ltcc2.tex Week 2 8 October 2012

Chapter II. CONDITIONING. STOCHASTIC PROCESSES.

§1. CONDITIONAL EXPECTATIONS.

Suppose that X is a random variable, whose expectation exists (i.e. $E|X| < \infty$, or $X \in L_1$). Then EX, the expectation of X, is a scalar (a number) – non-random. The expectation operator E averages out all the randomness in X, to give its mean (a weighted average of the possible value of X, weighted according to their probability, in the discrete case). It often happens that we have *partial information* about X – for instance, we may know the value of a random variable Y which is associated with X, i.e. carries information about X. We may want to average out over the remaining randomness. This is an expectation conditional on our partial information, or more briefly a conditional expectation. This idea will be familiar already from elementary courses, in two cases:
1. *Discrete case*, based on the formula

$$P(A|B) := P(A \cap B)/P(B) \text{ if } P(B) > 0.$$

If X takes values $x_1, \cdots, x_m$ with probabilities $f_1(x_i) > 0$, Y takes values $y_1, \cdots, y_n$ with probabilities $f_2(y_j) > 0$, (X, Y) takes values (x_i, y_j) with probabilities $f(x_i, y_j) > 0$, then
(i) $f_1(x_i) = \Sigma_j f(x_i, y_j), \qquad f_2(y_j) = \Sigma_i f(x_i, y_j)$,
(ii) $P(Y = y_j|X = x_i) = P(X = x_i, Y = y_j)/P(X = x_i) = f(x_i, y_j)/f_1(x_i)$

$$= f(x_i, y_j)/\Sigma_j f(x_i, y_j).$$

This is the *conditional distribution* of Y *given* $X = x_i$, written

$$f_{Y|X}(y_j|x_i) = f(x_i, y_j)/f_1(x_i) = f(x_i, y_j)/\Sigma_j f(x_i, y_j).$$

Its expectation is

$$E(Y|X = x_i) = \Sigma_j y_j f_{Y|X}(y_j|x_i) = \Sigma_j y_j f(x_i, y_j)/\Sigma_j f(x_i, y_j).$$

The problem here is that this approach only works when the events on which we condition have *positive* probability, which only happens in the *discrete* case.
2. *Density case.* If (X, Y) has density $f(x, y)$,

$$X \text{ has density } f_1(x) := \int_{-\infty}^{\infty} f(x, y)dy, \qquad Y \text{ has density } f_2(y) := \int_{-\infty}^{\infty} f(x, y)dx.$$

We *define* the *conditional density* of Y *given* $X = x$ by the continuous analogue of the discrete formula above:

$$f_{Y|X}(y|x) := f(x,y)/f_1(x) = f(x,y)/\int_{-\infty}^{\infty} f(x,y)dy.$$

Its expectation is

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = \int_{-\infty}^{\infty} y f(x,y)dy/\int_{-\infty}^{\infty} f(x,y)dy.$$

Example: Bivariate normal distribution, $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

$$E(Y|X=x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),$$

the familiar *regression line* of statistics (linear model).

The problem here is that joint densities need not exist – do not exist, in general.

One of the great contributions of Kolmogorov's classic book of 1933 [Kol] was the realization that measure theory – specifically, the Radon-Nikodym theorem – provides a way to treat conditioning in general, without making assumptions that we are in one of the two cases – discrete case and density case – above.

Recall that the probability triple is $(\Omega, \mathcal{F}, P)$. Suppose that $\mathcal{B}$ is a sub-σ-field of $\mathcal{F}$, $\mathcal{B} \subset \mathcal{F}$ (recall that a σ-field represents information; the big σ-field $\mathcal{F}$ represents 'knowing everything', the small σ-field $\mathcal{B}$ represents 'knowing something').

Suppose that Y is a non-negative random variable whose expectation exists: $EY < \infty$. The set-function

$$Q(B) := \int_B Y dP \qquad (B \in \mathcal{B})$$

is non-negative (because Y is), σ-additive – because

$$\int_B Y dP = \Sigma_n \int_{B_n} Y dP$$

if $B = \cup_n B_n$, B_n disjoint – and defined on the σ-algebra $\mathcal{B}$, so is a *measure* on $\mathcal{B}$. If $P(B) = 0$, then $Q(B) = 0$ also (the integral of anything over a null set is zero), so $Q << P$. By the Radon-Nikodym theorem (Week 1), there exists a Radon-Nikodym derivative of Q with respect to P on $\mathcal{B}$, which is $\mathcal{B}$-measurable [in the Radon-Nikodym theorem as stated in Week 1, we had $\mathcal{F}$ in place of $\mathcal{B}$, and got a random variable, i.e. an $\mathcal{F}$-measurable function. Here, we just replace $\mathcal{F}$ by $\mathcal{B}$.] Following [Kol], we call this Radon-Nikodym

derivative the *conditional expectation* of Y *given* (or *conditional on*) $\mathcal{B}$, $E(Y|\mathcal{B})$: this is $\mathcal{B}$-measurable, integrable, and satisfies

$$\int_B Y dP = \int_B E(Y|\mathcal{B}) dP \qquad \forall B \in \mathcal{B}. \tag{$*$}$$

In the general case, where Y is a random variable whose expectation exists ($E|Y| < \infty$) but which can take values of both signs, decompose Y as

$$Y = Y_+ - Y_-$$

and define $E(Y|\mathcal{B})$ by linearity as

$$E(Y|\mathcal{B}) := E(Y_+|\mathcal{B}) - E(Y_-|\mathcal{B}).$$

Suppose now that $\mathcal{B}$ is the σ-field generated by a random variable X: $\mathcal{B} = \sigma(X)$ (so $\mathcal{B}$ represents the information contained in X, or what we know when we know X). Then $E(Y|\mathcal{B}) = E(Y|\sigma(X))$, which is written more simply as $E(Y|X)$. Its defining property is

$$\int_B Y dP = \int_B E(Y|X) dP \qquad \forall B \in \sigma(X).$$

Similarly, if $\mathcal{B} = \sigma(X_1, \cdots, X_n)$ ($\mathcal{B}$ is the information in $(X_1, \cdots, X_n)$) we write $E(Y|\sigma(X_1, \cdots, X_n)$ as $E(Y|X_1, \cdots, X_n)$:

$$\int_B Y dP = \int_B E(Y|X_1, \cdots, X_n) dP \qquad \forall \mathcal{B} \in \sigma(X_1, \cdots, X_n).$$

Note. 1. To check that something is a conditional expectation: we have to check that it integrates the right way over the right sets [i.e., as in (*)].

2. From (*): if two things integrate the same way over all sets $B \in \mathcal{B}$, they have the same conditional expectation given $\mathcal{B}$.

3. For notational convenience, we shall pass between $E(Y|\mathcal{B})$ and $E_{\mathcal{B}}Y$ at will.

4. The conditional expectation thus defined coincides with any we may have already encountered - in regression or multivariate analysis, for example. However, this may not be immediately obvious. The conditional expectation defined above – via σ-fields and the Radon-Nikodym theorem – is rightly called by Williams ([**W**], p.84) 'the central definition of modern probability'. It may take a little getting used to. As with all important but non-obvious definitions, it proves its worth in action: see below for properties of conditional expectations, and for its use in studying stochastic processes, particularly martingales [which are defined in terms of conditional expectations].

§2. PROPERTIES OF CONDITIONAL EXPECTATIONS.

1. $\mathcal{B} = \{\emptyset, \Omega\}$. Here $\mathcal{B}$ is the *smallest* possible σ-field (*any* σ-field of subsets of Ω contains $\emptyset$ and Ω), and represents 'knowing nothing'.

$$E(Y|\{\emptyset, \Omega\}) = EY.$$

Proof. We have to check (*) for $B = \emptyset$ and $B = \Omega$. For $B = \emptyset$ both sides are zero; for $B = \Omega$ both sides are EY. •

2. $\mathcal{B} = \mathcal{F}$. Here $\mathcal{B}$ is the *largest* possible σ-field, and represents 'knowing everything'.

$$E(Y|\mathcal{F}) = Y \qquad P - a.s.$$

Proof. We have to check (*) for *all* sets $B \in \mathcal{F}$. The only integrand that integrates like Y over *all* sets is Y itself, or a function agreeing with Y except on a set of measure zero.
Note. When we condition on $\mathcal{F}$ ('knowing everything'), we *know* Y (because we know everything). There is thus no uncertainty left in Y to average out, so taking the conditional expectation (averaging out remaining randomness) has no effect, and leaves Y unaltered.

3. If Y is $\mathcal{B}$-measurable, $E(Y|\mathcal{B}) = Y \qquad P - a.s.$
Proof. Recall that Y is *always* $\mathcal{F}$-measurable (this is the definition of Y being a random variable). For $\mathcal{B} \subset \mathcal{F}$, Y may not be $\mathcal{B}$-measurable, but if it is, the proof above applies with $\mathcal{B}$ in place of $\mathcal{F}$.
Note. If Y is $\mathcal{B}$-measurable, when we are given $\mathcal{B}$ (that is, when we condition on it), we *know* Y. That makes Y effectively a constant, and when we take the expectation of a constant, we get the same constant.

4. If Y is $\mathcal{B}$-measurable, $E(YZ|\mathcal{B}) = YE(Z|\mathcal{B}) \qquad P - a.s.$
We refer for the proof of this to [**W**], p.90, proof of (j).
Note. Williams calls this property 'taking out what is known'. To remember it: if Y is $\mathcal{B}$-measurable, then given $\mathcal{B}$ we know Y, so Y is effectively a constant, so can be taken out through the integration signs in (*) (with YZ in place of Y).

5. If $\mathcal{C} \subset \mathcal{B}$, $E[E(Y|\mathcal{B})|\mathcal{C}] = E[Y|\mathcal{C}] \qquad a.s.$
Proof. $E_{\mathcal{C}}E_{\mathcal{B}}Y$ is $\mathcal{C}$-measurable, and for $C \in \mathcal{C} \subset \mathcal{B}$,

$$\int_C E_{\mathcal{C}}[E_{\mathcal{B}}Y]dP = \int_C E_{\mathcal{B}}Y\,dP \qquad \text{(definition of } E_{\mathcal{C}} \text{ as } C \in \mathcal{C})$$

$$= \int_C Y\,dP \qquad \text{(definition of } E_{\mathcal{B}} \text{ as } C \in \mathcal{B}).$$

So $E_{\mathcal{C}}[E_{\mathcal{B}}Y]$ satisfies the defining relation for $E_{\mathcal{C}}Y$. Being also $\mathcal{C}$-measurable, it *is* $E_{\mathcal{C}}Y$ (a.s.). •

5'. If $\mathcal{C} \subset \mathcal{B}$, $E[E(Y|\mathcal{C})|\mathcal{B}] = E[Y|\mathcal{C}] \quad a.s.$
Proof. $E[Y|\mathcal{C}]$ is $\mathcal{C}$-measurable, so $\mathcal{B}$-measurable as $\mathcal{C} \subset \mathcal{B}$, so $E[.|\mathcal{B}]$ has no effect, by 3.

Note. 5, 5' are the two forms of the *iterated conditional expectations property*. When conditioning on two σ-fields, one larger (finer), one smaller (coarser), the coarser rubs out the effect of the finer, either way round. This may be thought of as the *coarse-averaging property*: we shall use this term interchangeably with the iterated conditional expectations property (Williams [**W**] uses the term *tower property*).

6. *Role of independence.* If Y is independent of $\mathcal{B}$,

$$E(Y|\mathcal{B}) = EY \qquad a.s.$$

Proof. See [**W**], p.88, 90, property (k).
Note. In the elementary definition $P(A|B) := P(A \cap B)/P(B)$ (if $P(B) > 0$), if A and B are independent (that is, if $P(A \cap B) = P(A).P(B)$), then $P(A|B) = P(A)$: conditioning on something independent has no effect. One would expect this familiar and elementary fact to hold in this more general situation also. It does – and the proof of this rests on the proof above.

7. *Conditional Mean Formula.* $E[E(Y|\mathcal{B})] = EY \quad P - a.s.$
Proof. Take $\mathcal{C} = \{\emptyset, \Omega\}$ in 5 and use 1. •
Example. Check this for the bivariate normal distribution considered above.

8. *Conditional Variance Formula.* $varY = E_X var(Y|X) + var_X E(Y|X)$.
Recall $varX := E[(X - EX)^2]$. Expanding the square,

$$varX = E[X^2 - 2X.(EX) + (EX)^2] = E(X^2) - 2(EX)(EX) + (EX)^2 = E(X^2) - (EX)^2.$$

Conditional variances can be defined in the same way. Recall that $E(Y|X)$ is constant when X is known ($= x$, say), so can be taken outside an expectation over X, E_X say. Then

$$var(Y|X) := E(Y^2|X) - [E(Y|X)]^2.$$

Take expectations of both sides over X:

$$E_X var(Y|X) = E_X[E(Y^2|X)] - E_X[E(Y|X)]^2.$$

Now $E_X[E(Y^2|X)] = E(Y^2)$, by the Conditional Mean Formula, so the right is, adding and subtracting $(EY)^2$,

$$\{E(Y^2) - (EY)^2\} - \{E_X[E(Y|X)]^2 - (EY)^2\}.$$

The first term is $varY$, by above. Since $E(Y|X)$ has E_X-mean EY, the second term is $var_X E(Y|X)$, the variance (over X) of the random variable $E(Y|X$ (random because X is). Combining, the result follows.

Interpretation. $varY$ = total variability in Y,

$E_X var(Y|X)$ = variability in Y not accounted for by knowledge of X,

$var_X E(Y|X)$ = variability in Y accounted for by knowledge of X.

Example: the bivariate normal.

$$Y|X = x \text{ is } N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2)), \qquad varY = \sigma_2^2,$$

$$E(Y|X = x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad E(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1),$$

which has variance $(\rho\sigma_2/\sigma_1)^2 varX = (\rho\sigma_2/\sigma_1)^2\sigma_1^2 = \rho^2\sigma_2^2$,

$$var(Y|X = x) = \sigma_2^2 \text{ for all } x, var(Y|X) = \sigma_2^2(1 - \rho^2), E_X var(Y|X) = \sigma_2^2(1 - \rho^2).$$

COROLLARY. *$E(Y|X)$ has the same mean as Y and smaller variance (if anything).*

Proof. From the Conditional Mean Formula, $E[E(Y|X)] = EY$. Since $var(Y|X) \geq 0$, $E_X var(Y|X) \geq 0$, so $varE[Y|X] \leq varY$ from the Conditional Variance Formula.

This result has important applications in estimation theory. Suppose we are to estimate a parameter θ, and are considering a statistic X as a possible estimator (or basis for an estimator) of θ. We would naturally want X to contain all the information on θ contained within the entire sample. What (if anything) does this mean in precise terms? The answer lies in the concept of *sufficiency* ('data reduction') – one of the most important contributions to statistics of the great English statistician R. A. (Sir Ronald) Fisher (1880-1962). In the language of sufficiency, the Conditional Variance Formula is seen as (essentially) the Rao-Blackwell Theorem, a key result in the area (see the index in your favourite Statistics book if you want more here).

§3. FILTRATIONS.

The Kolmogorov triples $(\Omega, \mathcal{F}, P)$, and the Kolmogorov conditional expectations

$E(X|\mathcal{B})$, give us all the machinery we need to handle *static* situations involving randomness. To handle *dynamic* situations, involving randomness which unfolds with *time*, we need further structure.

We may take the initial, or starting, time as $t = 0$. Time may evolve discretely, or continuously. We defer the continuous case; in the discrete case, we may suppose time evolves in integer steps, $t = 0, 1, 2, \cdots$. We wish to model a situation involving randomness unfolding with time. We suppose, for simplicity, that information is never lost (or forgotten): thus, as time increases we learn more. Recall that σ-fields represent information or knowledge. We thus need a sequence of σ-fields $\{\mathcal{F}_n : n = 0, 1, 2, \cdots\}$, which are increasing:

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \qquad (n = 0, 1, 2, \cdots),$$

with $\mathcal{F}_n$ representing the information, or knowledge, available to us at time n. We shall always suppose all σ-fields to be *complete*, i.e., to contain all subsets of null sets as null sets (this can be avoided, and is not always appropriate, but it simplifies matters and suffices for our purposes). Thus $\mathcal{F}_0$ represents the initial information (if there is none, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial σ-field). On the other hand,

$$\mathcal{F}_\infty := lim_{n\to\infty}\mathcal{F}_n$$

represents all we ever will know (the 'Doomsday σ-field'). Often, $\mathcal{F}_\infty$ will be $\mathcal{F}$ (the σ-field from Week 1, representing 'knowing everything'. But this will not always be so; see e.g. [**W**], §15.8 for an interesting example.

Such a family $\{\mathcal{F}_n : n = 0, 1, 2, \cdots\}$ is called a *filtration*; a probability space endowed with such a filtration, $\{\Omega, \{\mathcal{F}_n\}, \mathcal{F}, P\}$ is called a *filtered probability space*. (These definitions are due to P. A. MEYER of Strasbourg; Meyer and the Strasbourg (and more generally, French) school of probabilists have been responsible for the 'general theory of [stochastic] processes', and for much of the progress in stochastic integration, since the 1960s; see [Mey66], [Mey76].) Since the filtration is so basic to the definition of a stochastic process, the more modern term for a filtered probability space is a *stochastic basis*.

§4. DISCRETE-PARAMETER STOCHASTIC PROCESSES.

A *stochastic process* $X = \{X_t : t \in I\}$ is a family of random variables, defined on some common probability space, indexed by an index-set I. Usually (always in this course), I represents *time* (sometimes I represents *space*, and one calls X a spatial process). Here, $I = \{0, 1, 2, \cdots, T\}$ (finite horizon) or $I = \{0, 1, 2, \cdots\}$ (infinite horizon).

The (stochastic) process $X = (X_n)_{n=0}^\infty$ is said to be *adapted* to the filtration

$(\mathcal{F}_n)_{n=0}^{\infty}$ if

$$X_n \text{ is } \mathcal{F}_n - \text{measurable.}$$

So if X is adapted, we will know the value of X_n at time n. If $\mathcal{F}_n = \sigma(X_0, X_1, \cdots, X_n)$, we call $(\mathcal{F}_n)$ the *natural filtration* of X. Thus a process is always adapted to its natural filtration. A typical situation is that

$$\mathcal{F}_n = \sigma(W_0, W_1, \cdots, W_n)$$

is the natural filtration of some process $W = (W_n)$. Then X is adapted to $(\mathcal{F}_n)$, i.e. each X_n is $\mathcal{F}_n$- (or $\sigma(W_0, \cdots, W_n)$-) measurable, iff

$$X_n = f_n(W_0, W_1, \cdots, W_n)$$

for some measurable function f_n (non-random) of $n+1$ variables.

Notation. For a random variable X on $(\Omega, \mathcal{F}, P)$, $X(\omega)$ is the value X takes on ω (ω represents the randomness). Often, to simplify notation, ω is suppressed – e.g., we may write $EX := \int_\Omega X dP$ instead of $EX := \int_\Omega X(\omega) dP(\omega)$.

For a stochastic process $X = (X_n)$, it is convenient (e.g., if using suffices, n_i say) to use X_n, $X(n)$ interchangeably, and we shall feel free to do this. With ω displayed, these become $X_n(\omega)$, $X(n, \omega)$, etc.

§5. STOCHASTIC PROCESSES IN CONTINUOUS TIME

The underlying set-up is as before, but now *time is continuous rather than discrete*; thus the time-variable will be $t \geq 0$ in place of $n = 0, 1, 2, \ldots$. The information available at time t is the σ-field $\mathcal{F}_t$; the collection of these as $t \geq 0$ varies is the filtration, modelling the information flow. The underlying probability space, endowed with this filtration, gives us the stochastic basis (filtered probability space) on which we work,

We assume that the filtration is *complete* (contains all subsets of null-sets as null-sets), and *right-continuous*: $\mathcal{F}_t = \mathcal{F}_{t+}$, i.e.

$$\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$$

(the 'usual conditions' – right-continuity and completeness – in Meyer's terminology).

A *stochastic process* $X = (X_t)_{t \geq 0}$ is a family of random variables defined on a filtered probability space with X_t $\mathcal{F}_t$-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 $[0, \infty)$, $(X_{t_1}, \cdots, X_{t_n})$, or $(X(t_1), \cdots, X(t_n))$ (for typographical convenience, we use both notations interchangeably, with or without ω: $X_t(\omega)$, or $X(t, \omega)$) is a random n-vector, with a distribution,

$\mu(t_1, \cdots, t_n)$ say. The class of all such distributions as $\{t_1, \cdots, t_n\}$ 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:
(i) deletion of one point t_i can be obtained by 'integrating out the unwanted variable', as usual when passing from joint to marginal distributions,
(ii) permutation of the t_i permutes the arguments of the measure $\mu(t_1, \cdots, t_n)$ on $\mathbb{R}^n$.
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: P. J. Daniell in 1918, A. N. Kolmogorov in 1933).

Important though the Daniell-Kolmogorov theorem is as a general existence result, however, it does not take us very far. It gives a stochastic process 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 shall usually be able to confine attention to much smaller and more manageable spaces, of functions satisfying regularity conditions. The most important of these is *continuity*: we want to be able to realise $X = (X_t(\omega))_{t \geq 0}$ 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* (below), for example, and its relatives. Sometimes we need to allow our random function $X_t(\omega)$ 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* (below) and its relatives.

General results on realisability – whether or not it is possible to *realise*, or obtain, a process so as to have its paths in a particular function space – are known, but it is usually better to *construct* the processes we need directly on the function space on which they naturally live.

Given a stochastic process 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 results of this type (separability, measurability, versions, regularization, ...) see e.g. the classic book Doob [D]. The continuous-time theory is technically much harder than the discrete-time theory, for two reasons:
(i) questions of path-regularity arise in continuous time but not in discrete time,
(ii) *uncountable* operations (like taking sup over an interval) arise in continuous time. But measure theory is constructed using *countable* operations: uncountable operations risk losing measurability.

6. RENEWAL PROCESSES; POISSON PROCESS.

Suppose we use components – light-bulbs, say – whose lifetimes $X_1, X_2, \dots$ 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 we quote 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!).

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* (below – for predicting the future, knowing the present is enough: we do not need to know the past). 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 (processes with stationary independent increments) we shall find also:

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