Stochastic Processes

Rohan Hitchcock

31 August 2023

Our goal for the next few talks is to understand stochastic differential equations. For example, as a model of stochastic gradient descent [SST92, Equation 2.7] considers the weights $\mathbf{W}$ of a neural network — a random variable — evolving according to the equation

$$\frac{\partial \mathbf{W}}{\partial t} = -\nabla_{\mathbf{W}} L(\mathbf{W}) - \nabla_{\mathbf{W}} V(\mathbf{W}) + \eta(t) \tag{0.1}$$

where L is the loss function and $\eta(t)$ is "white noise". What does this mean? When say some dynamic process "evolving according to a (non-stochastic) differential equation" what we are really talking about is solutions to that differential equation. Does (0.1) have solutions, and if so how do we find them? More fundamentally, what does it mean for $\mathbf{W}$ to be a solution to (0.1)?

This talk has used two main references. For stochastic processes and Brownian motion we refer to [Øks13, Chapter 2] and for technical results about product measures we refer to [Tao11, Chapter 2.4]. Although not used directly in this talk, [Man13] gives an intuitive introduction to stochastic calculus on manifolds and will likely be referenced in the sequel to this talk.

1 What is a stochastic process?

In this talk we will focus on defining stochastic processes. Let M be a topological space equipped with its Borel topology $\mathcal{B}$, which will be the space in which our stochastic processes take values. Typically M will be a smooth manifold and in this talk nothing is lost by taking $M = \mathbb{R}^n$. Let T be any set, representing time for our stochastic process. Typically T will be some subset of $\mathbb{Z}$ or $\mathbb{R}$, and almost always will have a total order. Finally let $(\Omega, \mathcal{F}, \varphi)$ be a probability space.

Definition 1. An *M-valued stochastic process parametrised by T* is a function $X : T \times \Omega \to M$ such that the function $X_t : \Omega \to M$ is a random variable for all $t \in T$.

When T is also measurable space typically we will assume that a stochastic process $X : T \times \Omega \to M$ is measurable.

For each fixed $t \in T$ we have a random variable X_t, which by definition is a measurable function

$$X_t : \Omega \to M \qquad \omega \mapsto X_t(\omega) \ .$$

We could also fix an outcome $\omega \in \Omega$ and consider the function

$$\widetilde{\omega} : T \to M \qquad t \mapsto X_t(\omega) \ .$$

The function $\widetilde{\omega}$ is called a *path* of the stochastic process (X_t). The paths of a stochastic process are exactly analogous to the samples of a random variable.

Let M^T denote the set of all functions from T to M. The mapping $\omega \mapsto \widetilde{\omega}$ identifies Ω with a subset $\widetilde{\Omega} \subseteq M^T$ What do events look like in $\widetilde{\Omega}$? We consider the σ-algebra $\mathcal{C}$ generated by sets of the form

$$\left\{\widetilde{\omega} \in \widetilde{\Omega} \mid \widetilde{\omega}(t_1) \in B_1, \dots, \widetilde{\omega}(t_n) \in B_n\right\} \qquad \text{where each } B_i \subseteq M \text{ is measurable}$$

for $t_1, \dots, t_n \in T$. The σ-algebra $\mathcal{C}$ is contained in the σ-algebra $\widetilde{\mathcal{F}}$ by induced on $\widetilde{\Omega}$ by $\mathcal{F}$. Indeed,

$$\begin{aligned}\left\{\widetilde{\omega} \in \widetilde{\Omega} \mid \widetilde{\omega}(t_1) \in B_1, \dots, \widetilde{\omega}(t_n) \in B_n\right\} &\cong \{\omega \in \Omega \mid X_{t_1}(\omega) \in B_1, \dots, X_{t_1}(\omega) \in B_n\} \\ &\cong X_{t_1}^{-1}(B_1) \cap \dots \cap X_{t_n}^{-1}(B_n) \subseteq \mathcal{F} \ .\end{aligned}$$

So we can consider a stochastic process as the probability space $(\widetilde{\Omega}, \mathcal{C}, \widetilde{\varphi})$ where $\widetilde{\varphi}$ is induced from φ, or equivalently as the probability space $(M^T, \mathcal{C}', \widetilde{\varphi})$ where $\mathcal{C}' = \sigma(\mathcal{C} \cup \{\widetilde{\Omega}^c\})$ and $\widetilde{\varphi}(\widetilde{\Omega}^c) := 0$.

The stochastic process (X_t) can, in essence, be recovered by defining

$$\widetilde{X}_t : \widetilde{\Omega} \to M \qquad \widetilde{\omega} \mapsto \widetilde{\omega}(t)$$

for each $t \in T$. Indeed for each $t \in T$ and measurable $B \subseteq M$ we have

$$\begin{aligned}\mathbf{P}(\widetilde{X}_t \in B) &= \widetilde{\varphi}(\widetilde{X}_t^{-1}(B)) \\ &= \widetilde{\varphi}(\left\{\widetilde{\omega} \in \widetilde{\Omega} \mid \widetilde{\omega}(t) \in B\right\}) \\ &= \varphi(\{\omega \in \Omega \mid X_t(\omega) \in B\}) \\ &= \varphi(X_t^{-1}(B)) \\ &= \mathbf{P}(X_t \in B)\end{aligned}$$

and likewise for the joint distributions. This is not the only way to define a random variable on $(M^T, \mathcal{C}', \widetilde{\varphi})$, and in fact in some cases it is the *wrong* way. More on this later.

So, there is a relationship between stochastic processes and probability spaces on M^T. This perspective is very natural. The elements of M^T are trajectories in M parametrised by T and the choice of a probability measure on M^T amounts to choosing a distribution over these trajectories. This fits nicely with the ideas discussed in the first talk.

This is a useful perspective to have, so lets take a digression to discuss a sensible choice of σ-algebra on M^T. Given a function $f : T \to M$ and $t \in T$ we want to be able to talk about the event that $f(t) \in B$ for any measurable $B \subseteq M$. To formalise this, to each $t \in T$ we can consider the function which is evaluation at t:

$$\pi_t : M^T \to M \ , \qquad f \mapsto f(t) \ .$$

Given a measurable set $B \subseteq M$ we have

$$\pi_t^{-1}(B) = \left\{f \in M^T \mid f(t) \in B\right\}$$

so in other words, for all measurable $B \subseteq M$ and $t \in T$ we want $\pi_t^{-1}(B) \subseteq M^T$ to be measurable.

Definition 2. Given a measurable space $(M, \mathcal{B})$ and set T, the *product σ-algebra*[1] on M^T, which we denote $\mathcal{B}^T$, is the smallest σ-algebra containing all sets of the form $\pi_t^{-1}(B)$ where $t \in T$ and $B \subseteq M$ is measurable. In symbols:

$$\mathcal{B}^T = \sigma\left(\bigcup_{t \in T} \left\{\pi_t^{-1}(B) \mid B \in \mathcal{B}\right\}\right) \ .$$

[1]Also called the *cylindrical σ-algebra*.

It is natural to ask about the relationship between $\mathcal{B}^T$ and the Borel σ-algebra of the product topology on M^T. When the cardinality of T is at most countable these coincide, but in general $\mathcal{B}^T$ is only contained within the Borel σ-algebra. The product σ-algebra is relatively coarse; if $E \in \mathcal{B}^T$ then the characteristic function $\mathbf{1}(f \in E)$ can only depend on the value of f at countably many inputs.

A probability measure $\widetilde{\varphi}$ on $(M^T, \mathcal{B}^T)$ induces a stochastic process in the same way as above. In the next section we will see that specifying all finite joint distributions induces a unique probability measure on $(M^T, \mathcal{B}^T)$ which agrees with these joint distributions. Hence a typical recipe for constructing a stochastic process is:

- Specify the finite dimensional joint distributions of the process. I.e. for all $(t_1, \ldots, t_k) \in T^*$ specify the joint distribution of $(X_{t_1}, \ldots, X_{t_k})$.
- Use this to define a measure on $(M^T, \mathcal{B}^T)$, which in turn induces a stochastic process with the specified joint distributions.
- Modify the resulting stochastic process so it has any other desired properties.

The last step is necessary because the σ-algebra $\mathcal{B}^T$ cannot distinguish some properties we care about, such as continuity. Probability measures on $(M^T, \mathcal{B}^T)$ should be thought of as classifying stochastic processes up to equivalence of all finite dimensional joint distributions. More later.

2 Kolmogorov's extension theorem

Let $(M, \mathcal{B})$ be a measurable space and T a set. In this section we will discuss *Kolmogorov's extension theorem*, which allows us to uniquely define a probability measure on $(M^T, \mathcal{B}^T)$ — a stochastic process — from a tractable description. Specifically we will see that, under certain conditions, specifying a probability distribution on each $(M^F, \mathcal{B}^F)$ where $F \subseteq T$ is *finite* uniquely determines a distribution on $(M^T, \mathcal{B}^T)$.

Example 1. Take $M = \mathbb{R}$ and let $\mathcal{B}$ be the Borel σ-algebra. If F is a finite set then we can identify $\mathbb{R}^F \cong \mathbb{R}^n$ where $n = |F|$, and in this case $\mathcal{B}^F$ is the usual Borel σ-algebra on $\mathbb{R}^n$.

Recall the evaluation maps $\pi_t : M^T \to M$. Likewise, for each subset $S \subseteq T$ we can consider the map which restricts the domain of a function to S:

$$\pi_S : M^T \longrightarrow M^S \ , \quad f \longmapsto f|_S$$

where $f|_S$ is the function f with domain restricted to S. Given $S' \subseteq S$ we can likewise consider the restriction of functions from M^S to $M^{S'}$, which we denote by $\pi^S_{S'}$.

Lemma 3. *For any $S_1 \subseteq S_2 \subseteq S_3 \subseteq T$ we have $\pi^{S_3}_{S_1} = \pi^{S_2}_{S_1} \circ \pi^{S_3}_{S_2}$.*

Now lets consider how a measure φ on $(M^T, \mathcal{B}^T)$ interacts with the intermediate measurable spaces $(M^S, \mathcal{B}^S)$. For any $S \subseteq T$ we can define $\varphi_S(B) = \varphi(\pi_S^{-1}(B))$ where $B \subseteq M^S$ is measurable. It is easy to check that this is a measure on $(M^S, \mathcal{B}^S)$. Furthermore, given $S' \subseteq S$ we can consider the measure defined by

$$(\varphi_S)_{S'}(B') = \varphi_S((\pi^S_{S'})^{-1}(B'))$$

where $B' \subseteq M^{S'}$ is an event. By applying definitions we can see

$$\begin{aligned}(\varphi_S)_{S'}(B') &= \varphi_S((\pi^S_{S'})^{-1}(B')) \\ &= \varphi(\pi_S^{-1}(\pi^S_{S'})^{-1}(B')) \\ &= \varphi((\pi^S_{S'} \circ \pi_S)^{-1}(B')) \\ &= \varphi_{S'}(B')\end{aligned}$$

so $(\varphi_S)_{S'} = \varphi_{S'}$ and everything is works as we expect.

Definition 4. Let $S' \subseteq S$ and consider the measurable spaces $(M^{S'}, \mathcal{B}^{S'})$ and $(M^S, \mathcal{B}^S)$. Two measures ρ' and ρ on $M^{S'}$ and M^S respectively are *compatible* if $\rho_{S'} = \rho'$.

Now suppose that we also have a Hausdorff topology on M such that every compact set is measurable. In future applications this topology will always be 'nice' (eg. a complete metric space like $\mathbb{R}^n$), so there is no harm in thinking about M in this way now. To state the theorem we need to be able to state a technical condition on measures on M^F where $F \subseteq T$ is a finite subset. This condition will be essentially be automatically true for any topological space M we might be interested in.

Definition 5. A measure φ on M is *inner regular* if the measure of any measurable $B \subseteq M$ can be approximated by the measures of the compact sets it contains:

$$\varphi(B) = \sup\{\varphi(K) \mid K \subseteq B \text{ where } K \text{ is compact}\} \ .$$

Theorem 6 (Kolmogorov's Extension Theorem)**.** *Consider the set M^T of all functions $T \to M$ where T is any set and M is a measurable space with σ-algebra $\mathcal{B}$ and also a Hausdorff topological space where all compact sets are measurable. For each finite set $F \subseteq T$ consider the measurable space $(M^F, \mathcal{B}^F)$ and suppose we have a collection of measures $\{\rho_F \mid F \subseteq T \text{ finite}\}$ such that:*

(1) For any finite $F' \subseteq F$ the measures $\rho_{F'}$ and ρ_F are compatible in the sense of Definition 4.

(2) Each measure ρ_F on $(M^F, \mathcal{B}^F)$ is inner regular with respect to the product topology on M^F.

There is a unique measure φ on $(M^T, \mathcal{B}^T)$ which agrees with all the measures ρ_F. That is,

$$\varphi(\pi_F^{-1}(B)) = \rho_F(B)$$

for all finite $F \subseteq T$ and measurable $B \subseteq M^F$.

3 Brownian motion

We now use Kolmogorov's extension theorem to construct Brownian motion on $\mathbb{R}^n$, following [Øks13, Chapter 2.2]. In the notation of the above section we take $M = \mathbb{R}^n$ and $T = [0, \infty)$.

Fix $a \in \mathbb{R}^n$. This will be the starting point of our Brownian motion. Consider the function

$$p(t, x, y) = \frac{1}{(2\pi t)^{n/2}} \exp(\tfrac{-1}{2t}\|x - y\|) \qquad \text{where } x, y \in \mathbb{R}^n \ t > 0$$

which, for fixed x, is the n-dimensional Gaussian distribution with mean x and standard deviation $\sqrt{t}$. At $t = 0$ we define $p(0, x, y) = \delta_x(y)$ to be the Dirac delta distribution at x.

We want to define Brownian motion so that if we observe the particle at $x \in \mathbb{R}^n$, its position after a period of time Δt is given by a Gaussian distribution with mean x and standard deviation $\sqrt{\Delta t}$. Given times $\boldsymbol{t} = \{t_1 < t_2 < \cdots < t_k\} \subseteq [0, \infty)$ we define a measure $\rho_{\boldsymbol{t}}$ on $(\mathbb{R}^n)^k$ as

$$\rho_{\boldsymbol{t}}(B) = \int_B p(t_1, a, x_1)p(t_2 - t_1, x_1, x_2) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k)dx_1 \cdots dx_k$$

where $B \subseteq (\mathbb{R}^n)^k$ is measurable. One can show that this is a consistent sequence of probability measures (basically amounts to observing that $\int_{\mathbb{R}^n} p(t, x, y)dy = 1$) and so by Kolmogorov's extension theorem this induces a measure β which agrees with the above finite dimensional distributions. We call this the *Brownian motion measure*. This in turn asserts the existence of a stochastic process with the above finite dimensional joint distributions.

Theorem 7 (Kolmogorov's continuity theorem)**.** *Let (M, d) be a complete metric space and $(\Omega, \mathcal{F}, \varphi)$ a probability space. Let $X : [0, \infty) \times \Omega \to M$ be a stochastic process and suppose for all $t > 0$ there exist $\alpha, \beta, D > 0$ such that*

$$\mathbf{E}(d(X_{s_1}, X_{s_2})^\alpha) \leq D|s_1 - s_2|^{1+\beta} \qquad \text{for all } 0 \leq s_1 < s_2 \leq t\ .$$

Then there exists a stochastic process $X' : [0, \infty) \times \Omega \to M$ as above such that:

(1) For almost all $\omega \in \Omega$ the function $t \mapsto X'_t(\omega)$ is continuous.

(2) For all $t \geq 0$ we have $X_t = X'_t$ (almost surely).

Using this theorem, one can show that there is a version of Brownian motion which is almost surely continuous. This is called *canonical Brownian motion*.

4 Some wrinkles

The perspective of a stochastic process as a measure on $(M^T, \mathcal{C})$ where $\mathcal{C}$ is appealing: a stochastic process is simply distribution over all possible paths in our manifold M. Unfortunately the σ-algebra $\mathcal{B}^T$ is not sufficiently expressive for our purposes.

In the Brownian motion case: $M = \mathbb{R}^n$, $T = [0, \infty)$ and β the Brownian motion measure on M^T, it would be <u>nice</u> if we could reframe Kolmogorov's continuity theorem as saying that

$$\text{“}\beta(C(T, M)) = 1\text{”}$$

where $C(T, M)$ is the set of all continuous functions from T to M. Unfortunately $C(T, M)$ is not a $\mathcal{B}^T$-measurable set! Instead, Kolmogorov's continuity theorem turns up in *how* we associate a function $B : T \times \Omega \to M$ to the measure β, where $\Omega = M^T$. We now refer to the MathOverflow answer [MO]. Let $Q = T \cap \mathbb{Q}$, the important property of Q being that it is countable and dense in T. Define the set of functions

$$U = \left\{\omega \in M^T \mid \omega|_Q \text{ is uniformly continuous on bounded sets}\right\}$$

One can show that U is $\mathcal{B}^T$-measurable. For all $\omega \in U$ we can define

$$B(t, \omega) = \begin{cases} \omega(t) & t \in Q \\ \lim^Q_{s \to t} \omega(s) & t \notin Q \end{cases}$$

where $\lim_{s\to t}^{Q} \omega(s)$ denotes the limit taken within Q. The content of Kolmogorov's continuity theorem in this context is that $\beta(U) = 1$, and so for $\omega \notin U$ we can choose $B(t, \omega)$ to be anything. The map $t \mapsto B(t, \omega)$ is continuous for all $\omega \in U$ and so the resulting process is almost surely continuous.

A better approach, in my opinion, would be to use a finer σ-algebra on M^T which contains $C(T, M)$ (and other relevant function spaces) and try to construct a measure on that space directly. Or, perhaps, to replace M^T with $C(T, M)$. This way one could *define* a stochastic process as a distribution over paths in M.

References

[Man13] Jonathan H. Manton. 'A Primer on Stochastic Differential Geometry for Signal Processing'. *IEEE Journal of Selected Topics in Signal Processing* 7.4 (Aug. 2013), pp. 681–699. arXiv: `1302.0430 [math]`.

[MO] Nate Eldredge (https://mathoverflow.net/users/4832/nate-eldredge). 'Continuity of Brownian motion constructed from Kolmogorov extension theorem?' MathOverflow. URL: `https://mathoverflow.net/q/354715`.

[Øks13] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. 6th ed. Universitext. Springer, 2013.

[SST92] Hyunjune Sebastian Seung, Haim Sompolinsky and Naftali Tishby. 'Statistical Mechanics of Learning from Examples'. *Physical review A* 45.8 (1992), p. 6056.

[Tao11] Terence Tao. *An Introduction to Measure Theory*. Graduate Studies in Mathematics. American Mathematical Society, 2011.