Boundary behavior in stochastic differential equations used in Finance

A Thesis
Presented to

The Institute of Physics within the Faculty of Science
Eötvös Lóránd University

In Partial Fulfillment
of the Requirements for the Degree of
Master of Science

by
Báskay János
Supervisor: Fáth Gábor

May 2020

Table of Contents

1. Abstract 5

2. Theoretical Introduction 6

Boundary classification of stochastic processes 6

Regular Boundary 10

Exit Boundary 11

Entrance Boundary 11

Natural (Feller) Boundary 11

The Cox-Ingersoll-Ross model 12

Interest rates 12

Models for short-term interest rates 14

Properties of the CIR model 18

The CIR model as a squared Ornstein-Uhlenbeck process 21

The Ornstein-Uhlenbeck model 22

Connection to the CIR model 23

The case of the Absorbing boundary condition 24

3. Monte Carlo implementation of the Cox-Ingersoll-Ross process 28

Random number generation 28

Mersenne Twister Algorithm 29

Inverse transform sampling 31

Discretization schemes 31

Handling the boundaries 33

Improving performance 34

Parallel processing 35

Series acceleration: Richardson extrapolation 35

4. Simulation Results 36

Choosing the right discretization scheme 36

Runtime 36
Accuracy 38
Comparing the distributions 39
Measuring the time to cross barriers 42
Finding γ_c numerically 42
Reflective Boundary 42
Absorbing Boundary 44
Regarding Monte Carlo error 45
5. Conclusion 47
Appendix A.: Code 48
Appendix B.: Additional Figures 48
Time to cross the barriers 48
Convergence at the Reflective boundary 50
Convergence at the Absorbing boundary 53
Comparing the PDF at $\gamma = ½$ 57
Monte Carlo error 58
Comparing the PDF of OU Process with the simulation 58
References 59

1. Abstract

Mathematical finance uses certain stochastic models, that possess rich phase diagrams in the space spanned by their parameters. The different phases can be separated based on the processes' behavior near its boundary: whether or not the boundary can be reached from the bulk, or the bulk can be reached from the boundary.

This thesis will be focused on investigating the boundary behavior of the Cox-Ingersoll-Ross model, which was introduced in 1985 by John C. Cox, Johnathan E. Ingersoll and Stephen A. Ross to describe the evolution of interest rates. The condition commonly referred to as Feller condition ensures the solution of this process is bounded below by zero. The attainability of the zero-boundary is dependent on the parameters of the model, and well known analytically.

My goal is to show that numerical simulation methods can also reproduce the analytical boundary behavior for both Reflective and Absorbing boundary conditions. To do this I will have to investigate multiple discretization schemes of the CIR model, introduce suitable order parameters for both the boundary conditions, and then approximate the continuum limit $(dt \to 0)$ of the phase diagram.

2. Theoretical Introduction

This first chapter's aim is to provide a summary for the most important theoretical considerations used in the simulations, starting from the modern classification of boundary behavior, then moving on to introducing the Cox-Ingersoll-Ross model, and finally seeing a curious case when a simpler stochastic model can be used to describe a special case of a more complex process.

Boundary classification of stochastic processes

The phase transitions in the boundary behavior of diffusion processes is understood as follows: A change in the control parameter(s) of the system will change how the boundary of the process is classified. In this section we will review the modern classification of possible behavior near the boundary of a diffusion processes based on the work of Samuel Karlin and Howard M. Taylor. (Karlin & Taylor, 1981)

"A continuous time parameter stochastic process which possesses the (strong) Markov property and for which the sample paths X_t *are continuous functions of* t *with probability* 1 *is called a diffusion process."* All diffusion processes can take the following form:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \qquad t > 0$$

$$dW_t = z\sqrt{dt}, \qquad z \in \mathcal{N}(0,1)$$

$\{X(t), t \geq 0\}$ Shall be a regular diffusion process bounded from the left with l and from the right with r. For every x in the interval (l, r) the drift and variance of the process shall be continuous and infinitesimal denoted by $\mu(x)$ and $\sigma(x) > 0$ respectively. We will focus on the left boundary l the derivations for the right are similar.

We'll investigate the following quantities in the $a \to l$ limit:

$$u(x) = u_{a,b}(x) = \Pr\{T_b < T_a | X(0) = x\}, \qquad l < a < x < b < r \qquad \textit{2.1. Eq.}$$

and

$$v(x) = v_{a,b}(x) = E\left[T_{a,b} \middle| X(0) = x\right], \qquad l < a < x < b < r \qquad \textit{2.2. Eq.}$$

where $u(x)$ is the probability that the process reaches b before a, $v(x)$ is the mean time to reach either a or b, T_a is the hitting time to a and $T_{a,b} = \min\{T_a, T_b\}$.

Let's introduce the scale function $S(x)$ and speed density $m(x)$:

$$S(x) = \int_{x_0}^{x} s(\xi)d\xi\,, \qquad s(\xi) = \exp\left\{-\int_{\xi_0}^{\xi}\left[\frac{2\mu(\eta)}{\sigma^2(\eta)}\right]d\eta\right\} \qquad \textit{2.3. Eq.}$$

$$m(x) = \frac{1}{[\sigma^2(x)s(x)]}$$

x_0 and ξ_0 are arbitrary fixed points inside (l, r), and the particular choice has no relevance.

Using the scale function, we can define the scale measure of a closed interval $J = [c, d] \subset (l, r)$ as follows:

$$S[J] = S[c, d] = S(d) - S(c) \qquad \textit{2.4. Eq.}$$

Furthermore, the scale measure of an infinitesimal interval $[x, x + dx]$ will be $S[dx] = S(x + dx) - S(x) = dS(x) = s(x)dx$. This can be used to evaluate $\int_c^d f(x)dS(x) = \int_c^d f(x)s(x)dx$.

Similarly, we define the speed measure M using the speed density $m(x)$:

$$M[J] = M[c, d] = \int_c^d m(x)dx\,, \qquad J = [c, d] \subset (l.r) \qquad \textit{2.5. Eq.}$$

Both measures are positive and finite for $J = [c, d] \subset (l.r)$.

Eq.-s 2.1 and *2.2* can be rewritten in terms of the scale and speed measures as

$$u(x) = u_{a,b}(x) = \frac{S[a, x]}{S[a, b]}, \qquad l < a < x < b < r \qquad \textit{2.6. Eq.}$$

and

$$v(x) = v_{a,b}(x) = 2\left\{u(x)\int_x^b S[\eta, b]dM(\eta) + [1 - u(x)]\int_a^x S[a, \eta]dM(\eta)\right\} \qquad \textit{2.7. Eq.}$$

We may use the non-negativity of the measure S and the monotonic behavior in a for $S[0, b]$ for fixed b to define $S(l, b] \leq \infty$ with the limit

$$S(l, b] = \lim_{a \to l} S[a, b] \leq \infty, \qquad l < b < r \qquad \textit{2.8. Eq.}$$

Since $[a, b] \subset (l.r)$ and $0 \leq S[a, b] < \infty$ it follows that $S(l, b] = \infty$ for some $b \in (l, r)$ if and only if $S(l, b] = \infty$ for $\forall b \in (l, r)$.

Now we'll derive the hitting time for the boundary l as follows: First let's introduce $T_{l+} = \lim_{a \to l} T_a \leq \infty$. We have to see, that $T_l = T_{l+}$, in order to do that we have to separate two cases. In the first case $T_{l+} = \infty$. We know that $T_a \leq T_l$ and by definition $T_{l+} \leq T_l$, but $T_{l+} = \infty$ thus $T_l = \infty$. In the second case T_{l+} shall be finite. As the paths are continuous, $X(T_{l+}) = \lim_{a \to l} X(T_a) = \lim_{a \to l} a = l > -\infty$ as such $T_{l+} \geq T_l$.This means $T_l = T_{l+} \leq \infty$. It is important to note that T_l exists even when l is not a possible state for the diffusion process.

Lemma

1.) Suppose that $S(l, x_0] < \infty$ for some $x_0 \in (l, r)$. Then $\Pr\{T_l \leq T_b | X(0) = x\} > 0$, $\forall\, l < x < b < r$.
2.) Suppose that $S(l, x_0] = \infty$ for some $x_0 \in (l, r)$. Then $\Pr\{T_l \leq T_b | X(0) = x\} = 0$, $\forall\, l < x < b < r$.

Definition

The boundary l is *attracting* if $S(l, x] < \infty$. This applies independently $\forall x \in (l, r)$.

Based on this definition we can imagine processes, that have an attracting boundary at l, however the expected value of T_l is infinite. Thus, the question naturally arises, when is a boundary attainable in finite expected time?

Let l be an attracting boundary with $S(l, x] < \infty, \forall x \in (l, r)$. According to the previously stated lemma the left boundary can be reached prior to reaching any arbitrary state b with a probability > 0 from any $x < b$ starting point.

$$\lim_{a \to l} E_x[T_{a,b}] = \lim_{a \to l} E_x[\min\{T_a, T_b\}], \qquad x < b < r \qquad \textit{2.9. Eq.}$$

This is the expected value of the first hitting time of the boundary l or the level b starting from x. Substituting in *Eq.-s 2.6* and *2.7* to *Eq. 2.9* we will arrive to

$$\lim_{a \to l} E_x[T_{a,b}] = \lim_{a \to l} \frac{2S[a,x]}{S[a,b]} \int_x^b S[\xi, b] dM(\xi) + \lim_{a \to l} \frac{2S[x,b]}{S[a,b]} \int_a^x S[a, \xi] dM(\xi) \qquad \textit{2.10. Eq.}$$

We assumed that l is attracting, so the limits $\lim_{a \to l} \frac{S[a,x]}{S[a,b]}$ and $\lim_{a \to l} \frac{S[x,b]}{S[a,b]}$ are both positive and finite. Thus, we arrive to the following requirement:

$$\lim_{a \to l} E_x[T_{a,b}] < \infty \;\; \textit{if and only if} \;\; \lim_{a \to l} \int_a^x S[a, \xi] dM(\xi) < \infty \qquad \textit{2.11. Eq.}$$

This motivates the introduction of a new quantity:

$$\Sigma(l) = \lim_{a \to l} \int_a^x S[a,\xi] dM(\xi) = \int_l^x S(l,\xi] dM(\xi) = \int_l^x \left\{ \int_l^\xi s(\eta) d\eta \right\} m(\xi) d\xi$$
$$= \int_l^x \left\{ \int_\eta^l m(\xi) d\xi \right\} s(\eta) d\eta = \int_l^x M[\eta, x] dS(\eta)$$

2.12. Eq.

Definition

The boundary l is said to be

1) *attainable* if $\Sigma(l) < \infty$
2) *unattainable* if $\Sigma(l) = \infty$

It can be shown, that Attainable boundaries are always attractive $(\Sigma(l) < \infty \Rightarrow S(l,x] < \infty)$, however Unattainable boundaries may or may not be attracting.

Lemma

Let l be an attracting boundary and suppose $l < x < b < r$. In this case the following statements are equivalent:

1) $\Pr\{T_l < \infty | X(0) = x\} > 0$
2) $E\left[T_{l,b} \middle| X(0) = x\right] < \infty$
3) $\Sigma(l) < \infty$

The equivalence of statement 2) and 3) is what led to the introduction of $\Sigma(l)$. To see the connection between 1) and 2) let's suppose $E\left[T_{l,b} \middle| X(0) = x\right] < \infty$ then $\Pr\{T_{l,b} < \infty\} = 1$. We also know, that $T_l \neq T_b$ and l is Attainable which means $\Pr\{T_l \leq T_b | X(0) = x\} > 0$, so $\Pr\{T_l < T_b | X(0) = x\} > 0$. Putting these together we'll get $\Pr\{T_l < \infty | X(0) = x\} > \Pr\{T_l < T_b | X(0) = x\} > 0$, thus establishing the connection between statements 1) and 2).

Based on its definition $\Sigma(l)$ measures the time it takes to reach the boundary l or an interior point b starting from and other interior point $l < x < b$. We will now introduce a quantity to measure the speed of the process near l, denoted by $M(l,X]$ and another quantity that measures the time it takes to reach the interior point x starting form the boundary l. This shall be denoted by $N(l)$.

$$M(l,x] = \lim_{a \to l} M[a,x]$$
$$N(l) = \int_l^x S[\eta, x] dM(\eta) = \int_l^x M(l,\xi] dS(\xi)$$

2.13. Eq.

The modern boundary classification depends on the quantities introduced earlier in this section, namely $S(l,x];\Sigma(l);N(l);M(l,x]$. The important question will be whether these quantities are finite or not.

Lemma

The following relations hold between $S(l,x],\Sigma(l),N(l)$ and $M(l,x]$:

1) $S(l,x] = \infty \; implies \; \Sigma(l) = \infty$,
2) $\Sigma(l) < \infty \; implies \; S(l,x] < \infty$,
3) $M(l,x] = \infty \; implies \; N(l) = \infty$,
4) $N(l) < \infty \; implies \; M(l,x] < \infty$,
5) $\Sigma(l) + N(l) = S(l,x]M(l,x]$

Most of these statements have already appeared previously, so we will focus on seeing 5). First the definition of $\Sigma(l)$ and $N(l)$

$$\Sigma(l) = \int_l^x S(l,\xi]dM(\xi)$$

$$N(l) = \int_l^x S[\xi,x]dM(\xi)$$

Substituting them into statement 5)

$$\Sigma(l) + N(l) = \int_l^x \{S(l,\xi] + S[\xi,x]\}dM(\xi) = S(l,x]\int_l^x dM(\xi) = S(l,x]M(l,x] \qquad \textit{2.14. Eq.}$$

There's 16 possible combination of these quantities, but most of them are ruled out by the lemma stated above. Only six combinations can be realized, these have been grouped and labeled differently by different authors. We'll now go through the classifications used by most Americans first introduced by William Feller. Then we'll see a table, that summarizes the labels and their corresponding criteria.

Regular Boundary

At a Regular boundary the diffusion process can both enter the bulk or leave. In this case the speed at the boundary must be specified, which can range from absorption $(M[\{l\}] = \infty)$ to reflection $(M[\{l\}] = 0)$. When the speed at the boundary falls in between the Absorbing and Reflecting case sticky boundary phenomenon occurs, and the process can spend some time stuck at the boundary before moving back to the bulk. Another option is to restart the process from the bulk every time it arrives to the boundary.

A boundary point l is considered Regular, if $S(l,x] < \infty$ and $M(l,x] < \infty$.

Exit Boundary

To establish that l is an Exit boundary the following criteria has to be met

$$\lim_{b\to l}\lim_{x\to l} \Pr\{T_b < t | X(0) = x\} = 0 \qquad \forall\, t > 0 \qquad \textit{2.15. Eq.}$$

This means that starting at l it is not possible to reach any b state in the bulk, regardless of how close b is to l. This also means, that if the boundary l is reached no continuous path leads out from it. For exit boundaries T_l is often called the exit time.

A boundary point l is considered Exit, if $\Sigma(l) < \infty$ and $M = \infty$.

Entrance Boundary

The special propriety of Entrance boundaries is that they can't be reached from the bulk, but it is possible for a process started at the boundary to reach the interior, generally this means that the process leaves the boundary quickly, and then never returns.

A boundary point l is considered Entrance, if $S(l,x] = \infty$ and $N(l) < \infty$.

Natural (Feller) Boundary

We can construct a diffusion process, where the expected value of the hitting time for the boundary l is infinite, and the process cannot be started from there neither. Natural boundaries are not part of the state space of the process, so if l is a Natural boundary, the state space shall be taken from (l,r) or$(l,r]$.

To show that l is a Natural boundary in the Feller sense it is enough to establish $\Sigma(l) = \infty$ and $N(l) = \infty$.

Criteria				**Terminology**		
$S(l,x]$	$M(l,x]$	$\Sigma(l)$	$N(l)$			
$< \infty^*$	$< \infty^*$	$< \infty$	$< \infty$	Regular	Attracting	Attainable
$< \infty$	$= \infty^*$	$< \infty^*$	$= \infty$	Exit		
$< \infty^*$	$= \infty^*$	$= \infty^*$	$= \infty$	Natural		Unattainable
$= \infty^*$	$< \infty^*$	$= \infty$	$= \infty^*$		Nonattracting	
$= \infty^*$	$= \infty^*$	$= \infty$	$= \infty$			
$= \infty^*$	$< \infty$	$= \infty$	$< \infty^*$	Entrance		
*** marks the minimal sufficient criteria to establish the nature of the boundary**						

2.1. Table: The terminology of the Feller boundary classification schemes (Karlin & Taylor, 1981). Attracting and Attainable boundaries can be reached in finite amount of time, Attracting and Unattainable boundaries can only be reached in infinite amount of time and Nonattracting and Unattainable boundaries cannot be reached whatsoever.

The Cox-Ingersoll-Ross model

Interest rates

Since the CIR model was created to describe the short-term structure of interest rates, I will now introduce some crucial notions from the field of Finance.

The interest rate is defined as the amount of money a borrower promises for the lender. The amount of the interest rate determined in a situation depends on the credit risk of the borrowing party, this means the risk that there will be a default and the borrower will be unable to pay. Generally, the higher the credit risk, the higher the interest rate will be on a transaction.

There are many types of interest rates available on the Financial Markets, the most commonly used are

- **Treasury Rates**: Governments can issue in their own currency Treasury bills and Treasury bonds. The interest earned on these is called the Treasury Rate. Since these are issued in the government's own currency, there is very little risk that the government cannot comply to the payment obligations, and thus these rates are regarded as risk-free.
- **LIBOR**: Short for *London Interbank Offered Rate*. LIBOR is an unsecured short-term borrowing rate between banks. The borrowing periods span from a single day to a year, and it can be arranged in many currencies. LIBOR is now regarded as a less-than-ideal reference rate for derivatives, since it's determined from estimates made by the banks, and not as a measured property of the Market.
- **Overnight Rates**: All banks maintain a cash reserve at the central bank. The reserve that is required for a particular bank is determined by their assets and liabilities. It can happen, that at a given time some banks have more cash reserve, than it is required, and some have less. This leads to borrowing and lending these reserves overnight.
- **Repo Rates**: Repo rates are similar to Overnight rates and LIBOR-s in the sense, that they are inter-bank transactions, however a crucial difference is that Repo rates are secured. Repo stands for *repurchase agreements*. One financial institution sells securities to and other, and in return agrees to buy them back in the future. The interest in this loan is the difference between the price at which the securities were sold, and the amount payed when they were repurchased. Repo rates are considered low risk.

As we have mentioned the risk of a loan pays a crucial role in determining the corresponding interest rate, and when we're setting up an investment portfolio, we're aiming for it

to be risk-free. The return of this portfolio can be regarded as the risk-free rate, which has to be determined from the Market. Based on the descriptions of the most common types of Interest Rates one would assume, that Treasury Rates are used for risk-free rates, however in practice this seldom happens. Prior to the Financial Crisis of 2008 LIBOR was used as a stand in for risk-free rate, but during the Crisis it turned out to be a mistake. LIBOR rates soared, and most banks realized, that they can't be regarded as risk-free anymore. Since then OIS rates are used. OIS is short for Overnight Indexed Swaps, it is Swap product where the parties agree to exchange a fixed rate for a period to the geometric average of the overnight rates for the same period.

When measuring the interest rates the statement, that the bank gives 10% annually is quite ambiguous since the realized yield still depends on compounding. This means, if the interest rates are measured with yearly compounding for a $\$100$ investment we will receive $\$100 * 1.1 = \$110$. If the compounding is bi-annual the same $\$100$ will yield $\$100 * 1.05 * 1.05 = \$110.25$. The last simple example shall be daily compounding, in this case we shall receive $\$110.52$. From these examples it is easy to determine the general formula for the yield, depending on the compounding frequency:

$$A\left(1+\frac{R}{m}\right)^{mn} \qquad \textit{2.16. Eq.}$$

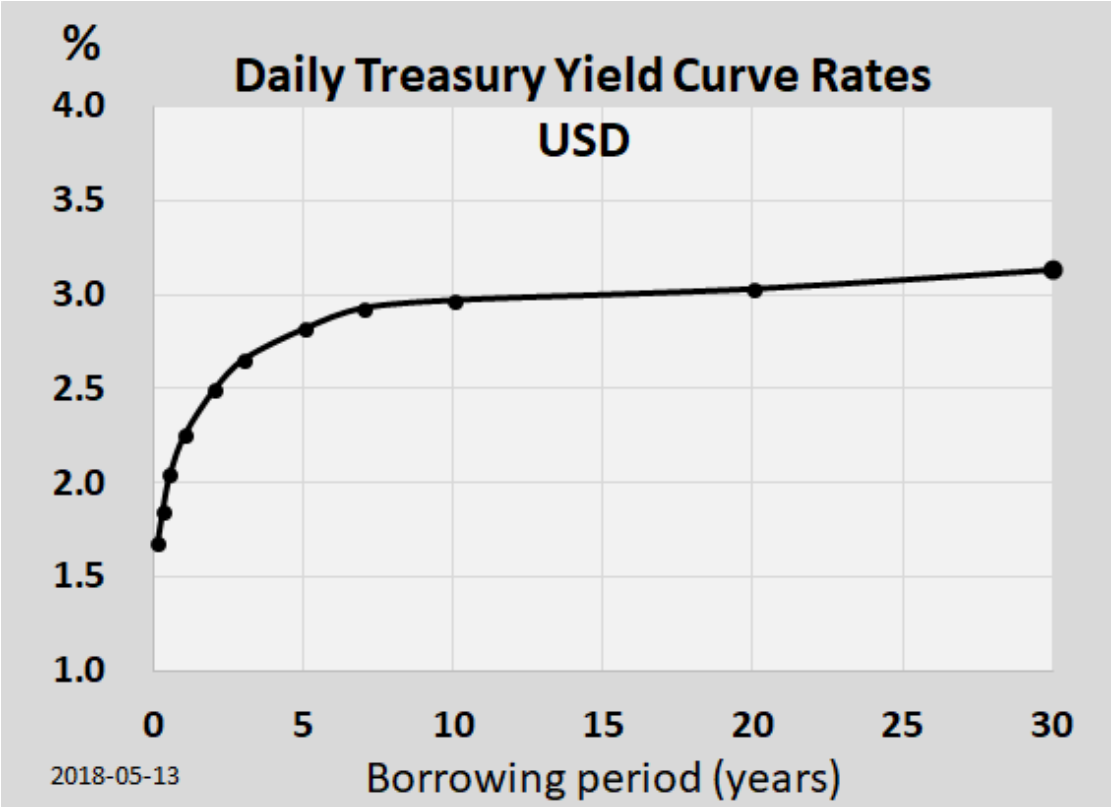
Where A is the amount invested, R is the interest rate per annum, n is the number of years and m is the compounding frequency.

From this it's easy to take the $m \to \infty$ limit, to introduce continuous compounding, we just have to use the definition for the exponential function:

$$\lim_{m\to\infty} A\left(1+\frac{R}{m}\right)^{mn} = Ae^{Rn} \qquad \textit{2.17. Eq.}$$

We can also see, that for the most cases daily compounding is a good approximation for continuous compounding. (At the $\$100$ example continuous compounding yields $\$110.52$, the difference is only in the third digit.)

When we're examining the Interest Rates for different durations (or maturities) of bonds, we see, that it is not a constant. The relationship between the Interest Rate and the maturity is called the term structure, and is often depicted by the yield curve, an example can be seen *Fig. 2.1.*



2.1. Figure: US Treasury yield curve from May, 2018 (Ldecola, 2018) based on the publicly available data on the US treasury website

The example depicts a typical yield curve, that has an upwards sloping tendency. The term-structure is generally understood based on the *liquidity preference theory*. (Keynes, 1936) The assumption of this theory is that investors like to preserve their liquidity thus prefer short-term investments, while borrowers tend to gravitate to fixed rate loans for long-term. This theory is consistent with empirical evidence, and most models concerning Interest Rates comply with it.

(Hull & Basu, Interest Rates, 2018)

Models for short-term interest rates

We will now turn our attention to models describing the short-term evolution of risk-free interest rate. These models approximately match the real-world term structure, but they do not provide an exact match. These models are useful tools for Monte Carlo simulations to perform scenario analysis for longer periods of time.

We will denote the risk-free short rate with r. At time t this rate applies to an infinitesimally short period. This means, that between t and $t + \Delta t$ investors will earn $r(t)\Delta t$. It is generally assumed that derivative products only depend on the evolution of $r(t)$, thus if we have a product with a payoff of f_T at time T the value of said derivative at time t shall be

$$E\left[e^{-\bar{r}(T-t)}f_T\right] \qquad \textit{2.18. Eq.}$$

Where $\bar{r}$ denotes the temporal average of $r(t)$ between times t and T. We can also define the price of a zero-coupon[1] bond that pays at time T as

$$P(t,T) = E\left[e^{-\bar{r}(T-t)}\right] \qquad \textit{2.19. Eq.}$$

If we denote the continuously compounded risk-free interest rate for a term of $T-t$ with $R(t,T)$ it follows from *Eq. 2.17* that

$$P(t,T) = e^{-R(t,T)(T-t)} \qquad \textit{2.20. Eq.}$$

This means

$$R(t,T) = -\frac{1}{T-t}\ln P(t,T) \qquad \textit{2.21. Eq.}$$

And using *Eq. 2.19*

$$R(t,T) = -\frac{1}{T-t}\ln E\left[e^{-\bar{r}(T-t)}\right] \qquad \textit{2.22. Eq.}$$

This equation lets us obtain the term structure of interest rates, given we have defined a risk neutral process that drives the evolution of the short rate $r(t)$.

The general form of a process for $r(t)$ looks like this:

$$dr = m(r,t)dt + s(r,t)dW_t \qquad \textit{2.23. Eq.}$$

Where the first part is often called the drift term and the second part can be called the diffusion term. dW_t denotes a Wiener-process, $dW_t = z\sqrt{dt}, z \in \mathcal{N}(0,1)$. (Wiener, 1976)

Itô's Lemma

Since we will use this Lemma multiple times, it will need a proper introduction. Ito's lemma concerns the differential equation of a process, that is the function of another stochastic process.

Lemma

Let's suppose that the variable x follows the Itô process (Itô, 1951)

[1] Zero-coupon means there is only one payoff at the end of maturity time T, there are no possibility of intermediate payments.

$$dx = a(x,t)dt + b(x,t)dW_t \qquad \textit{2.24. Eq.}$$

In this case the function $G(x,t)$ will obey the following Itô process

$$dG = \left(\frac{\partial G}{\partial t} + \frac{\partial G}{\partial x}a + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}b\,dW_t \qquad \textit{2.25. Eq.}$$

A nonrigorous derivation of the Lemma goes as follows. (Hull & Basu, A Nonrigorous Derivation of Itô's Lemma, 2018)

Let's suppose G is a continuous and differentiable function of x. From elementary calculus we know, that a small change in x will result in a small change in G:

$$\Delta G = \frac{dG}{dx}\Delta x \qquad \textit{2.26. Eq.}$$

We can of course obtain more corrections, by utilizing the Taylor-series for G:

$$\Delta G = \frac{dG}{dx}\Delta x + \frac{1}{2}\frac{d^2G}{dx^2}\Delta x^2 + \frac{1}{6}\frac{d^3G}{dx^3}\Delta x^3 + \cdots \qquad \textit{2.27. Eq.}$$

If G is a continuous and differentiable function of both x and y we can also expand it with Taylor series up to the second order:

$$\Delta G = \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial y}\Delta y + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\Delta x^2 + \frac{\partial^2 G}{\partial x \partial y}\Delta x \Delta y + \frac{1}{2}\frac{\partial^2 G}{\partial y^2}\Delta y^2 + \cdots \qquad \textit{2.28. Eq.}$$

We will now assume, that the variable x follows an Itô process, and change y to t:

$$dx = a(x,t)dt + b(x,t)z\sqrt{dt}, \qquad z \in \mathcal{N}(0,1) \qquad \textit{2.29. Eq.}$$

The discretized equations for Δx and Δx^2 will be

$$\Delta x = a\Delta t + bz\sqrt{\Delta t} \qquad \textit{2.30. Eq.}$$
$$\Delta x^2 = b^2 z^2 \Delta t + \mathcal{O}(\Delta t^2)$$

We can substitute these to *Eq. 2.28.* and only keep the linear terms in Δt:

$$\Delta G = \frac{\partial G}{\partial t}\Delta t + \frac{\partial G}{\partial x}\left(a\Delta t + bz\sqrt{\Delta t}\right) + \frac{\partial^2 G}{\partial x^2}(b^2 z^2 \Delta t) \qquad \textit{2.31. Eq.}$$

Collecting the Δt and $\sqrt{\Delta t}$ terms together:

$$\Delta G = \left(\frac{\partial G}{\partial t} + \frac{\partial G}{\partial x}a + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2 z^2\right)\Delta t + \frac{\partial G}{\partial x}bz\sqrt{\Delta t} \qquad \textit{2.32. Eq.}$$

Now we can substitute $z^2\Delta t$ with its expected value (Δt) and take the $\Delta t \to 0$ for the whole expression we'll obtain the form presented in the lemma:

$$dG = \left(\frac{\partial G}{\partial t} + \frac{\partial G}{\partial x}a + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bz\sqrt{\mathrm{d}t} \qquad \textit{2.33. Eq.}$$

Itô's lemma can be extended to cases where G is the function of multiple variables x_1, x_2, x_3 ..., and also to cases where the variable x has multiple sources of uncertainty.

Armed with Itô's lemma the process f that is the function of r will follow

$$df = \left(\frac{\partial f}{\partial t} + m\frac{\partial f}{\partial r} + \frac{1}{2}s^2\frac{\partial^2 f}{\partial r^2}\right)dt + s\frac{\partial f}{\partial r}dW_t \qquad \textit{2.34. Eq.}$$

In the risk-neutral world if f provides no income the process has to have the form of

$$df = rfdt + \cdots \qquad \textit{2.35. Eq.}$$

So, we arrive to the criteria

$$\frac{\partial f}{\partial t} + m\frac{\partial f}{\partial r} + \frac{1}{2}s^2\frac{\partial^2 f}{\partial r^2} = rf \qquad \textit{2.36. Eq.}$$

This criterion is the equivalent formulation of the Black-Scholes-Merton differential equation. (Black & Scholes, 1973)

We can apply this result to the zero-coupon bond price $P(t,T)$ introduced earlier, to get

$$\frac{\partial P(t,T)}{\partial t} + m\frac{\partial P(t,T)}{\partial r} + \frac{1}{2}s^2\frac{\partial^2 P(t,T)}{\partial r^2} = rP(t,T) \qquad \textit{2.37. Eq.}$$

(Mamon, 2004)

Most models for the short rate r contain some assumptions about economic variables, and then they explore what the said process implies for the bond and other product prices. The simplest of these are the so called one-factor models, which assume, that r has only a single source of uncertainty. It is also assumed in these models, that the parameters contain no explicit time dependence. These restrictions lead to the usual form of

$$dr = m(r)dt + s(r)dW_t \qquad \textit{2.38. Eq.}$$

A property of all one-factor models is that all rates move at the same direction for short times, but they don't move the same amount. This results in a rich variety of term-structures. The most common one-factor models are the following:

- **Rendleman and Bartter model**, (Richard J. Rendleman & Bartter, 1980):

$$m(r) = \mu r; s(r) = \sigma r$$

- **Vasicek model** (Vašíček, 1977):

$$m(r) = \kappa(\theta - r); s(r) = \sigma$$

- **Cox-Ingersoll-Ross model** (Cox, Ingersoll, & Ross, 1985):

$$m(r) = \kappa(\theta - r); s(r) = \sigma\sqrt{r}$$

(Hull & Basu, Equilibrium Models of the Short Rate, 2018)

Properties of the CIR model

From this point onwards r will be replaced by X in the notation of the process. This change is due to the fact that the focus of this thesis is to examine the numerical properties of this stochastic process, and the implication to the interest rate derivatives are not evaluated numerically. Thus, with the new notation the Cox-Ingersoll-Ross model follows the stochastic differential equation

$$dX = \kappa(\theta - X)dt + \sigma\sqrt{X}dW_t \qquad \textit{2.39. Eq.}$$

Where κ, θ and σ are non-negative constants. This process has mean reverting property, meaning X is pulled to the "reversion level" θ with a "reversion rate" κ. The diffusion term contains a square-root dependence on X, this will ensure that the process is bounded from below by zero.

Upon examining the boundary criteria, we will find, that if $2\kappa\theta \geq \sigma^2$ the zero boundary is unreachable. This is called the Feller condition (Feller, 1951) for the CIR process. This can be rephrased if we introduce a control parameter as $\gamma = \frac{2\kappa\theta}{\sigma^2}$, then the Feller condition will be $\gamma \geq 1$ for the zero-boundary to be Unattainable, for $\gamma < 1$ the zero boundary will be Attainable. We can see these properties by evaluating the quantities introduced with *Eq.-s 2.1,2.2,2.4,2.5,2.12,2.13.* (Fáth, 2011) We will find that

$$S(x) = \int^x \exp\left\{-\int^u \frac{2\kappa(\theta - y)}{\sigma^2 y}dy\right\}du \approx \begin{cases} \ln x & if\ \gamma = 1 \\ \dfrac{1}{1-\gamma}x^{1-\gamma} & otherwise \end{cases} \quad for\ x\ small \qquad \textit{2.40. Eq.}$$

Using this the probability to reach the zero-boundary from any arbitrary bulk point is

$$u(x) = \begin{cases} 0 & if\ \gamma \geq 1 \\ finite & if\ \gamma < 1 \end{cases} \qquad \textit{2.41. Eq.}$$

Upon investigating whether the boundary can be reached in a finite expected time

$$m(z) = \frac{1}{\sigma^2 z\, s(z)} \approx \frac{z^{\gamma-1}}{\sigma^2} \quad for\ z\ small \qquad \textit{2.42. Eq.}$$

Substituting this to *Eq. 2.5*

$$M[y,x] = \frac{1}{\sigma^2\gamma}[x^\gamma - y^\gamma] \qquad \textit{2.43. Eq.}$$

Thus

$$\Sigma(l) = \int_l^x \frac{1}{\sigma^2\gamma}[x^\gamma - y^\gamma]\, y^{-\gamma} dy = \frac{x^\gamma}{\sigma^2\gamma}\begin{Bmatrix} \frac{1}{1-\gamma}[x^{1-\gamma} - l^{1-\gamma}] & if\ \gamma \neq 1 \\ \ln\frac{x}{l} & if\ \gamma = 1 \end{Bmatrix} - \frac{x-l}{\sigma^2\gamma} \qquad \textit{2.44. Eq.}$$

Taking the $l \to 0$ limit

$$\Sigma(0) = \begin{cases} \infty & if\ \gamma \geq 1 \\ finite & if\ \gamma < 1 \end{cases} \qquad \textit{2.45. Eq.}$$

In an entirely similar fashion, we can evaluate whether the bulk can be reached in finite expected time, starting from the boundary

$$\mathrm{N}(l) = \int_l^x \frac{1}{\sigma^2\gamma}[y^\gamma - l^\gamma]\, y^{-\gamma} dy = \frac{x-l}{\sigma^2\gamma} - \frac{l^\gamma}{\sigma^2\gamma}\begin{cases} \frac{1}{1-\gamma}[x^{1-\gamma} - l^{1-\gamma}] & if\ \gamma \neq 1 \\ \ln\frac{x}{l} & if\ \gamma = 1 \end{cases} \qquad \textit{2.46. Eq.}$$

With $l \to 0$

$$N(0) = \begin{cases} \infty, & if\ \gamma < 0 \\ finite, & if\ 0 < \gamma < 1 \\ finite, & if\ \gamma \geq 1 \end{cases} \qquad \textit{2.47. Eq.}$$

Based on Table 2.1 the $l = 0$ boundary based on the value of the control parameter γ can be classified as

$$boundary(l=0) = \begin{cases} Exit, & if\ \gamma < 0 \\ Regular, & if\ 0 < \gamma < 1 \\ Entrance, & if\ \gamma \geq 1 \end{cases}$$

The cumulative distribution function for a CIR process given the initial conditions X_0 and t_0 at time t is given by the formula (Cox, Ingersoll, & Ross, 1985) (Malham & Wiese, 2012)

$$P(X < x | X_0) = F_{\chi^2_\nu(\lambda)}\big(x\,\eta(t-t_0)\big) \qquad \textit{2.48. Eq.}$$

Where $F_{\chi^2_\nu(\lambda)}(x)$ is the cumulative distribution function of the non-central chi-squared distribution with a non-centrality parameter of λ and ν degrees of freedom. In this specific case $\nu = \frac{4\kappa\theta}{\sigma^2} = 2\gamma$;

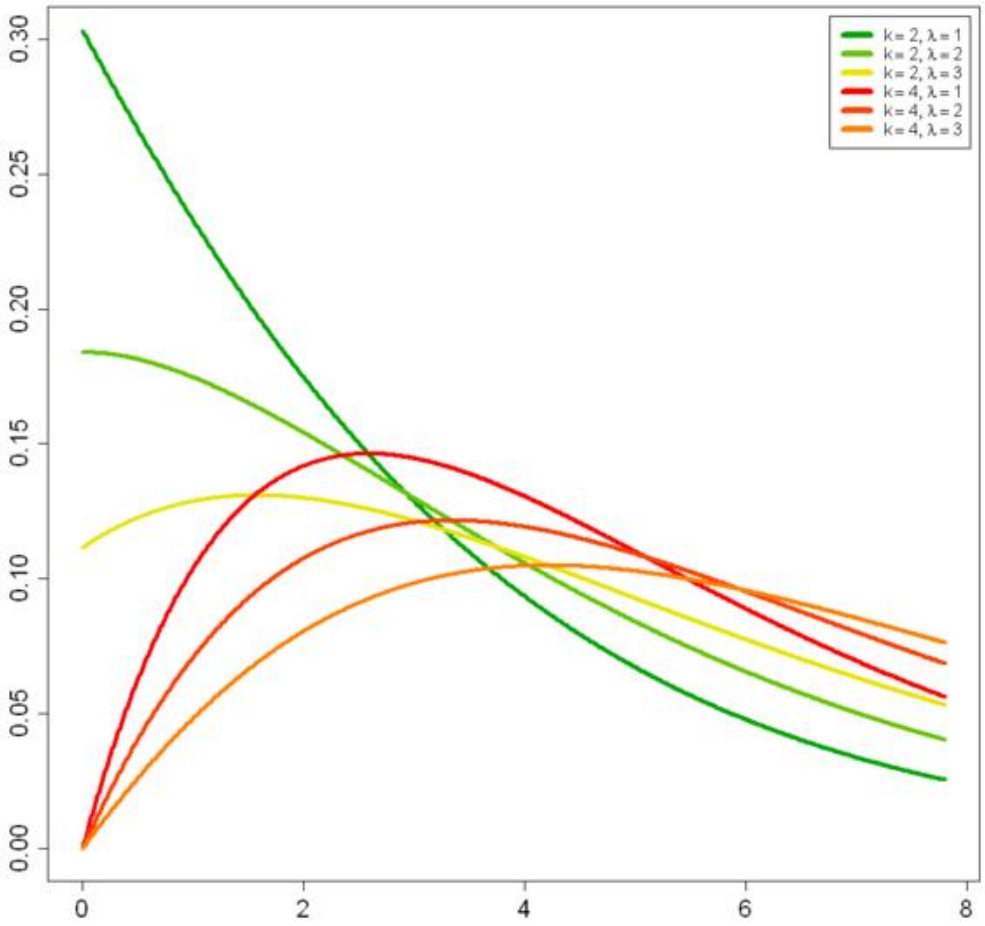
$\lambda = X_0 e^{-\kappa(t-t_0)}\,\eta(t-t_0)$, where we introduced the function $\eta(t-t_0) = \frac{4\kappa}{\sigma^2}\left(1-e^{-\kappa(t-t_0)}\right)$. This also implies that the probability density function for the same process will be

$$P(X,t|X_0,t_0) = f_{\chi^2_{\nu}(\lambda)}\left(x\,\eta(t-t_0)\right) \qquad \textit{2.49. Eq.}$$

Where $f_{\chi^2_{\nu}(\lambda)}(x)$ is the probability density function of the non-central chi-squared distribution with a non-centrality parameter of λ and ν degrees of freedom. This PDF is often written in the form

$$f_{\chi^2_{\nu}(\lambda)}(x) = \frac{1}{2}e^{-\frac{x+\lambda}{2}}\left(\frac{x}{\lambda}\right)^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}\left(\sqrt{\lambda x}\right) \qquad \textit{2.50. Eq.}$$

with $I_{\nu}(y)$ being a modified Bessel function of the first kind.



2.2. Figure: The probability density function of the Non-central χ^2 distribution for various values of degrees of freedom and non-centrality parameters. (Steiner, 2007)

For the sake of completeness, we'll now see some properties regarding the zero-coupon bond prices that assume a CIR process as the underlying driver for the short rate. The bond price will have the general form

$$P(t,T) = A(t,T)e^{-B(t,T)X(t)} \qquad \textit{2.51. Eq.}$$

With the functions of $A(t,T)$ and $B(t,T)$ being

$$A(t,T) = \left[\frac{2\sqrt{\kappa^2+2\sigma^2}e^{(\kappa+\sqrt{\kappa^2+2\sigma^2})(T-t)/2}}{\left(\sqrt{\kappa^2+2\sigma^2}+\kappa\right)\left(e^{\sqrt{\kappa^2+2\sigma^2}(T-t)}-1\right)+2\sqrt{\kappa^2+2\sigma^2}}\right]^{\frac{2\kappa\theta}{\sigma^2}}$$

$$B(t,T) = \frac{2\left(e^{\sqrt{\kappa^2+2\sigma^2}(T-t)}-1\right)}{\left(\sqrt{\kappa^2+2\sigma^2}+\kappa\right)\left(e^{\sqrt{\kappa^2+2\sigma^2}(T-t)}-1\right)+2\sqrt{\kappa^2+2\sigma^2}} \qquad \textit{2.52. Eq.}$$

Substituting *Eq. 2.51* into *Eq. 2.22* we'll get the zero rate at time t for the period $T-t$:

$$R(t,T) = -\frac{1}{T-t}\ln A(t,T) + \frac{1}{T-t}B(t,T)X(t) \qquad \textit{2.53. Eq.}$$

We can also use *Eq. 2.34* to obtain the stochastic process for $P(t,T)$:

$$dP(t,T) = X(t)P(t)dt - \sigma\sqrt{X(t)}B(t,T)P(t,T)dW_t \qquad \textit{2.54. Eq.}$$

With $t \to 0, B(t,T) \to 0$ meaning the volatility term vanishes as T is approached, leaving only the drift term. (Hull & Basu, Equilibrium Models of the Short Rate, 2018) (Cox, Ingersoll, & Ross, 1985)

In this section we have reviewed the most important analytical properties of the Cox-Ingersoll-Ross model: We started with an introduction to the Financial concept of Interest Rates, and how that relates to models predicting the evolution of the short rate. Then we've seen the boundary properties and the density function for the transition probability of the CIR model, and finally we've seen how the model is applied to the zero-bond coupon prices.

The CIR model as a squared Ornstein-Uhlenbeck process

In the $\gamma = \frac{1}{2}$ case we can connect the CIR process with the simpler Ornstein-Uhlenbeck process. We have seen in the previous section, that for $\gamma = \frac{1}{2}$ the $l = 0$ boundary behaves as a *Regular Boundary*, so we have to define the speed at the boundary $(M[\{l\}])$. This connection allows us to define $M[\{l\}]$ easily, both in the Reflective $(M[\{l\}] = 0)$ and the Absorbing $(M[\{l\}] = \infty)$

cases, and obtain the appropriate probability density function, which we can compare to the simulation results later on.

The Ornstein-Uhlenbeck model

This stochastic model was first proposed in 1930 by G. E. Uhlenbeck and L. S. Ornstein to describe the velocity of a heavy Brownian particle in an environment with friction. Since then it has seen varied use in both the field of Physics and Finance.

The corresponding SDE (Stochastic Differential Equation) (Uhlenbeck & Ornstein, 1930)

$$dY_t = -\kappa' Y_t dt + \sigma' dW_t \qquad \textit{2.55. Eq.}$$

With dW_t being the Wiener-process, and κ', σ' are positive constants. (The $'$ will help to set them apart from the similarly notated constant of the CIR process.) We can also introduce mean reversion to the drift term, and it's still considered to be an OU-process, however this version takes the form of the Vasicek model (Vašíček, 1977), which we briefly touched upon in the previous section

$$dY_t = \kappa'(\theta' - Y_t)dt + \sigma' dW_t \qquad \textit{2.56. Eq.}$$

Eq. 2.55. has a closed form solution in terms of an Itô integral (Gardiner, 2002)

$$Y_t = Y_0 e^{-\kappa' t} + \sigma' \int_0^t e^{-\kappa'(t-\tau)} dW_\tau \qquad \textit{2.57. Eq.}$$

Furthermore, we can express the Ornstein-Uhlenbeck process in terms of its transition probability, which will obey the Fokker-Planck equation (Risken, 1984)

$$\frac{\partial P}{\partial t}(Y,t) = \kappa' \frac{\partial}{\partial Y}\big(YP(Y,t)\big) + \frac{\sigma'^2}{2}\frac{\partial^2 P}{\partial Y^2}(Y,t) \qquad \textit{2.58. Eq.}$$

The solution for $P(Y,t)$ given the initial condition $P(Y_0,t_0) = \delta(Y - Y_0)$, meaning the process was started from Y_0 at time t_0

$$P(Y,t|Y_0,t_0) = \sqrt{\frac{\kappa'}{\pi\sigma'^2(1-e^{-2\kappa'(t-t_0)})}} \exp\left[-\frac{\kappa'}{\sigma'^2}\frac{\left(Y - Y_0 e^{-\kappa'(t-t_0)}\right)^2}{1-e^{-2\kappa'(t-t_0)}}\right] \qquad \textit{2.59. Eq.}$$

With general solution for any initial distribution $W(Y',t')$ the transition probability will be given by the following convolution

$$W(Y,t) = \int P(Y,t|Y',t')W(Y',t')dY' \qquad \textit{2.60. Eq.}$$

Connection to the CIR model

Generally, the two processes are connected the other way around (Jeanblanc, Yor, & Chesney, 2009) however we can also start from a Cox-Ingersoll-Ross process with constants κ, θ, σ

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t \qquad \textit{2.61. Eq.}$$

We introduce the function Y_t as $Y_t(X_t) = \sqrt{X_t}$. Note, that there is no explicit time dependence in Y_t. We can apply Itô's Lemma to obtain the SDE for Y_t. The relevant derivatives are

$$\frac{\partial Y_t}{\partial t} = 0$$

$$\frac{\partial Y_t}{\partial X_t} = \frac{1}{2}\frac{1}{\sqrt{X_t}} = \frac{1}{2}\frac{1}{Y_t} \qquad \textit{2.62. Eq.}$$

$$\frac{\partial^2 Y_t}{\partial X_t^2} = -\frac{1}{4}X_t^{-\frac{3}{2}} = -\frac{1}{4}Y_t^{-3}$$

With the mean and variance of the process being

$$a(X_t,t) = \kappa(\theta - X_t) = \kappa(\theta - Y_t^2)$$
$$b(X_t,t) = \sigma\sqrt{X_t} = \sigma Y_t \qquad \textit{2.63. Eq.}$$

Substituting in *Eq.-s 2.62* and *2.63* to the Lemma (*Eq. 2.25*), after simplifications will look like (Jeanblanc, Yor, & Chesney, 2009)

$$dY_t = \left[\left(\frac{\kappa\theta}{2} - \frac{\sigma^2}{8}\right)\frac{1}{Y_t} - \frac{\kappa}{2}Y_t\right]dt + \frac{\sigma}{2}dW_t \qquad \textit{2.64. Eq.}$$

In the special case of $\frac{\kappa\theta}{2} - \frac{\sigma^2}{8} = 0$ which in terms of $\gamma = \frac{2\kappa\theta}{\sigma^2} = \frac{1}{2}$ *Eq. 2.64* will take the form

$$dY_t = -\frac{\kappa}{2}Y_t dt + \frac{\sigma}{2}dW_t \qquad \textit{2.65. Eq.}$$

Which is an Ornstein-Uhlenbeck-type SDE with $\kappa' = \frac{\kappa}{2}$ and $\sigma' = \frac{\sigma}{2}$.

The probability density function (for the transition probability) is known for this type of SDE, it is written in *Eq. 2.59*. Substituting in the corresponding κ' and σ'

$$P(Y,t|Y_0,t_0) = \sqrt{\frac{2\kappa}{\pi\sigma^2(1-e^{-\kappa(t-t_0)})}}\exp\left[-\frac{2\kappa}{\sigma^2}\frac{\left(Y-Y_0e^{-\frac{\kappa}{2}(t-t_0)}\right)^2}{1-e^{-\kappa(t-t_0)}}\right] \qquad \textit{2.66. Eq.}$$

We can transform back this PDF to be the probability density function of X_t, which follows a CIR-type SDE using the well-known relation (Palla, 2019)

$$f_X(x) = f_Y\left(g^{-1}(x)\right)\left|\frac{d}{dx}g^{-1}(x)\right| \qquad \textit{2.67. Eq.}$$

In this case $X_t = Y_t^2$ so $g^{-1}(x) = \sqrt{x}$ and $\frac{d}{dx}g^{-1}(x) = \frac{1}{2}\frac{1}{\sqrt{x}}$. Thus, the PDF for the special case of $\gamma = \frac{1}{2}$ of the CIR process will be

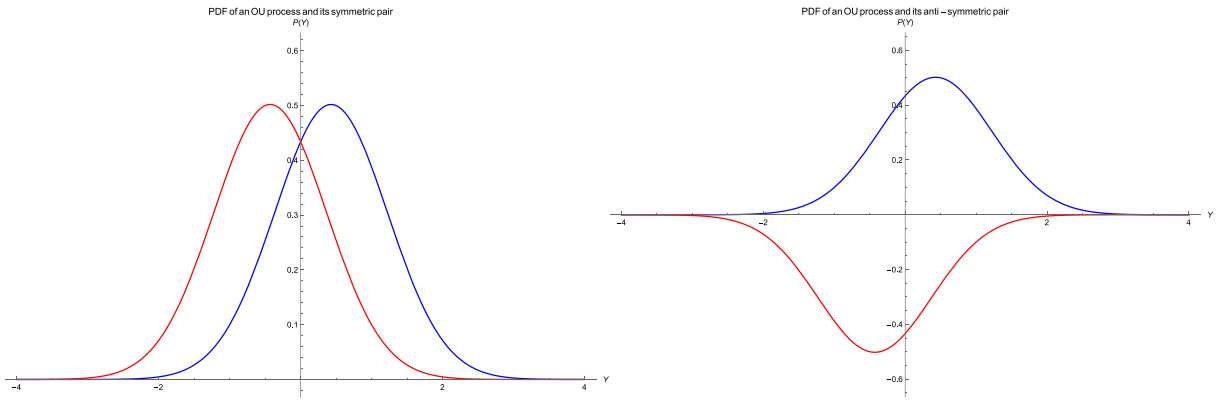
$$P(X,t|X_0,t_0) = \frac{1}{2}\frac{1}{\sqrt{X}}\sqrt{\frac{2\kappa}{\pi\sigma^2(1-e^{-\kappa(t-t_0)})}}\exp\left[-\frac{2\kappa}{\sigma^2}\frac{\left(\sqrt{X}-\sqrt{X_0}e^{-\frac{\kappa}{2}(t-t_0)}\right)^2}{1-e^{-\kappa(t-t_0)}}\right] \qquad \textit{2.68. Eq.}$$

Since this process is now the square of another process the $l = 0$ boundary will be consequently Regular with Reflective boundary condition. (Attainable in finite amount of time, and once the process reaches the boundary it bounces back to the bulk.) This can also be understood in an other way: The probability density function is inherently normalized, which in case of even partial absorption cannot happen.

The case of the Absorbing boundary condition

We have seen, that the Reflective boundary naturally arises from taking the square of the Ornstein-Uhlenbeck process, however we can also create the Absorbing boundary using this process.

We start with two OU processes, one with the initial condition of Y_0, and the other one starting from $-Y_0$. The second process will be the Anti-symmetric pair of the first one, so its PDF will have the property $-P(Y,t|-Y_0,t_0) < 0$. The two probability density functions are illustrated on *Fig. 2.3*.



2.3. Figure: The PDF of two symmetric OU processes (on the right) and the PDF of two anti-symmetric OU processes (on the left) , made with Wolfram Mathematica

We can add the two PDF-s together, the new function will take 0 value at the origin, because of the anti-symmetric construction. While the PDF-s integrated to 1 and -1 respectively, their sum

$$n := \int_0^\infty P(Y,t|Y_0,t_0) - P(Y,t|-Y_0,t_0)dY \; < 1 \qquad \textit{2.69. Eq.}$$

Will always be less than 1. Here we integrated from 0 to ∞, since the CIR process we aim to create is only defined on $[0,\infty)$[1]. In order to make this sum a proper PDF, we can use n defined in *Eq.2.69* to normalize it. Furthermore, we can define the rate of absorption at the boundary as

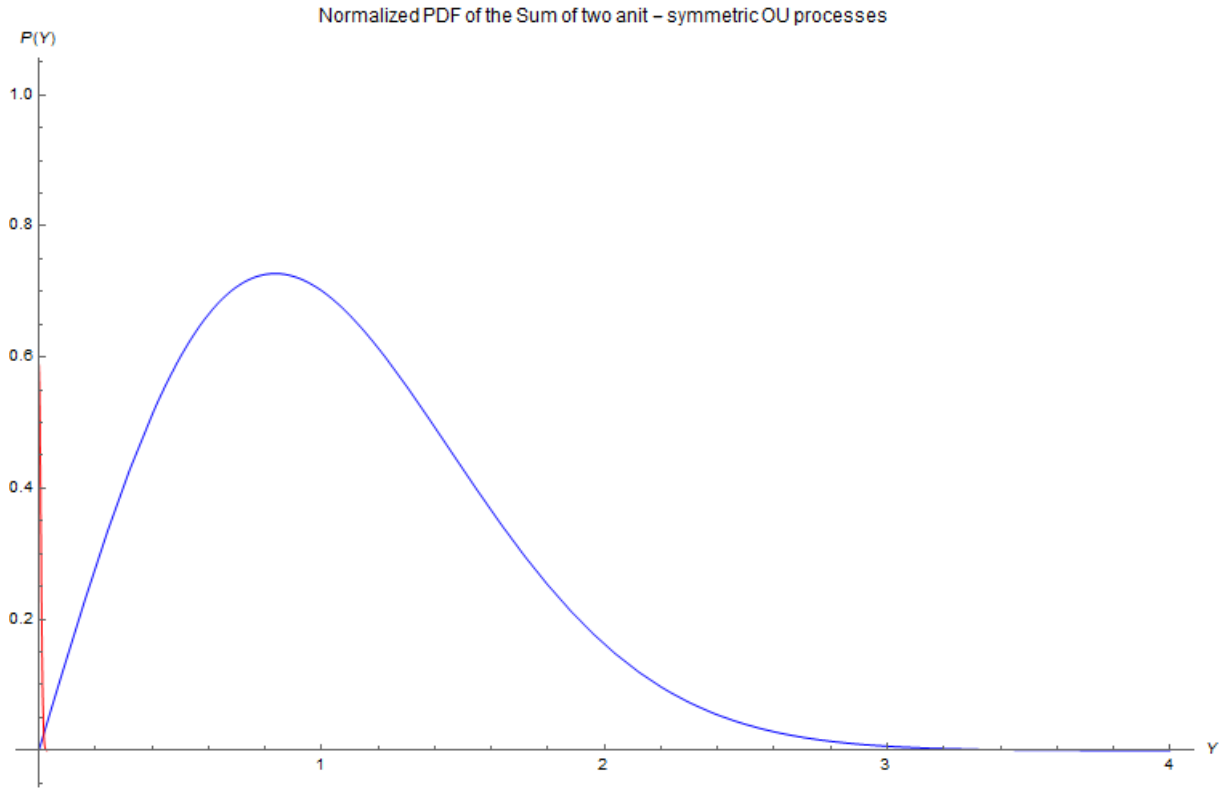
$$R_a = 1 - n \qquad \textit{2.70. Eq.}$$

Using this the "anti-symmetric" Ornstein-Uhlenbeck process' probability density function will be

$$P(Y,t|Y_0,t_0) = \frac{1}{n}\sqrt{\frac{2\kappa}{\pi\sigma^2(1-e^{-\kappa(t-t_0)})}}\left\{\exp\left[-\frac{2\kappa}{\sigma^2}\frac{\left(Y - Y_0 e^{-\frac{\kappa}{2}(t-t_0)}\right)^2}{1-e^{-\kappa(t-t_0)}}\right]\right. \\ \left. - \exp\left[-\frac{2\kappa}{\sigma^2}\frac{\left(Y + Y_0 e^{-\frac{\kappa}{2}(t-t_0)}\right)^2}{1-e^{-\kappa(t-t_0)}}\right]\right\} \qquad \textit{2.71. Eq.}$$

With n defined in *Eq.2.69*. This PDF is shown on *Fig. 2.4*.

[1] Or $(0,\infty)$ if $\gamma > 1$

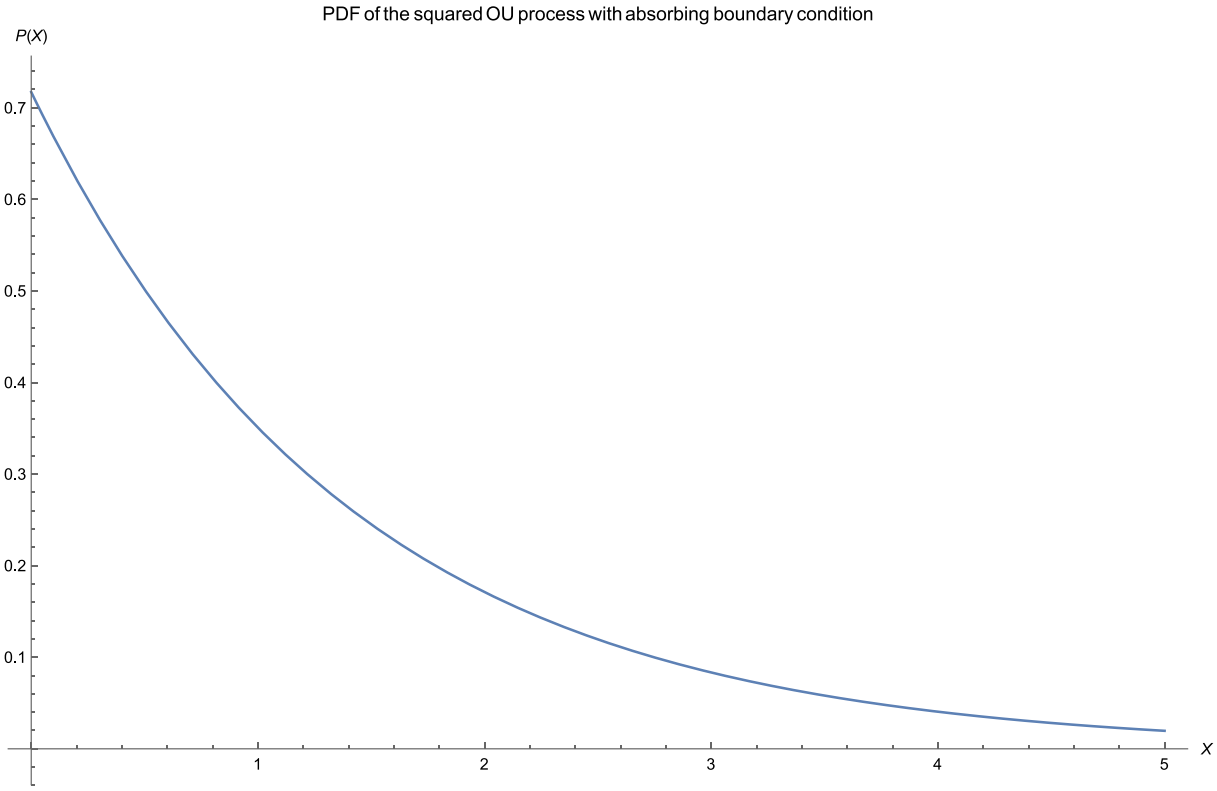


2.4. Figure: The PDF of the „anti-symmetric" OU process with the red Delta function at zero representing the trajectories absorbed in the boundary with parameters $X_0 = 0.5; \kappa = 1; \sigma = 2$ *made with Wolfram Mathematica*

Similarly, as in the Reflective case, we can transform the probability density function back to obtain the PDF for the Absorbing Cox-Ingersoll-Ross process.

$$P(X,t|X_0,t_0) = \\ = \frac{1}{2n}\frac{1}{\sqrt{X}}\sqrt{\frac{2\kappa}{\pi\sigma^2(1-e^{-\kappa(t-t_0)})}}\left\{\exp\left[-\frac{2\kappa}{\sigma^2}\frac{\left(\sqrt{X}-\sqrt{X_0}e^{-\frac{\kappa}{2}(t-t_0)}\right)^2}{1-e^{-\kappa(t-t_0)}}\right]\right. \\ \left. - \exp\left[-\frac{2\kappa}{\sigma^2}\frac{\left(\sqrt{X}+\sqrt{X_0}e^{-\frac{\kappa}{2}(t-t_0)}\right)^2}{1-e^{-\kappa(t-t_0)}}\right]\right\} \qquad \textit{2.72. Eq.}$$

This PDF can be seen on *Fig. 2.5*. Compared to the Reflective case the singularity in $X = 0$ cancels out, and the PDF goes to a constant.



2.5. Figure: PDF of the CIR Process at $\gamma = \frac{1}{2}$ *with Absorbing boundary conditions, made with Wolfram Mathematica*

It is crucial to note, that the PDF defined in Eq. 2.72 can only describe a Cox-Ingersoll-Ross process with Absorbing $l = 0$ boundary conditions if the parameters describing the process obey the criteria $\frac{\kappa\theta}{2} - \frac{\sigma^2}{8} = 0 \; or \; alternatively \; \gamma = \frac{1}{2}$.

3. Monte Carlo implementation of the Cox-Ingersoll-Ross process

I used *Python 3* to implement my simulation with commonly used libraries, such as *NumPy*, and *SciPy* for random number generation and other Mathematical operations, *Matplotlib* for visualization. I ran my code in an interactive Jupyter Notebook on Google Colab, the available computers had an Intel Xeon 2.20Ghz server grade CPU with 4 logical processor and 56 MB of Cache and 25 GB of RAM. While this was a powerful system for Monte Carlo simulations, the 12-hour runtime limit imposed on them was a hindrance to overcome.

Random number generation

When simulating stochastic processes, we have to introduce some randomness. Since computers are deterministic machines true random number generation is impossible using them. There are some examples when an external system is used to generate random numbers, a prominent instance is the lava lamps used by Cloudflare to help encrypt internet traffic (Scott, 2017) (Liebow-Feeser, 2017). The complex hydrodynamics driving a lava lamp is virtually impossible to predict, so digitalizing these patterns is a good way of obtaining true random numbers.

3.1. Figure: Lava lamps at the lobby of Cloudflare, source: (Levy, 2017)

The real problem with such systems is speed and quantity. In any Monte Carlo simulation, the user needs many random numbers, as fast as possible, so in such applications we have to resort to methods available without any external sources. Such methods are not truly random and are often called *pseudorandom number generation*. The simplest of these methods is the *linear congruent* or *power*

residue method. Using this we can generate a random sequence of numbers $r_i \in [0, M-1]$. To generate the next number in the sequence r_{i+1} one have to take the current number r_i multiply it with the constant a, add a constant c and then take the modulus by M.

$$r_{i+1} \coloneqq (ar_i + c) mod\ M = remainder\left(\frac{ar_i + c}{M}\right) \qquad \textit{3.1. Eq.}$$

This method produces uniformly distributed numbers, between 0 and $M-1$, and can be scaled to any desired range, for example to the commonly used $[0,1]$. This method can only be useful if the number M is sufficiently big, in the case of 32-bit numbers $M \approx 2^{32} = 2 \cdot 10^9$.

There are many ways to test how good a random number generation algorithm is, a simple way is calculating the correlation between near neighbors:

$$C(k) = \frac{1}{N}\sum_{i=1}^{N} r_i r_{i+k}\,, \qquad (k = 1,2,\dots) \qquad \textit{3.2. Eq.}$$

For independent and uniform random numbers with the joint probability distribution $P(r_i, r_{i+k})$ *Eq. 3.2* can be approximated as

$$\frac{1}{N}\sum_{i=1}^{N} r_i r_{i+k} \approx \int_0^1 dx \int_0^1 dy\, xyP(x,y) = \frac{1}{4} \qquad \textit{3.3. Eq.}$$

If the sequence of random numbers r_i fulfils the criterion their distribution is uniform. If the deviation from *Eq. 3.3* varies as $1/\sqrt{N}$ then the sequence is truly random.

(Landau, Páez, & Bordeianu, 2012)

Mersenne Twister Algorithm

There are more complex and accurate algorithms to generate pseudorandom numbers, than the ones outlined previously. In my simulation I used the MT19937 implementation of the Mersenne twister method[1]. It was developed in 1997 by Makato Matsumoto and Takuji Nishimura. The name comes from the fact, that the algorithm has a period length of a Mersenne prime. (Matsumoto & Nishimura, 1998)

While the linear congruent method can be easily understood with operations in the decimal number system, the Mersenne Twister algorithm uses bitwise operations. It starts with $n-1$[2] seed

[1] Since that's the one Python is using

[2] n is called the degree of recurrence

word vectors $\boldsymbol{x_0, x_1, \dots, x_{n-1}}$ that are w-dimensional row vectors over the two-element field $\mathbb{F}_2 = \{0,1\}$. The method can be summarized with the following recurrence:

$$x_{k+n} \coloneqq x_{k+m} \oplus \left(x_k^u \middle| x_{k+1}^l\right) A, \qquad (k = 0,1, \dots) \tag{3.4. Eq.}$$

In the right-hand side x_k^u denotes the "upper" r bits of the k-th word vector x_k, meaning if $x = (x_{w-1}, x_{w-2}, \dots x_0)$, then $x_k^u = (x_{w-1}, \dots x_r)$, similarly x_{k+1}^l are the "lower" r bits of the $k+1$-th word vector. This of course gives a constraint on r: $r \in [0, w-1]$. $\left(x_k^u \middle| x_{k+1}^l\right)$ notation means, that this word is "stiched together" from the upper part of the k-th seed word, and the lower part of the $k+1$-th seed word. This is then multiplied from the left by the constant $w \times w$ matrix A, which has its values in $\mathbb{F}_2$. A has the form

$$A = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ a_{w-1} & a_{w-2} & \dots & a_0 \end{pmatrix} \tag{3.5. Eq.}$$

With $\boldsymbol{a} = (a_{w-1}, a_{w-2}, \dots, a_0)$ also being a w-dimensional word vector. The matrix multiplication is carried out by the bit operation

$$\boldsymbol{x}A = \begin{cases} shiftright(\boldsymbol{x}), & if\ x_0 = 0 \\ shiftright(\boldsymbol{x}) \oplus \boldsymbol{a}, & if\ x_0 = 1 \end{cases} \tag{3.6. Eq.}$$

Then finally we add x_{k+m} to this vector bitwise to generate the new vector x_{k+n}.The subsequent vectors are taken with increasing k from 0 to 1 and so on…

Additionally, the algorithm is improved by multiplying each generated word with a $w \times w$ dimensional invertable matrix, called the "Tempering matrix" (T), which contains the following successive transformations:

$$\begin{gathered} \boldsymbol{y} \coloneqq \boldsymbol{x} \oplus (\boldsymbol{x} \gg u) \\ \boldsymbol{y} \coloneqq \boldsymbol{y} \oplus \left((\boldsymbol{y} \ll s)\ AND\ \boldsymbol{b}\right) \\ \boldsymbol{y} \coloneqq \boldsymbol{y} \oplus \left((\boldsymbol{y} \ll t)\ AND\ \boldsymbol{c}\right) \\ \boldsymbol{z} \coloneqq \boldsymbol{y} \oplus (\boldsymbol{y} \gg l) \end{gathered} \tag{3.7. Eq.}$$

Where l, s, t, u are integers and $\boldsymbol{b}, \boldsymbol{c}$ are suitable bitmasks of word size. $(\boldsymbol{x} \gg u)$ denotes a u-bit shift to the right, and $(\boldsymbol{x} \ll u)$ denotes a u-bit shift to the left. In the end the created words can be divided by $2^w - 1$ to obtain random numbers in the $[0,1]$ range.

Inverse transform sampling

So far, we have discussed how to obtain uniformly generated pseudorandom numbers using a general-purpose computer, however when simulation the Cox-Ingersoll-Ross model, one needs at the very least[1] standard normally distributed pseudorandom numbers.

In a case where the desired distribution has a known cumulative density function and said function is invertible inverse transform sampling can be used to transform the uniformly distributed numbers. (Deveroy, 1986)

The method works as follows:

1. We generate a random number u, with uniform distribution.
2. Invert the desired cumulative density function: $F_X^{-1}(x)$
3. Compute the values $F_X^{-1}(u)$, these will have the desired distribution F_X.

This method has many drawbacks, one of the most glaring is the fact that there are distributions with uninvertible CDF-s, or CDF-s that cannot be expressed in closed form. The other is the computationally expensive nature of the inversion. Several improved methods exist, such as the ziggurat algorithm, rejection sampling or the Box-Muller method.

Discretization schemes

Choosing the right discretization scheme has a huge effect on both the accuracy and the performance of the Monte Carlo simulations. When we have a one-factor SDE in the form

$$dX_t = m(X_t)dt + s(X_t)dW_t \qquad \textit{3.8. Eq.}$$

And we choose the discrete time steps to be $\Delta t = \frac{T}{N}$ where N is the size of the grid, and T is the time we want the values of X_T, the most commonly used way to create a recursion is with the Euler-Maruyama scheme:

$$X_{t+\Delta t} = X_t + \Delta X_t = X_t + m(X_t)\Delta t + s(X_t)\Delta W_t \qquad \textit{3.9. Eq.}$$

In the Case of the Cox-Ingersoll-Ross process the **Euler-Maruyama discretization** will take the form

$$X_{t+\Delta t} = X_t + \kappa(\theta - X_t)\Delta t + \sigma\sqrt{X_t}\Delta W_t \qquad \textit{3.10. Eq.}$$

[1] Some discretization schemes will call for different distributions.

Of course we can see an obvious problem with discrete equations right away: to reduce the error introduced by discretization we want $\Delta t \to 0$, which means we'll want as big of a grid as possible. The second issue is that this way we have to generate the entire trajectory, which in the case when only the value/distribution of X_T is needed seems like a waste. When examining the boundary behavior, this is of course not a problem, since the entire trajectory will be needed.

There are two approaches commonly used for measuring the error of a discretization scheme:

1. Strong Error Criterion

$$E\left[\left\|\hat{X}_{t_0+N\Delta t} - X_T\right\|\right] \qquad \textit{3.11. Eq.}$$

2. Weak Error Criterion

$$\left|E\left[f\left(\hat{X}_{t_0+N\Delta t}\right)\right] - E[f(X_T)]\right| \qquad \textit{3.12. Eq.}$$

In the case of the weak criterion only the distribution of the results is what matters.

Definition

$\hat{X}$ has a *strong order of convergence* of $\beta > 0$ if

$$E\left[\left\|\hat{X}_{t_0+N\Delta t} - X_T\right\|\right] \leq c\Delta t^{\beta} \qquad \textit{3.13. Eq.}$$

Where c is some constant and Δt is sufficiently small.

Definition

$\hat{X}$ has a *weak order of convergence* of $\beta > 0$ if

$$\left|E\left[f\left(\hat{X}_{t_0+N\Delta t}\right)\right] - E[f(X_T)]\right| \leq c\Delta t^{\beta} \qquad \textit{3.14. Eq.}$$

Where c is some constant (might be dependent on f) and Δt is sufficiently small.

In my application only the weak order of convergence matters, since the boundary behavior is tied closer to the distribution of X, than to the actual values. To that end instead of measuring the β for the weak order of convergence I will conduct a series of Kolmogorov-Smirnov test, to determine whether those distributions are coming from the expected distribution based on analytical properties.

To obtain a more complex discretization scheme, we can apply Itô's lemma to $s(X_t)$ to get a better approximation of the diffusion term in the range $[t, t + \Delta t]$. This will lead us to the so-called **Milstein Scheme** (Mil'shtein, 1974)

$$X_{t+\Delta t} = X_t + m(X_t)\Delta t + s(X_t)\Delta W_t + \frac{1}{2}s^2(X_t)\Delta t(\mathcal{N}(0,1)^2 - 1) \qquad \textit{3.15. Eq.}$$

In practice I will use an improved version of this second-order scheme called **Weighted Milstein Scheme**, which includes an extra weight scalar, α. For the CIR process the WMS discretization:

$$X_{t+\Delta t} = \frac{X_t + \kappa(\theta - \alpha X_t)\Delta t + \sigma\sqrt{X_t}\Delta W_t + \frac{1}{4}\sigma^2\Delta t(\mathcal{N}(0,1)^2 - 1)}{1 + (1-\alpha)\kappa\Delta t} \qquad \textit{3.16. Eq.}$$

(Haugh, 2017)

In the theoretical introduction chapter, I presented the transition probabilities for the CIR process with *Eq. 2.49*. Using this we can create a discretization scheme, where the random variable follows a non-central chi-square distribution instead of standard normal as in the previous examples. Based on the works of (Malham & Wiese, 2012) the discretization scheme, that I will refer to as **Non-central χ^2 Scheme**[1] will look like

$$X_{t+\Delta t} = \chi^2_\nu(\lambda)\frac{e^{-\kappa\Delta t}}{\eta(\Delta t)} \qquad \textit{3.17. Eq.}$$

Where $\nu = \frac{4\kappa\theta}{\sigma^2} = 2\gamma$ denotes the degrees of freedom, the non-centrality parameter $\lambda = X_t\eta(\Delta t)$ and the function $\eta(\Delta t) = \frac{4\kappa}{\sigma^2}e^{-\kappa\Delta t}(1 - e^{-\kappa\Delta t})$. Such model can present significant accuracy advantages compared to the previous two, however obtaining random numbers from the non-central χ^2distribution might be more time consuming. Furthermore, this discretization scheme is designed in a way, that the $l = 0$ boundary, when Attainable will be highly Reflective, so this model cannot be used to investigate Absorbing boundary conditions for $\gamma < 1$.

Handling the boundaries

As mentioned earlier for $\gamma < 1$, when the $l = 0$ boundary is Regular, boundary conditions must be specified. While the Non-central χ^2 scheme contains a strongly Reflective boundary condition, the Euler-Maruyama and Weighted Milstein schemes don't contain any conditions.

[1] Since this scheme is using the exact transition probability, it's often referred to as "Exact Monte Carlo" in the literature.

Absorbing Boundary

We can have Absorbing $l = 0$ boundaries in a simulation, if in the case when $X_{t+\Delta t} \leq 0$, we specify $X_{t+\Delta t} \coloneqq 0$ and terminate the process, since for every next time step it will stay 0, as it was absorbed by the boundary.

The rate of absorption can be calculated by dividing the number of terminated process with the number of all processes:

$$R_a = \frac{N_{absorbed}}{N_{all}} \qquad \textit{3.18. Eq.}$$

In the case of $\gamma = \frac{1}{2}$ we can also use the calculated transition probability of the squared Ornstein-Uhlenbeck process with Absorbing boundary condition (*Eq. 2.72.*), which again contains its condition by construction, so no "exception handling" is required.

Reflective Boundary

If we want to have a Reflective boundary condition for the Euler or Milstein schemes, when the trajectory $X_{t+\Delta t} \leq 0$ we restart the process from 0, meaning $X_{t+\Delta t} \coloneqq 0$ and then in the next step the drift term will immediately increase this value by $\kappa\theta\Delta t$. Since when the boundary is Attainable it is Regular, restarting from 0 poses no problem, as Regular boundaries have the property, that the bulk is reachable in finite time starting from them.

For the Non-central χ^2 and the squared Ornstein Uhlenbeck processes, the transition probabilities naturally contain the Reflective boundary property.

Improving performance

Choosing the right discretization scheme can aid or hinder the performance of the simulation, but it is not the only way to influence, and potentially improve the runtime.

One of the standard techniques is called the *Antithetic variable method*. Since generating a large enough sample size of standard normally distributed random numbers can get quite expensive, we can multiply the sample with (-1) to double the sample size and calculate every trajectory again, with the new random numbers.

Another common method is called *importance sampling*, in this case only the relevant part of the distribution is used, the rest is taken into account by weighing with its probability. (Sipos, 2018)

Parallel processing

Since in the one-factor SDE-s the generated trajectories do not interact with each other in any form, it is not required to generate them on the same logical processor, and the simulation process can be divided up to as many logical processors as possible. This will drastically increase the speed of the simulation and diminishing returns only surfaces on very high core counts, when the results have to be collected together for further evaluation.

Technically there are two different solutions[1][2], to divide up the workload of a program, one is called multi-threading, the other is multi-processing. The main difference between the two, is that in the case of multi-threading the program still runs as one process, and the different threads can trade variables with each other if necessary. This is advised to use, when there is some interaction involved between the trajectories. In the case of multi-processing the program spawns multiples processes, one for each available core. In this case it is crucial to terminate the processes, otherwise the program can get stuck. In practice modern CPU-s execute more processes than the number of their logical processors, as they can quickly change between multiple of them.

In my implementation I used Python 3's built in multiprocessing library, which felt more accessible, but this came at the cost, that my simulation code doesn't run in Windows environments, since the Windows implementation of said library differs from the UNIX one.

Series acceleration: Richardson extrapolation

As we've already seen on *Fig. 3.2* the shape and values of the CIR Phase diagram are dependent on the step size Δt. One way to determine the "true" shape and values in the $\Delta t \to 0$ limit, is to view every γ value as a series in Δt that converges to a constant. The limits then can be determined by a series acceleration algorithm, in my case I choose Richardson extrapolation. (Richardson, 1911)

If we want to approximate the value A^* and we have a series $A(h)$ that depends on a small parameter h polinomially

$$A(h) = A^* + Ch^n + \mathcal{O}(h^{n+1}) \qquad \textit{3.19. Eq.}$$

Then we can define the Richardson extrapolation of $A(n)$ as

[1] See https://docs.python.org/3/library/multiprocessing.html
[2] And https://docs.python.org/3/library/threading.html

$$R(h,t) \coloneqq \frac{t^n A\left(\frac{h}{t}\right) - A(h)}{t^n - 1} \qquad \textit{3.20. Eq.}$$

Where h and $\frac{h}{t}$ are distinct step sizes. Substituting in Eq. 3.19 to Eq. 3.20 we can see

$$R(h,t) = \frac{t^n\left(A^* + C\left(\frac{h}{t}\right)^n + \mathcal{O}(h^{n+1})\right) - \left(A^* + Ch^n + \mathcal{O}(h^{n+1})\right)}{t^n - 1} = \qquad \textit{3.21. Eq.}$$
$$= A^* + \mathcal{O}(h^{n+1})$$

that we transformed out the lowest order h dependence from $A(h)$ with this method.

4. Simulation Results

In this chapter I will discuss the results from my implementation of the Cox-Ingersoll-Ross process, starting from justifying the choice of discretization, then comparing the distributions with their analytical counterpart at $T = 1$. Afterwards I will Show why $P\left(T_{2X_0} > \tau_0\right)$ is a good order parameter for the Reflective boundary condition, and finally I will expound the Phase diagrams for both the Reflective and Absorbing boundary conditions.

Choosing the right discretization scheme

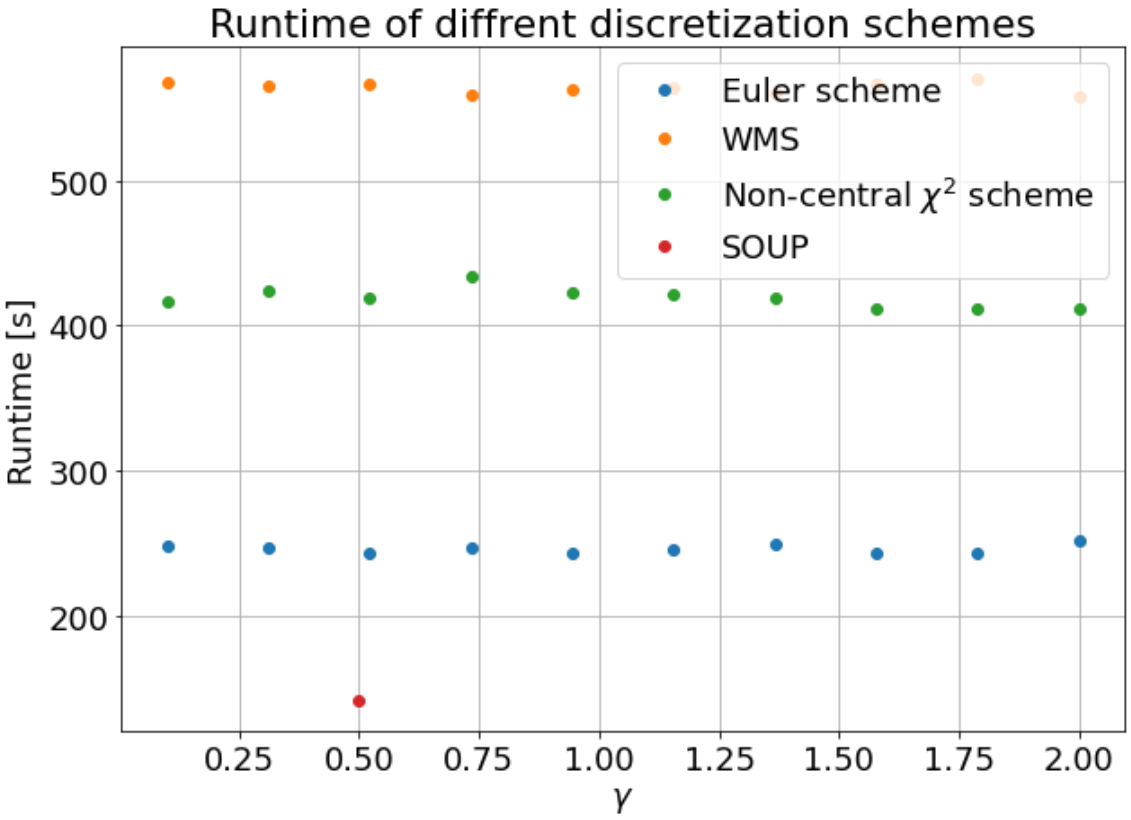
In an ideal world the choice of discretization should only have one goal: to get the most accurate[1] result. However in real world applications we have to take into account the limitations of the computing solutions, namely the 12-hour runtime limit of Google Colab. This means the right choice will be an appropriate compromise between speed and accuracy.

Runtime

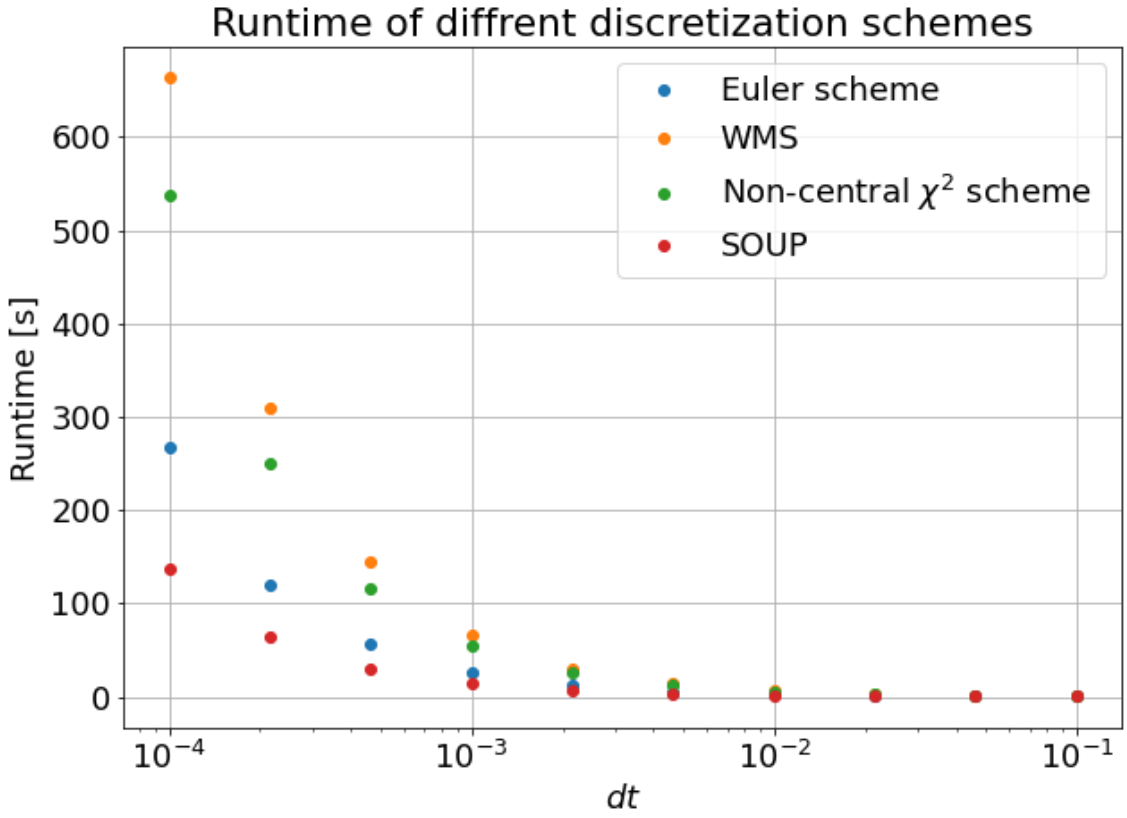
On *Fig.-s 4.1* and *4.2* we can see how the runtime depends on the value of γ for the different schemes, and how the choice of step size (and subsequently the size of the grid) influences the runtime.

We can see, that apart from minor fluctuations, the value of γ has little influence on the runtime of the simulation, however this gives us a great opportunity to compare the different schemes with one another.

[1] Closest to the analytical in the $\Delta t \to 0$ limit



4.1. Figure: Influence of γ on the runtime for different discretization schemes



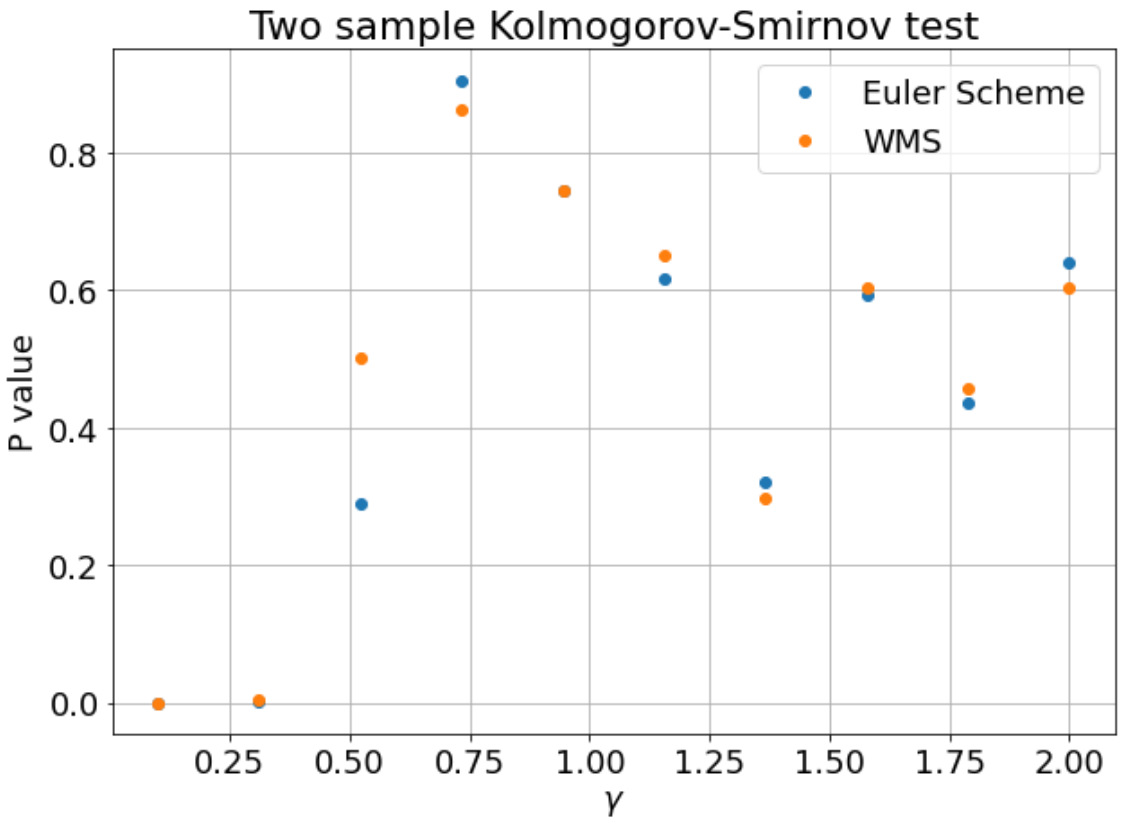
4.2. Figure: Influence of Δt on the runtime for different discretization schemes at $\gamma = 0.5$

The fastest scheme by an enormous margin is the Squared Ornstein-Uhlenbeck process (or SOUP for short), the only issue is, that this scheme cannot be used for other values of gamma, only for $\gamma = \frac{1}{2}$. This is followed by the Euler-scheme, then the Non-central χ^2 scheme, and finally the slowest one was the second order scheme.

As expected, the choice of the step size Δt exponentially increases the runtime for all schemes, and comparing them the same tendencies appear, as mentioned previously.

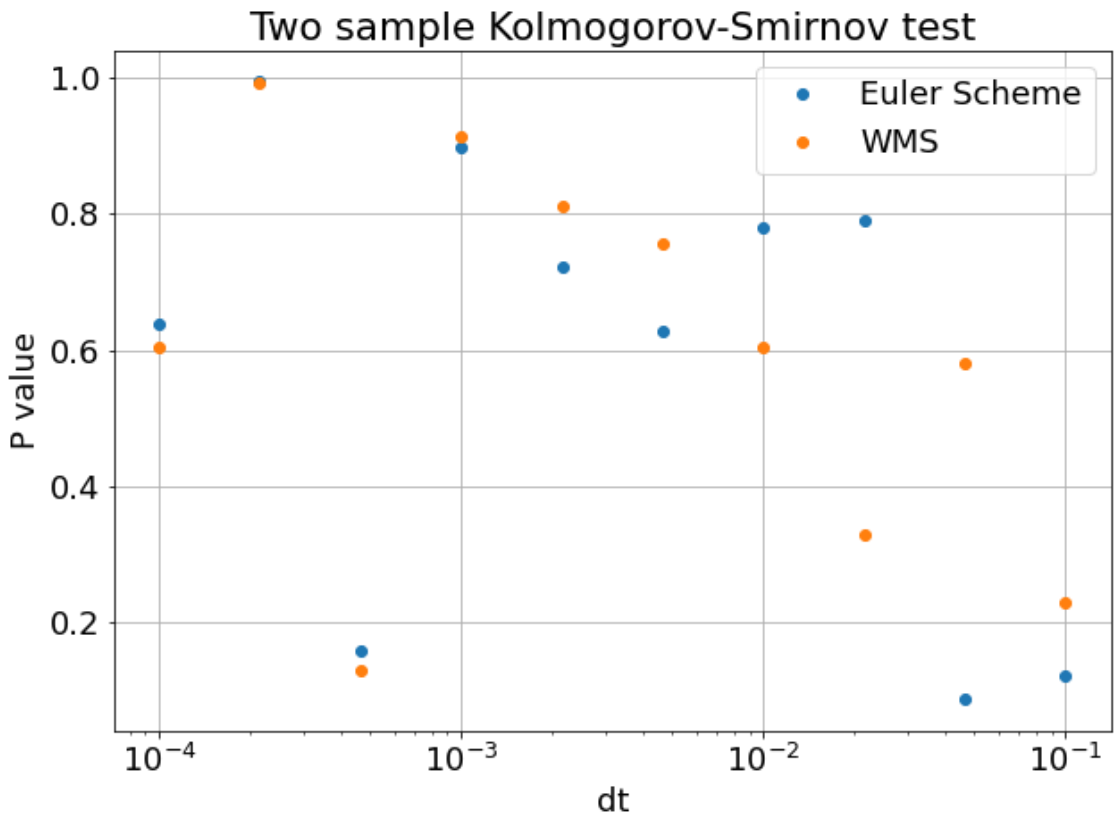
Accuracy

To quantify the difference between the accuracy of the models I used two-sample Kolmogorov-Smirnov statistical test, to determine whether the result of two simulations can originate from the same distribution. On every case one of the samples was the Non-central χ^2 scheme[1], as it has the exact transition probability, while the second sample was the model to be examined.



4.3. Figure: Influence of γ on the accuracy for different discretization schemes

[1] With this decision I assumed this scheme is the most accurate. This is not a far-fetched assumption, it is based on the works of (Malham & Wiese, 2012)



4.4. Figure: Influence of Δt on the accuracy for different discretization schemes at $\gamma = 2$

On *Fig. 4.3* we can see, that the "conventional" discretization schemes work better for $\gamma > 1$ then start to struggle, when the boundary becomes Attracting, and the boundary behavior has to be taken into account.

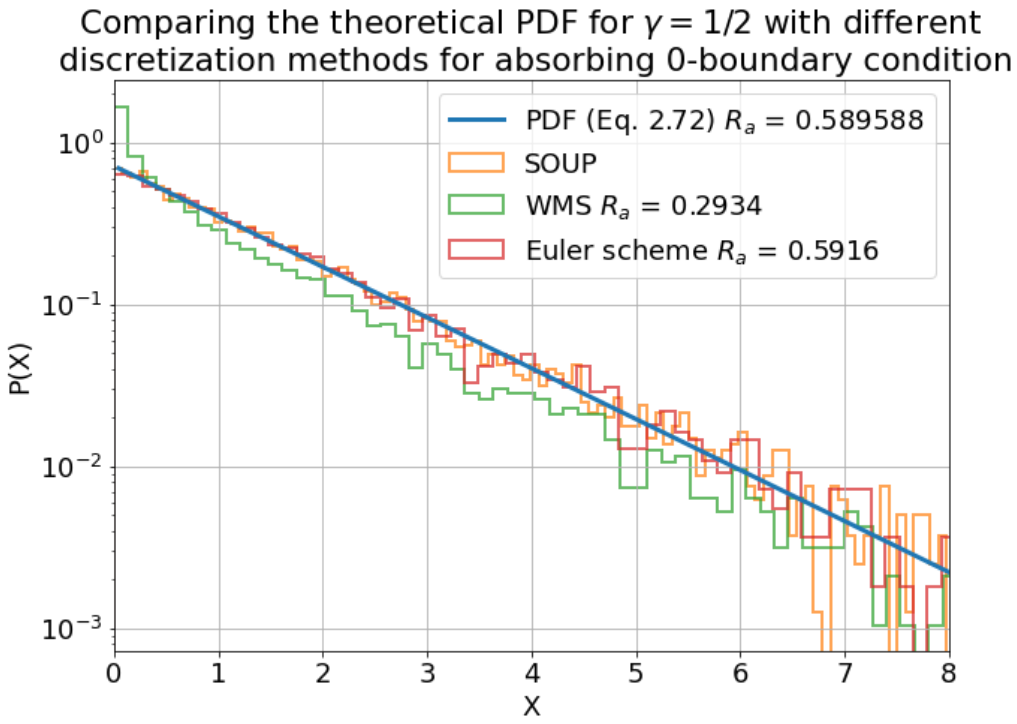
Fig. 4.4 shows that both models are getting better with the increase of the grid size, which can be easily understood as them getting closer-and closer to the exact transition probability, which the non-central χ^2 method knows for every step size.

Based on the findings presented here for the Reflective $l = 0$ boundary condtion the non-central χ^2 method shall be used, as it gives the best accuracy at the cost of approximately double the runtime of the Euler method. For Absorbing boundary conditions, the Euler method will be chosen, this choice shall become clear in the next section.

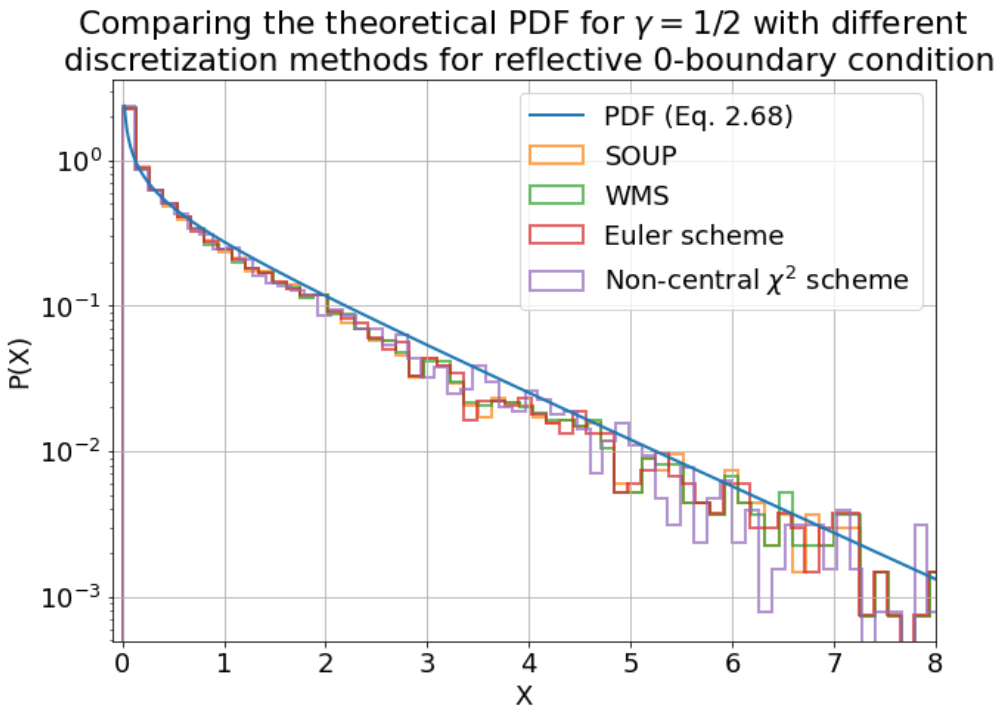
Comparing the distributions

Now we will compare the distributions of X after $T = 1$ with their analytical counterpart described in Chapter 2. At $\gamma = \frac{1}{2}$ we'll see both the reflecting and the Absorbing boundaries effect on the distribution, as well as we can see how the squared Ornstein-Uhlenbeck process holds up,

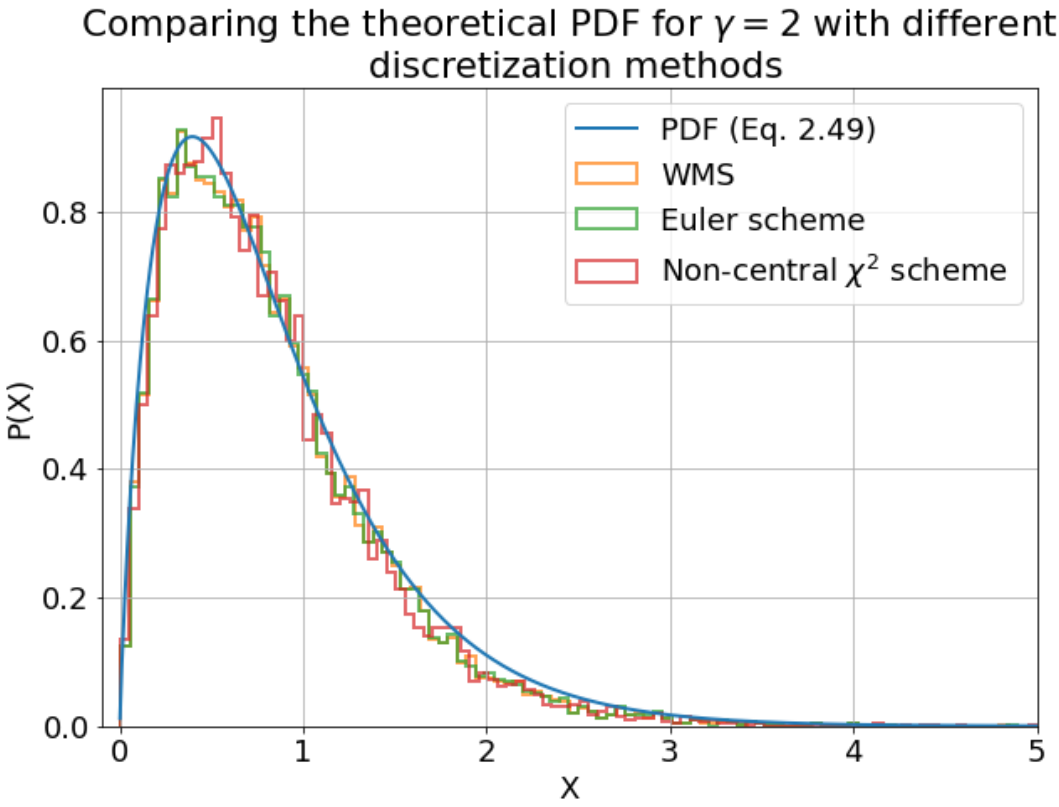
and at $\gamma = 2$ we'll see ad example of the Entrance type boundary, where extra conditions aren't required.



4.5. Figure: Comparing the analytical PDF for Absorbing 0-boundary.



4.6. Figure: Comparing the analytical PDF for Reflective 0-boundary.



4.7. Figure: Comparing the analytical PDF for $\gamma = 2$

At $\gamma = \frac{1}{2}$ for Absorbing boundary condition we can see, that the Euler-Murayama and the Squared-Ornstein-Uhlenbeck methods fit really well to *Eq. 2.72* with $\kappa = 1; \theta = 1; X_0 = 0.5, t_0 = 0, T = 1$. The second order process estimates more trajectories close to the boundary, than expected. Here we can also compare the rate of absorption, which for this case analytically $R_a^{analytical} = 0{,}589588$. The Euler method has $R_a^{Euler} = 0.5916$ and the second order scheme has a very low $R_a^{WMS} = 0.2934$. The Squared OU-process has no rate of absorption, as the transition probability already contains $n = 1 - R_a^{analytical}$ as a normalization factor. We can also see from this figure why R_a might be a good order parameter for the Absorbing case: It will have a finite value if the boundary is Attracting and it will always be zero for Non-attracting boundaries. With R_a chosen as the order parameter the Euler-Maruyama method seems like a good fit to simulate the entire Phase Diagram for this case, as it is already quite close to the analytical value at $\Delta t = 10^{-4}$.

On *Fig. 4.6* it is very hard to discern between the different discretization schemes, luckily from the Kolmogorov-Smirnov test we do know, that the Non-central χ^2 scheme is the best one to use for Reflective $l = 0$ boundary.

For the $\gamma = 2$ case we can see on *Fig. 4.7*, that all the methods seem to do quite well, however we also know, that the $\gamma > 1$ cases are less interesting, as it's almost guaranteed that the order parameter of our choice will be 0 here.

Measuring the time to cross barriers

For the Reflective boundary condition, I have chosen $P(\tau_{2X_0} > \tau_0)$ as the order parameter. This a special case of $u_{a,b}$ defined in *Eq. 2.1*. This quantity measures the probability, that the first time the trajectory reaches the value $2X_0$ is bigger, than the first time the trajectory reaches the $l = 0$ boundary. These hitting times are denoted with τ_{2X_0} and τ_0 respectively. This quantity behaves well, since if the boundary is Non-attracting $\tau_0 = \infty$, so the probability will always be 0, while when the boundary is Attracting this probability will be finite.

Such a choice allows us to improve the performance of the Monte Carlo simulation. We only need to run a simulation until it reaches one of the barriers, since the hitting time for the second barrier shall be larger. On *Additional_Fig.-s 1-4* we can see how much sooner the simulations can stop, saving a substantial amount of time. In the $\gamma < 1$ case both τ_{2X_0} and τ_0 are much smaller than $T = 1$, offering significant performance benefits. When $\gamma > 1$, $\tau_0 = \infty$, so the only benefits of early stopping can come from τ_{2X_0}. In this case there are trajectories that don't hit any of the barriers until $T = 1$, so the performance benefits are smaller, than compared to the other case.

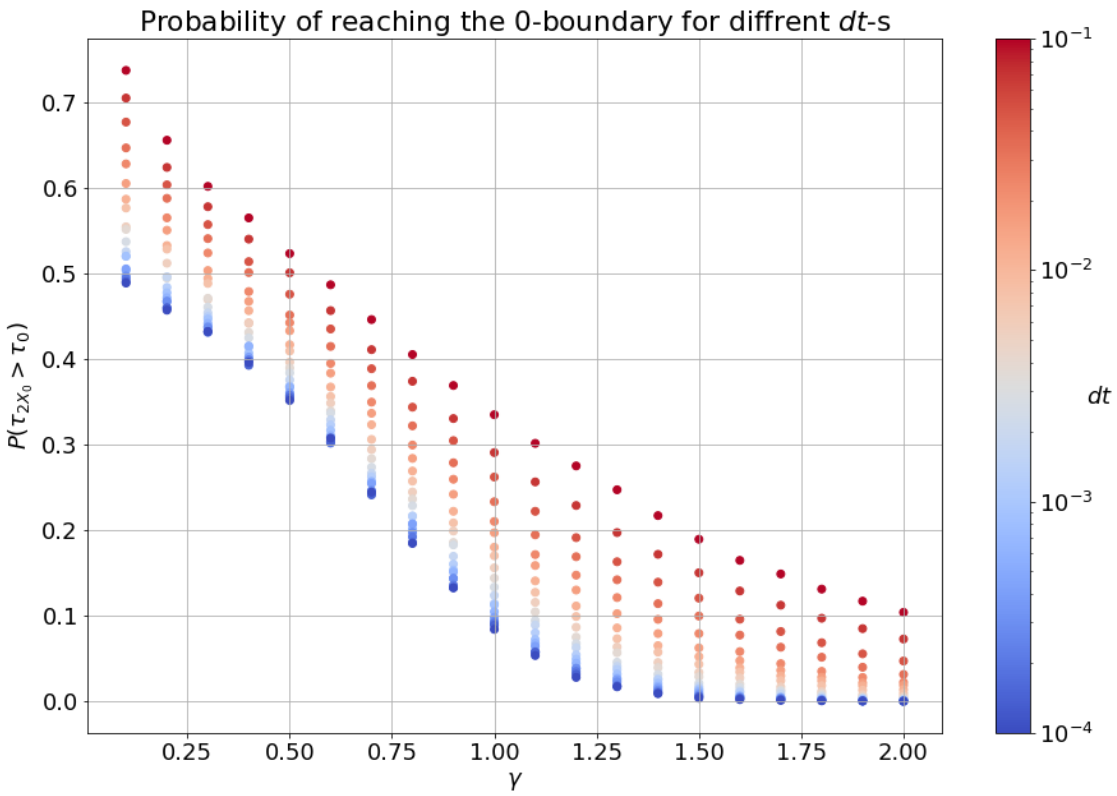
Finding γ_c numerically

Reflective Boundary

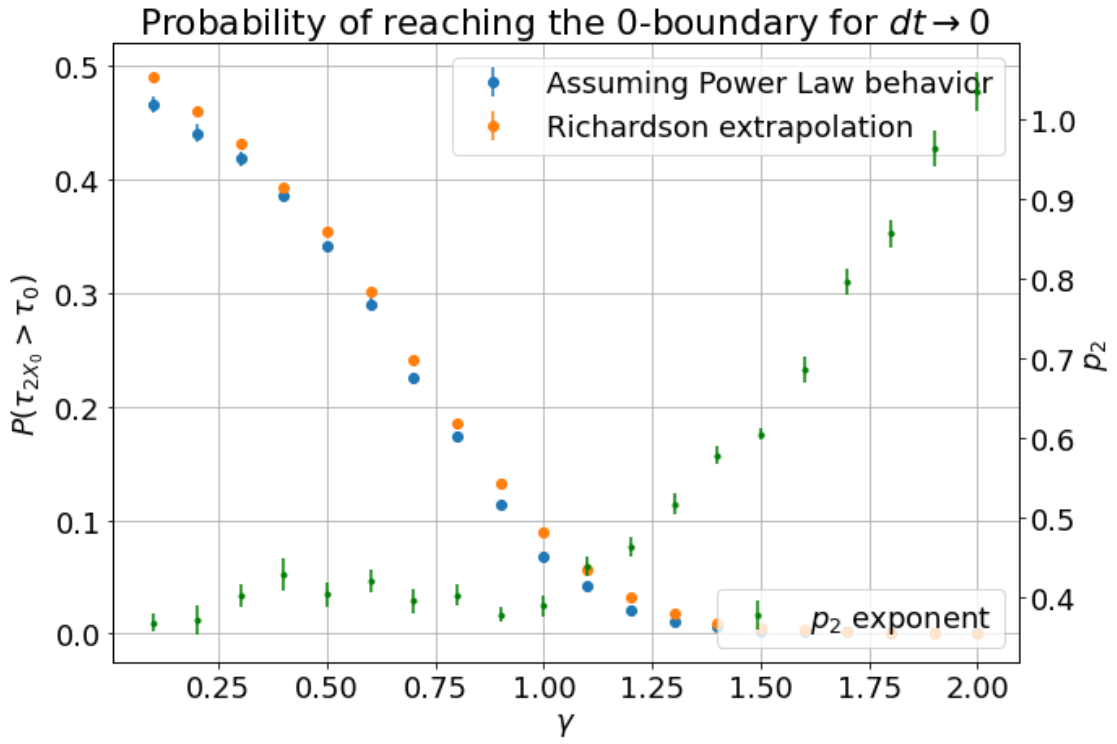
When we enforce a Reflective boundary condition for the $l = 0$ boundary the non-central χ^2 method with variables $\kappa = 1, \theta = 1, X_0 = 0.5, T = 1$ we'll get the discretization dependent Phase Diagram of *Fig. 4.8*. We can see, that as the resolution is improving, the Diagram edges ever closer to the $\Delta t \to 0$ limit.

Additional_Fig.-s 5-10. provide insight into the nature of this convergence. It was assumed, that this is a Power Law-like behavior in Δt, so the following function was fit:

$$P(\tau_{2X_0} > \tau_0 | \gamma, \Delta t) = P(\tau_{2X_0} > \tau_0 | \gamma) + p_1 \Delta t^{p_2} \qquad \textit{4.1. Eq.}$$



4.8. Figure: Phase Diagram for the Reflective boundary for different Δt-s



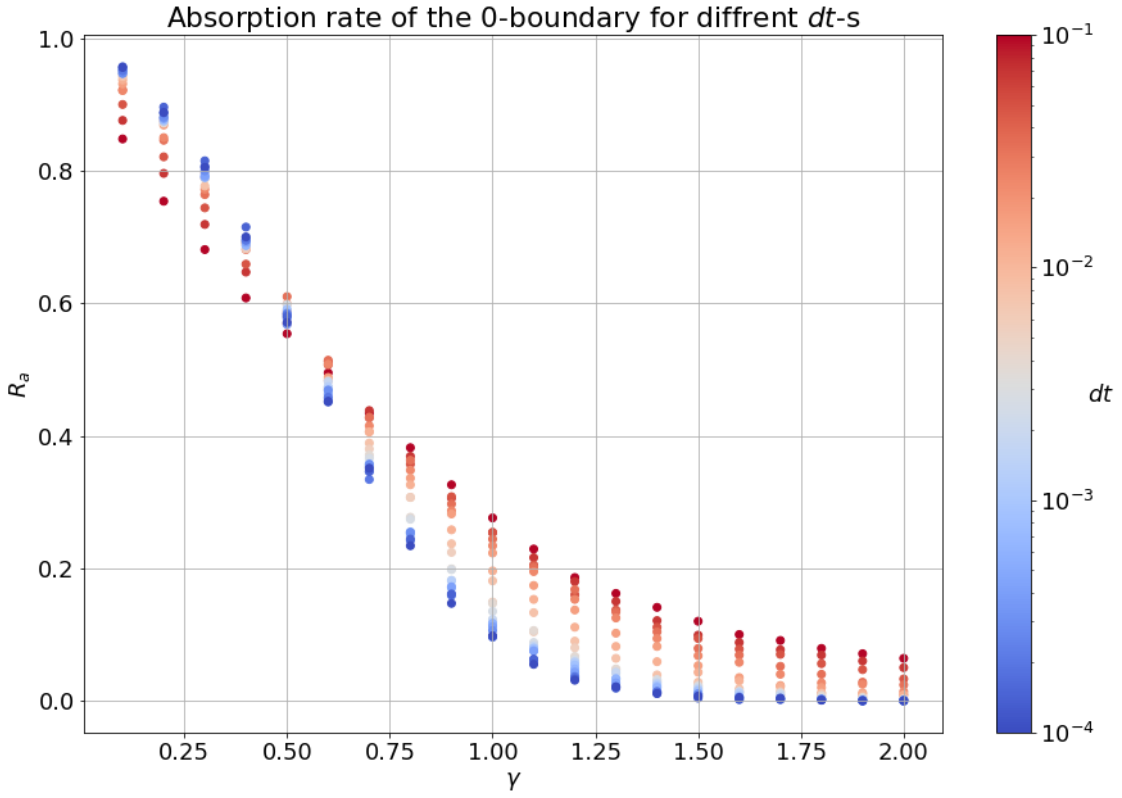
4.9. Figure: Extrapolated Phase Diagram for the Reflective boundary, with the right axis containing the p_2 exponent of the fitted Power Law. Note that it is always > 0 and starts linearly increasing around the analytical critical point. ($\gamma_c = 1$)

Since $p_2 > 0 \ \forall \gamma$, this Power Law term vanishes in the $\Delta t \to 0$ limit. To obtain the parameters of the Power Law (p_1, p_2) a linear fit was done on a log-log scale. Then using these as starting parameters the Power Law like form with a constant $\left(P\left(\tau_{2X_0} > \tau_0 | \gamma\right)\right)$ was fit. I have also performed Richardson extrapolation to obtain said limiting Phase Diagram

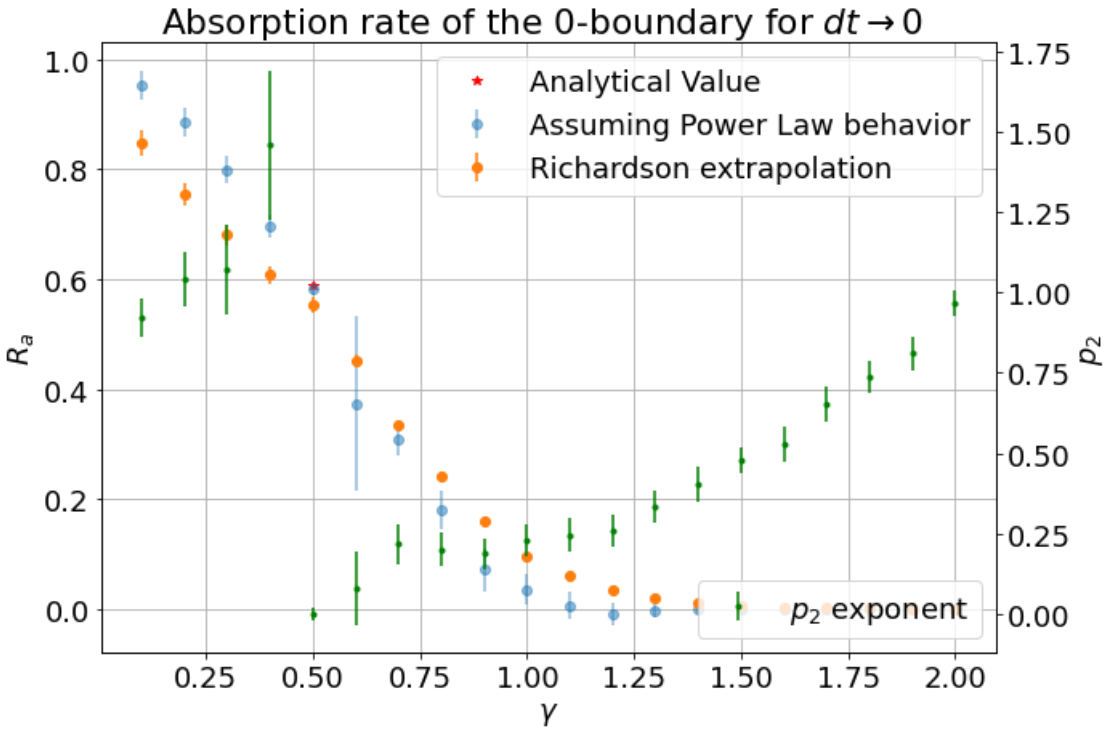
Fig. 4.9 contains the two different extrapolations to the analytical Phase Diagram. The points contain the Monte Carlo error, and in case of the function fit the error of the fitting was also displayed. We can see, that the Richardson extrapolation only ensures $\gamma = 1.5$ as a clear boarder between Attracting $l = 0$ boundary and Non-attracting boundary, while according to the function fit it is at $\gamma = 1.3$.

Absorbing Boundary

In the case of Absorbing $l = 0$ boundary using Euler-Maruyama discretization with the variables $\kappa = 1, \theta = 1, X_0 = 0.5, T = 1$ the discretization-dependent Phase Diagram is illustrated on *Fig. 4.10*. Here we can see, that the convergence changes direction from upwards to downwards at $\gamma = \frac{1}{2}$. This can also be seen on *Additional Fig.-s 11-18.* on a γ-per-γ basis.



4.10. Figure: Phase Diagram for the Absorbing boundary for different Δt-s



4.11. Figure: Extrapolated Phase Diagram for the Absorbing boundary, with the right axis containing the p_2 exponent of the fitted Power Law. Note that in this case the monotonic increase seem to start at $\gamma = 1/2$

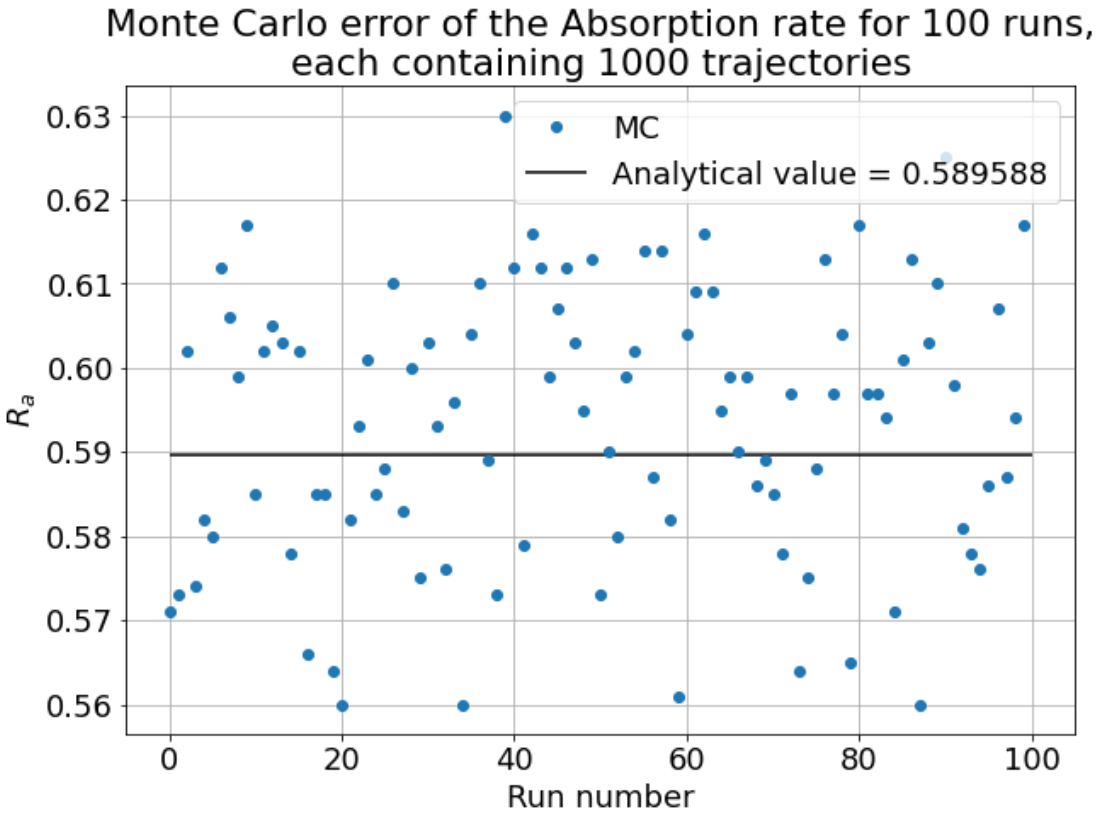
In this case similar Power Law fitting was used next to the Richardson extrapolation, exploiting the nature of the convergence. Here based on the Richardson extrapolation the Phase transition is at $\gamma = 1.5$, but it handles the $\gamma < \frac{1}{2}$ cases poorly, it couldn't deal with the change from downwards to upwards. The function fitting handled this better, putting the transition to $\gamma \approx 1.1$, and having no problem with the change, however since in this case less trajectories were used both the fitting and the Monte Carlo error were larger.

Regarding Monte Carlo error

Although I've already touched upon the accuracy of the models, the Absorbing case of $\gamma = \frac{1}{2}$ gives a great opportunity to estimate both the absolute error of the simulation, and the Monte Carlo error, which is dependent on the number of trajectories used in the simulation.

As I've mentioned the analytical value of the Absorption rate for $\gamma = \frac{1}{2}$ is 0.589588, the same value using Euler discretization is 0.571. This means an $\sim 3\%$ absolute error. Changing the seed and repeating the simulation with 1000 trajectories 100 times, the Standard Deviation from the Mean of this value is $2.717\ \%$. If we increase the number of trajectories to 10000 and repeat the experiment 10 times with different seeds, the Standard Deviation will reduce to 0.932%

precisely following the $1/\sqrt{N}$ rule, as expected. In the previous section it was assumed, that when a simulation was done with 10000 trajectories the Monte Carlo error will be $\sim 0.9\%$ and when 1000 trajectories were used the error will be $\sim 2.7\%$ based on these results, since repeating the error analysis for every point would be unfeasible.



4.12. Figure: The varying value of R_a when the seed of the simulation is changed.

5. Conclusion

In this thesis I set out to examine whether the analytical boundary properties are adequately replicated in the Monte Carlo simulations. First, I introduced the important notions of Feller's Boundary classifications, then I discussed the properties of the Cox-Ingersoll-Ross model and showed the analytical boundary behavior based on Feller's notions. I also showed how can one connect the CIR model to the Ornstein-Uhlenbeck process, and how this connection can be used to introduce boundary behavior into the PDF of this process.

I compared multiple discretization schemes, and introduced two different possible order parameters, the probability $P\left(\tau_{2X_0} > \tau_0\right)$ for Reflective boundaries and the absorption rate R_a for Absorbing boundaries. I've seen how the Phase Diagram is influenced by the step size of the discretization of the processes, and suggested two extrapolation methods, to find the $\Delta t \to 0$ limit.

For the Richardson extrapolation the results showed, that much finer step size would be required to obtain sensible results, while the Power Law fitting method, albeit it performed better, still failed to reproduce the expectations. The order parameters are continuous with a linear increase in the $\gamma < 1$ domain, meaning that this is a second-order phase transition. The methods suggested that the phase transition occurs at $\gamma = 1.1; 1.3; 1.5$ depending on method and boundary condition, the reason for such uncertainty in numerical simulations can be due to singularities at the critical point.

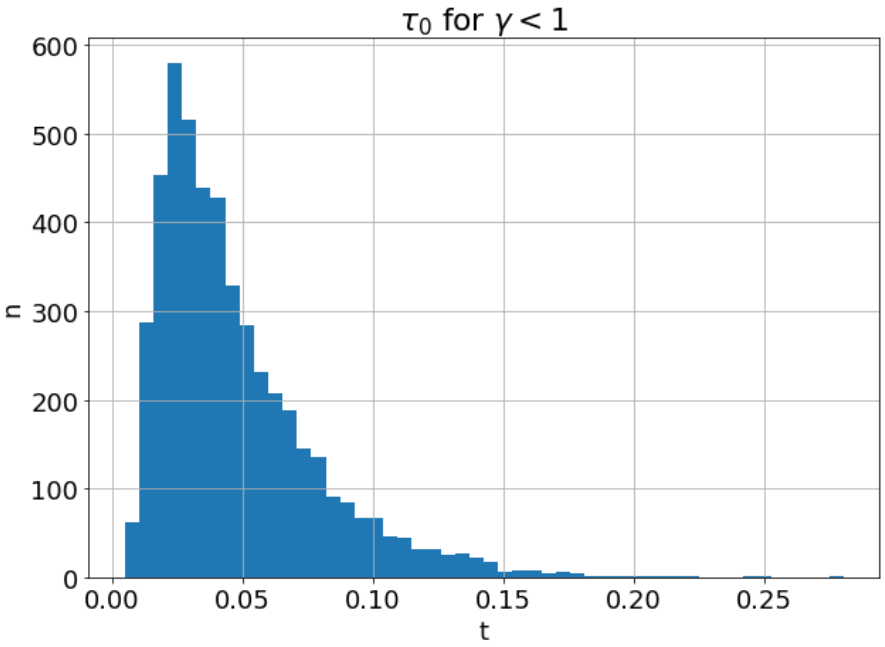
Future examinations would need a much higher focus on the way to extrapolate the $\Delta t \to 0$ limit of the Phase Diagram from the simulation data, as well as increasing the Monte Carlo trajectories to the order of 100k to reduce the Standard Deviation further, and smaller step sizes, so the extrapolation algorithm can use a wider span of data.

Appendix A.: Code

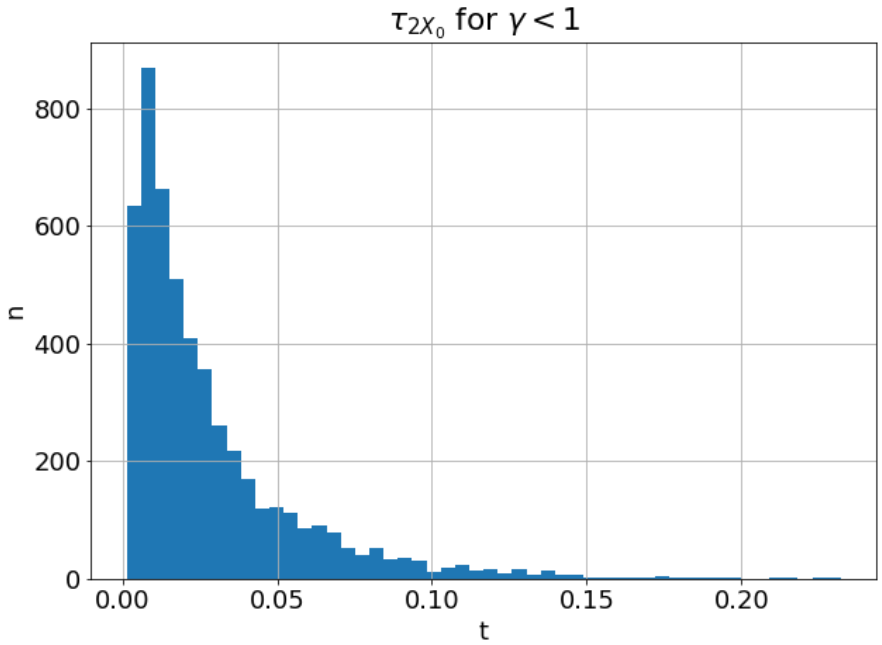
All simulation code is available on my GitHub at https://github.com/baskayj/Boundary-behaviour-in-stochastic-differential-equations-used-in-Finance

Appendix B.: Additional Figures

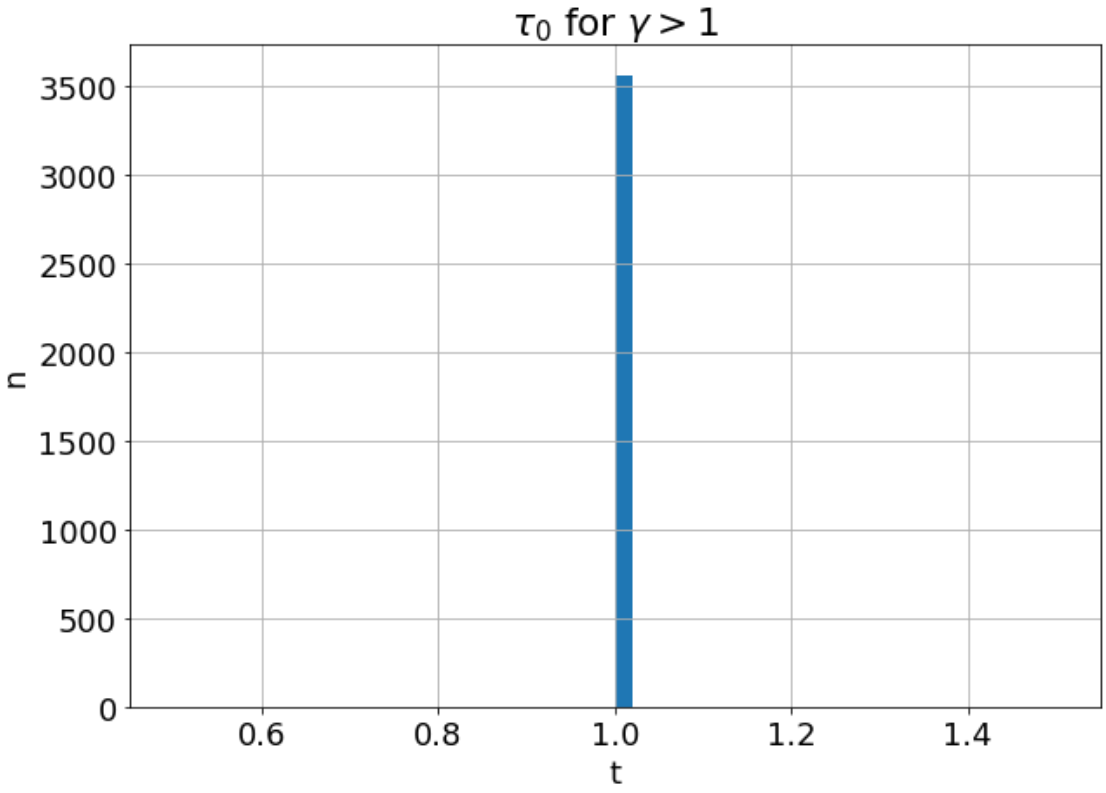
Time to cross the barriers



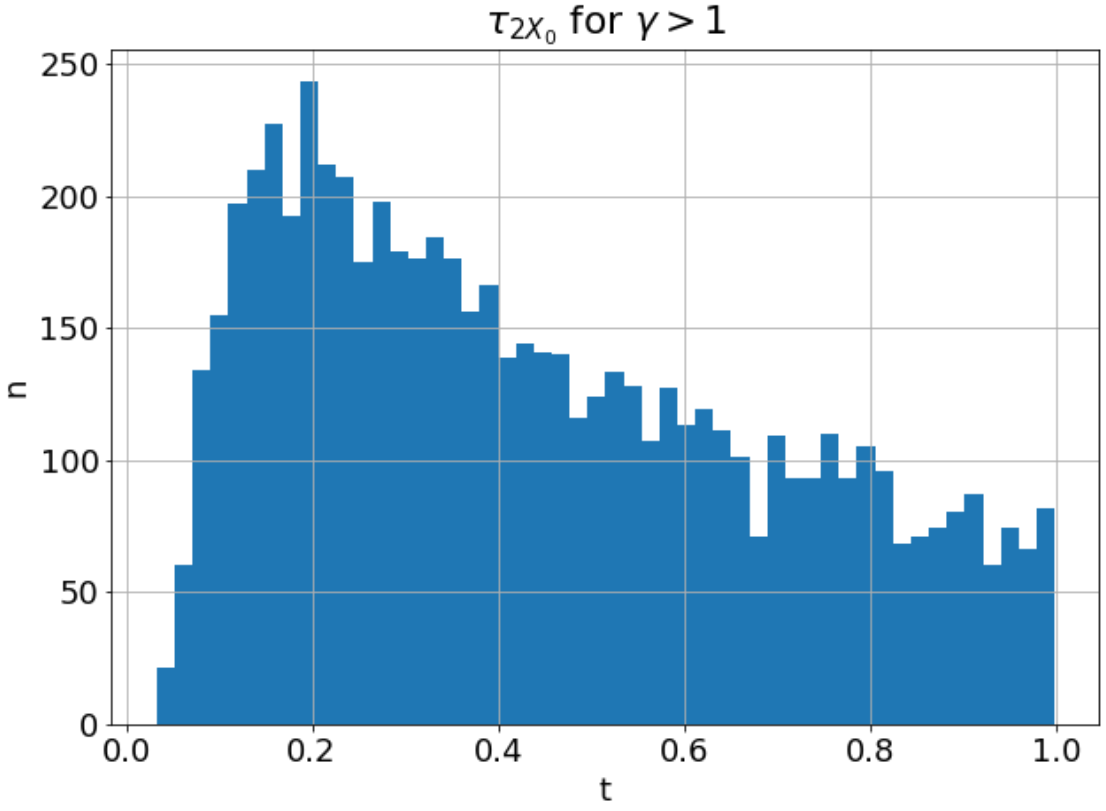
1. Additional Figure: τ_0 for $\gamma < 1$



2. Additional Figure: τ_{2X_0} for $\gamma < 1$

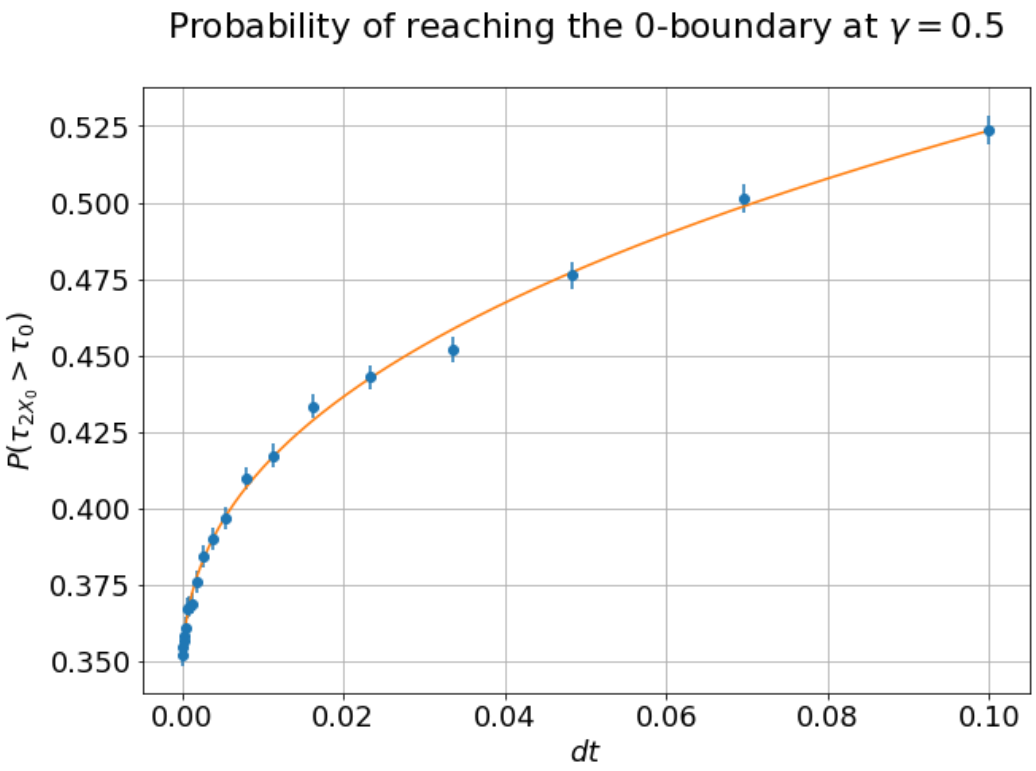


3. Additional Figure: τ_0 for $\gamma > 1$

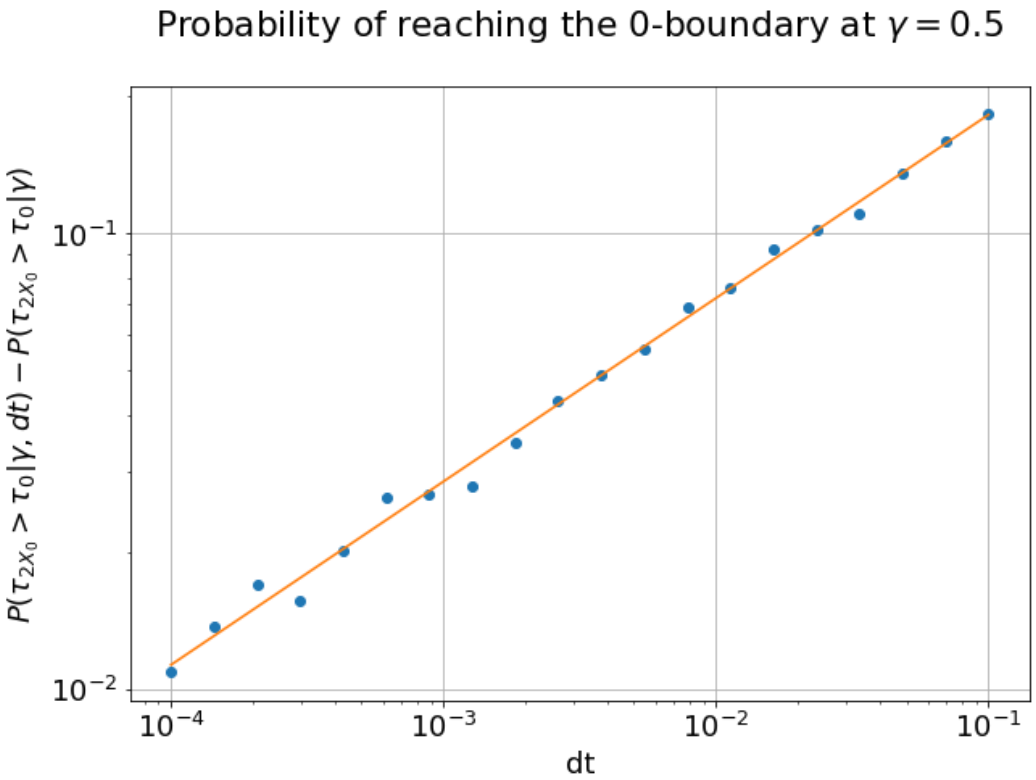


4. Additional Figure: τ_{2X_0} for $\gamma > 1$

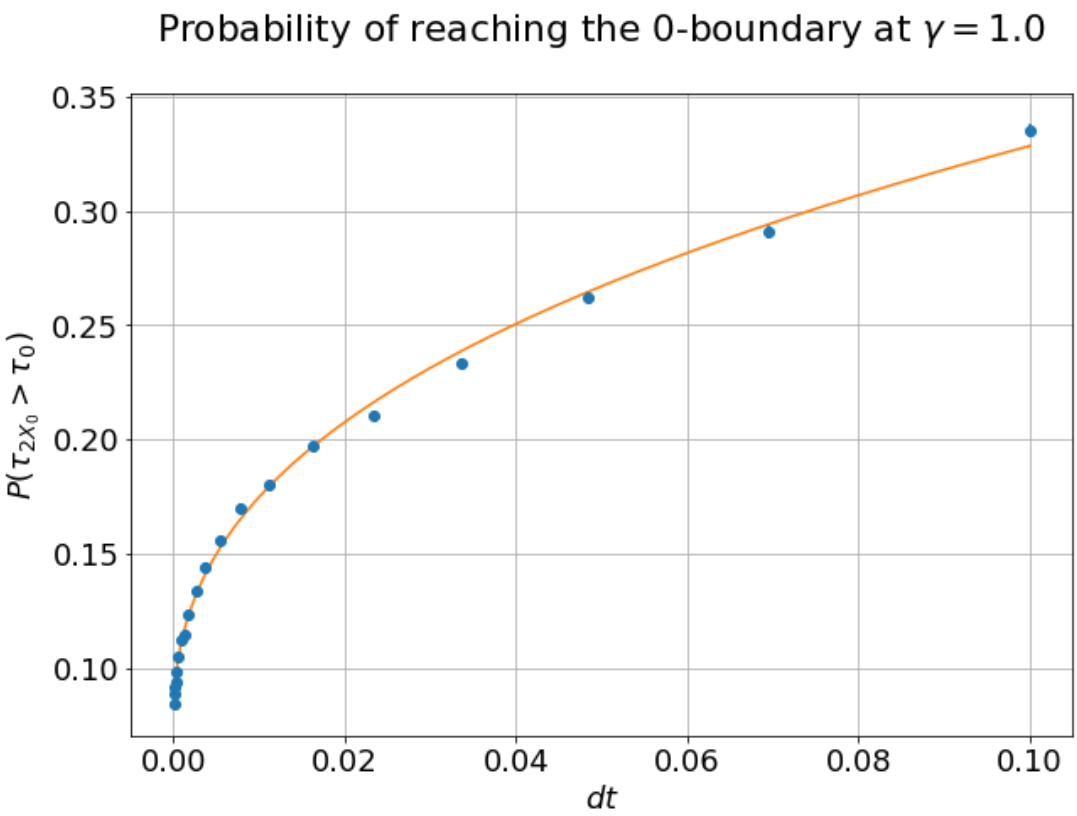
Convergence at the Reflective boundary



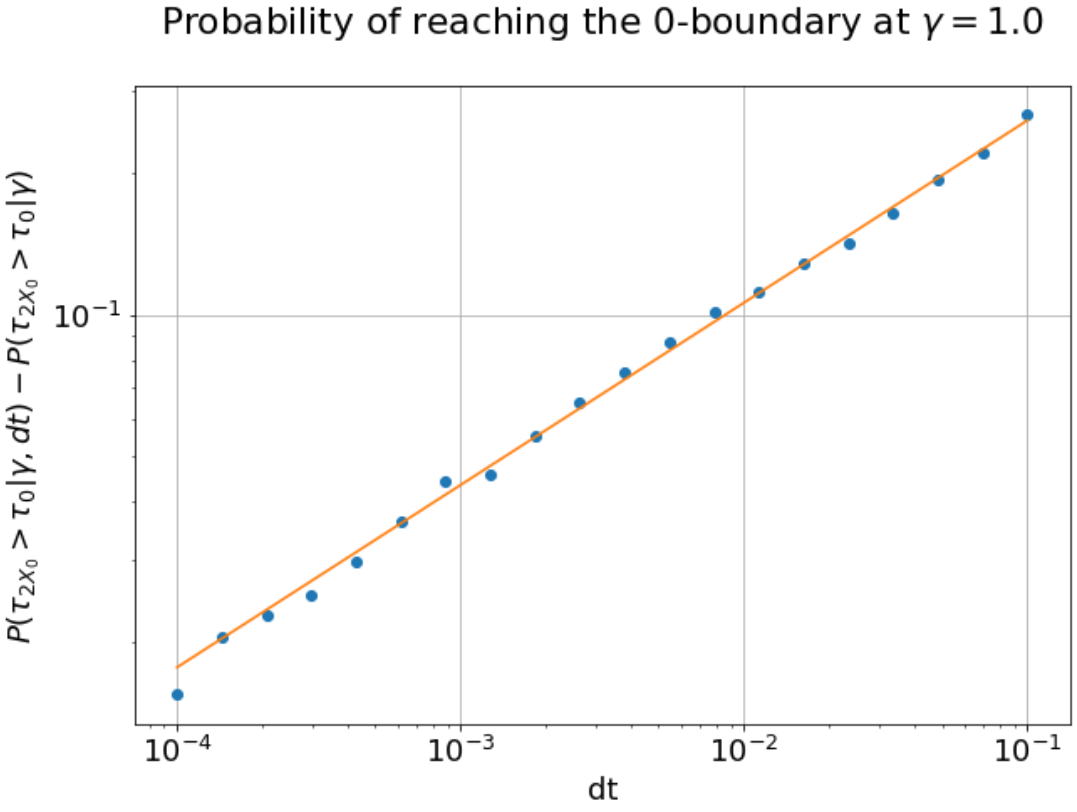
5. Additional Figure: Convergence of $P(\tau_{2X_0} > \tau_0)$



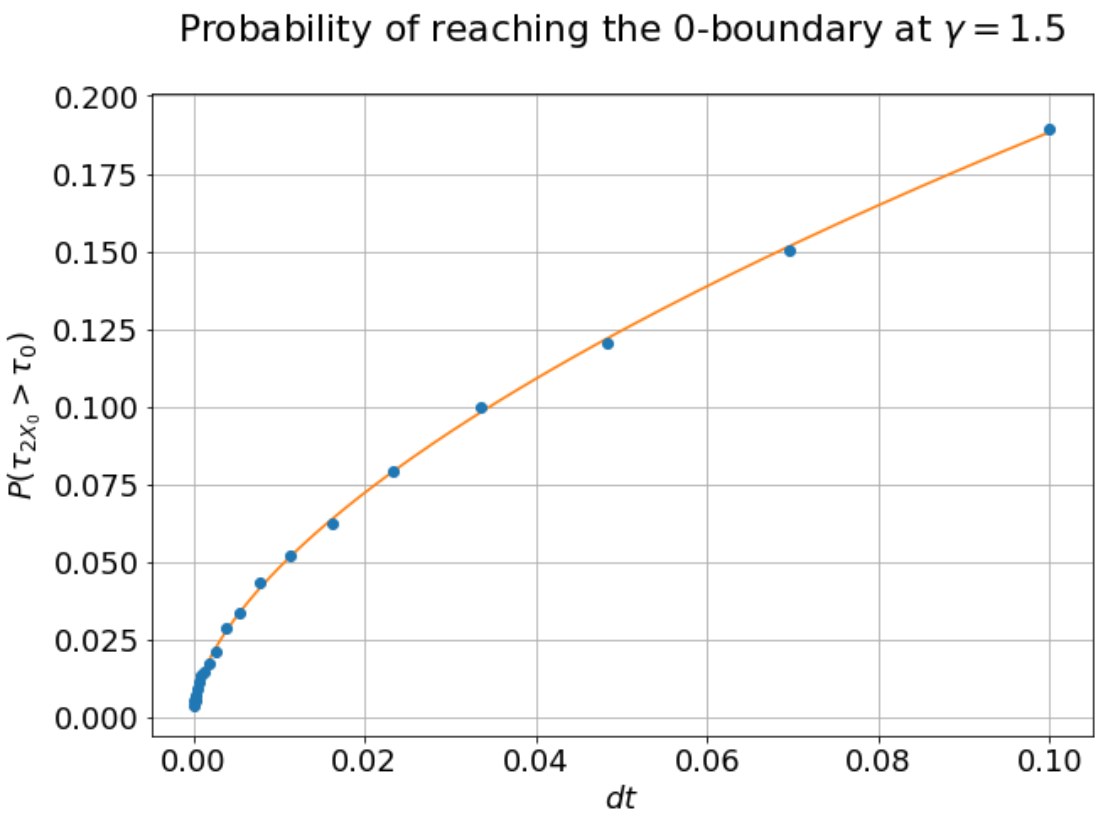
6. Additional Figure: Convergence of $P(\tau_{2X_0} > \tau_0)$



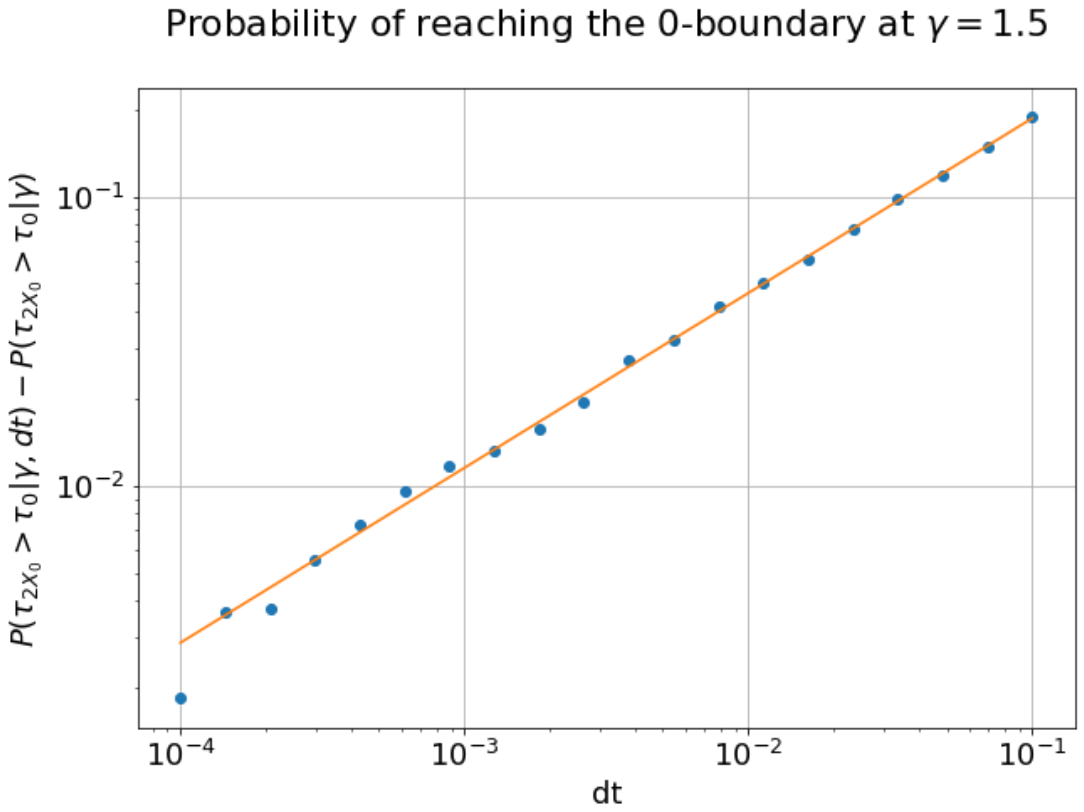
7. Additional Figure: Convergence of $P(\tau_{2X_0} > \tau_0)$



8. Additional Figure: Convergence of $P(\tau_{2X_0} > \tau_0)$

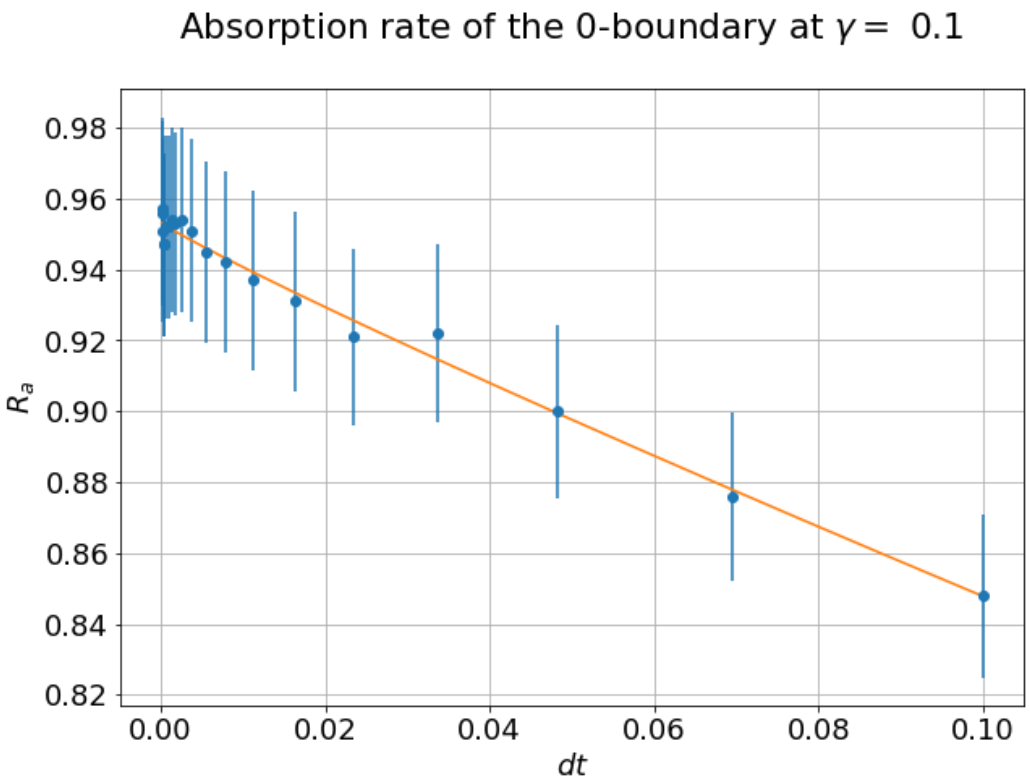


9. Additional Figure: Convergence of $P(\tau_{2X_0} > \tau_0)$

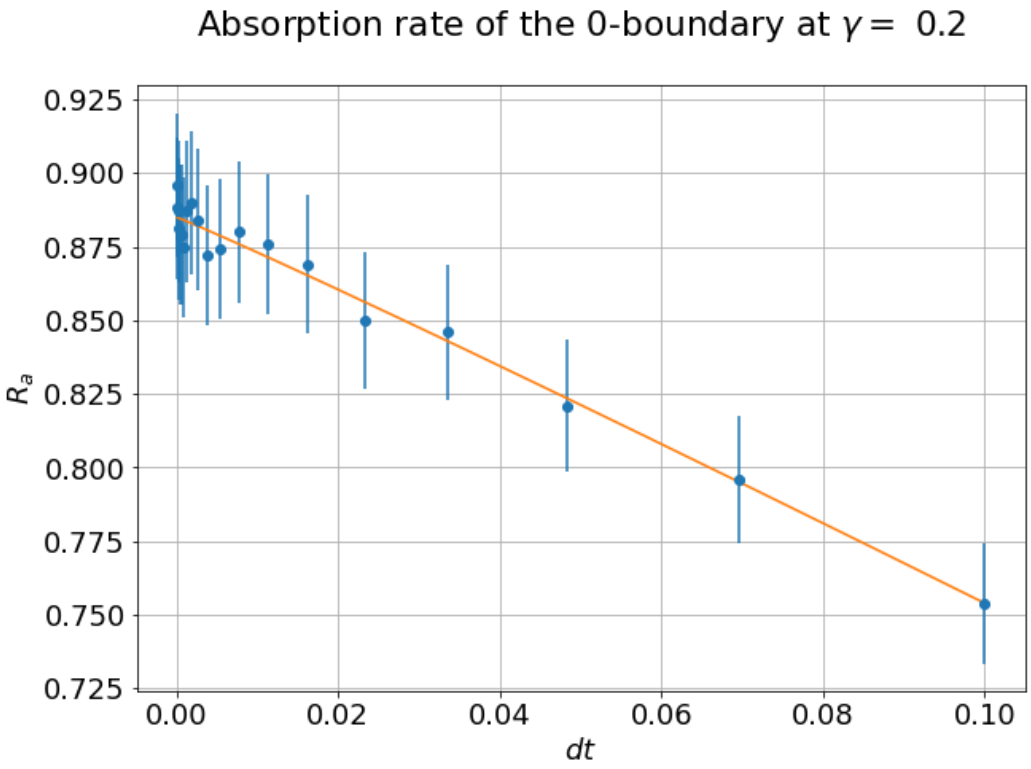


10. Additional Figure: Convergence of $P(\tau_{2X_0} > \tau_0)$

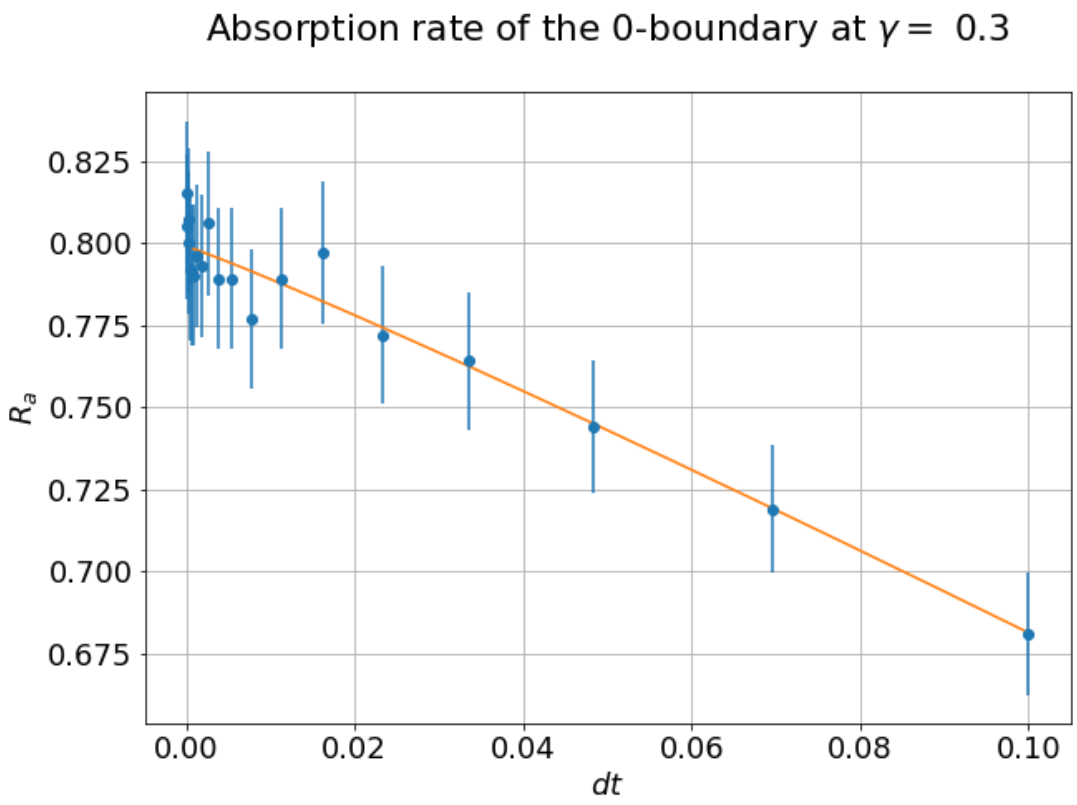
Convergence at the Absorbing boundary



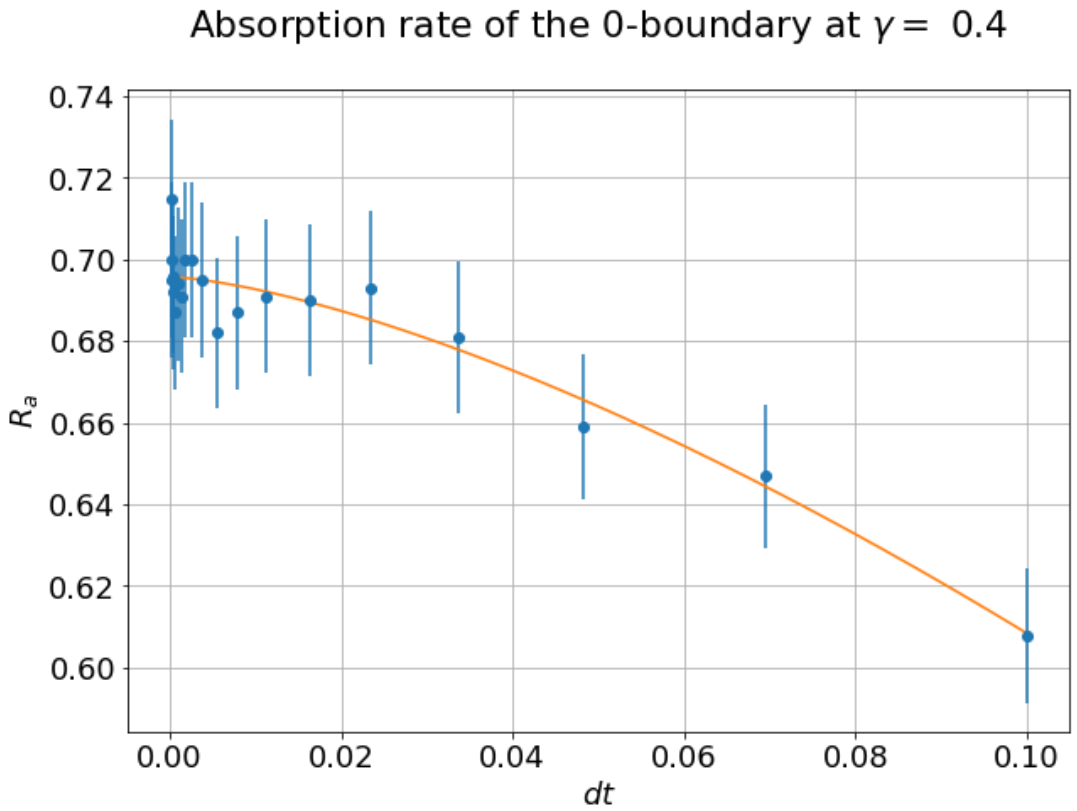
11. Additional Figure: Convergence of R_a



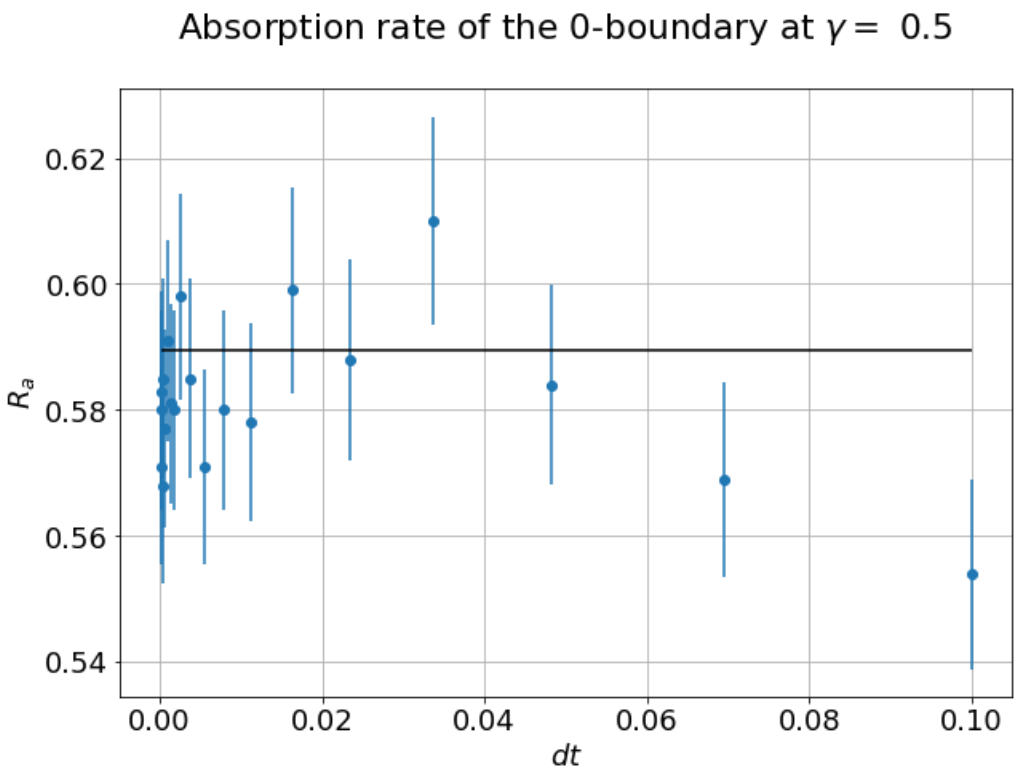
12. Additional Figure: Convergence of R_a



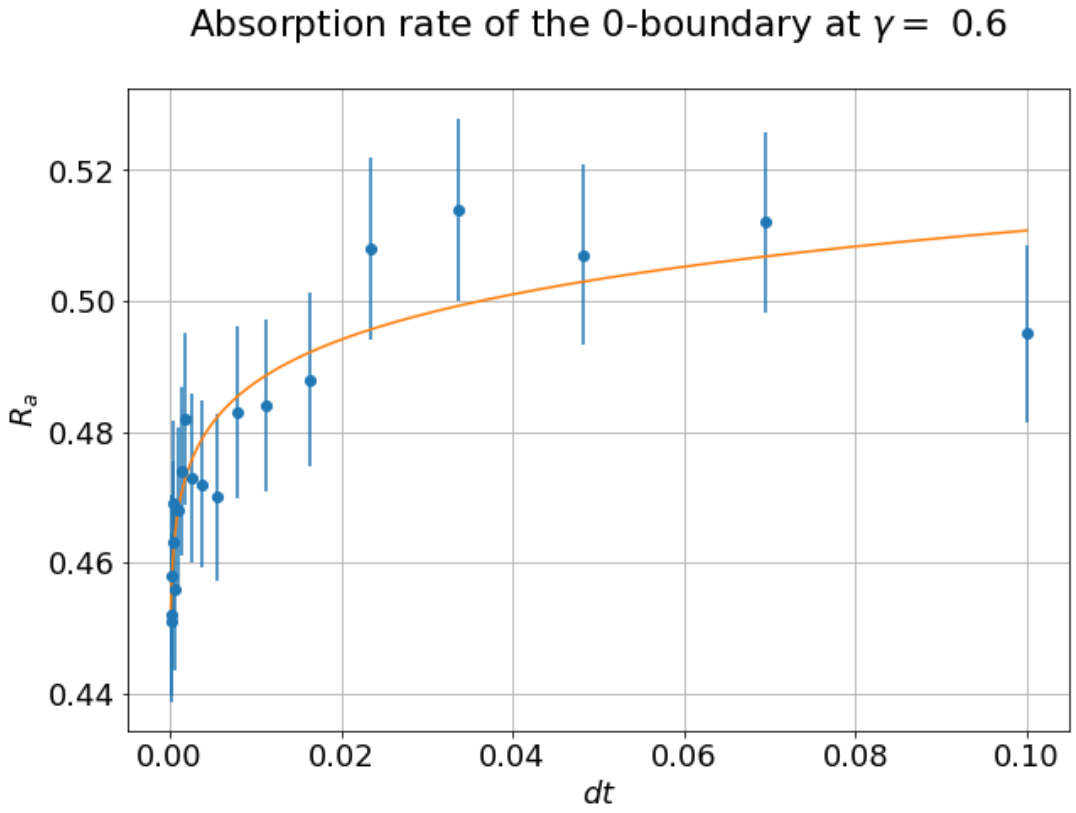
13. Additional Figure: Convergence of R_a



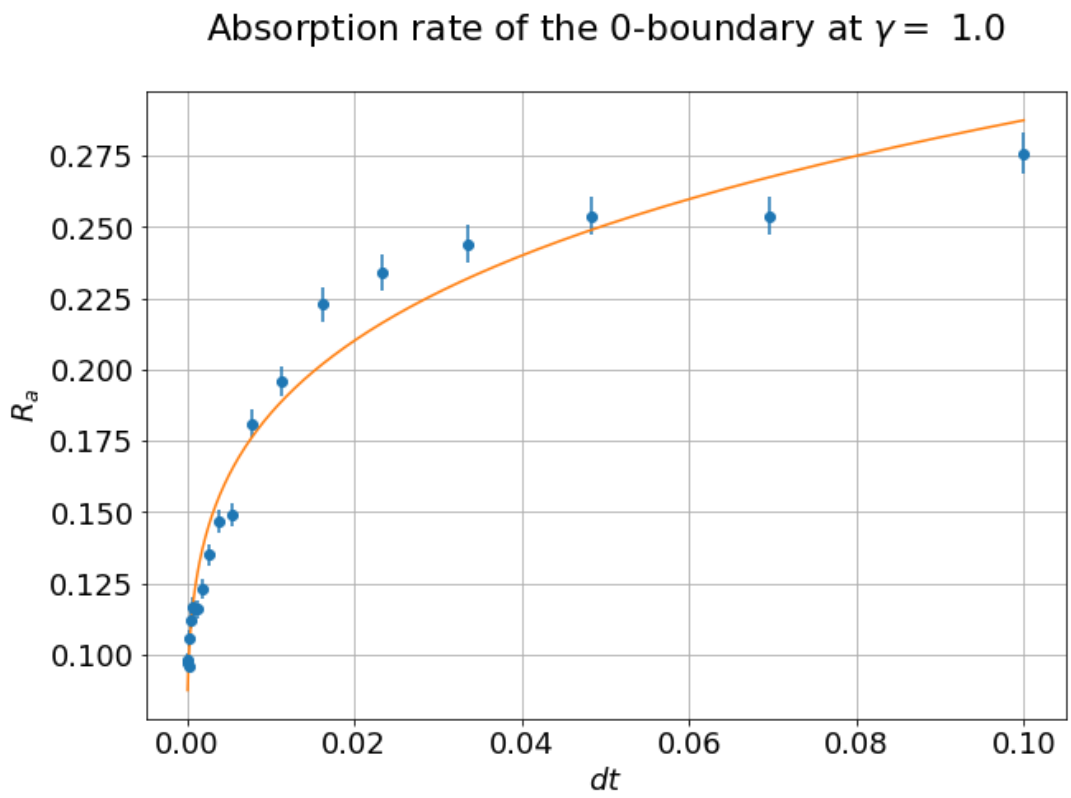
14. Additional Figure: Convergence of R_a



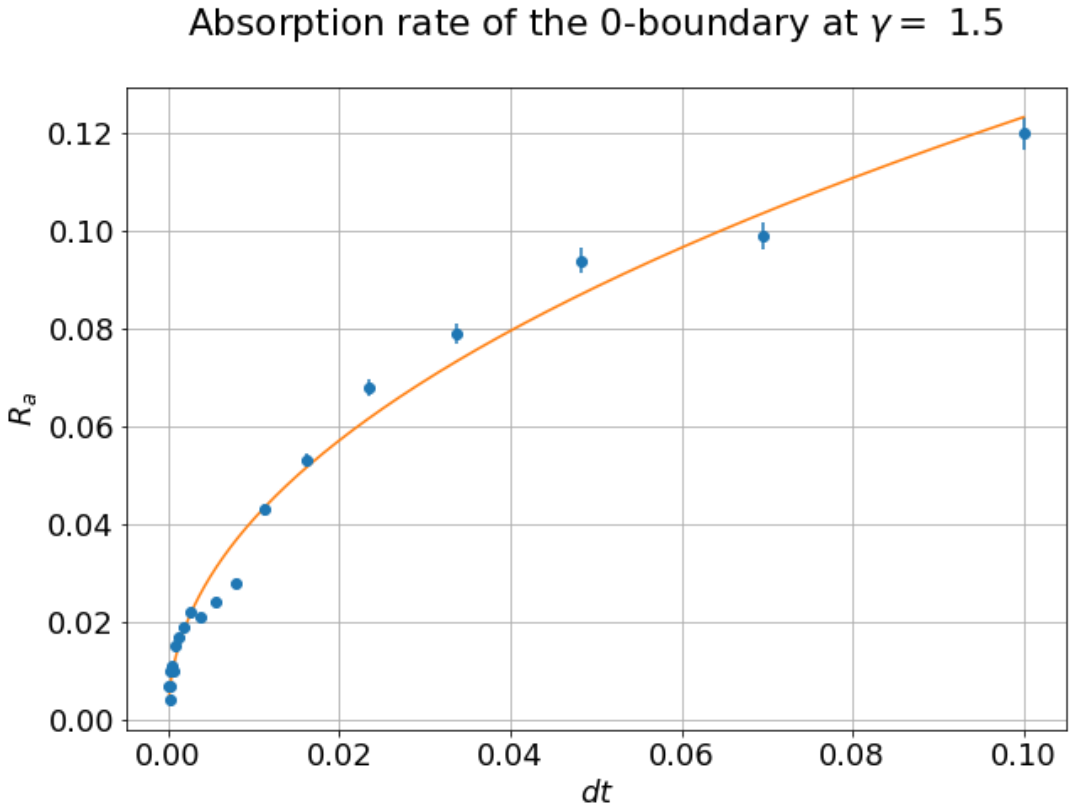
15. Additional Figure: Convergence of R_a



16. Additional Figure: Convergence of R_a

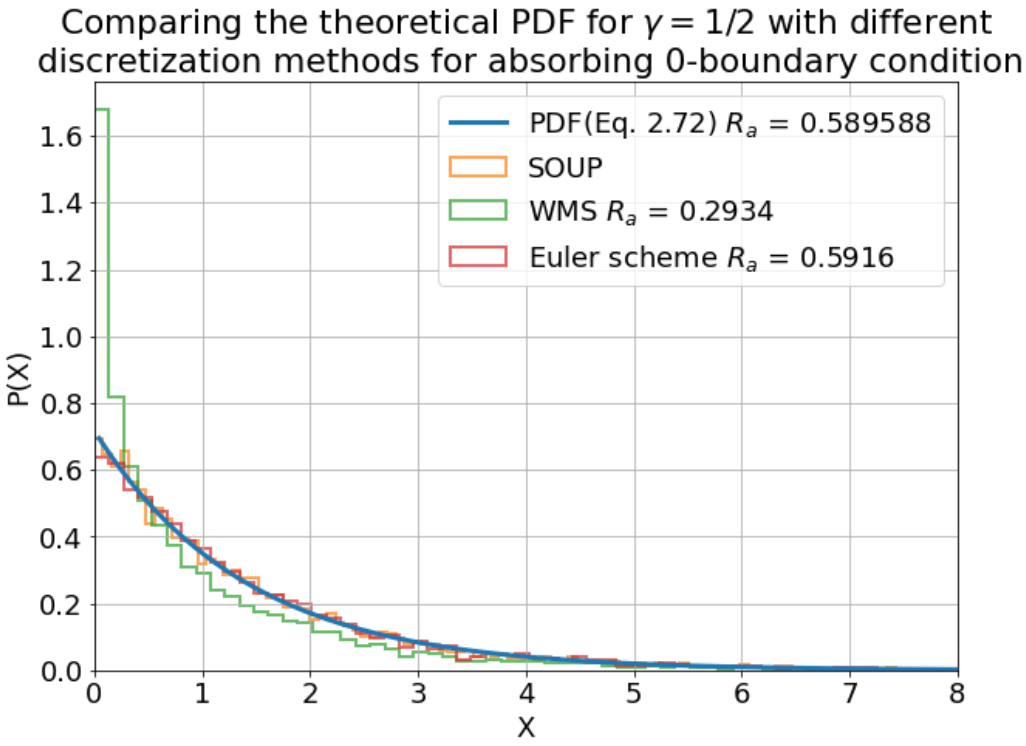


17. Additional Figure: Convergence of R_a

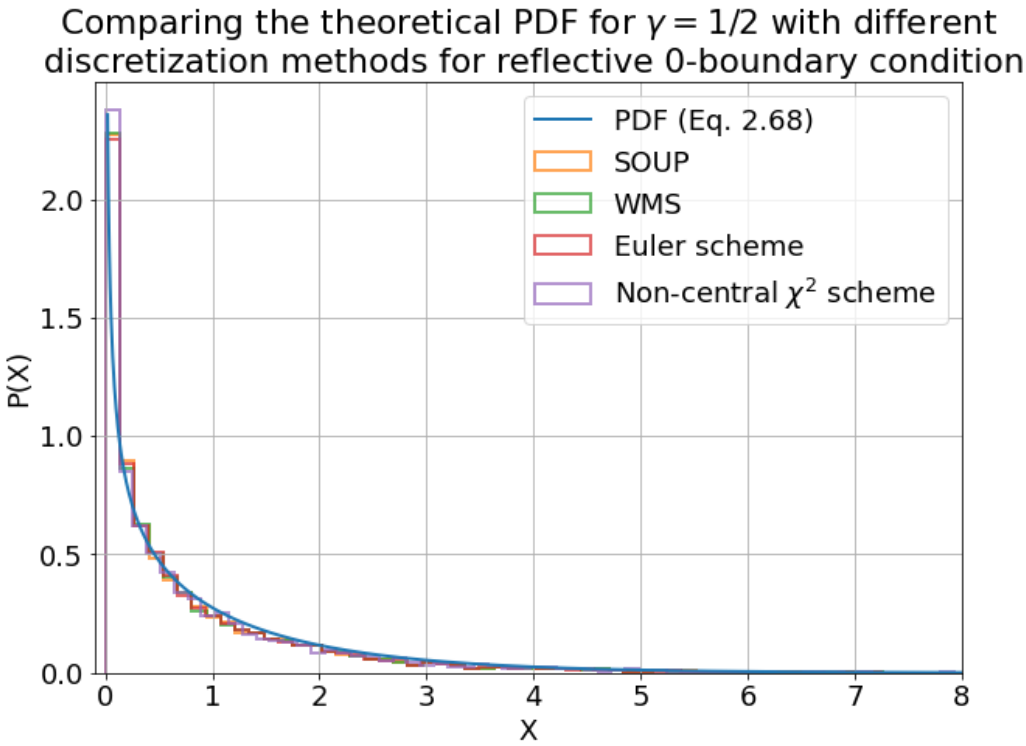


18. Additional Figure: Convergence of R_a

Comparing the PDF at $\gamma = ½$

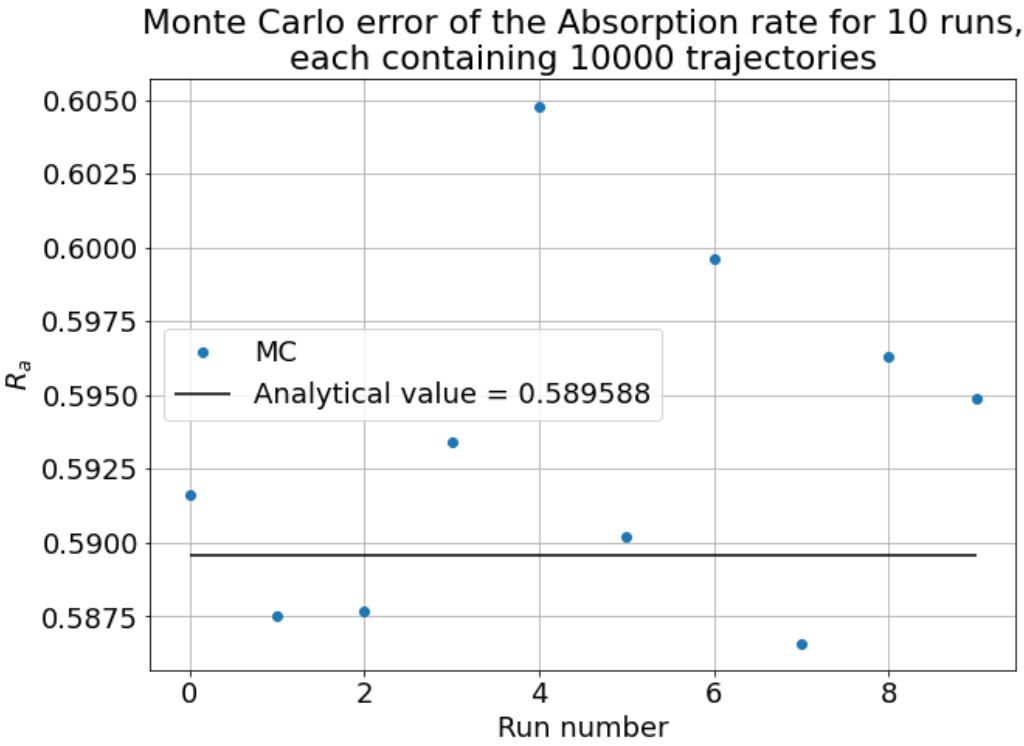


19. Additional Figure: Comparing the analytical PDF for Absorbing 0-boundary.



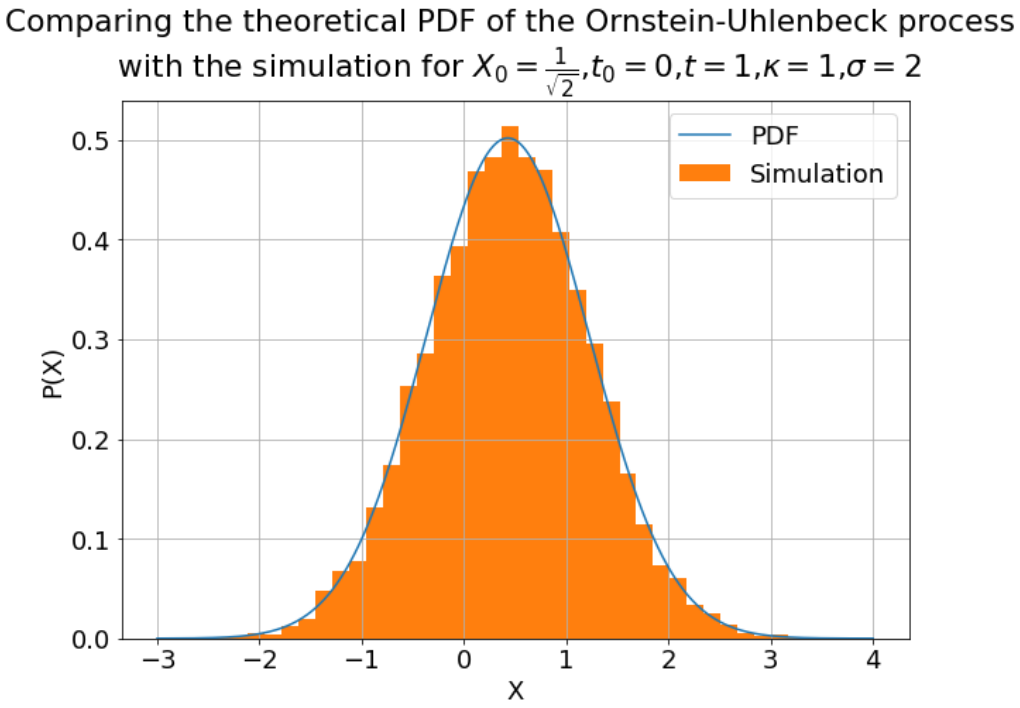
20. Additional Figure: Comparing the analytical PDF for Reflecting 0-boundary.

Monte Carlo error



21. Additional Figure: The varying value of R_a when the seed of the simulation is changed.

Comparing the PDF of OU Process with the simulation



22. Additional Figure: This was an important checkpoint in the implementation of the squared Ornstein-Uhlenbeck process, that the Analytical PDF matched with the simulation, nicely.

References

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy*, 637-654.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 385-407.

Deveroy, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer.

Fáth, G. (2011, August 8). Boundary classification of diffusion processes. Budapest.

Feller, W. (1951). Two Singular Diffusion Problems. *Annals of Mathematics*, 173-182.

Gardiner, C. W. (2002). *Handbook of Stochastic Methods.* Springer.

Haugh, M. (2017). *Monte-Carlo Simulation: Simulating Stochastic Differential Equations.* Retrieved from Columbia University: http://www.columbia.edu/~mh2078/MonteCarlo/MCS_SDEs_MasterSlides.pdf

Hull, J. C., & Basu, S. (2018). A Nonrigorous Derivation of Itô's Lemma. In J. C. Hull, & S. Basu, *Options, Futures, and Other Derivatives* (pp. 347-348). Noida: Pearson.

Hull, J. C., & Basu, S. (2018). Equilibrium Models of the Short Rate. In J. C. Hull, & S. Basu, *Options, Futures, and Other Drivatives* (pp. 742-753). Noida: Pearson.

Hull, J. C., & Basu, S. (2018). Interest Rates. In J. C. Hull, & S. Basu, *Options, Futures, and Other Derivatives* (pp. 94-127). Noida: Pearson.

Itô, K. (1951). *On Stochastic Differential Equations.* New York: American Mathematical Society.

Jeanblanc, M., Yor, M., & Chesney, M. (2009). *Mathematical Methods for Financial Markets.* Springer.

Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes.* New York: Academic Press, Inc.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* Cambridge: Palgrave Macmillan.

Landau, R. H., Páez, M. J., & Bordeianu, C. C. (2012). *Computational Physics: Problem Solving with Computers.* Wiley-VCH.

Ldecola. (2018, May 13). *Yield Curve.* Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Yield_curve#/media/File:Yield_curve_20180513.png

Levy, M. J. (2017, July 21). *Tweet*. Retrieved from Twitter: https://twitter.com/mahtin/status/888251632550424577

Liebow-Feeser, J. (2017, November 6.). *LavaRand in Production: The Nitty-Gritty Technical Details*. Retrieved from The Cloudflare Blog: https://blog.cloudflare.com/lavarand-in-production-the-nitty-gritty-technical-details/

Malham, S. J., & Wiese, A. (2012). Chi-square simulation of the CIR process and the Heston. *arxiv.org*.

Mamon, R. S. (2004). Three Ways to Solve for Bond Prices in the Vasicek Model. *Journal of Applied Mathematics and Decision Sciences*, 1-14.

Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 3-30.

Merton, R. C. (1973). Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 141-183.

Mil'shtein, G. N. (1974). Approximate integration of stochastic differential equations. *Teoriya Veroyatnostei i ee Primeneniya*, 583-588.

Palla, G. (2019). *Valószínűségszámítás és statisztika a fizikában előadás*. Retrieved from https://pallag.web.elte.hu/valszam/

Richard J. Rendleman, J., & Bartter, B. J. (1980). The Pricing of Options on Debt Securities. *Journal of Financial and Quantitative Analysis*, 11-24.

Richardson, L. F. (1911). The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society A.*, 307-357.

Risken, H. (1984). *The Fokker-Planck Equation: Methods of Solution and Applications*. New York: Springer.

Scott, T. (2017, November 6). *The Lava Lamps That Help Keep The Internet Secure*. Retrieved from YouTube: https://www.youtube.com/watch?v=1cUUfMeOijg

Sipos, R. (2018, May 2). Python for Finance: Monte Carlo simulations. ELTE, Budapest, Hungary.

Steiner, T. (2007, January). *Noncentral chi-squared distribution*. Retrieved from Wikipedia: https://commons.wikimedia.org/wiki/File:Chi-Squared-(nonCentral)-pdf.png

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the Theory of the Brownian motion. *Phisical Review*, 823-841.

Vašíček, O. A. (1977). An Equilibrium Charachterization of the Term Structure. *Journal of Financial Economics*, 177-188.

Wiener, N. (1976). *Collected Works, Volume 1*. Cambridge: The MIT Press.