Stochastic Processes and Stochastic Integration

Marcus Pivato

February 21, 1999

Contents

Chapter 1

Preface

1.1 How to Read this Document

This document is not a book. It is not intended to be read in a linear fashion. Instead, this document is divided into small, semi-independent "modules" of text. Rather than being arranged in a linear order, these modules are arranged in a loose heirarchy of "logical dependency"; modules dealing with more advanced topics often depend upon the reader's knowledge of material covered in certain previous modules. When such relationships of logical dependency exist, they will be explicitly indicated in the list of **prerequisites** which appear at the beginning of each module.

Modules also sometimes have **corequisites**. These are other modules which, although not logically *necessary* for comprehending the material in question, may be helpful in providing an intuitive context or motivation for what is discussed. Hence, if module A is a *prerequisite* for module B, then module A should be read, understood, and mastered by the student before module B is approached. If module A is, instead, a *corequisite*, then the reader is advised (but not mandated) to read or browse A simultaneously with reading B.

The **table of contents** portrays an organisational schema whereby modules are divided into "chapters", "sections" and "subsections". This organisational schema is useful for grouping related material together, but once again, it is not intended to suggest any sort of linear logical structure. The linearity of the contents is an artifact of a paper-oriented typesetting environment.

1.2 Notation

Symbol	Interpretation
Sets and Spaces	
$\mathbb{N}$	The natural numbers $\{0, 1, 2, \ldots\}$
$\mathbb{Z}$	The integers
$\mathbb{Q}$	The rational numbers
$\mathbb{R}$	The real numbers
$\mathbb{C}$	The complex numbers
$[a, b]$	The **closed** interval in $\mathbb{R}$ between a and b; $[a, b] := \{x \in \mathbb{R} \; ; \; a \leq x \leq b\}$
$[a, b)$	The **left-open** interval in $\mathbb{R}$ between a and b; $[a, b) := \{x \in \mathbb{R} \; ; \; a \leq x < b\}$
$(a, b]$	The **right-open** interval in $\mathbb{R}$ between a and b; $(a, b] := \{x \in \mathbb{R} \; ; \; a < x \leq b\}$
(a, b)	The **open** interval in $\mathbb{R}$ between a and b; $(a, b) := \{x \in \mathbb{R} \; ; \; a < x < b\}$
$[a..b]$	The **closed** interval in $\mathbb{Z}$ between a and b; $[a, b] := \{x \in \mathbb{Z} \; ; \; a \leq x \leq b\}$
$[a..b)$	The **left-open** interval in $\mathbb{Z}$ between a and b; $[a, b) := \{x \in \mathbb{Z} \; ; \; a \leq x < b\}$
$(a..b]$	The **right-open** interval in $\mathbb{Z}$ between a and b; $(a, b] := \{x \in \mathbb{Z} \; ; \; a < x \leq b\}$
$(a..b)$	The **open** interval in $\mathbb{Z}$ between a and b; $(a, b) := \{x \in \mathbb{Z} \; ; \; a < x < b\}$
$\mathbb{D}$	The unit disk (usually in the complex plane) $\mathbb{D} := \{x \in \mathbb{C} \; ; \; \|x\| \leq 1\}$
$\mathbb{B}^D(\vec{x}; R)$	The **closed** ball of radius R about a point $\vec{x}$ in $\mathbb{R}^D$ $\mathbb{B}^D(\vec{x}; R) := \left\{\vec{v} \in \mathbb{R}^D \; ; \; \|\vec{v} - \vec{x}\| \leq R\right\}$
$\square^D$	The unit cube in $\mathbb{R}^D$ $\square^D := [0, 1] \times [0, 1] \times \ldots \times [0, 1]$
$\mathbb{T}_{or}^D$	The D-dimensional torus $\mathbb{T}_{or}^1 := \{x \in \mathbb{C} \; ; \; \|x\| = 1\}$. $\mathbb{T}_{or}^D := \mathbb{T}_{or}^1 \times \mathbb{T}_{or}^1 \times \ldots \mathbb{T}_{or}^1$
$\mathbb{T}_{ime}$	The "time" line for a dynamical system or stochastic Normally $\mathbb{T}_{ime} := \mathbb{N}, \mathbb{Z}, \mathbb{R}$, or $[0, \infty)$

Spaces of Functions	
$\mathbf{C}_{bound}(X;Y)$	The space of **bounded continuous functions** from X to Y
$\mathbf{C}_{unif}(X;Y)$	The space of **uniformly continuous functions** from X to Y
Operators on Function Spaces	
$\mathcal{F}^{our}[f]$	The **Fourier Transform** of the function f.
$\mathcal{M}_{ax}[f]$	The **Hardy Maximal Function** associated to the function f.
Special Functions	
$\overbrace{x;R}$	The Gaussian distribution with **mean** x and **variance** R; in $\mathbb{R}^D$, $\overbrace{\vec{x};\vec{R}}(\vec{y}) := \frac{1}{(2\pi)^{D/2}\,R}\exp\left(\left\|\frac{\vec{x}-\vec{y}}{R}\right\|^2\right)$
$\mathcal{C}_{ard}[S]$	The **cardinality** of the set S.
$\ell_{ength}[\gamma]$	The **length** of the curve γ.
$\mathbf{Id}_X$	The **identity function** on the space X. $\mathbf{Id}_X(x) = x$, for all $x \in X$.
$\mathbf{pr}_n$	The **projection map** onto the nth coordinate of a product space. $\mathbf{pr}_n(x_1, x_2, \ldots, x_N) = x_n$.
$\mathbb{1}_U$	The **characteristic function** of the set U. $\mathbb{1}_U(x) = 1$ iff $x \in U$.
Measure theory	
$\forall_\mu\, x$	"For almost all x (with respect to the measure μ)".
$\mathcal{L}$	The **Lebesgue measure**.
$\mathcal{L}^D$	The **Lebesgue measure** on $\mathbb{R}^D$.
$\|\mu\|_{var}$	The **total variation norm** of the measure μ.
$\boxed{\mu}$	The **total variation measure** associated to the measure μ.

Probability Theory	
$(\Omega, \mathcal{W}, \mathbf{P})$	Almost universally used to represent the **universal sample space.**
$\mathbf{D}^{istr}[X]$	The **distribution** of the random variable X.
$X \overset{\cong}{_{distr}} Y$	"The random variables X and Y are **identically distributed**."
$X \overset{\perp}{\underset{distr}{\sim}} Y$	"The random variables X and Y are **independent** and **identically distributed**."
$\mathbf{P}_{rob}[U]$	The **probability** of the event U.
$\mathbf{P}_{rob}\left[\begin{array}{c\|c} U & V \end{array}\right]$	The **conditional probability** of event U, given event V.
$\mathbf{E}^{xpct}[f]$	The **expectation** of the random variable f.
$\mathbf{E}^{xpct}_{\mathcal{X}}[f]$	The **conditional expectation** of the random variable f, given the information in the sigma-algebra $\mathcal{X}$.
$\mathbf{E}^{xpct}\left[\frac{f}{V}\right]$	The **conditional expectation** of the random variable f, given the event V.

Part I

Stochastic Processes

Chapter 2

Introduction to Stochastic Processes

A **dynamical system** is a mathematical structure used to model the deterministic evolution of some physical system or phenomenon in time. For example, *ordinary differential equations* are normally interpreted as describing evolution in time, and hence, determine dynamical systems.

The key word here, however, is *deterministic* ; a dynamical system is deterministic because the future is (in principle) *completely predictable* from knowledge of the present state. What if this is not the case, however? What if there is some "intrinisic randomness" in the system, which makes perfect prediction of the future impossible? Perhaps very strong *trends* or *correlations* still exist, but there is always some element of uncertainty.

The mathematical structure used to model this sort of phenomenon is a **stochastic process** .

2.1 Timelines and Statespace

Informally, a stochastic process consists of three components.

- First, of course, there is a **statespace** ; the set of all "states" of the system we wish to model. The statespace is normally some topological space, or at least, a measure space.

  Normally, in our discussion, we will denote the statespace by X.

- Secondly, there is a **timeline** , which is the set used to index the passage of time.

  Normally, we will denote the timeline by $\mathbb{T}_{ime}$.

- Finally, there is a **probability measure** on the space of all possible "paths" from $\mathbb{T}_{ime}$ into X.

The timeline $\mathbb{T}_{ime}$ should have two basic properties:

1. $\mathbb{T}_{ime}$ should be linearly ordered, so that we can sensibly speak of things like "the future" and "the past".

2. $\mathbb{T}_{ime}$ should have some sort of additive structure, so that we can sensibly speak of "moving forwards" (or "backwards") by some amount.

Common choices for $\mathbb{T}_{ime}$ are:

- The real line, $\mathbb{R}$.
- The real "half-line", $[0, \infty)$
- The bounded time interval $[0, 1]$
- The integers, $\mathbb{Z}$.
- The natural numbers, $\mathbb{N}$.
- The rational numbers, $\mathbb{Q}$.

Note that we do not require that $\mathbb{T}_{ime}$ be closed under addition; this would essentially imply an "infinite future", which is not always possible. We also don't require the existence of arbitrary inverses; this would necessitate an "infinite past", which may also be impossible. Finally, we will sometimes chose $\mathbb{T}_{ime}$ to be **discrete** (ie. $\mathbb{Z}$ or $\mathbb{N}$) , and sometimes **continuous** (ie. $\mathbb{R}$, $[0, \infty)$, etc.) .

The space of all possible "histories" of our system is then the space of all possible $\mathbb{T}_{ime}$–indexed path through space; in other words, the space of all functions from $\mathbb{T}_{ime}$ into X. A dynamical system can be regarded as chosing a few, specific paths from this collection (called **orbits**) and declaring them to be the only "allowed" paths —furthermore, none of these paths can intersect. A stochastic process imposes a much weaker constraint; it involves simply placing a **probability measure** on the space of paths, declaring some histories to be more or less "likely" than others.

Formally, we define a stochastic process as follows....

2.1.1

Definition *Stochastic Process*

Let X be some set, $\mathbb{T}_{ime}$ some other set. Let $\mathcal{W}$ be some sigma-algebra on $X^{\mathbb{T}_{ime}}$. Then a $\mathcal{W}-$**measurable stochastic process** on the **statespace** X, over the **timeline** $\mathbb{T}_{ime}$, is a probability measure on $\mathcal{W}$.

Without further hypotheses about the statespace X, the properties of the desired paths, or the structure of $\mathcal{W}$, it is impossible to be any more specific.

For starters, what sigma-algebra should we use? Well, we want the sigma-algebra to be fine enough that, at the very least, for every time $t \in \mathbb{T}_{ime}$, the state of the system at time t is a measurable function. In other words, for all $t \in \mathbb{T}_{ime}$, we want:

$$\mathbf{pr}_t : X^{\mathbb{T}_{ime}} \ni \alpha \mapsto \alpha(t) \in X$$

to be measurable. Measurable with respect to what? We first need a sigma-algebra on X.

2.1.2

Definition *Cylinder algebra*

Let $(X, \mathcal{X})$ be a measurable space. The **cylinder sigma-algebra** on $X^{\mathbb{T}_{ime}}$ is the sigma-algebra $\mathcal{X}^{\otimes \mathbb{T}_{ime}}$ generated by all the set of all **coordinate functions** $\{\mathbf{pr}_t \ ; \ t \in \mathbb{T}_{ime}\}$.

In other words, $\mathcal{X}^{\otimes \mathbb{T}_{ime}}$ is the smallest sigma-algebra containing all **cylinder sets**

$$\left(\overline{\substack{t_1,t_2,\ldots,t_n \\ U_1,U_2,\ldots,U_n}}\right) := \left\{ \alpha : \mathbb{T}_{ime} \longrightarrow X \ ; \begin{array}{lll} \alpha(t_1) & \in & U_1, \\ \alpha(t_2) & \in & U_2, \\ & \vdots & \\ \alpha(t_N) & \in & U_N \end{array} \right\}$$

where $t_1,\ t_2,\ \ldots,\ t_n \in \mathbb{T}_{ime}$, and $U_1,\ U_2,\ \ldots,\ U_n \subset X$ are $\mathcal{X}-$measurable subsets.

2.1.3

Remark

In some contexts (for example, **continuous-time** processes) , the cylinder sigma-algebra may prove too "crude" for our purposes, and we may adopt a finer algebra instead. The cylinder algebra is in some sense the "minimum" amount of sigma-algebra necessary to do anything useful.

Sometimes, the formalism for stochastic processes is formulated in terms of **random variables** with respect to some universal **sample space**.

2.1.4

Definition *Stochastic Processes in Terms of Sample Spaces*

A **stochastic process** to be any *random variable* taking its values in $X^{\mathbb{T}_{time}}$.

In other words, fix an abstract probability space $(\Omega, \mathcal{W}, \mathbf{P})$. Then a stochastic process is a *measurable function*

$$\beta : \Omega \longrightarrow X^{\mathbb{T}_{time}}.$$

Exercise: Show the two definitions of **stochastic process** are equivilant.

2.2 Time Shifts and Stationarity

Now we want a way of moving the process "forwards" or "backwards" in time.

Note that, if $\alpha : \mathbb{T}_{time} \longrightarrow X$ is some path, and we "shift" α forward in time by some amount T, then the shifted path no longer a function from $\mathbb{T}_{time}$ into X. Instead, it is a function from $\mathbb{T}_{time} - T$ into X, where

$$\mathbb{T}_{time} - T \ := \ \{t - T \ ; \ t \in \mathbb{T}_{time}\}.$$

Of course, in many situations, these two sets may be the same –for example, if $\mathbb{T}_{time} = \mathbb{R}$. However, the distinction is worth bearing in mind...

2.2.1

Definition *Time Shift*

Using the additive property of our timeline, we can define a **shift action** on $X^{\mathbb{T}_{time}}$ by $\mathbb{R}$. For every $t \in \mathbb{T}_{time}$ and $\alpha \in X^{\mathbb{T}_{time}}$, define $\mathcal{S}^t_{hift}\,\alpha$ the function from $\mathbb{T}_{time} - t$ into X defined:

$$\mathcal{S}^t_{hift}\,\alpha(s) \;:=\; \alpha(t+s)$$

2.2.2

Definition *Stationarity*

Let μ be a measure on $X^{\mathbb{T}_{time}}$ defining a stochastic process on X. The stochastic process is called **stationary** if μ is **time-shift invariant** ; in other words, for all $t \in \mathbb{T}_{time}$,

$$\mu \circ \mathcal{S}^t_{hift} = \mu.$$

2.2.3

Remark

Clearly, in order for this to make sense, $\mu \circ \mathcal{S}^t_{hift}$ must also be a measure on $X^{\mathbb{T}_{time}}$. Thus, $\mathbb{T}_{time} - t \;=\; \mathbb{T}_{time}$. Thus, $\mathbb{T}_{time}$ must be closed under addition, so it must be a **group**; for example, $\mathbb{R}, \mathbb{Q}$, or $\mathbb{Z}$.

2.2.4

Discussion

Intuitively, what this says is that the process "looks the same", no matter when in time you look at it. The probability distribution of states in the system does not change over time. Furthermore, the probabilistic correlation between past and future events is *also* invariant over time.

Not all processes are stationary. For example, consider the random diffusion of a particle in an unbounded space. Intuitively, the particle starts near some "origin" point at time zero, and over time, wanders farther and farther from this point. Thus, near time zero, the probability distribution of the particle's position is concentrated near the origin, but as time goes

to infinity, the distribution becomes more and more "uniformly" spread out through the whole space.

On the other hand, the probability distribution of the particle's position can never become *completely* uniformly distributed, because the space is *unbounded* ; hence, the only "truly" uniform distribution is one which assigns every bounded region zero probability; a nonsensical condition. Hence, this stochastic process cannot approach any "stationary" distribution.

On the other hand, if the particle is diffusing in a *bounded* region, then, as time goes on, the probabilitiy distribution of its position will become uniformly distributed throughout the entire space. The limiting distribution is thus *stationary* .

Chapter 3

Discrete Time Stochastic Processes

3.1 Introduction

Prerequisites: [2]

The simplest sort of stochastic processes are those where we model the flow of time as a succession of discrete "steps".

3.1.1

Definition *Discrete Time Stochastic Process*

Let $(\Omega, \mathcal{W}, \mathbf{P})$ be a sample space, $(X, \mathcal{X})$ some measurable space. A **discrete time stochastic process** ranging over X is a measurable function

$$\Phi : \Omega \times \mathbb{T}_{ime} \longrightarrow X$$

where $\mathbb{T}_{ime}$ is either the integers or the natural numbers.

Alternately, consider $X^{\mathbb{T}_{ime}}$, the *space of paths from* $\mathbb{T}_{ime}$ *into* X, and endow it with the sigma-algebra:

$$\bigotimes_{n \in \mathbb{T}_{ime}} \mathcal{X}$$

A **discrete time stochastic process** is then probability measure on $\left(X^{\mathbb{T}_{time}}, \bigotimes_{n \in \mathbb{T}_{time}} \mathcal{X}\right)$.

If $\mathbb{T}_{time} = \mathbb{Z}$, then the process has an *infinite history.* On the other hand, if $\mathbb{T}_{time} = \mathbb{N}$, then the flow of time has an explicit beginning. Clearly, any process with an infinite past (ie. $\mathbb{T}_{time} = \mathbb{Z}$) can be restricted to one with a finite past (ie. $\mathbb{T}_{time} = \mathbb{N}$) . The converse, however, is not true; a process on $\mathbb{N}$ cannot always be extended to one on $\mathbb{Z}$.

Intuitively, this is because some processes are such that they become increasingly "random" over time. After an infinite amount of time has passed, the process has become "infinitely random" —something which makes sense as a limiting condition, but not as an actual starting probability distribution. If the process had an infinite history, then it would already be "infinitely random" by time zero, producing nonsensical results.

Because of this, processes where $\mathbb{T}_{time} = \mathbb{N}$ are sometimes called **irreversible.**

Discrete time stochastic processes can be ranked in increasing order of complexity, depending upon the degree of *causality* they embody. Often, we speak of the process as having a sort of "memory" of its own past; we imagine the process randomly "deciding" to assume a certain state at time n; this "decision" may be influenced, to a greater or lesser degree, by the process's memory of its previous states. The more "memory" a process has, the greater the potential complexity of its behaviour. Loosely speaking, this hierarchy is as follows:

- **Bernoulli Processes** are the simplest kind of stochastic process, representing systems with no memory whatsoever. In a Bernoulli process, each event in time occurs totally independently of its predecessors. In other words, the process has no "causality" whatsoever. A Bernoulli process thus represents a succession of totally independent, random events; for example, a sequence of coin-flipping trials performed with the same coin.

  Bernoulli processes are discussed in [3.2] .

- **Markov Processes** are processes possessing some "short term memory"; each random event is influenced, at least to some degree, by its *immediate predecessor.* That is, the the random event which occurs at time n may be dependent upon that which occured at time $(n-1)$. However, the process has no "direct" memory of earlier events; at time n, the only knowledge of time $(n-2)$ which remains is whatever knowledge is stored implicitly in its memory of time $(n-1)$.

Another way to look at this is in terms of *prediction*. If we do not know the state of the system at time $(n-1)$, then it may be quite difficult to predict its state at time n. Once we learn its state at time $(n-1)$, our predictive ability will be improved. However, if we already know the time $(n-1)$ state, then acquiring additional information about time $(n-2)$, time $(n-3)$, etc. *will not provide us with any further useful information* for making predictions at time n.

In a certain sense, almost any stochastic process can be "reformulated" as a kind of Markov process, if one choses a suitable statespace. Often, however, the technical complications introduced this way are not worth the benefits of the Markov property. Hence, we are often content to examine processes with simpler statespaces, even if they possess more complex kinds of causality.

Markov processes are discussed in [3.3] .

- **Finitary Processes** are the logical next step after Markov Processes; they possess a direct knowledge of their own past, but only up to some *finite* time. For example, a finitary process might "directly remember" its history over the previous five "steps", and possess only indirect "memory" of earlier history. Hence, a Markov Process is just a finitary process whose "amnesia horizon" is of length one.

  Finitary processes are discussed in [**??**] .

3.2 Bernoulli Processes

A Bernoulli process represents a succession of random events, occuring sequentially in time, each one of which is *totally independent* of all of the rest, but all of which are *identical in distribution.*

3.2.1

Example *A Fair Coin*

For example, suppose you flip a certain coin over and over again. Each coin toss is a random event, identical to all the rest, but independent of them. If the coin is "fair", then **heads** and **tails** both appear with probability $\frac{1}{2}$. If you flip the coin 7 times, then all 128 outcomes appear with probability $\frac{1}{128}$.

3.2.2

Example *A Biased Coin*.

Suppose instead the coin is not "fair"; suppose **heads** appears with probability 2/3. If you flip the coin 7 times, different outcomes have different probabilities;

- An outcome of **all heads** has probability $\frac{128}{2187}$.
- An outcome of **all tails** has probability $\frac{1}{2187}$.
- The outcome: "**H H T T H H T**" has probability $\frac{16}{2187}$.

Suppose the random event in question is like a random variable ranging over some topological space X, having distribution given by the Radon measure ν, and suppose that our timeline is called $\mathbb{T}_{ime}$. Then, really, for every point $t \in \mathbb{T}_{ime}$, we want a random variable X_t, with distribution ν, so that all the X_t are *independent* of one another.

We want our stochastic process to correspond to the successive "occurence" of each of these random variables in time. In other words, we want to build a probability measure μ on $X^{\mathbb{T}_{ime}}$ which projects down to ν on every single coordinate, and so that all of these projections are independent; in other words, μ should look like a "product" of a bunch of copies of μ. Formally:

$$\mu := \nu^{\mathbb{T}_{ime}}.$$

This sort of product measure is well-defined if $\mathbb{T}_{ime}$ is a finite set, but it is unclear what this equation is supposed to mean when $\mathbb{T}_{ime}$ is infinite. To clarify this, we will employ Kolmogorov's Theorem [4.2.1].

For every $t \in \mathbb{T}_{ime}$, let $\mu_{\{t\}} := \nu$. For every finite subset $A \subset \mathbb{T}_{ime}$, define the measure μ_A on the space X^A by:

$$\mu_A := \prod_{t \in A} \mu_t. \quad (1)$$

(**Exercise**: Check that the Kolmogorov Consistency conditions are met.)

Thus, by Kolmogorov's theorem, there is a unique measure μ, defined on $X^{\mathbb{T}_{ime}}$, so that its finite-dimensional marginals are given by the measures of the form (1).

3.3 Discrete Time Markov Proceses

Prerequisites: [3.1]

3.3.1 Introduction

In a Bernoulli process, events in the present are totally independent of the past, so there is no "causality" to speak of. However, suppose we were to introduce some minimal notion of causality, wherein the present moment has influence over at least immediate future.

Suppose that the event which occurs at time $(n+1)$ is at least partly determined by the event which occurs at time n. Hence, each event which occurs exerts some small "influence" on the next event which is to occur. Suppose, however, that the event which occurs at time $(n-1)$ exerts no *direct* influence on the event at time $(n+1)$; to the extent that time $(n-1)$ affects $(n+1)$, it does so only *indirectly* by way of time n; time $(n-1)$ influences time n, and then time n influences time $(n+1)$. Hence, time $(n+1)$ only "knows" about time $n-1$ by way of whatever information is relayed through time n.

For example, in a game of **Telephone**, each player relays the message to the *next* player in the line, introducing random changes due to misperception or misinterpretation at each step. Thus, this can be seen as a sort of Markovian process; Player $(n+1)$ only has knowledge of the words of Player $(n-1)$ via whatever he is told by Player n.

3.3.2 Transition Probabilities: Finite Statespace

Formally, we can model this by a stochastic process which ranges over some statespace X. Let us first assume that X is a finite set, to keep technicalities to a minimum.

The state of the system at time n will determine the *probability distribution* of the state at time $n+1$. Suppose the state at time n is called X_n. Then, for any x and y in X, we can speak of the **transition probability**

$$B_{x \to y} \;:=\; \mathbf{P}_{\!\mathscr{W}} \left[X_{n+1} = y \;\middle|\; X_n = x \right]$$

that is, *the probability that the system will be in state y at time $(n+1)$, given the knowledge that it was in state x at time n.*

We assume that the transition mechanism of the system, although random, remains constant over time. In other words, we assume:

The transition probabilities $B_{x\to y}$ is the same for all $n \in \mathbb{T}_{ime}$.

Clearly, this collection of probabilities forms a sort of matrix, indexed in each dimension by X. This is called the **transition probability matrix;**

$$\mathbf{B} \;:=\; [B_{x\to y}]_{x,y\in X}$$

Exercise

1. Suppose that state of the system at time n is a random variable whose probability distribution is given by a vector

   $$\vec{q} := [q_x]_{x\in X} \;\in\; \mathbb{R}^X$$

   Then the state of the system at time $(n+1)$ is given by the vector

   $$\mathbf{B}\cdot\vec{q}$$

   where this formula represents matrix multiplication.

2. Suppose that, more generally, we define, for any $m > 0$, the transition matrix

   $$\mathbf{B}^{[m]} \;:=\; \mathbf{B}^{[m]}_{x\to y}$$

   by

   $$\mathbf{B}^{[m]}_{[x\to y]} \;:=\; \mathbf{P}_{rob}\left[X_{n+m} = y \;\middle|\; X_n = x\right]$$

   (for any fixed time n –it doesn't matter which one).

   Then show that

   $$\mathbf{B}^{[m]} \;=\; \mathbf{B}^m \;=\; \underbrace{\mathbf{B}\cdot\mathbf{B}\cdot\ldots\cdot\mathbf{B}}_{m}$$

3. Suppose now that the transition probabilities were *not* the same at all times. This means we have to introduce more complicated notation. So, for any $m, n \in \mathbb{T}_{ime}$, with $m > n$, define the **transition probabilities from time n to time m** by:

$$\mathbf{B}_{[x\to y]}^{[n\to m]} := \mathbf{P}_{\omega}\left[X_m = y \;\middle|\; X_n = x\right]$$

and then define the associated **transition probability matrix** by:

$$\mathbf{B}^{[n\to m]} := \mathbf{B}_{x\to y}^{[n\to m]}$$

Now, show that, for any times $m > n$,

$$\mathbf{B}^{[n\to m]} = \mathbf{B}^{[(m-1)\to m]} \cdot \ldots \cdot \mathbf{B}^{[(n+1)\to(n+2)]} \cdot \mathbf{B}^{[n\to(n+1)]}$$

4. Suppose you are given an arbitrary matrix of positive numbers between 0 and 1, and asked to determine whether it, in fact *could* be the transition probability matrix of some Markov process. Such a matrix is called a **Stochastic Matrix** .
   Show that, if $\mathbf{A} := [A_{x,y}]_{x,y\in X}$, then $\mathbf{A}$ is a stochastic matrix if and only if:
   (a) All entries in $\mathbf{A}$ are in the interval $[0,1]$.
   (b) Every column vector of $\mathbf{A}$ is a **probability vector**; in other words, the entries in each column add up to 1.
5. Let $\mathbf{A}$ be a real-valued matrix indexed by $X \times X$, and think of $\mathbf{A}$ as inducing a linear transformation on the vector space $\mathbb{R}^X$. Let $\triangle^X$ be the **probability simplex** in $\mathbb{R}^X$; that is, the set of all vectors in $\mathbb{R}^X$ containing strictly nonegative entries that sum to 1. Show that
   $\mathbf{A}$ is a stochastic matrix if and only if it induces a linear transformation which maps $\triangle^X$ into itself.

3.3.1

Example *A Random Switch*

A stochastic dynamical system alternates between two states, called 0 and 1. If the system is in state q at time n, then it has a probability of 0.9 of remaining in this state at time $(n+1)$, and a 0.1 probability of changing states. This is a Markov process with transition probability matrix

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

3.3.2

Example *Random Walk on a Cyclic Group*

Imagine the cyclic group $G := \mathbf{Z}/N\mathbf{Z}$ as a set of N discrete points, arranged in a circle. Suppose that a particle "randomly walks" on G; during each time-step, the particle either moves one step forward, or one step back, with probability 0.5 either way. This is a Markov Process with transition probability matrix:

$$\begin{bmatrix} 0 & 0.5 & 0 & 0 & \dots & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & \dots & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.5 & 0 & 0 & 0 & \dots & 0.5 & 0 \end{bmatrix}$$

3.3.3

Example *Finite State Machine with Errors*

A computer executing a program can be regarded as a **finite state automaton** ; that is, it is a **dynamical system** with a finite number of states. Such an automaton can be represented by a **state transition matrix**. If the states of the system are numbered from 1 to N, then the entry a_{ij} in the matrix is 1 if and only if state i (at time n) yields to state j (at time $(n+1)$).

For example, consider a computer "clock", which has N distinct states, and simply counts through them in order, returning to state 1 after passing through state N. In other words, the clock's behaviour is:

$$1,\ 2,\ 3,\ \dots,\ N-1,\ N,\ 1,\ 2,\ 3,\ \dots$$

The state transition matrix for this machine would look like:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

Suppose that there is some sort of transient hardware error in the computer, so that occassionally, a random "misstep" occurs in the computation, and the system goes to the wrong state.

To be concrete, let's suppose the system has 100 states, and suppose that, during each time-step, it has a 0.1 probability of experiencing an error, in which case it will jump to a random state, with all 100 states being equally probable as outcomes of the error (including the "correct" state) . The transition probability matrix will then resemble:

$$\begin{bmatrix} 0.001 & 0.901 & 0.001 & 0.001 & \dots & 0.001 & 0.001 \\ 0.001 & 0.001 & 0.901 & 0.001 & \dots & 0.001 & 0.001 \\ 0.001 & 0.001 & 0.001 & 0.901 & \dots & 0.001 & 0.001 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.901 & 0.001 & 0.001 & 0.001 & \dots & 0.001 & 0.001 \end{bmatrix}$$

3.3.4

Example *Radioactive decay*

During each second, a certain atom of a radioactive material has a 0.001 chance of decaying. Once it has decayed, of course, it will never reconstitute itself into its original state. If the "decayed" state is called "1", and the "undecayed" state is called 0, then the transition probability matrix is:

$$\begin{bmatrix} 0.999 & 0 \\ 0.001 & 1 \end{bmatrix}$$

3.3.3 Transition Probabilities: Arbitrary Statespace

It appears that the proper way to think of a Markovian process on a discrete space is in terms of a linear transformation which acts to transform probability vectors into other probability vectors. This provides the appropriate avenue for the generalisation of the Markovian property to an arbitrary statespace. So, let $(X, \mathcal{X})$ be some measurable statespace, and let $\mathcal{M} := \mathcal{M}(X, \mathcal{X})$ be the vector space of all signed measures on $(X, \mathcal{X})$. Let $\triangle := \triangle(X, \mathcal{X}) \subset \mathcal{M}(X, \mathcal{X})$ be the set of all probability measures on X.

A linear transformation $\mathbf{B} : \mathcal{M} \longrightarrow \mathcal{M}$ will be called **stochastic** if $\mathcal{B}$ sends $\triangle$ to itself.

The transition mechanism of a general Markov process is defined by a stochastic linear transformations. This perhaps seems unnecessarily abstract. After all, naively, it seems we could just proceed as follows: In an arbitrary measure space, "events" do not correspond to points in X, but rather, measurable subsets. However, we can still define the concept of a **transition probability** ; given any measurable subsets U and $V \subset X$, we could define a transition probability:

$$B_{U\to V} \;:=\; \mathbf{P}_{rob}\left[X_{n+1}\in V \;\middle|\; X_n \in U \right].$$

The problem with this approach is that the actual probability of the event V occuring at time $n+1$ is not determined by the mere fact that the event U occured at time n; we must also have knowledge of the actual probability distribution (at time n) *within* U.

Hence, what we *really* want to do is this: given a probability distribution, μ_n, for the state at time n, we want to assign some corresponding probability distribution μ_{n+1} for the state at time $(n+1)$. A little consideration will reveal that this function $\mu_n \mapsto \mu_{n+1}$ must be **linear** as a function of μ_n. Hence, we are led to a linear function on the space of measures, as described above.

So, we have a linear transformation

$$\mathbf{B} : \mathcal{M} \longrightarrow \mathcal{M}$$

which preserves $\triangle$. This transform is defined by the property:

If the state of the system at time n is a random variable whose probability distribution is given by a measure μ on X, then the state of the system at time $(n+1)$ is given by the measure

$$\mathbf{B}\mu$$

Exercise

1. Suppose that, more generally, for any times $m > n$, we define stochastic linear transforms $\mathbf{B}^{[n\to m]}$ by the condition:

   If the state of the system at time n is a random variable whose probability distribution is given by a measure μ on X, then the state of the system at time m is given by the measure

   $$\mathbf{B}^{[n\to m]}\mu$$

   Then show that

   $$\mathbf{B}^{[n\to m]} \;=\; \mathbf{B}^{[(m-1)\to m]} \circ \ldots \circ \mathbf{B}^{[(n+1)\to(n+2)]} \circ \mathbf{B}^{[n\to(n+1)]}$$

2. How might we represent such a stochastic linear transformation as a sort of metaphorical "stochastic matrix"? Well, one way is to imagine it as a sort of "measurable function" taking its values in the space of probability measures. In other words, imagine a function

$$\mathbf{K} : X \longrightarrow \mathcal{M}$$

which is **measurable** in the sense that, for any set $U \in \mathcal{X}$, the function

$$\mathbf{K}[U] : X \ni x \mapsto \mathbf{K}(x)[U] \in \mathbb{R}$$

is a measurable function from X into $\mathbf{R}$.
The function $\mathbf{K}$ can be used to define a linear function $F : \mathcal{M} \longrightarrow \mathcal{M}$ by the formula:

$$F(\mu) \ := \ \int_X \mathbf{K}(x)\ d\mu[x].$$

What exactly is an *integral* of a measure-valued function? Well, $F(\mu)$ is a measure defined as follows: for any measurable subset $U \in X$, we have:

$$F(\mu)[U] \ := \ \int_X \mathbf{K}(x)[U]\ d\mu[x].$$

We then say that $\mathbf{K}$ is the **integral kernel** for F.
Suppose you are given an arbitrary integral kernel K and asked to determine whether it, in fact *could* be the integral kernel for some stochastic linear transform. Show that:

$\mathbf{K}$ *is the integral kernel for a stochastic linear transformation if and only if, for all*[1] x *in* $X, K(x) \in \triangle$.

3. Let $F : \mathcal{M} \longrightarrow \mathcal{M}$ be a linear transformation. The obvious way to define an integral kernel for F is to define $K : X \longrightarrow \mathcal{M}$ by:

$$\text{for all } x \in X, \qquad K(x) \ := \ F(\delta_x)$$

where δ_x is the **point mass** at x. Will this actually work?
Let $\mathcal{S}$ be the complex vector space spanned by the set of **point masses** in $\mathcal{M}$. In other words,

$$\mathcal{S} \ := \left\{ \sum_{n=1}^{N} c_n \cdot \delta_{x_n} \ ; \ N \in \mathbb{N};\ c_1, \ldots, c_N \in \mathbb{C};\ x_1, \ldots, x_N \in X \right\}.$$

[1] Note: *not* "almost all" x, but *all* x in X.

(a) Suppose $\mathcal{M}$ is equipped with a topological vector space structure so that $\mathcal{S}$ is **dense** in $\mathcal{M}$ with respect to this topology. Show that K is an integral kernel for F.

(b) Let X be a compact topological space, then let $\mathcal{M}$ be the space of **Radon Measures** on X, and endow it with the **weak topology** generated by all continuous functions. Then $\mathcal{S}$ is dense in $\mathcal{M}$.

(c) Let $(X, \mathcal{X})$ be a measurable space, and let $\mathcal{M}$ be the space of all measures on X of **bounded total variation**, endowed with the norm topology induced by the **total variation norm**. Then, in general, $\mathcal{S}$ is *not* dense in $\mathcal{M}$.

3.3.5

Example *A Random Walk on* $\mathbf{Z}$

Now, consider a **random walk** on the integers. During each time step, we move either forwards or backwards by one unit, moving in either direction with probability 0.5. We can again represent the transition probabilities via a sort of "integral kernel"; in this case, it will be a sort of infinite matrix, indexed by the integers. The central portion of this matrix will resemble:

$$\begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \dots & 0 & 0.5 & 0 & 0 & 0 & \dots \\ \dots & 0.5 & 0 & 0.5 & 0 & 0 & \dots \\ \dots & 0 & 0.5 & \boxed{0} & 0.5 & 0 & \dots \\ \dots & 0 & 0 & 0.5 & 0 & 0.5 & \dots \\ \dots & 0 & 0 & 0 & 0.5 & 0 & \dots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where the boxed entry has coordinates $(0, 0)$.

3.3.6

Example *A Gaussian Random Walk on* $\mathbb{R}$

Now suppose the particle randomly moves on the real line. During each successive time-step, the particle experiences a random spatial displacement which has a normal distribution with mean 0 and variance 1. We can again represent the transition probabilities in terms of an integral kernel. This time, the integral kernel will be an integrable function $\Phi : \mathbb{R} \longrightarrow \mathcal{M}(\mathbb{R})$. In

fact, for each fixed $x \in \mathbb{R}$, $\Phi(x)$ will be a Gaussian probability measure with mean x and variance 1. Hence, we can really visualise Φ as a measurable function $\Phi : \mathbb{R} \times \mathbb{R} \longrightarrow [0, \infty)$, where, for each $x \in \mathbb{R}$, the function $\Phi(x, \bullet) : \mathbb{R} \longrightarrow [0, \infty)$ is given by the function

$$\Phi(x, y) \;=\; \frac{1}{\sqrt{2\pi}} \exp(\|x - y\|^2)$$

We can then visualise this as a sort of "matrix", whose rows and columns are indexed by $\mathbb{R}$.

Notice that this integral kernel is closely related to the **Solution Kernel** for the **Heat Equation** . Can you guess why?

3.3.4 Instantaneous State Distributions

The **transition probabilities** of a Markov process describe how the state of the system changes randomly over time. However, suppose someone asks you, "What is the probability distribution of states at time n?"

If you had a probability distribution, μ_m, describing the state of the system at some time $m < n$, then you could answer them easily: The probability distribution of states at time n is just $\mathbf{B}^{n-m}(\mu)$, where $\mathbf{B}$ is the stochastic linear transformation describing the transition probabilities.

However, what if you *didn't* have such a μ_m? You'd be out of luck.

So, for our picture to be complete, we must provide probability distributions for the *instantaneous state* of the system. No matter what time n one is inquiring about, we must have some probability distribution μ_m, describing the probable state of the system at some time $m < n$. In other words, we must possess a collection of **instantaneous state distributions**, dating back into the arbitrarily distant past; that is, a collection of measures

$$\{\mu_n \;;\; n \in \mathbf{T}\}$$

where $\mathbf{T} \subset \mathbb{T}_{ime}$ is some subset of $\mathbb{T}_{ime}$ which has the property:

$$\forall n \in \mathbb{T}_{ime},\; \exists m \in T \text{ such that } m \leq n.$$

Furthermore, this collection of instantaneous state distributions must be *consistent*, in the sense that, if μ_m and μ_n are two such measures, with $m < n$, then

$$\mu_n \;=\; \mathbf{B}^{n-m}(\mu_m).$$

Clearly, once we possess such a consistent sequence of instantaneous state distributions, we can complete it to a unique, consistent collection,

$$\{\mu_n \; ; \; n \in \mathbb{T}_{ime}\},$$

of instantaneous state distributions defined for *all* times. Once we possess this data, we have sufficient data to completely describe the process.

3.3.7

Definition *Discrete Time Markov Process*

Let $(X, \mathcal{X})$ be a measurable space. A **Discrete Time Markov Process** on X is given by a pair $(\mathbf{B}, \{\mu_n \; ; \; n \in \mathbf{Z}\})$, where

$$\mathbf{B} : \mathcal{M}(X, \mathcal{X}) \longrightarrow \mathcal{M}(X, \mathcal{X})$$

is a stochastic linear transformation on the space of measures over X, and $\ldots, \; \mu_{-2}, \; \mu_{-1}, \; \mu_0, \; \mu_1, \; \mu_2, \; \ldots$ is a sequence of probability measures so that, for every $n \in \mathbf{Z}$,

$$\mathbf{B}(\mu_n) \; = \; \mu_{n+1}$$

Stationary Markov Processes

However, this infinite sequence of measures is a lot of information to keep track of. Wouldn't it be nicer if there were some *single* state probability distribution we could provide, that would answer the question for *all* times?

Naively, we might be inclined to merely give the state probability distribution at time zero, and expect that the transition probability transformations could be used to extrapolate probabilities both forwards and backwards. In general, however, this is not the case. For a **finite** statespace, it is easy to construct the "reverse time" transition probability matrix (**Exercise**: Do this. Hint: it is *not* just the inverse of the forward-time matrix). However, for an arbitrary statespace, things aren't this easy.

Ideally, we want to find a single state probability distribution that will be consistent at any time. In other words we want a probability measure μ which is **stationary** under transformation by $\mathbf{B}$. That is, we want a probability measure which is an **eigenvector** of $\mathbf{B}$.

3.3.8

Definition *Stationary Markov Process*

Let $(X, \mathcal{X})$ be a measurable space. A **Stationary Discrete Time Markov Process** on X is given by a pair $(\mathbf{B}, \mu)$, where

$$\mathbf{B} : \mathcal{M}(X, \mathcal{X}) \longrightarrow \mathcal{M}(X, \mathcal{X})$$

is a stochastic linear transformation on the space of measures over X, and μ is a *probability measure* on X that is an *eigenvector* of $\mathbf{B}$.

3.3.9

Example

Recall the example of the random walk on the cyclic group $G := \mathbf{Z}/N\mathbf{Z}$, having transition probability matrix

$$\begin{bmatrix} 0 & 0.5 & 0 & 0 & \dots & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & \dots & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0.5 & 0 & 0 & 0 & \dots & 0.5 & 0 \end{bmatrix}$$

The only probability vector which is an eigenvector of this matrix is the equidistributed probability distribution:

$$\mu := \left[\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right].$$

(**Exercise**: Check this).

Ahistorical Markov Processes

Sometimes, however, there *is* no stationary probability distribution. Indeed, if we allow a particle to randomly walk over an unbounded space, then, as time goes to infinity, the particle's probability distribution will become "infinitely diffused" –in other words, the probability measure for any compact region goes to zero. We might call such a Markov process **diffusive**....

3.3.10

Definition *Diffusive*

Let X be a topological space, and let $\mathcal{M}$ be the vector space of Borel measures on X. Let $\mathbf{B} : \mathcal{M} \longrightarrow \mathcal{M}$ be a stochastic linear transformation describing some markov process.

Say that $\mathbf{B}$ is **diffusive** if, for any $\mu \in \mathcal{M}$, the sequence $\left\{\mathbf{B}^n(\mu)|_{n \in \mathbb{N}}\right\}$ converges **weakly** to zero. That is, for any compact subset $K \subset X$, we have

$$\lim_{n \to \infty} \mathbf{B}^n(\mu)[K] \quad = \quad 0.$$

If we allow time to begin in the "infinite past", then we have a problem: by Time 0, the probability distribution is *already* infinitely diffused. An "infinitely diffused" probability measure makes sense as a limiting condition, but it definitely does not make sense as a probability distribution at time zero. Thus, in these sorts of systems, we *cannot* start the flow of time in the infinite past; we would get a contradiction if we did this.

Exercise: Show that the random walks described before on $\mathbb{Z}$ and $\mathbb{R}$ are both **diffusive**.

Thus, in a diffusive Markov process, $\mathbb{T}_{ime}$ itself must have a "beginning"; hence, rather than indexing Time with $\mathbb{Z}$, we must index it with $\mathbb{N}$.

A Markov process defined with $\mathbb{T}_{ime} = \mathbb{N}$ has "no history". Hence, it is called **ahistorical**.

3.3.11

Definition *Ahistorical Discrete Time Markov Process*

Let $(X, \mathcal{X})$ be a measurable space. An **ahistorical discrete time Markov process** on X is given by a pair $(\mathbf{B}, \{\mu_n \ ; \ n \in \mathbb{N}\})$, where

$$\mathbf{B} : \mathcal{M}(X, \mathcal{X}) \longrightarrow \mathcal{M}(X, \mathcal{X})$$

is a stochastic linear transformation on the space of measures over X, and $\mu_0, \ \mu_1, \ \mu_2, \ \ldots$ is a sequence of probability measures so that, for every $n \in \mathbb{N}$,

$$\mathbf{B}(\mu_n) \;=\; \mu_{n+1}$$

3.3.12

Example

1. Recall the random walk on $\mathbf{Z}$. Suppose that, at time zero, the particle is almost surely located at the origin. We then get a sequence of probability measures $\mu_0, \mu_1, \mu_2, \ldots$, which can be represented as $\mathbf{Z}$-indexed vectors.

   These vectors comprise the rows of a *renormalised* version of **Pascal's Triangle**:

$$\begin{array}{lcccccccccc}
\mu_0 & = & \ldots 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0\ldots \\
\mu_1 & = & \ldots 0 & 0 & 0 & \boxed{\frac{1}{2}} & 0 & \boxed{\frac{1}{2}} & 0 & 0 & 0\ldots \\
\mu_2 & = & \ldots 0 & 0 & \boxed{\frac{1}{4}} & 0 & \boxed{\frac{2}{4}} & 0 & \boxed{\frac{1}{4}} & 0 & 0\ldots \\
\mu_3 & = & \ldots 0 & \boxed{\frac{1}{8}} & 0 & \boxed{\frac{3}{8}} & 0 & \boxed{\frac{3}{8}} & 0 & \boxed{\frac{1}{8}} & 0\ldots \\
\mu_4 & = & \ldots & \boxed{\frac{1}{16}} & 0 & \boxed{\frac{4}{16}} & 0 & \boxed{\frac{6}{16}} & 0 & \boxed{\frac{4}{16}} & 0 & \boxed{\frac{1}{16}}\ldots \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

   We have placed boxes around some entries to highlight the **Pascal's Triangle** structure.

   Exercise: Verify this.

2. Recall the Gaussian random walk on $\mathbb{R}$. Again, suppose that, at time zero, the particle is almost surely located at the origin. We then get a sequence of probability measures, $\mu_0, \mu_1, \mu_2, \ldots$ which are Gaussian distributions of ever-increasing variance:

$$\begin{array}{lcl}
\mu_0 & = & \overbrace{0;0} \\
\mu_1 & = & \overbrace{0;1} \\
\mu_2 & = & \overbrace{0;2} \\
\mu_3 & = & \overbrace{0;3} \\
\vdots & \vdots & \vdots
\end{array}$$

Here, $\overbrace{0;0}$ represents the "point mass at zero", which physicists sometimes call the **Dirac Delta Function**.

Exercise: Verify this.

3.3.5 Formal Construction as a Stochastic Process

The formal definition of a stochastic process is in terms of a probability measure on the space of paths in some statespace. The definitions of Markov processes we have just given, while intuitively appealing, are not yet in this form. How can we reconceive a Markov Process as a stochastic process in the traditional sense?

Suppose we have a measurable statespace $(X, \mathcal{X})$, and a discrete timeline $\mathbb{T}_{ime}$, and suppose we have a general Markov process, given by a transition probability transform $\mathbf{B} : \mathcal{M}(X, \mathcal{X}) \longrightarrow \mathcal{M}(X, \mathcal{X})$, and a $\mathbb{T}_{ime}$–indexed sequence of probability measures on X, say, $\{\mu_n \ ; \ n \in \mathbb{T}_{ime}\}$. Suppose that $K : X \longrightarrow \mathcal{M}$ is an **integral kernel** for $\mathbf{B}$.

We want to construct a suitable probability measure on the space $X^{\mathbb{T}_{ime}}$, which somehow represents the information encoded in $\mathbf{B}$ and $\left\{\mu_n|_{n \in \mathbb{T}_{ime}}\right\}$. To do this, we will employ Kolmogorov's Theorem [4.2.1] .

By the results of [4.2.1] , it suffices to construct a consistent family of measures upon some **filter** inside $\mathbb{T}_{ime}$. We will use the filter of **finite intervals** in $\mathbb{T}_{ime}$:

$$\mathcal{A} \ := \ \{[n..m] \ ; \ n, m \in \mathbb{T}_{ime}\}$$

and, for all $A \in \mathcal{A}$, we will define measures μ_A as follows.

Singleton sets First, of course, when $A = \{n\}$ for some $n \in \mathbb{T}_{ime}$, define

$$\mu_A \ := \ \mu_n.$$

Pairs Next, consider $A = \{n-1, n\}$. Define the measure μ_A on the space $X^A \simeq X^2$ by:

$$\mu_A \ := \ \int_X (\delta_x \times K(x)) \ d\mu_{n-1}[x]$$

Exercise:

1. Verify that $\mathbf{pr}_1^*(\mu_A) \ = \ \mu_{n-1}$.
2. Verify that $\mathbf{pr}_2^*(\mu_A) \ = \ F(\mu_{n-1}) = \mu_n$.

3. In the case when X is a *finite* statespace, this formula is really a formula about matrices and vectors; express it in this form.

Induction More generally, suppose $A := [n...(m-1)]$, and assume inductively that μ_A has already been defined. Then define $\mu_{[n..m]}$ by:

$$\mu_{[n..m]} := \int_{X^A} \delta_{\vec{x}} \times K(x_{m-1})\, d\mu_A[\vec{x}]$$

(where $\vec{x} := (x_n,\ x_{n+1},\ \ldots, x_{m-1})$ ranges over X^A) .
Exercise:

1. Verify that $\mathbf{pr}_A^*(\mu_{[n..m]}) = \mu_A$.

2. Verify that $\mathbf{pr}_n^*(\mu_{[n..m]}) = F(\mu_{m-1}) = \mu_m$.

We have thus built a collection of marginal measures over the filter $\mathcal{A}$, satisfying the Kolmogorov Criterion. Hence, by Kolmogorov's Theorem, we can extend this to a well-defined probability measure on $X^{\mathbb{T}_{ime}}$, which defines the desired stochastic process.
□

Chapter 4

Continuous Time Stochastic Processes

4.1 Basic Ideas

4.1.1 Introduction

Prerequisites: [2]

A **continuous time stochastic process** is a probability distribution on a space of *continuous time* paths through a statespace. Formally...

4.1.1

Definition *Continuous Time Stochastic Process*

Let X be some set, and let $\mathbb{T}_{ime}$ be some open or closed interval in $\mathbb{R}$ representing an interval of time[1], and let $\mathcal{W}$ be some sigma-algebra on $X^{\mathbb{T}_{ime}}$,

Then a $\mathcal{W}-$**measurable continuous-time stochastic process** on the **statespace** X, over the **time interval** $\mathbb{T}_{ime}$, is a probability measure on $\mathcal{W}$.

[1] Common choices for $\mathbb{T}_{ime}$ are $[0,\infty)$, $[0,1]$, or all of $\mathbb{R}$

4.1.2 Informal Examples

Here we mention a few of the most important examples of continuous time stochastic processes, and provide references to informal introductory discussions of each process; these discussions can be read now, without any further preparation by the reader.

1. **Brownian Motion** [4.3.1].
2. **Stable Processes** [4.4.1].
3. **Poisson Processes** [4.5.1].
4. **Lévy Processes** [4.6.1].

4.1.3 Sample Path Properties

We often imagine a stochastic process as describing the path of a moving particle. It would be nice to know what sorts of topological or analytic properties the "typical" trajectory of such a particle possesses. For example, if we are to take Brownian motion seriously as describing the physical trajectory of a dust mote moving through space, then we would hope that the "typical" trajectory described by Brownian motion is *continuous*. Indeed, we would even like it to be *differentiable*, since it is generally accepted that "physical" phenomena always behave in a differentiable way.

Properties of a stochastic process which deal with the qualitative characteristics of a "typical" trajectory are called **sample path properties**. The investigation of sample path properties is actually more subtle than one might think, because it is possible to construct different "versions" of the same process whose sample paths actually are qualitatively different.

How can this be? How can two processes with identical statistical properties be different? Well, it depends on just what we mean by "identical statistical properties".

For example, often a stochastic process is "defined" by specifying its finite-dimensional marginal distributions. Thus, we might call two processes "statistically identical" if all of their finite-dimensional marginal distributions agree. Unfortunately, knowledge of the *finite* dimensional marginal distributions is insufficient to uniquely characterise the process; different processes may possess the same *finite* dimensional distributions, but differ at some more subtle level.

In particular, **sample-path properties** such as *continuity* or *differentiability* depend upon specification of the properties of the path at an infininite number of points. Thus, these properties may not be entirely fixed by knowledge of finite-dimensional behaviour.

Hence, a question like "Are Brownian motion paths almost surely continuous?" is in fact rather ill-posed. The proper question is: "Does there exist a *version* of Brownian motion where almost all paths are continuous?"

But what, exactly, is a "version" of Brownian motion?

4.1.4 Versions and Indistinguishibility

4.1.2

Definition *Version*

Let $\mathbb{T}_{ime}$ be some indexing set, X be some statespace, and let $\alpha, \beta : \Omega \times \mathbb{T}_{ime} \longrightarrow X$ be two stochastic processes indexed by $\mathbb{T}_{ime}$.

We say that α and β are **versions** of one another if, for every $t \in \mathbb{T}_{ime}$,

$$\mathbf{P}_{rob}\left[\alpha_t = \beta_t\right] \;=\; 1.$$

4.1.3

Remark

Thus, if β is a **version** of α, then all countable-dimensional projections of α and β are almost surely equal. In particular, all finite projections of α and β are almost surely equal.

4.1.4

Definition *Indistinguishible*

Let $\mathbb{T}_{ime}$ be some indexing set, X be some statespace, and let $\alpha, \beta : \Omega \times \mathbb{T}_{ime} \longrightarrow X$ be two stochastic processes indexed by $\mathbb{T}_{ime}$.

We say that α and β are **indistinguishible** if

$$\mathbf{P}_{rob}\left[\forall t \in \mathbb{T}_{ime},\ \alpha = \beta\right] \;=\; 1.$$

4.1.5

Remark

If $\mathbb{T}_{ime}$ is a **countable** set, then α and β are two versions if and only if they are indistinguishible.

However, $\mathbb{T}_{ime}$ is usually uncountable, in which case indistinguishibility is a much stronger relationship. As an illustration, consider the following:

4.1.6

Proposition

Suppose $\alpha : \Omega \times [0, \infty) \longrightarrow \mathbb{R}$ is a stochastic process with almost surely *continuous* sample paths.

1. Any process indistinguishible from α also will have almost surely continuous sample paths.
2. It is always possible to construct *versions* of α with almost surely *discontinuous* sample paths

Proof: 1 This follows immediately from the definition of indistinguisibility; if β is indistinguishible from α, then almost all of their paths are identical, point-for-point. Hence, almost all paths of β are continuous.

2 Let $\Phi : \Omega \longrightarrow [0, \infty)$ be some function so that, for any $t \in \mathbb{T}_{ime}$, the preimage $\Phi^{-1}(t)$ is at most countable in Ω.

Define β, a **version** of α, as follows:

$$\beta(t, \omega) \;=\; \begin{cases} \alpha(\omega, t) & if \quad t \neq \Phi(\omega) \\ \alpha(\omega, t) + 1 & if \quad t = \Phi(\omega) \end{cases}$$

Now, for any fixed $t \in [0, \infty)$, there are at most a countable collection of ω so that β_t differs from α_t. Thus,

$$\mathbf{P}_{rob}\left[\beta_t = \alpha_t\right] \;=\; 1.$$

However, by construction, for *every* ω, there is at least one point where the path $\beta(\omega)$ makes a sudden discontinuous jump; hence, β has almost surely discontinuous paths.

□

This example may seem artificial, since the discontinuities we have introduced are clearly isolated and removable. However, it is not much harder to build a version of α so that almost every sample path possesses a countable set of discontinuities which are **dense** in $[0, \infty)$ (**Exercise**: Do this.). Hence, we can construct two versions of the same process, one with paths which are almost surely continuous everywhere on $[0, \infty)$, the other with paths which are almost surely continuous nowhere on $[0, \infty)$!

4.1.5 Stochastic Processes on Topological Spaces

Prerequisites: [4.1.1]

Now, suppose our statespace has some topological structure, and suppose that we want to impose the condition that all sample paths are **continuous.**

A **continuous stochastic process** is a continuous-time stochastic process on X which is almost surely continuous. Formally...

4.1.7

Definition *Continuous Stochastic Process (I)*

Let X be a topological space, and $\mathbb{T}_{ime} \subset \mathbb{R}$. A **continuous stochastic process** on statespace X, with timeline $\mathbb{T}_{ime}$, is a **probability measure** on $X^{\mathbb{T}_{ime}}$ whose support is confined entirely to $\mathbf{C}(\mathbb{T}_{ime}, X)$, the space of **continuous** functions from $\mathbb{T}_{ime}$ into X.

Unfortunately, it may be very difficult, in general, to explicitly construct a probability measure on $X^{\mathbb{T}_{ime}}$ with this property. For one thing, the subset $\mathbf{C}(\mathbb{T}_{ime}, X) \subset X^{\mathbb{T}_{ime}}$ may not even be *measurable* with respect to the sigma-algebra $\mathcal{W}$.

Hence, we will obviate the technical problems by simply bypassing the space $X^{\mathbb{T}_{ime}}$ (which is much too large), and defining the measure directly on $\mathbf{C}(\mathbb{T}_{ime}, X)$. This motivates a second definition:

4.1.8

Definition *Continuous Stochastic Process (II)*

A **continuous stochastic process** is a probability measure on $\mathbf{C}(\mathbb{T}_{ime}, X)$.

The sigma algebra on $\mathbf{C}(\mathbb{T}_{ime}, X)$ is usually the **Borel sigma-algebra generated** by some some topology The most common choices for this topology are listed in order from coarsest to finest:

- The **Tychonoff Topology** on $\mathbf{C}(\mathbb{T}_{ime}, X)$ is generated by all **finite dimensional cylinder sets** of the form

$$\left(\overline{\substack{t_1,\ldots,t_N \\ U_1,\ldots,U_N}}\right) \;:=\; \{\alpha : \mathbb{T}_{ime} \longrightarrow X \;;\;\; \forall n \in [1..N],\; \alpha(t_n) \;\subset U_n\}$$

  where $\{t_1, \ldots, t_N\}$ is allowed to be any **finite** subset of $\mathbb{T}_{ime}$, and $U_1, \ldots, U_N$ are any **open** subsets of X.

- The **Compact-Open Topology** on $\mathbf{C}(\mathbb{T}_{ime}, X)$ is generated by all **CO-cylinder sets** of the form

$$\left(\overline{\substack{K \\ U}}\right) \;:=\; \{\alpha : \mathbb{T}_{ime} \longrightarrow X \;;\;\; \alpha(K) \;\subset U\}$$

  where K ranges over all **compact** subsets of $\mathbb{T}_{ime}$, while U ranges over all **open** subsets of X.

- **Uniform topology:** If X is a compact metric space, then we can consider the **uniform metric**, d_∞, on $\mathbf{C}(\mathbb{T}_{ime}, X)$, where

$$d_\infty(\alpha, \beta) \;:=\; \sup_{t \in \mathbb{T}_{ime}} d\left(\alpha(t), \beta(t)\right).$$

  the corresponding topology on $\mathbf{C}(\mathbb{T}_{ime}, X)$ is called the **uniform topology**.

Exercise:

1. Show that the topologies listed above appear in order of increasing fineness.

2(a) If $\mathbb{T}_{ime}$ is **discrete**, then the **Compact-Open** topology is identical with the **Tychonoff** topology

(b) On the other hand, if X is a metric space and $\mathbb{T}_{ime}$ is compact, then the **Compact-Open** topology is identical with the **uniform** topology

3(a) The Borel sigma-algebra generated by the Tychonoff product topology is identical with the **cylinder algebra**, $,^{\otimes \mathbb{T}_{ime}; \mathcal{B}(X)}$ where $\mathcal{B}(X)$ is the Borel sigma-algebra on X.

(b) If X be a metric space, then the **Tychonoff Topology** and the **uniform topology** on $\mathbf{C}(\mathbb{T}_{ime}; X)$ induce the *same* Borel sigma-algebras.

Hint: Show that every neighbourhood in the uniform norm can be written as a $\mathbf{G}_\delta$ subset with respect to the Tychonoff topology. Use the fact that $\mathbb{T}_{ime}$ is **second-countable**.

4(a) The set $\mathbf{C}(\mathbb{T}_{ime}; X)$ is *not* a measurable subset of $X^{\mathbb{T}_{ime}}$ with respect to the cylinder sigma algebra.

(b) On the other hand, the set $\mathbf{C}^1(\mathbb{T}_{ime}; X)$ *is* a measurable subset of $\mathbf{C}(\mathbb{T}_{ime}; X)$ with respect to the Borel sigma algebra on $\mathbf{C}(\mathbb{T}_{ime}; X)$

(c) On what level of the **Borel Heirarchy** does $\mathbf{C}^1(\mathbb{T}_{ime}; X)$ lie, within the Borel sigma-algebra on $\mathbf{C}(\mathbb{T}_{ime}; X)$?

4.1.6 Stochastic Processes on Vector Spaces

Prerequisites: [4.1.1]

Let V be some real vector space, with sigma-algebra $\mathcal{V}$. If $(\Omega, \mathcal{W}, \mathbf{P})$ is our universal sample space, then a **stochastic process** on V is a measurable function

$$\beta : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

The advantage of working on a vector space is that we can now think of a sample path as being the summation of a large number of random spacial displacements or **increments**, occuring one after another in time.

4.1.9

Definition *Increments*

Let β be a stochastic process on V. For any times t_0 and t_1, the **increment** of β between times t_0 and t_1 is the random variable, $\beta(t_0 \to t_1)$, defined:

$$\beta(t_0 \to t_1) \ := \beta(t_1) - \beta(t_0)$$

In other words, $\beta(t_0 \to t_1)$ is a measurable function from Ω into V, defined:

$$\beta(t_0 \to t_1)(\omega) \ := \ \beta_\omega(t_1) - \beta_\omega(t_0),$$

for any $\omega \in \Omega$.

The increments contain information describing the "dynamics" of the stochastic process. If we knew the initial position of a particle in space, and we knew its *changes in position* over all subsequent time intervals, we could use this information to reconstruct its entire trajectory. In a similar fashion, if we have random variables describing the "initial state" of the process β, along with random variables describing all subsequent increments, and we know the *joint distribution* of all of these variables, then we can use this information to reconstruct the entire process.

If a stochastic process describes the random motion of some particle in space, does this particle have "momentum"? If it does, then its past movements should exert some lingering effect on its future movements. If not, then we expect that its past and future motions would be completely independent....

4.1.10

Definition *Independent Increments*

The stochastic process β has **independent increments** if, for any times $t_0 < T_0 < t_1 < T_1 \in \mathbb{T}_{ime}$, the the increments $\beta(t_0 \to T_0)$ and $\beta(t_1 \to T_1)$ are **independent** random variables.

4.1.11

Proposition

Suppose that β has independent increments. Then, for any $t_0 < t_1 < t_2 \in \mathbb{T}_{ime}$, we have:

$$\mathbf{D}^{istr}\left[\beta(t_0 \to t_2)\right] \quad = \quad \mathbf{D}^{istr}\left[\beta(t_0 \to t_1)\right] \ * \ \mathbf{D}^{istr}\left[\beta(t_1 \to t_2)\right]$$

Proof: Exercise. □

Imagine that the stochastic process β describes the behaviour of some physical system. A basic scientific assumption is that the "laws of physics"

are unchanging over time. Thus, if the increments describe the "dynamics" of β, then these dynamics should be **invariant** under **time translation**.

4.1.12

Definition *Stationary Increments*

The stochastic process β has **stationary increments** if, for any times t_0 and $t_1 \in \mathbb{T}_{ime}$, and any displacement $t \in \mathbb{R}$ such that $t_0 - t$ and $t_1 - t$ are also in $\mathbb{T}_{ime}$, we have

$$\beta(t_0 \to t_1) = \beta\Big((t_0 - t) \to (t_1 - t)\Big)$$

So, for example, if $0 \in \mathbb{T}_{ime}$, then we can write:

$$\forall t_0, t_1 \in \mathbb{T}_{ime}, \quad \beta(t_0 \to t_1) = \beta\Big(0 \to (t_1 - t_0)\Big)$$

Convention:
The hypothesis of stationary increments says that it doesn't really matter *when* in time we begin our stochastic process. Thus, from now on, *we will assume that* $\mathbb{T}_{ime}$ *always contains the point* 0 (either at its beginning, or somewhere in its middle). If $\mathbb{T}_{ime}$ did not contain 0, then we could simple "shift" the whole process in time until it did, without introducing any technical complications.

The increments of β are often stationary, even when β itself is not.

4.1.13

Example *Brownian Motion*

Brownian Motion [4.3.2] is a stochastic process on $\mathbb{R}^D$ which **originates at zero**, and has **independent, Gaussian increments**. In particular, for any $t, T \in [0, \infty)$, the increment $\beta(t \to T)$ is a **normal variable** with **variance** $(T - t)^{1/2}$. Hence, Brownian motion has stationary increments.

Brownian motion can be interpreted as describing the "random diffusion" of a single particle away from the origin. Because this particle can diffuse throughout the unbounded extent of V, as time progresses, its expected distance from 0 becomes greater and greater. Hence, the process β cannot be stationary. However, by construction, its increments *are* stationary.

4.1.14

Example *Brownian Motion with Drift*

Brownian Motion with drift [4.3.3] also has stationary increments. If the **drift** vector is $\vec{v} \in \mathbb{R}^D$, then, for any $t, T \in [0, \infty)$, the increment $\beta(t \to T)$ is a **normal variable** with **variance** $(T-t)^{1/2}$ and **mean** $\vec{v}$.

4.1.15

Definition L

A **Lévy process** on a vector space V is any continuous time stochastic process with *stationary, independent increments.*

For details on Lévy processes, see [4.6.1].

As well as being unchanging over *time*, physical laws are normally hypothesised to be the same everywhere in *space.* In other words, the dynamics driving our stochastic process should be insensitive to its spacial position.

4.1.16

Definition *Spacially Homogeneous Increments*

The stochastic process β has **spacially homogeneous increments** if, for all $t, s \in \mathbb{T}_{time}$, , the random variable $\beta(t \to s)$ is **independent** of the random variable $\beta(t)$.

4.1.17

Remark

Do not confuse **spacial homogeneous**, **stationary**, and **independent** increments; although the three properties are related, they are fundamentally different.

4.1.7 Expansion and Origination

Prerequisites: [4.1.6]

4.1.18

Definition *Origin*

Let X be any statespace, and let $\beta : \mathbb{T}_{ime} \times \Omega \longrightarrow X$ be a stochastic process. Let $x_0 \in X$ and $t_0 \in \mathbb{T}_{ime}$. We say that β **originates at location** x_0 **at time** t_0 if

$$\beta(t_0) \;=\; x_0 \;\; \textit{almost surely.}$$

Normally, we chose to parameterise time so that $t_0 = 0 \in \mathbb{T}_{ime}$. Also, if X is a vector space, we can translate the stochastic process so that $\vec{x}_0 = \vec{0} \in X$. We will say that a stochastic process β on a vector space **originates at zero** if

$$\beta(0) \;=\; \vec{0} \;\; \textit{almost surely.}$$

Intuitively, we can imagine that the sample paths of the stochastic process slowly "spread out" from this origin point as time passes.

More generally, if we condition upon the state $\{\beta(t_0) = x_0\}$ for any fixed t_0 and x_0, we can imagine the paths of the process "expanding" away from this point as time passes. We might then ask: what is the rate of this expansion?

Suppose X was a **metric space**, and T and t were in $\mathbb{T}_{ime}$, with $t < T$. Intuitively, we could measure the *rate of expansion* of β between t and T:

$$\sigma_\beta(t \to T) \ := \ \overset{stan}{dev}\left[\beta(t \to T)\right]$$

In other words,

$$\sigma_\beta(t \to T) \ := \ \left(\mathbf{E}^{xpct}\left[\|\beta(t \to T)\|^2\right]\right)^{\frac{1}{2}}$$

Thus,

$$\sigma_\beta^2(t \to T) \ = \ \mathbf{var}\left[\beta(t \to T)\right] \ = \ \mathbf{E}^{xpct}\left[\|\beta(t \to T)\|^2\right]$$

4.1.19

Remark

1. Suppose that β **originates at zero**. Then, for any time t,

$$\sigma_\beta(0 \to t) \ = \ \overset{stan}{dev}\left[\beta(t)\right].$$

2. If β has **independent increments**, then, for any times $t_0 < t_1 < t_2$ in $\mathbb{T}_{ime}$,

$$\sigma_\beta^2(t_0 \to t_2) \ = \ \sigma_\beta^2(t_0 \to t_1) + \sigma_\beta^2(t_1 \to t_2).$$

3. If β has **spacially homogeneous increments**, then, for any times $t < T$ in $\mathbb{T}_{ime}$,

$$\mathbf{var}\left[\beta(t)\right] \ = \ \mathbf{var}\left[\beta(T)\right] + \sigma_\beta^2(t \to T).$$

Exercise: Check this.

4.1.20

Example Let β be Brownian motion. Then β originates at zero, and, for any t and T in $\mathbb{T}_{time}$, with $t < T$,

$$\sigma_\beta(t \to T) \;=\; \sqrt{t - T}$$

It turns out that this sort of expansion curve is typical of Lévy processes.

4.1.21

Theorem

Let V be a normed vector space, and let $\mathcal{L}$ be a **Lévy process** on V.

1. Suppose that, for some T' and t' in $\mathbb{T}_{time}$, with $t' < T'$, the value $\sigma^2_{\mathcal{L}}(t' \to T')$ is **finite**. Then $\sigma^2_{\mathcal{L}}(t \to T)$ is finite for all T and t, and is in fact a **linear function** of the difference $t - T$, (independently of T or t).

   Indeed, if we define:

   $$\lambda \;:=\; \frac{\sigma^2_{\mathcal{L}}(t' \to T')}{T' - t'}$$

   then, for all t and T in $\mathbb{T}_{time}$, we have:

   $$\sigma^2_{\mathcal{L}}(t \to T) \;=\; \lambda \cdot (t - T).$$

2. Suppose $\mathcal{L}$ also has **spacially homogeneous** increments, and suppose that, for some time $T \in \mathbb{T}_{time}$, we have $\mathbf{var}\,[\mathcal{L}(T)]$ is finite, with value C. Then:

   (a) The function:

   $$\mathbb{T}_{time} \ni t \mapsto \mathbf{var}\,[\mathcal{L}(t)] \in [0, \infty)$$

   is an affine function of t, with slope λ.

   (b) The process $\mathcal{L}$ **originates** at time $T_0 \;:=\; T - \frac{C}{\lambda}$ (at some unspecified location in V).

(c) As a consequence, the process $\mathcal{L}$ *cannot be well-defined before time* T_0, at least, as long as we are to maintain the hypothesis of stationarity of increments[2] Hence, such a process can be well-defined on at most the real *half*-line, $[T_0, \infty)$.

Proof: $\boxed{1}$

Claim 1: For any times $t < T$ in $\mathbb{T}_{time}$, $\sigma^2_{\mathcal{L}}(t \to T)$ must be finite.

Proof: First suppose $T, t \in [t', T')$. Then

$$\sigma^2_{\mathcal{L}}(t' \to T') \;=\; \sigma^2_{\mathcal{L}}(t' \to t) \;+\; \sigma^2_{\mathcal{L}}(t \to T) \;+\; \sigma^2_{\mathcal{L}}(T \to T')$$

(by Remark 4.1.7)

$$\geq \;\; \sigma^2_{\mathcal{L}}(t \to T)$$

so this quantity must be finite.

Next, suppose that T and t are arbitrary. Find some large number so that

$$\frac{T-t}{M} \;<\; T' - t'.$$

and define $\Delta t \;:=\; \frac{T-t}{M}$. Then

$$\begin{aligned}&\sigma^2_{\mathcal{L}}(t \to T)\\ &\quad= \;\; \sigma^2_{\mathcal{L}}(t \to (t+\Delta t)) \;+\; \sigma^2_{\mathcal{L}}((t+\Delta t) \to (t+2\Delta t)) \;+\; \ldots \;+\; \sigma^2_{\mathcal{L}}((t+(m-1)\Delta t) \to T)\end{aligned}$$

(by Remark 4.1.7)

$$= \;\; \underbrace{\sigma^2_{\mathcal{L}}(t \to t+\Delta t) \;+\; \sigma^2_{\mathcal{L}}(t \to (t+\Delta t)) \;+\; \ldots \;+\; \sigma^2_{\mathcal{L}}(t \to (t+\Delta t))}_{M}$$

[2] Theoretically, we could get out of this by specifying that the increments are stationary with one behaviour *after* time T_0, and trivially stationary with constant value 0 for all times *before* T_0.

(by stationarity)

$$\leq \quad M \cdot \sigma^2_{\mathcal{L}}(t' \to T' + \Delta t)$$

(again by stationarity), which is finite, by the previous case.
□[1]

Claim 2:
If we fix $T \in \mathbb{T}_{ime}$, then the function

$$\sigma_T : [0, \infty) \mapsto \sigma^2_{\mathcal{L}}(t \to T + t)$$

is **linear**.

Proof:

Claim 2.1: σ_T is additive.

Proof: For any $t_1, t_2 > T$,

$$\begin{aligned} \sigma_T(t_1 + t_2) &= \sigma^2_{\mathcal{L}}(T \to (T + t_1 + t_2)) \\ &= \sigma^2_{\mathcal{L}}(T \to (T + t_1)) \ + \ \sigma^2_{\mathcal{L}}((T + t_1) \to (T + t_1 + t_2)) \\ &= \sigma^2_{\mathcal{L}}(T \to (T + t_1)) \ + \ \sigma^2_{\mathcal{L}}(T \to (T + t_2)) \end{aligned}$$

(by stationarity)

$$= \quad \sigma_T(t_1) \ + \ \sigma_T(t_2).$$

□[2.1]

Claim 2.2: σ_T is nondecreasing.

Proof: Since σ_T is always nonegative, this follows from 2.1.
□[2.2]

Claim 2.3: Any nondecreasing, finite, additive function must be linear on its domain of definition.

Proof: Exercise. □2.3

□2

Thus, from Claim 2, we conclude that, for any fixed $T \in \mathbb{T}_{ime}$, there is some (nonegative) constant λ_T so that

$$\sigma_T(t) = \lambda_T \cdot t$$

for all t where $\sigma_T(t)$ is well-defined.

By stationarity, this λ_T must be the same for any choice of T. Thus, there is some constant λ so that

$$\sigma^2_{\mathcal{L}}(t \to T) \;=\; \lambda \cdot (t - T).$$

in particular, of course, we must have:

$$\lambda \;=\; \frac{\sigma^2_{\mathcal{L}}(t' \to T')}{T' - t'}.$$

2

This follows from part 1, and the contents of Remark 4.1.7.

□

4.1.22

Corollary

If $\mathcal{L}$ is a **Lévy process**, with increments of **finite variation**, then there is some constant λ so that, for any t and T in $\mathbb{T}_{ime}$, with $t < T$,

$$\sigma_{\mathcal{L}}(t \to T) \;=\; (\sqrt{\lambda}) \cdot \sqrt{t - T}$$

□

4.1.23

Definition *Expansion Coefficient*

Suppose $\mathcal{L}$ is a Lévy process with increments of finite variation. Then the constant λ described in the Theorem 4.1.7 and its corollary is called the **expansion coefficient** of $\mathcal{L}$.

If $\mathcal{L}$ does not have increments of finite variation, then we simply define the expansion coefficient to be ∞.

4.1.8 Balanced Processes

Prerequisites: [4.1.7]

4.1.24

Definition *Balanced increments*

Let V be a vector space, and let β be a stochastic process on V.

The process β has **balanced increments** if, for all $t_0, t_1 \in \mathbb{T}_{ime}$, the random variable $\beta(t_0 \to t_1)$ has zero expectation:

$$\mathbf{E}^{epct}\left[\beta(t_0 \to t_1)\right] \;=\; 0$$

4.1.25

Example *Brownian Motion*

- **Brownian motion** [4.3.1] has balanced increments, because of course Gaussian distributions are symmetrical around zero.
- On the other hand, **Brownian motion with drift** [4.3.3] does *not* have balanced increments.

4.1.26

Remark

If β is a process with **balanced, spacially homogeneous increments**, then β is a **martingale**.

(**Exercise**: Verify this).

Symmetry and Independence

4.1.27

Proposition

Let V be a normed vector space, and let β is a stochastic process on V with **spacially homogeneous, balanced increments** and having **finite expansion coefficient**. Then,

1. In fact, β has **independent** increments.

2. Furthermore, for all $t < s \in \mathbb{T}_{ime}$,

$$\mathbf{E}^{xpct}\left[\langle \beta(t), \beta(s)\rangle\right] \;=\; \overset{stan}{dev}\left[\beta(t)\right].$$

3. Thus, if β **originates at zero**, then, for all $t < s \in \mathbb{T}_{ime}$,

$$\mathbf{E}^{xpct}\left[\langle \beta(t), \beta(s)\rangle\right] \;=\; \lambda t^2.$$

where λ is the **expansion coefficient** of β.

Proof:

□

For fixed t, we can interpret $\beta_t := \beta(t) : \Omega \longrightarrow V$ as a random variable. Let $\mathcal{V}$ be the sigma-algebra on V, and let $\mathcal{W}_t := \beta_t^{-1}(\mathcal{V})$ be the **pulled-back sigma-algebra** on Ω. Thus, $\mathcal{W}_t$ contains all information about the value of β at time t.

Then:

$$\begin{aligned}
\mathbf{E}^{xpct}\left[\langle \beta(t),\ \beta(s)\rangle\right] &= \mathbf{E}^{xpct}\left[\mathbf{E}^{xpct}_{\mathcal{W}_t}\left[\langle \beta(t),\ \beta(s)\rangle\right]\right] \\
&= \mathbf{E}^{xpct}\left[\left\langle \beta(t),\ \mathbf{E}^{xpct}_{\mathcal{W}_t}\left[\beta(s)\right]\right\rangle\right]
\end{aligned}$$

(since $\beta(t)$ is $\mathcal{W}_t$-measurable, by definition)

$$
\begin{aligned}
&= \mathbf{E}^{xpct}\left[\left\langle \beta(t),\ \ \mathbf{E}^{xpct}_{\mathcal{W}_t}\left[\beta(s)-\beta(t)+\beta(t)\right]\right\rangle\right]\\
&= \mathbf{E}^{xpct}\left[\left\langle \beta(t),\ \ \mathbf{E}^{xpct}_{\mathcal{W}_t}\left[\beta(s)-\beta(t)\right]+\mathbf{E}^{xpct}_{\mathcal{W}_t}\left[\beta(t)\right]\right\rangle\right]\\
&= \mathbf{E}^{xpct}\left[\left\langle \beta(t),\ \ \mathbf{E}^{xpct}_{\mathcal{W}_t}\left[\beta(t\to s)\right]\ +\ \beta(t)\right\rangle\right]\\
&= \mathbf{E}^{xpct}\left[\left\langle \beta(t),\ \ \mathbf{E}^{xpct}\left[\beta(t\to s))\right]\ +\ \beta(t)\right\rangle\right]
\end{aligned}
$$

(because the increments are **spacially homogeneous**)

$$= \mathbf{E}^{xpct}\left[\langle \beta(t),\ \ 0\ +\ \beta(t)\rangle\right]$$

(because the increments are **balanced**)

$$
\begin{aligned}
&= \mathbf{E}^{xpct}\left[\langle \beta(t),\ \ \beta(t)\rangle\right]\\
&= \mathbf{E}^{xpct}\left[\left\|\beta(t)\right\|^2\right]\\
&= \sigma_\beta(t)^2.
\end{aligned}
$$

1

Now let $t_1 < t_2 < t_3 \in \mathbb{T}_{time}$. Then

$$
\begin{aligned}
&\mathbf{E}^{xpct}\left[\langle \beta(t_1\to t_2)\ ,\ \beta(t_2\to t_3)\rangle\right]\\
&\quad= \mathbf{E}^{xpct}\left[\langle (\beta(t_2)-\beta(t_1))\ \ ,\ \ (\beta(t_3)-\beta(t_2))\rangle\right]\\
&\quad= \mathbf{E}^{xpct}\left[\langle \beta(t_2)\ ,\ \beta(t_3)\rangle\ -\ \langle \beta(t_2)\ ,\ \beta(t_2)\rangle\ -\ \langle \beta(t_1)\ ,\ \beta(t_3)\rangle\ +\ \langle \beta(t_1)\ ,\ \beta(t_2)\rangle\right]\\
&\quad= \mathbf{E}^{xpct}\left[\langle \beta(t_2)\ ,\ \beta(t_3)\rangle\right]\ -\ \mathbf{E}^{xpct}\left[\langle \beta(t_2)\ ,\ \beta(t_2)\rangle\right]\\
&\qquad\qquad -\ \mathbf{E}^{xpct}\left[\langle \beta(t_1)\ ,\ \beta(t_3)\rangle\right]\ +\ \mathbf{E}^{xpct}\left[\langle \beta(t_1)\ ,\ \beta(t_2)\rangle\right]\\
&\quad= \sigma_\beta(t_2)^2\ -\ \sigma_\beta(t_2)^2\ -\ \sigma_\beta(t_1)^2\ +\ \sigma_\beta(t_1)^2\\
&\quad= 0
\end{aligned}
$$

Thus, since the random variables $\beta(t_1 \to t_2)$ and $\beta(t_1 \to t_2)$ are **balanced** and **orthogonal** they are independent.

3
This follows from part 1, and Theorem 4.1.7.
□

4.1.28

Example

In the case of **Brownian Motion** [4.3.1], we have

$$\forall s > t \geq 0, \quad \mathbf{E}^{xpct}\left[\langle \beta(s),\ \beta(t)\rangle\right] \ = \ t.$$

4.1.29

Corollary

Brownian motion is the unique stochastic process (modulo affine transformations) having **finite variance** *and possessing* **stationary, spacially homogeneous Gaussian increments**.

To be precise, suppose that β was some stochastic process so that:

1. For at least one $t_0 \in \mathbb{T}_{ime}$, **var** $[\beta(t_0)]$ was finite.
2. For all $t, T \in Time$, the random variables $\beta(t)$ and $\beta(t \to T)$ are *independent.*
3. For any t and T in $\mathbb{T}_{ime}$,

$$\mathbf{D}^{istr}\left[\beta(t \to T)\right] \ = \ \sqrt{0; R_{(t-T)}}$$

where $R_{(t-T)}$ is some function of the difference $(t - T)$.

Then there is some affine transformation $F : V \longrightarrow V$ so that $F(\beta)$ is Brownian motion.

Proof:

By construction, the increments of β are **spacially homogeneous** and **balanced**. Thus, by the previous theorem, they are **independent.** Also, by hypothesis, the increments are **stationary**.

Now, we will invoke Theorem 4.1.7. The number $R_{(t-s)}$ is just the **variance** of the increment $\beta(t \to s)$. By Theorem 4.1.7, this variance grows

linearly with time; that is, there is some constant $\lambda > 0$ (the **expansion coefficient**) so that.

$$R_{(t-s)} = \lambda \cdot (t - s)$$

Now we have, for any t and s in $\mathbb{T}_{time}$,

$$\mathbf{D}^{istr}[\beta(t \to s)] = \overbrace{0; \lambda \cdot (t - s)}$$

By hypothesis, β has finite variance at time t_0 –say $\mathbf{var}[\beta(t_0)] = C$. Then by part 2 of Theorem 4.1.7, β must **originate** at time $T_0 := t_0 - \frac{C}{\lambda}$.

Now, *where* does β originate? Let $\vec{x}_0 := \mathbf{E}^{xpct}[\beta(T)]$ for any fixed time T. Then, since the process has balanced, spacially homogeneous increments, $\vec{x}_0 := \mathbf{E}^{xpct}[\beta(t)]$ for *all* $t \in \mathbb{T}_{time}$.

Thus, consider the process

$$\hat{\beta} : (\mathbb{T}_{time} - T_0) \times \Omega \longrightarrow V$$

defined:

$$\hat{\beta}(t, \omega) := \frac{1}{\lambda} \cdot (\beta(t + T_0, \omega) - \vec{x}_0)$$

Then $\hat{\beta}$ is a process originating at zero with balanced, independent, stationary Gaussian increments whose variance grows linearly in time. In other words, $\hat{\beta}$ is Brownian motion.

□

4.2 Constructing Continuous Time Stochastic Processes

4.2.1 Kolmogorov's Theorem: Constructing a Stochastic Process with specified Marginals

Prerequisites: [4.1.1]

How might we go about actually *constructing* a stochastic process? The most obvious way would be to simply specify all the finite-dimensional *marginal* probability distributions of this process, throughout time.

So, to construct the process $\{\alpha_t; t \in \mathbb{T}_{time}\}$, first, for every time $t \in \mathbb{T}_{time}$, we would specify the probability distribution of the random variable α_t.

This would provide instantaneous "snapshots" of the statistics of α throughout history, but would provide us with no information about *causality*; that is, about how the states of α at different times are *correlated* together. Thus, we would have to specify correlation data; for each pair of times t and s, we would have to specify the *joint* probability distribution of the random variable (α_t, α_s).

More generally, for every finite sequence of times $t_1, t_2, \ldots, t_n \in \mathbb{T}_{ime}$, we would want to specify the $n-$fold joint probability distribution of the random variable $(\alpha_{t_1}, \alpha_{t_2}, \ldots, \alpha_{t_n})$.

Supposing we provided all this information in some sort of "internally consistent" way, we would hope that, as a consequence, we would ultimately "induce" a stochastic process. Kolmogorov's Theorem says that this is exactly what happens. Indeed, Kolmogorov's theorem says much more.

A stochastic process on a statespace X is really just a probability distribution on the product space $X^{\mathbb{T}_{ime}}$. Kolmogorov's theorem provides a general recipe for constructing probability measures on *infinite*-dimensional product spaces, via the specification of their *finite*-dimensional marginals.

First we review some technical terminology.

4.2.1

Definition *Projection Map; Marginal*

Let $X := \prod_{t \in T} X_t$ be a **Cartesian product** of sets $\{X_t|_{t \in T}\}$, where T is some indexing set.

For any $s \in T$, the **projection map** onto the $s-$th coordinate is the function:

$$\mathbf{pr}_s : X \ni \vec{x} \mapsto x_s \in X_s$$

where $\vec{x} := (x_t)_{t \in T}$ is an arbitrary element of X.

More generally, for any subset $S \subset \mathbb{T}_{ime}$, the **projection map** onto the coordinates of S is the function

$$\mathbf{pr}_S : X \ni \vec{x} \mapsto (x_s)_{s \in S} \in \prod_{s \in S} X_s$$

Now, suppose that, for all $t \in T$, the set X_t is endowed with a sigma-algebra $\mathcal{X}_t$, and let $\mathcal{X} := \bigotimes_{t \in T} \mathcal{X}_t$ be the **product** sigma-algebra on X. Suppose that μ is some measure on $\mathcal{X}$.

Then for any subset $S \subset T$, we define the **marginal** of μ on the coordinates of S to be the measure $\mathbf{pr}_S^*[\mu]$ on the space $\left(\prod_{s\in S} X_s, \bigotimes_{s\in S} \mathcal{X}_s\right)$ defined as follows: For any measurable subset $U \subset \prod_{s\in S} X_s$,

$$\mathbf{pr}_S^*(\mu)[U] \ := \ \mu\left[\mathbf{pr}_S^{-1}(U)\right] .$$

4.2.2

Theorem *Kolmogorov's Consistency Theorem*

Let X be a compact Hausdorff space, and let T be an arbitrary set. Suppose that, for every finite subset $A \subset T$, we specify a Radon measure μ_A on the product space X^A. Suppose further that this collection of measures satisfies the following *Kolmogorov consistency condition* :

If $A, B \subset T$ are two finite subsets, and $A \subset B$, then the marginal *of μ_B onto X^A is equal to μ_A:*

$$\mathbf{pr}^*(\mu_B) \ = \ \mu_A$$

Then there exists a unique Radon measure μ, defined on the space X^T, so that, for all finite subsets $A \subset T$,

$$\mathbf{pr}_A^*(\mu) := \mu_A$$

Proof:

A proof can be found in most standard probability textbooks. An approach appealing to analysts can be found in [6] , Section 9.4(p.301).

□

The "compactness" condition on the space X is a technicality which can easily be evaded.

4.2.3

Corollary *Kolmogorov Theorem for Noncompact Spaces*

Let X be a Hausdorff space, and let $\hat{X}$ be a **compactification** of X. Let T be an arbitrary set. Suppose that, for every finite subset $A \subset T$, we specify a Radon measure μ_A on the product space X^A, so that this collection of measures satisfies the *Kolmogorov consistency condition* .

Then there exists a unique Radon measure μ, defined on the space $\hat{X}^T$, so that, for all finite subsets $A \subset T$,

$$\mathbf{pr}_A^*(\mu) := \mu_A$$

Proof:

Simply treat the measures μ_A as being defined on $\hat{X}^A$ instead, and apply the previous theorem.

□

Theoretically, it is not necessary to define the marginals on *all* finite-dimensional subsets. Specification of marginals upon a suitably chosen subcollection of finite sets will suffice.

4.2.4

Definition *Filter*

Let T be a set. A **filter** on T is a collection $\mathcal{A}$ of subsets of T possessing the following properties:

1. For any subset $B \subset T$, there is some $A \in \mathcal{A}$ so that $B \subset A$.
2. If $A_1, A_2 \in \mathcal{A}$, then $A_1 \cap A_2 \in \mathcal{A}$.

4.2.5

Example *The Filter of Intervals*

Suppose $T = \mathbf{Z}$. Then a good choice of filter for T is the set of all finite **intervals** in $\mathbf{Z}$;

$$\mathcal{A} := \{[n..m] \ ; \ n, m \in \mathbf{Z}\}$$

where, for any $n < m$ in $\mathbf{Z}$,

$$[n..m] \ := \ \{n, \ n+1, \ n+2, \ldots, \ m-1, \ m\}.$$

4.2.6

Corollary *Filter Version of Kolmogorov's Theorem*

Let X be a Hausdorff space, and let $\hat{X}$ be a **compactification** of X. Let T be an arbitrary set, and let $\mathcal{A}$ be a **filter** of subsets of T.

Suppose that, for every $A \in \mathcal{A}$, we specify a Radon measure μ_A on the product space X^A, so that this collection of measures satisfies the *Kolmogorov consistency condition*[3].

Then there exists a unique Radon measure μ, defined on the space $\hat{X}^T$, so that, for all $A \in T$,

$$\mathbf{pr}_A^*(\mu) \ = \ \mu_A.$$

Proof:

We will first extend the collection of measures $\{\mu_A \ ; \ A \in \mathcal{A}\}$ to a collection of measures $\{\mu_S \ ; \ S \subset T, \ \ S \text{ finite. }\}$ which satisfy the Kolmogorov Consistency Condition. The way to do this is simple: for any subset $S \subset T$, find some $A \in \mathcal{A}$ so that $S \subset A$. Then define

$$\mu_S \ := \mathbf{pr}_S^* \mu_A$$

Claim 1: The measure μ_S is well-defined independent of the choice of superset A we use to define it.

Proof:

First find the *minimal* element $A \in \mathcal{A}$ such that $S \subset A$ (this can be done by using the fact that $\mathcal{A}$ is closed under intersections) .

Now, suppose that $B \in \mathcal{A}$ is such that $S \subset B$ as well.

Claim 1.1: $A \subset B$

[3]That is, whenever we have two sets elements $A, B \in \mathcal{A}$, with $A \subset B$, then $\mathbf{pr}_A^*(\mu_B) = \mu_A$.

Proof: $A \cap B \in \mathcal{A}$, and $S \subset A \cap B$; hence, since A is minimal, we conclude that $A \subset A \cap B$. But of course $A \cap B \subset A$; thus, $A \cap B = A$. Thus, $A \subset B$.

□1.1

Thus, by the Kolmogorov Condition, $\mathbf{pr}^*_A(\mu_B) = \mu_A$. Thus, $\mathbf{pr}^*_S(\mu_B) = \mathbf{pr}^*_S \circ \mathbf{pr}^*_A(\mu_B) = \mathbf{pr}^*_S(\mu_A)$.

□1

Claim 2: The collection of measures $\{\mu_S \; ; \; S \subset T, \;\; S \text{ finite. }\}$ satisfies the Kolmogorov Criterion . **Proof: Exercise.**

□2

Now apply the original Kolmogorov Theorem.

□

4.2.7

Example

Some examples of actual applications of Kolmogorov's Theorem to the construction of real stochastic processes are:

- **Bernoulli processes** [3.2].
- **Markov Processes** [3.3.5].
- **Brownian Motion** [4.3.2].

4.2.8

Remark *The Problem of Sample Path Properties*

The problem with the Kolmogorov approach to construction of stochastic processes is that, in practice, it is often very difficult to prove that the stochastic process in question has desired *sample path properties* . We might want to assert, for example, that almost every sample path of a certain stochastic process is **continuous** , or possesses a certain fractal dimension, etc. The Kolmogorov construction takes place in an extremely "large" space; the space $X^{\mathbb{T}_{ime}}$, of *all* possible functions from $\mathbb{T}_{ime}$ into X. Inside this space, the set of paths possessing a certain property (for example, continuity) may be extremely "small" and hard-to-find; indeed, it may not even be in the Borel sigma-algebra of $X^{\mathbb{T}_{ime}}$. Thus, the Kolmogorov method is not suitable our objective is to construct stochastic processes with such properties; in practice, other methods are more appropriate.

4.2.2 Constructing a Stochastic Process with Specified Increments

Prerequisites: [4.1.6]

Fashioning a Process from Increments

Sometimes, it is most convenient or natural to describe the trajectory of a particle by describing its *initial position*, and then specifying *velocity* at each instant. We can then reconstruct its trajectory by *integrating* the velocity function over time.

The corresponding construction of a stochastic process proceeds by specifying its *increments* in position over time as random variables, and describing the *initial position* of the process by some other random variable. We can use this data to build the process.

Suppose that α is a stochastic process on some vector space V, and let $\alpha(0)$, be a random variable describing the initial position of α. For each pair of times $s < t$ in $\mathbb{T}_{ime}$, suppose we are given another random variable $\alpha(s \to t)$, which purports to be the *increment* from time s to time t:

$$\alpha(s \to t) \ := \ \alpha(t) - \alpha(s)$$

As long as the increment data is sufficiently "consistent", we would expect that we could use it to *construct* the stochastic process α.

What does "sufficiently consistent" mean? Well, for one thing, note that, if $r < s < t$ are three different times, then

$$\alpha(r \to t) \ = \ \alpha(r \to s) + \alpha(s \to t)$$

so any collection of random variables purporting to be the increments of some process must satisfy this additivity property. Is this sufficient?

4.2.9

Lemma

Let V be a vector space, and $\mathbb{T}_{ime} \subset [0, \infty)$, and suppose that, for each pair of elements $s < t$ in $\mathbb{T}_{ime}$, we have a random variable $\alpha_{s,t}$ ranging over V, so that, for any $r < s < t$ in $\mathbb{T}_{ime}$, we have:

$$\alpha_{r,t} \ = \ \alpha_{r,s} + \alpha_{s,t} \qquad (1)$$

Let α_0 another random variable, ranging over V.

Then there exists a stochastic process α on V, with timeline $\mathbb{T}_{ime}$, so that

1. $\alpha(0) = \alpha_0$.
2. For all $s, t \in \mathbb{T}_{ime}, \alpha(s \to t) = \alpha_{s,t}$.

Proof:

For all t, define $\alpha(t) := \alpha_0 + \alpha_{0,t}$. The rest follows automatically.

□

Fashioning Increments from Distribution Data

That wasn't so hard; in some sense, all the work was done for us because of condition (1). But now we might ask a slightly subtler question. Suppose, instead of knowing that the random variables $\alpha_{r,t}$ and $\alpha_{r,s} + \alpha_{s,t}$ *are* the same, all we know is that $\alpha_{r,t}$ and $\alpha_{r,s} + \alpha_{s,t}$ *look* the same. In other words, instead of assuming

$$\alpha_{r,t} \;=\; \alpha_{r,s} + \alpha_{s,t}, \quad (1)$$

suppose instead we merely had:

$$\alpha_{r,t} \quad \underset{distr}{\cong} \quad \alpha_{r,s} + \alpha_{s,t}. \quad (2)$$

Would it still be possible to build a stochastic process possessing the desired increments?

The trick is to first build a **sample space** containing random variables resembling $a_{s,t}$, for all s, t. Only, in *this* space, rather than merely having the same distributions, as in (2), these random variables will actually satisfy the identity properties given by (1).

4.2.10

Lemma

Let V be a vector space, and $\mathbb{T}_{ime} \subset [0, \infty)$, with $0 \in \mathbb{T}_{ime}$. Suppose that, for each pair of elements $s, t \in \mathbb{T}_{ime}$, we have a probability measure $\mu_{s,t}$ on V, so that, the collection $\left\{ \mu_{s,t}|_{s,t \in \mathbb{T}_{ime}} \right\}$ satisfies the **cocycle**[4] **property:**

[4]This condition just says that these probability measures look like the distributions of random variables satisfying equation (1) above. For those who enjoy abstract algebra, it says that we have a **semigroup homomorphism** from the semigroup of subintervals of $[0, \infty)$ into the convolution monoid of measures on V.

For any $r < s < t$ in $\mathbb{T}_{ime}$,

$$\mu_{r,t} \;=\; \mu_{r,s} \,*\, \mu_{s,t}.$$

Then we can construct a sample space $(\Omega, \mathcal{W}, \mu)$, and random variables $\beta_{s,t} : \Omega \to V$, for all $s, t \in \mathbb{T}_{ime}$, so that

1. If $s_1 < t_1 < s_2 < t_2$, then the random variables β_{s_1,t_1} and β_{s_2,t_2} are **independent**.

2. For all s, t,

$$\mathbf{D}^{istr}\left[\beta_{s,t}\right] \;=\; \mu_{s,t}$$

3. For all $r < s < t$,

$$\beta_{r,t} \;=\; \beta_{r,s} + \beta_{s,t} \qquad (1)$$

Proof:

To construct this sample space, we will make use of **Kolmogorov's Theorem** [4.2.1] .

Let $\mathcal{V}$ be the sigma algebra on the vector space V, and define:

$$\Omega := V^{\mathbb{T}_{ime}}, \quad \textit{and} \quad \mathcal{W} := \bigotimes_{t \in \mathbb{T}_{ime}} \mathcal{V}.$$

For every finite subset $S \subset \mathbb{T}_{ime}$, we want to construct a measure μ_S on V^S, so that these measures satisfy the **Kolmogorov Consistency Condition.** First we'll do singleton sets and pairs to provide some intuitive motivation.

Singleton Sets

First suppose $S := \{s_0\}$, with $s_0 > 0$. and let $U \subset V$. The measure $\mu_{\{S\}}[U]$ should just represent the probability of that the increment from time 0 to time s_1 lies inside the set U. In other words, it makes sense to define:

$$\mu_{\{S\}}[U] \;:= \mu_{0,s_0}[U].$$

What if $s_0 = 0$, though? We want to say that the increment from *time zero to itself* is almost surely zero, let's just define an additional measure:

$$\mu_{0,0} \;:=\; \delta_0$$

(the point mass at 0) , and then define $\mu_{\{0\}} := \mu_{0,0}$.

Pairs

Next, suppose $S := \{s_0 \;<\; s_1\}$. It suffices[5] to define the measure μ_S on **measurable rectangles** of the form:

$$U \;:=\; U_0 \;\times\; U_1$$

where $U_0, U_1 \subset V$ are $\mathcal{V}-$measurable. Intuitively, the measure $\mu_S\,[U_0 \times U_1]$ should be the probability that the increment from *time* 0 to *time* s_0 is a point in U_0, and that the increment from *time* s_0 to *time* s_1 then carries us from *this* point into U_1. In other words:

$$\mu_S[U] \quad := \quad \int_{V^S} (\mathbb{1}_{U_0}\,[\vec{x}]) \;\cdot\; (\mathbb{1}_{U_1}\,[\vec{x}+\vec{v}_1]) \;\; d\mu_{0,s_0}[\vec{x}] \;\; d\mu_{s_0,s_1}[\vec{v}_1].$$

Arbitrary Sets

So suppose that $S := \{s_0 \;<\; s_1 \;<\; s_3 \;<\; \ldots \;<\; s_N\}$. Again it suffices define the measure μ_S first on **measurable rectangles** of the form:

$$U \;:=\; U_0 \;\times\; U_1 \;\times \ldots \times\; U_N$$

where $U_0,\; \ldots,\; U_N \subset V$ are $\mathcal{V}-$measurable.

Again, the measure $\mu_S\,[U]$ should be the probability that the increment from *time* 0 to *time* s_0 is a point in U_0, that the increment from *time* s_0 to *time* s_1 then carries us from *this* point into U_1, that the increment from *time* s_1 to *time* s_2 then carries us from here into U_2, and so on, until finally, the sum of all N increments puts us somewhere in U_N. Thus, define

$$\mu_S[U] \quad := \quad \int_{V^S} F\,(\vec{x},\, \vec{v}_1,\, \ldots,\, \vec{v}_N) \;\; d\mu_{0,s_0}[\vec{x}]\, d\mu_{s_0,s_1}[\vec{v}_1] \;\ldots\; d\mu_{s_{N-1},s_N}[\vec{x}_N],$$

where

$$F\,(\vec{x},\, \vec{v}_1,\, \ldots,\, \vec{v}_N) \;:=\; (\mathbb{1}_{U_0}\,[\vec{x}]) \;\cdot\; (\mathbb{1}_{U_1}\,[\vec{x}+\vec{v}_1]) \;\cdot\; (\mathbb{1}_{U_2}\,[\vec{x}+\vec{v}_1+\vec{v}_2]) \;\ldots$$
$$\ldots (\mathbb{1}_{U_N}\,[\vec{x}+\vec{v}_1+\vec{v}_2+\ldots+\vec{v}_N])\,.$$

[5] See [6] for details.

Claim 1: This collection of measures satisfies the **Kolmogorov Consistency Criterion** .

Proof: Exercise.
□1

Now, let μ be the measure on $\Omega := V^{\mathbb{T}ime}$ induced by this collection of finite marginals. We now have the desired sample space. Define the random variables $\beta_{s,t}$, for all $s, t \in$ Time, as follows:

$$\beta_{s,t} := \mathbf{pr}_t - \mathbf{pr}_s$$

where $\mathbf{pr}_s : V^{\mathbb{T}ime} \longrightarrow V$ is just **projection** onto the $s-$th coordinate.

Claim 2: These random variables have the desired properties.

Proof: Exercise.
□2
□

4.2.11

Remark

A much more concrete version of the construction employed in this proof appears in the **Kolmogorov** construction of **Brownian Motion** [4.3.2].

Combining these two lemmas together, we have:

4.2.12

Theorem

Let V be a vector space, and $\mathbb{T}ime \subset [0, \infty)$ Suppose that, for each pair of elements $s, t \in \mathbb{T}ime$, we have a probability measure $\mu_{s,t}$ on V, so that the collection $\left\{ \mu_{s,t}|_{s,t \in \mathbb{T}ime} \right\}$ satisfies the **cocycle property**. Let α_0 be some other random variable ranging over V.

Then there exists a stochastic process α on V, indexed by $\mathbb{T}ime$, so that

1. $\alpha(0) = \alpha_0$.

2. For all $s, t \in \mathbb{T}_{ime}, \mathbf{D}^{istr}[\alpha(s \to t)] = \mu_{s,t}$.

3. α is a process with **independent increments**.

□

4.2.13

Example

Some examples of actual applications of the Method of Increments to the construction of real stochastic processes are:

- **Brownian Motion**: See [4.3.2].
- **Ornstein-Uhlenbeck Process**: See [**??**].

Interpolating from Partial Increment Data

We can now make the question harder again. Suppose that, instead of being provided with probability distributions for *all* time increments $s \to t$, we are only given *some* of them. In particular, suppose we are given distributions for time increments lying on some refining sequence of discrete subsets of $\mathbb{T}_{ime}$.

Intuitively, this should be enough information to determine the stochastic process, as long as we are willing to assume that that the process behaves *continuously*, so that we can *interpolate* the behaviour of paths in the spaces between the elements of our partitions.

For this sort of "interpolative" construction succeed, we need to have a topology on our vector space V so we can define convergence. Also, in addition to stipulating an *algebraic* consistency condition for the increments, as we did in [4.2.2] , we must now also stipulate a *topological* consistency condition, to guarantee that our estimates converge.

First some technical remarks.

4.2.14

Remark *The Banach Algebra of Measures under Convolution*

Let $\mathcal{M}$ be the space of measures of **bounded variation** over V. The **variation norm** on $\mathcal{M}$ provides it a **Banach space** structure. The operation of **convolution** is continuous with respect to this norm, and satisfies the appropriate distribution properties with respect to addition. Hence,

the Banach space $\mathcal{M}$, together with the "multiplication" operator $*$, is a **Banach Algebra**. (For more details on this, see [4], Chapter VII, section 9, or see [7], Chapter VIII.)

For our purposes, the important fact is that $\mathcal{M}$, together with the Banach topology and the convolution operator, is a **topological monoid**. In other words, convolution is continuous with respect to this topology.

4.2.15

Remark *Cocycles, Stationarity, and Monoid Homomorphisms*

Assume $\mathbb{T}_{ime} \subset [0, \infty)$ be closed under addition –in other words, assume $(\mathbb{T}_{ime}, +)$ is a **monoid**.

Suppose $\left\{ \mu_{s,t}|_{s,t \in \mathbb{T}_{ime}} \right\}$ is a collection of measures satisfying the previously described **cocycle property**:

$$\forall r, s, t \in \mathbb{T}_{ime}, \text{ with } r < s < t, \quad \mu_{r,t} \ = \ \mu_{r,s} * \mu_{s,t}. \quad (1)$$

Suppose we wanted to use this collection of measures to define a process with **stationary increments** [4.1.6]. Then we want the measures themselves to be "stationary" in the sense that:

$$\forall s, t \in \mathbb{T}_{ime}, \text{ with } s < t, \quad \mu_{s,t} \ = \ \mu_{0,t-s}. \quad (2)$$

Hence, we really only need keep track of the measures

$$\mu_t \ := \ \mu_{0,t}$$

for all $t \in \mathbb{T}_{ime}$, since all the rest can be generated using (2). But now, if we apply property (1), we have:

$$\forall s, t \in \mathbb{T}_{ime} \mu_{s+t} \ = \ \mu_s * \mu_t.$$

in other words, we have a **monoid homomorphism** from $(\mathbb{T}_{ime}, +)$ into $(\mathcal{M}, *)$.

4.2.16

Remark *Extending the Homomorphism*

Now, suppose that $\mathbb{T}_{ime} \subset [0, \infty)$ is a monoid under addition, and is **dense** as a subset of $[0, \infty)$. Suppose that

$$\mathbb{T}_{ime} \ni t \mapsto \mu_t \in \mathcal{M}$$

is some monoid homomorphism, as described in the previous remark, and assume further that the map $(t \mapsto \mu_t)$ is **continuous** with respect to the topologies on $\mathbb{T}_{ime}$ and $\mathcal{M}$. We want to extend $(t \mapsto \mu_t)$ to a homomorphism from $[0, \infty)$ into $\mathcal{M}$.

But $\mathcal{M}$ is a **complete metric space**. Thus, any continuous function $F : S \longrightarrow \mathcal{M}$ from a dense subset $S \subset [0, \infty)$ can be extended in a unique fashion to a map $\hat{F} : [0, \infty) \longrightarrow \mathcal{M}$. (**Exercise**[6]) Furthermore, if F was a monoid homomorphism, then $\hat{F}$ will be one as well (**Exercise**[7]).

Combining all these considerations, we get:

4.2.17

Theorem

Let V be a vector space, and $\mathcal{M}$ the Banach algebra of measures on V having bounded variation.

Let $\mathbb{T}_{ime} \subset [0, \infty)$ be a monoid under addition, and dense as a subset of $[0, \infty)$. Suppose that

$$\mathbb{T}_{ime} \ni t \mapsto \mu_t \in \mathcal{M}$$

is a **continuous monoid homomorphism**, where, for each t, the measure μ_t is a probability measure. Finally, let μ_0 be some other probability measure on V.

Then there exists a stochastic process α, with statespace V and timeline $[0, \infty)$, so that:

1. $\mathbf{D}^{istr}[\alpha_0] = \mu_0$.
2. For any $t < T$ in $\mathbb{T}_{ime}$,

$$\mathbf{D}^{istr}[\alpha(t \longrightarrow T)] = \mu_{T-t}$$

3. Hence, α has **stationary increments**.

[6] Hint: it is important that $[0, \infty)$ is a metric space, and that $\mathcal{M}$ is complete.
[7] Hint: the topological monoid property is key.

Proof: By the previous remarks, extend the map $(t \mapsto \mu_t)$ to a continuous monoid homomorphism on all of $[0, \infty)$.

Next, for any s, t in $\mathbb{T}_{ime}$ with $s < t$, define

$$\mu_{s,t} := \mu_{t-s}$$

Check that this collection satisfies the **cocycle property**. Then apply the previous theorem.

□

4.2.18

Remark

Is there any relationship between the continuity of the monoid homomorphism $(t \longrightarrow \mu_t)$ and the continuity of the sample paths of the process α?

4.2.3 Constructing processes via Polygonal Approximation

Prerequisites: [4.1.5]

The Problem of Continuous Sample Paths

As we have seen, it is easy to construct a continuous-time stochastic process, by means, for example, of the Kolmogorov Consistency theorem. However, it is often difficult to prove that such a process actually has a version with *continuous sample paths*. Intuitively, this is because the function space $X^{\mathbb{T}_{ime}}$ is too "big", and the space of continuous functions $\mathbf{C}(\mathbb{T}_{ime}; X)$ is too "small" a subset of it; confining the support of the probability measure to $\mathbf{C}(\mathbb{T}_{ime}; X)$ is too difficult a task.

A better approach to constructing *continuous* stochastic processes on topological spaces is to work *from the inside*, constructing the process as a probability measure *within* $\mathbf{C}(\mathbb{T}_{ime}; X)$. This way, we never have to worry about proving that the sample paths are almost-surely continuous; they are continuous by construction.

So, to build a *continuous* version of some stochastic process β, we will build a version which lies entirely *inside* $\mathbf{C}(\mathbb{T}_{ime}, X)$. Think of this stochas-

tic process process as a *random variable*, taking its values in the spaceof continuous functions. In other words, it is a measurable function

$$\beta : \Omega \longrightarrow \mathbf{C}(\mathbb{T}_{ime}; X)$$

where $(\Omega, \mathcal{W}, P)$ is some universal sample space.

The actual process β we wish to construct may be quite complex; we may want to build it as a limit of a sequence of *approximations*. That is, we want to construct a sequence of stochastic processes

$$\beta_n : \Omega \longrightarrow \mathbf{C}(\mathbb{T}_{ime}; X)$$

for all $n \in \mathbb{N}$, and then define β to be the **pointwise limit** of $\left\{\beta_n|_{n\in\mathbb{N}}\right\}$. In other words, for all $\omega \in \Omega$, we will define:

$$\beta(\omega) \ := \lim_{n\to\infty} \beta_n(\omega)$$

where we take this limit in the **uniform** topology[8] on the space $\mathbf{C}(\mathbb{T}_{ime}; X)$. The difficulty, then, is to prove that the sequence of functions $\left\{\beta_n(\omega)|_{n\in\mathbb{N}}\right\}$ *does*, in fact, converge uniformly, at least for almost all $\omega \in \Omega$.

4.2.19

Lemma

Let X be a **complete metric** space, and $(\Omega, \mathcal{W}, P)$ a sample space, and, for all $n \in \mathbf{N}$, let

$$\beta_n : \Omega \longrightarrow \mathbf{C}(\mathbb{T}_{ime}; X)$$

be a stochastic process.

Suppose we could find a **summable** sequence $\left\{b_n|_{n\in\mathbb{N}}\right\}$ of positive numbers so that, if, for each $n \in \mathbb{N}$, we define

$$a_n \ := \ \mathbf{P}_{rob}\left[\mathbf{d}_\infty\left(\beta_n,\ \beta_{n+1}\right) > b_n\right]$$

then the sequence $\left\{a_n|_{n\in\mathbb{N}}\right\}$ is also summable.

Then for almost all $\omega \in \Omega$, the sequence of functions $\left\{\beta_n(\omega)|_{n\in\mathbb{N}}\right\}$ converges uniformly to some element of $C(\mathbb{T}_{ime}; X)$.

[8] Be careful here; we are talking about a **pointwise** limit over Ω, but a **uniform** limit over $\mathbb{T}_{ime}$. Also note that it is important to take these limits in the right order; we are looking at a pointwise limit of uniform limits, *not* a uniform limit of pointwise limits.

Proof:

For all $N \in \mathbf{N}$, define $B_N := \sum_{n=N}^{\infty} b_n$. Thus, since $\left\{ b_n|_{n \in \mathbf{N}} \right\}$ is summable, we know that $\lim_{N \to \infty} B_N = 0$.

For fixed N, consider the event

$$E_N \ := \ \left\{ \sum_{n=N}^{\infty} \mathbf{d}_\infty(\beta_n, \ \beta_{n+1}) \ < B_N \right\}$$

Claim 1: For almost all $\omega \in \Omega$, there is some N so that $\omega \in E_N$.

Proof:

For every M, let C_N be the complement of E_N. Then we are essentially trying to show that almost every ω manages to *avoid* being in at least one C_M.

In other words, we are trying to show:

$$\mathbf{P}_{rob} \left[\bigcap_{M=1}^{\infty} C_M \right] \ = \ 0.$$

However,

$$\begin{aligned}
\mathbf{P}_{rob}\left[C_N\right] &= \mathbf{P}_{rob}\left[\sum_{n=N}^{\infty} \mathbf{d}_\infty(\beta_n, \ \beta_{n+1}) \ > B_N \right] \\
&\leq \mathbf{P}_{rob}\left[\textit{There exists some } n > N \textit{ so that } \mathbf{d}_\infty(\beta_n, \ \beta_{n+1}) \ > b_N \right] \\
&= \mathbf{P}_{rob}\left[\bigcup_{n=N}^{\infty} \left\{ \mathbf{d}_\infty(\beta_n, \ \beta_{n+1}) \ > b_N \right\} \right] \\
&\leq \sum_{n=N}^{\infty} \mathbf{P}_{rob}\left[\mathbf{d}_\infty(\beta_n, \ \beta_{n+1}) \ > b_N \right] \\
&= \sum_{n=N}^{\infty} a_n,
\end{aligned}$$

by definition of a_n. Thus,

$$\begin{aligned} \mathbf{P}_{rob}\left[\bigcap_{N=1}^{\infty} C_N\right] &\leq \inf_{N\in\mathbb{N}} \mathbf{P}_{rob}\left[C_N\right] \\ &\leq \inf_{N\in\mathbb{N}} \left(\sum_{n=N}^{\infty} a_n\right) \\ &= 0 \end{aligned}$$

because the series $\sum_{n=1}^{\infty} a_n$ converges, by hypothesis.
□[1]

Claim 2:
Suppose that $\{x_n|_{n\in\mathbf{N}}\}$ is a sequence on some metric space, and suppose that, for some $N \in \mathbb{N}$,

$$\sum_{n=N}^{\infty} d(x_n, x_{n+1})$$

is a finite sum. Then the sequence $\{x_n|_{n\in\mathbf{N}}\}$ is **Cauchy**.

Proof:
Exercise.
□[2]
Thus, combining Claims 1 and 2, we conclude:

For almost all $\omega \in \Omega$, *the sequence* $\{\beta_n(\omega)|_{n\in\mathbf{N}}\}$ *is Cauchy in* $\mathbf{C}(\mathbb{T}_{ime}; X)$.

Since X is a complete metric space, the space $\mathbf{C}(\mathbb{T}_{ime}; X)$ is also complete with respect to the uniform metric. Thus, we have:

For almost all $\omega \in \Omega$, *the sequence* $\{\beta_n(\omega)|_{n\in\mathbf{N}}\}$ *converges uniformly in* $\mathbf{C}(\mathbb{T}_{ime}; X)$.

□

Polygonal Approximations

4.2.20

Definition

Let V be a vector space, and

$$\beta : \mathbb{T}_{ime} \times \Omega \longrightarrow V$$

some stochastic process. A **polygonal approximation sequence** for β is given by the following data:

1. A an ascending sequence

$$\mathbf{T}_1 \subset \mathbf{T}_2 \subset \ldots \ \mathbb{T}_{ime}$$

of **discrete subsets** of $\mathbb{T}_{ime}$, so that $\mathbf{T}_\infty := \bigcup_{n=1}^{\infty} \mathbf{T}_n$ is **dense** in $\mathbb{T}_{ime}$. This sequence is called the **skeleton** of the approximation.

2. For each $n \in \mathbb{N}$ a stochastic process

$$\beta_n : \mathbb{T}_{ime} \times \Omega \longrightarrow V$$

so that

(a) For almost all $\omega \in \Omega$, the function $\beta_{n;\omega} : \mathbb{T}_{ime} \longrightarrow V$ is **peicewise affine**[9] with respect to $\mathbf{T}_n$

(b) For almost all $\omega \in \Omega$, the function $\beta_{n;\omega}$ agrees in value with $\beta_{(n-1);\ \omega}$ on all points in $\mathbf{T}_{n-1}$

Thus, intuitively, we achieve each process β_n by taking its predecessor β_{n-1}, (which is peicewise affine with respect to the set $\mathbf{T}_{n-1}$,) and introducing additonal affine "perturbations" to the sample paths along all the points in $\mathbf{T}_n \setminus \mathbf{T}_{n-1}$, while keeping the sample paths anchored at their former values along all points in $\mathbf{T}_{n-1}$. (see picture).

[9]That is, the function $\beta_{n;\omega} : \mathbb{T}_{ime} \longrightarrow V$ is affine on the interval between any two elements of $\mathbf{T}_n$, although breaks in its affine behaviour can possibly occur at any point in $\mathbf{T}_n$.

It will be convenient to assume that the skeleton $\mathbf{T}_1 \subset \mathbf{T}_2 \subset \ldots$ is **dyadically refining**, in the sense that each set contains twice as many elements as its predecessor. Formally, if $\mathbb{T}_{ime} = [t_0, t_1]$, then we want:

$$\begin{aligned}
\mathbf{T}_1 &= \left\{t_{[0]} < t_{[1]}\right\} \\
\mathbf{T}_2 &= \left\{t_{[0]} < t_{[\frac{1}{2}]} < t_{[1]}\right\} \\
\mathbf{T}_3 &= \left\{t_{[0]} < t_{[\frac{1}{4}]} < t_{[\frac{1}{2}]} < t_{[\frac{3}{4}]} < t_{[1]}\right\} \\
\mathbf{T}_3 &= \left\{t_{[0]} < t_{[\frac{1}{8}]} < t_{[\frac{1}{4}]} < t_{[\frac{3}{8}]} < t_{[\frac{1}{2}]} < < t_{[\frac{5}{8}]} < t_{[\frac{3}{4}]} < t_{[\frac{7}{8}]} < t_{[1]}\right\} \\
&\vdots
\end{aligned}$$

etc.

4.2.21

Remark

We can always assume without loss of generality that the skeleton is dyadically refining; given a polygonal approximation with an arbitrary skeleton, we can use it to construct a polygonal approximation with a dyadically refining skeleton (**Exercise**: Check this).

So, we can define our polygonal approximation by means of a dyadically indexed collection of **perturbations**. Let $\mathbb{Q}_2$ be the **dyadic rational numbers** in the interval $(0, 1]$. In other words, $\mathbb{Q}_2 := \left\{\frac{k}{2^n} \; ; \; n \in \mathbb{N} \; and \; k \in [1..2^n]\right\}$. For each $q \in \mathbb{Q}_2$, let

$$\phi_{[q]} : \Omega \longrightarrow V$$

be a random variable. We will use these random variables as *perturbations* to create the sequence of dyadic polygonal approximations.

In addition, let Ψ be some *other* random variable. Ψ will determine the "origin point" of the stochastic process. In many cases, Ψ may just be a constant value (for example, if the process **originates at zero**, then we would want $\Psi = 0$ almost surely.)

4.2.22

Notation

Let $\mathcal{L}_{ine}\left[\begin{array}{c|c} t_1 & t_2 \\ \hline \vec{v}_1 & \vec{v}_2 \end{array}\right]$ denote the unique *peicewise affine function* from $\mathbb{T}_{ime}$ into V, which passes through the point $\vec{v}_1$ at time t_1 and the point $\vec{v}_2$ at time t_2, and which takes the constant value zero outside the interval $[t_1, t_2]$. In other words, for all $t \in \mathbb{T}_{ime}$,

$$\mathcal{L}_{ine}\left[\begin{array}{c|c} t_1 & t_2 \\ \hline \vec{v}_1 & \vec{v}_2 \end{array}\right](t) = \begin{cases} \vec{v}_1 + \frac{(t-t_1)}{t_2-t_1} \cdot (\vec{v}_2 - \vec{v}_1) & if \quad t \in [t_1, t_2] \\ 0 & if \quad t \notin [t_1, t_2] \end{cases}$$

(see fig)

4.2.23

Example

If $t_0 < t_1 < t_2$, and $\vec{v} \in V$, then

$$\mathcal{L}_{ine}\left[\begin{array}{c|c} t_0 & 0 \\ \hline t_1 & \vec{v} \end{array}\right] + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_1 & \vec{v} \\ \hline t_2 & 0 \end{array}\right]$$

is a function which is *zero* before time t_0, moves in a *straight line* from zero to $\vec{v}$ between times t_0 and t_1, *returns* to zero from time t_1 to t_2, and is *zero* after t_2.

Now we can define a polygonal approximation sequence as follows.

$$\begin{aligned}
\beta_{-1}(\omega) &= \Psi \\
\beta_0(\omega) &= \beta_{-1}(\omega) + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[0]} & t_{[1]} \\ \hline 0 & \phi_{[1]}(\omega) \end{array}\right] \\
\beta_1(\omega) &= \beta_0(\omega) + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[0]} & t_{[\frac{1}{2}]} \\ \hline 0 & \phi_{[\frac{1}{2}]}(\omega) \end{array}\right] + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[\frac{1}{2}]} & t_{[1]} \\ \hline \phi_{[\frac{1}{2}]}(\omega) & 0 \end{array}\right] \\
\beta_2(\omega) &= \beta_1(\omega) + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[0]} & t_{[\frac{1}{4}]} \\ \hline 0 & \phi_{[\frac{1}{4}]}(\omega) \end{array}\right] + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[\frac{1}{4}]} & t_{[\frac{1}{2}]} \\ \hline \phi_{[\frac{1}{4}]}(\omega) & 0 \end{array}\right] \\
&\quad + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[\frac{1}{2}]} & t_{[\frac{1}{4}]} \\ \hline 0 & \phi_{[\frac{3}{4}]}(\omega) \end{array}\right] + \mathcal{L}_{ine}\left[\begin{array}{c|c} t_{[\frac{3}{4}]} & t_{[1]} \\ \hline \phi_{[\frac{3}{4}]}(\omega) & 0 \end{array}\right]
\end{aligned}$$

and, inductively, for any $N \in \mathbb{N}$

$$\beta_N \;=\; \beta_{N-1}(\omega) + \sum_{n=0}^{2^{N-1}} \left(\mathcal{L}_{ine}\left[\frac{t_{[\frac{2n}{2^N}]}}{0} \middle| \frac{t_{[\frac{2n+1}{2^N}]}}{\phi_{[\frac{2n+1}{2^N}]}(\omega)} \right] + \mathcal{L}_{ine}\left[\frac{t_{[\frac{2n+1}{2^N}]}}{\phi_{[\frac{2n+1}{2^N}]}(\omega)} \middle| \frac{t_{[\frac{2n+2}{2^N}]}}{0} \right] \right)$$

(See fig)

Under what circumstances will this polygonal approximation series converge?

4.2.24

Lemma

Let V be a **Banach space**. Suppose we could find a **summable** sequence $\left\{ b_N|_{N\in\mathbb{N}} \right\}$ of positive numbers so that, if, for each $\frac{n}{2^N} \in \mathbb{Q}_2$, we define

$$a_{[\frac{n}{2^N}]} \;:=\; \mathbf{P}_{rob}\left[\left\| \phi_{[\frac{n}{2^N}]} \right\| > b_n \right]$$

then the sequence $\left\{ a_{[q]}|_{q\in\mathbb{Q}_2} \right\}$ is also summable[10].

Then for almost all $\omega \in \Omega$, the sequence of functions $\left\{ \beta_n(\omega)|_{n\in\mathbb{N}} \right\}$ **converges uniformly** to some element of $C(\mathbb{T}_{ime}; V)$.

Proof:

This is will be an application of the previous lemma [4.2.3]. For all $N \in \mathbb{N}$, define

$$A_N \;:=\; \sum_{n=0}^{2^{n-1}} a_{\left[\frac{2n+1}{2^N}\right]}$$

then clearly, $\left\{ a(q)|_{q\in\mathbb{Q}_2} \right\}$ is summable if and only if $\left\{ A_N|_{N\in\mathbb{N}} \right\}$ is summable. (These are all positive numbers, so all sums here converge absolutely). Now note that, for any $N \in \mathbb{N}$,

$$\mathbf{P}_{rob}\left[\|\beta_n \;-\; \beta_{n+1}\|_\infty > b_n \right] \;=\; \mathbf{P}_{rob}\left[\textit{For some } n \in [0..2^{N-1}],\;\; \phi_{[\frac{2n+1}{2^N}]} > b_n \right]$$

[10] This is a **countable** collection of **nonegative** numbers, so the property of summability is well-defined independent of the specific ordering of the indexing set.

$$\begin{aligned} &\leq \sum_{n\in[0..2^{N-1}]} \mathbf{P}_{rob}\left[\phi_{[\frac{2n+1}{2^N}]} > b_n\right] \\ &= \sum_{n=0}^{2^{n-1}} a_{[\frac{2n+1}{2^N}]} \\ &= A_N. \end{aligned}$$

at this point, we apply Lemma [4.2.3] to conclude that the processes β_n almost surely converge uniformly.

□

Properties of the Limit Process

Now we will investigate the properties of the limit process we have constructed. First, let's look at its *incremements*. Since we have built our process out of a sequence of dyadic approximations, it makes sense to try to get control of arbitrary increments by means of dyadic increments.

We will continue to use all the notation of the previous section.

4.2.25

Lemma

Let β be the limit of a sequence of dyadic polygonal approximations, as before, and suppose that β has **almost-surely continuous sample paths.**

Let $\underline{t}$ and $\bar{t} \in \mathbb{T}_{ime}$, and suppose that $\left\{t_{[\underline{q}_i]}|_{i\in\mathbb{N}}\right\}$ and $\left\{t_{[\bar{q}_i]}|_{i\in\mathbb{N}}\right\}$ are sequences converging to $\underline{t}$ and $\bar{t}$ respectively, where, for all i, $\underline{q}_i$ and $\bar{q}_i$ are **dyadic rational numbers**, given by:

$$\underline{q}_i = \frac{\underline{n}_i}{2^{\underline{N}_i}} \quad and \quad \bar{q}_i = \frac{\bar{n}_i}{2^{\bar{N}_i}}$$

in lowest terms, and let $N_i = \max\left\{\underline{N}_i, \bar{N}_i\right\}$.

Then, as a function from Ω into V, $\beta(\underline{t} \to \bar{t})$ is almost everywhere the **pointwise limit** of the sequence $\left\{\beta_{N_i}\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right)|_{i\in\mathbb{N}}\right\}$

Proof:

Claim 1: $\beta(\underline{t} \to \bar{t})$ is almost everywhere the pointwise limit of the sequence $\left\{\beta\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right) |_{i \in \mathbb{N}}\right\}$

Proof:

This follows from the fact that the sample paths of β are continuous, and because $\underline{t} = \lim_{i \to \infty} t_{[\underline{q}_i]}$ and $\bar{t} = \lim_{i \to \infty} t_{[\bar{q}_i]}$. The details are left as an **Exercise**.

□[1]

Claim 2:

For fixed $i \in \mathbb{N}$,

$$\beta_{N_i}\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right) = \beta_{N_{i+1}}\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right) = \beta_{N_{i+2}}\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right) =$$

$$\ldots = \beta\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right)$$

Proof:

After the Nth stage, the value of β at times $t_{[\frac{n}{2^N}]}$ (for every n) are fixed; further perturbations act only to modify the sample paths "between" these points.

□[2]

Hence, combining these two claims,

$$\begin{aligned}\beta(\underline{t} \to \bar{t}) &= \lim_{i \to \infty} \beta\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right) \\ &= \lim_{i \to \infty} \beta_{N_i}\left(t_{[\underline{q}_i]} \to t_{[\bar{q}_i]}\right).\end{aligned}$$

□

So, what do the dyadic increments look like, anyways? Suppose that $\underline{q}, \bar{q} \in \mathbb{Q}_2$, with $\underline{q} < \bar{q}$. Then we can always express $\underline{q}$, and $\bar{q}$ as sums:

$$\underline{q} = \sum_{n=0}^{N} \underline{d}_n \frac{1}{2^n}$$

and

$$\bar{q} \;=\; \sum_{n=0}^{\bar{N}} \bar{d}_n \frac{1}{2^n}$$

where, for all n, the numbers $\underline{d}_n$ and $\bar{d}_n$ are either 0 or 1. [11]
Next, throw out all the zero terms in these summands, and write

$$\underline{q} \;=\; \sum_{j=1}^{\underline{J}} \frac{1}{2^{\underline{n}_j}} \qquad \bar{q} \;=\; \sum_{j=1}^{\bar{J}} \frac{1}{2^{\bar{n}_j}},$$

for some sequences $\underline{n}_1 < \underline{n}_2 < \ldots < \underline{n}_{\underline{J}}$ and $\bar{n}_1 < \bar{n}_2 < \ldots < \bar{n}_{\bar{J}}$

4.2.26

Lemma

Suppose that

$$\underline{q} \;=\; \sum_{j=1}^{\underline{J}} \frac{1}{2^{\underline{n}_j}} \;\; \textit{and} \;\; \bar{q} \;=\; \sum_{j=1}^{\bar{J}} \frac{1}{2^{\bar{n}_j}},$$

and, for all $j_0 \in [1..\underline{J}]$, define

$$\underline{q}_{j_0} \;=\; \sum_{j=1}^{j_0} \frac{1}{2^{\underline{n}_j}}$$

and similarly, for all $j_0 \in [1..\bar{J}]$, define

$$\bar{q}_{j_0} \;=\; \sum_{j=1}^{j_0} \frac{1}{2^{\bar{n}_j}}$$

Then, as random variables,

$$\beta\left(t_{[\underline{q}]} \to t_{[\bar{q}]}\right) \;=\; \sum_{j=1}^{\bar{J}} \phi_{[\bar{q}_j]} \;-\; \sum_{j=1}^{\underline{J}} \phi_{[\underline{q}_j]}$$

[11] In other words, if written in **binary** form, the numbers $\underline{q}$ and $\bar{q}$ have expansions $0.\underline{d}_1\underline{d}_2\ldots\underline{d}_{\underline{N}}$ and $0.\bar{d}_1\bar{d}_2\ldots\bar{d}_{\bar{N}}$ respectively.

Proof:
Exercise.
□

4.2.27

Lemma

Let V be a vector space with an **inner product**, and suppose that:

1. For all $q \in \mathbb{Q}_2$, $\mathbf{E}^{xpct}\left[\phi_{[q]}\right] = 0$.
2. Each successive "wave" of perturbations are about half the size of the preceeding wave. Formally, for every dyadic number $\frac{2n+1}{2^N} \in \mathbb{Q}_2$, we have

$$\mathbf{var}\left[\phi_{[\frac{2n+1}{2^N}]}\right] = \frac{1}{4}\mathbf{var}\left[\phi_{[\frac{n+1}{2^{N-1}}]}\right]$$

Then the limit process β has **balanced, independent increments.**

Proof:
By Lemma [4.2.3], it suffices to check increments over dyadic times intervals.

Balanced Increments

Let $\underline{q}$ and $\bar{q}$ be dyadic numbers. Then, by Lemma [4.2.3], we can express the increment between them:

$$\beta\left(t_{[\underline{q}]} \to t_{[\bar{q}]}\right) = \sum_{j=1}^{\bar{J}} \phi_{[\bar{q}_j]} - \sum_{j=1}^{\underline{J}} \phi_{[\underline{q}_j]}$$

(refer to that lemma for explanation of notation).
Thus,

$$\begin{aligned}
\mathbf{E}^{xpct}\left[\beta\left(t_{[\underline{q}]} \to t_{[\bar{q}]}\right)\right] &= \mathbf{E}^{xpct}\left[\sum_{j=1}^{\bar{J}} \phi_{[\bar{q}_j]} - \sum_{j=1}^{\underline{J}} \phi_{[\underline{q}_j]}\right] \\
&= \sum_{j=1}^{\bar{J}} \mathbf{E}^{xpct}\left[\phi_{[\bar{q}_j]}\right] - \sum_{j=1}^{\underline{J}} \mathbf{E}^{xpct}\left[\phi_{[\underline{q}_j]}\right] \\
&= \sum_{j=1}^{\bar{J}} 0 - \sum_{j=1}^{\underline{J}} 0 \\
&= 0.
\end{aligned}$$

Independence

We want to show that, for any times $t_0 < T_0 < t_1 < T_1$, the increments $\beta(t_0 \to T_0)$ and $\beta(t_1 \to T_1)$ are independent random variables. Since the increments are all symmetric, this is equivilant to showing that $\beta(t_0 \to T_0)$ and $\beta(t_1 \to T_1)$ are **orthogonal**. We will illustrate the general argument with a special case.

Claim 1:
$\beta(t_{[0]} \to t_{[\frac{1}{2}]})$ and $\beta(t_{[\frac{1}{2}]} \to t_{[1]})$ are orthogonal.

Proof:

$$\begin{aligned}
\mathbf{E}^{xpct}\left[\left\langle \beta(t_{[0]} \to t_{[\frac{1}{2}]}),\ \beta(t_{[\frac{1}{2}]} \to t_{[1]})\right\rangle\right] &= \mathbf{E}^{xpct}\left[\left\langle \frac{1}{2}\phi_{[1]} + \phi_{[\frac{1}{2}]},\ \frac{1}{2}\phi_{[1]} - \phi_{[\frac{1}{2}]}\right\rangle\right] \quad \text{(By [4.2.3])} \\
&= \mathbf{E}^{xpct}\left[\frac{1}{4}\left\|\phi_{[1]}\right\|_2 - \left\|\phi_{[\frac{1}{2}]}\right\|_2\right] \\
&= \mathbf{E}^{xpct}\left[\frac{1}{4}\left\|\phi_{[1]}\right\|_2\right] - \mathbf{E}^{xpct}\left[\left\|\phi_{[\frac{1}{2}]}\right\|_2\right] \\
&= \frac{1}{4}\mathbf{var}\left[\phi_{[1]}\right] - \mathbf{var}\left[\phi_{[\frac{1}{2}]}\right] \\
&= 0. \quad \text{(By hypothesis 2)}
\end{aligned}$$

□[1]

Claim 2:
For any N in $\mathbb{N}$ and $n \in [1..2^{N-1}]$, the functions $\beta(t_{[\frac{2n}{2^N}]} \to t_{[\frac{2n+1}{2^N}]})$ and $\beta(t_{[\frac{2n+1}{2^N}]} \to t_{[\frac{2n+2}{2^N}]})$ are orthogonal.

Proof:
Exactly the same as the previous case.
□[2]

The general case is left as an **Exercise**. The idea is this: for an arbitrary quadruple of dyadic times $t_0 < T_0 < t_1 < T_1$, find the highest power of 2 common to all denominators —let's say its N —and break the time intervals down into subintervals of size 2^N. Now use Lemma 4.2.3 to express the increments $\beta(t_0 \to T_0)$ and $\beta(t_1 \to T_1)$ as sums of random variables, so that all variables in the *first* sum are orthogonal to all variables in the

second sum. $\beta(t_0 \to T_0)$ and $\beta(t_0 \to T_0)$ are therefor orthogonal to each other.

□

What about stationarity?

4.2.28

Lemma

Suppose that

1. For all $q \in \mathbb{Q}_2$, the probability distribution of $\phi_{[q]}$ is **symmetric:**

$$\mathbf{D}^{istr}\left[\phi_{[q]}\right] \;=\; \mathbf{D}^{istr}\left[-\phi_{[q]}\right]$$

2. For every $N \in \mathbb{N}$, the random variables

$$\phi_{[\frac{1}{2^N}]},\ \phi_{[\frac{3}{2^N}]},\ \phi_{[\frac{5}{2^N}]},\ \cdots,\ \phi_{[\frac{2^N-1}{2^N}]}$$

are **identically distributed.**

Then the limit process β has **stationary increments.**

Proof:
Exercise.
□

Combining all of the preceeding lemmas, we can now provide the following general construction.

4.2.29

Theorem L

Let V be a Hilbert space, and let $\Phi : \Omega \longrightarrow V$ be a random variable For all $q \in \mathbb{Q}_2$, let $\Phi_{[q]}$ be a copy of Φ, so that all the random variables $\left\{\Phi_{[q]}\big|_{q \in \mathbb{Q}_2}\right\}$ are independent yet identically distributed.

Now, for every $q \in \mathbb{Q}_2$, if $q = \frac{n}{2^N}$, then define:

$$\phi_{[q]} := \frac{1}{2^N}\Phi_{[q]}$$

and use the collection $\left\{\phi_{[q]}|_{q\in\mathbb{Q}_2}\right\}$ as a **perturbation series** to build a stochastic process β on V through a sequence of **dyadic polygonal approximations** as described above. Then

1. β originates at zero.
2. If Φ is **symmetric**, then β has **stationary increments.**
3. If, in addition, Φ has **finite variance**, then β also has **balanced, independent increments.**
4. Finally, suppose we could find a **summable** sequence $\left\{b_N|_{N\in\mathbb{N}}\right\}$ of positive numbers so that, if, for each $\frac{n}{2^N}\in\mathbb{Q}_2$, we define

$$a_{[\frac{n}{2^N}]} := \mathbf{P}_{rob}\left[\|\Phi\| > 2^N b_N\right]$$

then the sequence $\left\{a_{[q]}|_{q\in\mathbb{Q}_2}\right\}$ is also summable.

Then the sample paths of β are **almost surely continuous.**

Proof:
Combine the preceeding lemmas.
□

4.2.30

Example L

An excellent example application of the method of Polygonal Approximations can be found in Lévy 's construction of Brownian Motion, discussed in [4.3.2].

4.3 Brownian Motion

4.3.1 Informal Discussion

Prerequisites: [4.1.1]

Brownian motion is one of the most important stochastic processes; indeed, in some sense, it is the "canonical" random process. Brownian motion can be thought of as as process describing the random movements of an erratically jiggling particle. As time passes, the particle slowly drifts through space, gradually wandering away from its starting point. The original physical inspiration for Brownian motion was the random jiggling motions of a dust mote, being randomly jostled by the continual bombardment of millions of much smaller, fast-moving air molecules.

If we know that the particle is at position p at time 0, then, at time t, the particle's position will be a **normal random variable**, with a **mean** of p, and a **variance** of t. Thus, the particle's "mean distance" from its starting point is proportional to the *square root* of the amount of time which has passed.

One way to see Brownian motion is as the **continuum limit** of a discrete-time random walk. In a **simple random walk**, the particle walks through a lattice in discrete time; during each second, the particle takes a random step of size ± 1 in each dimension of the lattice.

We want to use this discrete model to approximate a continuous-time process; hence, we want to use the lattice as a sort of "discrete approximation" of Euclidean space, and regard the integers as a "discrete approximation" of continuous time. To improve the quality of the approximation, we increase the "resolution" of the lattice, by building it on a smaller and smaller scale.

So, imagine we *rescale* time by a factor of n; thus, during each second, the particle takes not one, but n random steps. After one second, the position of the particle is thus a sum of n independent, identical random variables. The **Central Limit Theorem** says that, as n gets large, the probability distribution of the particle's position after one second will begin to resemble a **Gaussian distribution** of **variance** n; hence, its "mean distance" from the origin will be $\sqrt{n}$.

Thus, to prevent the particle's motion from "exploding off to infinity" as we make n large, we must also rescale our *spacial* lattice by a factor of $\sqrt{n}$. A smaller rescaling value will allow the particle to escape to infinity as n gets large; a larger value will "crush" it into the origin; hence, $\sqrt{n}$ is the only sensible choice.

In a certain sense, we can construct a continuous-time process as the "limit" as n goes to infinity of a random walk, with time rescaled by n, and space rescaled by $\sqrt{n}$. The resulting process is Brownian motion.

4.3.2 Formal Construction(s)

Corequisites: [4.3.1]

There are various ways to formally construct Brownian motion, which provide different perspectives and insights into its properties.

So, let V be some **finite-dimensional vector space.** Since we want to endow Brownian motion with assorted "Gaussian" properties, we want to know what it means to say that a random variable ranging over V has a "normal" distribution. Because of this, it is necessary to endow V with some **inner product** structure. In practice, V is almost always just going to be $\mathbb{R}^D$ with the standard inner product. Indeed, in many cases, V is just the real line.

Via the Kolmogorov Consistency Theorem

Prerequisites: [4.2.1]

The simplest way to construct Brownian motion is by application of **Kolmogorov's Theorem**. This approach has the advantage of being relatively simple in terms of its technical requirements. Its disadvantages are twofold: at an intuitive level, it provides little insight into what Brownian motion "really is", while, at a mathematical level, it is not a very useful perspective for studying the actual properties of the process.

In order to apply Kolmogorov's theorem, we must define a **consistent** set of **finite-dimensional marginal probability distributions**.

So, let $\mathbb{T}_{ime} := [0, \infty)$, and let

$$\mathcal{A} := \{A \subset \mathbb{T}_{ime} \ ; \ \text{A finite}\}.$$

For every $A \in \mathcal{A}$, we will define will define a probability measure μ_A on V^A.

Let $A \in \mathcal{A}$, and suppose $A := \{a_1 < a_2 < \ldots < a_N\}$. It suffices[] to define the measure μ_A on **measurable rectangles** of the form:

$$U \ := U_1 \times U_2 \times \ldots \times U_N.$$

Now, intuitively, $\mu_A[U]$ should be the probability that the Brownian particle, starting at the origin at time zero, passes through the set U_1 at *time* a_1, set U_2 at *time* a_2, etc. , and eventually ends up in set U_N at *time* a_N.

Let $\vec{X} := \left(\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_N\right)$ be a random variable ranging over the space V^A, where, for all $n \in [2..N]$, the random variable $\vec{X}_n \in V$ is supposed to be the *increment* of Brownian motion from time a_{n-1} to time a_n, and $\vec{X}_1 \in V$ is the increment from time 0 to time a_1.

The Brownian particle's increments in time are **independent normal random variables** , whose variances are equal to the lengths of their respective time intervals. Thus,

$$\begin{aligned}
\mathbf{D}^{istr}\left[\vec{X}_1\right] &= \overbrace{0;\ a_1}\\
\mathbf{D}^{istr}\left[\vec{X}_2\right] &= \overbrace{0;\ a_2 - a_1}\\
\mathbf{D}^{istr}\left[\vec{X}_2\right] &= \overbrace{0;\ a_3 - a_2}\\
&\ddots\\
\mathbf{D}^{istr}\left[\vec{X}_N\right] &= \overbrace{0;\ a_N - a_n{-}1}
\end{aligned}$$

V is an **inner product space**; assume we have equipped it with some orthonormal basis. It doesn't really matter what we call this basis, but all the matrices described below will be written with respect to it.

First let's define **covariance matrices** for $\mathbf{D}^{istr}\left[\vec{X}_1\right], \ldots \mathbf{D}^{istr}\left[\vec{X}_N\right]$, as follows:

$$\begin{aligned}
\boxed{A_1} &= \begin{bmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_1 \end{bmatrix}\\
\boxed{A_2} &= \begin{bmatrix} a_2 - a_1 & 0 & \ldots & 0 \\ 0 & a_2 - a_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_2 - a_1 \end{bmatrix}\\
\boxed{A_3} &= \begin{bmatrix} a_3 - a_2 & 0 & \ldots & 0 \\ 0 & a_3 - a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_3 - a_2 \end{bmatrix}\\
&\vdots
\end{aligned}$$

$$\boxed{A_N} = \begin{bmatrix} a_N - a_{N-1} & 0 & \dots & 0 \\ 0 & a_N - a_{N-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_N - a_{N-1} \end{bmatrix}.$$

Thus, $\vec{X}$ itself is a normal random variable with **covariance matrix**

$$\boxed{\mathrm{A}} := \begin{bmatrix} \boxed{A_1} & \boxed{0} & \boxed{0} & \dots & \boxed{0} \\ \boxed{0} & \boxed{A_2} & \boxed{0} & \dots & \boxed{0} \\ \boxed{0} & \boxed{0} & \boxed{A_3} & \dots & \boxed{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \boxed{0} & \boxed{0} & \boxed{0} & \dots & \boxed{A_N} \end{bmatrix}$$

where $\boxed{0}$ refers to the **zero matrix** on V.

If $\vec{B} := (\vec{B}_1, \vec{B}_2, \dots, \vec{B}_N) \in V^A$ is the random vector indicating the actual **positions** of the Brownian particle at times $a_1, a_2, \dots, a_N$, then clearly,

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \\ Y_3 &= X_1 + X_2 + X_3 \\ &\vdots \\ Y_N &= X_1 + X_2 + \ldots + X_N \end{aligned}$$

That is, $Y = F(X)$, where $F : V^A \longrightarrow V^A$ is the **linear transformation** having **lower triangular matrix** :

$$\boxed{\mathrm{B}} := \begin{bmatrix} \boxed{\mathbf{Id}_V} & \boxed{0} & \boxed{0} & \dots & \boxed{0} \\ \boxed{\mathbf{Id}_V} & \boxed{\mathbf{Id}_V} & \boxed{0} & \dots & \boxed{0} \\ \boxed{\mathbf{Id}_V} & \boxed{\mathbf{Id}_V} & \boxed{\mathbf{Id}_V} & \dots & \boxed{0} \\ \vdots & \vdots & \vdots & \ddots & \ddots \\ \boxed{\mathbf{Id}_V} & \boxed{\mathbf{Id}_V} & \boxed{\mathbf{Id}_V} & \dots & \boxed{\mathbf{Id}_V} \end{bmatrix}$$

where $\boxed{\mathbf{Id}_V}$ is the **identity** matrix for V.

Claim 1: The random variable Y is be **normal** , having **covariance matrix** :

$$\boxed{\mathrm{C}} := \boxed{\mathrm{B}}^{-1} \cdot \boxed{\mathrm{A}} \cdot \boxed{\mathrm{B}}.$$

Proof: Exercise.
□1
So, define the probability measure μ_A on V^A to simply be the **normal** distribution with covariance matrix $\boxed{C}$:

$$\mu_A := \overbrace{\int 0; \boxed{C}}.$$

Do this for all finite subsets $A \subset \mathbb{T}_{ime}$.

Claim 2: The collection of measures

$$\{\mu_a \; ; \; A \subset \mathbb{T}_{ime}, \; \text{A finite}\}$$

satisfies the **Kolmogorov Consistency Criterion.**

Proof: Exercise.
□2
Now let μ be the probability measure on $V^{\mathbb{T}_{ime}}$ induced by this collection. This defines a stochastic process, which, by construction, must be **Brownian Motion.**

4.3.1

Remark

This construction is really just a special case of the proof of Lemma [4.2.2].

Via Independent Gaussian Increments

Prerequisites: [4.2.2]

Recall that the physical situation that originally motivated Brownian motion involves a dust mote being jostled about by the random bombardment of invisible, high-velocity microscopic particles. Mathematically, this

suggests that a Brownian motion process should be one built out of a random succession of independent, random **increments**.

We thus construct Brownian motion by specifying a process B with the properties:

1. $B(0) = 0$ almost surely.
2. For all $T > t > 0, B(t \to T)$ is a **normal random variable** ranging over V with **mean** zero and **variance** $(T - t)$.
3. B has **independent increments**.

Exercise: Check that this produces a process whose finite-dimensional distributions are identical with those given by the Kolmogorov Construction.

Via Polygonal Approximation

Prerequisites: [4.2.3]

The "polygonal approximation" approach is similar to, but not the same as, the "lattice approximation" described informally in [4.3.1] . This construction has the advantage that it immediately implies the *continuity* of the particle's motion. It was first developed by Paul Lévy .

We will apply Theorem [**??**] . In the notation of this theorem, let Φ be a **normal** random variable, ranging over V, and having distribution:

$$\mathbf{D}^{istr}[\Phi] = \overbrace{0;1}$$

Then the resulting process has **independent, stationary Gaussian increments**, and **originates at zero**. Thus, by Corollary [4.1.8] , it must be Brownian Motion.

4.3.2

Theorem

The sample paths of Brownian Motion are almost surely continuous.

Proof:

Claim : For any $a > 0$,

$$\mathbf{P}_{rob}\left[\|\Phi\|_2 > a\right] \leq \exp\left(\frac{-a^2}{2}\right)$$

Proof: Exercise. □

Thus, if we define $b_n := 2^{\frac{-n}{2}}$, then $\left\{b_n|_{n\in\mathbb{N}}\right\}$ is a summable sequence, and, if, for all $\frac{n}{2^N} \in \mathbb{Q}_2$, we let

$$\begin{aligned} a_{[\frac{n}{2^N}]} &:= \mathbf{P}_{rob}\left[\|\Phi\| > 2^N b_N\right] \\ &= \mathbf{P}_{rob}\left[\|\Phi\| > 2^N 2^{\frac{-N}{2}}\right] \\ &= \mathbf{P}_{rob}\left[\|\Phi\| > 2^{\frac{N}{2}}\right] \\ &\leq \exp\left(\frac{-2^N}{2}\right) \quad \text{(by Claim)} \\ &\leq \exp\left(-2^{N-1}\right) \end{aligned}$$

Then $\left\{b_q|_{q\in\mathbb{Q}_2}\right\}$ is also summable, so we are done, by Lévy 's theorem. □

4.3.3 Brownian Motion with Drift

Consider a *modified* Brownian process, with a "drift" term added. Specifically, for any vector $\vec{v} \in \mathbb{R}^D$, define the process $B^{\vec{v}}$ by:

- $B^{\vec{v}}(0) = 0$, almost surely.
- $B^{\vec{v}}$ has **independent increments**.
- For all $t_0, t_1 \in \mathbb{T}_{ime}$,

$$B(t_0 \to t_1) \;=\; \mathcal{N}\left((t_1 - t_0)\cdot\vec{v};\;\; |t_1 - t_0|\right)$$

If **Brownian motion** represents a particle randomly diffusing in $\mathbb{R}^D$, then $B^{\vec{v}}$ represents a particle *trying* to randomly diffuse, while being subjected to a steady "wind" in the direction of $\vec{v}$.

This process has stationary increments, although it itself is not stationary.

4.4 Stable Processes

4.4.1 Informal Discussion

Prerequisites: [4.1.1]

Corequisites: [4.3.1]

Symmetric Stable Random Variables:

Stable random variables are generalisations of Gaussians. The Gaussian distribution is very "natural" or "universal", because the **Central Limit Theorem** says that it appears ubiquitously as the limiting distribution of any sum of an infinite collection of independent, identically distributed random variables of finite variance, if we perform a suitable "renormalisation".

Specifically, if $X_1, X_2, \ldots$ are *i.i.d.* variables on $\mathbb{R}^D$ of finite variance and zero mean, then the random variable

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} X_n$$

converges in distribution to a Gaussian on $\mathbb{R}^D$ as N tends to infinity.

Suppose, however, that the variables $X_1, X_2, \ldots$ did *not* have finite variance. What then? It turns out that more general theorems exist, similar in flavour to the Central Limit Theorem, and asserting that, for a suitable choice of "renormalisation exponent" $\alpha \in (0, 2]$, we can often make the sum:

$$\frac{1}{N^{\frac{1}{\alpha}}} \sum_{n=1}^{N} X_n$$

converge to some limiting distribution. In the case when $\alpha = 2$, this is of course just the Central Limit Theorem.

These limiting distributions are called **stable random variables** . For a fixed exponent $\alpha \in (0, 2]$, a **symmetric $\alpha-$stable random variable** is described by a single parameter, the **variation** , which is a number $\sigma \geq 0$. In the case $\alpha = 2$, an $\alpha-$stable random variable is a Gaussian, and the variation corresponds to the **standard deviation** of the distribution (**Note:** The variation is *not* the *variance*) . The characteristic function

of a Gaussian random variable of standard deviation σ and mean zero is a function $\Phi : \mathbf{R}^D \longrightarrow \mathbf{C}$ given by:

$$\Phi(\vec{\lambda}) \ := \ \exp\left(\left|\sigma \cdot \vec{\lambda}\right|^2\right)$$

The characteristic function of a general $\alpha-$stable distribution with variation σ will be:

$$\Phi(\vec{\lambda}) \ := \ \exp\left(\left|\sigma \cdot \vec{\lambda}\right|^\alpha\right)$$

Unfortunately, in general, no explicit formula exists describing the **density** of $\alpha-$stable random variables, for $\alpha < 2$.

Skewed Stable Random Variables:

So far we have only discussed *symmetric* $\alpha-$stable random variables. Of course, the general Gaussian random variable is not symmetric about zero, but rather, has a nonzero **mean** . In the case of a general $\alpha-$stable random variable, there is also a *fourth* parameter, the **skewness** , which can influence the symmetry of the distribution. For simplicity, we will consider only the one-dimensional case.

If $\alpha \in (0,2], \sigma \in [0,\infty), v \in \mathbf{R}$, and $\beta \in [-1,1]$, then an $\alpha-$stable random variable of **variation** σ, **bias** v, and **skewness** β is one possessing the characteristic function $\Phi : \mathbf{R} \longrightarrow \mathbf{C}$ given by:

$$\Phi(\lambda) \ := \exp(\Psi^{[\alpha]}_{[\sigma;v;\beta]}(\lambda))$$

where

$$\Psi^{[\alpha]}_{[\sigma;v;\beta]}(\lambda) \ := \ iv\lambda \ - \ |\sigma \cdot \lambda|^\alpha \cdot \left(1 - \text{sign}(\lambda) i\beta \tan\left(\alpha\frac{\pi}{2}\right)\right)$$

if $\alpha \neq 1$, and

$$\Psi^{[1]}_{[\sigma;v;\beta]}(\lambda) \ := \ iv\lambda \ - \ |\sigma \cdot \lambda| \cdot (1 + \text{sign}(\lambda) i\beta \log|\lambda|)$$

If $\alpha > 1$, then the "bias" term v is identical with the **mean** of the random variable (as one might expect from the Gaussian case) . If $\alpha \leq 1$, however, the two are no longer necessarily the same.

The multidimensional case is more complicated. One might expect from the Gaussian case that the general multidimensional $\alpha-$stable random variable would simply be a *sum* of distinct one-dimensional random variables living in linearly independent dimensions. The presence of the *skewness*

terms, however, introduces additional complications; in general, the skewness of a $D-$dimensional $\alpha-$stable random variable cannot be represented by simply a $D-$tuple of numbers representing the "skewnesses" in different directions, but must instead be represented by a *measure* on the $D-$dimensional unit sphere, commonly called a **spectral measure**[12]

The Stability Property:

Gaussian random variables have an important **stability** property: if X_1 and X_2 are independent Gaussians of variances σ_1^2 and σ_2^2, then the sum, $X_1 + X_2$ is *also* a Gaussian, with variance $\left(\sigma_1^2 + \sigma_2^2\right)^{1/2}$.

This property carries over to general $\alpha-$stable random variables. For fixed $\alpha \in (0, 2]$, if X_1 and X_2 are independent $\alpha-$stable random variables, with variations σ_1 and σ_2, then the sum, $X_1 + X_2$ is *also* $a\alpha-$stable, with variance $(\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}$.

Stable Processes:

Just as we built a Brownian motion process by subjecting a particle's position to Gaussian perturbations, we can build a stochastic process by subjecting it to $\alpha-$stable random perturbations[13] . The resulting process is called a **stable process**.

For more information on $\alpha-$stable random variables and $\alpha-$stable processes, the definitive reference is [8] . An earlier development of the theory of $\alpha-$stable random variables (in the one-dimensional case) can be found in [5] , chapt. XVII. 4

4.5 Poisson Processes

4.5.1 Informal Discussion

Imagine a bank machine on a street corner. At random times, people walk up and use the bank machine. Some of them deposit money, and some

[12]This choice of terminology is most unfortunate. *These* "spectral measures" have no relation with the **spectral measures** associated to linear operators.

[13]In terms of the physical model of Brownian motion as the jiggling of a dust mote under random bombardment by smaller high-energy particles, we can imagine that this corresponds to bombarding the dust mote with extremely energetic particles, capable of delivering much more powerful "kicks"

of them withdraw it, and different people deposit or withdraw different amounts of cash.

Thus, over the course of the day, the amount of cash in the machine will fluctuate. Given that we cannot predict the arrival times or financial needs of the customeers, we can regard these fluctuations as "random". How might we model thus?

The fluctuating cash levels of the bank machine are an example of a **Poisson Process**. A Poisson process is a stochastic process which remains constant over randomly distributed intervals of time, only to suddenly change by discrete, random amounts.

Exponentially Distributed Random Variables

The lengths of the time interval between two customers is a random amount, but one thing we can say is that it is likely to be shorter, rather than longer. Suppose that there is a constant probability, p, that a customer will arrive during any given second. Then the odds of *no* customers for a period of ten seconds is $(1-p)^{10}$ In general, the odds of an idle period of n seconds duration is $(1-p)^n$ –in other words, an amount which decreases exponentially with time. The time interval is an **exponentially distributed random variable**.

Simple Poisson Processes

Now what about the withdrawals? First, assume all the customers do exactly the same thing. Let's say each one of them withdraws $ 100. Thus, over time, the amount of money in the machine will steadily decrease[14].

The amount of money withdrawn from the machine after 5 hours will then be $100 \cdot N$, where N is the number of customers who have visited. N is a random number, however; it is the number of independent, exponentially distributed time intervals that have transpired in the 5 hour period. A process like this is called a **Simple Poisson Process**.

Compound Poisson Processes

It's a bit unrealistic to assume all the customers withdraw exactly $100. Let's give them a bit more character, by letting them withdraw/deposit random amounts. However, let's assume that all the customers are "statistically identical", in the sense that all of them are equally likely to with-

[14] Of course, it will eventually go "Out of service", but never mind that.

draw/deposit money of any given quantity[15] Let's say that each customer's withdrawal/deposit is a random variable having distribution μ, where μ is some probability measure on $\mathbb{R}$. Let's further assume that the different customers are all *independent* of one another.

Thus, as time passes, the amount of money in the machine will be a random variable which is a sum of N *independent* random variables, each having distribution ν. The number N itself is also random; it is the number of exponentially distributed time intervals which have passed since we began observing. This is an example of a **compound Poisson process.** The process is described by two parameters; the distribution ν and the **intensity parameter**, p, which describes the "average frequency" of customers.

Compound Poisson Processes of Infinite Intensity

Now, this bank machine is only one of many millions of bank machines throughout the galaxy which belong to the Bank of Orion. Every millisecond[16], millions of customers throughout the galaxy are depositing or withdrawing money; hence, the total assets of the Bank of Orion are fluctuating continuously.

We can treat the assets of the Bank as one huge Poisson process, built by combining many small processes. Indeed, if we allow the size of our galaxy to become "infinitely big", then the **intensity parameter** p, describing the frequency of transactions, will become infinite. The only way we can prevent the assets of the Bank from fluctuating wildly all over the place is to assume that the "vast majority" of transactions which take place are very, very small. Hence, the total asset level of the Bank is subjected to infinitely many perturbations, but most of them are so tiny that they only cause the process to wiggly imperceptibly. Withdrawals which are large enough to cause noticeable jumps only happen once in a while (ie. with **finite** intensity).

The Bank's assets are also a **Compound Poisson Process** –this time, one of **infinite intensity**.

The Role of Poisson Processes

Poisson processes have many applications in stochastic modelling. For example, they are of central importance in **Queuing theory**, as the "Bank

[15] In practice, of course, this may not be true; it may be that customers arriving in the late afternoon have statistically different behaviour than those making withdrawals at 10:00 PM. However, let's keep it simple.

[16] Set aside relativistic concerns about "simultaneity" for the moment.

machine" example shows. They also appear in many other phenomena involving independent, randomly timed events. For example:

- The total radiation output, over time, from a radioactive material (a simple Poisson Process).
- The total Coloumbs of charge that pass from the sky to the ground through lightening strikes (a compound Poisson process).

Poisson processes are also of central *theoretical* importance in the study of stochastic processes. One of the most important and "canonical" types of process is a Lévy process [4.6.1], and any Lévy process can be expressed as a sum of three components [4.6.5], one of which is a compound Poisson process (of possibly infinite intensity).

4.5.2 Simple Poisson Processes

Prerequisites: [??] [4.1.1] **Corequisites:** [4.5.1]

We have seen in [??] how a exponential random variable describes the stochastic behaviour of a single event occuring randomly in time.

Now, suppose that we have a random event X, whose successive occurences are described by an exponential distribution. In other words, there is some negative real number q so that, for any time interval $[a, b)$,

$$\mathbf{P}_{rob}\left[X \text{ happens during } [a,b)\right] \quad = \quad 1 - e^{q(b-a)}.$$

Suppose that we are "counting" the occurences of X over time. We thus have a tally, $N(t)$, representing the number of times X has occured during the time interval $[0, t)$. Every time X occurs, we add one to our tally.

$N(t)$ is thus a random variable, taking its values in the natural numbers. We expect N to grow over time, and we further expect that, for any successive times $a < b$:

$$\mathbf{P}_{rob}\left[N(b) > N(a)\right] \quad = \quad 1 - e^{q(b-a)}.$$

The function N is thus a stochastic process; it is called the **Simple Poisson Process** with **parameter** q. (Sometimes, q is called the **intensity** of the process).

If we fix a time t, then the random variable $N(t)$ is called *a* **Poisson random variable**, and its properties are well-understood.

Formally....

4.5.1

Definition *Simple Poisson Process*

Let $q > 0$, and let $\alpha_1,\ \alpha_2,\ \alpha_3, \ldots$ be **independent, exponentially distributed** random variables with parameter q.

Define a stochastic process $N : \Omega \times [0, \infty) \longrightarrow \mathbb{N}$ as follows:

For any $t > 0$,

$$N(t) \ := \ \sup \left\{ N \ ; \ \sum_{n=1}^{N} \alpha_n \ < \ t \right\}$$

In other words, $N(t)$ is the number of exponentially distributed time intervals which have elapsed prior to time t.

Then N is the **simple Poisson process** with **parameter** q.

4.5.2

Theorem *Distribution Properties of the Simple Poisson Process*

Let $N : \Omega \times [0, \infty) \longrightarrow \mathbb{N}$ be a **simple Poisson process** of parameter q. Then, for any fixed $t > 0$,

1. The random variable $N(t)$ has probability distribution (on $\mathbb{N}$) given by:

$$\forall n \in \mathbb{N}, \ \mathbf{P}_{\text{rob}} \left[N(t) = n \right] \ = \ \frac{(q \cdot t)^n e^{-qt}}{n!}.$$

2. The expected value of $N(t)$ is $q \cdot t$.

3. The **characteristic function** of $N(t)$ (a function on $\mathbb{R}$) is given by:

$$\lambda \ \mapsto \ \exp\left(-q(1 - e^{i\lambda})\right)$$

Proof: Exercise.

□

4.5.3

Definition *Poisson Random Variable*

A **Poisson random variable** of **parameter** q is a random variable γ taking values in $\mathbb{N}$, so that,

$$\forall n \in \mathbb{N},\ \mathbf{P}_{rob}\left[N(t) = n\right] \ = \ \frac{(q \cdot t)^n e^{-qt}}{n!}.$$

4.5.4

Remark

1. Thus, the theorem says that $N(t)$ is a Poisson random variable of parameter $q \cdot t$.

2. A simple Poisson process, by definition, has **independent increments**. Thus, for any times $s < t$, the random variable $N(t) - N(s)$ is *also* a Poisson random variable, with parameter $q \cdot (t - s)$.

Vector Valued Poisson Processes

Now consider a slightly different version of the same process. Now, however, instead of adding unit amounts to some "tally", we will move through a vector space by some fixed amount.

Let $\vec{v} \in \mathbb{R}^D$ be a fixed vector. Begin at the origin in $\mathbb{R}^D$, and, whenever the random event X occurs, move by a displacement of $\vec{v}$ through $\mathbb{R}^D$.

Formally, we have a stochastic process,

$$V : \Omega x[0, \infty) \longrightarrow \mathbb{R}^D,$$

defined:

$$V(\omega, t) \ = \ N(\omega, t) \cdot \vec{v}.$$

This is called the **vector-valued Poisson process** of **intensity** q, in **direction** $\vec{v}$.

The next theorem is then an immediate corollary of the previous result:

4.5.5

Theorem

1. The random variable $V(t)$ has probability distribution (on $\mathbb{N} \cdot \vec{v}$) given by:

$$\forall n \in \mathbb{N},\ \mathbf{P}_{rob}\left[N(t) = n \cdot \vec{v}\right] \ = \ \frac{(q \cdot t)^n e^{-qt}}{n!}.$$

2. The expected value of $N(t)$ is $qt \cdot \vec{v}$.

3. The **characteristic function** of $N(t)$ (a function on $\mathbb{R}^D$) is given by:

$$\vec{\lambda} \ \mapsto \ \exp\left(q(1 - e^{i\langle \vec{\lambda}, \vec{v} \rangle})\right).$$

4.5.3 Compound Poisson Processes of Finite Intensity

Prerequisites: [4.5.2]

Now, suppose that X is some event which occurs randomly in time, whose arrival times are exponentially distributed, as described in [**??**] . In the case of a *simple* Poisson process (described in [4.5.2]) , every time the event X occured, we added 1 to our "tally". Now, however, suppose we add a *random* amount to our tally instead. Further, rather than merely adding a random scalar, we can imagine that each time, we add some random **vector** which lives in a vector space V.

4.5.6

Definition *Compound Poisson Process (of finite intensity)*

Let ν be some probability measure on the vector space V, and suppose that $q > 0$ is some intensity parameter.

Let $\alpha_1, \alpha_2, \alpha_3, \ldots$ be a sequence of independent, identically distributed random variables, distributed according to ν. Consider the simple Poisson process

$$N : \Omega \times [0, \infty) \longrightarrow \mathbb{N}$$

of intensity q. Now, define a new process,

$$S : \Omega \times [0, \infty) \longrightarrow V$$

as follows: for any t and ω,

$$S(\omega, t) := \sum_{n=1}^{N(\omega,t)} \alpha_n(\omega)$$

(where we consider the sum to be empty if $N(t, \omega) = 0$)

The stochastic process S is called a **compound Poisson process** . S is clearly determined by *two* parameters: q and ν. However, we know that ν must be a probability measure, and q a positive real number. Thus, we can combine these two parameters together, multiplying ν by q to create a new measure, ν_0, of total mass q. We then say that S is the compound Poisson process of **intensity measure** ν_0.

We will see later that there is reason why it in fact makes good sense to combine q and ν into a single parameter in this fashion.

What does this process look like? The name *compound* Poisson process suggests that S is rather like several distinct Poisson processes "compounded" together, and indeed, this is exactly the case.

First, suppose that ν is a very simple probability measure, focussing all of its mass on only two elements, $\vec{v}_1$ and $\vec{v}_2$, in $\mathbb{R}^D$. Suppose that $\nu(\vec{v}_1) = p_1$ and $\nu(\vec{v}_2) = p_2$. Then we can imagine the compound Poisson process S as being generated by the following procedure:

Procedure 1:

We start with a "random" alarm clock, which, once set, will go off at a random time in the future –a time which is exponentially distributed with parameter q. Whenever the alarm rings, we flip a biased coin, which lands "heads up" with probability p_1, and "tails up" with probability p_2. If it lands "heads up", then we add $\vec{v}_1$ to our tally; if it lands "tails up", we add $\vec{v}_2$.

The flipping of the coin, however, is an event which is *totally independent* of the timing of the alarm. A proportion of exactly p_1 of the alarms will result in the addition of $\vec{v}_1$, and a the remaining proportion, of p_2, of the alarms, will result in the addition of $\vec{v}_2$. Thus, in a sense, the addition of

copies of $\vec{v}_1$ and $\vec{v}_2$ can be regarded as two **independent** Poisson processes, occuring at rates $p_1 \cdot q$ and $p_2 \cdot q$, respectively. In other words, we could produce the identical stochastic process through. . .

Procedure 2:

We start with *two* random alarm clock, which go off at exponentially distributed times. The first one has alarms exponentially distributed with intensity $p_1 q$, the second with intensity $p_2 q$. The two alarm clocks operate independently. Whenever Alarm 1 goes off, we add $\vec{v}_1$ to our tally; whenever Alarm 2 goes off, we add $\vec{v}_2$ to our tally.

Thus, in a sense, we can regard the compound Poisson process S to be a sort of "linear combination" of two independent simple Poisson processes. Formally, if N_1 and N_2 are independent, simple Poisson processes of intensity $p_1 q$ and $p_2 q$ respectively, then we could write:

$$S(\omega, t) \;=\; \vec{v}_1 N_1(\omega, t) \;+\; \vec{v}_2 N_2(\omega, t)$$

This allows us to express the characteristic function for the random variable $S(t)$ in terms of those for $N_1(t)$ and $N_2(t)$.

Fix $t \in [0, \infty)$, and, for $k = 1, 2$, let Φ_k denote the characteristic function of $N_k(t)$. Let Ψ denote the characteristic function of $S(t)$ (these are all functions from V into $\mathbf{C}$) . Then we know by the results of [4.5.2] that

$$\Phi_k(\vec{\lambda}) = \exp\left(-p_k q(1 - e^{i\langle \vec{v}_k, \vec{\lambda}\rangle})\right)$$

Since Ψ is a sum of two independent random variables, its characteristic function is a **product** of theirs. That is:

$$\Psi = \Phi_1 \cdot \Phi_2$$

so that, for any $\vec{\lambda} \in \mathbf{R}^D$,

$$\begin{aligned}
\Psi(\vec{\lambda}) &= \exp\left(-p_1 q(1 - e^{i\langle \vec{v}_1, \vec{\lambda}\rangle})\right) \cdot \exp\left(-p_2 q(1 - e^{i\langle \vec{v}_2, \vec{\lambda}\rangle})\right) \\
&= \exp\left(-p_1 q(1 - e^{i\langle \vec{v}_1, \vec{\lambda}\rangle}) - p_2 q(1 - e^{i\langle \vec{v}_2, \vec{\lambda}\rangle})\right) \\
&= \exp\left(-q \sum_{k=1}^{2} p_k (1 - e^{i\langle \vec{v}_k, \vec{\lambda}\rangle})\right)
\end{aligned}$$

Indeed, if we let $\nu_0 := q \cdot \nu$ denote the **intensity measure** of the process, then we can rewrite this last expression as:

$$\exp\left(-\int_{\mathbf{R}^D}(1-e^{i\langle\vec{v},\vec{\lambda}\rangle})d\nu_0[\vec{v}]\right)$$

This simple example motivates the following theorem. The proof of the theorem follows exactly the same logic.

4.5.7

Theorem Let μ be a finite measure on $\mathbf{R}^D$, and let $S:\Omega\times[0,\infty)\longrightarrow\mathbf{R}^D$ be a compound Poisson process, with intensity measure μ. Then, for any $t>0, S(t)$ is a random variable whose characteristic function Ψ_t is given by:

$$\Psi(\vec{\lambda}) \;:=\; \exp\left(-\int_{\mathbf{R}^D}(1-e^{i\langle\vec{v},\vec{\lambda}\rangle})d\mu[\vec{v}]\right)$$

□

4.5.8

Remark

This characteristic function is the best description we have for the distribution of an arbitrary compound Poisson process; in general, there is no closed form expression for the *density* of such a process.

4.5.4 Poisson Measures

Prerequisites: [5.2] , [4.5.3]

In [5.2] , we showed how a simple Poisson process can be represented as the "integral" of a certain *random measure* ; one consisting of isolated "atoms" of unit mass, distributed randomly in time with exponentially distributed distances between them. In fact, *any* Poisson process can be represented as the integral of a random measure; these measures are called **Poisson Measures** .

4.5.9

Definition *Poisson Measure (of finite intensity)*

Let V be a Banach space[17] , and let μ be a finite measure on V. We define η, the **Poisson measure** on $[0,\infty)$ of **intensity** μ as follows:

Suppose $\mu(V) := c$. Let $\nu := \frac{1}{c}\mu$. Then ν is a probability measure.

Let $X_1, X_2, X_3, \ldots$ be an infinite collection of independent random variables ranging over V, all distributed according to ν. Let $T_1, T_2, T_3, \ldots$ be an infinite collection of independent random variables ranging over $[0,\infty)$, all exponentially distributed with parameter c.

For every $n \in \mathbb{N}$, let δ_n be a point mass placed randomly in $[0,\infty)$ as follows:

- The location of δ_n is $T_1 + T_2 + \ldots + T_n$.
- The mass of δ_n is X_n.

Now let $\eta := \sum_{n=1}^{\infty} \delta_n$.

4.5.10

Proposition

The Poisson measure η of intensity μ is just the **derivative** (in the sense of [5.2]) of the **compound Poisson process** of intensity μ. In other words, if we define the stochastic process N by:

$$N(t) \ := \ \eta[0,t]$$

then $N(t)$ is a compound Poisson process with intensity measure μ

Proof: Exercise□

The restriction to *finite* intensity measure seems somewhat artificial in the definition of a Poisson measure. After all, if μ was a **sigma-finite**

[17] Again, the "Banach" condition is just a distracting technicality. We want a vector space so that we can add, and we want a topology so that infinite sums will converge. For most normal purposes, you can assume $V = \mathbb{R}^D$.

measure, then perhaps we could express μ as $\mu := \sum_{n=1}^{\infty} \mu_n$, where each μ_n is a finite measure. For each μ_n, we could construct a corresponding Poisson measure η_n of intensity μ_n, as in part 1, and assume that, as random variables, $\eta_1, \eta_2, \ldots$ are all **independent**. Then we could define $\eta := \sum_{n=1}^{\infty} \eta_n$ and call it the Poisson measure of intensity μ.

The problem with this approach is that there is no guarantee that this infinite sum of random measures *converges*. Indeed, we haven't even specified a topology for the convergence. We must place some restrictions on μ in order to obviate this caveat.

4.5.11

Definition

Let μ be a measure on a Banach space V. Say that μ has **finite first moment** if the integral

$$\int_V \|v\|_d \, \mu[v]$$

has a finite value. If this is the case, denote this value by $\langle\langle \mu \rangle\rangle$.

Lemma I:

If $\langle\langle \mu \rangle\rangle$ is finite, then, for any open ball B around zero, the mass that μ assigns to the complement of B must be finite.

Proof: Suppose B has radius $\epsilon > 0$ with respect to the norm $\|\cdot\|$. Then

$$\begin{aligned} \epsilon\mu\left[V - B\right] &= \int_{V-B} \epsilon d\mu \\ &\leq \int_{V-B} \|v\| \, d\mu[v] \\ &\leq \int_V \|v\| \, d\mu[v] \end{aligned}$$

which is finite by hypothesis.
□

Hence, although the measure μ may have infinite mass, it is in a sense "finite away from zero", since the conentration of this infinite mass must take place at the origin.

Lemma II:

If μ_1 and μ_2 have **disjoint support** , then

$$\langle\langle \mu_1 + \mu_2 \rangle\rangle \;=\; \langle\langle \mu_1 \rangle\rangle + \langle\langle \mu_2 \rangle\rangle$$

Proof: : **Exercise**. □

Suppose μ is a finite measure, with **finite first moment**, and consider the Poisson measure η of intensity μ. Let U be a time interval of length L. Let $\eta_{|U}$ denote the **restriction** of η to a measure on U. Then $\eta_{|U}$ is an **almost surely finite random measure**. Hence, the **total variation norm** of $\eta_{|U}$ is an almost surely finite random variable. What is its expected value?

Lemma III:

The expected total variation norm of $\eta_{|U}$ is $\langle\langle \mu \rangle\rangle \cdot L$.

Proof:
Recall the construction of the Poisson measure η. If $c := \mu[V]$, then we define $X_1, X_2, X_3, \ldots$ as an infinite collection of independent random variables ranging over V, all distributed according to $\frac{1}{c}\mu$, and define $T_1, T_2, T_3, \ldots$ as an infinite collection of independent random variables ranging over $[0, \infty)$, all **exponentially distributed** with **parameter** c.

For every $n \in \mathbb{N}$, let δ_n be the point mass located at the random position $T_1 + T_2 + \ldots + T_n$, and define

$$\eta \;:=\; \sum_{n=1}^{\infty} X_n \delta_n.$$

Fix ω in Ω, and consider the measure $\eta^{(\omega)}$. Thus,

$$\eta^{(\omega)} \;=\; \sum_{n=1}^{\infty} X_n^{(\omega)} \delta_n^{(\omega)}$$

Where $\delta_n^{(\omega)}$ is a unit point mass located at position $T_1^{(\omega)} + \ldots + T_n^{(\omega)}$. Thus, the **total variation measure** associated to $\eta^{(\omega)}$ is defined:

$$\boxed{\eta^{(\omega)}} \;=\; \sum_{n=1}^{\infty} \left\| X_n^{(\omega)} \right\| \delta_n^{(\omega)},$$

which is yet *another* random measure on $[0, \infty)$.
Hence, the **expected value** of $\left\|\eta_{|U}\right\|_{var}$ is given by:

$$\begin{aligned}
\mathbf{E}^{xpct}\left[\left\|\eta_{|U}\right\|_{var}\right] &= \mathbf{E}^{xpct}\left[\boxed{\eta}[U]\right] \\
&= \mathbf{E}^{xpct}\left[\sum_{n=1}^{\infty} \|X_n\|\, \delta_n[U]\right] \\
&= \sum_{n=1}^{\infty} \mathbf{E}^{xpct}\left[\|X_n\|\right] \mathbf{E}^{xpct}\left[\delta_n[U]\right]
\end{aligned}$$

(since these random variables are all independent)

$$\begin{aligned}
&= \sum_{n=1}^{\infty} \left(\frac{1}{c}\int_V \|v\|\, d\mu[v]\right) \mathbf{E}^{xpct}\left[\delta_n[U]\right] \\
&= \sum_{n=1}^{\infty} \frac{1}{c}\langle\langle\mu\rangle\rangle\, \mathbf{E}^{xpct}\left[\delta_n[U]\right] \\
&= \frac{1}{c}\langle\langle\mu\rangle\rangle \sum_{n=1}^{\infty} \mathbf{E}^{xpct}\left[\delta_n[U]\right] \\
&= \frac{1}{c}\langle\langle\mu\rangle\rangle\, \mathbf{E}^{xpct}\left[\sum_{n=1}^{\infty} \delta_n[U]\right] \\
&= \frac{1}{c}\langle\langle\mu\rangle\rangle\, c \cdot L \\
&= \langle\langle\mu\rangle\rangle \cdot L
\end{aligned}$$

(since this is just the expected value of a simple Poisson process of intensity c.)

□

Now, finally, we can proceed

4.5.12

Definition *Poisson Measure*

Let V be a Banach space, and let μ be a *sigma-finite measure* on V of *finite first moment.* We can define η, the **Poisson measure** of **intensity** μ, as follows:

1. Let $\left\{\epsilon_n|_{n\in\mathbb{N}}\right\}$ be a decreasing sequence of radii converging to zero, and, for all $n \in \mathbb{N}$, let $B_n := \mathbb{B}(0, \epsilon_n)$. Then define $\mu_n := \mu_{|B_n \setminus B_{n-1}}$ (where we define $B_0 := V$) . Thus,

   (a) For all $n \in \mathbb{N}$, μ_n is a **finite measure**.

   (b) The measures $\left\{\mu_n|_{n\in\mathbb{N}}\right\}$ have **disjoint support**.

   (c) Thus, $\langle\langle\mu\rangle\rangle = \sum_{n=1}^{\infty} \langle\langle\mu_n\rangle\rangle$

2. Now, for each n, let η_n be the Poisson measure of intensity μ_n, and assume these measures are constructed so that, as random variables, they are independent.

3. Define $\eta := \sum_{n=1}^{\infty} \eta_n$. That is, for each $\omega \in \Omega$, and for any finite interval $U \subset [0, \infty)$, define the Borel measure $\eta_{|U}^{(\omega)}$ on U as the sum:

$$\eta_{|U}^{(\omega)} := \sum_{n=1}^{\infty} \left(\eta_n^{(\omega)}\right)_{|U} \quad (1)$$

   where convergence takes place with respect to the **total variation norm** on the space of $\mathcal{M}(U, V)V-$valued Borel measures over U.

Claim 1: Fix U. Then the sum (1) converges in $\mathcal{M}(U, V)$ for almost all ω.

Proof: It suffices to show that, for almost all ω, the sequence of partial sums

$$\sum_{n=1}^{N} \left(\eta_n^{(\omega)}\right)_{|U}$$

is *Cauchy* as a sequence indexed by N. To do this, it suffices to show that that the sum of norms

$$\sum_{n=1}^{\infty} \left\| \left(\eta_n^{(\omega)}\right)_{|U} \right\|_{var}$$

converges almost surely. Hence, we want to show:

Claim 1.1: The sum

$$\sum_{n=1}^{\infty} \left\| (\eta_n)_{|U} \right\|_{var}$$

is almost surely finite

Proof: This is equivilant to showing that the sum has a finite *expected value.*[18] But the summands are *independent* random variables, by construction, so

$$\begin{aligned}
\mathbf{E}^{expct}\left[\sum_{n=1}^{\infty} \left\| (\eta_n)_{|U} \right\|_{var}\right] &= \sum_{n=1}^{\infty} \mathbf{E}^{expct}\left[\left\| (\eta_n)_{|U} \right\|_{var}\right] \\
&= \sum_{n=1}^{\infty} \langle\!\langle \mu_n \rangle\!\rangle \, \ell_{ength}[U] \quad \text{(by Lemma III)} \\
&= \ell_{ength}[U] \sum_{n=1}^{\infty} \langle\!\langle \mu_n \rangle\!\rangle \\
&= \ell_{ength}[U] \, \langle\!\langle \mu \rangle\!\rangle \, . \quad \text{(by Lemma I)}
\end{aligned}$$

□[1.1] □[1]

Thus, the measure η_U is well-defined by expression (1), for any finite interval U. We can then define η as a limit of η_U, letting U grow larger to cover all of $[0, \infty)$.

Claim 2:

1. For fixed U, the measure η_U is well-defined independent of how we chose the sequence $\left\{\epsilon_n|_{n \in \mathbb{N}}\right\}$ of "decreasing radii".
2. The measure η is well-defined independent of the increasing sequence $\left\{U_n|_{n \in \mathbb{N}}\right\}$ we use to approximate $[0, \infty)$.

Proof:

1 If $\left\{\epsilon_n^{(1)}|_{n \in \mathbb{N}}\right\}$ and $\left\{\epsilon_n^{(2)}|_{n \in \mathbb{N}}\right\}$ are two descending sequences of radii, then define $\left\{\epsilon_n|_{n \in \mathbb{N}}\right\}$ to be their "common refinement". That is, define

$$E := \left\{\epsilon_n^{(k)} \; ; \; k \in \{1, 2\}, \; n \in \mathbb{N}\right\}$$

[18] If it was *not* almost surely finite, then of course its expected value would be *infinite.*

and then define:

$$\begin{aligned} \epsilon_1 &:= \max[E], \\ \epsilon_2 &:= \max[E \setminus \{\epsilon_1\}], \\ \epsilon_2 &:= \max[E \setminus \{\epsilon_1, \epsilon_2\}] \\ &\vdots \end{aligned}$$

Let $\left\{B_n^{(1)}|_{n\in\mathbb{N}}\right\}, \left\{B_n^{(2)}|_{n\in\mathbb{N}}\right\}$, and $\left\{B_n|_{n\in\mathbb{N}}\right\}$ be the three corresponding sequences of shrinking balls around the origin, inducing sequences of Poisson measures $\left\{\eta_n^{(1)}|_{n\in\mathbb{N}}\right\}, \left\{\eta_n^{(2)}|_{n\in\mathbb{N}}\right\}$, and $\left\{\eta_n|_{n\in\mathbb{N}}\right\}$ respectively. Then it is an easy **Exercise**to check that, almost surely,

$$\sum_{n=1}^{\infty} \left(\eta_n^{(1)}\right)_{|U} = \sum_{n=1}^{\infty} (\eta_n)_{|U} = \sum_{n=1}^{\infty} \left(\eta_n^{(2)}\right)_{|U}.$$

2 This is clear.
□

4.5.5 Compound Poisson Processes of Bounded Variation

Prerequisites: [5.2] , [4.5.4]

We now have the machinery to define a compound Poisson process of with almost arbitrary intensity measure.

4.5.13

Definition *Compound Poisson Process*

Let V be a Banach space, and let μ be a measure on V having finite first moment. A **compound Poisson process** of **intensity** μ is just the *integral* of a Poisson measure of intensity μ.

That is, let η be a Poisson measure defined on $[0, \infty)$, having intensity measure μ. Then the process $P : \Omega \times [0, \infty) \longrightarrow V$ is defined as follows: for fixed ω in Ω,

$$P_\omega(0) := 0$$

and, for all $t > 0$,

$$P_\omega(t) \ := \ \eta_\omega[0,t].$$

4.5.14

Remark

- If the measure μ is finite, then this definition agrees with the earlier one provided in [4.5.3] .
- Since these processes are defined as the integrals of certain **stochastic measures**, they are of **almost surely bounded variation** by definition.

4.5.15

Theorem *Characteristic Function of a Compound Poisson Process*

Let μ be a measure on $\mathbb{R}^D$ of finite first moment. Let $P : \Omega \times [0,\infty) \longrightarrow \mathbb{R}^D$ be a compound Poisson process, with intensity measure μ. Then, for any $t > 0, P(t)$ is a random variable whose characteristic function Ψ_t is given by:

$$\Psi_t(\vec{\lambda}) \ := \ \exp\left(-t\int_{\mathbf{R}^D}(1-e^{i\langle\vec{v},\vec{\lambda}\rangle})d\mu[\vec{v}]\right)$$

□

4.6 Lévy Processes

4.6.1 Informal Introduction

Let V be a vector space, and suppose α is a stochastic process ranging over V. Imagine α as describing the random jigglings of some particle moving in space.

Suppose that the jiggling of the particle results from a succession of independent *perturbations* occuring over time. We assume that these perturbations all occur independently of one another, and we assume that, as random variables, they are all identically distributed. In other words, we assume that the stochastic process α has two properties:

1. **Independent increments:** The future movements of the particle occur totally independently of its past history.
2. **Stationary increments:** The movement patterns of the particle (ie. the sorts of perturbations it is subjected to) remain unchanging over time.

A stochastic process possessing both of these properties is called a **Lévy process**. Specific examples of Lévy processes include:

- **Brownian Motion** [4.3.1].
- **Stable Processes** [4.4.1].
- **Poisson Processes** [4.5.1].

If α is a Lévy process, then its position at any time can be represented as a sum of an arbitrarily large number of independent, identical perturbations. In other words, for fixed $t \in \mathbb{T}_{time}$, the random variable $\alpha(t)$ can be written as sum of an arbitrarily large number of independent, identically distributed random variables. A random variable of this type is called **infinitely divisible.** Because of this, the theory of Lévy processes is intimately connected with the theory of infinitely divisible random variables.

4.6.2 Infinitely Divisible Random Variables...

We may often want to represent a random variable as the sum of two "smaller" independent random variables. For example, $X(t)$ is some stochastic process in some vector space V, then, for any times $r < t$, the increment $X(0 \to 1) := X(1) - X(0)$ is a random variable taking values in V. However, notice that we also have:

$$X(0 \to 1) \quad = \quad X\left(0 \to \frac{1}{2}\right) + X\left(\frac{1}{2} \to 1\right).$$

If we assume that the process has **independent, stationary increments** , then this sum is a sum of two independent, yet identically distributed random variables.

More generally, if we divided the time interval $(0, 1)$ up into n smaller intervals, then we could express $X(0 \to 1)$ as a sum of n independent, identically distributed random variables. A random variable that behaves this way is called **infinitely divisible**.

Definition *Infinitely Divisible*

Let Y be a random variable taking values in some vector space V. Y is **infinitely divisible** if, for every $N \in \mathbb{N}$, there exists a random variable Y_N, also taking its values in V, so that, if $Y_N^{(1)}, Y_N^{(2)}, \ldots Y_N^{(N)}$, are independent copies of Y_N, then the random variable

$$Y_N^{(1)} + Y_N^{(2)} + \ldots + Y_N^{(N)}$$

has the same distribution as Y.

4.6.2

Remark

The following conditions are equivilant to Y being infinitely divisible:

1. If μ is the density of Y in V, then, for every $N \in \mathbb{N}$, there is a measure μ_N so that

$$\mu = \mu_N * \mu_N * \ldots * \mu_N$$

We say μ is an **infinitely divisible distribution.**

2. If Φ is the characteristic function of Y, then, for every $N \in \mathbb{N}$, we can find a random variable whose characteristic function is $\Phi^{\frac{1}{N}}$. We then say that Φ is an **infinitely divisible characteristic function.**

4.6.3

Remark

1. The set of infinitely divisible random variables has a certain vector-space like property. Fix a sample space $(\Omega, \mathcal{W}, \mathbf{P})$ and a vector space V, and let X_1 and X_2 be infinitely divisible random variables taking their domains in Ω and ranging over V. Suppose X_1 and X_2 are **independent** [19]. Then for any scalars c_1 and $c_2 \in \mathbb{R}$, the random variable $c_1X_1 + c_2X_2$ is also infinitely divisible.

[19] This is important, and is the reason why we can't just say that the class of infinitely divisible distributions *"is"* a vector space.

2. Let $\vec{X} := (X_1, \ldots X_D)$ be a random variable ranging over $\mathbb{R}^D$. Then X is infinitely divisible as a random variable in $\mathbb{R}^D$ if and only if, for all $d \in [1..D]$, the component X_d is infinitely divisible as a random variable ranging over $\mathbb{R}$.

4.6.4

Example

1. **Point Mass:**

   The simplest infinitely divisible distribution is, of course, a **point mass** . Let $\vec{v} \in V$, and let $X_{\vec{v}}$ be a random variable whose value is almost surely $\vec{v}$. Then of course, for any $N \in \mathbb{N}$, if we define $\vec{w} := \frac{1}{N}\vec{v}$, then

   $$X_{\vec{v}} \quad \underset{distr}{\cong} \quad X_{\vec{w}}^{(1)} + X_{\vec{w}}^{(2)} + \ldots + X_{\vec{w}}^{(N)}$$

   where $X_{\vec{w}}^{(1)}, \ldots, X_{\vec{w}}^{(N)}$ are N independent random variables, all having the value $\vec{w}$ almost surely.

2. **Gaussian:**

   Let $\sigma > 0, \vec{v} \in V$, and let X be a random variable ranging over V, so that

   $$\mathbf{D}^{istr}[X] \;=\; \overbrace{\vec{v}; \sigma}$$

   That is, X is a Gaussian random variable, of **mean** $\vec{v}$, and **variance** σ. Thus, for any $N > 0$, we can write:

   $$X \quad \underset{distr}{\cong} \quad X^{(1)} + X^{(2)} + \ldots + X^{(N)}$$

   where $X^{(1)}, \ldots, X^{(N)}$ are N **independent normal random variables**, of **mean** $\frac{\vec{v}}{N}$ and **variance** $\frac{\sigma}{N}$

3. **Symmetric, Stable Random Variables**

   Stable random variables [4.4.1] get their name from the property of **stability** property; the sum of two independent $\alpha-$stable random variables is also an $\alpha-$stable random variable.

   For fixed $\alpha \in (0, 2]$, if X_1 and X_2 are independent, **symmetric** $\alpha-$stable random variables, with variations σ_1 and σ_2, then the sum, $X_1 + X_2$ is *also* $\alpha-$stable, with variance $(\sigma_1^\alpha + \sigma_2^\alpha)^{1/\alpha}$.

4. **Skewed Stable Random Variables**

   Skewed random variables [4.4.1] also possess the **stability** property, although the formula expressing the **skewness** and **bias** of the sum of two α-stable variables is more complicated.

5. **Simple Poisson Distributions**

   Let $c > 0$. A **Simple Poisson random variable** of parameter c is a random variable N_c, ranging over $\mathbb{N}$, whose density given by the formula:

$$\forall n \in \mathbb{N}, \quad \mathbf{P}_{rob}\left[N_c = n\right] \; = \; \frac{(ct)^n e^{-ct}}{n!}.$$

   The standard interpretation is that N_c represents the random value, at time 1, of a **Simple Poisson Process** of **intensity** c. This is a process ranging over the natural numbers, which intermittently increases by increments of $+1$; the time durations between increments are independent, **exponentially distributed** random variables of parameter c. For more about Poisson processes, see [4.5.1] .

Nonexamples:

1. **Two Point Masses**: Let x and y be distinct points in a vector space V, and let X be a random variable which has probability $\frac{1}{2}$ of assuming either value. Then X is not infinitely divisible.

2. **Uniform Distribution**: Let X be a random variable on V whose probability is **uniformely distributed** over some compact subset $U \subset V$. Then X is not infinitely divisible.

Exercise: Verify these assertions.

For more information about infinitely divisible random variables, see [5] , chap. XVII. Further information, in connection with Lévy distributions, can be found in [2] .

4.6.3 ...And Lévy processes

Prerequisites: [4.6.2],[4.1.6]

4.6.5

Definition L

Let V be a vector space. A **Lévy process** on V is a continuous time stochastic process on the timeline $[0, \infty)$, ranging over V and originating at the origin, with *stationary, independent increments* .

4.6.6

Remark

1. By definition, Lévy processes originate at zero. Thus, if P is a Lévy process, then, for any $T > t > 0$,

$$P(t \to T) \quad \underset{distr}{\cong} \quad P(0 \to (T-t)) \quad \underset{distr}{\cong} \quad P(T-t).$$

2. Let V be a fixed vector space, $\mathbb{T}_{ime}$ a fixed time-line, and $(\Omega, \mathcal{W}, \mathbf{P})$ a fixed sample space. Then the class of Lévy processes from $\mathbb{T}_{ime}$ into V, varying over Ω, acts, in a certain sense, like a vector space:

$$P_1 : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

and

$$P_2 : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

are two *independent*[20] Lévy processes, then for any real scalars r_1 and r_2, the process

$$r_1 P_1 + r_2 P_2 : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

is also a Lévy process. (The reason we can't just say that the class of Lévy processes *is* a vector space is because we are here requiring *independence* of the two processes) .

3. Let $\vec{X}(t) := (X_1(t), \ldots, X_D(t))$ be a stochastic process on $\mathbb{R}^D$. Then $\vec{X}$ is a Lévy process on $\mathbb{R}^D$ if and only if, for all $d \in [1..D]$, the process $X_d(t)$ is a Lévy process on on $\mathbb{R}$.

[20] This is important.

4. A Lévy process $X(t)$ is entirely characterised by the distribution of $X(1)$, in the sense that, if X_1 and X_2 are two Levy processes such that

$$X_1(1) \quad \underset{distr}{\cong} \quad X_2(1)$$

then, for all $T > t \geq 0$, we have:

$$X_1(t \to T) \quad \underset{distr}{\cong} \quad X_2(t \to T)$$

so that X_2 is just a **version** of X_1.

4.6.7

Example

1. **Deterministic Drift** :

The simplest Lévy process isn't even stochastic; it is just **deterministic drift** . V is a vector space, and we fix a vector $\vec{\delta} \in V$, then deterministic drift in the direction $\vec{\delta}$ is just the process

$$D : \Omega \times [0, \infty) \longrightarrow V$$

so that, for all $\omega, D(\omega, t) := t \cdot \vec{\delta}$.

2. **Brownian Motion:** [4.3.1]

Brownian motion on $\mathbb{R}^D$ is a Lévy process

$$B : \Omega \times [0, \infty) \longrightarrow \mathbb{R}^D$$

where, for any fixed times $T > t > 0$, the increment $B(t \to T)$ is a Gaussian random variable, of variance $(T - t)$, and independent of all previous increments.

3. **Symmetric $\alpha-$stable Motion:** [4.4.1]

Fix $\alpha \in (0, 2]$, and let $\sigma > 0$. Then **symmetric $\alpha-$stable motion** of **growth rate** σ on $\mathbb{R}^D$ is a Lévy process

$$S : \Omega \times [0, \infty) \longrightarrow \mathbb{R}^D$$

where, for any fixed times $T > t > 0$, the increment $S(t \to T)$ is an $\alpha-$**stable random variable**, of **variation** $(\frac{T-t}{\sigma})^{1/\alpha}$. That is, $S(t \to T)$ is a random variable ranging over $\mathbb{R}^D$, with **characteristic function** $\Phi : \mathbb{R}^D \longrightarrow \mathbb{C}$ given by:

$$\Phi(\vec{\lambda}) \ := \ \exp(-\sigma \left|\vec{\lambda}\right|^{\alpha}),$$

Brownian motion is of course the special case when $\alpha := 2$ and $\sigma := 1$.

4. **Skewed $\alpha-$stable Motion:** [4.4.1]

   Fix $\alpha \in (0, 2]$, and let $\sigma > 0$ as before. Now, however, we introduce two additional parameters, δ and τ, which we will call the **drift** and the **skewness** of the motion, respectively.

   Then **α-stable motion** on $\mathbb{R}^D$ of **growth rate** σ, **skewness** τ, and **drift** δ, is a Lévy process

   $$S^{\alpha}_{\sigma,\delta,\tau} : \Omega \times [0, \infty) \longrightarrow \mathbb{R}^D$$

   where, for any fixed times $T > t > 0$, the increment $S^{\alpha}_{\sigma,\delta,\tau}(t \to T)$ is an $\alpha-$stable random variable, of **variation** $(\frac{T-t}{\sigma})^{1/\alpha}$, **bias** $(T - t) \cdot \delta$, and **skewness** τ.

5. **Poisson Processes**: [4.5.1]

   Let μ be a measure on $\mathbb{R}^D$ with **finite expected norm** . Let P be the **compound Poisson Process** [4.5.5] with intensity measure μ. Then P is a Lévy process

   $$P : \Omega \times [0, \infty) \longrightarrow \mathbb{R}^D$$

   where, for any fixed times $T > t > 0$, the increment $P(t \to T)$ is a random variable ranging over V, with characteristic function $\Psi : \mathbb{R}^D \longrightarrow \mathbb{C}$

   $$\Psi(\vec{\lambda}) \ := \ \exp\left(-|T - t| \int_{\mathbf{R}^D} (1 - e^{i\langle \vec{v}, \vec{\lambda} \rangle}) d\mu[\vec{v}]\right)$$

   (for more details, see

The astute reader will notice a close correspondence between the examples of Lévy processes just listed, and the examples of **infinitely divisible** random variables given in [4.6.2] . We now make this correspondence precise.

4.6.8

Theorem

Let V be a vector space.

1. Suppose $P : \Omega \times [0, \infty) \longrightarrow V$ is a **Lévy process.** Then:

   (a) For any $T > t > 0$, the random increment $P(t \to T)$ is **infinitely divisible.**

   (b) Thus, for any $t > 0$, the random variable $P(t)$ is **infinitely divisible.**

2. Let X be an **infinitely divisible random variable** ranging over V. Then

   (a) There exists a **Lévy process** $P : \Omega \times [0, \infty) \longrightarrow V$, having almost surely **right-continuous** sample paths, so that

   $$X \quad \underset{distr}{\cong} \quad P(0 \to 1) \quad (1)$$

   (b) This process is **unique**: any two Lévy processes satisfying (1) are **versions** of one another.

Proof: 1 These facts follow immediately from the definition of Lévy processes, and from the fact that the real line itself is "infinitely divisible".

To be concrete, consider the increment $P(0 \to 1)$. Then clearly, for any $N > 0$,

$$P(0 \to 1) \;=\; P\left(0 \to \frac{1}{N}\right) + P\left(\frac{1}{N} \to \frac{2}{N}\right) + P\left(\frac{2}{N} \to \frac{3}{N}\right) + \ldots + P\left(\frac{N-1}{N} \to 1\right),$$

a sum of N independent, identically distributed random variables.

2 (Sketch)

Since X is **infinitely divisible**, for any $N \in \mathbb{N}$, we can find a random variable X_N so that, if $X_N^{(1)}, X_N^{(2)}, \ldots, X_N^{(2^N)}$ are 2^N independent copies of this variable, then

$$X \quad \underset{distr}{\cong} \quad X_N^{(1)} + X_N^{(2)} + \ldots + X_N^{(2^N)} \quad (1)$$

if $T := \cup_{N=1}^{\infty} T_N$, then of course T is just the set of all **dyadic rational numbers** in the unit interval, and T is dense in $[0, 1]$.

We want to make use of Theorem [4.2.2]. To do this, we need to build a **continuous monoid homomorphism** from $\mathbb{Q}_2^+$ into $\mathcal{M}$, where $\mathbb{Q}_2^+$ is the monoid of **nonegative dyadic rational numbers** under **addition**, and $\mathcal{M}$ is the monoid of **Radon measures on** V under **convolution**.

For all $N \in \mathbb{N}$, define

$$\mu_{\frac{1}{2^N}} \ := \ \mathbf{D}^{istr}\left[X_N\right]$$

and then for any time $t \in \mathbb{Q}_2^+$, express t as

$$t \ = \ \frac{m_1}{2} + \frac{m_2}{4} + \frac{m_3}{8} + \ldots + \frac{m_N}{2^N} \qquad (2.1)$$

and then define

$$\mu_t \ := \ \left(\underbrace{\mu_{\frac{1}{2}} * \ldots * \mu_{\frac{1}{2}}}_{m_1}\right) * \left(\underbrace{\mu_{\frac{1}{4}} * \ldots * \mu_{\frac{1}{4}}}_{m_2}\right) * \left(\underbrace{\mu_{\frac{1}{8}} * \ldots * \mu_{\frac{1}{8}}}_{m_3}\right) * \ldots$$
$$\ldots \ * \left(\underbrace{\mu_{\frac{1}{2^N}} * \ldots * \mu_{\frac{1}{2^N}}}_{m_N}\right) \qquad (2.2)$$

Claim 1:

1. The measure μ_t is well-defined by expression (2.2), independent of the decomposition (2.1) we chose.
2. The resulting map

$$\mathbb{Q}_2^+ \ni t \mapsto \mu_t \in \mathcal{M}$$

is a **monoid homomorphism**.

Proof: : Both facts follow from equation (1). The details are left as an **exercise**.

□1

Claim 2: The map $\mathbb{Q}_2^+ \ni t \mapsto \mu_t \in \mathcal{M}$ is **continuous** with respect to

the **weak topology** on $\mathcal{M}$.

Proof: [21] This is equivilant to showing:

$$wk - \lim_{t \to 0} \mu_t = \delta_0$$

.

Suppose that this was *not* the case. Thus, we could find some compact set $K \subset V$, bounded away from zero, and a $\delta > 0$, so that

$$\limsup_{t \to 0} \mu_t[K] \ > \ \delta.$$

In other words, for every $n \in \mathbb{N}$, we can find some time $t_n \in \mathbb{Q}_2^+$, with $t_n < \frac{1}{n}$, so that

$$\mu_{t_n}[K] \ > \ \delta.$$

Now, find a **linear basis** $\mathcal{B} := \{\vec{b}_1, \ \ldots, \ \vec{b}_D\}$ for V, such that

$$K \ \subset \ K_1 \ := \ \left\{ \vec{v} \in V \ ; \ \vec{v} := v_1 \vec{b}_1 + \ldots + v_D \vec{b}_D, \ and \ v_1 > \epsilon \right\}$$

for some $\epsilon > 0$. Thus,

$$\mu_{t_n}[K_1] \ > \ \delta.$$

Thus, if we define

$$\begin{aligned} K_n \ &:= \ \underbrace{K_1 + K_1 + \ldots + K_1}_{n} \\ &= \ \left\{ \vec{v} \in V \ ; \ \vec{v} := v_1 \vec{b}_1 + \ldots + v_D \vec{b}_D, \ and \ v_1 > n \cdot \epsilon \right\} \end{aligned}$$

Then

$$\mu_{n \cdot t_n}[K_n] \ > \ \delta^n.$$

(**Exercise**: Check this).
But by construction, $n \cdot t_n \leq 1$. Thus,

$$\mu_1[K_n] \ > \ \delta^n.$$

That is, μ_1 is a measure whose support around zero decreases no slower than at an inverse-exponential rate. Check that this is incompatible with

[21] This is the sketchy part.

the type of characteristic function μ_1 must have as an **infinitely divisible** distribution, according to the Lévy -Khintchine theorem.

□2

Thus, using Theorem [4.2.2] , we can build a stochastic process P from $[0, \infty)$ into V possessing the desired increments, and originating at zero. By construction, P has independent, stationary increments; thus, P is a Lévy process.

□

4.6.4 The Lévy -Khintchine Formula....

Prerequisites: [4.6.2] **Corequisites:** [4.5.5]

Unfortunately, no closed form expression is known for the **density function** of a general infinitely-divisible random variable. However, there *is* a general form for the **characteristic function** of such variable; this is the famous Lévy -Khintchine formula.

The Lévy -Khintchine formula says that an infinitely divisible random variable can be thought of as the independent sum of a **random variable** (possibly with a rather warped **covariance matrix**) , along with a **compound Poisson random variable** (possible with infinite **intensity measure**) . Thus, an infinitely divisible *distribution* will look like the convolution of a **normal distribution** with an infinite-intensity **Poisson distribution**. Thus, the characteristic function of an infinitely divisible random variable will be a product of two terms: A "Gaussian" term, and a "Poisson" term.

4.6.9

Theorem L

Let $\Phi : \mathbf{R}^D \longrightarrow \mathbb{C}$. Then Φ is the characteristic function of an **infinitely divisible distribution** if, and only if, Φ is of the form:

$$\Phi(\vec{\lambda}) \ := \ \exp(\Psi(\vec{\lambda}))$$

where Ψ is given by:

$$\Psi(\vec{\lambda}) \;:=\; i\left\langle \vec{a},\vec{\lambda}\right\rangle \;+\; \frac{1}{2}Q(\lambda) \;+\; \int_{\mathbf{R}^D}\left(1-e^{\langle \vec{\lambda},\vec{x}\rangle}\right)d\mu[\vec{x}]. \quad (LK1)$$

Here,

- $\vec{a}\in\mathbb{R}^D$,
- Q is a **positive semi-definite quadratic form** on $\mathbf{R}^D$,
- μ is a nonegative measure on $\mathbb{R}^D$, so that

  1. μ decays faster than $\|x\|^2$ around zero; in other words:

     $$\int_{\mathbf{B}^D(0;1)}\|x\|^2\, x d\mu[x]$$

     is finite.

  2. μ allocates *finite mass* away from zero; in other words, $\mu\left[\mathbb{R}^D\setminus\mathbb{B}^D(0;1)\right]$ is finite.

Proof:
A proof of the one-dimensional case can be found in [5] , Chap. XVII (see the Remarks below) . The general case follows by combining one-dimensional infinitely divisible variables living in different dimensions.
□

4.6.10

Remark

1. The characteristic function represents the independent sum of a Gaussian random variable, having **mean** $\vec{a}$ and **covariance matrix** Q, along with a compound Poisson random variable with **intensity measure** μ.

2. Different sources express the Lévy -Khintchine formula in slightly different ways. For example, in [2] , the formula for Ψ is given as:

$$\Psi(\vec{\lambda}) \;:=\; i\left\langle \vec{a}_0,\vec{\lambda}\right\rangle + \frac{1}{2}Q(\lambda) + \int_{\mathbf{R}^D}\left(1-e^{\langle \vec{\lambda},\vec{x}\rangle} i\left\langle \vec{\lambda},\vec{x}\right\rangle \mathbb{1}_{\mathbf{B}^D(0;1)}\right)d\mu[\vec{x}]. \quad (LK2)$$

(where $\vec{a}_0, Q$, and μ have the same properties as in (LK1)) .

Clearly, the formula (LK2) can be made equivilant to (LK1) be simply defining

$$\vec{a}_0 \quad := \quad \vec{a} \; + \; \int_{\mathbf{B}^D(0;1)} \left\langle \vec{\lambda}, \vec{x} \right\rangle d\mu[\vec{x}].$$

3. On the other hand, in [5] , the (one-dimensional) Lévy -Khintchine formula is given with

$$\Psi(\lambda) \quad := \quad ia\lambda \; + \; \int_{\mathbf{R}} \frac{1 - e^{\lambda x} - i\lambda \sin(x)}{x^2} d\mu_0[\times]. \qquad (LK3)$$

Here, μ_0 is clearly a different measure from μ, since the factor $\frac{1}{x^2}$ is already included in its definition. The sign change on the term $1-e^{\lambda x}$ may seem puzzling at first, but this can be accounted for by the fact that we are subtracting off a factor of $-i\lambda \sin(x)$, and of course $\sin(x) = \frac{e^{ix} - e^{-ix}}{2i}$; thus, for suitably chosen μ_0, the two integrands will have equal integrals near the origin. Of course, the measure μ_0 must also differ far away from the origin, to account for the oscillations introduced by the $-i\lambda \sin(x)$ term.

4.6.5 ....and the Canonical Representation of a Lévy process

Prerequisites: [4.6.3] , [4.6.4] , [4.5.5]

The Lévy -Khintchine formula [4.6.4] provides a canonical decomposition of any infinitely divisible distribution into "Gaussian" and "Poisson" components. Meanwhile, in [4.6.3] , we observed the close correspondence between Lévy processes and infinitely divisible distributions. Thus, we might expect that the Lévy -Khintchine formula can be used to provide a corresponding decomposition of the general Lévy process.

4.6.11

Theorem

Let V be a finite-dimensional vector space, and let $\mathcal{L} : [0, \infty) \longrightarrow V$ be some Lévy process. Then $\mathcal{L}$ has a unique decomposition of the form:

$$\mathcal{L} \;=\; \alpha + \beta + \gamma$$

where

1. α is a deterministic **drift** term;
2. β is **Brownian motion** with respect to some choice of linear basis on V;
3. γ is a **compound Poisson process**.

In other words, we can find:

1. A vector $\vec{a} \in V$;
2. A basis $\mathcal{B} := \{\vec{v}_1, \ldots \vec{v}_D\}$ of V;
3. A measure μ on V so that:

$$\int_{\mathbb{B}^D(0;1)} \|x\|^2 \, d\mu[x] \qquad \text{and} \qquad \mu\left[\mathbb{R}^D \setminus \mathbb{B}^D(0;1)\right]$$

are finite.

so that, if we define:

1. $\alpha(t) := t \cdot \vec{a}$;
2. $\beta(t) := B_1(t) \cdot \vec{v}_1 \;+\; B_2(t) \cdot \vec{v}_2 \;+\; \ldots \;+\; B_D(t) \cdot \vec{v}_D$, where $B_1, \ldots B_D$ are D independent one-dimensional **Brownian motions**;
3. γ a **compound Poisson process** on V with **intensity measure** μ, independent of $B_1, \ldots, B_D$;

then we can write, for all $t \in [0, \infty)$,

$$\mathcal{L}(t) \quad = \quad \alpha(t) + \beta(t) + \gamma(t) \qquad (2)$$

Proof:
Assume WOLOG that $V = \mathbb{R}^D$. A Lévy process is entirely determined by its distribution at time 1 (see the remarks in [4.6.3]) , so it is sufficient to consider the random variable $\mathcal{L}(1)$. This random variable is infinitely

divisible; thus, by the Lévy -Khintchine formula, its characteristic function is of the form

$$\Phi(\vec{\lambda}) \ := \ \exp(\Psi(\vec{\lambda}))$$

where Ψ is given by:

$$\Psi(\vec{\lambda}) \ := \ i\left\langle \vec{a}, \vec{\lambda} \right\rangle \ + \ \frac{1}{2}Q(\lambda) \ + \ \int_{\mathbf{R}^D} \left(1 - e^{\langle \vec{\lambda}, \vec{x} \rangle}\right) d\mu[\vec{x}]. \quad (2)$$

Here,

- $\vec{a} \in \mathbb{R}^D$,
- Q is a positive semi-definite quadratic form on $\mathbf{R}^D$,
- μ is a nonegative measure on $\mathbb{R}^D$ so that

$$\int_{\mathbf{B}^D(0;1)} \|x\|^2 \, d\mu[x] \qquad \text{and} \qquad \mu\left[\mathbb{R}^D \setminus \mathbb{B}^D(0;1)\right]$$

are finite.

Now, let $\sqrt{Q}$ be any matrix[22] with the property that $\sqrt{Q}\sqrt{Q} = Q$. Then we can think of $\sqrt{Q}$ as being the change-of-basis matrix from the canonical basis $\{\vec{e}_1, \ldots, \vec{e}_D\}$ of $\mathbb{R}^D$ into some other basis, $\{\vec{v}_1, \ldots, \vec{v}_D\}$.

Let $\vec{X}$ be a **normal random variable** on $\mathbf{R}^D$, with **mean** 0 and **variance** 1. Then $\sqrt{Q}(\vec{X})$ is a normal random variable on $\mathbb{R}^D$ with **mean** 0 and **covariance matrix** $\sqrt{Q}$. In other words, we can write

$$\sqrt{Q}(\vec{X}) \ = \ X_1\vec{v}_1 + X_2\vec{v}_2 + \ldots X_D\vec{v}_D$$

where $X_1, \ldots, X_D$ are D independent normal random variables in $\mathbb{R}$, all having means zero and variances equal to 1.

At the same time, let γ be a **Poisson process** having **intensity measure** μ. Define $\alpha(t) := t \cdot \vec{a}$, and let $\beta(t) := B_1(t)\vec{v}_1 + B_2(t)\vec{v}_2 + \ldots B_D(t)\vec{v}_D$, where $B_1(t)$, $B_2(t), \ldots,$ $B_D(t)$ are D independent Brownian motions. It is clear by construction that the random variable $\alpha(1) + \beta(1) + \gamma(1)$ has a characteristic function identical to (2). Thus,

$$\mathbf{D}^{istr}\left[\mathcal{L}(1)\right] \ = \ \mathbf{D}^{istr}\left[\alpha(1) + \beta(1) + \gamma(1)\right].$$

[22] A matrix of this form can always be found, because Q is **positive definite**. See, for example, [1].

which is sufficient to show that, as a stochastic process, $\alpha + \beta + \gamma$ is a **version** of $\mathcal{L}$.

□

4.6.12

Remark

Using the canonical representation, it is possible to immediately "read off" certain properties of the Lévy process $\mathcal{L}$. For example:

1. $\mathcal{L}$ is a process of **almost surely bounded variation** *if and only if*

   (a) $Q = 0$. That is, the **Gaussian** component is trivial.

   (b) μ is a measure of finite expected norm. That is, the **Poisson** component has **bounded variation**. (see [4.5.5])

2. $\mathcal{L}$ is a **martingale** *if and only if*

$$\vec{a} = \int_{\mathbf{R}^D} \vec{x} d\mu[\vec{x}]$$

   —in other words, the "drift component" cancels out the "expected growth rate" of the Poisson component, so that the expected value of increments is always zero.

3. $\mathcal{L}$ is a process with $\mathbf{L}^2$ increments *if and only if*

$$\int_{\mathbf{R}^D} \|\vec{x}\|^2 d\mu[\vec{x}]$$

   is finite —in other words, if and only if the **Poisson component** has $\mathbf{L}^2$ increments.

Chapter 5

Stochastic Integration

5.1 Introduction

Modulated Noise

Stochastic integration is of central importance in stochastic modelling with applications to mathematical physics, mathematical finance, and mathematical biology. Stochastic integration provides a method of constructing and studying a very broad class of stochastic processes. In particular, it allows us to define stochastic processes while look like deterministic dynamical systems "perturbed" by some sort of random noise. The stochastic integral allows us to *shape* what this random noise is going to look like, by taking a well-understood *canonical* noise process (like Brownian motion), and *modulating* it to emulate the stochasticity of some real system.

As we have seen in [4.1.6], a stochastic process on a vector space is entirely characterised by its **increments**. These increments can be thought of as providing a means of representing the process via a sequence of successive random *impulses* occuring over time. These impulses cause the sample path to *jiggle* about randomly, and this *jiggling* is the source of its stochasticity.

Suppose we then manipulated the process by making some of these impulses larger in magnitude, and others smaller. Thus, at some point, the random jiggling of the sample path will be very *large* (ie. there will be "large" stochasticity"), while at other times, the jiggling will be *small* (the path will behave "almost" deterministically). It is this sort of manipulation we intend to achieve via stochastic integration.

Suppose that we start with some stochastic process, $X(t)$ which we imagine is the result of an infinite succession of infinitesmal impulses. We might denote the impulse at time t by $dX(t)$. If $X(t)$ represents the *position*

of some particle at time t, then $dX(t)$ would be its *velocity* at this time. Thus, formally, we might write, for any time $T > 0$,

$$X(T) = \int_0^T dX(t)\, dt.$$

We might then take a second stochastic process, $f(t)$, and use it as the factor by which we magnify the "impulses" of X. In other words, we might define the "stochastic integral" of f with respect to X to be the stochastic process $Y(t)$ defined by the formal integration:

$$Y(T) := \int_0^T f(t) dX(t)\, dt. \quad (1)$$

Deterministic Context: Functions of Bounded Variation and the Lebesgue-Stieltjes Integral

Let's first examine this idea in a purely *deterministic* context. Forget about f and X being stochastic processes, and imagine that they are just deterministic functions:

$$X : [0, \infty) \longrightarrow \mathbb{R}^D$$

and

$$f : [0, \infty) \longrightarrow \mathbb{R}$$

What might the equation (1) mean? Well, suppose we were to approximate this equation by a finite sum. First, pick a **partition** of the form:

$$\{0 =: t_0 < t_1 < t_2 < \ldots < t_N < t_{N+1} := T\}$$

and then use the approximation:

$$\begin{aligned} Y(1) &= \int_0^1 f(t) dX(t)\, dt \\ &\sim \sum_{n=0}^{N} f(t_n) \cdot [X(t_{n+1}) - X(t_n)] \end{aligned}$$

Then take the limit as the partition becomes **infinitely fine**[1]. If the function X is of **bounded variation**, and the function f satisfies certain

[1] For example, define $\left\{ t_n^{(N)} \; ; \; n \in [0..(N+1)] \right\}$ by $t_n^{(N)} := \frac{n}{N}$, and then let $N \to \infty$.

mild integrability conditions, then these sums will converge; their limit is called the **Lebesgue-Stieltjes integral** of f with respect to X.

Indeed, if the functions f and X are **differentiable**, then we can apply classical integral calculus, and compute the integral via **Integration by Parts**:

$$\begin{aligned} Y(T) &= \int_0^T f(t)dX(t)\,dt \\ &= f(T)X(T) - f(0)X(0) + \int_0^T \frac{df}{dt}(t)X(t)\,dt. \end{aligned}$$

The Lebesgue-Stieltjes integral can also be defined by:

$$\int_0^1 f(t)dX(t)\,dt = \int_{[0,1]} f d\mu_X$$

where μ_T is a Borel measure on $[0,\infty)$, called the **Lebesgue-Stieltjes measure**[2] associated with the function X. μ_X is defined by the property that, for any **half-open** interval $(a,b] \subset [0,\infty)$,

$$\mu(a,b] := X(b) - X(a)$$

The measure μ_X is **absolutely continuous** with respect to the Lebesgue measure if and only if X is **differentiable** as a function $X : [0,\infty) \longrightarrow \mathbb{R}^D$. In this case, we have:

$$\frac{d\mu_X}{d\mathcal{L}}(t) = \frac{dX}{dt}(t)$$

(**Exercise**: Verify this.)

Even if X is *not* differentiable, however, it is clear that μ_X provides complete information for measuring the "rate of change" of X. Thus, the measure μ_X can be thought of as a generalisation of the classical notion of **derivative**.

Details on this construction can be found in [6] sections 1.5 and 3.5.

[2] The use of the term "measure" to describe μ_X may seem a bit peculiar, since it takes its values over the vector space $\mathbb{R}^D$, but it has all the normal countable additivity properties of any other measure. Just think of μ_X as being a $D-$tuple of real-valued measures: $\mu_X := (\mu_1, \mu_2, \ldots, \mu_D)$. For fixed $d \in [1..D]$, the measure μ_d is a real-valued, **signed measure**. μ_d can then be further decomposed as a sum: $\mu_d := \mu_d^+ - \mu_d^-$, where μ_d^+ and μ_d^- are **positively valued** measures, having **disjoint support** on $[0,\infty)$.

Stochastic Integration, pathwise or otherwise

Motivated by this approach, we might attempt to define the stochastic integral of one stochastic process with respect to another *pathwise*, by simply applying the definition of the **Lebesgue-Stieltjes integral** to *every* sample path, *seperately.*

If $X : \Omega \times [0, \infty) \longrightarrow \mathbb{R}^D$ and $f : \Omega \times [0, \infty) \longrightarrow \mathbb{R}$ are two stochastic processes, so that X has sample paths of almost-surely bounded variation, and f satisfies some weak boundedness conditions to guarantee that the Lebesgue-Stieltjes integral converges, we could define the process

$$Y(T) \quad := \quad \int_0^T f(t)\, dX[t]$$

seperately for each $\omega \in \Omega$, by the expression:

$$Y(\omega, T) \quad := \quad \int_0^T f(\omega, t)\, d\mu_{X_\omega}[t]$$

where μ_{X_ω} is the **Lebesgue-Stieltjes measure** associated to the path X_ω.

A more detailed exposition of pathwise integration is given in [5.3].

The problem is, of course, that this method only works if $X(t)$ is almost surely of **bounded variation**. However, some of the most interesting and natural processes, such as **Brownian motion**, are *not* of bounded variation. What then?

This problem was originally solved by Itô , who reasoned as follows. Abstractly speaking, the integration operation:

$$Y(T) \quad := \quad \int_0^T f(t) dX[t]$$

is a sort of *operator*, which takes a **stochastic process** f as *input*, and produces a **random variable** $Y(T)$ as *output.* f is thus a **measurable function**

$$f : \Omega \times [0, T] \longrightarrow \mathbb{R}^D,$$

and $Y(T)$ is a **measurable function**:

$$Y(T) : \Omega \longrightarrow \mathbb{R}^D.$$

So, treat f as living in some space of measurable functions $\mathbf{L}(\Omega \times [0,T];\, \mathbb{R}^D)$, and X as living in a space $\mathbf{L}(\Omega;\, \mathbb{R}^D)$, and treat the stochastic integral as an **operator**, which we will denote by $\int_0^T [\bullet]\, dX[t]$, then what we are looking at here is really an operator

$$\int_0^T [\bullet]\, dX[t] \,:\, \mathbf{L}(\Omega \times [0,T];\, \mathbb{R}^D) \longrightarrow \mathbf{L}(\Omega;\, \mathbb{R}^D)$$

Furthermore, since it represents some kind of "integration", this operator should be *linear* —as is the pathwise Lebesgue-Stieltjes integral.

Suppose that the operator $\int_0^T [\bullet]\, dX[t]$ preserved some suitably defined **norm** as it mapped between these two spaces. Then it would be a **bounded linear operator**, and it would suffice to define it only on a conveniently chosen **dense subset** of functions in $\mathbf{L}(\Omega \times [0,T];\, \mathbb{R}^D)$.

Recall that, in "classical" integration theory, we define the integral of a function

$$\int f d\mu$$

by *approximating* f with **simple** functions; that is, by functions which look like **finite linear combinations** of **characteristic functions**. If $\sum_{k=1}^K c_k \mathbb{1}_{A_k}$ is a simple function, then we define:

$$\int \left(\sum_{k=1}^K c_k \mathbb{1}_{A_k} \right) d\mu \;\; := \;\; \sum_{k=1}^K c_k \mu[A_k].$$

Next, if $\left\{ f_n|_{n \in \mathbb{N}} \right\}$ is a *sequence* of simple functions converging to f, then we define

$$\int f \, d\mu \;\; := \;\; \lim_{n \to \infty} \int f_n \, d\mu,$$

and a few technical exercises suffice to show that this integral is well-defined, independent of the approximating sequence chosen. The key idea is that, if two simple functions are "close" together, then their integrals are also close together.

In a similar fashion, the Itô integral is defined by approximating an arbitrary stochastic process f with **simple stochastic processes**. These are processes which are "peicewise constant", in much the same way that a simple function is peicewise constant. We would then like to define

$$\int_0^T f dX \;:=\; \lim_{n\to\infty} \int_0^T f_n dX$$

The problem is that now, instead of this limit referring to the convergence of *real numbers*, it refers to the convergence of a sequence of *random variables* –a much more subtle idea. In what sense does this convergence take place?

It turns out that we can define a certain measure $\nu_{[X]}$ on the space $\Omega \times [0,T]$, called the **Doléans measure**. $\nu_{[X]}$ is defined in terms of the *growth rate* of the process X. We can then look at the $\mathbf{L}^2$-space, with respect to $\nu_{[X]}$, of **simple stochastic processes**. It turns out that, under suitable conditions, the integration operator is actually an *isometry* ; if $\mathcal{S}$ is a simple stochastic process, then

$$\left\| \int_0^T [\mathcal{S}]\, dX[t] \right\|_2 \;=\; \|\mathcal{S}\|_2$$

where the norm on the *right hand side* is the $\mathbf{L}^2$ norm of $\mathbf{L}^2(\Omega \times [0,T];\, \mathbf{R}^D)$, while that on the *left hand side* is just the $\mathbf{L}^2$ norm of $\mathbf{L}^2(\Omega;\, \mathbf{R}^D)$ (ie. the **variance** of $\int_0^T [\mathcal{S}]\, dX[t]$ as a random variable).

Thus, we can use the density of simple stochastic processes to *extend* the definition of the integral to an isometry

$$\int_0^T [\bullet]\, dX[t] \;: \mathbf{L}^2_N(\Omega \times [0,T];\, \mathbb{R}^D) \longrightarrow \mathbf{L}(\Omega;\, \mathbf{R}^D)$$

Here, the subscript "N " in "$\mathbf{L}^2_N$ " refers to **nonanticipating**; a technical property of our integrands which we will discuss more in the next section.

In the case when X is Brownian Motion, the measure $\mu_{[X]}$ turns out to be nothing more than the **product measure**

$$\mu_{[X]} \;=\; \mathbf{P} \times \mathcal{L}$$

where $\mathbf{P}$ is the probability measure on Ω, and $\mathcal{L}$ is the **Lebesgue measure** on $[0,T]$.

For a general process X, given any half-open interval $[a,b)$, and any set $U \subset \Omega$, measurable with respect to the variable $X(a)$, the measure of the rectangle $U \times [a,b)$ is the **conditional variance** (on U) of the **increment** of X over the time interval $[a,b)$, weighted by the probability of μ:

$$\begin{aligned}\mu_{[X]}[U \times [a,b)] &:= \mathbf{E}^{xpct}\left[\mathbb{1}_U \cdot |X(a \to b)|^2\right] \\ &= \mathbf{E}^{xpct}\left[\frac{|X(a \to b)|^2}{X(a) \in U}\right] \cdot \mathbf{P}_{rob}[U]\end{aligned}$$

Itô 's method will work for a broad class of stochastic processes; in particular, it will work whenever X is a **martingale** with $\mathbf{L}^2$ increments. The best exposition of the Itô integral, in the context of $\mathbf{L}^2$-bounded martingales, is [3].

Nonanticipiting Integrands and Martingale Properties

We earlier motivated the construction of the **stochastic integral**

$$\int f\, dX$$

as a way of generating *modulated noise*; we take the stochastic process X (our "noise source"), and *attenuate* it by the process f to get the kind of noise we want.

If this noise is to be truly *random*, however, then we don't want the function f to be able to systematically "bias" the movement of the noise towards some directions and away from others. For example, suppose the function f was constructed using information about the future movements of X. Then we could build f so as to be "large" whenever X moved "up", and "small" whenever X moved "down". In this way, we could use the modulating function f to systematically *skew* the noise to favour movement in some directions and away from others. Ultimately, by carefully building f, we might even be able to make the integral process $\int f\, dX$ virtually *deterministic* in its movements. This would clearly defeat our entire intent.

To eliminate this possibility, we must demand that the process f be **nonanticipating**, in the sense that, at any given time t, the value of $f(t)$ can only depend on by information about the *past* movements of X, and must be totally *ignorant* of future movements.

There are many different ways to formally define **nonanticipating**, which lead to different technical developments, but all produce basically the same conclusion. The important idea is that f depends only on the *past* of X, and is independent of its *future* movements.

If f is **nonanticipating**, and the process X is a **martingale**, then the integral process $\int f\, dX$ is *also* martingale. This is one of the most important properties of the Itô integral.

(**Exercise**: Construct an intuitive argument why $\int f\,dX$ should be a martingale if f is nonanticipating.)

Stochastic Calculus

Stochastic integrals form the basis for defining an entire *calculus* of stochastic processes. The **Fundamental Theorem of Differential calculus** tells us that, if the function F is defined by:

$$F(T) \;:=\; \int_0^T f(t)\,dt,$$

then, for any T,

$$\frac{dF}{dt}(T) \;=\; f(T).$$

In a similar fashion, if a stochastic process F is *defined* via the stochastic integral

$$F(T) \;=\; \int_0^T f(t)dX[t] \quad (1)$$

then we might formally say that fdX is the "derivative" of F, and write:

$$dF = fdX \quad (2)$$

It turns out that this isn't just a cute abuse of notation; this notation actually turns out to be quite appropriate, because we can derive analogs to all of the usual rules of differentiation, such as the **Liebniz Rule** and the **Chain Rule**.

Ultimately, it becomes more convenient and more powerful to describe stochastic processes in the form of equation (2), rather than equation (1). Indeed, we can even formulate **stochastic differential equations** in this way, and then solve them, to construct *new* processes with subtle properties.

5.2 Random Measures as the Derivatives of Stochastic Processes

Prerequisites: [4.2.2]

We have seen in [4.2.2] how it is possible to build a stochastic process on a vector space V by specifying the random **increments** of the process over all time-intervals of nonzero length.

For every $s < t$ in $\mathbb{T}_{ime}$, we defined a random variable $\alpha_{s,t}$, so that this collection of variables satisfies the conditions:

$$\forall r < s < t \in \mathbb{T}_{ime},\ \alpha_{r,t} = \alpha_{r,s} + \alpha_{s,t}. \quad (1)$$

We then defined some random variable α_0 as the "starting point" of the process, and legislate that $\alpha_{s,t}$ is just the "increment", $\alpha(s \to t)$, then the process α naturally falls into place.

In some sense, in defining the random increment variable $\alpha_{s,t}$, we are really sort of assigning a random **measure**[3] to the interval $[s,t) \subset \mathbb{T}_{ime}$. Hence, the collection of random increments $\{\alpha_{s,t}\ ;\ s,t \in \mathbb{T}_{ime}\}$ is really a special case of what is called a **random measure**.

5.2.1

Definition *Random Measure*

Let $(X, \mathcal{X})$ be a measurable space. A **random measure** is a random variable taking its values in the space of measures on $(X, \mathcal{X})$.

Formally, we let $(\Omega, \mathcal{W}, \mathbf{P})$ denote some probability space, and V a Banach space[4] , then a **random measure** on X, taking values in V, is a function:

$$\mu : \Omega \times \mathcal{X} \longrightarrow V$$

so that:

1. For any fixed $\omega \in \Omega$, the function

$$\mu_\omega : \mathcal{X} \longrightarrow V$$

is a $V-$valued measure.

[3]This measure takes its values in V. For intuitive simplicity, imagine $V := \mathbb{R}$; then this is just a sort of signed measure in the classical sense. The equation (1) is then just the standard **finite additivity** condition which any measure must satisfy.

[4]The "Banach space" condition is a distracting technicality. We want a vector space so we can add things together, and we want a Banach topology so that they converge. Just think of V as $\mathbb{R}$

2. For any fixed measurable subset $U \in \mathcal{X}$, the function:

$$\mu[U] : \Omega \ni \ \omega \mapsto \mu_\omega[U] \ \in V$$

is *measurable.* (That is, it constitutes a *random variable.*)

5.2.2

Example *Simple Poisson Measures* **Prerequisites:** [4.5.2]

Let $N : \Omega \times [0,\infty) \longrightarrow \mathbb{N}$ be a Poisson process with intensity $\alpha > 0$. Thus, for fixed ω, the path $N_\omega : [O,\infty) \longrightarrow \mathbb{N}$ is a peicewise constant path which increases by unit "jumps" at irregular intervals.

We can think of N as being the result of **integrating** a random measure, as follows:

Let $X_1, X_2, X_3, \ldots$ be independent random variables on Ω, all **exponentially distributed** with parameter α. Define the random variables $Y_1, Y_2, Y_3, \ldots$ by:

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \\ Y_3 &= X_1 + X_2 + X_3 \\ &\vdots \end{aligned}$$

(Thus the variables $\{Y_k|_{k\in\mathbb{N}}\}$ represent the **jump times** of the Poisson process N.)

For any Borel subset $U \subset [0,\infty)$, define $\mu[U]$ to be the random variable:

$$\mu[U] \ := \ \mathcal{C}_{ard}\left[\{k \in \mathbb{N} \ ; \ Y_k \in U\}\right]$$

that is, the number of jump times that lie within U. Then $\mu : \Omega \times \mathcal{B}[0,\infty) \longrightarrow \mathbb{N}$ is a **random measure** on $[0,\infty)$, defined via the random placement of a countable collection of "atoms", each of unit mass, and

spaced apart with exponentially distributed distances between them. (**Exercise**: Verify this.)

In a sense, we can think of μ as being the *derivative* of the Poisson process N, because we can generate a version of N by by **integrating** μ to produce the stochastic process $M : \Omega\$x[0,\infty) \longrightarrow \mathbb{N}$, defined:

$$M(t) \ := \ \mu[0,t].$$

(**Exercise**: Check that M is a simple Poisson process of intensity α).

The measure μ is called a **simple Poisson measure**.

This example is very suggestive. *Any* random $V-$valued measure on $\mathbb{T}_{ime}$ can be used to define a stochastic process. Hence, we can think of the random measures as being the *derivatives* of stochastic processes.

5.2.3

Theorem *Integrating a Random measure*

Let V be a vector space, let $0 \in \mathbb{T}_{ime} \subset [0,\infty)$, and let μ be a $V-$valued **random measure** on $\mathbb{T}_{ime}$. Let α_0 be a random variable taking values in V, independent of μ. Then there exists a stochastic process $\alpha : \mathbb{T}_{ime} \longrightarrow V$ so that:

1. $\alpha(0) := \alpha_0$.
2. For all $s < t \in \mathbb{T}_{ime}$, $\alpha(t) - \alpha(s) = \mu(s,t]$.

Furthermore, the process α has sample paths of almost surely **bounded variation.**

Proof:

The stochastic process α is constructed by specifying its **increments** over half-open time intervals, as is described in [4.2.2]. For every $s < t$ in $\mathbb{T}_{ime}$, we define

$$\alpha_{s,t} \ := \ \mu(s,t]$$

Then the results of [4.2.2] ensure that there is a stochastic process α satisfying 1 and 2.

It remains to show that α has sample paths of almost surely bounded variation. But for fixed ω in Ω, the function $\alpha_\omega : \mathbb{T}_{ime} \longrightarrow V$ is obtained by simply integrating the measure μ_ω; that is:

$$\alpha_\omega(t) \ := \ \mu_\omega[0,t]$$

μ_ω is a $V-$valued Borel measure, so any function defined in this manner must have bounded variation, by basic results of real analysis (see [6] Theorem 3.29).

□

Can we reverse this process? Given a stochastic process, can we define its "derivative" as a suitable random measure?

5.2.4

Theorem *The Derivative of a Stochastic Process of Bounded Variation*

Let $\alpha : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ be a stochastic process, and suppose that the sample paths of α have almost surely **bounded variation**.

For each fixed ω in Ω, define the measure μ_ω on half-open intervals of $\mathbb{T}_{ime}$ as follows:

$$\forall s < t \in \mathbb{T}_{ime}, \ \mu_\omega(s,t] := \alpha(t) - \alpha(s)$$

and, for each ω, extend to the Borel sigma-algebra in the standard way. Then the resulting function $\mu : \Omega \times \mathcal{B}(\mathbb{T}_{ime}) \longrightarrow V$ is a $V-$valued **random measure**.

Proof:

Fix ω. The path $\alpha_\omega : \mathbb{T}_{ime} \longrightarrow V$ is a function of bounded variation. The measure μ_ω defined above is then just the corresponding **Lebesgue-Stieltjes integral** for this path. (For details on this construction, see [6] section 3.5, pp. 95).

□

5.2.5

Remark

These theorem show that random measures are an excellent way of describing stochastic processes having *almost surely bounded variation.* This is an important limitation, because many processes, such as **Brownian motion**, do *not* have almost surely bounded variation, and thus, *cannot* be represented as integrals of random measures.

5.3 Integrating against Processes of Bounded Variation

Corequisites: [5.1]

The easiest sort of stochastic integration is that with respect to a stochastic process whose paths are almost surely of **bounded variation.** In this case, we can just apply the classical **Lebesgue-Stieltjes integral** to each path seperately, in order to obtain a sensible definition of stochastic integration. The only requirement is that we must impose conditions on our *integrand* to guarantee that the Lebesgue Stieltjes integrals converge properly on almost every path.

5.3.1 Review of Lebesgue-Stieltjes Integral

5.3.1

Notation *Partitions*

Let $T \in \mathbb{T}_{ime}.A$ **partition** of $[0,T]$ is any finite collection

$$\{0 \ := \ t_0 \ < \ t_1 \ < \ \ldots \ < \ t_N \ := T\}$$

The set of all partitions of $[0,T]$ will be denoted by $\mathcal{P}[0,T]$. This set is a **directed set** under the ordering relation of inclusion.

5.3.2

Definition *Normalised Bounded Variation*

Let V be a Banach Space, and let $g : \mathbb{T}_{ime} \longrightarrow V$. We say g is a function of **normalised bounded variation** (NBV) if g is **right continuous** , and if, for all $T \in \mathbb{T}_{ime}$,

$$\sup \left\{ \sum_{n=1}^{N} |g(t_n) - g(t_{n-1})| \ ; \ \{t_0, t_1, \ldots, t_N\} \in \mathcal{P}[0,T] \right\}$$

is finite.

5.3.3

Definition *Lebesgue-Stieltjes Integral*

Let $g : \mathbb{T}_{ime} \longrightarrow V$ be **NBV** . Let $f : \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ be Borel-measurable. For any $T \in \mathbb{T}_{ime}$, we define the **Lebesgue-Stieltjes Integral** of f with respect to g:

$$\int_0^T f\, dg \;:=\; \lim \left\{ \sum_{n=1}^{N} f(t_{n-1}) \cdot (g(t_n) - g(t_{n-1})) \;;\; \{t_0, t_1, \ldots, t_N\} \in \mathcal{P}[0,T] \right\}$$

where the limit is taken over the directed set $\mathcal{P}[0,T]$, if this limit exists.

5.3.4

Remark

Actually, we don't need to use the value $f(t_{n-1})$ in the above sum; we could take $f(x_n)$ for any point $x_n \in [t_{n-1}, t_n)$. However, for what follows, the choice of $x_n := t_{n-1}$ will prove to be especially convenient.

For further information on this material, see [6] section 3.5.

5.3.2 Definition of the Integral

Prerequisites: [5.3.1], [4.1.1]

For the remainder of this section, V is a **Banach Space** , $\mathbb{T}_{ime}$ is some subset of $[0, \infty)$ with $0 \in \mathbb{T}_{ime}$, and $(\Omega, \mathcal{W}, \mathbf{P})$ is some universal **sample space** . Finally,

$$\alpha : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

is a stochastic process, such that the sample paths of α are almost surely **NBV** .

5.3.5

Definition *Pathwise Integrability*

Let $f : \Omega \times \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ be some real-valued stochastic process. Then f is **integrable** with respect to g if, for almost all $\omega \in \Omega$, and for all $T \in \mathbb{T}_{ime}$, the Lebesgue-Stieltjes integral

$$\int_0^T f_\omega \, d\alpha_\omega$$

exists.

5.3.6

Example

1. Suppose f is such that, for almost all $\omega \in \Omega$ the path $f_\omega : \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ is **bounded** [5] . Then f is integrable with respect to any NBV process α.

2. Let $p \in [1, \infty]$, and suppose α has sample paths which are almost surely in $\mathbf{L}^p$. Let p^* be the **conjugate exponenent** to p —-that is, $\frac{1}{p} + \frac{1}{p^*} = 1$. If f is a process whose sample paths are almost surely in $\mathbf{L}^{p^*}$, then f is integrable with respect to α.

   Exercise: Verify this.

5.3.7

Definition *Pathwise Stochastic Integral* .

Let $\alpha : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ be a stochastic process whose sample paths are almost surely NBV. Let $f : \Omega \times \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ be another stochastic process, so that f is *integrable* with respect to α. Then we define the **pathwise stochastic integral** of f with respect to α to be the stochastic process

$$\int f \, d\alpha \; : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

defined as follows. For every $T \in \mathbb{T}_{ime}$, and every $\omega \in \Omega$,

$$\left(\int f \, d\alpha \right) (\omega, T) \quad := \quad \int_0^T f_\omega \, d\alpha_\omega$$

[5] Note: This is stronger than just saying $f_\omega \in \mathbf{L}^\infty$. We need f_ω to be bounded *everywhere* in $\mathbb{T}_{ime}$, not just "almost everywhere".

where the expression on the *right hand side* is a *Lebesgue-Stieltjes integral* . Normally, for fixed T, we will denote this random variable by the symbol:

$$\int_0^T f \, d\alpha.$$

5.3.3 Definition in Terms of Random Measures

Prerequisites: [5.2]

We can also define the pathwise integral in terms of the **random measure** obtained by **differentiating** the process α.

Again, let $\alpha : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ be a stochastic process whose sample paths are almost surely NBV. Let $\mu := d\alpha$ be the **derivative** of α; thus, μ is a **random measure**

$$\mu : \Omega \times \mathcal{T} \longrightarrow V$$

where $\mathcal{T}$ is the Borel sigma-algebra on $\mathbb{T}_{ime}$.

5.3.8

Definition *Pathwise Stochastic Integral* .

Let $f : \Omega \times \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ be stochastic process, so that f is *integrable* with respect to α. Then we define the **pathwise stochastic integral** of f with respect to α to be the stochastic process

$$\int f \, d\alpha \; : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

defined as follows. For every $T \in \mathbb{T}_{ime}$,

$$\left(\int f \, d\alpha \right)(T) \;\; := \;\; \int_0^T f \, d\mu$$

Exercise: Verify that this is the same as the previous definition.

5.3.4 Martingale Properties

Prerequisites: [martingales], [5.3.2]

5.3.9

Definition *Slowly Growing Integral*

Let $f : \Omega \times \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ and $\alpha : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ be stochastic processes as before, and consider the stochastic integral process $\int f \, d\alpha$. We will say that the integral $\int f \, d\alpha$ **grows slowly** if, for every $T \in \mathbb{T}_{ime}$,

$$\lim_{\epsilon \to 0} \mathbf{E}^{xpct} \left[\left\| \int_T^{T+\epsilon} f \, d\alpha \right\|_V \right] \quad = \quad 0$$

where $\|\bullet\|_V$ is the norm on the Banach space V.

5.3.10

Remark

1. The importance of this property lies in the fact that it allows us to approximate the continuous time process $\int f \, d\alpha$ with a sort of "discrete time mesh". This will be used in a proof below.

2. If we regard $\int_T^{T+\epsilon} f \, d\alpha$ as a **random variable**, then we can consider its $\mathbf{L^1}$**-norm** as a function on the measure space $(\Omega, \mathcal{W}, \mathbf{P})$. The **slow growth** condition can then be written

$$\lim_{\epsilon \to 0} \left\| \int_T^{T+\epsilon} f \, d\alpha \right\|_1 \quad = \quad 0.$$

3. Suppose we define the measure μ on $\mathbb{T}_{ime}$ as follows: for any $a, b \in \mathbb{T}_{ime}$,

$$\mu[a,b) \quad := \quad \mathbf{E}^{xpct} \left[\left\| \int_a^b f \, d\alpha \right\| \right]$$

Then the stochastic integral **grows slowly** if and only if μ is a **continuous** measure —that is, it has no atoms.

4. Let $p \in [1, \infty]$, and let p^* be the **dual exponent** [6] to p. For any time interval $[t, T) \subset \mathbb{T}_{ime}$, we can consider the random variables:

$$\left\| \alpha_{|[t,T)} \right\|_{var}$$

and

$$\left\| f_{|[t,T)} \right\|_{\infty}$$

where $\|\bullet\|_{var}$ is the **total variation** of a path in V.

Suppose that we can find an increasing sequence $\{0 := T_0 < T_1 < T_2 < \ldots\} \subset \mathbb{T}_{ime}$ so that, for all $n \in \mathbb{N}$, $\left\| \alpha_{|[T_n, T_{n+1})} \right\|_{var}$ has **finite $p-$th moment** , and $\left\| f_{|[T_n, T_{n+1})} \right\|_{\infty}$ has **finite p^*-th moment** . That is:

$$\mathbf{E}^{xpct}\left[\left\| \alpha_{|[T_n, T_{n+1})} \right\|_{var}^{p} \right] \quad and \quad \mathbf{E}^{xpct}\left[\left\| f_{|[T_n, T_{n+1})} \right\|_{\infty}^{p^*} \right] \text{ are finite.}$$

Then the integral grows slowly. (**Exercise**) .

5.3.11

Definition *History Filtration*

The **history filtration** of the stochastic process $\alpha : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ is the filtration of sigma algebras $\left\{ \mathcal{F}_t|_{t \in \mathbb{T}_{ime}} \right\}$, where, for any $T \in \mathbb{T}_{ime}$, $\mathcal{F}_T$ is the sigma-algebra on Ω generated by the random variables $\left\{ \alpha_t|_{t \in [0,T]} \right\}$.

5.3.12

Remark

1. Thus, the sigma-algebra $\mathcal{F}_T$ contains all information about the behaviour of the process α up to time T, but none thereafter.

[6] That is, $\frac{1}{p} + \frac{1}{p^*} = 1$.

2. The process α is a **martingale** *if and only if* , for any $t < T$ in $\mathbb{T}_{ime}$,

$$\mathbf{E}^{xpct}_{\mathcal{F}_t}[\alpha(T)] \;=\; \alpha(t),$$

which, in turn, is true *if and only if*

$$\mathbf{E}^{xpct}_{\mathcal{F}_t}[\alpha(t \to T)] \;=\; 0.$$

5.3.13

Definition *Nonanticipating*

Let $f : \Omega \times \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ be stochastic process. f is called **nonanticipating** of α if, for all $t < T$ in $\mathbb{T}_{ime}$, the random variable $f(t)$ is *measurable* with respect to $\mathcal{F}_t$.

This basically means that the process f cannot "predict" the future increments of α.

When we stochastically integrate f with respect to α, we are basically using f as a control function to *modulate* the increments of α. The **nonanticipating** property says that this modulation must always be "unbiased"; we can't use f to "skew" the evolution of the integral. The next theorem makes this precise.

5.3.14

Theorem

Let α be a stochastic process with paths of almost surely NBV, and let f be **integrable** with respect to α, Suppose that:

1. α is a **martingale** .
2. f is **nonanticipating** of α.
3. The integral $\int f\, d\alpha$ **grows slowly.**

Then the process $\int f\, d\alpha$ is also a **martingale** .

Proof:
Let F denote the stochastic process:

$$F := \int f \, d\alpha.$$

For any $N \in \mathbb{N}$, let $F_N : \Omega \times \mathbb{T}_{time} \longrightarrow V$ be the stochastic process defined as follows. For any fixed $T \in \mathbb{T}_{time}$, let $\bar{T}_N$ be the *largest multiple of* $\frac{1}{2^N}$ *less than* T. That is,

$$\bar{T}_N \ := \ \lfloor 2^N \cdot T \rfloor$$

Then define:

$$\begin{aligned} F_N(T) \ := \ & \sum_{n=1}^{\bar{T}_N} f\left(\frac{n-1}{2^N}\right) \cdot \alpha\left[\frac{n-1}{2^N} \to \frac{n}{2^N}\right] \\ & + \ f\left(\frac{\bar{T}_N}{2^N}\right) \cdot \alpha\left[\frac{\bar{T}_N}{2^N} \to T\right]. \qquad (1.1). \end{aligned}$$

In other words, for any $\omega \in \Omega$, and $T \in \mathbb{T}_{time}$,

$$\begin{aligned} F_N(\omega, T) \ := \ & \sum_{n=1}^{\bar{T}_N} f\left(\omega, \frac{n-1}{2^N}\right) \cdot \left[\alpha\left(\omega, \frac{n}{2^N}\right) - \alpha\left(\omega, \frac{n-1}{2^N}\right)\right] \\ & + \ f\left(\omega, \frac{\bar{T}_N}{2^N}\right) \cdot \left[\alpha\left(\omega, T\right) - \alpha\left(\omega, \frac{\bar{T}_N}{2^N}\right)\right]. \qquad (1.2) \end{aligned}$$

Thus, F_N is a sort of "discrete time" approximation of F.

Claim 1: For almost all ω, and all T,

$$\lim_{N \to \infty} F_N(\omega, T) \ = \ F(\omega, T).$$

Proof: Fix ω. Then the expression (1.2) for $F_N(\omega, T)$ is just an approximation for the **Lebesgue-Stieltjes Integral**

$$\int_0^T f_\omega(t) \, d\alpha_\omega[t]$$

As $N \to \infty$, this will converge to the correct value.
□[1]

Claim 2:
For any $T \in \mathbb{T}_{ime}$,

$$\mathbf{L}^1\text{-}\lim_{N \to \infty} F_N(T) \;=\; F(T),$$

where the limit takes place in the $\mathbf{L}_1$ norm on random variables.

Proof: This follows immediately from Claim 1 and the Lebesgue Dominated Convergence Theorem.
□[2]

Claim 3:
Think of F_N as defining a **discrete time process** :

$$\hat{F}_N : \Omega \times \mathbb{N} \longrightarrow V$$

where $\hat{F}_N(n) \;:=\; F_N\left(\frac{n}{2^N}\right)$.
Then, for any fixed $N \in \mathbb{N}$, the process $\hat{F}_N$ is a **discrete time martingale** .

Proof:
Let $\left\{\mathcal{F}_t|_{t \in \mathbb{T}_{ime}}\right\}$ be the **history filtration** of α.
For any $n \in \mathbb{N}$, let $\mathcal{F}_{[n]} \;:=\; \mathcal{F}_{\frac{n}{2^N}}$. Thus, $\left\{\mathcal{F}_{[n]}|_{n \in \mathbb{N}}\right\}$ is the discrete-time **history filtration** of the discrete-time martingale

$$\{\alpha_0,\ \alpha_{\frac{1}{2^N}},\ \alpha_{\frac{2}{2^N}},\ \alpha_{\frac{3}{2^N}},\ \ldots\}$$

So, for any natural numbers $M_0 < m$,

$$\begin{aligned}
&\mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[f\left(\frac{m}{2^N}\right) \cdot \alpha\left[\frac{m}{2^N} \to \frac{m+1}{2^N}\right]\right] \\
&\quad= \mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[\mathbf{E}^{xpct}_{\mathcal{F}_{[m]}}\left[f\left(\frac{m}{2^N}\right) \cdot \alpha\left[\frac{m}{2^N} \to \frac{m+1}{2^N}\right]\right]\right] \\
&\quad= \mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[f\left(\frac{m}{2^N}\right) \cdot \mathbf{E}^{xpct}_{\mathcal{F}_{[m]}}\left[\alpha\left[\frac{m}{2^N} \to \frac{m+1}{2^N}\right]\right]\right] \qquad \text{(since } f \text{ is } \textbf{nonanticipating} \text{)} \\
&\quad= \mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[f\left(\frac{m}{2^N}\right) \cdot 0\right] \qquad \text{(since } \alpha \text{ is a } \textbf{martingale} \text{)} \\
&\quad= 0
\end{aligned}$$

Thus, for any $M_0 < M_1$,

$$\begin{aligned}
\mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[\hat{F}_N(M_1)\right] &= \mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[\hat{F}_N(M_0) + \sum_{m=M_0}^{M_1} f\left(\frac{m}{2^N}\right)\cdot\alpha\left[\frac{m}{2^N}\to\frac{m+1}{2^N}\right]\right] \\
&= \hat{F}_N(M_0) + \sum_{m=M_0}^{M_1}\mathbf{E}^{xpct}_{\mathcal{F}_{[M_0]}}\left[f\left(\frac{m}{2^N}\right)\cdot\alpha\left[\frac{m}{2^N}\to\frac{m+1}{2^N}\right]\right] \\
&= \hat{F}_N(M_0) + \sum_{m=M_0}^{M_1} 0.
\end{aligned}$$

□[3]

Claim 4:
Let N, M_0, and M_1 be in $\mathbb{N}$, and suppose $T_0 < \frac{M_0}{2^N}$. Then:

$$\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right]\right] = 0$$

Proof:
We will do this via the $\mathbf{L}^1$ norm; we will show that:

$$\left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right]\right]\right\|_1 = 0$$

Fix $\epsilon > 0$. By Claim 2, find $K > N$ large enough so that

$$F\left(\frac{M_0}{2^N}\right) \underset{\epsilon}{\approx} F_K\left(\frac{M_0}{2^N}\right) \quad and \quad F\left(\frac{M_1}{2^N}\right) \underset{\epsilon}{\approx} F_K\left(\frac{M_1}{2^N}\right).$$

where distance is measured in the $\mathbf{L}^1$ norm. Thus:

$$F\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right] \underset{2\cdot\epsilon}{\approx} F_K\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right]$$

Thus,

$$\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right]\right] \underset{2\cdot\epsilon}{\approx} \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}[F]_K\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right]$$

But by Claim 3, $\hat{F}_K$ is a martingale, so

$$\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F_K\left[\frac{M_0}{2^N}\to\frac{M_1}{2^N}\right]\right] = 0.$$

(**Exercise.**)
Thus,

$$\mathbf{E}^{\mathit{apot}}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_0}{2^N} \to \frac{M_1}{2^N}\right]\right] \quad \underset{2\cdot\epsilon}{\approx} \quad 0.$$

This is true for any $\epsilon > 0$, so we're done.
$\square$[4]

Claim 5:
F is a **continuous time martingale.**

Proof:
Let $T_0 < T_1$ in $\mathbb{T}_{\mathit{ime}}$. We want to show:

$$\mathbf{E}^{\mathit{apot}}_{\mathcal{F}_{T_0}}\left[F[T_0 \to T_1]\right] \;=\; 0.$$

We will do this via the $\mathbf{L}^1$ norm; we will show that:

$$\left\| \mathbf{E}^{\mathit{apot}}_{\mathcal{F}_{T_0}}\left[F[T_0 \to T_1]\right] \right\|_1 \;=\; 0.$$

Now, by hypothesis, the stochastic integral **grows slowly** . Thus, for any $\epsilon > 0$, there is a $N \in \mathbb{N}$, so that, if $\delta < \frac{1}{2^N}$, then

$$\left\| \int_{T_0}^{T+\delta} f \; d\alpha \right\|_1 \;<\; \epsilon$$

and

$$\left\| \int_{T_1-\delta}^{T_1} f \; d\alpha \right\|_1 \;<\; \epsilon.$$

where we are taking the $\mathbf{L}^1$-norms of these integrals as random variables on the measure space $(\Omega, \mathcal{W}, \mathbf{P})$.
Thus, if we find numbers M_0 and M_1 so that

$$\frac{M_0}{2^N} > T_0 \quad \textit{and} \quad \frac{M_0}{2^N} \;\underset{\delta}{\approx}\; T_0,$$

and

$$\frac{M_1}{2^N} < T_1 \quad \textit{and} \quad \frac{M_1}{2^N} \;\underset{\delta}{\approx}\; T_1,$$

then

$$\left\| F\left[T_0 \to \frac{M_0}{2^N}\right] \right\|_1 \quad and \quad \left\| F\left[\frac{M_1}{2^N} \to T_1\right] \right\|_1 \qquad (A)$$

Now, note that:

$$\begin{aligned}
\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}[F[T_0 \to T_1]] &= \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[T_0 \to \frac{M_0}{2^N}\right] + F\left[\frac{M_0}{2^N} \to \frac{M_1}{2^N}\right] + F\left[\frac{M_1}{2^N} \to T_1\right]\right] \\
&= \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[T_0 \to \frac{M_0}{2^N}\right]\right] + \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_0}{2^N} \to \frac{M_1}{2^N}\right]\right] \\
&\qquad + \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_1}{2^N} \to T_1\right]\right]. \\
&= \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[T_0 \to \frac{M_0}{2^N}\right]\right] + 0 + \mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_1}{2^N} \to T_1\right]\right].
\end{aligned}$$

by Claim 4. Thus,

$$\begin{aligned}
\left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}[F[T_0 \to T_1]]\right\|_1 &\leq \left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[T_0 \to \frac{M_0}{2^N}\right]\right]\right\|_1 + \left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_1}{2^N} \to T_1\right]\right]\right\|_1 \\
&= \left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[T_0 \to \frac{M_0}{2^N}\right]\right]\right\|_1 + \left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}\left[F\left[\frac{M_1}{2^N} \to T_1\right]\right]\right\|_1 \qquad \text{(By C} \\
&\leq \left\|F\left[T_0 \to \frac{M_0}{2^N}\right]\right\|_1 + \left\|F\left[\frac{M_1}{2^N} \to T_1\right]\right\|_1 \qquad \text{(By Triangle Inequal} \\
&\leq \epsilon + \epsilon \qquad (by\ (A)) \\
&= 2 \cdot \epsilon.
\end{aligned}$$

We can, of course, make ϵ arbitrarily small. Hence, we conclude:

$$\left\|\mathbf{E}^{xpct}_{\mathcal{F}_{T_0}}[F[T_0 \to T_1]]\right\|_1 = 0.$$

□[4]
□

5.4 Integrating against Tame Levy Processes

Prerequisites: [5.3] , [4.5.4] [4.6.5]

Corequisites: [5.1]

So far, we have three approaches to stochastic integration. If α is a **deterministic** process of bounded variation, then of course we can define stochastic integrals with respect to α just by means of the classical Lebesgue-Stieltjes Integral. If β is a **martingale** whose increments have **finite variance** , then we can use the machinery of the Itô integral to integrate with respect to β. Finally, if γ is a process whose sample paths have *a.s* **bounded variation** , then we can apply the Lebesgue-Stieltjes technique seperately to each path, as discussed in [5.3] .

By combining these three approaches, we can define stochastic integration for a broad class of **Levy processes** . To do this, we will employ the **canonical representation** developed in [4.6.5] .

5.4.1

Definition *Tame L*

Let $\mathcal{L} : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ be a Lévy process, and suppose that $\mathcal{L}$ has **canonical representation** :

$$\mathcal{L} \;=\; \alpha + \beta + \gamma$$

1. $\alpha(t) := t \cdot \vec{a}$ for some fixed (nonrandom) vector $\vec{a} \in V$;
2. $\beta(t) := B_1(t)\cdot\vec{v}_1 + B_2(t)\cdot\vec{v}_2 + \ldots + B_D(t)\cdot\vec{v}_D$, and where $B_1, \ldots B_D$ are D independent, one-dimensional **Brownian motions** , where $\vec{v}_1, \ldots \vec{v}_D$ are fixed (nonrandom) vectors in V;
3. γ a **compound Poisson process** on V with **intensity measure** μ, independent of $B_1, \ldots, B_D$;

Say that $\mathcal{L}$ is **tame** if the Poisson process γ is of almost surely **bounded variation** .

Equivilantly, $\mathcal{L}$ is **tame** if the measure μ has **finite first moment** [4.5.4] .

5.4.2

Remark

Intuitively, this says the process isn't "too" discontinuous. The Brownian component has **continuous** sample paths, and the deterministic component is actually **linear** . The Poisson process introduces a potentially

infinite collection of discontinuities, but we require that "most" of these be relatively "small" in size.

5.4.3

Definition *Integration with respect to Tame L*

Let $\mathcal{L} : \Omega \times \mathbb{T}_{ime} \longrightarrow V$ be a **tame Lévy process**, and let $\mathcal{L} = \alpha + \beta + \gamma$, where α, β, and γ are as before.

Let $f : \Omega \times \mathbb{T}_{ime} \longrightarrow \mathbb{R}$ be some real-valued stochastic process, and suppose:

1. f is **nonanticipating** of β.
2. $f \in \mathbf{L}^2\left(\Omega \times \mathbb{T}_{ime};\ \mu_{[\beta]}\right)$, where $\mu_{[\beta]}$ is the **Doléans measure** associated to β.
3. f is **integrable** with respect to γ.

Define the **stochastic integral** of f with respect to $\mathcal{L}$ to be the stochastic process:

$$\int f\, d\mathcal{L} \quad : \Omega \times \mathbb{T}_{ime} \longrightarrow V$$

defined, for all $T \in \mathbb{T}_{ime}$, by:

$$\int_0^T f\, d\mathcal{L} \quad := \quad \left(\int_0^T f\, dt\right) \cdot \vec{a} \ + \ \int_0^T f\, d\beta \ + \ \int_0^T f\, d\gamma,$$

where the *first* integral on the right hand side is just the standard **Lebesgue integral** on $\mathbb{T}_{ime}$, the *second* integral on the right hand side is an **Itô integral**, and the *third* integral on the right hand side is a **pathwise stochastic integral.**

Bibliography

[1] Michael Artin. *Algebra.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 07632, first edition, 1991.

[2] Jean Bertoin. *Lévy Processes.* Cambridge University Press, 40 West 20th Street, New York, NY, 10011-4311, USA, first edition, 1996.

[3] Kai Lai Chung and Ruth J. Williams. *Introduction to Stochastic Integration.* Birkhauser Boston Inc., Boston, Basel, Stuttgart, first edition, 1983.

[4] John B. Conway. *A Course in Functional Analysis.* Springer-Verlag, New York, second edition, 1990.

[5] William Feller. *An Introduction to probability theory and its applications.* John Wiley and Sons, Inc., New York, second edition, 1966.

[6] Gerald B. Folland. *Real Analysis.* John Wiley and Sons, New York, 1984.

[7] Yitzhak Katznelson. *An Introduction to Harmonic Analysis.* Dover, 180 Varick Street, New York, NY, 10014, USA, first edition, 1976.

[8] Gennady Samorodnitsky and Murad S. Taqqu. *STable Non-Gaussian random processes: stochastic models with infinite variance.* Chapman and Hall, 29 West 35th Street, New York, NY, 10001, USA, 1994.