Martingales

by D. Cox

December 2, 2009

1 Stochastic Processes.

Definition 1.1 *Let* $\mathbf{T}$ *be an arbitrary* index set*. A* stochastic process *indexed by* $\mathbf{T}$ *is a family of random variables* $(X_t : t \in \mathbf{T})$ *defined on a common probability space* $(\Omega, \mathcal{F}, P)$*. If* $\mathbf{T}$ *is clear from context, we will write* (X_t)*. If* $\mathbf{T}$ *is one of* $\mathbb{Z}$*,* $\mathbb{N}$*, or* $\mathbb{N} \setminus \{0\}$*, we usually call* (X_t) *a* discrete time *process. If* $\mathbf{T}$ *is an interval in* $\mathbb{R}$ *(usually* $\mathbb{R}$ *or* $[0, \infty)$*), then we usually call* (X_t) *a* continuous time *process.*

In a sense, all of probability is about stochastic processes. For instance, if $\mathbf{T} = \{1\}$, then we are just talking about a single random variable. If $\mathbf{T} = \{1, \ldots, n\}$, then we have a random vector $(X_1, \ldots, X_n)$. We have talked about many results for i.i.d. random variables X_1, X_2, ... Assuming an inifinite sequence of such r.v.s, $\mathbf{T} = \mathbb{N} \setminus \{0\}$ for this example. Given any sequence of r.v.s X_1, X_2, ..., we can define a *partial sum* process

$$S_n = \sum_{i=1}^{n} X_i, \quad n = 1, 2, \ldots.$$

One important question that arises about stochastic processes is whether they exist or not. For example, in the above, can we really claim there exists an infinite sequence of i.i.d. random variables? The product measure theorem tells us that for any valid marginal distribution P_X, we can construct any finite sequence of r.v.s with this marginal distribution. If such an infinite sequence of i.i.d. r.v.sr does not exist, we have stated a lot of meaniningless theorems. Fortunately, this is not the case. We shall state a theorem that shows stochastic processes exist as long as certain basic consistency properties hold.

In order to show existence, we will have to construct a probability space on which the r.v.s are defined. This requires us to first mathematically construct the underlying set Ω. The following will serve that purpose.

Definition 1.2 *Let* $\mathbf{T}$ *be an arbitrary index set. Then*

$$\mathbb{R}^{\mathbf{T}} = \{f : f \text{ is a function mapping } \mathbf{T} \longrightarrow \mathbb{R}\}.$$

Note that in the general definition of a stochastic process, for any "realization" $\omega \in \Omega$, $(X_t(\omega))$ is basically an element of $\mathbb{R}^{\mathbf{T}}$. Thus, a stochastic process may be thought of as a "random function" with domain $\mathbf{T}$ and range $\mathbb{R}$. Next, we need a σ-field.

Definition 1.3 *A* finite dimensional cylinder set $C \subset \mathbb{R}^{\mathbf{T}}$ *is a set of the form*

$$\exists\{t_1, \ldots, t_n\} \subset \mathbb{R}^{\mathbf{T}}, \quad \exists B_1, \ldots, B_n \in \mathcal{B}, \quad C = \left\{f \in \mathbb{R}^{\mathbf{T}} : f(t_i) \in B_i, 1 \leq i \leq n\right\}$$

Let $\mathcal{C}$ *denote the collection of all finite dimensional cylinder sets in* $\mathbb{R}^{\mathbf{T}}$. *Then the (canonical)* σ*-field on* $\mathbb{R}^{\mathbf{T}}$ *is*

$$\mathcal{B}_{\mathbf{T}} = \sigma(\mathcal{C}).$$

Before we can show the existence of probability measures on the measurable space $\left(\mathbb{R}^{\mathbf{T}}, \mathcal{B}^{\mathbf{T}}\right)$, we need to state the basic consistency properties such measures must satisfy. Any subsets $R \subset S \subset \mathbf{T}$, consider the *projection map* π_{SR} from $\mathbb{R}^S \longrightarrow \mathbb{R}^R$ defined by as the restriction of $f \in \mathbb{R}^S$ to R. More explicitly, if $f : S \longrightarrow \mathbb{R}$, and $g = \pi_{SR}(f) : R \longrightarrow \mathbb{R}$, then $g(t) = f(t)$ for all $t \in R$. We will denote $\pi_{\mathbf{T}R}$ by just π_R.

Definition 1.4 *A* consistent family of finite dimensional distributions *on* $I\!R^{\mathbf{T}}$ *is a family of probability measures* $\{P_S : S \subset \mathbf{T},\ S \text{ finite }\}$ *satisfying the property that for all* $R \subset S \subset \mathbf{T}$ *with both* S *and* R *finite,* $P_S \circ \pi_{RS}^{-1} = P_R$.

To explain the basic idea here, let $S = \{t_1, \ldots, t_n\}$. Then, if a process $(X_t : t \in \mathbf{T})$ exists, P_S is simply the (marginal) distribution of $(X_{t_1}, \ldots, X_{t_n})$. If $R = \{t_1, \ldots, t_k\} \subset S$, then the property above simply says that the marginal distribution P_R is consistent with P_S. The next result tells us that if this consistency condition holds, then there is a stochastic process with the given finite dimensional distributions.

Theorem 1.1 (Kolmogorov's Extension Theorem). *Let* $\{P_S : S \subset \mathbf{T},\ S \text{ finite }\}$ *be a consistent family of finite dimensional distributions. Then there exists a unique probability* P *measure on* $(I\!R^{\mathbf{T}}, \mathcal{B}_{\mathbf{T}})$ *such that for all finite* $S \subset \mathbf{T}$, $P \circ \pi_S^{-1} = P_S$.

For a proof, see either Ash or Billingsley. In fact, one may replace $I\!R$ by any complete and separable metric space. The theorem basically says that a stochastic process is determined by all of its finite dimensional distributions.

It is easy to show, for example, that if all of the finite dimensional distributions are measure products of a common distribution (i.e., everything is i.i.d.) then the consistency condition holds. Thus, we certainly have i.i.d. processes (with any index set!).

We close this section by noting that the above theorem does not solve all of the problems concerning stochastic processes. For example, if $\mathbf{T}$ is an interval of real numbers, we might be interested in whether (X_t) is a continuous function of t. It turns out that the set of continuous functions is not an element of $\mathcal{B}_{\mathbf{T}}$, i.e., it is not a measurable set in the probability space we constructed above.

2 Martingales: Basic Definitions.

For the rest of these notes, we will only consider discrete time stochastic processes indexed by either $I\!N$ or $I\!N \setminus \{0\}$. We shall use the subscript n to denote "time" rather than t.

Definition 2.1 *Given a probability space $(\Omega, \mathcal{F}, P)$, a* (discrete time) filtration *is an increasing sequence of sub-σ-fields $(\mathcal{F}_n : n \in I\!N)$ (or $(\mathcal{F}_n : n \in I\!N \setminus \{0\})$) of $\mathcal{F}$; i.e., all $\mathcal{F}_n \subset \mathcal{F}$ are σ-fields and $\mathcal{F}_n \subset \mathcal{F}_m$ if $n \leq m$.*

Given a process (X_n), we say (X_n) is adapted *to a filtration $(\mathcal{F}_n)$ (with the same index set) iff for all n, $X_n \in \mathcal{F}_n$ (i.e., X_n is $\mathcal{F}_n$-measurable, meaning $X_n^{-1}(B) \in \mathcal{F}_n$ for all Borel sets $B \subset I\!R$.*

Given any stochastic process (X_n), the filtration generated by (X_n), *or the* minimal filtration for (X_n), *is the filtration given by $\mathcal{F}_n = \sigma(X_m : m \leq n)$.*

When discussing processes, we will in general assume there is a filtration and the process is adapted; we can always use the minimal filtration for the given given process. For martingale theory, we will generally use $I\!N$ for the index set, and we assume $\mathcal{F}_0$ is an almost trivial σ-field, i.e. for all $A \in \mathcal{F}_0$, either $P(A) = 0$ or $P(A) = 1$. As the process will be adapted, this implies X_0 is constant, a.s.

Definition 2.2 *A process $(M_n : n \geq 0)$ is a* martingale *w.r.t. a filtration $(\mathcal{F}_n : n \geq 0)$ iff the following hold:*

(i) *(X_n) is adapted to $(\mathcal{F}_n)$;*

(ii) *For all n, $E[|X_n|] < \infty$;*

(iii) *For all n, $E[X_{n+1}|\mathcal{F}_n] = X_n$, a.s.*

We say (M_n) is a submartingale *iff properties (i) and (ii) hold, and property (iii) is replaced by*

$$\forall n, \quad E[X_{n+1}|\mathcal{F}_n] \;\geq\; X_n, \quad a.s.$$

We say (M_n) is a supermartingale *iff (i) and (ii) hold, and the reverse inequality above holds (i.e., $(-M_n)$ is a submartingale.)*

Note that to check a process is a martingale, it suffices to check property (iii) (which is usually called "the martingale property") since if it holds, then the conditional expectation makes sense, so (ii) holds, and since the conditional expectation is measurable with respect to the σ-field being conditioned on, it follows X_n is $\mathcal{F}_n$-measurable (up to sets of measure 0, which can always be finessed away; i.e., we can change the definition of X_n on a null set so as to make it $\mathcal{F}_n$ measurable). For sub- and supermartingales, it is necessary to check (i) and (iii) (since (iii) won't make sense unless (ii) holds). Some authors use the term "smartingale" to refer to a process which is either a martingale, a submartingale, or a supermartingale.

A martingale may be thought of as a "fair game" in the following sense: if X_n denotes the total amount you have won on the n^{th} play of a game, then, given all of the information in the current and previous plays (represented by $\mathcal{F}_n$), you don't expect to change your total winning. A submartingale would be a game which is not fair to your opponent (if X_n denotes the total amount you have won), and a supermartingale would be not fair to you.

One of the main reasons that martingale theory has become so useful is that martingales may be "found" in many probability models. Here are a few examples.

Example 2.1 Let X_1, X_2, ..., be independent r.v.s with $E[X_i] = \mu_i$. Define the partial sum process

$$S_0 = 0, \quad S_n \;=\; \sum_{i=1}^{n} X_i, \quad n = 1, 2, \ldots.$$

Let $\mathcal{F}_n$ be the minimal filtration for X_n (with $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial σ-field). If $\mu = 0$, then we claim S_n is a martingale. To check this, note that

$$\begin{aligned} E[S_{n+1}|X_1, \dots X_n] &= E\left[X_{n+1} + S_n \middle| X_1, \dots X_n\right] \\ &= E\left[X_{n+1} \middle| X_1, \dots X_n\right] + S_n \\ &= E[X_{n+1}] + S_n \\ &= S_n. \end{aligned}$$

The second line follows since $S_n \in \sigma(X_1, \dots, X_n)$ (see Theorem 1.5.7(f)), and next line by the independence assumption (see Theorem 1.5.9). Clearly, in general $M_n = S_n - \sum_{i=1}^n \mu_i$ is a martingale.

Example 2.2 Another construction which is often used is what might be called "partial product" processes. Suppose X_1, X_2, ... are independent with $E[X_i] = 1$. Let $M_n = \prod_{i=1}^n X_i$. Again using the minimal filtration for the (X_n) process, we have

$$\begin{aligned} E\left[M_{n+1} \middle| X_1, \dots, X_n\right] &= E\left[X_{n+1} M_n \middle| X_1, \dots, X_n\right] \\ &= M_n E\left[X_{n+1} \middle| X_1, \dots, X_n\right] \\ &= M_n. \end{aligned}$$

Again, at the second line we used one of the basic results on conditional expectation (see Theorem 1.5.7(h)).

Example 2.3 Let X be a r.v. with $E[|X|] < \infty$ and let $(\mathcal{F}_n)$ be any filtration (with $\mathcal{F}_0$ an almost trivial σ-field). Let $X_n = E\left[X \,|\mathcal{F}_n\right]$. Then (X_n) is martingale. See Exercise 3.

Example 2.4 Let $(X_n : n \geq 0)$ be an arbitrary process adapted to a filtration $(\mathcal{F}_n : n \geq 0)$. Assume that for all n, $E[|X_n|] < \infty$. For $n > 0$ define

$$Y_n = X_n - E[X_n|\mathcal{F}_{n-1}].$$

Put $M_0 = 0$ and for $n > 0$ let

$$M_n = \sum_{i=1}^{n} Y_i.$$

Then $(M_n : n \geq 0)$ is a martingale w.r.t. the filtration $(\mathcal{F}_n : n \geq 0)$. See Exercise 4.

3 The Optional Stopping Theorem.

Our main result in this section is not difficult and shows the power of martingale theory. We first need a very important definition.

Definition 3.1 *Let $(\mathcal{F}_n : n \in I\!N)$ be a filtration and let T be an $(I\!N \cup \{\infty\})$-valued random variable. Then T is called a* stopping time *w.r.t $(\mathcal{F}_n)$ iff for all $n \in I\!N$, the event $[T \leq n]$ is in $\mathcal{F}_n$.*

If (X_n) is adapted and $P[T < \infty] = 1$, then the stopped value of the process *is*

$$X_T = \sum_{n=0}^{\infty} I[T = n] X_n.$$

(We will write $I[\cdots]$ for the indicator $I_{[\ldots]}$ sometimes.)

The process $(X_{n \wedge T} : n \geq 0)$ is called the stopped process. *(Recall that $a \wedge b = \min\{a, b\}$.)*

Proposition 3.1 *If T_1 and T_2 are stopping times w.r.t $(\mathcal{F}_n)$, then so are $T_1 + T_2$, $T_1 \wedge T_2$, and $T_1 \vee T_2$.*

Proposition 3.2 *T is a stopping time if and only if for all $n \in I\!N$, $[T = n] \in \mathcal{F}_n$.*

Proof: $(\Rightarrow)$ Assume T is a stopping time. We have $[T = n] = [T \leq n] \cap [T \leq n-1]^c$, and both events in the last expression are in $\mathcal{F}_n$, so their intersection is also.

($\Leftarrow$) Assume for all n $[T = n] \in \mathcal{F}_n$. Then

$$[T \leq n] \ = \ \bigcap_{i=0}^{n} [T = i].$$

All of the events in the intersection are in $\mathcal{F}_n$, so also is $[T \leq n]$.

$\square$

Many of our stopping times will be of the following type.

Definition 3.2 *Suppose* (X_t) *is adapted to* $(\mathcal{F}_t)$*, and let* $B \subset I\!R$ *be a Borel set. The* hitting time *or* first entry time *to* B *is*

$$T_B \ = \ \inf\{n \in I\!N : X_n \in B\}.$$

Recall that by convention, $\inf \emptyset = \infty$*.*

Proposition 3.3 *A hitting time is a stopping time.*

Proof: Note that

$$[T_B = n] \ = \ [X_n \in B] \cap \bigcap_{i=0}^{n-1} [X_i \in B]^c.$$

Of course $[X_n \in B] \in \mathcal{F}_n$, and for $i < n$, $[X_i \in B]^c \in \mathcal{F}_i \subset \mathcal{F}_n$, so $[T_B = n] \in \mathcal{F}_n$.

$\square$

Before stating and proving the big result, it is useful to have the next one, which has many useful ramifications. First, a couple of definitions.

Definition 3.3 *A process* $(A_n : n \geq 1)$ *is called* non-anticipating *(or* pridictable*, or sometimes* previsible*) iff for all* $n \geq 1$, $A_n \in \mathcal{F}_{n-1}$*; i.e., the process* $X_n = A_{n+1}$, $n \geq 0$ *is adapted.*

We will also need the backwards difference operator defined by

$$\Delta M_n \;=\; M_n - M_{n-1}, \quad n \geq 1.$$

The process (ΔM_n) is sometimes called a *martingale difference process.* The defining property for such a process is

$$E[\Delta M_{n+1}|\mathcal{F}_n] \;=\; E[M_{n+1} - M_n|\mathcal{F}_n] \;=\; M_n - M_n \;=\; 0, \quad a.s.$$

Theorem 3.4 *Suppose $(M_n : n \geq 0)$ is a martingale w.r.t. $(\mathcal{F}_n : n \geq 0)$ and $(A_n : n \geq 1)$ is bounded non-anticipating w.r.t. $(\mathcal{F}_n)$. Then the process*

$$\tilde{M}_n \;=\; \sum_{i=1}^{n} A_i \Delta M_i,$$

(with $M_0 = 0$), which is called the martingale transform of (M_n) w.r.t. (A_n), *is a martingale w.r.t. $(\mathcal{F}_n)$.*

Proof: Using the boundedness of A_n (say, $|A_n| \leq K$), we have

$$E[|\tilde{M}_n \;\leq\; K \sum_{i=1}^{n} (E[|M_i|] + E[|M_{i-1}|]) \;<\; \infty.$$

Checking the martingale property

$$\begin{aligned}
E\left[\tilde{M}_{n+1}\middle|\,\mathcal{F}_n\right] &= E\left[A_{n+1}\Delta M_{n+1} + \sum_{i=1}^{n} A_i \Delta M_i\middle|\,\mathcal{F}_n\right] \\
&= A_{n+1} E\left[\Delta M_{n+1}|\,\mathcal{F}_n\right] \;+\; \sum_{i=1}^{n} A_i \Delta M_i \\
&= 0 \;+\; \sum_{i=1}^{n} A_i \Delta M_i \\
&= \tilde{M}_n.
\end{aligned}$$

The second line follows from the facts about conditional expectation and that (A_n) is non-anticipating and (M_n) is adapted. The third line is the martingale difference property.

$\square$

Now we can state our big result.

Theorem 3.5 (Optional Stopping Theorem.) *Let T be a stopping time and (M_n) a martingale w.r.t. $(\mathcal{F}_n)$. Then the stopped process $(M_{n\wedge T})$ is also a martingale.*

Proof: We begin with the assumption that $E[M_n] = 0$. Note that $I[T \geq n] = 1 - I[T \leq n-1]$ is bounded and non-anticipating. Thus

$$\tilde{M}_n \;=\; \sum_{i=1}^{n} I[T \geq n]\Delta M_n$$

is a martingale by the previous theorem. We will show in fact that $\tilde{M}_n = M_{n\wedge T}$, which will prove the result. This claim follows by "partial summation," which is analogous to integration by parts. If one lists out the summands as (note that $M_0 = 0$ by our assumption)

$$\begin{aligned}
\tilde{M}_n \;&=\; I[T \geq 1]M_1 + I[T \geq 2](M_2 - M_1) + I[T \geq 3](M_3 - M_2) + \ldots \\
&\qquad I[T \geq n-1](M_{n-1} - M_{n-2}) + I[T \geq n](M_n - M_{n-1}) \\
&=\; (I[T \geq 1] - I[T \geq 2])M_1 + (I[T \geq 2] - I[T \geq 3])M_2 + \ldots \\
&\qquad (I[T \geq n-1] - I[T \geq n])M_{n-1} + I[T \geq n]M_n \\
&=\; \sum_{i=1}^{n-1} I[T = i]M_i \;+\; I[T \geq n]M_n.
\end{aligned}$$

Of course, if $T \geq n$, then $T \wedge n = n$, and if $T = i < n$, then $T \wedge n = i$, so the last expression is equal to

$$\sum_{i=1}^{n} I[(T \wedge n) = i]M_i \;=\; M_{n\wedge T}.$$

If $E[M_n] \neq 0$, then apply the above argument to $M'_n = M_n - E[M_n]$. The resulting $\tilde{M}'_n$ is a mean 0 martingale, and it is clear that the corresponding $\tilde{M}_n = \tilde{M}'_n + E[M_n]$ $= M_{n\wedge T}$.

$\square$

More general versions of the optimal stopping theorem can be found; see e.g. Ash. This version is relatively elementary to prove and still very powerful, as we shall see in some examples.

Example 3.1 (Unbiased Gambler's Ruin.) Suppose you play a game with your opponent. The plays are i.i.d. with your winning on each play either ± 1 with equal probability. You begin with a total wealth of a and your opponent with b. We assume a and b are positive integers. Let us calculate the probability that you bankrupt your opponent before he bankrupts you. Letting X_n denote the outcome of the n^{th} play, we have $P[X_n = \pm 1] = 1/2$. The total winning is

$$S_n \;=\; \sum_{i=1}^{n} X_i.$$

Since $E[X_i] = 0$ and they are independent, we have already seen this is a martingale. The game will stop at the time

$$T \;=\; \inf\{n : S_n = -a \text{ or } S_n = b\}. \tag{1}$$

As this is a hitting time (of $(-\infty, -a] \cup [b, \infty)$) for an adapted process (we are using the filtration generated by the (X_n)), it is a stopping time. We claim $T < \infty$ a.s. Then, as $n \to \infty$, $S_{T\wedge n} \to S_T$ a.s. (Simply note that for ω in $[T < \infty]$, $T(\omega) \wedge n$ $= T(\omega)$ for all $n \geq T(\omega)$.) Also, $\forall n$, $|S_{T\wedge n}| \leq a \vee b$, so by dominated convergence we have $E[S_{T\wedge n}] \to E[S_T]$. Now S_T only takes on two values. Let $w = P[S_T = b]$ (the probability you win and your opponent is ruined). Then, since $S_{T\wedge n}$ is a mean 0 martingale, we have

$$0 = \lim E[S_{T\wedge n}] = wb + (1-w)(-a) \quad\Longrightarrow\quad w = a/(a+b).$$

Thus, if your initial fortune is larger than your opponent's (i.e., $a > b$), then you have more than 1/2 probability of ruining your opponent.

To complete the argument, we must show $T < \infty$ a.s. Let $N = a + b$ and for $k = 1, 2, \ldots$ define events

$$A_k \;=\; [X_j = 1 \text{ for } (k-1)N < j \leq kN].$$

Note that this entails of run of $a + b$ play where you win all of them. Clearly if A_k occurs, and if both players are still not ruined, you will ruin your opponent, so either there was a ruin prior to the event occuring or it will occur during or after the event. Note that the A_k are independent events (they involve non-overlapping blocks of the X_i), $P(A_k) = 2^{-N}$ for all k, and so $\sum_k A_k = \infty$. By the Borel-Cantelli Lemma, part II, the A_k must occur infinitely often, so they occur at least once, and hence ruin is assured with probability 1.

Example 3.2 (Biased Gambler's Ruin.) Now we consider the same problem as in the previous example except we change the probability of you winning a play of from 1/2. Let $P[X = 1] = p$ and $P[X = -1] = q$ where $q = 1 - p$. We assume $p \neq 1/2$. Also, the cases $p = 0, 1$ are not interesting as they mean almost certain ruin for one player in a constant number of moves. Now the martingale used above no longer applies, but we can try to find a useful "partial product" martingale. Specifically, we seek a constant r such that

$$E\left[r^{X_i}\right] \;=\; 1.$$

If we can find such a constant, then

$$M_n \;=\; \prod_{i=1}^{n} r^{X_i} \;=\; r^{S_n}$$

will be a martingale, and we can try to use the Optional Stopping Theorem again. Such an r must satisfy the equation

$$1 \;=\; E\left[r^{X_i}\right] \;=\; pr + qr^{-1}.$$

This is easily converted to a quadratic equation. Clearly $r = 1$ is one root of the equation (but one that doesn't help us), and it is easy to see $r = q/p$ is the other, and this works. Thus, but optional stopping and constancy of the expectation of a martingale

$$1 \;=\; E\left[r^{S_{n\wedge T}}\right].$$

It is easy to check that $T < \infty$ a.s. (the probability of the events A_k is now $p^N > 0$), so $M_{n\wedge T} \to M_T$ a.s. Also,

$$0 \;\le\; M_{T\wedge n} \;\le\; [(q/p) \vee (p/q)]^{(a\vee b)},$$

so dominated convergence applies again and we have

$$1 \;=\; E\left[M_{T\wedge n}\right] \;\to\; E\left[M_T\right].$$

But by direct calculation

$$E\left[M_T\right] \;=\; w(q/p)^b + (1-w)(q/p)^{-a} \;=\; 1.$$

Solving for w gives

$$w \;=\; \frac{(q/p)^a - 1}{(q/p)^{a+b} - 1}.$$

As a check, note that if $a = b$, this can be simplified to $w = 1/[(q/p)^a + 1]$, so if $q > p$ your chances of being ruined before your opponent is $> 1/2$, which is clearly correct. Also, as $a \to \infty$, your chances of ruin are almost certain, which makes sense, since if both you and your opponent are very wealthy, it will take a long time for ruin to occur and his advantage on each individual play will become more pronounced in the long run.

4 Martingale Convergence.

We will show that there are some simple, general conditions under which a martingale will converge a.s. to a fixed r.v. The proof involves the use of submartingales, which we haven't discussed too much up to this point. First, we consider a general way of constructing submartingales. We will need part (a) of the following proposition.

Proposition 4.1 *Assume the process X_n is a smartingale w.r.t the filtration $\mathcal{F}_n$. Let ϕ be a convex function defined on an interval (a,b), $-\infty < a < b < \infty$, and suppose $\forall n$, $P[X_n \in (a,b)] = 1$. Assume $\forall n$, $E[|\phi(X_n)|] < \infty$.*

(a) *If X_n is a martingale then $\phi(X_n)$ is a submartingale.*

(b) *If X_n is a submartingale and ϕ is nondecreasing, then $\phi(X_n)$ is a submartingale.*

Proof: Clearly $\phi(X_n)$ is adapted, and property (ii) in the definition of a smartingale holds by assumption. Jensen's inequality applies, so we have

$$E[\phi(X_{n+1})|\mathcal{F}_n] \ \geq \ \phi\left(E[X_{n+1}|\mathcal{F}_n]\right). \tag{2}$$

If X_n is a martingale, then the last expression is $\phi(X_n)$, thus showing the submartingale property. If X_n is a submartingale, then the submartingale property is that $E[X_{n+1}|\mathcal{F}_n] \geq X_n$. If ϕ is nondecreasing then it follows that the last expression in (2) is $\geq \phi(X_n)$, thus showing the submartingale property for $\phi(X_n)$.

$\square$

Example 4.1 It is easy to write down several transformations that might be interesting. If M_n is a martingale, then $|M_n|$ and $(M_n)_\pm$ (the positive or negative parts of M_n) are submartingales. Assuming integrability, M_n^2 and $\exp[aM_n]$ are also submartingales. For some of these transformations, if M_n is a submartingale, then so is the transformed process.

Theorem 4.2 (Martingale Convergence Theorem.) *If M_n is a martingale and there exists $\lambda > 0$ such $\forall n$, $E[|M_n|] \leq \lambda$, then there is a r.v. M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$ and $E[|M_\infty|] \leq \lambda$.*

Before giving the proof, we review some basic notions about convergence of a sequence of real numbers. The sequence $(a_n : n = 1, 2, \ldots)$ converges if and only if $\liminf_n a_n = \limsup_n a_n$, and the common value is $\lim_n a_n$. Of course $\liminf_n a_n$ is the smallest limit point of the sequence (a_n) (a limit point is the limit of any subsequence), and $\limsup_n a_n$ is the largest limit point.

Therefore, if (a_n) doesn't converge, then $\liminf_n a_n < \limsup_n a_n$, and thus we can find rational numbers c and d such that

$$\liminf_n a_n \; < \; c \; < \; d \; < \; \limsup_n a_n.$$

Now, we can find subsequences, say a_{n_j} and a_{m_k} such that $\lim_j a_{n_j} = \liminf_n a_n$ and $\lim_k a_{m_k} = \limsup_n a_n$. By selecting further subsequences if necessary, we can in fact insure that

(i)

$$\forall j, \; a_{n_j} < c, \; \text{and} \; \forall k, \; a_{m_k} > d.$$

(ii)

$$n_1 < m_1 < n_2 < m_2 < \cdots < n_j < m_j < n_{j+1} < m_{j+1} < \cdots.$$

The basic notion is that if sequence (a_n) doesn't have a limit, then there exist rationals $c < d$ such that infinitely often the sequence is below c but then at some later value is above d. This motivates the following definition. Given and numbers $c < d$, the *number of upcrossings of $[c, d]$ by the finite sequence a_0, a_1, …, a_N* is the largest k such that there exists $2k$ integers $0 \leq n_1 < m_1 < \cdots < n_k < m_k \leq N$ such that for all j, $1 \leq j \leq k$, $a_{n_j} < c$ and $a_{m_j} > d$. The sequence (a_n) converges if and only if the number of upcrossings of any rational interval is finite. (We can limit ourselves to rational intervals so in a proof that something happens with probability 1, we have only countably many null events to add up.) Note that the limit may be $\pm\infty$.

Lemma 4.3 (Upcrossing Inequality.) *Given a submartingale (M_n), define the r.v. $U_n([c,d])$ to be the number of upcrossings of $[c,d]$ by the finite sequence M_0, M_1, ..., M_n. Then*

$$(d-c)E[U_n([c,d])] \ \leq \ E[(M_n - c)_+].$$

Proof: The proof relies on constructing a non-anticipating process A_n and formally applying a martingale transform to the submartingale M_n w.r.t. A_n. (One canshow that the transform is in fact a submartingale.) The process A_n will be essentially an indicator of an upcrossing currently in progress. We will actually count upcrossings of $(c,d]$ rather than $[c,d]$; clearly there will be more of the former than the latter. Note that $(M_n - c)_+$ is a nonnegative submartingale, and the upcrossings by this process of $(0, d-c]$ are the same as the upcrossings by the original process of $(c,d]$. Thus, without loss of generality we may assume $M_n \geq 0$ and $c = 0$. We define A_n recursively (recall that the index of a non-anticipating process begins at $n = 1$).

$$A_1 \quad = \quad \begin{cases} 0 & \text{if } M_0 \geq 0; \\ 1 & \text{if } M_0 = 0. \end{cases}$$

For $n \geq 1$,

$$A_{n+1} \ = \ \begin{cases} 0 & \text{if } A_n = 0 \,\&\, M_n > 0, \text{ or } A_n = 1 \,\&\, M_n > d; \\ 1 & \text{if } A_n = 1 \,\&\, M_n \leq d, \text{ or } A_n = 0 \,\&\, M_n = 0. \end{cases}$$

It is not clear if explaining in words will make matters clearer, or if the reader should simply stare at the above to make sure A_n is 0 if an upcrossing is not in progress and is 1 if an upcrossing is underway. An upcrossing begins right after the first time (after beginning or after the last upcrossing ends) that M_n hits the level 0. It continues until the first time M_n goes above d. It is clear that A_n is non-anticipating since it only depends on A_{n-1} and M_{n-1}.

Now let $\tilde{M}_n$ be given by the martingale transform

$$\tilde{M}_n = \sum_{i=1}^{n} A_i \Delta M_i. \tag{3}$$

Let $0 \le n_1 < m_1 \le n_2 < m_2, \ldots,$ denote the beginning and ending times of the upcrossings (upcrossings begin at the n_j and end at the m_j). Then $A_i = 1$ if and only if for some j, $n_j < i \le m_j$, and otherwise $A_i = 0$. Thus the sum defining $\tilde{M}_n$ may be written as sums of blocks of the form

$$\sum_{i=n_j+1}^{m_j} A_i \Delta M_i = \sum_{i=n_j+1}^{m_j} (M_i - M_{i-1}) = M_{m_j} - M_{n_j} = M_{m_j} \ge d.$$

Note that for any n, it may happen that for some j, $n_j < n < m_j$, i.e., an upcrossing is underway but not yet completed at time n. In this case $\tilde{M}_n$ will involve an additional block whose value is $M_n - M_{n_j}$. Note that $M_n \ge 0$ and $M_{n_j} = 0$ (after our modification of the original process by replacing it with $(M_n - c)_+$), so $M_n - M_{n_j}$ is nonnegative and leaving it out of the summation simply makes the result possibly smaller. In summary, each upcrossing contributes no more than d to $\tilde{M}_n$, and we may ignore an upcrossing underway at time n to get

$$\tilde{M}_n \ge dU_n((0, d]).$$

Once we show $E[\tilde{M}_n] \le E[M_n]$, the lemma will be proved.

We have

$$\begin{aligned} E\left[\tilde{M}_n\right] &= \sum_{i=1}^{n} E[A_i \Delta M_i] \\ &= \sum_{i=1}^{n} E\left[E\left[A_i \left(M_i - M_{i-1}\right) \middle| \mathcal{F}_{i-1}\right]\right] \\ &= \sum_{i=1}^{n} E\left[A_i \left(E[M_i | \mathcal{F}_{i-1}] - M_{i-1}\right)\right] \end{aligned}$$

The second line uses the "law of total expectation," (Theorem 1.5.7(d)), and the third line uses uses another basic result on conditional expectation (Theorem 1.5.7(h)). By the submartingale property, $E[M_i|\mathcal{F}_{i-1}] - M_{i-1} \geq 0$. Since $A_i \in \{0,1\}$ we have

$$A_i \left(E[M_i|\mathcal{F}_{i-1}] - M_{i-1}\right) \;\leq\; \left(E[M_i|\mathcal{F}_{i-1}] - M_{i-1}\right).$$

Thus,

$$\begin{aligned} E\left[\tilde{M}_n\right] &\leq \sum_{i=1}^{n} E\left[E[M_i|\mathcal{F}_{i-1}] - M_{i-1}\right] \\ &= E\left[\sum_{i=1}^{n} E\left[M_i - M_{i-1}|\,\mathcal{F}_{i-1}\right]\right] \\ &= E\left[M_n - M_0\right] \\ &\leq E\left[M_n\right]. \end{aligned}$$

The last line follows since $M_0 \geq 0$. This completes the proof.

$\square$

Theorem 4.4 (Martingale Convergence Theorem.) *Let M_n be a martingale and suppose there is a $B < \infty$ such that $\forall n$, $E[|M_n|] \leq B$. Then there is a r.v. M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$ and $E[|M_\infty|] \leq B$.*

Proof: We will show that the number of upcrossings of any interval with rational endpoints is finite a.s., which will imply the existence of an extended r.v. M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$. By the upcrossing inequality, if $c < d$

$$E[U_n([c,d])] \;\leq\; E[(M_n - c)_+]/(d-c) \;\leq\; (B + |c|)/(d-c).$$

Note that the last expression is independent of n. Now as n increases, $0 \leq U_n([c,d])$ increases, so by Monotone Convergence Theorem $U_n \to U_\infty$ and $E[U_n] \to E[U_\infty]$. But our bound on $E[U_n]$ implies $E[U_\infty]$ is finite, and hence U_∞ is finite a.s., i.e., the total number of upcrossings if finite a.s., as claimed.

Now we show that M_∞ is finite a.s., and the bound on $E[|M_\infty|]$ holds. Note that by continuous mapping, $|M_n| \stackrel{a.s.}{\rightarrow} |M_\infty|$, and since $0 \leq |M_n|$, we have by Fatou's lemma that $E[|M_\infty|] \leq \liminf E[|M_n|] \leq B$. This establishes that $|M_\infty|$ is finite a.s., and the bound on its expectation.

$\square$

.

Example 4.2 Let M_n be an arbitrary martingale, and for any $a < b$, define the stopping time

$$T \;=\; \inf\{n : M_n \geq b \text{ or } M_n \leq a\}.$$

Now we know $M_{n\wedge T}$ is a martingale by the optional stopping theorem, but this martingale is also bounded, hence satisfies the conditions of the martingale convergence theorem. Thus, on the event $[\forall n, a < M_n < b] = [T = \infty]$, the process must converge a.s. to a constant.

If M_n is integer valued, the above implies that on the event $[T = \infty]$, M_n must eventually be a constant. In particular, if $\forall n$, $P[M_{n+1} = M_n] = 0$ (as was the case in the gambler's ruin example), we must have $T < \infty$ a.s. Thus, with a few simple assumptions, we can get some very general results about a martingale.

Exercises

1 Let $\mathbf{T}$ be an arbitrary index set and let $\mu : \mathbf{T} \to I\!R$ and $V : \mathbf{T} \times \mathbf{T} \to I\!R$. Assume that V satisfies the property that for any finite subset $S = \{t_1, \ldots, t_n)\} \subset \mathbf{T}$, the $n \times n$ matrix

$$V_{ij} \;=\; V(t_i, t_j), \quad 1 \leq i, j \leq n,$$

is symmetric and nonnegative definite. Now consider the family of finite dimensional distributions which for any finite S as above are multivariate normal with mean $(\mu(t_1), \ldots \mu(t_n))$ and covariance matrix V as above. Show that the family satisfies the consistency property, and conclude that there is a stochastic process with these as the finite dimensional distributions. This process is called the Gaussian process with mean function μ and covariance function V.

2 **(a)** Assume $(X_n : n \geq 0)$ is a martingale w.r.t. the filtration $(\mathcal{F}_n : n \geq 0)$ where all $A \in \mathcal{F}_0$ satisfy $P(A) = 0$ or 1. Show the following results:

(i) For all $k \geq 0$, $E[X_{n+k}|\mathcal{F}_n] = X_n$, a.s.

(ii) For all $n \geq 0$, $E[X_n] = X_0$, a.s., and X_0 is constant a.s.

(b) Give appropriate extensions of the properties in part (a) to submartingales.

3 Prove that the process (X_n) in Example 2.3 is indeed a martingale.

4 Prove that the process (M_n) in Example 2.4 is a martingale.

5 Prove Proposition 3.1.

6 Let $\mathcal{F}_n$ be a filtration and let A_1, A_2, ... be a sequence if independent events such that $\forall n$, $A_n \in \mathcal{F}_n$, and

$$\phi(n) \;=\; \sum_{i=1}^{n} P(A_i) \;\rightarrow\; \infty, \text{ as } n \rightarrow \infty.$$

Let $X_n = \sum_{i=1}^{n} I_{A_i}$. Fix a positive integer k and let

$$T \;=\; \inf\{n \geq 1 : X_n \;=\; k\}.$$

That is, T is the first time k of the events have occurred. Show that $T < \infty$ a.s., and

$$E[\phi(T)] = k.$$

7 **(a)** Let X_1, X_2, ... be i.i.d. r.v.s with $E[X_i] = 0$ and $0 < \sigma^2 = E[X_i^2] < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Show that $M_n = S_n^2 - n\sigma^2$ is a martingale w.r.t. the minimal filtration of the X_ns.

(b) Suppose that $P[X_i = 1] = P[X_i = -1] = 1/2$. Let a and b be positive integers and define the stopping time T as in equation (1). Show that $E[T] = ab$.

8 Let X_n denote the number of organisms in a population. Note that if $X_n = 0$ at some time, the population becomes extinct (i.e. $X_{n+m} = 0$ for all $m \geq 0$). Suppose that for every integer $N \geq 0$, there exists $\delta > 0$ such that for all n,

$$P[X_{n+1} = 0 | X_1 = x_1, \ldots, X_n = x_n] \geq \delta, \text{ if } x_n \leq N.$$

Let F be the event of extinction, i.e. $F = \bigcup_{n=1}^\infty [X_n = 0]$. Let G be the event $[X_n \to \infty]$. Show that $P(F) + P(G) = 1$. (We leave it to the reader to ponder the philosophical meaning of this if the environment is bounded so that $X_n \to \infty$ can't occur in practice.)

9 **(Doob's Martingale)** Let $\mathcal{F}_n$ be a filtration and let Y be any r.v. satisfying $E[|Y|] < \infty$. Put $M_n = E[Y|\mathcal{F}_n]$.

(a) Show that M_n is a martingale w.r.t. $\mathcal{F}_n$.

(b) Show that there exists a r.v. M_∞ such that $M_n \stackrel{a.s.}{\to} M_\infty$, and $E[|M_\infty|] \leq E[|Y|]$.

(c) Suppose there is a $K > 0$ such that $|Y| \leq K$ a.s. Show that $M_\infty = E[Y|\mathcal{F}_\infty]$ a.s., where $\mathcal{F}_\infty = \sigma\left(\bigcup_{n=1}^\infty \mathcal{F}_n\right)$. (Note: the result holds without assuming $|Y|$ is bounded a.s. but the proof requires results we have not given here.)

(d) **(Consistency of Bayesian Estimators.)** Suppose Θ is a random parameter, and there is a $K > 0$ such that $|\Theta| \leq K$ a.s. Once Θ is selected, data X_1, X_2, ... are generated, whose distribution depends on Θ. (We make no particular assumptions about these data.) Let $\mathcal{F}_n$ be the filtration generated by the X_ns. Assume there is a strongly consistent estimator of Θ, i.e., a sequence of functions $\hat{\theta}_n : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that

$$\hat{\theta}_n(X_1, \ldots, X_n) \stackrel{a.s.}{\to} \Theta.$$

Show that the posterior mean is a consistent estimator of Θ, i.e.

$$E[\Theta|\mathcal{F}_n] \stackrel{a.s.}{\to} \Theta.$$