rss1.tex

PROBABILITY MODELS BASED ON PROCESSES WITH INDEPENDENT INCREMENTS

N. H. BINGHAM

Ten-Lecture Postgraduate Course, Royal Statistical Society

Nuffield College, Oxford, 23-27 September 2002

Wolfson Court, Girton College Cambridge, 20-24 September 2004

Lecture 1.

Chapter 1. BACKGROUND ON STOCHASTIC PROCESSES

Before proceeding to the dynamic setting of stochastic processes, we briefly review what we shall need in a static setting.
Probability Spaces [Kol].

To describe a 'random experiment' - a situation generating randomness - we need:
a *sample space* Ω, representing the set of all possible outcomes (each individual outcome ω is a *sample point*); certain distinguished subsets of Ω, called *events*, for which a probability

$$P(A) = P(\{\omega : \omega \in A\})$$

is defined. If A, B are events, we need sets such as $A \cup B$, $A \cap B$, $A \setminus B$ to be events; if $A_1, A_2, \ldots$ are events, $\bigcup_{r=1}^{n} A_r$ is an event by above, and we further assume $\bigcup_{n=1}^{\infty} A_n$ is also an event. The class $\mathcal{F}$ of events is thus closed under countable set-theoretic operations, and is called a σ-field (sigma-field: S $\leftrightarrow$ Summe = sum, union in German). Note that $\emptyset$ ($= A \setminus A$) $\in \mathcal{F}$, and Ω ($= \emptyset^c$) $\in \mathcal{F}$.

We need a set-function P defined in $\mathcal{F}$ (so that $P(A)$ is defined for each event $A \in \mathcal{F}$), satisfying

$P(\emptyset) = 0, \qquad P(\Omega) = 1;$

$P(A) \geq 0;$

$P(\cup_1^n A_r) = \sum_1^n P(A_r)$ for $A_i \in \mathcal{F}$ disjoint (finite additivity),
which we shall strengthen to
$P(\cup_1^\infty A_n) = \sum_1^\infty P(A_n)$ for $A_n \in \mathcal{F}$ disjoint (countable additivity).
Such a P is called a *probability measure*, and such a triple $(\Omega, \mathcal{F}, \mathcal{P})$ is called a *probability space*.

Random Variables and the σ-Fields they Generate.

The σ-field *generated* by a class of sets is the smallest s-field containing them (or, the intersection of all σ-fields containing them).

A *random variable* X is a measurable function $X : \Omega \to \mathbf{R}$ - that is, a function such that for each $x \in \mathbf{R}$, $\{\omega : X(\omega) \leq x\}$ or $\{X \leq x\} \in \mathcal{F}$ (is an event), and so has a probability $P(X \leq x)$ defined. The collection of these probabilities as x varies is called the *distribution function* (distribution, law) of X. So:

X is a random variable (rv) iff its distribution function is defined.

The σ-field generated by the events $\{X \leq x\}$, $\forall x \in \mathbf{R}$ (equivalently, $\{X \in I\}$ for all intervals I, or $\{X \in B\}$ for all 'Borel sets' B (the Borel sets form the σ-field generated by the intervals) is called the σ-field *generated by* X, $\sigma(X)$.

Interpretation: $\sigma(X)$ represents the information contained in X, or 'what we know when we know X'.

Justification. By a theorem of Doob, for random variables X, Y,

$\sigma(X) \subset \sigma(Y)$ iff Y is a (measurable) function of X, $Y = f(X)$.

For, applying a function loses information in general (no loss iff the function is injective, so has an inverse, and we can go back by applying the inverse function).

Filtrations. Now we feed in time $t \geq 0$, discrete ($t = 1, 2, \ldots$ - we usually then write n for t) or continuous. As time progresses, we learn more. Write $\mathcal{F}_t$ for the σ-field representing our knowledge at time t ($\mathcal{F}_t$ will usually be the σ-field generated by all random variables observed by time t). Then

$$s \leq t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t :$$

$\{\mathcal{F}_t : t \geq 0\}$ is an increasing family of σ-fields, called a *filtration.* With the probability space $(\Omega, \mathcal{F},\ P)$ augmented by the filtration, we have a *filtered probability space* or *stochastic basis* - so called because it provides an adequate basis on which to define a stochastic process.

Stochastic Processes. A *stochastic process* $X = \{X_t : t \geq 0\}$ (or $\{X(t) : t \geq 0\}$) on a stochastic basis as above is a family X_t of random variables on the probability space such that for each t, X_t is $\mathcal{F}_t$-measurable. We say that X is *adapted* to the filtration. Often $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$, the *natural filtration* of X. Any process is adapted to its natural filtration.

Path Properties.

We will only need to look at stochastic processes X whose paths - the random function $t \mapsto X_t = X_t(\omega)$ or $X(t, \omega)$ - have some regularity, a.s. (two processes whose finite- dimensional distributions are the same are said to be *versions* of each other, and naturally we prefer to work with as regular a version as possible). Sometimes one has *continuous* paths (as with Brownian motion, Ch. 3). One will usually have paths which are *right-continuous with left limits* (RCLL), or in French (since we owe much of this to P.-A. Meyer and the French, particularly Strasbourg, school), continu à droite, limite à gauche (càdlàg). In important special cases, the paths are *jump-functions*, that change only at jump discontinuities: $\Delta X = X_t - X_{t-}$ (using the càdlàg property of paths), so $X_t = X_0 + \sum_{0<s\leq t} \Delta X_s$. There can be at most countably many jumps in such sums, but some processes we shall meet jump infinitely often in finite time (in consequence, it is impossible to draw their paths accurately!).

An exception to our usually working with càdlàg processes arises when we deal with stochastic integrals, $\int_0^t H_s dX_s$ (Ch. 4) with H, X stochastic processes. For reasons which will emerge later, integrands H are taken *left*-continuous with *right*- limits, continu à *gauche*, limite à *droite* or càglàd. For emphasis:

Integrators: càdlàg,
Integrands: càglàd.

Conditional Expectations.

In elementary probability, with X, Y discrete random variables, starting with the joint distribution of (X, Y), we can form the conditional distribution of one given the other, and so the conditional expectation of one given the other (when the - ordinary - expectation exists). Similarly when (X, Y) has a joint density (replace sums by integrals). We need a general framework containing both, and more.

Let $\mathcal{A}$ be a σ-field (as before, $\mathcal{A}$ represents an amount of information, or partial knowledge), $\mathcal{A} \subset \mathcal{F}$. As before, call X $\mathcal{A}$-measurable if $\{X \leq x\} \in \mathcal{A}$ for all x (as $\mathcal{A}$ is smaller than $\mathcal{F}$, this is a stronger restriction than $\mathcal{F}$-measurability, which we abbreviate to measurability).

Definition (Kolmogorov, 1933 [Kol]).

For a random variable X with $E|X| < \infty$, the *conditional expectation* of

X *given* $\mathcal{A}$ is the $\mathcal{A}$-measurable random variable, written $E(X|\mathcal{A})$, such that

$$\int_A X dP = \int_A E(X|\mathcal{A})\, dP \qquad \forall A \in \mathcal{A}, \qquad \text{a.s. } (*).$$

Note. 1. If $Q(A) := \int_A X\ dP$ $(A \in \mathcal{A})$, Q is a measure (non-negative, σ- additive set-function) if $X \geq 0$, and a signed measure in general. If $P(A) = 0$, $Q(A) = 0$, $\forall A \in \mathcal{A}$. Then Q is called *absolutely continuous* w.r.t. P, $Q << P$. By the Radon-Nikodým Theorem (which we quote from measure theory), Q has the form $Q(A) = \int_A Y dP$ for some $\mathcal{A}$-measurable Y.
2. Kolmogorov's definition $(*)$ of conditional expectation is rightly called by David Williams [Wil91] 'the central definition of modern probability theory'. It is important but non-obvious, and reveals its value in its properties and ease of handling in proofs.

Properties.

1. For $\mathcal{A}$ the trivial σ-field, containing only $\emptyset$ and Ω - 'knowing nothing'), $E(X|\mathcal{A}) = EX$.

Proof. For $A = \emptyset$, both sides of $(*)$ are 0. For $A = \Omega$, both sides are $\int_\Omega X dP = EX$.

2. For $\mathcal{A} = \mathcal{F}$ ('knowing everything'), $E(X|\mathcal{F}) = X$.

Proof. Now $E(X|\mathcal{F})$ has to integrate like X over *every* set A $(\in \mathcal{F})$, and this forces $E(X|\mathcal{F})$ to be X a.s.

3. If $\mathcal{A} \subset \mathcal{B}$,

$$E[E(X|\mathcal{A})|\mathcal{B}] = E(X|\mathcal{A}),$$
$$E[E(X|\mathcal{B})|\mathcal{A}] = E(X|\mathcal{A})$$

(iterated conditional expectations, or tower property, or coarse-averaging property). We omit the proof (an excellent exercise - recommended!) for brevity; see e.g. [Wil91] or [B-K98].
Interpretation: the coarser (smaller) σ-field rubs out the effect of the larger (finer) one, either way round.

4 (Conditional Mean Formula). $E[E(X|\mathcal{A})] = EX$.

Proof: Take $\mathcal{A}$ the trivial σ-field above, use (1), and then write $\mathcal{A}$ for $\mathcal{B}$.

5. If X is $\mathcal{A}$-measurable,

$$E(XY|\mathcal{A}) = X\ E(Y|\mathcal{A})$$

('taking out what is known'). Interpretation: as X is $\mathcal{A}$-measurable, given $\mathcal{A}$ we know X. So X now counts as a constant, and can be taken out through expectations (or integral signs).

6. If X is independent of $\mathcal{A}$ - that is, if $\{X \leq x\}$ and A are independent events for all $x \in \mathbf{R}$ and $A \in \mathcal{A}$ - $E(X|\mathcal{A}) = X$.
Interpretation: independence means that information in $\mathcal{A}$ is irrelevant to X, and so the conditioning has no effect.

Note. Take $\mathcal{B} = \mathcal{A}$ in (3): $E[E(X|\mathcal{A})|\mathcal{A}] = E(X|\mathcal{A})$. So $E[.|\mathcal{A}]$ is *idempotent.* It is also *linear.* So it is a *projection.* We can (and should) think of $E[.|\mathcal{A}]$ as projecting onto what we know given $\mathcal{A}$. This point of view enables us to think geometrically, as in Euclidean space - or in Hilbert space (and we will be working in Hilbert space with the square-integrable martingales below and in Ch. 2, 3).

Conditional expectation also has the properties one might expect of an integral - linearity and positivity - and conditional forms of the monotone and dominated convergence theorems, Fatou's lemma, Jensen's inequality etc. hold. For details, see e.g. [Wil91].

In statistics, conditional expectations corresponds to *regression* - a regression function is a conditional mean $E(Y|X = x)$. Furthermore, handling *sufficiency* - particularly *minimal sufficiency* - rigorously needs the machinery above. So this is not merely mathematical abstraction, but necessary tools for the statistician!