Basic Mathematical Details of Stochastic Processes

Ryan E. Scott

October 9, 2025

Contents

1 Introduction

1.1 A Thought Experiment

Let's consider briefly the sort of object we want to study. Fundamentally, we plan to sit down in front of a computer, browse to a website or log into a Bloomberg terminal, and look at an object that looks like this:

We want to analyze objects that look like this. Conceptually, this is a graph specifying price as a function of time. Now we know that in the case of these price graphs, the price only updates at discrete time intervals, so when we select a date range we are actually selecting a sequence of values $\langle p_1, p_2, ..., p_N \rangle$ and there is a 1-1 relationship between the subscript and the time. In the above, the initial time starts around the 31st of October and ends on the 29th of November. It turns out that the appropriate mathematical tool to analyze this sort of structure is called a *(discrete) stochastic process* and each p_k should be thought of as a *random variable*. In order to get a grasp on how this works, let's consider a simpler example.

Suppose I have a black box, and inside of it there is a single barrel of WTI crude oil. Now suppose there is a digital LED display mounted to the black box that conveys two pieces of information: first, the ticker for the product (e.g., USO), and second, the current (market) price. For the sake of the thought experiment, we don't need to worry about additional complications. Now suppose I take a stop watch and every 30 seconds I record the value of the price of the barrel of oil, and in between these 30 seconds I cannot see the display. Every time I measure the value, I can think of it as a simple statistical experiment. Doing this, I can consider the possible range of values the price can take on at each interval. In this case, I can assume that the price could be anywhere from 0$ all the way up to ∞ (which in this case would be interpreted as there being no finite quantity of resources that could combine to value equal to one barrel of WTI crude oil). Of course this latter case is a mathematical idealization, but it is nonetheless useful for framing our problem. If we treat the problem this way, we see that the price satisfies the relationship $0 \leq p_k < \infty$ and each experiment charts where the price lies in this range at each time interval. From here, we can start to ask interesting statistical questions: what is the probability that the price will lie between two values at the next click? How does the average value trend? Are there any global patterns which characterize the data? What is the distribution of probabilities over the price range? Does it change in time? It is the purpose of these notes to offer some basic details an insight into these questions, to elucidate the foundations of this interesting branch of mathematics, and to render its concepts intuitive and useful to the reader. We hope you find them interesting and helpful.

2 Preliminaries

We will introduce everything constructively, from first principles, assuming the reader has no background except the basics of algebra, pre-calculus, and a first course in calculus. We also suppose he has a mathematical disposition and a willingness to construct additional examples for himself as he reads whenever he is confused about a concept. We include ample examples as well as sketches of formal proofs for most claims.

2.1 Set Theory

A *set* is an object A with *elements* $x \in A$ (read 'x belongs to A' or 'x (is) in A' or 'x is an element of A'). Sets can be built in two ways: 1) listing the elements of the set: $A = \{a, 0, 7, q,), |, ?\}$ (notice you can throw *anything you want* into the set–it's just a collection of distinct *things*); 2) specify a *rule* that tells the reader how to judge whether an element of some 'larger' (elementary) set[1] belongs to the set in question. When we do this, we usually write $A = \{x|\phi(x)\}$, which should be read as 'the set A of elements x such that [that's the bar] $\phi(x)$ is true.' As an example of this latter method, if we take the plane of points (x, y) (i.e., ordered pairs), then we can define the unit circle as $S = \{(x, y)|x^2 + y^2 = 1\}$. To see if some $(x_0, y_0) \in S$, we just calculate $x_0^2 + y_0^2$ and compare it to 1.

You'll recall that in algebra, you need to follow rules which allow you to change an equation from one form to another–for example if $x = y$, then $x + 1 = y + 1$ or $x - y = 0$. Sets can be manipulated in this way as well, using the rules of union, intersection, and complements. One can also define the notion of equality, as well as the idea of a subset. Since we know sets have elements, we can define these operations in terms of what happens to the elements. The *union* of a pair of sets A and B is denoted $A \cup B$ and is given by $A \cup B = \{x|x \in A \text{ or } x \in B\}$. The *intersection* of the same pair of sets is $A \cap B = \{x|x \in A \text{ and } x \in B\}$. The *(relative) complement* of the pair of sets is $A - B = \{x|x \in A \text{ and not } x \in B\}$. Here's a discrete example: let $\overline{n} = \{0, 1, ..., n - 1\}$ be the set of the first n non-negative integers (also called natural numbers). Then $\overline{3} \cup \overline{2} = \overline{3}$, $\overline{3} \cap \overline{2} = \overline{2}$, and $\overline{3} - \overline{2} = \{2\}$. If the reader can follow these, he should be able to come up with other examples on his own. Here's an exercise: write the set S we defined above as the union of two semicircles. Can you write a quarter circle as the intersection of two semicircles? How are the two semicircles you just wrote down related? (Hint: think about the definition of the complement.)

Two sets are *equal* if and only if each element of one belongs to the other–that is, $A = B$ if and only if for all $x \in A$ implies $x \in B$ and for all $y \in B$ implies $y \in A$.

For the sake of these notes, we assume the reader is familiar with the basics of set algebra. That is, we suppose he knows something about set membership, unions, intersections, (relative) complements, and the like. These will be indispensable in the following theory, so we recall their definitions briefly and direct the reader to any good book on basic set theory to fill in the details, such as [Ha60]. Listing these basic definitions out, let A, B be sets. Then

[1] It's too far from our current purpose to analyze this in detail; suffice it to say that there are problems at the foundation of mathematics and whenever we specify a rule for building a set, we need to tacitly suppose there is an 'ambient' set we can pull from–the most common ones are $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. The interested reader is pointed to, say, Kleene's *Introduction to Metamathematics*.

$$
\begin{aligned}
A \subseteq B &\Leftrightarrow \forall x[(x \in A) \text{ implies } (x \in B)] \\
A \cup B &= \{x|(x \in A) \text{ or } (x \in B)\} \\
A \cap B &= \{x|(x \in A) \text{ and } (x \in B)\} \\
B - A &= \{x|(x \in B) \text{ and } \text{not}(x \in A)\}
\end{aligned}
$$

All other operations can be specified in terms of these. Some important notes: first, we use $\overline{A}$ when it is obvious that complements are taken with respect to B. It is also common to see $B \setminus A$ for complements. It is straight forward to show that De Morgan's laws and the distributive identity of $\cup$ over $\cap$ (and vice versa) hold. (If these aren't obvious to you, they make great elementary exercises in set theory.) Equality can always be proven either by establishing a bijection between sets or by proving $A \subseteq B$ and $B \subseteq A$. We use the notation $\oslash$ for the empty set–i.e., the set which contains no elements. The *Cartesian product* of two sets A and B is denoted $A \times B$ and is defined by $A \times B = \{(x, y)|x \in A \text{ and } y \in B\}$. For example, the Cartesian product of the set $\{0, 1\}$ with itself is $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Lastly, we mention the idea of a *power set* which is the set of all subsets of a given set. If A is the set in question, then the power set is often denoted $\mathcal{P}(A)$ or 2^A. As a simple example, for the set $\mathbf{2} = \{0, 1\}$, the power set $\mathcal{P}(\mathbf{2}) = \{\oslash, \{0\}, \{1\}, \{0, 1\}\}$.

2.1.1 Functions

Functions are essential for contemporary mathematical understanding and analysis, so we mention them explicitly. A *function* f is a mathematical entity which sends *each* element of a domain to a *single* element of the codomain, and is usually denoted $f : A \to B$, where f is the function, A is the *domain*, and B is the *codomain*. A function $f : A \to B$ is *well-defined* when it: 1) is defined on the whole domain (i.e., for every $x \in A$ there is a $y \in B$ such that $y = f(x)$), and; 2) sends every $x \in A$ to a *single* $y = f(x) \in B$ (i.e., for every $x \in A$, if $y_1 = f(x)$ and $y_2 = f(x)$ then $y_1 = y_2$). Domains and codomains do not necessarily need to be *sets*, but frequently they are. (When they are not, we usually use the word *arrow* or *map* rather than function.) If $x \in A$ and $f : A \to B$, then $f(x)$ is called the *(direct) image* of x under f. Functions *compose* whenever the domain of the following is the same as the codomain of the preceding, and their composition is associative. For example, if $f : A \to B$ and $g : B \to C$, then $h = g \circ f : A \to C$ defined by $h = g \circ f(x) = g(f(x))$. Functions can often come equipped with some structure-preserving property. For example, consider the function $f : \mathbb{R} \to \mathbb{R}_{>0}$ that sends $x \mapsto e^x$ (can you see that this function is well-defined?). Then if we think of the real numbers together with the operation of addition, this map actually sends addition to multiplication by the observation $x + y \mapsto f(x+y) = e^{x+y} = e^x e^y = f(x) \cdot f(y)$. This is an example of a *morphism*, which is a function with a structure-preserving property. In this case, we've mapped the group of real numbers under addition to the group of positive real numbers under multiplication. See for example [Pi10] for more details. Morphisms are ubiquitous in mathematics and can arguably thought of as *the very heart of mathematics itself*[2]. We also note that whenever sets $A \subseteq B$ there is a set function called *inclusion* or *injection* defined by the map $i : A \hookrightarrow B$ given by the function $i(x) = x$, but the image of x under i is to be thought of as an element of B rather than A. This trick comes in handy quite a bit, and this particular collection of notes is no exception. Finally, we mention the idea of *inverse image*. The inverse image of a function $f : A \to B$ is defined at each point $y \in B$ by $f^*(y) = f^{-1}(y) = \{x|f(x) = y\}$. Thus for a function $f : \{0, 1, 2\} \to \{0, 1\}$ defined by $\{0, 1\} \mapsto 0$ and $\{2\} \mapsto 1$, we have $f^*(0) = \{0, 1\}$ and $f^*(1) = \{2\}$. Interestingly, this definition implies that a function $f : A \to B$ induces a related function $f^* : \mathcal{P}(B) \to \mathcal{P}(A)$; you can demonstrate to yourself a very strong command of set theoretic knowledge if you can show this is the case. (Remember: a function is well-defined if whenever it takes a single input, it gives a single output, and this is true for the *whole* domain of the function.) Interestingly, inverse images preserve set-algebra operations.

$$
\begin{aligned}
A \subseteq B &\Rightarrow f^*(A) \subseteq f^*(B) \\
f^*(A \cup B) &= f^*(A) \cup f^*(B) \\
f^*(A \cap B) &= f^*(A) \cap f^*(B) \\
f^*(B - A) &= f^*(B) - f^*(A)
\end{aligned}
$$

This will be important below when we start to look at random variables and distributions. As with the previous

[2]Philosophically, and interestingly, you might also think that this means the heart of mathematics is really the property of *distributivity* or *linearity* (the general case is called *functoriality*), but it would take us too far from our current purpose to discuss this. See for example [Go14] for more.

fact, if you can show these properties, you will demonstrate a very strong understanding of set algebra.

2.1.2 Differential Equations and their Discrete Representations

One of the key mathematical tools for analyzing systems is the theory of differential equations. If we consider a real-valued function $f : \mathbb{R} \to \mathbb{R}$ (of one variable, i.e., $x \mapsto f(x)$), then a differential equation can be regarded as an algebraic function of two variables, the function f itself and a differential operator ∂_x. When we consider a function of only one variable, $\partial_x = d/dx$, the usual derivative operator. If G is such a function, then a *differential equation* can be defined as an equation satisfying

$$G(f, \partial_x) = 0$$

Some examples are

$$\begin{aligned} \partial_x f &= \frac{df}{dx} = 0 \\ \partial_x^2 f^2 + (\partial_x f)^3 &= \frac{d^2}{dx^2}(f^2) + \left(\frac{df}{dx}\right)^3 = 0 \\ \sum_{k=1}^{N} a_k \partial_x^k f &= \sum_{k=1}^{N} a_k \frac{d^k f}{dx^k} = 0 \end{aligned}$$

If we want to think of the associated functions G, they would be

$$\begin{aligned} G(x, y) &= yx \\ G(x, y) &= y^2 x^2 + (yx)^3 \\ G(x, y) &= \sum_k a_k y^k x \end{aligned}$$

where we aren't allowed to move the variables around. We can *discretize* derivatives using the traditional definition. Recall from calculus

$$\left.\frac{df}{dx}\right|_x = \lim_{\delta x \to 0} \frac{f(x + \delta x) - f(x)}{\delta x}$$

If we truncate this limit, and consider a sequence of points in an interval, then we can *define* the (forward) finite difference at a point x_k as

$$\frac{\Delta f}{\Delta x} = \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} \tag{1}$$

Notice in particular this defines a recursion relation

$$f_{k+1} := f(x_{k+1}) = f(x_k) + \frac{\Delta f}{\Delta x}(x_{k+1} - x_k) = f_k + \frac{\Delta f}{\Delta x}\Delta x \tag{2}$$

which will be useful in gaining an intuition for regression models later.

2.2 Probability Theory

Because the theory of probability will be foundational for our purposes, we review this particular section in detail and rely a bit more heavily on the mathematical theory because it will elucidate the whole theory. At the bottom, things are foundationally described in terms of the event space and the probability measure. We can illustrate all of these things in terms of basic properties and morphisms. We think this presents a coherence to the theory that usual approaches neglect. The interested reader is directed to basic texts on category theory. See for example [Mc92] (the author's personal favorite), but there are many good introductory texts ([Go14] is also commonly cited as a solid introductory text, but we would recommend otherwise). We will not rely explicitly on the theory, however, and the mathematical novice should lose nothing in terms of accessibility of these notes by knowing no category theory.

2.2.1 Sample Spaces, Event Spaces, and σ-Algebras

A *sample space* is to be regarded as a set of elements each of which is to be thought of as an elementary event that is the potential outcome of an experiment. The prototypical discrete examples are the tossing of a coin or the rolling of a die. In the former, the sample space is $\{H, T\}$ while the sample space of the latter is $\{1, 2, 3, 4, 5, 6\}$. A generic sample space will be denoted Ω and the elementary events will be denoted $\omega_0 \in \Omega$. Subsets of Ω will be denoted $\{\omega_0\} \subseteq \Omega$ or $\omega \subseteq \Omega$ (note the missing subscript!). An *event space* over the given sample space is a space of outcomes that are in general subsets of Ω. For example, in the case of a die, the event $\{2, 4, 6\}$ represents the outcome of rolling an even number. More specifically, an event space is a subset of $\mathcal{P}(\Omega)$ subjected to some algebraic conditions and in particular it is a *σ-algebra*, or a *measurable space*. A σ-algebra is specified in terms of the computational rules one is allowed to follow within it[3]. If Ω is our set, we can define a σ-algebra on Ω, which we denote $\mathcal{A}(\Omega)$, by the following rules:

(S1) $\mathcal{A}(\Omega) \subseteq \mathcal{P}(\Omega)$

(S2) $\oslash \in \mathcal{A}(\Omega)$

(S3) Whenever $\omega \in \mathcal{A}(\Omega)$, so is $\overline{\omega} = \Omega - \omega \in \mathcal{A}(\Omega)$

(S4) If $\{\omega_k\}$ is a collection (perhaps infinite) of subsets of Ω (i.e., for each integer k, $\omega_k \subseteq \Omega$) and for all k, $\omega_k \in \mathcal{A}(\Omega)$, then their union is also in $\mathcal{A}(\Omega)$, which we denote by $\bigcup_k \omega_k \in \mathcal{A}(\Omega)$

If these rules are confusing, the best thing to do is to consider some examples. Working with finite sets is always helpful because they are easy to manipulate and illustrate the idea. As a basic example, consider $\mathbf{2}$ again. One possible algebra we can choose to assign to $\mathbf{2}$ is the set $\{\oslash, \mathbf{2}\}$. We just run through the definitions to verify. (S1) Yes, this is a subset of the power set of $\mathbf{2}$; (S2) Yes, $\oslash \in \mathcal{A}(\mathbf{2})$; (S3) By this rule, we consider $\overline{\oslash} = \mathbf{2}$ (which is easy to check) and furthermore $\overline{\mathbf{2}} = \oslash$, so that this is closed under complements; (S4) since the sets we have are just $\oslash$ and $\mathbf{2}$, we can only consider at most $\oslash \cup \mathbf{2} = \mathbf{2}$ (since $\oslash \cup A = A$ for any set A). Thus, this is a σ-algebra, albeit an uninteresting one. From a probability perspective, this would be the example of an event space in which either nothing happens or something happens (from a sample space of two elementary outcomes). But since we assume that there will be an outcome for the experiment, then something always happens, which corresponds to the whole space. There is one other possible algebra definable on $\mathbf{2}$, which we leave for the reader to work out for himself. (If you can do it, it is a solid demonstration of your understanding so far.)

One can work out other properties of algebras. For example, by (S3), since $\oslash \in \mathcal{A}(\Omega)$ then $\Omega \in \mathcal{A}(\Omega)$ necessarily (can you see how?). Additionally by (S3) and (S4), $\bigcap_k \omega_k \in \mathcal{A}(\Omega)$ as well (you can show this by using (S3), (S4) and De Morgan's laws–which you should do now if it isn't obvious to you). See [GW14] (or any other introductory mathematical text on probability) for other examples, properties, and exercises. [Bo06] is another good text, albeit less rigorous and more intuitive. Furthermore, since $\oslash \in \mathcal{A}(\Omega)$ and $\Omega \in \mathcal{A}(\Omega)$, we have for every subset $\omega \subseteq \Omega$ that $\oslash \subseteq \omega \subseteq \Omega$. The reader should note that $\subseteq$ here appears to play the same role as $\leq$ does for the real numbers $\mathbb{R}$ (and in particular, he should be thinking of x such that $0 \leq x \leq 1$). In fact, one can show that for any sets A, B, C, $A \subseteq A$ (reflexivity), $A \subseteq B$ and $B \subseteq C$ implies $A \subseteq C$ (transitivity), and $A \subseteq B$ and $B \subseteq A$ implies $A = B$ (antisymmetry). Any relation which satisfies these three properties is called a *total ordering* and one can think of it as exactly analogous to $\leq$ in the real numbers (the reader should check for himself that $\leq$ is a total ordering if it isn't clear–consider three arbitrary real numbers x, y, z and go from there). This observation, together with the rules that define a σ-algebra, will be crucial for making mathematical sense out of the idea of a random variable, which will be the foundational concept in a stochastic process.

We mention in passing the canonical example of an event space, which is ubiquitous in applied statistics. If we take $\Omega = \mathbb{R}$ and the σ-algebra generated by all open intervals in $\mathbb{R}$ (i.e., intervals of the form (a, b)), we obtain the

[3] In general, this is true for any collection of mathematical entities–consider again addition in $\mathbb{R}$; there you know that addition is commutative, associative, for every x there is a y such that $x + y = 0$, and for any x, $x + 0 = x$–these rules define an *abelian group*.

algebra called the *Borel algebra* (on $\mathbb{R}$), denoted $\mathcal{B}(\mathbb{R})$. It has all intervals of the form (a,b), $(a,b]$, $[a,b)$, and $[a,b]$ (and their countable unions). As an example of how this might be, we show that we can obtain the singleton set $\{a\}$ from the open intervals and the rules characterizing a σ-algebra. Begin with (a,b). By the fact that we have all open intervals, the Borel algebra must also have the intervals $(-\infty,a)$ and (b,∞). By closure under unions, then $(-\infty,a)\cup(b,\infty)$ is also in the algebra. By complements, the complement of this is also in the algebra, i.e., $[a,b]$. Next consider the case when $b = n\cdot a + c$ for some real c and positive integer n with $0 \le c < n\cdot a \le b$ [4]. By the same line of reasoning, for any positive integer n, the closed interval $[a,b/n] = [a,a+c/n]$ is also in the algebra. By the derived property of closure under countable intersections, the set $\bigcap_n [a,a+c/n]$ is also in the algebra. But what is this set? If we consider the limit as $n\to\infty$, we see that $c/n \to 0$, but then the only element that is in this set for every n is a, so that $\bigcap_n [a,a+c/n] = [a,a] = \{a\}$. From this, the half-open and half-closed intervals easily follow (specifically using (S4)). We're done. We also mention in passing that the Borel algebra is the canonical measurable space on $\mathbb{R}$.

2.2.2 Probability Measures and Probability Spaces

So far, we've considered the construction of event spaces. The interpretation of these spaces is to view their members as indicating composite outcomes of some experiment (for example, the event that we measure the value of a pressure gauge between two specified values, or that we measure one of a collection of possible outcomes when rolling a die), but this structure alone is insufficient for assigning *probability*. To do this, we need a way of associating to a given event the likelihood of its outcome. To do this, we introduce the notion of a *probability measure*. We again specify it according to the rules it must satisfy. Let μ denote a probability measure. Then

(M1) $\mu : \mathcal{A}(\Omega) \to [0,1]$

(M2) For all $\omega \in \mathcal{A}(\Omega)$, $\mu(\omega) \ge 0$

(M3) $\mu(\Omega) = 1$

(M4) If $\{\omega_k\}$ is a (possibly infinite) collection of events in $\mathcal{A}(\Omega)$ such that for every pair of events $\omega_i,\ \omega_j \in \mathcal{A}(\Omega)$ it is the case that $\omega_i \cap \omega_j = \oslash$, then $\mu(\bigcup_k \omega_k) = \sum_k \mu(\omega_k)$

As with σ-algebras, one can work out some basic details from these rules. For example, one can show using (M2) and (M4) that $\mu(\oslash) = 0$. Recalling from the basic theory of probability, if an event ω has probability $p = \mu(\omega)$ then $\mu(\overline{\omega}) = 1-p$. We are in a position to *prove* this now from the definitions. Since this result is so useful, we show how to do this now. Consider an event $\omega \subseteq \Omega$. Then $\omega \cap \overline{\omega} = \oslash$. Then by (M4), consider the collection $\{\omega,\overline{\omega}\}$, we have $1 = \mu(\Omega) = \mu(\omega\cup\overline{\omega}) = \mu(\omega)+\mu(\overline{\omega}) = p+\mu(\overline{\omega})$. But then it immediately follows that $\mu(\overline{\omega}) = 1-p$. This just says that for any event which occurs with probability p, the probability of the event not occurring is $1-p$. There are other useful properties of measures. For example, they are *monotonic*. That is, if $\omega \subseteq \omega'$, then $\mu(\omega) \le \mu(\omega')$. To show this, use the fact that $\omega' = \omega \cup (\omega' - \omega)$[5], then apply (M4) and (M2). A *probability space* is an event space endowed with a probability measure. See [GW14] for additional basic properties.

We now show that this definition is a generalization of the usual definition of thinking of probability as the number of ways a specific event can occur relative to the number of all possible events. We can do this by showing that such a definition yields a probability measure in the above sense. We'll consider a discrete sample space, but this can be generalized to continuous sample spaces as well. If A is a set, let $|A|$ denote the number of elements in A, called its *cardinality*. By definition, $|\oslash| = 0$. Let D be our sample space, and $|D| = N$ be its cardinality. We'll consider the event space $\mathcal{P}(D)$. Define our probability measure to be

$$P(A) = \frac{|A|}{N}$$

Now we just need to check the various defining properties of the measure. (M1) Let $A \subseteq D$, then $|A| \le |D|$, which means that $|A|/N \le 1$. Since the cardinality of a set is always greater than or equal to 0, $A \mapsto P(A) = |A|/N$ sends A into $[0,1]$ for each $A \subseteq D$. Furthermore, $P(A)$ gives a single output for each A so that this is a well-defined function. (M2) We just pointed out that $|A| \ge 0$ for any discrete set A, so then $|A|/N \ge 0$ for every $A \in \mathcal{P}(D)$. (M3) We compute $P(D) = |D|/N = N/N = 1$. (M4) We only need to show that this holds for a pair of disjoint

[4] A more complete illustration would consider the other possible cases; the interested reader can work these out for himself.

[5] This is called the *relative complement* of ω with respect to ω'. Can you show that a σ-algebra has relative complements?

subsets of D since the set is finite, which means the general case follows by induction. Let A and B be disjoint subjects of D. Then $P(A \cup B) = |A \cup B|/N$. Since A and B are disjoint, they have no elements in common. That means $|A \cup B| = |A| + |B|$. Applying this here, we have $P(A \cup B) = |A \cup B|/N = |A|/N + |B|/N = P(A) + P(B)$, so that this property also holds. This means our usual idea of probability also holds in this case.

Finally, we discuss independent events and product spaces. Let $\langle \Omega, \mathcal{A}, \mu \rangle$ and $\langle \Omega', \mathcal{A}', \mu' \rangle$ be probability spaces. Then the *product space* of these two probability spaces is defined by $\langle \Omega, \mathcal{A}, \mu \rangle \times \langle \Omega', \mathcal{A}', \mu' \rangle = \langle \Omega \times \Omega', \mathcal{A} \times \mathcal{A}', \mu \times \mu' \rangle$. Here $\Omega \times \Omega'$ is the usual Cartesian product we defined above; the σ-algebra is just given by the products of the σ-algebras where events in $\mathcal{A}(\Omega) \times \mathcal{A}'(\Omega')$ are just ordered pairs of events, (ω, ω') and $\mu \times \mu' : (\omega, \omega') \mapsto \mu(\omega) \cdot \mu'(\omega')$. Whenever a probability space can be decomposed into a product space, the events in the respective factored spaces are called *(mutually) independent* and their probabilities are just given by the product of the respective probabilities.

2.2.3 Measurable Functions, Random Variables, and Distributions

A *measurable function* is a function $\xi : \Omega \to \mathbb{R}$ subjected to the condition that if $U \in \mathcal{B}(\mathbb{R})$ then $\xi^*(U) \in \mathcal{A}(\Omega)$. Recall that we mentioned in passing that a generic function induces a function between the power sets in the opposite direction. In this case, we are imposing an additional restriction that turns the function into a morphism. That is, we claim that $\xi^* : \mathcal{B}(\mathbb{R}) \to \mathcal{A}(\Omega)$ is a morphism. This means that we just need to show that ξ^* is structure-preserving in the sense that we outline above. By definition, (S1) is preserved so that we only need to consider (S2) - (S4). Because inverse images preserve set-algebraic operations, (S3) is satisfied. (S4) is satisfied for the same reason together with the principle of induction. To see (S2), observe that inverse images preserve intersections, and consider U, $\overline{U} \in \mathcal{B}(\mathbb{R})$. Then $\xi^*(\oslash) = \xi^*(U \cap \overline{U}) = \xi^*(U) \cap \overline{\xi^*(U)} = \oslash$. Since all other expressions in the σ-algebra can be expressed in terms of these facts, the map is structure preserving and a measurable function is a morphism between event spaces. A *random variable* is a measurable function from a space with a probability measure equipped. Schematically we can think about it like this

$$\begin{tikzcd}
\Omega \arrow[r, "\xi"] \arrow[d, "\overline{\mu}"'] & \mathbb{R} & \Longrightarrow & \mathcal{A}(\Omega) \arrow[d, "\mu"'] & \mathcal{B}(\mathbb{R}) \arrow[l, "\xi^*"'] \arrow[dl, "\mu\circ\xi^*"] \\
{[0,1]} & & & {[0,1]} &
\end{tikzcd}$$

so that the measure on Ω induces a measure on $\mathbb{R}$ through the inverse image. Whenever this diagram holds for a sample space Ω, the resulting map $\xi : \Omega \to \mathbb{R}$ is called a random variable. The composition $\mu \circ \xi^*$ is called a *distribution*.[6] A general distribution will be denoted $\delta : \mathcal{B}(\mathbb{R}) \to [0,1]$ with $U \mapsto \delta(U)$ when $U \subseteq \mathbb{R}$ and $\delta(U) \in [0,1]$.

There are two simple examples of distributions that are common: the first is the *uniform distribution*, and the second is the *Gaussian distribution*. These can be defined as

$$\begin{cases} D(x) = \begin{cases} \frac{1}{2a} & -a \leq x \leq a \\ 0 & \text{Otherwise} \end{cases} & \text{Uniform} \\ D(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_0)^2}{2\sigma^2}} & \text{Gaussian} \end{cases}$$

These are special cases of more general distributions, which we will touch on later. These two distributions have two key properties that make them particularly useful: (1) $D(x) = D(-x)$, so that they are symmetric about the origin; (2) $\lim_{x\to\infty} D(x) = 0$, so that they are bounded. These will be key properties when we discuss random walks and Brownian motion. The probability weights assigned by these distributions are

$$D(U) = \int_U d\mu(x) D(x)$$

[6] The common terminology here is to call this function a *(probability) density* and to refer to $\int_{-\infty}^{x} D(x')dx'$ as the *distribution*–the technical name of this latter object is the *cumulative distribution*.

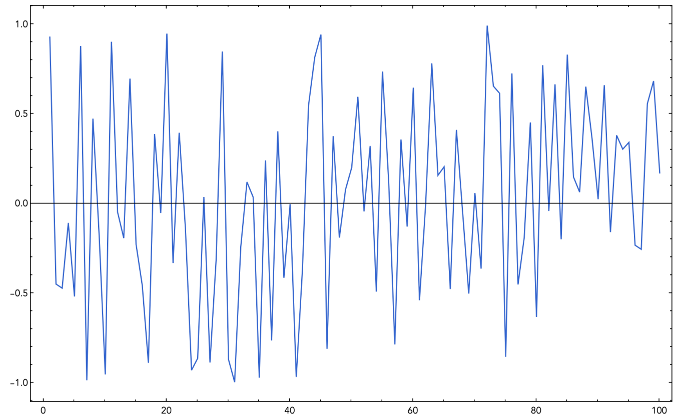

Figure 1: A uniformly distributed sequence of (pseudo)random variables with distribution of width 2.

For example, the weight assigned to $[x_1, x_2] \subseteq [-a, a]$ for the uniform distribution is just

$$D([x_1, x_2]) = \int_{x_1}^{x_2} \frac{dx}{2a} = \frac{x_2 - x_1}{2a}$$

which we would expect for a such a geometry.

2.2.4 Optional: Integration and Expectation

This is a pain in the ass.

3 Stochastic Processes

3.1 Basic Definitions and Overview

A *stochastic process* can now be defined to be a sequence of random variables. This can be conceptualized either as a discrete or continuous process, but the discrete process allows for building up of solid intuition to realize the continuous case. We will denote a (possibly infinite, discrete) stochastic process by the sequence $\langle \xi_1, ..., \xi_n, ... \rangle$, where the ξ_k are all (distinct) random variables. Notice that because each of these is a random variable, there is an associated sequence of distributions $\langle \delta_1, ..., \delta_n, ... \rangle$. Notice that this means we've implicitly made the distributions depend on the indexing parameter, which is typically regarded as time. If we consider a set of axes, (t, x), and we make $x(t)$ a random variable at each time t, then we can think of a distribution living at each time slice. This distribution may or may not depend on the previous time slices. In particular, this implies the possibility of a recursion relation (which in the continuous case will imply the existence of a differential equation). Let's assume it has the form

$$\delta_k = \sum_{j=1}^{k-1} a_j \delta_j + \sum_{j=1}^{k} \sum_{m=1}^{N} b_m^{(j)} \varepsilon_m^{(j)}$$

where the ε_m are additional (auxiliary) distributions which can also have contributions from previous time slices. In particular, if $\delta_k = \delta_{k-1}$, then $\delta_k - \delta_{k-1} = 0$ and the distribution does not depend on which time slice is chosen. Clearly the continuous analogue of this is $d(\delta)/dt = 0$. Making particular choices for the coefficients (which in our case are real-valued, but can be generalized) determines the construction of a specific stochastic process. For example, in the case that $\delta_k = \delta_{k-1} = \delta$ is the uniform distribution of width $2a$, then a sequence of random variables will essentially present a plot appearing like that of a *white noise process*. See Fig. 1

Analyzing in terms of the distributions may be more intuitive (specifically because it corresponds to a *definite* or *causal* description[7]), but it is less common. For example, *time series analysis* instead specifies the random

[7] This is actually the causal content of the theory of quantum mechanics. That theory just specifies differential equations which

variables appearing in a (discrete) stochastic process as linear combinations of other random variables. In that case, one has

$$\xi_k = \sum_{j=1}^{k-1} a_j \xi_j + \sum_{j=1}^{k} \sum_{m=1}^{N} b_m^{(j)} \eta_m^{(j)}$$

If we consider the special case where $a_j = 0$ for all $j \neq k-1$, then we can rewrite this equation

$$\Delta\xi_k = \xi_k - \xi_{k-1} = \sum_m c_m \eta_m$$

which allows us to think of $\sum_m b_m \eta_m$ as a sort of driving function for the random variable ξ_k. If the second term expresses itself like a difference in the random variables η_m (so in particular, it appears like $b_k(\eta_k - \eta_{k-1})$, then we have

$$\Delta\xi_k = c_k \Delta\eta_k \tag{3}$$

which is a finite-difference representation of a *stochastic differential equation*

$$d\xi(t) = c(t)d\eta(t)$$

In any case, the goal of time series analysis is to treat the form of the equations above as hypotheses which model the data (these are linear, but one can consider more exotic functions, clearly) and then apply statistical analysis to given data to determine whether a given form of relation between the random variables characterizes the relationship effectively. This process is called *regression analysis*. See for example [Ma07].

3.1.1 Example: Random Walks and Brownian Motion

One of the simplest stochastic processes we can consider is the *random walk*, which is a discrete representation of *Brownian motion*. We can present this in an intuitive way if we think about a one-dimensional object jostling back and forth about the origin, say. We can think about things like a box in the bed of a truck that moves back and forth while the truck drives on a bumpy road, dust particles floating around in the air (although in that case, there is also a drift term), or a leaf on a tree brach in gusty winds. Recalling our definition, we first need to specify the sample space. In this case, it would consist of the set of all displacements from the origin. These are real-valued, so in general we can think of the possible state space as just $\mathbb{R}$, with the associated Borel algebra describing the event space. We can infer our measure by considering what we want our distribution to be. In this case, any distribution which is symmetric about the origin and is bounded will be satisfactory. In particular, the Gaussian distribution is instructive for our purposes. Now if $\Delta x_k = x_k - x_{k-1}$ is the displacement the particle moves at the k^{th} time step, then the *position* of the particle at the k^{th} time step is just

$$x_k = \sum_{j=1}^{k} \Delta x_j$$

Any random variable satisfying this expression is called a *random walk*. Notice that the distribution which x_k satisfies can be reduced to the distribution on the Δx_j, which are all the same. The linearity of the expectation operator means that we can work just with the displacements themselves. For example, the average value of x_k can be calculated

$$E[x_k] = E\left[\sum_j \Delta x_j\right] = \sum_j E[\Delta x_j]$$

detail the time evolution of these distributions (with some caveats).

so that this reduces to finding the expectation of Δx_j. This is just

$$E[\Delta x_j] = \int_{-\infty}^{\infty} dx[\Delta x_j]\delta(x)$$

Now since we've chosen $\delta(x) = \delta(-x)$ and since $\Delta x_j \in \mathbb{R}$, this is the same as the integral

$$E[\Delta x_j] = \int_{-\infty}^{\infty} dx[x]\delta(x) = 0$$

which is easy to check (since it is an odd function over a symmetric interval). This is expected from a physical point of view–we are just as likely to observe the system to the left of the origin as we are to see it to the right of the origin. This means for all k $E[x_k] = 0$. Another interesting quantity to compute is the *variance*, which is defined as $\mathrm{Var}[x_k] = E[(x_k - E[x_k])^2]$. Since $E[x_k] = 0$, we have

$$\mathrm{Var}[x_k] = E[x_k^2] = E\left[\left(\sum_{j=1}^{k} \Delta x_j\right)^2\right] = \sum_{j_1, j_2} E[\Delta x_{j_1} \Delta x_{j_2}]$$

so that again we only need to work out $E[\Delta x_{j_1} \Delta x_{j_2}]$. We can observe that these two are independent of one another when $j_1 \neq j_2$, which means they decompose to a product of subspaces of the event space by $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. Then taking the expectation is just the product of the expectations of the individual variables, $E[\Delta x_{j_1} \Delta x_{j_2}] = E[\Delta x_{j_1}]E[\Delta x_{j_2}] = 0$. Now the special case of $j_1 = j_2$, we have

$$E[(\Delta x_{j_1})^2] = \int_{-\infty}^{\infty} dx[x^2]\delta(x) = \mathrm{Var}[\Delta x_j]$$

so that

$$\mathrm{Var}[x_k] = k\mathrm{Var}[\Delta x]$$

which means the average value remains 0 but the variance grows linearly in time. Compare Fig. 2. *Brownian motion* is just a generalization of this process to the case that the motion becomes continuous and the time steps become infinitesimal. In that case, we can put $\Delta x = v\Delta t$ then take the continuous limit. This defines the new random variable $v_k = \Delta x_k / \Delta t$ (assuming our time steps are uniform). Since Δt is a constant, the distributions of these two variables are the same, but only change by a factor. Then we obtain

$$x_k = \sum_j \Delta x_j \longrightarrow x(t) = \int_0^t dt'[v(t')]$$

All the results stay the same but are replaced by continuous counterparts. So $E[x(t)] = 0$ for all t and $\mathrm{Var}[x(t)] = t\mathrm{Var}[v]$.

References

[Bo06] M. L. Boas, *Mathematical Methods in the Physical Sciences*, John Wiley & Sons (2006).

[Go14] R. Goldblatt, *Topoi: The Categorial Analysis of Logic*, Elsevier (2014).

[Gr13] M. Grigoriu, *Stochastic Calculus: Applications in Science and Engineering*, Springer Science & Business Media (2013).

[GW14] G. Grimmett and D. Welsh, *Probability: An Introduction*, Oxford University Press (2014).

[Ha60] P. Halmos, *Naive Set Theory*, van Nostrand (1960).

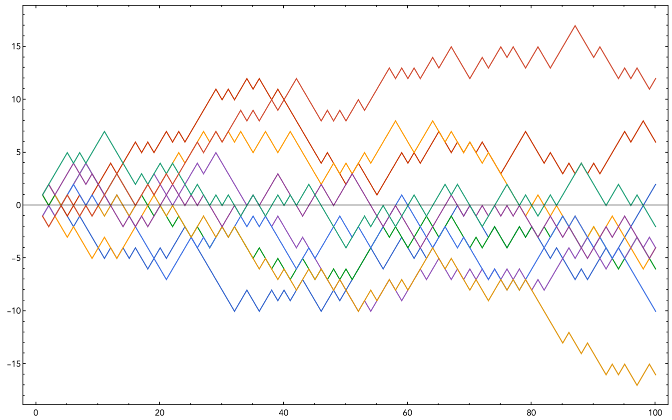

Figure 2: 10 different examples of a (pseudo)random walk with 100 time steps.

[Ma07] H. Madsen, *Time Series Analysis*, Chapman and Hall/CRC (2007).
[Mc92] C. McLarty, *Elementary Categories, Elementary Toposes*, Calrendon Press (1992).
[Pi10] C. C. Pinter, *A Book of Abstract Algebra*, Courier Corporation (2010).