MTH 5220 - The theory of martingales in discrete time

Summary

This document is in three sections, with the first dealing with the basic theory of discrete-time martingales, the second giving a number of examples and applications, and the third, an appendix, containing a number of useful results from general probability theory and analysis.

1 Theory

A *discrete time stochastic process* is a sequence of r.v.'s $S_1, S_2, S_3, \ldots$ and its corresponding increasing collection of σ-fields $\sigma(S_1) \subseteq \sigma(S_1, S_2) \subseteq \sigma(S_1, S_2, S_3) \ldots$. The increasing collection of σ-fields is called the *filtration* of the process, and represents the information available to an observer at any time. Often times the filtration is the *natural filtration*, which is formed by the σ-fields $\mathcal{F}_n = \sigma(S_1, \ldots, S_n)$; if the S_n's are discrete random variables, then $\mathcal{F}_n$ is generated by all sets of the form $\{S_1 = r_1, S_2 = r_2, \ldots, S_n = r_n\}$ (if the S_n's are continuous then the definition of $\mathcal{F}_n$ is somewhat more technical, see the section on the Radon-Nikodym Theorem below).

Let X be a $\mathcal{F}$-measurable random variable on a space Ω, and let $\mathcal{G}$ be a σ-field on Ω with $\mathcal{G} \subseteq \mathcal{F}$, so that X is not necessarily $\mathcal{G}$-measurable. There is a $\mathcal{G}$-measurable random variable, denoted $E[X|\mathcal{G}]$ and referred to as the *conditional expectation of X with respect to* $\mathcal{G}$, such that $E[X1_A] = E[E[X|\mathcal{G}]1_A]$ for all $A \in \mathcal{G}$. There are a few rules for this:

- $E[aX + bY|\mathcal{G}] = aE[X|\mathcal{G}] + bE[Y|\mathcal{G}]$.
- If $\mathcal{G} = \{\emptyset, \Omega\}$, then $E[X|\mathcal{G}] = E[X]$.
- If X is $\mathcal{G}$-measurable, then $E[X|\mathcal{G}] = X$, and more generally $E[XY|\mathcal{G}] = XE[Y|\mathcal{G}]$
- If $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $E[E[X|\mathcal{G}_2]|\mathcal{G}_1] = E[X|\mathcal{G}_1]$.
- If $\sigma(X)$ and $\mathcal{G}$ are independent, then $E[X|\mathcal{G}] = E[X]$.

We can also condition on a set A: for example, $E[X|A] = \sum_{r=-\infty}^{\infty} rP(X = r|A)$ when X is discrete. We will also define $E[X|Y] = E[X|\sigma(Y)]$ when Y is discrete to be the random variable which is equal to $E[X|Y = s]$ on the set $\{Y = s\}$, with the analogous definition for $E[X|Y_1, Y_2, \ldots]$. Alternatively, $E[X|Y]$ can be expressed as a function of Y, so $E[X|Y] = g(y)$, and it is the unique function such that

$$E[(X - E[X|Y])^2] \leq E[(X - f(Y))^2]$$

for any function f. Note: that the conditional expectation exists for any $X, \mathcal{G}$ is immediate from the Radon-Nikodym theorem (see the Applications section).

If M_n is a stochastic process with filtration $\mathcal{F}_n$ such that $E[M_n|\mathcal{F}_{n-1}] = M_{n-1}$ (along with the technical condition $E[|M_n|] < \infty$), then we say that M_n is a martingale. Usually, though not always, $\mathcal{F}_n$ is taken to be the natural filtration $\sigma(M_1, \ldots, M_n)$. Related notions include *supermartingales*, which are stochastic processes such that that $E[S_n|F_{n-1}] \leq S_{n-1}$, and *submartingales*, for which $E[S_n|F_{n-1}] \geq S_{n-1}$. We can give two examples of martingales immediately.

- Suppose $X_1, X_2, X_3, \ldots$ is a sequence of independent random variables with $E[X_i] = 0$. Then the process $S_n = X_1 + \ldots + X_n$ is a martingale with respect to the natural filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.
- Suppose $X_1, X_2, X_3, \ldots$ is a sequence of independent positive random variables with $E[X_i] = 1$. Then the process $S_n = X_1 X_2 \ldots X_n$ is a martingale with respect to the natural filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.

Given a stochastic process $S_1, S_2, \ldots$, a *stopping time* τ is a r.v. taking values in the nonnegative integers and ∞ such that

$$\{\tau = n\} \in \sigma(S_1, S_2, \ldots, S_n) \tag{1}$$

for all n. Intuitively, this condition roughly translates to "the decision to stop must be made only with information from the past and present, not the future".

We can think of a martingale as a fair game. One of the fundamental results in the theory is that it's not possible to make or lose money while playing such a fair game, provided that one stops at a reasonable time, i.e. a stopping time which satisfies certain conditions. In particular

Theorem 1 (Optional stopping theorem)**.** *Suppose M_n is a martingale and τ is a stopping time with at least one of the following conditions*

(i) $\tau < C < \infty$ *for some constant* C*.*

(ii) $|M_n| < C < \infty$ *for some constant* C *and all* n*, and* $\tau < \infty$ *a.s.*

(iii) $E[\tau] < \infty$ *and* $|M_n - M_{n-1}| < C < \infty$ *for some constant* C*.*

Then $E[M_\tau] = E[M_0]$*.*

With this in mind, let us now interpret our martingale M_n with filtration $\mathcal{F}_n$ as a stock price. Is there a strategy C_n of the number of shares of the stock to hold at time n which will allow us to make money? A reasonable assumption is that C_n is based on the values of $M_1, M_2, \ldots, M_{n-1}$, or in other words is measurable with respect to $\mathcal{F}_{n-1}$ (if the M_n's are discrete this means C_n is constant on all events in $\sigma(M_1, M_2, \ldots, M_{n-1})$, which are sets of the form $\{M_1 = r_1, \ldots, M_{n-1} = r_{n-1}\}$). We will call any process C_n satisfying this *previsible*. The amount of money we make at time n is $C_n(M_n - M_{n-1})$, and thus our total earnings at time n is $\sum_{j=1}^{N} C_j(M_j - M_{j-1})$. It may seem that the freedom to choose the C_j's will allow us to make money, however we have the following:

Theorem 2. *Under the given assumptions,* $S_n = \sum_{j=1}^{n} C_j(M_j - M_{j-1})$ *is itself a martingale.*

Thus, the optional stopping theorem applies to S_n, and we see that $E[S_\tau] = E[S_0]$ for any reasonable stopping time. Let us now consider the following strategy applied to a martingale M_n. Let $a < b$ be given, and let $s_1 = \inf\{n \geq 0 : M_n \leq a\}$, $t_1 = \inf\{n > s_1 : M_n \geq b\}$, $s_2 = \inf\{n > t_1 : M_n \leq a\}$, $t_2 = \inf\{n > s_2 : M_n \geq b\}$, and so forth. Let C_n be 0 for $0 \leq n \leq s_1$, then 1 for $s_1 + 1 \leq n \leq t_1$, then 0 again for $t_1 + 1 \leq n \leq s_2$, then 1 again for $s_2 + 1 \leq n \leq t_2$, and so forth. We can see that C_n is previsible, and thus the process

$$Y_n = \sum_{j=1}^{n} C_j(M_j - M_{j-1})$$

is a martingale. Let $U_n(a, b)$ be the number of *upcrossings* by time n; that is, $U_n(a, b) = \max\{j : t_j \leq n\}$. We can see that

$$Y_n \geq (b - a)U_n(a, b) - (a - M_n)^+,$$

where $(x)^+ = \max(x, 0)$. This implies

Lemma 1.

$$(b - a)E[U_n(a, b)] \leq E[(a - M_n)^+].$$

Let $U_\infty(a,b) = \lim_{n\nearrow\infty} U_n(a,b)$. Then

Corollary 1. *If* $\sup_n E[|M_n|] < \infty$*, then* $P(U_\infty(a,b) < \infty) = 1$.

If a stochastic process doesn't converge then essentially it must oscillate indefinitely. These upcrossing results imply that, if we have a bound on the expectation of the modulus of a martingale, or a lower bound for the martingale, then it can't oscillate indefinitely. We therefore have the following major results.

Corollary 2. *[Martingale Convergence Theorem] If* $\sup_n E[|M_n|] < \infty$*, then* $M_\infty = \lim_{n\longrightarrow\infty} M_n$ *exists almost surely, and* $P(|M_n| < \infty) = 1$.

Corollary 3. *If* M_n *is a martingale that is bounded above or below, then* $M_\infty = \lim_{n\longrightarrow\infty} M_n$ *exists almost surely, and* $P(|M_n| < \infty) = 1$. *In particular, if* M_n *is a non-negative martingale then it converges.*

We saw examples in class when a martingale M_n converged to M_∞, but $E[M_\infty] \neq E[M_n]$ (the martingale associated with the biased random walk is a good example, see Section 2). A natural question is to give sufficient conditions for $E[M_n] \longrightarrow E[M_\infty]$. A useful way to address this question is to look at the second moments, $E[M_n^2]$, if it is known that they are finite. One reason for the simplicity of the L^2 theory is that the increments of a martingale are orthogonal in L^2, and furthermore

Lemma 2. *If* M_n *is a martingale, then*

$$E[M_{n+1}^2 - M_n^2|\mathcal{F}_n] = E[(M_{n+1} - M_n)^2|\mathcal{F}_n]$$

This implies

Lemma 3. *If* M_n *is a martingale, then*

$$E[M_n^2] = E[M_0^2] + \sum_{j=1}^{n} E[(M_j - M_{j-1})^2]$$

This lemma gives us a stronger convergence theorem.

Theorem 3. *Suppose* M_n *is a martingale with* $\sup_n E[M_n^2] < \infty$*. Then* M_n *converges to* M_∞*,* $E[(M_n - M_\infty)^2] \longrightarrow 0$*, and* $E[M_n] \longrightarrow E[M_\infty]$.

In general, if S_n is any stochastic process with respect to a filtration $\mathcal{F}_n$, then

Theorem 4 (Doob decomposition)**.** *There is a decomposition*

$$S_n = S_0 + M_n + A_n,$$

where M_n *is a martingale with respect to* $\mathcal{F}_n$*, and* A_n *is previsible with respect to* $\mathcal{F}_n$*. This decomposition is unique in the sense that if we have another decomposition* $S_n = S_0 + M'_n + A'_n$*, then* $M'_n = M_n$ *and* $A'_n = A_n$ *a.s.*

Jensen's conditional inequality (see Appendix) implies that if M_n is a martingale, then M_n^2 is automatically a submartingale (provided $E[M_n^2] < \infty$), so that the previsible process A_n in the Doob decomposition of M_n^2 is a.s. nondecreasing. This process is often denoted $A_n = \langle M\rangle_n$, and is the discrete time analog of the quadratic variation in stochastic calculus. In other words, $M_n^2 - \langle M\rangle_n$ is a martingale.

If M_n is a martingale, C_n is previsible, and $Y_n = \sum_{j=1}^n C_j(M_j - M_{j-1})$, then

$$\langle Y\rangle_n = \sum_{j=1}^{n} C_j^2 E[(M_j - M_{j-1})^2|\mathcal{F}_{j-1}].$$

Note that $E[M_n^2] = E[\langle M\rangle_n]$. Thus, M is bounded in L^2 (and converges, etc.) if $E[\langle M\rangle_\infty] < \infty$. Furthermore, (not shown in class)

Theorem 5. *If M_n is an L^2 martingale, then $M_n \longrightarrow M_\infty$ a.s. on the set $\langle M \rangle_\infty < \infty$.*

The following general result is known as *Doob's inequality*.

Theorem 6. *If M_n is a nonnegative submartingale, then*

$$P(M_n^* \geq C) \leq \frac{E[M_n 1_{\{|M_n^*| \geq C]\}}]}{C} \leq \frac{E[M_n]}{C}$$

In practice, this is often applied to $M_n = \phi(S_n)$, where S_n is a martingale and ϕ is a nonnegative convex function, since then M_n is a submartingale. For instance, we have

Corollary 4. *If M_n is a martingale, then*

$$P(M_n^* \geq C) \leq \frac{E[|M_n|^p 1_{\{M_n^* \geq C]\}}]}{C^p} \leq \frac{E[|M_n|^p]}{C^p}$$

for $p \geq 1$.

A consequence of this is *Doob's L^p inequality*, which gives a bound on the moments of M_n^*:

Corollary 5. *If M_n is a martingale, then for any $p > 1$ we have*

$$E[|M_n|^p] \leq E[(M_n^*)^p] \leq \left(\frac{p}{p-1}\right)^p E[|M_n|^p].$$

Doob's L^p inequality shows that if a martingale is bounded in L^p, then there is a random variable in L^p (M_∞^*) which bounds it. In order to bring L^p and other considerations into martingale convergence theorems, we need a new concept, which is *uniform integrability*. A collection $\mathcal{C}$ of random variables is uniformly integrable if, for each (small) $\varepsilon > 0$ there is a (big) $K > 0$ such that

$$E[|X| 1_{\{|X| > K\}}] < \varepsilon$$

for every $X \in \mathcal{C}$. A martingale M is uniformly integrable if the collection of random variables M_n is uniformly integrable.

The next result is the final word on martingale convergence.

Theorem 7. *Suppose M_n is a uniformly integrable martingale with filtration $\mathcal{F}_n$. Then M_n converges a.s. and in L^1 as $n \longrightarrow \infty$ to a random variable M_∞, and $M_n = E[M_\infty|\mathcal{F}_n]$.*

Note we also showed that $M_n = E[M_\infty|\mathcal{F}_n]$ is a uniformly integrable martingale provided that $E[|M_\infty|] < \infty$, so this result is essentially the best possible. Uniform integrability allows us to bring the p-th moment into our results, as we have the following (we already had this for the especially simple case $p = 2$):

Corollary 6. *Suppose M_n is a martingale with $\sup_n E[|M_n|^p] < \infty$, for $p > 1$. Then M_n converges to M_∞ a.s., $E[|M_n - M_\infty|] \longrightarrow 0$, and $E[M_n] \longrightarrow E[M_\infty]$.*

2 Examples and applications

2.1 Simple and biased random walk

Arguably the simplest nontrivial example of a martingale is simple random walk. Let $M_n = X_1 + X_2 + \ldots + X_n$, where $X_1, X_2, \ldots$ is a sequence of independent random variables with $P(X_i = -1) = P(X_i = 1) = \frac{1}{2}$. M_n is a martingale, so $E[M_\tau] = 0$ for any stopping time τ which satisfies the conditions of the optional stopping theorem. Also, $M_n^2 - n$ is a martingale as well (that is, $\langle M \rangle_n = n$), and applying the optional stopping theorem to that process allows us to show for instance $E[T_{ab}] = ab$, $P(X_{T_{ab}} = b) = \frac{a}{a+b}$, $P(X_{T_{ab}} = a) = \frac{b}{a+b}$, where $T_{ab} = \inf_{n \geq 0}\{M_n = -a \text{ or } M_n = b\}$ for $a, b \geq 0$.

An interesting formula is the Doob decomposition of $f(M_n)$, where f is any function:

$$f(M_n) = f(M_0) + \sum_{j=1}^{n} \frac{1}{2}(f(M_{j-1}+1) - f(M_{j-1}-1))(M_j - M_{j-1})$$
$$+ \frac{1}{2}\sum_{j=1}^{n}(f(M_{j-1}+1) - 2f(M_{j-1}) + f(M_{j-1}-1)).$$

Note the similarity with Itó's formula.

We have also the biased random walk, $S_n = X_1 + X_2 + \ldots + X_n$ with $S_0 = 0$, where $X_1, X_2, \ldots$ is a sequence of independent random variables with $P(X_i = -1) = q, P(X_i = 1) = p$, where $p + q = 1$ and $p, q \neq \frac{1}{2}$. We saw that S_n is not a martingale, but $Y_n = r^{S_n}$ is one, where $r = \frac{q}{p}$, and furthermore, $Y_n \geq 0$, so $Y_\infty = \lim_{n\longrightarrow\infty} Y_n$ exists. However, $Y_\infty = 0$ a.s., so that $E[Y_\infty] \neq E[Y_n]$. This is a good example for the need for uniform integrability or some other condition.

It is easy to see that $S_n - (p-q)n$ is a martingale, and thus S_n has the Doob decomposition

$$S_n = (S_n - (p-q)n) + (p-q)n.$$

To any stochastic process S_n we can associate its *supremum process* $S_n^* = \sup_{0\leq j\leq n} S_j$. There is a financial reason to consider this process, as it is important in the analysis of *barrier options*, which generally take one of two forms: *knock-out* and *knock-in*. Knock-out options become worthless if the stock price reaches a certain level before the payoff time, while knock-in options only take on value if the stock prices reaches the level before payoff. Both types require knowledge of the supremum process.

Let us return to the simple random walk, $S_n = X_1 + X_2 + \ldots + X_n$ with $S_0 = 0$, where $X_1, X_2, \ldots$ is a sequence of independent random variables with $P(X_i = -1) = P(X_i = 1) = \frac{1}{2}$, and $S_n^* = \sup_{0\leq j\leq n} S_j$. A natural question is, what is the distribution of S_n^*? It is clear that S_n^* is a nonnegative process, and if $C \geq 0$ we can apply a *reflection principle* to show that $P(S_n^* \geq C) = P(S_n = C) + 2P(S_n > C)$. Note: the analogous principle applies to Brownian motion, and shows that $P(\sup_{0\leq s\leq t} B_s > C) = 2P(B_t > C)$ for $C \geq 0$.

Return now to the biased random walk, $S_n = X_1 + X_2 + \ldots + X_n$ with $S_0 = 0$, where $X_1, X_2, \ldots$ is a sequence of independent random variables with $P(X_i = -1) = q, P(X_i = 1) = p$, where $p + q = 1$ and $p, q \neq \frac{1}{2}$, and $S_n^* = \sup_{0\leq j\leq n} S_j$. How can we now determine the distribution of S_n^*? For $C \geq 0$ we can adapt the reflection principle to show that

$$\begin{aligned} P(S_n^* \geq C) &= P(S_n = C) + \sum_{r=1}^{\infty}(1 + (\frac{q}{p})^r)P(S_n = C + r) \\ &= P(S_n \geq C) + \sum_{r=1}^{\infty}(\frac{q}{p})^r P(S_n = C + r) \end{aligned} \tag{2}$$

That last expression includes something that looks suspiciously like the expectation of our martingale $M_n = (\frac{q}{p})^{S_n}$, and if we let $M_n^* = \sup_{0\leq j\leq n} M_j$ and manipulate a bit we get

$$P\Big(M_n^* \geq (\frac{q}{p})^C\Big) \leq \frac{2E[(\frac{q}{p})^{S_n}]}{(\frac{q}{p})^C}.$$

This is a (weakened) form of Doob's inequality.

2.2 Polya's Urn

Let us now consider *Polya's Urn*: we have an urn with 1 white ball and 1 black one in it. At each step, we choose a ball at random from the urn and then return it along with another ball of the same color. We therefore form two increasing stochastic processes $w_0, w_1, \ldots$ and $b_0, b_1, \ldots$, and it can be shown that the proportion process $M_n = \frac{w_n}{b_n+w_n}$ is a martingale. Since it is nonnegative it must converge a.s. to a limit M_∞, but what does this limit look like? M_∞ is uniformly distributed on $(0, 1)$.

We may generalize Polya's Urn by supposing we have a white balls and b black balls to begin with. At each step, we choose a ball at random from the urn and then return it along with another ball of the same color. As before we form two increasing stochastic processes $w_0, w_1, \ldots$ and $b_0, b_1, \ldots$, and it can be shown that the proportion process $M_n = \frac{w_n}{b_n+w_n}$ is a martingale. Since it is nonnegative it must converge a.s. to a limit M_∞, but what does this limit look like? We have

$$\begin{aligned} P(w_n = a + r) &= \binom{n}{r} \frac{a(a+1)\ldots(a+r-1)b(b+1)\ldots(b+(n-r)-1)}{(a+b)(a+b+1)\ldots(a+b+n-1)} \\ &= \binom{n}{r} \frac{\beta(a+r, b+(n-r))}{\beta(a,b)}. \end{aligned} \tag{3}$$

Using this, it was shown in the homework that

$$P(M_\infty \in A) = \frac{1}{\beta(a,b)} \int_A p^{a-1}(1-p)^{b-1} dp,$$

for any set $A \subseteq [0, 1]$. This is a good example of a martingale which converges a.s. to a non-trivial limit.

2.3 The Radon-Nikodym Theorem

We used martingale techniques to prove the *Radon-Nikodym theorem*:

Theorem 8. *Suppose P and Q are probability measures on a σ-field $\mathcal{F}$, and Q is absolutely continuous with respect to P; this means that $Q(A) = 0$ whenever $P(A) = 0$. Then there is a random variable $X = \frac{dQ}{dP}$ measurable with respect to $\mathcal{F}$ such that $Q(A) = E_P[X 1_A]$ for every set $A \in \mathcal{F}$. X is unique almost surely.*

This result immediately implies the existence of conditional expectation in the general case, since if we define a measure Q on the σ-field $\mathcal{F}$ by $Q(A) = E_P[X 1_A]$, then $E[X|\mathcal{F}] = \frac{dQ}{dP}$ (there are proofs of the Radon-Nikodym Theorem which do not use martingales). It is also of fundamental importance in real analysis and financial mathematics.

2.4 Kakutani's Theorem and the likelihood ratio test

The following is a powerful result when dealing with product martingales.

Theorem 9 (Kakutani's Theorem)**.** *Suppose $X_1, X_2, \ldots$ are independent non-negative random variables with $E[X_j] = 1$. Let $M_0 = 1$ and $M_n = X_1 X_2 \ldots X_n$. Then M_n is a non-negative martingale, and so converges to M_∞ a.s. Then M is uniformly integrable if, and only if, $\prod_{n=1}^\infty a_n > 0$, where $a_n = E[\sqrt{X_n}] \le 1$. This is equivalent to $\sum_{n=1}^\infty (1 - a_n) < \infty$. If these do not occur, then $M_\infty = 0$ a.s.*

Note that this shows immediately that $r^{S_n} \longrightarrow 0$ a.s., where S_n is biased random walk and r^{S_n} is its associated product martingale. Another good application of Kakutani's Theorem comes from statistics, the *likelihood ratio test*. Suppose we have a population, and we want to test the hypothesis that some measurement from the population admits the density f vs. the hypothesis that it admits the density g, where f and g are two positive functions on $\mathbb{R}$ with $\int_{\mathbb{R}} f(x)dx = \int_{\mathbb{R}} g(x)dx = 1$. Independent samples will be represented by an i.i.d. sequence of random variables $X_1, X_2, \ldots$, with common density either $f(x)$ or $g(x)$. If g is the true density, then the stochastic process

$$M_n = \prod_{j=1}^{n} \frac{f(X_j)}{g(X_j)}$$

is a martingale. Kakutani's Theorem allows us to conclude that $M_n \longrightarrow 0$ a.s., and in fact it can be shown that in most cases this occurs quite rapidly. On the other hand, if f is the true density, then M_n is not a martingale, but $\frac{1}{M_n}$ is, and the same argument allows us to conclude that $\frac{1}{M_n} \longrightarrow 0$ a.s., which means that $M_n \longrightarrow \infty$ a.s.

2.5 Pricing claims in financial mathematics

We consider a model in which there are two ways in which a person can invest their money. One is in a stock, S_n, which is a stochastic process which possesses risk, or randomness, and the second is in a bond or savings account, β_n, which is risk free, i.e. deterministic. We will generally take $\beta_n = (1+r)^n$, where r is the interest rate corresponding to unit time. We will create a *portfolio*, which is a trading strategy of buying a_n units of stocks and b_n units of bonds, and a_n and b_n must both be predictable (a.k.a. previsible). The value of the portfolio at any time t is

(4)
$$V_n = a_n S_n + b_n \beta_n$$

We require this process to be nonnegative, so $V_n \geq 0$ a.s. for every n, although a_n and b_n are each allowed to be negative (corresponding to borrowing money and short-selling stocks). We also require that the process be *self-financing*, that is, any change in the amount of money invested can only be funded by money earned or lost by the portfolio. We express this mathematically as

$$a_n S_n + b_n \beta_n = a_{n+1} S_n + b_{n+1} \beta_n$$

Any sort of predictable strategy a_n for holding shares of S_n can be fit into a self-financing one:

Lemma 4. *If a_n is predictable and V_0 is any $\mathcal{F}_0$-measurable r.v., then there is a unique predictable process b_n such that $V_n = a_n S_n + b_n \beta_n$ is a self-financing process which agrees with V_0 at $n = 0$.*

In practice, stock prices and portfolios of this type are not likely to be martingales, however an assumption which arises in modelling is that the quotient $\frac{S_n}{\beta_n}$ is one. We will write $\tilde{S}_n = \frac{S_n}{\beta_n}$, and $\tilde{V}_n = \frac{V_n}{\beta_n} = a_n \tilde{S}_n + b_n$. We then have

Lemma 5. *If $V_n = a_n S_n + b_n \beta_n$ is a self-financing strategy and $\tilde{S}_n = \frac{S_n}{\beta_n}$ is a martingale, then $\tilde{V}_n = \frac{V_n}{\beta_n}$ is a martingale as well.*

Another way of looking at the previous result is the following.

Lemma 6. *If $V_n = a_n S_n + b_n \beta_n$ is a self-financing strategy, then*

(i)

$$V_n = V_0 + \sum_{j=1}^{n} a_j (S_j - S_{j-1}) + \sum_{j=1}^{n} b_j (\beta_j - \beta_{j-1})$$

.

(ii)

$$\tilde{V}_n = \tilde{V}_0 + \sum_{j=1}^{n} a_j (\tilde{S}_j - \tilde{S}_{j-1})$$

.

Self-financing strategies V_n which satisfy $V_n \geq 0$ a.s. for every n are called *admissible*, and these are the strategies that we be will concerned with. A *claim at time* T is simply a non-negative random variable which is measurable with respect to $\mathcal{F}_T$, and which represent some sort of payoff at time T. We will mainly be interested in *attainable claims*. An attainable claim is a claim X for which there is an admissible portfolio such that $V_T = X$. One of the biggest problems in financial mathematics is pricing claims; that is, how much should we be willing to pay at time 0 for a claim X at time T?

Claims are priced under the principle of *no-arbitrage*. Arbitrage is essentially risk-free profit. That is, an arbitrage is an admissible trading strategy such that $V_0 = 0$ a.s. but $E[V_T] > 0$ (remember $V_n \geq 0$ for all n). We call the set of all strategies for a given S_n a *market*, and a market is *viable* if it contains no arbitrage strategies.

Theorem 10 (First Fundamental Theorem of Asset Pricing)**.** *A market is viable if, and only if, there exists a probability measure Q equivalent to P under which $\tilde{S}_n = \frac{S_n}{\beta_n}$ is a martingale.*

We call Q the *equivalent martingale measure* (EMM). Let us suppose that $V_n = a_n S_n + b_n \beta_n$ is an admissible strategy and X, which is a claim at time T, is given by V_T. If we assume no arbitrage, then there is a measure Q equivalent to P such that $\tilde{S}_n = \frac{S_n}{\beta_n}$ is a martingale with respect to Q. Since we can generate claim X by following the strategy, a fair price for the claim at time 0 would be $E_Q[\frac{X}{\beta_N}|\mathcal{F}_0]$, and for time n would be $E_Q[\frac{X}{\beta_{N-n}}|\mathcal{F}_n]$. Thus, claims which can be realized by admissible strategies, which we have called *attainable claims*, are of special importance. Markets in which every claim is attainable are called *complete*.

Theorem 11 (Second Fundamental Theorem of Asset Pricing)**.** *A viable market is complete if, and only if, the EMM Q is unique.*

The following is the *binomial options pricing model*, and is also referred to as the *Cox, Ross, and Rubinstein model*. Suppose $\beta_n = (1+r)^n$ and $S_0 = 1, S_n = S_{n-1}X_n$, where $X_1, X_2, \ldots$ is an i.i.d. sequence of r.v.'s, each taking values in $\{d, u\}$ with positive probability. We can find an EMM Q for $\frac{S_n}{\beta_n}$ if, and only if, $d < 1 + r < u$. The required EMM is given by $q_d = Q(X_n = d) = \frac{u-(1+r)}{u-d}$ and $q_u = Q(X_n = u) = \frac{(1+r)-d}{u-d}$. Thus, any claim X realized at time N can be priced by the formula

$$E_Q[\frac{X}{\beta_n}|\mathcal{F}_0] = (1+r)^{-N} E_Q[X].$$

For example, if X is a European call option, then $X = (S_N - K)_+$, and the value of X at time 0 is

$$(1+r)^{-N} E_Q[(S_N - K)_+] = \sum_{j=0}^{N} \frac{N!}{j!(N-j)!} q_d^j q_u^{N-j} (d^j u^{N-j} - K)_+.$$

An *American option* is like a European one, except that the buyer has the right to exercise the option at any point up to and including time N. In order to fit this idea into our model we require the buyer to choose a stopping time τ, and the value of the option is calculated based on S_τ. For example, if it a call option with strike price K then the buyer would receive $(S_\tau - K)_+$. What is a fair price for the option?

In order to handle the American options, we need to be able to analyze claims which depend on n. So let Y_n be such a time-dependent claim; that is, Y_n is a non-negative stochastic process for $0 \leq n \leq N$ adapted to the filtration $\mathcal{F}_n$ which represents the amount of money received if the option is exercised at time n. Let V_n be the value process at the same time of the corresponding European claim; that is, V_n is the value at time n (obtained under the no-arbitrage assumption) of the claim Y_N. Let V_n^A be the value process of Y_n. It is clear that $V_n^A \geq v_n$, but it is surprising that in some cases (for example, the call option) we have $V_n^A = v_n$!

Given a time-dependent claim Y_n, define a stochastic process Z_n by $Z_N = Y_N, Z_n = \max\{Y_n, \frac{1}{(1+r)} E[Z_{n+1}|\mathcal{F}_n]\}$. This is the *Snell envelope*, and helps us to price American options.

Theorem 12. *(i)* $Z_n = \max_\tau \{(1+r)^n E_Q[\frac{Y_\tau}{(1+r)^\tau}|\mathcal{F}_n]$*, where the maximum is taken over all stopping times τ with* $0 \leq \tau \leq N$.

(ii) *The maximum in* (i) *is realized by the stopping time* $\tau = \min\{n' \geq n : Z_{n'} = Y_{n'}\}$.

(iii) $\tilde{Z}_n = \frac{Z_n}{(1+r)^n}$ *is a* Q*-supermartingale, and is the smallest* Q*-supermartingale which dominates* $\tilde{Y}_n = \frac{Y_n}{(1+r)^n}$.

(iv) *The correct no-arbitrage value for an American option is* $V_n^A = Z_n$*, and the optimal exercise strategy is given by the stopping time* τ *defined above (with* $n = 0$*).*

The reason the American call has the same value as a European one is the following theorem.

Theorem 13. *If* Y_n *is a* Q*-submartingale, then the optimal strategy is* $\tau = N$*, and* $V_n^A = V_n$.

Corollary 7. *The optimal strategy for an American call option is* $\tau = N$.

2.6 The Kalman filter

Suppose we are given two processes (X_n, Y_n), $n = 0, \pm 1, \pm 2, \ldots$, where Y_n is the observations of a signal X_n contaminated by noise, e.g. $Y_n = X_n + Z_n$, where X_n is a signal and Z_n is noise. A good example of this would be in telecommunications, where transmissions will generally arrive with static. We want to find a filter which will give us a good estimate of the signal, $\hat{X}_n$. We will look at a famous model for filtering in this manner, the Kalman filter.

Before tackling the problem, we need to understand Bayes's Theorem. Recall that $P(A|B) = \frac{P(A\cap B)}{P(B)}$.

Theorem 14 (Bayes' Theorem)**.** *For two events* A, B*, with* $P(B) \neq 0$*, we have*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

We will interpret Bayes' Theorem for random variables in regards to their pdf's. Let us suppose that X, Y are two random variables which have a joint density $f_{X,Y}(x,y)$ which is strictly positive on $\mathbb{R}^2$. Then

$$P((X,Y) \in B \subseteq \mathbb{R}^2) = \int\int_B f_{X,Y}(x,y)dxdy.$$

Also

$$P(X \in B \subseteq \mathbb{R}^1) = \int_B \int_{\mathbb{R}} f_{X,Y}(x,y)dydx.$$

We see that the density for f is given by

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)dy.$$

We define

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Note that

$$P(Y \in B \subseteq \mathbb{R}^1 | X) = \int_B f_{Y|X}(y|X)dy.$$

In relation to pdf's, Bayes' Theorem takes the following form:

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

Let us suppose that X is fixed and has distribution $N(0, \sigma^2)$, and $Y_n = X + c_n Z_n$, are our "noisy" observations of X, where Z_n are i.i.d. $N(0,1)$ random variables, and $\{c_n\}$ is a sequence of constants. We wish to estimate X, which is unknown, by the values of Y_n, which are known. Let $\mathcal{F}_n = \sigma(Y_1, Y_2, \ldots, Y_n)$. We know that $M_n = E[X|\mathcal{F}_n]$, which is our best estimate for X based on information available at time n, is a u.i. martingale, and thus converges in L^1, but what does it converge to? The answer is given by the following.

Theorem 15. *Suppose X is a r.v. with $E[|X|] < \infty$, and $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ is an increasing sequence of σ-fields. Let $M_n = E[X|\mathcal{F}_n]$. Then M_n is a u.i. martingale, and $M_n \longrightarrow M_\infty = E[X|\mathcal{F}_\infty]$ a.s. and in L^1.*

The question then is, does $X = E[X|\mathcal{F}_\infty]$? And, as a practical matter, how do we calculate M_n in terms of the measurements $Y_1, Y_2, \ldots$?

In filtering theory, as in this case, very often one is dealing with normal random variables. When we say $X \sim N(\mu, \sigma^2)$, we mean that X admits a pdf of the form $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. We let $C_X(Y)$ denote the distribution of X conditioned on Y. The following is Bayes' formula for bivariate normal distributions:

Theorem 16. *Suppose $X \sim N(\mu, U)$ and $C_X(Y) = N(X, W)$. Then $C_Y(X) = N(\hat{X}, V)$, where*

$$\frac{1}{V} = \frac{1}{U} + \frac{1}{W}, \qquad \frac{\hat{X}}{V} = \frac{\mu}{U} + \frac{Y}{W}$$

The last theorem says that sampling Y gives the best estimate of $\hat{X} = V(\frac{\mu}{U} + \frac{Y}{W})$ for X, where $\frac{1}{V} = \frac{1}{U} + \frac{1}{W}$. We also have

Corollary 8.

$$E[(X - \hat{X})^2] = V$$

Let us recursively define $V_0 = \sigma^2, \frac{1}{V_n} = \frac{1}{V_{n-1}} + \frac{1}{c_n^2}$, so in fact $V_n = (\sigma^{-2} + \sum_{j=1}^n c_j^{-2})^{-1}$. Let also $\hat{X}_0 = 0$, and then $\frac{\hat{X}_n}{V_n} = \frac{\hat{X}_{n-1}}{V_{n-1}} + \frac{Y_n}{c_n^2}$. Then it was shown in class that $M_n = E[X|\mathcal{F}_n] = \hat{X}_n$, and $E[(X - \hat{X}_n)^2] = V_n$. Our estimate therefore converges to X in L^2 if, and only if, $\sum_{n=1}^\infty c_n^{-2} = \infty$. This would include the case when the c_n's are constant, and even allows c_n to grow as long as they don't grow too fast.

In practice it is more common that we are trying to estimate a sequence X_n that is changing over time, but which is evolving according to some rule. For instance suppose that

$$X_n - X_{n-1} = AX_{n-1} + HZ_n + g,$$

where the Z_n's are i.i.d. $N(0,1)$ random variables (X_n is known as an *autoregressive process*. Suppose again that we can't observe X_n directly, but can only observe Y_n, where

$$Y_n - Y_{n-1} = CX_n + KZ'_n,$$

where again the Z'_n's are i.i.d. $N(0,1)$ random variables. In this case, extending our previous techniques a bit, we arrive at the following Kalman filter equations:

$$\frac{1}{V_n} = \frac{1}{\alpha^2 V_{n-1} + H^2} + \frac{C^2}{K^2},$$

$$\frac{\hat{X}_n}{V_n} = \frac{\alpha \hat{X}_{n-1} + g}{\alpha^2 V_{n-1} + H^2} + \frac{C(Y_n - Y_{n-1})}{K^2}$$

where $\alpha = 1 + A$. It can be shown that V_n approaches the unique positive solution of $\frac{1}{x} = \frac{1}{\alpha^2 x + H^2} + \frac{C^2}{K^2}$.

2.7 The Galton-Watson process

Suppose $Z_{i,j}$ are a collection of i.i.d., nonnegative integer valued random variables which have the same distribution as a random variable Z. Form a stochastic process by $X_0 = 1$, and then $X_{n+1} = \sum_{j=1}^{X_n} Z_{n,j}$. This the *Galton-Watson process*, and is used to model family names, biological processes, and nuclear fission, among other things. Having a random variable as a limit in the sum causes some difficulties in calculations, but we were able to show

Theorem 17. *Suppose that $E[Z] = \mu$ and $Var(Z) = \sigma^2$. Then $E[X_n] = \mu^n$, and $Var(X_n) = \frac{\sigma^2\mu^{n-1}(\mu^n-1)}{\mu-1}$ unless $\mu = 1$, in which case $Var(X_n) = n\sigma^2$.*

These can be proved using the probability function, as discussed later in this subsection and in Section 3, but the moment formula is immediate from the fact that

$$E[X_{n+1}|\mathcal{F}_n] = \mu X_n,$$

where again $\mu = E[Z]$ and $\mathcal{F}_n$ is the natural filtration generated by X_n. Thus, $M_n = \frac{X_n}{\mu^n}$ is a martingale. It is also nonnegative, so $M_n \longrightarrow M_\infty$ a.s. as $n \longrightarrow \infty$. But does $E[M_\infty] = E[M_0] = 1$? Or is $E[M_\infty] = 0$ a.s.?

In order to address the previous question, let us first answer the following: what is the probability that the process eventually goes extinct? That is, what is $\lim_{n\longrightarrow\infty} P(X_n = 0)$? In order to calculate this, we can note that the generating function of X_n is simply the generating function of Z composed with itself n times (see Section 3). We will also make use of the following facts about the generating functions $f(s)$ of Z and $f_n(s)$ of X_n (in fact, all four properties apply to all generating functions):

- $f_n(0) = P(X_n = 0)$ and $f(0) = P(Z = 0)$.
- $f(s)$ is convex on $[0, 1]$ (so $f'(s)$ is increasing).
- $E[Z] = f'(1)$.
- $f(1) = 1$.

To avoid trivial cases we will assume $P(Z = 0) > 0$ and $P(Z \geq 2) > 0$. With these assumptions, the extinction probability is determined by the following theorem.

Theorem 18. *The extinction probability is the smallest fixed point of $f(s)$ (i.e. the smallest solution to the equation $f(s) = s$) in $[0, 1]$. $f(s)$ possesses exactly one fixed point in $[0, 1)$ if $E[Z] > 1$, and none if $E[Z] \leq 1$. Thus, the population has positive probability of survival if $E[Z] > 1$, but goes extinct a.s. if $E[Z] \leq 1$.*

2.8 Insurance modelling

Let us suppose that an insurance policy is sold which costs the buyer c dollars per each unit of time. Let us suppose further that the customer makes a claim in each unit of time which is represented by a nonnegative random variable X_n, and the random variables X_n are i.i.d. We define the *surplus process* U_n to be

$$U_n = x + cn - \sum_{j=1}^{n} X_j.$$

x represents the initial surplus that the insurer has, and U_n at any time represents the surplus at that time. The biggest question is to determine the probability that $T < \infty$, where $T = \inf\{n > 0 : U_n < 0\}$ is the *ruin time*.
So we would like to say something about $P_x(T < \infty) = P(T < \infty|U_0 = x)$. We begin by noting $E[U_n] = x + cn - nE[X]$, where X has the same distribution as the X_j's. Note that if $E[X] > c$ then $E[U_n] \longrightarrow -\infty$, and it can be shown to follow from this under most conditions on X that $P_x(T < \infty) = 1$ for any x. We therefore assume $E[X] < c$. We also will assume $P(X > c) > 0$, since otherwise $P(T < \infty) = 0$.

Calculating $P_x(T < \infty)$ can be difficult, however there is a nice way to get a good upper bound on this quantity, under the assumption that the moment generating function $M_X(r) = E[e^{rX}]$ of X exists. It can be shown that in this case there is a unique $R > 0$ such that $E[e^{-R(c-X)}] = 1$. This R is called the *adjustment coefficient* of the model.

Lemma 7. *e^{-RU_n} is a martingale with respect to the natural filtration.*

Theorem 19 (Lundberg's Inequality)**.**

$$P_x(T < \infty) < e^{-Rx}$$

We see that the adjustment coefficient is some sort of measure of the risk of an insurance policy: larger R means a lower probability of eventual ruin, while smaller R means that the policy is more risky (for the insurer).

3 Appendix

3.1 Modes of convergence and integral/expectation convergence theorems

In this course, we discussed four major types of convergence of random variables:

(i) *Almost sure convergence*, abbreviated as a.s. This is when $P(X_n \longrightarrow X) = 1$, that is, $X_n(\omega) \longrightarrow X(\omega)$ for all ω in a set of measure 1.

(ii) *Convergence in probability*. This is when $P(|X_n - X| > \varepsilon) \longrightarrow 0$ for any $\varepsilon > 0$.

(iii) *L^p convergence*. This is when $E[|X_n - X|^p] \longrightarrow 0$ for some fixed $p > 0$.

(iv) *Convergence in distribution.* This is when $F_n(x) \longrightarrow F(x)$ for all x at which F is continuous, where $F(x) = P(X \leq x)$ is the distribution function for X (and similarly for F_n).

a.s. and L^p convergence imply convergence in probability, though not conversely, although if $X_n \longrightarrow X$ in probability then there exists a subsequence X_{n_k} which converges to X a.s. Convergence in distribution is often proved by the following:

Theorem 20 (Lévy's Continuity Theorem)**.** *$X_n \longrightarrow X$ in distribution if, and only if, $\phi_{X_n}(t) \longrightarrow \phi_X(t)$ for all t, where ϕ denotes the characteristic functions (see below).*

The condition $E[X_n] \longrightarrow E[X]$ is often required, and is a consequence of L^p convergence for $p > 1$, but not of the other types of convergence. This makes the following results important.

Theorem 21 (Monotone Convergence Theorem)**.** *If $0 \leq X_n, X$ and $X_n \nearrow X$ a.s., then $E[X_n] \nearrow E[X]$.*

Theorem 22 (Fatou's Lemma)**.** *If $0 \leq X_n, X$ and $X_n \longrightarrow X$ a.s., then $E[X] \leq \liminf_{n \longrightarrow \infty} E[X_n]$.*

Theorem 23 (Dominated Convergence Theorem)**.** *If $X_n \longrightarrow X$ a.s. and there is $Y \geq 0$ with $E[Y] \leq \infty$ and $|X_n|, |X| \leq Y$, then $E[X_n] \nearrow E[X]$.*

The notion of uniform integrability discussed in Section 1 can be used to extend the dominated convergence theorem, as follows.

Theorem 24. *Suppose X is a r.v., and X_n is a sequence of r.v.'s. Then $X_n \longrightarrow X$ in L^1 (that is, $E[|X_n - X|] \longrightarrow 0$) if, and only if,*

(i) $X_n \longrightarrow X$ in probability.

(ii) the set of r.v.'s X_n is uniformly integrable.

3.2 Generating/characteristic functions

A few important tools in probability theory are the following.

- If X is a r.v. taking values in only in $0, 1, 2, \ldots$, then we define the *probability generating function* as

$$G(z) = G_X(z) = E[z^X] = \sum_{j=0}^{\infty} z^j P(X = j).$$

- For more general r.v., we define the *moment generating function* as

$$M(t) = M_X(t) = E[e^{tX}],$$

  whenever it exists (it does not always exist).

- For any r.v., we define the *characteristic function* as

$$\phi(t) = \phi_X(t) = E[e^{itX}].$$

  This always exists.

These objects satisfy the following useful properties.

- All three functions uniquely characterize distributions.
- All three turn sums of independent random variables into products. For example, if $X_1, X_2, \ldots, X_n$ are independent, then

$$\begin{aligned} G_{X_1+\ldots+X_n}(z) &= E[z^{X_1+\ldots+X_n}] = E[z^{X_1} \ldots z^{X_n}] \\ &= E[z^{X_1}] \ldots E[z^{X_n}] = G_{X_1}(z) \ldots G_{X_n}(z). \end{aligned} \tag{5}$$

- $M_X^{(n)}(0) = E[X^n]$, $G_X^{(n)}(0) = n!P(X = n)$, $G'_X(1) = E[X]$, and so forth.

These tools are important in many contexts, but for us one of the most valuable instances of their use was the analysis of the Galton-Watson process, because of the following facts. Suppose $Y = \sum_{j=0}^{T} Z_j$, where the Z_j's are i.i.d. and T is a r.v. taking values in the nonnegative integers which is independent of the Z's. Then, if $G_T(s) = E[s^T]$ is the generating function of T, we have

- if Z takes values in the nonnegative integers and has generating function $G_Z(s) = E[s^Z]$, then Y has generating function $G_Y(s) = G_T(G_Z(s))$.
- otherwise, suppose Z has a moment generating function $M_Z(s) = E[e^{sZ}]$. Then $M_Y(s) = G_T(M_Z(s))$.
- otherwise, if $M_Z(s)$ doesn't exist, suppose Z has characteristic function $\phi_Z(s) = E[e^{isZ}]$. Then $\phi_Y(s) = G_T(\phi_Z(s))$.

This allowed us to show easily for instance

Theorem 25 (Wald's identity)**.** *Suppose $Y = \sum_{j=0}^{T} Z_j$, where the Z_j's are i.i.d. with $E[|Z|] < \infty$ and T is a r.v. taking values in the nonnegative integers which is independent of the Z's. Then $E[Y] = E[T]E[Z]$.*

Returning to the Galton-Watson process in Section 2, we see that if we let $f_n(s) = G_{X_n}(s)$, then $f_n(s)$ is just $f(s) = G_Z(s)$ composed with itself n times. This was the key to the calculation of the extinction probability.

3.3 Inequalities

The L^p norm is

$$||X||_p = E[X^p]^{1/p}.$$

Of special importance are the L^2 norm, $||X||_2 = E[X^2]^{1/2}$, and the L^1 norm, which is simply $||X||_1 = E[|X|]$. They are related by the Cauchy-Schwarz inequality:

Theorem 26 (Cauchy-Schwarz inequality)**.**

$$||XY||_1 \leq ||X||_2 ||Y||_2.$$

In particular, taking $Y = 1$ gives

$$||X||_1 \leq ||X||_2.$$

The Cauchy-Schwarz inequality can be proved directly by a famous argument, but it is also a special case the following result, known as *Hölder's Inequality*, which is fundamental to the study of L^p spaces.

Theorem 27 (Hölder's inequality)**.** *Suppose $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, and let X and Y be any two random variables. Then*

$$E[|XY|] \leq E[|X|^p]^{\frac{1}{p}} E[|Y|^q]^{\frac{1}{q}}.$$

In other words, $||XY||_1 \leq ||X||_p ||Y||_q$

The Hölder and Cauchy-Schwarz inequalities, suitable formulated, apply to more arbitrary integrals and sums. Jensen's inequality, on the other hand, is more probabilistic in nature, since it requires a probability measure (rather than an arbitrary one):

Theorem 28 (Jensen's inequality)**.** *For any convex function $c(x)$ and any random variable X, we have $E[c(X)] \geq c(E[X])$*

There is a conditional form of Jensen's inequality, under the assumption that c is convex:

$$E[c(X)|\mathcal{F}] \geq c(E[X|\mathcal{F}]).$$