STOCHASTIC VOLATILITY MODELS AND MEMORY EFFECTS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2011

Hunida Malaikah
School of Mathematics

Contents

Word count 31820

List of Tables

List of Figures

The University of Manchester

Hunida Malaikah

Doctor of Philosophy

Stochastic Volatility Models And Memory Effects

June 2, 2011

According to financial empirical data, volatility has short memory (Markovian) effect during one day and long memory (non-Markovian) effect over a long time. We aim in this work to derive Markovian and non-Markovian models to model the volatility. To find these models, we introduce the theory of conditional arrival probability. We apply this theory when time is continuous and the state space is discrete, and derive the equations in Laplace domain then convert them to time domain. We obtain the master equations of the transition probability corresponding to different waiting time probability density function (PDF). In the case of non-Markovian we get fractional-time model. As examples, we find the master equations of the counting process corresponding to different waiting time PDF. Moreover, the formulas for first and second moments of the process. Then, we generalize the theory of conditional arrival probability to the case where time and state space are continuous. We derive the equations in Laplace-Fourier domain, then convert them to time-space domain to obtain the master equations of the processes in different forms corresponding to different waiting time PDF and different transition PDF. In the case of non-Markovian we get fractional time-space model. We propose to use these models for volatility. In addition, the non-Markovian master equation of transition probability includes the memory effect which appears as the kernel function. In order to find the kernel function we need to find the waiting time density function. Therefore, we introduce the age model to show how can we find the waiting time density from the hazard function. Then, we find the master equation of the probability regarding the age. Finally, we give examples to hazard function to get Markovian and non-Markovian models for volatility.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Dedication

To My Mother...

Acknowledgements

In the name of God I started my work. All my prayers to give me the success and the acceptance.

Then, from my heart I would like to thank my supervisor SERGEI FEDOTOV who is a really kind, helpful, patient professor. Honestly, I learnt a lot from him and I really appreciate his standing with me through all the circumstances I faced during my work. I extend my thanks to the University of Manchester, it is a very calm, warm, stimulating environment in which to work.

All of my love and my respect to my great family for their unlimited support. At the head, my mother who hasn't stopped praying for me and awaiting my success. My brothers and sisters who usually motivate and push me forward. My uncles and aunts for their sympathy and thanks to every one who has wished me success.

Special thanks to my daughter HALA for her understanding and her tolerance of my work and being abroad.

A big thanks to my friends Hanan and Alex who have brought happiness to my life in Manchester.

Last but not least, I would like to thank my sponsor the Saudi bureau and my university King Abdul Aziz university for giving me the chance to take my higher education abroad.

Chapter 1

Introduction

Derivatives securities are contracts whose value is derived from the underlying asset upon which they have been written. They are financial instruments which permit both the efficient transfer of risk that emanates from the holding of a financial asset. They also provide efficient ways to speculate in the global market [48]. The ability to price these financial instruments relies on the law of one price. However, in efficient markets, if a contract pays a certain amount of cash under a certain state of the world and another contract exists, then both contracts must trade at the same price, otherwise arbitrage opportunities exist.

Futures and options, also called contingent claims, are the basic contracts of all derivative securities, hence the importance of pricing options. A futures contract is an agreement between two parties to buy or sell (by obligation) an asset at a certain time in the future (called the expiration date or maturity) for a certain price called the exercise price (or strike price). An option gives the holder the right to buy (sell) the asset by maturity for the exercise price. An American option can be exercised at any time up to maturity, while a European option can be exercised only at maturity. Of course the holder of the option will exercise his right only if he is going to gain from the transaction. With no possibility of losses, a premium is demanded by the seller of the option. The importance of option has prompted researchers from finance, economics, mathematics, and physics to analyze option pricing from both theoretical and practical perspectives.

The long history of option pricing began in 1900 when Bachelier [42] assumed that the successive changes in the price asset are independent Gaussian random variables with zero mean and a variance proportional to time. However, Mandelbrot (1963) [74] found the market price has non-Gaussian nature of the distribution's tails and introduced the concept of Lévy flight and stable distribution in finance. Later on, [75],[43], they found that the past record should make a different prediction of the future price, so the price changes need not be independent as random walk. The sophistication of pricing options increased which increased the need for a good mathematical model. In the early 1970s, Fischer Black, Myron Scholes and Robert Merton made a major breakthrough in the pricing of stock options. This involved the development of what has become known as the Black-Scholes (BS) model [14]; by utilizing no arbitrage arguments, they derived a closed-form solution for the price of European options. In the same year Merton [88] introduced additional assumptions to examine and extend the seminal BS theory of option pricing and found an explicit formula, so that the model had a wider application. In considering the importance of the BS model, Merton (1976)[89] pointed out the critical assumption in their model derivation that the price dynamics of the stock have a continuous sample path with probability one. As a solution for this assumption he introduced for the stock price a jump stochastic process defined in continuous time. His model assumes that the stock price is described by Brownian motion with drift as in BS together with a compounded Poisson process to model the jumps of the stock price. Cox and Ross [32] implemented different types of jump processes to value the stock. Aurell [4] used discrete time process to price the option. Furthermore, Herzel[56] included the possibility of jumps for the path of stock price and for those of its volatility.

Similarly, the BS model has been widely used by traders and investors all over the world to price and hedge the option. Despite the popularity of the BS model it was limited by many assumptions, such as market completeness: a complete market is one where for every derivative contract there exists a self-financing portfolio of underlying assets which perfectly replicates it, but in the real world markets are not complete [44]. The reason for incompleteness may be due to factors such as transaction costs, stochastic volatility, jumps, arbitrage opportunities, etc. Moreover, a number of empirical findings have been compiled that cast doubt on another key assumption of the BS model, namely that returns are normally distributed with constant mean and volatility. It has been discovered that empirical returns are far from normally distributed as they exhibit fat tails and/or non-zero skewness; also the mean and volatility are not constant [48]. Non-zero skewness implies that returns are not distributed symmetrically around the mean, and it also explains the correlation of the volatility with the stock price process. In particular the assumption of constant volatility has attracted much debate and research. One way to modify this assumption was made by Rubinstein [106], who derived formulae in which volatility can be a function of stock; however these formula do not give more accurate results than the standard BS model. Put another way, researchers proposed option pricing models that explicitly take into account the stochastic nature of stock return volatility, but it is uncorrelated with stock market volatility. Early studies include those by Scott [110], Wiggins [119] and Hull and White [63]. Their attempts to solve the problem of option pricing under stochastic volatility concentrated mainly on deriving the partial differential equation that a derivative security must satisfy in the case of random changing volatility. Subsequently, numerical methods were used to calculate the option price. A notable exception is the approach of Hull and White, who provided an explicit solution to the problem of option pricing under stochastic volatility that is uncorrelated with the stock price, and a numerical solution when the volatility correlated with the stock price. Amin [1] considered that stock returns follow a lognormal jump diffusion process when the stock return volatility is stochastic and follows diffusion process. Moreover, the volatility correlated with the volatility of market return. However, perfect replication is not possible as volatility is not a traded asset. After that, researchers managed to provide a closed-form solution for the price of European options when volatility is stochastic and correlated with the stock price, by assuming that volatility follows a mean reverting process: for example Heston [57], Ball and Roma [7], and Bakshi [6]. Consistently, all stochastic volatility models outperform the BS model by providing a better fit for real option market values; in particular they achieved the volatility smile phenomena. In another view, the stochastic volatility should be modeled by a discrete stochastic process instead of a continuous one because of the market close after each trading day. As a case of implementing discrete stochastic volatility process in option pricing model, Ritchey [105] modeled a stochastic volatility as a binomial probabilities tree that spans the component normal distribution. Later, Gue [55] replaced Ritchey's binomial tree with a finite Markov chain with k discrete volatility states, and Ritchey's binomial probabilities were replaced by an n-step Markov transition probability matrix.

Despite the success of the stochastic volatility models, there exists strong empirical evidence suggesting that the volatility has long memory effect i.e., its correlation function decays very slowly. Accordingly, the stochastic processes used to describe volatility, such as mean reverting process or standard Brownian motion process, are not appropriate to describe correctly the long memory property of volatility due to the diffusion process, which has a Markovian nature. To remedy this problem it is more convenient to incorporate a long memory stochastic volatility process into option pricing model.

Engle [38] introduced another class of stochastic process, including the memory phenomena: autoregressive conditional heteroscedastic, or ARCH(p). This is a mean zero uncorrelated process with nonconstant variances conditional on the past, but constant unconditional variances. In other words, this process allows for the variance to change over time as a function of past errors. When this process modeled daily returns there was no autocorrelation, whereas their squares had a noticeable autocorrelation. However this autocorrelation decays exponentially to zero as time goes to infinity and this property does not exhibit the long memory. Many extensions to the ARCH model have been proposed, including the generalized ARCH process (GARCH(p,q)) process introduced by Bollerslev [15]. However, the exponential decay of autocorrelation of the GARCH process is still too fast to describe correctly the persistent dependence between the series observation as the time lag increases [45]. It turns out that the model with hyperbolic decay has slowly decaying autocorrelation, providing better fitting to financial time series. This led Engle and Bollerslev [39] to allow the variance to be a non-linear function of the squared innovations by proposing the Integrated GARCH process (IGARCH). Also, Nelson [97] tried to extend the GARCH process infinitely far into the past and proposed the Exponential GARACH process (EGARCH). Later on, Engle [40] allows different news to have a different impact on volatility, and also allows big news to have much more impact on volatility than the standard GARCH process. Moreover, he found the partially nonparametric ARCH model (PNP) successfully reveals the shape of the news impact curve [41].

In addition, Duan [35] developed GARCH asset return process. Moreover, a new class of Fractional Integrated GARCH (FIGARCH) was proposed by Bollerslev and Mikkelsen [18] and Billie [12], to modify a long-term dependence model in conditional variance. In the last process they found that the conditional variance, which may be considered as the volatility, dies out at a slow hyperbolic rate of decay and this is evidence of the long memory effect.

In the econophysics world the researchers conclude that not only the prices and returns can be considered as random variables, but also the waiting time between two transactions varies randomly. This led to incorporating the continuous time random walk (CTRW) model in finance: see [107],[109],[108],[83]. CTRW was first introduced by Montroll and Weiss (1965) [92]; it is used in physics to model anomalous diffusion, by incorporating a random time between particle jumps. In finance, the particle jumps are zero mean log-return and waiting time is measured between transactions. The advantage of using CTRW instead of Brownian motion to model the return is that the CTRW exhibits the skewness effect and the heavy tail distribution. Heavy-tailed appears due to power-law distribution for price jump and/or waiting time. This distribution led to fractional differential equation. By finding the moments of the process we can find the volatility of stock return.

This work is aiming to use the CTRW itself to model the volatility as the particle jumps. We introduce two cases of CTRW, when it has discrete and continuous state space. In each case we use different waiting time distributions in order to derive Markovian and non-Markovian models. We introduce the theory of conditional arrival probability to get the master equations of conditional transition probability and probability density in both cases of discrete and continuous random walk with continuous time. We derive our equations in Laplace domain in case of discrete random walk, and in Fourier-Laplace domain in the case of continuous random walk, then convert them to space-time domain. Furthermore, we find the moments of the process corresponding to different waiting time and jump distributions.

Also we propose the theory of age model (hazard function) again to find the waiting time density and derive the master equation for the probability when the process has discrete state space. The specific selection of hazard function leads us to Markovian or non-Markovian models.

1.1 Outline of Thesis

In this introductory chapter we mention the importance of option pricing theory, and the modification of stock return and volatility process to incorporate in option pricing models. The next chapter discusses the mathematical background that includes all the tools we are going through in our work. First, we introduce the definition and properties of stochastic process. Second, we explain the concept of Markov process and how can we classify it regarding time space and state space, explaining in each case how can we find the transition probability and illustrated by an example. We then explain the importance of Markov process as a short memory stochastic process which can be modified to get a long memory process. Third, we show how to find the memory of the process, illustrating the different methods which indicate the existence of long memory effect. We go on to review some famous models and evaluate their memory. Then we discuss a famous physical process which is nowadays used widely in finance: continuous time random walk (CTRW). This process may have short memory or long memory depending on the waiting time distribution. Furthermore, we explain the derivation of the master equation in Fourier-Laplace domain in order to find the master equation of the probability density of the process under different assumption of waiting time distribution. The assumption of long memory CTRW process leads to fractional derivative, which is given in detail in the last section of chapter two.

In chapter 3 we review the history of volatility models, starting from constant model then moving on to the deterministic model. We then present a short review of the most famous stochastic volatility models, predictable models and long memory models, in addition to the advantages and disadvantages of these models. We also show how we can find the volatility of the stock's return when it follows CTRW.

In chapter 4 we generalize CTRW, we derive the master equation corresponding to different waiting time and jump distributions. We implement our theory of conditional arrival probability when the state space is discrete and continuous. To confirm our result, we apply our result to the counting process to get the distribution of the process when it is Markovian; this gives us Poisson process, or non-Markovian.

Chapter 5 presents the theory of age model or hazard function. We illustrate how to choose this function in order to get Markovian and non-Markovian models.

The general conclusion of our work and recommendation for possible future work is given in chapter 6.

Chapter 2

Mathematical Background

2.1 Introduction

Our task in this chapter is to demonstrate all the mathematical tools that we are going to use in the dissertation. We start by giving background about the stochastic process and its statistical properties. This is important in order to provide some information about the process distribution and/or the memory of the process. Also we show the Markov chain with different types regarding time space and state space, in addition to the transition probability by using Chapman-Kolmogorov equation and examples for each type. The diffusion processes which are the solutions of stochastic differential equations that are used widely in stochastic volatility models and are considered as non-memory processes. Furthermore, we show how to find the memory of the process from its statistical properties. Then we classify the process regarding its memory to short memory process (Markovian) and long memory process (non-Markovian), giving the difference between them and examples of each. Additionally, we introduce Continuous Time Random Walk process (CTRW) which may be considered a Markovian or non-Markovian process depending on its waiting time distribution. In addition to the derivation of its master equation corresponding to different waiting times PDF, the moments and the distribution of the process correspond to special cases. The case where CTRW process is non-Markovian leads us to fractional derivative, which is illustrated in the final section of this chapter.

2.2 Stochastic Process

Stochastic process is a description of random phenomena changing with time [103]. Such phenomena occur in physics, biology, economics, engineering, sociology, and several other fields. In other words, a stochastic process $\{X_t, t \in R_+\}$ may be thought as a family of random variables where the parameter t represents the time. The time may be discrete when we use the parameter $n, n \in Z_+ = \{1, 2, 3, ...\}$, or continuous when we use $t, t \in [0, \infty)$. For example; throwing a true die letting X_n be the outcome of the n-th throw, then $\{X_n, n \geq 1\}$ is a stochastic process. The number of telephone calls received at a switchboard, if X_t is the random variables which represent the number of incoming calls in an interval $(0, t)$, then$\{X_t, t \geq 0\}$ is a stochastic process. Other examples would be: the fluctuation of current in an electrical circuit in the presence of so-called thermal noise; the random changes in the level of received radio-signals in the presence of random weakening of radio-signals (fading) created by meteorological or other disturbances; the turbulent flow of a liquid or gas. Stochastic processes are distinguished by their state space S, the range of possible values of the random variable X_n or X_t. The state space is either discrete or continuous.

2.2.1 Stationary Stochastic Process

A stochastic process $X = \{X_t, t \geq 0\}$, taking values in R, [54] is called **strongly stationary** if the families

$$\{X_{t_1}, X_{t_2}, ..., X_{t_n}\} \quad \text{and} \quad \{X_{t_1+h}, X_{t_2+h}, ..., X_{t_n+h}\}$$

have the same joint distribution for all $t_1, t_2, ..., t_n$ and $h > 0$.

Also, the stochastic process is called **weakly stationary** if the mean and covariance, if they exist, do not depend on time or do not change over time or position, i.e.

$$E[X_{t_1}] = E[X_{t_2}],$$

and

$$cov[X_{t_1}, X_{t_2}] = cov[X_{t_1+h}, X_{t_2+h}], \quad \text{for all} \quad t_1, t_2, \text{and} \quad h > 0.$$

The mean or the expected value is given by

$$\mu = E[X_t] =< X_t >= \int x f(x,t) dx,$$

where $f(x,t)$ is the probability density function of x observes at time t. While the variance is

$$var[X_t] = \sigma_x^2 = E[(X_t - \mu)^2],$$

if $\mu = 0$ then the variance is equal to the second moment

$$var[X_t] = \sigma_x^2 = E[X_t^2] =< X_t^2 >= \int x^2 f(x,t) dx.$$

The covariance function (or autocovariance in the case of stationary process)[1]

$$cov[X_{t_1}, X_{t_2}] = E[(X_{t_1} - \mu_1)(X_{t_2} - \mu_2)].$$

Again if: $\mu_1 = \mu_2 = 0$ then

$$cov[X_{t_1}, X_{t_2}] = E[X_{t_1} X_{t_2}] = \int\int x_1 x_2 f(x_1, x_2, t_1, t_2) dx_1 dx_2,$$

where$f(x_1, x_2, t_1, t_2)$ is the joint probability density function that x_1 observes at time t_1 and x_2 observes at time t_2.

The correlation function (or autocorrelation in the case of stationary process) $R(t_1, t_2)$ is equal to the autocovariance if the process has zero mean and unit variance, otherwise it is

$$R(t_1, t_2) = \frac{cov[X_{t_1}, X_{t_2}]}{\sigma_{x_1}\sigma_{x_2}} = \frac{E[(X_{t_1} - \mu_1)(X_{t_2} - \mu_2)]}{\sigma_{x_1}\sigma_{x_2}}.$$

If the mean and the variance are time-independent, i.e. the process is stationary, then the autocorrelation depends only on the difference between t_1 and t_2. In other words, the correlation depends only on the time-distance between the pair of values h where $h = t_2 - t_1$ but not on their position in time, and h is called the lag. This further implies that the autocorrelation can be expressed as a function of the time-lag h, and that this would be an even function of the lag h. The more familiar form of the autocorrelation function is given by

$$R(h) = \frac{E[(X_t - \mu)(X_{t+h} - \mu)]}{\sigma^2}.$$

[1]The integration in the statistical properties definition are used in the case of continuous time process, in the discrete time process we change the integration to summation.

The autocorrelation function is sensitive to the expected value of the stochastic process if it is different from zero. Hence it is useful to consider the autocovariance as [77]

$$R(h) = cov(h) - \mu^2.$$

The typical shape of $cov(h)$ [or $R(h)$] for positively correlated stochastic variables is a decreasing function starting from $cov(0) = \sigma^2$ and ending at $cov(h) = 0$ for large value of h.

2.2.2 Self-Similar Process

A stochastic process $\{X_t, t \geq 0\}$ is called self-similar with index component $H > 0$, if for any $a > 0$, $\{X_{at}\}$ and $a^H\{X_t\}$ have the same distribution. The index H is called the Hurst's exponent or scaling exponent for the process.

2.3 Markov Processes

A Markov process is a stochastic process whose future behavior can be determined by the present, independently of the past, so that it is considered a non-memory process [98]. It is categorized depending on the time and the state spaces; if the state space is discrete, $S \subset Z_+$, then it is called a Markov chain with discrete time or continuous time. It is also called Markov chain with continuous state space if the time is discrete. Finally, if both time and state space are continuous, it is called a Markov process.

2.3.1 Markov Property

The Markov property asserts that the distribution of the following state, $X_{n+1} = j, n = 0, 1, 2..., j \in S$, depends only on the current state $X_n = i, i \in S$ not on the whole history, i.e.

$$Pr\{X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, ..., X_n = i\} = Pr\{X_{n+1} = j \mid X_n = i\}. \tag{2.1}$$

We call

$$Pr\{X_{n+1} = j \mid X_n = i\} = p_{ij},$$

the transition probability. The same for continuous time, given the value X_s, the value of $X_t, t > s$, does not depend on the value of $X_u, u < s$, then the process $\{X_t, t \in R_+\}$ is said to be Markov process. A definition of Markov process is given by
If, for $u < s < t$

$$p_{ij} = Pr\{X_t \leq y \mid X_u = z, X_s = x\} = Pr\{X_t \leq y \mid X_s = x\}, \quad z < x < y, \quad x, z, y \in R_+.$$

2.3.2 Transition Matrix

The transition probability matrix or matrix of transition of Markov chain P; is a stochastic or Markov matrix, it is a square matrix with non-negative elements and unit row sums. Its elements are the transition probabilities p_{ij} satisfy

$$p_{ij} \geq 0, \quad \sum_j p_{ij} = 1 \quad \text{for all} \quad i, j \in S.$$

The transition matrix is formed in the case where the time is discrete and the state space too whenever it is finite or denumerably infinite, and also in the case of continuous time when the waiting time between the transactions is exponentially distributed, which satisfies the Markovian property of time.

2.3.3 Markov Chain

The stochastic process $\{X_n, n = 0, 1, 2...\}$ is called a Markov chain if it satisfies Markov property (2.1) whenever the first member is defined. The conditional probability p_{ij} is the probability of transition from state i at n-th trial to the state j at $n + 1$ trial, and it is the basis for studying the structure of the Markov chain. If the transition probability is independent of n, the Markov chain is said to be homogenous (or to have stationary transition probability) and is defined by

$$Pr\{X_{n+1} = j \mid X_n = i\} = Pr\{X_n = j \mid X_{n-1} = i\}.$$

If the transition probability is dependent on n, the chain is said to be non-homogenous.

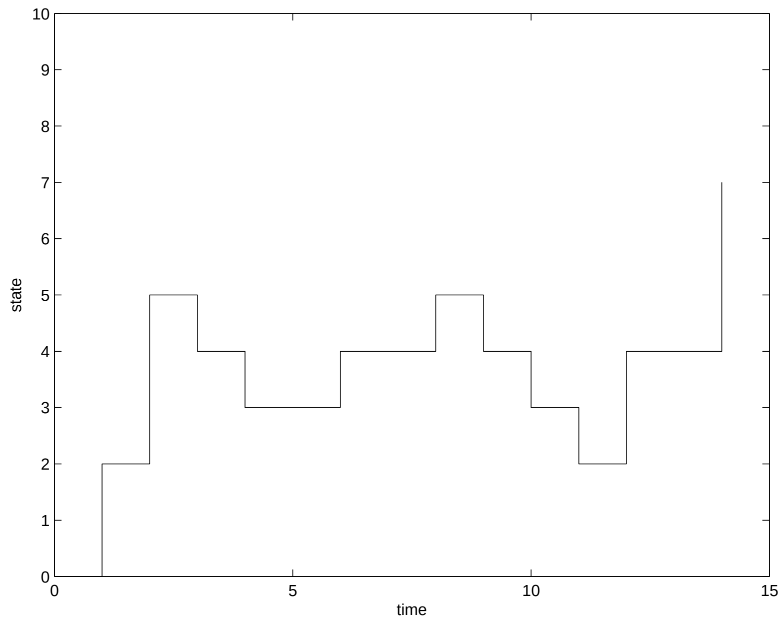

Figure 2.1: Example path of Markov chain

Stationary distribution (invariant distribution)

Let X_n be a homogeneous Markov chain with set S of finite states and transition probabilities $p_{ij} = Pr\{X_n = j \mid X_0 = i\}$. A stationary distribution is a set of numbers $\{\pi_j, j \in S\}$ such that

$$\pi_j \geq 0, \quad \sum_{j \in S} \pi_j = 1, \tag{2.2a}$$

$$\sum_{j \in S} \pi_i p_{ij}(t) = \pi_j, \quad j \in S, t > 0. \tag{2.2b}$$

The second property signifies that a stationary distribution is invariant in time; i.e. if $Pr\{X_0 = j\} = \pi_j$, then $Pr\{X_t = j\} = \pi_j$, for any $j \in S$, $t > 0$. Also the limit

$$lim_{t \to \infty} p_{ij}(t) = \pi_j, \quad i \geq 0, j \in S,$$

exists. In the matrix form we can write:

$$\Pi P = \Pi^T. \tag{2.3}$$

Chapman-Kolmogorov equation

Property (2.1) of p_{ij} gives the probability of unit-step transition from state i at a trial n to state j at the next following trial $n + 1$. The m-step transition probability is given by

$$p_{ij}^{(m)} = Pr\{X_{n+m} = j \mid X_n = i\} \quad n, m = 1, 2, ...,$$

$p_{ij}^{(m)}$ gives the probability of transition from the state i at trial n to the state j at trial $n + m$ in exactly m-steps.

The special case of the Chapman-Kolmogorov equation is given by

$$p_{ij}^{(n+m)} = \sum_r p_{ir}^{(n)} p_{rj}^{(m)} = \sum_r p_{rj}^{(n)} p_{ir}^{(m)}.$$

In a matrix form

$$P^{(n+m)} = P^{(n)} P^{(m)},$$

where $P^{(m)} = p_{ij}^{(m)}$ denotes the transition matrix of m-step transition, obtained by multiplying the matrix P by itself m times.

$$P^{(2)} = P.P = P^2$$

$$P^{(m+1)} = P^m.P = P.P^m$$

$$P^{(n+m)} = P^m.P^n = P^n.P^m$$

Example 2.1: Two-state Markov Chain

Let the state space labeled $\{1, 2\}$ and the time is $\{n, n = 1, 2, 3, ...\}$, with transition matrix given by

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \qquad 0 < \alpha, \beta < 1.$$

So the stationary distribution given by equation (2.3) is

$$\Pi^T = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right), \tag{2.4}$$

The m-step transition matrix can be computed from, [66], [31],

$$P^{(m)} = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^m}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}.$$

2.3.4 Markov Chain With Continuous Time

In the previous section we encountered the Markov chain (where the time and the states are discrete) which is characterized by the transition matrix P. In this section we introduce the Markov chain with continuous time and the state space is still discrete; we may call it the continuous time Markov chain [98] . This chain is denoted by the transition probability $p^t_{ij}(h)$ from time t to time $t+h$ over h time period and from state i to state j. When the transition probabilities are independent of the initial time t, the chain is called time-homogenous, and we must have $\sum_j p_{ij} = 1$.

Chapman-Kolmogorov equation

In the case of continuous time and discrete state space, the transition will not happen in a number of steps, but it will happen in a small time interval $(t, t+h)$. So we will use $p'_{ij}(t)$, which denote the instantaneous rate of change. Then the Chapman-Kolmogorov equation is given by

$$p'_{ij}(t) = \sum_{r \neq j} p_{ir}(t) p'_{rj}(0) + p_{ij}(t) p'_{jj}(0),$$
$$p'_{ij}(t) = \sum_{r \neq j} p'_{ir}(0) p_{rj}(t) + p'_{ii}(0) p_{ij}(t).$$

The first one is called Chapman-Kolmogorov forward equation, and the second is called Chapman-Kolmogorov backward equation. To solve these equations we will solve the equivalent forms, which needs to introduce the Q matrix.

If the state space is finite, we define a Q matrix as the following

$$Q = (q_{ij}), \quad q_{ij} := p'_{ij}(0), \quad -q_{ii} = p'_{ii}(0).$$

$q_{ij}, i \neq j$ models the rate that the chain enters j from i, q_{ii} models the rate that the chain leaves the state i. In terms of matrix Q the diagonal elements can be negative, where the other elements must be non-negative, and the sum of each row must equal zero. Then the equivalent forms for Chapman-Kolmogorov equations are given by

$$p'_{ij}(t) = \sum p_{ir}(t) q_{rj} \rightarrow \qquad P'(t) = P(t)Q,$$
$$p'_{ij}(t) = \sum q_{ir} p_{rj}(t) \rightarrow \qquad P'(t) = QP(t).$$

Thus, the solution of

$$P'(t) = P(t)Q, \quad P(0) = I,$$

is

$$P(t) = e^{tQ} = \sum_{m=0}^{\infty} \frac{(tQ)^m}{m!}.$$

If Q can be diagonalized by $Q = MDM^{-1}$ where D is the matrix of eigenvalues of matrix Q, this yields

$$P(t) = Me^{tD}M^{-1}.$$

Example 2.2: Two-state Markov Chain

In example 1.1 we explained in detail the two-state Markov chain where the time and the state are discrete. In this example we consider time is continuous and the two states are labeled again as $\{1, 2\}$. Given that state 1 is occupied at time t, there is a probability $\alpha \triangle t + o(\triangle t)$ of a transition to state 2 in $(t, t + \triangle t)$ independent of all occupance before t. If state 2 is occupied we assume a similar probability $\beta \triangle t + o(\triangle t)$ for transition to state 1, then the transition rate matrix Q is

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}. \quad 0 < \alpha, \beta < 1$$

The stationary distribution is the same as given by (2.4). Let the initial condition $p_{12}(0) + p_{21}(0) = 1$. Then the forward Chapman-Kolmogorov equations will be

$$p'_{21}(t) = -\alpha p_{21}(t) + \beta p_{12}(t),$$
$$p'_{12}(t) = -\beta p_{12}(t) + \alpha p_{21}(t).$$

The solutions which satisfy the initial condition are

$$p_{21}(t) = \frac{\beta}{\alpha+\beta} + [p_{21}(0) - \frac{\beta}{\alpha+\beta}]e^{-(\alpha+\beta)t},$$
$$p_{12}(t) = \frac{\alpha}{\alpha+\beta} + [p_{12}(0) - \frac{\alpha}{\alpha+\beta}]e^{-(\alpha+\beta)t}.$$

Whereas the probability of staying at the same state is

$$p_{11}(t) = 1 - p_{12}(t), \tag{2.5}$$

$$p_{22}(t) = 1 - p_{21}(t). \tag{2.6}$$

Hence, the stationary distribution is $\Pi^T = (p_{21}, p_{12})$ the same as the distribution specified at $t = 0$.

Finite Markov Chain

It is a generalization of the two-state Markov chain where the state space is finite $S = \{1, 2, ..., n\}$ and the time is continuous [65]. The general form of the transition probability from state i to state j at time $t > 0$ is given by [66]

$$p_{ij}(t) = \delta_{ij} e^{-q_i t} + \int_0^t q_i e^{-q_i \tau} \sum_{k \neq i} p_{ik} p_{kj}(t - \tau) d\tau,$$

where

$$\delta_{ij} = p_{ij}(0) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The stationary distribution which satisfies (2.2), will be given by

$$\sum_{k \neq j} \pi_k q_{kj} = -\pi_i q_{ij}.$$

Counting Process and Poisson Process

A counting process is a stochastic process that gives the number N_t of occurrences of the event in an interval of duration t, i.e. we start from an initial time $t = 0$, N_t will denote the number of occurrences up to t, and in addition it considers as right continuous process. The values of N_t given here are discrete and integral, and the time is continuous. The process is constant between events and jumps one unit at each event time. A Poisson process is an example of a counting process which has independent time between events [82], so that it is Markovian. Its distribution is determined by the intensity λ, the rate of transition in each event time. To find the process distribution let $p_n(t)$ be the probability that the random variable N_t assumes the value n; i.e.

$$p_n(t) = Pr\{N_t = n\}. \tag{2.7}$$

This probability is a function of time t and represents the probability distribution of the random variable N_t for every value of t. Since the only possible values of n are $n = 0, 1, 2, 3, ...$

$$\sum_{n=0}^{\infty} p_n(t) = 1.$$

Poisson process has mean equal to λt (λ being a constant). Additionally, it is a homogenous process because the probability $p_n(t)$ depends on the length of the interval $(t, t+h)$ which is h. Finally, the probability of one occurrence in the interval of length h is $\lambda h + o(h)$, and it is $o(h)$ for more than one occurrence, [96]. In other words

$$p_1(h) = \lambda h + o(h),$$

$$p_k(h) = o(h), k \geq 2,$$

$$p_0(h) = 1 - \lambda h + o(h),$$

whereas

$$h \to 0, \quad \frac{o(h)}{h} = 0.$$

Accordingly, the probability for Poisson process (2.7) is given by

$$p_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2, 3, ...$$

2.3.5 Markov Chain With Continuous State Space

The last two sections have discussed the Markov chain where the time is discrete and continuous while the state space is discrete, i.e., with $0, 1, 2, ...$ as possible values of X_n, or X_t. Here we consider chains X_n with continuous state space $(-\infty, \infty)$ as possible values of X_n. We shall have to use probability distribution function, or probability density function. X_n is said to be a Markov chain with continuous state space [2], if for all n and for all possible values of X_n in $(-\infty, \infty)$

$$Pr\{X_{n+1} \leq y \mid X_n = x, X_{n-1} = x_1, ..., X_0 = x_n\} = Pr\{X_{n+1} \leq y \mid X_n = x\} = P(y, x).$$

This formula gives the conditional probability distribution, which performs a one-step transition probability distribution function. More generally, the m-step transition probability distribution function is defined by

$$P_m(y, x) = Pr\{X_{n+m} \leq y \mid X_n = x, ..., X_0 = x_n\} = Pr\{X_{n+m} \leq y \mid X_n = x\}.$$

Chapman Kolmogorov equation

In the case of discrete time, continuous state space Chapman-Kolmogorov equation takes the form

$$P_{n+m}(y;x) = \int_{-\infty}^{\infty} dz P_n(y;z) P_m(z;x), \quad m,n \geq 0, \quad x < z < y,$$

which corresponds to $p_{ij}^{(n+m)} = \sum_r p_{ir}^{(n)} p_{rj}^{(m)}$ for Markov chain with discrete state space.

Example 2.3: Discrete-time Random Walk (Simple Random Walk)

Consider the random walk with independent identically distributed jumps $Z_1, Z_2, ...,$ which have a probability density function

$$w(z) = \frac{d}{dz} Pr(Z_n \leq z), \quad -\infty < z < \infty. \tag{2.8}$$

Suppose an unrestricted random walk that starts at the origin and is free to move indefinitely in either direction [82]. Then we have

$$\begin{aligned} X_n &= \sum_{i=1}^{n} Z_i, \\ X_{n+1} &= X_n + Z_{n+1}. \end{aligned} \tag{2.9}$$

Let us define the PDF for the particle position X_n at time n is given by [87]

$$p(x,n) = \frac{\partial}{\partial x} Pr(X_n \leq x).$$

It follows from (2.8) and(2.9) that the PDF $p(x,n)$ obeys the Kolmogorov forward equation

$$p(x,n+1) = \int_R p(x-z,n) w(z) dz, \quad n = 0, 1, ...$$

If Z_i has zero mean and finite variance σ^2, the central limit theorem ensures that the PDF for the rescaled particle position $X_n/\sqrt{n}$ tends to a Gaussian as $n \to \infty$. If the jumps Z_i have a symmetric heavy-tailed PDF with power-law index $\beta < 2$, then the variance σ^2 is infinite. According to the generalized central limit theorem, the rescaled position $X_n/n^{\frac{1}{\beta}}$ converges in distribution to the distribution of Z as $n \to \infty$.

2.3.6 Markov Process

The Markov process is a process whose state space is the continuum of real numbers in which changes of state are occurring all the time. In a small time interval such a process can only undergo a small displacement or change of state. The transition probability densities $p(x, s, y, t)$ giving the PDF of X_t (as a function of y) conditional on $X_s = x$. Thus

$$p(x, s, y, t) = Pr(X_t < y \mid X_s = x)$$

The variables x, s are the backward variables since they refer to the earlier time, and for the corresponding reason y, t are called the forward variables.
For the homogenous process, the transition probability density depends only on the length of the time interval (t, s) and can be written $p(x, y, t - s)$ giving the PDF of X_t (as a function of y) conditional on $X_s = x$ for any s.

Chapman-Kolmogorov equation

The Chapman-Kolmogorov equation connects the distribution function of the process at time t with that at an earlier time s where r is an intermediate time $(s < r < t)$

$$p(x, s, y, t) = \int_{-\infty}^{\infty} p(x, s, z, r)p(z, r, y, t)dz.$$

Example 2.4: Wiener Process

The simple random walk has the property that one-step transitions are permitted only to the nearest neighboring states [31]. Such local changes of state may be regarded as the analogue for discrete states of the phenomenon of continuous changes for continuous states. Thus if we imagine a small step of magnitude $\triangle$ taking place at small time intervals of length h, then the limit as $\triangle$ and h approach zero may be expected to produce a process whose realizations are a continuous function of time coordinates as shown in figure 2.2. This process is called Brownian motion. It has normal distribution (thin-tailed) with mean μt and variance $\sigma^2 t$ are finite. In the case of $\mu = 0$ and $\sigma^2 = 1$, the process is called a Wiener process. The PDF of the Wiener process is symmetrical about the origin for all t. The increment $X_{t+h} - X_t$ in small time interval h is independent of X_t and has zero mean and

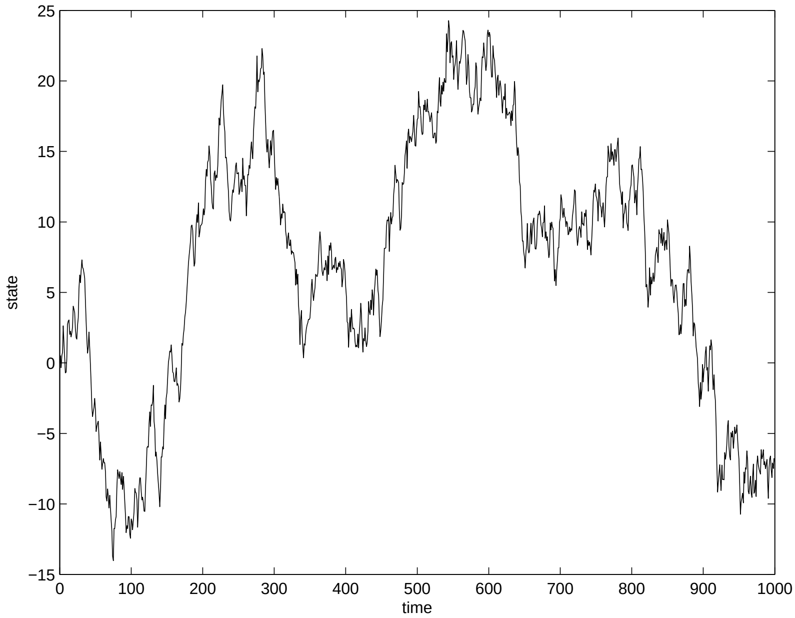

Figure 2.2: Example path of Wiener process

variance proportional to h. The Wiener process has been used widely in physics to describe the motion of a particle that is subject to a large number of small molecular shocks [62]. It is also used widely in economics in the modeling of a stock price.

2.4 The Memory Of The Process

Section 2.2.1 describes some statistical approaches of the stochastic process such as the mean, variance, covariance and correlation function. To go through the time memory of the process [77], [11], this can be evaluated from the integral of the autocorrelation function which gives the area below $R(h)$ and may take one of two possible values

$$\int_0^\infty R(h) = \begin{cases} finite \\ infinite. \end{cases}$$

If it is finite, the integral gives a typical time scale of the memory called the correlation time of the process, and in this case the process has short-range memory or is weakly dependent. If it is infinite the process has long-range memory or it is strongly dependent.

2.4.1 Short-Range Memory Stochastic Process

In the previous section we noted that the short-range memory stochastic process are characterized by a typical time memory. It can also be observed in the stochastic process characterized by an exponential decaying autocorrelation function.

$$R(h) = exp(\frac{-h}{h_c}), \quad h_c \quad \text{constant.}$$

The same statistical properties might be investigated in the frequency domain. Considering the power spectrum $S(k)$ of a random variable, which in case of stationary random process is the Fourier transform of its autocorrelation function

$$S(k) = \int_{-\infty}^{\infty} R(h)exp(ikh)dh,$$

in non-stationary process

$$S(k) = \frac{1}{k^{\beta}}.$$

If $\beta = 0$ the power spectrum corresponds to white noise, if $\beta = 2$ it corresponds to the Wiener process which is the integral of white noise. These are further evidence of a short memory range stochastic process. The same result can be obtained from the spectral density $f(w)$

$$f(\omega) = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} R(h)cos(h\omega)d\omega,$$

when the integration is finite.

In summary, short memory stochastic processes can be characterized with respect to their second order statistical properties by investigating the autocorrelation function and/or by the power spectrum. A fast decaying autocorrelation function and power spectrum resembling white noise or Wiener process for the integrated variable are fingerprints of short range correlated stochastic process (or short memory or weakly dependent).

2.4.2 Long-Range Memory Stochastic Process

If the stochastic process is characterized by a power-law autocorrelation function such as

$$R(h) = h^{\beta-1}, \quad 0 < \beta < 1, \tag{2.10}$$

or the integral of the autocorrelation function (2.10) is infinite[2], then the stochastic process is said to be strongly dependent or has long-range memory. Also if the power spectrum of the form

$$S(k) = \frac{const}{|k^{\beta}|}, \tag{2.11}$$

with $0 < \beta < 2$. The stochastic process characterized by a spectral density such as (2.11) is called $\frac{1}{f^{\beta}}$noise.

Slowly varying function

For another definition of long-range memory, see [70] [93]: Let $\{X_n\}$ be a discrete time stationary stochastic process. Introducing the concept of slowly varying function.
A real valued function $L : R \to R$ is called slowly varying function in zero (infinity) if it is bounded on any finite interval $I \subseteqq R$, and if for each $a > 0$, one has

$$\frac{L(ax)}{L(x)} \to 1 \quad , x \to 0 \quad \text{or} \quad \infty.$$

Such a function varies more slowly than any power function. Logarithms and constants are typical examples of slowly varying functions. Let $cov(h)$ be the autocovariance function of a stationary process X_n and L is a slowly varying function at infinity. The following definitions are for long-range memory (LRM):

1. The process X has LRM if there exists $0 < \alpha < 1$ such that

$$\sum_{h=-b}^{h=b} cov(h) \sim b^{\alpha} L(b), \quad b \to \infty.$$

2. The process X has LRM if there exists $0 < \beta < 1$ such that

$$cov(h) = h^{-\beta} L(h), \quad h \to \infty.$$

The parameters α,β measure the LRM intensity, in the sense that the greater the long range memory the greater the value of α and the smaller the values of β.
The topic of long memory and persistence has recently been considered the second moment (variance in the case of zero mean) of process. In this case the process X_t does not have

[2]The summation in the case of discrete time.

to be a stationary process. For the long memory process, the second moment seems to increase without limit as follows

$$E[X_t^2] \sim t^{2d+1}, \qquad 0 < d < \frac{1}{2}.$$

while for a short memory process, the second moment grows linearly with time according to

$$E[X_t^2] \sim t. \tag{2.12}$$

2.4.3 Short-Range Compared With Long-Range

Back to a Markov process which is a non-memory process and has conditional probability in the form

$$p(x_{n+1}; t_{n+1}|x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) = p(x_{n+1}; t_{n+1}|x_n; t_n).$$

For this process only the first- and second-order conditional probability densities $f(x_1, t_1)$ and $f(x_{n+1}; t_{n+1}|x_n, t_n)$ are needed to fully characterize the process and any higher order joint probability density can be determined from them. For a non-Markovian process (a process with long memory) this knowledge is not sufficient to fully characterize the stochastic process.

2.4.4 Examples of Short Memory Models

Example 2.5: Autoregressive AR model

The notation AR(p) refers to the autoregressive model of order p [22], which is a discrete model defined by

$$X_n = a_0 + \sum_{i=1}^{p} a_i X_{n-i} + \xi_n,$$

where $a_1, ..., a_p$ are the parameters of the model, ξ_n is the white noise.[3] The simplest AR(p) is an AR(1) which is given by

$$X_n = a_0 + a_1 X_{n-1} + \xi_n,$$

[3]White noise is continuous time random process with zero mean and constant variance σ^2, and its auto-correlation function is dirac delta function $R(\xi_1, \xi_2) = \delta(\xi_1 - \xi_2)$

with $|a_1| < 1$ for stationary. The process AR(1) is a stationary process with unconditional mean equal to zero and conditional mean[4] equal to $a_1 X_{n-1}$, [38]. Additionally, the unconditional variance can be found as follows

$$\begin{aligned} var[X_n] &= var(a_1 X_{n-1}) + var(\xi_n) \\ &= a_1^2 \quad var(X_{n-1}) + \sigma_\xi^2 \\ &= \frac{\sigma_\xi^2}{1 - a_1^2}. \end{aligned}$$

If $p > 1$, then

$$var[X_n] = \frac{\sigma_\xi^2}{1 - \sum_{i=1}^{p} a_i^2},$$

while the conditional variance is given by [39]

$$var[X_n] = E[X_n - E[X_n]]^2 = \sigma_\xi^2.$$

Again to evaluate the unconditional covariance when the mean is zero

$$\begin{aligned} cov[X_n, X_{n-1}] &= E[(X_n - \mu)(X_{n-1} - \mu)] \\ &= E[X_n X_{n-1}] = E[(a_1 X_{n-1} + \xi_n) X_{n-1}] \\ &= a_1 E(X_{n-1}^2) + E(\xi_n X_{n-1}) \\ &= a_1 \quad var[X_n] = a_1 \frac{\sigma_\xi^2}{1 - a_1^2}. \end{aligned}$$

$E[X_{n-k}\xi_n] = 0$ for $k > 0$, and $E[X_{n-k}\xi_n] = \sigma_\xi^2$ for $k = 0$.

$$\begin{aligned} corr[X_n, X_{n-1}] &= \frac{cov[X_n, X_{n-1}]}{\sqrt{var[X_n]}\sqrt{var[X_{n-1}]}} = \frac{cov[X_n, X_{n-1}]}{var[X_n]} \\ &= a_! \frac{\sigma_\xi^2}{1 - a_1^2} \div \frac{\sigma_\xi^2}{1 - a_1^2} = a_1. \end{aligned}$$

In general

$$corr[X_n, X_{n-k}] = a_1^k = e^{ck},$$

where $c = ln a_1 < 0$, since $|a_1| < 1$. Due to the exponential decaying of the autocorrelation function, the process has short memory.

[4]If the random variable X_n is drawn from conditional density function $f(X_n|X_{n-1})$, then conditional mean and variance depend upon past information.

Example 2.6: Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process is a stationary continuous process, [101], which solves the equation

$$dX_t = -\beta X_t dt + \sigma dW_t.$$

The explicit form of this process is given by [46]

$$X_t = X_0 e^{-\beta t} + \sigma e^{-\beta t} \int_0^t e^{\beta t'} dW(t').$$

We can find the statistical properties explicitly from the solution. So the mean is

$$\mu = E(X_t) = X_0 e^{-\beta t}.$$

To find the variance, the autocovariance and the autocorrelation function we assume for simplicity $X_0 = 0 \Rightarrow \mu = 0$. Consider X_s is

$$X_s = X_0 e^{-\beta s} + \sigma e^{-\beta s} \int_0^s e^{\beta t''} dW(t''),$$

then

$$\begin{aligned} X_t X_s &= X_0 e^{-\beta(t+s)} + \sigma X_0 e^{-\beta(t+s)} \int_0^s e^{\beta t''} dW(t''), \\ &+ \sigma X_0 e^{-\beta(t+s)} \int_0^t e^{\beta t'} dW(t'), \\ &+ \sigma^2 e^{-\beta(t+s)} \int_0^t \int_0^s e^{\beta(t'+t'')} dW(t') dW(t''). \end{aligned}$$

$$\begin{aligned} var[X_t] &= E(X_t^2) = E(X_t X_t) = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t}). \\ cov[X_t, X_S] &= corr[X_t, X_s] = E[X_t X_s] = \frac{\sigma^2}{2\beta} e^{-\beta|t-s|}. \end{aligned}$$

These result because $W(t)$ is a Wiener process, so $E(dW(t) = 0)$ and $E(dW(t')dW(t'')) = dt'\delta_{t',t''}$.

Again the autocorrelation function (which is equal to the autocovariance, since the mean is zero) is exponential decaying. Accordingly Ornstein-Uhlenbeck is a short memory stochastic process.

Example 2.7: Random telegraph process

Consider that a variable X_t can have either one of two values, a or b, and switch from one to the other with certain probabilities per unit time [46]. Thus we have the master equation

$$\partial_t P(a,t|x,t_0) = -\lambda P(a,t|x,t_0) + \mu P(b,t|x,t_0),$$

$$\partial_t P(b,t|x,t_0) = \lambda P(a,t|x,t_0) - \mu P(b,t|x,t_0),$$

where x is the initial value takes the value a or b and $P(x',t_0|x,t_0) = \delta_{x,x'}$.
Using the law of decaying exponentially: $\frac{dN}{dt} = -\lambda N \Rightarrow N = N_0 e^{-\lambda t}$, we obtain

$$\lambda P(a,t \mid x,t_0) - \mu P(b,t \mid x,t_0) = e^{-(\lambda+\mu)(t-t_0)}(\lambda\delta_{a,x} - \mu\delta_{b,x}).$$

Also we have

$$P(a,t \mid x,t_0) + P(b,t \mid x,t_0) = 1.$$

This yields

$$\begin{aligned} P(a,t \mid x,t_0) &= \frac{\mu}{\lambda+\mu} + e^{-(\lambda+\mu)(t-t_0)}\left(\frac{\lambda}{\lambda+\mu}\delta_{a,x} - \frac{\mu}{\lambda+\mu}\delta_{b,x}\right), \\ P(b,t \mid x,t_0) &= \frac{\lambda}{\lambda+\mu} - e^{-(\lambda+\mu)(t-t_0)}\left(\frac{\lambda}{\lambda+\mu}\delta_{a,x} - \frac{\mu}{\lambda+\mu}\delta_{b,x}\right). \end{aligned}$$

The statistical properties can be found

$$\begin{aligned} \mu = E[X_t] &= E(X_t|[x_0,t_0]) = aP(a,t|x_0,t_0) + bP(b,t|x_0,t_0) \\ &= \frac{a\mu+b\lambda}{\mu+\lambda} + \left(x_0 - \frac{a\mu+b\lambda}{\mu+\lambda}\right)e^{-(\lambda+\mu)(t-t_0)}, \end{aligned}$$

as $t \to \infty$ then

$$\mu = E[X_t] = \frac{a\mu+b\lambda}{\mu+\lambda}.$$

$$\begin{aligned} var[X_t] &= E[(X_t)^2] - E[X_t]^2 \\ &= \frac{(a-b)^2\lambda\mu}{(\lambda+\mu)^2}. \end{aligned}$$

Since the process is stationary the correlation function will be

$$\begin{aligned} cov[X_t,X_s] = E[X_tX_s] &= \sum xx'P(x,t|x',s)P_s(x') \\ &= \sum_{x'} P_s(x')E(x(t)|[x',s]) \\ &= \left(\frac{a\mu+b\lambda}{\lambda+\mu}\right)^2 + \frac{\mu\lambda(a-b)^2}{(\lambda+\mu)^2}e^{-(\lambda+\mu)(t-s)}, \\ corr[X_t,X_s] = cov[X_t,X_s] - \mu^2 &= \frac{\mu\lambda(a-b)^2}{(\lambda+\mu)^2}e^{-(\lambda+\mu)(t-s)}. \end{aligned}$$

So, the random telegraph process (switch process) has short memory.

Example 2.8: Two-state Markov chain

In this section we are giving the statistical properties of the two-state Markov chain explained in example 2.2, where the time is continuous and the states are labeled as 0, 1 [32].

$$E[X_t] = \frac{\alpha}{\alpha+\beta}.$$

$$var[X_t] = E[(X_t - E[X_t])^2] = E[(X_t)^2] - (E[X_t])^2 = \frac{\alpha\beta}{(\alpha+\beta)^2}.$$

Then to find the covariance and the correlation functions we need to find

$$\begin{aligned}
E[X_tX_{t+h}] &= \sum x(t)x(t+h)Pr(x(t+h), x(t)) \\
&= Pr(x(t) = x(t+h) = 1) = \frac{\alpha}{\alpha+\beta}P_1(h). \\
cov[X_t, X_{t+h}] &= E[(X_t - E(X_t)(X_{t+h} - E(X_{t+h}))] \\
&= E[X_tX_{t+h}] - [E(X_t)]^2 \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2}e^{-(\alpha+\beta)t}. \\
corr[X_t, X_{t+h}] &= \frac{cov}{var} = e^{-(\alpha+\beta)t}.
\end{aligned}$$

Consequently, a two-state Markov chain is a short memory process due to its exponential decay autocorrelation function.

2.4.5 Examples of Long Memory Models

Long memory processes appear to model the physical data and precede interests from economists and may also illustrate climate change. Long memory models have been used by econometricians since 1980. The presence of long memory can be defined from empirical data in terms of the persistence of observed autocorrelations. The extent of the persistence is consistent with an essentially stationary process, where the autocorrelations take far longer to decay than the exponential rate.

Example 2.9: Integrated process

The process X_t is said to be integrated of order α or $I(\alpha)$ [12] if

$$(1 - L)^{\alpha} X_t = \xi_t,$$

where L is the lag operator:

$$(1 - L)^{\alpha} = \sum_{j=0}^{\infty} \begin{pmatrix} \alpha \\ j \end{pmatrix} (-L)^j$$

α is a real number, usually $\frac{-1}{2} < \alpha < \frac{1}{2}$, ξ_t is a stationary and ergodic process.[5]
For $0 < \alpha < \frac{1}{2}$ the process X_t has long memory because its autocorrelations are all positive and decay at a hyperbolic rate. For $\frac{-1}{2} < \alpha < 0$ the sum of the absolute values of the autocorrelations tends to a constant, so the process has short memory.

Example 2.10: Fractional Brownian motion

Regular Brownian motion is a continuous time stochastic process $B(t)$ [12],[118]. Conversely, fractional Brownian motion $B_H(t)$ is Gaussian H-index self-similar stationary increment stochastic process where $0 < H < 1$, with $E(B_H(t)) = 0$ and $\sigma^2 = E((B_H(1))^2) = 1$.

- if $H = 1/2$, $\{B_{\frac{1}{2}}(t)\}$ is the ordinary Brownian motion with autocovariance function

$$E[B(t)B(s)] = \sigma^2 min(s, t)$$

- if $H \in (0, 1)$, then it is fractional Brownian motion and its autocovariance function is given by, see [24], [99]

$$E[B_H(t)B_H(s)] = \frac{1}{2}[t^{2H} + s^{2H} - \mid t - s \mid^{2H}]$$

For high lag h the autocovariance will be decreased as

$$E[B_H(t)B_H(s)] = \mid h \mid^{2H-2} .$$

[5] A stochastic process is said to be ergodic if its statistical properties (such as its mean and variance) can be deduced from a single, sufficiently long sample (realization) of the process.

However, it has a long range memory when $H \in (1/2, 1)$, and no memory when $H \in (0, 1/2)$. [37] Fractional Brownian motion can be regarded as the approximate $(\frac{1}{2} - H)$ fractional derivative of regular Brownian motion defined by

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t - \tau)^{H - \frac{1}{2}} dB(\tau), \quad H \in (0, 1)$$

$\Gamma(.)$ gamma function $B(t)$ is ordinary Brownian motion with unit variance.

Example 2.11 :The Switching process

The switching model is similar to the two-state Markov chain but the waiting time between the transitions is not exponentially distributed. We will generalize the process for any waiting time distribution with finite mean (Gamma waiting time as an example), and for power-law waiting time which has infinite mean. These waiting time distributions give the process memory property. We will demonstrate this process in detail later in chapter 4.

2.5 Continuous Time Random Walk

2.5.1 The Simple Random Walk

Returning to example 2.3 to illustrate the idea behind the continuous time random walk, we consider the simple random walk first with independent jump Z_i where for $i = 1, 2, ...$, $Pr(Z_i = 1) = p$, $Pr(Z_i = -1) = q$, $Pr(Z_i = 0) = 1 - p - q$. Suppose that the random walk starts at the origin and that the particle is free to move indefinitely in either direction [27]. Then we have

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

The possible positions of the particle at time n are $k = 0 \pm 1, ..., \pm n$. In order to reach the point k at time n the particle has to make i_1 positive jumps, i_2 negative jumps and i_3 zero jumps, where i_1, i_2, i_3 may be any non-negative integers satisfying the simultaneous equalities

$$i_1 - i_2 = k, i_3 = n - i_1 - i_2.$$

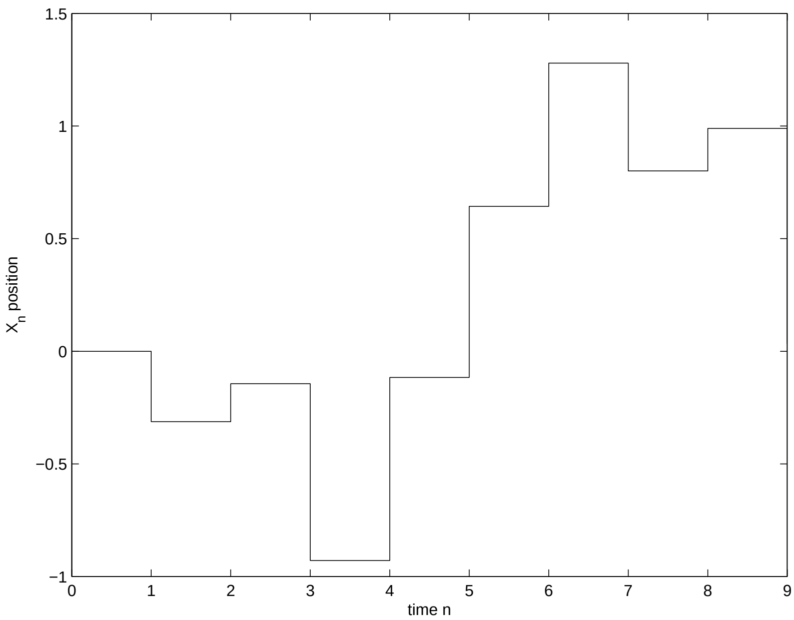

Figure 2.3: The simple random walk

Hence the probability that $X_i = k$ is given by the summation of multinomial probabilities:

$$Pr(X_n = k) = \sum \frac{n!}{i_1!i_2!i_3!} p^{i_1} q^{i_2} (1 - p - q)^{i_3}.$$

Let μ and σ^2 denote the mean and the variance of a jump X_i. Then $\mu = p - q$ and $\sigma^2 = p + q - (p - q)^2$ and hence

$$E(X_n) = n\mu, \qquad var[X_n] = n\sigma^2.$$

To find the probability of the particle at time n to be at states $j, j+1, ...k$, where j and k are possible values of $X_n (j < k)$, we resort to an approximation provided by the central limit theorem, that X_n will be approximately normally distributed with mean $n\mu$ and variance $n\sigma^2$ for large n. Thus

$$Pr(j < X_n < k) \simeq \frac{1}{\sqrt{2\pi n\sigma^2}} \int_j^k e^{-\frac{(x-n\mu)^2}{2n\sigma^2}} dx.$$

2.5.2 Continuous Time Random Walk

The name continuous time random walk (CTRW) became popular in physics after Montroll and Weiss in 1960s and 1970s [92] published a celebrated series of papers on random walks for modeling diffusion process in lattices. A spatially one-dimensional CTRW is generated

by a sequence of independent identically distributed (iid) positive random waiting times $\tau_1, \tau_2, \tau_3, ...$, such that $\tau_i = t_i - t_{i-1}$ the waiting time between two successful jumps time, each having the same probability density function $\phi(t), t > 0$. In addition, a sequence of iid random jumps length $X_1, X_2, X_3, ...$, in R each having the same probability density function $w(x), x \in R$ called the jump PDF or the transition PDF. The joint probability density function of the X_n and the waiting time τ_n is $\Phi(x,t)$ satisfies the normalization condition $\int\int \Phi(x,t)dxdt = 1$. If the waiting time and the jump length are independent then $\Phi(x,t) = \phi(t)w(x)$. Recall that $\phi(t) \geq 0$ with $\int_0^\infty \phi(t)dt = 1$ and $w(x) \geq 0$ with $\int_{-\infty}^\infty w(x)dx = 1$. Moreover, $w(x) = \int_0^\infty \Phi(x,t)dt$ and $\phi(t) = \int_{-\infty}^\infty \Phi(x,t)dx$.
Setting $t_0 = 0, t_n = \tau_1 + \tau_2 + ... + \tau_n$ for $n \in N$, the particle makes a jump of length X_n in instant t_n, so that its position is $x_0 = 0$ for $0 \leq t \leq \tau_1 = t_1$, and $x_n = X_1 + X_2 + ... + X_n$, for $t_n \leq t < t_{n+1}$.
We aim to obtain the probability density of X_t. This function is the propagator [79], defined by

$$p(x,t)dx = Pr\{x < X_t \leq x + dx\}.$$

From this definition we see that the propagator prior to the first jump, denoted by $p_0(x,t)$ is equal to

$$p_0(x,t) = \Psi(t)\delta(x),$$

where $\delta(x)$ is the dirac delta function which depending on a real parameter such that it is zero for all values of the parameter except when the parameter is zero, such as

$$\int_{-\infty}^{\infty} \delta(x)dx = 1.$$

Besides the survival function $\Psi(t)$ defined by

$$\Psi(t) = \int_t^\infty \phi(t')dt' = \int_0^\infty \phi(t')dt' - \int_0^t \phi(t')dt' = 1 - \int_0^t \phi(t')dt',$$

which denotes the probability that at instant t the particle is still sitting in its starting position $x = 0$. The waiting time is related to the survival function by the formula

$$\phi(t) = \frac{-d}{dt}\Psi(t).$$

If the distribution of the waiting times and that of the jumps are independent of each other, then by natural probabilistic arguments we arrive at the integral equation for the probability density $p(x,t)$ of the particle being in position x at time t, see [71],[72],[73]

$$p(x,t) = \delta(x)\Psi(t) + \int_0^t \phi(t-t')[\int_{-\infty}^{\infty} w(x-x')p(x',t')dx']dt', \tag{2.13}$$

where we have assumed that the initial jump occurred at $t = 0$. $p(x,t)$ satisfies the initial condition $p(x,0) = \delta(x)$. The first term of (2.13) expresses the persistence of the initial position $x = 0$. The second term gives the contribution to $p(x,t)$ from the walker sitting in point $x' \in R$ at instant $t' < t$ jumping to point x just at instant t after waiting time $t - t'$. Note that the special choice

$$w(x) = \delta(x-1)$$

gives the pure renewal process (counting process), with position $X_t = N(t)$, denoting the counting function. This function has jumps all of length 1 in positive direction happening at the renewal instants.

For simplicity the integral equation (2.13) of CTRW can be reformed in the Laplace-Fourier domain, by using the transforms [112],[30]

$$L\{\phi(t);s\} = \tilde{\phi}(s) = \int_0^{\infty} e^{-st}\phi(t)dt. \tag{2.14}$$

$$F\{w(x);k\} = \hat{w}(k) = \int_{-\infty}^{\infty} e^{ikx}w(x)dx. \tag{2.15}$$

Where $L[\delta(t)] = 1$ and $\tilde{\Psi}(s) = \frac{1-\tilde{\phi}(s)}{s}$. Thus, equation (2.13) becomes

$$\tilde{\hat{p}}(k,s) = \tilde{\Psi}(s) + \tilde{\phi}(s)\hat{w}(k)\tilde{\hat{p}}(k,s).$$

$$or$$

$$\tilde{\hat{p}}(k,s) = \tilde{\Psi}(s).\frac{1}{1-\tilde{\phi}(s)\hat{w}(k)} = \frac{1-\tilde{\phi}(s)}{s(1-\tilde{\phi}(s)\hat{w}(k))}, \tag{2.16}$$

recalling that $|\ \hat{w}(k)\ | < 1$ and $|\ \tilde{\phi}(s)\ | < 1$. This Laplace-Fourier representation is known in physics as the Montroll-Weiss equation, so named after the authors who derived it in 1965 as the basic equation for CTRW. Introducing formally in the Laplace domain the memory function

$$\tilde{H}(s) = \frac{\tilde{\Psi}(s)}{\tilde{\phi}(s)} = \frac{1-\tilde{\phi}(s)}{s\tilde{\phi}(s)}, \rightarrow \tilde{\phi}(s) = \frac{1}{1+s\tilde{H}(s)}. \tag{2.17}$$

Substituting the memory function into (2.16), we get the equivalent equation for

$$\tilde{H}(s)[s\tilde{\hat{p}}(k,s) - 1] = [\hat{w}(k) - 1]\tilde{\hat{p}}(k,s). \tag{2.18}$$

Assuming the Laplace inverse $H(t)$ exists, we get in the space-time domain the generalized Kolmogorov-Feller equation or the most familiar master equation for CTRW

$$\int_0^t H(t-t')\frac{\partial}{\partial t'}p(x,t')dt' = -p(x,t) + \int_{-\infty}^{\infty} w(x-x')p(x',t)dx', \tag{2.19}$$

with $p(x,0) = \delta(x)$, where $H(t)$ acts as a memory function. If the Laplace inverse $H(t)$ of the formally introduced function $\tilde{H}(s)$ does not exist, we can formally set the kernel function

$$\tilde{K}(s) = \frac{1}{\tilde{H}(s)},$$

and multiply (2.18) by $\tilde{K}(s)$. Then if $K(t)$ exists we get in place of (2.19) the alternative form of the generalized Kolmogorov-Feller equation, or the space-time master equation

$$\frac{\partial}{\partial t}p(x,t) = \int_0^t K(t-t')[-p(x,t') + \int_{-\infty}^{\infty} w(x-x')p(x',t')dx']dt'. \tag{2.20}$$

This equation allows us to compute $p(x,t)$ from the knowledge of the jump PDF $w(x)$ and of the waiting time $\phi(t)$. Scalas [107], demonstrates that the CTRW which is shown by the last equation is a non-Markovian model. The non-Markovian property results from the fact that at any time one has to know the value of the diffusion quantity as well as the time at which the last step took place, in order to predict the further course of walk. The non-Markovian property arises because the time of the previous step does vary and could be even $t = 0$, so the complete history of the process must be taken into account at all times. The only Markovian version of CTRW is the one in which the waiting time PDF is negative exponential.

The solution of the master equation (2.20) is evaluated in terms of $P(n,t)$, the probability of n jumps occurring up to time t, and of n-fold convolution of the jump density $w_n(x)$,

$$P(n,t) = \int_0^t \phi_n(t-\tau)\Psi(\tau)d\tau,$$

The Laplace transform of $P(n,t)$is

$$\tilde{P}(n,s) = [\tilde{\phi}(s)]^n\tilde{\Psi}(s). \tag{2.21}$$

Recalling $|\hat{w}(k)| < 1$ and $|\tilde{\phi}(s)| < 1$, if $k \neq 0$ and $s \neq 0$ equation (2.16) becomes

$$\tilde{\hat{p}}(k,s) = \tilde{\Psi}(s)\sum_{n=0}^{\infty}[\tilde{\phi}(s)\hat{w}(k)]^n, \tag{2.22}$$

inverting the Fourier-Laplace transforms gives the solution of the master equation

$$p(x,t) = \sum_{n=0}^{\infty} P(n,t)w_n(x). \tag{2.23}$$

For more details about the solution see [108].

2.5.3 Special Choices of Waiting Time Distributions

Exponential waiting time

When the waiting time is exponentially distributed, it has the following PDF

$$\phi(t) = me^{-mt}, \quad \Psi(t) = e^{-mt}, \tag{2.24}$$

$$\tilde{\phi}(s) = \frac{m}{m+s}, \quad \tilde{\Psi}(s) = \frac{1}{m+s}. \tag{2.25}$$

Thus

$$\tilde{H}(s) = \frac{1}{m},$$

given m is the average time between successive steps. Only for this form of the waiting time PDF, the probability that a step of the random walk will take a place in $(t, t+dt)$ is mdt, as $dt \to 0$, independent of the time at which the immediately preceding step occurred. This is not true of any other form of waiting time PDF, so that this is only the Markovian version of the CTRW. In this case equation (2.18) will be

$$s\tilde{\hat{p}}(k,s) - 1 = m[\hat{w}(k) - 1]\tilde{\hat{p}}(k,s),$$

inverting the Fourier- Laplace transform gives the master equation

$$\frac{\partial}{\partial t}p(x,t) = -mp(x,t) + m\int_{-\infty}^{\infty} w(x-x')p(x',t)dx' \tag{2.26}$$

If $m = 1$ this leads to classic Kolmogorov-Feller equation

$$\frac{\partial}{\partial t}p(x,t) = -p(x,t) + \int_{-\infty}^{\infty} w(x-x')p(x',t)dx'$$

Also, to find the solution of the master equation, we use (2.24) in equation (2.21), leading to

$$\tilde{P}(n,s) = \frac{m^n}{(m+s)^n}\frac{1}{m+s},$$

hence

$$P(n,t) = \frac{(mt)^n}{n!}e^{-mt}.$$

Clearly from (2.23) the solution of (2.26) is

$$p(x,t) = \sum_{n=0}^{\infty}\frac{(mt)^n}{n!}e^{-mt}w_n(x).$$

Gamma waiting time

When the waiting time has gamma distribution, its PDF will be

$$\phi(t) = \frac{\beta^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\beta t}, \quad t > 0, \alpha, \beta > 0.$$

$$\tilde{\phi}(s) = \frac{\beta^\alpha}{(s+\beta)^\alpha} \longrightarrow \tilde{H}(s) = \frac{(s+\beta)^\alpha - \beta^\alpha}{s\beta^\alpha}.$$

For simplicity, choose $\alpha = 2$. Consequently, the memory function is

$$\tilde{H}(s) = \frac{(s+\beta)^2 - \beta^2}{s\beta^2} = \frac{s+2\beta}{\beta^2}.$$

Substituting $\tilde{H}(s)$ into (2.18), yields

$$\begin{aligned} \frac{s+2\beta}{\beta^2}[s\tilde{\hat{p}}(k,x) - 1] &= -\tilde{\hat{p}}(k,x) + \hat{w}(k)\tilde{\hat{p}}(k,x) \\ &\textit{or} \\ s\tilde{\hat{p}}(k,x) - 1 &= -\frac{\beta^2}{s+2\beta}\tilde{\hat{p}}(k,x) + \frac{\beta^2}{s+2\beta}\hat{w}(k)\tilde{\hat{p}}(k,x). \end{aligned}$$

By the inverse Fourier-Laplace transform, the master equation is

$$\frac{\partial}{\partial t}p(x,t) = -\beta^2\int_0^t e^{-2\beta(t-t')}p(x,t')dt' + \beta^2\int_0^t e^{-2\beta(t-t')}[\int_0^\infty w(x-x')p(x',t')dx,]dt'.$$

The solution of this equation in the Fourier-Laplace domain is obtained from (2.22)

$$\tilde{\hat{p}}(k,s) = \frac{s+2\beta}{(s+\beta)^2}\sum_{n=0}^{\infty}\Big(\frac{\beta^2}{(s+\beta)^2}\hat{w}(k)\Big)^n.$$

Power-law memory function

In the case of a memory function $H(t)$ exhibiting a power-law decay, we get a long memory process. Such a choice is

$$H(t) = \frac{t^{-\beta}}{\Gamma(1-\beta)}, \quad 0 < \beta < 1,$$

corresponding to $\tilde{H}(s) = s^{\beta-1}$, giving

$$\tilde{\phi}(s) = \frac{1}{1+s^{\beta}}, \quad \tilde{\Psi}(s) = \frac{s^{\beta-1}}{1+s^{\beta}},$$

Likewise, the time forms

$$\phi(t) = -\frac{d}{dt}E_{\beta}(-t^{\beta}), \quad \Psi(t) = E_{\beta}(-t^{\beta})$$

where

$$E_{\beta}(-t^{\beta}) = \sum_{n=0}^{\infty}(-1)^n \frac{t^{\beta n}}{\Gamma(\beta n+1)},$$

is the Mittag-Leffler function of order β. For $0 < \beta < 1$ the Mittag-Leffler function $E_{\beta}(-t^{\beta})$ is known to be, for $t > 0$, a completely monotonic function of t, decreasing from 1 (at $t = 0$) to 0 like $t^{-\beta}$ as $t \to \infty$. For more details about the Mittag-Leffler function and its properties see[51],[107],[108],[109].

In this case we obtain the Fourier-Laplace equation (2.18) in such a way

$$s^{\beta-1}[s\tilde{\hat{p}}(k,x) - 1] = [\hat{w}(k) - 1]\tilde{\hat{p}}(k,x).$$

By converting it to the space-time domain, we get the time-fractional master equation

$$\frac{\partial^{\beta}}{\partial t^{\beta}}p(x,t) = -p(x,t) + \int_{-\infty}^{\infty} w(x-x')p(x',t)dx', \tag{2.27}$$

where $\frac{\partial^{\beta}}{\partial t^{\beta}}$ is the pseudo-differential operator, that we call the Caputo fractional derivative of order β (this will be given in detail in the next section).

The time fractional master equation can also be expressed via the Riemann-Liouville fractional derivative (the definition is explained in the next section), that is

$$\frac{\partial}{\partial t}p(x,t) = \frac{\partial^{1-\beta}}{\partial t}[-p(x,t) + \int_{-\infty}^{\infty} w(x-x')p(x',t)dx'].$$

To solve (2.27), we need to find (2.21)

$$\tilde{P}(n,s) = \frac{1}{(1+s^{\beta})^{n}} \frac{s^{\beta-1}}{1+s^{\beta}} = \frac{s^{\beta-1}}{(1+s^{\beta})^{n+1}},$$

Using the Laplace properties of the Mittag-Leffler function, [102], such as

$$L(t^{\beta n} E_{\beta}^{(n)}(-t^{\beta}; s)) = \frac{n! s^{\beta-1}}{(1+s^{\beta})^{n+1}},$$

where

$$E_{\beta}^{(n)}(z) = \frac{d^n}{dz^n} E_{\beta}(z).$$

This yields

$$P(n,t) = \frac{t^{\beta n}}{n!} E_{\beta}^{(n)}(-t^{\beta}).$$

Then the solution of (2.27) is

$$p(x,t) = \sum_{n=0}^{\infty} \frac{t^{\beta n}}{n!} E_{\beta}^{(n)}(-t^{\beta}) w_n(x).$$

2.5.4 Moments According To Characteristic Function

The Fourier transform given by equation (2.15) can be considered the characteristic function of the function $w(x)$

$$\hat{w}(k) = \int_{-\infty}^{\infty} e^{ikx} w(x) dx, \tag{2.28}$$

with associated inverse relation

$$w(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{w}(k) e^{-ikx} dk. \tag{2.29}$$

The characteristic function plays a significant role in the theory of random walk [117]. Firstly, the expression of the probability density function of the jump can be written in terms of the inverse transform of the characteristic function, equation (2.29). Secondly, the characteristic function defined by (2.28) can be expressed as the expected value (mean or the first moment). Since the expected value of the function $f(x)$ is given by

$$< f(x) >= \int_{-\infty}^{\infty} f(x) w(x) dx,$$

therefore

$$\hat{w}(k) \equiv < e^{ikx} > .$$

The characteristic function for the sum of n-steps, $\hat{w}_n(k)$ can therefore be written as

$$\hat{w}_n(k) \equiv < e^{ik(x_1+x_2+...x_n)} >=< e^{ikx_1} >< e^{ikx_2} > .... < e^{ikx_n} >= \hat{w}^n(k).$$

The same case can be considered for the Laplace transform: If T is the sum of n random times, $T = t_1 + t_2 + ... + t_n$, and if $\phi(t)$ denote the probability of a single one of the variable t. Then $\tilde{\phi}(s)$ is the Laplace transform of $\phi(t)$ defined by

$$\hat{\phi}(s) = \int_0^\infty e^{-st}\phi(t)dt \equiv < e^{-st} > .$$

This implies that the characteristic function of T, which can be written as $\tilde{\phi}_n(s)$ is

$$\tilde{\phi}_n(s) =< e^{-sT} >=< e^{-s(t_1+t_2+...+t_n)} >=< e^{-st_1} >< e^{-st_2} > ... < e^{-st_n} >= \tilde{\phi}^n(s).$$

The third property of the characteristic function is that the knowledge of this function makes it possible to calculate non-negative integer moments by differentiation rather than integral defining moments. To see how this is done, consider first that the probability density in one dimension is defined by Fourier transform (2.28). Let us differentiate this function n times giving

$$\frac{d^n\hat{w}(k)}{dk^n} = i^n \int_{-\infty}^\infty x^n e^{ikx} w(x)dx,$$

setting $k = 0$, we see the right hand side defines the n-th moment

$$< x^n >= (i)^{-n}\frac{d^n\hat{w}(k)}{dk^n} |_{k=0} .$$

We can generalize the last equation in the case of the CTRW where the probability density is $p(x,t)$ Let $< X_t^n >$. be the nth moment of the process

$$< X_t^n >= \int_{-\infty}^\infty x^n p(x,t)dx,$$

taking the Laplace transform of $< X_t^n >$ we get

$$\begin{aligned}
\int_0^\infty e^{-st} < X_t^n > dt &= \int_{-\infty}^\infty \int_0^\infty e^{-st} x^n p(x,t)dt, \\
\int_0^\infty e^{-st} < X_t^n > dt &= \int_{-\infty}^\infty x^n \tilde{p}(x,s)dx, \\
\int_0^\infty e^{-st}(i)^{-n}\frac{\partial^n \hat{p}(k,t)}{\partial k^n} |_{k=0}\, dt &\equiv < \tilde{X}^n(s) > .
\end{aligned}$$

This can be written in terms of the joint Fourier-Laplace transform of $p(x, t)$ by

$$< \tilde{X}^n(s) >= (i)^{-n} \frac{\partial^n \tilde{\hat{p}}(k, s)}{\partial k^n} |_{k=0} . \tag{2.30}$$

So, the first two moments of the CTRW by using (2.16), same results can be found in [117], are

$$< \tilde{X}(s) > \; = \; \frac{\partial \tilde{\hat{p}}(k, s)}{\partial k} |_{k=0} = \frac{< x > \tilde{\phi}(s)}{s[1 - \tilde{\phi}(s)]}, \tag{2.31}$$

$$< \tilde{X}^2(s) > \; = \; -\frac{\partial^2 \tilde{\hat{p}}(k, s)}{\partial k^2} |_{k=0} = \frac{< x^2 > \tilde{\phi}(s)}{s[1 - \tilde{\phi}(s)]} + \frac{2 < x >^2 \tilde{\phi}^2(s)}{s[1 - \tilde{\phi}(s)]^2}. \tag{2.32}$$

Equation (2.32) is the Laplace transform of the second moment that can be described as the mean squared displacement or the variance of the CTRW process. In the case of even jumps PDF, i.e. $w(x) = w(-x)$, the odd moments of jump process will vanish. Therefore, if σ_x^2 is the variance of the jump process, then the variance of the CTRW will take the following expression regarding the waiting time

- Exponential waiting time

  $$< \tilde{X}^2(s) >= \frac{\sigma_x^2 m}{s^2}.$$

  or

  $$< X_t^2 >= \sigma_x^2 mt. \tag{2.33}$$

  so that the variance always grows linearly with time as in Brownian motion with Gaussian distribution.

- Gamma waiting time with $\alpha = 2, \beta > 0$

  $$< \tilde{X}^2(s) >= \sigma_x^2 \frac{\beta^2}{s^2(s + 2\beta)},$$

  or

  $$< X_t^2 >= \sigma_x^2 [\frac{\beta t}{2} + \frac{1}{4}(e^{-2\beta t} - 1)]. \tag{2.34}$$

  Here the variance is a compound of linear proportional and exponentially decaying. However, for a long time $t \to \infty$, the variance takes the form of linearly increasing as the case of exponential waiting time.

- Power-law waiting time

$$< \tilde{X}^2(s) >= \frac{\sigma_x^2}{s^{\beta+1}}, \quad 0 < \beta < 1,$$

or

$$< X_t^2 >= \sigma_x^2 \frac{t^\beta}{\Gamma(\beta+1)}. \tag{2.35}$$

The variance in this case is proportional to a fractional power of time; in other words, it is a non-linear function of time and it is decaying slowly as $t \to \infty$.

Furthermore, this result is equivalent to the correlation function $< x(t_1)x(t_2) >$ of CTRW given by Barkai [9] for unbiased process $< x >= 0$. It is also equivalent to the mean square displacement in [8].

2.5.5 Moments of Waiting Time And Jump

The CTRW process is characterized by PDF $p(x,t)$, which provides maximal information about the evolution of the CTRW process, and its second moment. The knowledge of PDF can be followed by the existence or non-existence of the first moment (expected value) of time $< \tau >$ and the second moment $< x^2 >$ of the jump process; whereas to find the second moment of CTRW it suffices to know the waiting time PDF $\phi(t)$ and the first and the second moments of $w(x)$, as mentioned in the previous section.

We have essentially two different situations for the waiting time distribution according to its first moment being finite, or infinite, i.e. the waiting time has heavy-tails[51]. In the first case we assume the first moment is

$$< \tau >= \int_0^\infty t\phi(t)dt < \infty.$$

In the second case the first moment is infinite, so we assume the waiting time exhibits the power-law function

$$\phi(t) \sim t^{-(\beta+1)} \quad \text{for} \quad t \to \infty,$$

hence

$$\Psi(t) \sim \frac{1}{\beta} t^{-\beta}, \quad 0 < \beta < 1,$$

and the long-time limit $t \to \infty$, corresponds to

$$\tilde{\phi}(s) \sim 1 - \lambda s^{\beta} + o(s^{\beta}), as \qquad s \to 0^{+}, \quad 0 < \beta \leq 1, \tag{2.36}$$

where we have

$$\begin{aligned} \lambda &= < \tau >, \quad \text{if} \quad \beta = 1. \\ \lambda = \Gamma(-\beta) &= \frac{\pi}{\Gamma(\beta + 1) sin(\beta\pi)}, \quad \text{if} \quad 0 < \beta < 1. \end{aligned}$$

The diverging of first moment $< \tau >$ intimately related to the asymptotic behavior $\phi(t) \sim t^{-(1+\beta)}, 0 < \beta < 1$, leads to memory effects in time. This is characterized through the time fractional differential operator in the corresponding master equation as shown in equation (2.27). Besides, we have two different situations for the jump-length distribution according to its second moment being finite or infinite. In the first case we assume for the jump PDF $w(x)$

$$< x^2 >= \int_{-\infty}^{+\infty} x^2 w(x) dx < \infty, \tag{2.37}$$

whereas, in the second case when the second moment is infinite, we assume

$$w(x) \sim \mid x \mid^{-(\alpha+1)} \text{ for } \mid x \mid \to \infty, \quad 0 < \alpha < 2. \tag{2.38}$$

and the large-scale limit corresponds to

$$\hat{w}(k) \sim 1 - \sigma \mid k \mid^{\alpha} + o(\mid k \mid^{\alpha}), \quad \text{as} \quad k \to 0, \tag{2.39}$$

where

$$\begin{aligned} \sigma &= \frac{< x^2 >}{2} \quad \text{if} \quad \alpha = 2 \\ \sigma &= \frac{\pi}{\Gamma(\alpha + 1) sin(\frac{\alpha\pi}{2})} \quad \text{if} \quad 0 < \alpha < 2. \end{aligned}$$

The divergent of the second moment $< x^2 >$ intimately related to the asymptotic behavior $w(x) \sim \mid x \mid^{-(\alpha+1)}$, leads to space fractional derivative. Next we are giving some asymptotic expression of the CTRW distribution exemplified by its characteristic function $\hat{p}(k, t)$. These approximate expressions take into account the finite first moment for the waiting time $< \tau >$ and the even jump process PDF, so the first moment vanishes and its variance $\sigma_x^2 =< x^2 >$. The results can be found in [78],[79],[87],[91].

Case1: Finite jump variance

If the jump PDF $w(x)$ has a finite second moment (2.37), the asymptotic distribution of the CTRW for long time approaches to the Gaussian density

$$\hat{p}(k,t) \simeq e^{-\sigma_x^2 k^2 t/2<\tau>}, t \gg \quad <\tau> . \tag{2.40}$$

Case2: Diverging jump variance

This, due to $w(x)$, is a heavy-tailed density given by (2.38) as $|x| \to \infty$, then $\hat{w}(k)$ has an expansion of the form (2.39). Therefore, the asymptotic distribution for long time approaches L$\acute{e}$vy distribution

$$\hat{p}(k,t) \simeq e^{-\sigma_x |k|^\alpha t/<\tau>}, t \gg \quad <\tau> . \tag{2.41}$$

Case3: Finite time's expected value

At intermediate times, $t \approx <\tau>$, the behavior of $p(x,t)$ for large values of $|x|$ is the same as that of the jump distribution

$$p(x,t) \sim w(x)t/ <\tau> . \tag{2.42}$$

In the next chapter we are going to see how the CTRW process has been used in finance to model the stock returns and then find the return's volatility from the statistical property of the process.

2.6 Introduction To Fractional Calculus

Due to the fractional derivative we got from the CTRW with power-law waiting time, we introduce in this section the definition and properties of this derivative and how to define it in the time domain and in the Laplace domain. We start from the basic background of the fractional derivative [93]. Letting $X_n, n \in Z$ be a simple ARIMA$(0,k,0)$ (Auto Regressive Integrated Moving Average) process, with $k \in Z_+$, which has the following difference equation

$$(1-L)^k X_n = \sum_{j=0}^{k} \begin{pmatrix} k \\ j \end{pmatrix} (-1)^j X_{n-j} = \xi_n, \tag{2.43}$$

where ξ_n is white noise $\sim N(0, \sigma^2)$. On the left hand side we have a discrete $k-$order derivative, so that one can actually say that the $k-$th discrete derivative of an ARIMA$(0, k, 0)$ process is a white noise process.

Observe that the present value X_n depends on the past only up to time $n - k$. One can then heuristically think that an ARIMA$(0, k, 0)$ process is, in a certain sense, a short memory process. In order to make long memory appear, we let the parameter k be a real number α. Let $\alpha > 0$ be given, then ARIMA process X_n becomes a Fractional Auto Regressive Integrated Moving Average FARIMA$(0, \alpha, 0)$ process, with the following equation

$$(1 - L)^\alpha X_n = \sum_{j=0}^{\infty} \begin{pmatrix} \alpha \\ j \end{pmatrix} (-1)^j X_{n-j} = \xi_j,$$

where

$$\begin{pmatrix} \alpha \\ j \end{pmatrix} = \frac{\Gamma(1+\alpha)}{\Gamma(j+1)\Gamma(1-j-\alpha)}.$$

The value at the present time n now depends on the whole past history and the process is LRM for any $0 < \alpha < 1/2$, as mentioned in example 2.9.

The operator $(1 - L)^\alpha$ has to be interpreted as a certain type of discrete time-fractional derivative of order α. In the same way $(1 - L)^{-\alpha}$ would represent fractional integral of order α.

Consider now a smooth real function $p(t)$ defined on a closed interval $I \subset R$. The k-*th* derivative $D_t^k p(t) = p^{(k)}(t)$ could be written as

$$D_t^k p(t_n) = \frac{(1-L)^k}{\delta t^k} p(t_{n+1}), \quad k \geq 0,$$

where $\delta t = \frac{d(I)}{N-1}$, $d(I)$ is the diameter of I. Using (2.43) in the last equation one has

$$D_t^k p(t_n) = \frac{1}{\delta t^k} \sum_{j=0}^{k_*} (-1)^j \begin{pmatrix} k \\ j \end{pmatrix} p(t_{n+1-j}),$$

where $k_* = min(k, n + 1)$. In this formulation the approximation is indeed valid only in the points t_n, with $n \geq k - 1$.

Now consider $(0 < \alpha < 1)$ then,

$$D_t^\alpha p(t_n) = \frac{1}{\delta t^\alpha} \sum_{j=0}^{n+1} (-1)^j \begin{pmatrix} \alpha \\ j \end{pmatrix} p(t_{n+1-j}),$$

and it is called fractional derivative of order α. Then taking the limit $\delta t \to 0$, we got

$$D_t^\alpha p(t) = lim_{\delta t \to 0} \frac{1}{\delta t^\alpha} \sum_{j=0}^{[t/\delta t]} (-1)^j \begin{pmatrix} \alpha \\ j \end{pmatrix} p(t - j\delta t).$$

This is called the **Grunwald-Letnikov** (GL) fractional derivative of order α. It is possible to show the limit in the last equation converges to the integral equation, [93]

$$D_t^\alpha p(t) = \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t p(\tau)(t-\tau)^{-\alpha} d\tau,$$

for any $t > 0$, $0 < \alpha < 1$. This is so called Riemann-Liouville fractional derivative of order α. However the fractional derivative which we are concerned about is the Caputo fractional derivative and it will be derived in the next section.

2.6.1 Fractional Integral

First we introduce the **Riemann-Liouville** (RL) fractional integral, then the RL fractional derivative is the left inverse of the RL integral. The fractional integral of order α, where $0 < \alpha < 1$, is defined as

$$J^\alpha p(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} p(\tau) d\tau, \quad t > 0, \alpha \in R^+, \tag{2.44}$$

and $\Gamma(\alpha) = (\alpha - 1)!$, for more details see [25], [69], [102].
For complementarily $J^0 = I$ (Identity operator) i.e., $J^0 p(t) = p(t)$. Furthermore $J^\alpha p(0^+)$ it means the limit of $J^\alpha p(t)$ for $t \to 0^+$, this limit may exist or be infinite. Also

$$J^\alpha t^\gamma = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1+\alpha)} t^{\gamma+\alpha}, \quad \alpha > 0, \gamma > -1, t > 0.$$

2.6.2 Fractional Derivative

The fractional derivative of order $\alpha > 0$ is defined by $D^\alpha p(t) = D^m J^{m-\alpha} p(t)$. This implies

$$D^\alpha p(t) = \begin{cases} \frac{d^m}{dt^m} [\frac{1}{\Gamma(m-\alpha)} \int_0^t p(\tau)(t-\tau)^{m-\alpha-1} d\tau], & m-1 < \alpha < m, \\ \frac{d^m}{dt^m} p(t), & \alpha = m \end{cases} \tag{2.45}$$

which is namely the standard RL fractional derivative [102],[26]. Complementarily $D^0 = J^0 = I$, $D^\alpha J^\alpha = I$, D^α is left inverse but not right, i.e. $J^\alpha D^\alpha \neq I$. Also

$$D^\alpha t^\gamma = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1-\alpha)} t^{\gamma-\alpha}, \quad \alpha > 0, \gamma > -1, t > 0. \tag{2.46}$$

$$D^\alpha 1 = \frac{t^{-\alpha}}{\Gamma(1-\alpha)}, \quad \alpha \geq 0, t > 0.$$

The alternative definition of fractional derivative originally introduced by Caputo is called **Caputo fractional derivative** of order $\alpha > 0$: $D_*^\alpha p(t) = J^{m-\alpha} D^m p(t)$, given by

$$D_*^\alpha = \begin{cases} \frac{1}{\Gamma(m-\alpha)} \int_0^t p^{(m)}(\tau)(t-\tau)^{m-\alpha-1} d\tau, & m-1 < \alpha < m \\ \frac{d^m}{dt^m} p(t), & \alpha = m \end{cases} . \tag{2.47}$$

To distinguish between the two derivatives

$$D^\alpha p(t) = D^m J^{m-\alpha} p(t) \neq J^{m-\alpha} D^m p(t) = D_*^\alpha p(t),$$

unless the function $p(t)$ along with its first $m-1$ derivatives vanishes at $t = 0^+$. Otherwise for $m-1 < \alpha < m$, and $t > 0$

$$D^\alpha p(t) = D_*^\alpha p(t) + \sum_{k=0}^{m-1} \frac{t^{k-\alpha}}{\Gamma(k+1-\alpha)} p^{(k)}(0^+), \tag{2.48}$$

and therefore, recalling the fractional derivative of the power function (2.46)

$$D_*^\alpha p(t) = D^\alpha \left(p(t) - \sum_{k=0}^{m-1} \frac{t^k}{k!} p^{(k)}(0^+) \right).$$

In particular, the relevant property for which the fractional derivative of a constant is still zero

$$D_*^\alpha 1 \equiv 0, \quad \alpha > 0. \tag{2.49}$$

However, the most relevant difference between RL (2.45) and Caputo (2.47) fractional derivatives is the derivative of power function. This derivative is given by (2.46) in case of RL fractional derivative, and in case of Caputo fractional derivative is given by

$$D^\alpha t^{\alpha-1} \equiv 0, \quad , \alpha > 0, t > 0.$$

From (2.48) and (2.49) we thus recognize the following statements about functions which for $t > 0$ admit the same fractional derivative of order α, with $m-1 < \alpha \leq m$, [49]

$$D^\alpha p(t) = D^\alpha g(t) \Leftrightarrow p(t) = g(t) + \sum_{j=1}^{m} c_j t^{\alpha-j},$$

$$D_*^\alpha p(t) = D_*^\alpha g(t) \Leftrightarrow p(t) = g(t) + \sum_{j=1}^{m} c_j t^{m-j}.$$

where the coefficients c_j are arbitrary constants.

Again from (2.49), we can prove that D^α is not right inverse to J^α, since

$$J^\alpha D^\alpha t^{\alpha-1} \equiv 0, \text{but} \quad D^\alpha J^\alpha t^{\alpha-1} = t^{\alpha-1}, \quad \alpha > 0, t > 0.$$

We now consider the Laplace transform for the two fractional derivatives. For RL fractional derivative D^α the Laplace transform, assumed to exist, requires the knowledge of the initial values of fractional integral $J^{m-\alpha}$ and of its integer derivatives of order $k = 1, 2, .., m-1$. Firstly, from (2.44) we give the Laplace transform for the fractional integral

$$L\{J_t^\alpha p(t), s\} = s^{-\alpha}\tilde{p}(s) = \frac{\tilde{p}(s)}{s^\alpha}. \tag{2.50}$$

Hence, the Laplace transform for RL fractional derivative will be

$$L\{D^\alpha p(t), s\} = L\{D^m J^{m-\alpha} p(t), s\} = s^\alpha \tilde{p}(s) - \sum_{k=0}^{m-1} s^k D^{\alpha-1-k} p(0^+),$$

The Caputo fractional derivative appears more suitable to being treated by the Laplace transform in that it requires the knowledge of the initial values of the function, and of its integer derivatives of order $k = 1, 2, .., m-1$. In fact, by using (2.50) and noting that

$$J^\alpha D_*^\alpha p(t) = J^\alpha J^{m-\alpha} D^m p(t) = J^m D^m p(t) = p(t) - \sum_{k=0}^{m-1} \frac{t^k}{k!} p^{(k)}(0^+),$$

imply that the Laplace transform as follows

$$L\{D_*^\alpha p(t), s\} = s^\alpha \tilde{p}(s) - \sum_{k=0}^{m-1} s^{\alpha-1-k} p^{(k)}(0^+).$$

This is the Caputo derivative which appears in the master equation of the CTRW when the waiting time is power-law or heavy-tailed distributed. We may conclude that the appearance of the Caputo fractional derivative in an equation of the probability density of any process is evidence of memory effect. Therefore the process has long memory because, to calculate the present value of the variable, this needs the knowledge of the history of the variable.

Chapter 3

Review Of Volatility Models

3.1 Introduction

This chapter is concerned with the different volatility models. First, the Black-Scholes (BS) model assumes that volatility is constant; this contradicts the implied volatility. The implied volatility shows skewness and this suggests the assumption of constant volatility is not feasible. In other words, volatility shows an intermittent behavior with periods of high values and periods of low values this phenomena is called volatility smile. In addition, the asset volatility cannot be directly observed, hence stochastic volatility deals with these two facts. The stochastic volatility models employ the diffusion process to model the volatility which is different from the diffusion process that models the stock price. A review of famous stochastic volatility models is given, (see table 3.1), in addition to the assumptions and the shortcomings of these models. On the other hand, there have been some efforts to predict the change in the volatility by using heteroscedasticity models. However, these models failed to describe the volatility due to their short memory. Indeed, the empirical data has shown that volatility has long memory; this fact means the stochastic volatility models are not appropriate any longer, due to their Markovian nature. Later on, the efforts focused on employing long memory stochastic process to model volatility. Accordingly some extensions were applied to stochastic volatility models and heteroscedasticity models in order to get a more appropriate fitting for market data. First in this chapter we show how to model the stock price when it follows a geometric Brownian motion, or lognormal return. This model is widely used in stochastic volatility models; however, the distribution of this process is Gaussian and it has thin-tails and Markovian property. The empirical data exhibits fat tails which make the Gaussian distribution inappropriate, especially at the wings. Instead of Gaussian distribution the attempts suggest the distribution of returns or of logarithmic return followed a stable distribution; there was a barrier to the application of these concepts in the financial practice due to the infinite variance of stable distributions. The attempts to find a more appropriate model for stock price continued. A new market model proposed to fill the gap between Gaussian and stable distributions. The model, based on a continuous jump process, explains the appearance of fat tails and self-scaling but still keeps all moments finite. An alternate process is the CTRW; this process is non-Markovian, and in addition it exhibits skewness and heavy-tails. The idea behind using the CTRW in financial data, is that the randomness appears in waiting time between transactions as well as the stock price or stock's return change.

3.2 Modeling Stock Price

Suppose that a stock price S_t follows a geometric Brownian motion [62],[10] that has a constant drift rate μ and a constant variance σ^2

$$dS_t = \mu S_t dt + \sigma S_t dW, \tag{3.1}$$

where W is a standard Brownian motion (Wiener process). If S_t is the stock price at time t, the expected drift rate in S should be assumed to be μS_t, where parameter μ is the expected rate of return on the stock. This means that in a short interval of time $\triangle t$, the expected increase in S_t is $\mu S_t \triangle t$. If the volatility of the stock price (the variance) is zero, model (3.1) implies that

$$\frac{dS_t}{S_t} = \mu dt.$$

Integrating between time zero (the initial time) and time T (the maturity), we get

$$S_T = S_0 e^{\mu T},$$

where S_0, S_T are the stock price at time zero and time T. The last equation shows that when the volatility is zero, the stock price grows at a rate μ per unit of time. However, in practice, stock does exhibit volatility. A reasonable assumption is that the variability of the percentage return in a short period of time $\triangle t$ is the same regardless of the stock price. This suggests that the standard deviation of the change in a short period of time $\triangle t$ should be proportional to the stock price and leads to the model (3.1). This model is the most widely used model of stock price behavior, where σ is the volatility of the stock price and μ is its expected return. To price the option or any other derivative f_t contingent on S_t, the variable f_t must be a function of S_t and t. Then, applying Itô's lemma, it follows that the process defined by the function f will satisfy

$$df_t = (\frac{\partial f_t}{\partial S_t}\mu S_t + \frac{\partial f_t}{\partial t} + \frac{1}{2}\frac{\partial^2 f_t}{\partial^2 S_t^2}\sigma^2 S_t^2)dt + \frac{\partial f_t}{\partial S_t}\sigma S_t dW.$$

Lognormal Property of Stock Price

If we consider S_t follows the model (3.1) and we define $f_t = lnS_t$, then by applying Itô's lemma we can observe that

$$df_t = (\mu - \frac{\sigma^2}{2})dt + \sigma dW,$$

that is, f_t has constant drift rate $\mu - \frac{\sigma^2}{2}$ and constant variance σ^2. This means the change in lnS_t between time zero and some future time is normally distributed

$$LnS_T - lnS_0 \sim N\left[(\mu - \frac{\sigma^2}{2})T, \sigma\sqrt{T}\right],$$

or

$$Ln\frac{S_T}{S_0} \sim N\left[(\mu - \frac{\sigma^2}{2})T, \sigma\sqrt{T}\right]. \tag{3.2}$$

This gives what is known as lognormal property of stock price or lognormal return. Hence

$$S_T = S_0 e^{(\mu-\frac{1}{2}\sigma^2)T+\sigma\sqrt{T}\xi}, \quad \xi \sim N(0,1).$$

This, what is often referred to as the stock price follows a geometric Brownian motion, and is the solution of (3.1).

Stock's return

If we take the discrete time version of equation (3.1) we have [59]

$$\triangle S = S(t_{i+1}) - S(t_i) = \mu S(t_i)\triangle t + \sigma S(t_i)\xi\sqrt{\triangle t},$$ [1]

[1] $dW = \xi\sqrt{\triangle t}$

then the daily return is

$$\frac{S(t_{i+1}) - S(t_i)}{S(t_i)} = \mu\triangle t + \sigma\xi\sqrt{\triangle t}.$$

However, in finance we usually mention the log return instead of daily return. This is obtained from the following

$$log\frac{S(t_{i+1})}{S(t_i)} = log(1 + \frac{S(t_{i+1}) - S(t_i)}{S(t_i)}) \approx \frac{S(t_{i+1}) - S(t_i)}{S(t_i)}.$$

3.3 Volatility

In finance, a volatility σ of a stock is a measure of our uncertainty about the return provided by the stock, the uncertainty of the stock return is measured by its standard deviation [62], volatility can also be referred as the variance σ^2. The volatility is the main factor which affects the stock price and it is caused by the random arrival of new information about the future returns from the stock, or by trading or supply and demand. The imperative of volatility leads scientists' efforts to focus on evaluating and modeling it, or calculating it from any irregular-shaped distribution or stochastic process describing the return process. In this case, the volatility probability and cumulative distribution function can be derived empirically. Although we can estimate volatility empirically from the historical data over a long period to get an actual value for volatility, still this method is not accurate because the volatility does change over time, and data that are too old may not be relevant for predicting future volatility. However, it can be derived analytically only when volatility is attached to a standard distribution function, such as normal distribution. Our work in the next two chapters attempts to derive the probability distribution of volatility under a non-Markovian stochastic process.

Estimating Volatility from Historical Data

To estimate the volatility of a stock price empirically [62], the stock is usually observed at fixed intervals of time (e.g. every day, week or month). Let $n + 1$ be the number of observations. S_i is the stock price at the end of the i-th interval ($i = 0, 1, ..., n$), and τ the length of the time interval in years. Therefore, if the log return is

$$u_i = ln(\frac{S_{i+1}}{S_i}) \quad \text{for} \quad i = 1, 2, ..., n.$$

Then the usual estimate s of the standard deviation of the $u_i's$ is given by

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - \overline{u})^2},$$

or

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}u_i^2 - \frac{1}{n(n-1)}(\sum_{i=1}^{n}u_i)^2}.$$

The standard deviation of the u_i's is $\sigma\sqrt{\tau}$, since the standard deviation is increasing with the square root of time [62]. Therefore the variable s is an estimate of $\sigma\sqrt{\tau}$. It follows that σ itself can be estimated as $\hat{\sigma}$

$$\hat{\sigma} = \frac{s}{\sqrt{\tau}}.$$

Implied Volatility

Direct measurement of volatility is difficult in practice, and it is not obvious that the historic volatility is independent of the time series from which it is calculated, nor that it accurately predicts the future volatility that we require over the lifetime of an option [120]. All of these difficulties lead us to find another way to calculate the volatility, which takes the option price quoted in the market and observed from BS model, then, working backwards, deduce the market's opinion of the value of the volatility over the remaining life of the option. This volatility derived from the quoted price for a single option is called implied volatility and is not constant across exercise prices. That is, if we fixed the value of underlying, interest rate and maturity, the prices of options across exercise price reflect a U-shape for volatility. This curve is called volatility smile, (see figure 3.1). Another feature of implied volatility is that it is equal for call option and put option when they have the same exercise price K and maturity T,

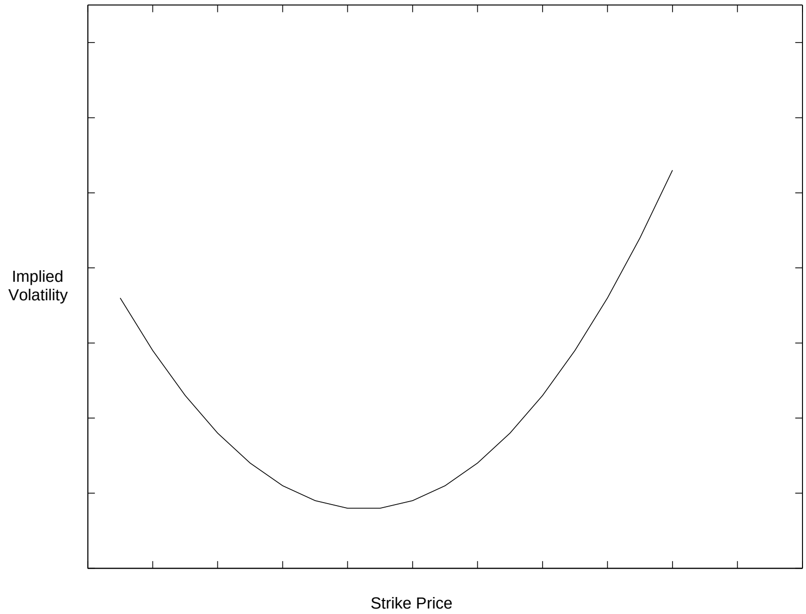

Figure 3.1: The smile curve of implied volatility

3.4 Modeling Volatility

In this section we show the most important volatility models, starting with the BS model that assumed a constant volatility, extending the BS model by considering volatility as a deterministic function, extending BS model by considering volatility as a stochastic variable and that it follows a diffusion process. On the other hand, the predictable models or autoregressive conditional heteroscedastic models include ARCH and GARCH processes, then the long memory models deduced from GARCH modification. Finally, we illustrate using the CTRW to model the log return, instead of geometric Brownian motion with normal distribution that was used in the stochastic volatility models to model the log return, and finding the return's volatility from the process moments.

3.4.1 Constant Model

It is obvious that when we refer to a constant volatility model we mean the Black-Scholes-Merton model, which was the first and widely used model for pricing options. This model was developed in 1970s by Fischer Black, Myron Scholes and Robert Merton. Black-Scholes-Merton derived their model under the following assumptions:

- The stock price follows the model (3.1).
- There are no transaction costs or taxes; all securities are perfectly divisible.
- There are no dividends during the life of the derivatives.
- There are no riskless arbitrage opportunities.
- Security trading is continuous.
- The risk free rate of interest, μ, is constant and the same for all maturities.
- Short selling is permitted.

Regarding these assumptions and their portfolio of option and portion of current stock price, $\Pi = -f_t + \frac{\partial f_t}{\partial S_t} S_t$, they built up their famous differential equation (model), that is

$$\frac{\partial f_t}{\partial t} + \mu S_t \frac{\partial f_t}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f_t}{\partial S_t^2} - \mu f_t = 0. \tag{3.3}$$

The model has had a huge influence on the way the traders price and hedge options. Despite the popularity of the BS model it is criticized for many reasons, [104], such as:

- The assumption of the geometric Brownian motion as the model of the stock price is not in accordance with the statistical data. For example, it may be shown that for much data the log returns ($log \frac{S_{t+i}}{S_t}$), are not normally distributed as it is in the case of the BS model.
- It is obvious that the volatility of much data is not constant. From the empirical observation there are periods with high variance for the stock and others with low variance.
- Often jumps in the stock price can be observed due to some extraordinary incidences. These discontinuities cannot be modeled by Brownian motion.
- The interest rate is not constant in time and is different between borrowing and lending.
- There are transaction costs on the market.

- The BS theory states there are no arbitrage arguments; this leads to formulation of a risk-free option price. However, it is well documented that arbitrage opportunities do exist.

Another attempt by Avellaneda *et al.* [5] assumed a constant volatility, but it is not known precisely. They assumed the volatility is bounded by two extreme values, σ_{min} and σ_{max}. The constants σ_{min} and σ_{max} represent upper and lower bounds on the volatility that should be input in the model according to the investor's expectation and uncertainty. The BS principle and model (3.3) still apply here. Instead of obtaining one value for option, they obtain a range of values with a maximum f_t^+ and minimum f_t^-. This model is reduced to the BS model when $\sigma_{min} = \sigma_{max}$ and does not provide any significant improvement to the BS model.

3.4.2 Deterministic Models

One of many attempts to relax the constant volatility is to assume that the volatility is driven by a deterministic function. One such popular approach is to modify the log-normal model by assuming volatility to be a deterministic function of stock price S_t and/or time t that is $\sigma(S_t, t)$ (see [32], [106],[36]) . The stochastic differential equation modeling the stock price is

$$dS_t = \mu S_t dt + \sigma(S_t, t) S_t dW. \tag{3.4}$$

Then the option price in a deterministic volatility environment satisfies

$$\frac{\partial f_t}{\partial t} + \mu S_t \frac{\partial f_t}{\partial t} + \frac{1}{2}\sigma^2(S_t, t) S_t^2 \frac{\partial^2 f_t}{\partial S_t^2} - \mu f_t = 0.$$

For the special case where volatility is just the time dependent that is $\sigma(S_t, t) = \sigma_t$, one can solve (3.4) and deduce that $ln(\frac{S_T}{S_t})$ is normally distributed with the mean $(\mu - \frac{1}{2}\bar{\sigma}^2)(T - t)$ and the variance $\bar{\sigma}^2(T - t)$ such that:

$$ln(\frac{S_T}{S_t}) \sim N((\mu - \frac{1}{2}\bar{\sigma}^2)(T - t), \bar{\sigma}^2(T - t)),$$

where the mean variance $\bar{\sigma}^2$ satisfies

$$\bar{\sigma}^2 = \frac{1}{T - t}\int_t^T \sigma^2(s) ds. \tag{3.5}$$

This option price still obeys the BS model (3.3) with σ replaced by $\sqrt{\bar{\sigma^2}}$. In general, empirical tests have shown that the constant or the deterministic volatility model failed to price the option more accurately than the BS model. Due to that failure, the variance of the stock return is empirically observed to be non-stationary, in other world the variance (volatility) is random.

Hence, the latter research focused on modeling volatility as a stochastic variable following another stochastic process different from the stock price process (3.1).

3.4.3 Stochastic Models

Because of the shortage of the BS model due to the constant volatility, and the failure of the deterministic volatility model to give a more accurate result, the following attempts are to model the volatility itself as a stochastic process. The Extension of the BS model began appearing after the original paper in 1973. For example, Merton [88] added further assumptions, such as:

- Trading takes place continually and borrowing and short selling are allowed without restriction; the borrowing rate equals the lending rate.
- Stock price dynamics: the instantaneous expected return on the common stocks is described by the stochastic differential equation

  $$\frac{dS_t}{S_t} = \mu dt + \sigma_t dW,$$

  where μ is the instantaneous return and σ_t^2 is the instantaneous variance of the return and W is the standard Wiener process. μ may be stochastic variable dependent on the stock price. So $\frac{dS_t}{S_t}$ is not an independent process or stationary, although W is. σ is restricted to be a function of time.

Merton incorporated jump diffusion models to the price of the underlying asset [89]. However, the result deviations from the BS option price are typically quite small, at the cost of having estimated more complex parameters. The same approach was introduced by Cox and Ross [32].

After that, many empirical studies by practitioners expressed a significant proof that the implied volatility of market price varies with strike price and time to maturity of the option contract. This particular shorting of the BS model is remedied by the stochastic volatility model to price option first studied by Hull and White (1987) [63], Scott(1987) [110], Wiggins (1987) [119], Stein and Stein (1991) [113], Heston (1993) [57], Ball and Roma (1994)[7] and Bouchaud (1994) [21]. The disadvantage of these models are: firstly, because some of them do not give a closed-form solution and require extensive use of numerical techniques to solve two-dimensional partial differential equations; secondly, the volatility processes are considered to exhibit the short memory property, while the empirical evidence exhibits that the volatility has long memory property; thirdly, the first three models assumed that volatility is uncorrelated with the stock price; however, this assumption is relaxed later by Heston. The following table gives a summary of the famous stochastic volatility models.

Table 3.1: Stochastic Volatility Models

Author(s)	Correlation	Risk-Premium	Volatility Process
Hull-White	$\rho = 0$	$\lambda* = 0$	$dV_t = \phi V_t dt + \beta V_t dW_2$
Scott	$\rho = 0$	$\lambda* = 0$	$dlnV_t = \phi(\theta - lnV_t)dt + \beta dW_2$
Stein-Stein	$\rho = 0$	$\lambda*$ =const	$dV_t = \phi(\theta - V_t)dt + \beta dW_2$
Ball-Roma	$\rho = 0$	$\lambda* = 0$	$dV_t = \phi(\theta - V_t)dt + \beta \sqrt{V_t} dW_2$
Heston	$\rho \neq 0$	$\lambda* = \bar{\lambda} V_t$	$dV_t = \phi(\theta - V_t)dt + \beta \sqrt{V_t} dW_2$

General Stochastic Volatility Model

The idea of incorporating the stochastic volatility model is to assume the stock return variance to be modeled by a diffusion process [111].

Consider the option f, whose price depends on the stock price S_t and its instantaneous variance, $V_t = \sigma_t^2$. The stock price is assumed to be modeled by the following stochastic process

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_1,$$

whereas the square of volatility process $V_t = \sigma_t^2$ is a continuous solution of

$$dV_t = \phi(V_t, t)dt + \beta(V_t, t)dW_2,$$

where W_1 and W_2 are the Wiener process. The drift rate μ is a parameter that may depend on S_t, σ_t and t. For simplicity μ is generally assumed to be constant in literature. Both parameters ϕ and β may depend on σ_t and t but are generally assumed not to depend on S_t. One also makes the simplifying assumption that the Wiener process W_1 and W_2 have constant correlation $\rho \in [-1, 1]$:

$$E[dW_1, dW_2] = \rho dt.$$

The main aim is to formulate the option price $f(S_t, V_t, t)$, as a smooth function of the stock price and the square of the volatility process. To achieve this one must construct a portfolio of assets using the no-arbitrage argument. Unlike the classical BS case, we need to eliminate an additional risk due to the Wiener process W_2. It is important to realize that the option price is now driven by two sources of randomness W_1, W_2. At any time, the option used in the BS case is no longer perfectly correlated with the stock price. Hence it is insufficient to hedge the portfolio solely with the underlying asset. To eliminate the randomness due to the volatility fluctuations W_2, we use another option which has the same strike price as the previous one but with a different expiry date. Consider the new hedge strategy for European call option

$$f_t^{(1)} = aS_t + bB_t + cf_t^{(2)},$$

where $f_t^{(1)}, f_t^{(2)}$ are options with different expiry dates, B_t is the price of a long term riskless bond at time t that must earn the risk-free interest rate r, such that $dB_t = rB_t dt$. To eliminate the randomness due to the Wiener process, [111] setting

$$\begin{aligned} a &= \frac{\partial f_t^{(1)}}{\partial S_t} - c\frac{\partial f_t^{(2)}}{\partial S_t}, \\ c &= \frac{\partial f_t^{(1)}}{\partial V_t} / \frac{\partial f_t^{(2)}}{\partial V_t}, \\ b &= (f_t^{(1)} - aS_t - cf_t^{(2)})/B_t. \end{aligned}$$

Then the general model for an option price in a stochastic volatility is

$$\begin{aligned} \frac{\partial f_t}{\partial t} + \frac{1}{2}V_t S_t^2 \frac{\partial^2 f_t}{\partial S_t^2} + \rho S_t \sqrt{V_t}\beta(V_t, t)\frac{\partial^2 f_t}{\partial S_t \partial V_t} + \frac{1}{2}\beta^2(V_t, t)\frac{\partial^2 f_t}{\partial V_t^2} + \\ (\phi(V_t, t) - \lambda^*\beta(V_t, t))\frac{\partial f_t}{\partial V_t} + \mu(S_t\frac{\partial f_t}{\partial S_t} - f_t) = 0, \qquad (3.6) \end{aligned}$$

where λ^* denotes the market price of the volatility risk also known as the risk-premium[2]. The call option must satisfy the following boundary conditions

$$C(S_T, V_T, T) = max(S_T - K), \quad C(0, v_t, t) = 0, \tag{3.7a}$$

$$C(S-t, V_t, t) \sim S_t - Ke^{-\mu(T-t)} \quad \text{as} \quad S_t \to \infty. \tag{3.7b}$$

Equation (3.6) was first suggested by Garman [47]. It is similar to the BS model but with additional terms representing the dependence of an option price on the volatility process. We should remark that the resulting option price does not depend on the expiry date but depends on the risk-premium λ^*. Since volatility is not a directly tradeable asset, a perfect replicating strategy ceases to exist. This implies that the risk in the market due to stochastic volatility cannot be completely hedged away. The parameter λ^* represents how the market values this unhedged risk. It is an unobservable parameter and in general $\lambda^* = \lambda(S_t, \sigma_t, t)$ depends on all three variables S_t, σ_t and t. Its presence is a major disadvantage in all the stochastic volatility models because it cannot be deduced accurately, owing to the risk-aversion level. Using the martingale method, a general solution to (3.6) for European call option can be written as

$$C(S_t, V_t, t) = e^{-\mu(T-t)} E^{Q(\lambda^*)}[max(S_t - K, 0)], \tag{3.8}$$

where Q is a martingale measure depending on the risk-premium λ^*. There is usually no closed-form solution for (3.6) due to mathematically intractable likelihood functions associated with the estimation of λ^*. In the event that λ^* is independent of stock price S_t, (3.8) can be simplified to

$$C(S_t, V_t, t) = E^{Q(\lambda^*)}[C_{BS}(\sqrt{\bar{V}_t})], \tag{3.9}$$

where C_{BS} denotes the Black-Scholes price and $\bar{V}_T$ is the mean variance which satisfies the stochastic integral (3.5).

[2] A risk premium is the minimum amount of money by which the expected return on a risky asset must exceed the known return on a risk-free asset, in order to induce an individual to hold the risky asset rather than the risk-free asset.

Hull and White Model

Hull and White (1987) consider the following model

$$dS_t = \mu S_t dt + \sigma S_t dW_1, \tag{3.10a}$$

$$dV_t = \phi V_t dt + \beta V_t dW_2, \tag{3.10b}$$

where S_t denotes the stock price at time t, $V_t = \sigma_t^2$ its instantaneous variance. W_1, W_2 are the Wiener process. The μ is a parameter that may depend on S_t, σ_t and t. ϕ, β may depend on σ_t and t but not on S_t. Since S_t, σ_t^2 are the only state variables affecting the price of the derivative asset (in this case call option) C_t, the parameter μ must be constant or at least deterministic. Similar models have been studied by Scott [110] and Wiggins [119].

The basic idea to price a call option in this model is to form a riskless portfolio containing the option, the stock, and the second call option with the same strike price, but different expiration date. If we denote $C(S_t, t)$ the value of the first call option at time t and stock price S_t, this approach yields a certain partial differential equation for the option pricing function $C(S_t, t)$

$$\frac{\partial C_t}{\partial t} + \frac{1}{2}\Big(\sigma_t^2 S_t^2 \frac{\partial^2 C_t}{\partial S_t^2} + 2\rho\sigma_t^3 \beta S_t \frac{\partial^2 C_t}{\partial S_t \partial V_t} + \beta^2 V_t^2 \frac{\partial^2 C_t}{\partial V_t^2}\Big) - \mu f_t = -\mu S_t \frac{\partial C_t}{\partial S_t} - \phi\sigma_t^2 \frac{\partial C_t}{\partial V_t}.$$

However, the solution of this equation is not unique unless one already knows the price function for the second option.

To recover uniqueness, Hull and White make the additional assumption that W_1, W_2 are independent, and the variance V_t has no systematic risk. This yields a unique option price which can be computed as the expectation of the discounted terminal payoff under a risk-neutral probability measure $\tilde{P}$[3].Then the option price is given by

$$C(S_t, \sigma_t^2, t) = e^{-\mu(T-t)} \int C(S_T, \sigma_T^2, T)\tilde{P}(S_T | S_t, \sigma_t^2) dS_T = e^{-\mu(T-t)} E^Q[C(S_T, \sigma_T^2, T)].$$

To obtain a more specific form for C_t, Hull and White use the additional assumption that the instantaneous variance V_t is not influenced by the stock price S_t. Setting $\bar{V}_t$ as the mean

[3]A risk-neutral measure, which is an equivalent martingale measure, or Q-measure is a probability measure that results when one assumes that the current value of all financial assets is equal to the expected future payoff of the asset discounted at the risk-free rate

variance over the life of the derivative security given by (3.5) with h is the probability density function in a risk-neutral world. This allow them to price a European call option as

$$C(S_t, \sigma_t^2) = \int C_{BS}(\bar{V}_t) h(\bar{V}_t | \sigma_t^2) d\bar{V}_t. \tag{3.11}$$

Equation (3.11) is always true in a risk-neutral world when the stock price and volatility are instantaneously uncorrelated. In addition if the volatility is uncorrelated with aggregate consumption, the equation is true in a risky world as well. Equation (3.11) states that the option price is the Black-Scholes price integrated over the distribution of the mean volatility. It is not possible to obtain an analytic form for the distribution h of $\bar{V}_t$. However, it is possible to calculate all the moments of $\bar{V}_t$ when ϕ and β are constants. Expanding $C_{BS}(\bar{V}_t)$ in a Taylor series about the expected value of $\bar{V}_t$, and using the moments of $\bar{V}_t$, Hull and White managed to derive a series of solutions for the price of a call option in a stock with a stochastic volatility that is uncorrelated with the stock price. It can be found in detail in their paper [63]. In the same paper they compare theoretical option value obtained from their model with the BS option price. They found in the case of uncorrelated volatility with stock price, options overvalues at-the-money $S = K$ and undervalues deep in-the-money $S < K$ and out-the-money $S > K$. In the case of volatility being correlated with the stock price, they examined their result by using numerical methods and found that when there is a positive correlation between the volatility and the stock price, out-of-the-money options are underpriced by the BS model, while in-the-money options are overpriced. This is very important because this means that when the correlation is negative the effect is reversed. These results are important in order that the stochastic volatility can allow excess kurtosis and thus successfully eliminate the volatility smile, which is a very clear phenomenon from empirical data.

Heston's Model

The previous works of Hull and White have generalized the model to allow volatility to be stochastic and uncorrelated with the stock price. However, their models do not have closed-form solution, especially when the volatility is correlated with the stock return and require extensive use of numerical techniques to solve two-dimensional partial differential equation. Heston offered a model of stochastic volatility that is not based on the BS model. His model provides a closed-form solution for the price of European call options when the stock price is correlated with volatility. Besides, his model incorporates stochastic interest rate. Heston's model assumes the stock price and volatility follow the diffusion process

$$dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_1, \tag{3.12a}$$

$$dV_t = \phi(\theta - V_t)dt + \beta\sqrt{V_t} dW_2, \tag{3.12b}$$

again W_1 , W_2 are Wiener process with correlation ρ. θ long term variance rate. ϕ means reversion parameter of the variance rate which denotes the speed at which the variance reverts to its long term average θ. V_t current volatility, β volatility of volatility parameter (ϕ, θ, β are constants). Also, Heston assumed the volatility risk premium $\lambda*$ to be non-zero since there is strong evidence to suggest that $\lambda*$ is non zero, see [68]. In fact, Heston assumed that $\lambda*$ is proportional to the instantaneous variance V_t i.e. $\lambda* = \lambda' V_t$ where λ' is a constant. By applying standard arbitrage arguments which demonstrate that the value of any asset $f_t(S_t, V_t, t)$ must satisfy the partial differential equation

$$\frac{1}{2} V_t S_t^2 \frac{\partial^2 f_t}{\partial S_t^2} + \rho\beta V_t S_t \frac{\partial^2 f_t}{\partial S_t \partial V_t} + \frac{1}{2}\beta^2 V_t \frac{\partial^2 f_t}{\partial V_t^2} + \mu S_t \frac{\partial f_t}{\partial S_t}$$
$$+\Big(\phi[\theta - V_t] - \lambda(S_t, V_t, t)\Big)\frac{\partial f_t}{\partial V_t} - \mu f_t + \frac{\partial f_t}{\partial t} = 0. \tag{3.13}$$

In the same spirit as the BS model, Heston formulated a solution of the form

$$C(S_t, V_t, t) = S_t P_1 - K e^{-\mu(T-t)} P_2, \tag{3.14}$$

where P_1, P_2 are modified cumulative distribution functions for a normal distribution. They also represent the probability that the option expires in the money. The first term in the last equation is the present value of the stock upon optimal exercise, and the second term is the present value of exercise price payment, both of these terms must satisfy (3.13). Heston utilizes the characteristic function of $P_j, j = 1, 2$ denoted by F_j to find a closed-form solution for P_j without making any assumption regarding the correlation ρ. Without going into detail, see [57], we state an explicit expression for P_j as

$$P_j = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\frac{e^{-i\psi lnK} F_j}{i\psi} d\psi, \quad j = 1, 2 \tag{3.15}$$

subject to the terminal condition

$$P_j = 1 \quad \text{for} \quad S \geq K.$$

Equations (3.13),(3.14) and (3.15) give the solution for European call options. The model achieves the imparting of almost any type of bias (skewness, kurtosis) to option prices, and links these biases to the dynamics of stock price and the distribution of stock return. Using the Heston stochastic volatility mode, one observes that the BS model overprices out-of-the-money and underprices in-the-money options when there is a negative correlation between the stock price and the variance process. The opposite is true when the correlation is positive.

3.4.4 Predictable Models

After the limitation of the BS model due to the constant volatility and the failure of the deterministic model to price the option more accurately than BS model, there were different attempts to model volatility such as the stochastic volatility models we discussed in the previous section. These models depend on the behavior of volatility as a random variable and use different diffusion models. On another side there were attempts to predict the apparent change in the volatility and results from a specific type of nonlinear dependence. This method will use the heteroscedasticity models, such as the ARCH and GARCH models first suggested by Engle [38] and Bollerslev [15], [16].

ARCH and GARCH processes

The ARCH(p) process is a discrete time stochastic process with autoregressive conditional heteroscedasticity [38]: ‘a stochastic process with non constant variances conditional on the past, but constant unconditional variances’. It has been used widely in the financial time series, and it is particularly useful for describing the dynamics of volatility fluctuation in a discrete time. ARCH(p) is defined by

$$\sigma_n^2 = a_0 + a_1\xi_{n-1}^2 + ... + a_p\xi_{n-p}^2 = a_0 + \sum_{i=1}^{p} a_i\xi_{n-p}^2, \tag{3.16}$$

where $a_0, a_1, ..., a_p$ are positive parameters such that $\sum_{i=0}^{p} a_i = 1$. ξ_n with zero mean and variance σ_n^2 characterized by a conditional PDF $f_n(\xi)$ which is usually taken to be a Gaussian PDF, but other choices are possible. The memory of the variance can be controlled by varying the number p of terms. Moreover, the stochastic nature of the ARCH(p) process is also changed by changing the form of the conditional PDF $f_n(\xi)$. An ARCH process is completely determined when p parameters and $f_n(\xi)$ are defined.

The simplest ARCH process, when $p = 1$ with Gaussian conditional PDF is denoted by ARCH(1), is given in example 2.5.

A large value of p is required to give a more accurate result, unless this poses a difficulty in calculating the $p + 1$ parameters $a_0, a_1, ..., a_p$. Overcoming this difficulty leads to the generalized ARCH process called GARCH(p,q) process, as follows

$$\sigma_n^2 = a_0 + a_1 \xi_{n-1}^2 + ... + a_p \xi_{n-p}^2 + b_1 \sigma_{n-1}^2 + ... + b_q \sigma_{n-q}^2, \tag{3.17}$$

with

$$\sum_{i=0}^{p} a_i + \sum_{j=0}^{q} b_j = 1,$$

where $a_0, a_1, ..., a_p, b_1, ..., b_q$ are controlled parameters. ξ_n as mentioned before in the ARCH process case.

The simplest GARCH process, namely the GARCH(1,1) process with Gaussian PDF, is defined by

$$\sigma_n^2 = a_0 + a_1 \xi_{n-1}^2 + b_1 \sigma_{n-1}^2. \tag{3.18}$$

Many researchers have used this process to forecast volatility, which can be obtained after estimating the parameters. After that, they use this volatility as an input to the usual Black-Scholes formula. According to Engle [41]: ‘Despite the apparent success of this simple parametrization, the ARCH and GARCH models can not capture some important features of the data. The most interesting feature not addressed by these models is the leverage or asymmetric effect discovered by Black [13]. Statistically, this effect occurs when unexpected drop in price increases predictable volatility more than unexpected increase in price’. Duan [35] devolved a theory with respect to which options can be priced when the evolution of the stock return follows the GARCH process. Unfortunately, analytical solutions for option prices under this theory are not generally available, and hence numerical procedures have to be invoked. Finally, Heston and Nandi [58] have devolved closed-form solutions for European options under a very specific GARCH like volatility updating scheme.

We can conclude that: Heteroscedasticity models explicitly designed to model time varying conditional variance often fail to fully capture the fat tails observed in asset return series. Heteroscedasticity explains some of the fat tail behavior, but typically not all of it [100], [114]. ARCH/GARCH models are able to describe the probability density function of price changes at a given time horizon, but both fail to describe the scaling properties of the probability density function for a short time horizon. The volatility of the market, modeled by GARCH, will die out fast (or the volatility autocorrelation function exhibits exponential decay), which means that the volatility has short memory. Indeed, the short memory is the main reason for the failure of the ARCH and GARCH models since it contradicts the empirical data which has shown that volatility has long memory or slowly decaying autocorrelation function. This reason leads to modification of the GARCH model in order to get a long memory or power-law decay autocorrelation, such as randomizing the time-scale of the continuous time ARCH model. Other approaches by integrated the GARCH model as following.

3.4.5 Long Memory Volatility Models

The topic of long memory and persistence has recently been considered the second moment of the process or volatility of stock price. There is substantial evidence that the long memory processes describe financial data well. The first contribution in this regard was by Taylor (1986) [115], who noticed the apparent fact that the absolute values of stock returns tended to have very slowly decaying autocorrelations. Ding *et al.* (1993) [34] and Anderson (1996) [3] noticed the same fact for the absolute or the powers of daily returns. Some long memory processes can be set up from the same foundation of conditional variance as the ARCH and GARACH models, for example: exponential general autoregressive conditional heteroscedasticity (EGARCH) and fractional integrated general autoregressive conditional heteroscedasticity (FIGARCH), which are introduced by Bollerslev and Mikkelson [17], [18] and Billie [12]. Another long memory process is set up from autoregressive moving average (ARMA) process, such as fractional integrated ARMA process and ARMA process with respect to fractional noise, as considered by Comte [28]. The empirical work of Bollerslev and Mikkelson [19] suggested that when they compared the risk-neutralized option pricing distribution from various ARCH type formulations, the degree of mean reversing in the volatility process implicit is best described by FIGARCH, because the FIGARCH process implies a slow hyperbolic rate of decay for lagged squared innovations. In another view, Hofmann *et al.*[60] considered the diffusion process model for stock price which allows for the description of stochastic and past-dependent volatility. Comte [29] extended Hull and White's option pricing to a continuous time long memory model of stochastic volatility by replacing the Wiener process in (3.10) with a fractional Brownian motion $B_H(t)$; see also [61], [95], [85] [86]. A variety of processes with long memory property are discussed in [53]. Furthermore, the stochastic Black-Scholes model involving volatility with long-range dependence was introduced by Fedotov and Tan [45]. Next, as an example of the long memory process used to model volatility, we give the modification of the GARCH model to get the long memory process FIGARCH.

Long memory GARCH models

Following Engle [38] and Bollerslev [15], a GARCH(p,q) model for the variance is specified as (3.17), or [23];

$$\sigma_t^2 = \omega + a(L)\xi_t^2 + b(L)\sigma_t^2, \tag{3.19}$$

with $\xi_t \mid I^{t-1} \sim N(0, \sigma_t^2)$, and I^{t-1} represents the information set up to time $t-1$, ω is a constant parameter. Moreover, L is the lag operator, such that

$$a(L) = \sum_{i=1}^{p} a_i L^i, \quad b(L) = \sum_{j=1}^{q} b_j L^j.$$

The stationarity of this process is achieved when all the roots of $[1 - a(L) - b(L)]$ are constrained to lie outside the unit circle, i.e.

$a(L) + b(L) < 1$. Defining $v_t = \xi_t^2 - \sigma_t^2$ the process (3.19) may be conveniently rewritten as an ARMA(m,p) process, with $m = max(p, q)$

$$[1 - a(L) - b(L)]\xi_t^2 = \omega + [1 - b(L)]v_t, \tag{3.20}$$

Starting from this formulation and allowing for the presence of a unit root in $[1 - a(L) - b(L)]$, i.e. $a(L) + b(L) = 1$, Engle and Bollerslev [39] defined the IGARCH(p,q) process as

$$\phi(L)(1 - L)\xi_t^2 = \omega + [1 - b(L)]v_t, \tag{3.21}$$

where $\phi(L) = [1 - a(L) - b(L)](1 - L)^{-1}$, and it is of order $m - 1$. Also, the IGARCH model is not able to adequately explain various findings of persistence (or long memory) in the volatility of financial instrument returns. Billie *et al.* [12] extend the IGARCH process, allowing the integration coefficient to vary in the range $[0, 1]$ in order to get a new process called the FIGARCH(p,d,m) process, defined as follows

$$\phi(L)(1 - L)^d\xi_t^2 = \omega + [1 - b(L)]v_t, \tag{3.22}$$

where all the roots of $\phi(L)$ and $[1 - b(L)]$ lie outside the unit circle. Analogously to (3.22) the FIGARCH process can also be represented as

$$[1 - b(L)]\sigma_t^2 = \omega + [1 - b(L) - \phi(L)(1 - L)^d]\xi_t^2. \tag{3.23}$$

The major feature of the FIGARCH model that it has hyperbolic decay, typical of long memory models. This mean that the impact of the innovation lies between the exponential decaying, typical to any GARCH, and infinite persistence, typical of any IGARCH. Davidson [33] gave some insight into the memory properties of the FIGARCH, pointing out the degree of persistence of FIGARCH models: as the d parameter gets closer to zero, the memory of the process increases. This is due to the inverse relation between the integration coefficient and the conditional variance.

3.5 CTRW In Finance

We show in this chapter that the log return has normal distribution (3.2). The normal distribution of the log return was the main concept behind the BS model, deterministic models and stochastic volatility models. However, the empirical returns are non-normally distributed. This has led some researchers to develop a theory of non-normal log return [20]. Other approaches use CTRW to model the log returns instead of normal distribution. The importance of random walk in finance has been known since the seminal work of Bachelier [42], who modeled the price dynamics as an ordinary random walk where price can go up and down owing to a variety of independent random causes. Consequently, the distribution of prices was Gaussian. However, the normal distribution does not fit financial data, especially at the wings of the distribution, and the tails of the price distribution are crucial in the analysis of financial risk. Therefore, obtaining a reliable distribution has significant consequences from a practical point of view. The ideas of Bachelier were further carried out by Mandelbrot [74],[75], who introduced the concepts of L*é*vy flights [4] and stable distribution in finance. The first attempts of Mandelbrot to explain the appearance of heavy tails in financial data based on L*é*vy stable law[5] obtained leptokurtic distribution[6]. Nevertheless, the price to pay is high: the resulting probability density function has infinite moments, except the first one. This is indeed a severe limitation, but still the Mandelbrot approach can be considered within the framework of the Central Limit Theorem, that is, the sum of independent random disturbances of infinite variance resulting in the L*é*vy distribution, which has infinite variance. On the other hand, the L*é*vy distribution has been tested against data in a great variety of situations, with the result always that the tails of the distribution are far too long compared with the actual data. In any case, Mantegna [76] showed that the L*é*vy distribution fits the center of the empirical distribution much better than the Gaussian density. Later on, Masoliver [80] explained speculative price dynamic in two different situations. If finite variance is assumed, the tails are too thin and the resulting

[4]L*é*vy flights is a type of random walk in which the increments are distributed according to a "heavy-tailed" probability distribution. Specifically, the distribution used is a power-law has an infinite variance.

[5]The stable law also called α-stable states: A sum of two independent random variables having an α-stable distribution with index α is again α-stable with the same index $\alpha \in (0, 2]$, which called the tail index or tail exponent or index of stability. the index α describes the distribution, when $\alpha = 2$ the Gaussian distribution result, $\alpha < 2$ the variance is infinite the distribution is a power-law

[6]A description of the kurtosis in a distribution in which the statistical value is positive. Leptokurtic distributions have higher peaks around the mean compared to normal distributions, which leads to thick tails on both sides. These peaks result from the data being highly concentrated around the mean, due to lower variations within observations.

Gaussian distribution only accounts for a narrow neighborhood at the center of the distribution. On the other hand, the assumption of infinite variance leads to the Lévy distribution which explains a wider neighborhood at the center of distribution but results in too fat tails. Although it was well known that the distribution of returns or of logarithmic returns approximately followed a stable law, there was a barrier to the application of these concepts in financial practice due to their infinite variance. Therefore, in mainstream finance, both theoreticians and practitioners prefer to use the more tractable continuous Wiener process instead of discontinuous Lévy flights. A way of overcoming these difficulties has been provided by empirical studies suggesting the use of truncated Lévy flights, characterized by probability density distributions with finite moments [76].
The trials to find a proper distribution for the stock price or log return continue. In financial markets not only prices and returns can be considered random variables, but also the waiting time between two transactions varies randomly.
Scalas *et al.* [107], [109] ,[108], presented CTRW to model high frequency price dynamics in financial markets. In general these dynamics are non-Markovian.

3.5.1 CTRW Models Return

Assuming the zero-mean log return X_t which is given by

$$X_t = ln[S(t)/S(t_0)] - < ln[S(t)/S(t_0)] >, \tag{3.24}$$

where $S(t)$ is the stock price, t_0 is the initial time, and $< ln[S(t)/S(t_0)] >$ is the return mean value. Suppose X_t is described by CTRW, therefore X_t changes at random times $t_0, t_1, ..., t_n, ...$. Also, we assume the waiting times $\tau_n = t_n - t_{n-1}, n = 1, 2, ..., n.$ which are the intervals between the successive changes, are independent identically distributed random variables with probability density function $\phi(t)$ defined by

$$\phi(t)dt = Pr\{t < \tau_n \leq t + dt\}.$$

The jump length can be defined as

$$\triangle X_n = X_{t_n} - X_{t_{n-1}}$$

whose probability density function is given by

$$w(x)dx = Pr\{x < \triangle X_n \leq x + dx\}.$$

The jumps are also assumed to be independent identically distributed random variables. Defining the joint probability function that $\triangle X_n$ is between x and $x + dx$ and the time interval τ_n between successive jumps is between t and $t + dt$, so

$$\Phi(x, t)dxdt = Pr\{x < \triangle X_n \leq x + dx; t < \tau_n \leq t + dt\}.$$

Assuming the function $\Phi(x, t)$ is an even function of x to ensure the absence of drift, i.e. $\Phi(-x, t) = \Phi(x, t)$, that implies $w(-x) = w(x)$. The jumps (zero mean log return) and waiting time may be considered independent random variables, as is the case in the literature [107],[108]; or dependent (coupled) such in the literature [64], [81], [79], [84]. When the jumps and the waiting times are independent the probability of the return's value X_t at time t is given by equation (2.13) or its equivalent master equation (2.20) with different forms regarding the waiting time, as motivated in chapter 2. The solution of the master equation is given by (2.16) in the Fourier-Laplace domain and by (2.23) in the space-time domain.

3.5.2 Return's Volatility

The volatility of zero mean log returns X_t is defined by (3.24) and follows the CTRW process, in the case of the even jump process, it can be considered the second moment of the process [81]. From (2.32), the second moment of the CTRW process in Laplace transform is given by

$$\begin{aligned} < \tilde{X}^2(s) > &= (i)^{-2}\frac{\partial^2 \tilde{\hat{p}}(k, s)}{\partial k^2} |_{k=0} \\ &= \frac{- < x^2 > \tilde{\phi}(s)}{s[1 - \tilde{\phi}(s)]}, \end{aligned} \tag{3.25}$$

where $\tilde{\hat{p}}(k, s)$ is the PDF of the CTRW process in the Fourier-Laplace transform, $\tilde{\phi}(s)$ is the PDF of the waiting times in the Laplace transform and $< x^2 >$ is the second moment of the jump process. Consequently, we convert equation (3.25) to the time domain form to get the volatility of zero mean log return X_t, recalling the variance of CTRW given by (2.33) when waiting time is exponential distribution. Therefore, the variance is linear function of time i.e., the variance or the volatility in this case has no memory. The variance given by (2.34) has short memory expressed by the exponential decay in the case of gamma distribution waiting time. Finally, the variance given by (2.35) when waiting time PDF has power-law form, appears also as a power-law decay of time, which is evidence of long memory effect. Kim *et al.* [67] used the theory of CTRW to analyze numerically the dynamical behavior of the PDF and the volatility from the data of some Korean government bond. Numerically they found that the waiting time density for the bond follows a power-law $\phi(t) \sim t^{-(\beta+1)}$ with scaling exponent $\beta = 0.51$. Also the jump probability density is given by a power-law $w(x) = x^{-\alpha}$ with $\alpha = 2.94$. The volatility is proportional anomalously to a power-law $vol = t^{-k}$with scale $k = 0.96$. Yoon *et al.*[121] applied again the CTRW theory on the yen-dollar exchange rate with power-law waiting time and power-law jump process, and found the same result regarding volatility with different scaling parameters.

Chapter 4

Generalized CTRW Model

4.1 Introduction

In chapter 2 we introduced the concept of continuous time random walk (CTRW) and its properties. Furthermore, we discussed the memory effects and how they can vary regarding the waiting time distribution. Moreover, we give examples of the probability distribution of the CTRW process regarding the first moment of waiting time and second moment of jump distribution. The use of the CTRW in finance is discussed in chapter 3 in addition to usage of the process to model the stock's return, then the stock's return volatility can be found as the second moment of the process. Furthermore, we infer from the numerical work of previous studies that volatility has long memory. In our study we aim to use the CTRW itself to model the volatility. We also need to find the master equations of conditional transition probability and the probability density of the process corresponding to different waiting time distributions and jump distributions. In order to find a method to derive the master equations, we introduce our theory of conditional arrival probability, and we derive the master equations in the Laplace domain. First, we develop our theory of conditional arrival probability in the discrete state space with continuous waiting time (discrete random walk), then generalize it to the continuous state space with continuous waiting time (continuous random walk). Furthermore, we apply our result when the process is Markovian as a result of exponential waiting time distribution and non-Markovian as a result of power-law waiting time. Also we calculate the first two moments for each case.

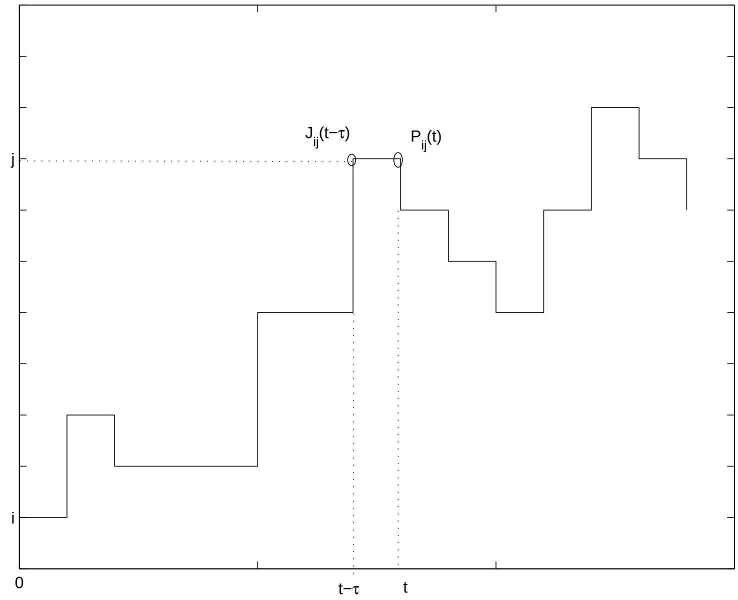

Figure 4.1: Conditional arrival probability $J_{ij}(t-\tau)$ from state i to state j, and conditional transition probability $P_{ij}(t)$.

We conclude that the moments are proportional to time for the Markovian process and non-linear function of time for the non-Markovian process. Moreover, we illustrate some numerical results for the counting process with different waiting time distributions and the switching process when it is Markovian and non-Markovian.

4.2 Conditional Arrival Probability

4.2.1 Discrete Random Walk

In order to find the master equations of the probability density for continuous time random walk with discrete state space, we introduce the theory of conditional arrival probability. First we assume the jump process X_t takes N states. We define the conditional transition probability as follows

$$P_{ij}(t) = Pr\{X_t = j \mid X_0 = i\}, \quad i, j = 1, 2, ..., N$$

which is the probability that the process X_t starts from state i at $t = 0$ and it is at state j at time t. We introduce the conditional arrival probability $J_{ij}(t)$ as the probability that the process X_t starts from state i at time $t = 0$ and arrives at state j at time t; see figure 4.1. Also, we introduce the probability density $n_j(t)$ such that

$$n_j(t) = Pr\{X_t = j\},$$

which is the probability to be at state j at time t. Consider $\phi_j(t)$ the probability density function for the waiting time at state j. The survival function $\Psi_j(t)$ is given by

$$\Psi_j(t) = \int_t^{\infty} \phi_j(t')dt' = 1 - \int_0^t \phi_j(t')dt', \tag{4.1}$$

which is the probability that no steps are taken in the time interval $[0, t)$. Moreover, the $n \times n$ transition matrix H with the matrix entries h_{ij} denotes the transition rate from state i to state j and satisfies[1]

$$\sum_{j=1}^{N} h_{ij} = 1. \tag{4.2}$$

If $h_{ij} = 1$ only when $j = i + 1$ and zero otherwise, then the discrete random walk process represents a counting process. The balance equation for $J_{ij}(t)$ is

$$J_{ij}(t) = \sum_{k \neq j}^{N} \int_0^t J_{ik}(t - \tau)\phi_k(\tau)h_{kj}d\tau + \phi_i(t)h_{ij}. \tag{4.3}$$

It follows from the law of total probabilities that $P_{ij}(t)$ obeys the equation

$$P_{ij}(t) = P_{ij}(0)\Psi_j(t) + \int_0^t J_{ij}(t - \tau)\Psi_j(\tau)d\tau, \tag{4.4}$$

where $P_{ij}(0)$ is the initial condition satisfies

$$P_{ij}(0) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{4.5}$$

The first term in the right-hand side (RHS) of equation (4.4) represents the probability of being at the initial state times the probability of no jump up to time t. The second term takes into account the probability of arriving at state j from state i at time $t - \tau$ and the probability of no jump during time τ. We assume that the jump process X_t is homogenous, thus

$$P_{ij}(t) = Pr\{X_{t+h} = j \mid X_t = i\} = Pr\{X_h = j \mid X_0 = i\},$$

[1] The transition rates must be given constants for each transition.

whereas

$$n_j(t) = \sum_{i=1}^{N} P_{ij}(t)n_i(0) = \sum_{i=1}^{N} P_{ij}(0)\Psi_j(t)n_i(0) + \sum_{i=1}^{N} \int_0^t J_{ij}(t-\tau)\Psi_j(\tau)n_i(0)d\tau,$$

using the definition of the initial condition (4.5), hence

$$n_j(t) = \Psi_j(t)n_j(0) + \sum_{i=1}^{N} \int_0^t J_{ij}(t-\tau)\Psi_j(\tau)n_i(0)d\tau. \tag{4.6}$$

In order to find the master equation of the conditional transition probability and the probability density for the process, we use the Laplace transform of equations (4.3), (4.4), (4.6)

$$\begin{aligned} \tilde{n}_j(s) &= \sum_{i=1}^{N} \tilde{P}_{ij}(s)n_i(0), && (4.7)\\ \tilde{J}_{ij}(s) &= \sum_{k\neq j}^{N} \tilde{J}_{ik}(s)\tilde{\phi}_k(s)h_{kj} + \tilde{\phi}_i(s)h_{ij}, && (4.8)\\ \tilde{P}_{ij}(s) &= P_{ij}(0)\tilde{\Psi}_j(s) + \tilde{J}_{ij}(s)\tilde{\Psi}_j(s). && (4.9) \end{aligned}$$

From (4.9) we obtain

$$\tilde{J}_{ij}(s) = \frac{\tilde{P}_{ij}(s)}{\tilde{\Psi}_j(s)} - P_{ij}(0)\frac{\tilde{\Psi}_j(s)}{\tilde{\Psi}_j(s)}, \tag{4.10}$$

substitution of (4.10) into (4.8) gives

$$\frac{\tilde{P}_{ij}(s)}{\tilde{\Psi}_j(s)} - P_{ij}(0) = \sum_{k\neq j}^{N} \left[\frac{\tilde{P}_{ik}(s)}{\tilde{\Psi}_k(s)} - \frac{P_{ik}(0)\tilde{\Psi}_i(s)}{\tilde{\Psi}_k(s)}\right]\tilde{\phi}_k(s)h_{kj} + \tilde{\phi}_i(s)h_{ij}. \tag{4.11}$$

By applying the definition of the initial condition (4.5), the last equation will be reduced to

$$\frac{\tilde{P}_{ij}(s)}{\tilde{\Psi}_j(s)} - P_{ij}(0) = \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\frac{\tilde{\phi}_k(s)}{\tilde{\Psi}_k(s)}h_{kj}. \tag{4.12}$$

To get the conditional transition probability, we rearrange equation (4.12)

$$\tilde{P}_{ij}(s) = P_{ij}(0)\tilde{\Psi}_j(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\frac{\tilde{\phi}_k(s)}{\tilde{\Psi}_k(s)}h_{kj}\tilde{\Psi}_j(s). \tag{4.13}$$

By using the definition of kernel function $\tilde{K}_k(s) = \frac{\tilde{\phi}_k(s)}{\tilde{\Psi}_k(s)}$, we obtain

$$\tilde{P}_{ij}(s) = P_{ij}(0)\tilde{\Psi}_j(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\tilde{K}_k(s)h_{kj}\tilde{\Psi}_j(s). \tag{4.14}$$

If the waiting time is independent identically distributed (iid) random variables for all states then (4.13) becomes

$$\tilde{P}_{ij}(s) = P_{ij}(0)\tilde{\Psi}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\tilde{\phi}(s)h_{kj}. \tag{4.15}$$

By inverting Laplace transform, it yields

$$P_{ij}(t) = P_{ij}(0)\Psi(t) + \sum_{k\neq j}^{N} \int_0^t P_{ik}(t-\tau)\phi(\tau)h_{kj}d\tau. \tag{4.16}$$

The first term in RHS represents the process starting from the origin and staying there until time t; the second term includes the contribution from the jump to state j from different states k and waiting up to time t.

To find the Laplace form of the probability density $\tilde{n}_j(s)$, we multiply equation (4.13) by $n_i(0)$ and take the summation. Firstly, in the case of varying waiting time with the states

$$\tilde{n}_j(s) = \sum_{i=1}^{N} \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\tilde{K}_k(s)h_{kj}\tilde{\Psi}_j(s)n_i(0) + n_j(0)\tilde{\Psi}_j(s),$$

by using (4.7), this implies

$$\tilde{n}_j(s) = \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj}\tilde{\Psi}_j(s) + n_j(0)\tilde{\Psi}_j(s).$$

Secondly, in the case when waiting time is the same at all states (invariant)

$$\tilde{n}_j(s) = \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{\phi}(s)h_{kj} + n_j(0)\tilde{\Psi}(s), \tag{4.17}$$

converting it to time domain, it will be

$$n_j(t) = \sum_{k\neq j}^{N} \int_0^t n_k(t-\tau)\phi(\tau)h_{kj}d\tau + n_j(0)\Psi(t). \tag{4.18}$$

To get the master equation of the conditional transition probability $P_{ij}(t)$ for discrete random walk, we move the first term in the left-hand side in equation (4.12) to the right and add $sP_{ij}(s)$ to both sides, thus

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = s\tilde{P}_{ij}(s) - \frac{\tilde{P}_{ij}(s)}{\tilde{\Psi}_j(s)} + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\frac{\tilde{\phi}_k(s)}{\tilde{\Psi}_k(s)}h_{kj}.$$

Using the definition of the survival function in Laplace domain, $\tilde{\Psi}(s) = \frac{1-\tilde{\phi}(s)}{s}$, we get

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = -\frac{\tilde{\phi}_j(s)}{\tilde{\Psi}_j(s)}\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\frac{\tilde{\phi}_k(s)}{\tilde{\Psi}_k(s)}h_{kj}.$$

Or by using the definition of kernel function, imply

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = -\tilde{K}_j(s)\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\tilde{K}_k(s)h_{kj}. \tag{4.19}$$

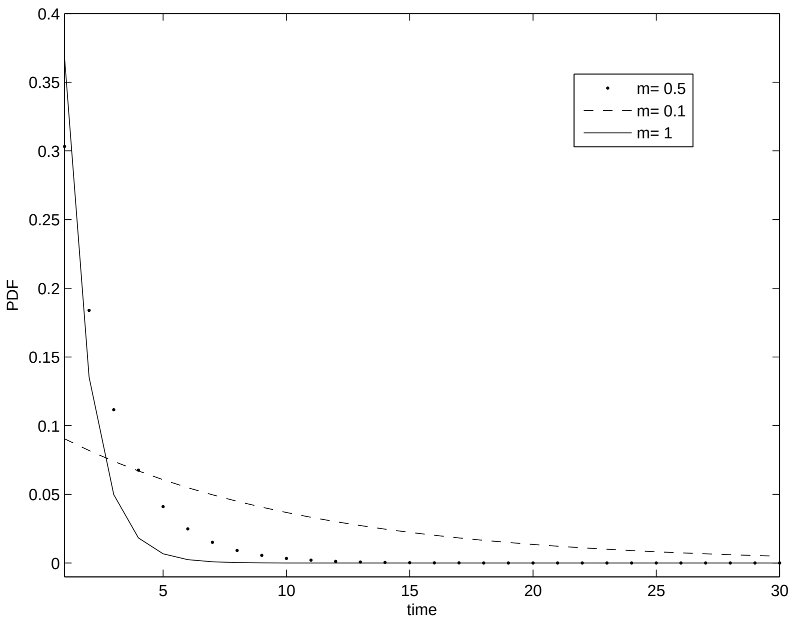

Figure 4.2: The probability density of exponential distribution with various rate parameter m.

Inverting Laplace transform with convolution theorem gives the master equation in time domain

$$\frac{dP_{ij}(t)}{dt} = -\int_0^t K_j(\tau)P_{ij}(t-\tau)d\tau + \sum_{k\neq j}^{N}\int_0^t K_k(\tau)P_{ik}(t-\tau)h_{kj}d\tau. \tag{4.20}$$

Next we are going to find the master equations for the conditional transition probability and the probability density corresponding to different waiting time distributions in the Laplace domain by using the previous equations (4.15), (4.17) and (4.19).

Exponential distribution waiting time:

Let the waiting time t follow an exponential distribution, i.e.

$$\phi_j(t) = m_j e^{-m_j t}, \quad t \geq 0, \quad j = 1, 2, ..., N \tag{4.21}$$

where m_j is a positive constant called the rate parameter. The expected value of the distribution is $1/m_j$ and the variance is $1/m_j^2$. From figure 4.2, we can see some features of an exponential distribution: the probability density declines monotonically as the value of waiting time increases, and the curve is steeper as the parameter m is larger. Therefore in this particular case, the waiting time is more likely to be very small and long waiting time seldom happens.

The Laplace transform of the distribution's PDF is

$$\tilde{\phi}_j(s) = \frac{m_j}{m_j + s}, \tag{4.22}$$

and for survival function

$$\tilde{\Psi}_j(s) = \frac{1}{m_j + s}. \tag{4.23}$$

Therefore, the kernel function will be

$$\tilde{K}_j(s) = m_j.$$

Substituting the value of kernel function in (4.19) gives the master equation in Laplace domain

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = -m_j\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} m_k\tilde{P}_{ik}(s)h_{kj}, \tag{4.24}$$

it follows by inverting to time domain

$$\frac{dP_{ij}(t)}{dt} = -m_jP_{ij}(t) + \sum_{k\neq j}^{N} m_kP_{ik}(t)h_{kj}.$$

If the waiting time is the same at all states, then the conditional transition probability (4.16) is

$$P_{ij}(t) = P_{ij}(0)e^{-mt} + \sum_{k\neq j}^{N} m\int_0^t e^{-m\tau}P_{ik}(t-\tau)h_{kj}d\tau.$$

Consequently, the probability density (4.18) is

$$n_j(t) = n_j(0)e^{-mt} + \sum_{k\neq j}^{N} m\int_0^t e^{-m\tau}n_k(t-\tau)h_{kj}d\tau,$$

while the master equation is a classic forward Kolmogorov equation

$$\frac{dP_{ij}(t)}{dt} = -mP_{ij}(t) + m\sum_{k\neq j}^{N} P_{ik}(t)h_{kj}. \tag{4.25}$$

The process in this case is Markovian because there is no memory effect. This means the current state of the process at time t does not depend on any previous states at the previous history.

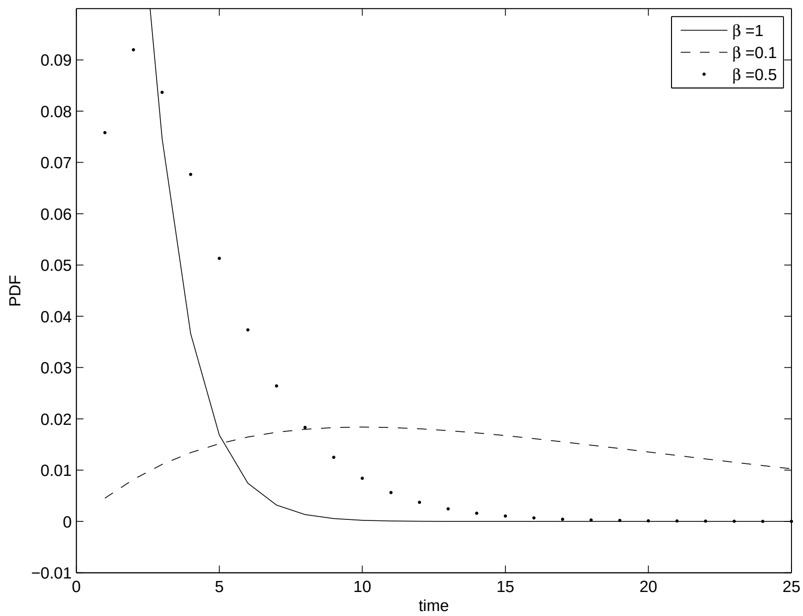

Figure 4.3: The probability density of gamma distribution with shape parameter $\alpha = 2$ and various scale parameter β.

Gamma distribution waiting time:

If the waiting time has gamma distribution then

$$\phi_j(t) = \frac{\beta_j^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta_j t}, \quad t > 0, j = 1, 2, ..., N \tag{4.26}$$

where α is the shape parameter and β is the scale parameter and both β, α are positive. The expected value of the distribution is $\alpha\beta$ and the variance is $\alpha\beta^2$. In addition the gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt,$$

and when α is a positive integer

$$\Gamma(\alpha) = (\alpha - 1)!.$$

Apparently, when $\alpha = 1$ the equation (4.26) reduces to the form of exponential distribution, hence exponential distribution is a special case of gamma distribution. It is clear from figure 4.3 that the curve of gamma distribution has only one peak and the peak moves as the parameters α, β vary. Thus gamma distribution can allow for longer waiting times if the proper parameters are chosen.

In order to derive the master equation for the general case of gamma distribution, firstly we consider a simple case when $\alpha = 2$ so the waiting time PDF is

$$\phi_j(t) = \beta_j^2 t e^{-\beta_j t}.$$

Using the Laplace transform for the PDF, we obtain the following functions

$$\tilde{\phi}_j(s) = \frac{\beta_j^2}{(s+\beta_j)^2}, \tag{4.27}$$

$$\tilde{\Psi}_j(s) = \frac{s+2\beta_j}{(s+\beta_j)^2}, \tag{4.28}$$

$$\tilde{K}_j(s) = \frac{\beta_j^2}{s+2\beta_j}.$$

Inserting them in (4.19) yields the master equation in the Laplace domain

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = -\frac{\beta_j^2}{s+2\beta_j}\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \frac{\beta_k^2}{s+2\beta_k}\tilde{P}_{ik}(s)h_{kj}. \tag{4.29}$$

Hence, in the time domain it will be

$$\frac{dP_{ij}(t)}{dt} = -\beta_j^2 \int_0^t e^{-2\beta_j\tau} P_{ij}(t-\tau)d\tau + \sum_{k\neq j}^{N} \beta_k^2 \int_0^t e^{-2\beta_k\tau} P_{ik}(t-\tau)h_{kj}d\tau.$$

If the waiting time is invariant, then from (4.16), (4.18), we have

$$P_{ij}(t) = \sum_{k\neq j}^{N} \beta^2 \int_0^t \tau e^{-\beta\tau} P_{ik}(t-\tau)h_{kj}d\tau + P_{ij}(0)e^{-\beta t}(\beta t+1),$$

$$n_j(t) = \sum_{k\neq j}^{N} \beta^2 \int_0^t \tau e^{-\beta\tau} n_k(t-\tau)h_{kj}d\tau + n_j(0)e^{-\beta t}(\beta t+1).$$

Accordingly, the master equation is

$$\frac{dP_{ij}(t)}{dt} = -\beta^2 \int_0^t e^{-2\beta\tau} P_{ij}(t-\tau)d\tau + \beta^2 \sum_{k\neq j}^{N} \int_0^t e^{-2\beta\tau} P_{ik}(t-\tau)h_{kj}d\tau. \tag{4.30}$$

Here we can notice the time integral in the right-hand side, which is evidence of memory effects. So the master equation in the case of gamma distributed waiting time is non-Markovian.

We are now in a position for general α, and when the waiting time is invariant. The Laplace form of equation (4.26) is given by

$$\tilde{\phi}(s) = \frac{1}{(\frac{s}{\beta}+1)^{\alpha}},$$

or by taking $\frac{1}{\beta} = \theta$:

$$\tilde{\phi}(s) = \frac{1}{(s\theta + 1)^{\alpha}} \tag{4.31}$$

So, the survival function is

$$\tilde{\Psi}(s) = \frac{(s\theta + 1)^{\alpha} - 1}{s(s\theta + 1)^{\alpha}}, \tag{4.32}$$

which implies,

$$\tilde{K}(s) = \frac{s}{(s\theta + 1)^{\alpha} - 1}.$$

First, we are going to give the master equations for the probability distribution in a Laplace form

$$\begin{aligned}
\tilde{P}_{ij}(s) &= \sum_{k\neq j}^{N} \frac{1}{(s\theta + 1)^{\alpha}} \tilde{P}_{ik}(s) h_{kj} + P_{ij}(0) \frac{(s\theta + 1)^{\alpha} - 1}{s(s\theta + 1)^{\alpha}}, \\
\tilde{n}_j(s) &= \sum_{k\neq j}^{N} \frac{1}{(s\theta + 1)^{\alpha}} \tilde{n}_k(s) h_{kj} + n_j(0) \frac{(s\theta + 1)^{\alpha} - 1}{s(s\theta + 1)^{\alpha}}, \\
s\tilde{P}_{ij}(s) - P_{ij}(0) &= -\frac{s}{(s\theta + 1)^{\alpha} - 1} \tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s) \frac{s}{(s\theta + 1)^{\alpha} - 1} h_{kj}.
\end{aligned}$$

To find the master equation in the time domain, it should be emphasized that it is impossible to find an explicit expression for the kernel function $K(t)$ for arbitrary choice of the waiting time PDF $\phi(t)$, such as this case. So, we define the memory function $\tilde{H}(s) = \frac{1}{\tilde{K}(s)}$, which for gamma waiting time PDF will be equal to

$$\tilde{H}(s) = \frac{(s\theta + 1)^{\alpha} - 1}{s} = \sum_{i=1}^{\alpha} \binom{\alpha}{i} \theta^i s^{i-1},$$

If we multiply equation (4.19), in the case of invariant kernel function, by the memory function $\tilde{H}(s)$, we get

$$\begin{aligned}
\tilde{H}(s)[s\tilde{P}_{ij}(s) - P_{ij}(0)] &= -\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s) h_{kj}, \\
\sum_{i=1}^{\alpha} \binom{\alpha}{i} \theta^i s^{i-1} [s\tilde{P}_{ij}(s) - P_{ij}(0)] &= -\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s) h_{kj}.
\end{aligned}$$

Therefore, by using the general differentiation property of the Laplace transform[2], the master equation for general gamma distribution in the time domain is

$$\sum_{i=1}^{\alpha} \binom{\alpha}{i} \theta^i \frac{d^i}{dt^i} P_{ij}(t) = -P_{ij}(t) + \sum_{k\neq j}^{N} P_{kj}(t) h_{kj}. \tag{4.33}$$

[2]The general differentiation property of Laplace transform is given by $f^{(n)}(t) = s^n \tilde{f}(s) - s^{n-1} f(0) - s^{n-2} f'(0) - ... - s f^{n-2}(0) - f^{(n-1)}(0)$

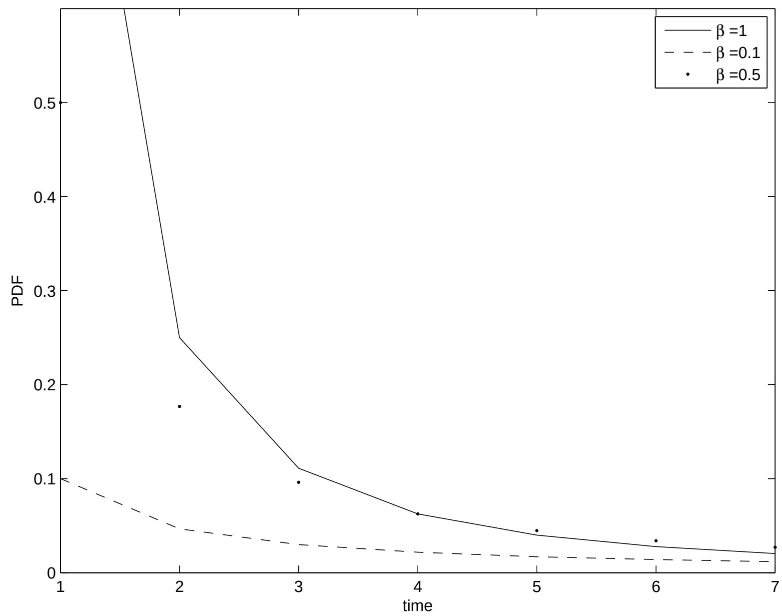

Figure 4.4: The probability density of Pareto distribution with various parameter β and minimum value equals one.

The master equation here is non-Markovian. However, it is different from equation (4.30).

Power-law distribution waiting time:

In the case when the waiting time has a heavy-tailed or power-law distribution, its PDF can be in the following form

$$\phi_j(t) \sim \frac{\beta_j}{\Gamma(1-\beta_j)} t^{-(\beta_j+1)}, \quad 0 < \beta_j < 1, \text{for} \quad t \to \infty. \tag{4.34}$$

Power-law distribution issued in the description of open system [77]. Power-law correlation is observed in a critical state of an infinite system, but if the system is finite, the finiteness limits the range within which the power-law behavior can be observed. One form of power-law distribution is called **Pareto distribution**, which has the following PDF

$$\phi(t) = \begin{cases} \frac{\beta b^\beta}{t^{\beta+1}} & \text{for} \quad t > b, \\ 0 & \text{for} \quad t < b, \end{cases} \tag{4.35}$$

where b is the minimum possible value of t and β is a positive parameter. Figure 4.4 shows Pareto PDF when the minimum value is one and β has various values. This distribution has expected value when $\beta > 1$ equal to $\beta b/(\beta - 1)$ or $\beta/(\beta - 1)$ for $b = 1$. Also it has variance equal to $(b/(\beta - 1))^2(\beta/(\beta - 2))$ that exists only for $\beta > 2$. Later on, we are going to use this distribution in our numerical application in order to compare it with another distribution that has finite expected value. Another example of power-law distribution is the Mittag-Leffler function that is given in section (2.5.3)

$$\phi_j(t) = -\frac{d}{dt}E_{\beta_j}(-t^{\beta_j}), \quad \Psi_j(t) = E_{\beta_j}(-t^{\beta_j}).$$

This function has a Laplace form such as

$$\tilde{\phi}_j(s) = \frac{1}{s^{\beta_j} + 1}, \tag{4.36}$$

therefore the survival function will be

$$\tilde{\Psi}_j(s) = \frac{s^{\beta_j - 1}}{1 + s^{\beta_j}}, \tag{4.37}$$

and the kernel function is

$$\tilde{K}_j(s) = \frac{1}{s^{\beta_j - 1}}.$$

By substituting the kernel function in (4.19), we obtain

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = -\frac{1}{s^{\beta_j - 1}}\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \frac{1}{s^{\beta_k - 1}}\tilde{P}_{ik}(s)h_{kj},$$

and multiplying the last equation by $s^{\beta_j - 1}$, implies

$$s^{\beta_j}\tilde{P}_{ij}(s) - s^{\beta_j - 1}P_{ij}(0) = -\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} s^{\beta_j - \beta_k}\tilde{P}_{ik}(s)h_{kj},$$

hence, by using the general differentiation property of the Laplace transform, the master equation takes the following form

$$\frac{d^{\beta_j}P_{ij}(t)}{dt^{\beta_j}} = -P_{ij}(t) + \sum_{k\neq j}^{N} \int_0^t \frac{(\tau)^{\beta_k - \beta_j - 1}}{\Gamma(\beta_k - \beta_j)} P_{ik}(t-\tau)h_{kj}d\tau.$$

Clearly, in the last equation the memory effect seems to be in both sides. In the right-hand side it happens because of the time integral, whereas in the-left hand side it happens due to the Caputo fractional derivative $\frac{d^{\beta}}{dt^{\beta}}$ of order β. This derivative is defined by [102]

$$\frac{d^{\beta}}{dt^{\beta}}P(t) = \frac{1}{\Gamma(1-\beta)}\frac{d}{dt}\int_0^t P(\tau)(t-\tau)^{-\beta}d\tau - \frac{t^{-\beta}}{\Gamma(1-\beta)}P(0). \tag{4.38}$$

So the master equation is non-Markovian. As with previous distributions, we are giving the master equations of the conditional transition probability and the probability density in the Laplace domain when the waiting time is invariant

$$\begin{aligned} \tilde{P}_{ij}(s) &= \sum_{k\neq j}^{N} \frac{1}{s^{\beta}+1}\tilde{P}_{ik}(s)h_{kj} + P_{ij}(0)\frac{s^{\beta-1}}{1+s^{\beta}}, \\ \tilde{n}_{j}(s) &= \sum_{k\neq j}^{N} \frac{1}{s^{\beta}+1}\tilde{n}_{k}(s)h_{kj} + n_{j}(0)\frac{s^{\beta-1}}{1+s^{\beta}}, \end{aligned}$$

Similarly, the master equation is

$$s^{\beta}\tilde{P}_{ij}(s) - s^{\beta-1}P_{ij}(0) = -\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)h_{kj}.$$

So, in the time domain we obtain

$$P_{ij}(t) = \sum_{k\neq j}^{N} -\int_0^t \frac{d}{d\tau}E_{\beta}(-\tau^{\beta})P_{ik}(t-\tau)h_{kj}d\tau + P_{ij}(0)E_{\beta}(-t^{\beta}), \tag{4.39}$$

$$\frac{d^{\beta}P_{ij}(t)}{dt^{\beta}} = -P_{ij}(t) + \sum_{k\neq j}^{N} P_{ik}(t)h_{kj}, \tag{4.40}$$

The memory effect appears in equation (4.39) due to the Mittag-Leffler function and in equation (4.40) due to the Caputo fractional derivative in the left-hand side.

Combined distribution waiting time

So far, we have discussed cases in which the waiting time follows exponential distribution, gamma distribution and power-law distribution. All of these distributions allow the probability density of waiting time to have only one peak. It means that the waiting time has a certain value t', for example, before and after which the probability has a monotonic and smooth increase and descent respectively. However, if the probability density of waiting time reveals some fluctuations (multiple peaks at $t'_1, t'_2, t'_3, ...$), the probability density curve may not be smooth enough (i.e. some undifferentiable points exist). For those reasons, we may combine several distributions to adjust the probability density of waiting time so as to get many peaks. In this section we assume waiting time follows a combined distribution of exponential distribution with parameter m and gamma distribution with parameter $\alpha = 2$ and β; see figure 4.5. Thus, the probability density for combined distribution waiting time

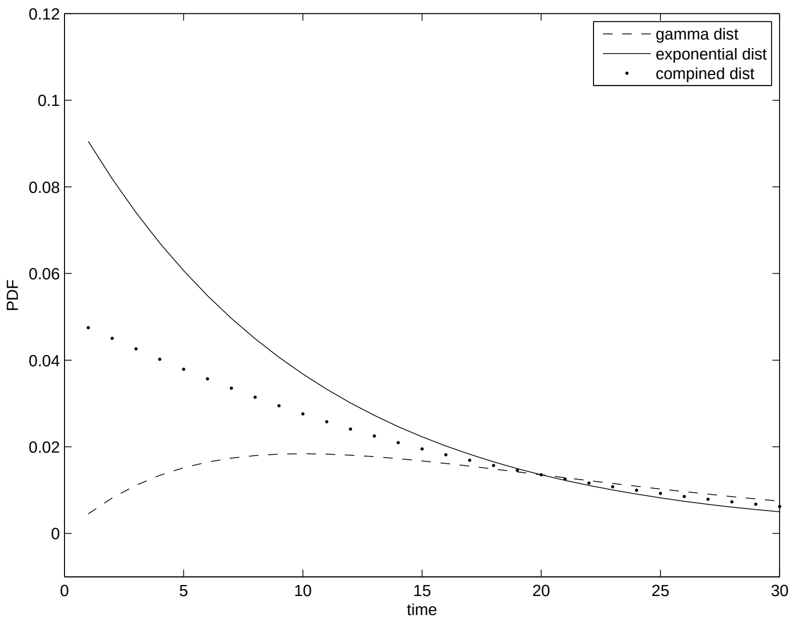

Figure 4.5: The probability density of combined distribution of exponential distribution with $m = 0.1$, gamma distribution with $\alpha = 2, \beta = 0.1$ and $q = 0.5$

that is invariant at all states will be

$$\phi(t) = qme^{-mt} + (1-q)\beta^2 te^{-\beta t}, \quad 0 < q < 1. \tag{4.41}$$

This can be explained as the waiting time has probability q to follow an exponential distribution and has a probability $1-q$ to follow gamma distribution.

In order to derive the master equations, we need the Laplace transform for (4.41), that is

$$\tilde{\phi}(s) = \frac{qm}{s+m} + \frac{(1-q)\beta^2}{(s+\beta)^2},$$

and the survival function of the distribution

$$\begin{aligned}\tilde{\Psi}(s) &= \frac{s^2 + 2s\beta + ms + 2\beta m - mqs - 2mq\beta + q\beta^2}{(s+m)(s+\beta)^2}, \\ or & \\ &= \frac{q}{s+m} + \frac{1-q}{s+\beta} + \frac{\beta(1-q)}{(s+\beta)^2}.\end{aligned}$$

After some calculations, the kernel function can be written as

$$\tilde{K}(s) = \frac{qm(s+\beta)^2 + (1-q)\beta^2(s+m)}{q(s+\beta)^2 + (1-q)(s+m)(s+2\beta)}.$$

By inserting the waiting and the survival functions in (4.15) and (4.17), we get

$$\begin{aligned}\tilde{P}_{ij}(s) &= \sum_{k\neq j}^{N}\tilde{P}_{ik}(s)\left(\frac{qm}{s+m}+\frac{(1-q)\beta^2}{(s+\beta)^2}\right)h_{kj}+P_{ij}(0)\left(\frac{q}{s+m}+\frac{1-q}{s+\beta}+\frac{\beta(1-q)}{(s+\beta)^2}\right)\\ \tilde{n}_j(s) &= \sum_{k\neq j}^{N}\tilde{n}_k(s)\left(\frac{qm}{s+m}+\frac{(1-q)\beta^2}{(s+\beta)^2}\right)h_{kj}+n_j(0)\left(\frac{q}{s+m}+\frac{1-q}{s+\beta}+\frac{\beta(1-q)}{(s+\beta)^2}\right),\end{aligned}$$

The master equations of the conditional transition probability and the probability density in time domain are

$$\begin{aligned}P_{ij}(t) &= \sum_{k\neq j}^{N}\int_0^t\left(qme^{-m\tau}+(1-q)\beta^2\tau e^{-\beta\tau}\right)P_{ij}(t-\tau)d\tau\\ &\quad +P_{ij}(0)\left(qe^{-mt}+(1-q)e^{-\beta t}+(1-q)\beta te^{-\beta t}\right).\\ n_j(t) &= \sum_{k\neq j}^{N}\int_0^t\left(qme^{-m\tau}+(1-q)\beta^2\tau e^{-\beta\tau}\right)n_i(t-\tau)h_{kj}d\tau\\ &\quad +n_j(0)\left(qe^{-mt}+(1-q)e^{-\beta t}+\beta(1-q)te^{-\beta t}\right).\end{aligned}$$

As is clear from the last equations, the memory effects due to the time integral and the time functions appear at each term of the RHS of the equations.

Similarly, inserting the kernel function in (4.19) gets the master equation

$$\begin{aligned}s\tilde{P}_{ij}(s)-P_{ij}(0) &= -\Big(\frac{qm(s+\beta)^2+(1-q)\beta^2(s+m)}{q(s+\beta)^2+(1-q)(s+m)(s+2\beta)}\Big)\tilde{P}_{ij}(s)\\ &\quad +\sum_{k\neq j}^{N}\tilde{P}_{ik}(s)\Big(\frac{qm(s+\beta)^2+(1-q)\beta^2(s+m)}{q(s+\beta)^2+(1-q)(s+m)(s+2\beta)}\Big)h_{kj}. \qquad (4.42)\end{aligned}$$

We may notice from the last equation, if $q = 0$, the distribution takes the gamma distribution form and the master equation will reduce to (4.29). Also if $q = 1$, the distribution will be exponential and the master equation takes the form (4.24). Going back to the master equation of combined distribution waiting time (4.42), and rearranging it

$$\begin{aligned}&\Big(q(s+\beta)^2+(1-q)(s+m)(s+2\beta)\Big)\Big(s\tilde{P}_{ij}(s)-P_{ij}(0)\Big)=\\ &\quad -\Big(qm(s+\beta)^2+(1-q)\beta^2(s+m)\Big)\tilde{P}_{ij}(s)+\sum_{k\neq j}^{N}\tilde{P}_{ik}(s)\Big(qm(s+\beta)^2+(1-q)\beta^2(s+m)\Big)h_{kj}.\end{aligned}$$

After some calculations, it may be written as

$$\begin{gathered}\Big(s^3\tilde{P}_{ij}(s)-s^2P_{ij}(0)\Big)+\Big(2\beta+m-qm\Big)\Big(s^2\tilde{P}_{ij}(s)-sP_{ij}(0)\Big)+\\ \Big(q\beta^2+2m\beta-2m\beta q\Big)\Big(s\tilde{P}_{ij}(s)-P_{ij}(0)\Big)=\\ \Big(s^2qm+s[\beta^2-\beta^2q+2qm\beta]+\beta^2m\Big)\Big(\textstyle\sum_{k\neq j}^{N}\tilde{P}_{ik}(s)h_{kj}-\tilde{P}_{ij}(s)\Big),\end{gathered}$$

or

$$\Big(s^3\tilde{P}_{ij}(s) - s^2P_{ij}(0)\Big) + \Big(2\beta + m - mq)\Big)\Big(s^2\tilde{P}_{ij}(s) - sP_{ij}(0)\Big)$$
$$+ \Big(q\beta^2 + 2m\beta - 2qm\beta\Big)\Big(s\tilde{P}_{ij}(s) - P_{ij}(0)\Big) =$$
$$qm\sum_{k\neq j}^{N}\Big(s^2\tilde{P}_{ik}(s) - sP_{ik}(0)\Big)h_{kj} - qm\Big(s^2\tilde{P}_{ij}(s) - sP_{ij}(0)\Big)$$
$$+ \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)\sum_{k\neq j}^{N}\Big(s\tilde{P}_{ik}(s) - P_{ik}(0)\Big)h_{kj} - \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)\Big(s\tilde{P}_{ij}(s) - P_{ij}(0)\Big)$$
$$+ \beta^2 m\Big(\sum_{k\neq j}^{N}\tilde{P}_{ik}(s)h_{kj} - \tilde{P}_{ij}(s)\Big)$$
$$+ qm\sum_{k\neq j}^{N} sP_{ik}(0)h_{kj} - qmsP_{ij}(0) + \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)\sum_{k\neq j}^{N}P_{ik}(0)h_{kj} - \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)P_{ij}(0).$$

Convert the last equation to the time domain, using the initial condition (4.5), and when $i \neq k$, $i \neq j$

$$\frac{d^3}{dt^3}P_{ij}(t) + \Big(2\beta + m - qm\Big)\frac{d^2}{dt^2}P_{ij}(t) + \Big(q\beta^2 + 2m\beta - 2m\beta q\Big)\frac{d}{dt}P_{ij}(t) =$$
$$qm\sum_{k\neq j}^{N}\frac{d^2}{dt^2}P_{ik}(t)h_{kj} - qm\frac{d^2}{dt^2}P_{ij}(t)$$
$$+ \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)\sum_{k\neq j}^{N}\frac{d}{dt}P_{ik}(t)h_{kj} - \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)\frac{d}{dt}P_{ij}(t)$$
$$+ \beta^2 m\sum_{k\neq j}^{N}\Big(P_{ik}(t)h_{kj} - P_{ij}(t)\Big).$$

After some calculation, it may be written as

$$\frac{d^3}{dt^3}P_{ij}(t) + (2\beta + m)\frac{d^2}{dt^2}P_{ij}(t) + (2m\beta + \beta^2)\frac{d}{dt}P_{ij}(t) + \beta^2 mP_{ij}(t) =$$
$$qm\sum_{k\neq j}^{N}\frac{d^2}{dt^2}P_{ik}(t)h_{kj} + \Big(\beta^2 - \beta^2 q + 2qm\beta\Big)\sum_{k\neq j}^{N}\frac{d}{dt}P_{ik}(t)h_{kj} + \beta^2 m\sum_{k\neq j}^{N}P_{ik}(t)h_{kj}. \tag{4.43}$$

The memory effects appear in the last equation due to the second and third derivative in time.

4.2.2 Continuous Random Walk

In the previous section we assumed the jump process X_t has discrete state space, and we derived the master equations for the transition probability and the probability density corresponds to different distributions of waiting time. These equations depend on the kernel function K and the transition matrix H. The Laplace transform and its inverse play the main role in the derivation. In this section we generalize the CTRW to include the continuous state space, i.e. the state space is the set of real numbers. This refers to the CTRW process given in chapter 2, and we will see by using our theory of conditional arrival probability how we can get the same result as shown in the literature review. Also the kernel function is the main factor to derive the master equations besides the jump distribution PDF. Unlike the previous section, here the Fourier-Laplace transform and its inverse will play a fundamental role in the whole processes of derivation and the summation sign will be replaced by the integral sign.

The conditional transition probability P may be defined as

$$\int_a^b P(y,t \mid x)dy = Pr\{a < X_t < b \mid X_0 = x\},$$

$P(y,t \mid x)$ gives the probability of finding the process X_t in the interval $(y, y+dy)$ provided $X_0 = x$. Here x is the backward variable and y is the forward variable. Again we consider the homogenous jump process as in the discrete state space. We define the probability density as follows

$$\int_a^b n(y,t)dy = Pr\{a < X_t < b\},$$

$$n(y,t) = \int_{\Re} P(y,t \mid x)n_0(x)dx. \tag{4.44}$$

The initial condition is

$$P(y,0 \mid x) = \delta(y-x).$$

The conditional arrival probability $J(y,t \mid x)$ means the process starts from x at time zero and arrives y at time t. To find it we need the waiting time PDF at point x which is $\phi(x,t)$, the survival function $\Psi(x,t)$ and the probability density for the jump process X_t from point x to point y which is $w(y \mid x)$. Now, let us write the balanc equation for $J(y,t \mid x)$

$$J(y,t \mid x) = \int_{\Re}\int_0^t J(z,t-\tau \mid x)\phi(z,\tau)w(y \mid z)dz + \phi(x,t)w(y \mid x), \quad x < z < y. \tag{4.45}$$

Accordingly, from the law of total probability we have the equation for conditional transition probability $P(y, t \mid x)$ such as

$$P(y, t \mid x) = \delta(y - x)\Psi(x, t) + \int_0^t J(y, t - \tau \mid x)\Psi(y, \tau)d\tau. \tag{4.46}$$

Let us consider the case when the waiting time distribution is the same over the state space, i.e. $\phi(x, t) = \phi(t)$. Transferring equations (4.44), (4.45) and (4.46) to the Fourier-Laplace transform give

$$\tilde{\hat{n}}(k, s) = \tilde{\hat{P}}(k, s)\hat{n}_0(k), \tag{4.47}$$

$$\tilde{\hat{J}}(k, s) = \tilde{\hat{J}}(k, s)\tilde{\phi}(s)\hat{w}(k) + \tilde{\phi}(s)\hat{w}(k), \tag{4.48}$$

$$\tilde{\hat{P}}(k, s) = \tilde{\Psi}(s) + \tilde{\hat{J}}(k, s)\tilde{\Psi}(s). \tag{4.49}$$

From (4.49), we have

$$\tilde{\hat{J}}(k, s) = \frac{\tilde{\hat{P}}(k, s)}{\tilde{\Psi}(s)} - 1.$$

Substituting the last formula into (4.48), yields

$$\frac{\tilde{\hat{P}}(k, s)}{\tilde{\Psi}(s)} - 1 = \frac{\tilde{\hat{P}}(k, s)}{\tilde{\Psi}(s)}\tilde{\phi}(s)\hat{w}(k), \tag{4.50}$$

hence

$$\tilde{\hat{P}}(k, s) = \tilde{\hat{P}}(k, s)\tilde{\phi}(s)\hat{w}(k) + \tilde{\Psi}(s), \tag{4.51}$$

or

$$\tilde{\hat{P}}(k, s) = \frac{\tilde{\Psi}(s)\hat{P}_0(k)}{1 - \tilde{\phi}(s)\hat{w}(k)} = \frac{(1 - \tilde{\phi}(s))\hat{P}_0(k)}{s(1 - \tilde{\phi}(s)\hat{w}(k))}. \tag{4.52}$$

The last equation is equivalent to the Montroll-Weiss equation in section (2.5.2). By inverting Fourier-Laplace, the equivalent of equation (4.51) in the time domain is

$$P(y, t|x) = \delta(y - x)\Psi(t) + \int_{\Re} P(z, t - \tau|x)\phi(\tau)w(z|x)dz.$$

Recalling equation(4.50) and adding $s\tilde{\hat{P}}(k, s)$ to both sides then rearranging it to obtain the master equation

$$s\tilde{\hat{P}}(k, s) - 1 = s\tilde{\hat{P}}(k, s) - \frac{\tilde{\hat{P}}(k, s)}{\tilde{\Psi}(s)} + \frac{\tilde{\hat{P}}(k, s)}{\tilde{\Psi}(s)}\tilde{\phi}(s)\hat{w}(k),$$

$$s\tilde{\hat{P}}(k, s) - \hat{P}_0(k) = -\tilde{K}(s)\tilde{\hat{P}}(k, s) + \tilde{\hat{P}}(k, s)\tilde{K}(s)\hat{w}(k). \tag{4.53}$$

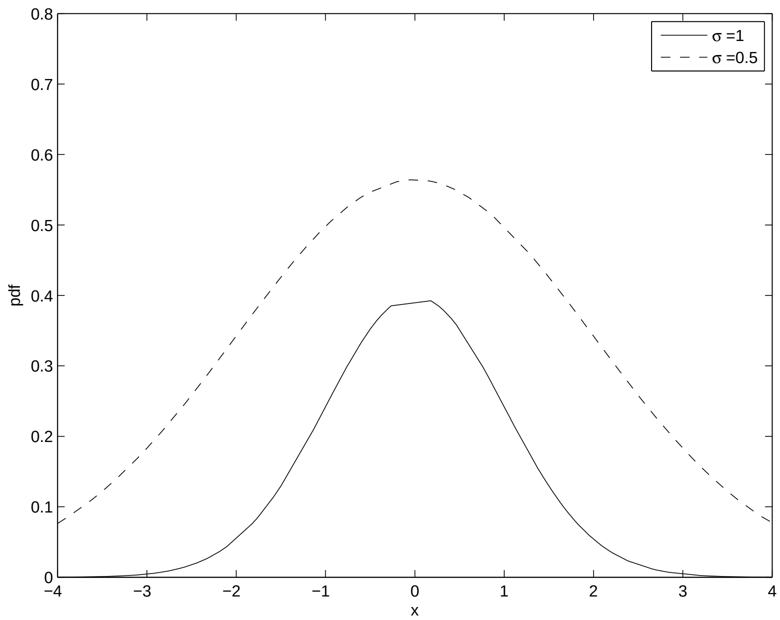

Figure 4.6: The probability density of zero-mean Gaussian distribution with various variance.

The inverse Fourier-Laplace transform gives the following master equation

$$\frac{\partial}{\partial t}P(y,t) = -\int_0^t K(\tau)P(y,t-\tau)d\tau + \int_0^t \int_{\Re} K(\tau)P(z,t-\tau)w(y-z)dzd\tau. \tag{4.54}$$

or

$$\frac{\partial}{\partial t}P(y,t) = \int_0^t K(\tau)\Big[-P(y,t-\tau) + \int_0^t \int_{\Re} P(z,t-\tau)w(y-z)dz\Big]d\tau.$$

This equation is equivalent to the master equation (2.20). Our following tasks consider the jump process with different continuous PDF, such as Gaussian distribution and Lévy distribution. Then we find the master equation for conditional transition probability in the case when the waiting time has no memory, like exponential distribution, and when it has memory, such as power-law distribution.

Gaussian distribution jump process

If the jump process follows a zero-mean Gaussian distribution, then its PDF has the form

$$w(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}}e^{\frac{-x^2}{2\sigma_x^2}},$$

where σ_x^2 is the variance of the distribution. From the properties of Gaussian distribution, (see figure 4.6), the process jumps to a state near to where it was at previous time instant and the large jumps seldom occur. In the hydrodynamic limit, we can get the approximation of $\hat{w}(k)$

$$\hat{w}(k) = 1 - \frac{\sigma_x^2 k^2}{2} + o(k^2) \sim 1 - \frac{\sigma_x^2 k^2}{2}, \quad k \to 0. \tag{4.55}$$

Here, in the hydrodynamic domain the limit of $k \to 0$ is equivalent to $x \to \infty$ in the space domain.

CTRW with Gaussian jump process and Exponential waiting time:

This case in long-time limit corresponds to Brownian motion when the waiting time's mean and jump's variance are finite. To obtain the master equation for the process in Fourier-Laplace domain equations (4.52) and (4.53), we use the definition of exponential distribution PDF in Laplace transform (4.22) in the case of invariant waiting time. For the jump process we use the Fourier transform of Gaussian jumps PDF (4.55). Therefore,

$$\tilde{\hat{P}}(k,s) = \frac{\hat{P}_0(k)}{s + m\frac{\sigma_x^2}{2}k^2}, \tag{4.56}$$

$$s\tilde{\hat{P}}(k,s) - \hat{P}_0(k) = -\left(\frac{m\sigma_x^2 k^2}{2}\right)\tilde{\hat{P}}(k,s). \tag{4.57}$$

By inverting the Fourier-Laplace transform, we obtain the diffusion equation when the jump's process follows Gaussian distribution with exponential waiting time

$$\frac{\partial}{\partial t}P(x,t) = \frac{m\sigma_x^2}{2}\frac{\partial^2}{\partial x^2}P(x,t), \tag{4.58}$$

where the Fourier transform $F\{\frac{\partial^2}{\partial x^2}P(x,t) = -k^2\hat{P}(k,t)\}$. The solution of equation (4.58) in the time domain is well-known Gaussian, due to the case of finite jump's variance and finite waiting time's mean as illustrated in chapter 2 (for more details see [90])

$$P(x,t) = \frac{1}{\sqrt{2\pi\sigma_x^2 mt}}\exp\Big(-\frac{x^2}{2\sigma_x^2 mt}\Big).$$

The second moment of the CTRW in Laplace form can be obtained from (2.32),

$$< \tilde{X}^2(s) >= -\frac{\partial^2\tilde{\hat{p}}(k,s)}{\partial k^2}\,|_{k=0}\,.$$

By using (4.56), this implies

$$\frac{\partial^2\tilde{\hat{p}}(k,s)}{\partial k^2}\,|_{k=0} = -\frac{m\sigma_x^2}{s^2},$$

which in the time domain will be

$$< X_t^2 >= m\sigma_x^2 t.$$

This result is equivalent to (2.33).

CTRW with Gaussian jump process and Power-law waiting time:

Here the waiting time PDF is heavy-tailed, so that the mean waiting time is infinite while the jump's variance is still kept finite. The asymptotic behavior of a heavy-tailed waiting time PDF is given by

$$\phi(t) \sim (t)^{-(1+\beta)} \quad \text{as} \quad t \to \infty. \tag{4.59}$$

Consequently, the long time limit corresponds to

$$\tilde{\phi}(s) \sim 1 - (\lambda s)^{\beta} \quad \text{as} \quad s \to 0, \tag{4.60}$$

where λ is a parameter with units of time, as given in section (2.5.5). Similarly, inserting these Laplace transforms of power-law PDF and the Fourier transform of jump distribution PDF (4.55) into (4.52) and (4.53) obtains the master equations as follows

$$\begin{aligned} \tilde{\hat{P}}(k,s) &= \frac{\lambda^{\beta} s^{\beta} \hat{P}_0(k)}{s[\lambda^{\beta} s^{\beta} + \frac{\sigma_x^2}{2} k^2]}, \\ &= \frac{s^{\beta-1} \hat{P}_0(k)}{s^{\beta} + K_{\beta} k^2} \\ s^{\beta} \tilde{\hat{P}}(k,s) - s^{\beta-1} \hat{P}_0(k) &= -\left(K_{\beta} k^2\right) \tilde{\hat{P}}(k,s), \end{aligned} \tag{4.61}$$

where $K_{\beta} = \sigma_x^2/(2\lambda^{\beta})$. Hence, the master equation is a time-fractional equation (see [50])

$$\frac{\partial^{\beta}}{\partial t^{\beta}} P(x,t) = K_{\beta} \frac{\partial^2}{\partial x^2} P(x,t), \tag{4.62}$$

where $\frac{\partial^{\beta}}{\partial t^{\beta}} P(x,t)$ is the Caputo fractional derivative defined by (4.38). On the other hand, the master equation can be written in the form of the Riemann-Liouville fractional derivative given by (2.45), such as

$$\frac{\partial}{\partial t} P(x,t) = K_{\beta} D_t^{1-\beta} [\frac{\partial^2}{\partial x^2} P(x,t)].$$

The time-fractional equation (4.62) is equivalent to the Fractional Fokker Planck equation given by Barkai *et al.*[8], with solution equivalent to (4.61) in the Laplace-Fourier domain.

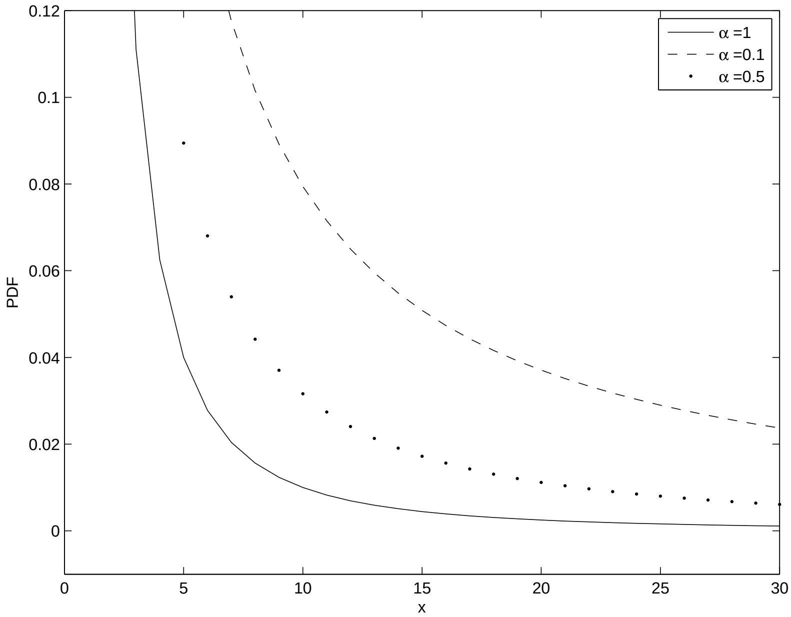

Figure 4.7: The probability density of Lévy distribution

A closed-form solution for (4.62) can be found in terms of the Wright function (see [73]) and in terms of the Fox function such as(see [90])

$$P(x,t) = \frac{1}{\sqrt{4\pi K_\beta t^\beta}} H^{2,0}_{1,2}\Big[\ \frac{x^2}{4K_\beta t^\beta} \ \Big| \ \begin{matrix} (1-\beta/2,\beta) \\ (0,1),(1/2,1) \end{matrix} \ \Big]$$

The second moment of CTRW can be found by substituting (4.61) into (2.32), hence

$$\frac{\partial^2 \tilde{\bar{p}}(k,s)}{\partial k^2} \,|_{k=0} = -\frac{2K_\beta}{s^{\beta+1}}.$$

Accordingly, in the time domain it will be

$$< X_t^2 >= \frac{2K_\beta t^\beta}{\Gamma(\beta+1)} = \frac{\sigma_x^2}{\lambda^\beta}\frac{t^\beta}{\Gamma(\beta+1)}.$$

This result is also equivalent to (2.35) when $\lambda = 1$.

Lévy distribution jump process

Lévy distribution looks similar to normal distribution in the center, but the tails are much flatter than those of Gaussian distribution. The variance of this PDF is infinite. There is no general explicit form for $w(x)$, but the Lévy distribution may be written as a power-law distribution for a large value of stochastic variable x [77]

$$w(x) \sim \sigma_x^{-\alpha}|x|^{-(\alpha+1)}, \quad \text{for} \quad |x| > \sigma_x, \quad 0 < \alpha < 2.$$

In a hydrodynamic limit of jump it will have the form, as given in section (2.5.5)

$$\hat{w}(k) = exp(-\sigma_x^\alpha |k|^\alpha) \sim 1 - \sigma_x |k|^\alpha, \quad \text{for} \quad k \to 0. \tag{4.63}$$

The fact that power-law distribution may lack a typical scale is reflected in Lévy processes, by the property that the variance of Lévy processes is infinite for $\alpha < 2$. Stochastic processes with infinite variance are extremely difficult to use and raise fundamental questions when applied to a real system. For example, in the finance system, an infinite variance would complicate the important task of risk estimation.

CTRW with Lévy jump process and Exponential waiting time:

Although the jump length variance is infinite, the process is of Markovian nature due to the finiteness of the waiting time's mean [87]. We apply the same procedure when the waiting time is exponentially distributed and has the Laplace forms (4.22) for its PDF and the jump process has Lévy distribution corresponding to (4.63) in a Fourier domain. Consequently, the master equations (4.52),(4.53) will be

$$\begin{aligned} \tilde{\hat{P}}(k,s) &= \frac{\hat{P}_0(k)}{s + m\sigma_x^\alpha |k|^\alpha}, \\ s\tilde{\hat{P}}(k,s) - \hat{P}_0(k) &= -K_\alpha |k|^\alpha \tilde{\hat{P}}(k,s), \end{aligned}$$

where $K_\alpha = m\sigma_x^\alpha$. Inverting the Fourier-Laplace transform to get the master equation in time-space domain

$$\frac{\partial}{\partial t} P(x,t) = K_\alpha \frac{\partial^\alpha}{\partial |x|^\alpha} P(x,t), \tag{4.64}$$

where $\frac{\partial^\alpha}{\partial |x|^\alpha}$ is space-fractional derivative known as **Riesz** fractional derivative of order $\alpha, 0 < \alpha < 2$ defined by [51]

$$\frac{\partial^\alpha}{\partial |x|^\alpha} P(x) = \Gamma(1+\alpha) \frac{sin(\alpha\pi/2)}{\pi} \int_0^\infty \frac{P(x+\xi) - 2P(x) + P(x-\xi)}{\xi^{1+\alpha}} d\xi.$$

The space-fractional derivative are obtained from the assumption that the random jump has a Lévy distribution, with power-law tails. The solution of the fractional space differential equation (4.64) can be obtained by using the Fox function, the result being [90]

$$P(x,t) = \frac{1}{\alpha |x|} H_{1,1}^{2,2} \Big[\tfrac{|x|}{(K_\alpha t)^{1/\alpha} t^\beta} \Big| \begin{matrix} (1, 1/\alpha), (1, 1/2) \\ (1,1), (1, 1/2) \end{matrix} \Big]$$

For limit $\alpha \to 2$, (4.64) goes to the diffusion equation and the classical Gaussian solution is recovered. When $m = \sigma_x = 1$, the solution of the space-fractional derivative equation $\frac{\partial}{\partial t}P(x,t) = \frac{\partial^\alpha}{\partial |x|^\alpha}P(x,t)$ as given by Scalas [107]

$$P(x,t) = t^{-1/\alpha}L_\alpha(xt^{-1/\alpha}),$$

where L_α is the Lévy standardized probability density function:

$$L_\alpha(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} exp\Big(-iqx - |q|^\alpha\Big)dq.$$

The second moment of CTRW in the case of exponential waiting time and Lévy jump process $< X_t^2 > \to \infty$ when $0 < \alpha < 2$, for $\alpha = 2$ corresponds to the case of the Gaussian jump process.

CTRW with Lévy jump process and Power-law waiting time:

In this case CTRW has heavy-tailed distribution for both waiting time and jump process. Accordingly, the waiting time's mean and the jump's variance are both infinite. Substituting the Laplace forms of power-law waiting time (4.60), the Fourier form of Lévy jump process (4.63) into (4.52) and (4.53) gets the master equations of CTRW, thus

$$\begin{aligned}
\tilde{\hat{P}}(k,s) &= \frac{\lambda^\beta s^\beta \hat{P}_0(k)}{s[\lambda^\beta s^\beta + \sigma_x^\alpha |k|^\alpha]}, \\
&= \frac{s^{\beta-1}\hat{P}_0(k)}{s^\beta + K_{\alpha,\beta}|k|^\alpha}, \\
s^\beta \tilde{\hat{P}}(k,s) - s^{\beta-1}\hat{P}_0(k) &= -K_{\alpha,\beta}|k|^\alpha \tilde{\hat{P}}(k,s),
\end{aligned}$$

where $K_{\alpha,\beta} = \sigma_x^\alpha/\lambda^\beta$. Hence, the master equation is a space-time fractional derivative equation [52]

$$\frac{\partial^\beta}{\partial t^\beta}P(x,t) = K_{\alpha,\beta}\frac{\partial^\alpha}{\partial |x|^\alpha}P(x,t). \tag{4.65}$$

This equation can be written in the form of Riemann-Liouville fractional derivative such as

$$\frac{\partial}{\partial t}P(x,t) = K_{\alpha,\beta}D_t^{1-\beta}[\frac{\partial^\alpha}{\partial |x|^\alpha}P(x,t)].$$

The solution of this equation, when $\lambda = \sigma_x = 1$, is defined by Scalas [108]

$$P(x,t) = t^{-\beta/\alpha}W_{\alpha,\beta}(xt^{-\beta/\alpha}),$$

where $W_{\alpha,\beta}(u)$ is given by

$$W_{\alpha,\beta}(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk e^{-iku} E_\beta(-|k|^\alpha),$$

that is the inverse Fourier transform of a Mittag-Leffler function. In the case $\beta = 1$ and $\alpha = 2$, the fractional equation reduces to the ordinary diffusion equation and the solution $P(x,t)$ becomes the Gaussian probability density function as demonstrated in (4.58) and its solution. In the general case $0 < \beta < 1$ and $0 < \alpha < 2$, the function $W_{\alpha,\beta}(u)$ is still a probability density function evolving in time and it belongs to the class of Fox function. Finally, the second moment of CTRW also diverges. In [90] they found another value for the second moment called imaginary mean squared displacement.

4.3 Example: Counting Process

In this section we are going to apply the master equation of continuous time random walk with discrete states (discrete random walk) (4.20), in order to get the distribution of the counting process and the first two moments when it is Markovian and non-Markovian.

4.3.1 Markovian Counting Process

In the Markovian case, we consider the kernel function which is defined in the Laplace form as $\tilde{K}_j(s) = \frac{\tilde{\phi}_j(s)}{\tilde{\Psi}_j(s)}, \quad j = 1, 2, ..., N$, to be a constant number. So $\tilde{K}_j(s) = m_j$, and in the time domain it is $K_j(\tau) = m_j\delta(\tau)$. Then the master equation (4.20) becomes as the system of differential equations

$$\frac{dP_j(t)}{dt} = \sum_{k\neq j}^{N} m_k P_k(t) h_{kj} - m_j P_j(t). \tag{4.66}$$

Let us now consider the counting process for which

$$h_{ij} = \begin{cases} 1 & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \tag{4.67}$$

When the waiting time is the same at all states, i.e. $m_j = m \quad \forall j$, the system becomes

$$P'_j(t) = mP_{j-1}(t) - mP_j(t).$$

Solving this equation using the induction method, we have

$$
\begin{aligned}
j = 0; \qquad & P_0'(t) = -mP_0(t) \\
& P_0(t) = e^{-mt}. \\
j = 1; \qquad & P_1'(t) = -mP_1(t) + mP_0(t), \\
& P_1'(t) + mP_1(t) = me^{-mt}, \\
& P_1(t) = mte^{-mt}. \\
j = 2; \qquad & P_2'(t) = -mP_2(t) + mP_1(t) \\
& P_2(t) = m^2\tfrac{t^2}{2}e^{-mt}. \\
j = 3; \qquad & P_3'(t) = -mP_3(t) + mP_2(t), \\
& P_3(t) = \tfrac{m^3}{2}\tfrac{t^3}{3}e^{-mt} = \tfrac{(mt)^3}{3!}e^{-mt}.
\end{aligned}
$$

and so on, at j-steps

$$P_j(t) = \frac{(mt)^j}{j!}e^{-mt}. \quad j = 1, 2, ... \tag{4.68}$$

This is the PDF of the Poisson process (2.7), an example of the counting process, where the expected value of the process is given by

$$< X_t >= \sum_{j=0}^{\infty} jP_j(t) = mt.$$

In fact, we reach a Poisson process as a Markovian case of the counting process, and it has an expected value that is proportional to time. This result is equivalent to (2.33) when the second moment (or the variance when jump PDF is even) of the jump process equals one. Likewise, we solve (4.66) again in the Laplace domain to see if we will get the same result, in order to use the Laplace transform in the rest of the applications. Starting from equation (4.19) under our assumptions (4.67) and invariant waiting time, we get

$$s\tilde{P}_j(s) - P_j(0) = m\tilde{P}_{j-1}(s) - m\tilde{P}_j(s),$$

Using the induction method again gives

$$j = 0; \qquad s\tilde{P}_0(s) - P_0(0) = -m\tilde{P}_0(s),$$
$$\tilde{P}_0(s) = P_0(0)\tilde{\Psi}(s),$$

$$j = 1; \qquad s\tilde{P}_1(s) - P_1(0) = m\tilde{P}_0(s) - m\tilde{P}_1(s),$$
$$\tilde{P}_1(s) = P_0(0)\tilde{\phi}(s)\tilde{\Psi}(s) + P_1(0)\tilde{\Psi}(s)$$

$$j = 2; \qquad s\tilde{P}_2(s) - P_2(0) = m\tilde{P}_1(s) - m\tilde{P}_2(s),$$
$$\tilde{P}_2(s) = P_0(0)\tilde{\phi}^2(s)\tilde{\Psi}(s) + P_1(0)\tilde{\phi}(s)\tilde{\Psi}(s) + P_2(0)\tilde{\Psi}(s).$$

For general j

$$\tilde{P}_j(s) = \sum_{i=0}^{j} P_i(0)\tilde{\phi}^{j-i}(s)\tilde{\Psi}(s).$$

$$\because P_i(0) = \begin{cases} 1 & \text{if} \quad i = 0 \\ 0 & \text{otherwise} \end{cases}$$

since there are no jumps at time zero. Therefore, the distribution of the process is

$$\tilde{P}_j(s) = \tilde{\phi}^j(s)\tilde{\Psi}(s),$$

this result is equivalent to (2.21), which in this Markovian case

$$\tilde{P}_j(s) = \frac{m^j}{(m+s)^{j+1}}.$$

This formula is the Laplace transform of the Poisson process given by equation (4.68). The expected value of this distribution in the Laplace domain is given by

$$< \tilde{X}(s) >= \sum_{j=0}^{\infty} j\tilde{P}_j(s) = \frac{m}{s^2}.$$

Again this is the Laplace formula of (mt), the expected value of the process in the time domain. In the same way we can find the second moment

$$< \tilde{X}^2(s) >= \sum_{j=0}^{\infty} j^2\tilde{P}_j(s) = \frac{m}{s^2} + \frac{2m^2}{s^3},$$

accordingly,

$$< X_t^2 >= mt + m^2 t^2.$$

Consequently, the variance will be

$$var[X_t] = mt,$$

which is the variance of the Poisson process. We can conclude that if we apply the master equation that we got from our theory of conditional arrival probability to the counting process, in the Markovian case we get a Poisson process either in the time domain or in the Laplace domain. Next, in the non-Markovian counting process we are going to solve the system of differential equations only in a Laplace domain.

4.3.2 Non-Markovian Counting Process

Now we are in a position to apply the master equation (4.20) to the non-Markovian counting process. This application aims to find the distribution of the counting process X_t when the process has a memory effect, i.e. the waiting time is not exponentially distributed. For example, the waiting time has gamma distribution or power-law distribution. Moreover, we are going to find the first two moments for these two distributions in the Laplace domain then convert them to the time domain. Back to the master equation

$$\frac{dP_j(t)}{dt} = -\int_0^t K_j(t-\tau)P_j(\tau)d\tau + \sum_{k\neq j}^{N}\int_0^t K_k(t-\tau)P_k(\tau)h_{kj}d\tau.$$

Assume that the waiting time is the same at all states , $K_j(t-\tau) = K(t-\tau) \quad \forall j$, and h_{ij} as given in (4.67). Then, the master equation will be

$$P_j'(t) = \int_0^t K(t-\tau)P_{j-1}(\tau)d\tau - \int_0^t K(t-\tau)P_j(\tau)d\tau.$$

Transfer it to the Laplace form:

$$s\tilde{P}_j(s) - P_0(0) = \tilde{K}(s)\tilde{P}_{j-1}(s) - \tilde{K}(s)\tilde{P}_j(s).$$

Again we use the induction method to find the probability density $\tilde{P}_j(s)$ in the Laplace form

$$j = 0; \qquad s\tilde{P}_0(s) - P_0(0) = -\tilde{K}(s)\tilde{P}_0(s)$$
$$\tilde{P}_0(s) = P_0(0)\tilde{\Psi}(s).$$

$$j = 1; \qquad s\tilde{P}_1(s) - P_0(0) = \tilde{K}(s)\tilde{P}_0(s) - \tilde{K}(s)\tilde{P}_1(s),$$
$$\tilde{P}_1(s) = P_0(0)\tilde{\phi}(s)\tilde{\Psi}(s) + P_0(0)\tilde{\Psi}(s).$$

$$j = 2; \qquad s\tilde{P}_2(s) - P_0(0) = \tilde{K}(s)\tilde{P}_1(s) - \tilde{K}(s)\tilde{P}_2(s),$$
$$\tilde{P}_2(s) = P_0(0)\tilde{\phi}^2(s)\tilde{\Psi}(s) + P_1(0)\tilde{\phi}(s)\tilde{\Psi}(s) + P_2(0)\tilde{\Psi}(s).$$

$$j = 3; \qquad s\tilde{P}_3(s) - P_0(0) = \tilde{K}(s)\tilde{P}_2(s) - \tilde{K}(s)\tilde{P}_3(s),$$
$$\tilde{P}_2(s) = P_3(0)\tilde{\phi}^3(s)\tilde{\Psi}(s) + P_1(0)\tilde{\phi}^2(s)\tilde{\Psi}(s) + P_2(0)\tilde{\phi}(s)\tilde{\Psi}(s) + P_3(0)\tilde{\Psi}(s).$$

For general j we have

$$\tilde{P}_j(s) = \sum_{i=0}^{j} P_i(0)\tilde{\phi}^{j-i}(s)\tilde{\Psi}(s),$$

since $P_i(0) = 0 \forall i \neq 0$, thus

$$\tilde{P}_j(s) = \tilde{\phi}^j(s)\tilde{\Psi}(s). \tag{4.69}$$

These are the same probability densities we got in the Markovian case. Therefore its expected value is

$$< \tilde{X}(s) >= \sum_{j=0}^{\infty} j\tilde{P}_j(s) = \tilde{\Psi}(s)\frac{\tilde{\phi}(s)}{(1-\tilde{\phi}(s))^2} = \frac{\tilde{\phi}(s)}{s(1-\tilde{\phi}(s))}, \tag{4.70}$$

and the second moment will be

$$< \tilde{X}^2(s) >= \sum_{j=0}^{\infty} j^2\tilde{P}_j(s) = \tilde{\Psi}(s)\frac{\tilde{\phi}(s)(1+\tilde{\phi}(s))}{(1-\tilde{\phi}(s))^3} = \frac{\tilde{\phi}(s)(1+\tilde{\phi}(s))}{s(1-\tilde{\phi}(s))^2}. \tag{4.71}$$

This result can be achieved from (2.31) and (2.32) when the first and second moments of the jump process are equal to one. This corresponds to the counting process because the average jump of the process is one jump per unit of time.

Counting process with Gamma distribution waiting time

If we substitute the Laplace form of gamma waiting time PDF (4.27) when $\alpha = 2$, and its survival function (4.28) into (4.69), the counting process will have the following Laplace form distribution

$$\tilde{P}_j(s) = \frac{\beta^{2j}(s + 2\beta)}{(s + \beta)^{2j+2}}. \tag{4.72}$$

From equation (4.70), the expected value is

$$< \tilde{X}(s) >= \frac{\beta^2}{s^2(s + 2\beta)}.$$

Similarly, we obtain the second moment from equation (4.71)

$$< \tilde{X}^2(s) >= \frac{\beta^2}{s^2(s + 2\beta)} + \frac{2\beta^4}{s^3(s + 2\beta)^2}.$$

Using the inverse Laplace transform with the convolution theorem, the expected value is

$$< X_t >= \frac{1}{4}[e^{-2\beta t} - 1] + \frac{\beta t}{2},$$

while the second moment will be

$$< X_t^2 >= \frac{1}{2} - \frac{\beta t}{2} + \frac{\beta^2 t^2}{2} - \frac{e^{-\beta t}}{2}[\beta t + 1].$$

The expected value and the second moment are the combined linear function of t and the monotonic exponentially decreasing function of t. However, for long time $t \to \infty$ they will be a proportional function of time.

Counting process with power-law distribution waiting time

In this case, we are going to use the Laplace transforms of the Mittag-Leffler function as an example of power-law waiting time PDF (4.36) and its survival function (4.37). Inserting them into (4.69), the counting process will have the following Laplace form distribution

$$\tilde{P}_j(s) = \frac{s^{\beta-1}}{(1 + s^\beta)} \frac{1}{(1 + s^\beta)^j}. \tag{4.73}$$

The expected value is obtained from (4.70)

$$< \tilde{X}(s) >= \frac{1}{s^{\beta+1}},$$

and the second moment can be found from (4.71)

$$< \tilde{X}^2(s) >= \frac{1}{s^{\beta+1}} + \frac{2}{s^{2\beta+1}}.$$

Inverting them to the time domain, yields :

$$< X_t >= \frac{t^\beta}{\Gamma(\beta+1)},$$

and

$$< X_t^2 >= \frac{t^\beta}{\Gamma(\beta+1)} + \frac{2t^{2\beta}}{\Gamma(2\beta+1)}, \quad 0 < \beta < 1.$$

The expected value and the second moment are non-linear functions of time. It is decaying very slowly as $t \to \infty$ due to the power-law form formula which is strong evidence of memory effects.

4.3.3 Numerical Results

Now we are going to display the difference between the three cases of the counting process with different waiting time distributions numerically, in order to notice the memory effects in each case. The numerical results are achieved by using Matlab codes. Firstly, we consider the three waiting time distributions with the same mean values of 10 units of time. Secondly, we fix the number of jumps (for example, 10 jumps). Then we find the time to reach ten jumps for each waiting time PDF; see figures 4.8, 4.9 and 4.10.

From these figures, we can see that the required time to reach ten jumps in a case of Pareto[3] waiting time distribution (as an example of power-law distribution, see (4.35)) is around 200 time units. This time is longer than the time required by gamma waiting time distribution (around 100 time units) and exponential waiting time distributions (around 85 time units).

In the next experiment: Firstly, we use the three waiting time distributions with the same mean values of 10 units of time as in the previous case. Secondly, we fix the time (30 units). Then we find the number of jumps during this time interval. We denote the number of jumps by X, and X is a random variable. We repeat the experiment N times. We count how many times during the N trials we get $X = j, j = 1, 2, ...$ For example, M times the number of jumps $X = j$. So we find the distribution of the number of jumps X as following:

$$Pr\{\text{X=j}\} = \frac{M}{N}, j = 1, 2, ...$$

[3] We use Pareto distribution with $\beta > 1$ to get a finite expected value, in order to compare it with the other distributions that have finite expected value.

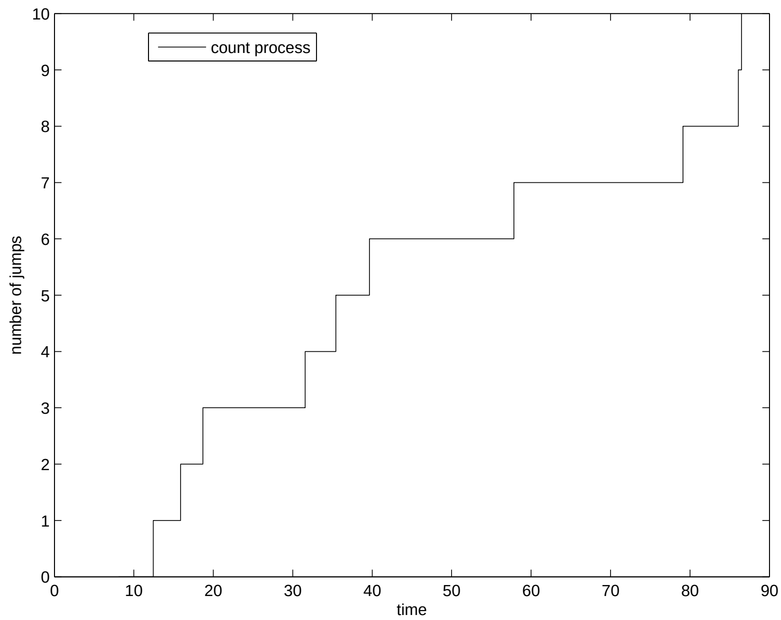

Figure 4.8: A sample path of the counting process has 10 jumps with exponential waiting time with mean=10 (Poisson process)

see figures 4.11, 4.12 and 4.13. From these figures the mean value of the number of jumps is noticeable. In the case of exponential waiting time distribution the mean value of the number of jumps is 10, while it is 5 in the case of gamma waiting time distribution and 8 in the case of Pareto waiting time distribution. In other words, the number of jumps has a larger mean value in the case of exponential distributed waiting. This is due to the short waiting time between jumps. Furthermore, for the same time interval the number of jumps takes values, in the case of gamma and Pareto waiting time distributions, less than in the case of exponential waiting time distribution. For example, the maximum number of jumps during the 30 units of time is 22 jumps in the case of exponential waiting time distribution, while it is 12 jumps for the other distributions.

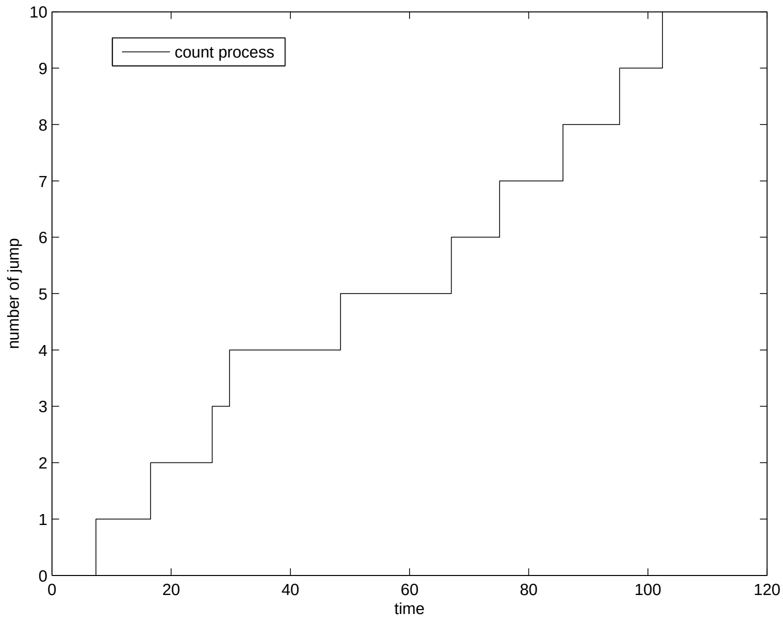

Figure 4.9: A sample path of the counting process has 10 jumps with Gamma waiting time with $\alpha = 2, \beta = 5$ and mean=10

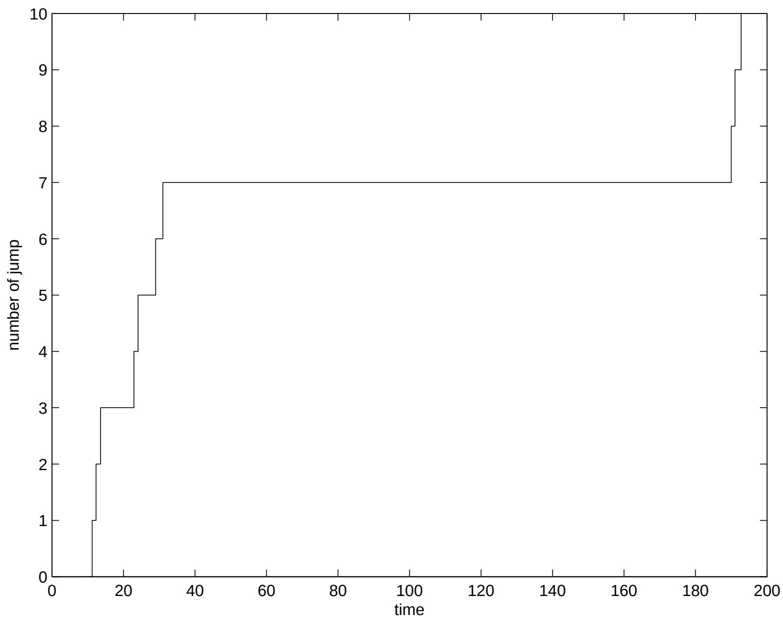

Figure 4.10: A sample path of the counting process has 10 jumps with Pareto waiting time with minimum value equal one, $\beta = 1.0961$ and mean =10

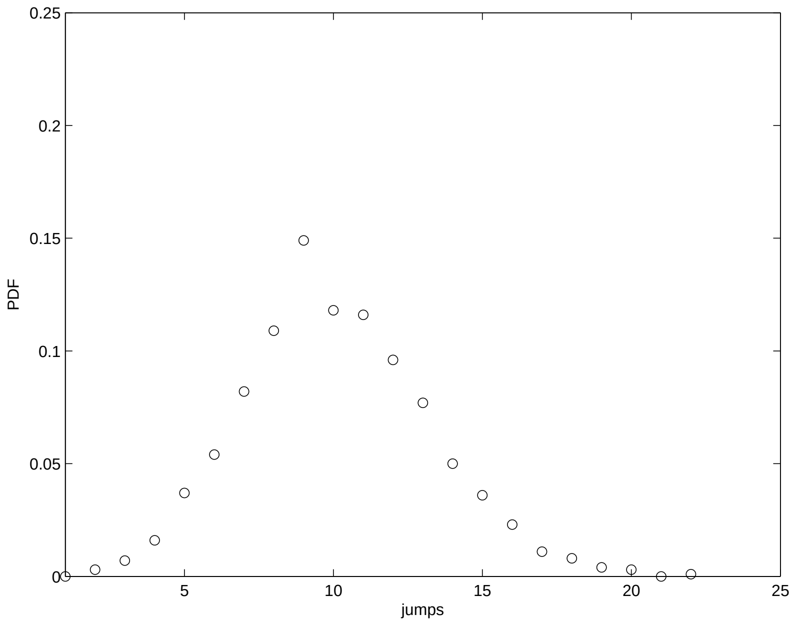

Figure 4.11: The distribution of number of jumps ($Prob\{X = j\} = \frac{M}{N}, N = 1000, M$ is number of trials give$X = j$) when the waiting time is exponentially distributed with mean=10

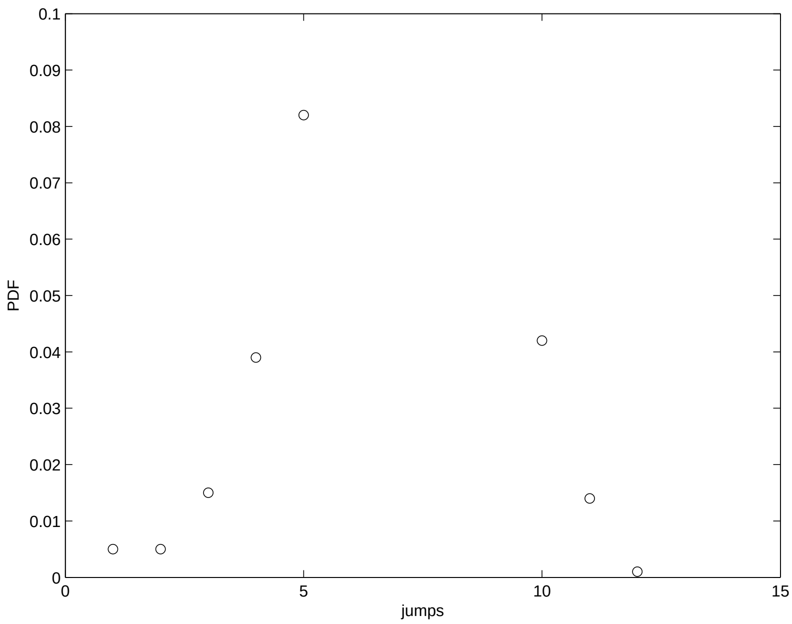

Figure 4.12: The distribution of number of jumps ($Prob\{X = j\} = \frac{M}{N}, N = 1000, M$ is number of trials give$X = j$) when the waiting time has gamma distribution with $\alpha = 2, \beta = 5$ and mean =10

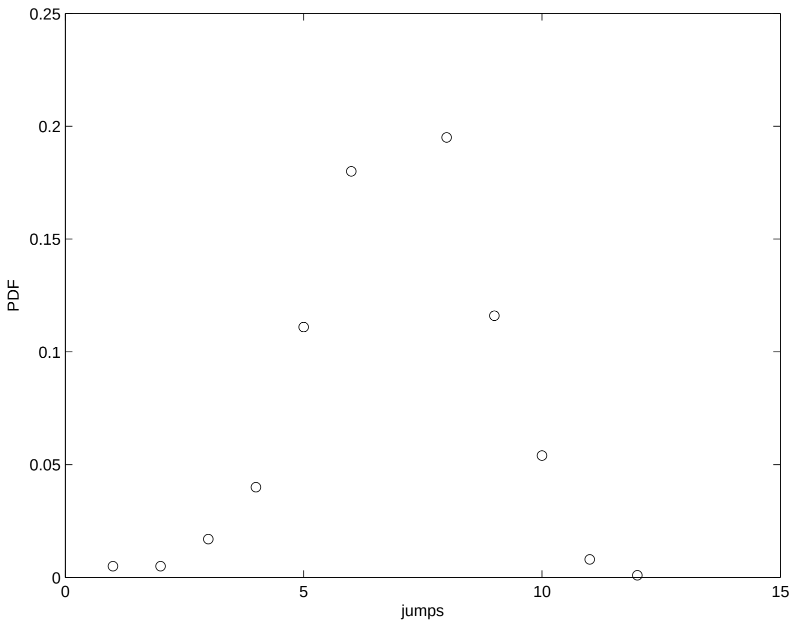

Figure 4.13: The distribution of number of jumps ($Prob\{X = j\} = \frac{M}{N}, N = 1000, M$ is number of trials give$X = j$) when the waiting time has Pareto distribution with $\beta = 1.0916$ and mean=10

4.3.4 Switching Model

The switching model is an example of N-state model when $N = 2$. In this subsection we consider the stochastic volatility with two random variables, for example the minimum value and the maximum value. Accordingly, we are going to find the stationary distribution for the switching process between these two values corresponding to two cases of waiting time. First, when the waiting time has a finite mean, for example, exponential distribution and gamma distribution. Second, when the waiting time has an infinite mean as is the case in power-law distribution, Pareto distribution with $\beta < 1$. The illustrations of these two cases are shown in figures 4.14, 4.15. Now let us consider again the master equation (4.19)

$$s\tilde{P}_{ij}(s) - P_{ij}(0) = -\tilde{K}_j(s)\tilde{P}_{ij}(s) + \sum_{k\neq j}^{N} \tilde{P}_{ik}(s)\tilde{K}_k(s)h_{kj}.$$

When $N = 2$, $i, j = 1, 2$. For the case $h_{11} = h_{22} = 0$ and $h_{12} = h_{21} = 1$, the process switches between the two states without waiting at either of them, thus

$$s\tilde{P}_{12}(s) - P_{12}(0) = -\tilde{K}_2(s)\tilde{P}_{12}(s) + \tilde{K}_1(s)\tilde{P}_{21}(s), \tag{4.74}$$

$$s\tilde{P}_{21}(s) - P_{21}(0) = -\tilde{K}_1(s)\tilde{P}_{21}(s) + \tilde{K}_2(s)\tilde{P}_{12}(s). \tag{4.75}$$

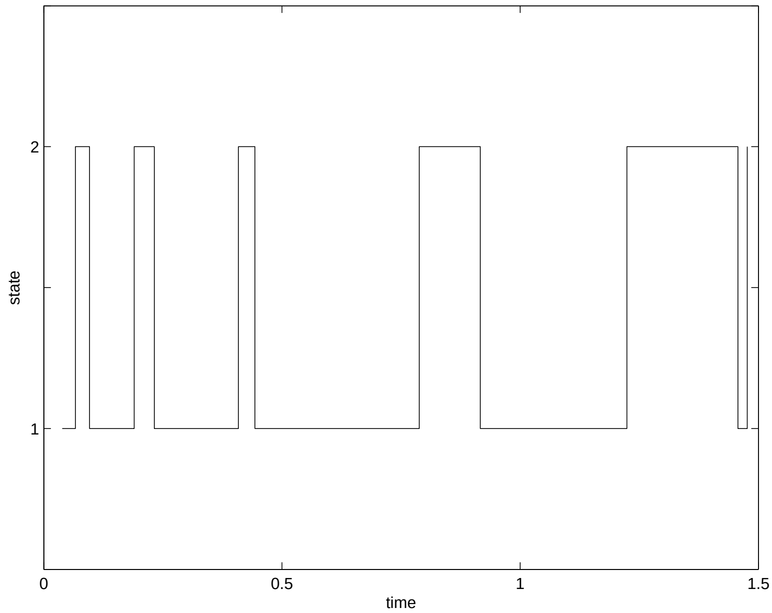

Figure 4.14: A sample path of the switching process with exponential waiting time with mean=10

By using $P_{12}(t)+P_{21}(t) = 1$, this implies $\tilde{P}_{12}(s)+\tilde{P}_{21}(s) = \frac{1}{s}$. Substituting $\tilde{P}_{21}(s) = \frac{1}{s}-\tilde{P}_{12}(s)$ in (4.74), we get

$$s\tilde{P}_{12}(s) - P_{12}(0) = -\tilde{K}_2(s)\tilde{P}_{12}(s) + \tilde{K}_1(s)[\frac{1}{s} - \tilde{P}_{12}(s)],$$

For stationary distribution P^{st}_{ij}, we have $\frac{dP^{st}_{ij}}{dt} = 0$. Therefore $s\tilde{P}_{ij}(s) - P_{ij}(0) = 0$, so the last equation may be rewritten as

$$\frac{\tilde{K}_1(s)}{s} = \tilde{P_{12}}(s)[\tilde{K}_2(s) + \tilde{K}_1(s)].$$

Hence,

$$s\tilde{P_{12}}(s) = \frac{\tilde{K}_1(s)}{\tilde{K}_2(s) + \tilde{K}_1(s)}. \tag{4.76}$$

Similarly, for $\tilde{P_{21}}(s)$, we have

$$s\tilde{P_{21}}(s) = \frac{\tilde{K}_2(s)}{\tilde{K}_2(s) + \tilde{K}_1(s)}. \tag{4.77}$$

Now let us find the stationary distribution for the different waiting time PDF using the fact that $lim_{t\to\infty}P(t) = lim_{s\to 0}s\tilde{P}(s)$.

Waiting time with finite mean

If the waiting time $\phi_j(t)$ has a finite mean τ_j, for a large time it can be expanded in the Laplace form to be written in the form

$$\tilde{\phi}_j(s) \sim 1 - \tau_j s + o(s), \quad j = 1, 2, \tag{4.78}$$

and therefore the Laplace form of the survival function is $\tilde{\Psi}_j(s) = \tau_j$. The kernel function is $\tilde{K}_j(s) = (1 - \tau_j s)/\tau_j$. Substituting the kernel functions in (4.76), (4.77) gives

$$\begin{aligned} s\tilde{P}_{12}(s) &= \frac{\tau_2 - \tau_2\tau_1 s}{\tau_1(1-\tau_2 s) + \tau_2(1-\tau_1 s)}, \\ lim_{s\to 0} s\tilde{P}_{12}(s) &= \frac{\tau_2}{\tau_1 + \tau_2}. \\ s\tilde{P}_{21}(s) &= \frac{\tau_1 - \tau_1\tau_2 s}{\tau_1(1-\tau_2 s) + \tau_2(1-\tau_1 s)}, \\ lim_{s\to 0} s\tilde{P}_{21}(s) &= \frac{\tau_1}{\tau_1 + \tau_2}. \end{aligned}$$

Hence, as $t \to \infty$, $P_{12}(t) \to P_{12}^{st}$

$$\begin{aligned} P_{12}^{st} &= \frac{\tau_2}{\tau_1 + \tau_2}, \\ P_{21}^{st} &= \frac{\tau_1}{\tau_1 + \tau_2}. \end{aligned}$$

If $\tilde{\phi}_1(s) = \tilde{\phi}_2(s)$, this implies

$$lim_{s\to 0} s\tilde{P}_{12}(s) = lim_{s\to 0} s\tilde{P}_{21}(s) = \frac{1}{2},$$

or

$$lim_{t\to\infty} P_{12}(t) = lim_{t\to\infty} P_{21}(t) = \frac{1}{2}$$

As we noticed from the previous results, as $t \to \infty$ the stationary distribution will be a specific constant for each transition and the total of these probabilities will equal one. In the case of invariant waiting time, the stationary distribution will be equal at each state and again their sum equal one. Next we are going to consider the case when waiting time has an infinite mean.

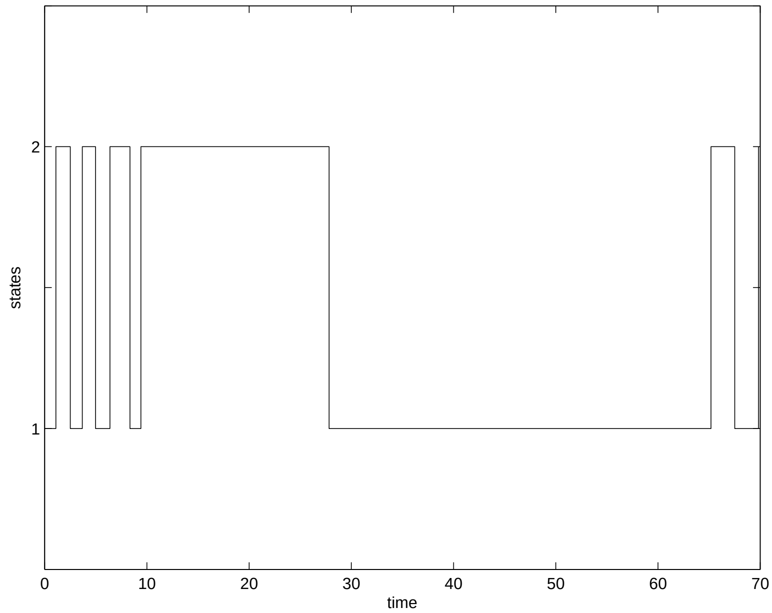

Figure 4.15: A sample path of the switching process with Pareto waiting time with $\beta = 0.8$

Waiting time with infinite mean

In the case when waiting time has an infinite mean, its PDF can be expanded in the following Laplace transform

$$\tilde{\phi}_j(s) \sim 1 - s^{\beta_j}, \quad 0 < \beta_j < 1, \quad j = 1, 2.$$

This formula corresponds to the Laplace transform of the Mittag-Leffler function (4.36), and the power-law formula (4.60) when $\lambda = 1$. The survival function will be $\Psi_j(s) = s^{\beta_j - 1}$, and the kernel function is $\tilde{K}_j(s) = (1 - s^{\beta_j})/s^{\beta_j - 1}$. By inserting them into (4.76) and (4.77) we obtain:

$$\begin{aligned} s\tilde{P}_{12}(s) &= \frac{s^{\beta_2 - 1}(1 - s^{\beta_1})}{s^{\beta_1 - 1}(1 - s^{\beta_2}) + s^{\beta_2 - 1}(1 - s^{\beta_1})}, \\ s\tilde{P}_{21}(s) &= \frac{s^{\beta_1 - 1}(1 - s^{\beta_2})}{s^{\beta_1 - 1}(1 - s^{\beta_2}) + s^{\beta_2 - 1}(1 - s^{\beta_1})}, \end{aligned}$$

or

$$\begin{aligned} s\tilde{P}_{12}(s) &= \frac{1 - s^{\beta_1}}{1 - 2s^{\beta_1} + s^{\beta_1 - \beta_2}}, \\ s\tilde{P}_{21}(s) &= \frac{1 - s^{\beta_2}}{1 - 2s^{\beta_2} + s^{\beta_2 - \beta_1}}, \end{aligned}$$

In the case of $\beta_1 > \beta_2$, then

$$\begin{aligned} P_{12}^{st} &= lim_{s\to 0} s\tilde{P}_{12}(s) = 1, \\ P_{21}^{st} &= lim_{s\to 0} s\tilde{P}_{21}(s) = 0, \end{aligned}$$

and vice-versa in the case of $\beta_2 > \beta_1$

$$\begin{aligned} P_{12}^{st} &= lim_{s\to 0} s\tilde{P}_{12}(s) = 0, \\ P_{21}^{st} &= lim_{s\to 0} s\tilde{P}_{21}(s) = 1. \end{aligned}$$

We can see from the last two equations that, as $t \to \infty$: when $\beta_1 > \beta_2$ the system starts from state 1 and is trapped at state 2, while $\beta_2 > \beta_1$ the system starts from state 2 and is trapped at state 1.

Chapter 5

Stochastic Volatility With Memory Effects

5.1 Introduction

Volatility is one of the most fundamental quantities in the financial markets. It changes all the time and it has memory effects. The main issue about volatility involves the estimation of volatility and its persistence [116]. We give in chapter 3 a short review of modeling volatility. The first attempt considered the constant volatility, as shown by the Black-Scholes model. The following attempt considered volatility as a time-dependent function. Later on, several extensions of the Black-Scholes model implemented the idea of stochastic volatility and stochastic volatility with jumps. Further classes of stochastic processes were introduced, called ARCH and GARCH processes, which were designed to investigate the persistence of volatility. The stochastic volatility models play an essential role in the field of mathematical finance. Although the stochastic volatility models are important, the empirical data shows that volatility has long memory phenomena that the stochastic models do not satisfy. Conversely, not only asset price and its returns or its volatility change randomly; the waiting times between the asset jumps also change randomly. So, we need to find a model that exhibits the change in both returns or volatility and the waiting time between the transaction. Accordingly, we propose using the CTRW process to model stochastic volatility. We showed at the end of chapter 3 how the CTRW

is used to model the logarithm of asset price return (3.24). In this chapter we aim to use CTRW itself to model the stochastic volatility σ instead of the logarithm of return.
We derive the master equation of CTRW when the jumps exhibit the volatility changes. Consequently, the master equation shows the memory effect via the kernel function. Therefore, to define the kernel function we need to define the waiting time density between the jumps. Hence, we propose another method to find the master equation of the probability of the random walk, in the case where the time space is continuous and the state space is discrete. This method depends on the age model, which studies the transition rate. In the previous chapter the transition rate from one state to another was given by the matrix H with entries h_{ij}. In this chapter the transition rate depends on the age τ (waiting time), i.e. the transition rate is a function of age τ and not constant as h_{ij}. τ is the waiting time in any state unlike t, which is the arrival time or the required time to find the transition probability in. First, we are going to apply this theory to the stochastic volatility model when it has two values (switching process), then we generalize it to multi-values.

5.2 Generalize CTRW To Model Stochastic Volatility

As mentioned in chapter 2, CTRW is characterized by the waiting time t and the jumps x. However, to replace the jump x with the volatility σ it is worth noting the difference between them. That is: σ is a positive random variable, while x can be either positive or negative. Now, we introduce the notations in the new CTRW model for volatility jumps and waiting time. We denote the volatility at time instant t, $(t \geq 0)$ by $\sigma(t)$. The times when volatility jumps happen are denoted by $t_1, t_2, ...,$ and $t_0 = 0$. The waiting times are defined as $\tau_i = t_i - t_{i-1}, i = 1, 2, ...,$ the same as those in chapter 2. The new random variable, the jump of volatility, is represented by ξ, i.e. $\xi(t_i) = \sigma(t_i) - \sigma(t_{i-1}), i = 1, 2, ...$ Furthermore, the joint probability density of waiting time and volatility jumps is written as $\Phi(\tau, \xi)$, such that

$$\phi(\tau) = \int_0^\infty \Phi(\xi, \tau) d\xi \quad \text{and} \quad w(\xi) = \int_0^\infty \Phi(\xi, \tau) d\tau$$

represent the definitions of waiting time PDF and volatility jump PDF respectively. In our study we focus on the case of the uncoupled CTRW model, so we have $\Phi(\xi, \tau) = w(\xi)\phi(\tau)$.

The survival probability is still: $\Psi(\tau) = 1 - \int_0^\tau \phi(\tau')d\tau'$.

Finally, we assume $p(\sigma, t)$ to be the PDF that represents the probability of the volatility being σ at time instant t. The initial condition is $\sigma = \sigma_0, \sigma_0 > 0$ at time $t = 0$, i.e. $p(\sigma, 0) = \delta(\sigma - \sigma_0)$.

The master equation of the new CTRW model can be written as

$$p(\sigma, t) = \delta(\sigma - \sigma_0)\Psi(t) + \int_0^t \phi(t - t') \int_0^\infty w(\sigma - \sigma')p(\sigma', t,)d\sigma' dt'. \tag{5.1}$$

The difference between (5.1) and (2.13) is the limit of the second integral. Recalling the same derivation in chapter 2, we can obtain the analog of the master equation in the Fourier-Laplace transform

$$\tilde{\hat{p}}(k, s) = \tilde{\Psi}(s).\frac{1}{1 - \tilde{\phi}(s)\hat{w}(k)} = \frac{1 - \tilde{\phi}(s)}{s(1 - \tilde{\phi}(s)\hat{w}(k))} \tag{5.2}$$

The memory function $H(t)$ has the same definition (2.17), and so on the kernel function

$$\tilde{K}(s) = \frac{1}{\tilde{H}(s)} = \frac{\tilde{\phi}(s)}{\tilde{\Psi}(s)} = \frac{s\tilde{\phi}(s)}{1 - \tilde{\phi}(s)}.$$

From equation (5.2), we have

$$s\tilde{\hat{p}}(k, s) - 1 = -\tilde{K}(s)[\tilde{\hat{p}}(k, s) - \hat{w}(k)\tilde{\hat{p}}(k, s)]. \tag{5.3}$$

By inverting the Fourier-Laplace transform, the last equation can be written in the time domain as:

$$\frac{\partial}{\partial t}p(\sigma, t) = \int_0^t K(t - t')[-p(\sigma, t') + \int_0^\infty w(\sigma - \sigma')p(\sigma', t')d\sigma']dt'. \tag{5.4}$$

Equation (5.4) is the generalized master equation for the new CTRW model, when the waiting times and volatility jumps are independent of each other. The memory effects clear at the RHS due to the kernel function $K(t - t')$.

Accordingly, we can apply the general master equations derived in the previous chapter by replacing x with σ. Firstly, in the case where volatility has N discrete values, it follows CTRW with discrete states and transition matrix H. The entries of the transition matrix denote the transition rates between the values of volatility. The transition probability of volatility to state j at time instant t (4.25) corresponds to exponential waiting time distribution, (4.33) when waiting time has gamma distribution, and (4.40) corresponds to power-law waiting time distribution.

Secondly, when volatility takes different values of positive real numbers, it follows CTRW with continuous state, and the values of volatility have continuous distribution such as Gaussian distribution or Lévy distribution. Consequently, the probability of volatility being at state σ at instant time t takes the form (4.58) when the volatility values are Gaussian and the waiting times are exponentially distributed, while it takes the form (4.62) when waiting times are power-law distributed. Similarly, the probability of volatility takes the form (4.64) when the volatility values have Lévy distribution and waiting times are exponentially distributed, while (4.65) when waiting times have power-law distribution.

5.3 Age Model

We see in (5.4) and in the master equation (4.19) the appearance of the kernel function shows the memory effect in modeling volatility. The memory effect happens because the transition between jumps is a function of time. To define the kernel function we need to find the waiting time density. Thus, we introduce the age model first, which illustrates how to find the waiting time distribution from the hazard function. Then we find the master equation for the probability when the state space is discrete. Therefore, we consider the case when volatility has finite values. We start with two values to derive the master equation, then we generalize the method to N values (multi-states).

5.3.1 Two-State Non-Markovian Process

We assume the volatility $\sigma(t)$ takes only two values σ_1 and σ_2. The rate of transition from state σ_1 to state σ_2 or vice-versa $\lambda_i(\tau), i = 1, 2$ is called the hazard function. It depends on the age τ (waiting time). The conditional transition probability is $P_i(t, \tau)$, which is the conditional probability that $\sigma(t)$ takes the value of σ_i at time t and the age is τ; see figure 5.1. Finally, the probability $n_i(t)$ that $\sigma(t)$ takes value σ_i at time t is

$$n_i(t) = Pr\{\sigma(t) = \sigma_i\} = \int_0^t P_i(t, \tau) d\tau, \tag{5.5}$$

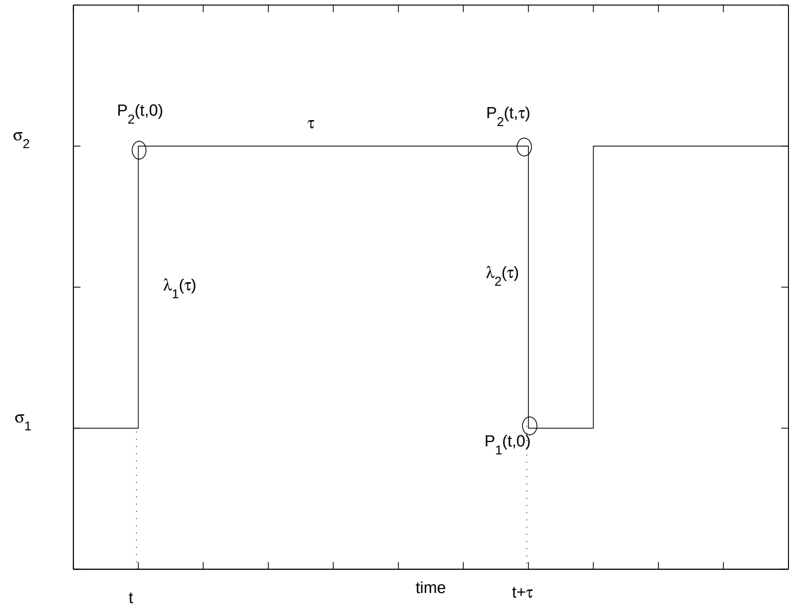

Figure 5.1: The conditional probability $P_2(t, 0)$ when the age $\tau = 0$ and $P_2(t, \tau)$ when the age $\tau > 0$. The transition rate from state 1 to state 2 is $\lambda_1(\tau)$, and from state 2 to state 1 is $\lambda_1(\tau)$

where

$$P_i(t, \tau)d\tau = Pr\{\sigma(t) = \sigma_i, \tau \leq T_i \leq \tau + d\tau\},$$

where T_i is the waiting time at state i from the arrival until time t. Now, let us formulate the equation for $P_i(t, \tau)$

$$P_i(t + h, \tau + h) = P_i(t, \tau)(1 - \lambda_i(\tau)h) + o(h), \quad i = 1, 2,$$

$$\begin{aligned} P_i(t + h, \tau + h) - P_i(t, \tau) &= -P_i(t, \tau)\lambda_i(\tau)h, \\ \frac{P_i(t + h, \tau + h) - P_i(t, \tau)}{h} &= -P_i(t, \tau)\lambda_i(\tau), \end{aligned} \tag{5.6}$$

which satisfies the boundary conditions at $\tau = 0$

$$P_1(t, 0) = \int_0^t \lambda_2(\tau)P_2(t, \tau)d\tau, \tag{5.7a}$$

$$P_2(t, 0) = \int_0^t \lambda_1(\tau)P_1(t, \tau)d\tau, \tag{5.7b}$$

and the initial conditions at $t = 0$

$$P_1(0, \tau) = P_1(0)\delta(\tau), \tag{5.8a}$$

$$P_2(0, \tau) = P_2(0)\delta(\tau). \tag{5.8b}$$

Equations (5.6) can be rewritten in the limit of $h \to 0$ as

$$\frac{\partial P_1}{\partial t} + \frac{\partial P_1}{\partial \tau} = -\lambda_1(\tau)P_1(t,\tau), \tag{5.9a}$$

$$\frac{\partial P_2}{\partial t} + \frac{\partial P_2}{\partial \tau} = -\lambda_2(\tau)P_2(t,\tau). \tag{5.9b}$$

Differentiation of equations (5.5) with respect to t give[1]

$$\frac{dn_1(t)}{dt} = P_1(t,t) + \int_0^t \frac{\partial P_1(t,\tau)}{\partial t} d\tau, \tag{5.10a}$$

$$\frac{dn_2(t)}{dt} = P_2(t,t) + \int_0^t \frac{\partial P_2(t,\tau)}{\partial t} d\tau. \tag{5.10b}$$

If we integrate equations (5.9), we get

$$\int_0^t \frac{\partial P_1(t,\tau)}{\partial t} d\tau = -\int_0^t \frac{\partial P_1(t,\tau)}{\partial \tau} d\tau - \int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau,$$

$$\int_0^t \frac{\partial P_2(t,\tau)}{\partial t} d\tau = -\int_0^t \frac{\partial P_2(t,\tau)}{\partial \tau} d\tau - \int_0^t \lambda_1(\tau)P_2(t,\tau)d\tau,$$

which imply

$$\int_0^t \frac{\partial P_1(t,\tau)}{\partial t} d\tau = -[P_1(t,t) - P_1(t,0)] - \int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau,$$

$$\int_0^t \frac{\partial P_2(t,\tau)}{\partial t} d\tau = -[P_2(t,t) - P_2(t,0)] - \int_0^t \lambda_1(\tau)P_2(t,\tau)d\tau.$$

Hence, we can rewrite (5.10) as

$$\frac{dn_1(t)}{dt} = P_1(t,0) - \int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau, \tag{5.11a}$$

$$\frac{dn_2(t)}{dt} = P_2(t,0) - \int_0^t \lambda_2(\tau)P_2(t,\tau)d\tau. \tag{5.11b}$$

Using the boundary conditions (5.7), the last equations become

$$\frac{dn_1(t)}{dt} = \int_0^t \lambda_2(\tau)P_2(t,\tau)d\tau - \int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau, \tag{5.12a}$$

$$\frac{dn_2(t)}{dt} = \int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau - \int_0^t \lambda_2(\tau)P_2(t,\tau)d\tau. \tag{5.12b}$$

---

[1] This from the rule

$$\frac{d}{dt}\int_0^t p(t,\tau)d\tau = \frac{dt}{dt}p(t,t) - \frac{d0}{dt}p(t,0) + \int_0^t \frac{\partial p(t,\tau)}{\partial t} d\tau.$$

If $\lambda_1(\tau), \lambda_2(\tau)$ are independent of time i.e. constants, we get the two-state Markov chain with continuous time

$$\frac{dn_1(t)}{dt} = \lambda_2 n_2(t) - \lambda_1 n_1(t), \tag{5.13a}$$
$$\frac{dn_2(t)}{dt} = \lambda_1 n_1(t) - \lambda_2 n_2(t). \tag{5.13b}$$

The solution of these equations as illustrated in example 2.2 (see chapter 2) is

$$n_1(t) = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \Big(n_1(0) - \frac{\lambda_2}{\lambda_1 + \lambda_2}\Big)e^{-(\lambda_1+\lambda_2)t},$$
$$n_2(t) = \frac{\lambda_1}{\lambda_1 + \lambda_2} + \Big(n_2(0) - \frac{\lambda_1}{\lambda_1 + \lambda_2}\Big)e^{-(\lambda_1+\lambda_2)t}.$$

In our assumption to get the non-Markovian process we need the case where the hazard function $\lambda_i(\tau)$ is a function of τ, so we get

$$Pr\{\tau < T_i < \tau + d\tau \mid T_i > \tau\} = \lambda_i(\tau)d\tau.$$

Therefore

$$\phi_i(\tau) = \lambda_i(\tau)\Psi_i(\tau), \tag{5.14}$$

from the definition of survival function (4.1) we have, $\frac{d\Psi(t)}{dt} = -\phi(t)$. Thus

$$\frac{d\Psi_i(\tau)}{d\tau} = -\lambda_i(\tau)\Psi_i(\tau).$$

This has a solution

$$\Psi_i(\tau) = exp\left(\int_0^\tau -\lambda_i(s)ds\right). \tag{5.15}$$

Inserting (5.15) into (5.14), the waiting time density will be

$$\phi_i(\tau) = \lambda_i(\tau)exp\left(\int_0^\tau -\lambda_i(s)ds\right). \tag{5.16}$$

Now our task is to solve equations (5.9) in order to find the master equations for the conditional probabilities P_1, P_2 and the master equations for the probabilities (5.12). Assume $\tau = \tau(t)$, so $P_1(t, \tau) = P_1(t, \tau(t))$. Then

$$\frac{dP_1}{dt} = \frac{\partial P_1}{\partial t} + \frac{\partial P_1}{\partial \tau}\frac{d\tau}{dt},$$

to get equation (5.9) we must have $\frac{d\tau}{dt} = 1$, therefore

$$\frac{dP_1}{dt} = \frac{\partial P_1}{\partial t} + \frac{\partial P_1}{\partial \tau} = -\lambda_1(\tau)P_1(t, \tau).$$

This equation has a solution

$$P_1(t,\tau(t)) = P_1(t_0,\tau_0)exp\left(-\int_0^t \lambda_1(\tau)d\tau\right).$$

However, we have two cases for $\frac{d\tau}{dt} = 1$ (see [94])

Case 1 $\tau(t) = t - t_0$ **if** $\tau < t$

$$P_1(t,\tau(t)) = P_1(t_0,0)exp\left(-\int_{t_0}^t \lambda_1(\tau(s))ds\right),$$
$$P_1(t,\tau) = P_1(t-\tau,0)exp\left(-\int_0^\tau \lambda_1(s)ds\right).$$

Case 2 $\tau(t) = t + \tau_0$ **if** $\tau > t$

$$P_1(t,\tau(t)) = P_1(0,\tau_0)exp\left(-\int_{\tau_0}^\tau \lambda_1(\tau(s))ds\right),$$
$$P_1(t,\tau) = P_1(0,\tau-t)exp\left(-\int_{\tau-t}^\tau \lambda_1(s)ds\right).$$

Then

$$P_1(t,\tau) = P_1(t-\tau,0)exp\left(-\int_0^\tau \lambda_1(s)ds\right) + P_1(0,\tau-t)exp\left(-\int_{\tau-t}^\tau \lambda_1(s)ds\right). \quad (5.17)$$

Similarly, for $P_2(t,\tau)$, we have

$$P_2(t,\tau) = P_2(t-\tau,0)exp\left(-\int_0^\tau \lambda_2(s)ds\right) + P_2(0,\tau-t)exp\left(-\int_{\tau-t}^\tau \lambda_2(s)ds\right). \quad (5.18)$$

By using the definition of the survival function (5.15), the last two equations are equivalent to

$$\begin{aligned} P_1(t,\tau) &= P_1(t-\tau,0)\Psi_1(\tau) + P_1(0,\tau-t)\Psi_1(t), \\ P_2(t,\tau) &= P_2(t-\tau,0)\Psi_2(\tau) + P_2(0,\tau-t)\Psi_2(t). \end{aligned}$$

They give the conditional probability at state σ_1, σ_2 when the age or waiting time at these states is τ. Substituting them in the boundary condition (5.7), we get

$$\begin{aligned} P_1(t,0) &= \int_0^t \lambda_2(\tau)P_2(t-\tau,0)\Psi_2(\tau)d\tau + \int_0^t \lambda_2(\tau)P_2(0,\tau-t)\Psi_2(t)d\tau, \\ P_2(t,0) &= \int_0^t \lambda_1(\tau)P_1(t-\tau,0)\Psi_1(\tau)d\tau + \int_0^t \lambda_1(\tau)P_1(0,\tau-t)\Psi_1(t)d\tau. \end{aligned}$$

Applying the initial condition $P_i(0, \tau - t) = P_i(0)\delta(\tau - t)$, and the property of dirac delta function, we obtain

$$P_1(t, 0) = \int_0^t \lambda_2(\tau)P_2(t - \tau, 0)\Psi_2(\tau)d\tau + \lambda_2(t)P_2(0)\Psi_2(t), \tag{5.19a}$$

$$P_2(t, 0) = \int_0^t \lambda_1(\tau)P_1(t - \tau, 0)\Psi_1(\tau)d\tau + \lambda_1(t)P_1(0)\Psi_1(t). \tag{5.19b}$$

The LHS of the first equation means the probability of being at state σ_1 at time t and there is no waiting time, in other words, it just arrives. The first term of the RHS is equal to the summation of the transition rate from state σ_2 to state σ_1 multiplied by the probability of being at state σ_2 at the intermediate time $t-\tau$, and the probability of no jump during the age τ. Also the second term of RHS is the transition rate multiplied by the initial probability of starting from state σ_2 multiplied by the survival function. Using (5.16), we can rewrite equations (5.19) as

$$P_1(t, 0) = \int_0^t P_2(t - \tau, 0)\phi_2(\tau)d\tau + P_2(0)\phi_2(t), \tag{5.20a}$$

$$P_2(t, 0) = \int_0^t P_1(t - \tau, 0)\phi_1(\tau)d\tau + P_1(0)\phi_1(t). \tag{5.20b}$$

Consequently, (5.12) becomes

$$\frac{dn_1(t)}{dt} = P_1(t, 0) - P_2(t, 0) \tag{5.21a}$$

$$\frac{dn_2(t)}{dt} = P_2(t, 0) - P_1(t, 0). \tag{5.21b}$$

Then we need to solve these equations in order to find the probability $n_i(t)$. First, we assume $P_i(0) = n_i(0)$, then transfer (5.20) and (5.21) to Laplace transforms

$$\begin{aligned}
\tilde{P}_1(s, 0) &= \tilde{P}_2(s, 0)\tilde{\phi}_2(s) + n_2(0)\tilde{\phi}_2(s), \\
&= [\tilde{P}_2(s, 0) + n_2(0)]\tilde{\phi}_2(s), \\
\frac{\tilde{P}_1(s, 0)}{\tilde{\phi}_2(s)} &= \tilde{P}_2(s, 0) + n_2(0) && (5.22) \\
\tilde{P}_2(s, 0) &= \tilde{P}_1(s, 0)\tilde{\phi}_1(s) + n_1(0)\tilde{\phi}_1(s), \\
&= [\tilde{P}_1(s, 0) + n_1(0)]\tilde{\phi}_1(s), \\
\frac{\tilde{P}_2(s, 0)}{\tilde{\phi}_1(s)} &= \tilde{P}_1(s, 0) + n_1(0) && (5.23) \\
s\tilde{n}_1(s) - n_1(0) &= \tilde{P}_1(s, 0) - \tilde{P}_2(s, 0), && (5.24) \\
s\tilde{n}_2(s) - n_2(0) &= \tilde{P}_2(s, 0) - \tilde{P}_1(s, 0). && (5.25)
\end{aligned}$$

Inserting (5.23) in (5.24), we obtain

$$s\tilde{n}_1(s) = \frac{\tilde{P}_2(s,0)}{\tilde{\phi}_1(s)} - \tilde{P}_2(s,0),$$

this equation gives

$$\tilde{P}_2(s,0) = \tilde{K}_1(s)\tilde{n}_1(s). \tag{5.26}$$

Similarly, if we insert (5.22) into (5.25), it follows

$$\tilde{P}_1(s,0) = \tilde{K}_2(s)\tilde{n}_2(s). \tag{5.27}$$

Now we use the new values we got for $\tilde{P}_1(s,0), \tilde{P}_2(s,0)$ in (5.24) and (5.25), therefore

$$\begin{aligned} s\tilde{n}_1(s) - n_1(0) &= \tilde{K}_2(s)\tilde{n}_2(s) - \tilde{K}_1(s)\tilde{n}_1(s), \\ s\tilde{n}_2(s) - n_2(0) &= \tilde{K}_1(s)\tilde{n}_1(s) - \tilde{K}_2(s)\tilde{n}_2(s). \end{aligned}$$

Inverting the Laplace transform gives the master equations for the probabilities of the two-state non-Markovian process

$$\frac{dn_1(t)}{dt} = \int_0^t K_2(t-\tau)n_2(\tau)d\tau - \int_0^t K_1(t-\tau)n_1(\tau)d\tau \tag{5.28a}$$

$$\frac{dn_2(t)}{dt} = \int_0^t K_1(t-\tau)n_1(\tau)d\tau - \int_0^t K_2(t-\tau)n_2(\tau)d\tau. \tag{5.28b}$$

5.3.2 Multi-States Non-Markovian Process

To get the master equation of the probability for the non-Markovian multi-states process, we first apply the previous method for the case of three states, then we can generalize it for multi-states N. Consider the following equations for $P_i(t,\tau)$

$$\frac{\partial P_i}{\partial t} + \frac{\partial P_i}{\partial \tau} = -\lambda_i(\tau)P_i(t,\tau), \quad i = 1,2,3. \tag{5.29}$$

The initial conditions are

$$P_i(0,\tau) = P_i(0)\delta(\tau), \quad i = 1,2,3, \tag{5.30}$$

while the boundary conditions will be

$$P_1(t,0) = h_{21}\int_0^t \lambda_2(\tau)P_2(t,\tau)d\tau + h_{31}\int_0^t \lambda_3(\tau)P_3(t,\tau)d\tau, \tag{5.31a}$$

$$P_2(t,0) = h_{12}\int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau + h_{32}\int_0^t \lambda_3(\tau)P_3(t,\tau)d\tau, \tag{5.31b}$$

$$P_3(t,0) = h_{13}\int_0^t \lambda_1(\tau)P_1(t,\tau)d\tau + h_{23}\int_0^t \lambda_2(\tau)P_2(t,\tau)d\tau, \tag{5.31c}$$

where h_{ij} satisfies (4.2). To solve (5.29), we assume as before $\frac{d\tau}{dt} = 1$. This implies

$$\frac{dP_i}{dt} = \frac{\partial P_i}{\partial t} + \frac{\partial P_i}{\partial \tau}\frac{d\tau}{dt} = \lambda_i(\tau)P_i(t,\tau),$$

which has the solution

$$P_i(t,\tau) = P_i(t_0,\tau_0)exp\left(-\int_0^t \lambda_i(\tau)d\tau\right).$$

Likewise, applying the two cases of τ regarding t: $\tau > t$ and $\tau < t$, we obtain

$$\begin{aligned} P_i(t,\tau) &= P_i(t-\tau,0)exp\left(-\int_0^\tau \lambda_i(\tau)d\tau\right) + P_i(0,\tau-t)exp\left(-\int_{\tau-t}^\tau \lambda_i(\tau)d\tau\right) \\ &\quad or \\ P_i(t,\tau) &= P_i(t-\tau,0)\Psi_i(\tau) + P_i(0,\tau-t)\Psi_i(t). \end{aligned} \tag{5.32}$$

Using the initial conditions (5.30) we can rewrite the last equation as

$$P_i(t,\tau) = P_i(t-\tau,0)\Psi_i(\tau) + P_i(0)\delta(\tau-t)\Psi_i(t). \tag{5.33}$$

This gives the formula for the conditional probability at time t after age τ spending at the state i, while the boundary condition (5.31) can be rewritten as

$$\begin{aligned} P_1(t,0) = h_{21}\Big(\int_0^t \lambda_2(\tau)P_2(t-\tau,0)\Psi_2(\tau)d\tau + P_2(0)\lambda_2(t)\Psi_2(t)\Big) \\ +h_{31}\Big(\int_0^t \lambda_3(\tau)P_3(t-\tau,0)\Psi_3(\tau) + P_3(0)\lambda_3(t)\Psi_3(t)\Big). \end{aligned}$$

By using the definition of the waiting time given by equation (5.16)

$$P_1(t,0) = h_{21}\left[\int_0^t P_2(t-\tau,0)\phi_2(\tau)d\tau + P_2(0)\phi_2(t)\right]+h_{31}\left[\int_0^t P_3(t-\tau,0)\phi_3(\tau)d\tau + P_3(0)\phi_3(t)\right].$$

or

$$P_1(t,0) = \sum_{k\neq 1}^{3} h_{k1}\left[\int_0^t P_k(t-\tau,0)\phi_k(\tau)d\tau + P_k(0)\phi_k(t)\right],$$

and so on; for any state i the master equation for the transition probability when age $\tau = 0$ is

$$P_i(t,0) = \sum_{k\neq i}^{3} h_{ki}\left[\int_0^t P_k(t-\tau,0)\phi_k(\tau)d\tau + P_k(0)\phi_k(t)\right]. \tag{5.34}$$

Generalize the master equation for N states

The previous result (5.34) can be generalized for any number of states N to get

$$\begin{aligned} P_j(t,0) &= \sum_{k\neq j}^{N}\left[\int_0^t P_k(t-\tau,0)\phi_k(\tau)h_{kj}d\tau + P_k(0)\phi_k(t)h_{kj}\right], \\ &= \sum_{k\neq j}^{N}\int_0^t P_k(t-\tau,0)\phi_k(\tau)h_{kj}d\tau + \phi_i(t)h_{ij}, \end{aligned} \tag{5.35}$$

such that i is the initial state, so $P_k(0) = 1$ only when $k = i$ and zero otherwise. If we compare (5.35) with (4.3), we find they are the same, i.e. when age $\tau = 0$ conditional probability $P_j(t,0)$ what we aim to find it in this chapter is the conditional arrival probability $J_{ij}(t)$ we introduced in a previous chapter. Because when age equal zero that means the process has just arrived at state j at time t. However, $P_j(t,\tau)$ means the process arrives at state j at time t and stays there for age τ. Now to find the probability $n_j(t)$, we recall equation (5.11) for state j

$$\frac{dn_j(t)}{dt} = P_j(t,0) - \int_0^t \lambda_j(\tau)P_j(t,\tau)d\tau. \quad j = 1,2,...,N$$

Inserting (5.33) in the last equation gives

$$\begin{aligned} \frac{dn_j(t)}{dt} &= P_j(t,0) - \int_0^t \lambda_j(\tau)P_j(t-\tau,0)\Psi_j(\tau)d\tau - \int_0^t \lambda_j(\tau)P_j(0)\delta(t-\tau)\Psi_j(t)d\tau \\ &= P_j(t,0) - \int_0^t P_j(t-\tau,0)\phi_j(\tau)d\tau - P_j(0)\phi_j(t). \end{aligned} \tag{5.36}$$

To find the value of $n_j(t)$ we write equations (5.35) and (5.36) in Laplace form, considering $P_j(0) = n_j(0)$

$$\tilde{P}_j(s,0) = \sum_{k\neq j}^{N}\tilde{P}_k(s,0)\tilde{\phi}_k(s)h_{kj} + \sum_{k\neq j}^{N}\tilde{\phi}_k(s)h_{kj}n_k(0), \tag{5.37}$$

$$\begin{aligned} s\tilde{n}_j(s) - n_j(0) &= \tilde{P}_j(s,0) - \tilde{P}_j(s,0)\tilde{\phi}_j(s) - n_j(0)\tilde{\phi}_j(s), \\ s\tilde{n}_j(s) &= \tilde{P}_j(s,0)[1-\tilde{\phi}_j(s)] + n_j(0)[1-\tilde{\phi}_j(s)], \\ \tilde{n}_j(s) &= \tilde{P}_j(s,0)\tilde{\Psi}_j(s) + n_j(0)\tilde{\Psi}_j(s). \end{aligned} \tag{5.38}$$

From the last equation we find a new value for $\tilde{P}_j(s,0)$, which is

$$\tilde{P}_j(s,0) = \frac{\tilde{n}_j(s)}{\tilde{\Psi}_j(s)} - n_j(0). \tag{5.39}$$

Inserting (5.39) into (5.37) gives

$$\tilde{P}_j(s,0) = \sum_{k\neq j}^{N} \frac{\tilde{n}_k(s)}{\tilde{\Psi}_k(s)}\tilde{\phi}_k(s)h_{kj} = \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj}.$$

Back to equation (5.38), this implies

$$\tilde{n_j}(s) = \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj}\tilde{\Psi}_j(s) + n_j(0)\tilde{\Psi}_j(s), \tag{5.40}$$

In the case of waiting times is the same at all states

$$\tilde{n_j}(s) = \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{\phi}(s)h_{kj} + n_j(0)\tilde{\Psi}(s), \tag{5.41}$$

which in the time domain is

$$n_j(t) = \sum_{k\neq j}^{N} \int_0^t \phi(\tau)n_k(t-\tau)h_{kj}d\tau + n_j(0)\Psi(t). \tag{5.42}$$

Equations (5.40) and (5.41) are the same result as in chapter 4 (4.14) and (4.15) for the transition probability $\tilde{P}_{ij}(s)$. Now to get the differential form of the master equation, we rearrange equation (5.40) in such a way

$$\frac{\tilde{n_j}(s)}{\tilde{\Psi}_j(s)} - n_j(0) = \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj}.$$

Then we move the first term in the left-hand side to the right, and we add $s\tilde{n}_j(s)$ to both sides, thus

$$s\tilde{n}_j(s) - n_j(0) = s\tilde{n}_j(s) - \frac{\tilde{n}_j(s)}{\tilde{\Psi}_j(s)} + \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj}.$$

This is equivalent to

$$s\tilde{n}_j(s) - n_j(0) = -\frac{\tilde{\phi}_j(s)}{\tilde{\Psi}_j(s)}\tilde{n}_j(s) + \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj},$$

by using the definition of the kernel function $\tilde{K}_j(s) = \frac{\tilde{\phi}_j(s)}{\tilde{\Psi}_j(s)}$

$$s\tilde{n}_j(s) - n_j(0) = -\tilde{K}_j(s)\tilde{n}_j(s) + \sum_{k\neq j}^{N} \tilde{n}_k(s)\tilde{K}_k(s)h_{kj}. \tag{5.43}$$

In the time domain, it yields

$$\frac{dn_j(t)}{dt} = -\int_0^t K_j(\tau)n_j(t-\tau)d\tau + \sum_{k\neq j}^{N} \int_0^t K_k(\tau)n_k(t-\tau)h_{kj}. \tag{5.44}$$

5.4 Example of The Hazard Function

5.4.1 Markovian Case

If we choose the hazard function as a function of the age τ we will get a different type of waiting time density. In order to get a Markovian model, the hazard function must be a constant, i.e. $\lambda_j(\tau) = \lambda_j$. Accordingly, the waiting time distribution, given by equation (5.16), takes the form of exponential distributions

$$\phi_j(t) = \lambda_j e^{-\lambda_j t}.$$

Consequently, the conditional probability when the age τ equals zero $P_j(t, 0)$, which is given by (5.35), and the waiting time is the same at all states, is

$$P_j(t,0) = \sum_{k\neq j}^{N} \lambda \int_0^t P_k(t-\tau,0)e^{-\lambda\tau} h_{kj} d\tau + \lambda e^{-\lambda t} h_{ij}, \quad j = 1, 2, ..., N$$

while the conditional probability when the age $0 < \tau \leq t$ is given by (5.33)

$$P_j(t,\tau) = P_j(t-\tau,0)e^{-\lambda\tau} + P_j(0)\delta(\tau - t)e^{-\lambda t}.$$

Finally, the integral form of the master equation for the probability (5.42)

$$n_j(t) = \sum_{k\neq j}^{N} \lambda \int_0^t e^{-\lambda\tau} n_k(t-\tau) h_{kj} d\tau + n_j(0)e^{-\lambda t},$$

while the differential form (5.44) is

$$\frac{dn_j(t)}{dt} = -\lambda n_j(t) + \lambda \sum_{k\neq j}^{N} n_k(t) h_{kj}. \tag{5.45}$$

This equation is equivalent to (4.25), which is derived by using the conditional arrival probability with exponential waiting time.

5.4.2 Non-Markovian Case

Now, to have a non-Markovian model we need the waiting time density to be in a form of power-law. To get a power-law form we may choose the hazard function to be a decreasing function of τ in the form such as

$$\lambda(\tau) = \frac{\beta}{\alpha + \tau}, \quad \alpha, \beta > 0.$$

Consequently, the waiting time density obtaining from (5.16) when it is the same at all states will be in the form of power-law distribution

$$\phi(t) = \frac{\beta\alpha^{\beta}}{(\alpha + t)^{\beta+1}}.$$

Therefore, the conditional probability when the age τ equals zero $P_j(t, 0)$ is

$$P_j(t,0) = \sum_{k\neq j}^{N} \int_0^t \frac{\beta\alpha^{\beta}}{(\alpha+\tau)^{\beta+1}} P_k(t-\tau,0) h_{kj} d\tau + \frac{\beta\alpha^{\beta}}{(\alpha+\tau)^{\beta+1}} h_{ij}, \quad j = 1, 2, ..., N$$

Besides, the conditional probability when the age $0 < \tau \leq t$ is $P_j(t, \tau)$, given as

$$P_j(t,\tau) = P_j(t-\tau,0)\frac{\alpha^{\beta}}{(\alpha+\tau)^{\beta}} + P_j(0)\delta(\tau - t)\frac{\alpha^{\beta}}{(\alpha+t)^{\beta}},$$

and the master equation for the probability is

$$n_j(t) = \sum_{k\neq j}^{N} \int_0^t \frac{\beta\alpha^{\beta}}{(\alpha+\tau)^{\beta+1}} n_k(t-\tau) h_{kj} d\tau + n_j(0)\frac{\alpha^{\beta}}{(\alpha+t)^{\beta}}.$$

The asymptotic behavior of power-law waiting PDF when $t \to \infty$ is given by $\phi(t) \sim (\tau_0/t)^{\beta+1}$, where τ_0 is a parameter with units of time. Consequently, the Laplace transform $\tilde{\phi}(s)$ is approximated by

$$\tilde{\phi}(s) \sim 1 - (\tau_0 s)^{\beta} \quad \text{For small} \quad s \quad \text{and} \quad 0 < \beta < 1.$$

Therefore, the kernel function is

$$\tilde{K}(s) = \frac{s\tilde{\phi}(s)}{1 - \tilde{\phi}(s)} \sim s(\tau_0 s)^{-\beta}.$$

Inserting the kernel function in (5.43), to get the master equation

$$s\tilde{n}_j(s) - n_j(0) = -\frac{s^{(1-\beta)}}{\tau_0^{\beta}}\tilde{n}_j(s) + \sum_{k\neq j}^{N} \frac{s^{(1-\beta)}}{\tau_0^{\beta}}\tilde{n}_k(s) h_{kj},$$

$$s^{\beta}\tilde{n}_j(s) - s^{\beta-1} n_j(0) = -\frac{1}{\tau_0^{\beta}}\tilde{n}_j(s) + \frac{1}{\tau_0^{\beta}}\sum_{k\neq j}^{N} \tilde{n}_k(s) h_{kj},$$

In the time domain, it will be

$$\frac{d^{\beta}}{dt^{\beta}} n_j(t) = -\frac{1}{\tau_0^{\beta}} n_j(t) + \frac{1}{\tau_0^{\beta}} \sum_{k\neq j}^{N} n_k(t) h_{kj}. \tag{5.46}$$

In this chapter we derived the master equation of the probability $n_j(t)$ defined in (5.5). The derivation in this chapter used the conditional probability $P_i(t, \tau)$, which is the probability to be at state i at instant t and wait there for age τ, while in chapter 4 we used the conditional arrival probability $J_{ij}(t)$, which is the probability of arriving state j from state i. The master equations we obtain in this chapter for the probability $n_j(t)$ are similar to the master equations we obtained in the previous chapter for the transition probability $P_{ij}(t)$.

In this chapter we consider $n_j(t)$ is the probability of the volatility $\sigma(t)$ to be at state j at instant t. This probability shows the memory effect, either in the integral form (5.42) or in the differential form (5.44). The memory effect exists because the transition rate (hazard function) between the values of volatility is a function of time. Different forms of the hazard function lead us to different forms of waiting time distribution. Constant hazard function is the only form that leads to exponential waiting time distribution then to the Markovian master equation. Otherwise, any form of the waiting time distribution leads us to the non-Markovian master equation. The memory effect appears in non-Markovian master equation via kernel function, which does not exist in the Markovian case. The kernel function can be found from the waiting time density, and the waiting time density is found from the hazard function (5.16).

Chapter 6

Conclusion

In this work we derived non-Markovian models for volatility. In chapter 4, we introduced the theory of conditional arrival probability. We used this theory to derive the master equation when the time is continuous and the space is discrete. The master equation takes the form (4.19) in the Laplace domain, and the form (4.20) in the time domain. We also derived the different forms of the master equations corresponding to different waiting time PDF (4.25), (4.33), (4.40), (4.43). The master equation is usually non-Markovian except when waiting time has exponential PDF. Then we generalized the theory of conditional arrival probability for the case of continuous time and continuous state. We got the Laplace form of the process PDF equivalent to the Montroll-Wiess equation, while the master equation takes the Laplace form (4.53) and the time domain form (4.54). Similarly, we derived the different forms of the master equations corresponding to different waiting time PDF and jump PDF (4.58), (4.62), (4.64), (4.65). When the waiting time is exponentially distributed, the process has a Markovian nature wherever the jump PDF is thin-tailed (Gaussian distribution) or heavy-tailed (Lévy distribution). In the case of power-law waiting time, we got a fractional differential master equation which can be written in the form of a Caputo fractional derivative or a Riemann-Liouville fractional derivative. In addition, the master equation is a time fractional derivative when the waiting time PDF is power-law and the jump PDF is Gaussian (4.62), while the master equation is a time-space fractional derivative when waiting time PDF is power-law and the jump PDF is Lévy (4.65).

We applied the master equation of continuous time discrete state (4.20) to the counting process, in order to get the distribution of the process and the first two moments when it is Markovian or non-Markovian. In the Markovian case of the counting process, we got the Poisson process (4.68), and the first two moments are proportional to time. In the non-Markovian case, we found the Laplace distribution form of the counting process when the waiting time has gamma distribution (4.72) or power-law distribution (4.73). Moreover, the first two moments are combined of linear function and exponentially decreasing function of time in the case of gamma distributed waiting time, while the first two moments are non-linear function of time and decay very slowly as a hyperbolic rare. Furthermore, numerical results were given to illustrate the difference between the counting process with different waiting time PDF. At the end of chapter 4, we derived the stationary distribution of the switching model when it has waiting time with finite mean and waiting time with infinite mean. In chapter 5, we generalized CTRW to model stochastic volatility. We derived the master equation when the waiting time and volatility are independent of each other (5.4). The master equation shows the memory effects which appears as the kernel function. To define the kernel function we need to find the waiting time distribution. Hence, we introduced the age model which illustrates how to find the waiting time distribution from the hazard function (5.16). First, we derived the master equation of the probability when volatility takes two states (5.28) by using the age model in the Markovian and the non-Markovian case. When the process is Markovian we get two-state Markov chain (5.13). Then we generalized our derivation when volatility has multi-states, and it is non-Markovian. The master equation of the probability takes the integral form (5.42) and the differential form (5.44). At the end of chapter 5, we gave an example of hazard function to get exponential waiting time density in order to get the Markovian model (5.45), as well as an example of hazard function to get a power-law waiting time in order to get a non-Markovian model (5.46). Our future works are going to apply these models to empirical data.

Bibliography

[1] K. Amin and V. Ng. Option valuation with systematic stochastic volatility. *Journal of Finance*, 48:881–910, 1993.

[2] J. Anderson. *Continuous-Time Markov Chains*. Springer, 1991.

[3] T. Anderson and T. Bollerslev. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance*, 52:975–1005, 1997.

[4] E. Auerll and S. Simdyankin. Pricing risky option simply. *Journal of Theoretical and Applied Finance*, 1:1–23, 1998.

[5] M. Avellaneda, A. Levy, and A. Paras. Pricing and hedging derivative securities in markets with uncertain volatility. *Applied Mathematical Finance*, 1(2):73–88, 1995.

[6] G. Bakshi, C. Cao, and Z. Chen. Pricing and hedging long-term option. *Journal of Econometrics*, 94:277–318, 2000.

[7] C. Ball and A. Roma. Stochastic volatility option pricing. *Journal of Financial and Quantitative Analysis*, 29(4):589–607, 1994.

[8] E. Barkai, R. Metzler, and J. Klafter. From continuous time random walk to the fractional fokkerplanck equation. *Physical Review E*, 61(1):132–138, 2000.

[9] E. Barki and I.M. Sokolov. Multi-point distribution function for the continuous time random walk. *Journal of Statistical Mechanics*, 2007, August 2007.

[10] M. Baxter and A. Rennie. *Financial Calculus: An Introduction To Derivative Pricing*. Cambridge University Press, 1996.

[11] J. Beran. *Statistics For Long-Memory Processes*. Chapman and Hall, 1994.

[12] R. Billie. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73:5–59, 1996.

[13] F. Black. Studies in stock price volatility changes. *Proceeding of the 1976 meeting of Business and Economic statistics, American Statistical Association*, pages 177–181, 1976.

[14] F. Black and M. Scholes. The pricing of option and corporate liabilities. *the Journal of Political Economy*, 81(5), 1973.

[15] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[16] T. Bollerslev. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized Arch model. *Review of Economics and Statistics*, 72:489–505, 1990.

[17] T. Bollerslev, R. Engle, and J. Wooldridge. A capital asset pricing model with time varying covariance. *Journal of Political Economy*, 96:116–131, 1988.

[18] T. Bollerslev and H. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73:151–184, 1996.

[19] T. Bollerslev and H. Mikkelsen. Long-term equity anticipation securities and stock market volatility dynamics. *Journal of Econometrics*, 92:75–99, 1999.

[20] L. Borland. Option pricing formula based on a non-Gaussian stock price model. *Quantitative Finance*, 2:415–431, 2002.

[21] J. Bouchaud and D. Sornette. The black-scholes option pricing problem in mathematical finance: Generalization and extensions for large class of stochastic process. *J. Phys*, 4(6):863–881, 1994.

[22] G. Box and G. Jenkins. *Time Series Analysis: Forecasting And Control*. Holden-Day; San Francisco, 1976.

[23] M. Caporin and G. Associati. Stationary, memory and parameter estimation of FIGARCH models. Technical report, GRETA, Venice, 2003.

[24] P. Carmona, L. Coutin, and G. Montseny. Stochastic integration with respect to fractional brownian motion. *Ann.I.H.Poincare*, Pr39:27–68, 2003.

[25] A. Carpinteri and F. Mainardi. *Fractals And Fractional Calculus in Continuum Mechanics*. Springer Verlag, 1997.

[26] C. Chen, F. Liu, I. Turner, and V. Anh. A Fourier method for the fractional diffusion equation describing sub-diffusion. *Journal of Computational Physics*, 277:886–897, 2007.

[27] J. Cohen. *Analysis Of Random Walks*. Ios Press, 1992.

[28] F. Comte and E. Renault. Long memory continuous time models. *Journal of Econometrics*, 73:101–149, 1996.

[29] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

[30] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 2 of *Partial Differential Equations*. Wiley, 1989.

[31] C. Cox and H. Miller. *The Theory of Stochastic Process*. Chapman and Hall, 1965.

[32] J. Cox and S. Ross. The valuation of option for alternative stochastic process. *Journal of Financial Economics*, 3:145–166, 1976.

[33] J. Davidson. Bootstrap tests for fractional cointegration: A reappraisal of the relationship between government popularity and economic performance in the UK. *Econometric Society*, (0324), 2000.

[34] Z. Ding, C. Granger, and R. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.

[35] J. Duan. The Garch option pricing model. *Mathematical Finance*, 5:13–32, 1995.

[36] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, 1996.

[37] T. Duncan, Y. Hu, and B. Duncan. Stochastic calculus for fractional brownian motion i:theory. *Siam J Control*, 38:582–616, 1999.

[38] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *The Econometric Society*, 50(4):987–1007, 1982.

[39] R. Engle and T. Bollerslev. Modeling the persistence of conditional variance. *Econometric Reviews*, 5:1–50, 1986.

[40] R. Engle and C. Mustafa. Implied Arch models from option prices. *Journal of Econometrics*, 52:289–311, 1992.

[41] R. Engle and V. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5):1749–1778, 1993.

[42] A. Etheridge and M. Davis. *Theory of Speculation*. Princeton, NJ. ; Oxford : Princeton University Press, 2006.

[43] E. Fama. The behavior of stock market prices. *The Journal of Business*, 38:34–105, 1965.

[44] S. Fedotov and S. Mikhailov. Option pricing for incomplete markets via stochastic optimization: Transaction costs, adaptive control and forecast. *International Journal of Theoretical and Applied Finance*, 4:179–195, 2001.

[45] S. Fedotov and A. Tan. Long memory stochastic volatility in option pricing. *International Journal of Theoretical and Applied Finance*, 8(3):381–392, 2005.

[46] C. Gardiner. *Handbook of Stochastic Methods For Physics*. Springer, 2004.

[47] M. Garman. A general theory of asset valuation under diffusion state process. *Center For Research in Management Science*, Technical Report 50, 1976.

[48] D. Gkamas. Stochastic volatility and option pricing. *Quantitative Finance*, 1:292–297, 2001.

[49] R. Gorenflo and F. Mainardi. Fractional calculus. *Lecture Notes, Department of Mathematics and Information, University of Berlin*, pages 223–276, 2000.

[50] R. Gorenflo, F. Mainardi, D. Moretti, and P. Paradisi. Time fractional diffusion: A discrete random walk approach. *Nonlinear Dynamics*, 29:129–143, 2002.

[51] R. Gorenflo, F. Mainardi, and A. Vivoli. The asymptotic universality of the Mittage-Leffler waiting time law in continuous time random walk. $Http : //Anotrans.Physik.Hu - Berlin.De/Talks/Gorenflo_Bad - Honnef$06.$Pdf$, 2006.

[52] R. Gorenflo, F. Mainardi, and A. Vivoli. Continuous-time random walk and parametric subordination in fractional diffusion. *Chaos, Solitons and Fractals*, 34:87–103, 2007.

[53] C. Granger and Z. Ding. Varieties of long memory models. *Journal of Econometrics*, 73:61–77, 1996.

[54] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford Science Publications, 1982.

[55] C. Guo. Option pricing with stochastic volatility following a finite Markov chain. *International Review of Economic and Finance*, 7(4):407–415, 1998.

[56] S. Herzel. Option pricing with stochastic volatility models. *Decision in Economics and Finance*, 23:75–99, 2000.

[57] S. Heston. A closed form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Study*, 6(2):327–343, 1993.

[58] S. Heston and S. Nandi. A closed form GARCH option valuation model. *Review of Future Market*, 13(3):585–625, 2000.

[59] D. Higham. *An Introduction To Financial Option Valuation*. Cambridge University Press, 2004.

[60] N. Hofmann, E. Platen, and M. Schweizer. Option pricing under incompleteness and stochastic volatility. *Mathematical Finance*, 2(3):153–187, 1992.

[61] Y. Hu and B. Oksendal. Fractional white noise calculus and application to finance. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 6(1):1–32, 2003.

[62] J. Hull. *Option, Futures, And Other Derivatives*. Pearson Prentice Hall, 2006.

[63] J. Hull and A. White. The pricing of option on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.

[64] A. Jurlewicz, A. Wylomanska, and P. Zebrowski. Financial data analysis by means of coupled continuous time random walk in Rachev-Ruschendrof model. *Acta Physica Polinica A*, 114:629–635, 2008.

[65] J. Kemeny and J. Snell. *Finite Markov Chains*. Princeton (N.J.), London, 1960.

[66] M. Kijima. *Markov Processes For Stochastic Modeling*. Chapman And Hall, 1997.

[67] K. Kim, J. Chaoi, S-M. Yoon, and K. Chang. Dynamical volatilities for yen-dollar exchange rates. *Journal of Korean Physical Society*, 44(4):1015–1018, 2004.

[68] C. Lamoureux and W. Lastrapes. Forecasting stock return variance: Toward an understanding of stochastic implied volatility. *Review of Financial Studies*, 6:293–326, 1993.

[69] Y. Luchko and R. Gorenflo. An operational method for solving fractional equation with caputo derivatives. *Acta Mathematica Vietnamica*, 24(2):207–233, 1999.

[70] C. Ma. Long memory continuous-time correlation models. *Applied Probability Trust*, 40(4):1133–1146, 2003.

[71] F. Mainardi. The time fractional diffusion-wave equation. *Radiophysics And Quantum Electronics*, 38:1–12, 1995.

[72] F. Mainardi and G. Pagnini. The Wright function as solution of the time-fractional diffusion equation. *Applied Mathematics and Computation*, 141:51–62, 2003.

[73] F. Mainardi, G. Pagnini, and R. Gotenflo. Some aspects of fractional diffusion equations of single and distributed order. *Applied Mathematics and Computation*, 187:295–305, 2007.

[74] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4):394, 1963.

[75] B. Mandelbrot. Forecasts of future prices, unbiased markets, and 'martingale'models. *The Journal of Business*, 39(1):242–255, 1966.

[76] R. Mantegna and H. Stanley. Stochastic process with ultraslow convergence to a Gaussian: The truncated lévy flight. *Phys. Rev. Lett.*, 73(22):2946–2949, 1994.

[77] R. Mantegna and H. Stanley. *An Introduction To Econophysics Correlation and Complexity in Finance*. Cambridge University Press, 200.

[78] J. Masoliver, M. Montero, J. Perello, and G. Weiss. The continuous time random walk formalism in financial markets. *Journal of Economic Behavior and Organization*, 61:577–598, 2006.

[79] J. Masoliver, M. Montero, J. Perello, and G. H. Weiss. The CTRW in finance: Direct and inverse problems with some generalizations and extensions. *Physica A*, 379:151–167, 2007.

[80] J. Masoliver, M. Montero, and J.M. Porra. A dynamical model describing stock market price distribution. *Physica A*, 283:559–567, 2000.

[81] J. Masoliver, M. Montero, and G. H. Weiss. Continuous time random walk model for financial distribution. *Physica Review E*, 67(021112):1–10, 2003.

[82] J. Medhi. *Stochastic Processes*. Wiley Eastern Limited, 1982.

[83] M. Meerschaert and E Scalas. Coupled continuous time random walks in finance. *Physica A*, 370:114–118, 2006.

[84] A. Meershaert and E. Scalas. Coupled continuous time random walk in finance. *Physica A*, 370:114–118, 2006.

[85] R. Mendes and M. Oliveria. Option pricing with fractional volatility. *The SAO/NASA Astrophysics Data system*, eprint arXiv:cond-mat/0404684(2004 cond.mat..4684V), 2004.

[86] V. Mendes. A fractional calculus interpretation of the fractional volatility model. *Springer Science + Business Media*, 55:359–399, 2009.

[87] V. Mendez, S. Fedotov, and W. Horsthemke. *Reaction Transport System*. Springer, 2010.

[88] R. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

[89] R. Merton. Option pricing when underlying stock return are discontinuous. *Journal of Financial Economics*, 3:152–144, 1976.

[90] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: A fractional dynamics approach. *Physics Report*, 339:1–77, 2000.

[91] R. Metzler, J. Klafter, and I. Sokolov. Anomalous transport in external fields: Continuous time random walks and fractional diffusion equation extend. *Physical Review E*, 58(2):1621–1633, 1998.

[92] E. Montroll and G.H. Weiss. Random walk on latices ii. *Journal of Mathematical Physics*, 6(2):167–181, 1965.

[93] A. Mura. *Thesis: Non-Markov Stochastic Processes And Their Applications: From Anomalous Diffusion To Time Series Analysis*. PhD thesis, Bologna University, 2008.

[94] J. Murray. *Mathematical Biology I*. Springer, Berlin, 2003.

[95] C. Necula. Option pricing in financial Brownian motion environment. *Draft version, Academy of Economic Studies, Bucharest, Romania*, 2002.

[96] S. Neftci. *An Introduction To The Mathematics of Financial Derivatives*. Academic Press, 1996.

[97] D. Nelson. Stationarity and persistence in the Garch(1,1). *Cambridge University Press*, 6(3):318–334, 1990.

[98] J. Norris. *Markov Chains*. Cambridge University Press, 1997.

[99] D. Ostry. Synthesis of accurate fractional Gaussian noise. *Ieee Transaction of Information Theory*, 52:1609–1623, 2006.

[100] S. Panayides. *Derivative Pricing And Hedging For Incomplete Market; Stochastic Arbitrage And An Adaptive Procedure For Stochastic Volatility*. PhD thesis, The School of Mathematics, University of Manchester, 2005.

[101] W. Paul and J. Baschna. *Stochastic Processes : From Physics To Finance*. Springer, 1999.

[102] I. Podlubny. *Fractional Differential Equation*. Academic Press, San Diego, 1999.

[103] N. Prabhu. *Stochastic Process*. The Macmillan, New York, 1965.

[104] M. Riedle. Lecture notes f stochastic modeling in finance. *The University of Manchester*, 2009.

[105] R. Ritchey. Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13:285–296, 1990.

[106] M. Rubinstein. Displaced diffusion option pricing. *Journal of Finance*, 38:213–217, 1983.

[107] E. Scalas, R. Gorenflo, and F. Mainardi. Fractional calculus and continuous-time finance i. *Physica A*, 284:376–384, 2000.

[108] E. Scalas, R. Gorenflo, and F. Mainardi. Uncoupled continuous-time random walks: Solution and limiting behavior of the master equation. *Physical Review*, E69(011107):1–8, 2004.

[109] E. Scalas, R. Gorenflo, F. Mainardi, and M. Raberto. Fractional calculus and continuous-time finance iii. *Mathematical Finance*, 5, 2001.

[110] L. Scott. Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis*, 22(4):419–438, 1987.

[111] K. Sircar and G. Papanicolaou. Stochastic volatility, smile and asymptotic. *Applied Mathematical Finance*, 6:107–145, 1999.

[112] M. Spiegel. *Schaum's Outline of Theory And Problems of Laplace Transforms*. Mcgraw-Hill, 1995.

[113] E. Stein and J. Stein. Stock price distributions with stochastic volatility: An analytic approach. *The Review of Financial Studies*, 4(4):727–752, 1991.

[114] A. Tan. *Long Memory Stochastic Volatility And A Risk Minimization Approach For Derivative Pricing And Hedging*. PhD thesis, The School of Mathematics, University of Manchester, 2006.

[115] S. Taylor. *Modeling Financial Time Series*. Wiley, Chichester, 1986.

[116] S. Taylor. Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4:183–204, 1994.

[117] G. Weiss. *Aspects And Applications of The Random Walk*. Elsevier Science, 1994.

[118] U. Wiersma. *Brownian Motion Calculus*. John Wiely And Sons, Ltd, 2008.

[119] J. Wiggins. Option values under stochastic volatility, theory and empirical estimation. *Journal of Financial Economics*, 19:351–372, 1987.

[120] P. Wilmott, S. Howison, and J. Dewynne. *The Mathematics of Financial Derivatives: A Student Introduction*. Cambridge University Press.

[121] S-M. Yoon, J. Chaoi, C. Lee, M-K. Yum, and K. Kim. Dynamical volatilities for yen-dollar exchange rates. *Physica A*, 359:569–575, 2006.