Technische Universität München

Zentrum Mathematik

Pearson diffusions with jumps

Diplomarbeit

von

Mathias Frederik Schmidt

Themenstellerin:	Prof. Dr. C. Klüppelberg
Betreuer:	Prof. Dr. M. Sørensen
Abgabetermin:	01.Juni 2008

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 30.April 2008

Acknowledgment

I would like to take this opportunity to thank the people how supported me during my years of study.

First of all I would like to thank Professor Doctor Claudia Klüppelberg for giving me the opportunity to write my thesis in Copenhagen. Without her support this thesis would not have been possible.

Furthermore, I am most thankful to my supervisor Professor Doctor Michael Sørensen, who had always time for me and helped my throughout this thesis. I am indebted to him for his huge patience and his great job.

My special thanks goes to Professor Doctor Martin Jacobsen for teaching me a lot about solutions to stochastic differential equations and much more.

Especially, I am deeply greatful for my lovely parents, who never hesitated to support my aims, in particular the choice of my place of study. Without you, I would have never been able to come so far.

Last but not least I want to thank David Drennan, Christoph Ferstel, Julia Huus Schmidt and Christopher Hörster for reading my the thesis and searching for proof-reading my thesis.

Contents

Introduction

In this thesis we will a take a closer look at a particular type of diffusion. The Pearson diffusions are stationary solutions to a stochastic differential equation for the form

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}dW_t$$

where $\theta > 0$ and a, b and c are chosen in such a way, that the square root is well defined. In particular, μ is the mean level of the process. We will see in chapter 4, that the drift and the diffusion coefficient are of course crucial for the existence of a solution (see chapter 3). Under certain conditions to the scale and speed density we can derive the existence of a solution. Since the Pearson diffusions are closed under linear transformation, we gain six different types of Pearson diffusions. Fortunately in all six possible cases for the Pearson diffusion is the existence of a solution guaranteed.
For this particular type of diffusions it is possible to derive a recursive formula for the moments and the conditional expectations. In this context we will see that the *generator* and its *eigenfunctions* play an important role. We will see that eigenfunctions of Pearson diffusions are polynomials and therefore especially comfortable. Further can we assume that the parameter for the highest order is equal to one. Thus can we derive a recursive formula for the conditional moments.
Another interesting fact about Pearson diffusions is, that the maximum likelihood estimator is not the most simple estimator to calculate. Instead, with the knowledge about their conditional moments we can derive an optimal estimating function, which has got some comfortable properties and converges under certain assumption.
The main part of this thesis deals with the extension of those processes by adding jumps. Assume that $\mathcal{N}_t$ is a Poisson process with a constant parameter λ, then $\mathcal{N}_t$ is the number of jumps in the time interval $(0, t)$. Further we postulate that the random variable J_k is distributed with distribution $F_{X_{\tau_k-}}$ where X_{τ_k-} is the last state of the process before the kth jump. Then our new process looks like

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta(aX_s^2 + bX_s + c)}dW_s + \sum_{n=1}^{N_t} g(J_n, X_{n-}), \quad t \geq 0$$

with g being a function with certain properties (see chapter 6). Then, under certain conditions, the existence of a unique solution can be proved.
The existence of the moments on the other hand is rather tricky. Under the assumption of a global Lischitz condition (which is satisfied in some cases), we can prove the existence of the first two moments. Since you can not apply the global Lipschtiz condition for the

proof of higher moments, as we will apply it for the proof of the first two moments, we have to find other conditions under which the proof of the existence of higher moments is possible. Those assumptions are nevertheless quiet similar.
Under the assumption of the existence of moments, we can derive an ordinary differential equation to calculate the conditional moments of X_t. We gain as in the case without jumps a recursive formula for the conditional moments.
Similar to the case without jumps we can derive a recursive formula for the eigenfunctions. In this case this recursive formula is of course a bit more complicated than in the easier case, but we can clearly see the connection to the jump less case. At first we need to derive the generator first to gain the eigenfunctions of the process. Coincidentally the eigenfunctions are again polynomials like in the jump less case.

Chapter 1

Estimating functions

Suppose as a model for the data $X_1, X_2, \ldots, X_n \in \Psi \subset \mathbb{R}^n$ that they are observations from a stochastic process indexed by a parameter $\theta \in \Theta \subset \mathbb{R}^p$. We define an *estimating function*, as a p-dimensional function of the parameter θ and the data, as follows

$$G_n(\theta; X_1, X_2, \ldots, X_n) : \Theta \times \Psi \to \mathbb{R}.$$

For ease of notation we write $G_n(\theta)$. From the estimation function we can derive an estimator for θ by solving the equation

$$G_n(\theta) = 0.$$

Depending on the situation, there could be more than one solution as well as none at all. We call an estimating function *unbiased* if and only if $E_\theta(G_n(\theta)) = 0$.

1.1 Martingale estimating functions

An estimating function $G_n(\theta)$ which is also a martingale, i.e. an estimating function satisfying

$$E_\theta(G_n(\theta)|\mathcal{F}_{n-1}) = G_{n-1}(\theta), \quad n = 1, 2, \ldots$$

where $\mathcal{F}_{n-1}$ is the σ-field generated by the observations $X_1, \ldots, X_{n-1}$ ($G_0 = 0, \mathcal{F}_0 = \{\Omega, \emptyset\}$), have in particular nice properties and relatively simple asymptotic theory based on well-developed martingale limit theory as summarized in the Appendix. As we will see later on, the score function is usually a martingale estimating function. In the case where it is difficult to calculate the score function, we want to approximate the score function by much simpler martingale estimating functions. The following example is a typical one

$$G_n(\theta) = \sum_{i=1}^{n} a(X_{i-1}; \theta)\big(f(X_i) - E_\theta(f(X_{i-1})|\mathcal{F}_{i-1})\big) \tag{1.1}$$

where f is a suitable function of the data, while a (typically a matrix) is chosen to ensure that the dimension of $G_n(\theta)$ equals the dimension of the parameter and to optimise the estimator.

Proposition 1.1. *The estimating function in (1.1) is in fact a martingale, if f is measurable, $E_\theta(f(X_i)) < \infty$ and all $a(x;\theta) < \infty$.*

Proof.

$$\begin{aligned} E_\theta(G_n(\theta) - G_{n-1}(\theta)|\mathcal{F}_{n-1}) =& a(X_{n-1};\theta)E_\theta\big(\big(f(X_n) - E_\theta(f(X_{n-1})|\mathcal{F}_{n-1})\big)|\mathcal{F}_{n-1}\big) \\ =& a(X_{n-1};\theta)\big(E_\theta(f(X_{n-1})|\mathcal{F}_{n-1}) \\ & - E_\theta(E_\theta(f(X_{n-1})|\mathcal{F}_{n-1})|\mathcal{F}_{n-1})\big) \\ =& 0 \end{aligned}$$

which is with the linearity of the expectation equivalent to $E_\theta(G_n(\theta)|\mathcal{F}_{n-1}) = G_{n-1}(\theta)$. Furthermore

$$E(|G_n(\theta)|) = \sum_{i=1}^{n} \underbrace{a(X_{i-1};\theta)}_{<\infty}\big(\underbrace{E(f(X_i))}_{<\infty} - \underbrace{E(E_\theta(f(X_{i-1})|\mathcal{F}_{i-1}))}_{<\infty}\big) < \infty$$

and since f is measurable and the conditional expectation is also measurable (definition), we gain that the estimating function in (1.1) is also measurable. □

The function a is usually called an *instrument* and it is discussed in Bibby, Jacobsen & Sørensen (2004), chapter 4, how it should be chosen to be optimal. Now we want to quote a version of the central limit theorem for martingales.

We can write $G_n(\theta)$ in the form $G_n(\theta) = \sum_{i=1}^{n} H_i(\theta)$ with $H_i(\theta) = G_i(\theta) - G_{i-1}(\theta)$. If we assume that $G_n(\theta)$ has variance, then we can define the *quadratic characteristic* of $G_n(\theta)$ as the random positive semi-definite $p \times p$-matrix

$$\langle G(\theta)\rangle_n = \sum_{i=1}^{n} E_\theta\big(H_i(\theta)H_i(\theta)^T|\mathcal{F}_{i-1}\big).$$

Theorem 1.2. *Assume that for $n \to \infty$*

$$\frac{1}{n}\sum_{i=1}^{n} E_\theta\big(H_i(\theta)H_i(\theta)^T\big) \to \Sigma_\theta \tag{1.2}$$

$$\frac{\langle G(\theta)\rangle_n}{n} \xrightarrow{P_\theta} \Sigma_\theta \tag{1.3}$$

and

$$\frac{1}{\sqrt{n}}\sup_{i\leq n}|H_i(\theta)| \xrightarrow{P_\theta} 0, \tag{1.4}$$

where Σ_θ is a positive definite $p \times p$-matrix. Then

$$\langle G(\theta)\rangle_n^{-\frac{1}{2}} G_n(\theta) \xrightarrow{\mathcal{D}} N(0, I_p). \tag{1.5}$$

Here I_p denotes the $p \times p$ identity matrix. A proof of a one-dimensional version of the theorem can be found in Hall & Heyde (1980). The multivariate version follows by the Cramér-Wold device. You can see from the conditions of the theorem, that we have a weak law of large numbers

$$n^{-1}G_n(\theta) \xrightarrow{P_\theta} 0,$$

which is a necessary condition for consistency.

1.2 Limit results for diffusion processes

Let us now consider (mainly one-dimensional) diffusion models, which are solutions of stochastic differential equations (short SDEs) of the form

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \tag{1.6}$$

where W_t is a standard Wiener process at time point t. We assume that the drift b and the diffusion coefficient σ depend on the parameter θ which varies in a subset $\Theta \subset \mathbb{R}^p$. The parameter θ will not be discussed in this section. We also assume that b and σ are smooth enough to garantee the existence of a solution for all $\theta \in \Theta$ (see chapter 3 for the conditions on b and σ). The *state space* of X is the interval (l, r).
In condition 4.1 you can find a condition to ensure that the solution X_t is ergodic.
We consider X_t is ergodic with invariant probability measure π_θ. We further assume that $X_0 \sim \pi_\theta$ (i.e. X_0 has the same distribution like π_θ), so that X is a stationary process with $X_t \sim \pi_\theta$ for all $t \geq 0$. Therefore the density of the distribution of (X_t, X_{t+s}) for $t, s > 0$ is given by

$$Q_\theta^s(x,y) = \pi_\theta(x)p(s,x,y;\theta),$$

where $y \mapsto p(s,x,y;\theta)$ is the transition density, i.e. the conditional density of X_{t+s} given that $X_t = x$. For a function $f : (l,r)^2 \to \mathbb{R}$, we use the notation

$$Q_\theta^s(f) = \int_{(l,r)^2} f(x,y)p(s,x,y;\theta)\pi_\theta(x)dydx,$$

assuming that the integral exists. Similarly we define for a function $f : (l,r) \mapsto \mathbb{R}$

$$\pi_\theta(f) = \int_l^r f(x)\pi_\theta(x)dx.$$

Suppose that a unique weak ergodic solution of (1.6) exists and that $f : (l,r) \to \mathbb{R}$ satisfies $\pi_\theta(|f|) < \infty$, and that $g : (l,r)^2 \to \mathbb{R}$ satisfies $Q_\theta^\Delta(|g|) < \infty$ for a $\Delta > 0$. Then

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{i\Delta}) \overset{a.s.}{\to} \pi_\theta(f) \tag{1.7}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta}) \overset{a.s.}{\to} Q_\theta^\Delta(g) \tag{1.8}$$

as $n \to \infty$, see Billingsley (1961b). Note that the first result is a particular case of the second one also and that equidistant observations, i.e. $t_i = \Delta i$, are required to ensure the stationarity of $f(X_{i\Delta})$ and $g(X_{(i-1)\Delta}, X_{i\Delta})$.
Further, if we assume the sum $\sum\limits_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta})$ to be a martingale, i.e. that

$$\int_l^r g(x,y)p(\Delta,x,y;\theta)dy = 0$$

for all $x \in (l, r)$, with finite variance $Q_\theta^\Delta(g^2)$, then under Condition 4.1 from Chapter 4

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta}) \xrightarrow{\mathcal{D}} N\big(0, Q_\theta^\Delta(g^2)\big)$$

as $n \to \infty$, see Billingsley (1961a).

Chapter 2

Estimating functions for diffusions

In this chapter will take a more intensive look at estimating functions for diffusion models. We assume $X_{t_0}, \ldots, X_{t_n}$ to be observations of a stochatical differential equation like (1.6). Here we also assume that b and σ are known apart from the parameter θ and that they are smooth enough to ensure the existence of a unique weak solution for all θ in Θ. From now on we will denote by $p(s, x, y, \theta)$ ($y \to p(s, x, y, \theta)$) the transition density, i.e. the conditional density of X_{t+s} given that $X_t = x$ ($s > 0$).

2.1 Maximum likelihood estimation

The diffusion process X is a Markov process, therefore we can derive the likelihood function (conditional on X_0) as

$$L_n(\theta) = \prod_{i=1}^{n} p(t_i - t_{i-1}, X_{t_{i-1}}, X_{t_i}; \theta).$$

The vector $U_n(\theta) = \frac{\partial \log L_n(\theta)}{\partial \theta}$ is called the *score function*. To get a maximum likelihood estimator, you solve the equation $U_n(\theta) = 0$. Therefore we affiliate

$$U_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \tag{2.1}$$

based on the observations $X_{t_0}, X_{t_1}, \ldots, X_{t_n}$ and with $\Delta_i = t_i - t_{i-1}$.

Proposition 2.1. *The score function (2.1) is a martingale estimating function if we assume that the interchanging of differentiation and intergration is allowed.*

Proof. First we want to show that

$$E(U_n(\theta) - U_{n-1}(\theta) | X_{t_1}, \ldots, X_{t_{n-1}}) = 0.$$

Let $n \in \mathbb{N}$, then we gain

$$\begin{aligned} E(U_n(\theta) - U_{n-1}(\theta)|X_{t_1}, \ldots, X_{t_{n-1}}) &= E_\theta(\partial_\theta \log p(\Delta_n, X_{t_{n-1}}, X_{t_n}; \theta)|X_{t_1}, \ldots, X_{t_{n-1}}) \\ &= E_\theta(\frac{\partial_\theta p(\Delta_n, X_{t_{n-1}}, X_{t_n}; \theta)}{p(\Delta_n, X_{t_{n-1}}, X_{t_n}; \theta)}|X_{t_{n-1}}) \\ &= \int_l^r \frac{\partial_\theta p(\Delta_n, X_{t_{n-1}}, y; \theta)}{p(\Delta_n, X_{t_{n-1}}, y; \theta)} p(\Delta_n, X_{t_{i-1}}, y; \theta) dy \\ &= \partial_\theta \underbrace{\int_l^r p(\Delta_n, X_{t_{n-1}}, y; \theta) dy}_{=1} = 0. \end{aligned}$$

With $\mathcal{F}_n = \sigma\{X_{t_s}|0 \leq s \leq n\}$ we get $U_n(\theta)$ is $\mathcal{F}_n$-measurable. $\square$

Under weak regularity conditions the maximum likelihood estimator is efficient. Since the transiton density is only rarely explicitly known, there are a number of numerical approaches that render likelihood inference feasible for diffusion models. Some of them are mentioned in Bibby, Jacobsen & Sørensen (2004), as well as some estimators which are not based on the score function.

2.2 Martingale estimating functions for diffusion models

The score function is a martingale estimating function of the form

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta). \tag{2.2}$$

If we want to approximate the score function, it is most natural to do this by martingale estimating functions of the general function (2.2) with

$$g(\Delta, x, y; \theta) = \sum_{j=1}^{N} a_j(\Delta, x; \theta) h_j(\Delta, x, y; \theta) \tag{2.3}$$

where $h_j(\Delta, x, y; \theta), j = 1, \ldots, N$ are given real valued functions satisfying

$$E_{p(\Delta,x,y;\theta)}(h_j) = \int_l^r h_j(\Delta, x, y; \theta) p(\Delta, x, y; \theta) dy = 0$$

for all $\Delta > 0, x \in (l, r)$ and $\theta \in \Theta$. It is most efficient to obtain estimators by combining the h_j's in an optimal way instead of defining each h_j as an estimating function of the form (2.2). It is straightforward how to choose the *weights* a_j (in the econometric literature also known as *instruments*) in an optimal way (see Bibby, Jacobsen & Søresen (2004), Chapter 4). The correct choice of the h_j, on the other hand, is rather difficult. The ability to tailor these functions to a given model or to particular parameters of interests is a considerable strength of the estimating functions methodology.

Martingale estimating functions have turned out to be very useful in obtaining estimators for discretely sampled diffusion-type models.
In the case of $N = 1$ the *linear estimating function* is a simple one and is given by

$$h_1(\Delta, x, y; \theta) = y - F(\Delta, x; \theta),$$

with

$$F(\Delta, x; \theta) = E_\theta(X_\Delta | X_0 = x) = \int_l^r y p(\Delta, x, y; \theta) dy.$$

Since the conditional expectation $F(\Delta, x; \theta)$ as well as the conditional variance $\phi(\Delta, x; \theta)$ are unknown in most models, they have to be determined by simulations which usually can be done easily.
Linear martingale estimating functions for diffusion models were used by Bibby & Sørensen(1995) to approximate the continuous time likelihood function. Its advantage is, that it is very robust to model misspecification as well as the esimator being consistent if only the first moment F of the transition distribution is correctly specified.
The linear estimating function is too simple to be useful, when the diffusion coefficient (the volatility) σ depends on a parameter, whereas the *quadratic estimating functions* are a natural applicable choice. In the one-dimensional casethey are given by $N = 2$ and

$$\begin{aligned} h_1(\Delta, x, y; \theta) &= y - F(\Delta, x; \theta) \\ h_2(\Delta, x, y; \theta) &= (y - F(\Delta, x; \theta))^2 - \phi(\Delta, x; \theta), \end{aligned}$$

where

$$\phi(\Delta, x; \theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x) = \int_l^r (y - F(\Delta, x; \theta))^2 p(\Delta, x, y; \theta) dy.$$

Analogously you can derive the version for multivariate diffusions.
In the case where Δ is small, the Gaussian density function with expectation $F(\Delta, x; \theta)$ and variance $\phi(\Delta, x; \theta)$ is a good approximation to the transition density p. If we insert this Gaussian density in (2.1) we get the approximate score function as follows

$$\begin{aligned} U_n(\theta) &= \sum_{i=1}^n \partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \\ &= -\sum_{i=1}^n \partial_\theta \Big(\frac{(X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)^2)}{2\phi(\Delta_i, X_{t_{i-1}}; \theta)} - \frac{1}{2} \log(\phi(\Delta_i, X_{t_{i-1}}; \theta)^{-1}) \Big) \\ &= \sum_{i=1}^n \Big(\frac{\partial_\theta F(\Delta_i, X_{t_{i-1}}; \theta)}{\phi(\Delta_i, X_{t_{i-1}}; \theta)} (X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)) \\ &\quad + \frac{\partial_\theta \phi(\Delta_i, X_{t_{i-1}}; \theta)}{2\phi(\Delta_i, X_{t_{i-1}}; \theta)^2} (X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta))^2 - \frac{\partial_\theta \phi(\Delta_i, X_{t_{i-1}}; \theta)}{2\phi(\Delta_i, X_{t_{i-1}}; \theta)} \Big) \\ &= \sum_{i=1}^n \Big(\frac{\partial_\theta F(\Delta_i, X_{t_{i-1}}; \theta)}{\phi(\Delta_i, X_{t_{i-1}}; \theta)} (X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)) \\ &\quad + \frac{\partial_\theta \phi(\Delta_i, X_{t_{i-1}}; \theta)}{2\phi(\Delta_i, X_{t_{i-1}}; \theta)^2} \big((X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta))^2 - \phi(\Delta_i, X_{t_{i-1}}; \theta) \big) \Big). \end{aligned}$$

If the first two moments of the transition density are correctly specified, the obtained estimator is consistent and are again rather robust to model misspecifications.

Example 2.2. Let us take a look at a *mean-reverting diffusion model*

$$dX_t = -\beta(X_t - \alpha)dt + \sigma(X_t)dW_t \tag{2.4}$$

where $\beta > 0$. Under weak conditions on σ it can be shown that

$$F(t, x; \alpha, \beta) = xe^{-\beta t} + \alpha(1 - e^{-\beta t}). \tag{2.5}$$

This can also be seen by the fact that $f(t) = F(t, x; \alpha, \beta)$ solves the ordinary differential equation $f' = -\beta(f - \alpha)$ (remember that $F(\Delta, x; \theta) = E_\theta(X_\Delta|X_0 = x)$ and $E(\int \sigma(X_t)dW_t) = 0$). Therefore the linear estimating function can be easily calculated.
If we consider $\sigma(x) = \tau\sqrt{x}$ ($\tau > 0$) we get the CIR process (proposed by Cox, Ingersoll, Jr & Ross (1985)). In this case $\phi(x; \alpha, \beta, \tau)$ can be derived explicitly

$$\phi(x; \alpha, \beta, \tau) = \frac{\tau^2}{\beta}\big((\frac{1}{2}\alpha - x)e^{-2\beta} - (\alpha - x)e^{-\beta} + \frac{1}{2}\alpha\big).$$

We can also derive the quadratic estimating function explicitly when we assume $\sigma^2 = \beta + x^2$ and $\alpha = 0$. Then we get

$$\phi(x; \beta) = x^2 e^{-2\beta}(e - 1) + \frac{\beta}{2\beta - 1}\big(1 - e^{1-2\beta}\big).$$

Δ

By choosing the h_js of the following form

$$h_j(\Delta, x, y; \theta) = f_j(y; \theta) - \xi^\theta_\Delta(f_j(\theta))(x) = f_j(y; \theta) - E_\theta(f_j(X_\Delta; \theta)|X_0 = x)$$

for suitable chosen functions f_j, we obtain a natural generalisation of the quadratic martingale estimating functions. We will refer to the functions f_j, $j = 1, \ldots, N$ as the *base* of the estimating function. It may be useful to choose a base that is tailored to a particular diffusion model, in particular if the tranisition density is far from Gaussian. A good example are estimating functions based on eigenfunctions of the generator of the diffusion (see chapter 2.3).
In most of the cases the conditional moments are unknown and therefore they have to be calculated numerically, if interested see Bibby, Jacobsen & Sørensen (2004).

2.3 Explicit estimating functions

In this section we consider a martingale estimating function of the form

$$G_n(\theta) = \sum_{i=1}^{n} a(\Delta_i, X_{t_{i-1}}, \theta)\big(f(X_{t_i}; \theta) - \xi^\theta_\Delta(f(\theta))(X_{t_{i-1}})\big) \tag{2.6}$$

with $f = (f_j)_{1\leq j\leq N}$ a (column) vector of given functions, the *base*, and $a = (a_{kj})_{1\leq k\leq p, 1\leq j\leq N}$ a $p \times N$-matrix of given functions, the *weights*. If all the f_j and a_{kj} are given explicitly *and* if the conditional expectations $\xi_\Delta^\theta(f(\theta))(x)$ can be determined explicitly, then we call $G_n(\theta)$ *explicit*.
The most simple case where $\xi_\Delta^\theta(f(\theta))(x)$ can be found explicitly, is where the base consists of eigenfunctions for the generator of the diffusion as proposed by Kessler & Sørensen (1999) for one-dimensional diffusions. In that case we take a look at the *eigenfunctions* of the *generator*. The generator is given by

$$A_\theta = b(x;\theta)\frac{d}{dx} + \frac{1}{2}\sigma^2(x;\theta)\frac{d^2}{dx^2}$$

and we call a twice differentiable function $\psi(x;\theta)$ *eigenfunction* for the generator if a real number $\lambda(\theta) \geq 0$ exists such that

$$A_\theta\psi(x;\theta) = -\lambda(\theta)\psi(x;\theta).$$

The number $\lambda(\theta)$ is called *eigenvalue*. Under weak regularity conditions, see Section 4.3,

$$\xi_\Delta^\theta(\psi(\theta))(x) = E_\theta(\psi(X_\Delta;\theta)|X_0 = x) = e^{-\lambda(\theta)\Delta}\psi(x;\theta)$$

and therefore we can derive

$$h_j(\Delta, x, y;\theta) = \psi_j(y;\theta) - e^{-\lambda(\theta)\Delta}\psi(x;\theta).$$

The special case of Pearson diffusions will be discussed in Section 4.3.

Chapter 3

Solutions to SDEs

Let us now examine a real-valued process X which is a solution to a stochastic differential equation (SDE) given by

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0. \tag{3.1}$$

We assume that b, σ are continuous functions on $(l, r) \subset \mathbb{R}$ and $\sigma > 0$,

$$P \bigcap_{t>0} (X_t \in (l, r)) = 1,$$

and in particular $x_0 \in (l, r)$.
It is our aim to find conditions on the functions b and σ to ensure that X stays away from the boundary point l and r even in the case where $l = -\infty$ or $r = +\infty$.
For any twice differentiable function $S : (l, r) \to \mathbb{R}$ we obtain with the Itô formula

$$dS(X_t) = AS(X_t)dt + S'(X_t)\sigma(X_t)dW_t$$

where A is the second order differential operator

$$Af(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x).$$

We want to achieve that $S(X_t)$ is a continuous local martingale (see Appendix for further information). Therefore we gain the equation $AS = 0$, i.e.

$$S''(X_t) = -\frac{2b(X_t)}{\sigma^2(X_t)}S'(X_t).$$

If we apply the theory for ordinary differential equations, we gather the solution given by

$$S'(x) = c \exp\left(-\int_{x_0}^{x} \frac{2b(u)}{\sigma^2(u)} du \right)$$

for some c. In particular, $S(x)$ is strictly increasing if $c > 0$. The function $S(x)$ is called the scale function for the diffusion X_t.

For a, b with the property $l < a < x_0 < b < r$, we define the stopping time $\tau_{ab} = \inf\{t : X_t = a \text{ or } X_t = b\}$. Since $(S(X_t))^{\tau_{ab}} = (S(X_{t\wedge\tau_{ab}}))$ is a bounded local martingale, it is a true martingale (see the Appendix for a proof). Therefore we can derive for each t

$$E_{x_0} S(X_{\tau_{ab}\wedge t}) = E_{x_0} S(X_0) = S(x_0)$$

and for $t \to \infty$, by dominated convergence

$$E_{x_0} S(X_{\tau_{ab}}) = S(x_0),$$

where

$$S(X_{\tau_{ab}}) = \begin{cases} S(b) & \text{if } \tau_b < \tau_a \\ S(a) & \text{if } \tau_a < \tau_b \\ \lim\limits_{t\to\infty} S(X_t) & \text{if } \tau_{ab} = \infty \end{cases}$$

exists by the martingale convergence theorem and $\tau_c := \inf\{t : X_t = c\}$ for $c \in (l, r)$.

Theorem 3.1. *If $l < a \leq x_0 \leq b < r$, we get*

$$P(\tau_{ab} < \infty) = 1$$

with τ_{ab} as defined above.

Proof. Since $\lim\limits_{t\to\infty} S(X_t)$ exists almost surely (a.s.). on $(\tau_{ab} = +\infty)$ and because S is strictly increasing and continuous, $\lim\limits_{t\to\infty} X_t$ also exists a.s. on $(\tau_{ab} = +\infty)$.
Let $\phi : [a, b] \to \mathbb{R}$ be continuous and let f denote the unique solution to

$$Af(x) = -\phi(x) \quad (a \leq x \leq b), \quad f(a) = f(b) = 0.$$

Then

$$\begin{aligned} df(X_t) =& Af(X_t)dt + f'(X_t)\sigma(X_t)dW_t \\ =& -\phi(X_t)dt + f'(X_t)\sigma(X_t)dW_t \end{aligned}$$

and therefore

$$\eta_t(f) := f(X_t) + \int_0^t \phi(X_s)ds$$

is a continuous, local martingale and especially $\eta(f)^{\tau_{ab}}$. Since

$$\sup_{s\leq t} |\eta_{\tau_{ab}\wedge s}(f)| \leq \sup_{a\leq x\leq b} |f(x)| + t \sup_{a\leq x\leq b} |\phi(x)| < \infty$$

, $\eta(f)^{\tau_{ab}}$ is a true martingale and in particular

$$\begin{aligned} f(x_0) =& E(\eta_0(f)) = E(\eta_{t\wedge\tau_{ab}}(f)) \\ =& E(f(X_{t\wedge\tau_{ab}})) + E(\int_0^{\tau_{ab}\wedge t} \phi(X_s)ds), \end{aligned}$$

which is equal to

$$E(\int_0^{\tau_{ab}\wedge t} \phi(X_s)ds) = f(X_0) - E(f(X_{t\wedge\tau_{ab}})).$$

We assumed f to be continuous, therefore $\lim_{t\to\infty} f(X_{t\wedge\tau_{ab}}) = f(X_{\tau_{ab}})$ exists a.s. and by dominated convergence we get

$$E(f(X_{\tau_{ab}})) = \lim_{t\to\infty} E(f(X_{t\wedge\tau_{ab}})).$$

Further, if we assume $\phi \geq 0$ or $\phi \leq 0$ (otherwise write $\phi = \phi^+ - \phi^-$ where $\phi^+ = \max(0, \phi)$ and $\phi^- = \max(0, -\phi)$), we get by the monotone convergence (i.e. the theorem by Beppo-Levi)

$$\lim_{t\to\infty} E(\int_0^{t\wedge\tau_{ab}} \phi(X_s)ds) = E(\lim_{t\to\infty} \int_0^{t\wedge\tau_{ab}} \phi(X_s)ds) = E(\int_0^{\tau_{ab}} \phi(X_s)ds)$$

so that for a continuous ϕ

$$E(\int_0^{\tau_{ab}} \phi(X_s)ds) = f(x_0) - E(f(X_{\tau_{ab}})).$$

Taking $\phi \equiv 1$ on $[a, b]$, this gives us

$$E(\tau_{ab}) = f_0(x_0) - E(f_0(X_{\tau_{ab}}))$$

where f_0 solves $Af_0 \equiv -1$, $f_0(a) = f_0(b) = 0$. Because the expression on the right hand side is finite, we get $E(\tau_{ab}) < \infty$ and in particular $P(\tau_{ab} < \infty) = 1$. Also for general ϕ we get

$$E(\int_0^{\tau_{ab}} \phi(X_s)ds) = f(x_0)$$

since $E(f(X_{\tau_{ab}})) = 0$ because $f(a) = f(b) = 0$. □

Now we can show that

$$S(x_0) = E(S(X_{\tau_{ab}})) = S(b)P(\tau_b < \tau_a) + S(a)P(\tau_a < \tau_b)$$

i.e.

$$P(\tau_b < \tau_a) = 1 - P(\tau_a < \tau_b) = \frac{S(x_0) - S(a)}{S(b) - S(a)}$$

, the basic formula.

Theorem 3.2. *Let S be an arbitrary scale function with*

$$S'(x) = c\exp(-\int_{x_0}^x \frac{2b(y)}{\sigma^2(y)}dy)$$

for some $c > 0$ and define $k : (l, r) \to \mathbb{R}_+$ by

$$k(x) = \frac{2}{\sigma^2(x)S'(x)}.$$

Then the unique solution f to $Af = -\phi$ on $[a, b], f(a) = f(b) = 0$, where ϕ is a given continuous function, is

$$f(x) = \int_a^b G_{ab}(x, y)\phi(y)k(y)dy$$

where G_{ab} is the Green function $G_{ab}(x, y) = G_{ab}(y, x)$ and

$$G_{ab}(x, y) = \frac{(S(x) - S(a))(S(b) - S(y))}{S(b) - S(a)} \quad (a \le x \le y \le b).$$

Proof. We get $f(a) = f(b) = 0$ easily from $G_{ab}(a, y) = G_{ab}(x, b) = 0$. If $x < z \in [a, b]$, we obtain

$$\begin{aligned} f(z) - f(x) =& \frac{S(x) - S(z)}{S(b) - S(a)} \int_a^x (S(y) - S(a))\phi(y)k(y)dy \\ &+ \frac{S(z) - S(x)}{S(b) - S(a)} \int_z^b (S(b) - S(y))\phi(y)k(y)dy \\ &+ \frac{1}{S(b) - S(a)} \int_x^z \Big((S(y) - S(a))(S(b) - S(z)) \\ &- (S(x) - S(a))(S(b) - S(y))\Big)\phi(y)k(y)dy. \end{aligned}$$

The integrand in the last term is

$$\le (S(z) - S(a))(S(b) - S(z)) - (S(x) - S(a))(S(b) - S(z)) = (S(z) - S(x))(S(b) - S(z))$$

and, by similar reasoning

$$\ge -(S(z) - S(x))(S(x) - S(a)).$$

Then the integral itself is

$$\le (S(z)-S(x))(S(b)-S(z)) \int_x^z \phi(y)k(y)dy, \quad \ge -(S(z)-S(x))(S(x)-S(a)) \int_x^z \phi(y)k(y)dy.$$

By reading, it follows that, by dividing the whole equation by $S(z) - S(x)$ and by taking the limit $z \to x$ (i.e. the last integral disappears), that

$$\begin{aligned} \frac{f'(x)}{S'(x)} =& - \frac{1}{S(b) - S(a)} \int_a^x (S(y) - S(a))\phi(y)k(y)dy \\ &+ \frac{1}{S(b) - S(a)} \int_x^b (S(b) - S(y))\phi(y)k(y)dy. \end{aligned}$$

By differentiating this equation with respect to (w.r.t.) x, we obtain

$$\begin{aligned} \left(\frac{f'}{S'}\right)'(x) =& - \frac{1}{S(b) - S(a)} \Big((S(x) - S(a))\phi(x)k(x) + (S(b) - S(x))\phi(x)k(x)\Big) \\ =& - \phi(x)k(x). \end{aligned}$$

On the other hand we get

$$\begin{aligned}\left(\frac{f'}{S'}\right)' =& \frac{1}{S'}\Big(f'' - (logS')'f'\Big)\\ =& \frac{k\sigma^2}{2}(f'' + \frac{2b}{\sigma^2}f')\\ =& kAf\end{aligned}$$

and thus $Af(x) = -\phi(x)$. □

The measure m on (l, r) with density k, $m(dx) = k(x)dx$ is called the *speed measure* for the diffusion X. Note that if the scale function S is replaced by $c_1 + c_2 S$ (where $c_1 \in \mathbb{R}, c_2 > 0$), k is replaced by $\frac{1}{c_2}k$.
We summarize the results obtained so far in the following theorem.

Theorem 3.3. *Let X be given by*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0$$

a (l, r)-valued diffusion, where b and $\sigma > 0$ are continuous. It holds for $a < x < b$, $a, b \in (l, r)$ that $P_x(\tau_{ab} < \infty) = 1$,

$$P_x(\tau_b < \tau_a) = \frac{S(x) - S(a)}{S(b) - S(a)},$$

and for $\phi : [a, b] \to \mathbb{R}$ bounded and measurable, that

$$E_x(\int_0^{\tau_{ab}} \phi(X_s)ds) = \int_a^b G_{ab}(x, y)\phi(y)k(y)dy$$

in particular

$$E_x(\tau_{ab}) = \int_a^b G_{ab}(x, y)k(y)dy.$$

In the formula above, S, given by (apart from an additive constant)

$$S'(x) = exp(-\int_{x_0}^x \frac{2b(y)}{\sigma^2(y)}dy)$$

for some $x_0 \in (l, r)$, is an arbitrary scale function and

$$k(x) = \frac{2}{\sigma^2(x)S'(x)}$$

is the corresponding speed measure density.

Proof. The statements are clear, see theorems before. □

Keep in mind that we have assumed that

$$P_x \bigcap_{t\geq 0} (X_t \in (l,r)) = 1,$$

i.e. that $\tau_r = \tau_l \equiv \infty$ P_x-a.s. The next results will tell us what the properties of S and k are that prevent X from reaching either of the boundaries l and r. Throughout S is a given scale density and k the matching density for the speed measure. Define

$$S(r) = \lim_{y\to r} S(y) \leq \infty \text{ and } S(l) = \lim_{y\to l} S(y) \geq -\infty.$$

Theorem 3.4. *(i) Either $S(r) = \infty$ or $\int_y^r (S(r) - S(z))k(z)dz = \infty$ $(y \in (l,r))$ and similarly $S(l) = -\infty$ or $\int_l^y (S(z) - S(l))k(z)dz = \infty$ $(y \in (l,r))$.*

(ii) For $a < x < b$ we gain

$$P_x(\tau_a < \infty) = \frac{S(r) - S(x)}{S(r) - S(a)}, \quad P_x(\tau_b < \infty) = \frac{S(x) - S(l)}{S(b) - S(l)},$$

in particular $P_x(\tau_y < \infty) > 0$ for all $x, y \in (l,r)$. Further, $P_x(\tau_a < \infty) = 1$ is true iff $S(r) = \infty$ and $P_x(\tau_b < \infty) = 1$ iff $S(l) = -\infty$.

(iii) If $S(r) < \infty$, then

$$\lim_{t\to\infty} X_t = r \quad P_x\text{-a.s. on } A_-,$$

where $A_- = \bigcup_{a:a<x} (\tau_a = \infty)$, $P_x(A_-) = \frac{S(x)-S(l)}{S(r)-S(l)}$, and if $S(l) > -\infty$, then

$$\lim_{t\to\infty} X_t = l \quad P_x\text{-a.s. on } A_+,$$

where $A_+ = \bigcup_{b:b>x} (\tau_b = \infty)$ and $P_x(A_+) = \frac{S(r)-S(x)}{S(r)-S(l)}$.

(iv) If $S(r) < \infty$ and $S(l) = -\infty$, then $P_x(\lim_{t\to\infty} X_t = r) = 1$ and if $S(r) = \infty$ and $S(l) > -\infty$, then $P_x(\lim_{t\to\infty} X_t = l) = 1$.

(v) If $S(r) < \infty$ and $S(l) > -\infty$, then

$$P_x(\lim_{t\to\infty} X_t = r) = 1 - P_x(\lim_{t\to\infty} X_t = l) = \frac{S(x) - S(l)}{S(r) - S(l)}.$$

(vi) If $S(r) = \infty$ and $S(l) = -\infty$, then X is recurrent in the sense that

$$P_x \bigcap_{y\in(l,r)} \bigcap_{t>0} \bigcup_{s>t} (X_s = y) = P_x \bigcap_{y\in(l,r)} \limsup_{t\to\infty} (X_t = y) = 1,$$

i.e. X hits any level infinitely often in any interval $[t,\infty), t \geq 0$.

Proof. Let $l < a < x < b < r$. For $b \nearrow r$, $\tau_b \nearrow \tau_r \equiv \infty$ (by assuming that X_t never hits r), since $P(\tau_b < \tau_a) + P(\tau_a < \tau_b) = 1$, we gain $1_{(\tau_b<\tau_a)} \to 1_{\{\tau_a=\infty\}}$ and therefore we get

$$\begin{aligned} P_x(\tau_a = \infty) =& \lim_{b\to r} P_x(\tau_b < \tau_a) = \lim_{b\to r} \frac{S(x) - S(a)}{S(b) - S(a)} \\ =& \frac{S(x) - S(a)}{S(r) - S(a)} \end{aligned}$$

and then

$$\begin{aligned} P_x(\tau_a < \infty) =& 1 - P_x(\tau_a = \infty) = 1 - \frac{S(r) - S(a)}{S(b) - S(a)} \\ =& \frac{S(r) - S(x)}{S(r) - S(a)}. \end{aligned}$$

That is the proof for (ii), because the other statements follow directly or analogously from this one.
If $S(r) < \infty$, we know from (ii) that $P_x(\tau_a = \infty) > 0$. Then we consider that $S(X_t)^{\tau_a}$is a bounded local martingale, hence a true martingale, so the r.v.

$$S(X_{\tau_a}) = \begin{cases} S(a) & \text{on } (\tau_a < \infty) \\ \lim_{t\to\infty} S(X_t) & \text{on } (\tau_a = \infty) \end{cases}$$

is well defined P_x-a.s. and satisfies

$$E_x(S(X_{\tau_a})) = S(x).$$

On the other hand

$$S(x) = E_x(S(X_{\tau_a})) = S(a)\frac{S(r) - S(x)}{S(r) - S(a)} + E_x(S(X_{\tau_a})1_{\tau_a=\infty})$$

implying that

$$E_x(S(X_{\tau_a})1_{\tau_a=\infty}) = S(r)P_x(\tau_a = \infty) = S(r)\frac{S(x) - S(a)}{S(r) - S(a)}.$$

Since $S(X_{\tau_a}) \le S(r)$, it follows that $S(X_{\tau_a}) = S(r)$ P_x-a.s. on $(\tau_a = \infty)$, i.e. $\lim_{t\to\infty} X_t = r$ P_x-a.s. on $(\tau_a = \infty)$ and (iii) follows since $P(\tau_a = \infty) \nearrow A_-$ as $a \searrow l$ so

$$P_x(A_-) = \lim_{a\searrow l} P_x(\tau_a = \infty) = \frac{S(x) - S(l)}{S(r) - S(l)}.$$

Now we can prove (i). If $S(r) < \infty$, then we get with (ii) that $\lim_{b\uparrow r} \tau_{ab} = \tau_a$ P_x-a.s. and so by monotone convergence

$$E_x(\tau_a) = \lim_{b\uparrow r} \int_a^b G_{ab}(x, y)k(y)dy.$$

But because of (ii), $P_x(\tau_a = \infty) > 0$, the left hand side is infinite. The right hand side is equal to

$$\begin{aligned}\lim_{b\uparrow r}\Big(\int_x^b \frac{(S(x)-S(a))(S(b)-S(y))}{S(b)-S(a)}k(y)dy + \int_a^x \frac{(S(y)-S(a))(S(b)-S(x))}{S(b)-S(a)}k(y)dy\Big) \\ = \frac{S(x)-S(a)}{S(r)-S(a)}\int_x^b (S(r)-S(y))k(y)dy + \frac{S(r)-S(x)}{S(r)-S(a)}\int_a^x (S(y)-S(a))k(y)dy\end{aligned}$$

with the last term finite, hence the first integral is positive infinite and (i) is proved.
The statements in (iv) and (v) follow directly from (i) and (iii).
It remains to establish (vi). From (iii) we know that $P_x(\tau_a < \infty) = P_x(\tau_b < \infty) = 1$ for all $a < x, b > x$. Let $a_n \searrow l$ and $b_n \nearrow r$, then $\tau_{a_n} \nearrow \infty$ and $\tau_{b_n} \nearrow \infty$ P_x-a.s. and between τ_{a_n} and τ_{b_n} X_t passes through all levels $y \in [a_n, b_n]$ since it is continuous. Now (vi) follows easily from this. □

Instead of starting with a solution of the SDE (3.1), assume we have given an open interval (l,r) and continuous functions $b : (l,r) \to \mathbb{R}$, $\sigma : (l,r) \to \mathbb{R}_+$, that satisfy the conditions from Theorem 3.4 (i), i.e.

$$\begin{array}{lll} S(r) = +\infty & \text{or} & \int_y^r (S(r)-S(z))k(z)dz = +\infty \\ S(l) = +\infty & \text{or} & \int_l^y (S(z)-S(l))k(z)dz = +\infty \end{array}$$

with $y \in (l,r)$ and where for some $x_0 \in (l,r)$

$$S'(x) = \exp(-\int_{x_0}^x \frac{2b(y)}{\sigma^2(y)}dy) \text{ and } k(x) = \frac{2}{\sigma^2(x)S'(x)}.$$

Theorem 3.5. *Let $(l,r), b, \sigma$ be as above, let W_t be a Wiener process in $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and let $U \in \mathcal{F}_0$ be a given (l,r)-valued r.v. Then the SDE*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 \equiv U,$$

has a unique solution, which is a diffusion. If $U \equiv x_0$, the distribution Π_{x_0} of X does not depend on the the choice of $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and W_t, and with an arbitrary initial condition $U \in \mathcal{F}_0$, the distribution of X_t is the mixture $\int_{(l,r)} \Pi_x P(U \in dx)$.

Proof. This follows from the classical charackterization of welldefined regular diffusions through their scale function and speed measure, see Freedman (1971). □

Let X again be a diffision on (l,r)

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad X_0 \equiv U,$$

with scale function S, speed measure density k, satisfying the critical conditions from Theorem 3.4 (i). As usual b, σ are continuous with $\sigma > 0$. The problem we shall now study is that of investigating whether there exists a probability π on (l,r), such that if U has distribution π, X_t is *stationary* for all t, and X_t has distribution π. The parameter π is also called an *invariant probablility* for X. If it exists, π is uniquely determined and typically, for all x, $p_t(x,.) \to \pi$ as $t \to \infty$, with $p_t(x,.)$ the transition probability from x.

Theorem 3.6. *The process X_t has an invariant probability π if and only if $\kappa := \int_l^r k(x)dx < \infty$ and in that case*

$$\pi(dx) = \frac{1}{\kappa} k(x)dx.$$

In particular, in order for the invariant probability to exist, it is necessary that X_t is recurrent, i.e. $S(r) = \infty$ and $S(l) = -\infty$.

Proof. Suppose first that the invariant probability π exists. Let φ denote the set of C^2-functions $f : (l, r) \to \mathbb{R}$ with compact support. By the Itô formula, for any $f \in \varphi$,

$$f(X_t) = f(X_0) + \int_0^t Af(X_s)ds + \int_0^t f'(X_s)\sigma(X_s)dW_s. \tag{3.2}$$

Since $f \in \varphi$, it follows that $f'\sigma$ is bounded and therefore the local martingale $M_t := \int_0^t f'(X_s)\sigma(X_s)dW_s$ is a true martingale with $M_0 = 0$. Writing $\pi(g) = \int g d\pi$ and taking conditional expectations we get, since π is invariant,

$$\begin{aligned} \pi(f) =& E_\pi(f(X_t)) = E_\pi(f(X_0)) + E_\pi(\int_0^t Af(X_s)ds) + \underbrace{E_\pi(\int_0^t f'(X_s)\sigma(X_s)dW_s)}_{=0} \\ =& \pi(f) + t\pi(Af), \end{aligned}$$

i.e. $\pi(Af) = 0$. But then we identify $\pi(dx) = u(x)dx$ and by using partial integration, we get

$$\begin{aligned} 0 &= \int_l^r (b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x))u(x)dx \\ &= \int_l^r (b(x)u(x) - \frac{1}{2}(\sigma^2 u)'(x))f'(x)dx. \end{aligned}$$

This holds for all $f \in \varphi$ iff

$$\frac{b(x)}{\sigma^2(x)}\big(\sigma^2(x)u(x)\big) - \frac{1}{2}\big(\sigma^2(x)u(x)\big)' \equiv 0,$$

i.e.

$$\sigma^2(x)u(x) = \exp(\int_{x_0}^x \frac{2b(y)}{\sigma^2(y)}dy),$$

and it follows that u is proportional to k.
If conversely $\int_l^r k(y)dy < \infty$, define $\pi(dx) = u(x)dx$, $u(x) = \frac{k(x)}{\int_l^r k(y)dy}$. Then certainly $\pi(Af) = 0$ for all $f \in \varphi$, the last term in the expansion of f in (3.2) is still a true martingale, and consequently, if X_0 has distibution π,

$$E_\pi(f(X_t)) = E_\pi(f(X_0)),$$

i.e. π is the invariant probability.
It remains to verify that if $\int_l^r k(y)dy < \infty$, then $S(r) = \infty, S(l) = -\infty$. But for an arbitrary $x \in (l, r)$,

$$\infty = \int_x^r (S(r) - S(y))k(y)dy \leq \int_l^r k(y)dy(S(r) - S(x))$$

so $\int_l^r k(y)dy < \infty$ forces $S(r) = \infty$. $\square$

Chapter 4

The Pearson diffusions

A stationary solution of a stochastic differential equation of the form

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}dW_t, \tag{4.1}$$

where $\theta > 0$ and where the square root is well defined, is called a *Pearson diffusion*. The parameter $\theta > 0$ is a scaling parameter of time which determines how fast the process moves. The other parameters (μ, a, b, c) determine the state space of the diffusion, the invariant distribution and especially, μ is the expectation of the diffusion.
We can obtain a lot of informations about the solution by looking at the scale and speed density which are defined as follows

$$s(x) = \exp(\int_{x_0}^{x} \frac{u-\mu}{au^2 + bu + c} du) \text{ and } m(x) = \frac{1}{s(x)(ax^2 + bx + c)}$$

where $ax_0^2 + bx_0 + c > 0$ with a fixed x_0. Assume that (l, r) is an interval such that $ax^2 + bx + c > 0$ for all $x \in (l, r)$.

Condition 4.1. *The following holds for a $x_0 \in (l, r)$*

$$\int_{l}^{x_0} s(x)dx = \infty, \quad \int_{x_0}^{r} s(x)dx = \infty$$

and

$$\int_{l}^{r} m(x)dx =: B < \infty.$$

It has been proved (see Chapter 3), that there exists a unique weak solution with an invariant probability measure to the stochastic differential equation (4.1) if and only if Condition 4.1 is satisfied. It also can be shown, that this soulution is ergodic. Its invariant density is given by

$$\pi(x) = B^{-1}m(x) \propto m(x), \quad x \in (l, r)$$

i.e. the invariant distribution is proportional to the speed density. Especially, it follows from

$$\frac{dm(x)}{dx} = -\frac{(2a+1)x - \mu + b}{ax^2 + bx + c} m(x), \tag{4.2}$$

that the invariant distribution belongs to the Pearson system, which is defined as the class of probability densities, which solve the differential equation (4.2).

4.1 Classification of the stationary solutions

Proposition 4.2. *The Pearson class of diffusions is closed under translations and scale-transformations.*

Proof. Let$(X_t)_{t\geq 0}$ be an ergodic Pearson diffusion with the stochastic differential equation (4.1) and $\tilde{X}_t = \gamma X_t + \delta$ with $\gamma \neq 0$. Then

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}dW_t$$

and with $X_t = \gamma^{-1}\tilde{X}_t - \delta$ we obtain

$$\frac{1}{\gamma}d\tilde{X}_t = -\theta(\frac{\tilde{X}_t - \delta}{\gamma} - \mu)dt + \sqrt{2\theta\big(a(\frac{\tilde{X}_t - \delta}{\gamma})^2 + b(\frac{\tilde{X}_t - \delta}{\gamma}) + c\big)}dW_t.$$

Then the last equation is equal to

$$d\tilde{X}_t = -\tilde{\theta}(\tilde{X}_t - \tilde{\mu})dt + \sqrt{2\tilde{\theta}(\tilde{a}\tilde{X}_t^2 + \tilde{b}\tilde{X}_t + \tilde{c})}dW_t$$

where $\tilde{\theta} = \theta, \tilde{\mu} = \gamma\mu + \delta, \tilde{a} = a, \tilde{b} = b\gamma - 2a\delta$ and $\tilde{c} = c\gamma^2 - b\gamma\delta + a\delta^2$. □

Therefore there are six cases of diffusions according to whether the squared diffusion coefficient ($\sigma^2(x)$) is constant, linear, a convex parabola with either zero, one or two roots, or a concave parabola with two roots.

Case 1. $\sigma^2(x) = 2\theta$
It is not difficult to see that $s(x) \propto \exp(\frac{1}{2}x^2 - \mu x) \propto \exp(\frac{(x-\mu)^2}{2})$ and furthermore $m(x) \propto \exp(-\frac{(x-\mu)^2}{2}) \propto \mathcal{N}(\mu, 1)$ and therefore $(l, r) = (-\infty, \infty)$. From this we can conclude that Condition 4.1 is satisfied for all $\mu \in \mathbb{R}$. In this special case the Pearson diffusion is an Ornstein-Uhlenbeck process, which in the finance literature is often referred to as the Vasiček model.

Case 2. $\sigma^2(x) = 2\theta x$
Here we get $s(x) \propto \exp(x - \mu \ln x) = x^{-\mu}e^x$ and $m(x) \propto x^{\mu-1}e^{-x}$. In this case a unique ergodic solution to (4.1) exists on the interval $(0, \infty)$ if and only if $\mu \geq 1$. Then the invariant distribution is the gamma distribution with scale parameter 1 and shape parameter μ, thus $[l, r) = [0, \infty)$. In the case where $0 < \mu < 1$ the boundary 0 can be reached at finite time with positive probability. On the other hand, if the boundary is made instantaneously reflecting, we also obtain a stationary process with the gamma distribution just like in the case with $\mu \geq 1$. In the finance literature the process is often refered to as the CIR-process; Cox, Ingersoll & Ross (1985).

Case 3. $a > 0$ **and** $\sigma^2(x) = 2\theta ax^2$
If you calculate the scale and speed density in this case, you get $s(x) = x^{\frac{1}{a}} \exp(\frac{\mu}{ax})$ and $m(x) = x^{-\frac{1}{a}-2} \exp(-\frac{\mu}{ax})$. If and only if $\mu > 0$ there exists a unique ergodic solution for all $a > 0$ on the positive halfline, i.e. $[l, r) = [0, \infty)$. Especially the invariant distribution is an inverse gamma distribution with shape parameter $1 + \frac{1}{a}$ and the scale parameter $\frac{a}{\mu}$. Its mean is μ and it has moments of order k for $k < 1 + \frac{1}{a}$. This process is also known as the GARCH diffusion model.

Case 4. $a > 0$ **and** $\sigma^2(x) = 2\theta a(x^2+1)$
In this case the scale density and the speed density are given by $s(x) = (x^2+1)^{\frac{1}{2a}} \exp(-\frac{\mu}{a}\tan^{-1}x)$ and $m(x) = (x^2+1)^{-\frac{1}{2a}-1}\exp(\frac{\mu}{a}\tan^{-1}x)$. For all $a > 0$ a unique solution of (4.1) is given for all $\mu \in \mathbb{R}$ on the real line, i.e. $(l,r) = (-\infty,\infty)$. For $\mu = 0$ the invariant distribution is a scaled t-distribution (scale parameter $\nu^{-\frac{1}{2}}$) with $\nu = 1+a^{-1}$ degress of freedom. Furthermore if $\mu \neq 0$ the invariant distribution is skew and has a tail decaying at the same rate as the t-distribution with $1+a^{-1}$ degress of freedom. Therefore this distribution is called the skew t-distribution and it is also known as the Pearson's type IV distribution. In both cases the mean is μ and the invariant distribution has moments of order k where $k < 1+a^{-1}$. As a result of the skewness and the heavy tailed marginal distribution the class of diffusions with $\mu \neq 0$ is potentially very useful in many applications e.g. finance. It was studied and fitted to the Nikkei 225 index, the TOPIX index and the Standard and Poors 500 index by Nagahara (1996) using the linearization method of Ozaki (1985). The skew t-distribution mean zero, ν degrees of freedom, and skewness parameter ρ has (unnormalised) density

$$f(z) \propto \big((\frac{z}{\sqrt{\nu}}+\rho)^2+1\big)^{-\frac{(\nu+1)}{2}} \exp\big(\rho(\nu-1)\tan^{-1}(\frac{z}{\sqrt{\nu}}+\rho)\big), \tag{4.3}$$

which is the invariant density of the diffusion $Z_t = \sqrt{\nu}(X_t - \rho)$ with $\nu = 1+a^{-1}$ and $\rho = \nu$. By the transformation result (see Proposition 4.2), the corresponding stochastic differential equation is

$$dZ_t = -\theta Z_t dt + \sqrt{2\theta(\nu-1)^{-1}(Z_t^2 + 2\rho\nu^{\frac{1}{2}}Z_t + (1+\rho^2)\nu)}dW_t. \tag{4.4}$$

For $\rho = 0$ the invariant distribution is the t-distribution with ν degrees of freedom.

Case 5. $a > 0$ **and** $\sigma^2(x) = 2\theta a x(x+1)$
Here the scale and speed density are given by $s(x) = (1+x)^{\frac{\mu+1}{a}}x^{-\frac{\mu}{a}}$ and $m(x) = (1+x)^{-\frac{\mu+1}{a}-1}x^{\frac{\mu}{a}}$. You can derive that if and only if $\mu \geq a$ and $a > 0$ (i.e.$\frac{\mu}{a} \geq 0$), there exists a unique ergodic solution on the positive halfline, i.e. $[l,r) = [0,\infty)$. This invariant distribution is a scaled F-distribution with $\frac{2\mu}{a}$ and $\frac{2}{a}+2$ degrees of freedom, scale parameter $\frac{\mu}{1+a}$, mean μ and it has moments of order k for $k < 1+\frac{1}{a}$. In the case where $0 < \mu < 1$ the boundary 0 can be reached with positive probability at finite time, but if we make the boundary 0 instantaneously reflecting, a stationary process is obtained with the indicated F-distribution as the invariant distribution.

Case 6. $a < 0$ **and** $\sigma^2(x) = 2\theta a x(x-1)$
For this model the scale and speed density are $s(x) = (1-x)^{\frac{1-\mu}{a}}x^{\frac{\mu}{a}}$ and $m(x) = (1-x)^{-\frac{1-\mu}{a}-1}x^{-\frac{\mu}{a}-1}$. We can see that Condition 4.1 holds if and only if $\frac{\mu}{a} \leq -1$ and $\frac{1-\mu}{a} \leq -1$. Therefore for all $a < 0$ and $\mu > 0$ such that $\min(\mu, 1-\mu) \geq -a$, a unique solution of (4.1) exists on the interval $(l,r) = (0,1)$. In particular, the invariant distribution is a Beta distribution with shape parameters $\frac{\mu}{-a}, \frac{1-\mu}{-a}$ and mean μ. If $0 < \mu < -a$ the boundary 0 can be reached with positive probability at finite time, but if we make the boundary 0 instantaneously reflecting, a stationary process is obtained with the indicated Beta distribution as the invariant distribution. A similar argument applies to the boundary 1 when $0 < 1-\mu < -a$. These diffusion are often referred to as the Jacobi diffusion, since their eigenfunctions are Jacobi polynomials.

4.2 Moments

In the case of a linear drift and the existence of the second order moment, it can be shown that the *autocorrelation function* is given by

$$\rho(t) = \mathrm{Cor}(X_s, X_{t+s}) = e^{-\theta t}.$$

Now we want to find an expression for the moments and conditional moments of the marginal distribution. In the previous section we saw that $E(|X_t|^k) < \infty$ if and only if $a < (k-1)^{-1}$. Therefore all moments exists if $a \leq 0$ and we get with Itô's formula

$$\begin{aligned} dX_t^n = &\Big(-\theta n X_t^{n-1}(X_t - \mu) + \theta n(n-1)X_t^{n-2}(aX_t^2 + bX_t + c)\Big)dt \\ &+ nX_t^{n-1}\sigma(X_t)dW_t. \end{aligned} \tag{4.5}$$

By using (4.5) we can get a recursive formula to calculate the nth moment of the marginal distribution. If $E(X_t^{2n})$ is finite (i.e. if $a < (2n-1)^{-1}$), we gain the following proposition.

Proposition 4.3. *In the case of $a < (2n-1)^{-1}$ we get*

$$E(X_t^n) = a_n^{-1}(b_n E(X_t^{n-1}) + c_n E(X_t^{n-2})), \tag{4.6}$$

where $a_n = \theta n(1-(n-1)a), b_n = \theta n(\mu + (n-1)b)$ and $c_n = \theta n(n-1)c$ for $n = 0, 1, 2, \ldots$. The initial conditions are given by $E(X_t^0) = 1$ and $E(X_t) = \mu$.

Proof. With (4.5) it is easy to see that

$$dX_t^n = -a_n X_t^n dt + b_n X_t^{n-1}dt + c_n X_t^{n-2}dt + nX_t^{n-1}\sigma(X_t)dW_t,$$

with a_n, b_n and c_n as defined in the statement. Since $a < (2n-1)^{-1}$, it is important to notice that the latter term of the previous formula is a martingale (since $E(n^2 X_t^{2(n-1)}\sigma^2(X_t)) = 2n^2\theta E(aX_t^{2n} + bX_t^{2n-1} + cX_t^{2(n-1)}) < \infty$ and by using the theory of stochastic integration; see in the Appendix for further information). For that reason and the fact, that X_t is stationary $((X_s, \ldots, X_{s+k}) \sim (X_t, \ldots, X_{t+k}))$ we conclude

$$\begin{aligned} \mu_n := E(X_t^n) = &\underbrace{E(X_0^n)}_{=\mu_n} - a_n \int\int_0^t X_s^n ds dP + b_n \int\int_0^t X_s^{n-1} ds dP \\ &+ c_n \int\int_0^t X_s^{n-2} ds dP + \int\int_0^t nX_s^{n-1}\sigma(X_s)dW_s dP. \end{aligned}$$

This is with Fubini equivalent to

$$\begin{aligned} \mu_n =& \mu_n - a_n \int_0^t E(X_s^n)ds + b_n \int_0^t E(X_s^{n-1})ds + c_n \int_0^t E(X_s^{n-2})ds \\ 0 =& -ta_n E(X_s^n) + tb_n E(X_s^{n-1}) + tc_n E(X_s^{n-2}) \\ E(X_t^n) =& a_n^{-1}(b_n E(X_t^{n-1}) + c_n E(X_t^{n-2})). \end{aligned}$$

The term $\int\int_0^t nX_s^{n-1}\sigma(X_s)dW_s dP$ disappears by the fact that it is a martingale with the value zero at time zero. $\square$

Example 4.4. Now we can take (4.6) to find the moments of the skewed t-distribution (see Case 4) even if we do not know the normalising constant of the density of (4.3). In particular, for the diffusion (4.4) where $\mu = 0$,

$$\begin{aligned} E(Z_t^2) &= \operatorname{Var}(Z_t) = \frac{(1+\gamma^2)\nu}{\nu-2}, \\ E(Z_t^3) &= \frac{4\gamma(1+\gamma^2)\nu^{\frac{3}{2}}}{(\nu-3)(\nu-2)} \end{aligned}$$

and

$$E(Z_t^4) = \frac{24\gamma^2(1+\gamma^2)\nu^2 + 3(\nu-3)(1+\gamma^2)^2\nu^2}{(\nu-4)(\nu-3)(\nu-2)}.$$

Δ

It is not far to get form Proposition 4.3 a recursive formula like (4.6) for the *conditional moments*. If we denote $q_n(x,t) := E(X_t^n|X_0 = x)$ and the $2n$th moment exists, then the recursion is given by

$$\frac{d}{dt} q_n(x,t) = -a_n q_n(x,t) + b_n q_{n-1}(x,t) + c_n q_{n-1}(x,t) + c_n q_{n-2}(x,t). \tag{4.7}$$

Solving this differential equation with initial $q_n(x,0) = x^n$ we get

$$q_n(x,t) = x^n e^{-a_n t} + b_n I_{n-1}(a_n,x,t) + c_n I_{n-2}(a_n,x,t)$$

with $I_\eta(\alpha,x,t) = e^{-\alpha t}\int_0^t e^{\alpha s} q_\eta(x,s)ds$. We also get a recursion for $I_\eta(\alpha,x,t)$ if we apply partial integration and (4.7).

$$\begin{aligned} I_\eta(\alpha) =& I_\eta(\alpha,x,t) \\ =& e^{-\alpha t}\int_0^t e^{\alpha s} q_\eta(x,s)ds \\ =& e^{-\alpha t}\Big(\frac{1}{\alpha} e^{\alpha s} q_\eta(x,s)\Big|_0^t + \frac{a_\eta}{\alpha}\int_0^t e^{\alpha s} q_\eta(x,s)ds \\ &- \frac{b_\eta}{\alpha}\int_0^t e^{\alpha s} q_{\eta-1}(x,s)ds - \frac{c_\eta}{\alpha}\int_0^t e^{\alpha s} q_{\eta-2}(x,s)ds\Big) \\ =& \frac{e^{-\alpha t}}{\alpha}\big(e^{\alpha t} q_\eta(x,t) - x^\eta\big) + \frac{a_\eta}{\alpha} I_\eta(\alpha) - \frac{b_\eta}{\alpha} I_{\eta-1}(\alpha) - \frac{c_\eta}{\alpha} I_{\eta-2}(\alpha) \end{aligned}$$

which is equivalent to

$$I_\eta(\alpha) = \frac{x^\eta(e^{-a_\eta t} - e^{-\alpha t}) + b_\eta(I_{\eta-1}(a_\eta) - I_{\eta-1}(\alpha)) + c_\eta(I_{\eta-2}(a_\eta) - I_{\eta-2}(\alpha))}{\alpha - a_\eta}.$$

To get $I_1(\alpha,x,t)$ we use the fact that $I_0(\alpha,x,t) = \alpha^{-1}(1-e^{-\alpha t})$ as $q_0(x,t) = 1$ and $c_1 = 1$. We see that $q_n(x,t)$ is a polynomial of order n in x for fixed t. A somewhat easier derivation of this result comes by means of the eigenfunctions considered below.

4.3 Eigenfunction

For any diffusion process

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{4.8}$$

where W_t is the Wiener process, we define the generator as follows

$$A = b(x)\frac{d}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2}. \tag{4.9}$$

We call a twice continuous diffential function h an eigenfunction of the diffusion process if there exists a positive number λ, the eigenvalue, such that

$$Ah(x) = -\lambda h(x).$$

Proposition 4.5. *Assume h is an eigenfunction with eigenvalue λ. Further, the condition*

$$\int_0^t E(h'(X_s)^2\sigma^2(X_s))ds < \infty \tag{4.10}$$

holds for all $t > 0$. Then

$$E(h(X_t)|X_0 = x) = e^{-\lambda t}h(x). \tag{4.11}$$

Proof. Set $Y_t = e^{\lambda t}h(X_t)$, then apply Itô's formula. Then we get

$$\begin{aligned}
dY_t =& \lambda e^{\lambda t}h(X_t)dt + e^{\lambda t}h'(X_t)dX_t + \frac{1}{2}e^{\lambda t}h''(X_t)\sigma^2(X_t)dt \\
=& \lambda e^{\lambda t}h(X_t)dt + e^{\lambda t}h'(X_t)b(X_t)dt + \frac{1}{2}e^{\lambda t}h''(X_t)\sigma^2(X_t)dt \\
& + e^{\lambda t}h'(X_t)\sigma(X_t)dW_t \\
=& \lambda e^{\lambda t}h(X_t)dt + e^{\lambda t}\underbrace{\left(h'(X_t)b(X_t) + \frac{1}{2}h''(X_t)\right)}_{=Ah(X_t)=-\lambda h(X_t)}dt + e^{\lambda t}h'(X_t)\sigma(X_t)dW_t \\
=& e^{\lambda t}h'(X_t)\sigma(X_t)dW_t.
\end{aligned}$$

Since the condition (4.10) holds, $M_t := \int_0^t e^{\lambda s}h'(X_s)\sigma(X_s)dW_s$ is a martigale with $M_0 = 0$. Therefore we obtain,

$$E(e^{\lambda t}h(X_t)|X_0 = x) = h(X_0) + \underbrace{E(\int_0^t e^{\lambda s}h'(X_s)\sigma(X_s)dW_s)}_{=0}$$

which is equivalent to

$$E(h(X_t)|X_0 = x) = e^{-\lambda t}h(x).$$

□

Property (4.11) can be used to construct martingale estimation functions. In the case of Pearson diffusions that have linear drift and a quadratic squared diffusion coefficient the generator maps polynomials into polynomials. It is not diffcult to show that $p_n(x) = \sum_{j=0}^{n} p_{n,j} x^j$ is an eigenfunction if there exists an eigenvalue $\lambda_n > 0$ that satisfies

$$\theta(ax^2 + bx + c)p_n''(x) - \theta(x-\mu)p_n'(x) = -\lambda_n p_n(x), \text{ i.e.}$$
$$\sum_{j=0}^{n}(\lambda_n - a_j)p_{n,j}x^j + \sum_{j=0}^{n-1} b_{j+1}p_{n,j+1}x^j + \sum_{j=0}^{n-2} c_{j+2}p_{n,j+2}x^j = 0$$

where $a_j = j\theta(1-(j-1)a), b_j = j\theta(\mu+(j-1)b)$ and $c_j = j(j-1)c\theta$ for $j \in \mathbb{N}$. Without loss of generality we can assume $p_{n,n} = 1$ and therefore we get $\lambda_n = a_n = \theta n(1-(n-1)a)$. Further, if we define $p_{n,n+1} = 0$, then the coefficients $(p_{n,j})_{j=0,\ldots,n-1}$ solve the linear system

$$(a_j - a_n)p_{n,j} = b_{j+1}p_{n,j+1} + c_{j+2}p_{n,j+2}. \tag{4.12}$$

Note that λ_n is positive if and only if $a < (n-1)^{-1}$, which is exactly the condition ensuring that $p_n(x)$ is integrable with respect to the invariant distribution. If even the stronger condition $a < (2n-1)^{-1}$ holds, the first n eigenfunctions belong to the space of the functions that are square intregrable with respect to the invariant distribution, and they are orthogonal with respect to the inner product. The space of square integrable and twice continuously differentiable functions is often taken as the domain of the generator. By using the linearity of the expectation and with (4.11), we can show that

$$E(X_t^n|X_0 = x) = e^{-a_n t}\sum_{j=0}^{n} p_{n,j}x^j - \sum_{j=0}^{n-1} p_{n,j}E(X_t^j|X_0 = x), \tag{4.13}$$

provided that the drift and diffusions coefficient are of linear growth and that the eigenfunction is of polynomial growth. The conditional expectation is for any fixed t a polynomial of order n in x with coefficients which are linear combinations of $1, e^{-\lambda_1 t}, \ldots, e^{-\lambda_n t}$. Set $\lambda_0 = 0$ and

$$E(X_t^n|X_0 = x) = q_n(x,t) = \sum_{j=0}^{n} q_{n,j}x^j = \sum_{j=0}^{n}\sum_{l=0}^{n} q_{n,j,l}e^{-\lambda_l t}x^j. \tag{4.14}$$

Especially $q_{n,j,n} = p_{n,j}$ and $q_{n,j,l} = -\sum_{k=l}^{n-1} p_{n,k}q_{k,j,l}$ for $l = 0,\ldots,n-1$.

If $a \leq 0$ in (4.1), then there exist infinitely many polynomial eigenfunctions and they are well-known families of orthogonal polynomials (see the table below). In particular all moments exists for these marginal distributions.

case	**marginal distribution**	**eigenfunction**
1	Normal	Hermite polynomials
2	Gamma	Laguerre polynomials
6	Beta	Jacobi polynomials on $(0,1)$

On the other hand if $a > 0$ (remaining cases) implies that there is only a finite number of polynomial eigenfunctions. The number is the integer part of $1 + a^{-1}$, which is also the order of the highest moment of the marginal distribution. Wong (1964) has shown that the spectrum of these diffusions is not discrete. The spectrum has a continuous part with eigenvalues that are larger than those in the discrete part of the spectrum.
It is known for instance that polynomial associated with the inverse gamma distribution are the Bessel polynomials.

Example 4.6. We want to derive the first two conditional moments for the skew t-diffusion (4.4). Here is $a = (\nu - 1)^{-1}$, therefore we get $\lambda_n = n(\nu - n)(\nu - 1)^{-1}\theta$ for $n < \nu$. Using (4.12) and (4.13) we can calculate the first eigenfunctions

$$\begin{aligned}
p_1(z) &= z, \\
p_2(z) &= z^2 - \frac{4\gamma\nu^{\frac{1}{2}}}{\nu - 3} z - \frac{(1 + \gamma^2)\nu}{\nu - 2}, \\
p_3(z) &= z^3 - \frac{12\gamma\nu^{\frac{1}{2}}}{\nu - 5} z^2 + \frac{24\gamma^2\nu + (1 + \gamma^2)\nu(\nu - 5)}{(\nu - 5)(\nu - 4)} z + \frac{8\gamma(1 + \gamma^2)\nu^{\frac{3}{2}}}{(\nu - 5)(\nu - 3)},
\end{aligned}$$

and

$$\begin{aligned}
p_4(z) =& z^4 - \frac{24\gamma\nu^{\frac{1}{2}}}{\nu - 7} z^3 + \frac{144\gamma^2\nu + 6(1 + \gamma^2)\nu(\nu - 7)}{(\nu - 7)(\nu - 6)} z^2 \\
&+ \frac{8\gamma(1 + \gamma^2)\nu^{\frac{3}{2}}(\nu - 7) + 48\gamma(1 + \gamma^2)\nu^{\frac{3}{2}}(\nu - 6) - 192\gamma^3\nu^{\frac{3}{2}}}{(\nu - 7)(\nu - 6)(\nu - 5)} z \\
&+ \frac{3(1 + \gamma^2)^2\nu(\nu - 7) - 72\gamma^2(1 + \gamma^2)\nu^2}{(\nu - 7)(\nu - 6)(\nu - 4)}
\end{aligned}$$

assuming $\nu > 4$. Now we can calculate the first and second conditional moments, which are

$$E(Z_t | Z_0 = z) = z e^{-\theta t}$$

and

$$E(Z_t^2 | Z_0 = z) = e^{-\frac{2\nu - 4}{\nu - 1}\theta t} z^2 + \frac{4\gamma\nu^{\frac{1}{2}}}{\nu - 3}\left(e^{-\theta t} - e^{-\frac{2\nu - 4}{\nu - 1}\theta t}\right) z + \frac{(1 + \gamma^2)\nu}{\nu - 2}(1 - e^{-\frac{2\nu - 4}{\nu - 1}\theta t}).$$

Δ

Chapter 5

Estimating functions for Pearson diffusions

Now we want to find estimating functions for Pearson diffusions. Suppose we have a sequence of observations $Y_0, Y_1, \ldots, Y_n$ at time points $t_i = i\Delta$ for $i = 0, 1, \ldots, n$. Our aim is to estimate a parameter ψ belonging to the parameter space $\Psi \subset \mathbb{R}^d$. The parameter ψ might be the parameter (θ, μ, a, b, c) (with θ, μ, a, b and c defined as in (4.1)) of the full class of Pearson diffusions, or it might be a subclass, e.g. a class corresponding to one of the Pearson types. In this chapter we discuss estimations methods that are simpler for the Pearson diffusion than for general diffusions.

5.1 Maximum likelihood estimation

The maximum likelihood estimation is well known for the different case of Pearson diffusions as seen in Chapter 4. Especially well known is the case of the Ornstein-Uhlenbeck process, where the transition is Gaussian with a simple expression for the first and second conditional moments. In the case of the CIR process (Case 2) the transition density is a non-cetral $\mathcal{X}^2$-distribution which can be expressed in terms of a modified Bessel function. Therefore exact likelihood interference is relatively easy for these models. For a diffusion with a discrete spectrum representation of the transition density exists with the help of the eigenfunction and we get

$$\pi(\Delta, x, y) = m(y) \sum_{j=1}^{\infty} e^{-\Delta\lambda_j} p_j(x) p_j(y) c_j$$

with $y \mapsto \pi(\Delta, x, y)$ the transition density. It is also given that p_j is the jth eigenfunction with the eigenvalue λ_j and $c_j^{-1} = \int_l^r p_j(x)^2 m(x) dx$. In the case of the Ornstein-Uhlenbeck process this is just Mehler's formula for Hermite polynomials and in the case of the CIR process it is just a classical expansion of the modified Bessel function.
The remaining Pearson diffusions have only a finite amountof discrete eigenvalues and the spectrum comprises a continuous part. A number of general techniques for maximum likelihood interference for discretely obeserved diffusions are available, some of them are

mentioned in Forman & Sørensen (2007). Due to the fact that these methods do not simplify the Pearson diffusion, they will not be considered furtheron.

5.2 Estimation based on conditional moments

We have seen that maximum likelihood interference is, at least approximately, feasible for the Pearson diffusions. However, much simpler estimators can be found by using the explicit expression available for the conditional polynomial moments. Those estimators have, if properly chosen, an efficiency close to the maximum likelihood, if the sampling frequency is not too small.
From the last chapter we know, that the conditional moments of Pearson diffusion are polynomials. If the first N moments exists, then the first N eigen-polynomials $p_1(.,\psi),\dots,p_N(.,\psi)$ are well defined. Therefore we can apply a martingale estimating function of the type (2.6) as following,

$$G_n(\psi)=\sum_{i=1}^{n}\sum_{j=1}^{N}\alpha_j(Y_{i-1},\psi)\{p_j(Y_i,\psi)-e^{-\lambda_j(\psi)\Delta}p_j(Y_{i-1},\psi)\} \tag{5.1}$$

where $\alpha_1,\dots,\alpha_N$ are weight functions and $\lambda_1(\psi),\dots,\lambda_N(\psi)$ are eigenvalues. We can also write it in a matrix form, then we get the associated estimating equation in the form

$$G_n(\psi)=\sum_{i=1}^{n}\alpha(Y_{i-1},\psi)h(Y_{i-1},Y_i,\psi)=0 \tag{5.2}$$

with $h_j(x,y,\psi)=p_j(y,\psi)-e^{-\lambda_j(\psi)\Delta}p_j(x,\psi)$ for $j=1,\dots,N$ and α is the $n\times N$ weight matrix. Later on we will assume that the diffusions has finite moments of the order of $2N$ in order to apply a central limit theorem.
We shall focus on the optimal estimating function of the type (5.2). In this case we define the optimal estimating function in the way that the weight function α is chosen to minimise the asymptotic variance of the related estimator.

5.3 Optimal martingale estimating function

An advantage of the Pearson diffusion in terms of efficient interference is that the optimal weights in the sense of Godambe & Heyde (1987) are simple and explicit. We assume that the Pearson diffusion is ergodic, has moments of order $2N$ (in particular $a<(2N-1)^{-1}$) and that the mapping $\psi\mapsto\tau=(\theta,\mu,a,b,c)$ is differentiable. The optimal weights for (5.1) are by

$$a^*(x,\psi)=-S(x,\psi)^TV(x,\psi)^{-1} \tag{5.3}$$

where

$$S_{j,k}(x,\psi)=-\,E_\psi(\partial_{\psi_k}p_j(Y_i,\psi)|Y_{i-1}=x)+\partial_{\psi_k}(e^{-\lambda_j(\psi)\Delta}p_j(x,\psi)) \tag{5.4}$$

$$V_{j,k}(x,\psi)=E_\psi(p_j(Y_i,\psi)p_k(Y_i,\psi)|Y_{i-1}=x)-e^{-(\lambda_j(\psi)+\lambda_k(\psi))\Delta}p_j(x,\psi)p_k(x,\psi), \tag{5.5}$$

i.e. $V_{j,k}(x,\psi) = \mathrm{COV}_x(h)$ with $|j-k| = h$. Note that the indicated conditions imply that S and V are well defined.

Proposition 5.1. *The matrix $V_{j,k}(x,\psi)$ as it is defined in (5.5) is invertible.*

Proof. We want to show that the matrix is positive definite. Since $V_{j,k}(x,\psi) = \mathrm{COV}_x(h)$ is $V_{j,k}(x,\psi)$ positive semidefinite for all (x,ψ). Moreover, for $z \in \mathbb{R}^N$ it holds that $z^T V(x,\psi) z = 0$ iff

$$0 = \sum_{j=0}^{N} z_j (p_j(y,\psi) - e^{-\lambda_j(\psi)\Delta} p_j(x,\psi))$$

for almost every y with respect to the conditional distribution of Y_i given $Y_{i-1} = x$ under ψ. However, the equation above is a polynomial in y with order j and leading coefficient $p_{j,j} = 1$. Therefore we can deduce that $z^T V(x,\psi) z = 0$ iff $z = 0$. □

With the help of equations (4.12) and (4.13) we can get the optimal weight explicitly. Then we get

$$S_{j,k}(x,\psi) = p_j(Y_i,\psi) e^{-\lambda_j(\psi)\Delta} \partial_{\psi^T} \lambda_j(\psi) + \sum_{j'=0}^{j} (q_{j'}(x,\Delta,\psi) - e^{-\lambda_j(\psi)\Delta} x^{j'}) \partial_{\psi^T} p_{j,j'}(\psi)$$

$$V_{j,k}(x,\psi) = \sum_{j'=0}^{j} \sum_{k'=0}^{k} p_{j,j'}(\psi) p_{k,k'}(\psi) q_{j'+k'}(x,\Delta,\psi) - e^{-(\lambda_j(\psi)+\lambda_k(\psi))\Delta} p_j(x,\psi) p_k(x,\psi)$$

where $q_j(x,t,\psi) = E_\psi(X_t^j | X_0 = x)$ is specified by equations (4.13) and (4.14). The j,k'th element of $S(x,\psi)$ is the polynomial of the order j with $s_{j,k}(x) = \sum_{l=0}^{j} s_{j,k,l} x^l$ and coefficients

$$s_{j,k,l} = e^{-\lambda_j \Delta}(p_{j,l} \delta_{\psi_k} \lambda_j - \delta_{\psi_k} p_{j,l}) + \sum_{j'=0}^{l} \partial_{\psi_k} p_{j,j'} q_{j',l}(\Delta)$$

and similar for the j,k'th element of $V(x,\psi)$ is a polynomial $v_{j,k}(x) = \sum_{l=0}^{j+k} v_{j,k,l} x^l$ with coefficients given by

$$v_{j,k,l} = \sum_{j'=0}^{0} \sum_{k'=0}^{k} p_{j,j'} p_{k,k'} (q_{j'+k',l}(\Delta) - e^{-(\lambda_j+\lambda_k)\Delta} I_{j'+k'=l}).$$

It is important to notice that we can derive from (4.12) that the derivatives $d_{j,l} = \partial_{\psi^T} p_{j,l}$ satisfy the recursion

$$d_{j,l} = \frac{b_{l+1} d_{j,l+1} + c_{l+2} d_{j,l+2} + p_{j,l} \partial_{\psi^T}(a_l - a_j) + p_{j,l+1} \partial_{\psi^T} b_{l+1} + p_{j,l+2} \partial_{\psi^T} c_{l+2}}{a_l - a_j}$$

for $l = j-1, j-2, \ldots, 0$ where initially $d_{j,j} = d_{j,j+1} = 0$.

5.4 Asymptotic theory

Under mild regularity conditions, the optimally martingale estimating function (5.1) provides consistent and asymptotically normal estimators of the parameter of the Pearson diffusion. From now on ψ_0 denotes the true parameter.

Theorem 5.2. *Suppose that the following hold true:*

(i) The Pearson diffusion is ergodic and has moments of the order $2N$ where $N \geq 2$.

(ii) The true parameter ψ_0 belongs to the interior of Ψ.

(iii) The mapping $\psi \mapsto \tau = (\theta, \mu, a, b, c)$ is differentiable and $\partial_\psi \tau(\psi_0)$ has full rank d.

Then with probability tending to one as $n \to \infty$ there exists a solution $\hat{\psi}_n$ to the estimating equation (5.2) with weights specified by (5.3) such that $\hat{\psi}_n$ converges to ψ_0 in probability and

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W(\psi_0)^{-1})$$

where $W(\psi_0) = E_{\psi_0}(S(Y_i, \psi_0)^T V(Y_i, \psi_0)^{-1} S(Y_i, \psi_0))$.

The proof of this theorem is given in the appendix of Forman & Sørensen (2007). By the condition (i) in the theorem above is ensured, that the eigenfunctions are well defined and that $h_1, \ldots, h_N$ have finite variance so that $G_n(\psi_0)$ is indeed a martingale, i.e. implies (i) that $G_n(\psi)$ is a square integrable martingale.

Example 5.3. Let us consider (4.4), the skewed t-diffusion, where the parameter $\psi = (\theta, \nu, \rho)$ is the canonical parameter,

$$\tau = (\theta, \nu, a, b, c) = \big(\theta, 0, \frac{1}{\nu - 1}, \frac{2\rho\nu^{\frac{1}{2}}}{\nu - 1}, \frac{(1 + \rho^2)\nu}{\nu - 1}\big)$$

and

$$\frac{\partial \tau}{\partial \psi^T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{(\nu-1)^2} & \frac{\rho}{\nu^{\frac{1}{2}}(\nu-1)} - \frac{2\rho\nu^{\frac{1}{2}}}{(\nu-1)^2} & \frac{1+\rho^2}{\nu-1} - \frac{\nu(1+\rho^2)}{(\nu-1)^2} \\ 0 & 0 & 0 & \frac{2\nu^{\frac{1}{2}}}{\nu-1} & \frac{2\nu\rho}{\nu-1} \end{pmatrix}$$

has full rank three. Hence, consistent and asymptotically normal estimators are obtained by means of the optimally weighted martingale estimating functions under the further assumption that $\nu_0 > 2N$.

Δ

Chapter 6

Pearson diffusions with jumps

Now we want to extend the process (4.1) by adding jumps. Although Zhou (2003) gives us a good guideline, we want to consider a more general process.
We still consider the process to be one dimensional and extend the diffusion (4.1) to the following process

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta(aX_s^2 + bX_s + c)}dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_n -}) \qquad (6.1)$$

where W_t is the standard Brownian motion and $W_t, \mathcal{N}_t$ are independent. We assume that $\mathcal{N}_t$ is a Poisson process with a constant jump rate $\lambda > 0$. Then we define the stopping time $\tau_k := \inf\{t > \tau_{k-1} | \mathcal{N}_t - \mathcal{N}_{t-} > 0\}$ which is the kth jump for $k \in \mathbb{N}$ and $\tau_1 = \inf\{t > 0 | \mathcal{N}_t - \mathcal{N}_{t-} > 0\}$ and $\mathcal{N}_{t-} = \lim_{s \to t, s<t} \mathcal{N}_s$. Furthermore $J_t \in \mathbb{R}\backslash\{0\}$ is a random variable and we assume that they are distributed with some distribution $F_{X_{t-}}$ and a density function $f_{X_{t-}}$, i.e. the distribution of $J_{X_{s-}}$ does only depend on the value of the process just before the jump. Furthermore we assume that the function $x \to F_x$ is continuous with respect to x. Let

$$g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad (y, x) \mapsto g(y, x)$$

be a function and let g be continuously differential with respect to y. We further assume that the inverse function of g with respect to y exists, i.e. for all $x \in \mathbb{R}\backslash\{0\}$ there exists a function $g_x^{-1}(y)$ such that $g(g_x^{-1}(y), x) = 1$, and that $|\partial_y g(y, x)^{-1}| \neq 0$ for all x and y. Furthermore we assume that the first two moments of $g(y, x)$ with respect to y exist, i.e. that

$$\int_{\mathbb{R}} |g(y, x)| f_x(y) dy < \infty \quad \text{and} \quad \int_{\mathbb{R}} |g(y, x)|^2 f_x(y) dy < \infty.$$

We can for example assume that g is a product of a function w, where

$$w : \mathbb{R} \to \mathbb{R}, \quad x \mapsto w(x)$$

and w is of such a form that the properties of g as described above are satisfied, and y, i.e.

$$g(y, x) = yw(x). \qquad (6.2)$$

The parameters a, b and c have to be discussed later on. To ensure a solution to the ordinary differential equation (6.61), we assume that

$$\lambda E(g(J_{\mathcal{N}_t}, X_{t-})^k | X_{t-} = x) = \lambda \int g(y,x)^k f_x(y)dy = \sum_{j=0}^{k} \alpha_{j,k} x^j \tag{6.3}$$

with $\alpha_{jk} \in \mathbb{R}$ for all $k \in \mathbb{N}$ and $0 \leq j \leq k$ holds. It is worth to mention, that $\alpha_{j,k} = 0$ if $j \notin [0, \ldots, k]$. We will see that (6.3) is an important condition if we want to derive a formula to calculate the conditional moments.
In order to simplify the writing, we will now use $b(X_t)$ for $-\theta(X_s - \mu)$ and $\sigma(X_t)$ for $\sqrt{2\theta(aX_s^2 + bX_s + c)}$.
For further use (see Subsection 6.2) we shall convert (6.1). We shall introduce two measures

$$\Psi((0,t] \times A) = \sum_{n=1}^{\mathcal{N}_t} 1_{(J_n \in A)} \quad \text{and} \tag{6.4}$$

$$\Lambda((0,t] \times A) = \lambda \int_0^t \int_A f_{X_{s-}}(y)dyds, \tag{6.5}$$

then we can convert (6.1) to

$$\begin{aligned} X_t =& X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_{n-}}) \\ =& X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\Psi(s,y) \\ =& X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\Psi(s,y) \\ &- \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\Lambda(s,y) + \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\Lambda(s,y) \\ =& X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\tilde{\mathcal{N}}(s,y) \\ &+ \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\Lambda(s,y) \end{aligned} \tag{6.6}$$

with $\tilde{\mathcal{N}}(s,y) := \Psi(s,y) - \Lambda(s,y)$. Then

$$\int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\tilde{\mathcal{N}}(s,y)$$

is a martingale (see the part about Poisson integration in the Appendix). If we define

$$h(x) = \int_{\mathbb{R}} g(y,x)f_x(y)dy,$$

then we can further convert (6.6) to

$$\begin{aligned}X_t =& X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\tilde{\mathcal{N}}(t,y)\\ &+ \lambda \int_0^t h(X_{s-})ds\\ =& X_0 + \int_0^t c(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_{s-})d\tilde{\mathcal{N}}(t,y) \end{aligned} \tag{6.7}$$

with $c(x) = b(x) + \lambda h(x)$.
Due to the addition of the jump function $g(y,x)$, a proposition like 4.2 is in the case of (6.1) not given for all $g(y,x)$. It is possible, that unpleasant occurences could appear such as $aX_t^2 + bX_s + c < 0$. If we assume for a moment, that we do assume that this case does not occur (later on will give some constrictions for the several cases), then we gain that a process of the form of (6.1) is closed under linear transformation, i.e. if we define $\tilde{X}_t = \gamma X_t + \delta$ with $\gamma \neq 0$ then we gain a new process of the type (6.1) where $\tilde{\theta} = \theta, \tilde{\mu} = \gamma\mu + \delta, \tilde{a} = a, \tilde{b} = b\gamma - 2a\delta, \tilde{c} = c\gamma^2 - b\gamma\delta + a\delta^2$ and $\tilde{g}(y,x) = g(y, \frac{x-\delta}{\gamma})$ (the proof is similar to the proof of proporsition 4.2). The linear transformation on x to get $\tilde{g}(y,x)$, does not change the properties that $g(y,x)$ satisfies.
Since a linear transformation is possible, like in the case without jumps, it is sufficient to briefly discuss the 6 different types like in Subsection 4.1. We also shall briefly discuss under which assumptions the Condition (6.28) holds, where we assume that the distribution of J is independent from the state of the process. In this case $\Lambda((0,t] \times A)$ simplifies to

$$\Lambda((0,t] \times A) = \lambda \int_0^t \int_A f_{X_{s-}}(y)dyds = \lambda t \int_A f(y)dy \tag{6.8}$$

and $h(x)$ to

$$h(x) = \int_{\mathbb{R}} g(y,x)f(y)dy.$$

Case 1. If we assume $\sigma^2(x) = 2\theta$, we get in this first case

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta}dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_{n-}}). \tag{6.9}$$

Since the term under the square root does not depend on X_t, we do not need constraints on the function $g(y,x)$.
Under the assumption that $g(y,x)$ is Lipschitz-continuous with respect to x, i.e. there exists a $L > 0$ such that for all $x_1, x_2 \in \mathbb{R}$ and for all y

$$|g(y,x_1) - g(y,x_2)| \leq Lr(y)|x_1 - x_2|, \tag{6.10}$$

with $r : \mathbb{R} \to \mathbb{R}$, $\int_{\mathbb{R}} r^2(y)f(y)dy := Z < \infty$, the existence of the second moment for the distribution of J and futher $\int_{\mathbb{R}} r^2(y)f^2(y)dy := O < \infty$, we derive that the condition

(6.28) is satisfied. For any $x_1, x_2 \in \mathbb{R}$, we gain that

$$\begin{aligned}
&|c(x_1) - c(x_2)|^2 + |\sigma(x_1) - \sigma(x_2)|^2 + \lambda \int_{\mathbb{R}} |g(y, x_1) - g(y, x_2)|^2 f(y) dy \\
&\leq |\theta(x_1 - x_2) + \lambda(h(x_1) - h(x_2))|^2 + |\sqrt{2\theta} - \sqrt{2\theta}|^2 + \lambda L^2 |x_1 - x_2|^2 \int_{\mathbb{R}} r(y)^2 f(y) dy \\
&\leq 2\theta^2 |(x_1 - x_2)|^2 + 2\lambda^2 | \int_{\mathbb{R}} (g(y, x_1) - g(y, x_2)) f(y) dy|^2 + \lambda Z L^2 |x_1 - x_2|^2 \\
&\leq K|x_1 - x_2|^2
\end{aligned}$$

with $K = 2\theta^2 + 2\lambda^2 O L^2 + \lambda Z L^2 < \infty$. Since all the three parts of (6.9) are Lipschitz continuous, we gather

$$\begin{aligned}
|g(y, x)| = |g(y, x) - g(y, 0) + g(y, 0)| &\leq |g(y, x) - g(y, 0)| + |g(y, 0)| \\
&\leq Lr(y)|x - 0| + |g(y, 0)| \leq Lr(y)|x| + |g(y, 0)| \\
&\leq K_g(1 + |x|)
\end{aligned}$$

where $K_g = \max\{|g(y, 0)|, Lr(y)\}$. We can do the same trick with $c(x)$ and $\sigma(x)$, and therefore the Condition (6.29) is satisfied.

Case 2. Like in Chapter 4, we assume in the second case $\sigma^2(x) = 2\theta x$ and then we gain

$$X_t = X_0 - \int_0^t \theta(X_s - \mu) ds + \int_0^t \sqrt{2\theta X_s} dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_n -}). \tag{6.11}$$

In this case we have to prevent that $X_t < 0$ and therefore we need a condition on the function $g(y, x)$ such as

$$P_x(g(J, x) > -x) = 1. \tag{6.12}$$

In order to satisfy (6.28), we need that $\sigma(x)$ satisfies

$$(\sigma(x_1) - \sigma(x_2))^2 \leq C|x_1 - x_2|^2 \tag{6.13}$$

for all $x_1, x_2 \geq 0$ and $C > 0$. In this case we have the situation that

$$(\sigma(x_1) - \sigma(x_2))^2 = 2\theta(\sqrt{x_1} - \sqrt{x_2})^2$$

and therefore we do not obtain a constant $C > 0$ such that (6.13) for all $x_1, x_2 \geq 0$ holds. One possible proof is to look at the derivative of $\sigma(x)$. Then you can see that $\sigma'(x)$ goes to infinity when x goes towards 0. In this case you only get a local Lipschitz condition for big x.

Case 3. In the third case we assume, that $a > 0$ and $\sigma^2(x) = 2\theta a x^2$ and therefore we get

$$X_t = X_0 - \int_0^t \theta(X_s - \mu) ds + \int_0^t \sqrt{2\theta X_s^2} dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_n -}). \tag{6.14}$$

Like in Case 1 we do not need any restrictions on $g(y,x)$ since the term below the sqare root is always positive.
But contrary to Case 2, we notice that for all $x_1, x_2 \geq 0$

$$(\sigma(x_1) - \sigma(x_2))^2 = 2\theta(x_1 - x_2)^2 \leq C(x_1 - x_2)^2$$

where $C = 2\theta$. Then we gain with the assumption like in Case 1, e.g. that $g(y,x)$ satisfies (6.10), that (6.14) satisfies (6.28).

Case 4. Now we consider $\sigma^2(x) = 2\theta a(x^2+1)$ and $a > 0$, then we gain

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta a(x^2+1)}dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_n -}). \tag{6.15}$$

Since $\sigma^2(x) \geq 0$ we do not need any constraints on $g(y,x)$.
If we take a closer look at the derivative of $\sigma(x)$ then we obtain

$$\partial\sigma(x) = \frac{2\theta a x}{\sqrt{2\theta a(x^2+1)}}$$

and the derivative is therefore bounded. Thus $\sigma(x)$ is Lipschitz continuous with $\sqrt{2\theta a}$ as Lipschitz constant. Then we gain, that Condition (6.28) and Conditon (6.29) are satisfied under the same assumptions as in Case 1.

Case 5. This time we assume $\sigma^2(x) = 2\theta a x(x+1)$ and $a > 0$, so we gain

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta a x(x+1)}dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_n -}). \tag{6.16}$$

In this case we need to assume like in Case 2, that $g(y,x)$ satisfies (6.12).
Like in Case 2 we do not obtain the condition (6.28) is satisfied for all $x_1, x_2 \geq 0$.

Case 6. In the last case we consider like in Chapter 4 $\sigma^2(x) = 2\theta a x(x-1)$ and $a < 0$. Therefore the process has the following form

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta a x(x-1)}dW_s + \sum_{n=1}^{\mathcal{N}_t} g(J_n, X_{\tau_n -}). \tag{6.17}$$

In the last case we need a restriction on $g(y,x)$ in such a way that $X_t \in (0,1)$. Thus we assume that

$$P_x(-x < g(J,x) < 1-x) = 1. \tag{6.18}$$

Like in Case 2 we do not obtain, that the Condition (6.28) is satisfied for all $0 \leq x_1, x_2 \leq 1$.

6.1 The existence of a solution

In this section we want to deal with the existence of a solution for (6.1). Sørensen (1991) gives us some conditions on (6.1) for the existence of a unique solution. Since the square root in (6.1) is defined in such a way, that $2\theta(aX_s^2 + bX_s + c) \geq 0$, we gain that

$$e(t,x) := \sigma(t,x)\sigma(t,x)^T = \sigma(x)^2 \tag{6.19}$$

is a continuous function on $\mathbb{R}_+ \times \mathbb{R}$ and is everywhere strictly positive definite. In particular, $b(x)$ is bounded on $\{x : ||x|| \leq n\}$ for all $n \in \mathbb{N}$. Those two statements are the first two conditions in Sørensen (1991) (condition C on page 73) to secure a unique solution. It is only left to check the last condition. In our case equation (2.7) in Sørensen (1991) is equal to

$$K(x,A) = \int_{\mathbb{R}} 1_A(g(y,x))F_x(dy) \tag{6.20}$$

for all Borel subsets A of $\mathbb{R}$. From (6.20) we can see that

$$K(x,A) = P_x(g(J,x) \in A).$$

From the last equation we gain that $K(x,\cdot)$ is the distribution of $g(x,J)$ when $J \sim F_x$. Since x is fixed, we can say that $g(x,y) = g_x(y)$. With the assumputions on g in the introduction of Chapter 6, we can gain the density function $h_x(y)$ from $g_x(y)$ with the transformation theorem. We gain

$$h_x(y) = f_x(g_x^{-1}(y))|\partial_y g_x^{-1}(y)|, \tag{6.21}$$

where f_x is the density function of F_x. Thus h_x is the density function of $K(x,\cdot)$. In order to gain the existence of a unique solution for (6.1), we need to show that the function

$$x \to \int_A (|y|^2 \wedge 1)h_x(y)dy \tag{6.22}$$

is continuous on $\mathbb{R}$ for all Borel subsets A of $\mathbb{R}$. We can simplify (6.22) to

$$\begin{aligned} l(x) &:= \int_A (|y|^2 \wedge 1)h_x(y)dy \\ &= \underbrace{\int_{-1}^{1} y^2 1_A(y)h_x(y)dy}_{a)} + \underbrace{\int_{-\infty}^{-1} 1_A(y)h_x(y)dy}_{b)} + \underbrace{\int_{1}^{\infty} 1_A(y)h_x(y)dy}_{c)}\,. \end{aligned} \tag{6.23}$$

It is left to show that (6.23) is a continuous function with respect to x, i.e. that for $x_n \to x$ we can derive $\lim_{n\to\infty} l(x_n) = l(x)$. We want to apply the dominated convergence theorem, because then we can interchange the integral and the limites, i.e.

$$\begin{aligned} \lim_{n\to\infty} l(x_n) &= \lim_{n\to\infty}\left(\int_{-1}^{1} y^2 1_A(y)h_{x_n}(y)dy + \int_{-\infty}^{-1} 1_A(y)h_{x_n}(y)dy + \int_{1}^{\infty} 1_A(y)h_{x_n}(y)dy\right) \\ &= \int_{-1}^{1} y^2 1_A(y)h_{\lim_{n\to\infty} x_n}(y)dy + \int_{-\infty}^{-1} 1_A(y)h_{\lim_{n\to\infty} x_n}(y)dy + \int_{1}^{\infty} 1_A(y)h_{\lim_{n\to\infty} x_n}(y)dy \\ &= l(\lim_{n\to\infty} x_n) = l(x) \end{aligned} \tag{6.24}$$

for any $x_n \to x$. Our next aim is to find dominating functions for $a), b)$ and $c)$ in (6.23) to apply the dominated convergence theorem to gather (6.24).
Under the assumption that $h_x(y)$ is bounded on $[-1, 1]$ we can dominate $1_A(y)y^2h_x(y)$ by

$$\max_{y\in[-1,1]} y^2 h_x(y)$$

and therefore the interchanging of the integral and the limes are allowed for $a)$ in (6.23). The assumpution that $h_x(y)$ is bounded on a compact interval is a rather small one compared to the following.
Since $b)$ and $c)$ in (6.23) need similar assumptions, we are going to discuss them together. There are two convergence theorems which can help us in this case. Both theorems can be used to derive (6.24). First of all the dominated convergence theorem, which has already been mentioned above. We assume that for any x there exists a neighbourhood U that for all $x \in U$ a function $h(y)$ exists, with $h_x(y) \leq h(y)$ and $h(y)$ is integrable on $[1,\infty)$ ($(-\infty,-1]$). Otherwise we can apply the monotone convergence theorem. If $x \to h_x(y)$ is a monotone increasing or decreasing function, then we also gather the desired property.

Example 6.1. Let us consider the case of (6.2). Since $g(x, y)$ still has the same properties, we can consider $w(x)$ is continuous and $w(x) \neq 0$ for all $x \in \mathbb{R}$. Then we get that

$$g_x^{-1}(y) = \frac{y}{w(x)}, \quad \partial_y g_x^{-1}(y) = \frac{1}{w(x)}$$

and

$$h_x(y) = f_x(\frac{y}{w(x)})|\frac{1}{w(x)}|. \tag{6.25}$$

Since $w(x)$ and $(y, x) \mapsto f_x(y)$ are continuous, we obtain that (6.25) is a continuous function. Then (6.23) is in this case equivalent to

$$\begin{aligned} l(x) = &\int_{-1}^{1} 1_A(y)y^2 f_x(\frac{y}{w(x)})|\frac{1}{w(x)}|dy + \int_{-\infty}^{-1} 1_A(y) f_x(\frac{y}{w(x)})|\frac{1}{w(x)}|dy \\ &+ \int_{1}^{\infty} 1_A(y) f_x(\frac{y}{w(x)})|\frac{1}{w(x)}|dy. \end{aligned} \tag{6.26}$$

Furthermore f_x is a density function and then we can dominate $h_x(y)$ by

$$\max_{y\in[-1,1]} y^2 f_x(\frac{y}{w(x)})|\frac{1}{w(x)}|$$

on $[-1, 1]$. In this particular example we do not need to assume a dominated function for the last two integrals in (6.26). We can simply do a transformation again with $y = g_x^{-1}(y)$ and then we gather for example for the third intergral

$$\int_{1}^{\infty} f_x(\frac{y}{w(x)})|\frac{1}{w(x)}|dy = \int_{g_x^{-1}(1)}^{g_x^{-1}(\infty)} f_x(y)dy = \int_{\frac{1}{w(x)}}^{\infty} f_x(y)dy. \tag{6.27}$$

Therefore we need some assumptions on $f_x(y)$ just like we do have them above. In the case where the distribution of jumps are independent from x we get that the last equation is continous in x, because it is equal to

$$1 - F(\frac{1}{w(x)})$$

and with $F(z)$ and $\frac{1}{w(x)}$ being continuous the respect to x we gain the desired property.

Δ

6.2 The existence of moments

Let us now take a look at the existence of moments. In order to get some results we have to constrain our model (6.1) a bit. We assume, that the distribution of the jumps do not depend on the state of the process, i.e. the J_t are i.i.d. with a density function $f(x)$.
In the case where the process satisfies a global Lipschitz condition, i.e. there exists a $K_1 > 0$ such that for all $x_1, x_2 \in \mathbb{R}$

$$\begin{aligned}(c(x_1) - c(x_2))^2 + (\sigma(x_1) - \sigma(x_2))^2 + \lambda \int_{\mathbb{R}} |g(y, x_1) - g(y, x_2)|^2 f(y) dy \\ \leq K_1 |x_1 - x_2|^2,\end{aligned} \tag{6.28}$$

and a growth condition, i.e. there exists a $K_2 > 0$ such that for all $y \in \mathbb{R}$

$$c(x)^2 + \sigma^2(x) + \lambda \int_{\mathbb{R}} |g(y, x)|^2 f(y) dy \leq K_2(1 + |x|^2), \tag{6.29}$$

we can gain, that the first two moments of X_t with $t \geq 0$ exist. Unfortunally the Conditions (6.28) and (6.29) are not given for all six different types of Pearson diffusions.

Theorem 6.2. *Let X_t satisfy the Conditions (6.28) and (6.29). Further we assume that $E(|X_0|^2) < \infty$ with $n \in \mathbb{N}$. Then the first two moments of X_t exist.*

Proof. We define for all $t \geq 0$ and $m \in \mathbb{N}$ the sequence

$$X_t^{(m+1)} = X_0^{(m)} + \int_0^t c(X_s^{(m)}) ds + \int_0^t \sigma(X_s^{(m)}) dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_{s-}^{(m)}) d\tilde{\mathcal{N}}(s, y) \tag{6.30}$$

and $X_t^0 = X_0$ for all $t \geq 0$. We can prove with induction that $X_t^{(m)}$ is adaptive and càdlàg (see Theorem 4.2.12 in Applebaum (2004)). For all $t \geq 0$, we have

$$\begin{aligned}X_t^{(m+1)} - X_t^{(m)} = &\int_0^t \Big(c(X_s^{(m)}) - c(X_s^{(m-1)})\Big) ds + \int_0^t \Big(\sigma(X_s^{(m)}) - \sigma(X_s^{(m-1)})\Big) dW_s \\ &+ \int_0^t \int_{\mathbb{R}} \Big(g(y, X_{s-}^{(m)}) - g(y, X_{s-}^{(m-1)})\Big) d\tilde{\mathcal{N}}(s, y).\end{aligned} \tag{6.31}$$

With induction we can easily proof that for all $x_1, x_2, \ldots, x_k \in \mathbb{R}$ with $k \in \mathbb{N}$

$$|x_1 + x_2 + \ldots + x_k|^2 \leq k(|x_1|^2 + |x_2|^2 + \ldots + |x_k|^2). \tag{6.32}$$

We need some inequalities about $E(\sup_{0\leq s\leq t} |X_s^{(m+1)} - X_s^{(m)}|^2)$ and we start with the simplest case. Then we get with (6.31) and (6.32) for $m = 0$

$$\begin{aligned}
|X_t^{(1)} - X_0|^2 &= \Big|\int_0^t c(X_0)ds + \int_0^t \sigma(X_0)dW_s + \int_0^t \int_{\mathbb{R}} (g(y, X_0)d\tilde{\mathcal{N}}(s,y)\Big|^2 \\
&\leq 3\Big(\Big|\int_0^t c(X_0)ds\Big|^2 + \Big|\int_0^t \sigma(X_0)dW_s\Big|^2 + \Big|\int_0^t \int_{\mathbb{R}} (g(y, X_0)d\tilde{\mathcal{N}}(s,y)\Big|^2\Big) \\
&= 3\Big(\big|tc(X_0)\big|^2 + \big|\sigma(X_0)W_t\big|^2 + \Big|\int_{\mathbb{R}} (g(y, X_0)\tilde{\mathcal{N}}(t,dy)\Big|^2\Big)
\end{aligned}$$

for each $t \geq 0$. Now we take the expectation to obtain

$$\begin{aligned}
E(\sup_{0\leq s\leq t} |X_s^{(1)} - X_0|^2) \leq& 3E\left(\sup_{0\leq s\leq t}\Big(\big(sc(X_0)\big)^2 + \big(\sigma(X_0)W_s\big)^2 + \Big(\int_{\mathbb{R}} g(y, X_0)d\tilde{\mathcal{N}}(t,y)\Big)^2\Big)\right) \\
=& 3E\big(t^2c(X_0)^2\big) + 3E\Big(\sup_{0\leq s\leq t}\big(\sigma(X_0)W_s\big)^2\Big) \\
&+ 3E\Big(\sup_{0\leq s\leq t}\Big(\int_{\mathbb{R}} (g(y, X_0)\tilde{\mathcal{N}}(s,dy)\Big)^2\Big).
\end{aligned} \tag{6.33}$$

With Doob's martingale inequality (see Theorem 11 in the Appendix) we get

$$3E(\sup_{0\leq s\leq t}\big(\sigma(X_0)W_s\big)^2) \leq 12E(|W_t|^2|\sigma(X_0)|^2) = 12tE(|\sigma(X_0)|^2) \tag{6.34}$$

and since $\int_{\mathbb{R}}(g(y, X_0)\tilde{\mathcal{N}}(t,dy)$ is a martingale we gain that $\Big(\int_{\mathbb{R}} g(y, X_0)d\tilde{\mathcal{N}}(t,y)\Big)^2$ is a submartingale. Again with Doob's martingale inequality we gather

$$3E\Big(\sup_{0\leq s\leq t}\Big(\int_{\mathbb{R}} (g(y, X_0)\tilde{\mathcal{N}}(s,dy)\Big)^2\Big) \leq 12E\bigg(\Big(\int_{\mathbb{R}} g(y, X_0)d\tilde{\mathcal{N}}(t,y)\Big)^2\bigg). \tag{6.35}$$

If we apply the Itô isometry (see p. 199 in Applebaum (2004)) we can simplify (6.35) to

$$3E\Big(\sup_{0\leq s\leq t}\Big(\int_{\mathbb{R}} (g(y, X_0)d\tilde{\mathcal{N}}(t,y)\Big)^2\Big) \leq 12t\int_{\mathbb{R}} E(|g(y, X_0)|^2)f(y)dy. \tag{6.36}$$

Then we finally get with (6.33), (6.34) and (6.36) that

$$\begin{aligned}
E(\sup_{0\leq s\leq t} |X_s^{(1)} - X_0|^2) \leq& 3t^2E\big(c(X_0)^2\big) + 12tE\big(|\sigma(X_0)|^2\big) \\
&+ 12t\lambda\int_{\mathbb{R}} E\big(|g(y, X_0)|^2\big)f(y)dy.
\end{aligned}$$

Then the last inequality is with (6.29) equivalent to

$$E(\sup_{0\le s\le t} |X_s^{(1)} - X_0|^2) \le C_1(t)tK_2(1 + E(|X_0|^2)) \tag{6.37}$$

where $C_1(t) = \max\{3t, 12\}$. We have shown that for $m = 0$

$$E(\sup_{0\le s\le t} |X_s^{(1)} - X_0|^2) < \infty.$$

Now we want to consider the case for general $m \in \mathbb{N}$. With similar arguments as above (Doob's martingale inequality) we obtain

$$\begin{aligned} E\Big(\sup_{0\le s\le t} |X_s^{(m+1)} - X_s^{(m)}|^2\Big) \le & 3E\Big(\sup_{0\le s\le t}\Big(\int_0^s (c(X_u^{(m)}) - c(X_u^{(m-1)}))du\Big)^2\Big) \\ & + 12E\Big(\Big(\int_0^t (\sigma(X_s^{(m)}) - \sigma(X_s^{(m-1)}))dW_s\Big)^2\Big) \\ & + 12E\Big(\Big(\int_0^t \int_{\mathbb{R}} g(y, X_{s-}^{(m)}) - g(y, X_{s-}^{(m-1)})\Big)^2 d\tilde{\mathcal{N}}(s,y)\Big). \end{aligned} \tag{6.38}$$

With the Cauchy-Schwarz inequality we get for all $s \ge 0$

$$\Big(\int_0^s (c(X_u^{(m)}) - c(X_u^{(m-1)}))du\Big)^2 \le s\int_0^s (c(X_u^{(m)}) - c(X_u^{(m-1)}))^2 du$$

and with the Itô isometry again we finally obtain

$$\begin{aligned} E(\sup_{0\le s\le t} |X_s^{(m+1)} - X_s^{(m)}|^2) \le & C_1(t)\Big(\int_0^t E(|c(X_s^{(m)}) - c(X_s^{(m-1)})|^2)ds \\ & + \int_0^t E\Big((\sigma(X_s^{(m)}) - \sigma(X_s^{(m-1)}))^2\Big)ds \\ & + \lambda\int_0^t \int_{\mathbb{R}} E\Big((g(y, X_{s-}^{(m)}) - g(y, X_{s-}^{(m-1)}))^2\Big) f(y)dyds\Big). \end{aligned}$$

Since the interchanging of the integrals is allowed and with the Lipschitz condition (6.28), we gather

$$E(\sup_{0\le s\le t} |X_s^{(m+1)} - X_s^{(m)}|^2) \le C_1(t)K_1\int_0^t E\Big(\sup_{0\le u\le s} |X_u^{(m)} - X_u^{(m-1)}|^2\Big)ds. \tag{6.39}$$

Thus we can prove via induction based on (6.37) and (6.39) that

$$E(\sup_{0\le s\le t} |X_s^{(m+1)} - X_s^{(m)}|^2) \le \frac{C_2(t)^m K_3^m}{m!} \tag{6.40}$$

for all $m \in \mathbb{N}$ and where $C_2(t) = tC_1(t)$ and

$$K_3 = \max\{K_1, K_2(1 + E(|X_0|^2))\}.$$

As we know $||\cdot||_2 := (E(|\cdot|^2))^{\frac{1}{2}}$ denotes a norm on $L^2(\mathbb{R}, \mathcal{B}, P)$. With (6.40) we can prove now, that $X_t^{(m)}$ is convergent in L^2. For all $m, l \in \mathbb{N}$ and for all $0 \leq s \leq t$ we observe that

$$||X_s^{(m)} - X_s^{(l)}||_2 \leq \sum_{r=l+1}^{m} ||X_s^{(r)} - X_s^{(r-1)}||_2 \leq \sum_{r=l+1}^{m} \frac{C_2(t)^{\frac{r}{2}} K_3^{\frac{r}{2}}}{(r!)^{\frac{1}{2}}}$$

and since the right series converges, we have that each $(X_s^{(m)}, m \in \mathbb{N})$ is Cauchy. Hence each $(X_s^{(m)}, m \in \mathbb{N})$ is convergent to some $X_s \in L^2(\mathbb{R}, \mathcal{B}, P)$. Then we yield the usefull estimate

$$||X_s - X_s^{(m)}||_2 \leq \sum_{r=m+1}^{\infty} ||X_s^{(r)} - X_s^{(r-1)}||_2 \leq \sum_{r=n+1}^{\infty} \frac{C_2(t)^{\frac{r}{2}} K_3^{\frac{r}{2}}}{(r!)^{\frac{r}{2}}} \tag{6.41}$$

for all $m \in \mathbb{N}, 0 \leq s \leq t$. We also need to establish almost sure convergence of X_t^m for $m \in \mathbb{N}$. If we apply the Chebyshev-Markov inequality ($P(|X| \geq C) \leq \frac{E(|X|^n)}{C^n}$) on (6.40), we gather

$$P\Big(\sup_{0\leq s\leq t} |X_s^{(n+1)} - X_s^{(n)}| \geq \frac{1}{2^n}\Big) \leq \frac{[4K_3 C_2(t)]^n}{n!}$$

and with this inequality we gather with Borel's lemma

$$P\Big(\limsup_{n\to\infty} \big\{ \sup_{0\leq s\leq t} |X_s^{(n+1)} - X_s^{(n)}| \geq \frac{1}{2^n} \big\}\Big) = 0.$$

With

$$P\Big(\liminf_{n\to\infty} D(n)^c\Big) = 1 - P\Big(\limsup_{n\to\infty} D(n)\Big),$$

where $D(n)$ with $n \in \mathbb{N}$ is a sequence of events in $\mathcal{B}$, we gather that

$$P\Big(\liminf_{n\to\infty} \big\{ \sup_{0\leq s\leq t} |X_s^{(n)} - X_s^{(n-1)}| < \frac{1}{2n} \big\}\Big) = 1.$$

Hence given any $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that for $n, l > n_0$, we have

$$\sup_{0\leq s\leq t} |X_s^{(n)} - X_s^{(l)}| \leq \sum_{r=l}^{n-1} \sup_{0\leq s\leq t} |X_s^{(r+1)} - X_s^{(r)}| < \sum_{r=l}^{n-1} \frac{1}{2^r} < \delta$$

with probability 1, from which we see that $X_t^{(n)}$ for $n \in \mathbb{N}$ is almost surely uniformly Cauchy on compact intervals and hence is almost surely convergent on compact intervals. From this we conclude that X_t is adapted and càdlàg.
It is left to show that X_t satisfies (6.7). Then we define $\tilde{X}_t$ for all $t \geq 0$ the stochastic process by

$$\tilde{X}_t = X_0 + \int_0^t c(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t \int_{\mathbb{R}} g(y, X_s)d\tilde{\mathcal{N}}(s, y).$$

Hence for each $n \in \mathbb{N}$ we gather

$$\begin{aligned}\tilde{X}_t - X_t^{(n)} &= \int_0^t \Big(c(X_s) - c(X_s^{(n)})\Big) ds + \int_0^t \Big(\sigma(X_s) - \sigma(X_s^{(n)})\Big) dW_s \\ &+ \int_0^t \int_{\mathbb{R}} \Big(g(y, X_s) - g(y, X_s^{(n)})\Big) d\tilde{\mathcal{N}}(s, y).\end{aligned}$$

Now we can apply the same arguments which led us to (6.39) and apply (6.41), and obtain for all $0 \leq s \leq t < \infty$

$$\begin{aligned}E(|\tilde{X}_s - X_s^{(n)}|^2) &\leq C_1(t) K_1 \int_0^t E(|X_u - X_u^{(n)}|^2) du \\ &\leq C_2(t) K_1 \sup_{0 \leq u \leq t} E(|X_u - X_u^{(n)}|^2) \\ &\leq C_2(t) K_1 \left(\sum_{r=n+1}^{\infty} \frac{C_2(t)^{\frac{r}{2}} K_3^{\frac{r}{2}}}{(r!)^{\frac{1}{2}}} \right)^2 \\ &\to 0 \quad \text{as} \quad n \to \infty.\end{aligned}$$

Hence each $\tilde{X}_s = L^2 - \lim_{n \to \infty} X_s^{(n)}$ and so, by uniqueness of limits $\tilde{X}_t = X_t$ as required. Now we can prove the existence of the first two moments. With (6.41) we gain especially

$$||X_t - X_0||_2 \leq \sum_{r=1}^{\infty} ||X_t^{(r)} - X_t^{(r-1)}||_2 \leq C(t)$$

where $C(t) \geq 0$ and then we get

$$E(|X_t|^2) \leq 2 \underbrace{E(|X_t - X_0|^2)}_{\leq C(t)^2} + 2 \underbrace{E(|X_0|^2)}_{< \infty} < \infty.$$

□

Due to the fact that (6.30) has to be derived for higher moments from the Itô formula with $f(x) = x^n$ for $n \in \mathbb{N}$ and the fact that a estimation like (6.39) is not possible, because the Lipschitz condition does not apply, we have to find another way to proof the existence of higher moments.
In the book of Gihman and Skorohod (1972) (Theorem 4 in Chaper 6) is a proof for the existence of higher moments in the case without jumps. We want to investigate now whether it is possible to extend the proof for our case. At first we calculate the SDE for X_t^n. Let f be a twice continuously differential function, then we get if we apply the Itô

formula (the version with jumps see the appendix) on (6.1)

$$
\begin{aligned}
f(X_t) =& f(X_0) + \int_0^t f'(X_{s-})dX_s + \frac{1}{2}\int_0^t f''(X_{s-})\sigma^2(X_s)ds \\
&+ \sum_{n=1}^{\mathcal{N}_t} (\underbrace{f(X_{\tau_{n-}} + g(J_n, X_{\tau_{n-}})) - f(X_{\tau_{n-}})}_{=:S_{\tau_n}^{J_n}} - f'(X_{\tau_n -})g(J_n, X_{\tau_{n-}})) \\
=& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\big)ds \\
&+ \int_0^t \sigma(X_s)f'(X_s)dW_s + \sum_{n=1}^{\mathcal{N}_t} S_{\tau_n}^{J_n}.
\end{aligned}
\tag{6.42}
$$

Now we apply the same trick as we have done it before and use the measures $\Psi(s, y)$ and $\Lambda(s, y)$ as we defined in (6.4) and (6.5).
Then we gather

$$
\begin{aligned}
f(X_t) =& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\big)ds \\
&+ \int_0^t \sigma(X_s)f'(X_s)dW_s + \int_0^t \int_{\mathbb{R}} S_s^y d\Psi(s, y) \\
=& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\big)ds \\
&+ \underbrace{\int_0^t \sigma(X_s)f'(X_s)dW_s + \int_0^t \int_{\mathbb{R}} S_s^y (d\tilde{\mathcal{N}}(s, y))}_{=:M_t} + \int_0^t \int_{\mathbb{R}} S_s^y d\Lambda(s, y) \\
=& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\big)ds \\
&+ M_t + \int_0^t \int_{\mathbb{R}} S_s^y d\Lambda(s, y),
\end{aligned}
\tag{6.43}
$$

where M_t is a martingale (see the Appendix for further information) with $M_0 = 0$ (note that $E(M_t) = 0$). In order for the next theorem we need some more definitions, therefore we define for some $N \in \mathbb{R}_+$

$$
\begin{aligned}
X_{0,N} &= \begin{cases} X_0 & \text{for } |X_0| \leq N \\ N\text{sign}(X_0) & \text{for } |X_0| > N \end{cases}, \\
b_N(x) &= \begin{cases} b(x) & \text{for } |x| \leq N \\ b(N\text{sign}(x)) & \text{for } |x| > N \end{cases}, \\
h_N(x) &= \begin{cases} h(x) & \text{for } |x| \leq N \\ h(N\text{sign}(x)) & \text{for } |x| > N \end{cases}, \\
\sigma_N(x) &= \begin{cases} \sigma(x) & \text{for } |x| \leq N \\ \sigma(N\text{sign}(x)) & \text{for } |x| > N \end{cases}
\end{aligned}
$$

and

$$g_N(y,x) = \begin{cases} g(y,x) & \text{for } |x| \leq N \\ g(y, N\text{sign}(x)) & \text{for } |x| > N \end{cases},$$

where sign $(x) = 1$ if $x > 0$ and sign $(x) = -1$ if $x < 0$. Then we can define the bounded process

$$X_{t,N} = X_{0,N} + \int_0^t b_N(X_{s,N})ds + \int_0^t \sigma_N(X_{s,N})dW_s + \int_0^t \int_{\mathbb{R}} g_N(y, X_{s,N})\Psi(ds, dy),$$

which is equivalent to

$$X_{t,N} = X_{0,N} + \int_0^t c_N(X_{s,N})ds + \int_0^t \sigma_N(X_{s,N})dW_s + \int_0^t \int_{\mathbb{R}} g_N(y, X_{s,N})\tilde{\mathcal{N}}(ds, dy), \tag{6.44}$$

where

$$c_N(x) = b_N(x) + \lambda h_N(x) = b_N(x) + \lambda \int_{\mathbb{R}} g_N(y,x) f(y)dy.$$

In addition to the proof we also need some other conditions than (6.28) and (6.29) but they are very similar. For $N > 0$ and for all $|x|, |y| \leq N$ there exist a $K_{1,N} > 0$ and a $K_2 > 0$ such that

$$|c(x) - c(y)| + |\sigma(x) - \sigma(y)| \leq K_{1,N}|x - y| \tag{6.45}$$

and

$$xb(x) + \sigma^2(x) \leq K_{2,N}(1 + x^2) \tag{6.46}$$

for all $t \in [0, T]$ holds. Then we gain the following theorem.

Theorem 6.3. *Assume that the coefficients of (6.1) satisfy the conditions (6.45) and (6.46), that J_t is independent from the state of the procees (i.e. $J \sim F$) and that $E(X_0^{2n}) < \infty$. Then we can find a C, depending only on n, K_2 and $D > 0$, for which*

$$E(X_t^{2n}) \leq E(1 + X_0^{2n})e^{Ct} < \infty. \tag{6.47}$$

Proof. In order to take expecatations on (6.43) for the bounded process (6.44) for $f(x) =$

x^{2n}, we need to prove that the expectation of $X_{t,N}^{2n}$ is bounded. We get with (6.44)

$$\begin{aligned}
E\Big((X_{t,N})^{2n}\Big) =& E\Bigg(4^{2n}\Big(\frac{1}{4}X_{0,N} + \frac{1}{4}\int_0^t c_N(X_{s,N})ds \\
&+ \frac{1}{4}\int_0^t \sigma_N(X_{s,N})dW_s + \frac{1}{4}\int_0^t\int_{\mathbb{R}} g_N(y,X_{s,N})\tilde{\mathcal{N}}(ds,dy)\Big)^{2n}\Bigg) \\
\leq& 4^{2n-1}E\Big(\underbrace{X_{0,N}^{2n}}_{a}\Big) + 4^{2n-1}\underbrace{E\Big(\Big(\int_0^t c_N(X_{s,N})ds\Big)^{2n}\Big)}_{b} \\
&+ 4^{2n-1}\underbrace{E\Big(\Big(\int_0^t \sigma_N(X_{s,N})dW_s\Big)^{2n}\Big)}_{c} \\
&+ 4^{2n-1}\underbrace{E\Big(\Big(\int_0^t\int_{\mathbb{R}} g_N(y,X_{s,N})\tilde{\mathcal{N}}(ds,dy)\Big)^{2n}\Big)}_{d}. \qquad (6.48)
\end{aligned}$$

Hence it is suffient to show that all part of (6.48) are finite. Then a in (6.48) is finte since $E(X_{0,N}^{2n}) \leq E(X_0^{2n}) < \infty$. Now let us take a look at b in (6.48), then we obtain

$$\Big(\int_0^t c_N(X_{s,N})ds\Big)^{2n} \leq \int_0^t \Big(c_N(X_{s,N})\Big)^{2n} ds.$$

Since $b_N(x)$ is bounded by a constant B and $h_N(x)$ is bounded by a constant H, we gain that $c_N(x) \leq B + \lambda H =: C$. Thus we gather

$$\Big(\int_0^t c_N(X_{s,N})ds\Big)^{2n} \leq \int_0^t \Big(c_N(X_{s,N})\Big)^{2n} ds \leq C^{2n}\int_0^t ds = C^{2n}t < \infty.$$

We gain the finiteness of the parts d and c of (6.48) on the same way (similarly arguments like in the proof of theorem 6.2), since both integrals are martingales. Thus we only consider part c. Then we gain

$$E\Big((\int_0^t \sigma_N(X_{s,N})dW_s)^{2n}\Big) \leq E\Bigg(\sup_{0\leq s\leq t}\Big(\Big(\int_0^s \sigma_N(X_{u,N})dW_u\Big)^2\Big)^n\Bigg)$$

and with the Doob martingale inequality (see the Appendix) we gather

$$E\Bigg(\sup_{0\leq s\leq t}\Big(\Big(\int_0^s \sigma_N(X_{u,N})dW_u\Big)^2\Big)^n\Bigg) \leq \Big(\frac{n}{n-1}\Big)^n E\Big(\Big(\int_0^t \sigma_N(X_{s,N})dW_s\Big)^2\Big).$$

Now we apply the Itô- Isometrie to gain

$$\Big(\frac{n}{n-1}\Big)^n E\Big(\Big(\int_0^t \sigma_N(X_{s,N})dW_s\Big)^2\Big) = \Big(\frac{n}{n-1}\Big)^n \int_0^t E\Big((\sigma(X_{s,N}))^2\Big)ds$$

and thus

$$E\Big(\big(\int_0^t \sigma_N(X_{s,N})dW_s\big)^{2n}\Big) \leq \left(\frac{n}{n-1}\right)^n \int_0^t \underbrace{E\Big((\sigma(X_{s,N}))^2\Big)}_{\leq A} ds$$
$$\leq \left(\frac{n}{n-1}\right)^n At < \infty.$$

Now we take expectation at both sides of (6.43) and then we gain with $f(x) = x^{2n}$

$$E(X_t^{2n}) = E(X_0^{2n}) + E\Bigg(\int_0^t \Big(2nb_N(X_s)X_s^{2n-1} + n(2n-1)\sigma^2(X_s)X_s^{2n-2}\Big)ds + \int_0^t \int_{\mathbb{R}} S_s^y \Lambda(ds, dy)\Big)\Bigg) \tag{6.49}$$

with

$$S_s^y = (X_{s-} + g(y, X_{s-}))^{2n} - X_{s-}^{2n} \tag{6.50}$$

and remember that the expectations of the martingales are zero. Then we take a closer look at (6.50) and we can simplify it to

$$\begin{aligned} S_s^y &= (X_{s-} + g(y, X_{s-}))^{2n} - X_{s-}^{2n} \\ &= \sum_{i=0}^{2n} \binom{2n}{i} X_{s-}^{2n-i} g(y, X_{s-})^i - X_{s-}^{2n} \\ &= \sum_{i=0}^{2n-1} \binom{2n}{i+1} X_{s-}^{2n-i-1} g(y, X_{s-})^{i+1} \end{aligned}$$

and then we get, if we use (6.3),

$$\begin{aligned} \int_0^t \int_{\mathbb{R}} S_s^y d\Lambda(s, y) &= \lambda \int_0^t \int_{\mathbb{R}} \sum_{i=0}^{2n-1} \binom{2n}{i+1} X_{s-}^{2n-i-1} g(y, X_{s-})^{i+1} f(y) dy ds \\ &= \int_0^t \sum_{i=0}^{2n-1} \binom{2n}{i+1} X_{s-}^{2n-i-1} \underbrace{\lambda \int_{\mathbb{R}} g(y, X_{s-})^{i+1} f(y) dy}_{=\lambda E(g(J, X_{s-})^{i+1})} ds \\ &= \int_0^t \sum_{i=0}^{2n-1} \binom{2n}{i+1} X_{s-}^{2n-i-1} \underbrace{E(g(J, X_{s-})^{i+1})}_{\sum_{j=0}^{i+1} \alpha_{j,i+1} X_{s-}^j} ds \\ &= \int_0^t \sum_{i=0}^{2n-1} \binom{2n}{i+1} \sum_{j=0}^{i+1} \alpha_{j,i+1} X_{s-}^{2n-1-i+j} ds \end{aligned}$$

since $j \leq i$ we see that the sums form a polynomial of X_{s-} with the $2n$ the highest order. Therefore we can define

$$\sum_{i=0}^{2n-1} \binom{2n}{i+1} \sum_{j=0}^{i+1} \alpha_{j,i+1} X_{s-}^{2n-1-i+j} = \sum_{k=0}^{2n} d_k X_{s-}^k,$$

where

$$d_k = \sum_{i=2n-k}^{2n} \binom{2n}{i} \alpha_{i-2n+k,i}.$$

Then we gain

$$\int_0^t \int_{\mathbb{R}} S_s^y d\Lambda(s,y) = \int_0^t \sum_{k=0}^{2n} d_k X_{s-}^k ds \tag{6.51}$$

Therefore we gain with the last equation and (6.49) that

$$E(X_t^{2n}) = E(X_0^{2n}) + E\Bigg(\int_0^t \Big(2nb_N(X_s) X_s^{2n-1} + n(2n-1)\sigma^2(X_s) X_s^{2n-2} + \sum_{k=0}^{2n} d_k X_{s-}^k \Big) ds \Bigg). \tag{6.52}$$

Then we get for the bounded process $X_{t,N}^{2n}$

$$E(X_{t,N}^{2n}) = E(X_{0,N}^{2n}) + E\Bigg(\int_0^t \Big(2nb_N(X_{s,N}) X_{s,N}^{2n-1} + n(2n-1)\sigma_N^2(X_{s,N}) X_{s,N}^{2n-2} + \sum_{k=0}^{2n} d_k X_{s-,N}^k \Big) ds \Bigg). \tag{6.53}$$

Let us take a look a the polynominal again. Then we can derive the following inequalitiy

$$\sum_{i=0}^{2n} d_i x^i \leq D(1 + 2x^{2n})$$

with $D > 0$ constant. Then we gain with (6.53)

$$E(X_{t,N}^{2n}) \leq E(X_{0,N}^{2n}) + E\Bigg(\int_0^t \bigg(\Big(2nb_N(X_{s,N}) X_{s,N} + n(2n-1)\sigma_N^2(X_{s,N}) \Big) X_{s,N}^{2n-2} + D(1 + 2X_{s,N}^{2n}) \bigg) ds \Bigg)$$

and then we derive with (6.46)

$$E(X_{t,N}^{2n}) \leq E(X_{0,N}^{2n}) + (2n+1)nK_2 \int_0^t E\Big((1+X_{s,N}^2)X_{s,N}^{2n-2} + D(1+2X_{s,N}^{2n})\Big)ds$$

and with

$$(1+x^2)x^{2n-2} = x^{2n-2} + x^{2n} \leq 1 + x^{2n} + x^{2n} = 1 + 2x^{2n}$$

we get

$$\begin{aligned} E(X_{t,N}^{2n}) &\leq E(X_0^{2n}) + (2n+1)nK \int_0^t (1 + 2E(X_{s,N}^{2n}))ds \\ &\leq E(X_0^{2n}) + (2n+1)nKt + (2n+1)2nK \int_0^t E(X_{s,N}^{2n}))ds. \end{aligned}$$

where $K = \max\{D, K_2\}$. Then we get with Lemma of Gronwall

$$\begin{aligned} E(X_{t,N}^{2n}) &\leq E(X_0^{2n}) + (2n+1)nKt \\ &+ (2n+1)2nK \int_0^t e^{2n(2n+1)K(t-s)}\Big(E(X_0^{2n}) + s(2n+1)nK\Big)ds. \end{aligned}$$

Hence follows the inequality (6.47) for $X_{t,N}^{2n}$ with $C = n(2n+1)K$.
The last step is to prove that equation also holds for X_t^{2n} and not only for $X_{t,N}^{2n}$. Therefore we define

$$\Omega_N = \{\omega \in \mathbb{R} | \sup_{0\leq s\leq t} |X_{s,N}^{2n}(\omega)| \leq N\}.$$

It is obvious that $\Omega_N \to \mathbb{R}$ for $N \to \infty$. Further we gain that

$$E(X_{t,N}^{2n}) = E(X_{t,N}^{2n} 1_{\Omega_N}) = E(X_t^{2n} 1_{\Omega_N}) \to E(X_t^{2n})$$

for $N \to \infty$. On the other hand

$$E(X_{t,N}^{2n}) \leq E(1 + X_0^{2n})e^{Ct} < \infty$$

where the constant C does not depend on N. □

6.3 Conditional moments

If we assume that the first n moments exist, then we can derive a formula for the conditional moments. In order to do so, we apply the Itô formula (the version with jumps see the appendix) on (6.1) and we gain, like in (6.42),

$$\begin{aligned} f(X_t) =& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma^2(X_s)f''(X_s)\big)ds \\ &+ \int_0^t \sigma(X_s)f'(X_s)dW_s + \sum_{n=1}^{\mathcal{N}_t} S_{\tau_n}^{J_n} \end{aligned}$$

with $f:\mathbb{R}\to\mathbb{R}$ twice continuous differentiable function and

$$S_s^y = f(X_{s-} - g(y, X_{s-})) - f(X_{s-}).$$

We would like to get an equation of the form

$$f(X_t) = f(X_0) + \int_0^t Af(X_s)ds + M_t \tag{6.54}$$

where M_t is a martingale (in particluar $M_0 = 0$) and A an operator. We still assume, like in Chapter 4, that $\int_0^t \sqrt{2\theta(aX_s^2 + bX_s + c)}f'(X_s)dW_s$ is a true martingale. In the case, that $f(x)$ is a polynomial of finite order, we gain that it is a true martingale. We still need to deal with the last term $(\sum_{n=1}^{\mathcal{N}_t} S_{\tau_n}^{J_n})$. With $\Lambda(s, y)$ like (6.5) and with $\Psi(s, t)$ like (6.4) we gain

$$\begin{aligned}
f(X_t) =& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma(X_s)^2 f''(X_s)\big)ds + \int_0^t \sigma(X_s)f'(X_s)dW_s \\
&+ \sum_{n=1}^{N_t} S_{\tau_n}^{J_n} \\
=& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma(X_s)^2 f''(X_s)\big)ds + \int_0^t \sigma(X_s)f'(X_s)dW_s \\
&+ \int_0^t \int_{\mathbb{R}} S_s^y \Psi(ds, dy) \\
=& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma(X_s)^2 f''(X_s)\big)ds + \int_0^t \sigma(X_s)f'(X_s)dW_s \\
&+ \int_0^t \int_{\mathbb{R}} S_s^y \tilde{\mathcal{N}}(ds, dy) + \int_0^t \int_{\mathbb{R}} S_s^y \Lambda(ds, dy)
\end{aligned}$$

where $\tilde{\mathcal{N}}(s, y) = \Psi(s, y) - \Lambda(s, y)$. It can be shown (see the Appendix), that $\int_0^t \int_{\mathbb{R}} S_s^y \tilde{\mathcal{N}}(ds, dy)$ is a martingale if S_s^y is predictable, which it is since S_s^y only depends on $X_{\tau_n -}$. If we further define

$$M_t := \int_0^t \sigma(X_s)f'(X_s)dW_s + \int_0^t \int_{\mathbb{R}} S_s^y \tilde{\mathcal{N}}(ds, dy), \tag{6.55}$$

we obtain

$$\begin{aligned}
f(X_t) =& f(X_0) + \int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma(X_s)^2 f''(X_s)\big)ds \\
&+ M_t + \int_0^t \int_{\mathbb{R}} S_s^y \Lambda(ds, dy).
\end{aligned} \tag{6.56}$$

Since we assumed that M_t is a martingale, we obtain from (6.56)

$$Af(x) = b(x)f'(x) + \frac{1}{2}\sigma(x)^2 f''(x) + \int_{\mathbb{R}} \lambda(f(x + g(y, x)) - f(x))F_x(dy). \tag{6.57}$$

If we take the conditional expectation on both sides of (6.56), we get

$$E(f(X_t)|X_0) = f(X_0) + E\Big(\int_0^t \big(b(X_s)f'(X_s) + \frac{1}{2}\sigma(X_s)^2 f''(X_s)\big)ds + \int_0^t \int_{\mathbb{R}} S_s^y \Lambda(ds, dy)|X_0\Big),$$

under the assumption that M_t is a true martingale with $M_0 = 0$. Then we get the conditional moments of X_t, by using $f(x) = x^k$ for $k \geq 1$ and $X_0 = x$,

$$E_x(X_t^k) = \underbrace{X_0^k}_{x^k} + E_x\bigg(\int_0^t \Big(b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} + \int_{\mathbb{R}} S_s^y \Lambda(ds, dy)\Big)ds\bigg).$$

We still can simplify the equation under those assumptions we made above.
Since $S_s^y = f(X_{s-} + g(y, X_{s-})) - f(X_{s-}) = (X_{s-} + g(y, X_{s-}))^k - X_{s-}^k$ and $\Lambda(ds, dy) = \lambda dF_{X_{s-}}(y)ds$, we get

$$E_x(X_t^k) = x^k + E_x\Bigg(\int_0^t b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} + \lambda \int_{\mathbb{R}} \Big((X_{s-} + g(y, X_{s-}))^k - X_{s-}^k\Big) dF_{X_{s-}}(y)ds\Bigg),$$

and finally

$$E_x(X_t^k) = x^k + E_x\Bigg(\int_0^t b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} + \lambda E_{F_{X_s}}\Big((X_s + g(J_s, X_s))^k - X_s^k\Big)ds\Bigg). \tag{6.58}$$

The last term of (6.58) we can simplify and it is then equivalent to

$$\begin{aligned}
\lambda E_{F_{X_s}}\Big(\big(X_s + g(J_s, X_s)\big)^k - X_s^k\Big) &= \lambda \sum_{j=0}^{k} \binom{k}{j} E_{F_{X_s}}(g(J_s, X_s)^j X_s^{k-j}) - E_{F_{x^s}}(X_s^k) \\
&= \lambda \sum_{j=1}^{k} \binom{k}{j} E_{F_{X_s}}(g(J_s, X_s)^j X_s^{k-j}) \\
&= \sum_{j=1}^{k} \binom{k}{j} X_s^{k-j} \lambda E_{F_{X_s}}(g(J_s, X_s)^j).
\end{aligned} \tag{6.59}$$

If the interchanging of expectation and integration is allowed we can derive an ordinary differential equation from (6.58) and (6.59)

$$\partial_s E_x(X_s^k) = E_x[b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} + \sum_{j=1}^{k} \binom{k}{j} X_{s-}^{k-j} \lambda E_{F_{X_s}}(g(J_s, X_s)^j)]$$

and by using the Condition (6.3) we get the following equation

$$\begin{aligned}\partial_s E_x(X_t^k) =& E_x(b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} \\ &+ \sum_{j=1}^{k} \binom{k}{j} X_s^{k-j} \sum_{i=0}^{j} \alpha_{i,j} X_s^i).\end{aligned} \tag{6.60}$$

Assumption (6.2) is a very important one and it is also large restriction. In order to get a recursive formula, there are a lot of constraints on the function $w(x)$. If for instance $w(x)$ is a polynomial of a order greater than one, then we can see from (6.60), that the kth moment would also depend on the $(k+1)$th moment and so on, i.e. we would not get a triangular matrix in (6.61). Therefore there are only limited types of functions, which $w(x)$ can take the form of.
If we denote with $E_x(\underline{X_t}) = E(\underline{X_t}|X_0 = x) = [E_x(X_t), E_x(X_t^2), \ldots, E_x(X_t^k)]'$ we can derive from (6.60) a linear differential equation system

$$\partial_s E_x(\underline{X_t}) = A(\beta)E_x(\underline{X_t}) + g(\beta) \tag{6.61}$$

where $A(\beta)$ is a lower-triangular $k \times k$ matrix, $g(\beta)$ is a $k \times 1$ vector and $\beta = (\theta, \mu, a, b, c, \alpha_{0,1}, \ldots, \alpha_{k,k})$. With the theory about differential equations we get the following solution

$$E_x(\underline{X_t}) = e^{tA(\beta)}\underline{X_0} + A(\beta)^{-1}(e^{tA(\beta)} - I)g(\beta) \tag{6.62}$$

where I is the $k \times k$ identical matrix and $\underline{X_0} = [x, x^2, \ldots, x^k]'$.

Example 6.4. Let us take a look again at Example 4.6 in Chapter 4. Here is the diffusion given by

$$dZ_t = -\theta Z_t dt + \sqrt{2\theta(\nu-1)^{-1}(Z_t^2 + 2\rho\nu^{\frac{1}{2}}Z_t + (1+\rho^2)\nu)}dW_t$$

with $\nu \neq 3$ and $Z_0 = z$. Since this is still a case without jumps, we should get the same results for the first two conditional moments as before if we apply (6.61) and (6.62). In this case we get

$$\begin{aligned}\partial_t E_z(Z_t^k) =& \left(-k\theta + \theta\frac{k(k-1)}{\nu-1}\right)E_z(Z_t^k) + 2\frac{\rho\nu^{\frac{1}{2}}}{\nu-1}k(k-1)E_z(Z_t^{k-1}) \\ &+ \frac{k(k-1)}{\nu-1}(1+\rho^2)\nu\theta E_z(Z_t^{k-2}).\end{aligned}$$

In paticular we obtain for $k = 1$

$$\partial_t E_z(Z_t) = -\theta E_z(Z_t)$$

and for $k = 2$

$$\partial_t E_z(Z_t^2) = (-2\theta + \frac{2\theta}{\nu-1})E_z(Z_t^2) + \frac{4\rho\nu^{\frac{1}{2}}}{\nu-1}E_z(Z_t) + \frac{1+\rho^2}{\nu-1}2\nu\theta.$$

Now we can read off $A(\beta)$ and $g(\beta)$ from (6.61)

$$A(\beta) = \begin{pmatrix} -\theta & 0 \\ \frac{4\rho\nu^{\frac{1}{2}}}{\nu-1} & -\theta(\frac{2\nu-4}{\nu-1}) \end{pmatrix},$$

$$A^{-1}(\beta) = \frac{\nu-1}{\theta^2(2\nu-4)} \begin{pmatrix} -\theta(\frac{2\nu-4}{\nu-1}) & 0 \\ -\frac{4\rho\nu^{\frac{1}{2}}}{\nu-1} & -\theta \end{pmatrix}$$

and

$$g(\beta) = \begin{pmatrix} 0 \\ \frac{1+\rho^2}{\nu-1}2\nu\theta \end{pmatrix}.$$

To gain the solution we just need to calculate

$$E_z(\underline{Z_t}) = e^{tA(\beta)}\underline{Z_0} + A(\beta)^{-1}(e^{tA(\beta)} - I)g(\beta). \tag{6.63}$$

Therefore we need $e^{tA(\beta)} = \sum_{i=0}^{\infty} \frac{(tA(\beta))^i}{i!}$. In order to do so we need the eigenvalues of $tA(\beta)$, which are given in this case by $\lambda_1 = -t\theta$ and $\lambda_2 = -t\theta\frac{2\nu-4}{\nu-1}$, and their eigenvectors e_1 and e_2, since $tAe_l = \lambda_l e_l$ and $(tA)^k e_l = \lambda_l^k e_l$ it follows that $e^{tA}e_l = e^{\lambda_l}e_l$ for $l = 1, 2$. We get the eigenvectors with simple calculation to be

$$e_1 = \begin{pmatrix} 1 \\ \frac{4\rho\nu^{\frac{1}{2}}}{3-\nu} \end{pmatrix} \text{ and } e_2 = \begin{pmatrix} 0 \\ \frac{4\rho\nu^{\frac{1}{2}}}{3-\nu} - z \end{pmatrix}.$$

Now we can express $\underline{Z_0} = \begin{pmatrix} z \\ z^2 \end{pmatrix}$ with e_1 and e_2 in the following way

$$\underline{Z_0} = z(e_1 - e_2).$$

With this information we can gain the first term of (6.63)

$$e^{tA(\beta)}\underline{Z_0} = \begin{pmatrix} ze^{-\theta t} \\ z\frac{4\rho\nu^{\frac{1}{2}}}{3-\nu}e^{-\theta t} + z^2 e^{-\theta t\frac{2\nu-4}{\nu-1}} - z\frac{4\rho\nu^{\frac{1}{2}}}{3-\nu}e^{-\theta t\frac{2\nu-4}{\nu-1}} \end{pmatrix}.$$

Similarly we get for the second term in (6.63)

$$\begin{pmatrix} 0 \\ \frac{(1-\rho^2)\nu}{\nu-2}(1 - e^{-t\theta\frac{2\nu-4}{\nu-1}}) \end{pmatrix}.$$

With these information we get

$$E_z(\underline{Z_t}) = \begin{pmatrix} ze^{-t\theta} \\ e^{-\frac{2\nu-4}{\nu-1}\theta t}z^2 + \frac{4\gamma\nu^{\frac{1}{2}}}{\nu-3}\left(e^{-\theta t} - e^{-\frac{2\nu-4}{\nu-1}\theta t}\right)z + \frac{(1+\gamma^2)\nu}{\nu-2}(1 - e^{-\frac{2\nu-4}{\nu-1}\theta t}) \end{pmatrix}$$

which is the same result as in Example 4.6 in Chapter 4.

Δ

Example 6.5. If we assume a process of the following form

$$X_t = X_0 - \int_0^t \theta(X_s - \mu)ds + \int_0^t \sqrt{2\theta(aX_s^2 + bX_s + c)}dW_s + \sum_{n=1}^{\mathcal{N}_t} J_n X_{\tau_{n-}} \tag{6.64}$$

where J is i.i.d. exponential distributed with a paramtere $\psi \geq 1$ and $X_0 = x$. Since the kth moment of an exponential distributed random variable is known to be $E(J^k) = \frac{k!}{\psi^k}$, thus (6.3) is equal to

$$\lambda E_{F_{X_{\tau_{n-}}}}(J_n^k X_{\tau_{n-}}^k) = \alpha_{k,k} X_{\tau_{n-}}^k,$$

i.e. $\alpha_{k,k} = \lambda \frac{k!}{\psi^k}$ and $\alpha_{j,k} = 0$ for $j \neq k$. Therefore we get

$$\begin{aligned}\partial_s E_x(X_s^k) =& E_x\left(b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} + \sum_{j=1}^{k} \begin{pmatrix} k \\ j \end{pmatrix} \alpha_{j,j} X_{s-}^k \right) \\ =& E_x\left(b(X_s)kX_s^{k-1} + \sigma(X_s)^2 k(k-1)X_s^{k-2} + \beta_k X_{s-}^k \right)\end{aligned}$$

with $\beta_k = \sum_{j=1}^{k} \begin{pmatrix} k \\ j \end{pmatrix} \alpha_{j,j}$. Furthermore we assume that $b(X_s) = -\theta(X_s - \mu)$, $\sigma(X_s) = \sqrt{2\theta(aX_s^2 + bX_s + c)}$ and $X_0 = x$. Then we get

$$\begin{aligned}\partial_s E(X_s^k) =& (\beta_k - \theta k + 2\theta k(k-1)a)E_x(X_s^k) + (\theta k\mu + 2b\theta k(k-1))E_x(X_s^{k-1}) \\ &+ 2\theta ck(k-1)E_x(X_s^{k-2})).\end{aligned}$$

Now we can take the same approach as in the example above. We get for $k = 1$

$$\partial_s E_x(X_s) = (\beta_1 - \theta)E_x(X_s) + \theta\mu$$

and for $k = 2$

$$\partial_s E_x(X_s^2) = (\beta_2 - 2\theta + 4a\theta)E_x(X_s^2) + (4b\theta - 2\mu\theta)E_x(X_s) + 4\theta c$$

and then we gain

$$\begin{aligned}A(\beta) &= \begin{pmatrix} \beta_1 - \theta & 0 \\ 4b\theta - 2\mu\theta & \beta_2 - 2\theta + 4a\theta \end{pmatrix}, \\ g(\beta) &= \begin{pmatrix} \theta\mu \\ 4c\theta \end{pmatrix}\end{aligned}$$

and

$$A(\beta)^{-1} = \frac{1}{(\beta_2 - 2\theta + 4a\theta)(\beta_1 - \theta)} \begin{pmatrix} \beta_2 - 2\theta + 4a\theta & 0 \\ -4b\theta + 2\mu\theta & \beta_1 - \theta \end{pmatrix}.$$

The eigenvalues of $tA(\beta)$ are then $\lambda_1 = (\beta_1 - \theta)t$ and $\lambda_2 = (\beta_2 - 2\theta + 4a\theta)t$. Then we gain with the same method as in the example before

$$e^{tA(\beta)}\underline{X_0} = \begin{pmatrix} xe^{\lambda_1} \\ dx(e^{\lambda_1} - e^{\lambda_2}) + x^2 e^{\lambda_2} \end{pmatrix}$$

where $d := \frac{t(2\mu\theta - 4b\theta)}{\lambda_1 - \lambda_2}$ and

$$(e^{tA(\beta)} - I)g(\beta) = \begin{pmatrix} \theta\mu(e^{\lambda_1} - 1) \\ d\theta\mu(e^{\lambda_1} - e^{\lambda_2}) - 4\theta c(e^{\lambda_2} + 1) \end{pmatrix}.$$

Now we finally get

$$E_x(X_t) = xe^{\lambda_1} + \frac{t}{\lambda_1}\theta\mu(e^{\lambda_1} - 1)$$

and

$$\begin{aligned} E_x(X_t^2) =& dx(e^{\lambda_1} - e^{\lambda_2}) + x^2 e^{\lambda_2} \\ & - \frac{\theta t}{\lambda_2}\left(\frac{2t\theta\mu}{\lambda_1}(e^{\lambda_1} - 1)(2b - \mu) - d\mu(e^{\lambda_1} - e^{\lambda_2}) - 4c(e^{\lambda_2} - 1))\right). \end{aligned}$$

Δ

6.4 The generator and eigenfunctions

We can now get the new generator from our thoughts in the last section. We already mentioned this generator in equation (6.57). Now we want to find eigenfunctions for those processes, i.e. a function f and a positive real number λ such that

$$Af(x) = -\lambda f(x)$$

with $Af(x)$ given by

$$Af(x) = b(x)f'(x) + \frac{1}{2}\sigma(x)^2 f''(x) + \int_{\mathbb{R}} \lambda(f(x + g(y,x)) - f(x))F_x(dy).$$

If we consider f to be polynomial then we can see that this generator also maps polynomials into polynomials. This is ensured by the assumption (6.3), because if f is polynomial, so is $f(x + g(y,x)) - f(x)$ and (6.3) ensures that the intergral in (6.57) is again a polynomial. Since the first two terms of (6.57) are also polynomials, we get a polynomial after all.

If we consider $f_n(x) = \sum\limits_{i=0}^{n} p_{n,i}x^i$ with $p_{n,n} = 1$, then we get

$$\begin{aligned} f_n(x + g(y,x)) - f_n(x) =& \sum_{i=0}^{n} p_{n,i}(x + g(y,x))^i - \sum_{i=0}^{n} p_{n,i}x^i \\ =& \sum_{i=0}^{n} p_{n,i} \sum_{k=0}^{i} \begin{pmatrix} i \\ k \end{pmatrix} x^{i-k} g(y,x)^k - \sum_{i=0}^{n} p_{n,i}x^i \\ =& \sum_{i=0}^{n} p_{n,i} \sum_{k=1}^{i} \begin{pmatrix} i \\ k \end{pmatrix} x^{i-k} g(y,x)^k \end{aligned}$$

and therefore we obtain with (6.3)

$$\begin{aligned}
\int_{\mathbb{R}} \lambda(f_n(x+g(y,x)) - f_n(x))F_x(dy) &= \lambda \sum_{i=0}^{n} p_{n,i} \sum_{k=1}^{i} \begin{pmatrix} i \\ k \end{pmatrix} x^{i-k} E_{F_x}(g(y,x)^k) \\
&= \sum_{i=0}^{n} p_{n,i} \sum_{k=1}^{i} \begin{pmatrix} i \\ k \end{pmatrix} x^{i-k} \sum_{j=0}^{k} \alpha_{j,k} x^j \\
&= \sum_{i=0}^{n} \sum_{k=1}^{i} \sum_{j=0}^{k} p_{n,i} \begin{pmatrix} i \\ k \end{pmatrix} \alpha_{j,k} x^{i-k+j}.
\end{aligned}$$

This last term is a polynomial of the maximum order of n. With this information we gather the equation

$$\begin{aligned}
-\lambda_n \sum_{i=0}^{n} p_{n,i} x^i =& -\theta(x-\mu) \sum_{i=0}^{n} i p_{n,i} x^{i-1} + \theta(ax^2+bx+c) \sum_{i=0}^{n} i(i-1) p_{n,i} x^{i-2} \\
&+ \sum_{i=0}^{n} \sum_{k=1}^{i} \sum_{j=0}^{k} p_{n,i} \begin{pmatrix} i \\ k \end{pmatrix} \alpha_{j,k} x^{i-k+j} \\
=& \sum_{i=0}^{n} \theta i(a(i-1)-1) p_{n,i} x^i + \sum_{i=0}^{n} \theta i(b(i-1)+\mu) p_{n,i} x^{i-1} \\
&+ \sum_{i=0}^{n} \theta c i(i-1) p_{n,i} x^{i-2} + \sum_{i=0}^{n} \sum_{k=1}^{i} \sum_{j=0}^{k} p_{n,i} \begin{pmatrix} i \\ k \end{pmatrix} \alpha_{j,k} x^{i-k+j} \\
=& -\sum_{i=0}^{n} a_i p_{n,i} x^i + \sum_{i=0}^{n-1} b_{i+1} p_{n,i+1} x^i + \sum_{i=0}^{n-2} c_{i+2} p_{n,i+2} x^i \\
&+ \sum_{i=0}^{n} \sum_{k=1}^{i} \sum_{j=0}^{k} p_{n,i} \begin{pmatrix} i \\ k \end{pmatrix} \alpha_{j,k} x^{i-k+j}
\end{aligned} \tag{6.65}$$

with $a_i := \theta i(1 - a(i-1)), b_i := \theta i(b(i-1)+\mu)$ and $c_i := \theta c i(i-1)$. It is not surprising, that a_i, b_i and c_i are defined in the same way as in Section 4.3. Then quation (6.65) is equivalent to

$$\sum_{i=0}^{n} (a_i - \lambda_n) p_{n,i} x^i - \sum_{i=0}^{n-1} b_{i+1} p_{n,i+1} x^i - \sum_{i=0}^{n-2} c_{i+2} p_{n,i+2} x^i = \sum_{i=0}^{n} \sum_{k=1}^{i} \sum_{j=0}^{k} p_{n,i} \begin{pmatrix} i \\ k \end{pmatrix} \alpha_{j,k} x^{i-k+j}. \tag{6.66}$$

Since it is easy to gather information about the coeffcients of the polynomial from the left side of the equation (6.66), the right side causes some inconvenience. Therefore we shall keep on simplifing the left side and we gain

$$\begin{aligned}\sum_{i=0}^{n}\sum_{k=1}^{i}\sum_{j=0}^{k}p_{n,i}\begin{pmatrix} i \\ k \end{pmatrix}\alpha_{j,k}x^{i-k+j} &= \sum_{i=0}^{n}\sum_{k=1}^{i}\sum_{l=0}^{k}p_{n,i}\begin{pmatrix} i \\ k \end{pmatrix}\alpha_{k-l,k}x^{i-l} \\ &= \sum_{i=0}^{n}\sum_{k=1}^{i}\sum_{r=i-k}^{i}p_{n,i}\begin{pmatrix} i \\ k \end{pmatrix}\alpha_{k-i+r,k}x^{r} \\ &= \sum_{i=0}^{n}p_{n,i}\sum_{r=0}^{i}x^{r}\sum_{k=(i-r)\vee 1}^{i}\begin{pmatrix} i \\ k \end{pmatrix}\alpha_{k-i+r,k} \\ &= \sum_{i=0}^{n}p_{n,i}\sum_{r=0}^{i}x^{r}\beta_{r,i} \\ &= \sum_{i=0}^{n}x^{i}\sum_{r=i}^{n}p_{n,r}\beta_{i,r}\end{aligned}$$

where

$$\beta_{r,i} = \sum_{k=(i-r)\vee 1}^{i}\begin{pmatrix} i \\ k \end{pmatrix}\alpha_{k-i+r,k} \tag{6.67}$$

and $(i-r)\vee 1 = \max\{(i-r),1\}$. If $i=r$, we denote $\beta_r := \beta_{r,r} = \sum_{k=1}^{r}\begin{pmatrix} r \\ k \end{pmatrix}\alpha_{k,k}$. Then we finally get that (6.66) is equivalent to

$$\sum_{i=0}^{n}(a_i-\lambda_n)p_{n,i}x^i - \sum_{i=0}^{n-1}b_{i+1}p_{n,i+1}x^i - \sum_{i=0}^{n-2}c_{i+2}p_{n,i+2}x^i = \sum_{i=0}^{n}x^i\sum_{r=i}^{n}p_{n,r}\beta_{r,i}. \tag{6.68}$$

Then we get the eigenvalue λ_n by taking a look at the highest order n (remember that we assumed $p_{n,n}=1$)

$$\lambda_n = a_n - \beta_n = a_n - \sum_{k=1}^{n}\begin{pmatrix} n \\ k \end{pmatrix}\alpha_{k,k}. \tag{6.69}$$

We get for $0 \le j < n$

$$(a_j-\lambda_n)p_{n,j} - b_{j+1}p_{n,j+1} - c_{j+2}p_{n,j+2} = \sum_{i=j}^{n}p_{n,i}\beta_{j,i}$$

which is equivalent to

$$p_{n,j} = \frac{b_{j+1}p_{n,j+1} + c_{j+2}p_{n,j+2} + \sum_{i=j+1}^{n}p_{n,i}\beta_{j,i}}{a_j - \beta_j - \lambda_n}. \tag{6.70}$$

In the case of Pearson diffusion with jumps, can we derive a similar property as for those without jumps as mentioned in Proposition 4.5.

Proposition 6.6. *Let h be an eigenfunction to the eigenvalue λ, then under the assumption that*

$$M_t := \int_0^t \sigma(X_s) f'(X_s) dW_s + \int_0^t \int_{\mathbb{R}} S_s^y \tilde{\mathcal{N}}(ds, dy) < \infty,$$

is a true martingale, it holds that

$$E(h(X_t)|X_0 = x) = e^{\lambda t}.$$

Proof. The proof is analogue to the proof of Proposition 4.5. We take a look at the process $Y_t = e^{\lambda t} h(X_t)$ and apply the Itô formula like we did in the Section 6.3. Then we get the equation $e^{\lambda t} h(X_t) = h(X_0) + \int_0^t M_s dW_s$ where M_t is martingale with $M_0 = 0$ by using the property of the eigenfunction. Then we gain $E(e^{\lambda t} h(X_t)|X_0 = x) = h(X_0)$ which is equivalent to the assumption in the proposition. □

With this last proposition we get like in the jumpless case, since the eigenfunction is again a polynomial with $p_{n,n} = 1$,

$$E(X_t^n|X_0 = x) = E_x(X_t^n) = e^{-\lambda_n t} \sum_{i=0}^{n} p_{n,i} x^i - \sum_{i=0}^{n-1} p_{n,i} E_x(X_t^i). \tag{6.71}$$

The conditional expectation is for any fixed t a polynomial of the order n in x with coefficients which are linear combinations of $1, e^{-\lambda_1 t}, \dots, e^{-\lambda_n t}$. Then we can derive that

$$E_x(X_t^n) = \sum_{i=0}^{n} q_{n,j} x^j = \sum_{j=0}^{n} \sum_{l=0}^{n} q_{n,j,l} e^{-\lambda_l t} x^j \tag{6.72}$$

with $q_{n,j,n} = p_{n,j}$ and $q_{n,j,l} = -\sum_{k=l}^{n-1} p_{n,k} q_{k,j,l}$ for $l = 0, \dots, n-1$. Then we derive for the first two moments

$$E_x(X_t) = e^{-\lambda_1 t} x + p_{1,0}(e^{-\lambda_1 t} - 1) \tag{6.73}$$

and

$$\begin{aligned} E_x(X_t^2) =& e^{-\lambda_2 t} x^2 + (p_{2,1} e^{-\lambda_2 t} - p_{2,1} e^{-\lambda_1 t}) x \\ &+ p_{2,0}(e^{-\lambda_2 t} - 1) - p_{2,1} p_{1,0}(e^{-\lambda_1 t} - 1). \end{aligned} \tag{6.74}$$

Example 6.7. Let us assume a process of the following form

$$X_t = X_0 - \int_0^t \theta(X_s - \mu) ds + \int_0^t \sqrt{2\theta} dW_s + \sum_{n=1}^{\mathcal{N}_t} J_n X_{\tau_n -} \tag{6.75}$$

where J_n is i.i.d. with an exponential distribution with the parameter $\psi \geq 1$ and $X_0 = x > 0$. Further we consider that $P(X_t > 0) = 1$ fo all $t \in \mathbb{R}$.We get with $g(y, x) = yx$, that

$$|g(y, x_1) - g(y, x_2)| = |y(x_1 - x_2)| = |y||x_1 - x_2|$$

and since

$$\int_{\mathbb{R}} |y|^2 f(y)dy < \infty,$$

where $f \sim \exp(\psi)$ and thus the process satisfies the global Lipschitz condition (6.28). With the Example 6.1 and $w(x) = x$ we gain that there exists a unique solution to this particular process. Furthermore it is well known, that the kth moment of an exponential distributed r.v. is $\frac{k!}{\psi^k}$ and therefore we get

$$\lambda E_{F_{X_{t-}}}((J_{\mathcal{N}_t} X_{t-})^k) = \lambda \frac{k!}{\psi^k} X_{t-}^k,$$

i.e. $\alpha_{k,k} = \lambda \frac{k!}{\psi^k}$ and $\alpha_{j,k} = 0$ for $j \neq k$. Then we get the eigenvalue with equation (6.69), i.e.

$$\lambda_n = a_n - \beta_n = \theta n(2-n) - \lambda \sum_{k=1}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{k!}{\psi^k}$$

and further it is worth to mention, that $\beta_{r,i} = 0$ if $r \neq i$. Now it is possible to write a recursive formula for the parameters of the eigenfunction. Thus we gather

$$\begin{aligned} p_{n,j} =& \frac{b_{j+1}p_{n,j+1} + c_{j+2}p_{n,j+2}}{a_j - \beta_j - \lambda_n} \\ =& \frac{\theta(j+1)\mu}{\theta j(2-j) - \beta_j - (a_n - \beta_n)} p_{n,j+1} \\ =& \frac{\theta j \mu}{\theta(j(2-j) - n(2-n)) - (\beta_j - \beta_n)} p_{n,j+1} \end{aligned}$$

for $0 \leq j < n$. Then we can calculate the first twoconditional moments with (6.73) and (6.74). We get

$$\begin{aligned} p_{1,0} &= \frac{\theta\mu\psi}{\lambda - \theta\psi}, \\ p_{2,0} &= \frac{\theta\mu\psi^2}{\lambda\psi + 2\lambda} \end{aligned}$$

and

$$p_{2,1} = \frac{\theta\mu\psi^2}{\theta\psi^2 + 2\lambda}.$$

Then the first two moments are

$$E_x(X_t) = e^{-\lambda_1 t} x + \frac{\theta\mu\psi}{\lambda - \theta\psi}(e^{-\lambda_1 t} - 1)$$

and

$$\begin{aligned} E_x(X_t^2) = e^{-\lambda_2 t} x^2 &+ (\frac{\theta\mu\psi^2}{\theta\psi^2 + 2\lambda} e^{-\lambda_2 t} - \frac{\theta\mu\psi^2}{\theta\psi^2 + 2\lambda} e^{-\lambda_1 t}) x \\ &+ \frac{\theta\mu\psi^2}{\lambda\psi + 2\lambda}(e^{-\lambda_2 t} - 1) - \frac{\theta\mu\psi^2}{\theta\psi^2 + 2\lambda} \frac{\theta\mu\psi}{\lambda - \theta\psi}(e^{-\lambda_1 t} - 1) \end{aligned}$$

with $\lambda_1 = \theta - \frac{\lambda}{\psi}$ and $\lambda_2 = 2\theta - (\frac{\lambda}{\psi} - \frac{2\lambda}{\psi^2})$.

Δ

Example 6.8. Let us extend Example 4.6 in Chapter 4. Let us assume again that $g(J_n, X_{\tau_{n-}}) = J_n X_{\tau_{n-}}$ and that J_n is i.i.d. Beta distributed with the parameters $p, q > 0$. Then we gain a process of the form

$$dZ_t = -\theta Z_t dt + \sqrt{2\theta(\nu-1)^{-1}(Z_t^2 + 2\rho\nu^{\frac{1}{2}} Z_t + (1+\rho^2)\nu)} dW_t + \sum_{n=1}^{\mathcal{N}_t} J_n X_{\tau_{n-}}$$

and $X_0 = x > 0$. With the assumption $P(X_t > 0) = 1$ for all $t \in \mathbb{R}$ we gain like in the previous example, that this process satisfies the global Lipschtiz condition (6.28) and that there exists a unique solution to this process. The characteristic function of the beta distribution is given by

$$\phi(t) = \sum_{k=0}^{\infty} \frac{\Gamma(p+k)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+k)} \frac{i^k t^k}{k!}$$

and due the fact that $E(x^l) = \frac{\phi^l(0)}{i^l}$ and $\Gamma(a+k) = (k-1)!\Gamma(a)$ for $k \in \mathbb{N}$, we gather the kth moment of the beta distribution to be

$$E(X^k) = \prod_{i=0}^{k-1} \frac{p+i}{p+q+i}.$$

Then (6.3) is equivalent to

$$\lambda E_{F_{X_{t-}}}(J_{\mathcal{N}_t} X_{t-}^k) = \lambda \prod_{i=0}^{k-1} \frac{p+i}{p+q+i} X_{t-}^j = \alpha_{k,k} X_{t-}^j$$

where $\alpha_{k,k} = \lambda \prod_{i=0}^{k-1} \frac{p+i}{p+q+i}$ and $\alpha_{l,k} = 0$ if $l \neq k$. Thus we get the nth eigenvalue by using (6.69)

$$\lambda_n = \theta n(1 - \frac{n-1}{\nu-1}) - \sum_{k=1}^{n} \begin{pmatrix} n \\ k \end{pmatrix} \alpha_{k,k}$$

and the coefficient with (6.70)

$$\begin{aligned} p_{n,j} =& \frac{b_{j+1}p_{n,j+1} + c_{j+2}p_{n,j+2}}{a_j - \beta_j - \lambda_n} \\ =& \frac{2\theta\rho\sqrt{\nu}\frac{j(j+1)}{(\nu-1)}p_{n,j+1} + \theta(1+\rho^2)\nu\frac{(j+1)(j+2)}{\nu-1}p_{n,j+2}}{\theta j(1-\frac{j-1}{\nu-1}) - \theta n(1-\frac{n-1}{\nu-1}) - \sum_{k=1}^{j}\begin{pmatrix} j \\ k \end{pmatrix}\alpha_{k,k} + \sum_{k=1}^{n}\begin{pmatrix} n \\ k \end{pmatrix}\alpha_{k,k}} \end{aligned}$$

for $0 \leq j < n$. Then we again can calculate the first twoconditional moments with (6.73) and (6.74). We get

$$p_{1,0} = p_{2,0} = 0$$

and

$$p_{2,1} = 4\frac{\theta\rho\frac{\sqrt{\nu}}{\nu-1}}{\theta(1-2\frac{\nu-2}{\nu-1}) - \lambda\frac{p(p+1)}{(p+q)(p+q+1)}}.$$

Then the first two moments are

$$E_x(X_t) = e^{-\lambda_1 t}x$$

and

$$\begin{aligned} E_x(X_t^2) = e^{-\lambda_2 t}x^2 + \Bigg(& 4\frac{\theta\rho\frac{\sqrt{\nu}}{\nu-1}}{\theta(1-2\frac{\nu-2}{\nu-1}) - \lambda\frac{p(p+1)}{(p+q)(p+q+1)}}e^{-\lambda_2 t} \\ & - 4\frac{\theta\rho\frac{\sqrt{\nu}}{\nu-1}}{\theta(1-2\frac{\nu-2}{\nu-1}) - \lambda\frac{p(p+1)}{(p+q)(p+q+1)}}e^{-\lambda_1 t}\Bigg)x \end{aligned}$$

with $\lambda_1 = \theta - \lambda\frac{p}{p+q}$ and $\lambda_2 = 2\theta(\frac{\nu-2}{\nu-1}) - \lambda\frac{p}{p+q}(1+\frac{p+1}{p+q+1})$.

Δ

Appendix

Martingale theory

In this section we want to state briefly some important conclusions from the martingale theory. We assume a probability space $(\Omega, \mathcal{F}, P)$ and an increasing sequence of σ-algebras $(\mathcal{F}_n)_{n\geq 0}$, having the property $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all $n \geq 0$.

Definition 1. Let $(X_n)_{n\geq 0} \subset \Omega$ be a sequence of random variables. We call $(X_n)_{n\geq 0}$ a *martingale* if

(i) $E(|X_n|) < \infty$ for all n, i.e. if X_n is integrable;

(ii) X_n is $\mathcal{F}_n$ measurable for all n;

(iii) $E(X_n|\mathcal{F}_m) = X_m$ for all $m \leq n$.

Example 2. Let $(X_n)_{n\geq 0}$ be i.i.d. ± 1 random viarables with $p = P(X_1 = 1)$ and $S_0 = 0, S_n = \sum_{i=1}^{n} X_i$ for $n > 0$ is the corresponding *random walk*. Then $Y_n = S_n - n(2p-1) = S_n - E(S_n)$ is a martingale, because

$$\begin{aligned} E(Y_n - Y_{n-1}|\mathcal{F}_{n-1}) &= E(X_n - (2p-1)|\mathcal{F}_{n-1}) \\ &= E(X_n|\mathcal{F}_{n-1}) - (2p-1) \\ &= E(X_n) - (2p-1) \\ &= 0 \end{aligned}$$

which is equivalent to

$$E(Y_n|\mathcal{F}_{n-1}) = E(Y_{n-1}|\mathcal{F}_{n-1}) = Y_{n-1}$$

and the other two properties are obvious.

Δ

Now we want to define a time continuous martingale, which we need later on.

Definition 3. Let X_t be a process and $\mathcal{F}_t = \sigma\{X_s|s \leq t\}$ for $t \geq 0$. We call X_t a *time continuous martingale* if

(i) $E(|X_t|) < \infty$ for all $t \geq 0$, i.e. if X_t is integrable;

(ii) X_t is $\mathcal{F}_t$ measurable;

(iii) $E(X_t|\mathcal{F}_s) = X_s$ for all $s \leq t$.

Example 4. Let $\mathcal{N}_t$ be a Poisson process with a parameter $\lambda > 0$. Further we define the *compensated Poisson process* $\tilde{\mathcal{N}_t} = \mathcal{N}_t - \lambda t$. Then $\tilde{\mathcal{N}_t}$ is a martingale. It is obvious that $\tilde{\mathcal{N}_t}$ is $\mathcal{F}_t$ measurable and further we know that

$$E(\tilde{\mathcal{N}_t}) = E(\mathcal{N}_t - \lambda t) = E(\mathcal{N}_t) - \lambda t = 0.$$

Then we get that for all $s \leq t$

$$E(\tilde{\mathcal{N}_t} - \tilde{\mathcal{N}_s}|\mathcal{F}_s) = E(\mathcal{N}_t - \mathcal{N}_s|\mathcal{F}_s) - \lambda(t-s) \overset{\text{ind.}}{=} E(\underbrace{\mathcal{N}_t - \mathcal{N}_s}_{\sim Poi(\lambda(t-s))}) - \lambda(t-s) = 0$$

and therefore is $\tilde{\mathcal{N}_t}$ a martingale. It can also be proved that $\tilde{\mathcal{N}}_t^2 - \lambda t$ is a martingale.

Δ

Definition 5. A random variable $T : \Omega \to \mathbb{N}$ is called *stopping time* if $\{T \leq n\} \in \mathcal{F}_n$ for all n. Futher we call a stopping time *bounded* by a constant c if $P(T \leq c) = 1$.

It is equivalent to show $\{T = n\} \in \mathcal{F}_n$ instead of $\{T \leq n\} \in \mathcal{F}_n$, because

$$\{T = n\} = \underbrace{\{T \leq n\}}_{\mathcal{F}_n} \cup \underbrace{\{T \leq n-1\}^c}_{\mathcal{F}_{n-1} \subset \mathcal{F}_n} \in \mathcal{F}_n.$$

If T is a finite stopping time, we denote by $X_T(\omega) = X_{T(\omega)}(\omega)$ the random variable with the value X_n whenever $T = n$.

Theorem 6. *(Stopping theorem)*
Let T be a c bounded stopping time and $(X_n)_{n\geq 0}$ a martingale. Then $E(X_T) = E(X_0)$.

Proof. Without loss of generality we assume that c is an integer. Therefore we can write

$$X_T = \sum_{n=0}^{\infty} X_n(\omega) 1_{\{T=n\}} = \sum_{n=0}^{c} X_n(\omega) 1_{\{T=n\}}.$$

With the fact that $\{T = n\} \in \mathcal{F}_n$, we gain

$$\begin{aligned}
E(X_T) &= E(\sum_{n=0}^{c} X_n 1_{\{T=n\}}) = \sum_{n=0}^{c} E(X_n 1_{\{T=n\}}) \\
&= \sum_{n=0}^{c} E(E(X_c|\mathcal{F}_n) 1_{\{T=n\}}) \\
&= \sum_{n=0}^{c} E(E(X_c 1_{\{T=n\}}|\mathcal{F}_n)) \\
&= \sum_{n=0}^{c} E(X_c 1_{\{T=n\}}) = E(X_c \sum_{n=0}^{c} 1_{\{T=n\}}) \\
&= E(X_c) = E(E(X_c|\mathcal{F}_0)) \\
&= E(X_0).
\end{aligned}$$

$\square$

Now let us assume that $(Y_n)_{n\geq 0}$ is an adaptive sequence of i.i.d. random viarables with

$$m = E(Y_1), \quad \sigma^2 = \mathrm{var}(Y_1) < \infty$$

and Y_{n+1} is independent to $\mathcal{F}_n$ for all $n \geq 0$. We define the partial sum

$$S_0 = 0, \quad S_n = Y_1 + \ldots Y_n, \quad n \geq 1.$$

Theorem 7. *(Wald's identity)*
Let T be a stopping time with $E(T) < \infty$. Then $S_T \in \mathcal{L}^1$ and $E(S_T) = mE(T)$.
Further, if $m = 0$ then $E(S_T^2) = \sigma^2 E(T)$ or if T independent from $(Y_n)_{n\geq 0}$ we get $Var(S_T) = \sigma^2 E(T) + m^2\, Var(T)$.

Proof. Let us take a look at the following process

$$X_n = S_n - nm, \quad n \geq 0, \tag{6.76}$$

then X_n is a martingale (see Example 2). Since T does not have to be bounded, we can achieve it by taking an integer k and by examing $(T \wedge k) = \min(T, k)$. If we now apply the stopping theorem (Theorem 6), we get

$$E(S_{T\wedge k}) = mE(T \wedge k) + \underbrace{E(X_{T\wedge k})}_{E(X_0)=0} = mE(T \wedge k) \tag{6.77}$$

for every stopping time T. The right side of (6.77) converges with the theorem about the monoton convergence to $mE(T)$. In the case where $T < \infty$ P-a.s., $S_{T\wedge k}$ converges a.s. to S_T. If we can show that $(S_{T\wedge k})_{k\geq 0}$ has a integrable major then we get $E(S_T) = mE(T)$ with theorem of Lebesque.
If we define $\bar{Y}_i = |Y_i|, i \in \mathbb{N}, \bar{S}_0 = 0, \bar{S}_n = \sum_{i=1}^{n} \bar{Y}_i, n \geq 0$, the $\bar{Y}_i, i \in \mathbb{N}$ are i.i.d. with $\bar{m} = E(|Y_i|)$ and analogue to above we can derive

$$E(\bar{S}_{T\wedge k}) = \bar{m}E(T \wedge k), \quad k \geq 0.$$

Now we can apply the theorem of monoton convergence onto both sides. Especially we derive from $E(T) < \infty$ that $E(\bar{S}_T) < \infty$, i.e. $\bar{S}_t \geq |S_{T\wedge k}|$ is the integrable major we were looking for. So far we have shown that $S_T \in \mathcal{L}^1$ and $E(S_T) = mE(T)$.
Let us examine now the martingale

$$M_n = X_n^2 - n\sigma^2.$$

It is a martingale since

$$E(M_{n+1} - M_n|\mathcal{F}_n) = E((Y_{n+1} - m)^2) - \sigma^2 + 2X_n(Y_{n+1} - m)|\mathcal{F}_n) = 0$$

and the other two properties are obvious. From $E(T) < \infty$ we can consider analogue to above that

$$\sigma^2 E(T) = E(X_T^2) - \underbrace{E(M_T)}_{=0} = E((S_T - Tm)^2) = E(S_T^2) - 2mE(TS_T) + m^2 E(T^2)$$

For $m = 0$ we get $E(S_T^2) = \sigma^2 E(T)$.
In the case where T is independent from $(Y_i)_{i\geq 1}$, we get

$$E(TS_T) = \sum_{i=1}^{\infty} \underbrace{iE(S_i 1_{\{T=i\}})}_{imP(T=i)} = mE(T^2)$$

and therefore

$$\begin{aligned}\mathrm{Var}(S_T) &= E(S_T^2) - (E(S_T))^2 \\ &= \sigma^2 E(T) + m^2 E(T^2) - (mE(T))^2 \\ &= \sigma^2 E(T) + m^2 \mathrm{Var}(T).\end{aligned}$$

□

There exist several theorems about the convergence of martingales. We just want to quote the probably most important ones.

Theorem 8. *(Martingale Convergence Theorem)*

a) Let $(X_n)_{n\geq 1}$ be a martingale and suppose $(X_n)_{n\geq 1}$ is a uniformly integrable collection of r.v.s. Then

$$\lim_{n\to\infty} X_n = X_\infty$$

exists a.s., X_∞ is in $\mathcal{L}^1$ and X_n converges to X_∞ in $\mathcal{L}^1$.
Moreover $X_n = E(X_\infty|\mathcal{F}_n)$.

b) Conversely let $Y \in \mathcal{L}^1$ and consider the martingale $X_n = E(Y|\mathcal{F}_n)$. Then $(X_n)_{n\geq 1}$ is a uniformly integrable collection of r.v.s.

Proof. See Jacod and Protter (2004), pages 231 to 232. □

There also exists a central limit theorem for martingales.

Theorem 9. *(Martingale Central Limit Theorem)*
Let $(X_n)_{n\geq 1}$ be a sequence of random variables satisfying

(i) $E(X_n|\mathcal{F}_{n-1}) = 0$

(ii) $E(X_n^2|\mathcal{F}_{n-1}) = 1$

(iii) $E(|X_n|^3|\mathcal{F}_{n-1}) \leq K < \infty$.

Let $S_n = \sum_{i=1}^{n} X_i$ and $S_0 = 0$. Then $\lim_{n\to\infty} \frac{1}{\sqrt{n}} S_n = Z$, where $Z \sim \mathcal{N}(0,1)$ and where the convergence is in distribution.

Proof. See Jacod and Protter (2004), pages 235 to 237. □

We shall also briefly discuss a very important property for submartingales.

Definition 10. Let $(X_n)_{n\geq 0} \subset \Omega$ be a sequence of random variables. We call $(X_n)_{n\geq 0}$ a *submartingale* (*supermartingale*), if

(i) $E(|X_n|) < \infty$ for all n, i.e. if X_n is integrable;

(ii) X_n is $\mathcal{F}_n$ measurable for all n;

(iii) $E(X_n|\mathcal{F}_m) \geq X_m$ ($E(X_n|\mathcal{F}_m) \leq X_m$) for all $m \leq n$.

For the proof of Theorem 6.2 we need the following theorem.

Theorem 11. *(Doob's martingale inequality)*
If X_t with $t \geq 0$ is a positive submartingale then, for any $p > 1$

$$E(\sup_{0\leq s\leq t} X_s^p) \leq q^p E(X_t^p)$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. See Dellacherie and Meyer (1980), page 18. □

Futher we briefly define a local martingale and mention one important proporsition.

Definition 12. Let τ_n be a stopping time for all $n \in \mathbb{N}$ with $\tau_n \to \infty$. We call a process M_t a *local martingale* if

$$M_t^{\tau_n} = M_{t\wedge\tau_n}$$

is a matrtingale.

Proposition 13. Let M_t be a bounded local martingale, then M_t is a true martingale.

Proof. Let (τ_n) be a sequence of stopping times with $\tau_n \to \infty$ for $n \to \infty$. For $s \leq t$ we know that

$$\int_F M_{\tau_n\wedge s} dP = \int_F M_{\tau_n\wedge t} dP,$$

for all $F \in \mathcal{F}_s$. Since $M_{\tau_n\wedge s} \to M_s$, $M_{\tau_n\wedge t} \to M_t$ a.s. for $n \to \infty$ and $\sup_s |M_{\tau_n\wedge s}| \leq C$, we gain with the dominated convergence theorem

$$\begin{aligned}
\int_F M_s dP &= \int_F \lim_{n\to\infty} M_{\tau_n\wedge s} dP \\
&= \lim_{n\to\infty} \int_F M_{\tau_n\wedge s} dP \\
&= \lim_{n\to\infty} \int_F M_{\tau_n\wedge t} dP \\
&= \int_F \lim_{n\to\infty} M_{\tau_n\wedge t} dP \\
&= \int_F M_t dP
\end{aligned}$$

for all $F \in \mathcal{F}_s$, i.e. $E(M_t|\mathcal{F}_s) = M_s$. The fact that M_t is $\mathcal{F}_t$-measurable follows from the definition of $\mathcal{F}_t = \sigma\{X_s | s \leq t\}$. You can show $E(|M_t|) < \infty$ with the dominated convergence theorem as above. □

Stochastic integration

In this very short section we want to announce a few very important statements from the theory of stochastic integration. It is our aim to define a stochatical integral $\int_0^t H_s dX_s$ for some random variable X_t.

Example 14. A typical example are trading earnings. Assume that $(X_t)_{t\in\mathbb{N}}, (H_t)_{t\in\mathbb{N}}$ timediscrete processes, where

$$X_t: \text{ stock prize at time point } t$$

$$H_t: \text{ the number of stocks at time point } t.$$

Then we get the trading earnings at time point t to be $H_t \underbrace{(X_t - X_{t-1})}_{=:\Delta X_t}$ and the total earnings in the time periode from 0 to t are $\sum\limits_{s=0}^{t} H_s \Delta X_s$.

The aim is to give a definition for the total earnings if we assume $(X_s)_{s\in\mathbb{R}_+}, (H_s)_{s\in\mathbb{R}_+}$.

Δ

To ease of notation we write W_t for the standard Brownian motion (or standard Wiener process) at time point t. For reasons of simplicity we consider a time periode $[0,T]$ with $T \in \mathbb{R}_+$, instead of $\mathbb{R}_+$. Further we define $0 = t_0 < t_1 < \cdots < t_p = T$.

Definition 15.

a) A process $(H_t)_{t\in[0,T]}$ is called a *simple process*, if it is of the form

$$H_t(\omega) = \sum_{i=1}^{p} \phi_{t_i}(\omega) 1_{(t_{i-1},t_i]}(t),$$

where ϕ_i are $\mathcal{F}_{t_{i-1}}$-measurable, bounded random variables.

b) For simple processes we define the *stochastic intergal* I

$$\Big(I(H)\Big)_{t\in[0,T]} = \Big(\int_0^t H_s dW_s\Big)_{t\in[0,T]}$$

by

$$\begin{aligned} I(H)_t &= \sum_{i=1}^{p} \phi_i (W_{t_i \wedge t} - W_{t_{i-1}\wedge t}) \\ &= \sum_{i=1}^{k} \phi_i (W_{t_i} - W_{t_{i-1}}) + \phi_{k+1}(W_t - W_{t_k}) \text{ for } t \in (t_k, t_{k+1}]. \end{aligned}$$

This integral has got several properties. A very important one is given by the next theorem.

Theorem 16. *Let $(H_t)_{t\in[0,T]}$ be a simple process and ϕ_n are $\mathcal{F}_{t_{n-1}}$ measurable random variables, then $\Big(I(H)\Big)_{t\in[0,T]}$ is a well defined, continuous martingale.*

Proof. The integral is well defined due to the fact that the representation does not depend on $t_0, \dots, t_p$. The process is adapted, because W_t is adaptive. The continuity and integrability are also clear. Therefore it is only left to show that $E(I(H)_t|\mathcal{F}_s) = I(H)_s$ for $s \le t$. Without loss of generality we assume that $s, t \in \{t_0, \dots, t_p\}$, otherwise add more supporting points. Since $E(E(X|\mathcal{F}_t)|\mathcal{F}_s) = E(X|\mathcal{F}_s)$ for $s \le t$, we deduce

$$E(I(H)_{t_n}|\mathcal{F}_{t_{n-2}}) = E(E(I(H)_{t_n}|\mathcal{F}_{t_{n-1}})|\mathcal{F}_{t_{n-2}}) = E(I(H)_{t_{n-1}}|\mathcal{F}_{t_{n-2}}).$$

It is only left to show that $E(I(H)_{t_n}|\mathcal{F}_{t_{n-1}}) = I(H)_{t_{n-1}}$.

$$\begin{aligned} E(I(H)_{t_n}|\mathcal{F}_{t_{n-1}}) =& E(I(H)_{t_{n-1}} + \phi_n(W_{t_n} - W_{t_{n-1}})|\mathcal{F}_{t_{n-1}}) \\ =& I(H)_{t_{n-1}} + \phi_n \underbrace{(\underbrace{E(W_{t_n}|\mathcal{F}_{t_{n-1}})}_{=W_{t_{n-1}}} - W_{t_{n-1}})}_{0} \\ =& I(H)_{t_{n-1}}. \end{aligned}$$

□

Now we want to extend the class of processes which are integrable with respect to the Wiener process. If we take a closer look at

$$\mathcal{H} := \{(H_t)_{t\in[0,T]}|(H_t) \text{ is measurable, adaptive and } E(\int_0^T H_s^2 ds) < \infty\},$$

we can prove that there exists an explicit mapping

$$J : \mathcal{H} \to \{(X_t)_{t\in[0,T]}|X \text{ is a continuous martingale}\}$$

where $J(H) = I(H)$ exept a nullset independent of t for all simple processes H. For $H \in \mathcal{H}$ we call $(\int_0^t H_s dW_s)_{t\in[0,T]} := (J(H)_t)_{t\in[0,T]}$ *stochastic integral of H with respect to W*.
It is very important to point out that if $X \in \mathcal{H}$, we know that $\int X_s dW_s$ is a martingale.
It is still possible to expand the class of processes that are integrable with respect to the Wiener process. We define this class as following

$$\tilde{\mathcal{H}} := \{(H_t)_{t\in[0,T]}|(H_t) \text{ is measurable, adaptive and } \int_0^T H_s^2 ds < \infty \quad a.s.\}.$$

In this case there exists an explicit mapping

$$\tilde{J} : \tilde{\mathcal{H}} \to \{(X_t)_{t\in[0,T]}|X \text{ is a continuous process}\},$$

where $\tilde{J}(H) = I(H)$ exept a nullset independent of t for all simple processes H. Again we call $\int_0^t H_s dW_s := \tilde{J}(H)_t$ *stochastic integral of H with respect to W* for all $H \in \tilde{\mathcal{H}}$.
It is important to point out again that $\tilde{\mathcal{H}}$ is usually not a martingale, but it is a *local martingale.*

Itô processes and the Itô formula

In the previous section we took a look at processes which are integrable with respect to the standard Brownian motion process (which we still denote with W_t). Under certain conditions we will find some process $(X_t)_{t\in[0.T]}$ with the property that they define an integral $\int L_s dX_s$ for some process L.

Definition 17. Let $(H_t)_{t\in[0.T]}, (K_t)_{t\in[0.T]}$ be adaptive, measurable processes with the properties $\int_0^T |H_s|^2 ds < \infty, \int_0^T |K_s| ds < \infty$ almost surely. Further X_0 is a $\mathcal{F}_0$-measurable random variable. We call a process *Itô process* if the process is of the following form

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s. \tag{6.78}$$

We sometimes use a different notation for (6.78)

$$dX_s = K_s ds + H_s dW_s$$

which is the same since

$$\underbrace{\int_0^t dX_s}_{=X_t - X_0} = \int_0^t K_s ds + \int_0^t H_s dW_s.$$

The processes K_s and H_s of a Itô process are explicit. Now we are able to define a stochastic integration with respect to an Itô process.

Definition 18. Let $(X_t)_{t\in[0,T]}$ be an Itô process with $(K_t)_{t\in[0,T]}$ and $(H_t)_{t\in[0,T]}$ like in Definition 17. Further we assume that $(L_t)_{t\in[0,T]}$ is an adaptive measurable process with $\int_0^T |L_s K_s| ds < \infty$ and $\int_0^T (L_s H_s)^2 ds < \infty$ almost surly. Then we can define the stochastic integral with respect to the Itô process $(X_t)_{t\in[0,T]}$

$$\int_0^t L_s dX_s := \int_0^t L_s K_s ds + \int_0^t L_s H_s dW_s.$$

If we want to go any further in the theory of Itô processes we need to define the quadratic variation process of X and the covariation process of X and Y.

Definition 19. Let $X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s$ and $Y_t = Y_0 + \int_0^t \breve{K}_s ds + \int_0^t \breve{H}_s dW_s$ be Itô processes.

1. The process $([X,X]_t)_{t\in[0,T]}$ with $[X,X]_t := \int_0^t H_s^2 ds$ is called *quadratic variation process of* X. Another notation is $\langle X, X\rangle$.

2. The process $([X,Y]_t)_{t\in[0,T]}$ with $[X,Y]_t := \int_0^t H_s \breve{H}_s ds$ is called *covariation process of* X *and* Y. Another notation is $\langle X, Y\rangle$.

The following theorem is called *the Itô formula* and it plays an important role if you deal with stochastic integration.

Theorem 20. *(Itô formula)*
Let $(X_t)_{t\in[0,T]}$ be an Itô process and $f:\mathbb{R}^2\to\mathbb{R}$ a twice continuously differentiable function. Then $(f(t,X_t))_{t\in[0,T]}$ is a Itô process of the form

$$f(t,X_t)=f(0,X_0)+\int_0^t \partial_1 f(s,X_s)ds+\int_0^t \partial_2 f(s,X_s)dX_s+\frac{1}{2}\int_0^t \partial_{22} f(s,X_s)H_s^2 ds.$$

Proof. You can proof it by using the Taylor expansion and the theorem about the quadratic variation of the Brownian motion. □

Example 21. Assuming that $X_t=W_t$ and $f(x)=x^2$, we can derive with Theorem 20

$$W_t^2=\underbrace{W_0^2}_{=0}+\int_0^t 2W_s dW_s+\frac{1}{2}\int_0^t 2ds$$

which leads to

$$\int_0^t W_s dW_s=\frac{1}{2}(W_t^2-t).$$

Δ

As we used it in Chapter 6 there also exists an Itô formula for diffusions with jumps. We call X_t a *semimartingale* if it has an composition of the form

$$X_t=X_0+M_t+A_t$$

where $A_t\in\Theta$ and M_t is a local martingale. The set Θ is the set of processes Y_t which are $\mathcal{F}_t$-measurable, right continuous with limits from the left, $Y_0=0$ and have finite variation in finite intervals e.g. a compounded Poisson process.

Theorem 22. *(Itô formula with jumps)*
Let $\underline{X_t}=(X_{1,t},\dots,X_{k,t})$ be a k-dimensional semimartingale, and let $f:\mathbb{R}^k\to\mathbb{R}$ be a twice continuously differentiable function. Then $f(\underline{X_t})$ is a semimartingale, and

$$\begin{aligned}f(\underline{X_t})=&f(\underline{X_0})+\sum_{i=1}^k\int_0^t\frac{\partial f}{\partial x_i}(\underline{X}_{s-})dX_{i,s}+\frac{1}{2}\sum_{i,j=1}^k\int_0^t\frac{\partial^2 f}{\partial x_i\partial x_j}(\underline{X}_{s-})d\left\langle X_i^c,X_j^c\right\rangle_s\\&+\sum_{s\le t}\Big(f(\underline{X}_s)-f(\underline{X}_{s-})-\sum_{i=1}^k\frac{\partial f}{\partial x_i}(\underline{X}_{s-})\Delta X_{i,s}\Big),\end{aligned}$$

where X^c is the continuous martingale part of X.

Example 23. Let us consider the following SDE

$$dX_t=X_{t-}dt+X_{t-}dW_t+X_{t-}dN_t,\quad X_0=x_0,$$

where N_t is a Poisson process that is independent of the standard Wiener process W_t. Let us apply the Itô formula with jumps on $Y_t=\ln(X_t)=f(X_t)$. Then we obtain

$$\begin{aligned}Y_t=&\ln(X_t)\\=&\ln(X_0)+\int_0^t\frac{1}{X_{s-}}dX_s-\frac{1}{2}\int_0^t\frac{1}{X_{s-}^2}X_{s-}^2ds+\sum_{s\le t}(\ln(X_s)-\ln(X_{s-})-\frac{1}{X_{s-}}\Delta X_s).\end{aligned}$$

Since $\Delta X_s = X_s - X_{s-}$, $X_s = X_{s-} + \Delta X_s$ and N_t the number of jumps in the interval $[0,t]$, we gain

$$\begin{aligned}\ln(X_t) =& \ln(x_0) + \int_0^t ds + \int_0^t dWs + \int_0^t dN_t - \frac{1}{2}\int_0^t ds + \sum_{j=1}^{N_t}(\ln(2X_{\tau_j-}) - \ln(X_{\tau_j-}) - 1)\\ =& \ln(x_0) + t + W_t - \frac{1}{2}t + N_t + \sum_{j=1}^{N_t}(\ln(2) - 1)\\ =& \ln(x_0) + \frac{1}{2}t + W_t + N_t + N_t(\ln(2) - 1)\\ =& \ln(x_0) + \frac{1}{2}t + W_t + N_t \ln(2)\end{aligned}$$

and the solution for the SDE is

$$X_t = x_0 \exp(W_t + \frac{1}{2}t)2^{N_t}.$$

Poisson integration

Now we want to define the integral with respect to a Poisson process and the integral with respect to compensated Poisson random measure. For simplicity reason we consider that the distribution of J_t is independent from the state of the process, i.e. $J \sim F$.
Let us consider that $\mathcal{N}_t$ is a Poisson process with a parameter $\lambda > 0$ for $t \geq 0$. We have seen that the compensated Poisson process $\tilde{\mathcal{N}_t}$ is a martingale (see example 4 in the Appendix).

Definition 24. Let $(\mathcal{S}, \mathcal{A})$ be a measurable space $(\Omega, \mathcal{F}, P)$ be probability space. A *random measure* M on $(\mathcal{S}, \mathcal{A})$ is a collection of random variables $M(B)$ with $B \in \mathcal{A}$ such that

(i) $M(\emptyset) = 0$,

(ii) given any sequence $(A_n, n \in \mathbb{N})$ of mutually disjoint sets in $\mathcal{A}$,

$$M\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} M(A_n) \quad \text{a.s.}$$

(iii) for each disjoint family $(B_1, \ldots, B_n)$ in $\mathcal{A}$, the random variables $M(B_1), \ldots, M(B_n)$ are independent.

Further we say that we have a *Poisson random measure* if each $M(B)$ has a Poisson distribution whenever $M(B) < \infty$.

As we defined in section 6 (equation (6.4)) we have that

$$\Psi((0,t] \times A) = \sum_{n=1}^{\mathcal{N}_t} 1_{(J_n \in A)} = \#\{0 \leq s \leq t | \Delta J_s \in A\}$$

is a Poisson random measure (especially $A \to \Psi((0,t]\times A)$ is a counting measre on $\mathcal{B}\backslash\{0\}$). We call

$$\Lambda((0,t]\times A) = \lambda \int_0^t \int_A f(y)dyds = \lambda t \int_A f(y)dy,$$

where f is a density function, as we defined in (6.8) a *intensity measure.*
Now we can define the *Poisson integral.* Let g be a Borel measureable function from $\mathbb{R}$ to $\mathbb{R}$, then we define the integral for any $t > 0, \omega \in \Omega$ by

$$\int_A g(x)\Psi(t,dx)(\omega) = \sum_{n=1}^{\mathcal{N}_t} g(J_n)1_{\{J_n\in A\}}(\omega).$$

Note that each $\int_A g(x)\Psi(t,dx)$ is a $\mathbb{R}$-valued random variable and gives rise to a càdlàg stochastic process as we vary t. Then we define for each $t \geq 0$ and A the *compensated Poisson random measure* by

$$\tilde{\mathcal{N}}(t,A) = \Psi(t,A) - \Lambda(t,A)$$

and therefore we define the *compensated Poisson integral* for $g \in L^1(A,\Lambda(1,A))$, i.e. all g with $E(\int_{\mathbb{R}} |g(y)|f(y)dy) < \infty$, by

$$\int_A g(x)\tilde{\mathcal{N}}(t,dx) = \int_A g(x)\Psi(t,dx) - \int_A g(x)\Lambda(t,dx).$$

Theorem 25. *Let $g \in L^1(A,\Lambda(1,A))$ and $t \geq 0$, then*

$$\int_A g(x)\tilde{\mathcal{N}}(t,dx)$$

is a martingale.

Proof. With $\mathcal{F}_t = \sigma\{X_s|s \leq t\}$ we get $\int_A g(x)\tilde{\mathcal{N}}(t,dx)$ is $\mathcal{F}_t$-measurable. Let us take a look at

$$\begin{aligned}
E\Big(\int_A g(x)\Psi(t,dx) - \int_A g(x)\Psi(s,dx)|\mathcal{F}_s\Big) =& E\Big(\sum_{n=1}^{\mathcal{N}_t} g(J_n)1_{\{J_n\in A\}} - \sum_{n=1}^{\mathcal{N}_s} g(J_n)1_{\{J_n\in A\}}|\mathcal{F}_s\Big)\\
=& E\Big(\sum_{n=\mathcal{N}_s+1}^{\mathcal{N}_t} g(J)1_{J\in A}|\mathcal{F}_s\Big)\\
=& E\Big(\sum_{n=\mathcal{N}_s}^{\mathcal{N}_t} g(J\in A)|\mathcal{N}_s\Big)\\
=& E\Big(\mathcal{N}_t - \mathcal{N}_s|\mathcal{N}_s\Big)E\Big(g(J\in A)|\mathcal{F}_s\Big)\\
=& E\Big(\mathcal{N}_{t-s}\Big)E\Big(g(J\in A)\Big)\\
=& \lambda(t-s)E\Big(g(J\in A)\Big)
\end{aligned}$$

with $s \leq t$ and then at

$$E\Big(\int_A g(x)\Lambda(t,dx) - \int_A g(x)\Lambda(s,dx)|\mathcal{F}_s\Big) = E\Big(\lambda(t-s)\int_A g(x)f(x)dx|\mathcal{F}_s\Big)$$
$$= E\Big(\lambda(t-s)E(g(J \in A))|\mathcal{F}_s\Big)$$
$$= \lambda(t-s)E\Big(g(J \in A)\Big).$$

For $s \leq t$ we get then

$$\begin{aligned} E\Big(\int_A g(x)\tilde{\mathcal{N}}(t,dx) - \int_A g(x)\tilde{\mathcal{N}}(s,dx)|\mathcal{F}_s\Big) &= E\Big(\int_A g(x)\Psi(t,dx) - \int_A g(x)\Psi(s,dx)|\mathcal{F}_s\Big) \\ &\quad - E\Big(\int_A g(x)\Lambda(t,dx) - \int_A g(x)\Lambda(s,dx)|\mathcal{F}_s\Big) \\ &= \lambda(t-s)E(g(J \in A)) \\ &\quad - \lambda(t-s)E(g(J \in A)) \\ &= 0 \end{aligned}$$

and therefore we get

$$E(\int_A g(x)\tilde{\mathcal{N}}(t,dx)|\mathcal{F}_s) = \int_A g(x)\tilde{\mathcal{N}}(s,dx).$$

Now we want to prove $E(|\int_A g(x)\tilde{\mathcal{N}}(t,dx)|) < \infty$. We get for $g \in L^1(A,\Lambda(1,A))$ that

$$E(|\int_A g(x)\Lambda(t,dx)|) = E(|\lambda t\int_A g(x)f(x)dx|) \leq \lambda t E(\int_a |g(x)|f(x)dx) < \infty$$

and

$$\begin{aligned} E(|\int_A g(x)\Psi(t,dx)|) &= E(|\sum_{n=1}^{\mathcal{N}_t} g(J)1_{J\in A}|) \\ &\leq E(|\sum_{n=1}^{\mathcal{N}_t} g(J)|) \\ &\leq E(\sum_{n=1}^{\mathcal{N}_t} |g(J)|) \\ &= \lambda t \underbrace{E(|g(J)|)}_{<\infty} < \infty. \end{aligned}$$

Then we gain with

$$|\int_A g(x)\tilde{\mathcal{N}}(t,dx)| \leq |\int_A g(x)\Psi(t,dx)|| + |\int_A g(x)\Lambda(t,dx)|,$$

that

$$E(|\int_A g(x)\tilde{\mathcal{N}}(t,dx)|) < \infty.$$

$\square$

Bibliography

[1] Appelbaum, D. (2004). "Lévy Processes and stochastic calculus". Cambridge studies in advanced mathematics

[2] Bibby, B.; Jacobsen, M.; Sørensen, M.(2004)."Estimating functions for discretely sampled diffusion-type models." AMS-preprint 4.

[3] Bibby, B.; Sørensen, M. (1995). "Martingale estimating functions for discretely observed diffusion processes". *Bernoulli*, 1:17-39.

[4] Billingsley, P. (1961a). "The Lindeberg-Lévy theorem for martingales". *Proc. Amer. Math. Soc.*, 12:788-792.

[5] Billingsley, P. (1961b). *Statistical interference for Markov processes.* The University of Chicago Press, Chicago.

[6] Cox, J. C.; Ingersoll, Jr., J. E. & Ross, S. A. (1985). "A theory of the term structure of interest rates". *Econometrica*, 53(2):385-407.

[7] Dellacherie, C.; Meyer, P. A. (1980). "Probabilités et Potential: Theorie des Martingales". Hermann

[8] Freedman, D. (1971). "Brownian motion and diffusion". Holden-Day, San Francisco

[9] Forman, J.L.; Sørensen, M. (2007). "The Pearson diffusions: a class of statistically tractable diffusion processes."

[10] Gihman, I.I.; Skorohod, A.V. (1972). "Stochastic differential equations". Springer, Berlin.

[11] Godambe, V.P.; Heyde, C. C. (1987). "Quasi likelihood and optimal estimation". *International Statistical Review*, 55:231-244.

[12] Hall, P.; Heyde, C.C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, *Econometrica* 50:1029-1054.

[13] Jacod, J.; Protter, P. (2004). "Probability essentials". Springer

[14] Kessler, M. & Sørensen, M. (1999). "Estimating equations based on eigenfunctions for a discretely observed diffusion process". *Bernoulli,* 5:299-314

[15] Nagahara, Y. (1996). "Non-Gaussian distribution for stock returns and related stochastic differential equation". *Financial Engineering and the Japanese Markets* 3:121-149

[16] Ozaki, T. (1985). "Non-linear time series models and dynamical systems". In Hannan, E. J.; Krishnaiah, P. R.; Rao, M. M., editors, *Handbook of Statistics, Vol. 5*, pages 25-83. Elsevier Science Publishers.

[17] Sørensen, M.(1991)."Likelihood methods for diffusions with jumps." *Statistical inference in stochastic processes* edited by Prabhu, N. U.; Basawa,I. V. pp. 67-105, Marcel Dekker, Inc.

[18] Wong, E.(1964). "The construction of a class of stationary Markoff processes". In Bellman, R., editor, *Stochastic Processes in Mathematical Physics and Engineering,* pages 264-276. American Mathematical Society, Rhode Island.

[19] Zhou, H. (2003). "Itô conditional moment generator and the estimation of short-rate processes". Journal of Financial Econometrics, Vol. 1, No. 2, pp. 250-271. Oxford University Press.