pfsl23(14).tex
Lecture 23. 2.12.2014

VI. STOCHASTIC PROCESSES

1. Filtrations; Finite-dimensional Distributions

For a random variable X, we call the σ-field generated by the inverse images (events) $X^{-1}(B) = \{\omega : X(\omega \in B\}$ as B varies over Borel sets B, the σ-field *generated by* X, written $\sigma(X)$. Equivalently, $\sigma(X)$ is the σ-field generated by the events $\{X \in I\}$ as I ranges over the intervals, or by the events $\{X \leq x\}$ as x varies. We quote

Doob's Lemma (see e.g. SP, L9). For two random variables X, Y, $\sigma(X) \subset \sigma(Y)$ iff $X = f(Y)$ for some measurable function f.

This gives us the right way to think about the (at first sight rather abstract) concept of a σ-field $\sigma(X)$: it represents the *information contained in* X. For, when we apply a function f, we in general *lose information*. There is no loss of information iff the function f is injective (one-to-one), i.e. the inverse function f^{-1} exists, so we can recover all previous information by applying f^{-1}.

(*Example*: $f(x) := x^2$: this is injective on $(0, \infty)$, but not on $(-\infty, \infty)$: when we take square roots, we introduce a sign ambiguity.

A *stochastic process (SP)* is a mathematical model of a random phenomenon unfolding with time. So for each t we have a random variable, X_t, and we have our current information, represented by a σ-field, $\mathcal{F}_t$, say. Bear in mind the *arrow of time*! We make the assumption that as time progresses, new information arrives, and no information is lost. (Of course, this is an idealisation! In real life, information is lost, by forgetting, and humanity finds itself endlessly having to re-invent the wheel, as it were, but we ignore this here for simplicity.) The set of these $\mathcal{F}_t$, which increase with time t as above, models our *information flow*. Following P.-A. MEYER (1934-2003), we call

$$\{\mathcal{F}_t\}_{t \geq 0}$$

a *filtration*. Adding a probability space, we obtain a *filtered probability space*

$$(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, \mathcal{P}).$$

We assume Meyer's *usual conditions* (conditions habituelles):
a. completeness: each $\mathcal{F}_t$ contains all P-null sets of $\mathcal{F}$;

b. the filtration is right-continuous, i.e. $\mathcal{F}_t = \mathcal{F}_{t+} := \cap_{s>t}\mathcal{F}_s$.

The alternative (and nowadays preferred) term for a filtered probability space is a *stochastic basis*, so called because it provides us with the basis on which to define a stochastic process – to which we now turn.

Definition. A *stochastic process* $X = (X(t))_{t\geq 0}$ is a family of random variables defined on a stochastic basis. Call X *adapted* if $X(t) \in \mathcal{F}_t$ (i.e. $X(t)$ is $\mathcal{F}_t$-measurable) for each t: thus $X(t)$ is known when $\mathcal{F}_t$ is known, at time t.

If $\{t_1, \cdots, t_n\}$ is a finite set of time points in $[0, \infty)$, $(X(t_1), \ldots, X(t_n))$ is a random n-vector, with a distribution, $\mu(t_1, \ldots, t_n)$ say. The class of all such distributions as $\{t_1, \ldots, t_n\}$ ranges over all finite subsets of $[0, \infty)$ is called the class of all *finite-dimensional distributions* of X. These satisfy certain obvious consistency conditions:

DK1. deletion of one point t_i can be obtained by 'integrating out the unwanted variable', as usual when passing from joint to marginal distributions;

DK2. permutation of the times t_i permutes the arguments of the measure $\mu(t_1, \ldots, t_n)$ on $\mathbb{R}^n$ in the same way.

Conversely, a collection of finite-dimensional distributions satisfying these two consistency conditions arises from a stochastic process in this way (this is the content of the *Daniell-Kolmogorov theorem*). This classical result (due to P.J. DANIELL (1889-1946) in 1918 and Kolmogorov in 1933) is the basic existence theorem for stochastic processes. For the proof, see e.g. [K].

Important though it is as a general existence result, however, the Daniell-Kolmogorov theorem does not take us very far. It gives an SP X as a random function on $[0, \infty)$, i.e. a random variable on $\mathbb{R}^{[0,\infty)}$. This is a vast and unwieldy space. We want to work on much smaller and more manageable spaces, of functions satisfying regularity conditions, such as continuity: we want to be able to realize $X = (X(t, \omega))_{t\geq 0}$ as a random continuous function, i.e. a member of $C[0, \infty)$. Such a process X is called *path-continuous* (since the map $t \to X(t, \omega)$ is called the sample path, or simply path, given by ω) – or more briefly, *continuous*. This is possible for the extremely important case of Brownian motion, for example (VI.1). Sometimes we need to allow $X(t, \omega)$ to have jumps. It is then customary, and convenient, to require $X(t)$ to be right-continuous with left limits (RCLL), or càdlàg (*continu à droite, limite à gauche*) – i.e. to have X in the space $D[0, \infty)$ of all such functions (the Skorohod space). This is the case, for instance, for the Poisson process and its relatives (Lévy processes – Ch. VI).

General results on realisability – whether or not it is possible to realize, or obtain, a process so as to have its paths in a particular function space –

are known, but for us it is usually better to construct the processes we need directly on the function space on which they naturally live.

Given an SP X, it is sometimes possible to improve the regularity of its paths without changing its distribution (that is, without changing its finite-dimensional distributions). For background on such results (separability, measurability, versions, regularization etc.) see e.g. [D].

There are several ways to define 'sameness' of two processes X and Y.
(i) X and Y have the *same finite-dimensional distributions* if, for any integer n and $\{t_1, \cdots, t_n\}$ a finite set of time points in $[0, \infty)$, the random vectors $(X(t_1), \ldots, X(t_n))$ and $(Y(t_1), \ldots, Y(t_n))$ have the same distribution;
(ii) Y is a *modification* of X if, for every $t \geq 0$, we have $P(X_t = Y_t) = 1$.

A process is called *progressively measurable* if the map $(t, \omega) \mapsto X_t(\omega)$ is measurable, for each $t \geq 0$. Progressive measurability holds for adapted processes with right-continuous (or left-continuous) paths – as here.

A random variable $\tau : \Omega \to [0, \infty]$ is a *stopping time* if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

For a set $A \subset \mathbb{R}^d$ and a stochastic process X, we can define the *hitting time* of A for X as

$$\tau_A := \inf\{t > 0 : X_t \in A\}.$$

For our usual situation (RCLL processes and Borel sets) hitting times are stopping times.

We will also need the *stopping time σ-algebra* $\mathcal{F}_\tau$ defined as

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

Intuitively, $\mathcal{F}_\tau$ represents the events known at time τ.

The continuous-time theory is technically much harder than the discrete-time theory, for two reasons:
1. questions of path-regularity arise with time continuous but not discrete;
2. uncountable operations (such as taking the supremum over an interval) arise in continuous time. But measure theory is constructed using countable operations: uncountable operations risk losing measurability.
This is why discrete and continuous time are often treated separately.
Conditional expectation.

The central definition of modern probability (Williams' phrase, [W]) is due to Kolmogorov in 1933, and explained in Doob's Lemma above. The *conditional expectation* of a random variable X *given* a σ-field $\mathcal{C}$, $E[X|\mathcal{C}]$, is any $\mathcal{C}$-measurable random variable that 'integrates the right sets the right

way', i.e. satisfies

$$\int_C E[X|\mathcal{C}]dP = \int_C X dP \qquad \forall C \in \mathcal{C}, \quad a.s. \qquad (CE)$$

This captures the idea of conditioning given known information, which you may have met in elementary Probability, or Statistics (e.g. regression). We will see it in use in V.2 below. For more on this, see e.g. SP, L 15, 16.

2. Martingales: discrete time.

We refer for a fuller account to [W]. The classic exposition is Ch. VII in Doob's book [D] of 1953.

Definition. A process $X = (X_n)$ in discrete time is called a *martingale* (mg) relative to $(\{\mathcal{F}_n\}, P)$ if

(i) X is adapted (to $\{\mathcal{F}_n\}$);

(ii) $E|X_n| < \infty$ for all n;

(iii) $[X_n|\mathcal{F}_{n-1}] = X_{n-1}$ P-a.s.

X is a *supermartingale* (supermg) if in place of (iii)

$$E[X_n|\mathcal{F}_{n-1}] \leq X_{n-1} \qquad P - a.s. \qquad (n \geq 1);$$

X is a *submartingale* (submg) if in place of (iii)

$$E[X_n|\mathcal{F}_{n-1}] \geq X_{n-1} \qquad P - a.s. \qquad (n \geq 1).$$

Martingales have a useful interpretation in terms of dynamic games: a mg is 'constant on average', and models a fair game; a supermg is 'decreasing on average', and models an unfavourable game; a submg is 'increasing on average', and models a favourable game.

Note. 1. Mgs have many connections with harmonic functions in probabilistic potential theory. Supermgs correspond to superharmonic functions, submgs to subharmonic functions.

2. X is a submg (supermg) iff $-X$ is a supermg (submg); X is a mg if and only if it is both a submg and a supermg.

3. (X_n) is a mg iff $(X_n - X_0)$ is a mg. So w.l.o.g. take $X_0 = 0$ if convenient.

4. If X is a martingale, then for $m < n$ using the iterated conditional expectation (tower) property and the martingale property repeatedly (all equalities are in the a.s.-sense)

$$E[X_n|\mathcal{F}_m] = E[E(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m] = E[X_{n-1}|\mathcal{F}_m] = \ldots = E[X_m|\mathcal{F}_m] = X_m,$$

and similarly for submgs, supermgs. See e.g. SP, L16, 17.