Lecture 5, 14.10.2022

Ch. 1: Stochastic Processes: Foundations

Filtrations

Much of Ch. 0 is *static*. We turn now to its dynamic counterpart, typically where randomness unfolds with time (as in life!)

Stochastic processes occur naturally even in finite situations. An example is in *mathematical finance*, when time is measured discretely (every day, say, or every hour – even every second or micro-second, depending on context), and space – here price (the value of a stock, say, to the nearest cent/penny, or (for interest rates and other percentages) basis point (bp – a hundredth of 1%). If the stock price at time n ($n = 0, 1, \cdots, N$) is S_n, we may be required to price *options* on the stock over time. Here we regard $S := (S_n)_{n=0}^{N}$as a stochastic process. Although everything here is finite (including the sample space Ω in the probability space needed to describe the model), a stochastic-process view point is needed here. The mathematics is fairly recent (1970s – e.g., the *Black-Scholes formula* of 1973), non-trivial, and of obvious practical importance. See e.g. [BinK] (above), Ch. 4.

Usually, however, the time-set will be infinite, e.g. $\mathbb{N}_0 := 0 \cup \mathbb{N}$, $\mathbb{R}_+ := [0, \infty)$, or some interval $I := [t_0, t_1]$. The 'space variable' (set of values taken) is also usually infinite, e.g. $\mathbb{Z}, \mathbb{N}, \mathbb{R}$.

Note.

1. Turning to the typical case where both time-set and value-set are infinite: we have an immediate split for each, between countable and uncountable. If both are countable, everyone calls a process with the Markov property (below) a *Markov chain*. Chung's classic (cited above) reserves the term *chain* for when the *state* variable is discrete, and divides his book between Part I: Discrete Parameter (recall the parameter $i \in I$ corresponds to time in Ch. 0) and Part II: Continuous Parameter. By contrast, for Revuz ([Rev], p. 13) a chain is in discrete time, whether state is discrete or continuous. It is more usual to use the term *Markov process* for the continuous-state case, and we shall follow this.
2. Recall that countability is built into Measure Theory because of the *countable additivity* property of measures. So we must expect extra difficulties when we pass from discrete to continuous, in either space or time, as the setting there is *uncountable*.

We begin in discrete time, with a random phenomenon producing, at each time $n = 0, 1, 2, \cdots$ (say) a random variable X_n. The information available to us at time n is the set of values $\{X_0, X_1, \cdots, X_n\}$. We can consider the *conditional* distribution of X_{n+1}, or X_m for $m > n$, *given* this information. Note that as n increases, the information available to us increases also.

Consider all events involving $\{X_0, X_1, \cdots, X_n\}$ – that is, all events of the form $\{(X_0, X_1, \cdots, X_n) \in A\}$, for A a (measurable) subset of $\mathbb{R}^{n+1}$ ('measurable' just means that the probability $P((X_0, X_1, \cdots, X_n) \in A)$ is defined). This class of events is called the σ-field *generated* by $(X_0, X_1, \cdots, X_n)$, written $\sigma(X_0, X_1, \cdots, X_n)$. It should be thought of as 'the information contained in $X_0, X_1, \cdots, X_n$', or 'what we know when we know $X_0, X_1, \cdots, X_n$'. For, knowing $X_0, X_1, \cdots, X_n$ we know exactly *which* of these events

$$\{(X_0, X_1, \cdots, X_n) \in A)\} \qquad (A \in \mathcal{B}(\mathbb{R}^{n+1}))$$

occur. We write

$$\mathcal{F}_n := \sigma(X_0, X_1, \cdots, X_n)$$

('$\mathcal{F}$ for field': recall we use script capitals for classes of sets). Then as the class of sets $\{(X_0, X_1, \cdots, X_n) \in A)\}$ increases with n,

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} :$$

this just says that we learn more as time progresses and new information becomes available.

Definition (P.-A. Meyer, 1970s). A *filtration* is an increasing family $\{\mathcal{F}_n\}$ of σ-fields.

Here $\mathcal{F}_\infty := \cup_{n=0}^{\infty} \mathcal{F}_n$ is also a σ-field (check). If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{F}_n,\ \mathcal{F}_\infty \subset \mathcal{F}$, then $(\Omega, \{\mathcal{F}_n\}, \mathcal{F}, \mathbb{P})$ is called a *filtered probability space.*

If $X = \{X\}_n$ is a stochastic process, its *natural filtration* is $\{\mathcal{F}_n\}$ with $\mathcal{F}_n := \sigma(X_0, X_1, \cdots, X_n)$. We will usually be concerned with natural filtrations in this course.

If X_n is $\mathcal{F}_n$-measurable for each n, the stochastic process $X = \{X\}_n$ is said to be *adapted* to this filtration. We will *always* deal with adapted filtrations in this course.

Just as a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a mathematical model for a *static* random experiment, so a filtered probability space $(\Omega, \{\mathcal{F}_n\}, \mathcal{F}, \mathbb{P})$ is a mathematical model for a *dynamic* random experiment. It can serve as a basis for the mathematical description of any stochastic process $\{X_n\}$ adapted to it – that is, to $\{\mathcal{F}_n\}$: X_n must be $\mathcal{F}_n$-measurable for each n. The filtered probability space may – should – be regarded as part of the definition of the stochastic process $\{X_n\}$. Think of it as the space on which the stochastic process 'lives'.

Now that we are reassured by the Daniell-Kolmogorov theorem that all the stochastic processes we need *exist*, and have filtered probability spaces on which they live, we will often (usually; whenever possible) omit explicit reference to the filtered probability space. We need to know that it's there, but don't need (or want) to see it normally (the analogy with underwear comes to my mind here; if you find that helpful, fine; if not, also fine – *but then think of your own analogy*, or the subject will seem too abstract, and it needn't, and shouldn't).

Now that we know what a stochastic process is and 'where one lives', we will usually drop the 'stochastic', and leave it to be understood from context. From now on, 'process' *means* 'stochastic process' (again: if this bothers you, fine – go on writing 'stochastic' in every time – unless/until you realise you no longer need it).

Dependence and independence

We have lots of experience of dealing with independence – in your first course(s) on Probability and (implicitly) on Statistics (the different readings in a sample being typically assumed to be independent copies from the same distribution).

Note on errors in Statistics

Independent errors tend to cancel. This is the essence of the Laws of Large Numbers (Week and Strong); see MATH70028/M4P6 PROBABILITY (Dr I. Krasowsky, Term 2).

It is also what makes Statistics work. This is basically why large samples (though more expensive to gather and analyse) are better than small ones – there is more cancellation). This is *not true* for dependent errors. *Correlated errors in Statistics are very dangerous.*

Example.

Imagine a Physics lesson at school, in which an experiment is to be performed. The teacher divides the class into 10 pairs, and then goes into his back room to catch up on exam marking. The two best experimenters are in the same pair; the others gang up on them and force them to do the experiment for them, while they revise for their exam. A physical constant is to be measured, correct to several significant figures. Unfortunately the instrument the 'good' pair are using reads way too high (it was dropped that morning; the culprit did not own up). When the 'good' pair have their result, the others copy the first three significant figures, but attempt to disguise their cheating by each of the 9 pairs inventing their own 'nonsense figures' after that. These, being independent, *will* tend to cancel. But the first three significant figures will not: they will be way too high for all 10 pairs. There is *no* cancellation there, only *replication* of a wrong result.

Independence is much easier to handle than dependence. It is here that the three classical limit theorems, LLN (Law of Large Number, Weak and Strong), CLT (Central Limit Theorem) and LIL (Law of the Iterated Logarithm) find their simplest forms. Beyond that, the two main areas in which *dependence* can be handled are *Markov* chains and processes (this course) and *martingales* (M4P6 next term again). The other main areas where much can be said are *weak dependence* (mainly a hierarchy of *mixing* conditions), *stationarity*, and *Gaussianity*.