liv4.tex **Week 4: 23.2.2016**

Kolmogorov's approach: conditional expectations via σ-fields

The problem with the approach above (Week 3: discrete and density cases) is that joint densities need not exist – do not exist, in general. One of the great contributions of Kolmogorov's classic book of 1933 was the realization that measure theory – specifically, the Radon-Nikodym theorem –provides a way to treat conditioning in general, without assuming that we are in the discrete case or 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: $E[Y] < \infty$. The set-function

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

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

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

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 (II.4), 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 II.4, 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 Kolmogorov (1933), 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

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

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 $E[Y|X_1, \cdots, X_n]$:

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

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 use $E[Y|\mathcal{B}]$ and $E_{\mathcal{B}}Y$ interchangeably.
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 II.6 below for properties of conditional expectations, and Chapter III for stochastic processes, particularly martingales [defined in terms of conditional expectations].

§6. 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'.

$$E[Y|\{\emptyset, \Omega\}] = E[Y].$$

Proof. We have to check (*) of §5 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: '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 (*), which is what we have to check (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 on it, 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. *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.

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

$$E[Y|\mathcal{B}] = E[Y] \qquad a.s.$$

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.

Projections.
In Property 5 (tower property), take $\mathcal{B} = \mathcal{C}$: we get the *Conditional Mean Formula*

$$E[E[X|\mathcal{C}]|\mathcal{C}] = E[X|\mathcal{C}].$$

This says that the operation of taking conditional expectation given a sub-σ-field $\mathcal{C}$ is *idempotent* – doing it twice is the same as doing it once. Also, taking conditional expectation is a *linear* operation (it is defined via an integral, and integration is linear). Recall from Linear Algebra that we have met such idempotent linear operations before. They are the *projections.* (Example: $(x, y, z) \mapsto (x, y, 0)$ projects from 3-dimensional space onto the (x, y)-plane.) This view of conditional expectation as projection is useful and powerful; see e.g. Neveu [N], [BK], [BF]. It is particularly useful when one has not yet got used to conditional expectation defined measure-theoretically as above, as it gives us an alternative (and perhaps more familiar) way to think.

Chapter III. STOCHASTIC PROCESSES IN DISCRETE TIME.

§1. 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 postpone the continuous case to Ch. V; in the discrete case, we may suppose time evolves in integer steps, $t = 0, 1, 2, \cdots$ (say, stock-market quotations daily, or tick data by the second). There may be a final time T, or *time horizon*, or we may have an infinite time horizon (in the context of option pricing, the time horizon T is the expiry time).

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* (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 Ch. II, 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.) 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.*

§2. 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 $E[X] := \int_\Omega X dP$ instead of $E[X] := \int_\Omega X(\omega) dP(\omega)$.

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.

§3. Discrete-Parameter Martingales.

We summarise what we need; for details, see [W], or e.g. [N]

Definition.

A process $X = (X_n)$ is called a *martingale* (mg for short) relative to $((\mathcal{F}_n), P)$ if
(i) X is adapted (to $(\mathcal{F}_n)$),
(ii) $E[|X_n|] < \infty$ for all n,
(iii) $E[X_n|\mathcal{F}_{n-1}] = X_{n-1} \qquad P-a.s. \qquad (n \geq 1)$;
X is a *supermartingale* 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* if in place of (iii)

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

Thus: 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. Martingales have many connections with harmonic functions in probabilistic potential theory. The terminology in the inequalities above comes from this: supermartingales correspond to superharmonic functions, submartingales to subharmonic functions.
2. X is a submg [supermg] iff $-X$ is a supermg [submg]; X is a mg iff it is both a submg and a supermg.
3. (X_n) is a mg iff $(X_n - X_0)$ is a mg. So we may without loss of generality take $X_0 = 0$ when convenient.
4. If X is a mg, then for $m < n$

$$\begin{aligned}
E[X_n|\mathcal{F}_m] &= E[E(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m] \qquad \text{(iterated conditional expectations)}\\
&= E[X_{n-1}|\mathcal{F}_m] \qquad a.s. \qquad \text{(martingale property)}\\
&= \cdots = E[X_m|\mathcal{F}_m] \qquad a.s. \qquad \text{(induction on } n\text{)},\\
&= X_m \qquad (X_m \text{ is } \mathcal{F}_m\text{-measurable})
\end{aligned}$$

and similarly for submartingales, supermartingales.
5. Examples of a mg include: sums of independent, integrable zero-mean random variables [submg: positive mean; supermg: negative mean].

From the OED: martingale (etymology unknown)
1. 1589. An article of harness, to control a horse's head.
2. Naut. A rope for guying down the jib-boom to the dolphin-striker.
3. A system of gambling which consists in doubling the stake when losing in order to recoup oneself (1815).
Thackeray: 'You have not played as yet? Do not do so; above all avoid a martingale if you do.'
Problem. Analyse this strategy.

Gambling games have been studied since time immemorial - indeed, the Pascal-Fermat correspondence of 1654 which started the subject was on a problem (de Méré's problem) related to gambling.

The doubling strategy above has been known at least since 1815.

The term 'mg' in our sense is due to J. VILLE (1939). Martingales were studied by Paul LÉVY (1886-1971) from 1934 on [see obituary, *Annals of Probability* **1** (1973), 5-6] and by J. L. DOOB (1910-2004) from 1940 on. The first systematic exposition was Doob's book [D], Ch. VII.
Example: Accumulating data about a random variable ([W], 96, 166-167).
If $X \in L_1(\Omega, \mathcal{F}, P)$, $M_n := E[X|\mathcal{F}_n]$ (so M_n represents our best estimate of X based on knowledge at time n), then

$$\begin{aligned} E[M_n|\mathcal{F}_{n-1}] &= E[E[X|\mathcal{F}_n]|\mathcal{F}_{n-1}] \\ &= E[X|\mathcal{F}_{n-1}] \quad \text{(iterated conditional expectations)} \\ &= M_{n-1}, \end{aligned}$$

so (M_n) is a mg. One has the convergence

$$M_n \to M_\infty := E[\xi|\mathcal{F}_\infty] \qquad a.s. \qquad \text{and in } L_1;$$

see II.4 below.

§4. Martingale Convergence.

A supermartingale is 'decreasing on average'. Recall that a decreasing sequence [of real numbers] that is bounded below converges (decreases to its greatest lower bound or infimum). This suggests that a supermartingale which is bounded below converges a.s. This is so [Doob's Forward Convergence Theorem: [W], §§11.5, 11.7].

More is true. Call X L_1-bounded if

$$\sup_n E[|X_n|] < \infty.$$

Theorem (Doob). An L_1-bounded supermartingale is a.s. convergent: there exists X_∞ finite such that

$$X_n \to X_\infty \qquad (n \to \infty) \qquad a.s.$$

In particular, we have

Doob's Martingale Convergence Theorem [W, §11.5]. An L_1-bounded martingale converges a.s.

We say that

$$X_n \to X_\infty \text{ in } L_1$$

if

$$E[|X_n - X_\infty|] \to 0 \qquad (n \to \infty).$$

For a class of martingales, one gets convergence in L_1 as well as almost surely [= with probability one]. Such martingales are called *uniformly integrable* (UI) [W], or *regular* [N], or *closed* (see below).

The following result is in [N], IV.2, [W], Ch. 14; cf. SP L18-19, SA L6.

Theorem (UI Martingale Convergence Theorem). The following are equivalent for martingales $X = (X_n)$:
(i) X_n converges in L_1,
(ii) X_n is L_1-bounded, and its a.s. limit X_∞ (which exists, by above) satisfies

$$X_n = E[X_\infty | \mathcal{F}_n],$$

(iii) There exists an integrable random variable X with

$$X_n = E[X | \mathcal{F}_n].$$

The random variable X_∞ above serves to "close" the martingale, by giving X_n a value at "$n = \infty$"; then $\{X_n : n = 1, 2, \ldots, \infty\}$ is again a martingale – which we may accordingly call a closed mg. The terms closed, regular and UI are used interchangeably here.

Notice that all the randomness in a closed mg is in the closing value X_∞ (so, although a stochastic process is an infinite-dimensional object, the randomness in a closed mg is one-dimensional). As time progresses, more is revealed, by "progressive revelation" – as in (choose your metaphor) a striptease, or the "Day of Judgement" (when all will be revealed).

As we shall see (Risk-Neutral Valuation Formula): closed mgs are vital in mathematical finance, and the closing value corresponds to the payoff of an option.

§5. Martingale Transforms.

Now think of a gambling game, or series of speculative investments, in discrete time. There is no play at time 0; there are plays at times $n = 1, 2, \cdots$, and

$$\Delta X_n := X_n - X_{n-1}$$

represents our net winnings per unit stake at play n. Thus if X_n is a martingale, the game is 'fair on average'.

Call a process $C = (C_n)_{n=1}^{\infty}$ *previsible* (or *predictable*) if

$$C_n \text{ is } \mathcal{F}_{n-1} - \text{measurable for all } n \geq 1.$$

Think of C_n as your stake on play n (C_0 is not defined, as there is no play at time 0). Previsibility says that you have to decide how much to stake on play n based on the history *before* time n (i.e., up to and including play $n-1$). Your winnings on game n are $C_n \Delta X_n = C_n(X_n - X_{n-1})$. Your total (net) winnings up to time n are

$$Y_n = \sum_{k=1}^{n} C_k \Delta X_k = \sum_{k=1}^{n} C_k (X_k - X_{k-1}).$$

We write

$$Y = C \bullet X, \qquad Y_n = (C \bullet X)_n, \qquad \Delta Y_n = C_n \Delta X_n$$

($(C \bullet X)_0 = 0$ as $\sum_1^0$ is empty), and call $C \bullet X$ the *martingale transform* of X by C.

Theorem. (i) If C is a bounded non-negative previsible process and X is a supermartingale, $C \bullet X$ is a supermartingale null at zero.

(ii) If C is bounded and previsible and X is a martingale, $C \bullet X$ is a martingale null at zero.

Proof. With $Y = C \bullet X$ as above,

$$E[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] = E[C_n(X_n - X_{n-1})|\mathcal{F}_{n-1}]$$
$$= C_n E[(X_n - X_{n-1})|\mathcal{F}_{n-1}]$$

(as C_n is bounded, so integrable, and $\mathcal{F}_{n-1}$-measurable, so can be taken out)

$$\leq 0$$

in case (i), as $C \geq 0$ and X is a supermartingale,

$$= 0$$

in case (ii), as X is a martingale. //

Interpretation. You can't beat the system!
In the martingale case, previsibility of C means we can't foresee the future (which is realistic and fair). So we expect to gain nothing – as we should.
Note. 1. Martingale transforms were introduced and studied by D. L. BURKHOLDER in 1966 [*Ann. Math. Statist.* **37**, 1494-1504]. For a textbook account, see e.g. [N], VIII.4.
2. Martingale transforms are the discrete analogues of stochastic integrals. They dominate the mathematical theory of finance in discrete time, just as stochastic integrals dominate the theory in continuous time.
3. In mathematical finance, X plays the role of a price process, C plays the role of our trading strategy, and the mg transform $C \bullet X$ plays the role of our gains (or losses!) from trading.

Proposition (Martingale Transform Lemma). An adapted sequence of real integrable random variables (M_n) is a martingale iff for any bounded previsible sequence (H_n),

$$E[\sum_{r=1}^{n} H_r \Delta M_r] = 0 \qquad (n = 1, 2, \cdots).$$

Proof. If (M_n) is a martingale, X defined by $X_0 = 0$, $X_n = \sum_1^n H_r \Delta M_r \qquad (n \geq 1)$ is the martingale transform $H \bullet M$, so is a martingale.

Conversely, if the condition of the Proposition holds, choose j, and for any $\mathcal{F}_j$-measurable set A write $H_n = 0$ for $n \neq j+1$, $H_{j+1} = I_A$. Then (H_n) is previsible, so the condition of the Proposition, $E[\sum_1^n H_r \Delta M_r] = 0$, becomes

$$E[I_A(M_{j+1} - M_j)] = 0.$$

As this holds for every $A \in \mathcal{F}_j$, the definition of conditional expectation gives

$$E[M_{j+1}|\mathcal{F}_j] = M_j.$$

Since this holds for every j, (M_n) is a martingale. //

§6. Stopping Times and Optional Stopping.

A random variable T taking values in $\{0, 1, 2, \cdots; +\infty\}$ is called a *stopping time* (or *optional time*) if

$$\{T \leq n\} = \{\omega : T(\omega) \leq n\} \in \mathcal{F}_n \qquad \forall n \leq \infty.$$

Equivalently,

$$\{T = n\} \in \mathcal{F}_n \qquad n \leq \infty.$$

Think of T as a time at which you decide to quit a gambling game: whether or not you quit at time n depends only on the history up to and including time n – NOT the future. [Elsewhere, T denotes the expiry time of an option. If we mean T to be a stopping time, we will say so.]

The following important classical theorem is discussed in [W], 10.10.

Theorem (Doob's Optional Stopping Theorem, OST). Let T be a stopping time, $X = (X_n)$ be a supermartingale, and assume that one of the following holds:
(i) T is bounded [$T(\omega) \leq K$ for some constant K and all $\omega \in \Omega$];
(ii) $X = (X_n)$ is bounded [$|X_n(\omega)| \leq K$ for some K and all n, ω];
(iii) $E[T] < \infty$ and $(X_n - X_{n-1})$ is bounded.

Then X_T is integrable, and

$$E[X_T] \leq E[X_0].$$

If here X is a martingale, then

$$E[X_T] = E[X_0].$$