Stochastic Calculus Notes, Lecture 3

Last modified February 20, 2007

1 Martingales and stopping times

1.1. Introduction: Martingales and stopping times are inportant technical tools used in the study of stochastic processes such as Markov chains and diffusions. A *martingale* is a stochastic process that is always unpredictable in the sense that $E[F_{t+t'} \mid \mathcal{F}_t] = F_t$ (see below) if $t' > 0$. A *stopping time* is a random "time", $\tau(\omega)$, so that we know at time t whether to stop, i.e. the event $\{\tau \leq t\}$ is measurable in $\mathcal{F}_t$. These tools work well together because stopping a martingale at a stopping also has mean zero: if $t \leq \tau \leq t'$, then $E\left[F_\tau \mid \mathcal{F}_t\right] = F_t$. A central fact about the Ito calculus is that Ito integrals with respect to Brownian motion are martingales. Time will be discrete throughout this lecture. The variable t will take values $t = 0, 1, 2, \ldots$.

1.2. Stochstic process: We have a probability space, Ω. The information available at time t is represented by the algebra of events $\mathcal{F}_t$. We assume that for each t, $\mathcal{F}_t \subset \mathcal{F}_{t+1}$; since we are supposed to gain information going from t to $t+1$, every known event in $\mathcal{F}_t$ is also known at time $t+1$. Such an expanding family of $\sigma-$algebras is a *filtration*. A *stochastic process* is simply a family of random variables, $X_t(\omega)$. The process is *adapted* to the filtration $\mathcal{F}_t$ if with $X_t \in \mathcal{F}_t$ (X_t measureable with respect to $\mathcal{F}_t$). Sometimes it happens that the random variables X_t contain all the information in the $\mathcal{F}_t$ in the sense that $\mathcal{F}_t$ is generated by X_0, …, X_t. This the *minimal algebra* in which the X_t form an adapted stochastic process. In other cases $\mathcal{F}_t$ contains more information. Economists use these possibilities when they distinguish between the "weak efficient market hypothesis" (the $\mathcal{F}_t$ are minimal), and the "strong hypothesis" ($\mathcal{F}_t$ contains all the public information in the world, literally). In the case of minimal $\mathcal{F}_t$, it may be possible to identify the outcome, ω, with the path $X = X_0, \ldots, X_T$. The probabilities $P(\omega)$ are not important in these definitions, only the algebras of sets and random variables X_t.

1.3. Notation: The value of a stochastic process at time t may be written X_t or $X(t)$. The subscript notation reminds us that the X_t are a family of functions of the random outcome (random variable) ω. In practical contexts, particularly in discussing multidimensional processes ($X(t) \in R^n$), we prefer $X(t)$ so that $X_k(t)$ can represent the k^{th} component of $X(t)$. When the process is a martingale, we often call it F_t. This will allow us to let $X(t)$ be a Markov chain and $F_t(X)$ a martingale function of X.

1.4. Example 1, Markov chains: In this example, the $\mathcal{F}_t$ are minimal and Ω is the path space of sequences of length $T+1$ from the state space, $\mathcal{S}$. The new information revealed at time t is the state of the chain at time t. The variables X_t are may be called "coordinate functions" because X_t is coordinate t (or entry t) in the sequence X. In principle, we could express this with the notation $X_t(X)$, but that would drive people crazy. Although we distinguish between Markov chains (discrete time) and Markov processes (continuous time), the term "stochastic process" can refer to either continuous or discrete time.

1.5. Example 2, diadic sets: This is a set of definitions for discussing averages over a range of length scales. The "time" variable, t, represents the amount of averaging that has been done. The new information revealed at time t is finer scale information about a function. The probability space is the positive integers from 1 to 2^T. A *diadic block* at level t is a sequence of integers of the form

$$B_{t,k} = \left\{1 + (k-1)2^{T-t}, 2 + (k-1)2^{T-t}, \ldots, k2^{T-t}\right\} \ . \tag{1}$$

Here k ranges from 1 to 2^{t-1}, so that there is one level one block, two level two blocks, four level 3 blocks, and so on. The blocks at level t form a partition of Ω,

$$\mathcal{P}_t = \left\{B_{t,k} \text{ with } k = 1, \ldots, 2^{T-t}\right\} \ .$$

which generates the algebra $\mathcal{F}_t$. Moving from level t to level $t+1$ splits each block into right and left halves:

$$B_{t,k} = B_{t+1,2k-1} \cup B_{t+1,2k} \ . \tag{2}$$

Therefore, the partition $\mathcal{P}_{t+1}$ is a refinement of $\mathcal{P}_t$, and the $\mathcal{F}_t$ form a filtration: $\mathcal{F}_t \subset \mathcal{F}_{t+1}$. the $\mathcal{P}_{t+1}$ is a refinement of $\mathcal{P}_t$. The function $X(\omega)$ is measurable in $\mathcal{F}_t$ if it is constant on sevel t diadic blocks. $X_1 \in \mathcal{F}_1$ only if $X_1(\omega)$ is constant. $X_2 \in \mathcal{F}_2$ if it has one constant value on the first half ($1 \leq \omega \leq 2^{T-1}$), and also on the second half. These constants can be different. We will return to this example after discussing martingales.

1.6. Martingales: A real valued adapted stochastic process, F_t, is a martingale[1] if

$$E[F_{t+1} \mid \mathcal{F}_t] = F_t \ . \tag{3}$$

The tower property then implies that

$$E[F_{t+2} \mid \mathcal{F}_t] = E\Big[\left(E\left[F_{t+2} \mid \mathcal{F}_{t+1}\right]\right) \mid \mathcal{F}_t\Big] = E\left[F_{t+1} \mid \mathcal{F}_t\right] = F_t \ .$$

More generally, if $s > 0$, $E\left[F_{t+s} \mid \mathcal{F}_t\right] = F_t$.

If we take the overall expectation of both sides we see that the expectation value does not depend on t, $E[F_{t+1}] = E[F_t]$. The martingale property says more. Whatever information you might have at time t notwithstanding, still the expectation of future values is the present value. There is a gambling interpretation: F_t is the amount of money you have at time t. No matter what

[1] For finite Ω this is the whole story. For countable Ω we also assume that the sums defining $E[X_t]$ converge absolutely, $E[|X_t|] < \infty$. This implies that the conditional expectations $E[X_t + 1 \mid \mathcal{F}_t]$ are well defined.

has happened, your expected winnings at between t and $t+1$, the "martingale difference" $Y_{t+1} = F_{t+1} - F_t$, has zero expected value. You can also think of martingale differences as a generalization of independent random variables. If the random variables Y_t were actually independent, then the sums $F_t = \sum_{k=1}^{t} Y_t$ would form a martingale (using the $\mathcal{F}_t$, generated by the $Y_1, \ldots, Y_t$). The reader should check this.

1.7. Examples: A random walk is a martingale if it has zero drift. One general way to get a martingale is to start with a random variable, $F(\omega)$, and define $F_t = E[F \mid \mathcal{F}_t]$. If we apply this to a Markov chain with the minimal filtration $\mathcal{F}_t$, and F is a final time reward $F = V(X(T))$, then $F_t = f(X(t), t)$ as in the previous lecture. If we apply this to the diadic filtration of Paragraph 5, with uniform probability $P(\omega) = 2^{-T}$ for $\omega \in \Omega$, we get the diadic martingale with $F_t(\omega)$ constant on the diadic blocks (1) and equal to the average of F over the block ω is in. In particular, if $\omega \in B_{t,k}$, then $F_t(\omega)$ is the average of the two values of F_{t+1} on the blocks $B_{t+1,2k-1}$ and $B_{t+1,2k}$, as in (2).

1.8. A lemma on conditional expectation: In working with martingales we often make use of a basic lemma about conditional expectation. Suppose $U(\omega)$ and $Y(\omega)$ are real valued random variables and that $U \in \mathcal{F}$. Then

$$E[UY \mid \mathcal{F}] = UE[Y \mid \mathcal{F}] \; . \tag{4}$$

We see this using classical conditional expectation over the sets in the partition defining $\mathcal{F}$. Let B be one of these sets. Let $y_B = E[Y \mid \omega \in B]$ be the value of $E[Y \mid \mathcal{F}]$ for $\omega \in B$. We know that $U(\omega)$ is constant in B because $U \in \mathcal{F}$. Call this value u_B. Then $E[UY \mid B] = u_B E[Y \mid B] = u_b y_b$. But this is the value of $UE[Y \mid \mathcal{F}]$ for $\omega \in B$. Since each ω is in some B, this proves (4) for all ω.

1.9. Doob's principle: This lemma lets us make new martingales from old ones. Let F_t be a martingale and $Y_t = F_t - F_{t-1}$ the *martingale differences* (called *innovations* by statisticians and *returns* in finance). We use the convention that $F_{-1} = 0$ so that $F_0 = Y_0$. The martingale condition is that $E[Y_{t+1} \mid \mathcal{F}_t] = 0$. Clearly $F_t = \sum_{t'=0}^{t} Y_{t'}$.

Suppose that at time t we are allowed to place a bet of any size[2] on the as yet unknown martingale difference, Y_{t+1}. Let $U_t \in \mathcal{F}_t$ be the size of the bet. The return from betting on Y_t will be $U_{t-1}Y_t$, and the total accumulated return up to time t is

$$G_t = U_0 Y_1 + U_1 Y_2 + \cdots + U_{t-1} Y_t \; . \tag{5}$$

Because of the lemma (4), the betting returns have $E[U_t Y_{t+1} \mid \mathcal{F}_t] = 0$, so $E[G_{t+1} \mid \mathcal{F}_t] = G_t$ and G_t also is a martingale.

The fact that G_t in (5) is a martingale sometimes is called *Doob's principle* or *Doob's theorem* after the probabilist who formulated it. A special case below for stopping times is *Doob's stopping time theorem* or the *optional stopping*

[2]We may have to require that the bet have finite expected value.

theorem. They all say that strategizing on a martingale never produces anything but a martingale. Nonanticipating strategies on martingales do not give positive expected returns.

1.10. Correlation and dependence: If F_t is a martingale, then the innovations $Y_t = F_t - F_{t-1}$ are uncorrelated. For example, since $E[Y_t] = E[Y_{t-1}] = 0$, we have $\mathrm{cov}[Y_t, Y_{t-1}] = E[Y_t Y_{t-1}] = 0$ (since $Y_{t-1} \in \mathcal{F}_{t-1}$). A martingale is called a random walk if the innovations are iid mean zero random variables. The lattice random walk of lecture 2 is a martingale if the drift is zero ($a = c$). If we modulate a random walk using (5), then the innovations $\widetilde{Y}_t = U_{t-1} Y_t$ will no longer be independent, though the correlation still is zero.

The lack of correlation gives the useful fact that the variance of a martingale is the sum of the variances of its innovations:

$$\mathrm{var}[F_t] = \sum_{t'=0}^{t} \mathrm{var}[Y_{t'}] \ . \tag{6}$$

In particular, if the Y_t are iid with variance $\sigma^2 = E[Y_t^2]$, then the modulated martingale (5) has

$$E\left[G_t^2\right] = \sigma^2 \sum_{t'=0}^{t-1} E\left[U_{t'}^2\right] \ . \tag{7}$$

1.11. Relation to the Ito integral: The Ito integral with respect to Brownian motion is a continuous time limit of sums like (5). Like these, the Ito integral turns one martingale, Brownian motion, into another. The continuous time version of (7) is the *Ito isometry formula*, which is very helpful in the theory of the Ito integral.

1.12. Weak and strong efficient market hypotheses: It is possible that the random variables F_t form a martingale with respect to their minimal filtration, $\mathcal{F}_t$, but not with respect to an enriched filtration $\mathcal{G}_t \supset \mathcal{F}_t$. The simplest example would be the algebras $\mathcal{G}_t = \mathcal{F}_{t+1}$, which already know the value of F_{t+1} at time t. Note that the F_t also are a stochastic process with respect to the $\mathcal{G}_t$. The *weak efficient market hypothesis* is that $e^{-\mu t} S_t$ is a martingale (S_t being the stock price and μ its expected growth rate) with respect to its minimal filtration. *Technical analysis* means using trading strategies that are nonanticipating with respect to the minimal filtration. Therefore, the weak efficient market hypothesis says that technical trading does not produce better returns than buy and hold. Any extra information you might get by examining the price history of S up to time t is already known by enough people that it is already reflected in the price S_t.

The *strong* efficient market hypothesis states that $e^{-\mu t} S_t$ is a martingale with respect to the filtration, $\mathcal{G}_t$, representing all the public information in the world. This includes the previous price history of S and much more (prices of related stocks, corporate reports, market trends, etc.).

1.13. Investing with Doob: Economists sometimes use Doob's principle and the efficient market hypotheses to make a point about active trading in the stock market. Suppose that F_t, the price of a stock at time t, is a martingale[3]. Suppose that at time t we all the information in $\mathcal{F}_t$, and choose an amount, U_t, to invest at time t. The fact that the resulting accumulated, G_t, has zero expected value is said to show that active investing is no better than a "buy and hold" strategy that just produces the value F_t. The well known book **A Random Walk on Wall Street** is mostly an exposition of this point of view. This argument breaks down when applied to non martingale processes, such as stock prices over longer times. Active trading strategies such as (5) may produce reduce the risk more than enough to compensage risk averse investors for small amounts of lost expected value. Merton's optimal dynamic investment analysis is a simple example of an active trading strategy that is better for some people than passive buy and hold.

1.14. Stopping times: We have Ω and the expanding family $\mathcal{F}_t$. A stopping time is a function $\tau(\omega)$ that is one of the times $1, \ldots, T$, so that the event $\{\tau \leq t\}$ is in $\mathcal{F}_t$. Stopping times might be thought of as possible strategies. Whatever your criterion for stopping is, you have enough information at time t to know whether you should stop at time t. Many stopping times are expressed as the first time something happens, such as the first time $X_t > a$. We cannot ask to stop, for example, at the last t with $X_t > a$ because we might not know at time t whether $X_{t'} > a$ for some $t' > t$.

1.15. Doob's stopping time theorem for one stopping time: Because stopping times are nonanticipating strategies, they also cannot make money from a martingale. One version of this statement is that $E[X_\tau] = E[X_1]$. The proof of this makes use of the events B_t, that $\tau = t$. The stopping time hypothesis is that $B_t \in \mathcal{F}_t$. Since τ has some value $1 \leq \tau \leq T$, the B_t form a partition of Ω. Also, if $\omega \in B_t$, $\tau(\omega) = t$, so $X_\tau = X_t$. Therefore,

$$
\begin{aligned}
E[X_1] &= E[X_T] \\
&= \sum_{t=1}^{T} E[X_T \mid B_t] P(B_t) \\
&= \sum_{t=1}^{T} E[X_\tau] P(\tau = t) \\
&= E[X_\tau] \; .
\end{aligned}
$$

In this derivation we made use of the classical statement of the martingale property, if $B \in \mathcal{F}_\sqcup$ then $E[X_T \mid B] = E[X_t \mid B]$. In our $B = B_t$, $X_t = X_\tau$.

This simple idea, using the martingale property applied to the partition B_t, is crucial for much of the theory of martingales. The idea itself was first used Kolmogorov in the context of random walk or Brownian motion. Doob

[3] This is a reasonable approximation for much short term trading

realized that Kolmogorov's was even simpler and more beautiful when applied to martingales.

1.16. Stopping time paradox: The technical hypotheses above, finite state space, bounded stopping times, may be too strong, but they cannont be completely ignored, as this famous example shows. Let X_t be a symmetric random walk starting at zero. This forms a martingale, so $E[X_\tau] = 0$ for any stopping time, τ. On the other hand, suppose we take $\tau = \min(t \mid X_t = 1)$. Then $X_\tau = 1$ always, so $E[X_\tau] = 1$. The catch is that there is no T with $\tau(\omega) \leq T$ for all ω. Even though $\tau < \infty$ "almost surely" (more to come on that expression), $E[\tau] = \infty$ (see previous lecture). Even that would be OK if the possible values of X_t were bounded. Suppose you choose T and set $\tau' = min(\tau, T)$. That is, you wait until $X_t = 1$ or $t = T$, whichever comes first, to stop. For large T, it is very likely that you stopped for $X_t = 1$. Sill, those paths that never reached 1 probably drifted just far enough in the negative direction so that their contribution to the overall expected value cancels the 1 to yield $E[X_{\tau'}] = 0$.

1.17. Another stopping times theorem: Suppose we have an increasing family of stopping times, $1 \leq \tau_1 \leq \tau_2 \cdots$. In a natural way the random variables $Y_1 = X_{\tau_1}$, $Y_2 = X_{\tau_2}$, etc. also form a martingale. This is a final elaborate way of saying that strategizing on a martingale is a no win game.

1.18. Wald's formula: