Summary of Stochastic Processes

Kui Tang

May 2013

Based on Lawler's *Introduction to Stochastic Processes*, second edition, and course slides from Prof. Hongzhong Zhang.

Contents

1 Difference/tial Equations

To solve

$$\begin{cases} y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \ldots + a_1 y'(t) + a_0 y(t) = 0 \\ y(0) = b_0, \ldots, y^{(n-1)}(0) = b_{n-1} \end{cases}$$

first find a basis of n linearly independent solutions, and then solve the coefficients by matching the initial conditions. Guess that the basis functions take the form $y(t) = e^{\lambda t}$. After all, what else *can* linear ODEs do? Then $y^{(k)}(t) = \lambda^k e^{\lambda t}$. Plugging in to the problem, we find that $y(t)$ is a solution if and only if

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = 0$$

Moreover, we can show that if λ is a multiple root, adding $t^j e^{\lambda t}$ solves the problem. In this way, we always get n linearly independent solutions.

A discrete differential equation is no different: we are given a recurrence

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \ldots + c_d a_{n-d}$$

and initial values $a_0, \ldots, a_{d-1}$. We form the characteristic polynomial $p(x) = x^d - c_1 t^{d-1} - \ldots - c_d$[1], and the general solution is

$$a_n = k_1 x_1^n + k_2 x_2^n + \ldots + k_d r_d^n$$

Note that the only difference is that the powers are taken of the characteristic roots themselves, rather than of e scaled by the root as in the continuous case.

One way to solve an inhomogeneous equation when the non-homogeneous term is constant is to use *symbolic differentiation:* Let's solve $a_{n+1} = 1 + \frac{a_n}{p}$. Write the equation $a_{n+2} = 1 + \frac{a_{n+1}}{p}$ and subtract to get the homogeneous $pa_{n+2} - pa_{n+1} = a_{n+1} - a_n$. Note that you turn a first-order equation into a second order one, but you only had one original IC. Manually iterate the second term to get the second IC.

On particular equation arises in random walks: you can either go left or right, giving a recurrence

$$f(n) = (1-p)f(n-1) + pf(n+1)$$

Applying the method above leads us to solve $px^2 - x - p + 1 = 0$, which has roots at $x = 1$ and $x = (1-p)/p$ when $p \neq 1/2$. This gives the solution

$$f(n) = c_1 + c_2 \left(\frac{1-p}{p}\right)^n$$

whereas for $p = 1/2$ we get a repeated root $x = 1$, giving

$$f(n) = c_1 + c_2 n$$

2 Discrete Markov Chains

We say $\mathbf{P}$ is a *stochastic matrix* if each of its rows sum to one (so that each row is a discrete probability measure). Let π be a row probability vector. If $\pi\mathbf{P} = \pi$, then we say π is an *invariant (stationary, equilibrium, steady-state) measure.* If $\lim_{n\to\infty} \mathbf{vP} = \pi$ for *any* row probability vector $\mathbf{v}$, then we say π is a *limiting distribution.*

We have

$$p_{n+m}(z,x) = P(X_{n+m} = z | X_0 = x) = \sum_{y \in S} P(X_{n+m} = z, X_m = y | X_0 = x) = \sum_{y \in S} p_m(x,z) p_n(z,y)$$

[1] The minus sign is *strictly* a consequence of writing a_n and the a_{n-1} on opposite sides of =, compared to the differential equation.

which is the Chapman-Kolmogorov equation. This is just ordinary matrix multiplication in the finite case, and defines matrix multiplication in the infinite case.

All limiting distributions are invariant because $\pi = \lim_n \mathbf{v}\mathbf{P}^{n+1} = (\lim_n \mathbf{v}\mathbf{P}^n)\,\mathbf{P} = \pi\mathbf{P}$, but not all invariant distributions are limiting distributions. Notably, a periodic Markov chain, we must take the limit of the average of powers of $\mathbf{P}$ within a period to obtain the invariant distribution. In fact, it takes a bit of work to justify the existence and uniqueness of the invariant distribution. How might we begin? To find π, we solve the eigenvector problem $\pi = \pi\mathbf{P}$, which if you write out, reduces to a homogeneous linear system. Sufficient conditions for the *existence and uniqueness* of invariant distributions are that

1. The eigenvalue 1 is simple and all other eigenvectors have norm less than 1; and
2. The eigenvector for 1 can be chosen to have nonnegative entries.

The *Perron-Frobenius* theorem implies that stochastic matrices with all strictly positive entries satisfy the sufficient conditions. In general, we can require just that $\mathbf{P}^n \succ 0$ for some $n > 0$.

2.1 Recurrence and Communication

Two states communicate if one can travel between both with positive probability. Communication is an equivalence relation, so we can partition the states into *communication classes.*

Communication classes are either *recurrent,* in which a chain starting in that class will never leave, or *transient*, in which a chain starting in that class will leave and never return with probability one. There are no other alternatives. Moreover, a chain must have at least one recurrent state. A chain is *irreducible* it contains just one (recurrent) communication class. Otherwise, the chain is reducible and we can decompose its transition matrix into

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 & & \mathbf{0} & \\ & \ddots & & \mathbf{0} \\ \mathbf{0} & & \mathbf{P}_r & \\ & \mathbf{S} & & \mathbf{Q} \end{bmatrix}$$

where each $\mathbf{P}_i$ is a recurrent subchain. Powers of this matrix are simply powers of the diagonal plus junk in $\mathbf{S}$.

2.2 Periodicity

Let $\mathbf{P}$ be irreducible. Define the set of possible return times to state i as

$$J_i = \{n \geq 0 : p_n(i,i) > 0\}$$

The the *period of state* i is the GCD of J_i. We can show that all states must have the same period because the J_i are closed under addition. (Why?)

The main idea is that irreducible periodic chains still have invariant distributions, but the invariant distribution is NOT its limiting distribution. Rather, a d-period chain has an invariant distribution equal to the limit of the period-average:

$$\pi = \lim_{n\to\infty} \frac{1}{d}\left[\pi_0\mathbf{P}^{n+1} + \ldots + \pi_0\mathbf{P}^{n+d}\right] = \pi$$

and the invariant distribution still represents the average time spent in each state. Further, $\mathbf{P}$ will have d simple eigenvalues around the circle of unity $z^d = 1$. Of course, to have the invariant probability, 1 is also a simple eigenvalue.

2.3 Return Times

Consider an irreducible Markov chain, so it has an invariant distribution π. Then the fraction of total time spent in state j is $\pi(j)$ by a simple law of large numbers argument. Now define the *return time to state i as* $T = \min\{n \geq 1 : X_n = i\}$. Consider a sequence k returns. Then the law of large numbers says

$$\frac{1}{k}(T_1 + \ldots + T_k) \to \mathbf{E}[T]$$

so that there are k visits to the chain in about $k\mathbf{E}[T]$ steps. But in n steps, we get $n\pi(i)$ visits. Putting together, we get $n = k\mathbf{E}[T]$ and $n\pi(i) = k$, so $\mathbf{E}[T] = 1/\pi(i)$.

2.4 Transient States and Absorbing Chains

The $\mathbf{Q}$ block of transient states has an interesting algebraic characterization as an infinite geometric series. By definition of transience, $\mathbf{Q}^n \to 0$, so we have $\sum_{n=1}^{\infty} \mathbf{Q}^n = (\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{M}$ in analogy with the scalar case. (The formal proof $(\mathbf{I} - \mathbf{Q}) \sum_{n=1}^{\infty} \mathbf{Q}^n = \mathbf{I}$ is "just" algebra; and analysis just tells us that the limit doesn't diverge, so the subtraction makes sense.) The matrix $\mathbf{I} - \mathbf{Q}$ is full rank because the $\mathbf{I}$ contributes full rank.

The (i, j) entry of $\mathbf{M}$ is the expected number of times a chain starting at i visits j because

$$\mathbf{E}\left[\sum_{n=0}^{\infty} I\left\{X_n = j\right\} | X_0 = i\right] = \sum_{n=0}^{\infty} \mathbf{P}\left\{X_n = j | X_0 = i\right\} = \sum_{n=0}^{\infty} p_n(i, j) = \mathbf{M}_{ij}$$

Note that just as (unconditional) expectations of indicators are (unconditional) probabilities, conditional expectations of indicators are conditional expectations. The expected time until the chain starting at i enters a recurrent state is just the sum of the expected times that the chain needs to visit each transient state.

An *absorbing chain* contains states that only loop to themselves, so absorbing chains have a transient class. One can answer questions about expected number of steps until or probabilities of ending up in a given state easily by modifying a given chain to be absorbing.

For instance, starting with an irreducible chain, we can answer, "What is the expected time that a chain starting from i visits j?" Make j an absorbing state. Now $\{j\}$ is its own recurrent class; the other states are transient, and we have only to sum row i of $\mathbf{M}$ to find the total expected time until the chain leaves the transient states (and enters j).

In general, we can collapse recurrent classes to individual points and reduce questions of entering a recurrent class to those of hitting particular points (which we've solved above).

2.5 Positive and Null-Recurrence

We say X_n is a *recurrent* if for each state x, $P(X_n = x \text{ for infinitely many } n) = 1$. A recurrent chain that returns to one state infinitely often necessarily returns to every state infinitely often, so is recurrent.

How do we characterize recurrence and transience? Define the *total number of returns to state x* as

$$R = \sum_{n=0}^{\infty} I\left[X_n = x\right]$$

Then

$$\mathbf{E}[R] = \sum_{n=0}^{\infty} P(X_n = x) = \sum_{n=0}^{\infty} p_n(x, x)$$

Now we characterize $\mathbf{E}[R]$ in terms of the geometric distribution: let $T = \min\{n > 0 : X_n = x\}$, with $T = \infty$ if the chain never returns to x, by convention. If $P\{T < \infty\} = 1$, then the chain always returns, and returns after that, and again, so the chain is recurrent. But if $P\{T < \infty\} = q < 1$, then the chain may only return (value of R) finitely many times. In particular, if $R = m$, then the chain returned $m-1$ times and failed to return the the mth time. That is, on the mth flip of the "go to infinity" coin, the coin said, "go to infinity," so we won't come back. Hence $P(R = m) = q^m(1-q)$ and in particular $R \sim \text{Geo}(1-q)$. Hence $\mathbf{E}[R] = 1/(1-q) < \infty$. In conclusion, **an irreducible Markov chain is transient if and only if** $E[R] = \sum_{n=0}^{\infty} p_n(x, x) < \infty$.

In the infinite case, the existence of a *limiting* probability distribution, i.e. a π such that for $x, y \in S$, we have $\lim_{n\to\infty} p_n(y, x) = \pi(x)$, becomes more subtle. As in the finite case, if this limit is zero, then there is no limiting distribution. But this limit can still be zero and the chain can be recurrent. We call this *null recurrence* to distinguish from *positive recurrence* (when the limit is positive, and behavior is very similar to a finite chain).

An irreducible, aperiodic chain is positive-recurrent if and only if its limiting distribution exists and is independent of y and is also the invariant distribution:

$$\sum_{y\in S} \pi(y)p(y, x) = \pi(x)$$

Recall that starting from state x, the time to first return to x is $T = \min\{n > 0 | X_n = x\}$ and $\mathbf{E}[T|X_n = x] = 1/\pi(x)$. To test for positive recurrence, it suffices to find an invariant distribution or show that one does not exist.

To discriminate null-recurrence and transience, in a null-recurrent chain, $T < \infty$ with probability 1, but $\mathbf{E}[T] = \infty$, whereas in transience, $T = \infty$ with positive probability.

The combination with trying to find an invariant distribution and looking at $\lim_n p_n(x, y)$ should suffice to classify the chain. In summary, a chain is

1. *Positive recurrent* if and only if there exists a π with $\sum_x \pi(x) = 1$ with $\sum_{y\in S} \pi(y)\pi(y, s) = \pi(x)$. In particular, this means $\mathbf{E}[T|X_n = x] = 1/\pi(x) < \infty$
2. *Null recurrent* if $\mathbf{P}(T < \infty) = 1$
3. *Transient* if $\mathbf{P}(T = \infty) = 1$

2.6 Random Walk

2.7 Branching Processes

In a branching process, each individual produces i offspring with probability p_i. Let $\mu = \sum_{i=0}^{\infty} ip_i$ denote the mean number of offspring produced by an individual. Then $\mathbf{E}[Y_1 + \ldots + Y_k] = k\mu$ and $\mathbf{E}[X_n] = \mu^n \mathbf{E}[X_0]$. Now if $\mu \le 1$ and $p_0 > 0$ (where p_0 is the probability that one (each) individual produces zero offspring), then the population dies with probability 1.

If on the other hand $\mu > 1$, then the extinction probability, a, is less than 1, and is the *smallest* root of

$$a = \phi(a) = \sum_{k=0}^{\infty} a^k \mathbf{P}[X = a]$$

3 Continuous-Time Markov Processes

3.1 Poisson Processes

In the Poisson process, X_t represents the customers who have arrived by time t. We assume that arrivals (and their times) are independent, the rate at which customers arrive are constant, and customers arrive one at a time. Mathematically,

$$\begin{aligned} P(X_{t+\Delta t}) &= 1 - \lambda\Delta t + o(\Delta t) \\ P(X_{t+\Delta t} = X_t + 1) &= \lambda\Delta t + o(\Delta t) \\ P(X_{t+\Delta t} \ge X_t + 2) &= o(\Delta t) \end{aligned}$$

By dividing t up into tiny bins and taking the Poisson approximation to the binomial, or by using the derivative

$$P_k'(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[P\left(X_{t+\Delta t} = k\right) - P\left(X_t = k\right) \right]$$

and substituting the conditioning for

$$\begin{aligned} P\left(X_{t+\Delta t} = k\right) &= P\left(X_t = k\right) P\left(X_{t+\Delta t} = k | X_t = k\right) \\ &\quad + P\left(X_t = k-1\right) P\left(X_{t+\Delta t} = k | X_t = k-1\right) \\ &\quad + P\left(X_t \leq k-2\right) P\left(X_{t+\Delta t} = k | X_t \leq k-2\right) \\ &= P_k(t)(1 - \lambda \Delta t) + P_{k-1}(t) \lambda \Delta t + o(\Delta t) \end{aligned}$$

And thus $P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t)$, which is a difference-differential equation with boundary conditions $P_0'(t) = \lambda P_0(t)$ and $P_0(0) = 1$. To solve this equation, first solve $P_0(t) = e^{-\lambda t}$. The key trick is the substitution $f_k(t) = e^{\lambda t} P_k(t)$. Then $f_k(0) = 1$ and $f_k'(t) = e^{-\lambda t}\left[\lambda P_k(t) + P_k'(t)\right] = \lambda f_{k-1}(t) = \lambda e^{\lambda t} P_{k-1}(t)$, which we can see is the original equation. Let's get the other boundary condition: $f_0(t) = e^{\lambda t} P_0(t) = 1$. Then $f_1'(t) = \lambda f_0(1) = \lambda$. By induction (successive integration of polynomials), we have $f_k(t) = \lambda^k t^k / k!$ so that

$$P_k(t) = P\left(X_t = k\right) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

which says that the number of customers who have arrived by time t is a Poisson variable with rate λt.

We can also consider waiting times. The waiting time between two successive arrivals is an exponential variable with rate λ. We could derive this from the Poisson distribution, and also directly from the definition of the Poisson process as requiring independent arrivals. This implies memoryless waiting times, and the exponential random variable is the only memoryless one, i.e. $P(T \geq s + t | T \geq t) = P(T \geq s)$.

3.2 Finite Space, Continuous-Time Markov Processes

Suppose we are in state i. To each other state to which we can transition, attach an "exponential alarm clock," i.e. an exponential random variable with rate b_i. The first exponential alarm clock that "rings" determines the state to which we will transition.

Two facts of exponential random variables make this problem easy. First, the minimum of a collection of independent exponential variables with rate $b_1, \ldots, b_n$ is itself exponential with rate $\sum_{i=1}^n b_i$: if $T = \min\{T_1, \ldots, T_n\}$, then

$$\begin{aligned} P(T \geq t) &= P(T_1 \geq t, \ldots T_n \geq t) \\ &= P(T_1 \geq t) \ldots P(T_n \geq t) \\ &= e^{-b_1 t} \ldots e^{-b_n t} \\ &= e^{-(b_1 + \ldots + b_n) t} \end{aligned}$$

Second, the probability that alarm i goes off first is the ratio of parameter i to the sum of all of the other parameters:

$$\begin{aligned} P(T_1 = T) &= \int_0^\infty P(T_2 > t, \ldots, T_n > t) dP(T_1 = t) \\ &= \int_0^\infty e^{-(b_2 + \ldots + b_n) t} b_1 e^{-b_1 t} dt \\ &= \frac{b_1}{\sum_{i=1}^n b_i} \end{aligned}$$

Analogously to the discrete case, we form an *infinitesimal generator* matrix A whose entries $\alpha(x, y)$ give the *rate* of the exponential clock governing a transition from x to y. Because the sum of exponential is exponential (whose rate is the sum of rates), we define the *total out-transition rate from* x as $\alpha(x) = \sum_{y \neq x} \alpha(x, y)$. We can then define a *time-homogeneous continuous-time Markov chain* as a stochastic process X_t satisfying

$$\begin{aligned} P\left(X_{t+\Delta t} = x | X_t = x\right) &= 1 - \alpha(x)\Delta t + o(\Delta t) \\ P\left(X_{t+\Delta t} = x | X_t = y\right) &= \alpha(y, x)\Delta t + o(\Delta t) \end{aligned}$$

Writing $p_x(t) = P\left(X_t = x\right)$ and doing the same differential-equation fu as for Poisson processes, we recover the differential equations (for each x)

$$p'_x(t) = -\alpha(x)p_x(t) + \sum_{y \neq x} \alpha(y, x)p_y(t)$$

Intuitively, this equation says that in a small interval of time, the probability that the chain stays in x . We express the entire system (remember, one equation for each state) as

$$\begin{aligned} \frac{dp(t)}{dt} &= p(t)\mathbf{A} \\ p(t) &= p(0)e^{t\mathbf{A}} \end{aligned}$$

We can demarginalize p to recover the equation in terms of the transition matrix:

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{P}_t\mathbf{A}$$

where $\mathbf{P}_t = e^{t\mathbf{A}}$. We call $\mathbf{A}$ the *infinitesimal generator* of the chain. (The rows of $\mathbf{A}$ sum to zero, the non-diagonals are nonnegative, the diagonals are nonpositive.)

As in linear ODEs, the solution is

$$\mathbf{P}_t = e^{t\mathbf{A}}$$

Of course, the time to transition from state x to y has distribution $\text{Exp}\left(\alpha(x, y)\right)$. If T is the time of out-transition, then the probability that the chain transits to state y is $\alpha(x, y)/\alpha(x)$.

The invariant distribution should be constant with respect to time, so $\pi\mathbf{A} = 0$. Then if the chain is irreducible, extensions of the results for stochastic matrices (essentially, normalizing the rows) shows that the π satisfying the above equation is unique, and all other eigenvalues of $\mathbf{A}$ have a negative real part.

3.2.1 Induced Discrete-Time Chain

A continuous time chain only transitions a discrete number of times. Its transition times are determined by the exponential clocks, and its transition values are determined by the ratios. So construct a new chain which disallows self-loops and for which $p(x, y) = \alpha(x, y)/\alpha(x)$. This chain replicates the state trajectory. Given a state trajectory, the corresponding dwell-time trajectory is an instantiation of $T_i \sim \text{Exp}(\alpha(i))$ where i is the ith state.

3.3 Birth and Death Processes and Queues

There be some definitions and equations for transience, positive, and null-recurrence.

WRITE THOSE GOD-DAMNED EQUATIONS!

4 Optimal Stopping

On a discrete time Markov chain, a *stopping time* is a random variable $T : \Omega \to \mathbb{N}$ such that for $n \in \mathbb{N}$, the event $\{T = n\}$ is $\mathcal{F}_n$-measurable.

An *optimal stopping rule* is a stopping time that maximizes the expected value of the payoff of a Markov chain (a payoff function f maps a state to a real). That a stopping rule must be a stopping time just means we cannot look into the future. Together with the fact that the process is stationary and Markov, we see that a stopping rule must depend only on the current state. So all stopping rules can be expressed in the following: partition the state space $\Omega = S \cup G$ (stop and go states), and stop if $X_n \in G$. If the chain begins at state x, the *value function* is the expected payoff of following the optimal stopping rule (note that we can define the value function without providing an algorithm, though an algorithm will be given later):

$$v(x) = \sup_T \mathbf{E}\left[f\left(X_T\right)|X_0 = x\right]$$

while the optimal stopping rule is

$$T = \inf\left\{n \geq 0 | X_n \in S\right\}$$

It is a theorem that *the value function is the smallest superharmonic ($v(x) \geq \mathbf{P}v(x)$) function that dominates the payoff.* This leads to the following algorithm to find the value function, with optional cost function $c(x)$ denoting the cost to continue if the current state is x and α discount, iterate

$$u_{n+1}(x) = \max\left\{f(x), \alpha\mathbf{P}u_n(x) - c(x)\right\}$$

5 Martingales

A stochastic process M_t is a *martingale* with respect to a filtration $\mathcal{F}_n$ if $\mathbf{E}\left[M_n|\mathcal{F}_m\right] = M_m$. It is a *continuous martingale* if the sample paths $t \mapsto M_t$ are continuous a.s. For discrete-time martingales, it is equivalent to $\mathbf{E}\left[M_{n+1}|\mathcal{F}_n\right] = M_n$. The proof is an illustrative use of the tower property:

$$\begin{aligned}
\mathbf{E}\left[M_{n+2}|\mathcal{F}_n\right] &= \mathbf{E}\left[\mathbf{E}\left[M_{n+1}|\mathcal{F}_{n+1}\right]|\mathcal{F}_n\right] \\
&= \mathbf{E}\left[M_{n+1}|\mathcal{F}_n\right] \\
&= M_n
\end{aligned}$$

and the rest follow by induction.

5.1 Conditional Expectation

The *conditional expectation* of X given $\mathcal{F}$, denoted $\mathbf{E}\left[X|\mathcal{F}\right]$ is the unique $\mathcal{F}$-measurable random variable such that for $A \in \mathcal{F}$, we have $\mathbf{E}\left[X\chi_A\right] = \mathbf{E}\left[\mathbf{E}\left[X|\mathcal{F}\right]\chi_A\right]$.

If Y is measurable with respect to $X_1, \ldots, X_n$, that is equivalent to saying that Y deterministic given $X_1, \ldots, X_n$. More precisely, the *Doob-Dynkin lemma* says that Y is $\sigma(X_1, \ldots, X_n)$ measurable iff $Y = g(X_1, \ldots, X_n)$ for some Borel-measurable $g : \mathbb{R}^n \to \mathbb{R}$.

Conditional expectation is linear in its first argument just like ordinary expectation. In addition, the following rules hold:

1. *Tower property:* If $\mathcal{E} \subset F$, then $\mathbf{E}\left[\mathbf{E}\left[X|\mathcal{F}\right]|\mathcal{E}\right] = \mathbf{E}\left[X|\mathcal{E}\right]$.
2. *Independence tells us nothing:* If $Y \perp \mathcal{F}$, then $\mathbf{E}\left[Y|\mathcal{F}\right] = \mathbf{E}\left[Y\right]$.
3. *Taking out what's known:* If Z is $\mathcal{F}$-measurable, then $\mathbf{E}\left[XZ|\mathcal{F}\right] = Z\mathbf{E}\left[X|\mathcal{F}\right]$.

5.2 Optional Sampling and Uniform Integrability

Let M_t be a martingale with respect $\mathcal{F}_t$ and T a stopping time with $\mathbf{P}(T < \infty) = 1$, and $\mathbf{E}\left[|M_T|\right] < \infty$ and (difficult condition) $\lim_{n\to\infty} \mathbf{E}\left[|M_n|\chi_{T>n}\right] = 0$. Then $\mathbf{E}\left[M_T\right] = \mathbf{E}\left[M_0\right]$. (We can also replace with conditional expectations (conditioned on the same event)).

A collection $\mathcal{C}$ of random variables is *uniformly integrable* if for $\epsilon > 0$, there exists K such that for $X \in \mathcal{C}$, we have

$$\mathbf{E}\left[|X|\,\chi_{|x|>K}\right] < \epsilon$$

1. UI *implies* that for $\epsilon > 0$, there exists a $\delta > 0$ such that if $\mathbf{P}(A) < \delta$, then $\mathbf{E}[|X|\chi_A] < \epsilon$ for all $X \in \mathcal{C}$.

   (a) If M_t is an uniformly integrable martingale with respect to X_t and T is a stopping time with $\mathbf{P}(T < \infty) = 1$, then $\lim_{n\to\infty} \mathbf{P}(T > n) = 0$. So for $\epsilon > 0$, we can choose n so large so that $\mathbf{P}(T > n) < \delta$ so that $\mathbf{E}[|M_n|\chi_{T>n}] < \epsilon$. Thus, a uniformly integrable martingale automatically satisfies the difficult condition.

2. If the L^p norms of elements in $\mathcal{C}$ are uniformly bounded, then $\mathcal{C}$ is UI for $p > 1$. In particular, bounded martingales are UI.

5.3 Martingale Convergence

Consider a martingale and an interval $[a, b]$. An *upcrossing* occurs when the martingale enters the interval from the bottom and leaves at the top.

A uniformly bounded martingale M_t converges a.s. to M_∞, which is measurable wrt $\mathcal{F}_\infty = \bigvee_{s\geq 0} \mathcal{F}_s$ And if M_t is UI (not necessarily uniformly bounded), then $M_\infty = \lim_{n\to\infty} M_n$ exists and $\mathbf{E}[M_\infty] = \mathbf{E}[M_0]$. To prove this, we fix arbitrary $a < b$ and denote by U_n the number of upcrossings (number of times when the martingale passed through the entire interval $[a, b]$). We show $P(U_\infty < \infty) = 1$. That is, the numbers of upcrossings is finite with probability 1. If our proofs allow us to take $b - a \to 0$, then

6 Brownian Motion

A *Brownian motion*, or *Wiener process with variance parameter* σ^2 is a stochastic process W_t such that

1. $W_0 = 0$,
2. Non-overlapping increments are independent, and for $s < t$, the variable $X_t - X_s \sim \mathcal{N}\left(0, (t-s)\sigma^2\right)$,
3. The paths $t \mapsto W_t$ are continuous.

The *quadratic variation* of a Brownian motion, denoted $\langle W \rangle_t$, is the limit as $n \to \infty$ of

$$Q_t^{(n)} = \sum_{j=0}^{n-1} \left[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}\right]^2$$

Since $W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t} \sim \mathcal{N}(0, t/n)$, the squared increments $\left[W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}\right]^2 \sim \frac{t}{n}\chi^2(1)$. Since increments are independent,

$$\begin{aligned}
\mathbf{E}\left[Q_t^{(n)}\right] &= \sum_{j=0}^{n-1} \mathbf{E}\left[\left(W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}\right)^2\right] \\
&= \frac{n}{nt}\mathbf{E}\left[\chi^2(1)\right] = t \\
\mathbf{V}\left[Q_t^{(n)}\right] &= \sum_{j=0}^{n-1} \mathbf{V}\left[\left(W_{\frac{j+1}{n}t} - W_{\frac{j}{n}t}\right)^2\right] \\
&= \frac{nt^2}{n^2}\mathbf{V}\left[\chi^2(1)\right] = \frac{2t^2}{n}
\end{aligned}$$

which goes vanishes as $n \to \infty$. Thus, the quadratic variation of Brownian motion is constantly t.

Brownian motion has the *strong Markov property.* That is, for a Brownian motion X_t and stopping time $T = \inf\{t : X_t = x\}$, the process $Y_t = X_{t+T} - X_T$ is a Brownian motion independent of $\mathcal{F}_T$.

6.1 Reflection Principle (Density of the Running Maximum)

We use this property to prove the *reflection principle:* If X_t is a BM starting at a with parameter σ^2, then for $b > a$ and $t > 0$, we have

$$\mathbf{P}(A) = \mathbf{P}\left(X_s \geq b \text{ for some } 0 \leq s \leq t\right) = 2\mathbf{P}\left(X_t \geq b | X_0 = a\right)$$

The trick is to characterize A as a stopping time: let T denote the first time X_s hits b. Consider the event $\{X_t \geq b, T \leq t\}$. In fact, the $T \leq t$ is redundant, because if $T > t$[2], then we could not have $X_t \geq b$ (due to continuity of sample paths: if the process never equals b before time t, then it cannot possibly be $\geq b$ at time t). So

$$\begin{aligned} \mathbf{P}\left(X_t \geq b\right) &= \mathbf{P}\left(X_t \geq b, T \leq t\right) \\ &= \mathbf{P}\left(T \leq t\right) \mathbf{P}\left(X_t \geq b | T \leq t\right) \end{aligned}$$

Now we will use the strong Markov property and "reflection:" Assume $T < t$ (otherwise $\mathbf{P}(A) = 0$). By definition, $X_T = b$ so that $X_t - X_T = X_t - b \sim \mathcal{N}(0, t - T)$. Although the variance is random, we have due to the zero mean, $\mathbf{P}\left(X_t - b \geq 0 | T < t\right) = \frac{1}{2}$. Plug in to the previous equations to get

$$\begin{aligned} \mathbf{P}\left(X_t \geq b\right) &= \mathbf{P}\left(T \leq t\right) \frac{1}{2} \\ 2\mathbf{P}\left(X_t \geq b\right) &= \mathbf{P}(T \leq t) = \mathbf{P}(A) \end{aligned}$$

(We have not tracked the condition $X_0 = a$ in each step, but the careful reader can just insert it back in appropriately.)

An example (somewhat contrived) of a Markov, but not strong Markov process: Let $\xi \sim \text{Exp}(1)$ and put $dX_t = \chi_{t \geq \xi} dt$ with $X_0 = 0$. Note that ξ is a stopping time because we can write it as $\xi = \inf\{t : X_t = 1\}$. Then $X_{t+\xi} \neq X_\xi$ a.s., so the process is not strong Markov. Yet it is Markov: either $t \geq \xi$, in which case the process is deterministic ODE (and hence Markov), or $t < \xi$, in which case $X_{t+s} = X_s$ a.s. due to the memoryless (Markov) property of the exponential distribution.

6.2 Examples of Symmetry

Let B_t be standard Brownian motion. Then $X = B_4 - B_2 \sim \mathcal{N}(0, 2)$ and $Y = B_6 - B_4 \sim \mathcal{N}(0, 2)$, so

$$\begin{aligned} \mathbf{P}\left(B_4 < B_2 < B_6\right) &= \mathbf{P}\left(B_4 - B_2 < 0, B_2 < B_6\right) \\ &= \mathbf{P}\left(B_4 - B_2 < 0, B_2 - B_4 < B_6 - B_4\right) \\ &= \mathbf{P}\left(X < 0, -X < Y\right) \\ &= \mathbf{P}\left(Y > -X | X < 0\right) \mathbf{P}\left(X < 0\right) \\ &= \mathbf{P}\left(Y > 0\right) \mathbf{P}\left(|Y| > |X|\right) \mathbf{P}\left(X < 0\right) \\ &= 1/8 \end{aligned}$$

The last few lines bear some explanation: The event $\{Y > -X | X < 0\}$ implies that $Y > 0$ and that the magnitude of Y is greater than that of X. But the event $|Y| > |X|$ is independent from the even that $Y > 0$ (although the variables are not).

7 Stochastic Integration

An *Ito process* has a time and Brownian-motion integral:

$$dZ_t = X_t dt + Y_t dW_t$$

[2]We interchange $T > t$ and $T \geq t$, because the two events differ only on a null set.

by which we mean

$$Z_t = Z_0 + \int_0^t X_s ds + \int_0^t Y_s dW_s$$

Now if $f(t,x)$ is C^2 in x and C^1 in t, then

$$df(t,Z_t) = \frac{\partial f}{\partial s} dW_s + \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial s} X_s + \frac{1}{2}\frac{\partial^2 f}{\partial s^2} Y_s^2 \right) ds$$

when $Z_t = W_t$, it simplifies to

$$df(t,W_t) = \frac{\partial f}{\partial s} dW_s + \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial s^2} \right) ds$$

For $f(t,W_t)$ to be martingale, we require zero drift (ds). This implies $0 = \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial s^2}$.

The integral of Brownian motion with respect to itself is $\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t$.

In ordinary calculus, we can imagine the Taylor expansions of $f(t+dt)$ and drop all terms $(dt)^n$ for $n > 1$. But the productive of two "derivatives" of Brownian motion is $(dW_t)^2 = dt$. This induces the Ito produce rule

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t$$

The Ito integral is a map $\int : L^2([0,t] \times \Omega) \to L_0^2(\Omega)$, taking a stochastic process adapted to $\mathcal{F}_t$ and outputting a random variable with zero mean and finite variance. That is, stochastic integrals are martingales. Further, the *Ito isometry* allows us to calculate variance (which is just the second moment, because the mean is zero):

$$\mathbf{E}\left[\left(\int_0^t Z_s dW_s \right)^2 \right] = \mathbf{E}\left[\int_0^t (Z_s)^2 \, ds \right]$$

Note that the RHS is a Lebesgue, not Ito integral. (It integrates wrt time, not BM).

For example, the integral of a *deterministic* integrand is a mean-zero normal with variance

$$\mathbf{E}\left[\left(\int_0^1 f(s) dW_s \right)^2 \right] = \mathbf{E}\left[\int_0^1 f^2(s) ds \right]$$

But more correctly, we need to show that the Laplace transform of $Z = \int_0^t f(s) dX_s$ agrees with that of the normal density, i.e. $\exp\left\{ \frac{1}{2}\lambda^2 \int_0^t f^2(s) ds \right\}$. Use the classic trick of the exponential martingale (below): define $M_t = \exp\left\{ -\lambda \int_0^t f(s) dX_s - \frac{1}{2}\lambda^2 \int_0^t f^2(s) ds \right\}$ so that $\mathbf{E}[M_t] = \mathbf{E}[M_0] = 1$ (view t as a bounded stopping time). Then $\mathbf{E}[M_0] = \mathbf{E}\left[e^{-\lambda Z} \right] = \exp\left\{ \frac{1}{2}\lambda^2 \int_0^t f^2(s) ds \right\}$.

7.1 Continuous/Exponential Martingales

Integration is a smoothing operator. Therefore, if Y_t is an $\mathcal{F}_t$-adapted process, then $M_t := \int_0^t Y_s dW_s$ is a continuous martingale (sample paths are continuous functions in time). The *quadratic variation* of a continuous martingale is $\langle M \rangle_t := \int_0^t M_s^2 ds$, and $(M_t)^2 - \langle M \rangle_t$ is martingale, because

$$\begin{aligned} d(M_t^2 - \langle M \rangle_t) &= 2M_t dM_t + (dM_t)^2 - d\langle M \rangle_t \\ &= 2M_t dM_t \end{aligned}$$

Note that Ito produce rule produces an extra $(dM_t)^2$ term. Of course, the derivative of quadratic variation is just "stripping off the integral:" $d\langle M \rangle_t = d\int_0^t Y_s^2 ds = (dM_t)^2$.

If Z_t is martingale driven by Brownian motion W_t, then $M_t := \exp\left\{Z_t - \frac{1}{2}\langle Z\rangle_t\right\}$ is the *exponential martingale,* and $dM_t = M_t dZ_t = M_t Y_t dW_t$

Continuous and continuous exponential martingales are useful because they convert arbitrary Ito processes into martingales, which can then be used with optional sampling. This is because the martingales are constructed to satisfy the zero-drift condition.

Note that for a **bounded** stopping time, optional sampling theorem always applies, while if it is just finite (i.e. $\mathbf{P}(T < \infty) = 1$), then you need the technical conditions.

7.2 Drifted Brownian Motion

The process $dS_t = \mu_t dt + \sigma_t dW_t$ is not a martingale. Instead, we look for a process Y_t such that for $M_t = \exp\left(\int_0^t Y_s dW_s - \frac{1}{2}\int_0^t Y_s^2 ds\right)$ such that $M_t S_t$ is martingale. Note that $dM_t = M_t Y_t dW_t$ (why?). This just means that

$$\begin{aligned} d(M_t S_t) &= M_t dS_t + S_t dM_t + dM_t dS_t \\ &= M_t\left(\alpha_t dt + \sigma_t dW_t\right) + S_t M_t Y_t dW_t + M_t Y_t \sigma_t dt \\ &= \left(M_t\alpha_t + M_t Y_t \sigma_t\right)dt + \left(M_t \sigma_t + S_t M_t Y_t\right)dW_t \end{aligned}$$

Define a new measure $\tilde{\mathbf{P}}$ such that $\tilde{\mathbf{E}}[A] := \mathbf{E}[AM_t]$ for $\mathcal{F}_t$-measurable A.

Now we want the above drift to be zero, i.e. $M_t\alpha_t + M_t Y_t\sigma_t = 0$, which implies $Y_t = -\alpha_t/\sigma_t$ and let M_t be the corresponding exponential martingale. Then we can define

$$d\tilde{W}_t = \frac{\alpha_t}{\sigma_t}dt + dW_t$$

which is a continuous martingale under $\tilde{\mathbf{E}}$ (because the Y term cancels the α_t/σ_t), and we have $dS_t = \sigma d\tilde{W}_t$, which is a martingale under $\tilde{\mathbf{E}}$.

We can generalize this to change the drift to S_t to be any arbitrary deterministic process β_t by defining instead $Y_t = \frac{\beta_t - \mu_t}{\sigma_t}$. This is *Girsanov's theorem.* Let M_t be as above. Then we can define a new measure, $dQ = M_t dP$, and the process $d\widetilde{W_t} = \frac{\mu_t}{\sigma_t}dt + dW_t$ is actually a martingale under dQ.

The *generator* of drifted Brownian motion is $\mathcal{A} = \mu\frac{\partial}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}$. That is, for suitable f, with $\mathbf{E}^x[\cdot] = \mathbf{E}[\cdot|S_0 = x]$ under the drift measure, we have

$$\mathcal{A}f(x) = \lim_{t\downarrow 0}\frac{\mathbf{E}^x[f(X_t)] - f(x)}{t} = \mu\frac{\partial}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}$$

To solve *drifted Gambler's ruin* (for μ, σ homogeneous) i.e. for the stopping time $T_{ab} = \inf\{t \geq 0 | S_t \notin (a,b)\}$, first consider the finitude and boundedness (recall that one-dimensional Brownian motion is recurrent; regardless, the process S_t is bounded in this interval. Construct a scale function f such that $f(S_t)$ is martingale. So use Ito's lemma to conclude

$$df(S_t) = \left(\mu\frac{\partial f}{\partial t} + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma\frac{\partial f}{\partial x}dW_t$$

so solve an ODE to set the inside integral to zero (hint hint: try exponentials, in particular, $f(x) = \exp\left\{-\frac{2\mu}{\sigma^{@}}x\right\}$. Then $f(0) = 1$ and we use optional sampling, blah blah blah, to get

$$\mathbf{P}\left(X_T = a | X_0 = x\right) = \frac{f(x) - f(b)}{f(a) - f(b)}$$

7.3 Duration Problems

A note on integral transforms: Let X be a random variable. Then its

1. *Moment generating function* is $\mathbf{E}\left[e^{tX}\right]$
2. *Laplace transform* is $\mathbf{E}\left[e^{-tX}\right]$
3. *Characteristic function* is $\mathbf{E}\left[e^{itX}\right]$

The difference is that the characteristic function always exists (because $\left|e^{itx}\right| = 1$). The Laplace transform is nothing but the negative moment generating function (its derivatives yield negative moments). When we use with stopping times, we generally use MGF/Laplace to avoid dealing with imaginary numbers (and we restrict to existing moments).

To solve duration problems, we find the Laplace transform of the stopping time (using an exponential martingale and optional sampling) and compare it to the normal distribution.

Example: Let $X_t = \mu t + \sigma dW_t$ be a drifted Brownian motion and $T = \inf\{t \geq 0 | X_t \leq a\}$ for some $a < 0$. Find the Laplace transform (negative moment generating function) $\mathbf{E}\left[e^{-\lambda T}\right]$. To do so, we make a bounded martingale with the term we care about, i.e. $M_t = e^{-\lambda t} f(X_t)$. Why do we need a bounded martingale? Because the stopping time is finite but not bounded (drifted Brownian motion is transient, regardless of direction!). Recall that only if the stopping time is *bounded* can we dispense with the technical conditions on the martingale, but this stopping time is not bounded. But we don't need to show uniform integrability or any of that silliness, since L^p-boundedness (and in particular, L^∞ boundedness, i.e. max bound) suffices for uniform integrability.

First, we choose f to make M_t martingale. The intuition, which I have not yet understood well enough to justify, is to note that $e^{-\lambda t}$ is a (deterministic/degenerate) Ito process with drift $-\lambda$, so we choose f to have drift $+\lambda$. This entails solving the ODE

$$\frac{1}{2}\sigma^2 f''(x) + \mu f'(x) = \lambda f(x)$$

plugging and chugging according to the formula sheet gives

$$c_{\pm} = -\frac{\mu}{\sigma^2} \pm \sqrt{\frac{\mu^2}{\sigma^4} + \frac{2\lambda}{\sigma^2}}$$

We have completed the martingale objective of the quest. Next is boundedness. By definition of our stopping time, X_t is bounded below by a, but it is not bounded above. We thus choose the negative exponent, and $f(x) = \exp\{c^{+x}\}$. We then apply optional sampling $\mathbf{E}[M_0] = \mathbf{E}[M_T]$ and retrieve the answer. And, we can take moments by differentiation (and negation!) (So that was actually baby Feynman-Kac?)

7.4 Feynman-Kac

The solution to the PDE

$$\begin{aligned} -\dot{V}(t,x) &= \frac{1}{2}b^2(t,Z_t)V''(t,x) + a(t,Z_t)V'(t,z) + v(t,Z_t)V(t,x) = 0 \\ V(t_0,x) &= f(x) \end{aligned}$$

is

$$V(t,x) = \mathbf{E}^x\left[f\left(Z_{t_0}\right)\exp\left\{\int_t^{t_0} v(s,Z_z)ds\right\}\right]$$

where $dZ_t = a(t,Z_t)dt + b(t,Z_t)dW_t$.