rss2.tex

Lecture 2

Filtering. Lest the introduction of filtrations seem fussy, we mention here an important practical subject in which it is of the essence (as the name suggests): *filtering*. Here one studies one process, given information unfolding with time about another process related to it. One then needs to carry both the filtration of the observed process, and of the 'signal' process under study. The language of filtrations and conditional expectations is essential here.

Filtering is intimately linked to stochastic control. As usual, the two principal prototypes are the Gaussian and Poisson cases. For monograph accounts, see e.g. [Kal] for the Gaussian case (including, in the special case of linear filtering, the Kalman filter), and [Bré] for the Poisson case.

Stopping Times. A random variable T is a *stopping time* for the filtration $\mathcal{F}_t$ iff

$$\{T \leq t\} \in \mathcal{F}_t \qquad \forall t.$$

Interpretation. Think of T as the time a gambler quits a gambling game. He has to take the decision whether to quit at t on the basis of what he knows at t - without foreknowledge of the future. Similarly, in finance an investor has to decide whether to change his holdings of a stock at t on the basis of what he knows already - he has no foreknowledge of future price movements.

If S, T are stopping times, so are $\max(S,T)$, $\min(S,T)$, etc.

Markov Processes. A process X is *Markov* if, for all $A \in \sigma\{X_s : s > t\}$ ('the future' at time t), $B \in \sigma\{X_s : s < t\}$ ('the past'),

$$P(A|X_t, B) = P(A|X_t). \qquad (MP)$$

In discrete time, a Markov process is usually called a Markov chain.

Interpretation. Knowing where you are at time t (X_t), knowing how you got there (B) is irrelevant for purposes of predicting the future (A).

Strong Markov Processes. A Markov process is *strong Markov* if the Markov property (MP) holds for stopping times T as well as fixed times t.

This is a definite restriction in continuous time, though not in discrete time.

Example: A Markov but not strong Markov process. Let T be an exponentially distributed random time, and let X_t be 0 for $t \leq T$, $T - t$ for $t \geq T$. Then X has graph below:

The Markov property *fails* at time T.

Another example is the left-continuous Poisson process; see below.

Martingales. A process $X = (X_t)$ is a *martingale* (mg) w.r.t. the filtration $\mathcal{F}_t$ - briefly, an $\mathcal{F}_t$-mg - if it is adapted to this filtration,

$$E|X_t| < \infty \qquad \forall t,$$

and for $s \leq t$,

$$E(X_t|\mathcal{F}_s) = X_s \qquad a.s.$$

X is a *submartingale* (submg) if $E(X_t|\mathcal{F}_s \geq X_s$ for $s \leq t$;
X is a *supermartingale* (supermg) if $E(X_t|\mathcal{F}_s \leq X_s$ for $s \leq t$.

Interpretation. Martingales model fair games. Submartingales model favourable games. Supermartingales model unfavourable games.

Martingales represent situations in which there is no drift, or tendency, though there may be lots of randomness. In the typical statistical situation where we have

$$\text{data} = \text{signal} + \text{noise},$$

martingales are used to model the noise component. It is no surprise that we will be dealing constantly with such decompositions later (with 'semimartingales').

Closed martingales. Some mgs are of the form

$$X_t = E[X|\mathcal{F}_t] \qquad (t \geq 0)$$

for some integrable random variable X. Then X is said to *close* (X_t), which is called a *closed* (or *closable*) mg, or a *regular* mg. It turns out that closed mgs have specially good convergence properties:

$$X_t \to X_\infty \qquad (t \to \infty) \text{ a.s. and in } L_1,$$

and then also

$$X_t = E[X_\infty|\mathcal{F}_t], \qquad a.s.$$

This property is equivalent also to *uniform integrability* (UI):

$$\sup_t \int_{\{|X_t|>x\}} |X_t| dP \to 0 \qquad (x \to \infty).$$

For proofs, see e.g. [Wil91], Ch. 14, [Nev], IV.2.

Square-Integrable Martingales. For $M = (M_t)$ a mg, write $M \in \mathcal{M}^2$ if M is L_2-bounded:

$$\sup_t E(M_t^2) < \infty,$$

and $M \in \mathcal{M}_0^2$ if further $M_0 = 0$. Write $c\mathcal{M}^2$, $c\mathcal{M}_0^2$ for the subclasses of continuous M.

For $M \in \mathcal{M}^2$, M is convergent:

$$M_t \to M_\infty \qquad \text{a.s. and in mean square}$$

for some random variable $M_\infty \in L_2$. One can recover M from M_∞ by

$$M_t = E[M_\infty | \mathcal{F}_t].$$

The bijection

$$M = (M_t) \leftrightarrow M_\infty$$

is in fact an isometry, and since $M_\infty \in L_2$, which is a Hilbert space, so too is $\mathcal{M}^2$. For proofs, see e.g. [R-W1]IV.4, §§23-28, or [Nev], VII.

Where there is a Hilbert-space structure, one can use the language of projections, of Pythagoras' theorem, etc., and draw diagrams as in Euclidean space. For a nice treatment of the Linear Model of statistics in such terms (analysis of variance = ANOVA, sums of squares, etc., see [Wil01], Ch. 8.

Local Martingales. For $X = (X_t)$ a stochastic process and T a stopping time, call X^T the process X *stopped* at T,

$$X^T := X_{t \wedge T}.$$

If X is a process and (T_n) an increasing sequence of stopping times, call X a *local martingale* if $(X^{T_n}) = (X_{T_n \wedge t})$ is a mg. The sequence (T_n) is called a *localizing sequence.*

Semi-Martingales. An adapted, càdlàg process $X = (X_t)$ is a *semi-martingale* if it can be decomposed as

$$X_t = X_0 + N_t + A_t,$$

with N a local mg and A a process locally of bounded variation (we will abbreviate 'locally of bounded variation' to BV). Such processes are also called *decomposable*, and the above is called the (semi-martingale, understood) *decomposition.* For background, see e.g. [Pro], III, [R-W1], IV.15.

(Recall: a function f is of *bounded variation* if the supremum of its variation $\sum_i |f(x_{i+i} - f(x_i)|$ over partitions x_i is finite; this happens iff f can be represented as the difference of two monotone functions.)
Previsible (= Predictable) Processes. The crucial difference between left-continuous (e.g., càglàd) functions and right-continuous (e.g., càdlàg) ones is that with *left-* continuity, one can 'predict' the value at t - 'see it coming' - knowing the values before t.

We write $\mathcal{P}$, called the *predictable* (or previsible) σ-algebra, for the σ-algebra on $\mathbf{R}_+ \times \Omega$ ($\mathbf{R}_+$ for time $t \geq 0$, Ω for randomness $\omega \in \Omega$ - we need both for a stochastic process $X = (X(t, \omega))$) for the σ-field generated by (= smallest σ-field containing) the adapted càglàd processes. (We shall almost always be dealing with adapted processes, so the operative thing here is the *left*-continuity.) We also write $X \in \mathcal{P}$ as shorthand for 'the process X is $\mathcal{P}$-measurable', and $X \in b\mathcal{P}$ if also X is bounded.
Predictability and Semi-Martingales. Anticipating greatly, let us confess here why we need to introduce the last two concepts. We will develop a theory of *stochastic integrals*, $\int_0^t H(s, \omega) dM(s, \omega)$ or $\int_0^t H_s dM_s$, where H, M are stochastic processes and the *integrator* M is a *semimartingale*, the *integrand* H is *previsible* (and bounded, or L_2, or whatever). This can be done (though we shall have to quote many proofs; see e.g. [Pro] or [R-W2]). More: this theory is the *most general* theory of stochastic integration *possible*, if one demands even reasonably good properties (appropriate behaviour under passage to the limit, for example). For emphasis:

Integrands: previsible,

Integrators: semimartingales.

Prototype: H is left-continuous (and bounded, or L_2, etc.); M is Brownian motion (Ch. 2, 3).
Economic Interpretation. Think of the integrator M as, e.g., a stock-price process. The increments over $[t, t+u]$ ($u > 0$, small) represent 'new information'. Think of the integrand H as the amount of stock held. The investor has no advance warning of the price change $M_{t+dt} - M_t$ over the immediate future $[t, t + dt]$, but has to commit himself on the basis of what he knows already. So H needs to be predictable at H before t (e.g., left- continuity will do), hence predictability of integrands. By contrast, $M_{t+dt} - M_t$ represents new price-sensitive information, or 'driving noise'. The value process of the portfolio is the limit of sums of terms such as $H_{t-}(M_{t+dt} - M_t)$, the stochastic integral $\int_0^t H_s dM_s$. This is the continuous-time analogue of the *martingale transform* in discrete time; see e.g. [Nev], VIII.4.

Note. Stochastic calculus with 'anticipating' integrands, 'backward' stochastic integrals, etc., have been developed, and are useful (e.g., in more advanced areas such as the *Malliavin calculus*. But let us learn to walk before we learn to run.

Renewal Processes; Lack of Memory Property; Exponential Distributions. Suppose we use components - light-bulbs, say - whose lifetimes $X_1, X_2, \ldots$ are independent, all with law F on $(0, \infty$. The first component is installed new, used until failure, then replaced, and we continue in this way. Write

$$S_n := \sum_1^n X_i, \qquad N_t := \max\{k : S_k < t\}.$$

Then $N = (N_t : t \geq 0)$ is called the *renewal process* generated by F; it is a *counting process*, counting the number of failures seen by time t.

The law F has the *lack-of-memory property* iff the components show no aging - that is, if a component still in use behaves as if new. The condition for this is

$$P(X > s + t | X > s) = P(X > t) \qquad (s, t > 0),$$

or

$$P(X > s + t) = P(X > s) P(X > t).$$

Writing $\overline{F}(x) := 1 - F(x)$ $(x \geq 0)$ for the *tail* of F, this says that

$$\overline{F}(s + t) = \overline{F}(s) \overline{F}(t) \qquad (s, t \geq 0).$$

Obvious solutions are

$$\overline{F}(t) = e^{-\lambda t}, \qquad F(t) = 1 - e^{-\lambda t}$$

for some $\lambda > 0$ - the exponential law $E(\lambda)$. Now

$$f(s + t) = f(s) f(t) \qquad (s, t \geq 0)$$

is a 'functional equation' - the *Cauchy functional equation* - and it turns out that these are the *only* solutions, subject to minimal regularity (such as one-sided boundedness, as here - even on an interval of arbitrarily small length!). For details, see e.g. [BGT], §1.1.1.

So the exponential laws $E(\lambda)$ are *characterized* by the lack-of-memory property. Also, the lack-of-memory property corresponds in the renewal context to the *Markov property*. The renewal process generated by $E(\lambda)$ is called the *Poisson (point) process* with *rate* λ, $Ppp(\lambda)$. So:

among renewal processes, the only Markov processes are the Poisson processes.

When we meet Lévy processes (Ch. 2 below) we shall find also:

among renewal processes, the only Lévy processes are the Poisson processes.

Historical Overview: Martingale Calculus.

Stochastic integrals can be found in the work of Wiener (e.g. Paley-Wiener-Zygmund, 1933, MZ) and Doob (1937, TAMS), but the origin of the subject in its modern setting and generality was the paper by Itô in 1944 (Proc. Imp. Acad. Tokyo), whence the name *Itô calculus.* Important further strides were taken by the Japanese school, in particular by Kunita and Watanabe (square-integrable mgs, Kunita- Watanabe inequality). The first classic textbook, on the Brownian case, was McKean in 1969 [McK]. The work of P.-A. Meyer and the French/Strasbourg school resulted in both the 'general theory of (stochastic) processes' and the first monograph account of stochastic calculus for semimartingales (the concept is due to Meyer) in 1976, [Mey]. The Poisson case, though less difficult, was developed later, the impetus being filtering theory (work of Wong, Snyder. Kushner and others in the 1970s). The term 'martingale calculus' is due to Wong; see [Bré] for the first monograph treatment of Poisson calculus, in the context of queueing theory, [BaBr] for its sequel, on 'Palm-martingale calculus' - again applied to queueing theory. The first monograph on stochastic calculus for Lévy processes (Ch. 2) - 'Lévy calculus' - is still to appear [A]; I thank Professor D. B. (Dave) Applebaum of NTU for access to a copy in advance.

Note. Although the Itô integral cannot be generalized further without losing its desirable properties, it is in fact not the only concept of stochastic integration that one needs. It does not transform well under change of coordinates in geometrical situations, for which one needs instead the Stratonovich integral (see e.g. [R-W2], IV.46, 47). Also, a new theory of stochastic integration with 'rough paths' (e.g., Brownian motion) has recently been developed by Professor T. J. (Terry) Lyons of Oxford:

T. J. LYONS & Zhongmin QIAN: *System control and rough paths*, OUP, 2002.