small, uniformly in b for fixed n (see Theorem 2.2 in Section 4). We obtain a uniform rate of convergence in n and also provide a closed-form expression for the threshold-dependent family of grids (see Definition (2.9) in Section 4).

3 Equidistant family of grids for Brownian Motion

This section is devoted to analyzing the asymptotic behavior of the relative bias for the equidistant family of grids. The methodology developed in this section will be used later to prove Theorem 2.2; in particular, the crucial part of the proof concerns bounds for the relative bias induced by an arbitrary finite grid, developed in Lemma 2.1.

The following theorem describes the asymptotic behaviour of the relative bias, under the equidistant family of grids.

Theorem 2.1. *Let $(B_t)_{t\in[0,1]}$ denote standard Brownian Motion and $\{T_n\}_{n\in\mathbb{N}}$ be the equidistant family of grids from Definition 2.2 with $\beta_n(b) := \beta_{T_n}(b)$.*

(a) Let b_0 be any positive, real number. There exist positive constants C_0, C_1, independent of b and n such that

$$\beta_n(b) \leq C_0 \cdot b n^{-1/2},$$

for all $b \geq b_0$, and

$$\beta_n(b) \leq C_1 \cdot n^{-1/2},$$

for all $b \in (0, b_0]$.

(b) Let $m : (0,\infty) \to (0,\infty)$ be such that $\lim_{b\to\infty} m(b)/b^2 = 0$. Then, as $b \to \infty$,

$$\inf_{n \leq m(b)} \beta_n(b) \longrightarrow \frac{1}{2}.$$

Part (a) of Theorem 2.1 states that $\beta_n(b) \leq C_0\, b n^{-1/2}$, so that in order to bound $\beta_n(b)$ uniformly in b it suffices to take $n = O(b^2)$. The second part of the Theorem 2.1 states that if $n = o(b^2)$ then $\beta_n(b) \to 1/2$, meaning that the relative bias cannot be bounded by an arbitrarily small number. Together, the two parts entail that the growth $n = O(b^2)$ is sufficient and there is no better (slower) growth which would guarantee a uniformly bounded relative bias.

The crucial part of the proof of Theorem 2.1 is the method of bounding the relative bias. Since no explicit expressions for $w_T(b)$ or $\beta_T(b)$ are known (even if T is an equidistant grid) we develop a general upper bound for $\beta_T(b)$ in the following lemma, in which we use the quantities

$$\begin{aligned} a_j(b) &:= \mathbb{P}\big(B_{t_j(b)-t_{j-1}(b)} < 0, \ldots, B_{t_n(b)-t_{j-1}(b)} < 0\big), \quad a_{n+1}(b) := 1/2, \\ w_j(b) &:= \mathbb{P}\big(\tau_b \in (t_{j-1}(b), t_j(b)] \;\big|\; \tau_b \in (0,1]\big), \\ \tau_b &:= \inf\{t \geq 0 : B_t > b\}. \end{aligned}$$

Notice that in this definition of $a_j(b)$ and $w_j(b)$ we allow grid points $t_1, \ldots, t_n$ to change their location with b. In the present section, which is on equidistant grids, the grid points obviously do not depend on b, but in later sections they do.

Lemma 2.1. *Let $T(b) = \{t_1(b), \ldots, t_n(b)\} \subset [0,1]$, where $0 < t_1(b) < \ldots < t_n(b) \leq 1$, and let $t_0(b) = 0$. The following lower and upper bounds for $\beta_T(b)$ apply:*

$$\underline{\beta}_T(b) \leq \beta_T(b) \leq \bar{\beta}_T(b)$$

with

$$\underline{\beta}_T(b) := \frac{1}{2}\sum_{j=1}^{n} a_{j+1}(b)\, w_j(b), \qquad \bar{\beta}_T(b) := \sum_{j=1}^{n} a_j(b)\, w_j(b).$$

A short proof of Lemma 2.1 is included in Section 8. The bounds consist of elements of two types: $a_j(b)$, the probability that B_t stays negative at times $t_j - t_{j-1}, \ldots, t_n - t_{j-1}$, and $w_j(b)$, the probability that B_t hits b for the first time in the interval $[t_{j-1}, t_j]$ given that its supremum over $[0,1]$ is greater than b.

For a general grid $T(b)$, the probabilities $a_j(b)$ are difficult to control. However, when $T(b)$ is equidistant (thus independent of b), then also the probabilities a_j are independent of b; we emphasize this independence by writing a_j instead of $a_j(b)$ throughout this section. As a result, there exists a tight asymptotic bound for them (see Lemma 2.2 below); we were inspired to look into such quantities while reading [78, Section 5]. The upper and lower bounds for the probabilities $w_j(b)$ are developed using a mean value theorem, see (B.V).

Lemma 2.2. *There exist constants $C_1^*, C_2^* > 0$ such that:*

$$C_1^* n^{-1/2} \leq \mathbb{P}\Big(B_1 > 0, \ldots, B_n > 0\Big) \leq C_2^* n^{-1/2}$$

for all $n \in \mathbb{N}$.

In fact, the assertion in Lemma 2.2 is true for any symmetric random walk; see [47, Theorem 4 in Section XII.7, and Lemma 1 in Section XII.8]. Before proving Theorem 2.1 we present one more lemma.

Lemma 2.3. *Let $T = \{t_1, \ldots, t_n\}$ be such that $t_k = \frac{k}{n}$ and let $t_0 = 0$. Then the upper bound $\bar{\beta}_T(b)$ developed in Lemma 2.1 is an increasing function of b.*

An important implication of Lemma 2.3 is that for any $b_0 > 0$ we have that $\beta_T(b) \leq \bar{\beta}_T(b) \leq \bar{\beta}_T(b_0)$ uniformly for all $b \leq b_0$, which completely covers the statement on the situation that $b \leq b_0$ in part (a) of Theorem 2.1. The proof of Lemma 2.3 is included in Appendix C.

Proof of Theorem 2.1(a). Thanks to Lemma 2.3 it suffices to prove the first part of Theorem 2.1(a), i.e. we assume that $b \geq b_0$. Without loss of generality we put $b_0 = 1$. Exploiting the upper bound developed in Lemma 2.1 we decompose the sum $\sum_{j=1}^{n} a_j \cdot w_j(b)$ into three parts, which we treat separately:

$$\beta_n(b) \leq a_1 \cdot w_1(b) + \sum_{j=2}^{n-1} a_j \cdot w_j(b) + a_n \cdot w_n(b), \tag{2.3}$$

Using the definition of the equidistant grid and the scaling property of Brownian motion we can see that $a_j = \mathbb{P}\big(B_{t_j - t_{j-1}} < 0, \ldots, B_{t_n - t_{j-1}} < 0\big) = \mathbb{P}\big(B_1 < 0, \ldots, B_{n-j+1} < 0\big)$ and the bound in Lemma 2.2 yields $a_j \leq C_2^*(n-j+1)^{-1/2}$ for all $j \in \{1, \ldots, n\}$. Since all $w_j(b) \leq 1$, we thus have a straightforward bound for the first term in (2.3):

$$a_1 \cdot w_1(b) \leq C_2^* \, n^{-1/2}$$

The second term we bound in the following fashion, relying on the upper bound stated in (B.V) that we have for $w_j(b)$,

$$\begin{aligned}
\sum_{j=2}^{n-1} a_j \cdot w_j(b) &\leq \sum_{j=2}^{n-1} C_2^* \, (n-j+1)^{-1/2} \cdot \tfrac{b(b+\sqrt{b^2+4})}{4} t \tfrac{\sqrt{n}}{(j-1)^{3/2}} e^{-\frac{b^2}{2} \cdot \left(\frac{n}{j} - 1\right)} \\
&\leq C_1 \cdot b n^{-1/2} \cdot \sum_{j=2}^{n-1} \frac{1}{n} \cdot \left(\tfrac{b}{\sqrt{1-j/n}} \cdot \left(\frac{j}{n} \right)^{-3/2} e^{-\frac{b^2}{2} \cdot \left(\frac{n}{j} - 1\right)} \right) \qquad (2.4) \\
&\leq C_1 \cdot b n^{-1/2} \cdot \int_0^1 \frac{b}{\sqrt{1-x}} \cdot x^{-3/2} \cdot e^{-\frac{b^2}{2}(1/x - 1)} \, \mathrm{d}x \qquad (2.5) \\
&\leq C_1 \cdot b n^{-1/2},
\end{aligned}$$

where C_2^* comes from Lemma 2.2 and C_1 is a positive constant, independent of b and n. To arrive at (2.4) we use that $2(j-1) \geq j$ for all relevant j. In the transition from (2.4) to (2.5) we use the definition of the Riemann sum for the function

$$f(b, x) := \frac{b}{\sqrt{1-x}} x^{-3/2} e^{-\frac{b^2}{2}(1/x - 1)};$$

note that, since $f(b, x)$ is an increasing function of x when $b \geq 1$, see (B.VI), the Riemann sum in (2.4) underestimates the integral, i.e., $\sum_{j=2}^{n} \frac{1}{n} f(b, \frac{j-1}{n}) \leq$

$\int_0^1 f(b,x)\,\mathrm{d}x = \sqrt{2\pi}$.
Lastly, since $a_n = \mathbb{P}(B_{t_n} < 0) = \frac{1}{2}$ we have

$$a_n \cdot w_n(b) \le \frac{1}{2} \cdot \frac{b(b+\sqrt{b^2+4})}{4} \frac{\sqrt{n}}{(n-1)^{3/2}} \le C_2 \frac{b^2}{n},$$

where C_2 is a positive constant independent of n and b. Since $w_n(b) \le 1$ this results in

$$a_n \cdot w_n(b) \le \min\left\{C_2 \frac{b^2}{n}, \frac{1}{2}\right\} \le \sqrt{\min\left\{C_2 \frac{b^2}{n}, \frac{1}{2}\right\}} \le \sqrt{C_2}\, b n^{-1/2}.$$

Combining the above bounds,

$$\begin{aligned}\beta_n(b) &\le a_1 \cdot w_1(b) + \sum_{j=2}^{n-1} a_j \cdot w_j(b) + a_n \cdot w_n(b) \\ &\le C_2^* n^{-1/2} + C_1\, b n^{-1/2} + \sqrt{C_2}\, b n^{-1/2} \le C_0\, b n^{-1/2},\end{aligned}$$

where C_0 is a positive constant independent of b and n. This concludes the proof. □

Proof of Theorem 2.1(b). Without loss of generality we can assume $m(b) \to \infty$ as $b \to \infty$. Similar to the proof of Lemma 2.1 in Section 8 we obtain:

$$\begin{aligned}w(b)\,\beta_T(b) &= \mathbb{P}\Big(\sup_{t\in T} B_t < b,\ \sup_{t\in[0,1]} B_t > b\Big) = \mathbb{P}\Big(\sup_{t\in T} B_t < b, \tau_b \in [0,1]\Big) \\ &= \sum_{j=1}^{n} \mathbb{P}\Big(\sup_{t\in\{t_j,\dots,t_n\}} B_t < b, \tau_b \in (t_{j-1}, t_j]\Big) \ge \mathbb{P}\Big(B_{t_n} < b, \tau_b \in (t_{n-1}, t_n]\Big) \\ &= \int_{t_{n-1}}^{t_n} \mathbb{P}(B_{t_n} < b | B_s = b)\mathbb{P}(\tau_b \in \mathrm{d}s) = \frac{1}{2}\,\mathbb{P}(\tau_b \in (t_{n-1}, t_n))\end{aligned}$$

Dividing both sides of the inequality by $w(b)$ yields a lower bound on $\beta_T(b)$

$$\frac{1}{2}\,\mathbb{P}\big(\tau_b \in (t_{n-1}, t_n] \;\big|\; \tau_b \in (0,1]\big) = \frac{1}{2} \cdot \frac{\Phi(-b/\sqrt{t_n}) - \Phi(-b/\sqrt{t_{n-1}})}{\Phi(-b)} \tag{2.6}$$

where $\Phi(\cdot)$ denotes the standard normal cdf, and we use the fact that $\mathbb{P}(\tau_b \le t) = 2\,\mathbb{P}(B_t > b)$. In our case $t_n = 1$ and $t_{n-1} = \frac{n-1}{n} \le \frac{m-1}{m}$, so that due to the

monotonicity of $\Phi(\cdot)$

$$\inf_{n\leq m(b)} \beta_n(b) \geq \frac{1}{2} - \frac{1}{2}\frac{\Phi(-b/\sqrt{(m-1)/m})}{\Phi(-b)}\,. \tag{2.7}$$

Taking the limit $b \to \infty$ on both sides of inequality (2.7) yields:

$$\begin{aligned}\lim_{b\to\infty} \inf_{n\leq m(b)} \beta_n(b) &\geq \frac{1}{2} - \frac{1}{2}\lim_{b\to\infty} \frac{\Phi(-b/\sqrt{(m-1)/m})}{\Phi(-b)} \\ &= \frac{1}{2} - \frac{1}{2}\lim_{b\to\infty} \frac{\frac{\sqrt{(m-1)/m}}{b}\varphi(b/\sqrt{(m-1)/m})}{\frac{1}{b}\varphi(b)} \\ &= \frac{1}{2} - \frac{1}{2}\lim_{b\to\infty} e^{-b^2/(2(m-1))} = \frac{1}{2}\end{aligned} \tag{2.8}$$

where $\varphi(\cdot)$ denotes the standard normal pdf, in (2.8) we apply (B.II), and the last equality is a consequence of the assumption that $\lim_{b\to\infty} m(b)/b^2 = 0$. □

In this section we have proven that in order to uniformly control the relative bias, the size of the equidistant grid must grow at least quadratically in b, as b approaches infinity. In the next section we present a threshold-dependent grid, which yields a uniform bound on the relative bias using a grid of given size. In other words, in order to control the relative bias with increasing b, instead of adding more and more points to the grid, it suffices to suitably shift their location.

4 Threshold-dependent grids for Brownian Motion

In this section we prove the main result of the chapter. We explicitly present a threshold-dependent family of grids which uniformly (in b) bounds the relative bias.

Before we introduce the result, we give some intuition why it is possible to control the relative bias as b grows, without increasing n. Firstly, for any given $\varepsilon > 0$, we have that

$$\mathbb{P}(\sup_{t\in[0,1-\varepsilon]} B_t > b) = 2\,\mathbb{P}(B_{1-\varepsilon} > b) = o(w(b)),$$

as $b \to \infty$. Therefore,

$$\mathbb{P}\Big(\sup_{t\in[1-\varepsilon,1]} B_t > b \,\Big|\, \sup_{t\in[0,1]} B_t > b \Big) \longrightarrow 1, \quad \text{as } b \to \infty.$$

It means that with growing b, the 'hitting of the threshold' occurs closer and closer to time $t = 1$. It indicates that the grid points should be gradually shifted towards the point $t = 1$, as b is increasing. Moreover, the result in Theorem 2.1 indicates how fast the points should be shifted. It states that for the family of equidistant grids, the uniform bound on the bias is achieved if the number of grid points grows quadratically in b. Equivalently, the distances between neighboring points are decreasing proportionally to b^{-2}. It turns out that this is indeed the pace at which the points should be shifted towards $t = 1$.

In the following result, $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the standard normal cdf and its inverse, respectively.

Theorem 2.2. *Let $(B_t)_{t\in[0,1]}$ be a standard Brownian Motion. Fix $b_0 > 0$ and let $\{T_n(b)\}_{n\in\mathbb{N},b>0}$ be a family of grids such that $T_n(b) = \{t_1^n(b), \ldots, t_n^n(b)\}$; here $t_k^n(b) := \frac{k}{n}$ for $b \le b_0$, and*

$$t_k^n(b) := \left(\frac{b}{\Phi^{-1}\left(\frac{k}{n}\,\Phi(-b)\right)} \right)^2, \tag{2.9}$$

for $b > b_0$. Denote $\beta_n(b) := \beta_{T_n(b)}(b)$. There exists a positive C, independent of b and n, such that

$$\beta_n(b) \le C\, n^{-1/4}$$

for all $b > 0$.

We emphasize that the bound for the relative bias $\beta_n(b)$ developed above does not depend on the threshold b and thus holds *uniformly*, for all b. Figure 2.2 shows the comparison between the relative bias of the equidistant and the threshold-dependent grid, both of size $n = 100$. The bias induced by the threshold-dependent grid remains uniformly bounded (by circa 0.1), while the former tends to 0.5, the worst possible relative bias, cf. Theorem 2.1, part (b).

Notice that for small b, $\{T_n(b)\}_{n\in\mathbb{N},b\in(0,b_0]}$ in Theorem 2.2 is identical to the equidistant family of grids. In fact this is exactly the setting of the second part

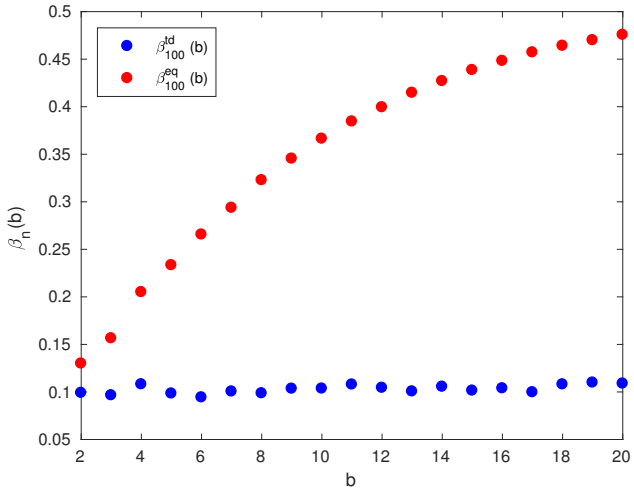


Figure 2.2: A plot of the relative bias of the equidistant grid $\beta^{\mathrm{eq}}_{100}(b)$ and the threshold-dependent grid $\beta^{\mathrm{td}}_{100}(b)$, both with fixed grid size $n = 100$, as a function of b. Notice that $\beta^{\mathrm{eq}}_{100}(b)$ tends to 0.5 with growing b (the largest possible bias), while $\beta^{\mathrm{td}}_{100}(b)$ remains bounded by about 0.1. The numerical results are computed with the algorithm discussed in Section 5. The relative error due to finite sample size is negligible (smaller than 0.006).

of the Theorem 2.1(a). The real contribution of Theorem 2.2 is the regime when $b > b_0$. The grid defined in (2.9) is the unique solution to the set of equations

$$\mathbb{P}\Big(\tau_b \in (t^n_{k-1}(b), t^n_k(b)] \;\Big|\; \tau_b \in (0,1]\Big) = \frac{1}{n} \tag{2.10}$$

for all $k \in \{1, \ldots, n\}$ and $t_0 := 0$. To see this, we sum up the first k equations in (2.10) and obtain an explicit equation for $t^n_k(b)$:

$$\mathbb{P}\Big(\tau_b \in (0, t^n_k(b)] \;\Big|\; \tau_b \in (0,1]\Big) = \frac{k}{n}. \tag{2.11}$$

Since for Brownian Motion it holds that

$$\mathbb{P}(\tau_b \in (0, t^n_k(b)]) = 2\,\mathbb{P}(B_{t^n_k(b)} > b) = 2\,\Phi(-b/\sqrt{t^n_k(b)}),$$

and in particular $\mathbb{P}(\tau_b \in (0,1]) = 2\,\mathbb{P}(B_1 > b) = 2\,\Phi(-b)$, Eqn. (2.11) can be

equivalently expressed as

$$\frac{\mathbb{P}(B_{t_k^n(b)} > b)}{\mathbb{P}(B_1 > b)} = \frac{k}{n} \tag{2.12}$$

or, in terms of the cdf $\Phi(\cdot)$,

$$\Phi\left(-b/\sqrt{t_k^n(b)}\right) = \frac{k}{n}\Phi(-b).$$

Finally, after taking the inverse $\Phi^{-1}(\cdot)$ from both sides of the equation above we see that $t_k^n(b)$ satisfies (2.9). Figure 2.3 shows the placement of the grid-points on the grid $T_5(b)$, as defined in (2.9), for increasing b. In fact, one can prove that

$$b^2\big(1 - t_k^n(b)\big) \xrightarrow{b\to\infty} -2\log(k/n) \tag{2.13}$$

and thus

$$t_k^n(b) \approx 1 - \frac{2\log(n/k)}{b^2}$$

for large b. It means that the points of the grid (2.9) are clustered around $t = 1$, with distances between the points proportional to b^{-2}. Here we see an important connection with Theorem 2.1(a), where the distances between grid-points decrease at the same pace, as already mentioned in the opening paragraph of this section.

For $b > b_0$, the points $t_1^n(b), \ldots, t_n^n(b)$ of the threshold-dependent grid (2.9) do not coincide with the equidistant grid, entailing that we can not directly use Lemma 2.2 to control the terms of type $a_j(b)$ in the upper bound developed in Lemma 2.1 in Section 3. The following lemma resolves this issue.

Lemma 2.4. *Let $t_0 = 0 < t_1 < t_2 < \ldots < t_n < \infty$. Then*

$$\mathbb{P}\Big(B_{t_1} > 0, \ldots, B_{t_n} > 0\Big) \le \mathbb{P}\Big(B_1 > 0, \ldots, B_N > 0\Big),$$

for any $N \le N_n$, where

$$N_n := \left(\frac{t_n}{\max_{k=1,\ldots,n}(t_k - t_{k-1})}\right)^{1/2}.$$

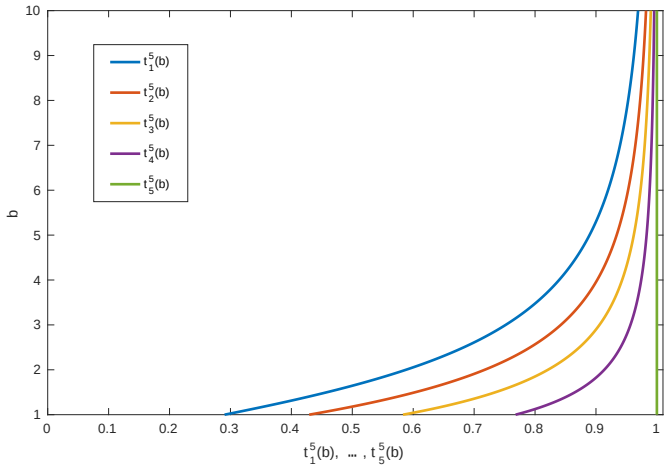


Figure 2.3: Location of the grid-points $t_1^5(b), \ldots, t_5^5(b)$ defined in (2.9) with increasing threshold b. Note that with growing b all the points are gradually shifted towards the end-point $t = 1$.

A proof of this lemma is provided in Section 8. Lemma 2.2 applied to the upper bound in Lemma 2.4 yields a simple upper bound for $\mathbb{P}\big(B_{t_1} > 0, \ldots, B_{t_n} > 0\big)$ for any choice of $t_0 = 0 < t_1 < t_2 < \ldots < t_n < \infty$. In our case, after applying Lemma 2.1 we have to control probabilities of the type $\mathbb{P}\big(B_{t_j - t_{j-1}} < 0, \ldots, B_{t_n - t_{j-1}} < 0\big)$, and thus we need a lower bound on

$$\frac{t_n - t_{j-1}}{\max_{k=j,\ldots,n}(t_k - t_{k-1})},$$

which we give in the following lemma.

Lemma 2.5. *For the grid in (2.9), for $k > j$, $b > 0$ and $n \in \mathbb{N}$ we have:*

$$\text{(a)} \qquad \frac{t_n^n(b) - t_j^n(b)}{t_k^n(b) - t_j^n(b)} \geq \frac{\log n - \log j}{\log k - \log j}.$$

and when additionally $b \geq \sqrt{3}$ we have

$$\text{(b)} \qquad \max_{k=j,\ldots,n} (t_k^n(b) - t_{k-1}^n(b)) = t_j^n(b) - t_{j-1}^n(b)$$

Lemma 2.5 is proven in Section 8. We note that the lower bound in part (a)

of Lemma 2.5 is in fact equal to

$$\lim_{b\to\infty} \frac{1 - t_j^n(b)}{t_k^n(b) - t_j^n(b)}.$$

With these lemmas we can prove Theorem 2.2.

Proof of Theorem 2.2. Part (a) of Theorem 2.1 states that for any choice of b_0 there exists positive C_1 such that $\beta_n(b) \le C_1 n^{-1/2}$ for $b \le b_0$ and thus also $\beta_n(b) \le C_1 n^{-1/4}$. Without the loss of generality, from now on we assume that $b > b_0 = \sqrt{3}$. Fix $n \in \mathbb{N}$ and denote $t_k := t_k^n(b)$ for notational simplicity. After combining the general upper bound from Lemma 2.1 with the equivalent definition (2.10) of the threshold-dependent grid (2.9) we obtain

$$\beta_n(b) \le \sum_{j=1}^{n} a_j(b) w_j(b) = \frac{1}{n} \sum_{j=1}^{n} a_j(b);$$

observe that in our setting $w_n(b) = \frac{1}{n}$. Moreover, Lemma 2.4 yields (recalling the definition of $a_n(b)$)

$$\beta_n(b) \le \frac{1}{2n} + \frac{1}{n} \sum_{j=2}^{n-1} \mathbb{P}\big(B_1 > 0, \dots, B_{N_n(j)} > 0\big) + \frac{1}{2n},$$

where

$$N_n(j) := \left\lceil \left(\frac{t_n^n(b) - t_{j-1}^n(b)}{\max_{k\ge j} |t_k^n(b) - t_{k-1}^n(b)|} \right)^{1/2} \right\rceil.$$

Combining Lemma 2.2 with Lemma 2.5 gives

$$\beta_n(b) \le \frac{1}{n} + C \frac{1}{n} \sum_{j=2}^{n-1} \widetilde{N}_n(j)^{-1/4}, \quad \text{where} \quad \widetilde{N}_n(j) := \frac{\log n - \log(j-1)}{\log j - \log(j-1)}$$

with a constant $C > 0$ that is independent of n and b. Notice that $\widetilde{N}_n(j)$ does not depend on b. For $b > b_0$ we thus obtain

$$\beta_n(b) \le \frac{1}{n} + C\, n^{-1} \sum_{j=2}^{n-1} \left(\frac{\log j - \log(j-1)}{\log n - \log(j-1)} \right)^{1/4}$$

$$
\begin{aligned}
&= \frac{1}{n} + C\, n^{-1} \sum_{j=2}^{n-1} \left(\frac{\log\left(1+\frac{1}{j-1}\right)}{\log\left(\frac{n}{j-1}\right)} \right)^{1/4} \\
&\leq \frac{1}{n} + C\ n^{-1/4} \sum_{j=2}^{n-1} \frac{1}{n} \left(\frac{\frac{n}{j-1}}{\log\left(\frac{n}{j-1}\right)} \right)^{1/4} \qquad (2.14) \\
&\leq \frac{1}{n} + C\ n^{-1/4} \int_0^1 \left(\frac{1}{-x \log x} \right)^{1/4} \mathrm{d}x \qquad (2.15) \\
&\leq C\, n^{-1/4}
\end{aligned}
$$

where C is a constant, independent from n and b, that might differ from line to line. In (2.14) we use the inequality $\log(1+x) \leq x$ and in (2.15) we use the convergence of the Riemann sum to the integral. This concludes the proof of Theorem 2.2. □

Remark 2.1. For the purpose of showing that for any confidence level α and bias ε, see also (2.16), the 'equiprobable' grid (as defined through (2.9)) requires a computational effort that is bounded in b, it suffices that the decay of the upper bound for $\beta_n(b)$ in Theorem 2.2 is of order $n^{-1/4}$; see Corollary 2.1 in Section 5. As an aside we remark that we hypothesize that this decay is actually of order $n^{-1/2}$. This is supported by numerical experiments; see Figure 2.4 where plots of $\beta_n(b)$ versus n are shown for the threshold-dependent grid (2.9). The step we expect to be 'loose', in obtaining the bound of Theorem 2.2, is the one corresponding to Lemma 2.4. We conjecture that Lemma 2.4 is valid with

$$N_n := \frac{t_n}{\max_{k=1,\ldots,n}(t_k - t_{k-1})}.$$

(i.e., without the square root), which suffices to yield the $n^{-1/2}$ decay of $\beta_n(b)$.

5 Numerical algorithm for estimation of $w(b)$

As mentioned in the introduction, the family of threshold-dependent grids (2.9) can be used to construct a strongly efficient algorithm for estimation of $w(b)$, see Corollary 2.1 below. In this chapter, by 'strongly efficient' we mean that for

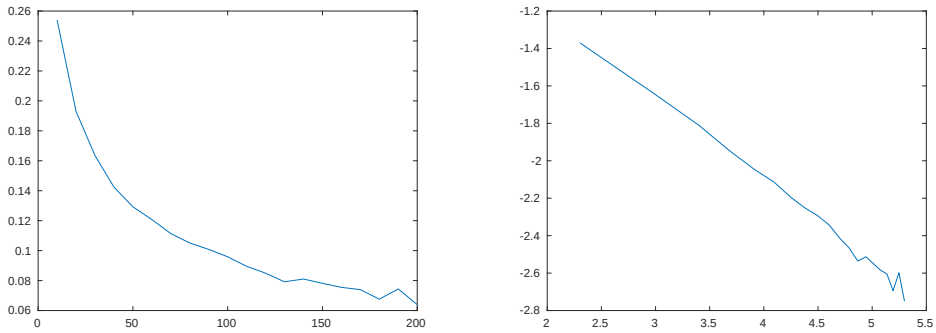


Figure 2.4: Relative bias $\beta_n(b)$ versus grid size n for the threshold-dependent grid (2.9). The threshold is fixed at $b = 3$. The right panel shows a loglog plot, left panel a linear plot. The results suggest that $\beta_n(b)$ decays proportionally to $n^{-1/2}$ rather than $n^{-1/4}$ (see also Remark 2.1).

any given accuracy $\varepsilon > 0$ and confidence level $\alpha > 0$ the computational time of an estimator $\widehat{w}(b)$ for $w(b)$ that satisfies

$$\mathbb{P}\left(\left|\frac{\widehat{w}(b)}{w(b)} - 1\right| > \varepsilon\right) < \alpha \tag{2.16}$$

is bounded independently of the threshold b.

In all numerical experiments throughout this chapter we used an algorithm developed by [3], see Algorithm 2.1 below. Although it is applicable for estimation of quantities such as $\mathbb{P}(\max_{i\in\{1,\dots,n\}} X_i > b)$, where $X \in \mathbb{R}^n$ is normally distributed with an arbitrary positive-definite covariance matrix, we present their algorithm for the specific case of Brownian Motion, as considered in this chapter.

Algorithm 2.1 ([3]). *Choose a threshold b and a finite grid $T = \{t_1, \dots, t_n\} \in [0, 1]$. The estimator $\widehat{w}_T(b)$, computed according to the following algorithm, is an unbiased estimator of $w_T(b)$.*

1. *Generate a random time τ on the grid, i.e. $\tau \in T$, according to the law*

$$\mathbb{P}(\tau = t_k) = \frac{\mathbb{P}(B_{t_k} > b)}{\sum_{j=1}^{n} \mathbb{P}(B_{t_j} > b)}.$$

2. *Generate B_τ under the condition $B_\tau > b$.*

3. *Generate a discrete path of the Brownian Motion $(B_{t_1}, \dots, B_{t_n})$ conditioned on the pair (τ, B_τ) generated in the previous steps.*

4. *Compute*
$$\widehat{w}_T(b) := \frac{\sum_{j=1}^n \mathbb{P}(B_{t_j} > b)}{\sum_{j=1}^n 1_{\{B_{t_j} > b\}}}.$$

[3] prove that the Algorithm 2.1 gives an *unbiased* estimator of $w_T(b)$ (not of $w(b)$) and that for a fixed T (independent of b), the relative variance $\frac{\mathbb{V}\mathrm{ar}(\widehat{w}_T(b))}{w_T^2(b)} \to 0$, as $b \to \infty$. The authors also propose an estimator for $w(b)$, which relies on a *random discretization*. However, with growing b, one needs increasingly many random grid-points in order to control the relative bias, therefore the continuous-time algorithm *is not* strongly efficient. In order to reduce the sampling error one generates multiple replicas of the estimator and takes their average. Since every replica is based on a different grid, one must repeatedly calculate the Cholesky decomposition (whose computational time is cubic in the number of grid-points) in order to sample discrete Gaussian paths in Step 3 of Algorithm 2.1. Choosing a predefined grid speeds up this computation, as in that case the Cholesky decomposition has to be performed *only once*, making its computational cost negligible.

Combining the threshold-dependent grids as proposed in Section 4 with Algorithm 2.1 yields a strongly efficient estimator for $w(b)$ which is given in the corollary below.

Corollary 2.1 (Strongly efficient algorithm for the estimation of $w(b)$)**.** *Fix an accuracy $\varepsilon > 0$ and a confidence level $\alpha > 0$. Choose a grid $T := T_n(b)$ from the family of grids defined in* (2.9) *such that $\beta_T(b) := \beta_n(b) < \varepsilon$ for all $b > 0$ (this is possible due to the result in Theorem 2.2). Let $\widehat{w}_T^{(1)}(b), \dots \widehat{w}_T^{(N)}(b)$ be i.i.d copies of the estimator from Algorithm 2.1, with*

$$N \geq \frac{n^2}{\alpha(\varepsilon - \beta_T(b))^2}.$$

Then

$$\widehat{w}(b) := \frac{1}{N}\sum_{i=1}^{N} \widehat{w}_T^{(i)}(b)$$

satisfies

$$\mathbb{P}\left(\left|\frac{\widehat{w}(b)}{w(b)} - 1\right| > \varepsilon\right) < \alpha, \tag{2.17}$$

and the computational effort to simulate $\widehat{w}(b)$ is bounded independently of b.

Proof. First notice that since $\beta_T(b)$ is uniformly bounded in b (see Theorem 2.2), so that N is fixed independently of b, it follows that $\widehat{w}(b)$ can be computed in bounded time, independently of b. It remains to prove that $\widehat{w}(b)$ satisfies the strong efficiency property (2.17). Note that $\widehat{w}(b)$ is an unbiased estimator of $w_T(b)$, not of $w(b)$. The relative variance of $\widehat{w}(b)$ with respect to $w_T(b)$ can be bounded independently of b for an arbitrary choice of the grid in terms of the grid size n,

$$\frac{\mathbb{V}\mathrm{ar}(\widehat{w}_T(b))}{(w_T(b))^2} \leq \frac{\mathbb{E}(\widehat{w}_T(b))^2}{(w_T(b))^2} \leq \left(\frac{\sum_{j=1}^{n}\mathbb{P}(B_{t_j} > b)}{\max_{j\in\{1,\dots,n\}}\mathbb{P}(B_{t_j} > b)}\right)^2 \leq n^2.$$

Due to Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{\widehat{w}(b)}{w(b)} - 1\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\frac{\widehat{w}(b) - w_T(b)}{w(b)} + \frac{w_T(b) - w(b)}{w(b)}\right| > \varepsilon\right) \\
&\leq \mathbb{P}\left(\left|\frac{\widehat{w}(b) - w_T(b)}{w(b)}\right| > \varepsilon - \beta_T(b)\right) \\
&\leq \frac{\mathbb{V}\mathrm{ar}(\widehat{w}(b))}{(\varepsilon - \beta_T(b))^2(w(b))^2} = \frac{1}{N}\cdot\frac{\mathbb{V}\mathrm{ar}(\widehat{w}_T(b))}{(\varepsilon - \beta_T(b))^2(w(b))^2} \\
&\leq \frac{1}{N}\cdot\frac{n^2}{(\varepsilon - \beta_T(b))^2} \leq \alpha.
\end{aligned}$$

This concludes the proof. □

We conclude this section by a remark on the simulation of the conditioned Brownian Motion in Step 3 of Algorithm 2.1. The naïve method would be to construct the covariance matrix of the conditioned process, calculate the Cholesky decomposition of that matrix (cubic in the number of grid points) and then simulate the process in a standard manner. Notice that this step must be

repeated for every replica $\widehat{w}_T^{(i)}(b)$ and thus its computational cost scales with the number of samples. The following algorithm, which can be found e.g. in [44], requires only a single calculation of the Cholesky decomposition for all replicas.

Algorithm 2.2 ([44]). *Let $X = (X_1, X_2)^T \in \mathbb{R}^n$, where $X_1 \in \mathbb{R}^{n-1}$ and $X_2 \in \mathbb{R}$, be normally distributed with mean μ and covariance matrix Σ,*

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \text{ where } \mu_1 \in \mathbb{R}^{n-1} \text{ and } \mu_2 \in \mathbb{R},$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix}, \text{ where } \Sigma_{11} \in \mathbb{R}^{(n-1)\times(n-1)}, \Sigma_{12} \in \mathbb{R}^{n-1} \text{ and } \Sigma_{22} \in \mathbb{R}.$$

The following algorithm generates a sample $\overline{X} \sim (X_1 | X_2 = x_2)$:

1. *Sample $Z = (X_1, X_2)^T \sim N(\mu, \Sigma)$*
2. *Compute $\overline{X} = X_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - X_2)$.*

Note that the computational effort to produce the conditioned Gaussian random variable $\overline{X}$ in Step 2 of Algorithm 2.2 is linear in the dimension n. Thus, this algorithm significantly reduces the computation time of Step 3 of Algorithm 2.1 when that step is repeated for each replica.

6 Efficient grids for a broad class of stochastic processes

In this section we discuss how the idea of threshold-dependent grids can be applied to stochastic processes other than Brownian Motion. We let $(X_t)_{t\in[0,1]}$ be a real-valued stochastic process and $t^*(b) := \arg\max_{t\in[0,1]} \mathbb{P}(X_t > b)$. For simplicity we here assume that $t \mapsto \mathbb{P}(X_t > b)$ is continuous and strictly increasing so that $t^*(b) = 1$ (but situations in which $t^*(b) \in (0,1)$ can be dealt with similarly, see also the discussion in Section 7).

As argued in the previous sections, it is efficient to let the position of the grid points depend on b. We constructed for Brownian Motion a grid by finding $T(b) = \{t_1(b), \ldots, t_n(b)\}$ such that

$$\mathbb{P}\big(\tau_b \in (0, t_k(b)] \,|\, \tau_b \in (0,1]\big) = \frac{k}{n}; \tag{2.18}$$

cf. (2.11). An inherent problem is that the class of processes for which the distribution of τ_b is known is very limited, so that the approach does not seem to be useful for relevant stochastic processes other than Brownian motion. We saw, however, that for Brownian Motion the $t_k(b)$ satisfying (2.18) also solve

$$\frac{\mathbb{P}(X_{t_k(b)} > b)}{\mathbb{P}(X_1 > b)} = \frac{k}{n}; \tag{2.19}$$

cf. (2.12). The idea now is to use the level-dependent (or: 'equiprobable') grid (2.19) for general real-valued processes. The major advantage of the grid (2.19) is that to calculate the position of the grid points t_k the sole prerequisite is that the process' *marginals* are known (rather than the distribution of τ_b). In addition, even if the marginal distributions of X_t are not available, but the *asymptotics* of $\mathbb{P}(X_t > b)$ (as $b \to \infty$) are, then a good approximation of this grid can be found. (In the sequel we write, for brevity, $T = \{t_1, \ldots, t_n\}$ instead of $T(b) = \{t_1(b), \ldots, t_n(b)\}$) and t^* instead of $t^*(b)$.)

We now provide the rationale behind the grid (2.19). Let T be a grid such that $t^* \in T$. Evidently, by the union bound,

$$\mathbb{P}(X_{t^*} > b) \leq w_T(b) \leq \sum_{t \in T} \mathbb{P}(X_t > b)$$

Now notice that if the grid T is such that for $t \in T \setminus \{t^*\}$

$$\mathbb{P}(X_t > b) = o\big(\mathbb{P}(X_{t^*} > b)\big), \quad \text{as } b \to \infty \tag{2.20}$$

then it does not make sense to include the point t for large b. Property (2.20) clearly compromises the performance of equidistant grids as $b \to \infty$. Considering however the grid points t_k of the threshold-dependent grid, as defined by (2.19), these will by design not experience (2.20).

To assess the performance of the above threshold-dependent grid (2.19), we introduce a measure of performance closely related to the relative bias. Note that when no formulas for $w(b)$ are available, nor it is known how to reliably approximate $w(b)$, we cannot determine the exact value of the relative bias. We now make the following two observations. (1) As $w_T(b) < w(b)$ for any choice

of T, the larger $w_T(b)$ is, the better; if $w_{T_1}(b) > w_{T_2}(b)$ for grids T_1, T_2, then also $\beta_{T_1}(b) < \beta_{T_2}(b)$. (2) The crude lower bound $w(b) \geq \mathbb{P}(X_{t^*} > b)$ provides us with a useful benchmark. Combining these two thoughts motivates the following performance measure of a grid T:

$$\gamma_T(b) := \frac{w_T(b)}{\mathbb{P}(X_{t^*} > b)}$$

Notice that for any T such that $t^* \in T$ we have

$$\gamma_T(b) \in \left[1, \frac{w(b)}{\mathbb{P}(X_{t^*} > b)}\right].$$

What is more, for any two grids T_1, T_2 we have $\gamma_{T_1}(b) \geq \gamma_{T_2}(b)$ if and only if $\beta_{T_1}(b) \leq \beta_{T_2}(b)$; this means that the bigger the $\gamma_T(b)$ is, the better. As our main aim is to efficiently approximate $w(b)$ using discrete-time approximations $w_T(b)$, we see that if $\gamma_T(b) \approx 1$ then there is little gain from using $w_T(b)$ over a deterministic estimator $\mathbb{P}(X_{t^*} > b)$.

In a series of examples we compare $\gamma_T(b)$ induced by (i) the threshold-dependent (equiprobable) grid and (ii) the equidistant grid of the same size; we consistently use $n = 100$ grid points. In all cases $t \mapsto \mathbb{P}(X_t > b)$ is a continuous, strictly increasing function (so that $t^* = 1$). The most important conclusion is that the experiments below uniformly indicate that the equiprobable grid outperforms the equidistant one, not only in the asymptotic regime, as threshold b grows large, but already for moderate values of b. This shows how the ideas the we developed earlier this chapter, that have provable optimality properties for Brownian motion, lead to an efficient estimation procedure for a much broader class of stochastic processes. In all examples, we observe that $\gamma_T(b)$ induced by the equidistant grid converges to 1, thus the corresponding $w_T(b)$ is asymptotically equivalent to $\mathbb{P}(X_{t^*} > b)$, as $b \to \infty$.

Example 2.1 (Brownian Motion with jumps)**.** Let $(X_t)_{t\in[0,1]}$ be a Brownian Motion with jumps, i.e.

$$X_t := B_t + N_t, \tag{2.21}$$

where B_t is a standard Brownian Motion and N_t is a standard Poisson process

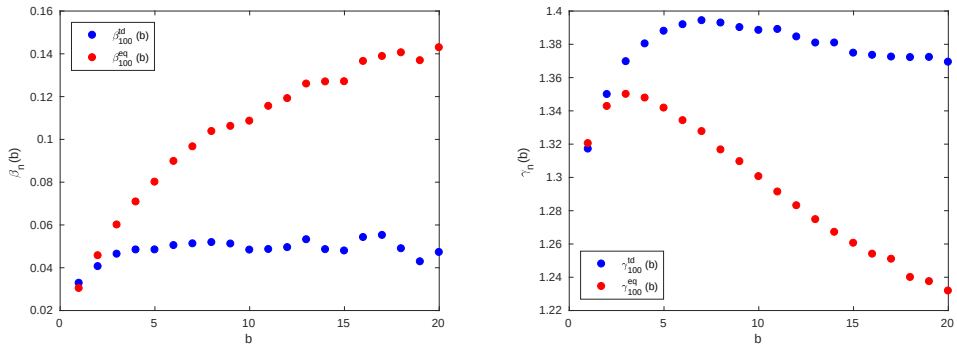


Figure 2.5: Brownian Motion with jumps. Plots of $\beta_n(b)$ (left) and $\gamma_n(b)$ (right) as a function of the threshold b for threshold-dependent and equidistant grids of size $n = 100$.

with intensity $\lambda = 1$.

Even though there are no closed-form expressions for $w(b)$, it is still possible to generate exact samples from $\sup_{t\in[0,1]} X_t$ (see [35, Section 10.1]). We can use this to construct an unbiased estimator of $w(b)$ and thus can estimate the relative bias of the tested grids. The results in Figure 2.5 show the substantial gain achieved by the level-dependent grid. The graphs look similar to those of Brownian Motion, which is indicative of the threshold-dependent grid having a uniformly bounded relative bias.

Example 2.2 (Ornstein-Uhlenbeck Process)**.** Let $(X_t)_{t\in[0,1]}$ be an Ornstein-Uhlenbeck process, i.e., a strong solution to the following SDE: with $X_0 = 0$,

$$\mathrm{d}X_t = -X_t\,\mathrm{d}t + \mathrm{d}W_t.$$

Then $(X_t)_{t\in[0,1]}$ is a zero-mean Markovian Gaussian process with covariance function

$$c(s,t) := \mathbb{C}\mathrm{ov}(X_s, X_t) = \frac{1}{2}\left(e^{-|t-s|} - e^{-(t+s)}\right).$$

The exact value of $w(b)$ is known only in terms of special functions, see [4] and it is not straightforwardly evaluated. However, the exact asymptotics of $w(b)$, as b grows large, *are* known:

$$w(b) = C\,\mathbb{P}(X_1 > b)(1 + o(1)), \text{ as } b \to \infty$$

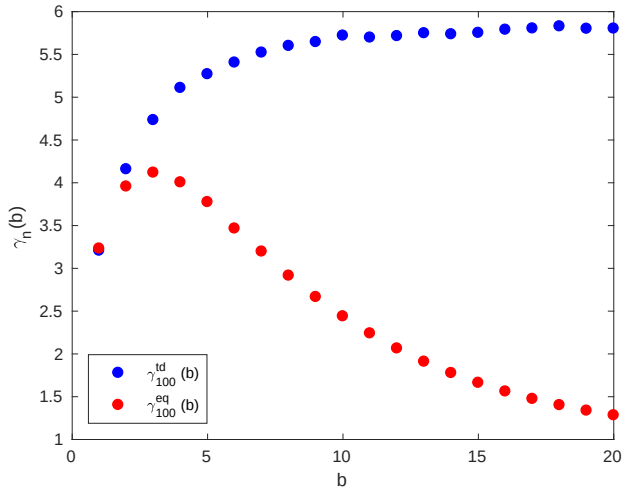


Figure 2.6: Ornstein-Uhlenbeck process. Plot of $\gamma_n(b)$ as a function of the threshold b for threshold-dependent and equidistant grids of size $n = 100$. Notice that with growing b the equidistant estimator converges to $\mathbb{P}(X_1 > b)$.

where C is a positive constant independent of b, see e.g. [34, Theorem 5.1] or the original theorem by [80]; this explains why for the level-dependent grid $\gamma_n(b)$ goes to a constant in Figure 2.6. Again the equidistant grid is significantly outperformed by the threshold-dependent grid.

Example 2.3 (Fractional Brownian Motion)**.** Let $(X_t)_{t\in[0,1]}$ be a fractional Brownian Motion (fBM) with a Hurst parameter $H \in (0, 1)$, that is a zero-mean Gaussian process with the covariance function

$$C_H(s,t) := \mathbb{C}\mathrm{ov}(X_s, X_t) = \frac{1}{2}\left(s^{2H} + t^{2H} - |t-s|^{2H}\right).$$

Observe that fBM with Hurst parameter $H = 1/2$ is a standard Brownian Motion. For any H we have $C_H(t,t) = t^{2H}$ (strictly increasing variance in time) and thus $t^* = 1$.

The exact value of the probability $w(b)$ for $H \neq 1/2$ remains unknown.

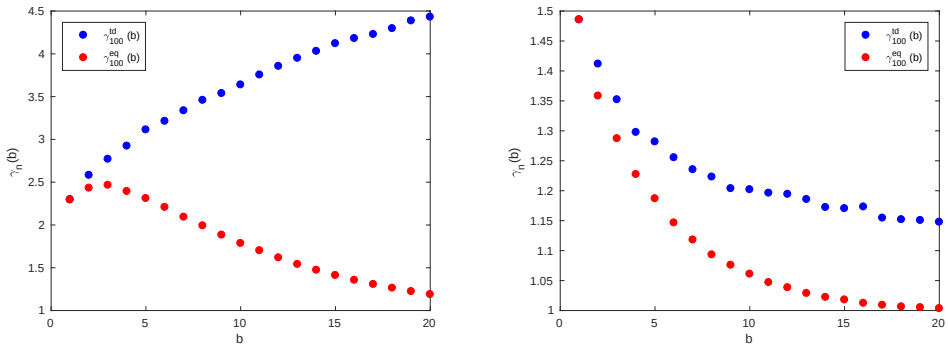


Figure 2.7: fBm with Hurst parameter $H = 0.4$ (left) and $H = 0.6$ (right). Plot of $\gamma_n(b)$ as a function of the threshold b for threshold-dependent and equidistant grids of size $n = 100$.

However, like in Example 2.2, the exact asymptotics of $w(b)$ are known:

$$w(b) = \begin{cases} C_H b^{1/H-2} \mathbb{P}(X_1 > b)(1 + o(1)), & \text{for } H \in (0, \frac{1}{2}) \\ \mathbb{P}(X_1 > b)(1 + o(1)), & \text{for } H \in (\frac{1}{2}, 1) \end{cases}$$

where C_H is a constant only depending on H; we again refer to [34, Theorem 5.1] or the original theorem by [80]. We apply threshold-dependent grids in these two different asymptotic regimes for $H = 0.4$ and $H = 0.6$, see the results in Figure 2.7. Again the threshold-dependent grid performs considerably better. In case $H = 0.4$ the above asymptotic result explains why for the level-dependent grid $\gamma_n(b)$ keeps increasing ($w(b)/\mathbb{P}(X_1 > b)$ behaves as the increasing function $b^{1/H-2}$). In case $H = 0.6$, again using the asymptotic result, $\gamma_n(b) \to 1$ as b grows large, both for the equidistant grid and for the threshold-dependent grid (equivalently, the relative bias vanishes for both as $b \to \infty$). Note however that with the threshold-dependent grid, $\gamma_n(b)$ tends to 1 slower than with the equidistant grid, as can be seen in Figure 2.7 (right panel), showing the more favorable performance of the threshold-dependent grid.

7 Concluding remarks and discussion

In this chapter we have demonstrated that the errors due to time discretization when estimating threshold-crossing probabilities $w(b)$ can be significantly re-

duced by using other grids than the commonly used equidistant grid. We have analyzed this in considerable detail for the case of standard Brownian Motion. In particular, we have shown that in order to control the error as b grows large, it suffices to properly *shift* the grid points instead of refining the grid with more and more points. At the same time, controlling the error using equidistant grids requires *quadratic* growth of the number of grid points, as b grows large.

Numerical estimation is evidently not needed for Brownian Motion due to the availability of analytical results. Our chapter however indicates that the underlying ideas can be used to construct efficient grids for a broad class of stochastic processes (notably, Lévy processes and Gaussian processes, such as fractional Brownian Motion). The results presented in this chapter are intended to develop valuable insight and useful heuristics for tackling the estimation of tail probabilities of these more general classes of processes. We have demonstrated such heuristics for several processes in Section 6. There, we presented a procedure, that is empirically shown to work well for stochastic process $(X_t)_{t\in[0,1]}$ of which the marginal distributions are known:

(i) Identify
$$t^*(b) := \arg\max_{t\in[0,1]} \mathbb{P}(X_t > b);$$
in case $(X_t)_{t\in[0,1]}$ is a zero-mean Gaussian process, t^* is a point of maximal variance, i.e., $\arg\max_{t\in[0,1]} \mathbb{V}\mathrm{ar}\, X_t$. As argued, for many key models we have that $t^* = 1$.

(ii) Construct a grid $T = \{t_1, \ldots, t_n\}$ clustered around $t^*(b)$, such that t_k solves (2.19), for $k \in \{1, \ldots, n\}$.

As we pointed out, even if the marginal distribution of X_t is not available but only the corresponding asymptotics, as $b \to \infty$, this procedure can be applied. It is also noted that it is straightforward to compare two different grids: the larger the value of $w_T(b)$, the closer it is to the target quantity $w(b)$.

A natural question that arises in relation to Theorem 2.2 is whether we can find a grid that is even better than the one defined in (2.9). Constructing an *optimal n-grid* $T_n^*(b)$, i.e. a grid of size n that minimizes the relative bias for a given b, remains elusive. However we have been able to find an explicit formula

for an optimal 2-grid, namely $T_2^*(b) = \{t_1^*(b), t_2^*(b)\}$, with

$$t_1^*(b) = \frac{\pi b^2}{4}\left(\sqrt{1+\frac{8}{\pi b^2}} - 1\right), \quad \text{and} \quad t_2^*(b) = 1$$

where $\lim_{b\to\infty} \beta_{T_2^*(b)}(b) = 1 - \frac{1}{2}\Phi(\sqrt{2/\pi}) - \frac{1}{4}e^{-1/\pi} \approx 0.4244$. For comparison, the threshold-dependent grid defined in (2.9) yields $\lim_{b\to\infty} \beta_2(b) = \frac{3}{8} + \frac{1}{2}\Phi(-\sqrt{2\log 2}) \approx 0.4348$, hence the grid (2.9) is not minimizing the bias (although the difference with the optimal 2-grid is small). Additionally, we were able to prove that for an optimal n-grid, $T_n^*(b) = \{t_1^*(b), \dots, t_n^*(b)\}$, the limits $\lim_{b\to\infty} b^2(1 - t_k^*(b))$ must exist, and are all finite and pairwise distinct. As a result we were able to numerically calculate the limit $\lim_{b\to\infty} \beta_{T_3^*(b)} \approx 0.3796$. Finding optimal grids for larger n remains an open problem. We note, however, that with the threshold-dependent grid we can bound the relative bias uniformly in b (see Theorem 2.2) and in this sense the grid (2.9) is already (asymptotically) optimal.

8 Proofs of Lemmas 2.1, 2.4 and 2.5

Proof of Lemma 2.1. Notice that the events $\{\sup_{t\in[0,1]} B_t > b\}$ and $\{\tau_b \in (0,1]\}$ are equivalent. We thus find

$$\begin{aligned}
w(b)\,\beta_T(b) &= \mathbb{P}\Big(\sup_{t\in T} B_t < b, \sup_{t\in[0,1]} B_t > b\Big) = \mathbb{P}\Big(\sup_{t\in T} B_t < b, \tau_b \in [0,1]\Big) \\
&= \sum_{j=1}^{n} \mathbb{P}\Big(\sup_{t\in\{t_j,\dots,t_n\}} B_t < b, \tau_b \in (t_{j-1}, t_j]\Big) \\
&= \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \mathbb{P}\Big(\sup_{t\in\{t_j,\dots,t_n\}} B_t < b \mid B_s = b\Big)\,\mathbb{P}(\tau_b \in \mathrm{d}s) \\
&= \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \mathbb{P}\Big(B_{t_j - s} < 0, \dots, B_{t_n - s} < 0\Big)\,\mathbb{P}(\tau_b \in \mathrm{d}s)
\end{aligned}$$

To prove the upper bound we use the fact that $\mathbb{P}(B_{t_j-s} < 0, \dots, B_{t_n-s} < 0)$ is a non-increasing function of $s \in [t_{j-1}, t_j]$ (see Appendix A, Transformation

(T2)), so that

$$\begin{aligned} w(b)\,\beta_T(b) &\le \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \mathbb{P}\Big(B_{t_j - t_{j-1}} < 0, \dots, B_{t_n - t_{j-1}} < 0\Big)\, \mathbb{P}(\tau_b \in \mathrm{d}s) \\ &= \sum_{j=1}^{n} \mathbb{P}\Big(B_{t_j - t_{j-1}} < 0, \dots, B_{t_n - t_{j-1}} < 0\Big) \cdot \mathbb{P}\big(\tau_b \in (t_{j-1}, t_j]\big). \end{aligned}$$

Dividing both sides of the inequality by $w(b) = \mathbb{P}(\tau_b \in (0,1])$ gives $\beta_T(b) \le \bar{\beta}_T(b)$. To prove the lower bound we use (B.IV), so as to obtain

$$\begin{aligned} w(b)\,\beta_T(b) &= \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \mathbb{P}\Big(B_{t_j - s} < 0, \dots, B_{t_n - s} < 0\Big)\, \mathbb{P}(\tau_b \in \mathrm{d}s) \\ &\ge \sum_{j=1}^{n-1} \int_{t_{j-1}}^{t_j} \frac{1}{2}\, \mathbb{P}\Big(B_{t_{j+1} - t_j} < 0, \dots, B_{t_n - t_j} < 0\Big)\, \mathbb{P}(\tau_b \in \mathrm{d}s) + \frac{1}{2}\, \mathbb{P}\big(\tau_b \in (t_{n-1}, t_n]\big) \\ &\ge \sum_{j=1}^{n-1} \frac{1}{2}\, \mathbb{P}\Big(B_{t_{j+1} - t_j} < 0, \dots, B_{t_n - t_j} < 0\Big) \mathbb{P}\big(\tau_b \in (t_{j-1}, t_j]\big) + \frac{1}{2}\, \mathbb{P}\big(\tau_b \in (t_{n-1}, t_n]\big). \end{aligned}$$

Dividing both sides of the inequality by $w(b)$ leads to $\beta_T(b) \ge \underline{\beta}_T(b)$ and concludes the proof. □

Proof of Lemma 2.4. Let $h := \max_{k=1,\dots,n}(t_k - t_{k-1})$. We transform the grid $T = \{t_1, \dots, t_n\}$ with Transformations (T1)–(T3), see Appendix A, in such a way that after all transformations we end up with $\{h, \dots, Nh\}$.

1. Using Transformation (T2), translate the grid to the right by $h - t_1$, i.e., put

$$t_j := t_j + h - t_1 \quad \text{for all } j \in \{1, \dots, n\}$$

2. Put $\sigma_1 := 1$, $c_1 = 1$ and $k := 2$. While $k \le N$ do:

   - Put $\sigma_k := \inf\{j : t_j \ge kh\}$.
   - Using Transformation (T3), contract the grid after time $t_{\sigma_{k-1}}$ by a

factor c_k, where c_k is defined by $h/(t_{\sigma_k} - t_{\sigma_{k-1}})$. Formally, we put

$$t_j := \begin{cases} t_j, & j \in \{1, \dots, \sigma_{k-1}\} \\ t_{\sigma_{k-1}} + c_k(t_j - t_{\sigma_{k-1}}), & j \in \{\sigma_{k-1}+1, \dots, n\} \end{cases}$$

Notice that after this operation $t_{\sigma_k} = kh$.

- Put $k := k + 1$.

3. Using Transformation (T1), delete all $t_k \notin \{h, \dots, hN\}$.

Now we prove that the algorithm is well-defined, more precisely, we confirm that all σ_k's exist. First, see that σ_1 is well-defined. By induction, assume that σ_k is well-defined and prove that σ_{k+1} is well-defined as well. Notice that after the kth loop in Step 2 of the algorithm, the distances between the points shrunk at most by a factor $p_k = \prod_{j=1}^{k} c_j$ compared with the initial maximal distance h. Moreover, we observe that

$$c_k = \frac{h}{t_{\sigma_k} - t_{\sigma_{k-1}}} \geq \frac{h}{h + \max_{j > \sigma_{k-1}} |t_{j+1} - t_j|} \geq \frac{1}{1 + \prod_{j=1}^{k-1} c_j} \tag{2.22}$$

We prove by induction that $p_k = \prod_{j=1}^{k} c_j \geq \frac{1}{k}$ for all $k \in \{1, \dots, N\}$. Obviously $p_2 = c_2 \geq \frac{1}{2}$. Assume that $p_{k-1} \geq \frac{1}{k-1}$. After multiplying inequality (2.22) by p_{k-1} we obtain

$$p_k \geq \frac{p_{k-1}}{1 + p_{k-1}} = 1 - \frac{1}{1 + p_{k-1}} \geq 1 - \frac{1}{1 + \frac{1}{k-1}} = \frac{1}{k},$$

which ends the inductive proof. Next, in order to show that σ_{k+1} is well defined for $k \in \{1, \dots N-1\}$ it suffices to prove that the endpoint t_n, after the kth loop of Step 2, is greater than $h(k+1)$. We prove a stronger statement, namely that the endpoint t_n after being shrunk by a factor p_k is still greater than $h(k+1)$, i.e. $h(k+1) \leq t_n p_k$. By the definition of N, h satisfies the inequality $h \leq t_n/N^2$, thus

$$h(k+1) \leq \frac{t_n(k+1)}{N^2} = \frac{t_n(k+1)}{N^2 p_k} p_k = \frac{k(k+1)}{N^2} t_n p_k \leq t_n p_k,$$

which concludes the proof that σ_{k+1} is well-defined. As all transformations used

in steps 1-3 satisfy (2.26) we have

$$\mathbb{P}\big(B_{t_1} > 0, \dots, B_{t_n} > 0\big) \leq \mathbb{P}\big(B_h > 0, \dots, B_{Nh} > 0\big)$$

We finish the proof by observing that $\mathbb{P}\big(B_h > 0, \dots, B_{Nh} > 0\big) = \mathbb{P}\big(B_1 > 0, \dots, B_N > 0\big)$, due to the scaling property of Brownian Motion. □

Proof of Lemma 2.5. Notice that the grid points $t_k^n(b)$ defined in (2.9) depend only on the threshold b and the *ratio* $\frac{k}{n} \in [0,1]$. We are able to extend the definition of $t_k^n(b)$ to $t : (0,1] \times (0,\infty) \to [0,1]$,

$$t(s,b) := \Big(\frac{b}{\Phi^{-1}\big(s\,\Phi(-b)\big)}\Big)^2$$

such that $t_k^n(b) = t(\frac{k}{n}, b)$. Equivalently, $t(s,b)$ can be defined as the unique solution to

$$\Phi\Big(-\frac{b}{\sqrt{t(s,b)}}\Big) = s\,\Phi(-b) \tag{2.23}$$

This extension makes it possible to inspect the derivative of $t_k^n(b)$ with respect to the ratio $\frac{k}{n}$. Using the extension function of $t_k^n(b)$, we aim to prove the more general statement that for $0 < s_1 < s_2 < 1$,

$$\frac{1 - t(s_1,b)}{t(s_2,b) - t(s_1,b)} \geq \frac{-\log s_1}{\log s_2 - \log s_1} \iff \frac{1 - t(s_1,b)}{-\log s_1} \leq \frac{1 - t(s_2,b)}{-\log s_2}. \tag{2.24}$$

Moreover, using the definition (2.23) we may substitute

$$s = \Phi\Big(-\frac{b}{\sqrt{t(s,b)}}\Big)/\Phi(-b)$$

and arrive at another equivalent form of inequality (2.24):

$$\frac{1-t(s_1,b)}{\log\big(\Phi(-b)\big) - \log\big(\Phi(-b/\sqrt{t(s_1,b)})\big)} \leq \frac{1-t(s_2,b)}{\log\big(\Phi(-b)\big) - \log\big(\Phi(-b/\sqrt{t(s_2,b)})\big)}, \tag{2.25}$$

which follows from (B.IX). For part (b) see that the density of the first hitting time,

$$\mathbb{P}(\tau_b \in \mathrm{d}s) = \frac{b}{\sqrt{2\pi}} s^{-3/2} e^{-b^2/(2s)}\,\mathrm{d}s, \quad \text{for } s > 0$$

is an increasing function on the interval $s \in [0, \frac{b^2}{3}]$ and thus part (b) follows from the second definition of the grid points $t_k^n(b)$ in (2.10). □

Appendix

A Grid transformations

Let $T = \{t_1, \ldots, t_n\}$ with $0 < t_1 < \ldots < t_n < \infty$. We introduce three *grid transformations*, i.e. operations $T \mapsto \widetilde{T}$ satisfying

$$\mathbb{P}\big(B_t > 0 \text{ for all } t \in T\big) \leq \mathbb{P}\big(B_t > 0 \text{ for all } t \in \widetilde{T}\big). \tag{2.26}$$

For abbreviation, in the following we denote $z_n := \mathbb{P}(B_{t_1} > 0, \ldots, B_{t_n} > 0)$.

(T1) **Deleting**. For any $k \in \{1, \ldots, n\}$

$$z_n \leq \mathbb{P}\Big(B_{t_1} > 0, \ldots, B_{t_{k-1}} > 0, B_{t_{k+1}} > 0, \ldots, B_{t_n} > 0\Big)$$

(T2) **Translation to the right of the whole sequence**. For any $s > 0$

$$z_n \leq \mathbb{P}\Big(B_{t_1+s} > 0, \ldots, B_{t_n+s} > 0\Big)$$

(T3) **Contraction of time after some point**. For $k < n$ and $c \in (0,1)$:

$$z_n \leq \mathbb{P}\Big(B_{t_1} > 0, \ldots, B_{t_k} > 0, B_{t_k + c(t_{k+1} - t_k)} > 0, \ldots, B_{t_k + c(t_n - t_k)} > 0\Big)$$

The proof that transformations (T1)-(T3) satisfy (2.26) is in Appendix C.

B Miscellaneous results

Let $\Phi(\cdot)$ denote the standard normal cumulative distribution function and $\varphi(\cdot)$ the standard normal density function. Below we list various results that we use throughout this chapter. Results (B.I) and (B.II) are standard, B.III is due to [91]; the proofs of the remaining results are included in Appendix C.

(B.I) For $x > 0$:

$$\frac{x}{1+x^2} \leq \frac{\Phi(-x)}{\varphi(x)} \leq \frac{1}{x} \tag{2.27}$$

(B.II) As $x \to \infty$,

$$\lim_{x \to \infty} \frac{\Phi(-x)}{\frac{1}{x}\varphi(x)} \to 1.$$

(B.III) For $x > -1$:

$$\frac{2}{x + (x^2+4)^{1/2}} \leq \frac{\Phi(-x)}{\varphi(x)} \leq \frac{4}{3x + (x^2+8)^{1/2}} \tag{2.28}$$

(B.IV) Let $0 < t_1 < \ldots < t_n < \infty$, then:

$$\mathbb{P}\Big(B_{t_1} > 0, \ldots, B_{t_n} > 0\Big) \geq \frac{1}{2}\,\mathbb{P}\Big(B_{t_2 - t_1} > 0, \ldots, B_{t_n - t_1} > 0\Big)$$

(B.V) Let $T = \{t_1, \ldots, t_n\}$, where $t_j := \frac{j}{n}$, $\tau_b := \inf\{t \geq 0 : B_t \geq b\}$ and $b > 0$, then:

$$\mathbb{P}\Big(\tau_b \in (t_{j-1}, t_j] \;\Big|\; \tau_b \in (0,1]\Big) \leq \frac{b\,(b + \sqrt{b^2+4})}{4} \cdot \frac{\sqrt{n}}{(j-1)^{3/2}} e^{-\frac{b^2}{2} \cdot \left(\frac{n}{j} - 1\right)}$$

and

$$\mathbb{P}\Big(\tau_b \in (t_{j-1}, t_j] \;\Big|\; \tau_b \in (0,1]\Big) \geq \frac{b\,(3b+\sqrt{b^2+8})}{8} \cdot \frac{\sqrt{n}}{j^{3/2}} e^{-\frac{b^2}{2}\cdot\left(\frac{n}{j-1}-1\right)}$$

for $j \in \{2, \dots n\}$.

(B.VI) Let $f : (0,\infty) \times (0,1) \to (0,\infty)$ such that

$$f(b,x) := \frac{b}{\sqrt{1-x}} x^{-3/2} e^{-\frac{b^2}{2}(1/x-1)}$$

Then $f(b,x)$ is an increasing function of x, when $b \geq 1$.

(B.VII) Let $f : (0,\infty) \to (0,\infty)$ such that

$$f(x) := \frac{\Phi(-x)}{\varphi(x)}$$

Then f is a strictly decreasing function.

(B.VIII) Let $f : (0,\infty) \to (0,\infty)$ such that

$$f(x) := \frac{\Phi(-x)}{\frac{1}{x}\varphi(x)}$$

Then f is a strictly increasing function.

(B.IX) Let $f : [0,1] \to [0,\infty)$ be such that

$$f(t) := \begin{cases} 0, & t = 0; \\ \dfrac{1-t}{\log\big(\Phi(-b)\big) - \log\big(\Phi(-b/\sqrt{t})\big)}, & t \in (0,1); \\ \dfrac{2\,\Phi(-b)}{b\,\varphi(b)}, & t = 1. \end{cases}$$

Then f is continuous and increasing.

C Supplementary Materials

Supplementary materials consist of (i) proof of Proposition 2.1, (ii) proof of Lemma 2.3, (iii) proof that Transformations (T1)-(T3) satisfy (2.26), and (iv) proofs of Results (B.IV)-(B.IX).

Proof of Proposition 2.1. In part (a) of Theorem 2.1 it has been proven already that $\beta_n(b) \leq C_0 b n^{-1/2}$. Thus, when b is fixed it is straightforward that the upper bound in the assertion of the theorem holds.
The lower bound developed in Lemma 2.1 reads $\beta_n(b) \geq \frac{1}{2}\sum_{j=1}^{n-1} a_{j+1} \cdot w_j(b) + \frac{1}{2} w_n(b)$. Since we have $a_j < a_{j+1}$ for the equidistant grid and all a_j and w_j are non-negative, we may use the weaker inequality

$$\beta_n(b) \geq \frac{1}{2}\sum_{j=2}^{n} a_j \cdot w_j(b).$$

In the following we use Lemma 2.2 for a lower bound on terms a_j and (B.V) for a lower bound on w_j.

$$\begin{aligned}
\sum_{j=2}^{n} a_j \cdot w_j(b) &\geq \frac{b\,(3b + \sqrt{b^2+8})}{8} \sum_{j=2}^{n} C_1^*(n-j+1)^{-1/2} \frac{\sqrt{n}}{j^{3/2}}\, e^{-\frac{b^2}{2}\left(\frac{n}{j-1}-1\right)} \\
&\geq C\, n^{-1/2} \sum_{j=2}^{n} \frac{1}{n} \left(\frac{b}{\sqrt{1-\frac{j-1}{n}}} \left(\frac{j-1}{n}\right)^{-3/2} e^{-\frac{b^2}{2}\left(\frac{n}{j-1}-1\right)} \right) \\
&\geq C\, n^{-1/2} \int_0^1 \frac{b}{\sqrt{1-x}}\, x^{-3/2}\, e^{-\frac{b^2}{2}(1/x-1)}\, \mathrm{d}x \qquad (2.29) \\
&\geq C\, n^{-1/2},
\end{aligned}$$

where C is a positive constant independent of n (but dependent on b) that may vary from line to line. To arrive at (2.29) we use the convergence of the Riemann sum, noting that b is fixed and that the function

$$f(x) := \frac{b}{\sqrt{1-x}}\, x^{-3/2}\, e^{-\frac{b^2}{2}(1/x-1)}$$

is integrable on $(0,1)$. This concludes the proof. □

Proof of Lemma 2.3. Recall the definitions of $a_j(b)$ and $w_j(b)$, and $\bar{\beta}_T(b) :=$

$\sum_{j=1}^{n} a_j(b)w_j(b)$. Notice that if we put $t_k = \frac{k}{n}$, then by the scaling property of Brownian Motion

$$a_j(b) = \mathbb{P}\big(B_1 < 0, \ldots, B_{1+n-j} < 0\big)$$

and thus $a_1 < a_2 < \ldots < a_n$ (since the $a_j(b)$ s are independent of b, we abbreviate $a_j := a_j(b)$).

Assume that for any $0 < b_1 < b_2$ there exists $k \in \{1, \ldots, n-1\}$ such that

$$w_j(b_1) \geq w_j(b_2), \text{ for } j \leq k \quad \text{and} \quad w_j(b_1) \leq w_j(b_2), \text{ for } j > k. \tag{2.30}$$

Since the weights $w_j(b)$ must satisfy $\sum_{j=1}^{n} w_j(b) = 1$ we have $\sum_{j=1}^{n} \big(w_j(b_2) - w_j(b_1)\big) = 0$ and thus

$$\sum_{j=k+1}^{n} \big(w_j(b_2) - w_j(b_1)\big) = \sum_{j=1}^{k} \big(w_j(b_1) - w_j(b_2)\big).$$

Finally,

$$\begin{aligned}
\bar{\beta}_T(b_2) - \bar{\beta}_T(b_1) &= \sum_{j=1}^{n} a_j\big(w_j(b_2) - w_j(b_1)\big) \\
&\geq a_{k+1} \sum_{j=k+1}^{n} \big(w_j(b_2) - w_j(b_1)\big) - a_k \sum_{j=1}^{k} \big(w_j(b_2) - w_j(b_1)\big) \\
&= \big(a_{k+1} - a_k\big) \sum_{j=k+1}^{n} \big(w_j(b_2) - w_j(b_1)\big) > 0.
\end{aligned}$$

For the remainder of the proof we prove the existence of $k \in \{1, \ldots, n-1\}$ satisfying (2.30). Let $\tau_b := \inf\{t \geq 0 : B_t \geq b\}$ be the first hitting time of level b and let $f(b,t)$ be the density of τ_b given that $\tau_b \leq 1$, i.e.,

$$f(b,t) := \frac{b}{2\Phi(-b)} t^{-3/2} \varphi\left(-\frac{b}{\sqrt{t}}\right),$$

where $b > 0$, $t \in (0,1)$, and $\varphi(\cdot)$ denotes the density of a standard normal random variable. We will prove that for any $0 < b_1 < b_2$ there exists t^* such

that:

$$f(b_1,t) > f(b_2,t) \text{ for } t \in (0,t^*) \quad \text{and} \quad f(b_1,t) < f(b_2,t) \text{ for } t \in (t^*,1]. \tag{2.31}$$

Then the weights

$$w_j(b) := \int_{t_{j-1}}^{t_j} f(b,t)\,\mathrm{d}t$$

are decreasing for all $\frac{j}{n} \leq t^*$, and increasing for all $\frac{j}{n} \geq \frac{1}{n} + t^*$. If nt^* is not an integer, it is not known whether $w_{[t^*n]+1}(b)$ increases or not, but for sure there exists $k \in \{1,\ldots,n-1\}$ satisfying (2.30). For the remainder we prove the existence of t^* satisfying (2.31). For $t \in (0,1)$:

$$\begin{aligned}
f(b_1,t) - f(b_2,t) &= \frac{1}{2}\,t^{-3/2}\left(\frac{b_1\,\varphi\left(-\frac{b_1}{\sqrt{t}}\right)}{\Phi(-b_1)} - \frac{b_2\,\varphi\left(-\frac{b_2}{\sqrt{t}}\right)}{\Phi(-b_2)}\right) \\
&= \frac{1}{2}\,t^{-3/2}\frac{b_2\,\varphi\left(-\frac{b_2}{\sqrt{t}}\right)}{\Phi(-b_2)}\left(\frac{b_1\,\varphi\left(-\frac{b_1}{\sqrt{t}}\right)\Phi(-b_2)}{b_2\,\varphi\left(-\frac{b_2}{\sqrt{t}}\right)\Phi(-b_1)} - 1\right) \\
&= \underbrace{\frac{1}{2}\,t^{-3/2}\frac{b_2\,\varphi\left(-\frac{b_2}{\sqrt{t}}\right)}{\Phi(-b_2)}}_{>0}\ \underbrace{\left(e^{\frac{b_2^2-b_1^2}{2t}}\,\frac{b_1\,\Phi(-b_2)}{b_2\,\Phi(-b_1)} - 1\right)}_{=:g(t)}
\end{aligned}$$

We have that $\lim_{t\to 0^+} g(t) = +\infty$ and $g(1) < 0$ (which follows from (B.VII)) and that $g(\cdot)$ is strictly decreasing, hence $g(\cdot)$ has exactly one zero t^* and $g(t) > 0$ for $t < t^*$ and $g(t) < 0$ for $t > t^*$. The observation that $\operatorname{sign}(f(b_1,t) - f(b_2,t)) = \operatorname{sign}(g(t))$ concludes the proof. □

Proof that Transformations (T1)-(T3) *satisfy (2.26).* Assertion (T1) is straightforward to verify. To show (T2) observe that

$$\begin{aligned}
z_n &= \int_0^\infty \mathbb{P}\big(B_{t_2} > 0,\ldots,B_{t_n} > 0 \mid B_{t_1} = x\big)\frac{1}{\sqrt{2\pi t_1}}e^{-x^2/(2t_1)}\,\mathrm{d}x \\
&= \int_0^\infty \mathbb{P}\big(B_{t_2-t_1} < x,\ldots,B_{t_n-t_1} < x\big)\frac{1}{\sqrt{2\pi t_1}}e^{-x^2/(2t_1)}\,\mathrm{d}x \\
&= \int_0^\infty \mathbb{P}\big(B_{t_2-t_1} < y\sqrt{t_1},\ldots,B_{t_n-t_1} < y\sqrt{t_1}\big)\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,\mathrm{d}y
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_0^\infty \mathbb{P}\big(B_{t_2-t_1} < y\sqrt{t_1+s}, \dots, B_{t_n-t_1} < y\sqrt{t_1+s}\big) \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,\mathrm{d}y \\
&= \int_0^\infty \mathbb{P}\big(B_{t_2-t_1} < x, \dots, B_{t_n-t_1} < x\big) \frac{1}{\sqrt{2\pi(t_1+s)}} e^{-x^2/(2(t_1+s))}\,\mathrm{d}x \\
&= \mathbb{P}\big(B_{t_1+s} > 0, \dots, B_{t_n+s} > 0\big),
\end{aligned}
$$

which ends the proof. For abbreviation we denote $z_k(x) := \mathbb{P}\big(B_{t_1} > 0, \dots, B_{t_{k-1}} > 0 \mid B_{t_k} = x\big)$. To show (T3) observe that

$$
\begin{aligned}
z_n &= \int_0^\infty z_k(x)\, \mathbb{P}\big(B_{t_{k+1}} > 0, \dots, B_{t_n} > 0 \mid B_{t_k} = x\big) \frac{1}{\sqrt{2\pi t_k}} e^{-x^2/(2t_k)}\,\mathrm{d}x \\
&= \int_0^\infty z_k(x)\, \mathbb{P}\big(B_{t_{k+1}-t_k} < x, \dots, B_{t_n-t_k} < x\big) \frac{1}{\sqrt{2\pi t_k}} e^{-x^2/(2t_k)}\,\mathrm{d}x \\
&\leq \int_0^\infty z_k(x)\, \mathbb{P}\left(B_{t_{k+1}-t_k} < \frac{x}{\sqrt{c}}, \dots, B_{t_n-t_k} < \frac{x}{\sqrt{c}}\right) \frac{1}{\sqrt{2\pi t_k}} e^{-x^2/(2t_k)}\,\mathrm{d}x \\
&= \int_0^\infty z_k(x)\, \mathbb{P}\big(B_{c(t_{k+1}-t_k)} < x, \dots, B_{c(t_n-t_k)} < x\big) \frac{1}{\sqrt{2\pi t_k}} e^{-x^2/(2t_k)}\,\mathrm{d}x \\
&= \mathbb{P}\big(B_{t_1} > 0, \dots, B_{t_k} > 0, B_{t_k+c(t_{k+1}-t_k)} > 0, \dots, B_{t_k+c(t_n-t_k)} > 0\big)
\end{aligned}
$$

which concludes the proof. □

Proof of (B.IV). Note that

$$
\begin{aligned}
z_n &= \int_0^\infty \mathbb{P}\Big(B_{t_2} > 0, \dots, B_{t_n} > 0 \mid B_{t_1} = x\Big) \frac{1}{\sqrt{2\pi t_1}} e^{-x^2/(2t_1)}\,\mathrm{d}x \\
&= \int_0^\infty \mathbb{P}\Big(B_{t_2-t_1} < x, \dots, B_{t_n-t_1} < x\Big) \frac{1}{\sqrt{2\pi t_1}} e^{-x^2/(2t_1)}\,\mathrm{d}x \\
&\geq \int_0^\infty \mathbb{P}\Big(B_{t_2-t_1} < 0, \dots, B_{t_n-t_1} < 0\Big) \frac{1}{\sqrt{2\pi t_1}} e^{-x^2/(2t_1)}\,\mathrm{d}x \\
&= \frac{1}{2}\,\mathbb{P}\Big(B_{t_2-t_1} > 0, \dots, B_{t_n-t_1} > 0\Big),
\end{aligned}
$$

which concludes the proof. □

Proof of (B.V). Using the mean value theorem and monotonicity of $\varphi(\cdot)$ on the negative half-line, we have $|\Phi(-x) - \Phi(-y)| \leq |x-y| \cdot \varphi(-y)$ for $0 < y < x$. Furthermore,

$$
\mathbb{P}\Big(\tau_b \in (t_{j-1}, t_j]\Big) = 2\Phi(-b/\sqrt{t_j}) - 2\Phi(-b/\sqrt{t_{j-1}}) \leq 2\left(\frac{b}{\sqrt{t_j}} - \frac{b}{\sqrt{t_j}}\right) \varphi\left(\frac{b}{\sqrt{t_j}}\right)
$$

Thus, for $b > 0$ and $j \in \{2, \dots, n\}$, after substituting $t_j = \frac{j}{k}$, the above combined with the inequality (B.III) yield:

$$\begin{aligned}
\mathbb{P}\Big(\tau_b \in (t_{j-1}, t_j] \;\Big|\; \tau_b \in (0,1]\Big) &\leq \frac{1}{\Phi(-b)} \left(\frac{b}{\sqrt{t_{j-1}}} - \frac{b}{\sqrt{t_j}}\right) \varphi\left(\frac{b}{\sqrt{t_j}}\right) \\
&= \frac{b\sqrt{n}}{\sqrt{2\pi}\Phi(-b)} \frac{\sqrt{j} - \sqrt{j-1}}{\sqrt{(j-1)j}} e^{-b^2 n/(2j)} \\
&\leq \frac{b\sqrt{n}}{2\sqrt{2\pi}\Phi(-b)} \frac{1}{(j-1)^{3/2}} e^{-b^2 n/(2j)} \\
&\leq \frac{b\,(b + \sqrt{b^2+4})}{4} \frac{\sqrt{n}}{(j-1)^{3/2}} e^{-\frac{b^2}{2}\left(\frac{n}{j} - 1\right)}
\end{aligned}$$

The proof of the second inequality is analogous. □

Proof of (B.VI). It suffices to prove that $\frac{\mathrm{d}}{\mathrm{d}x} f(b,x) \geq 0$ for $b \geq 1$. See that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} f(b,x) &= b e^{b^2/2} \frac{\mathrm{d}}{\mathrm{d}x} \frac{e^{-b^2/(2x)}}{\sqrt{1-x}\, x^{3/2}} = \frac{b e^{b^2/2}}{(1-x)x^3} \cdot \\
&\quad \cdot \Big(\frac{b^2}{2x^2} e^{-b^2/(2x)} \sqrt{1-x}\, x^{3/2} - e^{-b^2/(2x)} \Big(- \frac{x^{3/2}}{2\sqrt{1-x}} + \frac{3}{2}\sqrt{x(1-x)} \Big) \Big) \\
&= \frac{b e^{b^2/2\,(1-1/x)}}{2(1-x)^{3/2} x^{7/2}} \left(b^2(1-x) + x^2 - 3x(1-x) \right) \\
&= \underbrace{\frac{b e^{b^2/2\,(1-1/x)}}{2(1-x)^{3/2} x^{7/2}}}_{>0} \Big(\underbrace{4x^2 - (b^2+3)x + b^2}_{=:g(x)} \Big)
\end{aligned}$$

Note that $g(x)$ has at most one root when $b \in [1,3]$ thus $g(x) \geq 0$ for $b \in [1,3]$. Moreover, when $b > 3$, then $g'(x) = 8x - (b^2+3) < -1$ (for $x \in [0,1]$) thus $g(x)$ is strictly decreasing for $x \in [0,1]$. From the observation that $g(0) = b^2 > 0$ and $g(1) = 1 > 0$ we conclude that $g(x)$ is nonnegative on the interval $[0,1]$ for $b \geq 1$ and thus $\frac{d}{dx} f(b,x) \geq 0$, when $b \geq 1$. □

Proof of (B.VII). We have that

$$f'(x) = \frac{-\varphi(x) + \Phi(-x)\, x}{\varphi(x)},$$

thus $f'(x) \leq 0$ iff $-\varphi(x) + \Phi(-x)\, x \leq 0$, which is equivalent to Result (B.I). □

Proof of (B.VIII). See that

$$f'(x) = \frac{\Phi(-x) - x\varphi(x) + x^2\Phi(-x)}{\varphi(x)},$$

thus $f'(x) \geq 0$ iff $\frac{\Phi(-x)}{\varphi(x)} \geq \frac{x}{1+x^2}$, which is an implication of the lower bound from result (B.III). □

Proof of (B.IX). It is easy to see that $\lim_{t\to 0^+} f(t) = 0$. To see that $\lim_{t\to 1^-} f(t) = \frac{2\,\Phi(-b)}{b\,\varphi(b)}$ we expand $\log\big(\Phi(-b/\sqrt{t})\big)$ in a series around $t_0 = 1$ and obtain

$$\log\big(\Phi(-b/\sqrt{t})\big) = \log\big(\Phi(-b)\big) + \frac{b\,\varphi(b)}{2\Phi(-b)}(t-1) + o(t-1).$$

Thus

$$\lim_{t\to 1^-} f(t) = \lim_{t\to 1^-} \frac{1-t}{\frac{b\,\varphi(b)}{2\Phi(-b)}(1-t) + o(t-1)} = \frac{2\,\Phi(-b)}{b\,\varphi(b)}.$$

To prove that f is increasing we study the first derivative. For $t \in (0,1)$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} f(t) &= \frac{-\log\left(\frac{\Phi(-b)}{\Phi(-b/\sqrt{t})}\right) + \frac{(1-t)\Phi(-b/\sqrt{t})\Phi(-b)}{\Phi(-b)} \cdot \frac{bt^{-3/2}}{2\Phi(-b/\sqrt{t})^2}\varphi(-b/\sqrt{t})}{\Big(\log\big(\Phi(-b)\big) - \log\big(\Phi(-b/\sqrt{t})\big)\Big)^2} \\ &= \frac{\log\left(\frac{\Phi(-b/\sqrt{t})}{\Phi(-b)}\right) + \frac{b(1-t)}{2t^{3/2}} \cdot \frac{\varphi(-b/\sqrt{t})}{\Phi(-b/\sqrt{t})}}{\Big(\log\big(\Phi(-b)\big) - \log\big(\Phi(-b/\sqrt{t})\big)\Big)^2} \qquad (2.32)\end{aligned}$$

Due to Result (B.VII) we have the lower bound

$$\frac{\varphi(-b/\sqrt{t})}{\Phi(-b/\sqrt{t})} \geq \frac{\varphi(-b)}{\Phi(-b)}$$

and thus the numerator of the fraction in (2.32) can be bounded from below by the function $g : (0,1) \to \mathbb{R}$ defined as below:

$$g(t) := \log\left(\frac{\Phi(-b/\sqrt{t})}{\Phi(-b)}\right) + \frac{b(1-t)}{2\,t^{3/2}}\,\frac{\varphi(b)}{\Phi(-b)}$$

Notice that $g(t) \geq 0$ implies $\frac{d}{dt}f(t) \geq 0$ which is exactly what we want to

establish. For the remainder of the proof we show that $g(t)$ is non-negative. Since $\lim_{t\to 0^+} g(t) = +\infty$ and $g(1) = 0$, it suffices to show that $g'(t)$ is monotone (non-increasing). We study the first derivative

$$\begin{aligned} g'(t) &= \frac{b}{2t^{3/2}} \frac{\varphi(b/\sqrt{t})}{\Phi(-b/\sqrt{t})} + \frac{b}{4t^{3/2}} \frac{\varphi(b)}{\Phi(-b)} - \frac{3b}{4t^{5/2}} \frac{\varphi(b)}{\Phi(-b)} \\ &= \frac{b^2}{4t^2} \left(2 \frac{\frac{\sqrt{t}}{b}\varphi(b/\sqrt{t})}{\Phi(-b/\sqrt{t})} + \left(t^{1/2} - 3t^{-1/2} \right) \frac{\frac{1}{b}\varphi(b)}{\Phi(-b)} \right) \\ &\leq \frac{b^2}{4t^2} \left(2 \frac{\frac{\sqrt{t}}{b}\varphi(b/\sqrt{t})}{\Phi(-b/\sqrt{t})} - 2 \frac{\frac{1}{b}\varphi(b)}{\Phi(-b)} \right) \leq 0, \end{aligned}$$

where the last inequality is a consequence of the application of Result (B.VIII), that is

$$t \longmapsto \frac{\frac{\sqrt{t}}{b}\varphi(b/\sqrt{t})}{\Phi(-b/\sqrt{t})}$$

is an increasing function of t. □

Chapter 3

Zooming-in on a Lévy process: Failure to observe threshold exceedance over a dense grid

In this chapter we consider the detection error arising from time discretization when estimating the threshold crossing probability $\mathbb{P}(\sup_{t\in[0,1]} X_t > b)$, where X is a general Lévy process. Recall that in Chapter 2 we considered this problem with X a standard Brownian Motion. We establish exact asymptotic behavior of this error as the number of grid points on an equidistant (regular) grid tends to infinity. We assume that X has a zooming-in limit, which necessarily is $1/\alpha$-self-similar Lévy process with $\alpha \in (0, 2]$, and restrict to $\alpha > 1$. Moreover, the moments of the difference of the supremum and the maximum over the grid points are analyzed and their asymptotic behavior is derived.

1 Introduction

Consider a Lévy process $X = (X_t, t \geq 0)$ on the real line and let

$$M = \sup\{X_t : t \in [0,1]\}, \qquad \tau = \inf\{t \geq 0 : X_t \vee X_{t-} = M\}$$

be the supremum and its time, respectively, for the time interval $[0,1]$. For any $n \in \mathbb{N}_+$ consider also the maximum of X over the regular grid with step size $1/n$:

$$M^{(n)} = \max\{X_{i/n} : i = 0, \ldots, n\}.$$

In this chapter we derive exact asymtotic behavior of

$$\Delta_n(x) = \mathbb{P}(M > x, M^{(n)} \leq x) = \mathbb{P}(M > x) - \mathbb{P}(M^{(n)} > x) \tag{3.1}$$

as $n \to \infty$ for any fixed $x > 0$, which is the probability of failure in detecting threshold exceedance when restricting to the grid time-points. On the way towards this goal, we also provide asymptotics of the moments $\mathbb{E}(M - M^{(n)})^p$ of the discretization error in approximation of the supremum, markedly improving on the bounds in [54] and other works.

The motivation comes from various applications, where it is vital to understand if the process X has exceeded a fixed threshold $x > 0$ or not. Application areas include, among others, insurance and mathematical finance (pricing of barrier options), energy science (electric load), environmental science (pollution levels and exposure) and computer reliability. Normally, we observe the process of interest over a dense regular grid without having full knowledge about the continuous-time trajectory. It is then natural to base our judgment on $M^{(n)}$ instead of M. Thus $\Delta_n(x)$ is the probability of making an error: the process exceeds x but not over the grid points. Furthermore, our limit result can be used to provide a correction to $\mathbb{P}(M > x)$ when approximating it by $\mathbb{P}(M^{(n)} > x)$ derived from Monte Carlo simulation. Or vice versa, it can provide a correction to $\mathbb{P}(M_n > x)$ in cases when the formula for $\mathbb{P}(M > x)$ is available, see e.g. [23]. It is noted that $M^{(n)}$ always underestimates M, and so one can also consider more accurate estimators (but also more complicated, since these are potentially based on all the available information), which we leave for future work.

Our main vehicle is the zooming-in limit theory of [63], where it is shown

under a weak regularity assumption, see (3.8) below, that

$$\Big(V^{(n)} \mid \tau \in (0,1)\Big) \xrightarrow{\mathrm{d}} \widehat{V}, \qquad \text{with} \qquad V^{(n)} := b_n(M - M^{(n)}) \tag{3.2}$$

for some specific sequence $b_n > 0$ and a random variable $\widehat{V}$. More precisely, $\widehat{V}$ is defined in terms of the law of a self-similar Lévy process $\widehat{X}$ (the limit under zooming-in), see (3.9), and b_n is chosen such that $b_n X_{1/n} \xrightarrow{\mathrm{d}} \widehat{X}_1$.

The convergence in (3.2) readily suggests that $\mathbb{E}(M - M^{(n)})^p$ is of order b_n^{-p} and supplements it with exact asymptotics, but only when the underlying sequence of random variables $(V^{(n)})^p$ is uniformly integrable. Establishing the latter, however, is far from trivial and we could only do that thanks to [12] providing a representation of the pre- and post-supremum process using juxtaposition of the excursions in half-lines. Interestingly, the scaled moments would explode in some cases if we considered grid points to the right (or to the left) of τ only. Furthermore, certain conditions must be fulfilled, and the decay of the moments can never be faster than $1/n$ if X has jumps of both signs.

The intuition behind the asymptotics of the detection error probability $\Delta_n(x)$ is given by the following:

$$\begin{aligned} b_n\Delta_n(x) &= b_n\mathbb{P}(x < M \le x + b_n^{-1}V^{(n)}) \\ &\approx \int_0^\infty b_n\mathbb{P}(x < M \le x + b_n^{-1}y)\mathbb{P}(V^{(n)} \in \mathrm{d}y) \\ &\to f_M(x)\mathbb{E}\widehat{V}, \end{aligned}$$

where f_M is a density of M; we also show that $\mathbb{E}\widehat{V}$ has a simple explicit formula in terms of the basic parameters. The second line is suggested by the asymptotic independence of $V^{(n)}$ and M (the convergence in (3.2) is Rényi-mixing). Note also that uniform integrability of $V^{(n)}$ is needed in the last step, which forces us to assume that X has unbounded variation on compacts; we assume that $\alpha \in (1,2]$, see (3.7). Making $\approx$ precise turns out to be a major undertaking. In fact, asymptotic independence between M and $V^{(n)}$ is not enough – one can construct counterexamples resembling those in [10, Theorem 2.4(ii)]. In addition to exact asymptotics, we also provide bounds on both the moments and the error probability $\Delta_n(x)$ in the cases when the zooming-in assumption (3.8) is not satisfied.

Chapter structure: In Section 2 we set up the notation and provide a thorough discussion of the basic zooming-in assumption (3.8) including some important classes of examples. In Section 3 we show uniform integrability of $V^{(n)}$ for $\alpha > 1$ and provide moment asymptotics of $\mathbb{E}(M - M^{(n)})^p$ for $\alpha > p$ under the zooming-in assumption (3.8); we show a logarithmically tight upper bound otherwise. Section 4 establishes exact asymptotics of the error probability $\Delta_n(x)$. The proofs of many preparatory and auxiliary results are deferred to the Appendix.

Literature overview: The fundamental work in this area is [7], where the limit theorem for $M - M^{(n)}$ was established in the case of a linear Brownian motion X. This sparkled research in various application areas including mathematical finance, see [22] for approximations of option prices in discrete-time models using continuous-time counterparts. Various expansions and bounds on the expected error $\mathbb{E}(M - M^{(n)})$ were derived in [30, 54, 64] among others. The error probability $\Delta_n(x)$ asymptotics was identified in [23] in the case of a linear Brownian motion and later extended in [38] to Brownian motion perturbed by an independent compound Poisson process. An interested reader may consult [39] for an overview of the literature regarding discretization of Brownian motion, see also [16] for non-uniform grids.

There is also a large body of literature concerned with the supremum of a Lévy process, see [27, 29, 71] among many others, and with the small-time behavior of Lévy processes, see [13, 37, 40, 48] and references therein.

2 Preliminaries and examples

To set up the notation, recall the Lévy-Khintchine formula

$$\mathbb{E}e^{\theta X_t} = e^{\psi(\theta)t}, \quad \psi(\theta) = \gamma\theta + \frac{\sigma^2}{2}\theta^2 + \int_{\mathbb{R}} \left(e^{\theta x} - 1 - \theta x 1_{\{|x|<1\}}\right) \Pi(\mathrm{d}x)$$

with $\theta \in \mathrm{i}\mathbb{R}$ and parameters $\gamma \in \mathbb{R}, \sigma \geq 0, \Pi(\mathrm{d}x)$ where the latter is a Lévy measure satisfying $\int_{\mathbb{R}}(x^2 \wedge 1)\Pi(\mathrm{d}x) < \infty$. In the case of $\int_{-1}^{1} |x|\Pi(\mathrm{d}x) < \infty$ we have a simplified expression

$$\psi(\theta) = \gamma'\theta + \frac{\sigma^2}{2}\theta^2 + \int_{\mathbb{R}} \left(e^{\theta x} - 1\right) \Pi(\mathrm{d}x) \tag{3.3}$$

with $\gamma' \in \mathbb{R}$ being called the linear drift. We write ub.v. and b.v. for processes of unbounded and bounded variation on compacts, respectively. Recall that b.v. case corresponds to $\sigma = 0$ and $\int_{-1}^{1} |x|\Pi(\mathrm{d}x) < \infty$, so that (3.3) can be used.

Let us introduce some notation for the tails of Π:

$$\overline{\Pi}_+(x) = \Pi(x, \infty), \quad \overline{\Pi}_-(x) = \Pi(-\infty, -x), \quad \overline{\Pi}(x) = \overline{\Pi}_+(x) + \overline{\Pi}_-(x) \tag{3.4}$$

with $x > 0$. We also define the truncated mean and variance functions for $x \in (0, 1)$:

$$m(x) = \gamma - \int_{x<|y|<1} y\Pi(\mathrm{d}x), \quad v(x) = \sigma^2 + \int_{|y|<x} y^2\Pi(\mathrm{d}x), \tag{3.5}$$

which play a fundamental role in the study of small time behavior of X. Finally, we write $f \in \mathrm{RV}_\alpha$ to say that f is a function regularly varying at 0 with index α, see [15].

2.1 Important indices

Define the following indices:

$$\begin{aligned} \beta_0 &:= \inf\big\{\beta \geq 0 : \int_{|x|<1} |x|^\beta \Pi(\mathrm{d}x) < \infty\big\}, \\ \beta_\infty &:= \sup\big\{\beta \geq 0 : \int_{|x|>1} |x|^\beta \Pi(\mathrm{d}x) < \infty\big\}. \end{aligned} \tag{3.6}$$

The index $\beta_0 \in [0, 2]$ provides some basic information about the intensity of small jumps and is often called Blumenthal-Getoor index, whereas $\beta_\infty \in [0, \infty]$ is about integrability of big jumps. Moreover, let

$$\alpha = \begin{cases} 2, & \sigma \neq 0 \\ 1, & \text{b.v. with } \gamma' \neq 0 \\ \beta_0, & \text{otherwise} \end{cases} \tag{3.7}$$

and note that necessarily $\alpha \geq \beta_0$.

2.2 Attraction to self-similar processes under zooming in

Throughout most part of this work we assume that

$$X_\varepsilon/a_\varepsilon \xrightarrow{\mathrm{d}} \widehat{X}_1, \qquad \text{as } \varepsilon \downarrow 0 \tag{3.8}$$

for some function $a_\varepsilon > 0$ and a random variable $\widehat{X}_1$, not identically 0. Then necessarily [67, Thm. 15.12(ii)] $\widehat{X}_1$ is infinitely divisible, and the above weak convergence extends to convergence of the respective processes (in Skorokhod J_1 topology):

$$(X_{\varepsilon t}/a_\varepsilon)_{t\geq 0} \xrightarrow{\mathrm{d}} (\widehat{X}_t)_{t\geq 0}.$$

Furthermore, the Lévy process $\widehat{X}$ must be self-similar with Hurst parameter $1/\alpha$ for some $\alpha \in (0, 2]$, implying that it is either

(i) a (driftless) Brownian motion, and then $\alpha = 2$,

(ii) a linear drift, and then $\alpha = 1$,

(iii) a strictly α-stable Lévy process, and then $\alpha \in (0, 2)$.

Note that $\alpha = 1$ corresponds to two different classes: drift processes and strictly 1-stable process also known as a Cauchy process (symmetric and drifted), and this suggests that $\alpha = 1$ is often an intricate case. Moreover, $a_\varepsilon \in \mathrm{RV}_{1/\alpha}$, that is, the scaling function is regularly varying at 0 with index $1/\alpha$. The respective domains of attraction are completely characterized in terms of Lévy triplets in [63], which also provides a comprehensive overview of the related literature (the domains of attraction to the Brownian motion and linear drift have been characterized in [42] before). We emphasize that the parameter α is necessarily of the form (3.7), see [63, Cor. 1]. Moreover, the limit $\widehat{X}$ is unique up to scaling by a positive constant and a_ε is unique up to asymptotic proportionality: $a_\varepsilon \sim ca'_\varepsilon$.

In the following we will make extensive use of the positivity parameter of the attractor:

$$\rho = \mathbb{P}(\widehat{X}_1 > 0),$$

which can be easily derived from the Lévy triplet of $\widehat{X}$ using formulas in [100]. It is well known that the pair (α, ρ) specifies the self-similar process $\widehat{X}$ up to scaling by a positive constant. This is sufficient for our purpose since we may

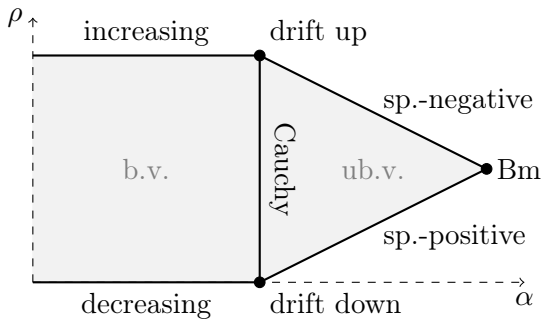


Figure 3.1: Self-similar Lévy processes with parameters (α, ρ)

always choose an appropriate scaling sequence a_ε. For $\alpha \in (1, 2]$ the range of ρ is given by $\rho \in [1 - 1/\alpha, 1/\alpha]$ with the left and right boundary values corresponding to the spectrally-positive and spectrally-negative processes, respectively, see Figure 3.1. For $\alpha \in (0, 1]$ we have $\rho \in [0, 1]$ with boundary values corresponding to a decreasing and increasing processes, respectively. Finally, $\alpha = 1, \rho \in (0, 1)$ specifies the class of drifted Cauchy processes, whereas $\alpha = 1, \rho = \pm 1$ corresponds to linear drifts with signs $\pm$. We often write

$$X \in \mathcal{D}_{\alpha,\rho}$$

to say that (3.8) holds with $1/\alpha$-self-similar process $\widehat{X}$ having positivity parameter ρ. Note that $X \in \mathcal{D}_{\alpha,\rho}$ implies that $\mathbb{P}(X_\varepsilon > 0) \to \rho$ as $\varepsilon \downarrow 0$, which follows from the fact that $\mathbb{P}(\widehat{X}_1 = 0) = 0$.

2.3 Examples

The trivial examples satisfying (3.8) are (i) $\sigma > 0$ and arbitrary $\gamma, \Pi(\mathrm{d}x)$ and (ii) b.v. process with $\gamma' \neq 0$ and otherwise arbitrary $\Pi(\mathrm{d}x)$. In case (i) $\widehat{X}$ is a Brownian motion and a_ε is asymptotically proportional to $\varepsilon^{1/2}$, and in case (ii) $\widehat{X}$ is a linear drift (having the same sign as γ') and a_ε is asymptotically proportional to ε.

Let us stress that (3.8) is a weak regularity assumption satisfied for almost every Lévy process of practical interest. The most notable exceptions are driftless compound Poisson process and its neighbors: driftless gamma and variance gamma processes, where in all cases $\sigma = \gamma' = 0, \overline{\Pi}(x) \in \mathrm{RV}_0$ as $x \downarrow 0$. In other

words, small jump activity is too weak to have a non-trivial limit. In the following we briefly consider some important classes of Lévy processes and establish their zooming-in limits (such examples are missing in [63]).

2.3.1 Tempered stable processes (CGMY)

A rather general class of Lévy processes is obtained by considering the Lévy measure $\Pi(\mathrm{d}x)$ of the form

$$\Pi(\mathrm{d}x) = c_+ x^{-1-\alpha_+} e^{-\lambda_+ x} 1_{\{x>0\}} \mathrm{d}x + c_- |x|^{-1-\alpha_-} e^{\lambda_- x} 1_{\{x<0\}} \mathrm{d}x,$$

where $c_\pm, \lambda_\pm \geq 0, \alpha_\pm < 2$ and $\alpha_\pm > 0$ if $\lambda_\pm = 0$. Particular examples are stable, gamma, inverse Gaussian and variance gamma processes. When $c_\pm = 0$ we put $\alpha_\pm = 0$. We assume that there is no Gaussian part ($\sigma = 0$) since otherwise $\widehat{X}$ is a Brownian motion, and that in b.v. case ($\alpha_+, \alpha_- < 1$) there is no linear drift ($\gamma' = 0$) since otherwise $\widehat{X}$ is a linear drift process.

We have the following cases according to [63, Thm. 2] (for clarity we avoid specifying ρ in some cases):

- if $\alpha_\mp \leq \alpha_\pm \in (0,1) \cup (1,2)$ then $\widehat{X}$ is a strictly $\alpha_\pm$-stable process; in the case of strict inequality $\widehat{X}$ has one-sided jumps of sign $\pm$.
- if $\alpha_\mp < \alpha_\pm = 1$ or $\alpha_+ = \alpha_- = 1$ with $c_\mp < c_\pm$ then $\widehat{X}$ is a linear drift process of sign $\mp$ (the sign might look counterintuitive at first, see Remark 3.1).
- if $\alpha_+ = \alpha_- = 1$ and $c_+ = c_-$ then $\widehat{X}$ is a Cauchy process.
- if $\alpha_+, \alpha_- \leq 0$ then such a process does not have a non-trivial limit under zooming-in; the intensity of jumps is too small.

In particular, the gamma process has no limit, and the same is true for variance gamma processes with 0 mean (in these cases we have $\alpha_\pm = 0$).

2.3.2 Generalized hyperbolic processes

Another important family of Lévy processes is a 5-parameter class of generalized hyperbolic Lévy motions introduced by Barndorff-Nielsen [11] and advocated for financial use in [46]. Note that this class includes normal inverse Gaussian

processes. Generalized hyperbolic processes have no Gaussian component, and their Lévy measure possesses a density behaving as $C/x^2+\mathcal{O}(1/|x|)$ at 0 for some constant $C > 0$, see [82, Prop. 2.18]. Thus such processes are ub.v. and satisfy $\overline{\Pi}_\pm \in \mathrm{RV}_{-1}$ with $\overline{\Pi}_+(x)/\overline{\Pi}_-(x) \to 1$ as $x \downarrow 0$. We also see that $x\overline{\Pi}_+(x) \to C$ and with a little more work we find that $\int_{x\le|y|<1} y\Pi(\mathrm{d}y)$ has a finite limit. Hence according to [63, Thm. 2] every generalized hyperbolic motion is attracted to a Cauchy process.

2.3.3 Subordination

Another popular way to construct a Lévy process is by considering $X_t = Y_{S_t}$, where Y and S are two independent Lévy processes and the latter is non-decreasing (a subordinator). In this case, it is sufficient to check (3.8) for the underlying processes:

Lemma 3.1. *Suppose that* $Y_\varepsilon/y_\varepsilon \xrightarrow{\mathrm{d}} \widehat{Y}_1, S_\varepsilon/s_\varepsilon \xrightarrow{\mathrm{d}} \widehat{S}_1$ *as* $\varepsilon \downarrow 0$ *with* $y_t, s_t > 0$ *and non-trivial limits. Then* (3.8) *is satisfied with* $\widehat{X}_1 = \widehat{Y}_{\widehat{S}_1}$ *and* $a_\varepsilon = y_{s_\varepsilon}$.

Proof. With obvious notation we have for any $\theta \in \mathrm{i}\mathbb{R}$ that $\psi_Y(\theta/y_\varepsilon)\varepsilon \to \psi_{\widehat{Y}}(\theta)$ and a similar statement with respect to S. Now compute

$$\psi(\theta/a_\varepsilon)\varepsilon = \psi_S(\psi_Y(\theta/y_{s_\varepsilon}))\varepsilon = \psi_S(\psi_{\widehat{Y}}(\theta)/s'_\varepsilon)\varepsilon \to \psi_{\widehat{S}}(\psi_{\widehat{Y}}(\theta)),$$

where $s'_\varepsilon \sim s_\varepsilon$ and such a change is irrelevant for the limit. □

Letting $\alpha_Y, \alpha_S \in (0, 1]$ be the corresponding indices, see (3.7), we find that $\alpha = \alpha_Y\alpha_S$; one can see this by recalling that $a \in \mathrm{RV}_{1/\alpha}$. Note also that $\rho = \rho_{\widehat{Y}}$. Thus subordination allows for a direct description of the limiting process without the need to identify the Lévy triplet of X. For example, the normal inverse Gaussian must belong to $\mathcal{D}_{1,1/2}$. That is, the limit is a Brownian motion subordinated by the 1/2-stable subordinator and this yields a Cauchy process as already observed in §2.3.2.

2.3.4 A sufficient condition

The following result provides an easy-to-check condition which implies (3.8). It does not allow, however, to check attraction to Cauchy processes. Note that one may also check the following condition for $-X$ instead of X.

Proposition 3.1. *Assume that $\sigma = 0$, $\gamma' = 0$ in b.v. case, and $\overline{\Pi}_+ \in \mathrm{RV}_{-\alpha}$ with $\alpha \in (0, 2]$. Then (3.8) holds true in the following cases:*

- *$\alpha = 2$ and $\liminf_{x \downarrow 0} \overline{\Pi}_+(x)/\overline{\Pi}_-(x) > 0$,*
- *$\alpha = 1$ and $\liminf_{x \downarrow 0} \overline{\Pi}_+(x)/\overline{\Pi}_-(x) > 1$ (positive/negative linear drift limit according to b.v./ub.v.),*
- *$\alpha \in (0, 1) \cup (1, 2)$ and $\lim_{x \downarrow 0} \overline{\Pi}_+(x)/\overline{\Pi}_-(x) \in (0, \infty]$.*

Proof. See Appendix A for the proof. □

Remark 3.1. In the case $\alpha = 1$ in Proposition 3.1 we assume that the positive jumps are dominant and show that $\widehat{X}$ is then a linear drift. One expects that this drift is positive, which is indeed true when X is b.v. If, however, X is ub.v. then the limiting drift process has a negative slope, which on intuitive level can be explained by the standard construction of X as the limit of compensated compound Poisson processes. It turns out that the compensating drift 'wins' – it determines the sign.

Let us stress that the case $\alpha = 1$ with $\overline{\Pi}_+(x) \sim \overline{\Pi}_-(x) \in \mathrm{RV}_{-1}$ does not guarantee (3.8). In this case, according to [63, Thm. 2], we need to check that $m(x)/(x\overline{\Pi}(x))$ has a limit in $[-\infty, \infty]$ with $m(x)$ defined in (3.5); finite limits correspond to $\widehat{X}$ being Cauchy. In the following we provide an example, where the latter does not hold true, and thus (3.8) is not satisfied.

We consider an ub.v. Lévy process with

$$\sigma = \gamma = 0, \quad \text{and} \quad \overline{\Pi}_+(x) = (1 + u(x))/x, \qquad \overline{\Pi}_-(x) = 1/x$$

for small enough x, where $u(x) = \sin(\log(-\log x))/\log x \to 0$. One can verify that $(1 + u(x))/x$ is decreasing for small enough x and so we have legal $\overline{\Pi}_\pm(x)$ functions, which are asymptotically equivalent and RV_{-1}. Now compute

$$m(x) = c + \int_x^h y \mathrm{d}\frac{u(y)}{y} = c' - u(x) - \int_x^h y^{-1} u(y) \mathrm{d}y,$$

where $h > 0$ is some small number and c, c' are constants. The latter integral evaluates to $\cos(\log(-\log x)) - c''$ which is an oscillating function as $x \downarrow 0$ and the same is true about $m(x)$. Finally, $x\overline{\Pi}(x) \to 2$ and we see that $m(x)/(x\overline{\Pi}(x))$ oscillates as well, which shows that (3.8) does not hold.

3 Moments of the discretization error

Consider the error $M - M^{(n)}$ in approximation of the supremum of a Lévy process over the interval $[0, 1]$ by the maximum over the uniform grid with step size $1/n$. Our first result provides an upper bound on the moments of this discretization error for a general process X, not necessarily satisfying the zooming-in assumption (3.8).

Theorem 3.1. *For any $p > 0$ satisfying $\int_{|x|>1} |x|^p \Pi(\mathrm{d}x) < \infty$ and any $\epsilon > 0$ we have*

$$\mathbb{E}(M - M^{(n)})^p = \begin{cases} \mathrm{O}(n^{-p/\alpha+\varepsilon}), & p \le \alpha, \\ \mathrm{O}(n^{-1}), & p > \alpha \end{cases}$$

as $n \to \infty$.

Moreover, the bound can be strengthened to $\mathrm{O}(n^{-p/\alpha})$ in the boundary cases: (i) $p \le \alpha = 2$ and (ii) $p \le \alpha = 1$ and X is b.v.

For $p = 1$ the result in Theorem 3.1 is close to [30, Theorem 5.2.1]; in the case of b.v. process X our bound $\mathrm{O}(n^{-1})$ is slightly sharper. Importantly, the result [30] cannot be generalized in a straightforward fashion to $p \neq 1$, since it crucially relies on Spitzer identity. Nevertheless, [54] provides some bounds for $p \neq 1$ in the particular case when $\sigma = 0, \gamma' = 0$ and $\Pi(\mathrm{d}x)$ has a density sandwitched between $c_1|x|^{-1-\alpha}$ and $c_2|x|^{-1-\alpha}$ for small $|x|$. These bounds, however, have suboptimal rates (in the logarithmic sense) unless $p > 2\alpha$ or X is spectrally negative.

Our main goal, however, is to provide exact moment asymptotics, which is possible under the regularity assumption (3.8). In the following, $V^{(n)} := b_n(M - M^{(n)})$, as defined in (3.2) and $b_n = 1/a_{1/n}$.

Theorem 3.2. *Assume that $X \in \mathcal{D}_{\alpha,\cdot}$. Then for any $p \in (0, \alpha)$ satisfying $\int_{|x|>1} |x|^p \Pi(\mathrm{d}x) < \infty$ the sequence $\mathbb{E}\left(V^{(n)}\right)^p$ is bounded.*

For completeness let us recall from [63] that

$$\widehat{V} = \min\{-\widehat{\xi}_{U+\mathbb{Z}}\}, \tag{3.9}$$

where $(\widehat{\xi}_t, t \in \mathbb{R})$ is the limit of $\widehat{X}$ over $[0, T]$ as seen from its supremum point as $T \to \infty$, and U is an independent uniform random variable. The weak convergence in (3.2) and Theorem 3.2 immediately yield the uniform integrability

of certain powers of $V^{(n)}$. Combining this result with the limiting expression for $\mathbb{E}\widehat{V}^{(n)}$ in [9, Prop. 2], we obtain the following result (recall the definition of β_∞ in (3.6)).

Corollary 3.1. *Let $X \in \mathcal{D}_{\alpha,\rho}$. Then for any positive $p < \alpha \wedge \beta_\infty$ we have*

$$\mathbb{E}\left(V^{(n)}\right)^p \to \mathbb{E}\widehat{V}^p\,\mathbb{P}(\tau \notin \{0,1\}) \in (0,\infty) \qquad \text{as } n \to \infty.$$

In particular, for $\alpha > 1$ and $\beta_\infty > 1$:

$$\mathbb{E}V^{(n)} \to \mathbb{E}\widehat{V} = -\zeta\left(\frac{\alpha-1}{\alpha}\right)\mathbb{E}\widehat{X}_1^+ \qquad \text{as } n \to \infty,$$

where ζ is the Riemann zeta function.

It is noted that $\mathbb{E}\widehat{X}_1^+$ has an explicit expression, see [100, Thm. 3] or [9].

Proof. Note that if $\tau \in \{0,1\}$ with positive probability then $(V^{(n)} \mid \tau \in \{0,1\}) = 0$ a.s., because of the nature of discretization. Now the first result follows from the weak convergence in (3.2) and uniform integrability of $(V^{(n)})^p$, where the latter is a consequence of Theorem 3.2 applied with a slightly larger p.

Next, we note that $\alpha \in (1,2]$ implies that X is ub.v. process and, in particular, $\mathbb{P}(\tau \in (0,1)) = 1$. Moreover, $\mathbb{E}\widehat{V}^{(n)} \to \mathbb{E}\widehat{V}$, where $\widehat{V}^{(n)}$ corresponds to the discretization of $\widehat{X}$ which is in its own domain of attraction. The limit of $\mathbb{E}\widehat{V}^{(n)}$ was obtained in [9, Prop. 2] using self-similarity and Spitzer's identity, see also [7]. □

3.1 Comments and extensions

Note that Theorem 3.1 is weaker than Corollary 3.1 (when the conditions of the latter are satisfied) providing the exact asymptotics of $\mathbb{E}(M - M^{(n)})^p$. In particular, we see that $\mathbb{E}(M - M^{(n)})^p$ is a sequence regularly varying at ∞ with index $-p/\alpha$, which is clearly upper-bounded by $n^{-p/\alpha+\epsilon}$ for large n.

Remark 3.2. Assuming that X has jumps of both signs, it is not hard to see that we may choose $c, h > 0$ such that $\mathbb{P}(M - M^{(n)} > h) \geq cn^{-1}$, see also . Hence we must have $\mathbb{E}(M - M^{(n)})^p \geq c'n^{-1}$ for any $p > 0$. Note that this complements the case $p > \alpha$ in Theorem 3.1 with a lower bound of the same

order. Furthermore, when $X \in \mathcal{D}_{\alpha,\rho}$, we get that $\mathbb{E}\left(V^{(n)}\right)^p \geq b_n^p c' n^{-1} \to \infty$ when $p > \alpha$, because b_n^p is regularly varying at ∞ with index $p/\alpha > 1$.
This question is more delicate for one-sided processes, and we have no complete answer here. In the case of no jumps (X is a Brownian motion with drift) boundedness of exponential moments was established in [7]. Furthermore, if X is a b.v. spectrally-negative (-positive) process, then the error $M - M^{(n)}$ is bounded from above by $|\gamma'|n^{-1}$, showing that $\mathbb{E}(M - M^{(n)})^p = \mathrm{O}(n^{-p})$, see also [54].

Remark 3.3. Letting $V_s^{(n)}$ be the analogue of $V^{(n)}$ but for a shifted grid $(i+s)/n$ with $i \in \mathbb{Z}$ and all points in $[0,1]$, we note that also $\{\mathbb{E}\big(V_s^{(n)}\big)^p : n \geq 1, s \in [0,1)\}$ is bounded under the assumptions of Thm. 3.2; the proof does not need any modifications. This readily implies (by letting $s \uparrow 1, s \downarrow 0$) that we may also exclude the endpoints from the standard grid without affecting the result of Thm. 3.2.

3.1.1 Dealing with big jumps

If $\int_1^\infty x\Pi(\mathrm{d}x) = \int_{-\infty}^{-1} |x|\Pi(\mathrm{d}x) = \infty$ then necessarily $\mathbb{E}V^{(n)} = \infty$ for all $n \geq 1$, and the analogous statement is true for all $p > 0$. When only the positive jumps, say, are non-integrable we may still arrive to an unbounded sequence of $\mathbb{E}V^{(n)}$. The problem is that the discretization error obtained by looking to the right of the supremum time exclusively may not be bounded in expectation, even when jumps exceeding 1 in absolute value are discarded (this can be shown using the lower bound on the entrance law in [19, Prop. 3]).

Importantly, we may remove the condition $\beta_\infty > 1$ on the absence of big jumps if we restrict to the event that no two big jumps are close to each other or to the endpoints of the interval, say. For this, let $T_1 < T_2 < \ldots$ be the times of jumps exceeding 1 in absolute value and let N be their number in the interval $(0,1)$; we also put $T_0 = 0$ and $T_{N+1} = 1$. Finally, define the event to be excluded:

$$A^{(n)} = \{\exists i \in \{0, \ldots, N\} : T_{i+1} - T_i < 1/n\}. \tag{3.10}$$

It is well-known that $\mathbb{P}(A^{(n)}) = \mathrm{O}(1/n)$ as $n \to \infty$, which is also easy to see using Slivnyak's formula from Palm theory; this observation will be used in the proof of Prop. 3.3.

Proposition 3.2. *Let $X \in \mathcal{D}_{\alpha,\cdot}$ with $\alpha > 1$. Then the family $\widetilde{V}^{(n)} = V^{(n)} 1_{\{A^{(n)c}\}}$ is uniformly integrable, and $\mathbb{E}\widetilde{V}^{(n)} \to \mathbb{E}\widehat{V}$.*

Proof. The proof is given in Appendix B. □

3.1.2 Conjecture for processes of bounded variation

Consider a b.v. process X with $\int_{|x|>1} |x|\Pi(\mathrm{d}x) < \infty$. Recall that we have an upper bound on $\mathbb{E}(M - M^{(n)})$ of order n^{-1}, see Theorem 3.1, and a lower bound of the same order when X has jumps of both signs, see Remark 3.2. It is thus natural to ask if $n\mathbb{E}(M - M^{(n)})$ has a finite positive limit as $n \to \infty$. We conjecture the following:

$$n\mathbb{E}(M - M^{(n)}) \to \tfrac{1}{2}|\gamma'| \cdot \mathbb{P}(\tau \in (0,1)) + \tfrac{1}{2} I, \tag{3.11}$$

where

$$I =: \int_0^1 \iint_{x,y,u,v\geq 0} ((x-u) \wedge (y-v))_+ \\ \Pi(\mathrm{d}x)\Pi(-\mathrm{d}y)\mathbb{P}(-\underline{X}_t \in \mathrm{d}u)\mathbb{P}(\overline{X}_{1-t} \in \mathrm{d}v)\mathrm{d}t.$$

Here the first term is suggested by the small-time behavior, and the second comes from the possibility of having a jump up (of size x) before τ and a jump down (of size $-y$) after τ, and no observations in-between. This result should not rely on the zooming-in assumption (3.8), and may require quite a different set of tools for its proof. Proposition 3.4 in Appendix B demonstrates (3.11) in the simple case of a compound Poisson process with drift. Note also that $I = 0$ in the case of one-sided jumps, and otherwise we can not expect $n(M - M^{(n)})$ to be uniformly integrable.

3.2 Proofs

The crucial step in the proof of the above results is given by the following lemma, where we write $\overline{X}_1$ and $\underline{X}_1$ for the supremum and infimum of X up to time 1, respectively. We write $\overline{X}$ and τ for the supremum and its time over the interval $[0,T]$ for some fixed $T > 0$.

Lemma 3.2. *Consider X on the interval $[0,T]$ for any fixed $T \geq 1$ and let*

$$Z_T = \sup_{t\in[0,1]} \{(\overline{X} - X_{\tau-t}) \wedge (\overline{X} - X_{\tau+1-t})\}$$

with convention that $X_s = -\infty$ if $s \notin [0,T]$. Then Z_T is first-order stochastically dominated by Z_1 and hence by $\overline{X}_1 - \underline{X}_1$.

Proof. According to Bertoin's [12] representation of the joint law of post- and pre-supremum processes on the interval $[0,T]$ we have $Z_T \stackrel{\mathrm{d}}{=} Z'_T$, where

$$Z'_T := \sup_{t\in[0,1]} \{X^{\Uparrow}_t \wedge -(X^{\Downarrow}_{1-t})\}, \tag{3.12}$$

where $X^{\Uparrow}_t = Y^+_{a^+_t}$ and $X^{\Downarrow}_t = Y^-_{a^-_t}$ for some processes $Y^{\pm}$ (which do not depend on the choice of T) and a^+_t, a^-_t being the right-continuous inverses of $A^+_t := \int_0^t 1_{\{X_s>0\}}\mathrm{d}s$ and $A^-_t := \int_0^t 1_{\{X_s\leq 0\}}\mathrm{d}s$, respectively. In particular, $X^{\Uparrow}$ and $X^{\Downarrow}$ jump into cemetery state at the times A^+_T and A^-_T, respectively, which is the only dependence on the time horizon T. We use the convention that the cemetery state in the above minimum is ignored so that minimum yields the other quantity. Thus for increasing T, the deaths of processes $X^{\Uparrow}$ and $X^{\Downarrow}$ can occur only later and hence $X^{\Uparrow}_t \wedge -(X^{\Downarrow}_{1-t})$ may only become smaller for each t, and so $Z'_1 \geq Z'_T$ a.s. Thus Z_T is stochastically dominated by Z_1, but from a simple sample path consideration we have $Z_1 = \overline{X}_1 - \underline{X}_1$ a.s. concluding the proof. □

In the above Lemma it is crucial to consider pre- and post-supremum processes together, since any of them can die arbitrarily early whereas the sum of life-times is exactly T. In particular, we can not conclude that $\sup_{t\in[0,1],t\leq\tau}\{\overline{X} - X_{\tau-t}\}$ is dominated by the analogous quantity for $T = 1$. In fact, the opposite is true. Furthermore, the results of this Section are false if we look only to the left (or to the right) of the time of supremum. This should not be confused with removal of the observations at the endpoints as discussed in Remark 3.3.

In the following we consider a family of Lévy processes

$$X^{(n)}_t = b_n X_{t/n} \tag{3.13}$$

with $b_n = 1/a_{1/n}$ whenever (3.8) holds. Let $(\sigma^{(n)}, \gamma^{(n)}, \Pi^{(n)})$ be the correspond-

ing Lévy triplets. It is noted that $\Pi^{(n)}(\mathrm{d}x) = \Pi(b_n^{-1}\mathrm{d}x)/n$,

$$\gamma^{(n)} = \frac{b_n}{n}\Big(\gamma - \int_{b_n^{-1}\le|x|<1} x\Pi(\mathrm{d}x)\Big) \quad \text{and} \quad \sigma^{(n)} = \sigma b_n/\sqrt{n}. \tag{3.14}$$

Lemma 3.3. *Let $X \in \mathcal{D}_{\alpha,\cdot}$. Then $\gamma^{(n)}, \sigma^{(n)}, \int_{|x|\le 1} x^2\Pi^{(n)}(\mathrm{d}x)$ have finite limits as $n \to \infty$. Moreover, for any $p < \alpha$ such that $\int_1^\infty x^p\Pi(\mathrm{d}x) < \infty$ we have*

$$\int_1^\infty x^p\Pi^{(n)}(\mathrm{d}x) \to \int_1^\infty x^p\widehat{\Pi}(\mathrm{d}x) < \infty.$$

Proof. The first part is a direct consequence of [63, Eq. (19)–(21)], see also [67, Thm. 15.14]. The second part relies on regular variation and conditions for the domains of attraction, and is given in Appendix B. □

Lemma 3.4. *If $\gamma_+^{(n)}, \sigma^{(n)}, \int_{|x|\le 1} x^2\Pi^{(n)}(\mathrm{d}x), \int_1^\infty x^p\Pi^{(n)}(\mathrm{d}x)$ are bounded then so is $\mathbb{E}(\overline{X}_1^{(n)})^p$.*

Proof. The following arguments seem to be rather standard. Let $Z_t^{(n)}$ be the process $X_t^{(n)}$, when the jumps exceeding 1 in absolute value are discarded and then the mean is subtracted; in other words we temporarily assume that $\Pi^{(n)}(-\infty,-1) = \Pi^{(n)}(1,\infty) = \gamma^{(n)} = 0$. Assume for the moment that $p > 1$ and note that $x^p \le ae^x$ for some $a > 0$ and all $x \ge 0$. Hence

$$\mathbb{E}\left|Z_1^{(n)}\right|^p/a \le \mathbb{E}\exp\left(\left|Z_1^{(n)}\right|\right) \le \mathbb{E}\exp\left(Z_1^{(n)}\right) + \mathbb{E}\exp\left(-Z_1^{(n)}\right)$$

showing that $\|Z_1^{(n)}\|_p$ is bounded if so are $\psi_{Z^{(n)}}(\pm 1)$, but the latter follows from the Lévy-Khintchine formula and boundedness of $\sigma^{(n)}, \int_{|x|\le 1} x^2\Pi^{(n)}(\mathrm{d}x)$. The process $Z^{(n)}$ is a martingale, and by Doob's martingale inequality we have

$$\|\overline{Z}_1^{(n)}\|_p \le \frac{p}{p-1}\|Z_1^{(n)}\|_p.$$

For $p \in (0,1]$ we simply use the inequality $x^p < 1 + x^2$ for all $x > 0$, and so $\mathbb{E}(\overline{Z}_1^{(n)})^p$ is bounded for all $n \ge 1$.

Note that

$$\overline{X}_1^{(n)} \le \overline{Z}_1^{(n)} + \gamma_+^{(n)} + P_1^{(n)},$$

where $P_1^{(n)}$ is an independent compound Poisson process with Lévy measure

$\Pi^{(n)}(\mathrm{d}x)1_{\{x\geq 1\}}$. But for any $p > 0$ we have the inequality $(x+y)^p \leq (1 \vee 2^{p-1})(x^p + y^p)$ for all $x, y > 0$. Thus it is left to show that $\mathbb{E}(P_1^{(n)})^p$ is bounded. By Minkowski inequality we find for $p \geq 1$ that

$$\|P_1^{(n)}\|_p^p \leq \|N^{(n)}\|_p^p \|\Delta^{(n)}\|_p^p \leq \mathbb{E}(N^{(n)})^{\lceil p \rceil} \int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x)/\lambda^{(n)} \tag{3.15}$$

where $N^{(n)}$ is Poisson with intensity $\lambda^{(n)} = \Pi^{(n)}(1,\infty)$, and the generic jump $\Delta^{(n)}$ is distributed according to $1_{\{x>1\}}\Pi^{(n)}(\mathrm{d}x)/\lambda^{(n)}$. But the moments of Poisson distribution are polynomial functions (with 0 free term) of its intensity. This shows that the right hand side of (3.15) is indeed bounded, because so are $\Pi^{(n)}(1,\infty) \leq \int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x)$. For $p \in (0,1)$ we use the simple bound: $(x+y)^p \leq x^p + y^p$ for all $x, y > 0$. □

Proof of Theorem 3.2. Observe that $V^{(n)}$ is the error made by considering the maximum of $X_t^{(n)}$ at the times $0, 1, \ldots, n$ as compared to its supremum on $[0, n]$. According to Lemma 3.2 we find that $V^{(n)}$ is first-order stochastically dominated by $\overline{X}_1^{(n)} - \underline{X}_1^{(n)}$; it is sufficient to look at the discretization epochs right next to the time of supremum. But from Lemma 3.3 and Lemma 3.4 (applied to X and $-X$) we readily see that the sequence $\mathbb{E}(\overline{X}_1^{(n)} - \underline{X}_1^{(n)})^p, n \geq 1$ is bounded given that $p < \alpha$ and $\int_{|x|>1} |x|^p \Pi(\mathrm{d}x) < \infty$. □

Proof of Theorem 3.1. Define a family of Lévy processes $X_t^{(n)}$ as in (3.13) with $b_n = n^{1/\alpha_+}$ where $\alpha_+ > \alpha \geq p$. Using Lemma 3.2 we obtain

$$\mathbb{E}(M - M^{(n)})^p = n^{-p/\alpha_+}\mathbb{E}(b_n(M - M^{(n)}))^p \leq n^{-p/\alpha_+}\mathbb{E}\left(\overline{X}_1^{(n)} - \underline{X}_1^{(n)}\right)^p.$$

Furthermore, we may apply Lemma 3.4 because according to Lemma 3.11 in Appendix B, all the relevant quantities have 0 limits. The proof is now complete for $p \leq \alpha$. In the cases (i) $p \leq \alpha = 2$ and (ii) $p \leq \alpha = 1$ with X b.v. we use Lemma 3.12 instead.

When $p > \alpha$ we simply take $\alpha_+ = p$ so that $b_n = n^{1/p}$. Note that

$$\int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x) = \int_{b_n^{-1}}^\infty x^p \Pi(\mathrm{d}x),$$

which is bounded since $p > \alpha \geq \beta_0$. The above reasoning now applies and we get the upper bound of order n^{-1}. □

4 Asymptotic probability of error in threshold exceedance

Consider the error probability $\mathbb{P}(M > x, M^{(n)} \leq x)$ in detection of threshold exceedance. The main aim of this section is to prove the following theorem.

Theorem 3.3. *Assume that $X \in \mathcal{D}_{\alpha,\cdot}$ with $\alpha > 1$. Then M has a continuous density, say $f_M(x)$, and for any $x > 0$*

$$b_n \mathbb{P}(M > x, M^{(n)} \leq x) \to f_M(x) \mathbb{E}\widehat{V}$$

as $n \to \infty$, where $b_n = 1/a_{1/n}$ and $\mathbb{E}\widehat{V}$ is given in Corollary 3.1.

The intuition behind this result is explained in Section 1. It is also noted that Theorem 3.3 has been established for a linear Brownian motion in [23], and later extended to an independent sum of a linear Brownian motion and a compound Poisson process in [38].

Remark 3.4. The result of Theorem 3.3 is also true for a shifted grid $(i+s)/n$ with all points in $[0, 1]$. Furthermore, with some additional effort one can show that the limit in Theorem 3.3 holds uniformly in all positive levels x away from 0 and all shifts $s \in [0, 1)$.

4.1 Preparatory results

The basic vehicle is the following result, which follows immediately from the generalized continuous mapping theorem [98, p. 2].

Lemma 3.5. *Consider a sequence of probability measures μ_n on a separable metric space S with a weak limit μ, and a sequence of measurable functions h_n on S such that $h_n(x_n) \to h(x)$ whenever $x_n \to x \in S$. If, moreover, the random variables $h_n(\mu_n)$, corresponding to the push-forward measures, are uniformly integrable then*

$$\int h_n(x) \mu_n(\mathrm{d}x) \to \int h(x) \mu(\mathrm{d}x).$$

It is noted that uniform integrability is needed to pass from the weak convergence to convergence of expectations. This result readily extends to defective

probability measures with $\mu_n(S) \leq 1$, in which case a proper random variables $h_n(\mu_n)$ is obtained by adding the missing mass at 0.

Throughout this section we assume that

$$X \in \mathcal{D}_{\alpha,\cdot} \qquad \text{with } \alpha > 1, \tag{3.16}$$

which implies the Orey's condition:

$$\liminf_{\epsilon \downarrow 0} \epsilon^{\gamma-2} \left(\sigma^2 + \int_{-\epsilon}^{\epsilon} x^2 \Pi(\mathrm{d}x) \right) > 0 \tag{3.17}$$

for some $\gamma \in (1, 2]$, since the function in brackets must be $\mathrm{RV}_{2-\alpha}$ according to [63] and then one can take $\gamma \in (1, \alpha)$; $\gamma = 2$ is possible only when $\sigma^2 > 0$. Therefore, X_t has a smooth bounded density, say $p(t, x)$, for each $t > 0$, see [86, Prop. 28.3] and [79, Thm. 3.1]. Moreover, $p(t, x)$ is continuous and strictly positive on $(0, \infty) \times \mathbb{R}$, see [88]. The following Lemma 3.6 shows that $p(t, x)$ must be bounded on any set away from the origin. We could not locate such a result in the literature, but see [73, Prop. III.6] and [49, Prop. 6.3] for some related results.

Lemma 3.6. *Assume that* (3.17) *holds with* $\gamma > 1$. *Then for any* $\delta > 0$ *the function* $p(t, x)$ *is upper bounded for all* $t > 0, x \in \mathbb{R}$ *such that* $t > \delta$ *or* $|x| > \delta$. *The conclusion may fail for any* $\gamma < 1$.

Proof. See Appendix C. □

It is noted that the proof of Lemma 3.6 also shows that $p(0+, x) = 0$ for any $x \neq 0$, when $\gamma > 1$ and that for $\gamma < 1$ this does not need to be the case. In the following we write $\mathbb{P}_z$ for the law of the shifted Lévy process X with $X_0 = z$.

Lemma 3.7. *Assume that* (3.17) *holds with* $\gamma > 1$. *Then, for any* $t > 0$ *the measure* $\mathbb{P}_z(X_t \in \mathrm{d}x, \underline{X}_t > 0)$ *has a continuous density* $f_{z,t}(x)$ *which is bounded and jointly continuous on* $\{(x, z) : x > \delta, z > 0\}$ *for any* $\delta > 0$.

Proof. We start as in the proof of [43, Lemma 8]. By the strong Markov property applied at $\tau_0 = \inf\{t \geq 0 : X_t < 0\}$, the first time the process becomes negative, we find that

$$\mathbb{P}_z(X_t \in \mathrm{d}x, \underline{X}_t \leq 0)/\mathrm{d}x = \int_{s \in (0,t), y \geq 0} \mathbb{P}_z(\tau_0 \in \mathrm{d}s, -X_s \in \mathrm{d}y) p(t - s, x + y),$$

where we use $\mathbb{P}_z(\underline{X}_t = 0) = 0$ and $\mathbb{P}_z(\tau_0 = t) = 0$. Note that it is enough to establish that the right hand side is jointly continuous for $x > \delta, z > 0$, because then

$$f_{z,t}(x) = p(t, x - z) - \int_{s\in(0,t),y\geq 0} \mathbb{P}_z(\tau_0 \in \mathrm{d}s, -X_s \in \mathrm{d}y)p(t - s, x + y)$$

must be bounded and jointly continuous.

For any $z_n \to z > 0, x_n \to x > \delta$ we need to show that

$$\int_{s\in(0,t),y\geq 0} \mathbb{P}_{z_n}(\tau_0 \in \mathrm{d}s, -X_s \in \mathrm{d}y)p(t - s, x_n + y)$$

has the corresponding limit. This readily follows from Lemma 3.5, joint continuity of $p(t, x)$ and the fact that it is bounded for all $t > 0$ and x away from 0, see Lemma 3.6. □

4.2 Proofs

Proposition 3.3. *Assume* (3.16) *and chose* $\delta \in (0, 1)$. *Then* $\mathbb{P}(M \in \mathrm{d}x, \tau \geq \delta)$ *has a continuous density, say* $f_M(x; \delta)$, *and*

$$b_n\mathbb{P}(M > x, M_n < x, \tau \geq \delta) \to f_M(x; \delta)\mathbb{E}\widehat{V}$$

as $n \to \infty$ *for any* $x > 0$.

Proof. The upper bound on $\mathbb{P}(M > x, M_n < x, \tau \geq \delta)$ is obtained by restricting the discretization grid to the times exceeding δ, see Figure 3.2. Considering the post-δ process $X_{t+\delta} - X_\delta, t \geq 0$ (independent of $X_t, t \leq \delta$ and having the same law) and its functionals

$$M_{[\delta,1]} := \sup_{t\in[\delta,1]} X_t - X_\delta, \qquad V^{(n)}_{[\delta,1]}/b_n := \sup_{t\in[\delta,1]} X_t - \max_{i/n\in[\delta,1]} X_{i/n}$$

we have the upper bound:

$$\begin{aligned}&\int_{z>0,y>0} \mathbb{P}(M_{[\delta,1]} \in \mathrm{d}z, V^{(n)}_{[\delta,1]}/b_n \in \mathrm{d}y)\mathbb{P}(x \vee \overline{X}_\delta < X_\delta + z < x + y)\\ =&\int_{z>0,y>0} \mathbb{P}(M_{[\delta,1]} \in \mathrm{d}z, V^{(n)}_{[\delta,1]} \in \mathrm{d}y)\mathbb{P}_z(X_\delta \in (x, x + y/b_n], \underline{X}_\delta > 0) \qquad (3.18)\end{aligned}$$

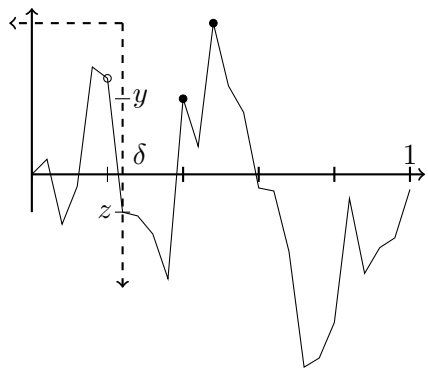


Figure 3.2: Schematic sample path and the reversal

where in the second line we have used the time and space reversal (yielding a process with the same law) at the time δ, see the dashed axes in Figure 3.2. The lower bound is obtained by restricting to the event when the supremum over $[0, \delta)$ is smaller than the discretized maximum over $[\delta, 1]$, implying that discretization epochs before δ do not matter. Thus the lower bound is given by (3.18) with a single change, where $\underline{X}_\delta > 0$ is replaced by $\underline{X}_\delta > y/b_n$.

According to [63] the measure $\mathbb{P}(M_{[\delta,1]} \in \mathrm{d}z, V^{(n)}_{[\delta,1]} \in \mathrm{d}y)$ has the weak limit $\mathbb{P}(M_{[\delta,1]} \in \mathrm{d}z) \times \mathbb{P}(\widehat{V} \in \mathrm{d}y)$, because we are discretizing the process with the same law and the limit does not depend on the time horizon neither on the grid shift. Moreover, for any positive $(y_n, z_n) \to (y, z)$ the mean value theorem and Lemma 3.7 show that

$$b_n \mathbb{P}_{z_n}(X_\delta \in (x, x + y_n/b_n], \underline{X}_\delta > 0) = f_{z_n,\delta}(x_n) y_n \to y f_{z,\delta}(x),$$

where $x_n \in (x, x + y_n/b_n)$. Moreover, the same limit holds true for

$$b_n \mathbb{P}_{z_n}(X_\delta \in (x, x + y_n/b_n], \underline{X}_\delta > y_n/b_n) \\ = b_n \mathbb{P}_{z_n - y_n/b_n}(X_\delta \in (x - y_n/b_n, x], \underline{X}_\delta > 0)$$

appearing in the lower bound. Assume $\beta_\infty > 1$, see (3.6). Now Lemma 3.5 applies, because the uniform integrability of the corresponding family of measures follows from that of $V^{(n)}_{[\delta,1]}$ and the boundedness of $f_{z,\delta}(x)$. Hence the limit of

interest is

$$\int y f_{z,\delta}(x)\mathbb{P}(M_{[\delta,1]} \in \mathrm{d}z) \times \mathbb{P}(\widehat{V} \in \mathrm{d}y) = f_M(x;\delta)\mathbb{E}\widehat{V},$$

where the continuity of $\int f_{z,\delta}(x)\mathbb{P}(M_{[\delta,1]} \in \mathrm{d}z) = f_M(x,\delta)$ follows from the boundedness of $f_{z,\delta}(x)$ and the dominated convergence theorem.

For $\beta_\infty \leq 1$ note that $b_n\mathbb{P}(A^{(n)}) = b_n\mathrm{O}(1/n) \to 0$, where the event $A^{(n)}$ is defined in (3.10). Thus we may work on the event $A^{(n)^c}$ for the post-δ process, and apply Proposition 3.2 to get uniform integrability. □

Lemma 3.8. *Assume* (3.16). *Then* $f_M(x) := \lim_{\delta\downarrow 0} f_M(x;\delta)$ *is a density of* M *continuous for* $x > 0$.

Proof. According to [29],

$$f_M(x;\delta) = \int_{s\in(\delta,1),y>0} \underline{n}(X_{s/2} \in \mathrm{d}y) f_{y,s/2}(x) n(1-s<\zeta)\mathrm{d}s,$$

which is also true for $\delta = 0$ yielding $f_M(x)$; here ζ denotes the life time.

It is left to show that $f_M(x)$ is continuous, for which it is sufficient to establish that

$$\int_{s\in(0,\delta),y>0} \underline{n}(X_s \in \mathrm{d}y) f_{y,s}(x) n(1-2s<\zeta)\mathrm{d}s \to 0$$

as $\delta \downarrow 0$ uniformly in $x \geq x_0 > 0$, because $f_M(x;\delta)$ is continuous according to Proposition 3.3. Recall from the proof of Lemma 3.7 that $f_{y,s}(x) \leq p(s, x-y)$ and the latter is bounded when $x - y$ is away from 0, see Lemma 3.6. Hence it is sufficient to show that

$$\int_0^\delta \underline{n}(X_s > x_0/2) \sup_x p(s,x)\mathrm{d}s \to 0,$$

where $\sup_x p(s,x) = \mathrm{O}(s^{-1/\gamma})$ with $\gamma \in (1,\alpha)$ according to [79, Thm. 3.1]. But $\underline{n}(X_s > x_0/2)$ is upper bounded by a function in RV_ρ as $s \downarrow 0$ according to [19, Prop. 3], implying that it is bounded for small s (converges to 0) since necessarily $\rho > 0$. The proof is now complete. □

Proof of Theorem 3.3. We need to show that

$$\limsup_{n\to\infty} b_n \mathbb{P}(M > x, M^{(n)} < x, \tau < \delta) < C_x(\delta), \tag{3.19}$$

where $C_x(\delta) \downarrow 0$ as $\delta \downarrow 0$, because then Proposition 3.3 shows that

$$\begin{aligned} f_M(x;\delta) &\le \liminf b_n \mathbb{P}(M > x, M^{(n)} < x) \\ &\le \limsup b_n \mathbb{P}(M > x, M^{(n)} < x) \le f_M(x;\delta) + C_x(\delta) \end{aligned}$$

implying the result with the help of Lemma 3.8.

We assume that $\beta_\infty > 1$, see (3.6), since the other case can be handled in exactly the same way as in Proposition 3.3. In order to remove the effect of shifting the grid (needed later) we observe that

$$\mathbb{P}(M > x, M^{(n)} < x, \tau < \delta) \le \mathbb{P}(M > x, \underline{M}^{(n)} < x, \tau < \delta),$$

where

$$\underline{M}^{(n)} = \inf_{t\in[0,1/n]} \{X_{\tau-t} \vee X_{\tau-t+1/n}\}$$

with the convention that $X_t = -\infty$ if $t \notin [0,1]$. Recall from the proof of Theorem 3.2 that $b_n(M - \underline{M}^{(n)})$ is uniformly integrable, see also Lemma 3.2. Moreover, it can be shown that $b_n(M - \underline{M}^{(n)})$ has a weak (Rényi) limit, call it $\widehat{\underline{V}}$, which corresponds to taking the same map of the limiting process seen from the supremum, see [63]; the limiting process is composed of two independent pieces neither of which can jump at a fixed time.

Consider a stopping time $\widehat{\tau} = \inf\{t \ge 1/2 : X_t = 0\}$ and note that $p = \mathbb{P}(\widehat{\tau} < 1 \wedge \tau_x) > 0$, because an ub.v. process hits 0 immediately [21, 68]. For $\delta \le 1/4$ and $n \ge 4$ we establish that

$$p\mathbb{P}(M > x, \underline{M}^{(n)} < x, \tau < \delta) \le \mathbb{P}^{[0,3/2]}(M > x, \underline{M}^{(n)} < x, \tau \ge 1/2, D_\delta > x) \tag{3.20}$$

where $\mathbb{P}^{[0,3/2]}$ indicates that we consider the process on the interval $[0,3/2]$ instead of $[0,1]$, and $D_\delta = \sup_{t\in(0,\delta], t\le\tau-1/2}\{M - X_{\tau-t}\}$. The left hand side (by the strong Markov property) is the probability that our process hits 0 in the interval $[1/2,1)$ (it has not yet crossed x) and in the following unit of time it

achieves its supremum exceeding x withing δ time units, while the corresponding $\underline{M}^{(n)}$ is below x, see Figure 3.3. It is not hard to see that this event implies the event on the right hand side ($\underline{M}^{(n)}$ may become larger but must still be below x), and thus the inequality follows. It is crucial here that the quantities do not depend on the grid shifting due to the random time $\widehat{\tau}$.

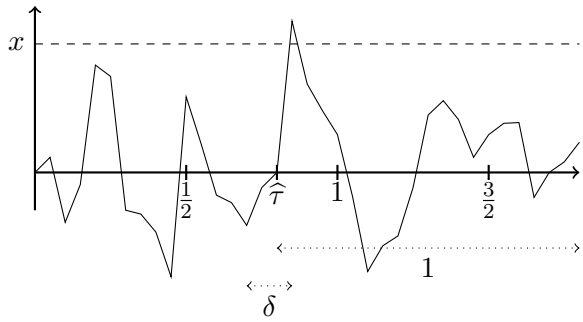


Figure 3.3: Schematic sample path explaining the bound in (3.20)

The same arguments as in Proposition 3.3 show that

$$b_n \mathbb{P}^{[0,3/2]}(M > x, \underline{M}^{(n)} < x, \tau \geq 1/2, D_\delta > x) \to \mathbb{E}\widehat{\underline{V}} \int f_{z,1/2}(x) \mu_{\delta,x}(\mathrm{d}z),$$

where $\mu_{\delta,x}(\mathrm{d}z) = \mathbb{P}(M \in \mathrm{d}z, \sup_{t\in(0,\delta], t\leq\tau}\{M - X_{\tau-t}\} > x)$. It is noted that now we are splitting the process at $1/2$, and the upper bound would be enough for what follows. Note also that restriction of $\underline{M}^{(n)}$ to the times larger than $1/2$ makes it only smaller due to the inner supremum in its definition.

Finally, since $f_{z,1/2}(x)$ is bounded for all $z > 0$ according to Lemma 3.7, we find that

$$\limsup_{n\to\infty} b_n \mathbb{P}(M > x, M^{(n)} < x, \tau < \delta) < C_x \mathbb{P}(\sup_{t\in(0,\delta], t\leq\tau} \{M - X_{\tau-t}\} > x)$$

for some constant C_x not depending on δ. Moreover, the probability on the right hand side must decay to 0 as $\delta \downarrow 0$, because ub.v. process does not jump at τ. The bound in (3.19) is now established and the proof is thus complete. □

4.3 Further bounds and comments

It is noted that the above arguments may also be used to provide an asymptotic upper bound on the detection error in the case when (3.8) is not satisfied. Assuming that a non-monotone X has a jointly continuous density $p(t,x)$ bounded for $|x|>\delta, t>0$ (e.g. (3.17) holds with $\gamma>1$), we find with the help of Theorem 3.1 that for any $\alpha_+ > \alpha \vee 1$ and $\delta \in (0,1)$:

$$\mathbb{P}(M > x, M^{(n)} \le x, \tau \ge \delta) = \mathrm{O}(n^{-1/\alpha_+}), \qquad \text{as } n \to \infty.$$

Furthermore, we may strengthen this bound to $\mathrm{O}(n^{-1/2})$ when $\alpha = 2$, and to $\mathrm{O}(n^{-1})$ when X is b.v. process, see Theorem 3.1.

Moreover, we may also take $\delta = 0$ apart from b.v. case when (i) $\gamma' = 0$ or (ii) point 0 is not in the support of $\Pi(\mathrm{d}x)$, because in these cases the trick in the proof of Theorem 3.3 does not apply; there may be other ways to establish such bounds for these processes, however.

There is also an asymptotic lower bound of order n^{-1} on the detection error under some minor conditions. We omit the analysis of one-sided processes and state the following:

Lemma 3.9. *Assume that X has jumps of both signs. Then*

$$\liminf_{n\to\infty} n\mathbb{P}(M > x, M^{(n)} \le x) > 0.$$

Proof. Let $\delta > 0$ be a point of continuity of $\overline{\Pi}_-(x)$ such that $\overline{\Pi}_-(\delta) > 0$. Then $\mathbb{P}(X_t < -\delta)/t \to \overline{\Pi}_-(\delta) > 0$ as $t \downarrow 0$, see e.g. [48]. Now consider a lower bound

$$\begin{aligned}
&\mathbb{P}(M > x, M^{(n)} \le x) \\
&\ge \mathbb{P}(\tau_x < 1, X_{\tau_x} - x < \delta/2, \{\tau_x n\} < 1/2, \sup_{t\in[\tau_x + 1/(2n),1]} X_t < x) \\
&\ge \mathbb{P}(\tau_x < 1, X_{\tau_x} - x < \delta/2, \{\tau_x n\} < 1/2)\mathbb{P}(X_{1/(2n)} < -\delta)\mathbb{P}(M < \delta/2).
\end{aligned}$$

Hence it is left to show that

$$\liminf_{n\to\infty} \mathbb{P}(\tau_x < 1, X_{\tau_x} - x < \delta/2, \{\tau_x n\} < 1/2) > 0. \tag{3.21}$$

The compensation formula applied to the Poisson point process of jumps

with intensity $\mathrm{d}t \times \Pi(\mathrm{d}y)$, see also [41], yields

$$\mathbb{P}(\tau_x \in A, X_{\tau_x} \in (x, x+\delta/2)) \\ = \int_A \int_0^\infty \mathbb{P}(\overline{X}_t \leq x, X_t + y \in (x, x+\delta/2))\Pi(\mathrm{d}y)\mathrm{d}t \quad (3.22)$$

showing that the corresponding measure is absolutely continuous. Hence $\{\tau_x n\}$ converges weakly to a uniform random variable on the event $\tau_x < 1, X_{\tau_x} \in (x, x+\delta/2)$; here we ignore the possibility of creeping over x. Hence (3.21) is lower bounded by

$$\mathbb{P}(\tau_x < 1, X_{\tau_x} \in (x, x+\delta/2))/2,$$

and it is left to show that this quantity is non-zero. Assume that 0 belongs to the support of Π. Then using the ideas from [86, §24] we find that the support of $\mathbb{P}(\overline{X}_t \leq x, X_t \in \mathrm{d}x)$ is given by $(-\infty, x]$, and the positivity easily follows from (3.22). The case when 0 is not in the support of $\Pi(\mathrm{d}x)$ and there is no Brownian component corresponds to a possibly drifted compound Poisson. In view of Theorem 3.3 it is left to consider the latter, and in this case the statement follows from tedious but trivial considerations. □

Appendix

A Proofs for Section 2

We will need the following Karamata's theorem in the boundary case, see [15, Thm. 1.5.9a–b]:

Lemma 3.10. *Let $\ell(x) \in \mathrm{RV}_0$ (as $x \downarrow 0$), such that $\int_x^1 \ell(t)\mathrm{d}t < \infty$ for any $x \in (0,1)$.*

(i) Then $\int_x^1 t^{-1}\ell(t)\mathrm{d}t/\ell(x) \to \infty$ and the numerator is RV_0.

(ii) If $\int_0^1 t^{-1}\ell(t)\mathrm{d}t < \infty$ then $\int_0^x t^{-1}\ell(t)\mathrm{d}t/\ell(x) \to \infty$ and the numerator is RV_0.

Proof of Proposition 3.1. Let us verify the conditions of [63, Thm. 2], which is trivial in the case $\alpha \in (0,1) \cup (1,2)$. In case $\alpha = 2$ our assumption implies that $\overline{\Pi}_-(x) \le C\overline{\Pi}_+(x)$ for some finite C and all x small enough. Hence for such x we have

$$\frac{x^2\overline{\Pi}(x)}{v(x)} \le (C+1)\frac{x^2\overline{\Pi}_+(x)}{\int_0^x y^2\Pi(\mathrm{d}y)}$$

and it is left to show that this fraction converges to 0. Letting $\ell(x) = x^2\overline{\Pi}_+(x) \in \mathrm{RV}_0$ we find using integration by parts that

$$\int_0^x y^2\Pi(\mathrm{d}y) = -\int_0^x y^2\mathrm{d}\overline{\Pi}_+(y) = 2\int_0^x y^{-1}\ell(y)\mathrm{d}y - \ell(x) + \ell(0+)$$

showing that $\ell(0+)$ is convergent. According to Lemma 3.10(ii) we must have $\ell(0) = 0$ and the above must explode when divided by $\ell(x)$ as $x \downarrow 0$. The proof in the case $\alpha = 2$ is now complete.

In the case $\alpha = 1$ we need to show that $m(x)/(x\overline{\Pi}(x)) \to \pm\infty$. Our assumption implies that $\overline{\Pi}_+(x) - \overline{\Pi}_-(x) > c\overline{\Pi}_+(x)$ for some $c > 0$ and all small enough x, as well as $\overline{\Pi}(x) \leq 2\overline{\Pi}_+(x)$. We let $\ell_\pm(x) = x\overline{\Pi}_\pm(x)$ and first consider b.v. case. As above, observe that $\ell_+(0) = 0$ and thus also $\ell_-(0) = 0$. Now

$$\begin{aligned} m(x) &= -\int_0^x y\mathrm{d}\overline{\Pi}_+(y) + \int_0^x y\mathrm{d}\overline{\Pi}_-(y) \\ &= \int_0^x y^{-1}(\ell_+(y) - \ell_-(y))\mathrm{d}y - (\ell_+(x) - \ell_-(x)) \\ &> c\int_0^x y^{-1}\ell_+(y)\mathrm{d}y - \ell_+(x) \end{aligned}$$

for small enough x. Hence for small x

$$\frac{m(x)}{x\overline{\Pi}(x)} > \frac{c\int_0^x y^{-1}\ell_+(y)\mathrm{d}y}{2\ell_+(x)} - 1/2 \to \infty$$

according to Lemma 3.10(ii). This shows that (3.8) holds true with the limit process being a positive drift.

In ub.v. case we have

$$\int_x^1 y\Pi(\mathrm{d}y) = \int_x^1 y^{-1}\ell(y)\mathrm{d}y - \ell(1) + \ell(x) \to \infty$$

showing that $\int_x^1 y^{-1}\ell(y)\mathrm{d}y \to \infty$, see Lemma 3.10(i). Now

$$\begin{aligned} m(x) &= \gamma - \int_x^1 y^{-1}(\ell_+(y) - \ell_-(y))\mathrm{d}y + (\ell_+(1) - \ell_-(1)) - (\ell_+(x) - \ell_-(x)) \\ &< c' - c\int_x^1 y^{-1}\ell_+(y)\mathrm{d}y < -\frac{c}{2}\int_x^1 y^{-1}\ell_+(y)\mathrm{d}y \end{aligned}$$

for some constant c' and small enough x, where the last line is implied by the divergence of the integral. Using Lemma 3.10(i) once again we find

$$\frac{m(x)}{x\overline{\Pi}(x)} < -\frac{c/2\int_x^1 y^{-1}\ell_+(y)\mathrm{d}y}{2\ell_+(x)} \to -\infty$$

and the proof is complete. □

B Proofs for Section 3

Recall that $(\gamma^{(n)}, \sigma^{(n)}, \Pi^{(n)})$ is the Lévy triplet of the rescaled process $b_n X_{t/n}$ as defined in (3.14).

Proof of Lemma 3.3. Using $\Pi^{(n)}(\mathrm{d}x) = \Pi(b_n^{-1}\mathrm{d}x)/n$ we find that

$$\int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x) = \frac{b_n^p}{n}\int_{b_n^{-1}}^\infty x^p \Pi(\mathrm{d}x).$$

Since b_n is regularly varying at ∞ with index $1/\alpha$ we see that $b_n^p/n \to 0$ and so it is sufficient to consider the limit of

$$\frac{b_n^p}{n}\int_{b_n^{-1}}^1 x^p \Pi(\mathrm{d}x) = \frac{\varepsilon}{a_\varepsilon^p}\int_{a_\varepsilon}^1 x^p \Pi(\mathrm{d}x),$$

where $\varepsilon = 1/n \downarrow 0$. Recall also the definitions of $m(x)$ and $v(x)$, the truncated mean and variance functions, given in (3.5).

Consider the case where $\widehat{X}$ is a Brownian motion, so that $p < 2$. From [63, Thm. 6 (i)] we see that $v \in \mathrm{RV}_0$ and $\varepsilon v(a_\varepsilon)/a_\varepsilon^2 \to \widehat{\sigma}^2$. Hence

$$\frac{\varepsilon}{a_\varepsilon^p}\int_{a_\varepsilon}^1 x^p \Pi(\mathrm{d}x) \le \frac{\varepsilon}{a_\varepsilon^p}\int_{a_\varepsilon}^1 x^{p-2}\mathrm{d}v(x) = \frac{\varepsilon v(a_\varepsilon)}{a_\varepsilon^2}\cdot\frac{\int_{a_\varepsilon}^1 x^{p-2}\mathrm{d}v(x)}{a_\varepsilon^{p-2}v(a_\varepsilon)} \to 0,$$

because the first ratio converges to $\widehat{\sigma}^2$ and the second to 0 according to the Karamata's theorem, see [15] or [63, Thm. 6].

Consider the case of a strictly α-stable process $\widehat{X}$. Let $f_+(x) := \Pi(x, 1)$, $f_-(x) := \Pi(-1, -x)$ and $f(x) = f_-(x) + f_+(x)$ then according to [63, Thm. 2] we have $f_\pm(x) \in \mathrm{RV}_{-\alpha}$ (or at least the dominating one) and $\varepsilon f_\pm(\alpha_\varepsilon) \to \frac{\widehat{c}_\pm}{\alpha}$.

Since

$$\int_x^1 y^p \Pi(\mathrm{d}y) = x^p f_+(x) + p \int_x^1 y^{p-1} f_+(y) \mathrm{d}y$$

thus we have

$$\begin{aligned}\frac{\varepsilon}{a_\varepsilon^p} \int_{a_\varepsilon}^1 x^p \Pi(\mathrm{d}x) &= \frac{\varepsilon}{a_\varepsilon^p} \left(a_\varepsilon^p f_+(a_\varepsilon) + p \int_{a_\varepsilon}^1 x^{p-1} f_+(x) \mathrm{d}x \right) \\ &= \varepsilon f_+(a_\varepsilon) \cdot \left(1 + \frac{p \int_{a_\varepsilon}^1 x^{p-1} f_+(x) \mathrm{d}x}{a_\varepsilon^p f_+(a_\varepsilon)} \right) \to \frac{\widehat{c}_+}{\alpha} \cdot \left(1 + \frac{p}{\alpha - p} \right) = \frac{\widehat{c}_+}{\alpha - p}\end{aligned}$$

and the result follows.

Finally, consider the case, where $\widehat{X}$ is a linear non-zero drift process. Then necessarily $\alpha = 1$ and according to [63, Thm. 6 (ii)] we must have $x\overline{\Pi}(x)/m(x) \to 0$ and $\varepsilon m(a_\varepsilon)/a_\varepsilon \to \widehat{\gamma}$. Letting $M(x) = \int_{x \le |y| < 1} |y|^p \Pi(\mathrm{d}y)$ note that it is sufficient to show that $\varepsilon M(a_\varepsilon)/a_\varepsilon^p \to 0$.

Let $f(x) = f_+(x) + f_-(x)$. The main difficulty here is that $f(x)$ is does not necessarily belong to the class RV_{-1} however we do have that $m \in \mathrm{RV}_0$ according to [63, Proof of Thm. 6], see also its proof. We have $xf(x)/m(x) \to 0$ and $\varepsilon m(a_\varepsilon)/a_\varepsilon \to \widehat{\gamma}$ thus $\varepsilon f(a_\varepsilon) \to 0$ and for any $\delta > 0$ there exists x_0 such that $xf(x) \le \delta m(x)$ for $x < x_0$. Then

$$\begin{aligned}\frac{\varepsilon}{a_\varepsilon^p} \int_{a_\varepsilon \le |x| < 1} |x|^p \Pi(\mathrm{d}x) &= \frac{\varepsilon}{a_\varepsilon^p} \left(a_\varepsilon^p f(a_\varepsilon) + p \int_{a_\varepsilon}^1 x^{p-1} f(x) \mathrm{d}x \right) \\ &= \varepsilon f(a_\varepsilon) + \frac{\varepsilon m(a_\varepsilon)}{a_\varepsilon} \cdot \frac{p \int_{a_\varepsilon}^{x_0} x^{p-2} m(x) \mathrm{d}x}{a_\varepsilon^{p-1} m(a_\varepsilon)} \cdot \delta + \frac{\varepsilon}{a_\varepsilon^p} \int_{x_0}^1 x^{p-1} f(x) \mathrm{d}x\end{aligned}$$

The first and the third term converge to 0. Since $m \in \mathrm{RV}_0$, then according to Karamata's Theorem we have $\frac{\int_{a_\varepsilon}^{x_0} x^{p-2} m(x) \mathrm{d}x}{a_\varepsilon^{p-1} m(a_\varepsilon)} \to \frac{1}{1-p}$ and since the choice of $\delta > 0$ was arbitrary, we conclude that

$$\frac{\varepsilon}{a_\varepsilon^p} \int_{a_\varepsilon \le |x| < 1} |x|^p \Pi(\mathrm{d}x) \to 0,$$

and the proof is complete. □

Next, we consider the general case where (3.8) does not necessarily hold, but redefine $b_n = n^{1/\alpha_+}$ for some $\alpha_+ > \alpha$, and thus also $X_t^{(n)} = b_n X_{t/n}$.

Lemma 3.11. *For any $\alpha_+ > \alpha$ and $b_n = n^{1/\alpha_+}$ we have $X_1^{(n)} \xrightarrow{\mathrm{d}} 0$. If, moreover, $p < \alpha_+$ and $\int_1^\infty x^p \Pi(\mathrm{d}x) < \infty$ then $\int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x) \to 0$ as $n \to \infty$.*

Proof. First, we consider the integral. As in the beginning of the proof of Lemma 3.3 it is sufficient to observe that

$$\tfrac{b_n^p}{n} \int_{1/b_n}^1 x^p \Pi(\mathrm{d}x) \le \tfrac{1}{n} \int_{1/b_n}^1 (b_n x)^{\beta_+} \Pi(\mathrm{d}x) \le \tfrac{b_n^{\beta_+}}{n} \int_0^1 x^{\beta_+} \Pi(\mathrm{d}x) \to 0,$$

where $\beta_+ \in (\alpha \vee p, \alpha_+)$ and the latter integral is finite because $\beta_+ > \alpha \ge \beta_0$.

According to [67, Thm. 15.14] the convergence $X_1^{(n)} \xrightarrow{\mathrm{d}} 0$ is equivalent to $m(b_n^{-1})b_n/n, v(b_n^{-1})b_n^2/n, \overline{\Pi}_\pm(ub_n^{-1})/n \to 0$ for all $u > 0$. All of these limits can be shown using the above trick, and we only consider the first quantity (the most tedious). If $\int_{|x|<1} |x| \Pi(\mathrm{d}x) < \infty$ then $m(b_n^{-1}) = \gamma' + \int_{|y|<b_n^{-1}} y \Pi(\mathrm{d}y)$. The case $\alpha = 1$ is trivial, but for $\alpha < 1$ we have

$$\frac{b_n}{n} \int_{|y|<b_n^{-1}} |y| \Pi(\mathrm{d}y) \le \frac{1}{n} \int_{|y|<b_n^{-1}} |b_n y|^{\beta_+} \Pi(\mathrm{d}y) \le \frac{b_n^{\beta_+}}{n} \int_{|y|<1} |y|^{\beta_+} \Pi(\mathrm{d}y) \to 0$$

with $\beta_+ \in (\alpha, \alpha_+ \wedge 1)$. When $\int_{|x|<1} |x| \Pi(\mathrm{d}x) = \infty$ we have $\alpha \ge 1$ and it is sufficient to note that

$$\frac{b_n}{n} \int_{b_n^{-1}<|y|<1} |y| \Pi(\mathrm{d}y) \le \frac{b_n^{\beta_+}}{n} \int_{|y|<1} |y|^{\beta_+} \Pi(\mathrm{d}y) \to 0$$

for $\beta_+ \in (\alpha, \alpha_+)$. The proof is concluded. □

Lemma 3.12. *In the cases (i) $p \le \alpha = 2$ and (ii) $p \le \alpha = 1$ with X b.v., the sequences $|\gamma^{(n)}|, \sigma^{(n)}, \int_{|x|\le 1} x^2 \Pi^{(n)}(\mathrm{d}x), \int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x)$ for the scaling $b_n = n^{1/\alpha}$ are bounded provided that $\int_1^\infty x^p \Pi(\mathrm{d}x) < \infty$.*

Proof. The statement is trivially true for $\sigma^{(n)}$ and the rest follows using the simple trick from Lemma 3.11. We only consider

$$\int_1^\infty x^p \Pi^{(n)}(\mathrm{d}x) = \frac{1}{n} \int_{n^{-1/\alpha}}^\infty (n^{1/\alpha} x)^p \Pi(\mathrm{d}x) \le C + \frac{1}{n} \int_{n^{-1/\alpha}}^1 (n^{1/\alpha} x)^\alpha \Pi(\mathrm{d}x),$$

where we used $p/\alpha - 1 \le 0$ and convergence of $\int_1^\infty x^p \Pi(\mathrm{d}x)$. The second term converges to $\int_0^1 x^\alpha \Pi(\mathrm{d}x) < \infty$, which is finite in both cases (i) and (ii). □

Proof of Proposition 3.2. It is clear that $1_{\{A^{(n)c}\}}$ converges to 1 in probability and so by Slutsky's Lemma we find that $\widetilde{V}^{(n)} \xrightarrow{\mathrm{d}} \widehat{V}$. Thus it is left to show uniform integrability.

Conditional on the event $\{N = k, A^{(n)c}\}$, split the process into $k+1$ pieces separated by the big jumps, where each piece starts at 0 and does not include the terminating big jump. Let $V_{i,k}^{(n)}$ be the (conditional) discretization error for the supremum of the i-th piece keeping the original grid. Since each piece is at least $1/n$ long and conditioning affects only the length of the piece, the same argument as in the proof of Theorem 3.2 based on Lemma 3.2 shows that the $1+\epsilon$ moment of $V_{i,k}^{(n)}$ is bounded by a constant C not depending on i, k. Note that by removing big jumps we still have a process in $\mathcal{D}_{\alpha,\cdot}$, but now $\beta_\infty = \infty$.

It is left to note that the conditional $V^{(n)}$ is bounded by the maximum over $V_{i,k}^{(n)}$ which in turn is bounded by the sum. Hence using Minkowski's inequality we find

$$\begin{aligned}\mathbb{E}\left(\widetilde{V}^{(n)}\right)^{1+\epsilon} &\leq \sum_{k=0}^{\infty} \mathbb{P}(N=k)\mathbb{E}\left(\left(V^{(n)}\right)^{1+\epsilon} | A^{(n)c}, N=k\right) \\ &\leq \sum_{k=0}^{\infty} \mathbb{P}(N=k)(k+1)^{1+\varepsilon} C \leq \mathbb{E}(1+N)^{1+\varepsilon} C < \infty\end{aligned}$$

and the proof is complete. □

Proposition 3.4. *Let X be a compound Poisson process with drift $\gamma \in \mathbb{R}$ satisfying $\int_{|x|>1} |x|\Pi(\mathrm{d}x) < \infty$. Then* (3.11) *holds true.*

Proof. For $n \in \mathbb{N}$, $k \in \{0, \ldots, n-1\}$ define

$$N_{k,n} := \#\left\{t \in \left(\tfrac{k}{n}, \tfrac{k+1}{n}\right) : X_{t-} \neq X_t\right\}, \quad A_{k,n} := \{N_{k,n} \leq 1\}$$

and put $A_n := \bigcap_{k=0}^{n-1} A_{k,n}$. We have the decomposition

$$n\mathbb{E}(M - M^{(n)}) = n\mathbb{E}(M - M^{(n)}; A_n) + n\mathbb{E}(M - M^{(n)}; A_n^c).$$

Step 1. Show that $n\mathbb{E}(M - M^{(n)}; A_n) \to \frac{1}{2}|\gamma|\mathbb{P}(\tau \in (0,1))$. This is clear when $\gamma = 0$; in the following we assume $\gamma \neq 0$. Since then X is in a domain of

attraction of linear drift, it follows that

$$\Big(n(M - M^{(n)}) \mid \tau \in (0,1)\Big) \xrightarrow{\mathrm{d}} \widehat{V} \stackrel{\mathrm{d}}{=} |\gamma| U,$$

where U is uniformly distributed over $[0,1]$. Combination of Slutsky Lemma with the uniform integrability of $(M - M^{(n)})1_{\{A_n\}}$ yields the result.

Step 2. Show that, with $B_k := \{N_{k,n} = 2, \cap_{i \neq k} A_{i,n}\}$, the following are equal up to o(1) term:

$$\begin{aligned} &n\mathbb{E}(M - M^{(n)}; A_n^c), \quad \textstyle\sum_{k=0}^{n-1} n\mathbb{E}(M - M^{(n)}; A_{k,n}^c), \\ &\textstyle\sum_{k=0}^{n-1} n\mathbb{E}(M - M^{(n)}; N_{k,n} = 2), \quad \textstyle\sum_{k=0}^{n-1} n\mathbb{E}(M - M^{(n)}; B_k). \end{aligned}$$

This step is a rather tedious, but also, a pretty straightforward application of inclusion-exclusion principle. We only show the first equivalence, as the rest is similar. Note that $\mathbb{P}(A_{k,n}) = O(n^{-2})$ and that we have a very crude upper bound $M - M^{(n)} \leq \gamma + \sum_{k=1}^{N} |J_k|$, where N is the number of jumps of CP process and $J_1, J_2, \ldots$ are iid jumps. For $j < k < n$ we have

$$\begin{aligned} \mathbb{E}(M - M^{(n)}; A_{j,n}^c \cap A_{k,n}^c) &\leq \mathbb{P}(A_{j,n}^c \cap A_{k,n}^c)\mathbb{E}\Big(\gamma + \textstyle\sum_{k=1}^{N} |J_k| \,\Big|\, A_{j,n}^c \cap A_{k,n}^c\Big) \\ &= \mathbb{P}(A_{k,n})^2 \mathbb{E}\Big(\gamma + \textstyle\sum_{k=1}^{N} |J_k| \,\Big|\, N \geq 4\Big) \leq Cn^{-4}, \end{aligned}$$

where $C > 0$ does not depend on j, k, n. This implies that

$$\textstyle\sum_{0 \leq j < k < n} n\mathbb{E}(M - M^{(n)}; A_{j,n}^c \cap A_{k,n}^c) \to 0.$$

Step 3. Notice that when X is a Compound Poisson process then I has an alternative representation:

$$I = \lambda^2 \mathbb{E}\Big((J_1 + \underline{X}_U) \wedge (-J_2 - \overline{X}'_{1-U})\Big)^+,$$

where $\lambda = \Pi(\mathbb{R})$, U is uniformly distributed over $[0,1]$, X' is a statistical copy of X, random variables J_1, J_2 have the law $\Pi(\mathrm{d}x)/\lambda$, and U, X, X', J_1, J_2 are independent.

Step 4. Show that $\sum_{k=0}^{n-1} n\mathbb{E}(M - M^{(n)}; B_k) \to \frac{1}{2} I$. Working on the event B_k, let J_1, J_2 be the two jumps in time interval $\left(\frac{k}{n}, \frac{k+1}{n}\right)$ in the order of occurrence.

Moreover, let

$$L_t := X_t - \sup_{s\in[0,t]} X_s, \quad R_t := \sup_{t\in[s,1]} X_s - X_t$$

and notice that L_{t_1}, R_{t_2} are independent when $t_1 \le t_2$. We have

$$(M - M^{(n)})1_{\{B_k\}} \ge \Big(\big((J_1 + L_{k/n}) \wedge (-J_2 - R_{(k+1)/n})\big)^+ - |\gamma|/n\Big)\, 1_{\{\cap_{i\neq k} A_{i,n}\}}$$

and an analogous upper bound holds true, with $+|\gamma|/n$ instead of $-|\gamma|/n$. Now, we denote

$$\begin{aligned} G^{(n-)}(t) &:= \mathbb{E}\left(\big((J_1 + L_t) \wedge (-J_2 - R_t)\big)^+ - |\gamma|/n; A_n\right) \\ G^{(n+)}(t) &:= \mathbb{E}\big((J_1 + L_t) \wedge (-J_2 - R_t)\big)^+ + |\gamma|/n. \end{aligned}$$

L_t and R_t are stochastically non-increasing and non-decreasing respectively since $L_t \overset{\mathrm{d}}{=} \underline{X}_t$, $R_t \overset{\mathrm{d}}{=} \overline{X}_{1-t}$ (this holds true also on the event A_n) thus

$$G^{(n-)}((k+1)/n) \le (M - M^{(n)})1_{\{B_k\}} \le G^{(n+)}(k/n).$$

It is clear that $G^{(n\pm)}(t) \to G(t)$ point-wise, where $G(t) := \mathbb{E}\big((J_1 + \underline{X}_t) \wedge (-J_2 - \overline{X}'_{1-t})\big)^+$. Since $\mathbb{P}(N_{k,n} = 2) = \frac{\lambda^2}{2n^2} e^{-\lambda/n}$, we have

$$\begin{aligned} \sum_{k=1}^{n-1} n\mathbb{E}(M - M^{(n)}; B_k) &\le \sum_{k=0}^{n-1} n\mathbb{P}(N_{k,n} = 2) G^{(n+)}(k/n) \\ &= \tfrac{\lambda^2}{2} \sum_{k=0}^{n-1} \tfrac{1}{n} G^{(n+)}(k/n) \to \tfrac{\lambda^2}{2} \int_0^1 G(t)\mathrm{d}t = \tfrac{1}{2} I, \end{aligned}$$

where we used dominated convergence. Analogous reasoning leads to the same lower bound, which concludes the proof. □

C Proofs for Section 4

Proof of Lemma 3.6. Letting $\varphi(\theta) = \psi(\mathrm{i}\theta)$, we note that the condition (3.17) ensures the following bound on the characteristic function of X_t: $|e^{\varphi(\theta)t}| \le \exp(-ct|\theta|^\gamma)$ for some $c > 0$ and $|\theta| > 1$, see [79, Lem. 2.3]. By the inversion

formula we have

$$p(t,x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}x\theta+\varphi(\theta)t}\mathrm{d}\theta,$$

because the characteristic function $e^{\varphi(\theta)t}$ is integrable. Thus $p(t,x)$ is bounded for all $t > \delta, x \in \mathbb{R}$, and so we need to consider $t \in (0,\delta], x > \delta$ since the case $x < -\delta$ is analogous.

Assume for a moment that X has no jumps larger than 1 in absolute value, and so $\varphi(\theta)$ is smooth. From the Lévy-Khintchine formula we find that $|\varphi'(\theta)| \leq c_0 + c_1|\theta|$ and $|\varphi^{(k)}(\theta)| \leq c_k$ for $k \geq 2$ and some positive constants c_k; for this we differentiated under the integral with respect to $\Pi(\mathrm{d}x)$ and used the inequality $|e^{\mathrm{i}a} - 1| \leq |a|$. Integration by parts gives

$$\int_0^\infty e^{-\mathrm{i}x\theta+\varphi(\theta)t}\mathrm{d}\theta = \frac{1}{\mathrm{i}x} + \int_0^\infty \frac{1}{\mathrm{i}x}\varphi'(\theta)te^{-\mathrm{i}x\theta+\varphi(\theta)t}\mathrm{d}\theta,$$

and it would be sufficient to establish that

$$\int_1^\infty t(c_0 + c_1\theta)\exp(-c\theta^\gamma t)\mathrm{d}\theta$$

is bounded for all $t \in (0,\delta)$. This, however, is only true for $\gamma = 2$. Nevertheless, we may apply integration by parts k times to arrive at the bound:

$$\int_1^\infty A(t,\theta)\exp(-c\theta^\gamma t)\mathrm{d}\theta,$$

where $A(t,\theta)$ is a weighted sum of the terms $\theta^i t^j$ with $i < j$ and $i = j = k$; one may also use Faà di Bruno's formula here. Note that

$$\int_0^\infty \theta^i t^j \exp(-c\theta^\gamma t)\mathrm{d}\theta = t^{j-(i+1)/\gamma}\int_0^\infty \theta^i\exp(-c\theta^\gamma)\mathrm{d}\theta,$$

which is bounded for small t when $\gamma \geq (i+1)/j$. Since $\gamma > 1$ this inequality is always satisfied for the integers $i < j$, whereas for $i = j = k$ we get $\gamma \geq (k+1)/k$ and so we simply need to ensure that k is sufficiently large.

Suppose now that $X_t = \widehat{X}_t + Y_t$ is an independent sum, where Y is a Poisson process with jumps larger than 1 . The density of X_t is given by

$$p(t,x) = \int \mathbb{P}(Y_t \in \mathrm{d}z)\widehat{p}(t,x-z) \leq \mathbb{P}(Y_t = 0)\widehat{p}(t,x) + \mathbb{P}(Y_t \neq 0)\sup_x \widehat{p}(t,x),$$

where $\widehat{p}(t,x)$ is bounded on the set away from the origin. It is thus sufficient to show that the second term stays bounded as $t \downarrow 0$. But $\mathbb{P}(Y_t \neq 0)$ is of order t and $\sup_x \widehat{p}(t,x) = \mathcal{O}(t^{-1/\gamma})$ according to [79, Thm. 3.1] completing the proof.

Finally, suppose that (3.17) is satisfied with $\gamma < 1$ but for some $\gamma' \in (\gamma, 1)$, we have that

$$\lim_{\epsilon \to 0} \epsilon^{\gamma'-2} \int_{-\epsilon}^{\epsilon} x^2 \Pi(\mathrm{d}x) = 0$$

which according to [79, Thm. 3.1(b)] implies $\sup_x p(t,x) \geq c t^{-1/\gamma'}$ for t small enough. We may assume that for small enough t the supremum is achieved by $x \in [-\delta, \delta]$, because otherwise we have a contradiction. Now suppose that $\Pi(\mathrm{d}x)$ has a point mass at 1, so that with probability of order t there is one jump of size 1. But then $\sup_{x \in [1-\delta, 1+\delta]} p(t,x) \geq c_1 t^{1-1/\gamma'} \to \infty$ as $t \to 0$ showing that $p(t,x)$ explodes away from $x = 0$. □

Chapter 4

Simulation-Based Assessment of the Stationary Tail Distribution of a Stochastic Differential Equation

A commonly used approach for analyzing stochastic differential equations (SDEs) relies on performing Monte Carlo simulation with a discrete-time counterpart. In this chapter we study the impact of such a time-discretization when assessing the stationary tail distribution. For a family of semi-implicit Euler discretization schemes with time-step $h > 0$, we quantify the relative discretization error, as a function of h and the exceedance level x. By studying the existence of certain (polynomial and exponential) moments, using a sequence of prototypical examples, we demonstrate that this error may tend to 0 or ∞. The results show that the original shape of the tail can be heavily affected by the discretization.

The cases studied indicate that one has to be very careful when estimating the stationary tail distribution using Euler discretization schemes.

1 Introduction

Let $(X_t)_{t\geq 0}$ solve the stochastic differential equation (SDE)

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W_t \tag{4.1}$$

with an initial condition $X_0 \sim \xi$. The functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are called *drift* and *volatility* respectively, while ξ follows an arbitrary, tight (possibly degenerate) probability law concentrated on $\mathbb{R}$. Under some conditions on f and g, X_t converges to a stationary distribution as $t \to \infty$. Let μ_0 be the corresponding stationary (or *invariant, ergodic*) measure, that is, the unique probability measure such that $X_0 \sim \mu_0$ implies $X_t \sim \mu_0$ for all $t > 0$. In the following, we abbreviate $\bar{\mu}_0(x) := \mu_0((x, \infty))$.

We are interested in determining the *shape of the tail* of μ_0, i.e., the way $\bar{\mu}_0(x)$ decays to 0 as $x \to \infty$. Besides the one-dimensional case, no explicit expressions for $\bar{\mu}_0(x)$ are available, thus motivating the use of simulation-based methods. Ideally, one would sample a path of $(X_t)_{t\geq 0}$ (in continuous time, that is), and estimate $\bar{\mu}_0(x)$ by the fraction of time it spends above level x in a time interval $[0, T]$ (which, by the ergodic theorem, converges to $\bar{\mu}_0(x)$ as $T \to \infty$). It is evidently impossible to sample a continuous and infinitely long path of a process $(X_t)_{t\geq 0}$ on a computer, explaining the need for time-discretization and truncation. Discretization schemes are not exact and may intrinsically change the dynamics of the original continuous-time process defined through (4.1). As a consequence, the stationary measure pertaining to the discretized process will generally differ from μ_0 (or might not even exist!) A few relevant references on this topic are [84], [90], and [76].

In this chapter we study the effect of discretization on the shape of the tail of the stationary distribution. In order to illustrate the problem that might occur, we use the Ornstein-Uhlenbeck (OU) process as an example. Let $(X_t)_{t\geq 0}$ solve $\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$; it can be shown that $\mu_0 \sim N(0, 1)$. The forward Euler discretization scheme with time-step $h < 2$ can be shown to have invariant mea-

sure $\mu_h \sim N(0, (1-h/2)^{-1})$. Both distributions may look similar, but for any fixed h the ratio $\bar{\mu}_h(x)/\bar{\mu}_0(x)$ explodes as $x \to \infty$, entailing that in this example the forward Euler scheme creates a stationary distribution with tails *heavier* than those of the original continuous-time process. Whereas in the OU case μ_h and μ_0 belong to the same class of distributions (i.e., Gaussian), we will identify other examples in which they belong to different classes of distributions; we even find an instance in which $\bar{\mu}_0(x)$ has essentially the shape $\mathbb{E}(-x^2)$, whereas $\bar{\mu}_h(x)$ has power-law decay; see Section 4.4. Thus, an important message from our work is that one should be extremely careful when estimating the stationary tail distribution using time-discretization based simulation.

In earlier studies, it has been observed that the discretization influences the stationary distribution. Talay [92] showed that under regularity conditions, for a Milstein scheme with time-step $h > 0$, (i) there exists a stationary distribution μ_h and (ii) for a class of functions $r : \mathbb{R} \to \mathbb{R}$, $\left|\int r(x)\mu_h(\mathrm{d}x) - \int r(x)\mu(\mathrm{d}x)\right| \to 0$, as $h \downarrow 0$. Recently Abdulle et al. [1] extended this result to higher order schemes. These results are reassuring, as they show that at least in some sense μ_h is close to μ_0, but unfortunately they do not provide any insight into the level of resemblance between the respective tails.

Schurz [87] proposes a family of θ-implicit Euler schemes to discretize a system of linear SDEs. The author proves that in case of additive noise, the trapezoidal rule (i.e., the semi-implicit scheme with implicitness parameter θ equal to $\frac{1}{2}$) gives the same stationary distribution as the original system, independent of the choice of h. This observation has motivated us to also consider a family of semi-implicit Euler schemes in this chapter. Mattingly et al. [76] provide general conditions under which the ergodicity properties of the original continuous-time process carry over to various discretization schemes. In this chapter we discuss one of those, viz. an implicit *split-step* Euler scheme.

To study $\bar{\mu}_h(x)$ under various discretization schemes, we use tools from the theory of random iterated functions, and results on the existence of moments of stationary distributions of Markov chains.

The chapter is organized as follows. In Section 2 we introduce the discretization schemes that are used throughout the chapter, We briefly discuss the existence and uniqueness of the stationary distribution of the continuous-

time system (4.1) and its discretized version, and we specify what we mean by the 'shape of the tail'. In Section 3 we introduce our key tools; most notably, we recall a theorem by Goldie [55] on random iterated functions, which we use to describe the behavior of the tail of a stationary distribution for linear systems of SDEs under the discretization, and a theorem based on [93], which gives conditions for existence and nonexistence of moments. In Section 4 we assess (analytically and numerically) the ratio $\gamma_h(x) := \bar{\mu}_h(x)/\bar{\mu}_0(x)$, using four illustrative examples and various discretization schemes.

2 Preliminaries

In this section we first introduce discretization schemes. Then we present results from the literature on the existence and uniqueness of the invariant measure for SDEs and corresponding discretized versions. Finally, we introduce the notion of the 'shape of the tail'.

2.1 Discretization Schemes for SDEs

Let $(X_t)_{t\geq 0}$ be driven by (4.1). In order to simulate X_t on a computer we use discretization schemes; for a survey on discretization methods for SDEs we refer to [69]. Let $(W_t)_{t\geq 0}$ be a standard Wiener process and $\Delta W_n := W_{n+1} - W_n$. As motivated in the introduction, we focus on semi-implicit discretization schemes (parametrized by $\theta \in [0,1]$) and split-step schemes. In particular we distinguish two schemes:

- The (standard) θ-implicit Euler-Maruyama scheme

$$X_{n+1} := X_n + \big(\theta f(X_{n+1}) + (1-\theta) f(X_n)\big)h + g(X_n)\sqrt{h}\,\Delta W_n. \tag{4.2}$$

- The split-step θ-implicit Euler-Maruyama scheme

$$\begin{cases} X^\star_{n+1} := X_n + \big(\theta f(X^\star_{n+1}) + (1-\theta) f(X_n)\big)h, \\ X_{n+1} := X^\star_{n+1} + g(X_n)\sqrt{h}\,\Delta W_n. \end{cases} \tag{4.3}$$

We denote by μ_h the stationary measure of the discretization scheme considered. When $\theta = 0$ and $\theta = 1$ we recover the fully explicit and fully implicit Euler

schemes, respectively. It is assumed that the schemes are well-defined, in that there exists a unique solution X_{n+1} to the implicit equations (4.2)-(4.3). In this framework, the dynamics of the discretized system can be described by, with $X_0 \sim \xi$,

$$X_{n+1} = H(X_n, \Delta W_n) \tag{4.4}$$

for some deterministic function $H : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Later in the chapter on one occasion we also study a Milstein scheme, but we do not introduce it here, to keep the presentation as clear as possible.

The main difference between a standard scheme (4.2) and a split-step scheme (4.3) is the moment the 'noise part' is added: in the former case one first adds the noise and then solves the implicit equation for X_{n+1}, while in the latter case this order is reversed. The main advantage of using the split-step scheme is the simplicity, in the sense that it can be rewritten as, for functions $a : \mathbb{R} \to \mathbb{R}$ and $b : \mathbb{R} \to \mathbb{R}$,

$$X_{n+1} = a(X_n) + b(X_n)\Delta W_n. \tag{4.5}$$

Later in the chapter we will observe that the scheme (4.5) is simpler to study than its standard counterpart (4.2). In particular, for split-step schemes it is easier to assess the existence of the stationary measure; see e.g. [58, Section 3] for ergodicity conditions of Markov chains with a Gaussian transition kernel. Our examples in Section 4 will reveal that (4.2) and (4.3) may lead to entirely different stationary tail distributions, which may also be very different from the tail behavior corresponding to μ_0.

2.2 Existence and Uniqueness of Stationary Distributions

In continuous time when the invariant measure μ_0 exists and has a density π_0, it is given by

$$\pi_0(x) \propto \frac{1}{g^2(x)} \mathbb{E}\left\{ 2 \int^x \frac{f(y)}{g^2(y)} \, \mathrm{d}y \right\}; \tag{4.6}$$

see e.g. [45, p. 210]. Hence it is straightforward to derive the asymptotic behavior of $\bar{\mu}_0(x)$. All SDEs considered in this chapter have a stationary density proportional to the right-hand side of (4.6).

We now move to discrete time. We say that the Markov chain driven by

(4.4) has a stationary distribution with law μ iff $X_0 \sim \mu$ implies $X_1 \sim \mu$ (which implies $X_n \sim \mu$ for all $n \in \mathbb{N}$). Contrary to the continuous-time case, there are no general formulas available for such distribution μ. In this chapter, we assess the ergodicity of discrete-time Markov chains using Theorem 4.2 (introduced later, in Section 3), which is based on results from Tweedie [93] and Meyn and Tweedie [77].

2.3 Shape of the Tail

In this chapter by the 'shape of the tail' of a random variable X we mean the rate of convergence of the complementary cumulative distribution function $\mathbb{P}(X > x)$ in x as $x \to \infty$. For example we might have *polynomial decay* (i.e., $\mathbb{P}(X > x) \sim Cx^{-\alpha}$ for some constants $C, \alpha > 0$), or *Weibullian* decay (i.e., $\mathbb{P}(X > x) \sim C\mathbb{E}(-s\,x^p)$ for some $C, s, p > 0$).

The tails of two random variables can be compared through their moments, as follows. Let X, Y be random variables and $r : \mathbb{R} \to \mathbb{R}_+$ such that $r(x) = 0$ for $x < 0$. If $\int r(x)\mathbb{P}(X \in \mathrm{d}x) < \infty$ and $\int r(y)\mathbb{P}(Y \in \mathrm{d}y) = \infty$, then necessarily $\mathbb{P}(X > x)/\mathbb{P}(Y > x) \to 0$ as $x \to \infty$ (given that the limit exists); we say that X has a lighter tail than Y.

If $\mathbb{P}(X > x) \sim Cx^{-\alpha}$, then α can be identified from the fractional moments $\mathbb{E}(X^\alpha; X > 0)$. More specifically,

$$\alpha = \sup\Big\{\beta > 0 : \int_0^\infty x^\beta\, \mathbb{P}(X \in \mathrm{d}x) < \infty\Big\} \tag{4.7}$$

Conversely, if α in (4.7) is finite, then $\mathbb{P}(X > x) \sim L(x)\, x^{-\alpha}$ for some 'sub-polynomial' function $L(\cdot)$.

If the random variable X admits all moments, so that $\alpha = \infty$ in (4.7), then $\mathbb{P}(X > x)$ decays to 0 faster than $x^{-\beta}$ for any $\beta > 0$. In this case, the tail can be described more precisely, via the existence of exponential moments such as $\mathbb{E}\big(\mathbb{E}(s\,X^p); X > 0\big)$. In particular, if $\mathbb{P}(X > x) \sim C\mathbb{E}(-sx^p)$, then p and s obey

$$\begin{aligned} p &:= \sup\Big\{q > 0 : \exists\, t > 0 : \int_0^\infty \mathrm{e}^{t\,x^q}\, \mathbb{P}(X \in \mathrm{d}x) < \infty\Big\}, \\ s &:= \sup\Big\{t > 0 : \int_0^\infty \mathrm{e}^{t\,x^p}\, \mathbb{P}(X \in \mathrm{d}x) < \infty\Big\}. \end{aligned} \tag{4.8}$$

Conversely, if the p, s in (4.8) are finite, then $\mathbb{P}(X > x) \sim L(x)\mathrm{e}^{-s\,x^p}$ for a 'sub-Weibullian' function $L(\cdot)$.

For example, an exponentially distributed random variable with mean λ^{-1} has $(p, s) = (1, \lambda)$, and a normally distributed random variable with variance σ^2 has $(p, s) = (2, (2\sigma^2)^{-1})$. Both distributions are light-tailed, but the Gaussian is clearly much lighter, as it corresponds to a higher p. One could say that distributions with the same p but different s belong to the same class (but the one with the higher s is lighter).

3 Tools for the study of the tails

The behavior of $\bar{\mu}_0(x)$ (for x large) follows from the density (4.6). Finding the tail behavior of μ_h is less straightforward. In this section we present two tools.

3.1 Random Iterated Functions

The following result, from [55, Theorem 4.1], describes the stationary tail distribution $\bar{\mu}$ corresponding to the stochastic recursion $X_{n+1} = M_n X_n + Q_n$, where $(Q_1, M_1), (Q_2, M_2), \ldots$ are i.i.d. It will be useful for analysing the stationary laws corresponding to discretizations of linear SDEs; see Section 4.2.

Theorem 4.1. *Let κ be such that $\mathbb{E}|M_1|^\kappa = 1$, $\mathbb{E}|M_1|^\kappa \log^+ |M_1| < \infty$ and $\mathbb{E}|Q_1|^\kappa < \infty$. Assume that the law of $\log |M_1|$ given $M_1 \neq 0$ is non-arithmetic. Then the stochastic recursion $X_{n+1} = M_n X_n + Q_n$ has a unique invariant measure μ satisfying $\bar{\mu}(x) \sim Cx^{-\kappa}$, for some $C > 0$, as $x \to \infty$.*

3.2 Existence and Non-Existence of Moments

In Section 2.3 we discussed the connection between existence and nonexistence of moments and the behavior of the tail. Theorem 4.2 presents sufficient conditions for existence (and nonexistence) of moments of the stationary measure of a Markov chain. Let $P(x, \mathrm{d}y)$ be a transition kernel of a Markov chain and $r : \mathbb{R} \to \mathbb{R}$ be such that $r \geq 1$ and $\int r(y)P(x, \mathrm{d}y) < \infty$. Define the following parameters associated with the function r:

$$L(r) = \limsup_{|x|\to\infty} \frac{\int r(y)P(x, \mathrm{d}y)}{r(x)} \quad \text{and} \quad L^+(r) := \lim_{x\to\infty} \frac{\int_0^\infty r(y)P(x, \mathrm{d}y)}{r(x)}. \quad (4.9)$$

Note that while L is well defined, $L^+(r)$ does not necessarily exist. Below, we use the concepts of aperiodicity, irreducibility and small sets as introduced in Appendix A.

Theorem 4.2. *Suppose that $(X_n)_{n\in\mathbb{N}}$ is an aperiodic, μ^{Leb}-irreducible Markov chain with a transition kernel $P(x, \mathrm{d}y)$. Suppose that all intervals $[-M, M]$ for $M > 0$ are small. Let $r : \mathbb{R} \to \mathbb{R}$ be such that $r \geq 1$, $\int r(y)P(x, \mathrm{d}y) < \infty$ and $L(r), L^+(r)$ are defined in (4.9). Then*

(i) If $L(r) < 1$ and $\sup_{x\in[-M,M]} \int r(y)P(x, \mathrm{d}y) < \infty$ for all $M > 0$ then there exists a unique stationary probability measure μ with $\int r(x)\mu(\mathrm{d}x) < \infty$.

(ii) If there exists a unique stationary probability measure μ, and r is non-decreasing for $x \in (0, \infty)$ then $L^+(r) > 2$ implies $\int_0^\infty r(x)\mu(\mathrm{d}x) = \infty$.

Proof. Notice that $L(r) < 1$ implies there exists $\delta, b > 0$ such that for all $x \in \mathbb{R}$,

$$\int r(y)P(x, \mathrm{d}y) < (1 - \delta)r(x) + b1_{\{x\in[-M,M]\}}\,.$$

Thus, part (i) follows from [77, Thm 14.0.1]. Part (ii) can be proven as follows. $L^+(r) > 2$ implies there exists $\delta, x_0 > 0$ such that $\int_0^\infty r(y)P(x, \mathrm{d}y) > (2+\delta)r(x)$ for all $x \geq x_0$. Thus,

$$r(x) + \int_x^\infty r(y)P(x, \mathrm{d}y) \geq \int_0^\infty r(y)P(x, \mathrm{d}y), \tag{4.10}$$

for all $x \geq x_0$, since r is non-decreasing on $(0, \infty)$ and $\int P(x, \mathrm{d}y) = 1$. Combining these elements we find

$$\int_{x_0}^\infty r(y)P(x, \mathrm{d}y) \geq \int_x^\infty r(y)P(x, \mathrm{d}y) > (1 + \delta)r(x),$$

for all $x \geq x_0$. Then, with the second equality due to 'Fubini',

$$\begin{aligned}\int_{x_0}^\infty r(x)\mu(\mathrm{d}x) &= \int_{x_0}^\infty r(x)\left(\int \mu(\mathrm{d}y)P(y, \mathrm{d}x)\right) = \int\left(\int_{x_0}^\infty r(x)P(y, \mathrm{d}x)\right)\mu(\mathrm{d}y)\\ &\geq \int_{x_0}^\infty \left(\int_{x_0}^\infty r(x)P(y, \mathrm{d}x)\right)\mu(\mathrm{d}y) > (1 + \delta)\int_{x_0}^\infty r(y)\mu(\mathrm{d}y).\end{aligned}$$

This entails $\int_0^\infty r(x)\mu(\mathrm{d}x) = \infty$. □

Remark 4.1. *Additionally, there is a very simple condition which implies nonexistence of a moment. If there exists a set A such that $\mu(A) > 0$ and $\int r(x)P(x, \mathrm{d}y) = \infty$ for all $x \in A$ then necessarily $\int r(x)\mu(\mathrm{d}x) = \infty$.*

In view of the considerations in Section 2.3, it is particularly convenient to work with Thm. 4.2 for classes of functions such as $r_\alpha(x) = 1 + |x|^\alpha$ and $r_{p,s}(x) = \mathbb{E}(s|x|^p)$. Note however, that there is a 'moment gap', that is when $L(r) > 1$ and $L^+(r) < 2$, it cannot be inferred from Theorem 4.2 whether the moment $\int_0^\infty r(x)\mu(\mathrm{d}x)$ exists or not. The gap can be decreased by considering bounds tighter than (4.10) based on the transition kernel; for the purposes of this chapter, considering the condition $L^+(r) < 2$ is sufficient.

4 Assessment of the tail in benchmark models

In this section we study four prototypical SDEs, comparing the tails of the continuous-time processes with the tails of their discrete-time counterparts. Thus, we compare the tails $\bar{\mu}_0(x)$ and $\bar{\mu}_h(x)$ for large x. We do so by studying the ratio

$$\gamma_h(x) := \bar{\mu}_h(x)/\bar{\mu}_0(x) \tag{4.11}$$

for large x, both analytically (using the tools presented in Section 3) and numerically. In particular, if $\gamma_h(x) \to 0$ as $x \to \infty$ then μ_h has a lighter tail than μ_0, whereas if $\gamma_h(x) \to \infty$ as $x \to \infty$ then μ_h has a heavier tail than μ_0.

The main message from this section is that $\gamma_h(x)$ can differ substantially from 1. In addition, the $\bar{\mu}_h(x)$ resulting from different discretization schemes can also be orders of magnitude different. This leads to the general conclusion that, when aiming at estimating the SDE's tail distribution, one has to be very careful with using discretization schemes. We focus in this section on one-dimensional SDEs, as for these μ_0 is available in closed-form, but obviously the above warning carries over to multi-dimensional SDEs.

The numerical results in Sections 4.2-4.4 concern the estimation of $\bar{\mu}_h(x)$ for various values of x and h. In all sections, the estimates were determined by a crude Monte Carlo method. The Markov chain has been divided in blocks of 10^4 time steps (of size h) separated by periods of 10^4 time steps. These blocks of 10^4 time steps are assumed to be independent, and based on that assumption we

derive the sample error. The estimation procedure is stopped when the sample relative error falls below 5%.

4.1 Ornstein-Uhlenbeck Process

It was already observed in e.g. [92] that the stationary distribution of an OU process is affected by discretization (motivating the use of a second-order scheme that has a stationary distribution closer to the continuous-time one). Schurz [87] proposes discretizing multidimensional linear SDEs (covering the OU case) using the implicit Euler methods; he shows that the trapezoidal rule (semi-implicit scheme with $\theta = \frac{1}{2}$) is the only discretization scheme in that family that preserves the correct stationary distribution.

The OU process is defined by the following SDE: for $\vartheta, \sigma > 0$,

$$\mathrm{d}X_t = -\vartheta X_t \,\mathrm{d}t + \sigma \,\mathrm{d}W_t. \tag{4.12}$$

It is well-known that (4.12) admits a Gaussian stationary distribution with mean zero and variance $\Sigma := \sigma^2/(2\vartheta)$. In this particular case, due to the drift function $f(x) = -\vartheta x$ and volatility $g(x) \equiv \sigma$ being both linear, both discretization schemes (4.2)-(4.3) yield the same numerical scheme:

$$X_{n+1} = X_n \,(1 - \vartheta \Gamma \, h) + \sigma \Gamma \sqrt{h} \, \Delta W_n, \tag{4.13}$$

with $\Gamma := (1 + \vartheta\theta h)^{-1}$. This is an AR(1) process, which admits a stationary distribution iff $|1 - \vartheta \Gamma h| < 1$, that is when $h < 2/(\vartheta(1 - 2\theta))$; see e.g. van der Vaart [94]. More specifically, this stationary distribution is Gaussian with mean zero and variance $\Sigma(1 + \vartheta h(\theta - \frac{1}{2}))^{-1}$. This shows that by taking $\theta = \frac{1}{2}$, the invariant measure of the discretized chain is the same as the invariant measure of the original, continuous-time one, cf. the findings of Schurz [87].

We take the analysis a bit further to assess what errors one makes for the OU case with other rules than the trapezoid rule. A first, naïve, choice could be the explicit Euler scheme ($\theta = 0$). As seen above, $\mu_0 \sim N(0, \Sigma)$ and $\mu_h \sim N(0, \Sigma/(1 - \frac{1}{2}\vartheta h))$. Now consider the ratio $\gamma_h(x)$. Let $\Phi(\cdot)$ and $\varphi(\cdot)$ denote a standard normal cdf and pdf respectively. For large x, $\Phi(-x) \sim \varphi(x)/x$, so that

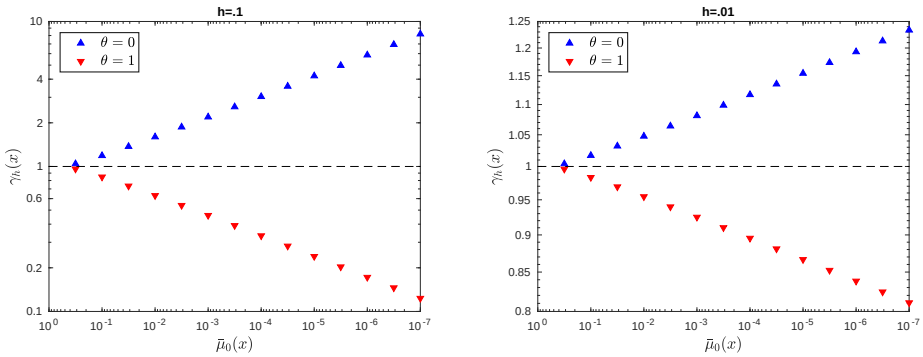


Figure 4.1: Ornstein-Uhlenbeck process with $\vartheta = 3, \sigma = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.1$ (left) and $h = 0.01$ (right) for the two extreme values of the 'implicitness parameter' θ.

for large x

$$\gamma_h(x) = \frac{\Phi\Big(-\dfrac{x}{\sqrt{\Sigma}}\sqrt{1-\frac{1}{2}\vartheta h}\Big)}{\Phi\Big(-\dfrac{x}{\sqrt{\Sigma}}\Big)} \sim \sqrt{\frac{2}{2-\vartheta h}}\,\mathbb{E}\left\{-\frac{x^2\vartheta}{2\sigma^2}\,(2-\vartheta h-2)\right\}$$
$$= \sqrt{\frac{2}{2-\vartheta h}}\cdot\mathbb{E}\left\{\frac{\vartheta^2\,x^2 h}{2\sigma^2}\right\}.$$

This means that $\gamma_h(x) \to \infty$ for any given $h > 0$, showing that the tail of μ_h is heavier than the tail of μ_0. Similar calculations show that in the fully implicit case ($\theta = 1$) the tail of μ_h is lighter than the one of μ_0. In general, increasing the 'implicitness parameter' θ makes the tail of μ_h lighter; cf. Fig. 4.1.

A pragmatic rule that can be used to control $\gamma_h(x)$ is to decrease the step h proportionally to x^2. More, concretely, let h satisfy $x^2 h \to \varepsilon > 0$, as x grows large. Then $\gamma_h(x) \approx 1 + \varepsilon\,\vartheta^2/(2\sigma^2)$. So in order to keep the ratio $\gamma_h(x)$ smaller than some given $1+\delta$, one could choose h in the following way:

$$h = x^{-2}\,\left(2\,\vartheta^{-2}\,\sigma^2\log(1+\delta)\right).$$

Conclusions: This example shows that, although μ_h still belongs to the same class of distributions as μ_0 (Weibull-tail with shape parameter 2), the ratio $\gamma_h(x)$ can deviate substantially from 1 (in fact, $\gamma_h(x)$ may even go to 0 or ∞, depending on the choice of the implicitness parameter). This effect can be

mitigated by choosing the step-size h sufficiently small. The tail of μ_h becomes lighter when increasing θ.

4.2 Linear Drift and Linear Volatility

We now consider the process $(X_t)_{t\geq 0}$ that solves the SDE

$$\mathrm{d}X_t = \vartheta(m - X_t)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \tag{4.14}$$

where $\sigma, \vartheta > 0$ and $m > 0$ (observe that if m would have been 0, then the stationary measure μ_0 is entirely concentrated at 0). Using (4.6) we see that μ_0 corresponds to an *Inverse-Gamma* distribution with density

$$\pi_0(x) = \beta^{\alpha_0}(\Gamma(\alpha_0))^{-1}\, x^{-(1+\alpha_0)}\, \mathrm{e}^{-\beta/x}, \quad x > 0,$$

with shape parameter $\alpha_0 = 1 + 2\vartheta/\sigma^2$ and rate parameter $\beta = 2\vartheta m/\sigma^2$. As a consequence, for large x, $\pi_0(x) \sim (\beta^{\alpha_0}/\Gamma(\alpha_0))x^{-(1+\alpha_0)}$ and hence $\bar{\mu}_0(x) \sim C_0 x^{-\alpha_0}$ for some $C_0 > 0$.

Now let us focus on the stationary measure μ_h corresponding to the various discretization schemes. Both semi-implicit Euler discretization schemes (4.2)-(4.3) are equivalent in this case. They are given by

$$X_{n+1} = \frac{\vartheta m h}{1 + \vartheta\theta h} + X_n\Big(1 - \frac{\vartheta h}{1 + \vartheta\theta h} + \frac{\sigma\sqrt{h}}{1 + \vartheta\theta h}\Delta W_n\Big) \tag{4.15}$$

Notice that this is exactly the setting of Thm. 4.1, so it can be applied to derive the tail behavior of μ_h. The following proposition, which is an immediate consequence of Thm. 4.1, states that $\bar{\mu}_h(x)$ decays polynomially. The corresponding rate α_h (which differs from α_0) solves

$$\mathbb{E}\big|1 - \frac{\vartheta h}{1 + \vartheta\theta h} + \frac{\sigma\sqrt{h}}{1 + \vartheta\theta h}\Delta W_1\big|^{\alpha_h} = 1.$$

Proposition 4.1. *The family of Markov Chains evolving according to* (4.15) *admits a stationary distribution μ_h for all $h > 0$ and for each fixed h, which satisfies $\bar{\mu}_h(x) \sim C_h x^{-\alpha_h}$, as $x \to \infty$, for some constant $C_h > 0$.*

We now analyze α_h as $h \downarrow 0$, again using Thm. 4.1. Using Taylor expansions

and taking the expectation (recalling that $\Delta W_1 \sim N(0,1)$), we obtain

$$\mathbb{E}\Big|1 - \frac{\vartheta h}{1+\vartheta\theta h} + \frac{\sigma\sqrt{h}}{1+\vartheta\theta h}\,\Delta W_1\Big|^{\alpha} = 1 + \Big(-\alpha\vartheta + \binom{\alpha}{2}\sigma^2\Big)h + \Big(\alpha\theta\vartheta^2 - 3\binom{\alpha}{3}\sigma^2\vartheta + 3\binom{\alpha}{4} + \binom{\alpha}{2}(\vartheta^2 - 2\theta\sigma^2\vartheta x^2)\Big)h^2 + o(h^2),$$

which is to be equated to 1. Now it can be verified that α_h has the following first-order Taylor expansion around $h = 0$: with α_0 the rate corresponding to $\bar{\mu}_0(x)$,

$$\alpha_h = \alpha_0 + \big(\vartheta^2(1+2\theta)\sigma^{-2} - \vartheta\big)\,h + o(h). \tag{4.16}$$

Remark 4.2. *An expansion similar to (4.16) can be obtained for other discretization schemes as well, for instance for the semi-implicit Milstein scheme given by*

$$\begin{aligned} X_{n+1} = X_n &+ \Big(\theta f(X_{n+1}) + (1-\theta)f(X_n)\Big)h \\ &+ \sigma g(X_n)\sqrt{h}\,\Delta W_n + \frac{1}{2}g(X_n)g'(X_n)h((\Delta W_n)^2 - 1). \end{aligned} \tag{4.17}$$

This scheme takes the explicit form

$$X_{n+1} = \tfrac{\vartheta\mu h}{1+\vartheta\theta h} + X_n\left(1 - \tfrac{\vartheta+\frac{1}{2}\sigma^2}{1+\vartheta\theta h}\,h + \tfrac{\sigma}{1+\vartheta\theta h}\,\sqrt{h}\,\Delta W_n + \tfrac{\frac{1}{2}\sigma^2}{1+\vartheta\theta h}\,h(\Delta W_n)^2\right)$$

Without presenting the computations, we claim that the Milstein-discretized stationary distribution is heavy tailed: $\bar{\mu}_h(x) \sim C_h x^{-\alpha_h}$. The power α_h has, as $h \downarrow 0$, the first-order Taylor expansion

$$\alpha_h = \alpha_0 + \big(\vartheta^2(2\theta - 3)\sigma^{-2}\big)\,h + o(h)\,.$$

The reasoning is as in the proof of Prop. 4.1 and the computation of α_h for the schemes (4.2)-(4.3). ◇

The expression (4.16) for α_h for $\theta = 0$ reveals that if $\vartheta/\sigma^2 > 1$ then the tail of μ_h is lighter than the tail of μ_0, but vice versa if $\vartheta/\sigma^2 < 1$ — a perhaps surprising result for an explicit Euler scheme. The expansion of α_h in (4.16) advocates choosing $\theta \in [0,1]$ such that the absolute value of $(\vartheta(1+2\theta)/\sigma^2 - 1)$

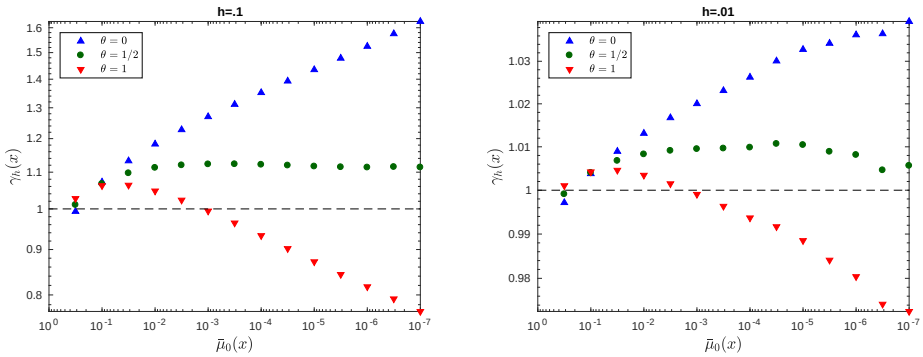


Figure 4.2: Linear drift and linear volatility process with $\vartheta = 1$, $\sigma = \sqrt{2}$, $m = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.1$ (left) and $h = 0.01$ (right) for $\theta \in \{0, \frac{1}{2}, 1\}$. From (4.16), the tail of μ_h should be heavier than the tail of μ_0 if $\theta = 0$, and lighter if $\theta = 1$. For $\theta = \frac{1}{2}$ the tails are more similar, as $\alpha_h = \alpha_0 + o(h)$.

is minimized. See Fig. 4.2 for illustrations.

Conclusions: In this example, the tail of μ_h and μ_0 are in the same class (power-law decay), and again the ratio $\gamma_h(x)$ can differ substantially from 1. Choosing $\theta = 0$ now does not necessarily lead to the situation of the tail of μ_h being heavier than the tail of μ_0 (cf. the OU process in Section 4.1). The tail of μ_h becomes lighter when increasing θ (also in the Milstein case).

4.3 Cubic Drift and Constant Volatility

We now consider a stochastic process with nonlinear drift, more specifically $(X_t)_{t\geq 0}$ solving

$$\mathrm{d}X_t = -\vartheta X_t^3 \,\mathrm{d}t + \sigma \,\mathrm{d}W_t \tag{4.18}$$

where $\vartheta, \sigma > 0$. From the formula for the density of the invariant measure (4.6), we see that μ_0 has a centered *Generalized Normal Distribution* with the density

$$\pi_0(x) = \beta \left(2\alpha\Gamma(1/\beta)\right)^{-1} \mathrm{e}^{-(|x|/\alpha)^\beta} \tag{4.19}$$

with scale parameter $\alpha = \left(2\sigma^2/\vartheta\right)^{1/4}$ and shape parameter $\beta = 4$. It has been shown in [76, Lemma 6.3] that the explicit Euler scheme (4.2) of the process

driven by (4.18) is transient, i.e., does not converge to a stationary distribution. After some calculations we can show that the θ-implicit Euler scheme (4.2) reads, for $\theta \in (0,1]$,

$$X_{n+1} := a\big(F_\theta(X_n) + \sigma\sqrt{h}\,\Delta W_n\big) \tag{4.20}$$

whereas the split-step θ-implicit Euler scheme (4.3) is given by, again for $\theta \in (0,1]$,

$$X_{n+1} := a\big(F_\theta(X_n)\big) + \sigma\sqrt{h}\,\Delta W_n, \tag{4.21}$$

where

$$a(x) := c^{-1}\sinh\big(\operatorname{arcsinh}\big(3cx\big)/3\big),\ F_\theta(x) := x - \vartheta(1-\theta)x^3h,\ c := \sqrt{3\vartheta\theta h}/2 \tag{4.22}$$

To study the ergodicity of the Markov chains driven by (4.20) and (4.21), it is useful to derive the asymptotics of $a(\cdot)$. The function $a(\cdot)$ is odd, satisfies $|a(x)| \le x^{1/3}(\vartheta\theta h)^{-1/3}$ for all $x \in \mathbb{R}$, and in addition $a(x)x^{-1/3} = (\vartheta\theta h)^{-1/3}$, as $|x| \to \infty$. From these properties, it follows that

$$a(F_\theta(x)) \sim \begin{cases} -x\,((1-\theta)/\theta)^{1/3}, & \theta \in (0,1) \\ x^{1/3}\,(\vartheta h)^{-1/3}, & \theta = 1 \end{cases} \tag{4.23}$$

as $x \to \infty$. Both schemes are ergodic for $\theta > \frac{1}{2}$. Below we give a shortened version of a proof of ergodicity, existence, and nonexistence of moments of the stationary measure; the interested reader is referred to Appendix B for the full proof.

For the split-step scheme, ergodicity follows from [58, Section 3] and the asymptotics of the function $a(\cdot)$ listed earlier. To show the ergodicity of the standard Euler scheme one can use Thm. 4.2(i). We note that the chain (4.20) satisfies the conditions of irreducibility, small sets and aperiodicity as stated in Theorem 4.2. Let $r_{p,s}(x) := \mathbb{E}(s|x|^p)$ and denote $Z \sim N(0,1)$. We verify the remaining assumptions of Thm. 4.2(i). Based on the properties of the function $a(\cdot)$, it is straightforward to verify that when $\theta = 1$,

$$\int r_{p,s}(y)P(x,\mathrm{d}y) \le \mathbb{E}\Big(s(\vartheta h)^{-p/3}\cdot\Big|x+\sigma\sqrt{h}Z\Big|^{p/3}\Big). \tag{4.24}$$

Now let $p = 6$. When $s < \vartheta^2 h/(2\sigma^2)$, the expression on the rhs of (4.24) is finite and there exist constants $A, B > 0$ such that it is equal to $A\mathbb{E}(Bx^2)$. Thus, $L(r_{p,s}) = 0$, where L is defined in (4.9). Lastly, as

$$\sup_{x\in[-M,M]} \int r_{p,s}(y)P(x, \mathrm{d}y) \leq A\mathbb{E}(BM^2) < \infty,$$

we conclude that when $\theta = 1$, the standard Euler scheme is ergodic with stationary measure μ_h and moreover, $r_{p,s}(x)\mu_h(\mathrm{d}x) < \infty$ for $p = 6$ and $s < \vartheta^2 h/(2\sigma^2)$. Proving ergodicity in case $\theta \in (1/2, 1)$ can be dealt with similarly using the test function $r(x) = 1 + x^6$.

Our next objective is to compare the tails of μ_h and μ_0 for a fully implicit scheme ($\theta = 1$). From (4.19), μ_0 has Weibullian decay with $(p, s) = (4, \vartheta/(2\sigma^2))$. For the standard Euler scheme μ_h is characterized by $(p, s) = (6, \vartheta^2 h/(2\sigma^2))$. This follows from the fact that $\int_0^\infty r_{6,s}(y)P(x, \mathrm{d}y) = \infty$ for $s < \vartheta^2 h/(2\sigma^2)$ (cf. Remark 4.1). The latter can be established using asymptotics of the function $a(\cdot)$, similar to the way we established the existence of moments. For the split-step scheme μ_h has parameters $(p, s) = (2, 1/(2\sigma^2 h))$, as can be shown in an analogous fashion.

The implication is that the standard Euler scheme gives a distribution in a lighter class than μ_0. It becomes heavier as $h \downarrow 0$ (since $s \downarrow 0$), but remains in the class with $p = 6$ (compared to $p = 4$ for μ_0). By contrast, the split-step scheme gives a distribution in a heavier class (viz. with $p = 2$), but its tail becomes lighter as $h \downarrow 0$. This behavior is reflected in our simulation results, see Fig. 4.3. We observe that the ratio $\gamma_h(x)$ goes to 0 as $x \to \infty$ for the standard Euler scheme, whereas $\gamma_h(x)$ explodes for the split-step scheme.

Conclusions: This example shows that μ_0 can be in a different class of distributions than μ_h. In addition, it strongly depends on the discretization scheme in which class μ_h is.

4.4 Cubic Drift and Linear Volatility

In our last example $(X_t)_{t\geq 0}$ solves

$$\mathrm{d}X_t = \vartheta(m - X_t^3)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t, \tag{4.25}$$

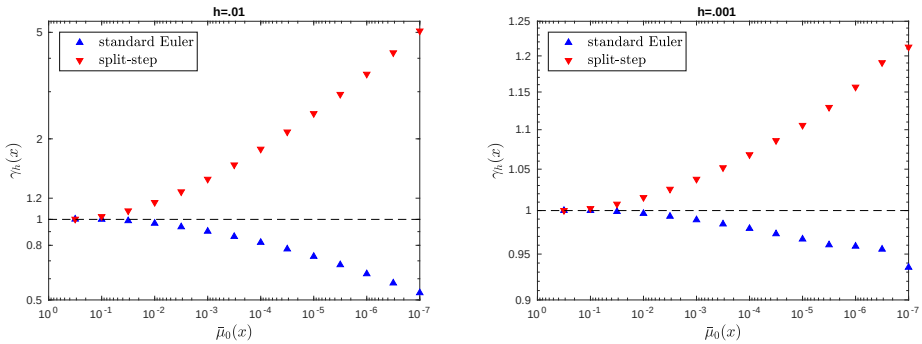


Figure 4.3: Cubic drift and constant volatility process with $\vartheta = 1, \sigma = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.01$ (left) and $h = 0.001$ (right) for fully implicit ($\theta = 1$) standard and split-step Euler schemes.

where $\sigma, \vartheta > 0$ and $m > 0$. Like in the previous examples, we can find the density of μ_0:

$$\pi_0(x) \propto x^{-2}\mathbb{E}\Big\{ -\frac{\vartheta}{\sigma^2}\Big(\frac{2m}{x} + x^2\Big)\Big\}, \quad x > 0 \tag{4.26}$$

Let $a(\cdot)$ and $F_\theta(\cdot)$ be as defined in (4.22). Similar to the example from the previous section, for $\theta \in (0, 1]$ the θ-implicit Euler scheme and the θ-implicit split-step scheme evolve according to

$$X_{n+1} := a\big(\vartheta m h + F_\theta(X_n) + \sigma X_n \sqrt{h}\, \Delta W_n\big), \tag{4.27}$$

$$X_{n+1} := a\big(\vartheta m h + F_\theta(X_n)\big) + \sigma X_n \sqrt{h}\, \Delta W_n, \tag{4.28}$$

respectively. Both schemes are ergodic for $\theta > 1/2$. Below we give a shortened version of a proof of ergodicity, existence, and nonexistence of moments of the stationary measure. In particular, we do not prove the regularity conditions (i.e. irreducibility, small sets, aperiodicity) of Thm. 4.2 here. Furthermore we focus on the fully implicit scheme $\theta = 1$; the case $\theta \in (1/2, 1)$ can be dealt with similarly. The interested reader is referred to Appendix C for the full proof.

Standard Euler case. We show, using Thm. 4.2(i), that the fully implicit standard Euler scheme has a stationary measure with Weibullian decay with

$p \in [3,6]$. Let $r_{p,s}(x) := \mathbb{E}(s|x|^p)$. We have

$$\int r_{p,s}(y)P(x,\mathrm{d}y) \leq \mathbb{E}\Big(\tfrac{s}{(\vartheta h)^{p/3}}\Big|\vartheta m h + x + \sigma x\sqrt{h}Z\Big|^{p/3}\Big),$$

with $Z \sim N(0,1)$. We see that for $p = 3$ the expression on the rhs can be bounded by $A\mathbb{E}(Bx^2)$ for some $A, B > 0$. Hence $L(r_{3,s}) = 0$ for all $s > 0$. In addition,

$$\sup_{x\in[-M,M]} \int r_{3,s}(y)P(x,\mathrm{d}y) < A\mathbb{E}(BM^2) < \infty.$$

Thus, based on Thm. 4.2(i), we can conclude that the chain is ergodic with stationary probability measure μ_h and admits exponential moments $\int r_{3,s}(x)\mu_h(\mathrm{d}x) < \infty$. Furthermore, applying Remark 4.1 to $r_{6,s}(\cdot)$, we see that $\int_0^\infty r_{6,s}(y)P(x,\mathrm{d}y) = \infty$ for x large enough and thus $\int_0^\infty r_{6,s}(x)\mu_h(\mathrm{d}x) = \infty$.

Split-step Euler case. We show that the fully implicit split-step Euler scheme has a stationary measure with polynomial decay. Let $r_\alpha(x) := 1 \vee |x|^\alpha$ and denote $Z \sim N(0,1)$. We have for $|x| > 1$:

$$\frac{\int r_\alpha(y)P(x,\mathrm{d}y)}{r_\alpha(x)} = \mathbb{E}\big|\tfrac{a(\vartheta m h + x)}{x} + \sigma\sqrt{h}Z)\big|^\alpha \xrightarrow{|x|\to\infty} (\sigma^2 h)^{\alpha/2}\cdot\mathbb{E}|Z|^\alpha = L(r_\alpha),$$

by 'dominated convergence'; moreover, $\sup_{x\in[-M,M]}\int r_\alpha(y)P(x,\mathrm{d}y) < \infty$. Using Thm. 4.2(i) we conclude that the chain is ergodic with stationary probability measure μ_h and has polynomial moments $\int r_\alpha(x)\mu_h(\mathrm{d}x) < \infty$ for $\alpha < \alpha_h$, where α_h solves $(\sigma^2 h)^{\alpha_h/2}\cdot\mathbb{E}|Z|^{\alpha_h} = 1$. The fractional moments are available in closed form: $\mathbb{E}|Z|^\alpha = 2^{\alpha/2}\Gamma((\alpha+1)/2)/\sqrt{\pi}$, see e.g. [99]. Hence, α_h can be found numerically. For instance, when $\sigma = 1$ (where ϑ and m are irrelevant) $\alpha_1 = 2$ and $\alpha_{1/2} \approx 4.75$. Conversely,

$$\begin{aligned}\frac{\int_0^\infty r_\alpha(y)P(x,\mathrm{d}y)}{r_\alpha(x)} &= \mathbb{E}\Big(\big|\tfrac{a(\vartheta m h + x)}{x} + \sigma\sqrt{h}Z\big|^\alpha; \tfrac{a(\vartheta m h + x)}{x} + \sigma\sqrt{h}Z > 0\Big)\\ &\to (\sigma^2 h)^{\alpha/2}\mathbb{E}(Z^\alpha; Z > 0),\end{aligned}$$

as $x \to \infty$. So $L^+(r_\alpha) = \frac{1}{2}(\sigma^2 h)^{\alpha/2}\mathbb{E}|Z|^\alpha$, which diverges to ∞, as $\alpha \to \infty$. According to Thm. 4.2(ii), $L^+(r_\alpha) > 2$ implies $\int_0^\infty r(x)\mu_h(\mathrm{d}x) = \infty$. From this we conclude that the tail of μ_h has a power law. Notably, $\alpha_h \to \infty$ as $h \downarrow 0$, so the tail indeed gets lighter as $h \downarrow 0$. Observe however that for all positive h, the

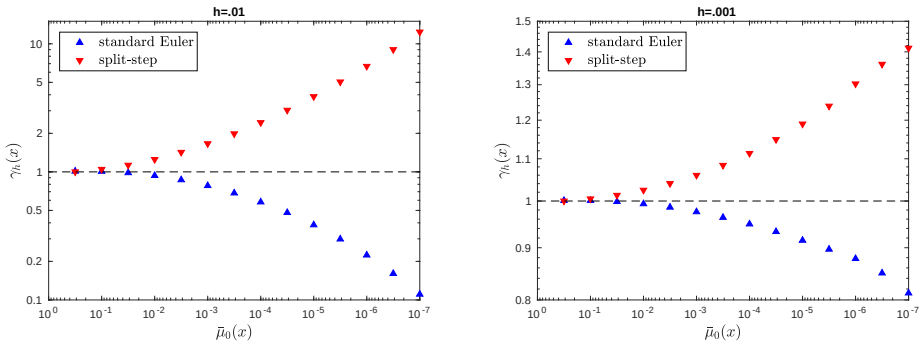


Figure 4.4: Cubic drift and linear volatility process with $\vartheta = 1, \sigma = 1, m = 1$. Plot of $\gamma_h(x)$ against $\bar{\mu}_0(x)$ for $h = 0.01$ (left) and $h = 0.001$ (right) for fully implicit ($\theta = 1$) standard and split-step Euler schemes.

invariant measure μ_h has still only finitely many moments, whereas μ_0 decays in a Weibullian way.

Conclusions: In this example μ_0 corresponds to Weibullian decay with $p = 2$. The fully implicit standard and split-step Euler schemes exhibit completely different behavior. The standard Euler scheme gives rise to Weibullian decay with some $p \in [3, 6]$, and the split-step scheme (strikingly) results in power-law decay. The tails become lighter in the split-step scheme as $h \downarrow 0$. Experimental results are presented in Fig. 4.4.

5 Conclusions

In this chapter we have studied simulation-based techniques for estimating the stationary distribution of SDEs, with a focus on the tail distribution. More specifically, we have considered how the estimate of the stationary tail distribution is affected by the time-discretization scheme chosen. The main conclusion is that estimation of the stationary tail is a highly delicate issue. Commonly accepted discretization techniques may very well lead to highly inaccurate estimates of the stationary tail distribution of the continuous-time process. For instance, we have identified cases in which the stationary tail of the SDE is Weibullian, while the discretizations lead to power-law tails. In our study we have focused on discretizations of one-dimensional SDEs, as for those the solution of the SDE is available in closed form and can be used as a benchmark;

evidently, the above warning carries over to multi-dimensional SDEs.

Given these findings, it is advisable to use different discretization schemes (e.g. standard Euler, split-step, different θ) for assessing tail behavior, and compare their outcomes. Furthermore, we have shown in the first example that the error can be reduced by adapting the stepsize h to the threshold x. For the OU case, we were able to show how to choose h to control the error. Similar results may be possible for other cases, but require further study.

Appendix

This appendix is organized as follows. In Section A we introduce definitions from a general theory of stability of discrete-time Markov chains. In sections B and C we verify the claims about ergodicity and the existence of moments from sections 4.3 and 4.4 respectively.

A Some Theory of Discrete Time Markov Chains

Below we introduce the notions from the stability theory of Markov chains needed to be fulfilled in Theorem 4.2. Necessarily, the theory we lay out here is simplified and incomplete; an interested reader is deferred to [77], on which we heavily rely as well.

Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain taking values in $\mathbb{R}$ with a Borel σ-algebra $\mathcal{B}(\mathbb{R})$ and transition probability kernel $P(x, \mathrm{d}y)$. For any $A \in \mathcal{B}(\mathbb{R})$, $x \in \mathbb{R}$ we define the *first return* time $\tau_A := \min\{n \geq 1 : X_n \in A\}$, and *return probability* $L(x, A) := P_x(\tau_A < \infty)$.

Definition 4.1. *A chain $(X_n)_{n\in\mathbb{N}}$ is called* μ^{Leb}-irreducible *if $L(x, A) > 0$ for all $x \in \mathbb{R}$.*

Definition 4.2. *A set C is called a* small set *if there exists an $m > 0$, and a non-trivial measure ν_m on $\mathcal{B}(\mathbb{R})$, such that for all $x \in C$, $B \in \mathcal{B}(\mathbb{R})$,*

$$P^m(x, B) \geq \nu_m(B) \tag{4.29}$$

When (4.29) holds we say that C is ν_m-small.

Assume that C is a ν_m-small set for some $m \in \mathbb{N}$, and $\nu_m(C) > 0$. For such set C define:

$$E_C := \{n \geq 1 : \text{the set } C \text{ is } \nu_n\text{-small, with } \nu_n = \delta_n \nu_m \text{ for some } \delta_n > 0\} \tag{4.30}$$

Definition 4.3. *Suppose a chain $(X_n)_{n\in\mathbb{N}}$ is μ^{Leb}-irreducible, and let $C \in \mathcal{B}(\mathbb{R})$ be a ν_m-small set with $\nu_m(C) > 0$, and d be the greatest common divisor of the set E_C defined in (4.30). If $d = 1$ then the chain is called* aperiodic.

In particular, from Definition 4.3 it follows that when C is ν_1-small, the chain is aperiodic.

B Supplementary Material for Section 4.3

We verify the ergodicity, existence and nonexistence of moments for Markov chains driven by (4.20) (standard Euler scheme) and (4.21) (split-step Euler scheme) from Section 4.3.

Standard Euler scheme

Regularity conditions for the application of Thm. 4.2. The one-step transition kernel of the chain (4.20) is given by:

$$P(x, \mathrm{d}y) = \frac{1 + 3\theta\vartheta y^2 h}{\sigma\sqrt{h}} \varphi\left(\frac{y + \theta\vartheta y^3 h - F_\theta(x)}{\sigma\sqrt{h}}\right),$$

where $\varphi(\cdot)$ denotes a standard normal pdf. Now, the chain (4.20) is μ^{Leb}-irreducible because $P(x, \mathrm{d}y) > 0$ for all $(x, y) \in \mathbb{R}^2$. All sets $[-M, M]$ for $M > 0$ are ν_1-small for a non-trivial measure

$$\nu_1(\mathrm{d}y) = \inf_{x\in[-M,M]} P(x, \mathrm{d}y) = \tfrac{1+3\theta\vartheta y^2 h}{\sigma\sqrt{h}} \varphi\left(\tfrac{y+\theta\vartheta y^3 h + \mathrm{sgn}(y)(M+\vartheta(1-\theta)M^3 h)}{\sigma\sqrt{h}}\right),$$

where $\operatorname{sgn}(y) := 1_{\{y\geq 0\}} - 1_{\{y<0\}}$. Moreover we have $\nu_1([-M,M]) > 0$, so the chain is aperiodic.

Ergodicity. We show that the scheme is ergodic for $\theta \in (1/2, 1)$. The ergodicity in case $\theta = 1$ is an implication of the existence of moments that we consider later in this section. Let $r(x) = 1 + |x|^6$. We will show that $L(r) < 1$ and $\sup_{x\in[-M,M]} \int r(y)P(x,\mathrm{d}y) < \infty$ for all $M > 0$, which, according to Thm. 4.2(i), entitles that the chain is ergodic. We have:

$$\begin{aligned}\int r(y)P(x,\mathrm{d}y) &\leq 1 + \tfrac{1}{(\vartheta\theta h)^2}\mathbb{E}|x - \vartheta x^3(1-\theta)h + \sigma\sqrt{h}Z|^2 \\ &= 1 + \frac{\sigma^2 h + x^2 + x^6(\vartheta(1-\theta)h)^2 - 2x^4(1-\theta)h}{(\vartheta\theta h)^2} \leq \left(\left(\tfrac{\vartheta(1-\theta)h}{(\vartheta\theta h)}\right)^2 + \varepsilon\right)\cdot x^6\end{aligned}$$

for $\varepsilon > 0$ for $|x|$ large enough. Since $\theta \in (1/2, 1)$ then $\frac{1-\theta}{\theta} < 1$ so we can find ε small enough such that the last expression is smaller than Ax^6 for some $A < 1$, which finally implies that $L(r) < 1$. Lastly, for any $M > 0$:

$$\sup_{x\in[-M,M]} \int r(y)P(x,\mathrm{d}y) \leq 1 + \frac{\sigma^2 h + M^2 + M^6(\vartheta(1-\theta)h)^2}{(\vartheta\theta h)^2} < \infty$$

Existence of moments in case $\theta = 1$. In order to show the existence of moments we apply Thm. 4.2(i) to a family of test functions $r_{p,s}(x) := \mathbb{E}(s|x|^p), x > 0$. For all $M > 0$:

$$\int r_{p,s}(y)P(x,\mathrm{d}y) = \mathbb{E}\Big(s\cdot\Big|a\big(x+\sigma\sqrt{h}Z\big)\Big|^p\Big) \leq \mathbb{E}\Big(\tfrac{s}{(\vartheta h)^{p/3}}\cdot|x+\sigma\sqrt{h}Z|^{p/3}\Big),$$

Now let $p = 6$. We have:

$$\begin{aligned}\mathbb{E}\Big(\tfrac{s}{(\vartheta h)^2}\cdot(x+\sigma\sqrt{h}Z)^2\Big) &= \int \mathbb{E}\Big(\tfrac{s}{(\vartheta h)^2}\cdot(x+\sigma\sqrt{h}z)^2\Big)\tfrac{1}{\sqrt{2\pi}}\mathbb{E}(-z^2/2)\,\mathrm{d}z \quad (4.31)\\ &= \int \tfrac{1}{\sqrt{2\pi}}\mathbb{E}\Big(-\tfrac{Az^2-2Bxz-Cx^2}{2D^2}\Big)\,\mathrm{d}z = \tfrac{D}{\sqrt{A}}\mathbb{E}\Big(\tfrac{(B/A)^2+(C/A)}{2(D^2/A)}\cdot x^2\Big),\end{aligned}$$

where

$$A = (\vartheta h)^2 - 2s\sigma^2 h, \quad B = 2s\sqrt{h}, \quad C = 2s, \text{ and } \quad D = \vartheta h$$

Notice that we quietly assumed that $A > 0$, that is $s < \frac{\vartheta^2 h}{2\sigma^2}$. This shows that $L(r_{p,s}) < 1$. Evidently, we have:

$$\sup_{x\in[-M,M]} \int r_{p,s}(y)P(x,\mathrm{d}y) \leq \tfrac{D}{\sqrt{A}}\mathbb{E}\Big(\tfrac{(B/A)^2+(C/A)}{2(D^2/A)} \cdot M^2\Big)\infty.$$

We conclude that when $\theta = 1$, the standard Euler scheme is ergodic with stationary measure μ_h and moreover, $r_{p,s}(x)\mu_h(\mathrm{d}x) < \infty$ for $p = 6$ and $s < \frac{\vartheta^2 h}{2\sigma^2}$.

Nonexistence of moments in case $\theta = 1$. In order to show that $\int_0^\infty r(x)\mu(\mathrm{d}x) = \infty$ it suffices to show that $\int_0^\infty r(y)P(x,\mathrm{d}y) = \infty$ for all $x \in \mathbb{R}$. We will show that this is the case for $r_{p,s}(\cdot)$ with $p = 6$ and $s > \frac{\vartheta^2 h}{2\sigma^2}$. From the asymptotic behavior of function $a(\cdot)$, we know that for any $\varepsilon \in (0,1)$ there exists $M_\varepsilon > 0$ large enough such that $a(x) > x^{1/3}(\vartheta h)^{-1/3} \cdot (1-\varepsilon)$ for all $x > M_\varepsilon$. Based on the previous calculations, we have:

$$\begin{aligned}\int_0^\infty \mathbb{E}(sy^6)P(x,\mathrm{d}y) \geq \int_{M_\varepsilon}^\infty \mathbb{E}(sy^6)P(x,\mathrm{d}y) &= \mathbb{E}\Big(s|a(x+\sigma\sqrt{h}Z)|^6; x+\sigma\sqrt{h}Z > M_\varepsilon\Big)\\ &\geq \mathbb{E}\Big(\tfrac{s(1-\varepsilon)^6}{(\vartheta h)^2}(x+\sigma\sqrt{h}Z)^2; x+\sigma\sqrt{h}Z > M_\varepsilon\Big)\end{aligned}$$

and the last expression is infinite for $\frac{s\sigma^2 h(1-\varepsilon)^6}{(\vartheta h)^2} > \frac{1}{2}$. Since the choice of $\varepsilon \in (0,1)$ was arbitrary, we conclude that $\int_0^\infty \mathbb{E}(sx^6)\mu(\mathrm{d}x) = \infty$ for $s > \frac{\vartheta^2 h}{2\sigma^2}$.

Split-step Euler scheme

Regularity conditions. See [58, Section 3].

Ergodicity. The fact that the split-step scheme is ergodic for $\theta \in (1/2, 1]$ is immediate due to [58, Section 3] and the asymptotics of function $a(\cdot)$.

Existence of moments in case $\theta = 1$. In the same fashion, as in (4.31), we show that for a family of test functions $r_{p,s} = \mathbb{E}(s|x|^p)$ with $p = 2$ we have

$$\int r_{p,s}(y)P(x,\mathrm{d}y) = \mathbb{E}\Big(s\big|a(x)+\sigma\sqrt{h}Z\big|^p\Big) \leq C_0 \mathrm{e}^{C_1 a^2(x)}$$

where we quietly assumed that $s < \frac{1}{2\sigma^2 h}$. This shows $L(r_{p,s}) < 1$. Lastly, for any $M > 0$

$$\sup_{x\in[-M,M]} \int r_{p,s}(y)P(x,\mathrm{d}y) \leq C_0 \mathrm{e}^{C_1 a^2(M)} < \infty.$$

We conclude that when $\theta = 1$, the stationary measure μ_h exists and moreover, $r_{p,s}(x)\mu_h(\mathrm{d}x) < \infty$ for $p = 2$ and $s < \frac{1}{2\sigma^2 h}$.

Nonexistence of moments in case $\theta = 1$. The reasoning that $\int r_{p,s}(x)\mu(\mathrm{d}x) = \infty$ for $p = 2$, $s > \frac{1}{2\sigma^2 h}$ is analogous to the nonexistence of moments in standard Euler case for $\theta = 1$.

C Supplementary Material for Section 4.4

We verify the ergodicity, existence and nonexistence of moments for Markov chains driven by (4.27) (standard Euler scheme) and (4.28) (split-step Euler scheme) from Section 4.3.

Standard Euler scheme

Regularity conditions. The one-step transition kernel of the chain (4.28) is $P(0,\mathrm{d}y) = \delta_{\{a(\vartheta mh)\}}$, i.e. a dirac measure concentrated at $a(\vartheta mh)$, and, for $x \neq 0$:

$$P(x,\mathrm{d}y) = \tfrac{1+3\theta\vartheta y^2 h}{\sigma x\sqrt{h}}\varphi\left(\tfrac{y+\theta\vartheta y^3 h-\left(a(m\vartheta h)+F_\theta(x)\right)}{\sigma x\sqrt{h}}\right),$$

where $\varphi(\cdot)$ denotes a standard normal pdf. Now, the chain (4.27) is μ^{Leb}-irreducible because $P(x,\mathrm{d}y) > 0$ for all $x \neq 0$, $y \in \mathbb{R}$, and $P^2(0,\mathrm{d}y) = P(a(\vartheta mh),\mathrm{d}y) > 0$ for all $y \in \mathbb{R}$. Now we prove apediodicity. Consider an interval $C := [l_1, l_2] = [a(\vartheta mh) - \varepsilon, 1 + a(\vartheta mh)]$ for any positive $\varepsilon < a(\vartheta mh)$. C is a ν_1-small set with:

$$\nu_1(\mathrm{d}y) = \inf_{x\in C} P(x,\mathrm{d}y) = \tfrac{1+3\theta\vartheta y^2 h}{\sigma l_2\sqrt{h}}\varphi\left(\tfrac{y+\theta\vartheta y^3 h+\mathrm{sgn}(y)\left(a(m\vartheta h)+l_2+\vartheta(1-\theta)l_2^3 h\right)}{\sigma l_1\sqrt{h}}\right),$$

where $\mathrm{sgn}(y) := 1_{\{y\geq 0\}} - 1_{\{\}} y < 0$. Since $\nu_1(\mathrm{d}y) > 0$ then also $\nu_1(C) > 0$ which shows that the chain is aperiodic.

Now we show that the intervals $[-M, M]$ for $M > 0$ are ν_2-small sets (we will introduce the measure ν_2 later). Let $Z \sim N(0,1)$ and denote $Y(x) = a\big(\vartheta mh + F_\theta(x) + \sigma x\sqrt{h}Z\big)$. Then $Y(x)$ is a random variable with law $P(x, \mathrm{d}y)$. By Continuous Mapping Theorem we see that, as $x \to 0$, $Y(x) \to a(\vartheta mh)$ a.s. It follows that for any $\delta > 0$, $\mathbb{P}(|Y(x) - a(\vartheta mh)| > \delta) \to 0$. In particular, since $a(\vartheta mh) \in C$, for any positive $\delta_0 < 1$ there exists $\delta_1 > 0$ such that $\int_C P(x, \mathrm{d}y) = \mathbb{P}(Y(x) \in C) > 1 - \delta_0$ for $|x| < \delta_1$. Additionally, for $x \in [-M, \delta_1] \cup [\delta_1, M]$ we have:

$$\begin{aligned}\int_C P(x,\mathrm{d}y) &= \int_C \tfrac{1+3\theta\vartheta y^2 h}{\sigma x\sqrt{h}}\varphi\left(\tfrac{y+\theta\vartheta y^3 h - \left(a(m\vartheta h)+F_\theta(x)\right)}{\sigma x\sqrt{h}}\right)\mathrm{d}y \\ &\geq \int_C \tfrac{1+3\theta\vartheta y^2 h}{\sigma M\sqrt{h}}\varphi\left(\tfrac{y+\theta\vartheta y^3 h + \mathrm{sgn}(y)\left(a(m\vartheta h)+M+\vartheta(1-\theta)M^3 h\right)}{\sigma\delta_1\sqrt{h}}\right)\mathrm{d}y\end{aligned}$$

We call the last expression c_0 and see that $c_0 > 0$, because the integrand is strictly positive. This shows that $\inf_{x\in[-M,M]} \int_C P(x, \mathrm{d}y) \geq \min\{1 - \delta_0, c_0\}$. Finally, we see that for $x \in [-M, M]$:

$$\begin{aligned}P^2(x,\mathrm{d}y) &= \int_{z\in\mathbb{R}} P(x,\mathrm{d}z)P(z,\mathrm{d}y) \geq \int_{z\in C} P(x,\mathrm{d}z)P(z,\mathrm{d}y) \\ &\geq \int_{z\in C} P(x,\mathrm{d}z)\nu_1(\mathrm{d}y) \geq \min\{1-\delta_0, c_0\}\cdot\nu_1(\mathrm{d}y).\end{aligned}$$

So $[-M, M]$ is ν_2-small for $\nu_2(\mathrm{d}y) := \min\{1 - \delta_0, c_0\}\nu_1(\mathrm{d}y)$.

Ergodicity. We show that the scheme is ergodic for $\theta \in (1/2, 1)$. The ergodicity in case $\theta = 1$ is an implication of the existence of moments that we consider later in this section. Let $r(x) = 1 + |x|^6$. We will show that $L(r) < 1$ and $\sup_{x\in[-M,M]} \int r(y)P(x, \mathrm{d}y) < \infty$ for all $M > 0$, which, according to Thm. 4.2(i), entitles that the chain is ergodic. The calculations are very similar to the standard Euler case in Section B.

$$\begin{aligned}\int r(y)P(x,\mathrm{d}y) &= 1 + \mathbb{E}\Big|a\big(\vartheta mh + F_\theta(x) + \sigma x\sqrt{h}Z\big)\Big|^6 \\ &\leq 1 + \tfrac{1}{(\vartheta\theta h)^2}\mathbb{E}|\vartheta mh + x - \vartheta x^3(1-\theta)h + \sigma x\sqrt{h}Z|^2 \\ = 1 + \frac{\sigma^2x^2h + \big(\vartheta mh + x - \vartheta x^3(1-\theta)h\big)^2}{(\vartheta\theta h)^2} &\leq \left(\left(\tfrac{\vartheta(1-\theta)h}{(\vartheta\theta h)}\right)^2 + \varepsilon\right)\cdot x^6\end{aligned}$$

for any $\varepsilon > 0$ for large enough $|x|$. Since $\theta \in (1/2, 1)$ then $\frac{1-\theta}{\theta} < 1$ so we can find ε small enough such that the last expression is smaller than Ax^6 for some $A < 1$, which finally implies that $L(r) < 1$. Lastly, for $M > 0$:

$$\sup_{x\in[-M,M]} \int r(y)P(x, \mathrm{d}y) \leq 1 + \frac{\sigma^2 M^2 h + \big(\vartheta m h + M + \vartheta M^3(1-\theta)h\big)^2}{(\vartheta\theta h)^2} < \infty$$

Existence of moments in case $\theta = 1$. We will show, using Thm. 4.2(i), that the fully implicit standard Euler scheme admits a stationary measure with exponential moments $r_{p,s} = \mathbb{E}(s|x|^p)$ with at least $p = 3$. Let $Z \sim N(0, 1)$. We have

$$\int r_{p,s}(y)P(x, \mathrm{d}y) \leq \mathbb{E}\Big(\tfrac{s}{(\vartheta h)^{p/3}} \cdot |\vartheta m h + x + \sigma x\sqrt{h}Z|^{p/3}\Big)$$

and we see that for $p = 3$, the expression on the rhs can be bounded by $A\mathbb{E}(Bx^2)$ for some $A, B > 0$. This shows that $L(r_{3,s}) = 0$ for all $s > 0$. Additionally we have $\sup_{x\in[-M,M]} \int r_{3,s}(y)P(x, \mathrm{d}y) < A\mathbb{E}(BM^2) < \infty$, which, according to Thm. 4.2(i), entitles that the chain is ergodic with stationary probability measure μ_h and admits exponential moments $\int r_{3,s}(x)\mu_h(\mathrm{d}x) < \infty$ for all $s > 0$.

Nonexistence of moments in case $\theta = 1$. Applying Remark 4.1 to $r_{p,s}$ with $p = 6$ we will see that $\int_0^\infty r_{6,s}(y)P(x, \mathrm{d}y) = \infty$ for x large enough and thus $\int_0^\infty r_{6,s}(x)\mu_h(\mathrm{d}x) = \infty$.

Split-step Euler scheme

Regularity conditions. Verifying that the chain driven by (4.28) is μ^{Leb}-irreducible, aperiodic, and intervals $[-M, M]$ are small sets is similar to the standard Euler case.

Ergodicity. We show that the scheme is ergodic for $\theta \in (1/2, 1)$. The ergodicity in case $\theta = 1$ is an implication of the existence of moments that we consider later in this section. Let $r_\alpha(x) = 1 \vee |x|^\alpha$. We will show that $L(r) < 1$ and $\sup_{x\in[-M,M]} \int r(y)P(x, \mathrm{d}y) < \infty$ for all $M > 0$, which, based on Thm. 4.2 (i), together entitle that there exists a unique stationary probability measure μ_h. Based on 4.23,

$$\frac{\int r_\alpha(y)P(x, \mathrm{d}y)}{r_\alpha(x)} \xrightarrow{|x|\to\infty} \mathbb{E}|\left(\tfrac{1-\theta}{\theta}\right)^{1/3} + \sigma\sqrt{h}Z|^\alpha = L(r_\alpha)$$

by Dominated Convergence Theorem. Since $\theta \in (1/2, 1)$ then $\frac{1-\theta}{\theta} < 1$, so there exist $\alpha > 0$ such that $L(r_\alpha) < 1$. Lastly:

$$\begin{aligned}\int r_\alpha(y)P(x, \mathrm{d}y) &\leq 1 \wedge \mathbb{E}\Big|a\big(\vartheta mh + F_\theta(x)\big) + \sigma x\sqrt{h}Z\Big|^\alpha \\ &\leq \mathbb{E}\Big(\big|a(\vartheta mh + |x| + \vartheta|x|^3(1-\theta)h + \sigma|x|\sqrt{h}|Z|\Big)^\alpha\end{aligned}$$

thus for $M > 0$,

$$\sup_{x\in[-M,M]} \int r(y)P(x, \mathrm{d}y) \leq \mathbb{E}\Big(\big|a(\vartheta mh + M + \vartheta M^3(1-\theta)h + \sigma M\sqrt{h}|Z|\Big)^\alpha < \infty.$$

Existence of moments in case $\theta = 1$. Ergodicity and existence of moments is analogous to the case $\theta \in (1/2, 1)$. Using Thm. 4.2(i) we conclude that the chain is ergodic with stationary probability measure μ_h and has polynomial moments $\int r_\alpha(x)\mu_h(\mathrm{d}x) < \infty$ for $\alpha < \alpha_h$, where α_h solves $(\sigma^2 h)^{\alpha_h/2} \cdot \mathbb{E}|Z|^{\alpha_h} = 1$.

Nonexistence of moments in case $\theta = 1$. With arguments analogous to the case of *existence of moments*, we have:

$$\frac{\int_0^\infty r(y)P(x, \mathrm{d}y)}{r(x)} \to (\sigma^2 h)^\alpha \mathbb{E}(Z^\alpha; Z > 0),$$

as $x \to \infty$. So $L^+(r_\alpha) = \frac{1}{2}(\sigma^2 h)^{\alpha/2}\mathbb{E}|Z|^\alpha$, which diverges to ∞, as $\alpha \to \infty$. According to Thm. 4.2(ii), $L^+(r_\alpha) > 2$ implies $\int_0^\infty r(x)\mu_h(\mathrm{d}x) = \infty$.

Chapter 5

Rare Event Simulation for Steady-State Probabilities via Recurrency Cycles

In Chapter 4 we have quantified the discretization error for the stationary distribution of a time-continuous Markov chain. In all examples considered, the stationary distribution of the discrete version of the Markov chain was unknown and the tail probabilities had to be estimated via simulation. Since the smallest probabilities were of order 10^{-7} the naïve Monte Carlo estimation was extremely inefficient and it took a couple of days to obtain a reasonably accurate estimator.

In this chapter we develop a new algorithm for the estimation of rare event probabilities associated with the steady-state of a Markov stochastic process with continuous state space $\mathbb{R}^d$ and discrete time steps (i.e. a discrete-time $\mathbb{R}^d$-valued Markov chain). The method could be used for the estimation of tail probabilities from Chapter 4. The algorithm, which we coin Recurrent Multilevel Splitting (RMS), relies on the Markov chain's underlying recurrent struc-

ture, in combination with the Multilevel Splitting method. Extensive simulation experiments are performed, including experiments with a nonlinear stochastic model that has some characteristics of complex climate models. The numerical experiments show that RMS can boost the computational efficiency by several orders of magnitude compared to the Monte Carlo method.

1 Introduction

Many stochastic processes have a 'stable regime', in the sense that with time their distribution converges to a so-called *steady-state*. The steady-state (or stationary, equilibrium, ergodic) probability distribution captures the long-term behavior of the process; the steady-state probability of an arbitrary event (or set) B is equal to the fraction of time the process spends in B in the long run (irrespective of the process' initial value). In many application domains steady-state probabilities are of crucial interest; think of physics (e.g. particle systems), chemistry (e.g. reaction networks), and operations research (e.g. queueing systems). Within this context of steady-state distributions, an important subdomain concerns the analysis of *rare events*. Particularly when it concerns rare events with a potentially catastrophic impact, there is a clear need to accurately estimate their likelihood (earthquakes, extreme weather conditions, simultaneous failure of multiple components of a machine, etc.). As examples, we refer to [81] for rare-event simulation methods in the climate context, and to [85] for a textbook treatment covering applications in e.g. engineering, chemistry, and biology.

Despite the evident importance of being able to estimate steady-state rare-event probabilities, relatively little attention has been paid to the development of efficient algorithms; rare-event simulation in a finite-time horizon context received considerably more attention (focusing e.g. on the estimation of the probability to hit a set B_1 before hitting another set B_2). The main contribution of this chapter concerns the development of a broadly applicable rare-event simulation method that is tailored to the estimation of small steady-state probabilities.

In our setup we focus on discrete-time $\mathbb{R}^d$-valued Markov chains. This frame-

work covers a wide class of intensively used stochastic models. It for instance includes the numerical solutions to stochastic differential equations (SDEs), see e.g. [69]. In addition, various (inherently discrete-time) standard models from e.g. finance, biology, and econometrics fall under this umbrella. The main advantage of our proposed algorithm is its broad applicability, the fact that it does not require detailed knowledge of the system under study, and that it is fairly straightforward to implement. In the sequel, we let $(X_n)_{n\in\mathbb{N}}$ be our d-dimensional Markov chain, which we assume to admit the stationary distribution μ. We are interested in the probability that in steady-state the process attains a value in the set B, i.e.,

$$\gamma := \mu(B) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} 1_{\{X_n \in B\}} \tag{5.1}$$

Throughout, the event B is assumed to be *rare*, entailing that γ is very small, typically of order 10^{-4} or less (depending on the application at hand).

Our interest lies in estimating rare-event probabilities in the context of *models*, so in principle we can do more than applying statistical methods of extreme value analysis to model data; cf. Coles et al. [31] for a textbook on Extreme Value Analysis. In our setup, the steady-state distribution is not explicitly known; one therefore has to resort to simulation. The naïve, Monte Carlo estimator for γ is

$$\widehat{\gamma}_{\mathrm{MC}} := \frac{1}{N} \sum_{n=1}^{N} 1_{\{X_n \in B\}},$$

i.e., *the average number of visits to set B until time N*, which is known to be extremely inefficient when B is rare; see e.g. Asmussen and Glynn [8]. Informally, one needs prohibitively many samples in order to obtain a reasonably accurate estimate of γ; the number of samples required to obtain an estimate of given precision is inversely proportional to γ. In many cases, especially while working with complex or high-dimensional systems, where the integration of the model is time consuming, such computation might not be feasible.

An additional complication is that sampling directly from the steady-state distribution can be challenging. In our new method, we settle this issue by

dissecting the paths of the underlying Markov chain into *recurrency cycles*. For an arbitrary set A, we say that a recurrency occurs each time $(X_n)_{n\in\mathbb{N}}$ crosses A *inwards*, i.e., each time the event $\{X_{n-1} \notin A, X_n \in A\}$ occurs. Assuming the process is in stationarity, γ is equal to *the average amount of time spent in B between two visits to the set A, divided by the average length of a recurrency cycle.*

An example of a recurrency cycle is shown in Figure 5.1. It starts at P_1 and ends at P_5; the time spent in set B is the time spent between states P_3 and P_4. Note that recurrency is defined with respect to A; it is not necessary that the system enters B during a recurrency cycle.

In our algorithm we separately estimate the numerator (expected time spent in B during a single recurrency cycle) and the denominator (expected length of a single recurrency cycle). Here, two challenges arise. The first concerns the choice of the set A. Any A could in principle be used, but in order to maximize the efficiency of the algorithm, it should be chosen so as to minimize the expected time spent between visits to the set A. The second challenge is posed by the rarity of visiting B within a cycle. To tackle this issue, we propose the use of Multilevel Splitting (MLS), see Garvels [51], [85], but we remark that instead of MLS other methods could be chosen. These alternatives include Genealogical Particle Analysis (see e.g. Del Moral and Garnier [36]), RESTART (see e.g. [95]), Adaptive Multilevel Splitting (see e.g. Cérou and Guyader [28]), fixed-effort and fixed number of successes versions of Multilevel Splitting (see e.g. Amrein and Künsch [5]) and Importance Sampling (see e.g. Heidelberger [59]). We emphasize that we do not seek to compete with any of the aforementioned methods but rather introduce a new overarching framework, in which all these methods can be used to assess stationary performance metrics. We have chosen to work with MLS mostly for its conceptual simplicity and intuitive use.

The algorithm we propose is inspired by expressions for steady-state probabilities resulting from the theory of *regenerative processes*. Regeneration instances dissect the path of the process into probabilistically identical, independent segments. For regenerative processes we have that γ equals the average amount of time spent in B in a regeneration cycle divided by the average length of a regeneration cycle. For more background we refer to Crane and Iglehart [32]

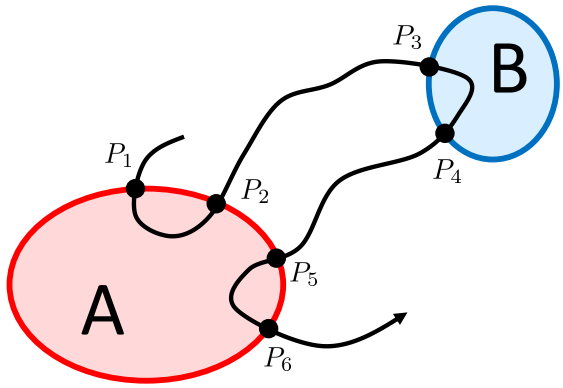


Figure 5.1: An example of a recurrency cycle. The cycle begins at P_1, where the Markov chain enters the set A from the outside, and ends at P_5 where the chain enters A again (and the next recurrency cycle begins).

and Asmussen [6], or (in a more informal language) Henderson and Glynn [60]. In our setup, with its uncountable state space and a steady-state distribution potentially lacking atoms, we cannot straightforwardly construct regeneration points. We therefore develop an approach that relies on the recurrency cycles introduced above, so as to set up a scheme that yields probabilistically identical (but not necessarily independent) cycles. We refer to Goyal et al. [56] for an algorithm corresponding to the setting in which the set A consists of finitely many elements (which inspired us to develop our algorithm). We also mention that a large subclass of general (continuous) state-space Markov chains, called *positive Harris*, is regenerative. However, constructing regeneration cycles in this context is typically technically difficult, and in addition the implementation may be computationally inefficient due to excessively long cycle lengths; see Henderson and Glynn [61].

The chapter is organized as follows. In Section 2 we discuss preliminaries, such as basic theory of general state-space Markov chains. We also give an alternative representation of the parameter γ based on the recurrent structure of a Markov Chain in Theorem 5.1. Relying on this alternative representa-

tion, in Section 3, we introduce a new algorithm for the estimation of γ, which we coin *Recurrent Multilevel Splitting* (RMS). In Section 4, we establish (in a simplified setting) the optimal parameters for the RMS algorithm, and provide implementation-related guidelines. Theorem 5.3 in Appendix C establishes the asymptotic efficiency of the RMS algorithm. A technical derivation of the optimal parameters is given in Appendix B. In Section 5 we test the method on a set of numerical examples, we discuss which factors affect the method's performance, and provide heuristics. Finally, in Section 6 we discuss possible extensions of the algorithm and give a summary. Appendix A consists of a collection of required technical results.

2 Preliminaries

Here we introduce concepts used later in Section 3 such as (Harris) recurrence, the stationary measure and *recurrency cycles*.

2.1 Continuous State-Space Markov Chains

In this subsection we provide some background on the (well-established) theory of stability of discrete-time Markov chains with a general (continuous) state-space. The underlying theory can be found in textbooks on Markov chains; our notation is in line with the one used in Meyn and Tweedie [77].

The theory of stability for general state-space time-discrete Markov chains differs from the one for its *finite* (or countable) state-space counterpart. Due to the continuous state space, multiple visits to the same state may happen with probability 0. This explains why the classic notion of *irreducibility* and *recurrence* of states has been generalized to *sets* (rather than states). In this setting one typically works with the concept of so-called *positive Harris recurrent* chains: sets of states are guaranteed to be visited infinitely often, with in addition a *finite expected return time*. Effectively all Markov chains with an invariant probability distribution are positive Harris (with an exception of pathological, custom-made examples); see [77, Section 9] for a rigorous treatment of the topic.

Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain taking values in $\mathbb{R}^d$ with a transition kernel

$P(x, \mathrm{d}y)$, meaning that the distribution of X_{n+1} conditional on $X_n = x$ is given by

$$\mathbb{P}(X_{n+1} \in A \mid X_n = x) = \int_A P(x, \mathrm{d}y) \tag{5.2}$$

for measurable sets $A \subseteq \mathbb{R}^d$. We denote $P(x, A) := \int_A P(x, \mathrm{d}y)$. Then, the stationary distribution μ satisfies the relation

$$\mu(A) = \int_{\mathbb{R}^d} \mu(\mathrm{d}x) P(x, A).$$

For an arbitrary probability measure ν, we define the conditional probability and expectation by $\mathbb{P}_\nu(\cdot) = \mathbb{P}(\cdot \mid X_0 \sim \nu)$ and $\mathbb{E}_\nu(\cdot) = \mathbb{E}(\cdot \mid X_0 \sim \nu)$, respectively. In particular, when ν corresponds to a point mass at x, we use the compact notations $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$ and $\mathbb{E}_x(\cdot) = \mathbb{E}(\cdot \mid X_0 = x)$, respectively.

2.2 Recurrent Structure of a Markov Chain

As mentioned in the introduction, a large class of general state-space Markov chains (more specifically, the class of positive Harris recurrent Markov chains) allows a regenerative structure; see e.g. [61]. However, for application purposes, it is often difficult to sample the regeneration times. Moreover, even when it is possible to sample these, the implementation is often inefficient due to the long cycle lengths — in fact, the regeneration may be a rare event itself.

There are many other ways to decompose a Markov chain into cycles. We propose to work with cycles that start with an inward crossing of a set A (i.e., entering A from the outside). We denote the time of the $(k+1)$-th inward crossing by S_k, i.e.,

$$S_k := \inf\{n > S_{k-1} : X_{n-1} \notin A, X_n \in A\}.$$

with $S_{-1} := 0$. Then, we define the paths within the cycles through

$$\mathcal{C}_k := \big(X_n : S_{k-1} \le n < S_k - 1\big). \tag{5.3}$$

With a k-th cycle we associate the *cycle length* and the *cycle origin* (or starting point),

$$L_k := S_k - S_{k-1}, \qquad X_k^A := X_{S_{k-1}}. \tag{5.4}$$

We call A the *recurrency set* and $\mathcal{C}_1, \mathcal{C}_2, \ldots$ *recurrency cycles*. Under the assumption that the process $(X_n)_{n\in\mathbb{N}}$ starts in a cycle-stationary regime (that is $X_0 \sim \mu$ and $S_0 = 1.$), the pairs $(\mathcal{C}_1, L_1)$, $(\mathcal{C}_2, L_2), \ldots$ are identically distributed. However, the cycles (5.3) are generally not independent, as two distinct cycle origins X_k^A, X_m^A separated by a short time period $S_{m-1} - S_{k-1}$ tend to be located within the same subregion of the recurrency set. Because of this dependence, the decomposition into recurrency cycles is neither *classic* nor *wide sense regenerative*, see Definition 3.1 and 3.3 in Kalashnikov [66]. The way we define cycles is a special case of the *almost regenerative cycles* introduced by Gunther and Wolff [57]. The interested reader is referred to the introduction of [26], where a more exhaustive account of different regeneration-type methods is outlined.

A single recurrency cycle reflects the behavior of the process in steady-state. To make this claim more precise, define *the total time spent in the set B within the k-th cycle*:

$$R_k := \sum_{n=S_{k-1}}^{S_k - 1} 1_{\{X_n \in B\}}. \tag{5.5}$$

Since (in a cycle-stationary regime) the cycles in (5.3) are identically distributed, so are $R_1, R_2, \ldots$. The following theorem states that the total fraction of time that the process (X_n) spends in the set B is proportional to the expected time spent in B between two consecutive inward crossings into A. Define the *frequency of recurrence* $\alpha_A := \mathbb{P}_\mu(X_0 \notin A, X_1 \in A)$.

Theorem 5.1. *Let $(X_n)_{n\in\mathbb{N}}$ be a positive Harris recurrent Markov chain and let μ denote its unique stationary probability measure. Let A, B be measurable sets such that $\mu(A) \in (0, 1)$. Let L_1 be as defined in* (5.4)*, R_1 as defined in* (5.5)*, and $T_B := \mathbb{E}_\mu R_1$. Then $\mathbb{E}_\mu L_1 < \infty$,*

$$\mu(B) = \alpha_A \cdot T_B \tag{5.6}$$

and $\alpha_A = (\mathbb{E}_\mu L_1)^{-1}$.

Proof. See Appendix A. □

The factorization (5.6) of γ from Theorem 5.1 is the starting point from which we develop our steady-state rare-event simulation algorithm in Section 3.

We note that an analogue of Theorem 5.1 holds for regenerative processes. Dissection of a Markov chain into *regeneration cycles* has one clear advantage over dissection into recurrency cycles, namely, the regeneration cycles are *independent*. Using this independence, one can easily infer the variance of an estimator based on regeneration cycles. Nonetheless, it is more attractive to use recurrency cycles than regeneration, as the latter is harder to implement and has a (much) longer expected cycle length. Moreover, in situations where it is possible to sample from the stationary distribution μ, one can simulate independent paths until the first recurrency cycle has ended, such that the resulting cycles will be independent as well.

3 Recurrent Splitting Algorithm

Our algorithm essentially relies on the result from Theorem 5.1, namely the representation of γ as a product of two quantities. Thus, we divide our algorithm into two stages: first there is the estimation of α_A (the frequency of recurrence, equal to the reciprocal of the expected cycle length), and secondly the estimation of T_B (the expected time spent in set B within a recurrency cycle).

3.1 Estimation of α_A

While it is relatively straightforward to estimate α_A (for example with a crude Monte Carlo method), the choice of the recurrency set A is non-trivial. In this section we assume that A has already been chosen; the choice of A is discussed in Section 4.2.

In typical situations one can generate sample paths of X_n by simulation but it is not possible to *exactly* sample from the stationary distribution. Even though the law of X_n converges to μ weakly, as $n \to \infty$, at any fixed time n, the law of X_n is not exactly μ. Perhaps the most straightforward method to estimate α_A in this setting is the method of batch-means. It relies on dissecting a path of the Markov chain of length N into $m \in \mathbb{N}$ batches of equal length, and calculating the sample frequency of entering the set A for each batch. More

specifically, with $M := [N/m]$,

$$\widehat{\alpha}_k := \frac{1}{M} \sum_{n=(k-1)M+1}^{kM} 1_{\{X_{n-1} \notin A, X_n \in A\}},$$

and then the batch-means estimator is

$$\widehat{\alpha}_A^{\mathrm{BM}} := \frac{1}{m} \sum_{k=1}^{m} \widehat{\alpha}_k. \tag{5.7}$$

Let s_{BM}^2 be the sample variance of $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_m$ and t_{m-1} a Student's t distribution with $m-1$ degrees of freedom. Then, due to the 'near independence' between the batches, under appropriate regularity assumptions,

$$\sqrt{m}(\widehat{\alpha}_A^{\mathrm{BM}} - \alpha)/s_{\mathrm{BM}} \xrightarrow{\mathrm{d}} t_{m-1}, \tag{5.8}$$

as $N \to \infty$, with '$\xrightarrow{\mathrm{d}}$' denoting convergence in distribution. For more details and background, we refer to e.g. Asmussen and Glynn [8].

We remark that when an exact sampling procedure from μ is available, then it might be more efficient to use the following Monte Carlo estimator. Generate M independent pairs

$$(X_0^{(1)}, X_1^{(1)}), \ldots, (X_0^{(M)}, X_1^{(M)})$$

with (for all $i = 1, .., M$) $X_0^{(i)} \sim \mu$ and $X_1^{(i)}$ distributed according to the dynamics of the Markov chain (5.2) conditional on the value of $X_0^{(i)}$. The Monte Carlo estimator

$$\widehat{\alpha}_A^{\mathrm{MC}} := \frac{1}{M} \sum_{i=1}^{M} 1_{\{X_0^{(i)} \notin A, X_1^{(i)} \in A\}} \tag{5.9}$$

is unbiased, $\mathbb{V}\mathrm{ar}\, \widehat{\alpha}_A^{\mathrm{MC}} = \alpha_A(1-\alpha_A)/M$, and, as $M \to \infty$,

$$\sqrt{M}(\widehat{\alpha}_A^{\mathrm{MC}} - \alpha)/s_{\mathrm{MC}} \xrightarrow{\mathrm{d}} N(0,1), \tag{5.10}$$

with s_{MC}^2 the sample variance.

Whether exact simulation from μ is available or not, both methods allow for the construction of confidence intervals based on the weak convergence results

(5.8) and (5.10). It should be clear that the set A should be chosen such that α_A is not prohibitively small, so that the methods (5.7) and (5.9) are computationally efficient. Otherwise, the estimation of α_A would be a rare event simulation problem itself (which we obviously want to avoid).

3.2 Estimation of T_B

The second stage of the algorithm concerns the estimation of T_B, as defined in Theorem 5.1. This step is the more challenging one, as the quantity T_B is expected to be very small. We resort to rare-event simulation methods. For clarity of exposition, throughout this section we assume that the chain $(X_n)_{n\in\mathbb{N}}$ is stationary, $S_0 = 0$ and we drop the subscript in $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ (i.e., we write simply $\mathbb{P}$ and $\mathbb{E}$, respectively). We also assume that we can sample from the distribution of the cycle starting point X_1^A (note that $X_1^A, X_2^A, ...$ are all identically distributed). If we can not, then we sample from X_1^A approximately; this is discussed in Section 3.3. We first introduce some notation; we define $p_B := \mathbb{P}(\tau_B < \tau_A^{\text{in}})$, with

$$\tau_B := \inf\{n > 0 : X_n \in B\}, \quad \tau_A^{\text{in}} := S_1 = \inf\{n > 0 : X_{n-1} \notin A, X_n \in A\},$$

and

$$R_+ :\overset{\text{d}}{=} \big(R_1 \mid R_1 > 0\big) \tag{5.11}$$

with '$\overset{\text{d}}{=}$' denoting equality in distribution. Note that $\tau_A^{\text{in}} - 1$ marks the end of the first recurrency cycle. Since $\{R_1 > 0\} = \{\tau_B < \tau_A^{\text{in}}\}$, p_B is the probability of reaching B within a cycle, and R_+ is a random variable distributed as the total time spent in the set B within a cycle conditioned on the cycle reaching set B. As was noted in Garvels [51],

$$\mathbb{E}(R_1) = \mathbb{P}(R_1 > 0) \cdot \mathbb{E}(R_1 \mid R_1 > 0).$$

This entails that

$$T_B = \mathbb{P}(\tau_B < \tau_A^{\text{in}}) \cdot \mathbb{E}(R_1 \mid \tau_B < \tau_A^{\text{in}}) = p_B \cdot \mathbb{E}R_+ \tag{5.12}$$

The estimation of p_B is a classic rare-event simulation problem, for which various methods have been developed. Following [51], we propose to use a Multilevel

Splitting (MLS) algorithm to estimate T_B (but, as we mentioned before, other approaches could be followed as well). There are a number of variations of the MLS algorithm; we chose to rely on its simplest version (called 'Fixed Splitting'). The following exposition aligns with Amrein and Künsch [5].

As mentioned, the naïve Monte Carlo method is inefficient for the estimation of small probabilities, because of the computational effort wasted on simulating irrelevant paths. The core idea behind the MLS method is to *split* the path of the process when it approaches B. This way, we have more control over the simulation, by forcing the process into interesting regions. In order to implement the MLS algorithm, one must first choose an *importance function* $H : \mathbb{R}^d \to [0, 1]$ which assigns an *importance value* to every possible state. H should be chosen such that $H(x) = 1$ if and only if $x \in B$ and $H(x) = 0$ for $x \in A$. We postpone the discussion about the choice of the importance function to Section 4.2.

We now formally introduce the MLS algorithm. First divide the interval $[0, 1]$ into m subintervals with endpoints:

$$0 = \ell_0 < \ell_1 < \ldots < \ell_m = 1,$$

and define the corresponding stopping times and events

$$\tau_k := \inf\{n \geq 0 : H(X_n) \geq \ell_k\}, \;\; D_k := \{\tau_k < \tau_A^{\text{in}}\}; \tag{5.13}$$

for $k \in \{0, \ldots, m\}$. Note that τ_k is the first time an importance value greater or equal to ℓ_k has been reached; in particular $\tau_m = \tau_B$ and $\tau_0 = 0$, so that $X_{\tau_0} \overset{\text{d}}{=} X_1^A$. Finally let

$$p_k := \mathbb{P}(D_k \,|\, D_{k-1}), \quad k \in \{1, \ldots, m\},$$

and $p_0 = 1$, to which we refer as *conditional probabilities*. From the definition (5.13) we have $\mathbb{P}(D_m) = p_B$ and since $D_0 \subseteq D_1 \subseteq \ldots \subseteq D_m$, we conclude

$$p_B = \prod_{k=0}^{m} p_k.$$

Finally, define *splitting factors* $n_0, n_1, \ldots, n_m \in \mathbb{N}$, representing the number of

independent continuations of the process that are sampled when reaching the respective importance levels. Here n_0 plays a special role, as it is a number of independent MLS estimators; the final estimator will be a mean of n_0 independent MLS estimators. By virtue of this independence, we are able to estimate the variance of the final estimator. For simplicity, in the following it is assumed that $n_0 = 1$.

Algorithm 5.1 (Multilevel Splitting)**.**

1. *Set* $k := 0$, $r_0 := 1$, *sample* $X_0^1 \sim X_1^A$.

2. *In the* k*-th stage we have a sample of* r_k *entrance states* $(X_k^1, \ldots, X_k^{r_k})$, *where we denote*

$$X_k^i := X_{\tau_k^i}^i.$$

   For each state X_k^i *generate* n_k *independent path continuations until* $\min\{\tau_{k+1}, \tau_A^{\text{in}}\}$. *The number of paths for which the event* D_{k+1} *occurred is denoted by* r_{k+1}. *Store all* r_{k+1} *states* X_{k+1}^i, *for which the event* D_{k+1} *occurred, in memory.*

3. *If* $r_{k+1} = 0$, *then stop the algorithm and put* $\widehat{p}_B := 0$, $\widehat{T}_B := 0$.

4. *If* $k < m - 1$, *then increase* k *by one and go back to step 2; otherwise put*

$$\widehat{p}_B := \frac{r_m}{\prod_{k=0}^{m-1} n_k}. \tag{5.14}$$

5. *If* $r_m = 0$, *then return* $\widehat{T}_B = 0$*; otherwise, for each state* X_m^i *generate* n_m *independent path continuations until* τ_A^{in}. *For each of these* $r_m n_m$ *continuations record the time spent in set* B*:*

$$\widehat{R}_+^{(j)} := \sum_{k=\tau_m}^{\tau_A^{\text{in}}-1} 1_{\{X_k \in B\}}.$$

   Calculate the total time spent in B *by*

$$r_{m+1} := \sum_{j=1}^{r_m n_m} \widehat{R}_+^{(j)}$$

6. The final estimator is

$$\widehat{T}_B := \frac{r_{m+1}}{\prod_{k=0}^{m} n_k} \tag{5.15}$$

Theorem 5.2. *The estimators $\widehat{p}_B$ and $\widehat{T}_B$, as defined in* (5.14) *and* (5.15)*, are unbiased estimators for p_B and T_B respectively.*

The following proof is based on notes of the *Summer School in Monte Carlo Methods for Rare Events* that took place at Brown University, Providence RI, USA in June 2016 (authored by J. Blanchet, P. Dupuis, and H. Hult). It is noted that various alternative derivations can be constructed; see e.g. [8].

Proof of Theorem 5.2. Let $\bar{X}_{i,j}$ be labeling all descendants of the original particle, with i indexing time and j indexing the descendant. All descendants $\bar{X}_{\cdot,j}$ are identically distributed (but not independent). Now suppose that each particle has an evolving weight $w_{i,j}$. Concretely, this means that when a particle crosses a threshold ℓ_k, it is split into n_k particles and its weight is divided equally among its descendants (i.e., each of them obtaining a share $1/n_k$ of $w_{i,j}$). Each particle that reaches the set B has been split m times, and its weight is thus $1/\prod_{k=1}^{m} n_k$. For particles that did not reach set B, we artificially split these particles (keeping them in A) for the remaining thresholds so that the total number of particles is $\prod_{k=1}^{m} n_k$, each of equal weight. Then, using the fact that the descendants are identically distributed, we obtain

$$\mathbb{E}\widehat{T}_B = \mathbb{E}\left(\sum_{j=1}^{\prod_{k=1}^{m} n_k} \frac{1}{\prod_{k=1}^{m} n_k} \sum_i 1_{\{\bar{X}_{i,j} \in B\}} \right) = \mathbb{E} \sum_i 1_{\{\bar{X}_{i,1} \in B\}} = T_B.$$

Analogously, $\mathbb{E}\widehat{p}_B = p_B$, which ends the proof. □

We remark that, with $r_1, \ldots, r_m$ as defined in Algorithm 5.1, the same arguments as the ones featuring in the proof of Thm. 5.2 imply the unbiasedness of the estimators for $\mathbb{P}(D_k)$:

$$\mathbb{E}\left(\frac{r_k}{\prod_{i=0}^{k-1} n_i} \right) = \mathbb{P}(D_k) = p_1 \cdots p_k. \tag{5.16}$$

3.3 Estimation of γ

As already mentioned at the beginning of Section 3, the final estimator for γ is the product $\widehat{\gamma} := \widehat{\alpha}_A \cdot \widehat{T}_B$. In the description the MLS algorithm, in Step 1, we tacitly assumed that we can sample the recurrency cycle origin X_1^A. As this is typically not the case, we sample X_1^A approximately, in the following way. During the estimation of α_A with the batch-means method (5.7) we store each inwards crossing to the set A and we bootstrap these states in Step 1 of Algorithm 5.1. We thus end up with the following algorithm for estimating the rare-event probability γ, as defined in (5.1).

Algorithm 5.2 (Recurrent Multilevel Splitting)**.**

1. *Choose a recurrency set A satisfying the assumptions of Theorem* 5.1 *and an importance function $H : \mathbb{R}^d \to [0, 1]$.*

2. *Estimate α_A using the batch-means method* (5.7)*, and return $\widehat{\alpha}_A$. Store the locations of the cycle origins in the set $\mathcal{S}_{\text{rec}} := \{X_1^A, X_2^A, \ldots\}$.*

3. *Estimate T_B using the Multilevel Splitting algorithm (Algorithm* 5.1*); in Step* 1 *sample the origin X_0^1 uniformly from S_{rec}. The output is $\widehat{T}_B$.*

4. *The final estimator is*
$$\widehat{\gamma} := \widehat{\alpha}_A \cdot \widehat{T}_B \tag{5.17}$$

It is assumed that the set S_{rec} is 'representative enough' to make sure that resampling from S_{rec} can be interpreted as taking i.i.d. samples of X_1^A in the stationary regime. Under this assumption, the estimators $\widehat{\alpha}_A$, $\widehat{T}_B$ are independent and the variance of $\widehat{\gamma}$ can be inferred using the sample variance of $\widehat{\alpha}_A$ and $\widehat{T}_B$. However, in our numerical experiments in Section 5 we do not assume this independence to get an estimate of the variance. Instead we run Algorithm 5.2 multiple times, resulting in multiple estimates $\widehat{\gamma}$ from which we obtain a reliable estimate for the variance of $\widehat{\gamma}$. For implementation details, see Section 5.1.

4 Choice of Parameters

In a rare-event setting, both the expectation and the variance of an estimator are very small, so that the variance itself is not a meaningful measure of accuracy.

Instead, it makes sense to look at its value relative to the expectation, i.e., the *Relative Error* (RE):

$$\mathrm{RE}^2(\widehat{\gamma}) := \mathbb{E}(\widehat{\gamma} - \gamma)^2/\gamma^2.$$

An estimator with a lower relative error is not necessarily preferred; a more meaningful criterion involves the corresponding total computational time (or: *workload*), which we denote $W(\widehat{\gamma})$; see the beginning of Section 5.1 for more details. In the following section we consider a setting, in which we can derive optimal parameters of the MLS estimator by minimizing the workload under a constraint on the relative error (i.e., $\mathrm{RE}^2(\widehat{\gamma}) \leq \rho$ for a given accuracy $\rho > 0$).

4.1 Simplified Setting

Due to possible dependencies between the number of successes $r_1, \ldots, r_m$, there is no tractable general expression for the variance of MLS estimator. A typical assumption made in the literature is to assume some sort of independence between them, and to study the variance afterwards. With τ_k, D_k defined as in (5.13) and R_+ as defined in (5.11), we assume

(I) for all $k \in \{1, \ldots, m\}$,

$$\mathbb{P}(D_k \mid D_{k-1}, X_{\tau_{k-1}}) \equiv \mathbb{P}(D_k \mid D_{k-1}) = p_k$$

(II) for all X_{τ_m},

$$\big(R_1 \mid R_1 > 0, X_{\tau_m}\big) \stackrel{\mathrm{d}}{=} (R_1 \mid R_1 > 0) =: R_+$$

Assumption (4.1) has been proposed in [5]. It states that the probability of reaching the k-th importance level, given the $(k-1)$-st level has been reached, is constant over all possible entrance states. Assumption (4.1) states that the time spent in the rare set B within a cycle, conditioned on the set B has been reached, does not depend on the position of the entrance state to B. In principle, we have the possibility to choose the set A and the importance function $H(\cdot)$ such that Assumption (4.1) is satisfied; see the discussion in Section 4.2. Whether Assumption (4.1) holds or not is effectively problem specific, in the sense that we do not have control over it due to the fact that the set B is given. We argue

that for a large class of problems there exists a most likely point of entry X_{τ_B} to B, which implies (4.1) approximately. We emphasize that Assumptions (4.1-4.1) are not required for the RMS algorithm to work, but if they are fulfilled, optimality results can be derived. Under (4.1-4.1) we find the squared relative error of $\widehat{T}_B$:

$$\mathrm{RE}^2(\widehat{T}_B) = \sum_{k=1}^{m} \frac{(1-p_k)/p_k}{\prod_{j=0}^{k-1} n_j p_j} + \frac{\mathrm{RE}^2(R_+)}{\prod_{j=0}^{m} n_j p_j}. \tag{5.18}$$

We derive (5.18) in Appendix A. Following the approach of Amrein and Künsch [5], in Appendix B we derive the optimal parameters $m, p_1, \ldots, p_m, n_0, \ldots, n_m$ for the MLS algorithm; here, optimality refers to the property that the expected computational time is minimized under the constraint for the relative error $\mathrm{RE}^2(\widehat{T}_B) \leq \rho$ for a given accuracy $\rho > 0$. It is worth noting that the optimal number of thresholds m is roughly equal to $|\log p_B|$ with conditional probabilities p_k all equal to approximately 0.2. What is more, the optimal solution satisfies $n_k p_{k+1} = 1$ for $k \in \{1, \ldots, m-1\}$, so we can choose $n_k = 5$. This so-called *balanced growth* (see [51]) ensures that, on average, n_0 paths are sampled in each stage of the algorithm (with an exception of the last stage, which corresponds to the estimation of R_+). The optimal workload reads

$$W(\widehat{T}_B) = \frac{1}{q}\left(\frac{c\,|\log p_B|}{\sqrt{2c-1}} + \mathrm{RE}(R_+)\right)^2 \tag{5.19}$$

with a constant c defined as below display (5.38). As already mentioned, a rigorous derivation of this result can be found in Appendix B, and the exact values of the optimal parameters $m, p_1, \ldots, p_m, n_0, \ldots, n_m$ in Eq. (5.38). In all our numerical experiments in Section 5, we spend an initial portion of computational time on a rough estimation of p_B and $\mathrm{RE}(R_+)$ in order to find a sufficiently accurate approximation of the optimal parameters. See Section 5.1 for a more detailed account of the implementation details.

The optimal workload in (5.19) is proportional to $(\log p_B)^2$, which offers a huge gain in efficiency, compared with the Monte Carlo method (5.42) (whose workload is inversely proportional to p_B). We derive efficiency results in Appendix C; in particular, Theorem 5.3 proves that RMS is logarithmically efficient under specific assumptions.

4.2 Choice of Recurrency Set and Importance Function

In Section 4.1 we have seen that under Assumptions (4.1-4.1), the MLS method is particularly efficient. As already mentioned, the level up to which Assumption (4.1) is fulfilled depends on both the choice of the recurrency set and the importance function; we thus aim to choose A and $H(\cdot)$ in such a way that (4.1) is approximately satisfied. At the same time, we would like to choose A so as to maximize α_A, so that the batch-means estimator $\widehat{\alpha}_A$ (as defined in (5.7)) is computationally efficient as well. These two requirements are often conflicting and one must in the end strike a proper balance between them.

For each k, Assumption (4.1) concerns the choices of both A and $H(\cdot)$. However, it implies a property that relates to the choice of A only, namely, the probability of reaching set B within a recurrency cycle is independent of the initial point:

$$\mathbb{P}(\tau_B < \tau_A^{\text{in}} \mid X_1^A) \equiv p_B.$$

Thus, Assumption (4.1) implies that

$$X_1^A \stackrel{\text{d}}{=} \left(X_1^A \mid R_1 > 0\right) =: X_+^A; \tag{5.20}$$

informally, there is independence between the origin of the cycle on one hand, and the random variable $1_{\{R_1>0\}}$ (indicating whether set B has been reached within a cycle) on the other hand. Intuitively, the smaller the set A is, the more closely (5.20) is satisfied but also, the smaller α_A is. In particular, (5.20) trivially holds when A consists of one point only, but then $\alpha_A = 0$. In Section 5.2.3 we give an example of a setting in which (5.20) is violated, but one can imagine that in many situations (5.20) 'roughly holds'. Thus, for practical purposes, it is desirable that *the set A maximizes α_A while it also approximately satisfies* (5.20). In full generality, it is not an easy task to fulfill both aims.

A poorly chosen importance function will lead the split particles into uninteresting regions, or it will force the paths to hit the rare set in an unlikely fashion. This potentially leads to low efficiency of the MLS algorithm. Given that we have already chosen a set A satisfying (5.20), there exists an importance function guaranteeing (4.1) to be satisfied:

$$H(x) := \mathbb{P}_x(\tau_B \leq \tau_A^{\text{in}}),$$

Of course this insight is of theoretical value only: if we knew the quantity on the right hand side, then we would not even have to use the MLS algorithm. However, also

$$H_g(x) := g(\mathbb{P}_x(\tau_B \leq \tau_A^{\text{in}})),$$

with $g : [0,1] \to \mathbb{R}$ any increasing function, satisfies (4.1). This already gives a helpful guideline for the choice of H. Namely, *the states from which it is more likely to visit B before returning to A should have larger importance.* When an approximation or asymptotic behavior of $\mathbb{P}_x(\tau_B \leq \tau_A^{\text{in}})$ is available it might be useful to use it as an importance function. In Dean and Dupuis [33] a large-deviations based approach to the choice of importance function is discussed.

Sometimes, a so-called *distance-based* importance function can be a good choice. This function is basically

$$H(x) := \text{dist}(x, B) = \inf\{\|x - a\| : a \in B\},$$

normalized in such a way that $H(x) = 1$ iff $x \in B$ and $H(x) = 0$ for $x \in A$. This importance function can be a good choice for systems whose paths conditioned on $\{\tau_B < \tau_A^{\text{in}}\}$ are effectively gradually driven towards B. In contrast, distance-based importance function will be a poor choice for systems for which it is most likely to reach rare set B by first getting away from it. In Section 5 we include examples of problems for which a distance-based importance function is a good choice, but also one in which it does not work well.

In some cases we may have already chosen a particular *shape* of the set A (e.g. an ellipsoid, half-space, or multidimensional cube) which can be parametrized by a single parameter $\ell \in \mathbb{R}$. Even better, if we have already chosen an importance function, then a level set

$$A(\ell) = \{x \in \mathbb{R}^d : H(x) \leq \ell\}$$

could be a good choice. In any case, we should choose ℓ to maximize $\alpha_{A(\ell)}$. We propose to use a crude estimator to find ℓ^*: we find a maximizer of $\alpha_{A(\ell)}$ by putting

$$\widehat{\ell^*} := \arg\max \left\{ \sum_{n=0}^{N} 1_{\{X_n \notin A(\ell), X_{n+1} \in A(\ell)\}} \right\}. \tag{5.21}$$

Quantile validation. While it is not clear in general how to choose A such that it satisfies (5.20), one can statistically test whether (5.20) holds after the choice of A has been made. We now propose one particular method to do so that can be used in combination with the RMS algorithm. In Step 2 of Algorithm 5.2 calculate and store the *maximum importance attained within cycles*, i.e.,

$$H_k^{\max} := \max\{H(x) : x \in \mathcal{C}_k\},$$

with $\mathcal{C}_k$ as defined in (5.3). Assuming a good importance function has been chosen, the cycle origins corresponding to the highest importance should also be approximately distributed as X_+^A. This gives us means of comparing the distributions of X_1^A and X_+^A. Let N_{rec} be the total number of pairs $(X_k^A, H_k^{\max})$ obtained in Step 2 of Algorithm 5.2. Let

$$\sigma : \{1, \ldots, N_{\text{rec}}\} \to \{1, \ldots, N_{\text{rec}}\}$$

be a permutation ordering $(H_k^{\max})_{1 \geq k \geq N_{\text{rec}}}$ into a non-decreasing sequence, i.e.,

$$H_{\sigma(1)}^{\max} \leq H_{\sigma(2)}^{\max} \leq \ldots \leq H_{\sigma(N_{\text{rec}})}^{\max}$$

Now choose a $q \in (0, 1)$ and let

$$S_{\text{rec}}^q := \left\{X_{\sigma([(1-q)N_{\text{rec}}])}^A, \ldots, X_{\sigma(N_{\text{rec}})}^A\right\} \tag{5.22}$$

That is, S_{rec}^q is a subset of S_{rec} which contains the cycle origins corresponding to the fraction q of values with highest importance. In particular $S_{\text{rec}}^1 = S_{\text{rec}}$. Then S_{rec} and S_{rec}^q (for small q) can be thought of as sets of samples from the random variables X_1^A and X_+^A, respectively. Various tests can now be performed, to compare e.g. the means or variances; alternatively QQ-plots can be made, or histograms can be compared.

5 Numerical Experiments

The aim of this section is to test the RMS method on a series of specific examples. The examples range from simple cases, where the ground truth is known, to more complicated dynamical systems, where the ground truth is unknown

and we can only compare to estimates obtained with Monte Carlo (MC) methods. In Section 5.2.3 we also carefully look into an example where the RMS method (with a naïve choice of the importance function) does not perform that well; we discuss why this was to be expected. It will be seen throughout that RMS is superior to MC in terms of the computational time needed to achieve a desired level of accuracy; in extreme cases, like in Section 5.3, the RMS method can be three orders of magnitude faster than MC (and the efficiency gain is expected to be even greater as γ decreases).

5.1 Implementation Details

As already mentioned in Section 4, the relative error of an estimator is not always a meaningful measure of its performance, as it does not take the workload into account. We therefore compare RMS with MC using the ratio of *work normalized squared relative errors*; see e.g. [70]. In particular, we define

$$\mathrm{Eff}(\widehat{\gamma}) = \frac{W(\widehat{\gamma}^{\mathrm{MC}})}{W(\widehat{\gamma})} \cdot \frac{\mathrm{RE}^2(\widehat{\gamma}^{\mathrm{MC}})}{\mathrm{RE}^2(\widehat{\gamma})}. \tag{5.23}$$

This value can be interpreted as the ratio of the computational cost of MC to the cost of RMS when both methods reach the same accuracy (same relative error). Clearly, the larger $\mathrm{Eff}(\widehat{\gamma})$ is, the more efficient the RMS method is in comparison with Monte Carlo.

In each of our experiments, the underlying Markov chain $(X_n)_{n\in\mathbb{N}}$ represents the numerical solution to a d-dimensional Stochastic Differential Equation (SDE) using an explicit Euler scheme, with time step $h > 0$; see e.g. [69]. We remark that the time discretization potentially has a significant effect on a the underlying value of γ, especially in the rare-event setting; see the recent systematic study [17]. However, in the context of this article we only focus on discrete recursions that arise from numerical time integration schemes. For these recursions we compare RMS with the corresponding Monte Carlo results; we do not aim at studying the behavior as $h \downarrow 0$.

Notice that our method relies on properties of discrete-time processes, in particular in the definition of the recurrency cycles. More specifically, in the corresponding continuous-time model recurrency cycles are ill-defined, as a set

may be entered and left infinitely often in a time interval of finite length. This feature could potentially lead to computational issues when working with a small time step h. However, one can easily circumvent the problem and still integrate the process with arbitrarily small h_0 but store values every $h > h_0$. Note that the discretization error depends only on h_0 (and not h), since h_0 determines the stationary distribution. In fact, this is what we do in Section 5.3, where the process is integrated with $h_0 = 10^{-4}$ but it is stored only every $h = 10^{-2}$.

In each experiment the rare event B is a half-space parametrized by $u \in \mathbb{R}$:

$$B(u) = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_1 \geq u\}.$$

In other words, the probability under consideration corresponds to the the first dimension attaining high values in stationarity:

$$\gamma(u) := \mathbb{P}_\mu(X_0 \in B(u)) \tag{5.24}$$

for large u. Furthermore, in each experiment we choose the recurrency set A to be a half-space parametrized by ℓ (where the value of ℓ is chosen depending on the particular experiment):

$$A(\ell) = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_1 \leq \ell\}. \tag{5.25}$$

We use a distance-based importance function, i.e.,

$$H(x_1, \ldots, x_d) = \begin{cases} 0, & x_1 \leq 0 \\ x_1/u, & x_1 \in (0, u) \\ 1, & x_1 \geq u \end{cases} \tag{5.26}$$

We now provide more details on our implementation of Algorithm 5.2. In Step 2, we estimate α_A using the method of batch means as in (5.7); the number of iterations of the Markov chain N is chosen such that S_{rec} consists of roughly 10^4 inwards crossings of A. In Step 3, we want to choose parameters $m, n_0, \ldots, n_m, \ell_1, \ldots, \ell_m$ for the Multilevel Splitting in such a way that the workload is minimized and the resulting estimator satisfies

$$\mathrm{RE}(\widehat{T}_B) = 5 \cdot 10^{-3}. \tag{5.27}$$

We run a pilot MLS with many intermediate thresholds ($m = 20$). The pilot gives us rough estimates of p_B, T_B and $\mathrm{RE}(R_+)$. We put the number of thresholds m and splitting factors $n_0, \ldots, n_m$ as in (5.38); we emphasize that the optimal n_0 is also determined by the desired squared relative error ρ. We find the intermediate thresholds $\ell_1, \ldots, \ell_m$ following the log-linear interpolation approach from Wadman et al. [96]. Assuming (4.1-4.1) are satisfied, the MLS method with these parameters should give the desired relative error, as in (5.27). We note that in the pilot we use the variant of MLS called 'Fixed Number of Successes' developed by Amrein and Künsch [5].

The final estimator $\widehat{\gamma}$ is the mean of $N = 100$ independent replicas $\widehat{\gamma}^{(1)}, \ldots, \widehat{\gamma}^{(N)}$ of the RMS estimator (5.17) with parameters as discussed above; i.e.

$$\widehat{\gamma} := \frac{1}{N} \sum_{i=1}^{N} \widehat{\gamma}^{(i)}$$

This additional 'Monte Carlo wrapper' around the RMS method enables us to approximate the relative error $\mathrm{RE}(\widehat{\gamma})$ with

$$\mathrm{RE}^2(\widehat{\gamma}) \approx \frac{1}{N-1} \sum_{i=1}^{N} \left(\frac{\widehat{\gamma}^{(i)}}{\widehat{\gamma}} - 1 \right)^2, \tag{5.28}$$

and we can approximate $\mathrm{RE}(\widehat{\alpha}_A)$ and $\mathrm{RE}(\widehat{T}_B)$ in a similar way. For each experiment we present a table with results corresponding to multiple values of the threshold u. Each table displays the final estimator $\widehat{\gamma}$ as well as its estimate for $\mathrm{RE}(\widehat{\gamma})$, as in (5.28), and $\mathrm{Eff}(\widehat{\gamma})$, as in (5.23) based on the run of an MC estimator $\widehat{\gamma}^{\mathrm{MC}}$.

Various checks can be done in order to assess the reliability of the estimator $\widehat{\gamma}$. In each table we additionally give the estimate for $\mathrm{RE}(\widehat{T}_B)$; if it matches the desired relative error, i.e. $\mathrm{RE}(\widehat{T}_B) \approx 5 \cdot 10^{-3}$, then this is an indication that Assumptions (4.1-4.1) are satisfied. When $\mathrm{RE}(\widehat{T}_B)$ is larger than desired, it might be a result of poorly chosen intermediate thresholds $\ell_1, \ldots, \ell_m$; we propose to verify, after the algorithm has been executed, whether the estimates for all the intermediate probabilities $p_1, \ldots, p_m$ roughly equal the optimal $p_{\mathrm{opt}} \approx 0.20$. If this is the case and we still get a particularly large $\mathrm{RE}(\widehat{T}_B)$, this is an indication

that either the recurrency set or the importance function have not been properly chosen. In case of violation of the former, in Section 4.2 we proposed a test for the appropriateness of the choice of the set A. Additional verification can be performed to assess whether resampling from the set S_{rec} obtained in Step 2 of the RMS algorithm is a good approximation of taking i.i.d. samples of X_1^A. This implies that $\widehat{\alpha}_A$ and $\widehat{T}_B$ are independent, but if they are independent then necessarily

$$\text{RE}^2(\widehat{\gamma}) = \text{RE}^2(\widehat{\alpha}_A) + \text{RE}^2(\widehat{T}_B)\,. \tag{5.29}$$

Thus, if (5.29) is not approximately satisfied, it is an indication that S_{rec} does not represent the distribution of X_1^A well. We emphasize that the relative error of $\widehat{\gamma}$ presented in the tables is calculated as in (5.28).

5.2 Ornstein-Uhlenbeck Process

Let $(X_t)_{t\geq 0}$ be a d-dimensional Ornstein-Uhlenbeck process (d-dim OU), i.e., a process taking values in $\mathbb{R}^d$ solving the SDE

$$\mathrm{d}X_t = -QX_t\mathrm{d}t + \mathrm{d}W_t \tag{5.30}$$

with $Q \in \mathbb{R}^{d\times d}$ and $(W_t)_{t\geq 0}$ denoting a standard d-dimensional Wiener process. Applying the explicit Euler numerical scheme to (5.30), with time step $h > 0$ yields

$$X_{n+1} = (I - Qh)X_n + Z_n, \tag{5.31}$$

with I the d-dimensional identity matrix I, and $Z_1, Z_2, \ldots$ i.i.d. d-dimensional standard normal random variables. It is known [87] that the stationary distribution μ of (5.31) exists if there exists a positive-definite matrix $M = (M_{ij})_{i,j\in\mathbb{N}}$ solving

$$M = (I - Qh)M(I - Qh)^\top + hI; \tag{5.32}$$

then the stationary distribution μ is d-dimensional centered normal with covariance matrix M. The rare event of our interest is the exceedance of a high threshold in the first dimension under the stationary distribution (of the discrete-time Markov chain in (5.31)), as in (5.24). Eq. (5.32) is a well-known Sylvester equation and its solution M can be found numerically, so that $\gamma(u)$ can be evaluated

as

$$\gamma(u) = \Phi(-u/\sqrt{M_{11}}), \tag{5.33}$$

with $\Phi(\cdot)$ the standard normal cdf. Knowing the ground truth $\gamma(u)$ gives us means to determine how accurate the RMS estimator $\widehat{\gamma}$ is.

In the following three subsections we study the OU process with different sets of parameters but with the same choice of the recurrency set and importance function, as in (5.25) and (5.26). First, we study the simplest case of a one-dimensional OU process. This is an 'ideal' example in the sense that Assumptions (4.1-4.1) are (approximately) satisfied. Second, we study a multidimensional OU process; while the simplifying assumptions do not seem to be satisfied, they are 'close enough' for the RMS method to give satisfactory results. The third case describes a two-dimensional OU process with the matrix Q chosen such that Assumptions (4.1-4.1) are not satisfied for our choice of the recurrency set and the importance function.

5.2.1 1-dim OU

In this experiment we put $d = 1$, $Q = 1$, $h = 0.01$. The recurrency set $A(\ell)$ and importance function $H(\cdot)$ are as in (5.25) with $\ell = 0$ and (5.26) respectively.

If we would study the stationary distribution of the original SDE driven by (5.30) (rather than the time-discrete numerical solution in (5.31)), then the paths of the process would be continuous and thus $X_1^A = 0$ a.s. Moreover, because of their continuity, these paths must cross all intermediate states $x \in (0, u)$ before reaching B. Therefore $x \mapsto \mathbb{P}_x(\tau_B < \tau_A^{\text{in}})$ is an increasing function, implying that the distance-based importance function satisfies (4.1) in the continuous-time case. By similar arguments, $X_{\tau_B} = u$ a.s., and hence (4.1) is satisfied as well in that case.

The Markov chain driven by (5.31) is a discrete-time approximation of (5.30), so the assumptions will not be satisfied *exactly*. In particular, we note that for any time step $h > 0$, the support of X_{τ_B} is the entire halfline $[u, \infty)$ because in principle the process can exceed the threshold u by any positive value upon the first entry. This shows that Assumption (4.1) is not satisfied. An analogous argument can be used to show that Assumption (4.1) is not satisfied either. Nonetheless, for a small time step $h > 0$, extreme overshooting upon

$\gamma(u)$	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
$\widehat{\gamma}$	$9.94 \cdot 10^{-4}$	$9.93 \cdot 10^{-5}$	$9.96 \cdot 10^{-6}$	$9.96 \cdot 10^{-7}$	$9.96 \cdot 10^{-8}$
$\mathrm{RE}(\widehat{\gamma})$	3.95e-03	5.45e-03	6.53e-03	6.31e-03	5.49e-03
$\mathrm{Eff}(\widehat{\gamma})$	4.1	8.9	45.2	378.9	1836.2
$\mathrm{RE}(\widehat{T}_B)$	3.90e-03	4.99e-03	6.42e-03	6.30e-03	5.32e-03

Table 5.1: RMS algorithm for an 1-dim OU process. Parameters: $Q = 1$, $A = \{x_1 \leq 0\}$, $B = \{x_1 \geq u\}$; u has been chosen using (5.33) to match the values of γ in the first row. We have $\widehat{\alpha}_A = 0.0225$ and $\mathrm{RE}(\widehat{\alpha}_A) = 1.66 \cdot 10^{-3}$.

the first entry (i.e., X_{τ_B} being significantly larger than u, or X_{τ_k} significantly larger than $\ell_k u$) is very unlikely. We conclude that the assumptions are satisfied approximately.

Since the value of $\gamma(u)$ can be evaluated using (5.33), we chose the thresholds u to match the desired value of $\gamma(u)$, as in Table 5.1. The results show that $\mathrm{RE}(\widehat{T}_B) \approx 5 \cdot 10^{-3}$, as desired in (5.27); this is a good indication that Assumptions (4.1-4.1) are satisfied. Also, the relative error calculated under the independence assumption via (5.29) matches the estimated $\mathrm{RE}(\widehat{\gamma})$.

Conclusions. In this setting the RMS algorithm is very efficient, as compared with MC. The numerical results agree very well with the theoretical outcomes, confirming our observation that Assumptions (4.1-4.1) are approximately satisfied.

5.2.2 10-dim OU, Q with real eigenvalues

In this experiment we put $d = 10$, $h = 0.01$. The matrix $Q = (Q_{ij})_{i,j \in \{1,\ldots,d\}}$ is randomly generated such that all its eigenvalues are real. The recurrency set $A(\ell)$ and importance function $H(\cdot)$ are as in (5.25) with $\ell = 0$ and (5.26) respectively.

In Fig. 5.2 we plot four randomly chosen recurrency cycles, projected onto the first and second dimension, which have reached the rare event B. These conditional paths seem to follow a linear pattern; similar behavior is seen in other projections (not shown). This indicates that attaining high values in the first dimension is coupled with attaining high values in the second dimension (and similar statements can be made about other dimensions). Therefore, the distance-based importance function is not expected to satisfy (4.1), as it does

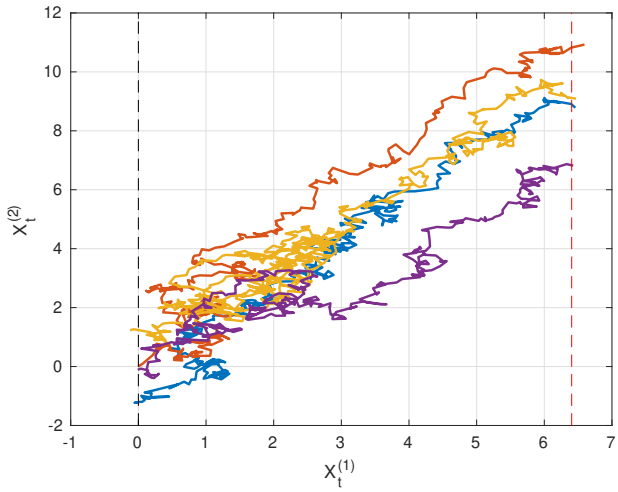


Figure 5.2: 10-dim OU process. Four random realizations of recurrency cycles conditioned on reaching the rare set. The cycles have been plotted until the first hitting time of B. Parameters: $A = \{x_1 \leq 0\}$, $B = \{x_1 \geq u\}$ with $u \approx 6.4$ such that and $\gamma(u) = 10^{-6}$.

not take this behavior into account; an ideal importance function should give larger importance to states which attain *simultaneously* high values in the first and second dimension. While the distance-based importance function is not the most appropriate choice, it is still expected to give satisfactory results, as it drives the paths gradually towards the rare event.

The results of the RMS algorithm are presented in Table 5.2. It can be seen that the values of $\mathrm{RE}(\widehat{T}_B)$ do not exactly match the desired value $5 \cdot 10^{-3}$ in (5.27), which in view of the earlier discussion is not surprising, as we did not expect Assumptions (4.1-4.1) to hold. However, the estimates $\widehat{\gamma}$ are still very accurate, and the efficiency is still excellent (relative to the MC method).

Conclusions. This experiment shows that the RMS algorithm can be effectively implemented in a multidimensional setting, even when Assumptions (4.1-4.1) are violated. This underscores the robustness of the distance-based importance function.

$\gamma(u)$	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
$\widehat{\gamma}$	$1.00 \cdot 10^{-3}$	$9.95 \cdot 10^{-5}$	$1.02 \cdot 10^{-5}$	$9.92 \cdot 10^{-7}$	$1.00 \cdot 10^{-7}$
$\mathrm{RE}(\widehat{\gamma})$	7.84e-03	$1.03{\cdot}10^{-2}$	$1.35{\cdot}10^{-2}$	$1.12{\cdot}10^{-2}$	$1.49{\cdot}10^{-2}$
$\mathrm{Eff}(\widehat{\gamma})$	0.8	2.4	9.3	34.9	180.5
$\mathrm{RE}(\widehat{T}_B)$	7.87e-03	$1.02{\cdot}10^{-2}$	$1.35{\cdot}10^{-2}$	$1.12{\cdot}10^{-2}$	$1.49{\cdot}10^{-2}$

Table 5.2: RMS algorithm for a 10-dim OU process. Parameters: Q is a matrix with only real eigenvalues, $A = \{x_1 \leq 0\}$, $B = \{x_1 \geq u\}$; u has been chosen using (5.33) to match the values of γ in the first row. We have $\widehat{\alpha}_A = 0.0124$, $\mathrm{RE}(\widehat{\alpha}_A) = 2.46 \cdot 10^{-3}$.

5.2.3 2-dim OU, Q with complex eigenvalues

In this experiment we put $d = 2$, $h = 0.01$. We choose Q to have non-real eigenvalues: for a positive θ,

$$Q(\theta) = \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix}. \tag{5.34}$$

The drift generates a rotating (or spiraling) motion of the paths, with the speed of rotation increasing as θ increases. We compare the efficiency of the RMS method for increasing values of θ. The recurrency set $A(\ell)$ and importance function $H(\cdot)$ are as in (5.25) with $\ell = 0$ and (5.26) respectively.

The results are presented in Table 5.3. We see that for most values of θ, RMS outperforms the Monte Carlo, but the larger θ is, the lower the efficiency ratio $\mathrm{Eff}(\widehat{\gamma})$ becomes. At the same time, as θ grows, the value of $\mathrm{RE}(\widehat{T}_B)$ deviates more and more from the desired target $5 \cdot 10^{-3}$, as in (5.27). This indicates a violation of Assumptions (4.1-4.1). We note that the estimates $\widehat{\gamma}$ are quite accurate nonetheless, with a minor relative error of a few percent visible for larger values of θ.

In Fig. 5.3 we plot five random recurrency cycles conditioned on reaching the rare set B. We see that the paths do not gradually drift towards B, but rather first move far away from B, due to the drift-induced rotation. This hints that the distance-based importance function might be a poor choice. Fig. 5.4 shows that even property (5.20) seems to be violated. In this figure we compare the histograms of S_{rec} and S^q_{rec} in order to compare the distributions of X^A_1 and X^A_+ (see the discussion Section 4.2). The figure shows that X^A_+ has more

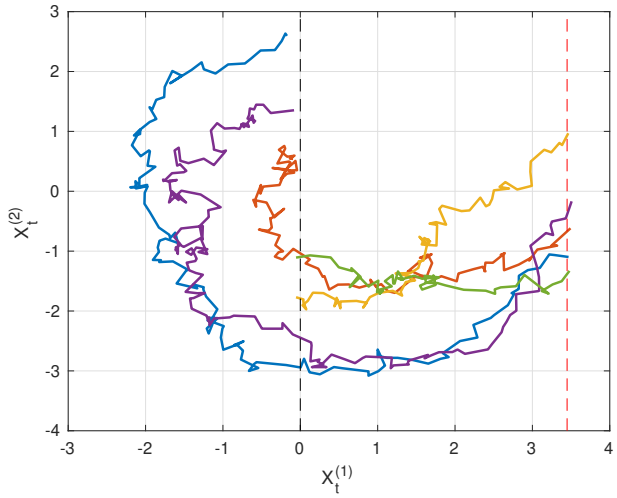


Figure 5.3: 2-dim OU process. Five random realizations of recurrency cycles conditioned on reaching the rare set. The cycles have been plotted until the first hitting time of B. Parameters: $A = \{x_1 \leq 0\}$, $\theta = 3$, $B = \{x_1 \geq u\}$ with $u \approx 3.4$ such that $\gamma(u) = 10^{-6}$.

probability mass in the sets $\{x_2 \leq -1\}$ or $\{x_2 \geq 1\}$ than X_1^A.

Conclusions. When Q has non-real eigenvalues, the naïve choice of the recurrency set and the distance-based importance function (i.e., (5.25) and (5.26)) seems inadequate and leads to a relative error higher than expected. This underscores the fact that one has to be careful with the choice of A and $H(\cdot)$ and verify whether Assumptions (4.1-4.1) are satisfied; this can be done e.g. by the means described in Section 4.2. Despite violation of Assumptions (4.1-4.1), RMS still gives rather accurate estimates of γ, and outperforms Monte Carlo for small θ.

5.3 Franzke (2012) Stochastic Climate Model

As our final example, we consider the low-order stochastic climate model presented by Franzke [50]. This is a 4-dimensional SDE with certain key features that are also present in more complex climate models, including nonlinear (quadratic) drift terms that are energy-conserving. We refer to Franzke [50] for a more detailed discussion of the physical interpretation of this model.

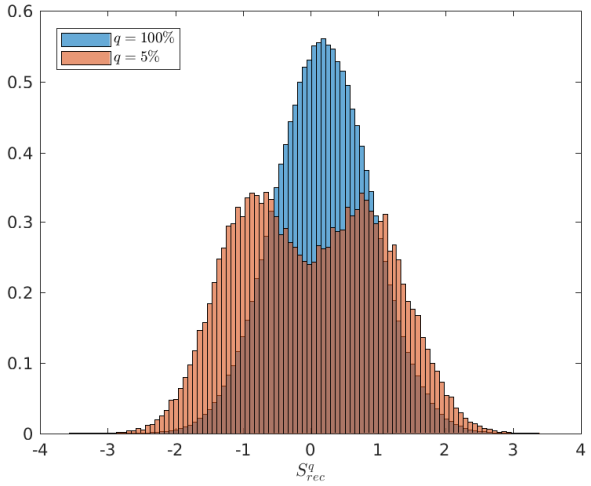


Figure 5.4: 2-dim OU process, $\theta = 3$. Marginal histograms of S^q_{rec} projected onto the second dimension. The histograms have been normalized to a probability density function.

θ	0.5	1	1.5	2	3
$\widehat{\gamma}$	$9.91 \cdot 10^{-7}$	$1.00 \cdot 10^{-6}$	$1.00 \cdot 10^{-6}$	$9.73 \cdot 10^{-7}$	$9.60 \cdot 10^{-7}$
$\text{RE}(\widehat{\gamma})$	$8.20{\cdot}10^{-3}$	$1.05{\cdot}10^{-2}$	$2.34{\cdot}10^{-2}$	$2.66{\cdot}10^{-2}$	$4.01{\cdot}10^{-2}$
$\text{Eff}(\widehat{\gamma})$	31.9	27.9	7.1	5.8	1.0
$\text{RE}(\widehat{T}_B)$	$7.63{\cdot}10^{-3}$	$1.05{\cdot}10^{-2}$	$2.37{\cdot}10^{-2}$	$2.67{\cdot}10^{-2}$	$4.01{\cdot}10^{-2}$

Table 5.3: RMS algorithm applied to 2-dim OU process. Parameters: $Q(\theta)$ as in (5.34), $A = \{x_1 \leq 0\}$, $B = \{x_1 \geq u\}$; u has been chosen depending on θ such that in every case $\gamma(u) = 10^{-6}$.

The model is given by the following set of SDEs. It uses a standard, two-dimensional Wiener process $(W_t^{(1)}, W_t^{(2)})$. We write $x_i := X_t^{(i)}$, $y_i := Y_t^{(i)}$ and $W_i := W_t^{(i)}$ to simplify notation. We consider the system

$$\begin{aligned}
\mathrm{d}x_1 &= \mu\big(- x_2(L_{12} + a_1x_1 + a_2x_2) + d_1x_1 + F_1 \\
&\qquad + L_{13}y_1 + B^1_{123}x_2y_1 + (B^2_{131} + B^2_{113})x_1y_1\big)\mathrm{d}t \\
\mathrm{d}x_2 &= \mu\big(+ x_1(L_{21} + a_1x_1 + a_2x_2) + d_2x_2 + F_2 \\
&\qquad + L_{24}y_2 + B^1_{213}x_1y_1 + (B^3_{242} + B^3_{224})x_2y_2\big)\mathrm{d}t \\
\mathrm{d}y_1 &= \mu\big(- L_{13}x_1 + B^1_{312}x_1x_2 + B^2_{311}x_1^2 + F_3 - \tfrac{\gamma_1}{\varepsilon}y_1\big)\mathrm{d}t + \tfrac{\sigma_1}{\sqrt{\varepsilon}}\mathrm{d}W_1
\end{aligned}$$

$$\mathrm{d}y_2 = \mu\big(- L_{24}x_2 + B^3_{422}x_2x_2 + F_4 - \tfrac{\gamma_2}{\varepsilon}y_2\big)\mathrm{d}t + \tfrac{\sigma_2}{\sqrt{\varepsilon}}\mathrm{d}W_2$$

When the parameter ε is set to a small value, a separation of timescales is created between the variables x_1, x_2 (slow) and y_1, y_2 (fast). The main interest is in the behavior of the slow variables x_1, x_2.

The parameters we use match those used in Franzke [50]. This means that we set $\mu = 1$, the B-coefficients are given by $B^1_{123} = 4$, $B^1_{213} = 4$, $B^1_{312} = -8$, $B^2_{131} = 0.25$, $B^2_{113} = 0.25$, $B^2_{311} = -0.5$, $B^3_{242} = -0.3$, $B^3_{224} = -0.4$, $B^3_{422} = 0.7$, the L-coefficients by $L_{13} = -L_{24} = -0.2$, and the other parameters by $\omega = 1$, $a_1 = 1$, $a_2 = -1$, $d_1 = -0.2$, $d_2 = -0.1$, $\gamma_1 = \gamma_2 = 1$, $\sigma_1 = 3$, $\sigma_2 = 1$. In addition we put $L_{12} = -L_{21} = 1, \varepsilon = 0.2$. The forcing vector (F_1, F_2, F_3, F_4) is given by $(-0.25, 0, 0, 0)$.

Since this process is non-standard, in order to build intuition, we first generated a contour plot of the estimated stationary density of (x_1, x_2); see Fig. 5.5. The process turns out to randomly switch between two modes: one mode with $x_1 \leq x_2$ and a second mode with $x_1 \geq x_2$. The estimated density function in Fig. 5.5 shows that the process is more likely to be in the second mode.

We use the explicit Euler scheme with $h_0 = 10^{-4}$ but we store the values of the process every $h = 0.01$. The small integration time step h_0 is needed for numerical stability. Similar to the previous examples, the rare event we study is the exceedance of a high threshold by x_1 under the stationary distribution, cf. (5.24). We choose the recurrency set $A(\ell)$ as in (5.25) with $\ell^* = 7.9$ suggested by the algorithm (5.21). The importance function $H(\cdot)$ is as in (5.26).

The results of the RMS method are outstanding, see Tab. 5.4. For $u = 18.5$, when $\gamma(u) \approx 10^{-7}$, we find $\mathrm{Eff}(\widehat{\gamma}) \approx 1522$. In other words, the RMS algorithm is more than 1500 times faster than MC. The values of $\mathrm{RE}(\widehat{T}_B)$ match the desired $5 \cdot 10^{-3}$ (see (5.27)) very closely even for very high thresholds, indicating that Assumptions (4.1-4.1) are satisfied. A random realization of a cycle reaching the rare event, shown in Fig. 5.6, is yet another indication that the distance-based importance function is a good choice, as the path seems to gradually drift towards the rare event.

Conclusions. This example shows a successful application of the RMS algorithm to a multidimensional nonlinear stochastic-dynamical model with charac-

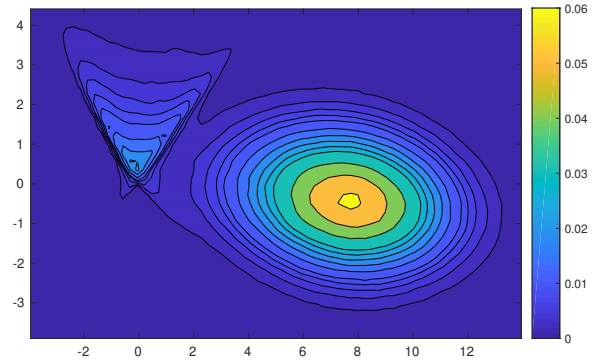


Figure 5.5: Contour plot of the marginal stationary density of slow variates (x_1, x_2) of the model of Franzke [50].

u	14	15	16	17.5	18.5
$\widehat{\gamma}$	$1.08 \cdot 10^{-3}$	$1.99 \cdot 10^{-4}$	$3.00 \cdot 10^{-5}$	$1.14 \cdot 10^{-6}$	$9.78 \cdot 10^{-8}$
$\mathrm{RE}(\widehat{\gamma})$	$6.1{\cdot}10^{-3}$	$7.2{\cdot}10^{-3}$	$7.4{\cdot}10^{-3}$	$7.4{\cdot}10^{-3}$	$5.8{\cdot}10^{-3}$
$\widehat{\gamma}^{\mathrm{MC}}$	$1.08 \cdot 10^{-3}$	$2.00 \cdot 10^{-4}$	$2.98 \cdot 10^{-5}$	$1.12 \cdot 10^{-6}$	$8.85 \cdot 10^{-8}$
$\mathrm{RE}(\widehat{\gamma}^{\mathrm{MC}})$	$1.4{\cdot}10^{-3}$	$2.9{\cdot}10^{-3}$	$6.5{\cdot}10^{-3}$	$2.7{\cdot}10^{-2}$	$8.5{\cdot}10^{-2}$
$\mathrm{Eff}(\widehat{\gamma})$	1.9	8.6	32.1	269.9	1521.8
$\mathrm{RE}(\widehat{T}_B)$	$5.1{\cdot}10^{-3}$	$6.4{\cdot}10^{-3}$	$7.2{\cdot}10^{-3}$	$6.6{\cdot}10^{-3}$	$5.4{\cdot}10^{-3}$

Table 5.4: RMS algorithm applied to the model of Franzke [50]. Parameters: $A = \{x_1 \leq 7.9\}$, $B = \{x_1 > u\}$. We have $\widehat{\alpha}_A = 0.0124$, $\mathrm{RE}(\widehat{\alpha}_A) = 2.83 \cdot 10^{-3}$.

teristics of complex climate models. We find that RMS is up to three orders of magnitude faster than MC in this example, and the efficiency gain is expected to be even larger for higher thresholds u.

6 Summary

In this chapter we have proposed a new algorithm for the estimation of small steady-state probabilities $\gamma = \mu(B)$, as in (5.1), of Markov processes with continuous state space. Our approach, which we have called the Recurrent Multilevel Splitting (RMS) algorithm, is based on the alternative representation (5.6) of γ (as given in Theorem 5.1). This representation is obtained by dissecting the path of the Markov process into recurrency cycles, each cycle beginning with an inwards crossing of a set A. It allows to transform the problem of estimating

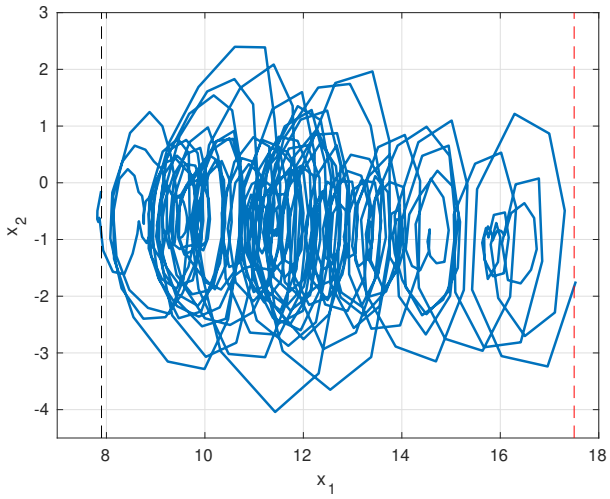


Figure 5.6: The model of Franzke [50]. A random realization of a recurrency cycle conditioned on reaching the rare set. The cycle has been plotted until the first hitting time of B. Parameters: $A = \{x_1 \leq 7.9\}$, $B = \{x_1 \geq 17.5\}$, $\gamma(17.5) \approx 1.14 \cdot 10^{-6}$.

γ essentially into the problem of estimating T_B, the expected time spent in the set B in a recurrency cycle.

In order to efficiently estimate T_B we use Multilevel Splitting (MLS), but we emphasize that other rare event simulation methods could have been used instead (such as Genealogical Particle Analysis or Importance Sampling). We have derived optimal parameters for the MLS in Appendix B, and we have shown (Theorem 5.3) that under simplifying assumptions, a suitable choice of the recurrency set A in combination with the optimal choice of the parameters leads to logarithmic efficiency of the RMS algorithm.

In Section 5, four numerical studies were presented, where we used the RMS algorithm to estimate steady state probabilities of high threshold exceedances for various SDEs discretized in time. The experiments demonstrate that RMS gives accurate results. Furthermore, they unanimously show the efficiency gain of RMS compared to Monte Carlo; in the most notable case of the [50] model (Section 5.3), RMS outperforms MC by up to three orders of magnitude.

One of the numerical experiments (Section 5.2.3) was designed to give subop-

timal results, with an SDE displaying rotating motion so that the most straightforward choices of the recurrency set and importance function (as used in the experiments) were expected to be not very suitable. Although the estimates obtained with RMS were still quite accurate, the efficiency gain of RMS compared to MC was decreasing as the rotation speed was increasing. This example showed how the choice of the recurrency set and the importance function can impact the performance of the algorithm.

In light of this example, an interesting topic for future research is the choice of the recurrency set A. As already mentioned in Section 4.2, a good choice of A should be a suitable compromise between visiting A relatively often and (5.20) being (approximately) met. We have proposed a method of optimizing $A(\ell)$ parametrized by ℓ in (5.21), and pointed out a method of testing whether A satisfies (5.20) through a quantile validation (5.22). Further development of these ideas to construct an optimal A is a challenging open research topic.

Appendix

A Technical Results

Proof of Theorem 5.1. Define a new Markov chain $Z_n := (X_{n-1}, X_n)$; it is also positive Harris with a stationary measure $\widetilde{\mu}$ satisfying, for measurable sets C_0, C_1,

$$\widetilde{\mu}(C_0, C_1) = \mathbb{P}(X_0 \in C_0, X_1 \in C_1 \mid X_0 \sim \mu).$$

We see that the stopping times S_n coincide with the times the process Z_n visits a set $\mathcal{A} := (A^c, A)$, with $A^c := \mathbb{R}^d \setminus A$. Since $\mu(A) \in (0,1)$ we have

$$\alpha_A = \mathbb{P}_\mu(X_0 \in A, X_1 \in A^c) > 0.$$

According to Meyn and Tweedie [77, Thm. 10.4.9] we have, with $\tau_{\mathcal{A}} := \inf\{n > 0 : Z_n \in \mathcal{A}\}$,

$$\widetilde{\mu}(\mathbb{R}^d, B) = \int_{\mathcal{A}} \widetilde{\mu}(\mathrm{d}x, \mathrm{d}y)\, \mathbb{E}_x \sum_{n=0}^{\tau_{\mathcal{A}}-1} 1_{\{Z_n \in (\mathbb{R}^d, B)\}}.$$

Due to $\widetilde{\mu}(\mathbb{R}^d, B) = \mu(B)$, $\{Z_n \in (\mathbb{R}^d, B)\} = \{X_n \in B\}$, and $\widetilde{\mu}(\mathcal{A}) = \alpha_A$, it follows that

$$\mu(B) = \alpha_A \cdot \mathbb{E}\Big(\sum_{n=0}^{\tau_{\mathcal{A}}-1} 1_{\{X_n \in B\}} \mid Z_0 \sim \widetilde{\mu}, Z_0 \in \mathcal{A} \Big).$$

Finally, we recognize that the conditioning above is equivalent to X_0 being distributed as an initial point of a recurrency cycle X_1^A in stationarity, so that we conclude (5.6). Similarly, one can show that $\alpha_A = (\mathbb{E}_\mu L_1)^{-1}$ by considering the expected time spent in $(\mathbb{R}^d, \mathbb{R}^d)$ within a recurrency cycle. □

Derivation of (5.18). Notice that (4.1) implies that the number of times the k-th threshold r_k is hit, is distributed as a sum of $n_{k-1}\, r_{m-1}$ *independent* Bernoulli trials, each with probability of success p_k:

$$\big(r_k \mid r_{k-1}\big) \stackrel{\mathrm{d}}{=} \mathrm{Bin}(n_{k-1} r_{k-1}, p_k); \tag{5.35}$$

here $\mathrm{Bin}(n, p)$ denotes a Binomial distribution with n trials with success probability p, with the convention that $\mathrm{Bin}(0, p) \equiv 0$. Similarly, (4.1) implies that the total time spent in the rare set is distributed as a sum of $n_m r_m$ *independent* copies from the distribution R_+:

$$\big(r_{m+1} \mid r_m\big) \stackrel{\mathrm{d}}{=} \sum_{k=1}^{n_m r_m} R_+^{(k)}, \tag{5.36}$$

where $R_+^{(1)}, R_+^{(2)}, \ldots$ are i.i.d. copies of R_+ (with the empty sum being defined as 0). Using (5.35) and the law of total variance we obtain, for $k \in \{1, \ldots m\}$,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(r_k) &= \mathbb{E}(\mathbb{V}\mathrm{ar}(r_k | r_{k-1})) + \mathbb{V}\mathrm{ar}(\mathbb{E}(r_k | r_{k-1})) \\
&= \mathbb{E}(n_{k-1} r_{k-1} p_k (1 - p_k)) + \mathbb{V}\mathrm{ar}(n_{k-1} r_{k-1} p_k) \\
&= n_{k-1} p_k (1 - p_k) \mathbb{E}(r_{k-1}) + n_{k-1}^2 p_k^2 \mathbb{V}\mathrm{ar}(r_{k-1}).
\end{aligned}$$

Similarly, using (5.36) we obtain

$$\mathbb{V}\mathrm{ar}(r_{m+1}) = n_m \mathbb{E}(r_m) \mathbb{V}\mathrm{ar}(R_+) + n_m^2 (\mathbb{E} R_+)^2 \mathbb{V}\mathrm{ar}(r_m).$$

Combining these results with (5.16) yields (5.18). □

B Derivation of Optimal Parameters

Following [5], we assume that the computational effort w_k in the k-th stage of Algorithm 5.1 (to sample a path starting from X_{τ_k} until $\min\{\tau_{k+1}, \tau_A^{\text{in}}\}$) does not depend on the entry state X_{τ_k}. Simplifying this further, we assume that w_k does not depend on k, so without loss of generality,

$$w_k \equiv 1, \quad k \in \{0, \dots, m\}. \tag{5.37}$$

A more general cost w_k can be considered for particular problems, see e.g. [72].

Let $N_k := n_k r_k$, for $k \in \{0, \dots, m\}$, be the number of paths simulated in the k-th stage of the algorithm, with $r_0 := 1$. Then the average total workload equals

$$W := \sum_{k=0}^{m} \mathbb{E} N_k,$$

and since $\mathbb{E} r_k = p_1 \cdots p_k$, cf. (5.16), we conclude

$$\mathbb{E} N_k = \prod_{j=0}^{k} n_j p_j.$$

Finally, we formulate the minimization problem

$$\begin{aligned} \text{minimize} \quad & W := \textstyle\sum_{k=0}^{m} \prod_{j=0}^{k} n_j p_j \\ \text{with respect to:} \quad & m, p_1, \dots, p_m, n_0, \dots, n_m \\ \text{subject to:} \quad & \begin{cases} \mathrm{RE}^2(\widehat{T}_B) \le \rho, \\ \prod_{k=1}^{m} p_k = p_B, \\ m \in \mathbb{N}, \\ p_k \in (0,1), \quad k \in \{1, \dots, m\}, \\ n_k \in \mathbb{N}, \quad k \in \{0, \dots, m\}. \end{cases} \end{aligned}$$

In our simplified setting, i.e., under Assumptions (4.1-4.1), we have derived a formula for the corresponding squared relative error in (5.18). We are able to solve the optimization problem above under the additional relaxation that the

n_k and m are real and positive. To this end, it is helpful to denote

$$\begin{aligned} c_k &:= \textstyle\prod_{j=0}^{k-1} n_j p_j, \quad k \in \{1, \ldots, m+1\}, \\ a_k &:= (1-p_k)/p_k, \quad k \in \{1, \ldots m\}, \\ a_{m+1} &:= \mathrm{RE}^2(R_+). \end{aligned}$$

Then we can write

$$W = \sum_{k=1}^{m+1} c_k \quad \text{and} \quad \mathrm{RE}^2(\widehat{T}_B) = \sum_{k=1}^{m+1} \frac{a_k}{c_k}.$$

We want to minimize the workload W under the constraint that

$$\mathrm{RE}^2(\widehat{T}_B) \leq \rho.$$

We do this in steps. First, we fix m and the conditional probabilities $p_1, \ldots, p_m$, so that $a_1, \ldots, a_m$ are fixed (recall that a_{m+1} is not a parameter of the algorithm). We relax the problem and let the splitting factors n_k be allowed to attain any real, positive value. This means that we wish to solve (over $c_1, \ldots, c_{m+1} > 0$)

$$\begin{aligned} &\text{minimize} && W(c_1, \ldots, c_{m+1}) := \textstyle\sum_{k=1}^{m+1} c_k \\ &\text{subject to:} && \begin{cases} g(c_1, \ldots, c_{m+1}) := \sum_{k=1}^{m+1} \frac{a_k}{c_k} \leq \rho, \\ c_k > 0, \quad k \in \{1, \ldots, m+1\}. \end{cases} \end{aligned}$$

The corresponding Karush–Kuhn–Tucker conditions are

$$\begin{cases} \nabla W + \mu \nabla g = 0, \\ \mu(g - \rho) = 0, \\ \mu \in [0, \infty). \end{cases}$$

with the gradient '∇' taken with respect to vector $(c_1, \ldots, c_{m+1})$. These are solved by

$$c_k := \frac{1}{\rho} \sqrt{a_k} \sum_{j=1}^{m+1} \sqrt{a_j},$$

with the optimal workload

$$W = \frac{1}{\rho}\Big(\sum_{j=1}^{m+1}\sqrt{a_j}\Big)^2.$$

In the next step, we keep m fixed and minimize over $a_1, \ldots, a_m$. Notice that $1 + a_k = 1/p_k$, so that our minimization problem takes the form

$$\begin{aligned} \text{minimize:} \quad & W(a_1,\ldots,a_m) := \tfrac{1}{\rho}\big(\textstyle\sum_{k=1}^{m+1}\sqrt{a_k}\big)^2 \\ \text{subject to:} \quad & \begin{cases} h(a_1,\ldots,a_m) := \prod_{k=1}^{m}(1+a_k) = p_B^{-1}, \\ a_k > 0, \quad k \in \{1,\ldots,m\}. \end{cases} \end{aligned}$$

Not surprisingly, this system is solved by

$$a_1 = \ldots = a_m = p_B^{-1/m} - 1,$$

so that the optimal intermediate probabilities coincide:

$$p_k = p_B^{1/m}, \quad k \in \{1,\ldots,m\},$$

with the optimal workload being

$$W(m) = \frac{1}{\rho}\left(m\sqrt{p_B^{-1/m}-1} + \sqrt{a_{m+1}}\right)^2.$$

The final step is finding the optimal number of thresholds m. We see that the minimizer of $W(m)$ is also a minimizer of

$$m\sqrt{\mathbb{E}(-\log(p_B)/m) - 1}.$$

Again, we relax this problem, allowing m to be any real, positive number.

Finally, the optimal parameters are:

$$\begin{aligned} m &= c\,|\log p_B| \\ p_k &= p_{\text{opt}} := \frac{2c-1}{2c} \approx 0.2032, \quad k \in \{1,\dots,m\}, \\ n_0 &= \frac{1}{\rho\sqrt{2c-1}} \cdot \left(\frac{c\,|\log p_B|}{\sqrt{2c-1}} + \text{RE}(R_+) \right), \\ n_k &= 1/p_{k+1} = 1/p_{\text{opt}}, \quad k \in \{1,\dots,m-1\}, \\ n_m &= \text{RE}(R_+) \cdot \frac{2c}{\sqrt{2c-1}}. \end{aligned} \tag{5.38}$$

with $c \approx 0.6275$ solving $\mathbb{E}(1/c) = 2c/(2c-1)$ and the optimal workload reads as in (5.19). Since m, n_k must be integers, we propose to simply round the optimal parameters to the closest integer. A similar result (but without the last splitting stage, in which we estimate the time spent in the set B) has been presented in [72, Example 3.2.].

C Logarithmic Efficiency of the RMS Algorithm

In this section we study the efficiency of the RMS method, in the asymptotic regime that the rare event probability (5.1) tends to 0 (i.e. $\gamma \to 0$). First, we notice that if we fix the recurrency set A, then α_A does not change as $\gamma \to 0$; hence we only have that $T_B \to 0$. This indicates that asymptotic efficiency properties of RMS will be closely related to those of MLS. In order to study the performance of the estimator, we first introduce the concepts of *strong* and *logarithmic efficiency*.

Let $\widehat{\Psi}_\ell$ be a family of unbiased estimators for $\Psi_\ell > 0$, parametrized by ℓ such that $\Psi_\ell \to 0$, as $\ell \to \infty$. Let $W(\widehat{\Psi}_\ell)$ denote the computation time corresponding to $\widehat{\Psi}_\ell$. The estimator Ψ_ℓ is called *strongly efficient* if

$$\limsup_{\ell\to\infty} \frac{W(\widehat{\Psi}_\ell) \cdot \mathbb{V}\text{ar}(\widehat{\Psi}_\ell)}{\Psi_\ell^2} < \infty; \tag{5.39}$$

and *logarithmically efficient* if

$$\lim_{\ell\to\infty} \frac{W(\widehat{\Psi}_\ell) \cdot \mathbb{V}\text{ar}(\widehat{\Psi}_\ell)}{\Psi_\ell^{2-\varepsilon}} = 0, \quad \text{for all } \varepsilon > 0. \tag{5.40}$$

Strong efficiency implies that the workload needed to estimate the quantity of interest Φ_ℓ with a desired accuracy $\mathrm{RE}^2(\Psi_\ell) \leq \rho$ is uniformly bounded as $\ell \to \infty$. Logarithmic efficiency implies that workload needed to achieve the accuracy $\mathrm{RE}^2(\Psi_\ell) = \rho$ is increasing slower than $\Psi_\ell^{-\varepsilon}$ for any $\varepsilon > 0$, as $\ell \to \infty$. Evidently, strong efficiency implies logarithmic efficiency.

Before we prove the logarithmic efficiency of RMS in Theorem 5.3 we show an inefficiency result for the Monte Carlo estimator for T_B. Let $\widehat{T}_B^{\mathrm{MC}}$ be a sample mean of N independent copies of R_1. We then have

$$\mathrm{RE}^2(\widehat{T}_B^{\mathrm{MC}}) = \frac{1 - p_B + \mathrm{RE}^2(R_+)}{p_B N}. \tag{5.41}$$

Now to achieve a desired level of accuracy $\mathrm{RE}^2(\widehat{T}_B^{\mathrm{MC}}) \leq \rho$, assuming (5.37), the total required workload is

$$W(\widehat{T}_B^{\mathrm{MC}}) := \frac{1}{q} \cdot \frac{1 - p_B + \mathrm{RE}^2(R_+)}{p_B}. \tag{5.42}$$

As already noted in Section 4.1, $W(\widehat{T}_B^{\mathrm{MC}})$ is inversely proportional to p_B and so it follows that the Monte Carlo estimator is not logarithmically efficient.

We have seen, cf. (5.19), that the workload of the MLS estimator with the optimal parameters $W(\widehat{T}_B)$ is proportional to $(\log(p_B))^2$. It turns out that under mild additional assumption, the MLS algorithm is logarithmically efficient and thus so is RMS. We make this rigorous in the following theorem.

Theorem 5.3 (Logarithmic Efficiency of RMS)**.** *Fix the recurrency set A and let the set B_ℓ be parametrized by ℓ, such that $\gamma_\ell := \mu(B_\ell) \to 0$ as $\ell \to \infty$. Assume*

- *that the estimators $\widehat{\alpha}_A$ and $\widehat{T}_{B_\ell}$ are independent;*
- *that Assumptions* (4.1-4.1) *are valid for each ℓ;*
- *that the workload satisfies* (5.37)*;*
- *and that, for $\delta > 0$ sufficiently small,*

$$\limsup_{\ell \to \infty} \frac{\mathbb{V}\mathrm{ar}(R_+)}{(\mathbb{E}R_+)^2} < \infty, \quad \lim_{\ell \to \infty} T_{B_\ell} \cdot p_{B_\ell}^{-\delta} = 0. \tag{5.43}$$

Then the RMS estimator $\widehat{\gamma}_\ell$ for γ_ℓ, with the optimal choice of the parameters (5.38)*, is logarithmically efficient.*

We point out that the first part of the assumption (5.43) is equivalent to strong efficiency of the crude Monte Carlo estimator for R_+, under the workload assumption (5.37). This is not too restrictive, as often the main difficulty when estimating T_B lies in the fact that p_B is extremely small (and does not relate to the large variance of R_+.) Since $\gamma_\ell \to 0$ and A is fixed then necessarily $T_{B_\ell} \to 0$. In the second part of (5.43) we require that there exists a $\delta > 0$ such that $\mathbb{E}R_+ p_{B_\ell}^{1-\delta} \to 0$. Loosely speaking, it means that p_{B_ℓ} converges to 0 at least polynomially faster than $\mathbb{E}R_+$ grows to infinity; this is trivially satisfied when $\mathbb{E}R_+$ is bounded.

Proof of Theorem 5.3. Since the recurrency set A is fixed, the quantities $\widehat{\alpha}_A$, $\mathrm{RE}(\widehat{\alpha}_A)$ and $W(\widehat{\alpha}_A)$ do not depend on ℓ. In addition, $\alpha_A \cdot T_{B_\ell} = \mu(B_\ell) \to 0$ is equivalent to $T_{B_\ell} \to 0$. Moreover, since $T_{B_\ell} = p_{B_\ell} \cdot \mathbb{E}R_+$, cf. (5.12), and $\mathbb{E}R_+ \geq 1$, we necessarily have $p_{B_\ell} \to 0$, as ℓ grows. Observe that

$$\begin{aligned}\frac{W(\widehat{\gamma}_\ell)\mathbb{V}\mathrm{ar}(\widehat{\gamma}_\ell)}{\gamma_\ell^{2-\varepsilon}} &= \frac{W(\widehat{\alpha}_A) + W(\widehat{T}_{B_\ell})}{\gamma_\ell^{-\varepsilon}} \cdot \frac{\mathbb{V}\mathrm{ar}(\widehat{\alpha}_A \cdot \widehat{T}_{B_\ell})}{\alpha_A^2 \cdot T_{B_\ell}^2} \\ &= \gamma_\ell^{\varepsilon}\big(W(\widehat{\alpha}_A) + W(\widehat{T}_{B_\ell})\big) \cdot \big(\mathrm{RE}(\widehat{\alpha}_A) + \mathrm{RE}(\widehat{T}_{B_\ell})\big) \qquad (5.44)\end{aligned}$$

We put $\mathrm{RE}(\widehat{T}_{B_\ell}) = q$. Then the workload $W(\widehat{T}_{B_\ell})$ is given as in (5.19), and we see that

$$\begin{aligned}\gamma_\ell^{\varepsilon} W(\widehat{T}_{B_\ell}) &= \alpha_A^{\varepsilon} T_{B_\ell}^{\varepsilon} \cdot \frac{1}{q}\left(\frac{c\,|\log p_{B_\ell}|}{\sqrt{2c-1}} + \mathrm{RE}(R_+)\right)^2 \\ &\sim \frac{c^2 \alpha_A^{\varepsilon} (T_{B_\ell} p_{B_\ell}^{-\delta})^{\varepsilon}}{q(2c-1)} \cdot p_{B_\ell}^{\delta\varepsilon} (\log p_{B_\ell})^2,\end{aligned}$$

where $\delta > 0$ is as in (5.43). Now since $p_{B_\ell} \to 0$, we also have

$$p_{B_\ell}^{\delta\varepsilon} (\log p_{B_\ell})^2 \to 0,$$

and $\gamma_\ell^{\varepsilon} W(\widehat{T}_{B_\ell}) \to 0$, which applied to (5.44) finishes the proof. □

Bibliography

[1] Abdulle, A., Vilmart, G., and Zygalakis, K. C. (2014). High order numerical approximation of the invariant measure of ergodic sdes. *SIAM Journal on Numerical Analysis*, 52(4):1600–1622.

[2] Adler, R. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS Lecture Series. Institute of Mathematical Statistics.

[3] Adler, R., Blanchet, J., and Liu, J. (2012). Efficient Monte Carlo for high excursions of Gaussian random fields. *The Annals of Applied Probability*, 22(3):1167–1214.

[4] Alili, L., Patie, P., and Pedersen, J. (2005). Representations of the first hitting time density of an Ornstein-Uhlenbeck process. *Stochastic Models*, 21(4):967–980.

[5] Amrein, M. and Künsch, H. R. (2011). A variant of importance splitting for rare event estimation: Fixed number of successes. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 21(2):13.

[6] Asmussen, S. (2008). *Applied Probability and Queues*, volume 51. Springer Science & Business Media.

[7] Asmussen, S., Glynn, P., and Pitman, J. (1995). Discretization error in simulation of one-dimensional reflecting Brownian motion. *The Annals of Applied Probability*, 5:875–896.

[8] Asmussen, S. and Glynn, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*, volume 57. Springer Science & Business Media.

[9] Asmussen, S. and Ivanovs, J. (2018). Discretization error for a two-sided reflected Lévy process. *Queueing Syst.*, 89(1-2):199–212.

[10] Avikainen, R. (2009). On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13(3):381–401.

[11] Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 353, pages 401–419. The Royal Society.

[12] Bertoin, J. (1993). Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic Process. Appl.*, 47(1):17–35.

[13] Bertoin, J., Doney, R. A., and Maller, R. A. (2008). Passage of Lévy processes across power law boundaries at small times. *Ann. Probab.*, 36(1):160–197.

[14] Beskos, A., Roberts, G. O., et al. (2005). Exact simulation of diffusions. *The Annals of Applied Probability*, 15(4):2422–2444.

[15] Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

[16] Bisewski, K., Crommelin, D., and Mandjes, M. (2018a). Controlling the time discretization bias for the supremum of Brownian motion. *ACM Trans. Model. Comput. Simul.*, 28(3):Art. 24, 25.

[17] Bisewski, K., Crommelin, D., and Mandjes, M. (2018b). Simulation-based assessment of the stationary tail distribution of a stochastic differential equation. In *Proceedings of the 2018 Winter Simulation Conference*, pages 1742–1753.

[18] Bisewski, K., Crommelin, D., and Mandjes, M. (2019). Rare event simulation for steady-state probabilities via recurrency cycles. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(3):033131.

[19] Bisewski, K. and Ivanovs, J. (2019). Zooming-in on a Lévy process: Failure to observe threshold exceedance over a dense grid. *arXiv preprint arXiv:1904.06162*.

[20] Blanchet, J. and Zhang, F. (2017). Exact simulation for multivariate it\^ o diffusions. *arXiv preprint arXiv:1706.05124*.

[21] Bretagnolle, J. (1971). Résultats de Kesten sur les processus à accroissements indépendants. In *Séminaire de Probabilités V Université de Strasbourg*, pages 21–36. Springer.

[22] Broadie, M., Glasserman, P., and Kou, S. (1997). A continuity correction for discrete barrier options. *Mathematical Finance*, 7(4):325–349.

[23] Broadie, M., Glasserman, P., and Kou, S. G. (1999). Connecting discrete and continuous path-dependent options. *Finance Stoch.*, 3(1):55–82.

[24] Calvin, J. (1997). Average performance of a class of adaptive algorithms for global optimization. *The Annals of Applied Probability*, 7:711–730.

[25] Calvin, J. and Glynn, P. (1997). Average case behavior of random search for the maximum. *Journal of Applied Probability*, 34(3):632–642.

[26] Calvin, J. M., Glynn, P. W., and Nakayama, M. K. (2006). The semi-regenerative method of simulation output analysis. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(3):280–315.

[27] Cázares, J. G., Mijatović, A., and Bravo, G. U. (2019). Geometrically convergent simulation of the extrema of Lévy processes. (submitted for publication, arXiv:1810.11039).

[28] Cérou, F. and Guyader, A. (2007). Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443.

[29] Chaumont, L. (2013). On the law of the supremum of Lévy processes. *Ann. Probab.*, 41(3A):1191–1217.

[30] Chen, A. (2011). *Sampling error of the supremum of a Lévy process*. PhD thesis, University of Illinois at Urbana-Champaign.

[31] Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.

[32] Crane, M. A. and Iglehart, D. L. (1975). Simulating stable stochastic systems: III. Regenerative processes and discrete-event simulations. *Operations Research*, 23(1):33–45.

[33] Dean, T. and Dupuis, P. (2009). Splitting for rare event simulation: A large deviation approach to design and analysis. *Stochastic Processes and their Applications*, 119(2):562–587.

[34] Dębicki, K. and Mandjes, M. (2003). Exact overflow asymptotics for queues with many gaussian inputs. *Journal of Applied Probability*, 40(3):704–720.

[35] Dębicki, K. and Mandjes, M. (2015). *Queues and Lévy fluctuation theory*. Springer.

[36] Del Moral, P. and Garnier, J. (2005). Genealogical particle analysis of rare events. *The Annals of Applied Probability*, 15(4):2496–2534.

[37] Deng, C.-S. and Schilling, R. L. (2015). On shift Harnack inequalities for subordinate semigroups and moment estimates for Lévy processes. *Stochastic Processes and their Applications*, 125(10):3851–3878.

[38] Dia, E. H. A. and Lamberton, D. (2011). Connecting discrete and continuous lookback or hindsight options in exponential Lévy models. *Adv. Appl. Probab.*, 43(4):1136–1165.

[39] Dieker, A. and Lagos, G. (2019). On the Euler discretization error of Brownian motion about random times. (submitted for publication, arXiv:1708.04356).

[40] Doney, R. A. (2007). *Fluctuation theory for Lévy processes*, volume 1897 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005, Edited and with a foreword by Jean Picard.

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

[42] Doney, R. A. and Maller, R. A. (2002). Stability and attraction to normality for Lévy processes at zero and at infinity. *J. Theoret. Probab.*, 15(3):751–792.

[43] Doney, R. A. and Savov, M. S. (2010). The asymptotic behavior of densities related to the supremum of a stable process. *Ann. Probab.*, 38(1):316–326.

[44] Doucet, A. (2010). A note on efficient conditional simulation of Gaussian distributions. *Departments of Computer Science and Statistics, University of British Columbia*.

[45] Durrett, R. (1996). *Stochastic Calculus: a Practical Introduction*. CRC Press.

[46] Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. In *Lévy processes*, pages 319–336. Birkhäuser Boston, Boston, MA.

[47] Feller, W. (1971). *An introduction to probability theory and its applications*, volume II (2. Ed.). John Wiley & Sons.

[48] Figueroa-López, J. E. (2008). Small-time moment asymptotics for Lévy processes. *Statistics & Probability Letters*, 78(18):3355–3365.

[49] Figueroa-López, J. E. and Houdré, C. (2009). Small-time expansions for the transition distributions of Lévy processes. *Stochastic Processes and their applications*, 119(11):3862–3889.

[50] Franzke, C. (2012). Predictability of extreme events in a nonlinear stochastic-dynamical model. *Physical Review E*, 85(3):031134.

[51] Garvels, M. J. J. (2000). *The splitting method in rare event simulation*. Universiteit Twente.

[52] Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*, volume 195. Springer Science & Business Media.

[53] Giles, M. (2008). Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617.

[54] Giles, M. B. and Xia, Y. (2017). Multilevel Monte Carlo for exponential Lévy models. *Finance Stoch.*, 21(4):995–1026.

[55] Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 1:126–166.

[56] Goyal, A., Shahabuddin, P., Heidelberger, P., Nicola, V. F., and Glynn, P. W. (1992). A unified framework for simulating Markovian models of highly dependable systems. *IEEE Transactions on Computers*, 41(1):36–51.

[57] Gunther, F. and Wolff, R. (1980). The almost regenerative method for stochastic system simulations. *Operations Research*, 28(2):375–386.

[58] Hansen, N. R. (2003). Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli*, 9(4):725–743.

[59] Heidelberger, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 5(1):43–85.

[60] Henderson, S. G. and Glynn, P. W. (1999). Can the regenerative method be applied to discrete-event simulation? In *Proceedings of the 31st Winter Simulation Conference*, pages 367–373.

[61] Henderson, S. G. and Glynn, P. W. (2001). Regenerative steady-state simulation of discrete-event systems. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 11(4):313–345.

[62] Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM review*, 43(3):525–546.

[63] Ivanovs, J. (2017). Zooming in on a Lévy process at its supremum. *The Annals of Applied Probability*, 28(2):912–940.

[64] Janssen, A. and Van Leeuwaarden, J. (2009). Equidistant sampling for the maximum of a Brownian motion with drift on a finite horizon. *Electronic Communications in Probability*, 14:143–150.

[65] Kahn, H. and Harris, T. E. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards applied mathematics series*, 12:27–30.

[66] Kalashnikov, V. V. (1994). *Topics on Regenerative Processes*. CRC Press.

[67] Kallenberg, O. (2002). *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition.

[68] Kesten, H. (1969). *Hitting probabilities of single points for processes with stationary independent increments*. Memoirs of the American Mathematical Society, No. 93. American Mathematical Society, Providence, R.I.

[69] Kloeden, P. E. and Platen, E. (1992). Numerical solution of stochastic differential equations.

[70] Kroese, D. P., Taimre, T., and Botev, Z. I. (2013). *Handbook of Monte Carlo methods*, volume 706. John Wiley & Sons.

[71] Kwaśnicki, M., Małecki, J., and Ryznar, M. (2013). Suprema of Lévy processes. *Ann. Probab.*, 41(3B):2047–2065.

[72] Lagnoux, A. (2006). Rare event simulation. *Probability in the Engineering and Informational Sciences*, 20(1):45–66.

[73] Léandre, R. (1987). Densité en temps petit d'un processus de sauts. In *Séminaire de probabilités XXI*, pages 81–99. Springer.

[74] L'Ecuyer, P., Demers, V., and Tuffin, B. (2007). Rare events, splitting, and quasi-monte carlo. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 17(2):9.

[75] Li, X. and Liu, J. (2015). Rare-event simulation and efficient discretization for the supremum of Gaussian random fields. *Advances in Applied Probability*, 47(3):787–816.

[76] Mattingly, J. C., Stuart, A. M., and Higham, D. J. (2002). Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185–232.

[77] Meyn, S. P. and Tweedie, R. L. (2012). *Markov Chains and Stochastic Stability*. Springer.

[78] Mörters, P. and Peres, Y. (2010). *Brownian Motion*, volume 30. Cambridge University Press.

[79] Picard, J. (1997). Density in small time for Lévy processes. *ESAIM: Probability and Statistics*, 1:357–389.

[80] Piterbarg, V. and Prisyazhnyuk, V. (1978). Asymptotic analysis of the probability of large excursions for a nonstationary Gaussian process. *Theory of Probability and Mathematical Statistics*, 18:121–134.

[81] Ragone, F., Wouters, J., and Bouchet, F. (2018). Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29.

[82] Raible, S. (2000). *Lévy processes in finance: Theory, numerics, and empirical facts*. PhD thesis, Universität Freiburg.

[83] Rhee, C.-h. and Glynn, P. W. (2015). Unbiased estimation with square root convergence for sde models. *Operations Research*, 63(5):1026–1043.

[84] Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363.

[85] Rubino, G. and Tuffin, B. (2009). *Rare Event Simulation using Monte Carlo Methods*. John Wiley & Sons.

[86] Sato, K.-I. (2013). *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

[87] Schurz, H. (1999). The invariance of asymptotic laws of stochastic systems under discretization. *ZAMM-Zeitschrift fur Angewandte Mathematik und Mechanik*, 79(6):375–382.

[88] Sharpe, M. (1969). Zeroes of infinitely divisible densities. *The Annals of Mathematical Statistics*, 40(4):1503–1505.

[89] Siegmund, D. (1985). *Sequential analysis: Tests and confidence intervals*. Springer-Verlag.

[90] Stramer, O. and Tweedie, R. (1999). Langevin-type models i: Diffusions with given stationary distributions and their discretizations. *Methodology and Computing in Applied Probability*, 1(3):283–306.

[91] Szarek, S. and Werner, E. (1999). A nonsymmetric correlation inequality for gaussian measure. *Journal of Multivariate Analysis*, 68(2):193–211.

[92] Talay, D. (1990). Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics*, 29(1):13–36.

[93] Tweedie, R. (1983). The existence of moments for stationary markov chains. *Journal of Applied Probability*, 20(1):191–196.

[94] van der Vaart, A. (2010). Time series. `http://www.math.leidenuniv.nl/~avdvaart/timeseries/dictaat.pdf`.

[95] Villén-Altamirano, M. and Villén-Altamirano, J. (2011). The rare event simulation method restart: efficiency analysis and guidelines for its application. In *Network performance engineering*, pages 509–547. Springer.

[96] Wadman, W., Crommelin, D., and Frank, J. (2014). A separated splitting technique for disconnected rare event sets. In *Proceedings of the 46th Winter Simulation Conference*, pages 522–532.

[97] Wadman, W. S., Crommelin, D. T., and Zwart, B. P. (2016). A large-deviation-based splitting estimation of power flow reliability. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 26(4):23.

[98] Whitt, W. (1980). Some useful functions for functional limit theorems. *Math. Oper. Res.*, 5(1):67–85.

[99] Winkelbauer, A. (2012). Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*.

[100] Zolotarev, V. M. (1957). Mellin-Stieltjes transforms in probability theory. *Theory of Probability & Its Applications*, 2(4):433–460.

List of Publications

- **Bisewski, K.**, Crommelin, D., and Mandjes, M. (2018). *Controlling the Time Discretization Bias for the Supremum of Brownian Motion.* ACM Transactions on Modeling and Computer Simulation (TOMACS), 28(3), 24.

  Chapter 2 is based on this research article. The research work, writing of the article and numerical simulations were carried out by K. Bisewski under the supervision of D. Crommelin and M. Mandjes.

- **Bisewski, K.**, and Ivanovs, J. (2019). *Zooming-in on a Levy process: Failure to Observe Threshold Exceedance over a Dense Grid.* Submitted.

  Chapter 3 is based on this research article. The research work and writing of the article were carried out jointly by K. Bisewski and J. Ivanovs.

- **Bisewski, K.**, Crommelin, D., and Mandjes, M. (2018). *Simulation-Based Assessment of the Stationary Tail Distribution of a Stochastic Differential Equation.* In Proceedings of Winter Simulation Conference 2018.

  Chapter 4 is based on this research article. The research work, writing of the article and numerical simulations were carried out by K. Bisewski under the supervision of D. Crommelin and M. Mandjes.

- **Bisewski, K.**, Crommelin, D., and Mandjes, M. (2019). *Rare Event Simulation for Steady-State Probabilities via Recurrency Cycles.* Chaos: An Interdisciplinary Journal of Nonlinear Science 29.033131.

  Chapter 5 is based on this research article. The research work, writing of the article and numerical simulations were carried out by K. Bisewski under the supervision of D. Crommelin and M. Mandjes.