1. Foundations of stochastic processes

MATH 275D is a course in stochastic calculus, whose prime concepts of interest are the stochastic integral and stochastic differential equations. Both of these are useful tools in the general theory of stochastic processes. We therefore start by explaining how stochastic processes are defined in the standard model of probability theory.

1.1 Stochastic processes in the Kolmogorov model.

We will always assume the standard setting for probability, which is the Kolmogorov model rooted in measure theory. Every probability setting or statement thus assumes, often tacitly, an underlying *probability space*, defined as follows:

Definition 1.1 *A probability space is a triplet $(\Omega, \mathcal{F}, P)$ with the following properties:*

(1) *Ω is a set representing, at least in our imagination, the collection of all potential outcomes of a random experiment that we are concerned with,*

(2) *$\mathcal{F}$ is a σ-*algebra *on Ω, which means that $\mathcal{F}$ is a collection of subsets of Ω that contains $\varnothing$ and Ω and is closed under countable unions,*

$$\forall \{A_n\}_{n\geqslant 1} \subseteq \mathcal{F}\colon \quad \bigcup_{n\geqslant 1} A_n \in \mathcal{F} \tag{1.1}$$

and complements,

$$\forall A \in \mathcal{F}\colon \quad A^{\mathrm{c}} := \Omega \smallsetminus A \in \mathcal{F} \tag{1.2}$$

(3) *P is a* probability measure *on $(\Omega, \mathcal{F})$, which means that P is a map $P\colon \Omega \to [0,1]$ that is* countably additive *in the sense*

$$\forall \{A_n\}_{n\geqslant 1} \subseteq \mathcal{F} \text{ disjoint}\colon \quad P\Big(\bigcup_{n\geqslant 1} A_n\Big) = \sum_{n\geqslant 1} P(A_n) \tag{1.3}$$

and normalized so that $P(\Omega) = 1$. (This is indeed a normalization condition since countable additivity implies monotonicity of $A \mapsto P(A)$ under the set inclusion.)

Elements of $\mathcal{F}$ are referred to as *events*; these are the "questions" that we are allowed to ask about our random experiment. We note that various σ-algebras may be assumed on the same probability space. The finer (meaning, larger) is the σ-algebra, the more events — and thus more accessible information — it contains. This underlies the following natural concept:

Definition 1.2 (Filtration) *Given a measurable space $(\Omega, \mathcal{F})$ and a linearly ordered set S, a* filtration *on $(\Omega, \mathcal{F})$ indexed by S is a family $\{\mathcal{F}_t\}_{t\in S}$ such that*

(1) $\forall t \in S\colon$ *$\mathcal{F}_t$ is a σ-algebra* $\wedge$ $\mathcal{F}_t \subseteq \mathcal{F}$

(2) $\forall t, s \in S\colon$ $t \leqslant s \Rightarrow \mathcal{F}_t \subseteq \mathcal{F}_s$

The most natural choices for the index set S are $S := \mathbb{N}$, which is used mainly for discrete-time processes, and $S := [0, \infty)$, which is pertinent to continuous-time processes. A number of powerful limit theorems are associated with filtrations; we will discuss these when they become relevant.

An important concept of probability is that of a *random variable* which, intuitively, refers to a quantity that takes random values. The precise mathematical meaning of this concept is nonetheless rather dry:

Definition 1.3 (Random variable) *Let $(\mathscr{X}, \Sigma)$ be a measurable space — meaning that $\mathscr{X}$ is a non-empty set and Σ is a σ-algebra on $\mathscr{X}$. An $\mathscr{X}$-valued random variable X is then a map $X\colon \Omega \to \mathscr{X}$ which is $\mathcal{F}/\Sigma$-*measurable *in the sense that*

$$\forall B \in \Sigma\colon \quad X^{-1}(B) \in \mathcal{F}. \tag{1.4}$$

Here $X^{-1}(B) := \{\omega\colon X(\omega) \in B\}$.

The most typical choice is $\mathscr{X} := \mathbb{R}$, which we simply refer to as a random variable, or $\mathscr{X} := \mathbb{R}^d$, when we talk of a *random vector*. In both of these cases we take Σ to be the σ-algebra of *Borel sets*, which is defined as the smallest σ-algebra containing all open sets. (This is achieved by intersecting all the σ-algebras with this property while noting that the intersection of any non-empty family of σ-algebras is a σ-algebra.) Notwithstanding, as we shall see, the above framework permits talking seamlessly about random functions, random measures or random linear operators with no difficulty.

The measurability requirement is needed so that "X takes values in B," abbreviated using set theoretical notation as $\{X \in B\}$ where

$$\{X \in B\} := X^{-1}(B), \tag{1.5}$$

are events for any measurable set $B \in \Sigma$. In particular, we are allowed to write expressions such as $P(X \in B)$, read as " probability that X lies in B." As is readily checked,

$$\sigma(X) := \big\{\{X \in B\}\colon B \in \Sigma\big\} \tag{1.6}$$

is a σ-algebra as soon as Σ is a σ-algebra. It can also be checked that $\sigma(X)$ is the least σ-subalgebra — or, equivalently, the intersection of all σ-subalgebras — of $\mathcal{F}$ that make X measurable. (The corresponding concept exists for multiple random variables as well; we will discuss this in Chapter 3 of these notes.)

An additional benefit of requiring measurability is that it enables the powerful theory of the *Lebesgue integral*. While measurability is often deemed typical for functions considered in analysis, the fact that probability regards the underlying σ-algebra as a variable entity makes measurability easy to break (as well as restore).

We are now ready for the first important new concept in this course:

Definition 1.4 (Stochastic process) *Let $(\mathscr{X}, \Sigma)$ be a measurable space and T a non-empty set. An $\mathscr{X}$-*valued stochastic process over T *is a family $\{X_t\colon t \in T\}$ of $\mathscr{X}$-valued random variables (defined on the same probability space).*

One way to think of a stochastic process is as a *random function* $X\colon T \to \mathscr{X}$ defined, for each $\omega \in \Omega$, as $t \mapsto X_t(\omega)$. However, some care is needed as not all the values of X may be accessible at the same time due to measurability limitations. (We will explain this in full depth in the next two lectures.)

A typical choice we will make is $\mathscr{X} := \mathbb{R}$ and $T := [0, \infty)$. In this case we simply talk about a stochastic process with t interpreted as *time*. (This is the setting where filtrations

are most naturally observed.) The above definition nonetheless includes a way to treat more complicated objects such as a *random field*, when $T := \mathbb{R}^d$ or a subset thereof, or even a *random measure*, in which case T is a σ-algebra and $A \mapsto X_A$ is countably additive in the sense specified above.

1.2 Key examples: Poisson process and Brownian motion.

We will now give two examples of stochastic processes that will be periodically returned to throughout the course. The first one of these is:

Definition 1.5 (Homogeneous Poisson process) *A* homogeneous Poisson process *is an* $\mathbb{N}$*-valued stochastic process* $\{N_t : t \in [0,\infty)\}$ *on the non-negative reals such that*

(1) $N_0 = 0$
(2) $\forall n \geqslant 1\, \forall 0 = t_0 < t_1 < \cdots < t_n$:
$$\{N_{t_i} - N_{t_{i-1}} : i = 1, \ldots, n\} \text{ are independent} \tag{1.7}$$
(3) $\forall 0 \leqslant s < t$: $N_t - N_s = \text{Poisson}(t-s)$
(4) *every realization of* $t \mapsto N_t$ *is right-continuous with left limits.*

We use the convention that the naturals $\mathbb{N}$ contain zero. Recall that a random variable Z is Poisson with parameter $\lambda > 0$ if it is $\mathbb{N}$-valued and

$$\forall n \in \mathbb{N}: \ P(Z = n) = \frac{\lambda^n}{n!} \mathrm{e}^{-\lambda} \tag{1.8}$$

We also recall that random variables $X_1, \ldots X_n$ are *independent* if for all $B_1, \ldots, B_n \in \Sigma$,

$$P\left(\bigcap_{i=1}^{n} \{X_i \in B_i\} \right) = \prod_{i=1}^{n} P(X_i \in B_i) \tag{1.9}$$

Intuitively, sampling one of several independent random variables does not affect the statistics of the values of the others.

By one of the standard interpretations, the quantity N_t models the number of customers arriving to the queue (or served) by time t (with no customers at time zero). The process is piecewise constants; all jumps are by $+1$ a.s. The times of the jumps are exactly the times when new customers arrive. The requirement in 4) technically means that, for all $\omega \in \Omega$, $t \mapsto N_t(\omega)$ is right-continuous with left limits (to be abbreviated RCLL in the sequel). At this point, this requirement is merely a convention.

The word "homogeneous" refers to the fact that the distribution of the law of the *increment* $N_t - N_s$ depends only on $t - s$. Another expression for this fact is that N has *stationary increments*. To depart from homogeneity, one replaces $t - s$ by $G(t) - G(s)$ for $G : [0,\infty) \to [0,\infty)$ a non-decreasing RCLL function. Customers then arrive according to the *intensity measure* $\mathrm{d}G$, which for the homogeneous process is simply the Lebesgue measure on $[0,\infty)$. All these considerations are cosmetic on a technical level; indeed, the inhomogeneous process is realized from the homogeneous one by taking $\{N_{G(t)} : t \in [0,\infty)\}$.

Another important example of a stochastic process that we will be concerned with in this course is *Brownian motion*. The name dates back to the observations by biologist

R. Brown in the 1820s of a jittery motion of pollen grains caused, as elucidated by A. Einstein in 1905, by collisions with molecules of an ambient fluid. The probabilistic model we will now present captures only some of the features of this problem; the full physics theory of this motion is rather complex and remains the subject of interesting mathematical research to the present day; see E. Nelson's "Dynamical Theories of Brownian Motion" for an exposition of this problem.

The mathematical concept of Brownian motion goes back to N. Wiener (and is therefore sometimes alternatively referred to as the *Wiener process*):

Definition 1.6 (Brownian motion) *A* standard Brownian motion *is a stochastic process* $\{B_t\colon t\in[0,\infty)\}$ *on the non-negative reals such that*

(1) $B_0 = 0$
(2) $\forall n \geqslant 1\, \forall 0 = t_0 < t_1 < \cdots < t_n$:

$$\big\{B_{t_i} - B_{t_{i-1}}\colon i = 1,\dots,n\big\} \text{ are independent} \tag{1.10}$$

(3) $\forall 0 \leqslant s < t\colon B_t - B_s = \mathcal{N}(0, t-s)$
(4) *every realization of* $t \mapsto B_t$ *is continuous.*

Here $\mathcal{N}(\mu,\sigma^2)$ denotes the *normal* (or *Gaussian*) random variable with mean μ and variance σ^2. For $\sigma^2 > 0$ this ($\mathbb{R}$-valued) random variable has a *probability density* (with respect to the Lebesgue measure) which means that, for any Borel $B \subseteq \mathbb{R}$,

$$P\big(\mathcal{N}(\mu,\sigma^2)\in B\big) = \int_B \frac{1}{\sqrt{2\pi\sigma^2}}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x. \tag{1.11}$$

In the degenerate case of zero variance, we have $\mathcal{N}(\mu,0) = \mu$ with probability one.

Both the homogeneous Poisson process and the standard Brownian motion are examples of processes with *independent increments*. If in addition the distribution of the increment only depends only the difference of the time, we talk about a *Lévy process*. Any Lévy process $\{X_t\colon t\in[0,\infty)\}$ thus obeys that X_t has the law of the sum of n independent copies of $X_{t/n}$ which means that the law of X_t is *infinitely divisible*. The Lévy-Khinchin formula then implies existence of numbers $\mu\in\mathbb{R}$ and $\sigma^2 \geqslant 0$ and a Borel measure λ on $\mathbb{R}$ subject to conditions $\lambda(\{0\}) = 0$ and $\int \frac{x^2}{1+x^2}\lambda(\mathrm{d}x) < \infty$ such that

$$E\big(\mathrm{e}^{\mathrm{i}zX_t}\big) = \exp\Big\{\mathrm{i}\mu z t - \frac{1}{2}\sigma^2 z^2 t + t\int\Big(\mathrm{e}^{\mathrm{i}zx} - 1 - \frac{\mathrm{i}zx}{1+x^2}\Big)\lambda(\mathrm{d}x)\Big\} \tag{1.12}$$

holds for all $t \geqslant 0$. In this context we call λ the *Lévy measure* of the process. The three terms in the exponent are then interpreted as a constant-speed term, Brownian term and a compound Poisson process, with the latter two independent of each other.

Besides the above definition which is based on independence of increments, the standard Brownian motion is also an example of a *Gaussian process*, due to the multivariate-normal nature of its finite dimensional distributions. This permits an alternative formulation of conditions 2) and 3) by way of covariance structure of the process, to be discussed in a homework assignment. What makes the Brownian motion "standard" is the condition 1), which forces B_t to be centered and 3), where the variance is equal $t-s$,

not just proportional to it. Changing to $B_0 := a$ and $B_t - B_s = \mathcal{N}(0, \sigma^2(t-s))$ results in a process that we refer to plainly as a Brownian motion.

The conditions 4) in the above definitions are hard to motivate *a priori*. The reasons for imposing them is that it is precisely these conditions that allow us to think of the realizations $t \mapsto N_t(\omega)$ and $t \mapsto B_t(\omega)$ as random functions. Moreover, the conditions are needed if we want to prove that the above processes, once constructed, are unique, in a suitable sense. Again, we will say more in the forthcoming lectures.

Further reading: Chapter 2 in Øksendal, Chapter 1.1 of Karatzas-Shreve