Lecture 2. 7.10.2022

*Measure and Probability*

We will freely abbreviate 'probability' [11 letters] to pr [or prob, if you prefer], and (probability) distribution function to (probability) d/n, d/n fn or *law*.

We have a well-known correspondence between sets of technical terms here. With Measure Theory of the left and Probability Theory on the right:
measure [as a set-function] $\leftrightarrow$ probability [when the measure has mass 1]
measure [of a measurable set] $\leftrightarrow$ probability [of an event]
measure space $\leftrightarrow$ probability space
measurable set $\leftrightarrow$ event
measurable function $\leftrightarrow$ random variable
integral [wrt a measure] $\leftrightarrow$ expectation [wrt a pr measure]
convergence in measure $\leftrightarrow$ convergence in probability
almost everywhere, a.e. [except on a set of measure 0] $\leftrightarrow$ almost surely, a.s. [except on a set of pr 0]
a.e. convergence $\leftrightarrow$ a.s. convergence
$L_2$-convergence $\leftrightarrow$ mean-square convergence
weak convergence $\leftrightarrow$ convergence in distribution

*Stochastic processes*

The term *stochastic* (coined in the 1920s by A. Ya. Khinchin/Ȟinčin (depending on transliteration from Cyrillic to Roman), or Khintchine (when he was writing in French), can generally be used interchangeably with'probabilistic', or even 'random' (though note, as we shall see, a *stochastic matrix* is one whose row-sums are 1, while a *random matrix* is one whose elements are random). The term *process* denotes a phenomenon evolving, or unfolding, with time (if with space, we speak of a spatial process, if both of a space-time process). ['Stochastic' comes from the Greek word meaning to aim at or guess; compare the title of the first classic book in probability,
Jacob Bernoulli, *Ars conjectandi*, 1713 ('The art of guessing').
Perhaps *random process* as in [GriS] would be better.]

To begin, we need a space to model where the randomness takes place: a *probability space*. This is a *triple* $(\Omega, \{\mathcal{F}\}, \mathbb{P})$, consisting of:
a *sample space*, $\Omega$ – the set whose elements are the individual random outcomes, the *sample points*, $\omega$;
a $\sigma$-field $\mathcal{F}$ of *events* – the subsets $A$ of $\Omega$ whose probability $\mathbb{P}(A)$ is defined;
a probability measure $\mathbb{P}$.

*General note on notation.*

Our first duty is to avoid ambiguity – a sin, in mathematics.

Our second duty is to use the best notation for the job in hand, which is usually the minimal one (all mathematicians are minimalists at heart!) This is *context-dependent*, just as ordinary language is. The same word can mean quite different things in different contexts.

Not changing notation to suit a changed context compares with not changing one's clothes to suit a changed temperature or changed degree of formality etc. There is no virtue in 'uniform consistency', in either context. Remember: *notation is our servant, not our master.*

You have probably been exposed by now to enough different lecturers using different notation to be used to this sort of thing by now. I hope so!

*Notation and sampling*

Suppose we have a random variable, $X$ say (below we shall use $X$ for a stochastic process, a whole collection of random variables indexed by an index set $I$, but the two are consistent: one random variable corresponds to $I$ being a singleton). The randomness here is in the sample point $\omega$, so in full this is $X(\omega)$. It is customary, and very convenient, to omit the $\omega$ unless we need to mention it specifically, and we shall usually do this. To begin with, for your own use: if in doubt, put them in; if not in doubt, leave them out.

*Note.*

Chung used '$\omega$s everywhere' in the 1960 edition of his classic book [Chu] *Markov chains*. In the Preface to the 1967 edition (p. X), he writes:

'Personal taste and habit not being stationary in time, I should have liked to make more radical departures such as deleting hundreds of the $\omega$'s in the cumbersome notation, but have generally decided to leave well alone'.

I too have moved in the same way (perhaps under his influence).

A *stochastic process* (in brief, stoch proc, SP, or just *process* ) $X$ is a mathematical model of a phenomenon evolving in time, the model taking the form of a *family* of random variables $\{X_i : i \in I\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and indexed by some index set $I$.

With a stochastic process indexed by time (below), the value of the process $X$ at time $t$ will be written $X_t(\omega)$, or just $X_t$. But, we frequently have multiple time-points $t_i$, say, and rather than have 'suffices within suffices', it is convenient to write $X_t(\omega) = X_t$ as $X(t, \omega) = X(t)$. I use them interchangeably, just as I do $L_p$ and $L^p$, and hardly notice the difference.