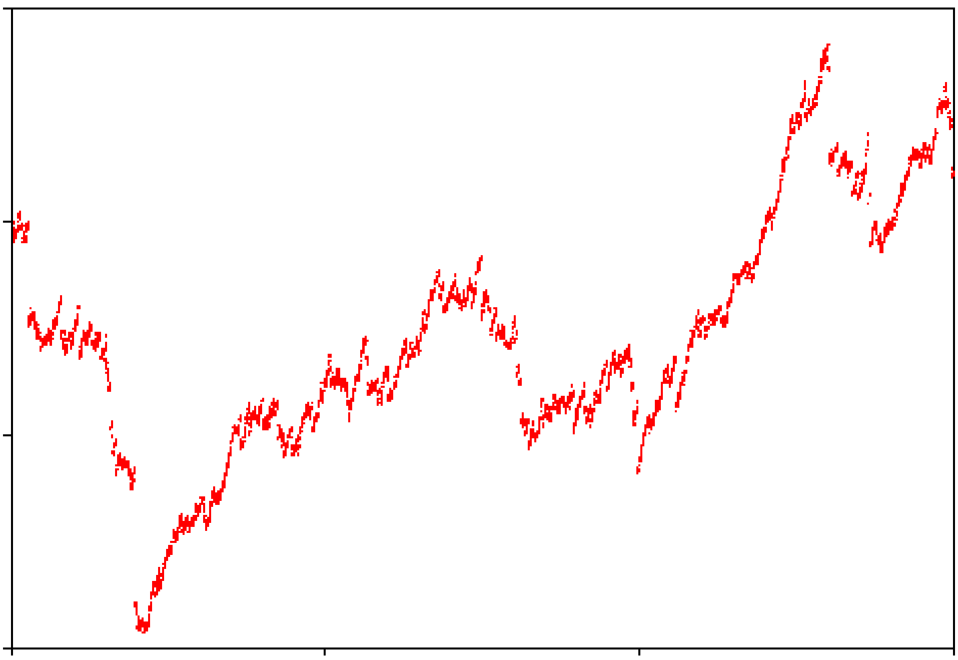
MS3b/MScMCF
Lévy Processes and Finance

Matthias Winkel
Department of Statistics
University of Oxford



HT 2010

MS3b (and MScMCF)
Lévy Processes and Finance

Matthias Winkel – 16 lectures HT 2010

Prerequisites

Part A Probability is a prerequisite. BS3a/OBS3a Applied Probability or B10 Martingales and Financial Mathematics would be useful, but are by no means essential; some material from these courses will be reviewed without proof.

Aims

Lévy processes form a central class of stochastic processes, contain both Brownian motion and the Poisson process, and are prototypes of Markov processes and semimartingales. Like Brownian motion, they are used in a multitude of applications ranging from biology and physics to insurance and finance. Like the Poisson process, they allow to model abrupt moves by jumps, which is an important feature for many applications. In the last ten years Lévy processes have seen a hugely increased attention as is reflected on the academic side by a number of excellent graduate texts and on the industrial side realising that they provide versatile stochastic models of financial markets. This continues to stimulate further research in both theoretical and applied directions. This course will give a solid introduction to some of the theory of Lévy processes as needed for financial and other applications.

Synopsis

Review of (compound) Poisson processes, Brownian motion (informal), Markov property. Connection with random walks, [Donsker's theorem], Poisson limit theorem. Spatial Poisson processes, construction of Lévy processes.

Special cases of increasing Lévy processes (subordinators) and processes with only positive jumps. Subordination. Examples and applications. Financial models driven by Lévy processes. Stochastic volatility. Level passage problems. Applications: option pricing, insurance ruin, dams.

Simulation: via increments, via simulation of jumps, via subordination. Applications: option pricing, branching processes.

Reading

- J.F.C. Kingman: *Poisson processes.* Oxford University Press (1993), Ch.1-5, 8
- A.E. Kyprianou: *Introductory lectures on fluctuations of Lévy processes with Applications.* Springer (2006), Ch. 1-3, 8-9
- W. Schoutens: *Lévy processes in finance: pricing financial derivatives.* Wiley (2003)

Further reading

- J. Bertoin: *Lévy processes.* Cambridge University Press (1996), Sect. 0.1-0.6, I.1, III.1-2, VII.1
- K. Sato: *Lévy processes and infinite divisibility.* Cambridge University Press (1999), Ch. 1-2, 4, 6, 9

Contents

Lecture 1

Introduction

Reading: Kyprianou Chapter 1
Further reading: Sato Chapter 1, Schoutens Sections 5.1 and 5.3

In this lecture we give the general definition of a Lévy process, study some examples of Lévy processes and indicate some of their applications. By doing so, we will review some results from BS3a Applied Probability and B10 Martingales and Financial Mathematics.

1.1 Definition of Lévy processes

Stochastic processes are collections of random variables X_t, $t \geq 0$ (meaning $t \in [0, \infty)$ as opposed to $n \geq 0$ by which means $n \in \mathbb{N} = \{0, 1, 2, \ldots\}$). For us, all X_t, $t \geq 0$, take values in a common state space, which we will choose specifically as $\mathbb{R}$ (or $[0, \infty)$ or $\mathbb{R}^d$ for some $d \geq 2$). We can think of X_t as the position of a particle at time t, changing as t varies. It is natural to suppose that the particle moves continuously in the sense that $t \mapsto X_t$ is continuous (with probability 1), *or* that it has jumps for some $t \geq 0$:

$$\Delta X_t = X_{t+} - X_{t-} = \lim_{\varepsilon \downarrow 0} X_{t+\varepsilon} - \lim_{\varepsilon \downarrow 0} X_{t-\varepsilon}.$$

We will usually suppose that these limits exist for all $t \geq 0$ and that in fact $X_{t+} = X_t$, i.e. that $t \mapsto X_t$ is *right-continuous with left limits* X_{t-} for all $t \geq 0$ almost surely. The path $t \mapsto X_t$ can then be viewed as a *random right-continuous function*.

Definition 1 (Lévy process) A real-valued (or $\mathbb{R}^d$-valued) stochastic process $X = (X_t)_{t \geq 0}$ is called a *Lévy process* if

(i) the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent for all $n \geq 1$ and $0 \leq t_0 < t_1 < \ldots < t_n$(*independent increments*),

(ii) $X_{t+s} - X_t$ has the same distribution as X_s for all $s, t \geq 0$ (*stationary increments*),

(iii) the paths $t \mapsto X_t$ are right-continuous with left limits (with probability 1).

It is implicit in (ii) that $\mathbb{P}(X_0 = 0) = 1$ (choose $s = 0$).

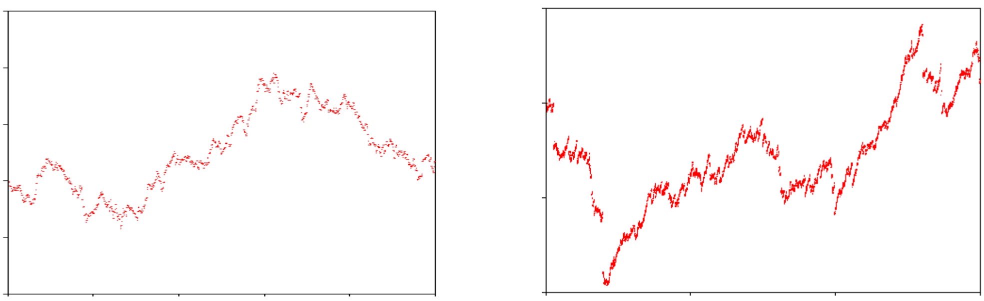


Figure 1.1: Variance Gamma process and a Lévy process with no positive jumps

Here the independence of n random variables is understood in the following sense:

Definition 2 (Independence) Let $Y^{(j)}$ be an $\mathbb{R}^{d_j}$-valued random variable for $j = 1, \ldots, n$. The random variables $Y^{(1)}, \ldots, Y^{(n)}$ are called *independent* if, for all (Borel measurable) $C^{(j)} \subset \mathbb{R}^{d_j}$

$$\mathbb{P}(Y^{(1)} \in C^{(1)}, \ldots, Y^{(n)} \in C^{(n)}) = \mathbb{P}(Y^{(1)} \in C^{(1)}) \ldots \mathbb{P}(Y^{(n)} \in C^{(n)}). \tag{1}$$

An *infinite* collection $(Y^{(j)})_{j \in J}$ is called independent if $Y^{(j_1)}, \ldots, Y^{(j_n)}$ are independent for *every* finite subcollection. *Infinite-dimensional* random variables $(Y_i^{(1)})_{i \in I_1}, \ldots, (Y_i^{(n)})_{i \in I_n}$ are called independent if $(Y_i^{(1)})_{i \in F_1}, \ldots, (Y_i^{(n)})_{i \in F_n}$ are independent for *all* finite $F_j \subset I_j$.

It is sufficient to check (1) for rectangles of the form $C^{(j)} = (a_1^{(j)}, b_1^{(j)}] \times \ldots \times (a_{d_j}^{(j)}, b_{d_j}^{(j)}]$.

1.2 First main example: Poisson process

Poisson processes are Lévy processes. We recall the definition as follows. An $\mathbb{N}(\subset \mathbb{R})$-valued stochastic process $X = (X_t)_{t \geq 0}$ is called a *Poisson process* with rate $\lambda \in (0, \infty)$ if X satisfies (i)-(iii) and

(iv)$^{\mathrm{Poi}}$ $\mathbb{P}(X_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$, $k \geq 0$, $t \geq 0$ (*Poisson distribution*).

The Poisson process is a continuous-time Markov chain. We will see that all Lévy processes have a Markov property. Also recall that Poisson processes have jumps of size 1 (spaced by independent exponential random variables $Z_n = T_{n+1} - T_n$, $n \geq 0$, with parameter λ, i.e. with density $\lambda e^{-\lambda s}$, $s \geq 0$). In particular, $\{t \geq 0 : \Delta X_t \neq 0\} = \{T_n, n \geq 1\}$ and $\Delta X_{T_n} = 1$ almost surely (short a.s., i.e. with probability 1). We can define more general Lévy processes by putting

$$C_t = \sum_{k=1}^{X_t} Y_k, \qquad t \geq 0,$$

for a Poisson process $(X_t)_{t \geq 0}$ and independent identically distributed Y_k, $k \geq 1$. Such processes are called *compound Poisson processes*. The term "compound" stems from the representation $C_t = S \circ X_t = S_{X_t}$ for the random walk $S_n = Y_1 + \ldots + Y_n$. You may think of X_t as the number of claims up to time t and of Y_k as the size of the kth claim. Recall (from BS3a) that its moment generating function, if it exists, is given by

$$\mathbb{E}(\exp\{\gamma C_t\}) = \exp\{\lambda t (\mathbb{E}(e^{\gamma Y_1} - 1))\}.$$

This will be an important building block of a general Lévy process.

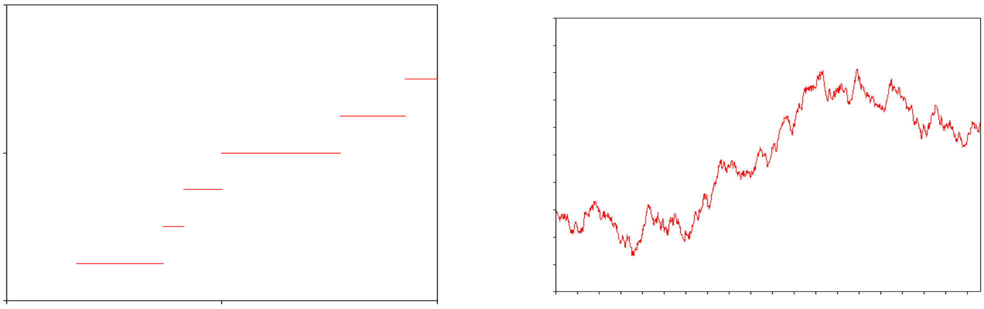


Figure 1.2: Poisson process and Brownian motion

1.3 Second main example: Brownian motion

Brownian motion is a Lévy process. We recall (from B10b) the definition as follows. An $\mathbb{R}$-valued stochastic process $X = (X_t)_{t\geq 0}$ is called Brownian motion if X satisfies (i)-(ii) and

(iii)$^{\rm BM}$ the paths $t \mapsto X_t$ are continuous almost surely,

(iv)$^{\rm BM}$ $\mathbb{P}(X_t \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi t}} \exp\left\{-\frac{y^2}{2t}\right\} dy$, $x \in \mathbb{R}$, $t > 0$. (*Normal distribution*).

The paths of Brownian motion are continuous, but turn out to be nowhere differentiable (we will not prove this). They exhibit erratic movements at all scales. This makes Brownian motion an appealing model for stock prices. Brownian motion has the *scaling property* $(\sqrt{c}X_{t/c})_{t\geq 0} \sim X$ where "$\sim$" means "has the same distribution as".

Brownian motion will be the other important building block of a general Lévy process.

The canonical space for Brownian paths is the space $C([0,\infty),\mathbb{R})$ of *continuous* real-valued functions $f : [0,\infty) \to \mathbb{R}$ which can be equipped with the topology of *locally uniform convergence*, induced by the metric

$$d(f,g) = \sum_{k\geq 1} 2^{-k} \min\{d_k(f,g),1\}, \qquad \text{where } d_k(f,g) = \sup_{x\in[0,k]} |f(x) - g(x)|.$$

This metric topology is complete (Cauchy sequences converge) and separable (has a countable dense subset), two attributes important for the existence and properties of limits. The bigger space $D([0,\infty),\mathbb{R})$ of *right-continuous* real-valued functions with left limits can also be equipped with the topology of locally uniform convergence. The space is still complete, but not separable. There is a *weaker metric topology*, called Skorohod's topology, that is complete and separable. In the present course we will not develop this and only occasionally use the familiar uniform convergence for (right-continuous) functions $f, f_n : [0,k] \to \mathbb{R}$, $n \geq 1$:

$$\sup_{x\in[0,k]} |f_n(x) - f(x)| \to 0, \qquad \text{as } n \to \infty,$$

which for stochastic processes $X, X^{(n)}$, $n \geq 1$, with time range $t \in [0,T]$ takes the form

$$\sup_{t\in[0,T]} |X_t^{(n)} - X_t| \to 0, \qquad \text{as } n \to \infty,$$

and will be as a convergence in probability *or* as almost sure convergence (from BS3a or B10a) *or* as L^2-convergence, where $Z_n \to Z$ in the L^2-sense means $\mathbb{E}(|Z_n - Z|^2) \to 0$.

1.4 Markov property

The Markov property is a consequence of the independent increments property (and the stationary increments property):

Proposition 3 (Markov property) *Let X be a Lévy process and $t \geq 0$ a fixed time, then the pre-t process $(X_r)_{r \leq t}$ is independent of the post-t process $(X_{t+s} - X_t)_{s \geq 0}$, and the post-$t$ process has the same distribution as X.*

Proof: By Definition 2, we need to check the independence of $(X_{r_1}, \ldots, X_{r_n})$ and $(X_{t+s_1} - X_t, \ldots, X_{t+s_m} - X_t)$. By property (i) of the Lévy process, we have that increments over disjoint time intervals are independent, in particular the increments

$$X_{r_1}, X_{r_2} - X_{r_1}, \ldots, X_{r_n} - X_{r_{n-1}}, X_{t+s_1} - X_t, X_{t+s_2} - X_{t+s_1}, \ldots, X_{t+s_m} - X_{t+s_{m-1}}.$$

Since functions (here linear transformations from increments to marginals) of independent random variables are independent, the proof of independence is complete. Identical distribution follows first on the level of single increments from (ii), then by (i) and linear transformation also for finite-dimensional marginal distributions. □

1.5 Some applications

Example 4 (Insurance ruin) A compound Poisson process $(Z_t)_{t \geq 0}$ with positive jump sizes A_k, $k \geq 1$, can be interpreted as a claim process recording the total claim amount incurred before time t. If there is linear premium income at rate $r > 0$, then also the gain process $rt - Z_t$, $t \geq 0$, is a Lévy process. For an initial reserve of $u > 0$, the reserve process $u + rt - Z_t$ is a shifted Lévy process starting from a non-zero initial value u.

Example 5 (Financial stock prices) Brownian motion $(B_t)_{t \geq 0}$ or linear Brownian motion $\sigma B_t + \mu t$, $t \geq 0$, was the first model of stock prices, introduced by Bachelier in 1900. Black, Scholes and Merton studied geometric Brownian motion $\exp(\sigma B_t + \mu t)$ in 1973, which is not itself a Lévy process but can be studied with similar methods. The Economics Nobel Prize 1997 was awarded for their work. Several deficiencies of the Black-Scholes model have been identified, e.g. the Gaussian density decreases too quickly, no variation of the volatility σ over time, no macroscopic jumps in the price processes. These deficiencies can be addressed by models based on Lévy processes. The Variance gamma model is a time-changed Brownian motion B_{T_s} by an independent increasing jump process, a so-called Gamma Lévy process with $T_s \sim \text{Gamma}(\alpha s, \beta)$. The process B_{T_s} is then also a Lévy process itself.

Example 6 (Population models) Branching processes are generalisations of birth-and-death processes (see BS3a) where each individual in a population dies after an exponentially distributed lifetime with parameter μ, but gives birth not to single children, but to twins, triplets, quadruplet etc. To simplify, it is assumed that children are only born at the end of a lifetime. The numbers of children are independent and identically distributed according to an offspring distribution q on $\{0, 2, 3, \ldots\}$. The population size process $(Z_t)_{t \geq 0}$ can jump downwards by 1 or upwards by an integer. It is not a Lévy process but is closely related to Lévy processes and can be studied with similar methods. There are also analogues of processes in $[0, \infty)$, so-called continuous-state branching processes that are useful large-population approximations.

Lecture 2

Lévy processes and random walks

Reading: Kingman Section 1.1, Grimmett and Stirzaker Section 3.5(4)
Further reading: Sato Section 7, Durrett Sections 2.8 and 7.6, Kallenberg Chapter 15

Lévy processes are the continuous-time analogues of random walks. In this lecture we examine this analogy and indicate connections via scaling limits and other limiting results. We begin with a first look at infinite divisibility.

2.1 Increments of random walks and Lévy processes

Recall that a random walk is a stochastic process in discrete time

$$S_0 = 0, \qquad S_n = \sum_{j=1}^{n} A_j, \qquad n \geq 1,$$

for a family $(A_j)_{j\geq 1}$ of independent and identically distributed real-valued (or $\mathbb{R}^d$-valued) random variables. Clearly, random walks have stationary and independent increments. Specifically, the A_j, $j \geq 1$, themselves are the increments over single time units. We refer to $S_{n+m} - S_n$ as an increment over m time units, $m \geq 1$.

While every distribution may be chosen for A_j, increments over m time units are sums of m independent and identically distributed random variables, and not every distribution has this property. This is not a deep observation, but it becomes important when moving to Lévy processes. In fact, the increment distribution of Lévy processes is restricted: any increment $X_{t+s} - X_t$, or X_s for simplicity, can be decomposed, for every $m \geq 1$,

$$X_s = \sum_{j=1}^{m} (X_{js/m} - X_{(j-1)s/m})$$

into a sum of m independent and identically distributed random variables.

Definition 7 (Infinite divisibility) A random variable Y is said to have an *infinitely divisible distribution* if for every $m \geq 1$, we can write

$$Y \sim Y_1^{(m)} + \ldots + Y_m^{(m)}$$

for some independent and identically distributed random variables $Y_1^{(m)}, \ldots, Y_m^{(m)}$.

We stress that the distribution of $Y_j^{(m)}$ may vary as m varies, but not as j varies.

The argument just before the definition shows that increments of Lévy processes are infinitely divisible. Many known distributions are infinitely divisible, some are not.

Example 8 The Normal, Poisson, Gamma and geometric distributions are infinitely divisible. This often follows from the closure under convolutions of the type

$$Y_1 \sim \text{Normal}(\mu, \sigma^2), Y_2 \sim \text{Normal}(\nu, \tau^2) \quad \Rightarrow \quad Y_1 + Y_2 \sim \text{Normal}(\mu + \nu, \sigma^2 + \tau^2)$$

for independent Y_1 and Y_2 since this implies by induction that for independent

$$Y_1^{(m)}, \ldots, Y_m^{(m)} \sim \text{Normal}(\mu/m, \sigma^2/m) \quad \Rightarrow \quad Y_1^{(m)} + \ldots + Y_m^{(m)} \sim \text{Normal}(\mu, \sigma^2).$$

The analogous arguments (and calculations, if necessary) for the other distributions are left as an exercise. The geometric(p) distribution here is $\mathbb{P}(X = n) = p^n(1 - p)$, $n \geq 0$.

Example 9 The Bernoulli(p) distribution, for $p \in (0, 1)$, is *not* infinitely divisible. Assume that you can represent a Bernoulli(p) random variable X as $Y_1 + Y_2$ for independent identically distributed Y_1 and Y_2. Then

$$\mathbb{P}(Y_1 > 1/2) > 0 \quad \Rightarrow \quad 0 = \mathbb{P}(X > 1) \geq \mathbb{P}(Y_1 > 1/2, Y_2 > 1/2) > 0$$

is a contradiction, so we must have $\mathbb{P}(Y_1 > 1/2) = 0$, but then

$$\mathbb{P}(Y_1 > 1/2) = 0 \ \Rightarrow \ p = \mathbb{P}(X = 1) = \mathbb{P}(Y_1 = 1/2)\mathbb{P}(Y_2 = 1/2) \ \Rightarrow \ \mathbb{P}(Y_1 = 1/2) = \sqrt{p}.$$

Similarly,

$$\mathbb{P}(Y_1 < 0) > 0 \quad \Rightarrow \quad 0 = \mathbb{P}(X < 0) \geq \mathbb{P}(Y_1 < 0, Y_2 < 0) > 0$$

is a contradiction, so we must have $\mathbb{P}(Y_1 < 0) = 0$ and then

$$1 - p = \mathbb{P}(X = 0) = \mathbb{P}(Y_1 = 0, Y_2 = 0) \quad \Rightarrow \quad \mathbb{P}(Y_1 = 0) = \sqrt{1 - p} > 0.$$

This is impossible for several reasons. Clearly, $\sqrt{p} + \sqrt{1 - p} > 1$, but also

$$0 = \mathbb{P}(X = 1/2) \geq \mathbb{P}(Y_1 = 0)\mathbb{P}(Y_2 = 1/2) > 0.$$

2.2 Central Limit Theorem and Donsker's theorem

Theorem 10 (Central Limit Theorem) *Let $(S_n)_{n \geq 0}$ be a random walk with $\mathbb{E}(S_1^2) = \mathbb{E}(A_1^2) < \infty$. Then, as $n \to \infty$,*

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mathbb{E}(A_1)}{\sqrt{n\text{Var}(A_1)}} \to \text{Normal}(0, 1) \qquad \textit{in distribution.}$$

This result as a result for one time $n \to \infty$ can be extended to a convergence of processes, a convergence of the discrete-time process $(S_n)_{n \geq 0}$ to a (continuous-time) Brownian motion, by scaling of both space and time. The processes

$$\frac{S_{[nt]} - [nt]\mathbb{E}(A_1)}{\sqrt{n\text{Var}(A_1)}}, \qquad t \geq 0,$$

where $[nt] \in \mathbb{Z}$ with $[nt] \leq nt < [nt] + 1$ denotes the integer part of nt, are scaled versions of the random walk $(S_n)_{n \geq 0}$, now performing n steps per time unit (holding time $1/n$), centred and each only a multiple $1/\sqrt{n\text{Var}(A_1)}$ of the original size. If $\mathbb{E}(A_1) = 0$, you may think that you look at $(S_n)_{n \geq 0}$ from further and further away, but note that space and time are scaled differently, in fact so as to yield a non-trivial limit.

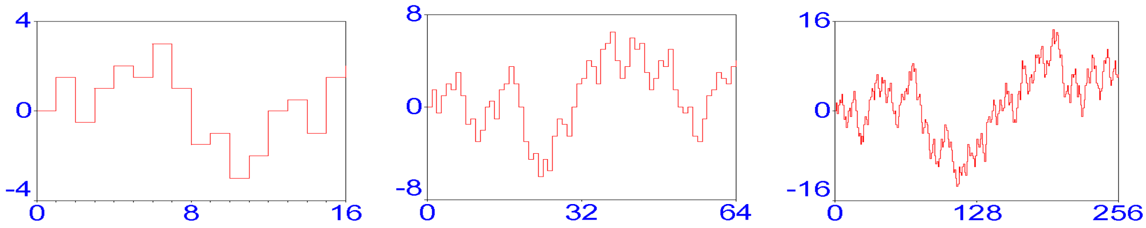


Figure 2.1: Random walk converging to Brownian motion

Theorem 11 (Donsker) *Let $(S_n)_{n\geq 0}$ be a random walk with $\mathbb{E}(S_1^2) = \mathbb{E}(A_1^2) < \infty$. Then, as $n \to \infty$,*

$$\frac{S_{[nt]} - [nt]\mathbb{E}(A_1)}{\sqrt{n\mathrm{Var}(A_1)}} \to B_t \qquad \text{locally uniformly in } t \geq 0,$$

"in distribution", for a Brownian motion $(B_t)_{t\geq 0}$.

Proof: [only for $A_1 \sim \mathrm{Normal}(0,1)$] This proof is a coupling proof. We are not going to work directly with the original random walk $(S_n)_{n\geq 0}$, but start from Brownian motion $(B_t)_{t\geq 0}$ and define a family of embedded random walks

$$S_k^{(n)} := B_{k/n}, \qquad k \geq 0, n \geq 1.$$

Then note using in particular $\mathbb{E}(A_1) = 0$ and $\mathrm{Var}(A_1) = 1$ that

$$S_1^{(n)} \sim \mathrm{Normal}(0, 1/n) \sim \frac{S_1 - \mathbb{E}(A_1)}{\sqrt{n\mathrm{Var}(A_1)}},$$

and indeed

$$\left(S_{[nt]}^{(n)}\right)_{t\geq 0} \sim \left(\frac{S_{[nt]} - [nt]\mathbb{E}(A_1)}{\sqrt{n\mathrm{Var}(A_1)}}\right)_{t\geq 0}.$$

To show convergence in distribution for the processes on the right-hand side, it suffices to establish convergence in distribution for the processes on the left-hand side, as $n \to \infty$.

To show locally uniform convergence we take an arbitrary $T \geq 0$ and show uniform convergence on $[0, T]$. Since $(B_t)_{0\leq t\leq T}$ is uniformly continuous (being continuous on a compact interval), we get a.s.

$$\sup_{0\leq t\leq T} \left|S_{[nt]}^{(n)} - B_t\right| \leq \sup_{0\leq s\leq t\leq T:|s-t|\leq 1/n} |B_s - B_t| \to 0$$

as $n \to \infty$. This establishes a.s. convergence, which "implies" convergence in distribution for the embedded random walks and for the original scaled random walk. This completes the proof for $A_1 \sim \mathrm{Normal}(0,1)$. □

Note that the almost sure convergence only holds for the embedded random walks $(S_k^{(n)})_{k\geq 0}$, $n \geq 1$. Since the identity in distribution with the rescaled original random walk only holds for fixed $n \geq 1$, not jointly, we cannot deduce almost sure convergence in the statement of the theorem. Indeed, it can be shown that almost sure convergence will fail. The proof for general increment distribution is much harder and will not be given in this course. If time permits, we will give a similar coupling proof for another important special case where $\mathbb{P}(A_1 = 1) = \mathbb{P}(A_1 = -1) = 1/2$, the simple symmetric random walk.

2.3 Poisson limit theorem

The Central Limit Theorem for Bernoulli random variables $A_1, \ldots, A_n$ says that for large n, the number of 1s in the sequence is well-approximated by a Normal random variable. In practice, the approximation is good if p is not too small. If p is small, the Bernoulli random variables count rare events, and a different limit theorem is relevant:

Theorem 12 (Poisson limit theorem) *Let W_n be binomially distributed with parameters n and $p_n = \lambda/n$ (or if $np_n \to \lambda$, as $n \to \infty$). Then we have*

$$W_n \to Poi(\lambda), \qquad \text{in distribution, as } n \to \infty.$$

Proof: Just calculate that, as $n \to \infty$,

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)\ldots(n-k+1)}{k!} \frac{(np_n)^k}{n^k} \frac{\left(1-\frac{np_n}{n}\right)^n}{(1-p_n)^k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

□

Theorem 13 *Suppose that $S_k^{(n)} = A_1^{(n)} + \ldots + A_k^{(n)}$, $k \geq 0$, is the sum of independent* Bernoulli(p_n) *random variables for all $n \geq 1$, and that $np_n \to \lambda \in (0, \infty)$. Then*

$$S_{[nt]}^{(n)} \to N_t \qquad \text{"in the Skorohod sense" as functions of } t \geq 0,$$

"in distribution" as $n \to \infty$, for a Poisson process $(N_t)_{t\geq 0}$ with rate λ.

The proof of so-called finite-dimensional convergence for vectors $(S_{[nt_1]}^{(n)}, \ldots, S_{[nt_m]}^{(n)})$ is not very hard but not included here. One can also show that the jump times $(T_m^{(n)})_{m\geq 1}$ of $(S_{[nt]}^{(n)})_{t\geq 0}$ converge to the jump times of a Poisson process. E.g.

$$\mathbb{P}(T_1^{(n)} > t) = (1-p_n)^{[nt]} = \left(1 - \frac{[nt]p_n}{[nt]}\right)^{[nt]} \to \exp\{-\lambda t\},$$

since $[nt]/n \to t$ (since $(nt-1)/n \to t$ and $nt/n = t$) and so $[nt]p_n \to t\lambda$. The general statement is hard to make precise and prove, certainly beyond the scope of this course.

2.4 Generalisations

Infinitely divisible distributions and Lévy processes are precisely the classes of limits that arise for random walks as in Theorems 10 and 12 (respectively 11 and 13) with different step distributions. Stable Lévy processes are ones with a scaling property $(c^{1/\alpha} X_{t/c})_{t\geq 0} \sim X$ for some $\alpha \in \mathbb{R}$. These exist, in fact, for $\alpha \in (0, 2]$. Theorem 10 (and 11) for suitable distributions of A_1 (depending on α and where $\mathbb{E}(A_1^2) = \infty$ in particular) then yield convergence in distribution

$$\frac{S_n - n\mathbb{E}(A_1)}{n^{1/\alpha}} \to \text{stable}(\alpha) \qquad \text{for } \alpha \geq 1, \qquad \text{or} \qquad \frac{S_n}{n^{1/\alpha}} \to \text{stable}(\alpha) \qquad \text{for } \alpha \leq 1.$$

Example 14 (Brownian ladder times) *For a Brownian motion B and a level $r > 0$, the distribution of $T_r = \inf\{t \geq 0 : B_t > r\}$ is 1/2-stable, see later in the course.*

Example 15 (Cauchy process) *The Cauchy distribution with density $a/(\pi(x^2 + a^2))$, $x \in \mathbb{R}$, for some parameter $c \in \mathbb{R}$ is 1-stable, see later in the course.*

Lecture 3

Spatial Poisson processes

Reading: Kingman 1.1 and 2.1, Grimmett and Stirzaker 6.13, Kyprianou Section 2.2
Further reading: Sato Section 19

We will soon construct the most general nonnegative Lévy process (and then general real-valued ones). Even though we will not prove that they are the most general, we have already seen that only infinitely divisible distributions are admissible as increment distributions, so we know that there are restrictions; the part missing in our discussion will be to show that a given distribution is infinitely divisible only if there exists a Lévy process X of the type that we will construct such that X_1 has the given distribution. Today we prepare the construction by looking at spatial Poisson processes, objects of interest in their own right.

3.1 Motivation from the study of Lévy processes

Brownian motion $(B_t)_{t\geq 0}$ has continuous sample paths. It turns out that $(\sigma B_t + \mu t)_{t\geq 0}$ for $\sigma \geq 0$ and $\mu \in \mathbb{R}$ is the only continuous Lévy process. To describe the full class of Lévy processes $(X_t)_{t\geq 0}$, it is vital to study the process $(\Delta X_t)_{t\geq 0}$ of jumps.

Take e.g. the Variance Gamma process. In Assignment 1.2.(b), we introduce this process as $X_t = G_t - H_t$, $t \geq 0$, for two independent Gamma Lévy processes G and H. But how do Gamma Lévy processes evolve? We could simulate discretisations (and will do!) and get some feeling for them, but we also want to understand them mathematically. Do they really exist? We have not shown this. Are they compound Poisson processes? Let us look at their moment generating function (cf. Assignment 2.4.):

$$\mathbb{E}(\exp\{\gamma G_t\}) = \left(\frac{\beta}{\beta-\gamma}\right)^{\alpha t} = \exp\left\{\alpha t \int_0^\infty (e^{\gamma x}-1)\frac{1}{x}e^{-\beta x}dx\right\}.$$

This is almost of the form of a compound Poisson process of rate λ with non-negative jump sizes Y_j, $j \geq 1$, that have a probability density function $h(x) = h_{Y_1}(x)$, $x > 0$:

$$\mathbb{E}(\exp\{\gamma C_t\}) = \exp\{\lambda t \int_0^\infty (e^{\gamma x}-1)h(x)dx\}$$

To match the two expressions, however, we would have to put

$$\lambda h(x) = \lambda_0 h^{(0)}(x) = \frac{\alpha}{x}e^{-\beta x}, \qquad x > 0,$$

and $h^{(0)}$ *cannot* be a probability density function, because $\frac{\alpha}{x}e^{-\beta x}$ is not integrable at $x \downarrow 0$. What we can do is e.g. truncate at $\varepsilon > 0$ and specify

$$\lambda_\varepsilon h^{(\varepsilon)}(x) = \frac{\alpha}{x}e^{-\beta x}, \quad x > \varepsilon, \qquad h^{(\varepsilon)}(x) = 0, \quad x \le \varepsilon.$$

In order for $h^{(\varepsilon)}$ to be a probability density, we just put $\lambda_\varepsilon = \int_\varepsilon^\infty \frac{\alpha}{x}e^{-\beta x}dx$, and notice that $\lambda_\varepsilon \to \infty$ as $\varepsilon \downarrow 0$. But λ_ε is the rate of the Poisson process driving the compound Poisson process, so jumps are more and more frequent as $\varepsilon \downarrow 0$. On the other hand, the average jump size, the mean of the distribution with density $h^{(\varepsilon)}$ tends to zero, so most of these jumps are very small. In fact, we will see that

$$G_t = \sum_{s \le t} \Delta G_s,$$

as an absolutely convergent series of infinitely (but clearly countably) many positive jump sizes, where $(\Delta G_s)_{s\ge 0}$ is a Poisson point process with intensity $g(x) = \frac{\alpha}{x}e^{-\beta x}$, $x > 0$, the collection of random variables

$$N((a,b] \times (c,d]) = \#\{t \in (a,b] : \Delta G_t \in (c,d]\}, \qquad 0 \le a < b, 0 < c < d$$

a Poisson counting measure (evaluated on rectangles) with intensity function $\lambda(t,x) = g(x)$, $x > 0$, $t \ge 0$; the random countable set $\{(t, \Delta G_t) : t \ge 0 \text{ and } \Delta C_t \ne 0\}$ a spatial Poisson process with intensity $\lambda(t,x)$. Let us now formally introduce these notions.

3.2 Poisson counting measures

The essence of one-dimensional Poisson processes $(N_t)_{t\ge 0}$ is the set of arrival ("event") times $\Pi = \{T_1, T_2, T_3, \ldots\}$, which is a random countable set. The increment $N((s,t]) := N_t - N_s$ counts the number of points in $\Pi \cap (s,t]$. We can generalise this concept to counting measures of random countable subsets on other spaces, say $\mathbb{R}^d$. Saying directly what exactly (the distribution of) random countable sets is, is quite difficult in general. Random counting measures are a way to describe the random countable sets implicitly.

Definition 16 (Spatial Poisson process) A random countable subset $\Pi \subset \mathbb{R}^d$ is called a *spatial Poisson process* with (constant) intensity λ if the random variables $N(A) = \#\Pi \cap A$, $A \subset \mathbb{R}^d$ (Borel measurable, *always, for the whole course, but we stop saying this all the time now*), satisfy

(a) for all $n \ge 1$ and disjoint $A_1, \ldots, A_n \subset \mathbb{R}^d$, the random variables $N(A_1), \ldots, N(A_n)$ are independent,

${}^{\text{hom}}$(b) $N(A) \sim \text{Poi}(\lambda|A|)$, where $|A|$ denotes the volume (Lebesgue measure) of A.

Here, we use the convention that $X \sim \text{Poi}(0)$ means $\mathbb{P}(X = 0) = 1$ and $X \sim \text{Poi}(\infty)$ means $\mathbb{P}(X = \infty) = 1$. This is consistent with $\mathbb{E}(X) = \lambda$ for $X \sim \text{Poi}(\lambda)$, $\lambda \in (0,\infty)$. This convention captures that Π does not have points in a given set of zero volume a.s., and it has infinitely many points in given sets of infinite volume a.s.

In fact, the definition fully specifies the joint distributions of the random set function N on subsets of $\mathbb{R}^d$, since for any non-disjoint $B_1, \ldots, B_m \subset \mathbb{R}^d$ we can consider all intersections of the form $A_k = B_1^* \cap \ldots \cap B_m^*$, where each B_j^* is either $B_j^* = B_j$ or $B_j^* = B_j^c = \mathbb{R}^d \setminus B_j$. They form $n = 2^m$ disjoint sets $A_1, \ldots, A_n$ to which (a) of the definition applies. $(N(B_1), \ldots, N(B_m))$ is a just a linear transformation of $(N(A_1), \ldots, N(A_n))$.

Grimmett and Stirzaker collect a long list of applications including modelling stars in a galaxy, galaxies in the universe, weeds in the lawn, the incidence of thunderstorms and tornadoes. Sometimes the process in Definition 16 is not a perfect description of such a system, but useful as a first step. A second step is the following generalisation:

Definition 16 (Spatial Poisson process, continued) A random countable subset $\Pi \subset D \subset \mathbb{R}^d$ is called a *spatial Poisson process* with (locally integrable) *intensity function* $\lambda : D \to [0, \infty)$, if $N(A) = \#\Pi \cap A$, $A \subset D$, satisfy

(a) for all $n \geq 1$ and disjoint $A_1, \ldots, A_n \subset D$, the random variables $N(A_1), \ldots, N(A_n)$ are independent,

${}^{\text{inhom}}$(b) $N(A) \sim \text{Poi}\left(\int_A \lambda(x)dx\right)$.

Definition 17 (Poisson counting measure) A set function $A \mapsto N(A)$ that satisfies (a) and ${}^{\text{inhom}}$(b) is referred to as a *Poisson counting measure* with intensity function $\lambda(x)$.

It is sufficient to check (a) and (b) for rectangles $A_j = (a_1^{(j)}, b_1^{(j)}] \times \ldots \times (a_d^{(j)}, b_d^{(j)}]$.

The set function $\Lambda(A) = \int_A \lambda(x)dx$ is called the intensity *measure* of Π. Definitions 16 and 17 can be extended to measures that are *not* integrals of intensity *functions*. Only if $\Lambda(\{x\}) > 0$, we would require $\mathbb{P}(N(\{x\}) \geq 2) > 0$ and this is incompatible with $N(\{x\}) = \#\Pi \cap \{x\}$ for a random countable set Π, so we prohibit such "atoms" of Λ.

Example 18 (Compound Poisson process) Let $(C_t)_{t\geq 0}$ be a compound Poisson process with independent jump sizes Y_j, $j \geq 1$ with common probability density $h(x)$, $x > 0$, at the times of a Poisson process $(X_t)_{t\geq 0}$ with rate $\lambda > 0$. Let us show that

$$N((a, b] \times (c, d]) = \#\{t \in (a, b] : \Delta C_t \in (c, d]\}$$

defines a Poisson counting measure. First note $N((a, b] \times (0, \infty)) = X_b - X_a$. Now recall

> **Thinning property of Poisson processes:** If each point of a Poisson process $(X_t)_{t\geq 0}$ of rate λ is of type 1 with probability p and of type 2 with probability $1 - p$, independently of one another, then the processes $X^{(1)}$ and $X^{(2)}$ counting points of type 1 and 2, respectively, are independent Poisson processes with rates $p\lambda$ and $(1 - p)\lambda$, respectively.

Consider the thinning mechanism, where the jth jump is of type 1 if $Y_j \in (c, d]$. Then, the process counting jumps in $(c, d]$ is a Poisson process with rate $\lambda\mathbb{P}(Y_1 \in (c, d])$, and so

$$N((a, b] \times (c, d]) = X_b^{(1)} - X_a^{(1)} \sim \text{Poi}((b - a)\lambda\mathbb{P}(Y_1 \in (c, d])).$$

We identify the intensity measure $\Lambda((a, b] \times (c, d]) = (b - a)\lambda\mathbb{P}(Y_1 \in (c, d])$.

For the independence of counts in disjoint rectangles $A_1, \ldots, A_n$, we cut them into smaller rectangles $B_i = (a_i, b_i] \times (c_i, d_i]$, $1 \leq i \leq m$ such that for any two B_i and B_j either $(c_i, d_i] = (c_j, d_j]$ or $(c_i, d_i] \cap (c_j, d_j] = \varnothing$. Denote by k the number of different intervals $(c_i, d_i]$, w.l.o.g. $(c_i, d_i]$ for $1 \leq i \leq k$. Now a straightforward generalisation of the thinning property to k types splits $(X_t)_{t\geq 0}$ into k independent Poisson processes $X^{(i)}$ with rates $\lambda\mathbb{P}(Y_1 \in (c_i, d_i])$, $1 \leq i \leq k$. Now $N(B_1), \ldots, N(B_m)$ are independent as increments of independent Poisson processes or of the same Poisson process over disjoint time intervals.

3.3 Poisson point processes

In Example 18, the intensity measure is of the product form $\Lambda((a,b]\times(c,d]) = (b-a)\nu((c,d])$ for a measure ν on $D_0 = (0,\infty)$. Take $D = [0,\infty)\times D_0$ in Definition 16. This means, that the spatial Poisson process is homogeneous in the first component, the time component, like the Poisson process.

Proposition 19 *If $\Lambda((a,b]\times A_0) = (b-a)\int_{A_0} g(x)dx$ for a locally integrable function g on D_0 (or $= (b-a)\nu(A_0)$ for a locally finite measure ν on D_0), then no two points of Π share the same first coordinate.*

Proof: If ν is finite, this is clear, since then $X_t = N([0,t]\times D_0)$, $t \ge 0$, is a Poisson process with rate $\nu(D_0)$. Let us restrict attention to $D_0 = \mathbb{R}^* = \mathbb{R}\setminus\{0\}$ for simplicity – this is the most relevant case for us. The local integrability condition means that we can find intervals $(I_n)_{n\ge1}$ such that $\bigcup_{n\ge1} I_n = D_0$ and $\nu(I_n) < \infty$, $n \ge 1$. Then the independence of $N((t_{j-1},t_j]\times I_n)$, $j = 1,\ldots,m$, $n\ge1$, implies that $X_t^{(n)} = N([0,t]\times I_n)$, $t \ge 0$, are independent Poisson processes with rates $\nu(I_n)$, $n \ge 1$. Therefore any two of the jump times $(T_j^{(n)}, j\ge1, n\ge1)$ are jointly continuously distributed and take different values almost surely:

$$\mathbb{P}(T_j^{(n)} = T_i^{(m)}) = \int_0^\infty \int_x^x f_{T_j^{(n)}}(x) f_{T_i^{(m)}}(y)dydx = 0 \qquad \text{for all } n \neq m.$$

[Alternatively, show that $T_j^{(n)} - T_i^{(m)}$ has a continuous distribution and hence does not take a fixed value 0 almost surely].

Finally, there are only countably many pairs of jump times, so almost surely no two jump times coincide. □

Let Π be a spatial Poisson process with intensity measure $\Lambda((a,b]\times(c,d]) = (b-a)\int_c^d g(x)dx$ for a locally integrable function g on D_0 (or $= (b-a)\nu((c,d])$ for a locally finite measure ν on D_0), then the process $(\Delta_t)_{t\ge0}$ given by

$$\Delta_t = 0 \quad \text{if } \Pi\cap\{t\}\times D_0 = \varnothing, \qquad \Delta_t = x \quad \text{if } \Pi\cap\{t\}\times D_0 = \{(t,x)\}$$

is a Poisson point process in $D_0\cup\{0\}$ with intensity function g on D_0 in the sense of the following definition.

Definition 20 (Poisson point process) Let g be locally integrable on $D_0 \subset \mathbb{R}^{d-1}\setminus\{0\}$ (or ν locally finite). A process $(\Delta_t)_{t\ge0}$ in $D_0\cup\{0\}$ such that

$$N((a,b]\times A_0) = \#\{t\in(a,b] : \Delta_t \in A_0\}, \qquad 0 \le a < b, A_0 \subset D_0 \text{ (measurable)},$$

is a Poisson counting measure with intensity $\Lambda((a,b]\times A_0) = (b-a)\int_{A_0} g(x)dx$ (or $\Lambda((a,b]\times A_0) = (b-a)\nu(A_0)$), is called a *Poisson point process with intensity g (or intensity measure ν).*

Note that for every Poisson point process, the set $\Pi = \{(t,\Delta_t) : t \ge 0, \Delta_t \ne 0\}$ is a spatial Poisson process. Poisson random measure and Poisson point process are representations of this spatial Poisson process. Poisson point processes as we have defined them always have a time coordinate and are homogeneous in time, but not in their spatial coordinates.

In the next lecture we will see how one can do computations with Poisson point processes, notably relating to $\sum \Delta_t$.

Lecture 4

Spatial Poisson processes II

Reading: Kingman Sections 2.2, 2.5, 3.1; Further reading: Williams Chapters 9 and 10

In this lecture, we construct spatial Poisson processes and study sums $\sum_{s\leq t} f(\Delta_s)$ over Poisson point processes $(\Delta_t)_{t\geq 0}$. We will identify $\sum_{s\leq t} \Delta_s$ as Lévy process next lecture.

4.1 Series and increasing limits of random variables

Recall that for two independent Poisson random variables $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ we have $X + Y \sim \text{Poi}(\lambda + \mu)$. Much more is true. A simple induction shows that

$$X_j \sim \text{Poi}(\mu_j), 1 \leq j \leq m, \text{ independent } \Rightarrow X_1 + \ldots + X_m \sim \text{Poi}(\mu_1 + \ldots + \mu_m).$$

What about countably infinite families with $\mu = \sum_{m\geq 1} \mu_m < \infty$? Here is a general result, a bit stronger than the convergence theorem for moment generating functions.

Lemma 21 *Let $(Z_m)_{m\geq 1}$ be an* increasing *sequence of $[0,\infty)$-valued random variables. Then $Z = \lim_{m\to\infty} Z_m$ exists a.s. as a $[0,\infty]$-valued random variable. In particular,*

$$\mathbb{E}(e^{\gamma Z_m}) \to \mathbb{E}(e^{\gamma Z}) = M(\gamma) \qquad \textit{for all } \gamma \neq 0.$$

We have

$$\mathbb{P}(Z < \infty) = 1 \iff \lim_{\gamma \uparrow 0} M(\gamma) = 1$$
$$\textit{and} \qquad \mathbb{P}(Z = \infty) = 1 \iff M(\gamma) = 0 \textit{ for all (one) } \gamma < 0.$$

Proof: Limits of increasing sequences exist in $[0,\infty]$. Hence, if a random sequence $(Z_m)_{m\geq 1}$ is increasing a.s., its limit Z exists in $[0,\infty]$ a.s. Therefore, we also have $e^{\gamma Z_m} \to e^{\gamma Z} \in [0,\infty]$ with the conventions $e^{-\infty} = 0$ and $e^{\infty} = \infty$. Then (by monotone convergence) $\mathbb{E}(e^{\gamma Z_m}) \to \mathbb{E}(e^{\gamma Z})$.

If $\gamma < 0$, then $e^{\gamma Z} = 0 \iff Z = \infty$, but $\mathbb{E}(e^{\gamma Z})$ is a mean (weighted average) of nonnegative numbers (write out the definition in the discrete case), so $\mathbb{P}(Z = \infty) = 1$ if and only if $\mathbb{E}(e^{\gamma Z}) = 0$. As $\gamma \uparrow 0$, we get $e^{-\gamma Z} \uparrow 1$ if $Z < \infty$ and $e^{-\gamma Z} = 0 \to 0$ if $Z = \infty$, so (by monotone convergence)

$$\mathbb{E}(e^{\gamma Z}) \uparrow \mathbb{E}(1_{\{Z<\infty\}}) = \mathbb{P}(Z < \infty)$$

and the result follows. □

Example 22 For independent $X_j \sim \text{Poi}(\mu_j)$ and $Z_m = X_1 + \ldots + X_m$, the random variable $Z = \lim_{m\to\infty} Z_m$ exists in $[0, \infty]$ a.s. Now

$$\mathbb{E}(e^{\gamma Z_m}) = \mathbb{E}((e^{\gamma})^{Z_m}) = e^{(e^{\gamma}-1)(\mu_1+\ldots+\mu_m)} \to e^{-(1-e^{\gamma})\mu}$$

shows that the limit is $\text{Poi}(\mu)$ if $\mu = \sum_{m\to\infty} \mu_m < \infty$. We do not need the lemma for this, since we can even directly identify the limiting moment generating function.

If $\mu = \infty$, the limit of the moment generating function vanishes, and by the lemma, we obtain $\mathbb{P}(Z = \infty) = 1$. So we still get $S \sim \text{Poi}(\mu)$ within the extended range $0 \le \mu \le \infty$.

4.2 Construction of spatial Poisson processes

The examples of compound Poisson processes are the key to constructing spatial Poisson processes with finite intensity measure. Infinite intensity measures can be decomposed.

Theorem 23 (Construction) *Let Λ be an intensity measure on $D \subset \mathbb{R}^d$ and suppose that there is a partition $(I_n)_{n\ge1}$ of D into regions with $\Lambda(I_n) < \infty$. Consider independently*

$$N_n \sim \text{Poi}(\Lambda(I_n)), \qquad Y_1^{(n)}, Y_2^{(n)}, \ldots \sim \frac{\Lambda(I_n \cap \cdot)}{\Lambda(I_n)}, \ \textit{i.e. } \mathbb{P}(Y_j^{(n)} \in A) = \frac{\Lambda(I_n \cap A)}{\Lambda(I_n)}$$

and define $\Pi_n = \{Y_j^{(n)} : 1 \le j \le N_n\}$. Then $\Pi = \bigcup_{n\ge1} \Pi_n$ is a spatial Poisson process with intensity measure Λ.

Proof: First fix n and show that Π_n is a spatial Poisson process on I_n

Thinning property of Poisson variables: Consider a sequence of independent Bernoulli(p) random variables $(B_j)_{j\ge1}$ and independent $X \sim \text{Poi}(\lambda)$. Then the following two random variables are independent:

$$X_1 = \sum_{j=1}^{X} B_j \sim \text{Poi}(p\lambda) \quad \text{and} \quad X_2 = \sum_{j=1}^{X} (1 - B_j) \sim \text{Poi}((1-p)\lambda).$$

To prove this, calculate the joint probability generating function

$$\begin{aligned}
\mathbb{E}(r^{X_1} s^{X_2}) &= \sum_{n=0}^{\infty} \mathbb{P}(X = n)\mathbb{E}(r^{B_1+\ldots+B_n} s^{n-B_1-\ldots-B_n}) \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} r^k s^{n-k} \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} (pr + (1-p)s)^n = e^{-\lambda p(1-r)} e^{-\lambda(1-p)(1-s)},
\end{aligned}$$

so the probability generating function factorises giving independence and we recognise the Poisson distributions as claimed.

For $A \subset I_n$, consider $X = N_n$ and the thinning mechanism, where $B_j = 1_{\{Y_j^{(n)} \in A\}} \sim$ Bernoulli($\mathbb{P}(Y_j^{(n)} \in A)$), then we get property (b):

$N_n(A) = X_1$ is Poisson distributed with parameter $\mathbb{P}(Y_j^{(n)} \in A)\Lambda(I_n) = \Lambda(A)$.

For property (a), disjoint sets $A_1, \ldots, A_m \subset I_n$, we apply the analogous thinning property for $m+1$ types $Y_j^{(n)} \in A_i$ $i = 0, \ldots, m$, where $A_0 = I_n \setminus (A_1 \cup \ldots \cup A_m)$ to deduce the independence of $N_n(A_1), \ldots, N_n(A_m)$. Thus, Π_n is a spatial Poisson process.

Now for $N(A) = \sum_{n\geq 1} N_n(A \cap I_n)$, we add up infinitely many Poisson variables and, by Example 22, obtain a $\text{Poi}(\mu)$ variable, where $\mu = \sum_{n\geq 1} \Lambda(A \cap I_n) = \Lambda(A)$, i.e. property (b). Property (a) also holds, since $N_n(A_j \cap I_n)$, $n \geq 1$, $j = 1, \ldots, m$, are all independent, and $N(A_1), \ldots, N(A_m)$ are independent as functions of independent random variables. □

4.3 Sums over Poisson point processes

Recall that a Poisson point process $(\Delta_t)_{t\geq 0}$ with intensity function $g : D_0 \to [0, \infty)$ – focus on $D_0 = (0, \infty)$ first but this can then be generalised – is a process such that

$$N((a,b] \times (c,d]) = \#\{a < t \leq b : \Delta_t \in (c,d]\} \sim \text{Poi}\left((b-a)\int_c^d g(x)dx\right),$$

$0 \leq a < b$, $(c,d] \subset D_0$, defines a Poisson counting measure on $D = [0,\infty) \times D_0$. This means that

$$\Pi = \{(t, \Delta_t) : t \geq 0 \text{ and } \Delta_t \neq 0\}$$

is a spatial Poisson process. Thinking of Δ_s as a jump size at time s, let us study $X_t = \sum_{0\leq s\leq t} \Delta_s$, the process performing all these jumps. Note that this is the situation for compound Poisson processes X; in Example 18, $g : (0,\infty) \to [0,\infty)$ is integrable.

Theorem 24 (Exponential formula) *Let $(\Delta_t)_{t\geq 0}$ be a Poisson point process with* locally *integrable intensity function $g : (0,\infty) \to [0,\infty)$. Then for all $\gamma \in \mathbb{R}$*

$$\mathbb{E}\left(\exp\left\{\gamma \sum_{0\leq s\leq t} \Delta_s\right\}\right) = \exp\left\{t \int_0^\infty (e^{\gamma x} - 1) g(x) dx\right\}.$$

Proof: Local integrability of g on $(0,\infty)$ means in particular that g is integrable on $I_n = (2^n, 2^{n+1}]$, $n \in \mathbb{Z}$. The properties of the associated Poisson counting measure N immediately imply that the random counting measures N_n counting all points in I_n, $n \in \mathbb{Z}$, defined by

$$N_n((a,b] \times (c,d]) = \{a < t \leq b : \Delta_t \in (c,d] \cap I_n\}, \qquad 0 \leq a < b, (c,d] \subset (0,\infty),$$

are independent. Furthermore, N_n is the Poisson counting measure of jumps of a compound Poisson process with $(b-a)\int_c^d g(x)dx = (b-a)\lambda_n \mathbb{P}(Y_1^{(n)} \in (c,d])$ for $0 \leq a < b$ and $(c,d] \subset I_n$ (cf. Example 18), so $\lambda_n = \int_{I_n} g(x)dx$ and (if $\lambda_n > 0$) jump density $h_n = \lambda_n^{-1} g$ on I_n, zero elsewhere. Therefore, we obtain

$$\mathbb{E}\left(\exp\left\{\gamma \sum_{0\leq s\leq t} \Delta_s^{(n)}\right\}\right) = \exp\left\{t \int_{I_n} (e^{\gamma x} - 1) g(x) dx\right\}, \text{ where } \Delta_s^{(n)} = \begin{cases} \Delta_s & \text{if } \Delta_s \in I_n \\ 0 & \text{otherwise} \end{cases}$$

Now we have

$$Z_m = \sum_{n=-m}^{m} \sum_{0\le s\le t} \Delta_s^{(n)} \uparrow \sum_{0\le s\le t} \Delta_s \qquad \text{as } m\to\infty,$$

and (cf. Lemma 21 about finite or infinite limits), the associated moment generating functions (products of individual moment generating functions) converge as required:

$$\prod_{n=-m}^{m} \exp\left\{ t\int_{2^n}^{2^{n+1}} (e^{\gamma x}-1)g(x)dx \right\} \to \exp\left\{ t\int_0^\infty (e^{\gamma x}-1)g(x)dx\right\}.$$

□

4.4 Martingales (from B10a)

A discrete-time stochastic process $(M_n)_{n\ge 0}$ in $\mathbb{R}$ is called a *martingale* if for all $n\ge 0$

$$\mathbb{E}(M_{n+1}|M_0,\ldots,M_n) = M_n, \quad \text{i.e. if } \mathbb{E}(M_{n+1}|M_0=x_0,\ldots,M_n=x_n) = x_n \text{ for all } x_j.$$

This is the principle of a fair game. What can I expect from the future if my current state is $M_n = x_n$? No gain and no loss, on average, whatever the past. The following important rules for conditional expectations are crucial to establish the martingale property

- If X and Y are independent, then $\mathbb{E}(X|Y) = \mathbb{E}(X)$.
- If $X = f(Y)$, then $\mathbb{E}(X|Y) = \mathbb{E}(f(Y)|Y) = f(Y)$ for functions $f:\mathbb{R}\to\mathbb{R}$ for which the conditional expectations exist.
- Conditional expectation is linear $\mathbb{E}(\alpha X_1 + X_2|Y) = \alpha\mathbb{E}(X_1|Y) + \mathbb{E}(X_2|Y)$.
- More generally: $\mathbb{E}(g(Y)X|Y) = g(Y)\mathbb{E}(X|Y)$ for functions $g:\mathbb{R}\to\mathbb{R}$ for which the conditional expectations exist.

These are all not hard to prove for discrete random variables. The full statements (continuous analogues) are harder. Martingales in continuous time can also be defined, but (formally) the conditioning needs to be placed on a more abstract footing. Denote by $\mathcal{F}_s$ the "information available up to time $s\ge 0$", for us just the process $(M_r)_{r\le s}$ up to time s – this is often written $\mathcal{F}_s = \sigma(M_r, r\le s)$. Then the four bullet point rules still hold for $Y = (M_r)_{r\le s}$ or for Y replaced by $\mathcal{F}_s$.

We call $(M_t)_{t\ge 0}$ a martingale if for all $s\le t$

$$\mathbb{E}(M_t|\mathcal{F}_s) = M_s.$$

Example 25 Let $(N_s)_{s\ge 0}$ be a Poisson process with rate λ. Then $M_s = N_s - \lambda s$ is a martingale: by the first three bullet points and by the Markov property (Proposition 3)

$$\mathbb{E}(N_t - \lambda t|\mathcal{F}_s) = \mathbb{E}(N_s + (N_t - N_s) - \lambda t|\mathcal{F}_s) = N_s + (t-s)\lambda - \lambda t = N_s - \lambda s.$$

Also $E_s = \exp\{\gamma N_s - \lambda s(e^\gamma - 1)\}$ is a martingale since by the first and last bullet points above, and by the Markov property

$$\begin{aligned}
\mathbb{E}(E_t|\mathcal{F}_s) &= \mathbb{E}(\exp\{\gamma N_s + \gamma(N_t - N_s) - \lambda t(e^\gamma - 1)\}|\mathcal{F}_s)\\
&= \exp\{\gamma N_s - \lambda t(e^\gamma - 1)\}\mathbb{E}(\exp\{\gamma(N_t - N_s)\})\\
&= \exp\{\gamma N_s - \lambda t(e^\gamma - 1)\}\exp\{-\lambda(t-s)(e^\gamma - 1)\} = E_s.
\end{aligned}$$

We will review relevant martingale theory when this becomes relevant.

Lecture 5

The characteristics of subordinators

Reading: Kingman Section 8.4

We have done the leg-work. We can now harvest the fruit of our efforts and proceed to a number of important consequences. Our programme for the next couple of lectures is:

- We construct Lévy processes from their jumps, first the most general increasing Lévy process. As linear combinations of independent Lévy processes are Lévy processes (Assignment A.1.2.(a)), we can then construct Lévy processes such as Variance Gamma processes of the form $Z_t = X_t - Y_t$ for two increasing X and Y.
- We have seen martingales associated with N_t and $\exp\{N_t\}$ for a Poisson process N. Similar martingales exist for all Lévy processes (cf. Assignment A.2.3.). Martingales are important for finance applications, since they are the basis of arbitrage-free models (more precisely, we need equivalent martingale measures, but we will assume here a "risk-free" measure directly to avoid technicalities).
- Our rather restrictive first range of examples of Lévy processes was obtained from known infinitely divisible distributions. We can now model using the intensity function of the Poisson point process of jumps to get a wider range of examples.
- We can simulate these Lévy processes, either by approximating random walks based on the increment distribution, or by constructing the associated Poisson point process of jumps, as we have seen, from a collection of independent random variables.

5.1 Subordinators and the Lévy-Khintchine formula

We will call (weakly) increasing Lévy processes "subordinators". Recall "$\nu(dx) \hat{=} g(x)dx$".

Theorem 26 (Construction) *Let $a \geq 0$, and let $(\Delta_t)_{t\geq 0}$ be a Poisson point process with intensity measure ν on $(0,\infty)$ such that*

$$\int_{(0,\infty)} (1 \wedge x)\nu(dx) < \infty,$$

then the process $X_t = at + \sum_{s\leq t} \Delta_s$ is a subordinator with moment generating function $\mathbb{E}(\exp\{\gamma X_t\}) = \exp\{t\Psi(\gamma)\}$, where

$$\Psi(\gamma) = a\gamma + \int_{(0,\infty)} (e^{\gamma x} - 1)\nu(dx).$$

Proof: Clearly $(at)_{t\geq 0}$ is a deterministic subordinator and we may assume $a=0$ in the sequel. Now the Exponential formula gives the moment generating function of $X_t = \sum_{s\leq t}\Delta_s$. We can now use Lemma 21 to check whether $X_t<\infty$ for $t>0$:

$$\mathbb{P}(X_t<\infty)=1 \iff \mathbb{E}\left(\exp\{\gamma X_t\}\right)=\exp\left\{t\int_0^\infty (e^{\gamma x}-1)\nu(dx)\right\}\to 1 \qquad \text{as } \gamma\uparrow 0.$$

This happens, by monotone convergence, if and only if for some (equivalently all) $\gamma<0$

$$\int_0^\infty (1-e^{\gamma x})\nu(dx)<\infty \iff \int_0^\infty (1\wedge x)\nu(dx)<\infty.$$

It remains to check that $(X_t)_{t\geq 0}$ is a Lévy process. Fix $0\leq t_0<t_1<\ldots<t_n$. Since $(\Delta_s)_{s\geq 0}$ is a Poisson point process, the processes $(\Delta_s)_{t_{j-1}\leq s<t_j}$, $j=1,\ldots,n$, are independent (consider the restrictions to disjoint domains $[t_{j-1},t_j)\times(0,\infty)$ of the Poisson counting measure

$$N((a,b]\times(c,d])=\{a\leq t<b:\Delta_t\in(c,d]\},\quad 0\leq a<b, 0<c<d),$$

and so are the sums $\sum_{t_{j-1}\leq s<t_j}\Delta_s$ as functions of independent random variables. Fix $s<t$. Then the process $(\Delta_{s+r})_{r\geq 0}$ has the same distribution as $(\Delta_s)_{s\geq 0}$. In particular, $\sum_{0\leq r\leq t}\Delta_{s+t}\sim\sum_{0\leq r\leq t}\Delta_r$. The process $t\mapsto\sum_{s\leq t}\Delta_s$ is right-continuous with left limits, since it is a random increasing function where for each jump time T, we have

$$\lim_{t\uparrow T}\sum_{s\leq t}\Delta_s=\lim_{t\uparrow T}\sum_{s<T}\Delta_s 1_{\{s\leq t\}}=\sum_{s<T}\Delta_s \quad\text{and}\quad \lim_{t\downarrow T}\sum_{s\leq t}\Delta_s=\lim_{t\downarrow T}\sum_{s\leq T+1}\Delta_s 1_{\{s\leq t\}}=\sum_{s\leq T}\Delta_s,$$

by monotone convergence, because each of the terms $\Delta_s 1_{\{s\leq t\}}$ in the sums converges. □

Note also that, due to the Exponential formula, $\mathbb{P}(X_t<\infty)>0$ already implies $\mathbb{P}(X_t<\infty)=1$. We shall now state but not prove the Lévy-Khintchine formula for nonnegative random variables.

Theorem 27 (Lévy-Khintchine) *A nonnegative random variable Y has an infinitely divisible distribution if and only if there is a pair (a,ν) such that for all $\gamma\leq 0$*

$$\mathbb{E}(\exp\{\gamma Y\})=\exp\left\{a\gamma+\int_{(0,\infty)}(e^{\gamma x}-1)\nu(dx)\right\},\tag{1}$$

where $a\geq 0$ and ν is such that $\int_{(0,\infty)}(1\wedge x)\nu(dx)<\infty$.

Corollary 28 *Given a nonnegative random variable Y with infinitely divisible distribution, there exists a subordinator $(X_t)_{t\geq 0}$ with $X_1\sim Y$.*

Proof: Let Y have an infinitely divisible distribution. By the Lévy-Khintchine theorem, its moment generating function is of the form (1) for parameters (a,ν). Theorem 26 constructs a subordinator $(X_t)_{t\geq 0}$ with $X_1\sim Y$. □

This means that the class of subordinators can be parameterised by two parameters, the nonnegative "drift parameter" $a\geq 0$, and the "Lévy measure" ν, or its density, the "Lévy density" $g:(0,\infty)\to[0,\infty)$. The parameters (a,ν) are referred to as the "Lévy-Khintchine characteristics" of the subordinator (or of the infinitely divisible distribution). Using the Uniqueness theorem for moment generating functions, it can be shown that a and ν are unique, i.e. that no two sets of characteristics refer to the same distribution.

5.2 Examples

Example 29 (Gamma process) The Gamma process, where $X_t \sim \text{Gamma}(\alpha t, \beta)$, is an increasing Lévy process. In Assignment A.2.4. we showed that

$$\mathbb{E}\left(\exp\{\gamma X_t\}\right) = \left(\frac{\beta}{\beta-\gamma}\right)^{\alpha t} = \exp\left\{t\int_0^\infty (e^{\gamma x}-1)\alpha x^{-1}e^{-\beta x}dx\right\}, \quad \gamma < \beta.$$

We read off the characteristics $a = 0$ and $g(x) = \alpha x^{-1}e^{-\beta x}$, $x > 0$.

Example 30 (Poisson process) The Poisson process, where $X_t \sim \text{Poi}(\lambda t)$, has

$$\mathbb{E}\left(\exp\{\gamma X_t\}\right) = \exp\left\{t\lambda(e^\gamma - 1)\right\}$$

This corresponds to characteristics $a = 0$ and $\nu = \lambda\delta_1$, where δ_1 is the discrete unit point mass in (jump size) 1.

Example 31 (Increasing compound Poisson process) The compound Poisson process $C_t = Y_1 + \ldots + Y_{X_t}$, for a Poisson process X and independent identically distributed *nonnegative* $Y_1, Y_2, \ldots$ with probability density function $h(x)$, $x > 0$, satisfies

$$\mathbb{E}\left(\exp\{\gamma C_t\}\right) = \exp\left\{t\int_0^\infty (e^{\gamma x}-1)\lambda h(x)dx\right\},$$

and we read off characteristics $a = 0$ and $g(x) = \lambda h(x)$, $x > 0$. We can add a drift and consider $\widetilde{C}_t = \widetilde{a}t + C_t$ for some $\widetilde{a} > 0$ to get a *compound Poisson process with drift*.

Example 32 (Stable subordinator) The stable subordinator is best defined in terms of its Lévy-Khintchine characteristics $a = 0$ and $g(x) = x^{-\alpha-1}$. This gives for $\gamma \le 0$

$$\mathbb{E}\left(\exp\{\gamma X_t\}\right) = \exp\left\{t\int_0^\infty (e^{\gamma x}-1)x^{-\alpha-1}dx\right\} = \exp\left\{t\frac{\Gamma(1-\alpha)}{\alpha}(-\gamma)^\alpha\right\}.$$

Note that $\mathbb{E}(\exp\{\gamma c^{1/\alpha}X_{t/c}\}) = \mathbb{E}(\exp\{\gamma X_t\})$, so that $(c^{1/\alpha}X_{t/c})_{t\ge 0} \sim X$. More generally, we can also consider e.g. tempered stable processes with $g(x) = x^{-\alpha-1}\exp\{-\rho x\}$, $\rho > 0$.

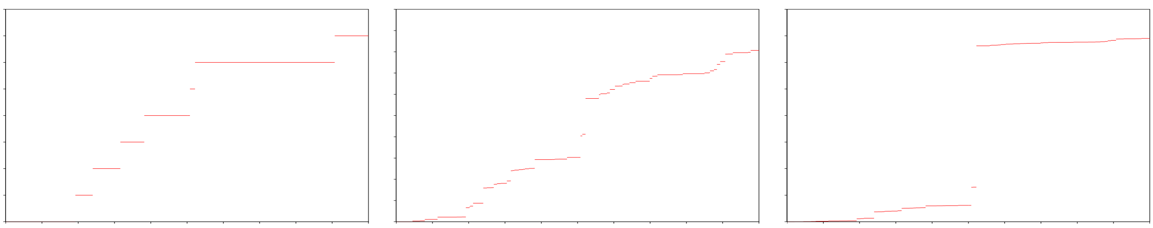


Figure 5.1: Examples: Poisson process, Gamma process, stable subordinator

5.3 Aside: nonnegative Lévy processes

It may seem obvious that a nonnegative Lévy process, i.e. one where $X_t \ge 0$ a.s. for all $t \ge 0$, is automatically increasing, since every increment $X_{s+t} - X_s$ has the same distribution X_t and is hence also nonnegative. Let us be careful, however, and remember that there is a difference between something never happening at a fixed time and something never happening at any time. We have e.g. for a (one-dimensional) Poisson process $(N_t)_{t\ge 0}$

$$\mathbb{P}(\Delta N_t \ne 0) = \sum_{n\ge 1}\mathbb{P}(T_n = t) = 0 \qquad \text{for all } t \ge 0, \text{ but} \qquad \mathbb{P}(\exists t : \Delta N_t \ne 0) = 1.$$

Here we can argue that if $f(t) < f(s)$ for some $s < t$ and a right-continuous function, then there are also two rational numbers $s_0 < t_0$ for which $f(t_0) < f(s_0)$, so

$$\mathbb{P}(\exists s, t \in (0,\infty), s < t : X_t - X_s < 0) > 0 \Rightarrow \mathbb{P}(\exists s_0, t_0 \in (0,\infty) \cap \mathbb{Q} : X_{t_0} - X_{s_0} < 0) > 0$$

However, the latter can be bounded above (by subadditivity $\mathbb{P}(\bigcup_n A_n) \le \sum_n \mathbb{P}(A_n)$)

$$\mathbb{P}(\exists s_0, t_0 \in (0,\infty) \cap \mathbb{Q} : X_{t_0} - X_{s_0} < 0) \le \sum_{s_0, t_0 \in (0,\infty)\cap\mathbb{Q}} \mathbb{P}(X_{t_0 - s_0} < 0) = 0.$$

Another instance of such delicate argument is the following: if $X_t \ge 0$ a.s. for one $t > 0$ and a subordinator X, then $X_t \ge 0$ a.s. for all $t \ge 0$. It is true, but to say if $\mathbb{P}(X_s < 0) > 0$ for some $s < t$ then $\mathbb{P}(X_t < 0) > 0$ may not be all that obvious. It is, however, easily justified for $s = t/m$, since then $\mathbb{P}(X_t < 0) \ge \mathbb{P}(X_{tj/m} - X_{t(j-1)/m} < 0$ for all $j = 1, \ldots, m) > 0$. We have to apply a similar argument to get $\mathbb{P}(X_{tq} < 0) = 0$ for all rational $q > 0$. Then we use again right-continuity to see that a function that is nonnegative at all rationals cannot take a negative value at an irrational either, so we get

$$\mathbb{P}(\exists s \in [0,\infty) : X_s < 0) = \mathbb{P}(\exists s \in [0,\infty) \cap \mathbb{Q} : X_s < 0) \le \sum_{s \in [0,\infty)\cap\mathbb{Q}} \mathbb{P}(X_s < 0) = 0.$$

5.4 Applications

Subordinators have found a huge range of applications, but are not directly models for a lot of real world phenomena. We can now construct more general Lévy processes of the form $Z_t = X_t - Y_t$ for two subordinators X and Y. Let us here indicate some subordinators as they are used/arise in connection with other Lévy processes.

Example 33 (Subordination) For a Lévy process X and an independent subordinator T, the process $Y_s = X_{T_s}$, $s \ge 0$, is also a Lévy process (we study this later in the course). The rough argument is that $(X_{T_s+u} - X_{T_s})_{u\ge 0}$ is independent of $(X_r)_{r \le T_s}$ and distributed as X, by the Markov property. Hence $X_{T_{s+r}} - X_{T_s}$ is independent of X_{T_s} and distributed as X_{T_r}. A rigorous argument can be based on calculations of joint moment generating functions. Hence, subordinators are a useful tool to construct Lévy processes, e.g. from Brownian motion X. Many models of financial markets are of this type. The operation $Y_s = X_{T_s}$ is called subordination – this is where subordinators got their name from.

Example 34 (Level passage) Let $Z_t = at - X_t$ where $a = \mathbb{E}(X_1)$. It can be shown that $\tau_s = \inf\{t \ge 0 : Z_t > s\} < \infty$ a.s. for all $s \ge 0$ (from the analogous random walk result). It turns out (cf. later in the course) that $(\tau_s)_{s \ge 0}$ is a subordinator.

Example 35 (Level set) Look at the zero set $\mathcal{Z} = \{t \ge 0 : B_t = 0\}$ for Brownian motion (or indeed any other centred Lévy process) B. $\mathcal{Z}$ is unbounded since B crosses zero at arbitrarily large times so as to pass beyond all s and $-s$. Recall that $(tB_{1/t})_{t\ge 0}$ is also a Brownian motion. Therefore, $\mathcal{Z}$ also has an accumulation point at $t = 0$, i.e. crosses zero infinitely often at arbitrarily small times. In fact, it can be shown that $\mathcal{Z}$ is the closed range $\{X_r, r \ge 0\}^{\mathrm{cl}}$ of a subordinator $(X_r)_{r \ge 0}$. The Brownian scaling property $(\sqrt{c}B_{t/c})_{t\ge 0} \sim B$ shows that $\{X_r/c, r \ge 0\}^{\mathrm{cl}} \sim \mathcal{Z}$, and so X must have a scaling property. In fact, X is a stable subordinator of index $1/2$. Similar results, with different subordinators, hold not just for all Lévy processes but even for most Markov processes.

Lecture 6

Lévy processes with no negative jumps

Reading: Kyprianou 2.1, 2.4, 2.6, Schoutens 2.2; Further reading: Williams 10-11

Subordinators X are processes with no negative jumps. We get processes that can decrease by adding a negative drift at for $a < 0$. Also, Brownian motion B has no negative jumps. A guess might be that $X_t + at + \sigma B_t$ is the most general Lévy process with no negative jumps, but this is false. It turns out that even a non-summable amount of positive jumps can be incorporated, but we will have to look at this carefully.

6.1 Bounded and unbounded variation

The (total) variation of a right-continuous function $f : [0, t] \to \mathbb{R}$ with left limits is

$$||f||_{\mathrm{TV}} := \sup\left\{\sum_{j=1}^{n} |f(t_j) - f(t_{j-1})| : 0 = t_0 < t_1 < \ldots < t_n = t, n \in \mathbb{N}\right\}.$$

Clearly, for an increasing function with $f(0) = 0$ this is just $f(t)$ and for a difference $f = g-h$ of two increasing functions with $g(0) = h(0) = 0$ this is at most $g(t)+h(t) < \infty$, so all differences of increasing functions are of bounded variation. There are, however, functions of infinite variation, e.g. Brownian paths: they have finite quadradic variation

$$\sum_{j=1}^{2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}|^2 \to t \qquad \text{in the } L^2 \text{ sense}$$

since

$$\mathbb{E}\left(\sum_{j=1}^{2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}|^2\right) = 2^n \mathbb{E}(B_{t2^{-n}}^2) = t$$

and

$$\begin{aligned}\mathbb{E}\left(\left(\sum_{j=1}^{2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}|^2 - t\right)^2\right) &= \mathrm{Var}\left(\sum_{j=1}^{2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}|^2\right) \\ &\leq 2^n(2^{-n}t)^2\mathrm{Var}(B_1^2) \to 0,\end{aligned}$$

but then assuming finite total variation with positive probability, the uniform continuity of the Brownian path implies

$$\sum_{j=1}^{2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}|^2 \le \left(\sup_{j=1,\dots,2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}| \right) \sum_{j=1}^{2^n} |B_{tj2^{-n}} - B_{t(j-1)2^{-n}}| \to 0$$

with positive probability, but this is incompatible with convergence to t, so the assumption of finite total variation must have been wrong.

Here is how jumps influence total variation:

Proposition 36 *Let f be a right-continuous function with left limits and jumps $(\Delta f_s)_{0\le s\ge t}$. Then*

$$||f||_{\mathrm{TV}} \ge \sum_{0\le s\le t} |\Delta f_s|$$

Proof: Enumerate the jumps in decreasing order of size by $(T_n, \Delta f_{T_n})_{n\ge 0}$. Fix $N \in \mathbb{N}$ and $\delta > 0$. Choose $\varepsilon > 0$ so small that $\bigcup [T_n - \varepsilon, T_n]$ is a disjoint union and such that $|f(T_n - \varepsilon) - f(T_n -)| < \delta/N$. Then for $\{T_n - \varepsilon, T_n : n = 1, \dots, N\} = \{t_1, \dots, t_{2N+1}\}$ such that $0 = t_0 < t_1 < \dots < t_{2N+1} < t_{2N+2} = t$, we have

$$\sum_{j=1}^{2N+2} |f(t_j) - f(t_{j-1})| \ge \sum_{n=1}^{N} \Delta f(T_n) - \delta.$$

Since N and δ were arbitrary, this completes the proof, whether the right-hand side is finite or infinite. □

6.2 Martingales (from B10a)

Three martingale theorems are of central importance. We will require in this lecture just the maximal inequality, but we formulate all three here for easier reference. They all come in several different forms. We present the L^2-versions as they are most easily formulated and will suffice for us.

A stopping time is a random time T such that for every $s \ge 0$ the information $\mathcal{F}_s$ allows to decide whether $T \le s$. More formally, if the event $\{T \le s\}$ can be expressed in terms of $(M_r, r \le s)$ (is measurable with respect to $\mathcal{F}_s$). The prime example of a stopping time is the first entrance time $T_A = \inf\{t \ge 0 : M_t \in A\}$. Note that

$$\{T \le s\} = \{M_r \notin A \text{ for all } r \le s\}$$

(and at least for closed sets A we can drop the irrational $r \le s$ and see measurability, then approximate open sets.)

Theorem 37 (Optional stopping) *Let $(M_t)_{t\ge 0}$ be a martingale and T a stopping time. If $\sup_{t\ge 0} \mathbb{E}(M_t^2) < \infty$, then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.*

Theorem 38 (Convergence) *Let $(M_t)_{t\ge 0}$ be a martingale such that $\sup_{t\ge 0} \mathbb{E}(M_t^2) < \infty$, then $M_t \to M_\infty$ almost surely.*

Theorem 39 (Maximal inequality) *Let $(M_t)_{t\ge 0}$ be a martingale. Then $\mathbb{E}(\sup\{M_s^2 : 0 \le s \le t\}) \le 4\mathbb{E}(M_t^2)$.*

6.3 Compensation

Let $g : (0, \infty) \to [0, \infty)$ be the intensity function of a Poisson point process $(\Delta_t)_{t\geq 0}$. If g is not integrable at infinity, then $\#\{0 \leq s \leq t : \Delta_s > 1\} \sim \mathrm{Poi}(\int_1^\infty g(x)dx) = \mathrm{Poi}(\infty)$, and it is impossible for a right-continuous function with left limits to have accumulation points in the set of such jumps (lower and upper points of a sequence of jumps will then have different limit points). If however g is not integrable at zero, we have to investigate this further.

Proposition 40 *Let $(\Delta_t)_{t\geq 0}$ be a Poisson point process with intensity measure ν on $(0, \infty)$.*

(i) If $\int_0^\infty x\nu(dx) < \infty$, then

$$\mathbb{E}\left(\sum_{s\leq t} \Delta_s\right) = t\int_0^\infty x\nu(dx).$$

(ii) If $\int_0^\infty x^2\nu(dx) < \infty$, then

$$\mathrm{Var}\left(\sum_{s\leq t} \Delta_s\right) = t\int_0^\infty x^2\nu(dx).$$

Proof: These are the two leading terms in the expansion with respect to γ of the Exponential formula: the first moment can always be obtained from the moment generating function by taking $\frac{\partial}{\partial\gamma}|_{\gamma=0}$, here

$$\frac{\partial}{\partial\gamma}\exp\left\{t\int_0^\infty (e^{\gamma x} - 1)\nu(dx)\right\}\bigg|_{\gamma=0} = t\int_0^\infty xe^{\gamma x}\nu(dx)\bigg|_{\gamma=0} = t\int_0^\infty x\nu(dx),$$

and the second moment follows from the second derivative in the same way. □

Consider compound Poisson processes, with a drift that turns them into martingales

$$Z_t^\varepsilon = \sum_{s\leq t} \Delta_s 1_{\{\varepsilon<\Delta_s\leq 1\}} - t\int_\varepsilon^1 x\nu(dx) \tag{1}$$

We have deliberately excluded jumps in $(1, \infty)$. These are easier to handle separately. What integrability condition on ν do we need for Z_t^ε to converge as $\varepsilon \downarrow 0$?

Lemma 41 *Let $(\Delta_t)_{t\geq 0}$ be a Poisson point process with intensity measure ν on $(0, 1)$. With Z^ε defined in (1), Z_t^ε converges in L^2 if $\int_0^1 x^2\nu(dx) < \infty$.*

Proof: We only do this for $\nu(dx) = g(x)dx$. Note that for $0 < \delta < \varepsilon < 1$, by Proposition 40(ii) applied to $g_{\delta,\varepsilon}(x) = g(x)1_{\{\delta\leq x<\varepsilon\}}$,

$$\mathbb{E}(|Z_t^\varepsilon - Z_t^\delta|^2) = t\int_\delta^\varepsilon x^2 g(x)dx$$

so that $(Z_t^\varepsilon)_{0<\varepsilon<1}$ is a Cauchy family as $\varepsilon \downarrow 0$, for the L^2-distance $d(X, Y) = \sqrt{\mathbb{E}((X - Y)^2)}$. By completeness of L^2-space, there is a limiting random variable Z_t as required. □

We can slightly tune this argument to establish a general existence theorem:

Theorem 42 (Existence) *There exists a Lévy process whose jumps form a Poisson point process with intensity measure ν on $(0, \infty)$ if and only if $\int_{(0,\infty)}(1 \wedge x^2)\nu(dx) < \infty$.*

Proof: The "only if" statement is a consequence of a Lévy-Khintchine type characterisation of infinitely divisible distributions on $\mathbb{R}$, cf. Theorem 44, which we will not prove. Let us prove the "if" part in the case where $\nu(dx) = g(x)dx$.

By Proposition 40(i), $\mathbb{E}(Z_t^\varepsilon - Z_t^\delta) = 0$. By Assignment A.2.3.(c), the procecss $Z_t^\varepsilon - Z_t^\delta$ is a martingale, and the maximal inequality (Theorem 39) shows that

$$\mathbb{E}\left(\sup_{0\le s\le t}\left|Z_t^\varepsilon - Z_t^\delta\right|\right) \le 4\mathbb{E}(|Z_t^\varepsilon - Z_t^\delta|^2) = 4t\int_\delta^\varepsilon x^2 g(x)dx$$

so that $(Z_s^\varepsilon, 0 \le s \le t)_{0<\varepsilon<1}$ is a Cauchy family as $\varepsilon \downarrow 0$, for the uniform L^2-distance $d_{[0,t]}(X,Y) = \sqrt{\mathbb{E}(\sup_{0\le s\le t}|X_s - Y_s|^2)}$. By completeness of L^2-space, there is a limiting process $(Z_s^{(1)})_{0\le s\le t}$, which as the uniform limit (in L^2) of $(Z_s^\varepsilon)_{0\le s\le t}$ is right-continuous with left limits. Also consider the independent compound Poisson process

$$Z_t^{(2)} = \sum_{s\le t}\Delta_s 1_{\{\Delta_s>1\}} \qquad \text{and set} \qquad Z = Z^{(1)} + Z^{(2)}.$$

It is not hard to show that Z is a Lévy process that incorporates all jumps $(\Delta_s)_{0\le s\le t}$. □

Example 43 Let us look at a Lévy density $g(x) = |x|^{-5/2}$, $x \in [-3,0)$. Then the compensating drifts $\int_\varepsilon^3 xg(x)dx$ take values 0.845, 2.496, 5.170 and 18.845 for $\varepsilon = 1$, $\varepsilon = 0.3$, $\varepsilon = 0.1$ and $\varepsilon = 0.01$. In the simulation, you see that the slope increases (to infinity, actually as $\varepsilon \downarrow 0$), but the picture begins to stabilise and converge to a limit.

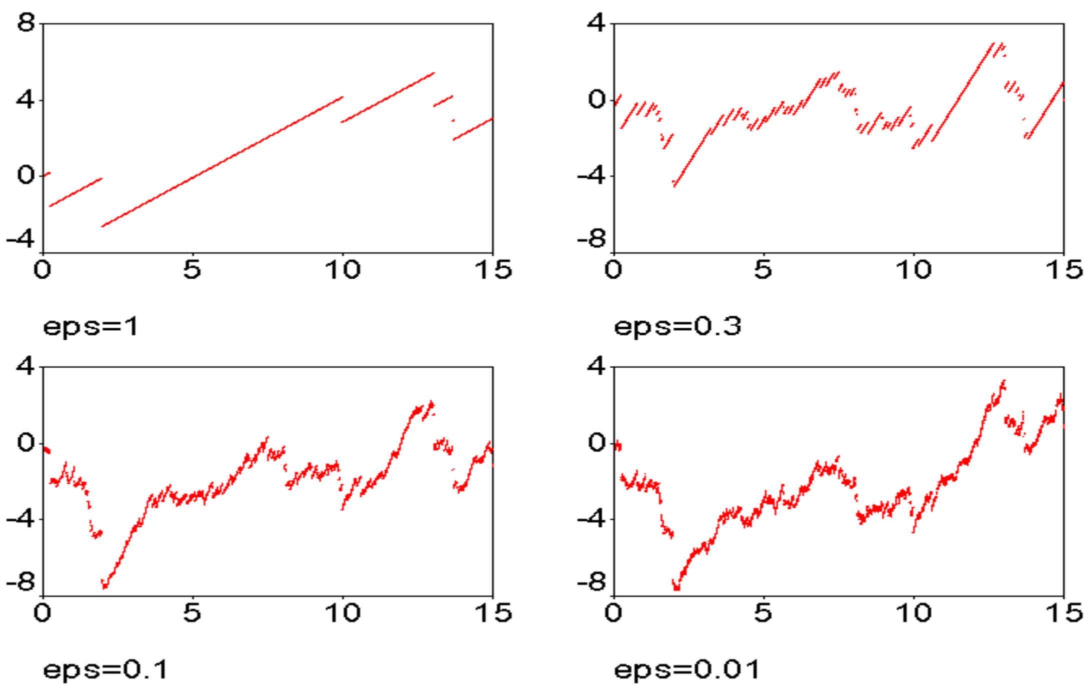


Figure 6.1: Approximation of a Lévy process with no positive jumps – compensating drift

Lecture 7

General Lévy processes and simulation

Reading: Schoutens Sections 8.1, 8.2, 8.4

For processes with no negative jumps, we compensated jumps by a linear drift and incorporated more and more smaller jumps while letting the slope of the linear drift tend to negative infinity. We will now construct the most general real-valued Lévy process as the difference of two such processes (and a Brownian motion). For explicit marginal distributions, we can simulate Lévy processes by approximating random walks. In practice, we often only have explicit characteristics (drift coefficient, Brownian coefficient and Lévy measure). We will also simulate Lévy processes based on the characteristics.

7.1 Construction of Lévy processes

The analogue of Theorem 27 for real-valued random variables is as follows.

Theorem 44 (Lévy-Khintchine) *A real-valued random variable X has an infinitely divisible distribution if there are parameters $a \in \mathbb{R}$, $\sigma^2 \geq 0$ and a measure ν on $\mathbb{R} \setminus \{0\}$ with $\int_{-\infty}^{\infty}(1 \wedge x^2)\nu(dx) < \infty$ such that $\mathbb{E}(e^{i\lambda X}) = e^{-\psi(\lambda)}$, where*

$$\psi(\lambda) = -ia\lambda + \frac{1}{2}\sigma^2\lambda^2 - \int_{-\infty}^{\infty}(e^{i\lambda x} - 1 - i\lambda x 1_{\{|x|\leq 1\}})\nu(dx), \qquad \lambda \in \mathbb{R}.$$

Lévy processes are parameterised by their *Lévy-Khintchine characteristics* (a, σ^2, ν), where we call a the *drift coefficient*, σ^2 the *Brownian coefficient* and ν the *Lévy measure* or *jump measure*. $\nu(dx)$ will often be of the form $g(x)dx$, and we then refer to g as the *Lévy density* or *jump density.*

Theorem 45 (Existence) *Let (a, σ^2, ν) be Lévy-Khintchine characteristics, $(B_t)_{t\geq 0}$ a standard Brownian motion and $(\Delta_t)_{t\geq 0}$ an independent Poisson point process of jumps with intensity measure ν. Then there is a Lévy process*

$$Z_t = at + \sigma B_t + M_t + C_t, \qquad \text{where } C_t = \sum_{s\leq t} \Delta_s 1_{\{|\Delta_s|>1\}},$$

is a compound Poisson process (of big jumps) and

$$M_t = \lim_{\varepsilon\downarrow 0}\left(\sum_{s\le t}\Delta_s 1_{\{\varepsilon<|\Delta_s|\le 1\}} - t\int_{\{x\in\mathbb{R}:\varepsilon<|x|\le 1\}} x\nu(dx)\right)$$

is a martingale (of small jumps – compensated by a linear drift).

Proof: The construction of $M_t = P_t - N_t$ can be made from two independent processes P_t and N_t with no negative jumps as in Theorem 42. N_t will be built from a Poisson point process with intensity measure $\overline{\nu}((c,d]) = \nu([-d,-c))$, $0<c<d\le 1$ (or $\overline{g}(y) = g(-y)$, $0<y<1$).

We check that the characteristic function of $Z_t = at + \sigma B_t + P_t - N_t + C_t$ is of Lévy-Khintchine type with parameters (a,σ,ν). We have five independent components. Evaluate at $t=1$ to get

$$\begin{aligned}
\mathbb{E}(e^{\gamma a}) &= e^{\gamma a}\\
\mathbb{E}(e^{\gamma\sigma B_1}) &= \exp\{\frac{1}{2}\gamma^2\sigma^2\}\\
\mathbb{E}(e^{\gamma P_1}) &= \exp\left\{\int_0^1 (e^{\gamma x}-1-\gamma x)\nu(dx)\right\}\\
\mathbb{E}(e^{-\gamma N_1}) &= \exp\left\{\int_0^1 (e^{-\gamma y}-1+\gamma y)\overline{\nu}(dy)\right\} = \exp\left\{\int_{-1}^0 (e^{\gamma x}-1-\gamma x)\nu(dx)\right\}\\
\mathbb{E}(e^{i\lambda C_1}) &= \exp\left\{\int_{|x|>1} (e^{i\lambda x}-1)\nu(dx)\right\}.
\end{aligned}$$

The last formula is checked in analogy with the moment generating function computation of Assignment A.1.3 (in general, the moment generating function will not be well-defined for this component). For the others, now "replace" γ by $i\lambda$. A formal justification can be obtained by analytic continuation, since the moment generating functions of these components are entire functions of γ as a complex parameter. Now the characteristic function of Z_1 is the product of characteristic functions of the independent components, and this yields the formula required. □

We stress in particular, that every Lévy process is the difference of two processes with only positive jumps. In general, these processes are not subordinators, but of the form in Theorem 42 plus a Brownian motion component. They can then both take positive and negative values.

Example 46 (Variance Gamma process) We introduced the Variance Gamma process as difference $X = G - H$ of two independent Gamma subordinators G and H. We can generalise the setting of Exercise A.1.2.(b) and allow $G_1 \sim \text{Gamma}(\alpha_+,\beta_+)$ and $H_1 \sim \text{Gamma}(\alpha_-,\beta_-)$. The moment generating function of the Variance Gamma process is

$$\begin{aligned}
\mathbb{E}(e^{\gamma X_t}) &= \mathbb{E}(e^{\gamma G_t})\mathbb{E}(e^{-\gamma H_t}) = \left(\frac{\beta_+}{\beta_+-\gamma}\right)^{\alpha_+ t}\left(\frac{\beta_-}{\beta_-+\gamma}\right)^{\alpha_- t}\\
&= \exp\left\{t\int_0^\infty (e^{\gamma x}-1)\alpha_+ x^{-1}e^{-\beta_+ x}dx\right\}\exp\left\{t\int_0^\infty (e^{-\gamma y}-1)\alpha_- y^{-1}e^{-\beta_- y}dy\right\}\\
&= \exp\left\{t\int_0^\infty (e^{\gamma x}-1)\alpha_+|x|^{-1}e^{-\beta_+|x|}dx + t\int_{-\infty}^0 (e^{\gamma x}-1)\alpha_-|x|^{-1}e^{-\beta_-|x|}dx\right\}.
\end{aligned}$$

and this is in Lévy-Khintchine form with $\nu(dx) = g(x)dx$ with

$$g(x) = \begin{cases} \alpha_+|x|^{-1}e^{-\beta_+|x|} & x > 0 \\ \alpha_-|x|^{-1}e^{-\beta_-|x|} & x < 0 \end{cases}$$

The process $(\Delta X_t)_{t\geq 0}$ is a Poisson point process with intensity function g.

Example 47 (CGMY process) Theorem 45 encourages to specify Lévy processes by their characteristics. As a natural generalisation of the Variance Gamma process, Carr, Geman, Madan and Yor (CGMY) suggested the following for financial price processes

$$g(x) = \begin{cases} C_+ \exp\{-G|x|\}|x|^{-Y-1} & x > 0 \\ C_- \exp\{-M|x|\}|x|^{-Y-1} & x < 0 \end{cases}$$

for parameters $C_\pm > 0$, $G > 0$, $M > 0$, $Y \in [0, 2)$. While the Lévy density is a nice function, the probability density function of an associated Lévy process X_t is not available in closed form, in general. The CGMY model contains the Gamma model for $Y = 0$. When this model is fitted to financial data, there is usually significant evidence against $Y = 0$, so the CGMY model is more appropriate than the Variance Gamma model.

We can construct Lévy processes from their Lévy density and will also simulate from Lévy densities. Note that this way of modelling is easier than searching directly for infinitely divisible probability density functions.

7.2 Simulation via embedded random walks

"Simulation" usually refers to the realisation of a random variable using a computer. Most mathematical and statistical packages provide functions, procedures or commands for the generation of sequences of pseudo-random numbers that, while not random, show features of independent and identically distributed random variables that are adequate for most purposes. We will not go into the details of the generation of such sequences, but assume that we a sequence $(U_k)_{k\geq 1}$ of independent Unif$(0, 1)$ random variables.

If the increment distribution is explicitly known, we simulate via time discretisation.

Method 1 (Time discretisation) Let $(X_t)_{t\geq 0}$ be a Lévy process so that X_t has probability density function f_t. Fix a time lag $\delta > 0$. Denote $F_t(x) = \int_{-\infty}^x f_t(y)dy$ and $F_t^{-1}(u) = \inf\{x \in \mathbb{R} : F_t(x) > u\}$. Then the process

$$X_t^{(1,\delta)} = S_{[t/\delta]}, \qquad \text{where } S_n = \sum_{k=1}^n Y_k \text{ and } Y_k = F_\delta^{-1}(U_k),$$

is called the *time discretisation* of X with time lag δ.

One usually requires numerical approximation for F_t^{-1}, even if f_t is available in closed form. That the approximations converge, is shown in the following proposition.

Proposition 48 *As $\delta \downarrow 0$, we have $X_t^{(1,\delta)} \to X_t$ in distribution.*

Proof: We can employ a coupling proof: t is a.s. not a jump time of X, so we have $X_{[t/\delta]\delta} \to X_t$ a.s., and so convergence in distribution for $X_t^{(1,\delta)} \sim X_{[t/\delta]\delta}$. □

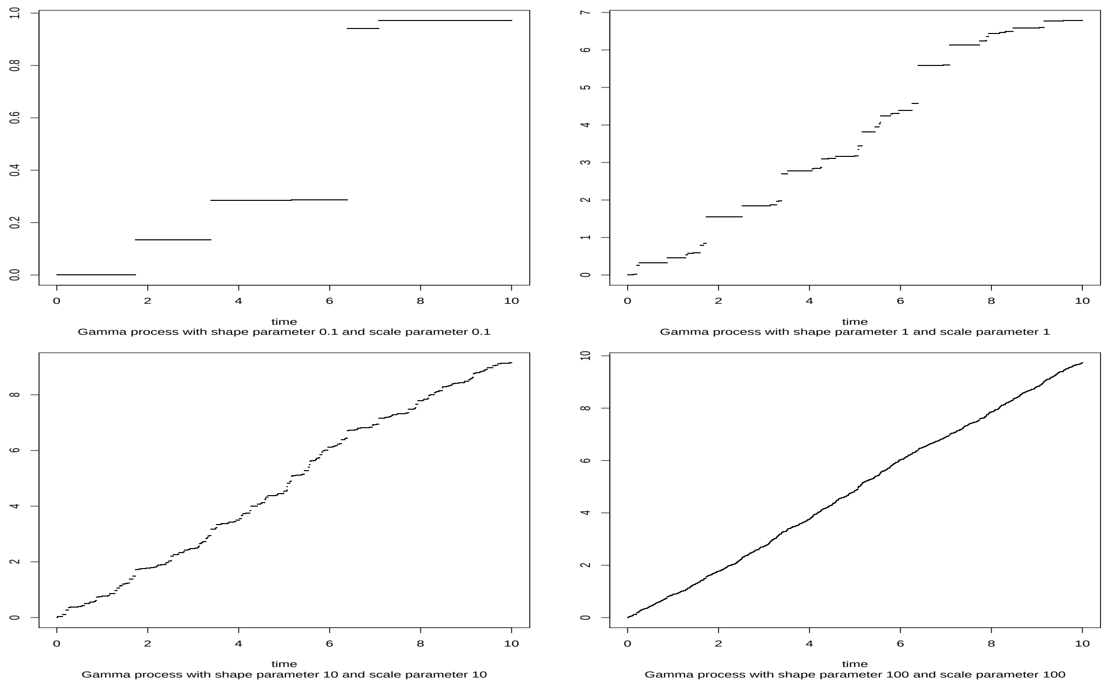


Figure 7.1: Simulation of Gamma processes from random walks with Gamma increments

Example 49 (Gamma processes) For Gamma processes, F_t is an incomplete Gamma function, which has no closed-form expression, and F_t^{-1} is also not explicit, but numerical evaluations have been implemented in many statistical packages. There are also Gamma generators based on more uniform random variables. We display a range of parameter choices. Since for a Gamma$(1,1)$ process X, the process $(\beta^{-1}X_{\alpha t})_{t\geq 0}$ is Gamma(α,β):

$$\mathbb{E}(\exp\{\gamma\beta^{-1}X_{\alpha t}\}) = \left(\frac{1}{1-\gamma\beta^{-1}}\right)^{\alpha t} = \left(\frac{\beta}{\beta-\gamma}\right)^{\alpha t},$$

we chose $\alpha=\beta$ (keeping mean 1 and comparable spatial scale) but a range of parameters $\alpha\in\{0.1,1,10,100\}$ on a fixed time interval $[0,10]$. We "see" convergence to a linear drift as $\alpha\to\infty$ (for fixed t this is due to the laws of large numbers).

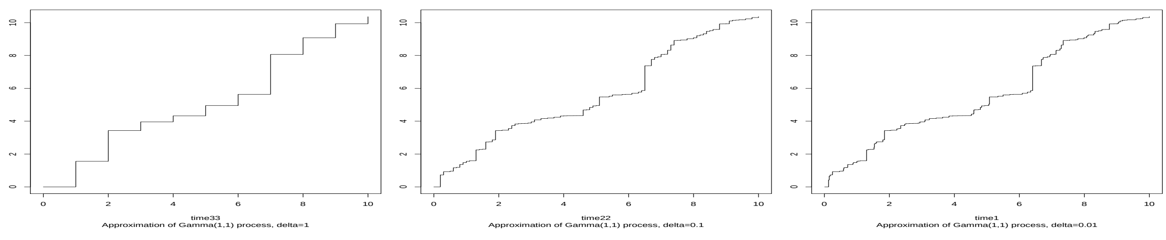


Figure 7.2: Random walk approximation to a Lévy process, as in Proposition 48

Example 50 (Variance Gamma processes) We represent the Variance Gamma process as the difference of two independent Gamma processes and focus on the symmetric case, so achieve mean 0 and fix variance 1 by putting $\beta = \alpha^2/2$; we consider $\alpha \in \{1, 10, 100, 1000\}$. We "see" convergence to Brownian motion as $\alpha \to \infty$ (for fixed t due to the Central Limit Theorem).

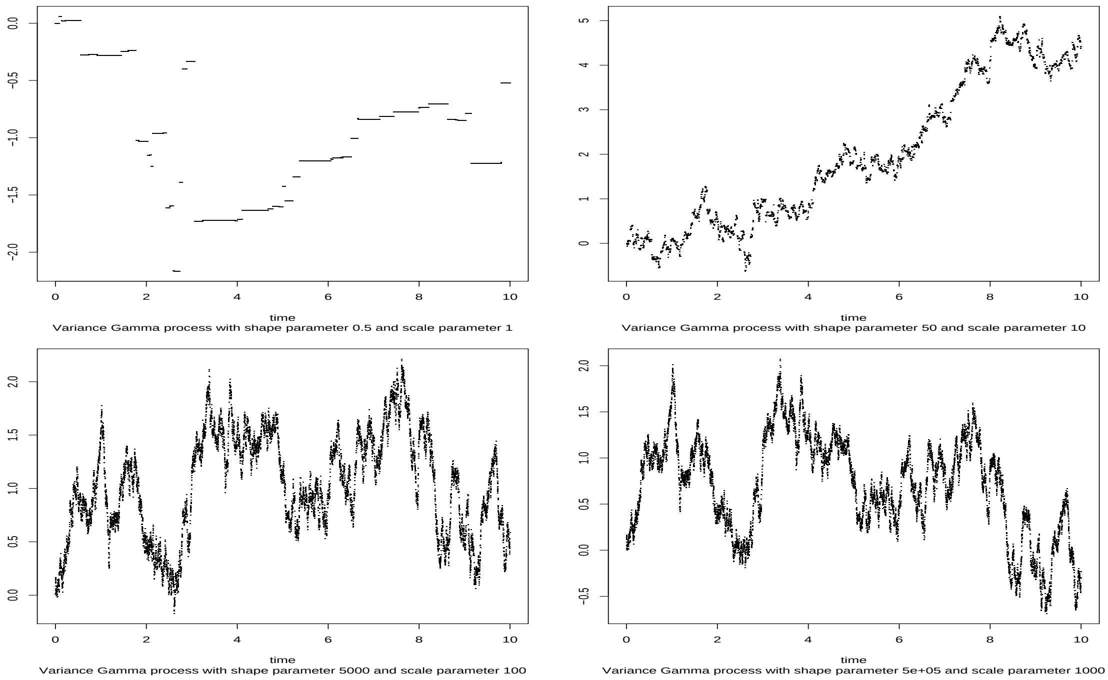


Figure 7.3: Simulation of Variance Gamma processes as differences of random walks

7.3 R code – not examinable

The following code is posted on the course website as `gammavgamma.R`.

```
psum <- function(vector){
b=vector;
b[1]=vector[1];
for (j in 2:length(vector)) b[j]=b[j-1]+vector[j]; b}

gammarw <- function(a,p){
unif=runif(10*p,0,1)
pos=qgamma(unif,a/p,a);
space=psum(pos);
time=(1/p)*1:(10*p);
plot(time,space,
pch=".",
sub=paste("Gamma process with shape parameter",a,"and scale parameter",a))}
```

```
vgammarw <- function(a,p){
unifpos=runif(10*p,0,1)
unifneg=runif(10*p,0,1)
pos=qgamma(unifpos,a*a/(2*p),a);
neg=qgamma(unifneg,a*a/(2*p),a);
space=psum(pos-neg);
time=(1/p)*1:(10*p);
plot(time,space,
pch=".",
sub=paste("Variance Gamma process with shape parameter",a*a/2,
"and scale parameter",a))}
```

Now you can try various values of parameters $a > 0$ and steps per time unit $p = 1/\delta$ in `gammarw(a,p)`, e.g.

```
gammarw(10,100)
vgammarw(10,1000)
```

Lecture 8

Simulation II

Reading: Ross 11.3, Schoutens Sections 8.1, 8.2, 8.4

In practice, the increment distribution is often not known, but the Lévy characteristics are, so we have to simulate Poisson point processes of jumps, by "throwing away the small jumps" and then analyse (and correct) the error committed.

8.1 Simulation via truncated Poisson point processes

Example 51 (Compound Poisson process) Let $(X_t)_{t\geq 0}$ be a compound Poisson process with Lévy density $g(x) = \lambda h(x)$, where h is a probability density function. Denote $H(x) = \int_{-\infty}^{x} h(y)dy$ and $H^{-1}(u) = \inf\{x \in \mathbb{R} : H(x) > u\}$. Let $Y_k = H^{-1}(U_{2k})$ and $Z_k = -\lambda^{-1}\ln(U_{2k-1})$, $k \geq 1$. Then the process

$$X_t^{(2)} = S_{N_t}, \qquad \text{where } S_n = \sum_{k=1}^{n} Y_k, \quad T_n = \sum_{k=1}^{n} Z_k, \quad N_t = \#\{n \geq 1 : T_n \leq t\},$$

has the same distribution as X.

Method 2 (Throwing away the small jumps) Let $(X_t)_{t\geq 0}$ be a Lévy process with characteristics $(a, 0, g)$, where g is not the multiple of a probability density function. Fix a jump size threshold $\varepsilon > 0$ so that $\lambda_\varepsilon = \int_{\{x\in\mathbb{R}:|x|>\varepsilon\}} g(x)dx > 0$, and write

$$g(x) = \lambda_\varepsilon h_\varepsilon(x), \quad |x| > \varepsilon, \qquad h_\varepsilon(x) = 0, \quad |x| \leq \varepsilon,$$

for a probability density function h_ε. Denote $H_\varepsilon(x) = \int_{-\infty}^{x} h_\varepsilon(y)dy$ and $H_\varepsilon^{-1}(u) = \inf\{x \in \mathbb{R} : H_\varepsilon(x) > u\}$. Let $Y_k = H_\varepsilon^{-1}(U_{2k})$ and $Z_k = -\lambda^{-1}\ln(U_{2k-1})$, $k \geq 1$. Then the process

$$X_t^{(2,\varepsilon)} = S_{N_t} - b_\varepsilon t, \qquad \text{where } S_n = \sum_{k=1}^{n} Y_k, \quad T_n = \sum_{k=1}^{n} Z_k, \quad N_t = \#\{n \geq 1 : T_n \leq t\},$$

and $b_\varepsilon = a - \int_{\{x\in\mathbb{R}:\varepsilon<|x|\leq 1\}} xg(x)dx$, is called the *process with small jumps thrown away.*

For characteristics (a, σ^2, g) we can now simulate $L_t = \sigma B_t + X_t$ by $\sigma B_t^{(1,\delta)} + X_t^{(2,\varepsilon)}$. The following proposition says that such approximations converge as $\varepsilon \downarrow 0$ (and $\delta \downarrow 0$). This is illustrated in Figure 6.3.

Proposition 52 *As $\varepsilon \downarrow 0$, we have $X_t^{(2,\varepsilon)} \to X_t$ in distribution.*

Proof: For a process with no negative jumps and characteristics $(0, 0, g)$, this is a consequence of the stronger Lemma 41, which gives a coupling for which convergence holds in the L^2 sense. For a general Lévy process with characteristics $(a, 0, g)$ that argument can be adapted, or we write $X_t = at + P_t - N_t$ and deduce the result:

$$\begin{aligned}\mathbb{E}(\exp\{i\lambda X_t^{(2,\varepsilon)}\}) &= e^{iat}\mathbb{E}(\exp\{i\lambda P_t^{(2,\varepsilon)}\})\mathbb{E}(\exp\{-i\lambda N_t^{(2,\varepsilon)}\})\\ &\to e^{iat}\mathbb{E}(\exp\{i\lambda P_t\})\mathbb{E}(\exp\{-i\lambda N_t\}) = \mathbb{E}(e^{i\lambda X_t}).\end{aligned}$$

□

Example 53 (Symmetric stable processes) Symmetric stable processes X are Lévy processes with characteristics $(0, 0, g)$, where $g(x) = c|x|^{-\alpha-1}$, $x \in \mathbb{R} \setminus \{0\}$ for some $\alpha \in (0, 2)$ (cf. Assignment 3.2.). We decompose $X = P - N$ for two independent processes with no negative jumps and simulate P and N. By doing this, we have

$$\lambda_\varepsilon = \int_\varepsilon^\infty g(x)dx = \frac{c}{\alpha}\varepsilon^{-\alpha}, \quad H_\varepsilon(x) = 1 - (\varepsilon/x)^\alpha \quad \text{and} \quad H_\varepsilon^{-1}(u) = \varepsilon(1-u)^{-1/\alpha}.$$

For the simulation we choose $\varepsilon = 0.01$. We compare $\alpha \in \{0.5, 1, 1.5, 1.8\}$. All processes are centred with infinite variance. Big jumps dominate the plots for small α. Recall that $\mathbb{E}(e^{i\lambda X_t}) = e^{-b|\lambda|^\alpha} \to e^{-b\lambda^2}$ as $\alpha \uparrow 2$, and we get, in fact, convergence to Brownian motion, the stable process of index 2.

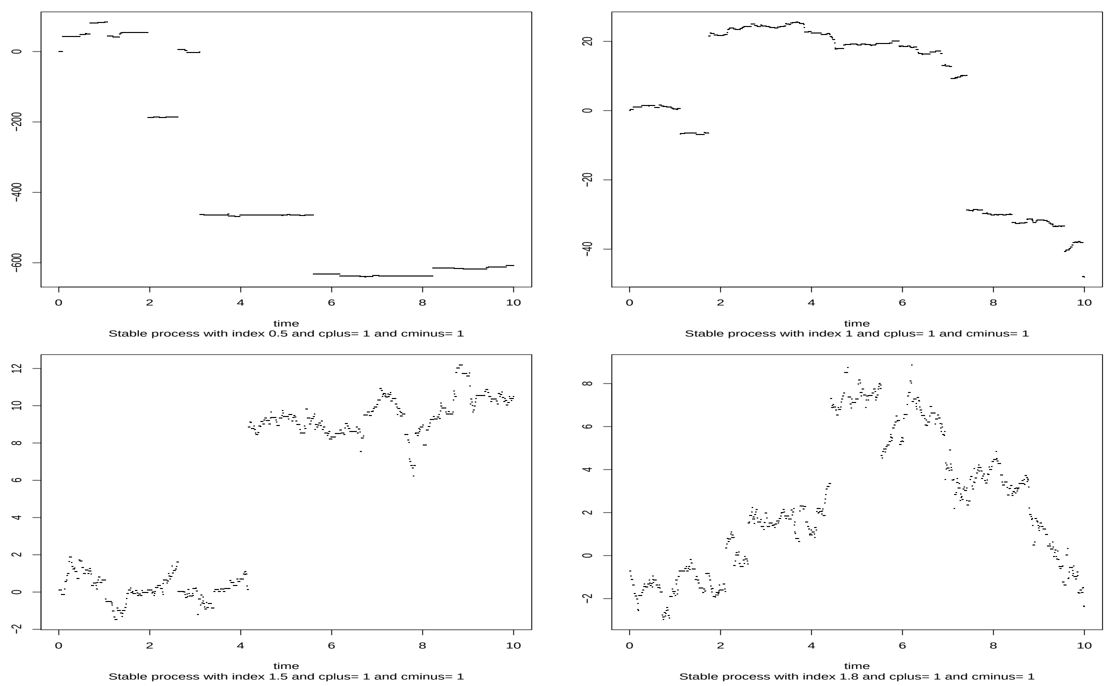


Figure 8.1: Simulation of symmetric stable processes from their jumps

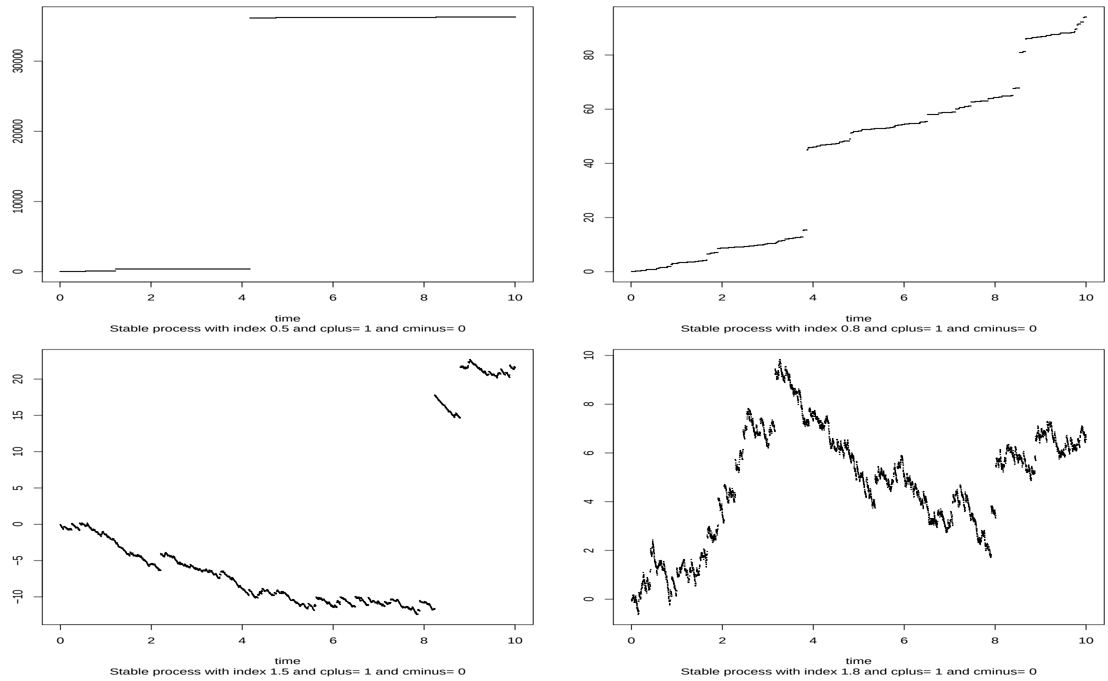


Figure 8.2: Simulation of stable processes with no negative jumps

Example 54 (Stable processes with no negative jumps) For stable processes with no negative jumps, we have $g(x) = c_+ x^{-\alpha-1}$, $x > 0$. The subordinator case $\alpha \in (0,1)$ was discussed in Assignment A.3.1. – $X_t = \sum_{s\le t} \Delta_s$. The case $\alpha \in [1,2)$, where compensation is required, is such that $\mathbb{E}(X_1) = 0$, i.e. $a = -\int_1^\infty xg(x)dx$. We choose $\epsilon = 0.1$ for $\alpha \in \{0.5, 0.8\}$ and $\varepsilon = 0.01$ for $\alpha \in \{1.5, 1.8\}$.

Strictly speaking, we take as triplet (a, σ^2, g) in Theorem 45 g as given, but for $\alpha \in (0,1)$ we take $a = \int_0^1 xg(x)dx$ so that

$$\begin{aligned}\mathbb{E}(e^{i\lambda X_t}) &= \exp\left\{t\int_0^\infty (e^{i\lambda x} - 1)g(x)dx\right\}\\ &= \exp\left\{-t\left(-i\lambda a - \int_0^\infty (e^{i\lambda x} - 1 - i\lambda x 1_{\{|x|\le 1\}})g(x)dx\right)\right\},\end{aligned}$$

since compensation of small jumps is not needed and we obtain a subordinator if we do not compensate, and for $\alpha \in (1,2)$, we take $a = -\int_1^\infty xg(x)dx$ so that

$$\begin{aligned}\mathbb{E}(e^{i\lambda X_t}) &= \exp\left\{t\int_0^\infty (e^{i\lambda x} - 1 - i\lambda x)g(x)dx\right\}\\ &= \exp\left\{-t\left(-i\lambda a - \int_0^\infty (e^{i\lambda x} - 1 - i\lambda x 1_{\{|x|\le 1\}})g(x)dx\right)\right\},\end{aligned}$$

since we can compensate all jump and achieve $\mathbb{E}(X_t) = 0$. Only with these choices we obtain the Lévy processes with no negative jumps that satisfy the scaling property.

This discussion shows that the representation in Theorem 45 is artificial and representations with different compensating drifts are often more natural.

8.2 Generating specific distributions

In this course, we will not go into the computational details of simulations. However, we do point out some principles here that lead to improved simulations, and we discuss some of the resulting modifications to the methods presented.

Often, it is not efficient to compute the inverse cumulative distribution function. For a number of standard distributions, other methods have been developed. We will here look at standard Normal generators. The Gamma distribution is discussed in Assignment A.4.2.

Example 55 (Box-Muller generator) Consider the following procedure

1. Generate two independent random numbers $U \sim \mathrm{Unif}(0,1)$ and $V \sim \mathrm{Unif}(0,1)$.
2. Set $X = \sqrt{-2\ln(U)}\cos(2\pi V)$ and $Y = \sqrt{-2\ln(U)}\sin(2\pi V)$.
3. Return the pair (X, Y).

The claim is the X and Y are independent standard Normal random variables. The proof is an exercise on the transformation formula. First, the transformation is clearly bijective from $(0,1)^2$ to $\mathbb{R}^2$. The inverse transformation can be worked out from $X^2+Y^2 = -2\ln(U)$ and $Y/X = \tan(2\pi V)$ as

$$(U, V) = T^{-1}(X, Y) = (e^{-(X^2+Y^2)/2}, (2\pi)^{-1}\arctan(Y/X))$$

(with an appropriate choice of the branch of arctan, which is not relevant here). The Jacobian of the inverse transformation is

$$J = \begin{pmatrix} -xe^{-(x^2+y^2)/2} & -ye^{-(x^2+y^2)/2} \\ -\frac{1}{2\pi}\frac{y}{x^2}\frac{1}{1+y^2/x^2} & \frac{1}{2\pi}\frac{1}{x}\frac{1}{1+y^2/x^2} \end{pmatrix} \qquad \Rightarrow |\det(J)| = \frac{1}{2\pi}e^{-(x^2+y^2)/2}$$

and so, as required,

$$f_{X,Y}(x,y) = f_{U,V}(T^{-1}(x,y))|\det(J)| = \frac{1}{2\pi}e^{-(x^2+y^2)/2}.$$

For a more efficient generation of standard Normal random variables, it turns out useful to first generate uniform random variables on on the disk of radius 1:

Example 56 (Uniform distribution on the disk) For $U_1 \sim \mathrm{Unif}(0,1)$ and $U_2 \sim \mathrm{Unif}(0,1)$ independent, we have that $(V_1, V_2) = (2U_1 - 1, 2U_2 - 1)$ is uniformly distributed on the square $(-1,1)^2$ centered at $(0,0)$, which contains the disk $D = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 < 1\}$, and we have, in particular $\mathbb{P}((V_1,V_2) \in D) = \pi/4$. Now, for all $A \subset \mathbb{R}^2$, we have

$$\mathbb{P}((V_1,V_2) \in A|(V_1,V_2) \in D) = \frac{\mathrm{area}(A \cap D)}{\pi},$$

so the conditional distribution of (V_1, V_2) given $(V_1, V_2) \in D$ is uniform on D. By the following lemma, this conditioning can be turned into an algorithm by repeated trials:

1. Generate two independent random numbers $U_1 \sim \mathrm{Unif}(0,1)$ and $U_2 \sim \mathrm{Unif}(0,1)$.

2. Set $(V_1, V_2) = (2U_1 - 1, 2U_2 - 1)$.

3. If $(V_1, V_2) \in D$, go to 4., else go to 1.

4. Return the numbers (V_1, V_2).

The pair of numbers returned will be uniformly distributed on the disk D.

Lemma 57 (Conditioning by repeated trials) *Let $X, X_1, X_2, \ldots$ be independent and identically distributed d-dimensional random vectors. Also let $A \subset \mathbb{R}^d$ such that $p = \mathbb{P}(X \in A) > 0$. Denote $N = \inf\{n \geq 1 : X_i \in A\}$. Then $N \sim \text{geom}(p)$ is independent of X_N, and X_N has as its (unconditional) distribution the conditional distribution of X given $X \in A$, i.e.*

$$\mathbb{P}(X_N \in B) = \mathbb{P}(X \in B | X \in A) \quad \textit{for all } B \subset \mathbb{R}^d.$$

Proof: We calculate the joint distribution

$$\begin{aligned} \mathbb{P}(N = n, X_n \in B) &= \mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A \cap B) \\ &= (1-p)^{n-1}\mathbb{P}(X_n \in A \cap B) = (1-p)^{n-1}p\mathbb{P}(X_n \in B | X_n \in A). \end{aligned}$$

□

We now get the following modification of the Normal generator:

Example 58 (Polar method) The following is a more efficient method to generate two independent standard Normal random variables:

1. Generate two independent random numbers $U_1 \sim \text{Unif}(0,1)$ and $U_2 \sim \text{Unif}(0,1)$.

2. Set $(V_1, V_2) = (2U_1 - 1, 2U_2 - 1)$ and $S = V_1^2 + V_2^2$.

3. If $S \leq 1$, go to 4., else go to 1.

4. Set $P = \sqrt{-2(\ln(S))/S}$

5. Return the pair $(X, Y) = (PV_1, PV_2)$.

The gain in efficiency mainly stems from the fact that no sine and cosine need to be computed. The method works because in polar coordinates $(V_1, V_2) = (R\cos(\Theta), R\sin(\Theta)$, we have independent $S = R^2 \sim \text{Unif}(0,1)$ and $\Theta \sim \text{Unif}(0, 2\pi)$ (as is easily checked), so we can choose $U = S$ and $2\pi V = \Theta$ in the Box-Muller generator.

8.3 R code – not examinable

The following code is posted on the course website as `stable.R`.

```
stableonesided <- function(a,c,eps,p){
f=c*eps^(-a)/a;
n=rpois(1,10*f);
t=runif(n,0,10);
y=(eps^(-a)-a*f*runif(n,0,1)/c)^(-1/a);
ytemp=1:n;res=(1:(10*p))/100;{
for (k in 1:(10*p)){{for (j in 1:n){
if(t[j]<=k/p)ytemp[j]<-y[j] else ytemp[j]<-0}}; res[k]<-sum(ytemp)}};
res}

stable <- function(a,cp,cn,eps,p){
pos=stableonesided(a,cp,eps,p);
neg=stableonesided(a,cn,eps,p);
space=pos-neg;time=(1/p)*1:(10*p);
plot(time,space,
pch=".",
sub=paste("Stable process with index",a,"and cplus=",cp,"and cminus=",cn))}

stableonesidedcomp <- function(a,c,eps,p){
f=(c*eps^(-a))/a;
n=rpois(1,10*f);
t=runif(n,0,10);
y=(eps^(-a)-a*f*runif(n,0,1)/c)^(-1/a);
ytemp=1:n;
res=(1:(10*p))/100;{ for (k in 1:(10*p)){{if (n!=0)for (j in 1:n){
if(t[j]<=k/p)ytemp[j]<-y[j] else ytemp[j]<-0}};{
if (n!=0)res[k]<-sum(ytemp)-(c*k/(p*(a-1)))*(eps^(1-a))
else res[k]<--c*k/(p*(a-1))*(eps^(1-a))}}};
res}
```

Lecture 9

Simulation III

Reading: Ross 11.3, Schoutens Sections 8.1, 8.2, 8.4; Further reading: Kyprianou Section 3.3

9.1 Applications of the rejection method

Lemma 57 can be used in a variety of ways. A widely applicable simulation method is the rejection method. Suppose you have an explicit probability density function f, but the inverse distribution function is not explicit. If $h \geq cf$, for some $c < 1$ is a probability density function whose inverse distribution function is explicit (e.g. uniform or exponential) or from which we can simulate by other means, then the procedure

1. Generate a random variable X with density h.
2. Generate an independent uniform variable U.
3. If $Uh(X) \leq cf(X)$, go to 4., else go to 1.
4. Return X.

Proposition 59 *The procedure returns a random variable with density f.*

Proof: Denote $p = \mathbb{P}(Uh(X) \leq cf(X))$. By Lemma 57 (applied to the vector (X, U)), the procedure returns a random variable with distribution

$$\begin{aligned} \mathbb{P}(X \leq x | Uh(X) \leq cf(X)) &= \frac{\mathbb{P}(X \leq x, Uh(X) \leq cf(X))}{p} \\ &= \frac{1}{p}\int_{-\infty}^{x} h(z)\mathbb{P}(U \leq cf(z)/h(z))dz = \frac{c}{p}\int_{-\infty}^{x} f(z)dz, \end{aligned}$$

and letting $x \to \infty$ shows $c = p$. □

Example 60 (Gamma distribution) *Note that the Gamma density for $\alpha > 1$ satisfies*

$$\frac{1}{\Gamma(\alpha)}\beta^{\alpha}x^{\alpha-1}e^{-\beta x} \leq \frac{\beta^{\alpha-1}}{\Gamma(\alpha)}\beta e^{-\beta x}$$

so we can apply the procedure with $f(x) = \beta e^{-\beta x}$ and $c = \Gamma(\alpha)/\beta^{\alpha-1}$.

It is important that c is not too small, since otherwise lots of iterations are needed until the random variable is returned. The number of iterations is geometrically distributed (first success in a sequence of independent Bernoulli trials) with success parameter c, so on average $1/c$ trials are required.

For the simulation via Poisson point processes (with truncation at ε, say), we can use properties of Poisson point processes to simulate separately jumps of sizes in intervals I_n, $n = 1, \ldots, n_0$ and can choose intervals I_n so that the intensity function g is almost constant.

Example 61 (Distribution on I_n) *Suppose that we simulate a Poisson point process on a bounded spatial interval $I_n = (a, b]$, with some intensity function $g : I_n \to [0, \infty)$. Then we can take uniform*

$$h(x) = 1/(b-a) \quad \textit{and} \quad c = \frac{\int_a^b g(x)dx}{(b-a)\max\{g(x) : a < x \leq b\}}.$$

and simulate $\mathrm{Exp}(\int_a^b g(x)dx)$*-spaced times T_n and spatial coordinates Δ_{T_n} by the rejection method with h and c as given.*

9.2 "Errors increase in sums of approximated terms."

Methods 1 and 2 are based on sums of many, mostly small, independent identically distributed random variables. As $\delta \downarrow 0$ or $\varepsilon \downarrow 0$, these are more and more smaller and smaller random variables. If each is affected by a small error, then adding up these errors makes the approximations worse whereas the precision should increase.

For Method 1, this can often be prevented by suitable conditioning, e.g. on the terminal value:

Example 62 (Poisson process) A Poisson process with intensity λ on the time interval $[0, 1]$ can be generated as follows:

1. Generate a Poisson random variable N with parameter λ.

2. Given $N = n$, generate n independent $\mathrm{Unif}(0, 1)$ random variables $U_1, \ldots, U_n$.

3. Return $X_t = \#\{1 \leq i \leq N : U_i \leq t\}$.

Clearly, this process is a Poisson process, since X_s (and indeed $(X_{t_1}, \ldots, X_{t_n - t_{n-1}})$) is obtained from $X_1 \sim \mathrm{Poi}(\lambda)$ by thinning as in Section 4.2.

This is not of much practical use since we would usually simulate a Poisson random variable by evaluating a unit rate Poisson process (simulated from standard exponential interarrival times) at λ. In the case of Brownian motion (and the Gamma process, see Assignment A.4.3.), however, such conditioning is very useful and can then be iterated, e.g. in a dyadic scheme:

Example 63 (Brownian motion) Consider the following method to generate Brownian motion on the time interval $[0, 1]$.

1. Set $X_0 = 0$ and generate $X_1 \sim \text{Normal}(0,1)$ hence specifying $X_{k2^{-n}}$ for $n = 0$, $k = 0, \ldots, 2^n$.
2. For $k = 1, \ldots, 2^n$, conditionally given $X_{(k-1)2^{-n}} = x$ and $X_{k2^{-n}} = z$, generate
$$X_{(2k-1)2^{-n-1}} \sim \text{Normal}\left(\frac{x+z}{2}, 2^{-n-2}\right)$$
3. If the required precision has been reached, stop, else increase n by 1 and go back to 2.

This process is Brownian motion, since the following lemma shows that Brownian motion has these conditional distributions. Specifically, the $n = 0$, $k = 1$ case of 2. is obtained directly for $s = 1/2$, $t = 1$. For $n \geq 1$, $k = 1, \ldots, 2^n$, note that $X_{(2k-1)2^{-n-1}} - X_{(k-1)2^{-n}}$ is independent of $X_{(k-1)2^{-n}}$ and so, we are really saying that for Brownian motion
$$X_{(2k-1)2^{-n-1}} - X_{(k-1)2^{-n}} \sim \text{Normal}\left(\frac{z-x}{2}, 2^{-n-2}\right),$$
conditionally given $Z_{k2^{-n}} - Z_{(k-1)2^{-n}} = z - x$, which is equivalent to the specification in 2.

A further advantage of this method is that $\delta = 2^{-n}$ can be decreased without having to start afresh. Previous less precise simulations can be refined.

Lemma 64 *Let $(X_t)_{t\geq 0}$ be Brownian motion and $0 < s < t$. Then, the conditional distribution of X_s given $X_t = z$ is* $\text{Normal}(zs/t, s(t-s)/t)$.

Proof: Note that $X_s \sim \text{Normal}(0,s)$ and $X_t - X_s \sim \text{Normal}(0, t-s)$ are independent. By the transformation formula (X_s, X_t) has joint density
$$f_{X_s,X_t}(x,z) = \frac{1}{2\pi\sqrt{s(t-s)}} \exp\left\{-\frac{x^2}{2s} - \frac{(z-x)^2}{2(t-s)}\right\},$$
and so the conditional density is
$$\begin{aligned} f_{X_s|X_t=z}(x) &= \frac{f_{X_s,X_t}(x,z)}{f_{X_t}(z)} = \frac{1}{\sqrt{2\pi}\sqrt{s(t-s)/t}} \exp\left\{-\frac{x^2}{2s} - \frac{(z-x)^2}{2(t-s)} + \frac{z^2}{2t}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{s(t-s)/t}} \exp\left\{-\frac{(x-zs/t)^2}{2s(t-s)/t}\right\}. \end{aligned}$$
□

For Method 2, we can achieve similar improvements by simulating $(\Delta_t)_{t\geq 0}$ in stages. Choose a strictly decreasing sequence $\infty = a_0 > a_1 > a_2 > \ldots > 0$ of jump size thresholds with $a_n \downarrow 0$ as $n \to \infty$
$$\Delta_t^{(k)} = \Delta_t 1_{\{a_k \leq \Delta_t < a_{k-1}\}}, \quad \Delta_t^{(-k)} = \Delta_t 1_{\{-a_k \geq \Delta_t > -a_{k-1}\}}, \qquad k \geq 1, t \geq 0.$$
Simulate the Poisson counting processes $N^{(k)}$ associated with $\Delta^{(k)}$ as in Example 62 and otherwise construct
$$Z_t^{(k)} = \sum_{s\leq t} \Delta_s^{(k)} - t\int_{a_k}^{a_{k-1}} x 1_{\{0<x<1\}} g(x) dx$$
as in Method 2 and include so many $k = \pm 1, \pm 2, \ldots$ as precision requires.

9.3 Approximation of small jumps by Brownian motion

Theorem 65 (Asmussen-Rosinski) *Let $(X_t)_{t\geq 0}$ be a Lévy process with characteristics $(a,0,g)$. Denote*

$$\sigma^2(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} x^2 g(x)dx$$

If $\sigma(\varepsilon)/\varepsilon \to \infty$ as $\varepsilon \downarrow 0$, then

$$\frac{X_t - X_t^{(2,\varepsilon)}}{\sigma(\varepsilon)} \to B_t \qquad \text{in distribution as } \varepsilon \downarrow 0$$

for an independent Brownian motion $(B_t)_{t\geq 0}$

If $\sigma(\varepsilon)/\varepsilon \to \infty$, it is well-justified to adjust Method 2 to set

$$X_t^{(2+,\varepsilon)} = X_t^{(2,\varepsilon)} + \sigma(\varepsilon)B_t$$

for an independent Brownian motion. In other words, we may approximate the small jumps by an independent Brownian motion.

Example 66 (CGMY process) The CGMY process is a popular process in Mathematical Finance. It is defined via its characteristics $(0,0,g)$, where

$$g(x) = C\exp\{-G|x|\}|x|^{-Y-1},\ x<0, \quad g(x) = C\exp\{-M|x|\}|x|^{-Y-1},\ x>0.$$

for some $C \geq 0$, $G > 0$, $M > 0$ and $Y < 2$. Let $(X_t)_{t\geq 0}$ be a CGMY process. We calculate

$$\sigma^2(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} x^2 g(x)dx \leq C\int_{-\varepsilon}^{\varepsilon} |x|^{1-Y}dx = \frac{2C}{2-Y}\varepsilon^{2-Y}$$

and for every given $\delta > 0$ and all $\varepsilon > 0$ small enough, the same quantity with C replaced by $C - \delta$ is a lower bound, so that

$$\frac{\sigma(\varepsilon)}{\varepsilon} \sim \sqrt{\frac{2C}{2-Y}}\varepsilon^{-Y/2} \to \infty \iff Y > 0$$

Hence an approximation of the small jumps of size $(-\varepsilon,\varepsilon)$ thrown away by a Brownian motion $\sigma(\varepsilon)B_t$ is appropriate if and only if $Y > 0$. In fact, for $Y < 0$, the process has finite jump intensity, so all jumps can be simulated. Therefore, only the case $Y = 0$ is problematic. This is the Variance Gamma process (and its asymmetric companions).

Whether or not we can approximate small jumps by a Brownian motion, we have to decide what value of ε to choose. By the independence properties of Poisson point processes, the remainder term that $X_t - X_t^{2,\varepsilon}$ is a (zero mean, for $\varepsilon < 1$) Lévy process with intensity function g on $[-\varepsilon,\varepsilon]$ and variance

$$\sigma^2(\varepsilon) = \mathrm{Var}(X_t - X_t^{(2,\varepsilon)})$$

(let $\delta \downarrow 0$ in the proof of Lemma 41). We can choose ε so that the accuracy of $X_t^{(2,\varepsilon)}$ is within an agreed deviation h, i.e. e.g. $2\sigma(\varepsilon) = h$. In the setting of Theorem 65, this means that a deviation of X_t from $X_t^{(2,\varepsilon)}$ by more than h would happen with probability about 0.05.

9.4 Appendix: Consolidation on Poisson point processes

This section and and the next should not be necessary this year, because the relevant material has been included in earlier lectures. They may still be useful as a reminder of key concepts.

We can consider Poisson point processes $(\Delta_t)_{t\geq 0}$ in very general spaces, e.g. (topological) spaces $(E, \mathcal{O})$ where we have a collection/notion of open sets $O \in \mathcal{O}$ (and an associated Borel σ-algebra $\mathcal{B} = \sigma(\mathcal{O})$, the smallest σ-algebra that contains all open sets, for which we also require that $\{x\} \in \mathcal{B}$ for all $x \in E$ and $G = \{(x,x) : x \in E\} \in \mathcal{B} \otimes \mathcal{B}$). We just require that $(\Delta_t)_{t\geq 0}$ is a family of $\Delta_t \in E \cup \{0\}$ such that there is an intensity (Borel) measure ν on E with $\nu(\{x\}) = 0$ for all $x \in E$,

(a) for disjoint $A_1 = (a_1, b_1] \times O_1, \ldots, A_n = (a_n, b_n] \times O_n$, $O_i \in \mathcal{O}$, the counts

$$N(A_i) = N((a_1, b_1] \times O_i) = \#\{t \in (a_i, b_i] : \Delta_t \in O_i\}, \quad i = 1, \ldots, n$$

are independent random variables and

inhom (b) $N(A_i) \sim \mathrm{Poi}((b_i - a_i)\nu(O_i))$.

For us, $E = \mathbb{R} \setminus \{0\}$ is the space of jump sizes, and $\nu(O_i) = \nu((c_i, d_i)) = \int_{c_i}^{d_i} g(x)dx$ for an intensity function $g : \mathbb{R} \setminus \{0\} \to [0, \infty)$. Property (a) for all open sets is then equivalent to property (a) for all measurable sets or all half-open intervals or all closed intervals etc. (all that matters is that the collection generates the Borel σ-algebra). It is an immediate consequence of the definition (and this discussion) that for (measurable) disjoint $B_1, B_2, \ldots \subset \mathbb{R} \setminus \{0\}$, the "restricted" processes

$$\Delta_t^{(i)} = \Delta_t 1_{\{\Delta_t \in B_i\}}, \qquad t \geq 0,$$

are also Poisson point processes with the restriction of g to B_i as intensity function, and they are independent. We used this fact crucially and repeatedly in two forms. Firstly, for $B_1 = (0, \infty)$ and $B_2 = (-\infty, 0)$ (and $B_3 = B_4 = \ldots = \varnothing$), we consider Poisson point processes of positive points (jump sizes) and of negative points (jump sizes). We constructed from them independent Lévy processes. Secondly, for a sequence $\infty = a_0 > a_1 > a_2 > \ldots$, we considered $B_i = [a_i, a_{i-1})$, $i \geq 1$, so as to simulate separately independent Lévy processes (in fact compound Poisson processes with linear drift) with jump sizes only in B_i.

9.5 Appendix: Consolidation on the compensation of jumps

The general Lévy process requires compensation of small jumps in its approximation by processes with no jumps in $(-\varepsilon, \varepsilon)$, as $\varepsilon \downarrow 0$. This is reflected in its characteristic function of the form

$$\mathbb{E}(e^{i\lambda X_t}) = e^{-t\psi(\lambda)}, \ \psi(\lambda) = -ia_1\lambda + \frac{1}{2}\sigma^2\lambda^2 - \int_{-\infty}^{\infty} (e^{i\lambda x} - 1 - i\lambda x 1_{\{|x|\leq 1\}})\nu(dx), \ \lambda \in \mathbb{R}, \ (1)$$

where usually $\nu(dx) = g(x)dx$. This is a parametrisation by (a_1, σ^2, ν) or (a_1, σ^2, g), where we require the (weak) integrability condition $\int_{-\infty}^{\infty}(1 \wedge x^2)\nu(dx) < \infty$.

The first class of Lévy processes that we constructed systematically were subordinators, where no compensation was necessary. We parametrised them by parameters $a_2 \geq 0$ and $g : (0, \infty) \to [0, \infty)$ (or ν measure on $(0, \infty)$) so that the moment generating function is of the form

$$\mathbb{E}(e^{\gamma X_t}) = e^{t\Psi(\gamma)}, \quad \Psi(\gamma) = a_2\gamma + \int_0^{\infty}(e^{\gamma x} - 1)\nu(dx), \qquad \gamma \leq 0. \tag{2}$$

We required the stronger integrability condition $\int_0^{\infty}(1 \wedge x)\nu(dx) < \infty$. Similarly, for differences of subordinators, we have a characteristic function

$$\mathbb{E}(e^{i\lambda X_t}) = e^{-t\psi(\lambda)}, \quad \psi(\lambda) = -ia_2\lambda + \frac{1}{2}\sigma^2\lambda^2 - \int_{-\infty}^{\infty}(e^{i\lambda x} - 1)\nu(dx), \tag{3}$$

under the stronger integrability condition $\int_{-\infty}^{\infty}(1 \wedge |x|)\nu(dx) < \infty$. Compensation in (1) is only done for small jumps. This is, because, in general the indicator $1_{\{|x|\leq 1\}}$ cannot be omitted. However, if $\int_{-\infty}^{\infty}(x \wedge x^2)\nu(dx) < \infty$, then we can also represent

$$\mathbb{E}(e^{i\lambda X_t}) = e^{-t\psi(\lambda)}, \quad \psi(\lambda) = -ia_3\lambda + \frac{1}{2}\sigma^2\lambda^2 - \int_{-\infty}^{\infty}(e^{i\lambda x} - 1 - i\lambda x)\nu(dx). \tag{4}$$

Equations (1), (3) and (4) are compatible whenever any two integrability conditions are fulfilled, since the linear (in λ) terms under the integral can be added to a_1 to give

$$a_2 = a_1 + \int_{\{|x|\leq 1\}} x\nu(dx) \qquad \text{and} \qquad a_3 = a_1 - \int_{\{|x|>1\}} x\nu(dx).$$

Note that then (by differentiation at $\lambda = 0$), we get $a_3 = \mathbb{E}(X_1)$. If $a_3 = 0$, then $(X_t)_{t\geq 0}$ is a martingale. On the other hand, for processes with finite jump intensity, i.e. under the even stronger integrability condition $\int_{-\infty}^{\infty} g(x)dx < \infty$, we get a_1 as the slope of the paths of X between the jumps. Both a_1 and a_3 are therefore natural parameterisations, but not available, in general. a_2 is available in general, but does not have such a natural interpretation.

We use characteristic functions for similar reasons: in general, moment generating functions do not exist. If they do, i.e. under a strong integrability condition $\int_1^{\infty} e^{\gamma x}\nu(dx) < \infty$ for some $\gamma > 0$ or $\int_{-\infty}^{-1} e^{\gamma x}\nu(dx) < \infty$ for some $\gamma < 0$, we get

$$\mathbb{E}(e^{\gamma X_t}) = e^{t\Psi(\gamma)}, \quad \Psi(\gamma) = a_1\gamma + \frac{1}{2}\sigma^2\gamma^2 + \int_{-\infty}^{\infty}(e^{\gamma x} - 1 - \gamma x 1_{\{|x|\leq 1\}})g(x)dx. \tag{5}$$

Moment generating functions are always defined on an interval I possibly including end points $\gamma_- \in [-\infty, 0]$ and/or $\gamma_+ \in [0, \infty]$, we always have $0 \in I$, but maybe $\gamma_- = \gamma_+ = 0$. If $1 \in I$ and a_1 is such that $\Psi(1) = 0$, then $(e^{X_t})_{t\geq 0}$ is a martingale.

Lecture 10

Lévy markets and incompleteness

Reading: Schoutens Chapters 3 and 6

10.1 Arbitrage-free pricing (from B10b)

By Donsker's Theorem, Brownian motion is the scaling limit of most random walks and in particular of the simple symmetric random walk $R_n = X_1 + \ldots + X_n$ where $X_1, X_2, \ldots$ are independent with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$.

Corollary 67 *For simple symmetric random walk $(R_n)_{n\geq 0}$, we have $e^{R_{[nt]}/\sqrt{n}} \to e^{B_t}$, geometric Brownian motion, in distribution as $n \to \infty$.*

Proof: First note that $\mathbb{E}(X_1) = 0$ and $\mathrm{Var}(X_1) = 1$. Now for all $x > 0$

$$\mathbb{P}(e^{R_{[nt]}/\sqrt{n}} \leq x) = \mathbb{P}(R_{[nt]}/\sqrt{n} \leq \ln(x)) \to \mathbb{P}(B_t \leq \ln(x)) = \mathbb{P}(e^{B_t} \leq x),$$

by the Central Limit Theorem. □

This was convergence of for fixed t. Stronger convergence, locally uniformly in t can also be shown. Note that $(R_n)_{n\geq 0}$ is a martingale, and so is $(B_t)_{t\geq 0}$. However,

$$\mathbb{E}(e^{R_n}) = \left(\frac{1}{2}e^{-1} + \frac{1}{2}e\right)^n \to \infty$$

Proposition 68 *For non-symmetric simple random walk $(R_n)_{n\geq 0}$ with $\mathbb{P}(X_i = 1) = p$, the process $(e^{R_n})_{n\geq 0}$ is a martingale if and only if $p = 1/(1+e)$.*

Proof: By the fourth and first rules for conditional expectations, we have

$$\mathbb{E}(e^{R_{n+1}}|e^{R_0}, \ldots, e^{R_n}) = \mathbb{E}(e^{R_n}e^{X_{n+1}}|e^{R_0}, \ldots, e^{R_n}) = e^{R_n}\mathbb{E}(e^{X_{n+1}})$$

and so, $(e^{R_n})_{n\geq 0}$ is a martingale if and only if

$$1 = \mathbb{E}(e^{X_{n+1}}) = pe + (1-p)e^{-1} \iff p(e - 1/e) = 1 - 1/e \iff p = 1/(e+1). \quad \square$$

The argument works just assuming that $R_n = X_1 + \ldots + X_n$, $n \geq 0$, satisfies $\mathbb{P}(|X_{n+1}| = 1|R_0 = r_0, \ldots, R_n = r_n) = 1$. Among all joint distributions, the non-symmetric exponentiated random walk with $p = 1/(1+e)$ is the only martingale.

The concept of *arbitrage-free* pricing in binary models is leaving aside any randomness. We will approach the Black-Scholes model from discrete models.

Suppose we have a *risky asset* with random price process S per unit and a *risk-free asset* with deterministic value function A per unit. Consider *portfolios* (U,V) of U units of the risky asset and V units of the risk-free asset. We allow that (U_t,V_t) depends on the performance of $(S_s)_{0\le s<t}$ but not on $(S_s)_{s\ge t}$. We denote the value of the portfolio at time t by $W_t=U_tS_t+V_tA_t$. The composition of the portfolio may change with time, but we consider only *self-financing* ones, for which any risky asset bought is paid for from the risk-free asset holdings and vice versa. We say that *arbitrage opportunities* exist if there is a self-financing portfolio process (U,V) and a time t so that $\mathbb{P}(W_0=0)=1$, $\mathbb{P}(W_t\ge 0)=1$ and $\mathbb{P}(W_t>0)>0$. We will be interested in models where no arbitrage opportunities exist.

Example 69 (One-period model) There are two scenarios "up" and "down" (to which we may later assign probabilities $p\in(0,1)$ and $1-p$). The model consists of (S_0,S_1) only, where S_0 changes to $S_1(\text{up})$ or $S_1(\text{down})<S_1(\text{up})$ after one time unit. The risk-free asset will evolve from A_0 to A_1. At time 0, we have $W_0=U_0S_0+V_0A_0$. At time 1, the value will change to either

$$W_1(\text{up})=U_0S_1(\text{up})+V_0A_1 \qquad \text{or} \qquad W_1(\text{down})=U_0S_1(\text{down})+V_0A_1. \tag{1}$$

It is easily seen that arbitrage opportunities occur if and only if $A_1/A_0\ge S_1(\text{up})/S_0$ or $S_1(\text{down})/S_0\ge A_1/A_0$, i.e. if one asset is uniformly preferable to the other.

A *derivative security* (or contingent claim) with maturity t is a contract that provides the owner with a *payoff* W_t dependent on the performance of $(S_s)_{0\le s\le t}$. If there is a self-financing portfolio process (U,V) with value W_t at time t, then such a portfolio process is called a *hedging portfolio process* replicating the contingent claim. The value $W_0=U_0S_0+V_0A_0$ of the hedging portfolio at time 0 is called the *arbitrage-free price* of the derivative security. It is easily seen that there would be an arbitrage opportunity, if the derivative security was available to buy and sell at any other price (as an additional asset in the model). In general, not all contingent claims can be hedged.

Example 69 (One-period model, continued) Consider any contingent claim, i.e. a payoff of $W_1(\text{up})$ or $W_1(\text{down})$ according to whether scenario "up" or "down" take place. Equations (1) can now be used to set up a hedging portfolio (U_0,V_0) and calculate the unique arbitrage-free price. Note that the arbitrage-free price is independent of probabilities $p\in(0,1)$ and $1-p$ that we may assign to the two scenarios as part of our model specification. Because of the linearity of (1), there is a unique $q\in(0,1)$ such that for *all* contingent claims $W_1:\{\text{up},\text{down}\}\to\mathbb{R}$

$$W_0=qW_1(\text{up})+(1-q)W_1(\text{down}).$$

If we refer to q and $1-q$ as probabilities of "up" and "down", then W_0 is the expectation of W_1 under this distribution. If $A_0=A_1=1$, $S_0=1$, $S_1(\text{up})=e$, $S_1(\text{down})=e^{-1}$, then we identify (for $W_1=S_1$ and hence $(U_0,V_0)=(1,0)$) that $q=1/(1+e)$.

The property that *every* contingent claim can be hedged by a self-financing portfolio process is called *completeness* of the market model.

Example 70 (n-period model) Each of n periods has two scenarios, "up" and "down", as is the case for the model $S_k = e^{R_k}$ for a simple random walk $(R_k)_{0\le k\le n}$. Then there are 2^n different combinations of "up" ($X_k = 1$) and "down" ($X_k = -1$). A contingent claim at time n is now any function W_n assigning a payoff to each of these combinations. By a backward recursion using the one-period model as induction step, we can work out hedging portfolios (U_k, V_k) and the value W_k of the derivative security at times $k = n-1, n-2, \dots, 0$, where in each case, (U_k, V_k) and W_k will depend on previous "up"s and "down"s $X_1, \dots, X_k$, so this is a specification of 2^k values each. W_0 will be the unique arbitrage-free price for the derivative security. The induction also shows that, for $A_0 = A_1 = \ldots = A_n = 1$, it can be worked out as

$$W_0 = \mathbb{E}(W_n(X_1, \dots, X_n)), \quad \text{where } X_k \text{ independent with } \mathbb{P}(X_k = 1) = 1/(1+e),$$

and that $(W_k)_{0\le k\le n}$ is a martingale with $W_k = \mathbb{E}(W_n|X_1, \dots, X_k)$, e.g. $(S_k)_{0\le k\le n}$. If $A_k = (1+i)^k = e^{\delta k}$, we get an arbitrage-free model if and only if $-1 < \delta < 1$, and then

$$W_0 = e^{-\delta n}\mathbb{E}(W_n), \quad \text{where } X_k \text{ independent with } \mathbb{P}(X_k = 1) = (e^{1+\delta} - 1)/(e^2 - 1),$$

where now $(e^{-\delta k}W_k)_{0\le k\le n}$ is a martingale. In particular, the n-period model is complete.

Example 71 (Black-Scholes model) Let $S_t = S_0 \exp\{\sigma B_t + (\mu - \frac{1}{2}\sigma^2)t\}$ for a Brownian motion $(B_t)_{t\ge 0}$ and two parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Also put $A_t = e^{\delta t}$. It can be shown that also in this model, every contingent claim can be hedged, i.e. the Black-Scholes model is complete. Moreover, the pricing of contingent claims W_t can be carried out using the risk-neutral process

$$R_t = S_0 \exp\{\sigma B_t + (\delta - \tfrac{1}{2}\sigma^2))t\}, \quad t \ge 0,$$

where the drift parameter is δ, not μ. The discounted process $M_t = e^{-\delta t}R_t$ is a martingale and has analogous uniqueness properties to the martingale for the n-period model, but they are much more complicated to formulate here.

The arbitrage-free price of $W_t = G((S_s)_{0\le s\le t})$ is now (for all $\mu \in \mathbb{R}$)

$$W_0 = e^{-\delta t}\mathbb{E}(G((R_s)_{0\le s\le t})).$$

Examples are $G((S_s)_{0\le s\le t}) = (S_t - K)^+$ for the European call option and $G((S_s)_{0\le s\le t}) = (K - S_t)^+$ for the European put option. We will also consider path-dependent options such as Up-and-out barrier options with payoff $(S_t - K)^+ 1_{\{\overline{S}_t < H\}}$, where $\overline{S}_t = \sup_{0\le s\le t} S_s$ and H is the barrier. The option can only be exercised if the stock price does not exceed the barrier H at any time before maturity.

10.2 Introduction to Lévy markets

The Black-Scholes model is widely used for option pricing in the finance industry, largely because many options can be priced explicitly and there are computationally efficient methods also for more complicated derivatives, that can be carried out frequently and for high numbers of options. However, its model fit is poor and any price that is obtained from the Black-Scholes model must be adjusted to be realistic. There are several models based on Lévy processes that offer better model fit, but the Black-Scholes methods for option pricing do not transfer one-to-one. Lévy processes give a very wide modelling freedom. For practical applications it is useful to work with parametric subfamilies. Several such families have been suggested.

Example 72 (CGMY process) Carr, Geman, Madan and Yor proposed a model with four parameters (the first letters of their names. It is defined via its intensity

$$g(x) = C\exp\{-G|x|\}|x|^{-Y-1},\ x < 0, \quad g(x) = C\exp\{-M|x|\}|x|^{-Y-1},\ x > 0.$$

Let $(X_t)_{t\geq 0}$ be a CGMY process. If $M > 1$, then a risk-neutral price process can be modelled as $R_t = R_0\exp\{X_t - t(\phi(1) - \delta)\}$. Then the discounted process $(e^{-\delta t}R_t)_{t\geq 0}$ is a martingale, and it can be shown that arbitrage-free prices for contingent claims $W_t = G((R_s)_{0\leq s\leq t})$ can be calculated as

$$W_0 = e^{-\delta t}\mathbb{E}(G((R_s)_{0\leq s\leq t})).$$

It can also be shown, however, that this is not the only way to obtain arbitrage-free prices, and other prices do not necessarily lead to arbitrage opportunities. Also, not every contingent claim can be hedged, the model is not complete.

10.3 Incomplete discrete financial markets

Essentially, arbitrage-free discrete models are complete only if the number of possible scenarios $\omega_0, \ldots, \omega_m$ (for one period) is the same as the number of assets $S_1^{(1)}, \ldots, S_1^{(m)} : \Omega = \{\omega_0, \ldots, \omega_m\} \to \mathbb{R}$ in the model, since this leads to a system of linear equations to relate a contingent claim $W_1 : \{\omega_0, \ldots, \omega_m\} \to \mathbb{R}$ to a portfolio $(U^{(1)}, \ldots, U^{(m)}, V)$

$$V_0A_1 + \sum_{i=1}^{m} U_0^{(i)}S_1^{(i)}(\omega_j) = W_1(\omega_j), \qquad j = 0, \ldots, m$$

that can usually be uniquely solved for $(U_0^{(1)}, \ldots, U_0^{(m)}, V_0)$, and we can read off

$$W_0 = V_0A_0 + \sum_{i=1}^{m} U_0^{(i)}S_0^{(i)}.$$

If the number of possible scenarios is higher, then the system does not have a solution, in general (and hedging portfolios will not exist, in general). If the number of possible scenarios is lower, there will usually be infinitely many solutions.

If the system has no solution in general, the model is incomplete, but this does not mean that there is no price. It means that there is not a unique price. We can, in general, get some lower and upper bounds for the price imposed by no-arbitrage. One way of approaching this is to add a derivative security to the market as a further asset with an initial price that keeps the no arbitrage property for the extended model. One can, in fact, add more and more assets until the model is complete. Then there exist unique probabilities $q_j = \mathbb{P}(\omega_j)$, $0 \leq j \leq m$, that make all discounted assets $(A_0/A_1)S^{(j)}$, $1 \leq j \leq m$ (including the ones added to complete the market) martingales.

Example 73 (Ternary model) Suppose there are three scenarios, but only two assets. The model with $S_0 = 1 = A_0 < A_1 = 2$, and $1 = S_1(\omega_0) < 2 = S_1(\omega_1) < 3 = S_1(\omega_2)$ is easily seen to be arbitrage-free since $S_1(\omega_0) < A_1 < S_1(\omega_2)$. The contingent claim $0 = W_1(\omega_0) = W_1(\omega_1) < W_1(\omega_2) = 1$ can be hedged if and only if

$$2V_0 + U_0 = 0, \quad 2V_0 + 2U_0 = 0, \quad 2V_0 + 3U_0 = 1,$$

but the first two equations already imply $U_0 = V_0 = 0$ and then the third equation is false. Therefore the contingent claim W_1 cannot be hedged. The model is not complete.

For the model (W, S, A) to be arbitrage-free we clearly require $W_0 > 0$ since otherwise we could make arbitrage with a portfolio $(1, 0, 0)$, just "buying" the security. Its cost at time zero is nonpositive and its value at time one is nonnegative and positive for scenario ω_2. Now note that (S, A) is arbitrage-free, so any arbitrage portfolio must be of the form $(-1/W_0, U_0, V_0)$ with zero value $-1 + U_0 + V_0 = 0$ at time 0 and values at time 1

$$U_0 + 2V_0 = 2 - U_0, \quad 2U_0 + 2V_0 = 2 > 0, \quad -1/W_0 + 3U_0 + 2V_0 = -1/W_0 + 2 + U_0,$$

so that we need $2 \geq U_0 \geq 1/W_0 - 2$, and this is possible if and only if $W_0 \geq 1/4$. Therefore, the range of arbitrage-free prices is $W_0 \in (0, 1/4)$.

We can get these prices as expectations under martingale probabilities:

$$1 = \frac{1}{2}\mathbb{E}_q(S_1) = \frac{1}{2}q_0 + q_1 + \frac{3}{2}q_2, \quad W_0 = \frac{1}{2}\mathbb{E}_q(W_1) = \frac{1}{2}q_2, \quad q_0 + q_1 + q_2 = 1.$$

This is a linear system for (q_0, q_1, q_2) that we solve to get

$$q_0 = q_2 = 2W_0, \qquad q_1 = 1 - 4W_0$$

and this specifies a probability distributon on all three scenarios iff $W_0 \in (0, 1/4)$. Since we can express every contingent claim as a linear combination of A_1, S_1, W_1, we can now price every contingent claim X_1 under the martingale probabities as $X_0 = \frac{1}{2}\mathbb{E}_q(X_1)$.

Lecture 11

Lévy markets and time-changes

11.1 Incompleteness and martingale probabilities in Lévy markets

By a Lévy market we will understand a model (S, A) of a risky asset $S_t = \exp\{X_t\}$ for a Lévy process $X = (X_t)_{t\geq 0}$ and a deterministic risk-free bank account process, usually $A_t = e^{\delta t}$, $t \geq 0$. We exclude deterministic $X_t = \mu t$ in the sequel.

Theorem 74 (No arbitrage) *A Lévy market allows arbitrage if and only if either $X_t - \delta t$ is a subordinator or $\delta t - X_t$ is a subordinator.*

Proof: We only prove that these cases lead to arbitrage opportunities. If $X_t - \delta t$ is a subordinator, then the portfolio $(1, -1)$ is an arbitrage portfolio. □

The other direction of proof is difficult, since we would need technical definitions of admissible portfolio processes and related quantities.

No arbitrage is closely related (almost equivalent) to the existence of martingale probabilities. Formally, an equivalent martingale measure $\mathbb{Q}$ is a probability measure which has the same sets of zero probability as $\mathbb{P}$, i.e. under which the same things are possible/impossible as under $\mathbb{P}$, and under which $(e^{-\delta t}S_t)_{t\geq 0}$ is a martingale. For simplicity, we will not bother about this passage to a so-called risk-neutral world that is different from the physical world. Instead, we will consider models where $(e^{-\delta t}S_t)_{t\geq 0}$ is already a martingale. Prices of the form $W_0 = e^{-\delta t}\mathbb{E}_{\mathbb{Q}}(W_t)$ are then arbitrage-free prices. The range of arbitrage-free prices is

$$\left\{e^{-\delta t}\mathbb{E}_{\mathbb{Q}}(W_t) : \mathbb{Q} \text{ martingale measure equivalent to } \mathbb{P}\right\}.$$

The proof of incompleteness is also difficult, but the result is not hard to state:

Theorem 75 (Completeness) *A Lévy market is complete if and only if $(X_t)_{t\geq 0}$ is either a multiple of Brownian motion with drift, $X_t = \mu t + \sigma B_t$ or a multiple of the Poisson process with drift, $X_t = at + bN_t$ (with $(a - \delta)b < 0$ to get no arbitrage).*

Completeness is closely related (almost equivalent) to the uniqueness of martingale probabilities. In an incomplete market, there are infinitely many choices for these martingale probabilities. This raises the question of how to make the right choice. While we can determine an arbitrage-free system of prices for all contingent claims, we cannot hedge the contingent claim, and this presents a risk to someone selling e.g. options.

11.2 Option pricing by simulation

If we are given a risk-neutral price process (martingale) $(S_t)_{t\geq 0}$, we can price contingent claims $G((S_s)_{0\leq s\leq t})$ as expectations

$$P = e^{-\delta t}\mathbb{E}(G((S_s)_{0\leq s\leq t})).$$

Often, such expectations are difficult to work out theoretically or numerically, particularly for path-dependent options such as barrier options. Monte-Carlo simulation always works, by the strong law of large numbers:

$$\frac{1}{n}\sum_{k=1}^{n} G((S_s^{(k)})_{0\leq s\leq t}) \to \mathbb{E}(G((S_s)_{0\leq s\leq t}) \qquad \text{almost surely,}$$

as $n\to\infty$, where $(S_s^{(k)})_{0\leq s\leq t}$ are independent copies of $(S_s)_{0\leq s\leq t}$. By simulating these copies, we can approximate the expectation on the right to get the price of the option.

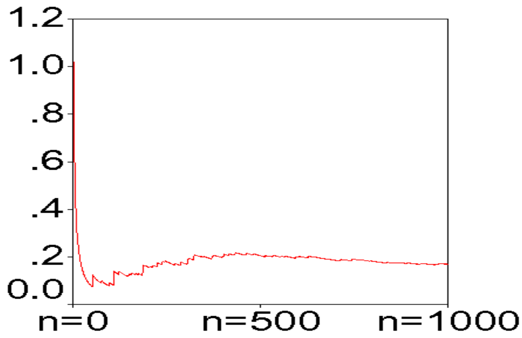


Figure 11.1: Option pricing by simulation

11.3 Time changes

Lévy markets are one way to address shortcomings of the Black-Scholes model. Particularly quantities such as one-day return distributions can be fitted well. Other possibilities include modifications to the Black-Scholes model, where the *speed* of the market is modelled separately. The rationale behind this is to capture days with increased activity (and hence larger price movements) by notions of operational versus real time. In operational time, the price process follows a Brownian motion, but in real time, a busy day corresponds to several days in operational time, while a quiet day corresponds to a fraction of a day in operational time.

The passage from operational to real time is naturally modelled by a time-change $y\mapsto\tau_y$, which we will eventually model by a stochastic process built from a Poisson point process. The price process is then $(B_{\tau_y})_{y\geq 0}$. This stochastic process cannot be observed directly in practice, but approximations of quadratic variation permit to estimate the time change.

The most elementary time change is for $\tau_y = f(y)$, a deterministic continuous strictly increasing function $f:[0,\infty)\to[0,\infty)$ with $f(0)=0$ and $f(\infty)=\infty$. In this case, the time-changed process $Z_y = X_{f(y)}$, $y \geq 0$, visits the same states as X in the same order as X, performing the same jumps as X, but travelling at a different speed. Specifically, if $f(y) << y$, then, by time y, the process X will have gone to X_y, but Z only to $Z_y = X_{f(y)}$. We say that Z has evolved more slowly than X, and faster if instead $f(y) >> y$. If f is differentiable, we can more appropriately make local speed statements according to whether $f'(y) < 1$ or $f'(y) > 1$. Note, however, that “speed” really is “relative speed” when comparing X and Z, since X is not “travelling at unit speed” in a sense of rate of spatial displacement; jumps and particularly unbounded variation make such notions useless. We easily calculate

$$\mathbb{E}(e^{i\lambda Z_y}) = \mathbb{E}(e^{i\lambda X_{f(y)}}) = e^{-f(t)\psi(\lambda)}, \quad \text{if } \mathbb{E}(e^{i\lambda X_t}) = e^{-t\psi(\lambda)}.$$

and see that Z is a stochastic process with independent increments and right-continuous paths with left limits, but will only have stationary increments if $f(y) = cy$ for all $y \geq 0$ and some $c \in (0, \infty)$.

Example 76 (Foreign exchange rates) Suppose that the EUR/USD-exchange rate today is S_0 and you wish to model the exchange rate $(S_t)_{t\geq 0}$ over the next couple of *days*. As a first model you might think of

$$S_t = S_0 \exp\{\sigma B_t - t\sigma^2/2\},$$

where B is a standard Brownian motion σ is a volatility parameter that measures the magnitude of variation. This magnitude is related to the amount of activity on the exchange markets and will be much higher for the EUR/USD-exchange rate than e.g. for the EUR/DKK-exchange rate (DansKe Kroner, Danish crowns are not traded so frequently in such high volumes. Also, DKK is closely aligned with EUR due to strong economic ties between Denmark and the Euro countries).

However, in practice, trading activity is not constant during the day. When stock markets in the relevant countries are open, activity is much higher than when they are all closed and a periodic function $f' : [0, \infty) \to [0, \infty)$ can explain a good deal of this variability and provide a better model

$$S_t = S_0 \exp\{\sigma B_{f(y)} - f(y)\sigma^2/2\} = S_0 \exp\{\widetilde{B}_{\widetilde{f}(y)} - \widetilde{f}(y)/2\},$$

where $\widetilde{B}_s = \sigma B_{s/\sigma^2}$, $s \geq 0$, is also a standard Brownian motion and $\widetilde{f}(y) = f(y)\sigma^2$ makes the parameter σ redundant – the flexibility for $\widetilde{f}$ retains all modelling freedom.

If we weaken the requirement of strict monotonicity to weak monotonicity and $f(y) = c$, $y \in [l, r)$, is constant on an interval, then $Z_y = X_c$, $y \in [l, r)$, during this interval. For a financial market model this can be interpreted as time intervals with no market activity, when the price will not change.

If we weaken the continuity of f to allow (upward) jumps, then $Z_y = X_{f(y)}$, $y \geq 0$, does not evaluate X everywhere. Specifically, if $\Delta f(y) > 0$ is the only jump of f, then Z will visit the same points as X in the same order until $X_{f(y-)-}$ and then skip over $(X_{f(y-)+s})_{0\leq s<\Delta f(y)}$ to directly jump to $X_{f(y)}$. In general, this is the behaviour at every jump of f.

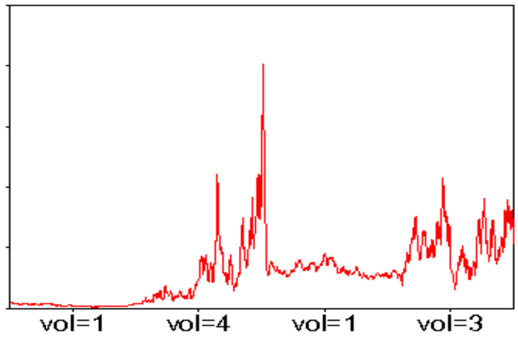


Figure 11.2: piecewise constant volatility

11.4 Quadratic variation of time-changed Brownian motion

In Section 6.1 we studied quadratic variation of Brownian motion in order to show that Brownian motion has infinite total variation (and is therefore not the difference of two increasing processes). Let us here look at quadratic variation of time-changed Brownian motion $Z_y = B_{f(y)}$ for an increasing function $f : [0, \infty) \to [0, \infty)$:

$$[Z]_t = p - \lim_{n\to\infty} [Z]_t^{(n)}, \quad \text{where } [Z]_t^{(n)} = \sum_{j=1}^{[2^n t]} (Z_{j2^{-n}} - Z_{(j-1)2^{-n}})^2$$

and $p - \lim$ denotes a limit of random variables in probability. One may expect that $[B]_t = t$ implies that $[Z]_y = f(y)$, and this is true *under suitable assumptions.*

Proposition 77 *Let B be Brownian motion and $f : [0, \infty) \to [0, \infty)$ continuous and increasing with $f(0) = 0$. Then $[Z]_y = f(y)$ for all $y \geq 0$.*

Proof: The proof (for Z) is the same as for Brownian motion (B) itself, see Section 6.1 and Assignment 6. □

Quadratic variation is accumulated locally. Under the continuity assumption of Brownian motion and its time change, it is the wiggly local behaviour of Brownian motion that generates quadratic variation. In Section 6.1 we showed that under the no-jumps assumption, positive quadratic variation implies infinite total variation. Hence, still under the no-jumps assumption, finite total variation implies zero quadratic variation. But what is the impact of jumps on quadratic variation? It can be shown as in Proposition 3 that

$$[f]_y \geq \sum_{s\leq y} |\Delta f_s|^2.$$

Example 78 Consider a piecewise linear function $f : [0, \infty) \to [0, \infty)$ with slope 0.1 and jumps $\Delta f_{2k-1} = 1.8$, $k \geq 1$. Then $f(2k) = 2k$, but by Proposition 77

$$[f]_{2k} \geq \sum_{s\leq 2k} |\Delta f_s|^2 = k(1.8)^2 = 3.24k$$

and, in fact, this is an equality, since

$$[f]^{(n)}_{2k} = k(1.8 + 2^{-n}0.1)^2 + (2^{n+1} - 1)k(2^{-n}0.1)^2 \to k(1.8)^2.$$

Now define $Z_y = B_{f(y)}$ and note that

$$\begin{aligned}
[Z]^{(n)}_{2k} &= \sum_{i=1}^{k} \Bigg(\sum_{j=1}^{2^n-1} (B_{(2i-2)+j2^{-n}0.1} - B_{(2i-2)+(j-1)2^{-n}0.1})^2 \\
&\qquad + (B_{2i-0.1} - B_{(2i-2)+0.1-2^{-n}})^2 \\
&\qquad + \sum_{j=1}^{2^n} (B_{2i-(j-1)2^{-n}0.1} - B_{2i-j2^{-n}0.1})^2 \Bigg) \\
&\to 2k0.1 + \sum_{i=1}^{k} (B_{2i-0.1} - B_{(2i-2)+0.1})^2,
\end{aligned}$$

as $n \to \infty$, which is actually $0.1(2k) + \sum_{s \le 2k} |\Delta Z_s|^2$. Note that this is a random quantity.

In general, quadratic variation consists of a continuous part due to Brownian fluctuations and the sum of squared jump sizes.

Lecture 12

Subordination and stochastic volatility

Subordination is the operation X_{τ_y}, $y \geq 0$, for a Lévy (or more general Markov) process $(X_t)_{t\geq 0}$ and a subordinator $(\tau_y)_{y\geq 0}$. One distinguishes subordination in the sense of Bochner, where X and τ are independent and subordination in the wide sense where τ_y is a stopping time for all $y \geq 0$. These are both special cases of the more general concept of *time change*, where $(\tau_y)_{y\geq 0}$ does not have to be a subordinator.

12.1 Bochner's subordination

Theorem 79 (Bochner) *Let $(X_t)_{t\geq 0}$ be a Lévy process and $(\tau_y)_{y\geq 0}$ an* independent *subordinator. Then the process $Z_y = X_{\tau_y}$, $y \geq 0$, is a Lévy process, and we have*

$$\mathbb{E}(e^{i\lambda Z_y}) = e^{-y\Phi(\psi(\lambda))}, \qquad \textit{where } \mathbb{E}(e^{i\lambda X_t}) = e^{-t\psi(\lambda)} \textit{ and } \mathbb{E}(e^{-q\tau_y}) = e^{-y\Phi(q)}.$$

Proof: First calculate by conditioning on τ_y (assuming that τ_y is continuous with probability density function f_{τ_y})

$$\begin{aligned}\mathbb{E}(\exp\{i\lambda Z_y\}) &= \mathbb{E}(\exp\{\gamma X_{\tau_y}\}) = \int_0^\infty f_{\tau_y}(t)\mathbb{E}(\exp\{i\lambda X_t\})dt \\ &= \int_0^\infty f_{\tau_y}(t)\exp\{-t\psi(\lambda)\}dt = e^{-y\Phi(\psi(\lambda))}.\end{aligned}$$

Now, for $r, s \geq 0$,

$$\begin{aligned}&\mathbb{E}(\exp\{i\lambda Z_y + i\mu(Z_{y+x} - Z_y)\}) \\ &\quad= \int_0^\infty \int_0^\infty f_{\tau_y,\tau_{y+x}-\tau_y}(t,u)\mathbb{E}(\exp\{i\lambda X_t + i\mu(X_{t+u} - X_t)\})dtdu \\ &\quad= \int_0^\infty \int_0^\infty f_{\tau_y}(t)f_{\tau_x}(u)e^{-t\psi(\lambda)}e^{-u\psi(\mu)}dtdu = e^{-y\Phi(\psi(\lambda))}e^{-x\Phi(\psi(\mu))},\end{aligned}$$

so that we deduce that Z_y and $Z_{y+x} - Z_y$ are independent, and that $Z_{y+x} - Z_y \sim Z_x$. For the right-continuity of paths, note that

$$\lim_{\varepsilon\downarrow 0} Z_{y+\varepsilon} = \lim_{\varepsilon\downarrow 0} X_{\tau_{y+\varepsilon}} = X_{\tau_y} = Z_y,$$

since $\tau_y + \delta := \tau_{y+\varepsilon} \downarrow \tau_y$ and therefore $X_{\tau_y+\delta} \to X_{\tau_y}$. For left limits, the same argument applies. □

Note that $\Delta Z_y = Z_y - Z_{y-} = X_{\tau_y} - X_{\tau_{y-}} \neq$ can be non-zero if either $\Delta\tau_y \neq 0$, or $\Delta X_{\tau_y} \neq 0$, so Z inherits jumps from τ and from X. We have, with probability 1 for all $y \geq 0$ that

$$\Delta Z_y = X_{\tau_y} - X_{\tau_{y-}-} = \begin{cases} (\Delta X)_{\tau_y} & \text{if } (\Delta X)_{\tau_y} \neq 0, \\ X_{\tau_y} - X_{\tau_{y-}} & \text{if } \Delta\tau_y \neq 0. \end{cases}$$

Note that we claim that $X_{\tau_{y-}} = X_{\tau_{y-}-}$, i.e. $(\Delta X)_{\tau_{y-}} \neq 0$ if $\Delta\tau_y \neq 0$, for all $y \geq 0$ with probability 1. This is due to the fact that the countable set of times $\{\tau_{y-}, \tau_y : y \geq 0 \text{ and } \Delta\tau_y \neq 0\}$ is a.s. disjoint from $\{t \geq 0 : \Delta X_t \neq 0\}$.

Note also that $X_{\tau_y} = X_{\tau_{y-}}$ is possible with positive probability, certainly in the case of a compound Poisson process X.

Heuristically, if X_t has density f_t and τ Lévy density g_τ, then Z will have Lévy density

$$g(z) := \int_0^\infty f_t(z) g_\tau(t) dt, \qquad z \in \mathbb{R}, \tag{1}$$

since every jump of τ of size $\Delta\tau_y = t$ leads to a jump $X_{\tau_y} - X_{\tau_{y-}} \sim X_t$, and the total intensity of jumps of size z receives contributions from τ-jumps of all sizes $t \in (0,\infty)$. We can make this precise as follows:

Proposition 80 *Let X be a Lévy process with probability density function f_t of X_t, $t \geq 0$, τ a subordinator with Lévy-Khintchine characteristics $(0, g_\tau)$, then $Z_y = X_{\tau_y}$ has Lévy-Khintchine characteristics $(0,0,g)$, where g is given by (1).*

Proof: Consider a Poisson point process $(\Delta_y)_{y\geq 0}$ with intensity function g, then, by the Exponential Formula

$$\begin{aligned}
&\mathbb{E}\left(\exp\left\{i\lambda \sum_{s\leq y} \Delta_s 1_{\{|\Delta_s|>\varepsilon\}}\right\}\right) \\
&= \exp\left\{y \int_{-\infty}^\infty (e^{i\lambda z} - 1) g(z) 1_{\{|z|>\varepsilon\}} dz\right\} \\
&= \exp\left\{y \int_0^\infty \int_{-\infty}^\infty (e^{i\lambda z} - 1) f_t(z) 1_{\{|z|>\varepsilon\}} dz g_\tau(t) dt\right\} \\
&\to \exp\left\{-y \int_0^\infty (1 - e^{-t\psi(\lambda)}) g_\tau(t) dt\right\} = \exp\left\{-y\Phi(\psi(\lambda))\right\},
\end{aligned}$$

as $\varepsilon \downarrow 0$, and this is the same distribution as we established for Z_y. □

Note that we had to prove convergence in distribution as $\varepsilon \downarrow 0$, since we have not studied integrability conditions for g. This is done on Assignment sheet 6.

Corollary 81 *If X is Brownian motion and τ has characteristics (b, g_τ), then $Z_y = X_{\tau_y}$ has characteristics $(0, b, g)$.*

Proof: Denote $\Phi_0(q) = \Phi(q) - bq$. Then the calculation in the proof of the proposition does not yield a characteristic exponent $\Phi(\psi(\lambda))$, but $\Phi_0(\psi(\lambda))$. Note that $\Phi(\psi(\lambda)) = \Phi(\frac{1}{2}\lambda^2) = \frac{1}{2}b\lambda^2 + \Phi_0(\psi(\lambda))$, so we consider

$$bB_t + \sum_{s\leq t} \Delta_s$$

for an independent Brownian motion B, which has characteristic exponent as required. □

Example 82 If we define the Variance Gamma process by subordination of Brownian motion B by a Gamma(α,θ) subordinator τ with Lévy density $g_\tau(y) = \alpha y^{-1}e^{-\theta y}$, then we obtain a Lévy density

$$g(z) = \int_0^\infty f_t(z)g_\tau(t)dt = \int_0^\infty \frac{1}{\sqrt{2\pi t}}e^{-z^2/(2t)}\alpha t^{-1}e^{-\theta t}dt$$

and we can calculate this integral to get

$$g(z) = \alpha|z|^{-1}e^{-\sqrt{2\theta}|z|}$$

as Lévy density. The Variance Gamma process as subordinated Brownian motion has an interesting interpretation when modelling financial price processes. In fact, the stock price is then considered to evolve according to the Black-Scholes model, but not in real time, but in *operational time* τ_y, $y \geq 0$. Time evolves as a Gamma process, with infinitely many jumps in any bounded interval.

Note that all Lévy processes that we can construct as subordinated Brownian motions are symmetric. However, not all symmetric Lévy processes are subordinated Brownian motions,

12.2 Ornstein-Uhlenbeck processes

Example 83 (Gamma-OU process) Let $(N_t)_{t\geq 0}$ be a Poisson process with intensity $a\lambda$ and jump times $(T_k)_{k\geq 1}$, $(X_n)_{n\geq 1}$ a sequence of independent Gamma$(1,b)$ random variables, $Y_0 \sim$ Gamma(a,b), consider the stochastic process

$$Y_t = Y_0 e^{-\lambda t} + \sum_{k=1}^{N_t} X_n e^{-\lambda(t-T_k)}$$

We use this model for the speed of the market and think of an initial speed of Y_0 which slows down exponentially, but at times of a Poisson process, events occur that make the speed jump up at times T_k, $k \geq 0$. Each of these also slow down exponentially. In fact, there is a strong equilibrium in that

$$\begin{aligned}
\mathbb{E}(e^{-qY_t}) &= \mathbb{E}(e^{-qe^{-\lambda t}Y_0})\mathbb{E}\left(\exp\left\{-q\sum_{k=1}^{N_t} X_k e^{-\lambda(t-T_k)}\right\}\right) \\
&= \left(\frac{b}{b+qe^{-\lambda t}}\right)^a \sum_{n=0}^\infty \frac{(\lambda a t)^n}{n!}e^{-\lambda a t}\left(\int_0^t \frac{1}{t}\frac{b}{b+qe^{-\lambda s}}ds\right)^n \\
&= \left(\frac{b}{b+qe^{-\lambda t}}\right)^a \left(\frac{b+qe^{-\lambda t}}{b+q}\right)^a = \left(\frac{b}{b+q}\right)^a,
\end{aligned}$$

so Y_t has the same distribution as Y_0. In fact, $(Y_t)_{t\geq 0}$ is a stationary Markov process. The process Y is called a Gamma-OU process, since it has the Gamma distribution as its stationary distribution. The time change process

$$\tau_y = \int_0^y Y_s ds$$

associated with speed Y is called integrated Ornstein-Uhlenbeck process. Note that the stationarity of Y implies that τ has stationary increments, but note that τ does not have independent increments. The associated stochastic volatility model is now the time-changed Brownian motion $(B_{\tau_y})_{y\geq 0}$.

In general, we can define Ornstein-Uhlenbeck processes associated with any subordinator Z or rather its Poisson point process $(\Delta Z_t)_{t\geq 0}$ of jumps as

$$Y_t = Y_0 e^{-\lambda t} + \sum_{s\leq t} \Delta Z_s e^{-\lambda(t-s)},$$

for any initial distribution for Y_0, but a stationary distribution can also be found.

We can always associate a stochastic volatility model $(B_{\tau_y})_{y\geq 0}$ using the integrated volatility $\tau_y = \int_y Y_s ds$ as time change. Note that, by the discussion of the last section, we can actually infer the time change from sums of squared increments for a small time lag 2^{-n}, even though the actual time change is not observed. In practice, the so-called market microstructure (piecewise constant prices) destroys model fit for small times, so we need to choose a moderately small 2^{-n}. In practice, 5 minutes is a good choice.

12.3 Simulation by subordination

Note that we can simulate subordinators using simulation Method 1 (Time discretisation) or Method 2 (Throwing away the small jumps). The latter consisted essentially in simulating the Poisson point process of jumps of the subordinator. Clearly, we can apply this method also to simulate an Ornstein-Uhlenbeck process.

Method 3 (Subordination) Let $(\tau_y)_{y\geq 0}$ be an increasing process that we can simulate, and let $(X_t)_{t\geq 0}$ be a Lévy process with cumulative distribution function F_t of X_t. Fix a time lag $\delta > 0$. Then the process

$$Z_y^{(3,\delta)} = S_{[y/\delta]}, \qquad \text{where } S_n = \sum_{k=1}^n A_k \text{ and } A_k = F^{-1}_{\tau_{k\delta}-\tau_{(k-1)\delta}}(U_k)$$

is the time discretisation of the subordinated process $Z_y = X_{\tau_y}$.

Example 84 We can use Method 3 to simulate the Variance Gamma process, since we can simulate the Gamma process τ and we can simulate the A_k. Actually, we can use the Box-Muller method to generate standard Normal random variables N_k and then use

$$\widetilde{A}_k \sim \sqrt{\tau_{k\delta} - \tau_{(k-1)\delta}} N_k, \qquad k \geq 1,$$

instead of A_k, $k \geq 1$.

Lecture 13

Level passage problems

Reading: Kyprianou Sections 3.1 and 3.3

13.1 The strong Markov property

Recall that a stopping time is a random time $T \in [0, \infty]$ such that for every $s \geq 0$ the information $\mathcal{F}_s$ up to time s allows to decide whether $T \leq s$. More formally, if the event $\{T \leq s\}$ can be expressed in terms of $(X_r, r \leq s)$ (is measurable with respect to $\mathcal{F}_s = \sigma(X_r, r \leq s)$). The prime example of a stopping time is the first entrance time $T_I = \inf\{t \geq 0 : X_t \in I\}$ into a set $I \subset \mathbb{R}$. Note that

$$\{T \leq s\} = \{\text{there is } r \leq s \text{ such that } X_r \in I\}$$

(for *open* sets I we can drop the irrational $r \leq s$ to show measurability.).

We also denote the information $\mathcal{F}_T$ up to time T. More formally,

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leq s\} \in \mathcal{F}_s \text{ for all } s \geq 0\},$$

i.e. $\mathcal{F}_T$ contains those events that, if $T \leq s$, can be expressed in terms of $(X_r, r \leq s)$, for all $s \geq 0$.

Recall the simple Markov property which we can now state as follows. For a Lévy process $(X_t)_{t\geq 0}$ and a fixed time t, the post-t process $(X_{t+s} - X_t)_{s\geq 0}$ has the same distribution as X and is independent of the pre-t information $\mathcal{F}_t$.

Theorem 85 (Strong Markov property) *Let $(X_t)_{t\geq 0}$ be a Lévy process and T a stopping time. Then given $T < \infty$, the post-T process $(X_{T+s} - X_T)_{s\geq 0}$ has the same distribution as X and is independent of the pre-T information $\mathcal{F}_T$.*

Proof: Let $0 < s_1 < \ldots < s_m$, $C_1, \ldots, C_m \subset \mathbb{R}$ closed, $A \in \mathcal{F}_T$. Then we need to show that

$$\begin{aligned}&\mathbb{P}(A, T < \infty, X_{T+s_1} - X_T \leq c_1, \ldots, X_{T+s_m} - X_T \leq c_m)\\&= \mathbb{P}(A, T < \infty)\mathbb{P}(X_{s_1} \leq c_1, \ldots, X_{s_m} \leq c_m).\end{aligned}$$

First define stopping times $T_n = 2^{-n}([2^n T] + 1)$, $n \geq 1$, that only take countably many values. These are the next dyadic rationals *after* time T. Note that $T_n \downarrow T$ as $n \to \infty$. Now note that $A \cap \{T_n = k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$ and the simple Markov property yields

$$\begin{aligned}
&\mathbb{P}(A, T_n < \infty, X_{T_n+s_1} - X_{T_n} \leq c_1, \ldots, X_{T_n+s_m} - X_{T_n} \leq c_m) \\
&\quad = \sum_{k=0}^{\infty} \mathbb{P}(A, T_n = k2^{-n}, X_{k2^{-n}+s_1} - X_{k2^{-n}} \leq c_1, \ldots, X_{k2^{-n}+s_m} - X_{k2^{-n}} \leq c_m) \\
&\quad = \sum_{k=0}^{\infty} \mathbb{P}(A, T_n = k2^{-n}) \mathbb{P}(X_{s_1} \leq c_1, \ldots, X_{s_m} \leq c_m) \\
&\quad = \mathbb{P}(A, T_n < \infty) \mathbb{P}(X_{s_1} \leq c_1, \ldots, X_{s_m} \leq c_m).
\end{aligned}$$

Now the right-continuity of sample paths ensures $X_{T_n+s_j} \to X_{T+s_j}$ as $n \to \infty$ and we conclude

$$\begin{aligned}
&\mathbb{P}(A, T < \infty, X_{T+s_1} - X_T \leq c_1, \ldots, X_{T+s_m} - X_T \leq c_m) \\
&\quad = \lim_{n\to\infty} \mathbb{P}(A, T_n < \infty, X_{T_n+s_1} - X_{T_n} \leq c_1, \ldots, X_{T_n+s_m} - X_{T_n} \leq c_m) \\
&\quad = \lim_{n\to\infty} \mathbb{P}(A, T_n < \infty) \mathbb{P}(X_{s_1} \leq c_1, \ldots, X_{s_m} \leq c_m) \\
&\quad = \mathbb{P}(A, T < \infty) \mathbb{P}(X_{s_1} \leq c_1, \ldots, X_{s_m} \leq c_m),
\end{aligned}$$

for all $(c_1, \ldots, c_m)$ such that $\mathbb{P}(X_{T+s_j} - X_T = c_j) = 0$, $j = 1, \ldots, m$. Finally note that $(X_{T+s} - X_T)_{s\geq 0}$ clearly has right-continuous paths with left limits. □

13.2 The supremum process

Let $X = (X_t)_{t\geq 0}$ be a Lévy process. We denote its supremum process by

$$\overline{X}_t = \sup_{0\leq s\leq t} X_s, \qquad t \geq 0.$$

We are interested in the joint distribution of $(X_t, \overline{X}_t)$, e.g. for the payoff of a barrier or lookback option. Moment generating functions are easier to calculate and can be numerically inverted. We can also take such a transform over the time variable, e.g.

$$q \mapsto \int_0^\infty e^{-qt} \mathbb{E}(e^{\gamma X_t}) dt = \frac{1}{q - \Psi(\gamma)} \quad \text{uniquely identifies } \mathbb{E}(e^{\gamma X_t}),$$

and the distribution of X_t. But $q \int_0^\infty e^{-qt} \mathbb{E}(e^{\gamma X_t}) dt = \mathbb{E}(e^{\gamma X_\tau})$ for $\tau \sim \text{Exp}(q)$.

Proposition 86 (Independence) *Let X be a Lévy process, $\tau \sim \text{Exp}(q)$ an independent random time. Then $\overline{X}_\tau$ is independent of $\overline{X}_\tau - X_\tau$.*

Proof: We only prove the case where $G_1 = \inf\{t > 0 : X_t > 0\}$ satisfies $\mathbb{P}(G_1 > 0) = 1$. In this case we can define successive record times $G_n = \inf\{t > G_{n-1} : X_t > \overline{X}_{G_{n-1}}\}$, $n \geq 2$, and also set $G_0 = 0$. Note that, by the strong Markov property at the stopping times G_n we have that $X_{G_n} > \overline{X}_{G_{n-1}}$ (otherwise the post-G_{n-1} process $\widetilde{X}_t = X_{G_{n-1}+t} - X_{G_{n-1}}$ would have the property $\widetilde{G}_1 = 0$, but the strong Markov property yields $\mathbb{P}(\widetilde{G}_1 > 0) =$

$\mathbb{P}(G_1 > 0) = 1$). So X can only reach new records by upward jumps, $\overline{X}_\tau \in \{X_{G_n}, n \geq 0\}$ and more specifically, we will have $\overline{X}_\tau = G_n$ if and only if $G_n \leq \tau < G_{n+1}$ so that

$$\begin{aligned}
\mathbb{E}(e^{\beta\overline{X}_\tau+\gamma(\overline{X}_\tau-X_\tau)}) &= \int_0^\infty qe^{-qt}\mathbb{E}(e^{\beta\overline{X}_t+\gamma(\overline{X}_t-X_t)})dt \\
&= q\mathbb{E}\left(\sum_{n=0}^\infty \int_{G_n}^{G_{n+1}} e^{-qt}e^{\beta\overline{X}_t+\gamma(\overline{X}_t-X_t)}dt\right) \\
&= q\sum_{n=0}^\infty \mathbb{E}\left(e^{\beta X_{G_n}}e^{-qG_n}\int_0^{G_{n+1}-G_n} e^{-qs}e^{-\gamma(X_{G_n+s}-X_{G_n})}ds\right) \\
&= q\sum_{n=0}^\infty \mathbb{E}\left(e^{-qG_n+\beta X_{G_n}}\right)\mathbb{E}\left(\int_0^{\widetilde{G}_1} e^{-qs-\gamma\widetilde{X}_s}ds\right)
\end{aligned}$$

where we applied the strong Markov property at G_n to split the expectation in the last row – note that $\int_0^{G_{n+1}-G_n} e^{-qs-\gamma(X_{G_n+s}-X_{G_n})}ds$ is a function of the post-G_n process, whereas $e^{-qG_n+\beta X_{G_n}}$ is a function of the pre-G_n process, and the expectation of the product of independent random variables is the product of their expectations.

This completes the proof, since the last row is a product of a function of β and a function of γ, which is enough to conclude. More explicitly, we can put $\beta = 0$, $\gamma = 0$ and $\beta = \gamma = 0$, respectively, to see that indeed the required identity holds:

$$\mathbb{E}(e^{\beta\overline{X}_\tau+\gamma(\overline{X}_\tau-X_\tau)}) = \mathbb{E}(e^{\beta\overline{X}_\tau})\mathbb{E}(e^{\gamma(\overline{X}_\tau-X_\tau)}).$$

□

13.3 Lévy processes with no positive jumps

Consider stopping times $T_x = \inf\{t \geq 0 : X_t \in (x, \infty)\}$, so-called first passage times. For Lévy processes with no positive jumps, we must have $X_{T_x} = x$, provided that $T_x < \infty$. This observation allows to calculate the moment generating function of T_x. To prepare this result, recall that the distribution of X_t has moment generating function

$$\mathbb{E}(e^{\gamma X_t}) = e^{t\Psi(\gamma)}, \quad \Psi(\gamma) = a_1\gamma + \frac{1}{2}\sigma^2\gamma^2 + \int_{-\infty}^0 (e^{\gamma x} - 1 - \gamma x 1_{\{|x|\leq 1\}})g(x)dx.$$

Let us exclude the case where $-X$ is a subordinator, i.e. where $\sigma^2 = 0$ and $a_1 - \int_{-1}^0 xg(x)dx \leq 0$, since in that case $T_x = \infty$. Then note that

$$\Psi''(\gamma) = \sigma^2 + \int_{-\infty}^0 x^2 e^{\gamma x}g(x)dx > 0,$$

so that Ψ is convex and hence has at most two zeros, one of which is $\Psi(0) = 0$. There is a second zero $\gamma_0 > 0$ if and only if $\Psi'(0) = \mathbb{E}(X_1) < 0$, since we excluded the case where $-X$ is a subordinator, and $\mathbb{P}(X_t > 0) > 0$ implies that $\Psi(\infty) = \infty$.

Theorem 87 (Level passage) *Let $(X_t)_{t\geq 0}$ be a Lévy process with no positive jumps and T_x the first passage time across level x. Then*

$$\mathbb{E}(e^{-qT_x}1_{\{T_x<\infty\}}) = e^{-x\Phi(q)},$$

where $\Phi(q)$ is the largest γ for which $\Psi(\gamma) = q$.

Proof: We only prove this for the case where $\mathbb{P}(T_x < \infty) = 1$, i.e. $\mathbb{E}(X_1) \geq 0$ and $\gamma_0 = 0$. By Exercise A.2.3.(a) the processes $M_t = e^{\gamma X_t - t\Psi(\gamma)}$ are martingales. We will apply the Optional stopping theorem to T_x. Note that $\mathbb{E}(M_t^2) = e^{t(\Psi(2\gamma)-2\Psi(\gamma))}$ is not such that $\sup_{t\geq 0}\mathbb{E}(M_t^2) < \infty$. However, if we put

$$M_t^{(u)} = M_t \quad \text{if } t \leq u \quad \text{and} \quad M_t^{(u)} = M_u \quad \text{if } t \geq u,$$

then $(M_t^u)_{t\geq 0}$ is a martingale which satisfies $\sup_{t\geq 0}\mathbb{E}((M_t^{(u)})^2) < \infty$. Also, $T_x \wedge u$ is a stopping time, so that for $\gamma \geq \gamma_0 = 0$ (so that $\Psi(\gamma) \geq 0$)

$$1 = \mathbb{E}(M_{T_x\wedge u}^{(u)}) = \mathbb{E}(M_{T_x\wedge u}) \to \mathbb{E}(M_{T_x}) = \mathbb{E}(e^{\gamma x - \Psi(\gamma)T_x}), \quad \text{as } u \to \infty,$$

by dominated convergence ($M_{T_x\wedge u} \leq \exp\{\gamma x - \Psi(\gamma)T_x\} \leq \exp\{\gamma x\}$). We now conclude that

$$\mathbb{E}(e^{-\Psi(\gamma)T_x}) = e^{-\gamma x}$$

which for $q = \Psi(\gamma)$ and $\Phi(q)$ the unique $\gamma \geq \gamma_0 = 0$ with $\Psi(\gamma) = q$. □

Corollary 88 *Let X be a Lévy process with no positive jumps and $\tau \sim \mathrm{Exp}(q)$ independent. Then $\overline{X}_\tau \sim \mathrm{Exp}(\Phi(q))$.*

Proof: $\mathbb{P}(\overline{X}_\tau > x) = \mathbb{P}(T_x \leq \tau) = \int_0^\infty \mathbb{P}(\tau \geq t) f_{T_x}(t)dt = \mathbb{E}(e^{-qT_x}) = e^{-\Phi(q)x}$. □

If we combine this with the Independence Theorem of the previous section we obtain.

Corollary 89 *Let X be a Lévy process with no positive jumps and $\tau \sim \mathrm{Exp}(q)$ independent. Then*

$$\mathbb{E}(e^{-\beta(\overline{X}_\tau - X_\tau)}) = \frac{q(\Phi(q)-\beta)}{\Phi(q)(q-\Psi(\beta))}$$

Proof: Note that we have from the Independence Theorem that

$$\mathbb{E}(e^{\beta\overline{X}_\tau})\mathbb{E}(e^{-\beta(\overline{X}_\tau - X_\tau)}) = \mathbb{E}(e^{\beta X_\tau}) = \int_0^\infty qe^{-qt}\mathbb{E}(e^{\beta X_t})dt = \frac{q}{q-\Psi(\beta)}$$

and from the preceding corollary

$$\mathbb{E}(e^{\beta\overline{X}_\tau}) = \frac{\Phi(q)}{\Phi(q)-\beta} \qquad \text{and so} \qquad \mathbb{E}(e^{-\beta(\overline{X}_\tau - X_\tau)}) = \frac{q}{q-\Psi(\beta)}\frac{\Phi(q)-\beta}{\Phi(q)}.$$

□

13.4 Application: insurance ruin

Proposition 86 splits the Lévy process at its supremum into two increments. If you turn the picture of a Lévy process by 180°, this split occurs at the infimum, and it can be shown (Exercise A.7.1) that $\underline{X}_\tau \sim X_\tau - \overline{X}_\tau$. Therefore, Corollary 89 gives $\mathbb{E}(e^{\beta\underline{X}_\tau})$, also for $q \downarrow 0$ if $\mathbb{E}(X_1) > 0$, since then

$$\mathbb{E}(e^{\beta\underline{X}_\infty}) = \lim_{q\downarrow 0}\frac{q(\Phi(q)-\beta)}{\Phi(q)(q-\Psi(\beta))} = \frac{\beta\mathbb{E}(X_1)}{\Psi(\beta)}$$

since $\Phi'(0) = 1/\Psi'(0) = 1/\mathbb{E}(X_1)$ and note that for an insurance reserve process $R_t = u + X_t$, the probability of ruin is $r(u) = \mathbb{P}(\underline{X}_\infty < -u)$, the distribution funcion of $\underline{X}_\infty$ which is uniquely identified by $\mathbb{E}(e^{\beta\underline{X}_\infty})$.

Lecture 14

Ladder times and storage models

Reading: Kyprianou Sections 1.3.2 and 3.3

14.1 Case 1: No positive jumps

In Theorem 87 we derived the moment generating function of $T_x = \inf\{t \geq 0 : X_t > x\}$ for any Lévy process with no positive jumps. We also indicated the complication that $T_x = \infty$ is a possibility, in general. Let us study this is in more detail in our standard setting

$$\mathbb{E}(e^{\gamma X_t}) = e^{t\Psi(\gamma)}, \quad \Psi(\gamma) = a_1\gamma + \frac{1}{2}\sigma^2\gamma^2 + \int_{-\infty}^{0}(e^{\gamma x} - 1 - \gamma x 1_{\{|x|\leq 1\}})g(x)dx.$$

The important quantity is

$$\mathbb{E}(X_1) = \frac{\partial}{\partial\gamma}\mathbb{E}(e^{\gamma X_t})\bigg|_{\gamma=0} = \Psi'(0) = a_1 - \int_{-1}^{0} xg(x)dx.$$

The formula that we derived was

$$\mathbb{E}(e^{-qT_x}1_{\{T_x<\infty\}}) = e^{-x\Phi(q)}$$

where for $q > 0$, $\Phi(q) > 0$ is unique with $\Psi(\Phi(q)) = q$. Letting $q \downarrow 0$, we see that

$$\mathbb{P}(T_x < \infty) = \lim_{q\downarrow 0}\mathbb{E}(e^{-qT_x}1_{\{T_x<\infty\}}) = e^{-x\Phi(0+)}.$$

Here the convexity of Ψ that we derived last time implies that $\Phi(0+) = 0$ if and only if $\mathbb{E}(X_1) = \phi'(0) \geq 0$. Therefore, $\mathbb{P}(T_x < \infty) = 1$ if and only if $\mathbb{E}(X_1) \geq 0$.

Part of this could also be deduced by the strong (or weak) law of large numbers. Applied to increments $Y_k = X_{k\delta} - X_{(k-1)\delta}$ it implies that

$$\frac{X_{n\delta}}{n\delta} = \frac{1}{\delta}\frac{1}{n}\sum_{k=1}^{n} Y_k \to \frac{1}{\delta}\mathbb{E}(Y_1) = \frac{1}{\delta}\mathbb{E}(X_\delta) = \mathbb{E}(X_1),$$

almost surely (or in probability) as $n \to \infty$. We can slightly improve this result to a convergence as $t \to \infty$ as follows

$$\mathbb{E}(e^{\gamma X_t/t}) = e^{t\phi(\gamma/t)} \to e^{\gamma\phi'(0)} = e^{\gamma\mathbb{E}(X_1)} \quad \Rightarrow \quad \frac{X_t}{t} \to \mathbb{E}(X_1),$$

in probability. We used here that $Z_t \to a$ in distribution implies $Z_t \to a$ in probability, which holds only because a is a constant, not a random variable. Note that indeed for all $\varepsilon > 0$, as $t \to \infty$,

$$\mathbb{P}(|Z_t - a| > \varepsilon) \le \mathbb{P}(Z_t \le a - \varepsilon) + 1 - \mathbb{P}(Z_t \le a + \varepsilon) \to 0 + 1 - 1 = 0.$$

From this, we easily deduce that $X_t \to \pm\infty$ (in probability) if $\mathbb{E}(X_t) \ne 0$, but the case $\mathbb{E}(X_t) = 0$ is not so clear from this method. In fact, it can be shown that all convergences hold in the almost sure sense, here.

By an application of the Strong Markov property we can show the following.

Proposition 90 *The process $(T_x)_{x\ge 0}$ is a subordinator.*

Proof: Let us here just prove that $T_{x+y} - T_x$ is independent of T_x and has the same distribution as T_y. The remainder is left as an exercise.

Note first that $X_{T_x} = x$, since there are no positive jumps. The Strong Markov property at T_x can therefore be stated as $\widetilde{X} = (X_{T_x+s} - x)_{s\ge 0}$ is independent of $\mathcal{F}_{T_x}$ and has the same distribution as X. Now note that

$$\begin{aligned} T_x + \widetilde{T}_y &= T_x + \inf\{s \ge 0 : \widetilde{X}_s > y\} = T_x + \inf\{s \ge 0 : X_{T_x+s} > x + y\} \\ &= \inf\{t \ge 0 : X_t > x + y\} = T_{x+y} \end{aligned}$$

so that $T_{x+y} - T_x = \widetilde{T}_y$, and we obtain

$$\mathbb{P}(T_x \le s, T_{x+y} - T_x \le t) = \mathbb{P}(T_x \le s, \widetilde{T}_y \le t) = \mathbb{P}(T_x \le s)\mathbb{P}(T_y \le t),$$

since $\{T_x \le s\} \in \mathcal{F}_{T_x}$. Formally, $\{T_x \le s\} \cap \{T_x \le r\} = \{T_x \le s \wedge r\} \in \mathcal{F}_r$ for all $r \ge 0$ since T_x is a stopping time. □

We can understand what the jumps of $(T_x)_{x\ge 0}$ are: in fact, X can be split into its supremum process $\overline{X}_t = \sup_{0\le s\le t} X_s$ and the bits of path below the supremum. Roughly, the times

$$\overline{\{T_x, x \ge 0\}} = \overline{\{t \ge 0 : X_t = \overline{X}_t\}}$$

are the times when the supremum increases. $T_x - T_{x-} > 0$ if the supremum process remains constant at height x for an amount of time $T_x - T_{x-}$. The process $(T_x)_{x\ge 0}$ is called "ladder time process". The process $(X_{T_x})_{x\ge 0}$ is called ladder height process. In this case, $X_{T_x} = x$ is not very illuminating. Note that $(T_x, X_{T_x})_{x\ge 0}$ is a bivariate Lévy process.

Example 91 (Storage models) Consider a Lévy process of bounded variation, represented as $A_t - B_t$ for two subordinators A and B. We interpret A_t as the amount of work arriving in $[0, t]$ and B_t as the amount of work that can *potentially* exit from the system. Let us focus on the case where A is a compound Poisson process and $B_t = t$ for a continuously working processor. The quantity of interest is the amount W_t of work waiting to be carried out and requiring storage, where $W_0 = w \ge 0$ is an initial amount of work stored.

Note that $W_t \neq w + A_t - B_t$, in general, since $w + A_t - B_t$ can become negative, whereas $W_t \geq 0$. In fact, we can describe as follows: if the storage is empty, then no work exits from the system. Can we see from $A_t - B_t$ when the storage will be empty? We can express the first time it becomes empty and the first time it is refilled thereafter as

$$L_1 = \inf\{t \geq 0 : w + A_t - B_t = 0\} \quad \text{and} \quad R_1 = \inf\{t \geq L_1 : \Delta A_t > 0\}.$$

On $[L_1, R_1]$, $X_t = B_t - A_t$ increases linearly at unit rate from w to $w + (R_1 - L_1)$, whereas W remains constant equal to zero. In fact,

$$W_t = w - X_t + \int_{L_1 \wedge t}^{t} 1 ds = w - X_t + \int_0^t 1_{\{X_s = \overline{X}_s \geq w\}} ds = (w \vee \overline{X}_t) - X_t, \quad 0 \leq t \leq R_1.$$

An induction now shows that the system is idle if and only if $X_t = \overline{X}_t \geq w$, so that $W_t = X_t + (w \vee \overline{X}_t)$ for all $t \geq 0$.

In this context, $(\overline{X}_t - w)^+$ is the amount of time the system was idle before time t, and $T_x = \inf\{t \geq 0 : X_t > x\}$ is the time by which the system has accumulated time $x - w$ in the idle state, $x \geq w$, and we see that $(x-w)/T_x \sim x/T_x \to 1/\mathbb{E}(T_1) = 1/\Phi'(0) = \phi'(0) = \mathbb{E}(X_1) = 1 - \mathbb{E}(A_1)$ in probability, if $\mathbb{E}(A_1) \leq 1$.

Example 92 (Dams) Suppose that the storage model refers more particularly to a dam that releases a steady stream of water at a constant intensity a_2. Water is added according to a subordinator $(A_t)_{t \geq 0}$. The dam will have a maximal capacity of, say, $M > 0$. Given an initial water level of $w \geq 0$, the water level at time t is, as before

$$W_t = (w \vee \overline{X}_t) - X_t, \qquad \text{where } X_t = a_2 t - A_t.$$

The time $F = \inf\{t \geq 0 : W_t > M\}$, the first time when the dam overflows, is a quantity of interest. We do not pursue this any further theoretically, but note that this can be simulated, since we can simulate X and hence W.

14.2 Case 2: Union of intervals as ladder time set

Proposition 93 *If $X_t = a_2 t - C_t$ for a compound Poisson process $(C_t)_{t \geq 0}$ and a drift $a_2 \geq 0 \vee \mathbb{E}(C_1)$, then the ladder time set is a collection of intervals. More precisely, $\{t \geq 0 : X_t = \overline{X}_t\}$ is the range $\overline{\{\sigma_y, y \geq 0\}}$ of a compound Poisson subordinator with positive drift coefficient.*

Proof: Define $L_0 = 0$ and then for $n \geq 0$ stopping times

$$R_n = \inf\{t \geq L_n : \Delta C_t > 0\}, \quad L_{n+1} = \inf\{t \geq R_n : X_t = \overline{X}_t\}.$$

The strong Markov property at these stopping times show that $(R_n - L_n)_{n \geq 0}$ is a sequence of Exp(λ) random variables where $\lambda = \int_0^\infty g(x) dx$ is the intensity of positive jumps, and $(L_n - R_{n-1})_{n \geq 1}$ is a sequence of independent identically distributed random variables. Now define $T_n = R_0 - L_0 + \ldots + R_{n-1} - L_{n-1}$ and $(\sigma_y)_{y \geq 0}$ to be the compount Poisson process with unit drift, jump times T_n, $n \geq 1$, and jump heights $L_n - R_{n-1}$, $n \geq 1$. □

The ladder height process $(X_{\sigma_y})_{y\geq 0}$ will then also have a positive drift coefficient and share some jump times with σ (whenever X jumps from below $\overline{X}_{t-}$ above $\overline{X}_{t-}$, but have extra jump times when X jumps from $\overline{X}_{t-}$ upwards, some jump times of $(\sigma_y)_{y\geq 0}$ are not jump times of $(X_{\sigma_y})_{y\geq 0}$ – if X reaches $\overline{X}_{t-}$ again without a jump.)

Example 94 (Storage models) In the context of the previous examples, consider more general subordinators B_t with unit drift. Interpret jumps of B as unfinished work *potentially* exiting to be carried out elsewhere. We should be explicit and make a convention that if the current storage amount W_t is not sufficient for a jump of B, then all remaining work exits. With this convention, the amount W_t waiting to be carried out is still

$$W_t = w - X_t + \int_0^t 1_{\{X_s = \overline{X}_s \geq w\}} ds \qquad t \geq 0,$$

but note that the latter integral cannot be expressed in terms of $\overline{X}_t$ so easily, but σ_y is still the time by which the system has accumulated time $y - w$ in the idle state, for $y \geq w$, so $I_t = \inf\{y \geq 0 : \sigma_y > t\}$ is the amount of idle time before t.

14.3 Case 3: Discrete ladder time set

If $X_t = a_2 t - C_t$ for a compound Poisson process (or indeed bounded variation pure jump process) $(C_t)_{t\geq 0}$ and a drift $a_2 < 0$, then the ladder time set is discrete. We can still think of $\overline{\{t \geq 0 : X_t = \overline{X}_t\}}$ as the range $\overline{\{\sigma_y, y \geq 0\}}$ of a compound Poisson subordinator with zero drift coefficient. More naturally, we would define successive ladder times $G_0 = 0$ and $G_{n+1} = \inf\{t > G_n : X_t = \overline{X}_t\}$. By the strong Markov property, $G_{n+1} - G_n$, $n \geq 0$, is a sequence of independent and identically distributed random variables, and for any intensity $\lambda > 0$, we can specify $(\sigma_y)_{y\geq 0}$ to be a compound Poisson process with rate $\lambda > 0$ and jump sizes $G_{n+1} - G_n$, $n \geq 0$.

Note that $(\sigma_y)_{y\geq 0}$ is not unique since we have to choose λ. In fact, once a choice has been made and $q > 0$, we have $\{\sigma_y : y \geq 0\} = \{\sigma_{qy}, y \geq 0\}$, not just here, but also in Cases 1 and 2. In Cases 1 and 2, however, we identified a natural choice (of q) in each case.

14.4 Case 4: non-discrete ladder time set and positive jumps

The general case is much harder. It turns out that we can still express

$$\overline{\{t \geq 0 : X_t = \overline{X}_t\}} = \overline{\{\sigma_y : y \geq 0\}}$$

for a subordinator $(\sigma_y)_{y\geq 0}$, but, as in Case 3, there is no natural way to choose this process. It can be shown that the bivariate process $(\sigma_y, X_{\sigma_y})_{y\geq 0}$ is a bivariate subordinator in this general setting, called the ladder process. There are descriptions of its distribution and relations between these processes of increasing ladder events and analogous processes of decreasing ladder events.

Lecture 15

Branching processes

Reading: Kyprianou Section 1.3.4

15.1 Galton-Watson processes

Let $\xi = (\xi_k)_{k\geq 0}$ be (the probability mass function of) an offspring distribution. Consider a population model where each individual gives birth to independent and identically distributed numbers of children, starting from $Z_0 = 1$ individual, the common ancestor. Then the $(n+1)$st generation Z_{n+1} consists of the sum of numbers of children $N_{n,1}, \ldots, N_{n,Z_n}$ of the nth generation:

$$Z_{n+1} = \sum_{i=1}^{Z_n} N_{n,i}, \qquad \text{where } N_{n,i} \sim \xi \text{ independent, } i \geq 1,\ n \geq 0.$$

Proposition 95 *Let ξ be an offspring distribution, $g(s) = \sum_{k\geq 0} \xi_k s^k$ its generating function, then*

$$\mathbb{E}(s^{Z_1}) = g(s),\ \mathbb{E}(s^{Z_2}) = g(g(s)), \ldots, \mathbb{E}(s^{Z_n}) = g^{\circ(n)}(s),$$

where $g^{\circ(0)}(s) = s$, $g^{\circ(n+1)}(s) = g^{\circ(n)}(g(s))$, $n \geq 0$.

Proof: The result is clearly true for $n = 0$ and $n = 1$. Now note that

$$\begin{aligned}\mathbb{E}(s^{X_{n+1}}) &= \mathbb{E}\left(s^{\sum_{i=1}^{Z_n} N_{n,i}}\right) = \sum_{j=0}^{\infty} \mathbb{P}(Z_n = j)\mathbb{E}\left(s^{\sum_{i=1}^{j} N_{n,i}}\right) \\ &= \sum_{j=0}^{\infty} \mathbb{P}(Z_n = j)(g(s))^j = \mathbb{E}((g(s))^{Z_n}) = g^{\circ(n)}(g(s)).\end{aligned}$$

□

Proposition 96 *$(Z_n)_{n\geq 0}$ is a Markov chain whose transition probabilities are given by*

$$p_{ij} = \mathbb{P}(N_1 + \ldots + N_i = j), \qquad \text{where } N_1, \ldots, N_i \sim \xi \text{ independent.}$$

In particular, if $(Z_n^{(1)})_{n\geq 0}$ and $(Z_n^{(2)})_{n\geq 0}$ are two independent Markov chains with transition probabilities $(p_{ij})_{i,j\geq 0}$ starting from population sizes k and l, respectively, then $Z_n^{(1)} + Z_n^{(2)}$, $n \geq 0$, is also a Markov chain with transition probabilities $(p_{ij})_{i,j\geq 0}$ starting from $k + l$.

Proof: Just note that the independence of $(N_{n,i})_{i\geq 1}$ and $(N_{k,i})_{0\leq k\leq n-1,i\geq 1}$ implies that

$$\begin{aligned}\mathbb{P}(Z_{n+1}=j|Z_0=i_0,\ldots,Z_{n-1}=i_{n-1},Z_n=i_n) &= \mathbb{P}(N_{n,1}+\ldots,N_{n,i_n}=j|Z_0=i_0,\ldots,Z_n=i_n)\\ &= \mathbb{P}(N_{n,1}+\ldots,N_{n,i_n}=j)=p_{i_nj},\end{aligned}$$

as required. For the second assertion note that

$$\begin{aligned}b_{(i_1,i_2),j} &:= \mathbb{P}(Z^{(1)}_{n+1}+Z^{(2)}_{n+1}=j|Z^{(1)}_n=i_1,Z^{(2)}_n=i_2)\\ &= \mathbb{P}(N^{(1)}_{n,1}+\ldots+N^{(1)}_{n,i_1}+N^{(2)}_{n,1}+\ldots+N^{(2)}_{n,i_2}=j)=p_{i_1+i_2,j}\end{aligned}$$

only depends on i_1+i_2 (not i_1 or i_2 separately) and is of the form required to conclude that

$$\mathbb{P}(Z^{(1)}_{n+1}+Z^{(2)}_{n+1}=j|Z^{(1)}_n+Z^{(2)}_n=i)=\frac{\sum_{i_1=0}^{i}\mathbb{P}(Z^{(1)}_n=i_1,Z^{(2)}_n=i-i_1)b_{(i_1,i-i_1),j}}{\mathbb{P}(Z^{(1)}_n+Z^{(2)}_n=i)}=p_{ij}.$$

□

The second part of the proposition is called the *branching property* and expresses the property that the families of individuals in the same generation evolve completely independently of one another.

15.2 Continuous-time Galton-Watson processes

We can also model lifetimes of individuals by independent exponentially distributed random variables with parameter $\lambda > 0$. We assume that births happen at the end of a lifetime. This breaks the generations. Since continuously distributed random variables are almost surely distinct, we will observe one death at a time, each leading to a jump of size $k-1$ with probability ξ_k, $k \geq 0$. It is customary to only consider offspring distributions with $\xi_1 = 0$, so that there is indeed a jump at every death time. Note that at any given time, if j individuals are present in the population, the next death occurs at a time

$$H=\min\{L_1,\ldots,L_j\}\sim \mathrm{Exp}(j\lambda),\qquad \text{where } L_1,\ldots,L_j\sim\mathrm{Exp}(\lambda).$$

From these observations, one can construct (and simulate!) the associated population size process $(Y_t)_{t\geq 0}$ by induction on the jump times.

Proposition 97 *$(Y_t)_{t\geq 0}$ is a Markov process. If $Y^{(1)}$ and $Y^{(2)}$ are independent Markov processes with these transition probabilities starting from k and l, then $Y^{(1)}+Y^{(2)}$ is also a Markov process with the same transition probabilities starting from $k+l$.*

Proof: Based on BS3a Applied Probability, the proof is not difficult. We skip it here.

□

$(Y_t)_{t\geq 0}$ is called a continuous-time Galton-Watson process. In fact, these are the only Markov processes with the branching property (i.e. satisfying the second statement of the proposition for all $k \geq 1$, $l \geq 1$).

Example 98 (Simple birth-and-death processes) If individuals have lifetimes with parameter μ and give birth at rate β to single offspring repeatedly during their lifetime, then we recover the case

$$\lambda = \mu + \beta \qquad \text{and} \qquad \xi_0 = \frac{\mu}{\mu+\beta}, \quad \xi_2 = \frac{\beta}{\mu+\beta}.$$

In fact, we have to reinterpret this model by saying each transition is a death, giving birth to either two or no offspring. These parameters arise since, if only one individual is present, the time to the next transition is the minimum of the exponential birth time and the exponential death time.

The fact that all jump sizes are independent and identically distributed is reminiscent of compound Poisson processes, but for high population sizes j we have high parameters to the exponential times between two jumps – the process Y moves faster than a compound Poisson process at rate λ. Note however that for $H \sim \mathrm{Exp}(j\lambda)$ we have $jH \sim \mathrm{Exp}(\lambda)$. Let us use this observation to specify a time-change to slow down Y.

Proposition 99 *Let $(Y_t)_{t\geq 0}$ be a continuous-time Galton-Watson process with offspring distribution ξ and lifetime distribution* $\mathrm{Exp}(\lambda)$. *Then for the piecewise linear functions*

$$J_t = \int_0^t Y_u du, \quad t \geq 0, \qquad \varphi_s = \inf\{t \geq 0 : J_t > s\}, \quad 0 \leq s < J_\infty,$$

the process

$$X_s = Y_{\varphi_s}, \quad 0 \leq s < J_\infty,$$

is a compound Poisson process with jump distribution $(\xi_{k+1})_{k\geq -1}$ and rate λ, run until the first hitting time of 0.

Proof: Given $Y_0 = i$, the first jump time $T_1 = \inf\{t \geq 0 : Y_t \neq i\} \sim \mathrm{Exp}(i\lambda)$, so

$$J_{T_1} = iT_1 \qquad \text{and} \qquad \varphi_s = s/i, \quad 0 \leq s \leq iT_1,$$

so we identify the first jump of $X_s = Y_{s/i}$, $0 \leq s \leq iT_1$ at time $iT_1 \sim \mathrm{Exp}(\lambda)$.

Now the strong Markov property (or the lack of memory property of all other lifetimes) implies that given k offspring are produced at time T_1, the process $(Y_{T_1+t})_{t\geq 0}$ is a continuous-time Galton-Watson process starting from $j = i + k - 1$, independent of $(Y_r)_{0\leq r\leq T_1}$. We repeat the above argument to see that $T_2 - T_1 \sim \mathrm{Exp}(j\lambda)$, and for $j \geq 1$,

$$J_{T_2} = iT_1 + j(T_2 - T_1) \qquad \text{and} \qquad \varphi_{iT_1+s} = T_1/i + s/j, \quad 0 \leq s \leq j(T_2 - T_1),$$

and the second jump of $X_{iT_1+s} = Y_{T_1+s/j}$, $0 \leq s \leq j(T_2 - T_1)$, happens at time $iT_1 + j(T_2 - T_1)$, where $j(T_2 - T_1) \sim \mathrm{Exp}(\lambda)$ is independent of iT_1. An induction as long as $Y_{T_n} > 0$ shows that X is a compound Poisson process run until the first hitting time of 0. □

Corollary 100 *Let $(X_s)_{s\geq 0}$ be a compound Poisson process starting from $l \geq 1$ with jump distribution $(\xi_{k+1})_{k\geq -1}$ and jump rate $\lambda > 0$. Then for the piecewise linear functions*

$$\varphi_s = \int_0^s \frac{1}{X_v} dv, \quad 0 \leq s < T_{\{0\}}, \qquad \textit{and} \qquad J_t = \inf\{s \geq 0 : \varphi_s > t\}, \quad t \geq 0,$$

the process

$$Y_t = X_{J_t}, \quad t \geq 0,$$

is a continuous-time Galton-Watson process with offspring distribution ξ and lifetime distribution $\mathrm{Exp}(\lambda)$.

15.3 Continuous-state branching processes

Population-size processes with state space $\mathbb{N}$ are natural, but for large populations, it is often convenient to use continuous approximations and use a state space $[0, \infty)$. In view of Corollary 100 it is convenient to define as follows.

Definition 101 (Continuous-state branching process) Let $(X_s)_{s\geq 0}$ be a Lévy process with no negative jumps starting from $x > 0$, with $\mathbb{E}(\exp\{-\gamma X_s\}) = \exp\{s\phi(\gamma)\}$. Then for the functions

$$\varphi_s = \int_0^s \frac{1}{X_v} dv, \quad 0 \leq s < T_{\{0\}}, \qquad \text{and} \qquad J_t = \inf\{s \geq 0 : \varphi_s > t\}, \quad t \geq 0,$$

the process

$$Y_t = X_{J_t}, \quad t \geq 0,$$

is called a continuous-state branching process with branching mechanism ϕ.

We interpret upward jumps as birth events and continuous downward movement as (infinitesimal) deaths. The behaviour is accelerated at high population sizes, so fluctuations will be larger. The behaviour is slowed down at small population sizes, so fluctuations will be smaller.

Example 102 (Pure death process) For $X_s = x - cs$ we obtain

$$\varphi_s = \int_0^s \frac{1}{x - cv} dv = -\frac{1}{c} \log(1 - cs/x), \quad \text{and} \quad J_t = \frac{x}{c}(1 - e^{-ct}),$$

and so $Y_t = xe^{-ct}$.

Example 103 (Feller diffusion) For $\phi(\gamma) = \gamma^2$ we obtain Feller's diffusion. There are lots of parallels with Brownian motion. There is a Donsker-type result which says that rescaled Galton-Watson processes converge to Feller's diffusion. It is the most popular model in applications. A lot of quantities can be calculated explicitly.

Proposition 104 *Y is a Markov process. Let $Y^{(1)}$ and $Y^{(2)}$ be two independent continuous-state branching processes with branching mechanism ϕ starting from $x > 0$ and $y > 0$. Then $Y^{(1)} + Y^{(2)}$ is a continuous-state branching process with branching mechanism ϕ starting from $x + y$.*

Lecture 16

The two-sided exit problem

Reading: Kyprianou Chapter 8, Bertoin Aarhus Notes, Durrett Sections 7.5-7.6

16.1 The two-sided exit problem for Lévy processes with no negative jumps

Let X be a Lévy process with no negative jumps. As we have studied processes with no positive jumps (such as $-X$) before, it will be convenient to use compatible notation and write

$$\mathbb{E}(e^{-\gamma X_t}) = e^{t\phi(\gamma)}, \quad \phi(\gamma) = a_{-X}\gamma + \frac{1}{2}\sigma^2\gamma^2 + \int_{-\infty}^{0} (e^{\gamma x} - 1 - \gamma x 1_{\{|x|\le 1\}}) g_{-X}(x)dx$$
$$= -a_X\gamma + \frac{1}{2}\sigma^2\gamma^2 - \int_0^{\infty} (1 - e^{-\gamma x} - \gamma x 1_{\{|x|\le 1\}}) g_X(x)dx,$$

where $a_{-X} = -a_X$ and $g_X(x) = g_{-X}(-x)$, $x > 0$. Then we deduce from Section 11.4 that, if $\mathbb{E}(X_1) < 0$,

$$\mathbb{E}(e^{-\beta \overline{X}_\infty}) = -\frac{\beta \mathbb{E}(X_1)}{\phi(\beta)}, \quad \beta \ge 0. \tag{1}$$

The two-sided exit problem is concerned with exit from an interval $[-a, b]$, notably the time

$$T = T_{[-a,b]^c} = \inf\{t \ge 0 : X_t \in [-a, b]^c\}$$

and the probability to exit at the bottom $\mathbb{P}(X_T = -a)$. Note that an exit from $[-a, b]$ at the bottom happens necessarily at $-a$, since there are no negative jumps, whereas an exit at the top may be due to a positive jump across the threshold b leading to $X_T > b$.

Proposition 105 *For any Lévy process X with no negative jumps, all $a > 0$, $b > 0$ and $T = T_{[-a,b]^c}$, we have*

$$\mathbb{P}(X_T = -a) = \frac{W(b)}{W(a+b)}, \qquad \text{where } W \text{ is such that } \int_0^{\infty} e^{-\beta x} W(x)dx = \frac{1}{\phi(\beta)}.$$

Proof: We only prove the case $\mathbb{E}(X_1) < 0$. By (1), we can identify (the right-continuous function) W, since

$$\begin{aligned} -\frac{\beta\mathbb{E}(X_1)}{\phi(\beta)} &= \mathbb{E}(e^{-\beta\overline{X}_\infty}) = \int_0^\infty e^{-\beta x} f_{\overline{X}_\infty}(x)dx \\ &= \int_0^\infty \beta e^{-\beta x}\mathbb{P}(\overline{X}_\infty) \le x)dx, \end{aligned}$$

by partial integration, and so by the uniqueness of moment generating functions, we have $W(x) = c\mathbb{P}(\overline{X}_\infty \le x)$, where $c = -\mathbb{E}(X_1) > 0$.

Now define $\tau_a = \inf\{t \ge 0 : X_t < -a\}$ and apply the strong Markov property at τ_a to get a post-τ_a process $\widetilde{X} = (X_{\tau_a+s} + a)_{s\ge 0}$ independent of $(X_r)_{r\le\tau_a}$, in particular of $\overline{X}_{\tau_a}$, so that

$$\begin{aligned} cW(b) &= \mathbb{P}(\overline{X}_\infty \le b) = \mathbb{P}(\overline{X}_{\tau_a} \le b, \overline{\widetilde{X}}_\infty \le a + b) \\ &= \mathbb{P}(\overline{X}_{\tau_a} \le b)\mathbb{P}(\overline{X}_\infty \le a + b) = \mathbb{P}(\overline{X}_{\tau_a} \le b)cW(a + b), \end{aligned}$$

and the result follows. □

Example 106 (Stable processes) Let X be a stable process of index $\alpha \in (1, 2]$ with no negative jumps, then we have

$$\int_0^\infty e^{-\lambda x}W(x)dx = \lambda^{-\alpha} \qquad \Rightarrow \qquad \Gamma(\alpha)W(x) = x^{\alpha-1}.$$

We deduce that

$$\mathbb{P}(X_T = -a) = \left(\frac{b}{a+b}\right)^{\alpha-1}.$$

16.2 The two-sided exit problem for Brownian motion

In the Brownian case, we can push the analysis further without too much effort.

Proposition 107 *For Brownian motion B, all $a > 0$, $b > 0$ and $T = T_{[-a,b]^c}$, we have*

$$\mathbb{E}(e^{-qT}|B_T = -a) = \frac{V_q(b)}{V_q(a+b)} \quad \textit{and} \quad \mathbb{E}(e^{-qT}|B_T = b) = \frac{V_q(a)}{V_q(a+b)},$$

where

$$V_q(x) = \frac{\sinh(\sqrt{2q}x)}{x}.$$

Proof: We condition on B_T and use the strong Markov property of B at T to obtain

$$\begin{aligned}
e^{-a\sqrt{2q}} &= \mathbb{E}(e^{-qT_{\{-a\}}}) \\
&= \mathbb{P}(B_T=-a)\mathbb{E}(e^{-qT_{\{-a\}}}|B_T=-a)+\mathbb{P}(B_T=b)\mathbb{E}(e^{-qT_{\{-a\}}}|B_T=b) \\
&= \frac{b}{a+b}\mathbb{E}(e^{-qT}|B_T=-a)+\frac{a}{a+b}\mathbb{E}(e^{-q(T+\widetilde{T}_{\{-a-b\}})}|B_T=b) \\
&= \frac{b}{a+b}\mathbb{E}(e^{-qT}|B_T=-a)+\frac{a}{a+b}\mathbb{E}(e^{-qT}|B_T=b)e^{-(a+b)\sqrt{2q}}
\end{aligned}$$

and, by symmetry,

$$e^{-b\sqrt{2q}}=\frac{a}{a+b}\mathbb{E}(e^{-qT}|B_T=b)+\frac{b}{a+b}\mathbb{E}(e^{-qT}|B_T=-a)e^{-(a+b)\sqrt{2q}}.$$

These can be written as

$$\begin{aligned}
\frac{b+a}{ab} &= a^{-1}\mathbb{E}(e^{-qT}|B_T=-a)e^{a\sqrt{2q}}+b^{-1}\mathbb{E}(e^{-qT}|B_T=b)e^{-b\sqrt{2q}} \\
\frac{b+a}{ab} &= a^{-1}\mathbb{E}(e^{-qT}|B_T=-a)e^{-a\sqrt{2q}}+b^{-1}\mathbb{E}(e^{-qT}|B_T=b)e^{b\sqrt{2q}}
\end{aligned}$$

and suitable linear combinations give, as required,

$$\begin{aligned}
2\sinh(a\sqrt{2q})\frac{b+a}{ab} &= 2\sinh((a+b)\sqrt{2q})b^{-1}\mathbb{E}(e^{-qT}|B_T=b) \\
2\sinh(b\sqrt{2q})\frac{b+a}{ab} &= 2\sinh((a+b)\sqrt{2q})a^{-1}\mathbb{E}(e^{-qT}|B_T=-a).
\end{aligned}$$

□

Corollary 108 *For Brownian motion B, all $a>0$ and $T=T_{[-a,a]^c}$, we have*

$$\mathbb{E}(e^{-qT})=\frac{1}{\cosh(a\sqrt{2q})}$$

Proof: Just calculate from the previous proposition

$$\mathbb{E}(e^{-qT})=\frac{V_q(a)}{V_q(2a)}=2\frac{e^{\sqrt{2q}a}-e^{-\sqrt{2q}a}}{(e^{\sqrt{2q}a})^2-(e^{-\sqrt{2q}a})^2}=\frac{1}{\cosh(a\sqrt{2q})}.$$

□

16.3 Appendix: Donsker's Theorem revisited

We can now embed simple symmetric random walk (SSRW) into Brownian motion B by putting

$$T_0=0,\quad T_{k+1}=\inf\{t\ge T_k:|B_t-B_{T_k}|=1\},\qquad S_k=B_{T_k},\quad k\ge 0,$$

and for step sizes $1/\sqrt{n}$ modify $T_{k+1}^{(n)}=\inf\{t\ge T_k^{(n)}:|B_t-B_{T_k^{(n)}}|=1/\sqrt{n}\}$.

Theorem 109 (Donsker for SSRW) *For a simple symmetric random walk* $(S_n)_{n\geq 0}$ *and Brownian motion* B*, we have*

$$\frac{S_{[nt]}}{\sqrt{n}} \to B_t, \qquad \textit{locally uniformly in } t \geq 0, \textit{ in distribution as } n \to \infty.$$

Proof: We use a coupling argument. We are not going to work directly with the original random walk $(S_n)_{n\geq 0}$, but start from Brownian motion $(B_t)_{t\geq 0}$ and define a family of embedded random walks

$$S_k^{(n)} := B_{T_k^{(n)}}, \quad k \geq 0, n \geq 1, \qquad \Rightarrow \qquad \left(S_{[nt]}^{(n)}\right)_{t\geq 0} \sim \left(\frac{S_{[nt]}}{\sqrt{n}}\right)_{t\geq 0}.$$

To show convergence in distribution for the processes on the right-hand side, it suffices to establish convergence in distribution for the processes on the left-hand side, as $n \to \infty$.

To show locally uniform convergence we take an arbitrary $T \geq 0$ and show uniform convergence on $[0, T]$. Since $(B_t)_{0\leq t\leq T}$ is uniformly continuous (being continuous on a compact interval), we get in probability

$$\sup_{0\leq t\leq T} \left|S_{[nt]}^{(n)} - B_t\right| = \sup_{0\leq t\leq T} \left|B_{T_{[nt]}^{(n)}} - B_t\right| \leq \sup_{0\leq s\leq t\leq T: |s-t|\leq \sup_{0\leq r\leq T} |T_{[nr]}^{(n)} - r|} |B_s - B_t| \to 0$$

as $n \to \infty$, if we can show (as we do in the lemma below) that $\sup_{0\leq t\leq T} |T_{[nt]}^{(n)} - t| \to 0$.

This establishes convergence in probability, which "implies" convergence in distribution for the embedded random walks and for the original scaled random walk. □

Lemma 110 *In the setting of the proof of the theorem,* $\sup_{0\leq t\leq T} |T_{[nt]}^{(n)} - t| \to 0$ *in probability.*

Proof: First for fixed t, we have

$$\mathbb{E}(e^{-qT_{[nt]}^{(n)}}) = \left(\mathbb{E}(e^{-qT_1^{(n)}})\right)^{[nt]} = \frac{1}{(\cosh(\sqrt{2q/n}))^{[nt]}} \to e^{-qt},$$

since $\cosh(\sqrt{2q/n}) \sim 1 + q/n + O(1/n)$. Therefore, in probability $T_{[nt]}^{(n)} \to t$. For uniformity, let $\varepsilon > 0$. Let $\delta > 0$. We find $n_0 \geq 0$ such that for all $n \geq n_0$ and all $t_k = k\varepsilon/2$, $1 \leq k \leq 2T/\varepsilon$ we have

$$\mathbb{P}(|T_{[nt_k]}^{(n)} - t_k| > \varepsilon/2) < \delta\varepsilon/2T,$$

then

$$\mathbb{P}\left(\sup_{0\leq t\leq T} |T_{[nt]}^{(n)} - t| > \varepsilon\right) \leq \mathbb{P}\left(\sup_{1\leq k\leq 2T/\varepsilon} |T_{[nt_k]}^{(n)} - t_k| > \frac{\varepsilon}{2}\right) \leq \sum_{k=1}^{2T/\varepsilon} \mathbb{P}\left(|T_{[nt_k]}^{(n)} - t_k| > \frac{\varepsilon}{2}\right) < \delta.$$

□

We can now describe the recipe for the full proof of Donsker's Theorem. In fact, we can embed every standardized random walk $((S_k - k\mathbb{E}(S_1))/\sqrt{n\mathrm{VarS_1}})_{k\geq 0}$ in Brownian motion X, by first exits from independent random intervals $[-A_k^{(n)}, B_k^{(n)}]$ so that $X_{T_k^{(n)}} \sim (S_k - k\mathbb{E}(S_1))/\sqrt{n\mathrm{Var}(S_1)}$, and the embedding time change $(T_{[nt]}^{(n)})_{t\geq 0}$ can still be shown to converge uniformly to the identity.