

Stochastic Processes II

Introduction
Renewal Theory
 Limit Theorems
 Key Renewal Theorem and Applications
 Alternating Renewal Processes
Martingales
 Applications
 Martingale Varieties
 Cauchy Martingales
 Concentration Inequalities
 Applications
Brownian Motion
 MST for Continuous Time
 Brownian Motion with Drift
 Reflections and Maxima
 Gaussian Processes
Stochastic Calculus
 Ito integral
 Ito's Formula
 Applications of Ito's Formula
Markov Chains
 Markov Chains on Finite State Spaces
 Long Run Behavior
 Markov Chains on Countable State Spaces
 Polya's Theorem

Stochastic Processes II

Introduction

- We have an experiment -> set of possible outcomes called the sample space Ω
- An event is a subset of outcomes
- Axioms of probability
 - $0 \leq P(E) \leq 1$ for all events E ($P(E)$ = probability of E)
 - $P(\Omega) = 1$
 - If E_1, E_2, \dots disjoint events, then $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$. Disjoint means $E_i \cap E_j = \emptyset$ whenever $i \neq j$. Probability of the union is the sum of probabilities for disjoint events.
- RV is a function from Ω into \mathbb{R} .
- A stochastic process is a collection of RVs $(X_t)_{t \in T} = X$. T can be $\{1, 2, 3, \dots\}$ or $[0, \infty)$; a sequence or continuum of RVs.

Renewal Theory

- A counting process for which the interarrival times are iid with an arbitrary distribution.
- Let X_1, X_2, \dots be iid nonnegative RVs. Assume $P(X_n = 0) < 1$, there is some chance that our RV is nonzero. Let $\mu = E(X_n)$ where we allow μ to be ∞ , but we have a well-defined expected value.
 - For example, if X is integer-valued and $P(X = k) = \frac{c}{k^2}$ where $c \sum \frac{1}{k^2} = 1$ then $E(X) = \infty$
- These variables are called **interarrival times**. Imagine clients arriving at a server, X_i = time between $(i-1)th$ and i th arrival. Clients are immediately served (no queueing involved). Let $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$. Then S_n = arrival time of the n th client.
- Let $N(t) = \sup\{n : S_n \leq t\}$ (\sup since theoretically there could be infinitely many) = number of clients arriving by time t . Maximum n such that the n th client has arrived by time t .
- The stochastic process $(N(t))_{t \geq 0}$ is called a **renewal process**. Since the interarrival times are iid, it follows that at each renewal the process probabilistically starts over. Think of replacing lightbulbs - it stays on for a certain time before being replaced. $N(t)$ is the number of bulbs you have to replace by time t . Plotted, we see piecewise constant function, jumping to the next integer i at each S_i . The times at which the jumps happen are random. A Poisson process is an example of a renewal process, limited to interarrival times with exponential random variables.
- Let $F_n(t) = P(S_n \leq t)$. This function is a CDF of a sum of RVs S_n , may not be easy to evaluate but can still be quite useful.
- **$N(t) / S_n$ Equivalence** that $N(t) \geq n \iff S_n \leq t$; the number of variables up to the time t iff the n th customer has arrived by time t . Therefore $P(N(t) = n) = P(N(t) \geq n) - P(N(t) \geq n+1) = P(S_n \leq t) - P(S_{n+1} \leq t) = F_n(t) - F_{n+1}(t)$. If we know the CDF then we can calculate the probabilities for $N(t)$.
- **Renewal function:** expected number of variables up to time t . The renewal function $m : [0, \infty) \rightarrow \mathbb{R}$ is defined as $m(t) = E(N(t))$ = expected number of arrivals by time t .
- Generally we observe the renewal process for a while, observe the distribution of interarrival times then try to say something about what we can expect in the future.
- **Theorem** (Proposition 3.2.1, 3.2.2): $m(t) = \sum_{n=1}^{\infty} F_n(t)$ and $m(t) < \infty$ for all t .
 - In words, $N(t)$, the number of renewals by time t , has finite expectation.
 - We can get F from the interarrival times, then calculate m and get the expected number of arrivals in some interval i in the future. $m(t)$ is finite no matter what distribution we use.
 - **Lemma 1:** Let X be a non-negative integer valued RV with $E(X) < \infty$ then $E(X) = \sum_{n=1}^{\infty} P(X \geq n)$
 - Proof: $E(X) = P(X = 1) + 2P(X = 2) + 3P(X = 3)$. Regrouping $= [P(X = 1) + P(X = 2) + \dots] + [P(X = 2) + P(X = 3) + \dots] + [P(X = 3) + P(X = 4) + \dots] + \dots$. This sum is equal to $\sum_{n=1}^{\infty} P(X \geq n)$. When summing non-negative quantities, we can rearrange the sum in any way and still get the same answer.
 - **Lemma 2** (Markov's Inequality): Let X be a non negative RV with $E(X) < \infty$ then for any $t > 0$, $P(X \geq t) \leq \frac{E(X)}{t}$
 - The LHS of the inequality is often much more complicated than the RHS, so this comes in handy
 - Proof: Let y be the following RV: $y = \begin{cases} 1 & \text{if } x \geq t \\ 0 & \text{else} \end{cases}$. Let $Z = X/t$. Then if $X < t$ we have $y = 0 \leq Z$ since Z is a nonnegative RV. If $X \geq t$ then $y = 1 \leq x/t = Z$. So y is always $\leq Z \implies E(y) \leq E(Z)$. Finally, $E(y) = 0P(y = 0) + 1P(y = 1) = P(X \geq t)$ and $E(Z) = \frac{E(X)}{t}$
 - **Lemma 3:** Let X be a non-negative RV st $P(X = 0) < 1$, then $E(e^{-X}) < 1$.
 - Note if X were always 0 then e^{-X} would always be 1.

- Proof: Suppose $E(e^{-X}) = 1$ then $E(1 - e^{-X}) = 0$. Since X is nonnegative then $1 - e^{-X} \geq 0$. So $1 - e^{-X}$ is a nonnegative RV whose expectation is 0. So by Markov's Inequality, $P(1 - e^{-X} \geq t) \leq 0$ for any $t > 0 \implies P(1 - e^{-X} = 0) = 1 \implies P(X = 0) = 1$. So we get a contradiction. Thus $E(e^{-X}) \neq 1$. But $e^{-X} < 1$ always so $E(e^{-X})$ cannot be > 1 thus $E(e^{-X}) < 1$
- **Lemma 4:** $m(t) < \infty$ for any t
 - Proof: $P(N(t) \geq n) = P(S_n \leq t)$ by above. $= P(e^{-S_n} \geq e^{-t}) \leq \frac{E(e^{-S_n})}{e^{-t}}$. Using the fact that X 's are iid, then $= e^t(E(e^{-X_1})^n) = e^t p^n$. By Lemma 1 $p = E(e^{-X_1})$. Then $m(t) = E(N(t)) = \sum_{n=1}^{\infty} P(N(t) \geq n) = \sum_{n=1}^{\infty} P(S_n \leq t) = \sum_{n=1}^{\infty} F_n(t)$. By the above $\sum_{n=1}^{\infty} P(N(t) \geq n) \leq e^t \sum_{n=1}^{\infty} p^n < \infty$ since $0 \leq p < 1$.
- Note the above also shows $E[N^r(t)] < \infty$ for all $t, r \geq 0$
- Recall: $\mu = E(X_1) = E(X_2) = \dots$ which could be infinite.

Limit Theorems

- **Theorem** (Proposition 3.3.1): With probability 1, $\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu}$
 - As we take larger times, the number of arrivals goes to infinity. The only way in which $N(\infty)$, the total number of renewals that occurs, can be finite is for one of the interarrival times to be infinite (the next arrival never occurs).
 - $\lim_{t \rightarrow \infty} \frac{N(t)}{t}$ is the rate at which $N(t)$ goes to infinity. Per unit of time, how does $N(t)$ change? $S_{N(t)}$ is the time of the last renewal up to time t and $S_{N(t)+1}$ is the time of the first renewal after time t .
 - We will use the Strong Law of Large Numbers (SLLN) which says that $\frac{S_n}{n} \rightarrow \mu$ with probability 1. Means that the $P(\lim_{t \rightarrow \infty} \frac{S_n}{n} = \mu) = 1$. This holds for mu finite or infinite.
 - **Lemma:** $\lim_{t \rightarrow \infty} N(t) = \infty$ with probability 1.
 - Proof: The limit always exists because N is an increasing process. $\lim_{t \rightarrow \infty} N(t) < \infty \iff X_n = \infty$ for some n . This is an integer valued process, so finite if it stops at some point. $P(X_n = \infty) = 0$ so $P(\lim_{t \rightarrow \infty} N(t) < \infty) = p(\cup_{n=1}^{\infty} \{X_n = \infty\}) \leq \sum_{n=1}^{\infty} P(X_n = \infty) = 0$ (Probability of union is bounded by sum of probabilities). Fact: for any events A_1, A_2, \dots $P(\cup_{n=1}^{\infty} A_i) \leq \sum_{n=1}^{\infty} P(A_i)$
 - Proof: Note $S_{N(t)} \leq t < S_{N(t)+1}$. The customer we count up to for $N(t)$ arrives by t but the next customer must be after t by the definition of $N(t)$. Recall S_n is the arrival time of the n th client, $N(t) = \#$ of arrivals by time t . This implies $\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)}$. $\frac{S_{N(t)}}{N(t)}$ is the average of the first $N(t)$ interarrival times, and by using our SLLN fact that $N(t) \rightarrow \infty$, $\frac{S_n}{n} \rightarrow \mu$ as $n \rightarrow \infty$ we get $\frac{S_{N(t)}}{N(t)} \rightarrow \mu$ as $t \rightarrow \infty$. This all happens with probability 1.
 - Then $\frac{S_{N(t)}}{N(t)} = \frac{N(t)+1}{N(t)} \frac{S_{N(t)}}{N(t)+1} \rightarrow 1 \times \mu$ as $t \rightarrow \infty$. Therefore $\frac{t}{N(t)} \rightarrow \mu$ as $t \rightarrow \infty$
- Note $S_{N(t)}$ is the time of the last renewal prior to or at time t . Therefore $S_{N(t)+1}$ is the time of the first renewal after time t .
- $\frac{1}{\mu}$ is the rate of the renewal process. By Prop 3.3.1, with probability 1 the long-run rate at which renewals occur is equal to $\frac{1}{\mu}$.
- **Stopping Time:** Let X_1, X_2, \dots denote a sequence of RVs. An integer-valued RV N is said to be a stopping time for the sequence if the event $\{N = n\}$ is independent of X_{n+1}, X_{n+2}, \dots for all n and N is positive integer valued. Said another way, the occurrence or non-occurrence of event $\{N = n\}$ is completely determined by the values of X_1, X_2, \dots for every n .

- Note we allow $N = \infty$ also. Whether you stop or not, our evaluation of whether to stop is completely deterministic at time n .
- Example: Let $N = \min\{n : X_1 + \dots + X_n \geq 5\}$. First n so the sum is at least 5. Then the event $\{N = n\}$ can be rewritten as the event $\{X_1 < 5, X_1 + X_2 < 5, \dots, X_1 + \dots + X_{n-1} < 5, X_1 + \dots + X_n \geq 5\}$. Up to $n-1$, our sum is always less than 5, then at n , the sum is greater or equal to 5. Rewritten this way, it is clear that the values of X_1, \dots, X_n completely determine whether $\{N = n\}$ has happened or not, making N a stopping time for this sequence of X 's.
- **Theorem 3.3.2: Wald's Equation:** If X_1, X_2, \dots are iid RVs having finite expectations ($E|X_i| < \infty$), and if N is a stopping time for X_1, X_2, \dots st $E(N) < \infty$ then $E\left[\sum_i^N X_n\right] = E[N]E[X_1]$
 - Normally with a finite sum, the expectation would just move inside. Here we have a finite sum, but the range of the sum is random (see 217 for theorems on random sums).
 - Fact: If Y_1, Y_2, \dots are RVs st $\sum_{i=1}^{\infty} E|Y_i| < \infty$ then $E(\sum_{i=1}^{\infty} Y_i) = \sum_{i=1}^{\infty} E(Y_i)$ (Consequence of dominated convergence theorem)
 - Proof of Wald: Let $I_n = \begin{cases} 1 & \text{if } N \geq n \\ 0 & \text{else} \end{cases}$. Let $Y_n = X_n I_n = \begin{cases} X_n & \text{if } N \geq n \\ 0 & \text{else} \end{cases}$. Then $\sum_{n=1}^N X_n = \sum_{n=1}^{\infty} Y_n$ since $Y_n = 0$ for $N \geq n$. We will show that $\sum_{n=1}^{\infty} E|Y_n| < \infty$ then we will conclude that $E(\sum_{n=1}^N X_n) = E(\sum_{n=1}^{\infty} Y_n) = \sum_{n=1}^{\infty} E(Y_n)$. Finally we will show that $\sum_{n=1}^{\infty} E(Y_n) = E(N)E(X_1)$
 - First note that the value of I_n is determined by the occurrence or non-occurrence of the event $\{N < n\}$. This event $\{N < n\} = \{N = 1\} \cup \{N = 2\} \cup \dots \cup \{N = n-1\}$. The occurrence or non-occurrence of each of the events on the right can be determined by the values of X_1, \dots, X_{n-1} . Therefore I_n itself is a function of X_1, \dots, X_{n-1} . Since these variables are independent, we conclude that X_n, I_n are independent.
 - So $E|Y_n| = E|X_n I_n| = E(|X_n| | I_n|) = E|X_n| E(I_n) \stackrel{iid}{=} E|X_1| E(I_n)$. Then $= E|X_1| P(N \geq n)$. So in summation, $\sum_{n=1}^{\infty} E|Y_n| = E|X_1| \sum_{n=1}^{\infty} P(N \geq n) = E|X_1| E(N) < \infty$.
 - Thus $E(\sum_{n=1}^N X_n) = E(\sum_{n=1}^{\infty} Y_n) = \sum_{n=1}^{\infty} E(Y_n)$
 - But again, $E(Y_n) = E(X_n I_n) \stackrel{\perp}{=} E(X_n)(I_n) = E(X_1)P(N \geq n)$ So $\sum_{n=1}^{\infty} E(Y_n) = E(X_1) \sum_{n=1}^{\infty} P(N \geq n) = E(X_1)E(N)$
 - Example: Stopping time $E(N) = \infty$ and Wald's equation fails. Let X_1, \dots iid with $P(X_1 = 1) = P(X_1 = -1) = 1/2$. Let $S_n = \sum_{i=1}^n X_i$, $S_0 = 0$. let $N = \min\{n : S_n = 1\}$. This is a random walk, looking for first time sum equal to 1. The event $\{N = n\} = \{S_1 \neq 1, \dots, S_{n-1} \neq 1, S_n = 1\}$, so N is a stopping time but Wald's equation does not hold because $E(N)E(X_1) = 0$ since $E(X_1) = 0$ but $E(\sum_{i=1}^N X_i) = E(S_N) = 1$ since $S_N = 1$. We can prove that the sum will definitely hit 1 at some point.
 - Example: Let X_1, X_2, \dots iid $P(X_i = 1) = P(X_i = -1) = 1/2$. Let $T = \begin{cases} 1 & \text{if } X_1 > X_1 + X_2 \text{ and} \\ 2 & \text{if } X_1 \leq X_1 + X_2 \end{cases}$ $S_n = \sum_{i=1}^n X_i$. Stop at 1 if somehow you know in the future this will yield you more money but continue to step 2 otherwise. T is not a stopping time since requires future knowledge. For all possible combinations of X_1, X_2 , $E(S_T) = 1/2 \neq E(T)E(X_1) = 0$. Wald's lemma fails because T is not a stopping time.
 - **Theorem (Elementary Renewal Theorem):** $\lim_{t \rightarrow \infty} \frac{m(t)}{t} = \frac{1}{\mu}$
 - In words, while the PP is the only renewal process whose renewal function $m(t) = \lambda t$ is exactly linear, all renewals are asymptotically linear.
 - Note that $\frac{m(t)}{t} = E\left(\frac{N(t)}{t}\right)$, so it looks like we have this result from above. However, $X_n \rightarrow a$ does not imply $E(X_n) \rightarrow a$ so we must prove this result.

- This tells us about the behavior of the expected number of variables. Blackwell's renewal theorem is a more useful version of this.
- **Lemma:** X_1, X_2, \dots iid interarrival times. Take any $t \geq 0$ then $N(t) + 1$ is a stopping time with respect to the sequence of X 's.
 - Proof: $\{N(t) + 1 = n\} = \{N(t) = n - 1\} = \{X_1 + \dots + X_{n-1} \leq t, X_1 + \dots + X_n > t\}$. The n th variable happens after time t , but the $n-1$ st variable happens before or at t .
- **Corollary 3.3.3:** $E(S_{N(t)+1}) = \mu(m(t) + 1)$ if $\mu < \infty$
 - Proof: We have shown that $E(N(t) + 1) = E(N(t) + 1) = m(t) + 1 < \infty$. So by Wald's equation, $E(S_{N(t)+1}) = E(\sum_{i=1}^{N(t)+1} X_i) = E(N(t) + 1)E(X_1) = \mu(m(t) + 1)$
- Proof: First we will show that $\liminf_{t \rightarrow \infty} \frac{m(t)}{t} \geq \frac{1}{\mu}$. If $\mu < \infty$ then by the corr, $\frac{\mu(m(t)+1)}{t} = \frac{E(S_{N(t)+1})}{t}$, the arrival time of the $N(t) + 1$ customer over t , which we know have to be bigger than 1 since the arrival always happens after t . But $S_{N(t)+1} > t \implies E(S_{N(t)+1}) > t$ always. This implies $\frac{\mu(m(t)+1)}{t} > 1 \implies \liminf_{t \rightarrow \infty} \frac{m(t)}{t} \geq \frac{1}{\mu}$. If $\mu = \infty$, then this is trivially true.
 - Next step: show $\limsup_{t \rightarrow \infty} \frac{m(t)}{t} \leq \frac{1}{\mu}$. Take some constant $M > 0$, then define

$$\bar{X}_n = \begin{cases} X_n & \text{if } X_n \leq M \\ M & \text{if } X_n > M \end{cases}$$

$$\bar{S}_n = \sum_{i=1}^n \bar{X}_i, \mu_M = E(\bar{X}_1), \bar{N}(t) = \sup\{n : \bar{S}_n \leq t\}, \bar{m}(t) = E(\bar{N}(t))$$

We can note that $\bar{S}_{\bar{N}(t)+1} \leq t + M$ - if the interarrival times are bounded by M , the arrival time of the customer that arrives right after t cannot exceed t by more than M . This is a nice result of truncation - we are supplied with an upper bound. Recall $\bar{N}(t) + 1$ is a stopping time and $E(\bar{N}(t) + 1) = \bar{m}(t) + 1 < \infty$ so by Wald, $(\bar{m}(t) + 1)\mu_M \leq t + M \implies \frac{\bar{m}(t)}{t} + \frac{1}{t} \leq \frac{t+M}{t\mu_M}$. Take limsup of both sides $\implies \limsup_{t \rightarrow \infty} \frac{\bar{m}(t)}{t} \leq \frac{1}{\mu_M}$.
 - Note that the n th customer in the new process always arrives before the n th customer in the old process since the interarrival times are equal or smaller. The new process has every interarrival time equal or less than the old process for all n . Thus $\bar{S}_n \leq S_n, \forall n$ and so $\bar{N}(t) \geq N(t)$ for all t . This implies that $\bar{m}(t) \leq m(t) \implies \limsup_{t \rightarrow \infty} \frac{m(t)}{t} \leq \limsup_{t \rightarrow \infty} \frac{\bar{m}(t)}{t} \leq \frac{1}{\mu_M}$. The LHS does not depend on M , so the proof will be complete if we can show that $\mu_M \rightarrow \mu$ as $M \rightarrow \infty$. This follows by a measure theoretic results known as the monotone convergence theorem.

Key Renewal Theorem and Applications

- A nonnegative RV X is said to be **lattice** if there exists $d \geq 0$ st $\sum_{n=0}^{\infty} P\{X = nd\} = 1$ - it only takes on integral multiples of some nonnegative number d . In other words, $X \in \{0, d, 2d, \dots\}$ with probability 1. The largest d is said to be the **period** of X . If X is lattice and F is the distribution function of X , then we say F is lattice.
- **Blackwell's Theorem:** (i) If F is not lattice, then $\lim_{t \rightarrow \infty} m(t+a) - m(t) = a/\mu$ for all $a \geq 0$. (ii) If F is lattice with period d , then $E(\text{number of arrivals at } nd) \rightarrow d/\mu$ as $n \rightarrow \infty$
 - F is the CDF of the interarrival times.
 - If F is not lattice (say X continuous), the expected number of renewals in an interval of length a far from the origin is approx a/μ - the further away from the origin we are, the less influence initial effects will have.
 - When F is lattice with period d , then the limit $g(a) \equiv \lim_{t \rightarrow \infty} [m(t+a) - m(t)]$ cannot exist. Renewals can only occur at integral multiples of d , so the expected number of renewals in an interval far from the origin depends on how many points of nd it contains. If interarrivals are always positive, then part (ii)

says $\lim_{n \rightarrow \infty} P\{\text{renewal at } nd\} = \frac{d}{\mu}$

- For a PP, $m(t)$ is exactly linear equal to λt , so its derivative wrt t is always λ . In general, $m(t)$ may start out wiggly, but for very large t , $m(t)$ will look approximately linear.
- Let h be a function defined on $[0, \infty) \rightarrow \mathbb{R}$. For any $a > 0$, let $\underline{h}_n(a)$ be the supremum and $\bar{h}_n(a)$ the infimum of $h(t)$ over the interval $(n - 1)a \leq t \leq na$. We say that h is directly Riemann integrable if $\sum_{n=1}^{\infty} \bar{h}_n(a)$ and $\sum_{n=1}^{\infty} \underline{h}_n(a)$ are finite for all $a > 0$. Basically these sums quickly go to zero for these sums to be finite. Also $\lim_{a \rightarrow 0} a \sum_{n=1}^{\infty} \bar{h}_n(a) = \lim_{a \rightarrow 0} a \sum_{n=1}^{\infty} h_n(a)$. A sufficient **condition for h to be directly Riemann integrable** is that
 - $h(t) \geq 0, \forall t \geq 0$ - non negative
 - $h(t)$ is nonincreasing
 - $\int_0^{\infty} h(t)dt < \infty$
- **Key Renewal Theorem:** If F is not lattice and if $h(t)$ is directly Riemann integrable then $\lim_{t \rightarrow \infty} \int_0^t h(t-x)dm(x) = \frac{1}{\mu} \int_0^t h(t)dt$ where $m(x) = \sum_{n=1}^{\infty} F_n(x)$ and $\mu = \int_0^{\infty} \bar{F}(t)dt$.
 - Aside: $\int \dots dm(x)$ is the Riemann-Stieltjes integral wrt m . Usual Riemann integral $\int_a^b f(x)dx = \lim \sum_{i=1}^n f(x_i)(x_{i+1} - x_i)$ as you take finer and finer partitions x_0, \dots, x_n of $[a, b]$. The RS integral $\int_a^b f(x)dg(x)$ is the limit of $\sum_{i=1}^n f(x_i)(g(x_{i+1}) - g(x_i))$. If g is differentiable, then since $g(x_{i+1}) - g(x_i) \approx g'(x_i)(x_{i+1} - x_i)$ we get $\int_a^b f(x)dg(x) = \int_a^b f(x)g'(x)dx$
 - Note $\int_a^b f(x)dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x)df(x)$
 - Blackwell and KRT can be shown to be equivalent.
 - Intuition: If m is differentiable, $\int_0^t h(t-x)dm(x) = \int_0^t h(t-x)m'(x)dx$. Now $h(t-x)$ is very small unless x is close to t since h decays rapidly (see conditions on h). So we can ignore the part of the integral where x is far away from t . OTOH, if t is large and x is close to t , Blackwell's thm implies that $m'(x) \approx \frac{1}{\mu}$. So if x close to t and large, $\int_0^t h(t-x)m'(x)dx \approx \int_0^t h(t-x)\frac{1}{\mu}dx = \frac{1}{\mu} \int_0^t h(y)dy \approx \frac{1}{\mu} \int_0^{\infty} h(y)dy$ if t large.
 - KRT is used when one wants to compute the limiting value of $g(t)$, so probability or expectation at time t . Generally derive an equation for $g(t)$ by conditioning on the time of the last renewal prior to t , which yields equations of the form $g(t) = h(t) + \int_0^t h(t-x)dm(x)$.
 - **Lemma 3.4.3:** $P\{S_{N(t)} \leq s\} = \bar{F}(t) + \int_0^t \bar{F}(t-y)dm(y)$ for $t \geq s \geq 0$

Alternating Renewal Processes

- Example: Lightbulb replacements. Want to know the distribution of the amount of time that a lightbulb remains lit. Say you have an inspector that comes in over long periods of time and records how long the bulb has been on. If most are very short lived but a few will live for a while, the inspector will not see the short lived bulbs typically - he will see the longer lasting bulbs since they are installed the majority of the time (the others are replaced quickly, despite their majority). KRT can make this distribution precise.
- The lifetime observed at time t tends to be longer than a random interval between points, X_n . There is length-biased sampling.
- Basic formula: Let X be nonneg RV with cdf F and let A be an event. Then $P(A \cap \{X \leq s\}) = \int_0^s P(A|X=t)dF(t)$.
 - Proof sketch: $P(A \cap \{X \leq s\}) = \sum_{i=1}^n P(A \cap \{s_i < X \leq s_{i+1}\})$ where $0 = s_0 \leq s_1 \leq \dots \leq s_n = s$. This can be written as $\sum_{i=1}^n P(A|s_i < X \leq s_{i+1})P(s_i < X \leq s_{i+1})$ and $P(s_i < X \leq s_{i+1}) = F(s_{i+1}) - F(s_i)$. The s_i are finer and finer partitions - x lies in a small interval so $P(A|s_i < X \leq s_{i+1}) \approx P(A|X = s_i)$ so the sum approaches this integral. When $F' = f$, then

$$P(A \cap \{X \leq s\}) = \int_0^s P(A|X=t)f(t)dt \text{ generally.}$$

- Using this idea, let's define some distributions. Consider process of replacing lightbulbs, let $A(t) = t - S_{N(t)}$ = age of the bulb that is on at time t , since $S_{N(t)}$ is the time of the last replacement. Then $Y(t) = S_{N(t)+1} - t$ = residual life of the bulb that is on at time t . Then $X_{N(t)+1}$ = total lifetime of the bulb that is on at time t . Therefore $X_{N(t)+1} = A(t) + Y(t)$. On a numberline, have a time t with $S_{N(t)}$ to the left and the space between them is $A(t)$, while the space from t to $S_{N(t)+1}$ to the right is $Y(t)$. The gap from $S_{N(t)}$ to $S_{N(t)+1}$ is $X_{N(t)+1}$.
 - We will compute the limiting distribution of $A(t)$, $Y(t)$, and $X_{N(t)+1}$ as $t \rightarrow \infty$. $A(t)$ we can calculate from data and use renewal theory for the other distributions.
 - The expected number of arrivals in a short period of time far in the future is $\frac{1}{\mu}$ times the length of the interval
- Distribution of $A(t)$:** $P(A(t) \leq x) = ?$ for $0 \leq x < t$. This event $\{A(t) \leq x\}$ is the disjoint union of the events $\{t - x \leq S_n \leq t, S_{n+1} > t\}, n = 1, 2, \dots$. If this event happens then these other events must happen; and if one of these events happens then $A(t) \leq x$. The n th event is that the lightbulb burns out between $t - x$, t and the $n+1$ st lightbulb burns out after t .
 - $A(t) \leq x$ means that exactly one of the following must have happened: either the first lightbulb goes out between time $t-x$ and t $S_1 \in [t-x, t]$, $S_2 > t$ and the second lightbulb goes out after t OR the second lightbulb out between $t - x$ and t and third goes out after t $S_2 \in [t-x, t]$, $S_3 > t$...etc. Disjoint events
 - So $P(A(t) \leq x) = \sum_{n=1}^{\infty} P(t - x \leq S_n \leq t, S_{n+1} > t)$ - the probability of $A(t)$ is the same as exactly one of these events has happened.
 - Then $P(t - x \leq S_n \leq t, S_{n+1} > t) = \int_{t-x}^t P(S_{n+1} > t | S_n = y) dF_n(y)$, recalling F_n is CDF of S_n . (Eg. if F_n differentiable and derivative is the pdf of S_n then $P(t - x \leq S_n \leq t, S_{n+1} > t) = \int_{t-x}^t P(S_{n+1} > t | S_n = y) f_n(y) dy$. This is a generalization of that.)
 - So we can write $P(A(t) \leq x) = \sum_{n=1}^{\infty} \int_{t-x}^t P(S_{n+1} > t | S_n = y) dF_n(y)$.
 - General measure-theoretic results imply that the infinite sum can be taken inside the integral:

$$P(A(t) \leq x) = \int_{t-x}^t \sum_{n=1}^{\infty} P(S_{n+1} > t | S_n = y) dF_n(y).$$
 - Then (for $y \leq t$) $P(S_{n+1} > t | S_n = y) = P(x_{n+1} > t - y | S_n = y) = P(x_{n+1} > t - y)$ since X_{n+1} and S_n are independent, so conditioning goes away. $P(x_{n+1} > t - y) = \bar{F}(t - y)$ where $\bar{F}(n) = 1 - F(n)$ and F is the CDF of the interarrival times.
 - Therefore $P(A(t) \leq x) = \int_{t-x}^t \bar{F}(t - y) \sum_{n=1}^{\infty} dF_n(y)$. In a R-S integral

$$\int f(x)dg(x) + \int f(x)dh(x) = \int f(x)d(g+h)(x) = \int f(x)dw(x)$$
 for $w = g + h$. Extends to infinite sums under conditions. We know that $m(t) = \sum_{n=1}^{\infty} F_n(t)$. Thus $\sum_{n=1}^{\infty} dF_n(y) = dm(y)$.
 - Then $P(A(t) \leq x) = \int_{t-x}^t \bar{F}(t - y) dm(y) = \int_0^t h(t - y) dm(y)$. Define $h(u) = \begin{cases} \bar{F}(u) & \text{if } u \leq x \\ 0 & \text{else} \end{cases}$. Since $y < t - x$ implies $t - y > x$ implies $h(t-y) = 0$.
 - Returning to the Key Renewal Thm, the h that we defined satisfies the sufficient conditions for KRT, so we can invoke the KRT to get the limit of our probability. h is nonnegative, non increasing, and integral is finite. Assume F is non lattice.
 - From KRT we get $\lim_{t \rightarrow \infty} P(A(t) \leq x) = \frac{1}{\mu} \int_0^{\infty} h(u) du = \frac{1}{\mu} \int_0^x \bar{F}(u) du$. Therefore the limiting CDF of $A(t)$ as $t \rightarrow \infty$ is given by $G(u) = \frac{1}{\mu} \int_0^x \bar{F}(u) du$. By FTC, this is a differentiable function:

$$G'(x) = \frac{1}{\mu} \bar{F}(x) = \frac{1}{\mu} (1 - F(x))$$
. Why is this a valid PDF? Because

$$\mu = E(X_1) = \int_0^{\infty} P(X_1 > t) dt = \int_0^{\infty} \bar{F}(t) dt$$
 (we did not prove but is a fact: we can write expectations as infinite integrals).
 - Example: $X_i \sim \exp(\lambda)$ (we have a PP). Then $F(x) = 1 - e^{-\lambda x}$ and $\mu = \lambda$. Then $G(x) = \lambda e^{-\lambda x}$, as $t \rightarrow \infty$ then $A(t) \sim \text{Exp}(\lambda)$.

- If instead $X_i \sim Unif(0, 1)$ then $F(x) = x$, $x \in [0, 1]$, $\mu = 1/2$. Then $G(x) = 2(1 - x)$. Our pdf is then more likely to be close to 0 than towards 1 - not uniformly distributed.
- **Limiting Distribution of $Y(t)$** , the residual lifetime of the bulb that is on at time t.
 - Want to understand $P(Y(t) \leq x) = ?$. Events $\{Y(t) \leq x\}$ is the union of disjoint events $\{S_n \leq t, t < S_{n+1} \leq t + x\}$ $n = 0, 1, 2, \dots$ Either the first lightbulb goes off between t and t + x OR second lightbulb goes off between t and t + x and first goes off up to t...etc. This means

$$P(Y(t) \leq x) = \sum_{n=0}^{\infty} P(S_n \leq t, t < S_{n+1} \leq t + x)$$

$$= P(t < S_1 \leq t + x) + \sum_{n=1}^{\infty} \int_0^t P(t < S_{n+1} \leq t + x | S_n = y) dF_n(y)$$
 - So the first term = $\bar{F}(t) - \bar{F}(t + x)$ for $\bar{F}(y) = P(X_1 > y)$. Then in total have

$$\bar{F}(t) - \bar{F}(t + x) + \sum_{n=1}^{\infty} \int_0^t P(t - y < x_{n+1} \leq t + x - y | S_n = y) dF_n(y)$$
 Note $S_1 = X_1$ and $S_{n+1} = S_n + X_n$. So $= \bar{F}(t) - \bar{F}(t + x) + \sum_{n=1}^{\infty} \int_0^t P(t - y < x_{n+1} \leq t + x - y) dF_n(y)$ and by iid X we have $P(t - y < x_{n+1} \leq t + x - y) = \bar{F}(t - y) - \bar{F}(t + x - y)$.
 - Then moving the sum inside = $\bar{F}(t) - \bar{F}(t + x) + \int_0^t \bar{F}(t - y) - \bar{F}(t + x - y) \sum_{n=1}^{\infty} dF_n(y)$ and noting $\sum_{n=1}^{\infty} dF_n(y) = dm(y)$: $P(Y(t) \leq x) = \bar{F}(t) - \bar{F}(t + x) + \int_0^t h(t - y) dm(y)$ where $h(u) = \bar{F}(u) - \bar{F}(u + x)$. Notice that $\bar{F}(t) \rightarrow 0$, $t \rightarrow \infty$
 - So by KRT, $\lim_{t \rightarrow \infty} P(Y(t) \leq x) = \frac{1}{\mu} \int_0^{\infty} h(u) du = \frac{1}{\mu} \int_0^{\infty} (\bar{F}(u) - \bar{F}(u + x)) du$. The splitting into two:

$$= \frac{1}{\mu} [\int_0^{\infty} \bar{F}(u) du - \int_0^{\infty} \bar{F}(u + x) du] = \frac{1}{\mu} [\int_0^{\infty} \bar{F}(u) du - \int_x^{\infty} \bar{F}(y) dy]$$
 So in total get in the limit $\lim_{t \rightarrow \infty} P(Y(t) \leq x) = \frac{1}{\mu} \int_0^x \bar{F}(u) du$. This is the same limit as the one we got for A(t)!
 - The age and the residual time have the same distribution. For our uniform example, both the age and the residual time are closer to 0 than to 1!
 - If we look backwards in time, the time between successive events is still independent with distribution F - we have a renewal process in both forward and backward directions.
- **Limiting distribution of $X_{N(t)+1}$** - the total lifetime of the bulb that is on at time t.
 - Look at event $\{X_{N(t)+1} > x\}$ is the disjoint union of the events $\{\{S_n \leq t, S_{n+1} > t, X_{n+1} > x\}\}$ for $n = 0, 1, 2, \dots$
 - Taking $t > x \geq 0$, $P(x_{N(t)+1} > x) = \bar{F}(t) + \sum_{n=1}^{\infty} P(S_n \leq t, S_{n+1} > t, X_{n+1} > x)$ Then

$$= \bar{F}(t) + \sum_{n=1}^{\infty} \int_0^t P(S_{n+1} > t, X_{n+1} > x | S_n = y) dF_n(y)$$
 - Suppose $S_n = y$ - there are two cases to consider. (1) $t - x \leq y \leq t$ and (2) $y < t - x$. Case 1: in this case, $x_{n+1} > x \implies S_{n+1} > t$ and then

$$P(S_{n+1} > t, X_{n+1} > x | S_n = y) = P(X_{n+1} > x | S_n = y) \stackrel{\perp}{=} P(X_{n+1} > x) = \bar{F}(x)$$
 In Case 2: in this case $S_{n+1} > t \implies x_{n+1} > x$. So $P(S_{n+1} > t, X_{n+1} > x | S_n = y) = P(S_{n+1} > t | S_n = y)$

$$= P(X_{n+1} > t - y | S_n = y) \stackrel{\perp}{=} P(X_{n+1} > t - y) = \bar{F}_n(t - y)$$
 - Combining these cases, you get

$$P(x_{N(t)+1} > x) = \bar{F}(t) + \int_0^t \sum_{n=1}^{\infty} P(S_{n+1} > t, X_{n+1} > x | S_n = y) dF_n(y)$$
 implies

$$= \bar{F}(t) + \int_0^t h(t - y) \sum dF_n(y)$$
 where $\sum dF_n(y) = dm(y)$. Here $h(u) = \begin{cases} \bar{F}(x) & \text{if } u \leq x \\ \bar{F}(u) & \text{if } u > x \end{cases}$.
 - By KRT, $\lim_{t \rightarrow \infty} P(X_{N(t)+1} > n) = \frac{1}{N} \int_0^{\infty} h(u) du = \frac{x \bar{F}(u)}{\mu} + \frac{1}{\mu} \int_x^{\infty} \bar{F}(u) du$.
 - Simplifying using integration by parts: $\int_x^{\infty} y dF(y) = - \int_x^{\infty} y d\bar{F}(y)$ (Think of this RS as integrating wrt the derivative of f). $= [-y \bar{F}(y)]_x^{\infty} + \int_x^{\infty} \bar{F}(y) dy = x \bar{F}(x) + \int_x^{\infty} \bar{F}(y) dy$. Thus

$$\lim_{t \rightarrow \infty} P(X_{N(t)+1} > n) = \frac{1}{\mu} \int_x^{\infty} y dF(y)$$
 This is known as the size biased distribution of F. If F has pdf f, then the size biased distribution has pdf $\frac{1}{\mu} x f(x)$ (Since it has expectation μ , dividing by μ gives us a pdf that integrates to one).
 - The lightbulbs have some distribution to their lifetimes, but if you look at the renewal process they are only replaced when they go out. At some large time t, the lightbulb in place will have a distribution biased towards the right - longer than the typical lightbulb. Why? If some lightbulbs live short and some

long, some time in far future it is more likely that you will see a long lived bulb than short.

- This has relevance to survey sampling. Want some distribution of family income in a population, select some phone numbers at random and survey them. The distribution will be the size biased distribution of the actual distribution - larger families will have more representatives - a family of ten people is more likely to get a call than a single person.

Martingales

- A stochastic process $\{Z_n\}_{n \geq 1}$ is called a martingale if
 - $E|Z_n| < \infty$ for all n
 - $E(Z_{n+1}|Z_1, \dots, Z_n) = Z_n$ for all n
- Let $f(Z_1, \dots, Z_n) = E(Z_{n+1}|Z_1 = z_1, \dots, Z_n = z_n)$ then the RV $f(Z_1, \dots, Z_n)$ is denoted by $E(Z_{n+1}|Z_1, \dots, Z_n)$.
- Generalized version of a fair game - if Z_n is a gambler's fortune after the n th gamble, then his fortune after the $(n+1)$ st gamble is equal to his fortune after the n th gamble: $E[Z_{n+1}] = E[Z_n] = E[Z_1]$.
 - Sums of zero mean random variables, products of mean 1 random variables.
- Example: Let X_1, \dots be independent RVs with $E(X_i) = 0$ and $E|X_i| < \infty$. Let $S_n = \sum_{i=1}^n X_i$, $S_0 = 0$ Then $\{S_n\}_{n \geq 0}$ is a martingale
 - Proof: $E(S_{n+1}|S_0 = s_0, \dots, S_n = s_n) = E(S_{n+1}|x_1 = s_1, X_2 = s_2 - s_1, \dots, X_n = s_n - s_{n-1})$. Then $= E(X_{n+1} + s_n|x_1 = s_1, X_2 = s_2 - s_1, \dots, X_n = s_n - s_{n-1})$. By independence, $E(X_{n+1}|x_1 = s_1, X_2 = s_2 - s_1, \dots, X_n = s_n - s_{n-1}) = E(X_{n+1}) = 0$. Thus $E(S_{n+1}|S_0 = s_1, \dots, S_n = s_n) = s_n$. Therefore $E(S_{n+1}|S_0, \dots, S_n) = S_n$. Also $E|S_n| \leq E|X_1| + \dots + E|X_n| < \infty$ by triangle ineq. Thus $\{S_n\}_{n \geq 0}$ is a martingale
- Example: Polya's Urn Model. An urn contains one white ball and one black ball. Pick a ball at random. Replace it back to the urn together with an extra ball of the same color. Repeat. Let X_n be the proportion of white balls at time n (start at time 1). We know $X_1 = 1/2$, and at any time the number of balls = $n + 1$. Claim: $\{X_n\}_{n \geq 1}$ is a martingale.
 - Proof: Note $E|X_n| \leq 1$ since a proportion. Now $E(X_{n+1}|X_1 = x_1, \dots, X_n = x_n) = P(\text{picking white ball}|X_1 = x_1, \dots, X_n = x_n) + P(\text{picking black ball}|X_1 = x_1, \dots, X_n = x_n)$. The first prob is $\frac{(n+1)x_n + 1}{n+2}$ - the number of white balls at time n is $(n+1)x_n$. Then the latter is $\frac{(n+1)x_n}{n+2}$. So this expression is $= X_n \frac{(n+1)x_n + 1}{n+2} + (1 - X_n) \frac{(n+1)x_n}{n+2}$. This simplifies to $\frac{(n+2)x_n}{n+2} = X_n$. Thus $E(X_{n+1}|X_1, \dots, X_n) = X_n$ and we have a martingale.
- A **bounded martingale** will always converge to a limit - which is what we have with Polya. With simulation we will see some volatility in the proportion but over time may look like convergence to a limit. But if you run the experiment again, it will also converge, but not necessarily to the same proportion - the limit itself is a random variable.
- Let $\{Z_n\}_{n \geq 1}$ be a martingale. Recall that a RV N is called a stopping time wrt the sequence if for any n , the occurrence or non-occurrence of the event $\{N = n\}$ is determined by the values of Z_1, \dots, Z_n .
- **Martingale Stopping Theorem** (Simple Version): Let $\{Z_n\}_{n \geq 1}$ be a martingale and N be a stopping time st there is some n for which $N \leq n$ always (bounded by some deterministic number). Then $E(Z_N) = E(Z_1)$.
 - Note that N is random, so Z_N is not one of the RVs Z_1, \dots, Z_n . What is Z_N ? Recall that there is an experiment with a set of outcomes Ω st all of our RVs are functions from $\Omega \rightarrow \mathbb{R}$. The RV Z_N is defined as $Z_N(\omega) := Z_{N(\omega)}(\omega)$ - this is called a **stopped random variable**. For example, if for some ω , $N(\omega) = 5$, $Z_5(\omega) = 0.1$ then $Z_N(\omega) = Z_{N(\omega)}(\omega) = Z_5(\omega)$
 - This may not be very useful on its own since many Martingales are not bounded. But combined with

other results will be more useful.

- Proof of MST: Claim: $E(Z_n | Z_1, \dots, Z_k) = z_k$ for any $1 \leq k \leq n$. Proof:

$$E(Z_n | Z_1 = z_1, \dots, Z_{n-2} = z_{n-2})$$

$$= \sum_j E(Z_n | Z_1 = z_1, \dots, Z_{n-2} = z_{n-2}, Z_{n-1} = j) P(Z_{n-1} = j | Z_1 = z_1, \dots, Z_{n-2} = z_{n-2})$$

- So given the first n-2, we can write this decomposition via law of total probability. Then since martingale, can say $= \sum_j j P(Z_{n-1} = j | Z_1 = z_1, \dots, Z_{n-2} = z_{n-2}) = E(Z_{n-1} | Z_1 = z_1, \dots, Z_{n-2} = z_{n-2}) = z_{n-2}$. Proceed by induction. Notice, we used that we have a martingale to take a claim from n-1 to n-2, allowing induction to show its generality for k in our range.

- Next claim: For any $1 \leq k \leq n$ let $Y_k = \begin{cases} 1 & \text{if } N = k \\ 0 & \text{else} \end{cases}$. Then $E(Z_n Y_k) = E(Z_k Y_k)$. Proof: Note that

Y_k is a function of Z_1, \dots, Z_k since N is a stopping time (determined by these Z's). So given

$Z_1 = z_1, \dots, Z_k = z_k$, Y_k is just a constant, say y_k . Thus $E(Z_n Y_k | Z_1 = z_1, \dots, Z_k = z_k) = y_k z_k$ by the first claim, using the Martingale property. Thus

$$E(Z_n Y_k) = \sum_{Z_1, \dots, Z_k} E(Z_n Y_k | Z_1 = z_1, \dots, Z_k = z_k) P(Z_1 = z_1, \dots, Z_k = z_k) \text{ and then}$$

$$= \sum_{Z_1, \dots, Z_k} z_k y_k P(Z_1 = z_1, \dots, Z_k = z_k) = E(Z_k Y_k) \text{ (since the expectation of a function of an RV is that function times its pdf, summed over all possibilities).}$$

- Completing the proof: Note that N must always take one of the values 1,2,...,n. Thus $\sum_{k=1}^n Y_k = 1$, since one point in the sample space gives value 1 and the rest 0. Therefore

$$E(Z_n) = E(Z_n \sum_{k=1}^n Y_k) = \sum_{k=1}^n E(Z_n Y_k) = \sum_{k=1}^n E(Z_k Y_k) = E(\sum_{k=1}^n Z_k Y_k) = E(Z_N) \text{ since for any } \omega \in \Omega, \sum_{k=1}^n Z_k(\omega) Y_k(\omega) = Z_j(\omega) \text{ if } Y_j(\omega) = 1 \text{ and } Y_k(\omega) = 0 \text{ for } j \neq k. \text{ In other words}$$

$$\sum_{k=1}^n Z_k(\omega) Y_k(\omega) = Z_{N(\omega)}(\omega) = Z_N(\omega).$$

- Thus $E(Z_n) = E(Z_N)$. But we also know by the first claim that $E(Z_n | Z_1 = z_1) = z_1$. Then

$$E(Z_n) = \sum_{z_1} E(Z_n | Z_1 = z_1) P(Z_1 = z_1) = \sum_{z_1} z_1 P(Z_1 = z_1) = E(Z_1). \text{ Therefore } E(Z_N) = E(Z_1).$$

- In a fair game, if a gambler uses a stopping time to decide when to quit, then his expected final fortune is equal to his expected initial fortune. We can view the MST as another proof of Wald's equation. Given the Wald conditions, $Z_n = \sum_{i=1}^n (X_i - \mu)$ is a Martingale and therefore $E[Z_N] = E\left[\sum_{i=1}^N X_i\right] - E[N]\mu$.

- Example: Let X_1, \dots be independent mean 0 RVs and $S_n = \sum_{i=1}^n X_i$. We know that $\{S_n\}_{n \geq 0}$ is a martingale. Let $T = \min\{n = S_n \notin (a, b)\}$, $a < 0 < b$. First time it leaves an interval like this (fixed a,b).

- Easy to see T is a stopping time. $\{T = n\} = \{S_0 \in (a, b), S_1 \in (a, b), \dots, S_{n-1} \in (a, b), S_n \notin (a, b)\}$. Completely determined by past events, so T is a stopping time but not a bounded stopping time. Simple MST does not apply here as a note.
- Suppose X_i 's are ± 1 with probability 1/2 each. Then S_n is a simple symmetric random walk. Suppose also that a,b are integers then at time T, the walk is either at a or at b. So S_T can only take one of two values - a or b. So $S_T(\omega) = S_{T(\omega)}(\omega)$. At that time $T(\omega)$ the walk is either a or b.

- Computing a mean time until a given pattern occurs: sequence of iid discrete RVs observed sequentially, expected number until a specified pattern is observed.

- Let N denote the time until the sequence of length X appears. At the end of day N, each of the gamblers 1,...,N-X would have lost 1 unit, gambler N-(X-1) would have won the full better sequence, and the rest to N win depending on how the sequence repeats itself.

- **Dominated Convergence Theorem:** Suppose that $\{X_n\}_{n \geq 1}$ is a sequence of RV converging to a limit RV X. Suppose that there exists a RV Y with $E|Y|$ finite st $|X_n| \leq |Y|$, $\forall n$. Then $\lim_{n \rightarrow \infty} E(X_n) = E(X)$. Special case: **bounded convergence theorem** where Y is a constant.

- **Monotone Convergence Theorem:** If X_n is a sequence of non-negative RVs increasing to a RV X, then the $\lim_{n \rightarrow \infty} E(X_n) = E(X)$. (Alternative version: If Y_1, Y_2, \dots are non-neg RVs then $E(\sum_{i=1}^{\infty} Y_i) = \sum_{i=1}^{\infty} E(Y_i)$)

- Let Z_1, Z_2, \dots be any sequence of RVs, let N be a stopping time for this sequence. Take any $n \geq 1$. Define a new RV N_n as $N_n = \begin{cases} N & \text{if } N \leq n \\ n & \text{if } N > n \end{cases}$. Claim: N_n is a stopping time (for any n), producing sequence of stopping times.
 - Proof: Take any k . If $k < n$ then the event $\{N_n = k\}$ is the same as the event $\{N = k\}$ and so its occurrence or non-occurrence is determined by the values of Z_1, \dots, Z_k . Why? This can only happen if $N = k$, ie. strictly less than n . If $k > n$ then $\{N_n = k\}$ is an impossible event and is therefore vacuously determined by Z_1, \dots, Z_k . Finally, if $k = n$, then $\{N_n = k\} = \{N_n = n\}$, is the same as $\{N \geq n\}$. But $\{N \geq n\} = \{N < n\}^C = \{N = 1\}^C \cap \{N = 2\}^C \cap \dots \cap \{N = n - 1\}^C$. Occurrence or non-occurrence of each of these events can be determined by the values of Z_1, \dots, Z_{n-1} . Thus $\{N_n = n\}$ is also determined by Z_1, \dots, Z_n .
- Lemma:** If $P(N < \infty) = 1$ (recall stopping times can be infinite), then the sequence N_n converges to N as $n \rightarrow \infty$. Moreover $N_1 \leq N_2 \leq \dots$ (an increasing sequence).
 - Proof: The inequality $N_n \leq N_{n+1}$ is simple. If $N \leq n$ then $N_n = N_{n+1} = N$. If $N \geq n + 1$ then $N_n = n$, $N_{n+1} = n + 1$ and so $N_n \leq N_{n+1}$. Next suppose that $N(\omega) < \infty$ for some $\omega \in \Omega$. Say $N(\omega) = k$ then $N_n(\omega) = n$ for $n < k$ and $N_n(\omega) = k$ for $n \geq k$. Therefore $\lim_{n \rightarrow \infty} N_n(\omega) = k = N(\omega)$. Thus if $P(N < \infty) = 1$ then $P(\lim_{n \rightarrow \infty} N_n = N) = 1$.
 - Say you decide to buy a stock if index hits a certain value - this is a stopping time. Say instead you buy a stock if index hits a certain value but within a 10 day window, if doesn't hit by day 10 you buy the stock anyway - this is N_n .
 - Note for any n , N_n is a bounded stopping time, because $N_n \leq n$. This allows us to use the MST.
- Theorem: Better MST.** Let Z_1, Z_2, \dots be a martingale. Let N be a stopping time for this sequence. Suppose that there is a constant C such that $|Z_n| \leq C$ whenever $n \leq N$ (the sequence may not be bounded, but you know there is a non random constant C , before you hit the stopping time the abs val of the sequence is bounded by C). Assume that $P(N < \infty) = 1$. Then $E(Z_N) = E(Z_1)$
 - With the stock buying behavior - the index can hit any value, but we have set a price level at which we buy and then hit our stopping time. Then we know that as long as the stopping time has not been hit, the price level is below our trigger price level.
 - Proof: Take any n . Then N_n is a bounded stopping time, and so $E(Z_{N_n}) = E(Z_1)$. Let $Y_n = Z_{N_n}$. Since $P(N < \infty) = 1$ we know that $N_n \rightarrow N$ as $n \rightarrow \infty$ and since these are integer-valued RVs, this implies that $N_n = N$ for all sufficiently large n , where suff. large is random. Therefore $Y_n \rightarrow Z_N$ as $n \rightarrow \infty$ or $\lim_{n \rightarrow \infty} Y_n = Z_N$. Y is a sequence and Z_N is a single random variable.
 - The value of the index on the day it crosses our limit value is Z_N . Then Y_n is the value of the index on the day it crosses the threshold which could be before Z_N or equal Z_N . If it crosses on day 100, then Y_1 equals value on the first day etc..., Y_{100} equals the value of the index on day 100, but then $Y_{101} = Y_{102} = Y_{103} = \dots = Y_{100}$ since we have crossed the threshold.
 - Moreover by assumption, $|Y_n| = |Z_{N_n}| \leq C$ since $|Z_k| \leq C$ when $k \leq N$ and $N_n \leq N$. Therefore by the bounded convergence theorem, $E(Z_N) = E(\lim_{n \rightarrow \infty} Y_n) = \lim_{n \rightarrow \infty} E(Y_n) = \lim_{n \rightarrow \infty} E(Z_{N_n}) = E(Z_1)$
 - Aside: example of N where there is no such C . Let X_1, \dots iid $N(0,1)$ and $S_n = \sum_{i=1}^n X_i$. Then S_0, S_1, \dots is a martingale. Let $N = \min\{n : |S_n| \geq 1\}$. Then you know that $|S_n| < 1$ for all $n < N$ but no such guarantee can be given for $n = N$ since the X_i 's are unbounded RVs

Applications

- Example: Suppose that a gambler starts with a dollars in his pocket. At each turn he wins 1 dollar with probability $1/2$ and loses with probability $1/2$. He sets a target $b > a$ for himself, planning to continue playing until he either reaches b or goes broke. (a, b integers)
 - (1) What is the probability of reaching b before going broke? Let $S_0 = a$ and $S_n =$ dollar worth of the gambler after playing n games. Let $T = \min\{n : S_n = 0 \text{ or } b\}$ = stopping time of the gambler. (You can check S_0, S_1, \dots is a martingale, since it is just a RW shifted by a). Claim: $P(T < \infty) = 1$ - note this is not obvious, that you always stop eventually.
 - Proof: Divide up the positive integers into disjoint blocks of length b . If the gambler goes on playing indefinitely (allowing negative dollars, ie. borrowing) then if any block has all wins for the gambler (a block where he wins every turn) then T must happen in that block or before it. Say b is 100 dollars, if there is ever a block where the gambler wins all 100 turns, if his dollar amount at the beginning a were between 0 - 100 then he certainly would have more than a within that block or the stopping time (0 or 100) would have already been hit in a prior block. (Why? Suppose that the gambler wins every toss from step $kb + 1$ to $(k + 1)b$, ie. the k th block. Then either (1) $S_{kb} \in [0, b]$. In this case, $S_{(k+1)b} \geq b$ and so $T \leq (k + 1)b$ or (2) $S_{kb} \notin [0, b]$ In this case we know that $T \leq kb$ by the definition of T . So in either case $T \leq (k + 1)b$) But the chance that this happens in any given block of b bets is $(\frac{1}{2})^b = 2^{-b}$ (and not happening $(1 - 2^{-b})$) so $P(\text{this does not happen in blocks } 1, \dots, k) = (1 - 2^{-b})^k$ by independence of blocks. This goes to 0 as k goes to infinity - $P(\text{this never happens}) = 0$. Therefore $P(T = \infty) = 0$ (since T is only infinity when there is no such block). You can think of each block as a toss of a coin - even when there is small chance of flipping heads, given a long enough time horizon you are guaranteed to flip heads in the limit.
 - Next note that if $n \leq T$ then $|S_n| \leq b$ - the winning of the gambler bounded by b . This a random walk with bounded increments. Therefore by the theorem, $E(S_T) = E(S_0) = a$ but

$$S_T = \begin{cases} b & \text{if gambler reaches } b \text{ before } 0 \\ 0 & \text{else} \end{cases}$$
 The value of the winnings of the gambler at the stopping time is either b or 0.
 - So $E(S_T) = bP(\text{reaching } b \text{ before } 0) + 0P(\text{reaching } 0 \text{ before } b) = bP(\text{reaching } b \text{ before } 0)$. But we now know that $E(S_T) = a$ so $P(\text{reaching } b \text{ before } 0) = \frac{a}{b}$
 - For example if a gambler starts with 900 dollars and target is 1000, then $P(\text{reaching target}) = \frac{900}{1000} = 0.9$ so there is a 90% chance of reaching the target before going broke.
 - We will see this takes a large number of steps and the result is very sensitive to the probabilities of winning / losing.
 - (2) What is $E(T)$? Claim: Define $Z_n = (S_n - a)^2 - n$. This sequence $Z_0 = 0, Z_1, \dots$ is also a martingale.
 - Aside: Suppose you have a sequence Z_1, Z_2, \dots and another sequence X_1, \dots st each Z_i is a function of X_1, \dots, X_i and $E(Z_{n+1}|X_1, \dots, X_n) = Z_n$ for each n . Then also we say that Z_1, \dots is a martingale and all theorems apply.
 - Proof: Expected value given the past $E(Z_{n+1}|X_1, \dots, X)$? X_i are iid ± 1 representing the wins and losses. So

$$E(Z_{n+1}|X_1, \dots, X_n) = E((S_{n+1} - a)^2 - (n + 1)|X_1, \dots, X) = E((X_{n+1} + S_n - a)^2 | X_1, \dots, X) - (n + 1)$$
 . Expanding the squared terms and using expectation linearity

$$= E(X_{n+1}^2 | X_1, \dots, X) + 2E(X_{n+1}(S_n - a) | X_1, \dots, X) + E((S_n - a)^2 | X_1, \dots, X) - (n + 1).$$
 - Then $E(X_{n+1}^2 | X_1, \dots, X) = E(X_{n+1}^2) = 1$.
 - And given X_1, \dots, X_n , $(S_n - a)$ is a constant so $2E(X_{n+1}(S_n - a) | X_1, \dots, X) = 2(S_n - a)E(X_{n+1} | X_1, \dots, X) = 2(S_n - a)E(X_{n+1}) = 0$.
 - Finally by similar logic, $E((S_n - a)^2 | X_1, \dots, X) = (S_n - a)^2$. So overall $E(Z_{n+1}|X_1, \dots, X) = 1 + 0 + (S_n - a)^2 - (n + 1) = (S_n - a)^2 - n = Z_n$ so we have a

martingale.

- So if we let $T_n = \begin{cases} T & \text{if } T \leq n \\ n & \text{else} \end{cases}$ then $E(Z_{T_n}) = E(Z_0) = 0$. But $Z_{T_n} = (S_{T_n} - a)^2 - T_n$ by the definition of Z. So $E((S_{T_n} - a)^2) = E(T_n)$. Now $T_n \rightarrow T$ as $n \rightarrow \infty$ and T_n is an increasing sequence, so by monotone convergence $E(T_n) \rightarrow E(T)$.
- Next note that $S_{T_n} \rightarrow S_T$ as $n \rightarrow \infty$ and $(S_{T_n} - a)^2 \leq b^2$ so by bounded convergence, $E((S_{T_n} - a)^2) \rightarrow E(S_T - a)^2$. Thus $E(S_T - a)^2 = E(T)$.
- But we know that $P(S_T = b) = a/b$ and $P(S_T = 0) = 1 - a/b$ and thus $E(T) = E(S_T - a)^2 = (b - a)^2 P(S_T = b) + (0 - a)^2 P(S_T = 0)$ and simplifying $= (b - a)^2 a/b + a^2 \frac{b-a}{b} = a(b - a)$ So expected time to hit b or go broke $E(T) = a(b - a)$. If $a = 900$, $b = 1000$, we saw $P(\text{reaching b before 0}) = 0.9$, but $E(T) = 900(1000 - 900) = 90,000$. It takes a very long time despite favorable odds for winning b; many oscillating steps to reach 1000.
- (3) What if the chance of winning at each turn is not exactly $1/2$? Say it is $p \neq 1/2$ instead. In Roulette, $p=18/38$.
 - How to calculate $P(\text{reaching b before 0})$? In the case $p = 1/2$, recall that we used the fact that S_n is a martingale to prove that $E(S_T) = E(S_0) = a$ and then deduced the distribution of S_T from this (which is possible since S_T can only take two values 0, b. Otherwise cannot deduce a distribution from an expected value). When $p \neq 1/2$, S_n is not a martingale; here actually $S_n - (2p - 1)n$ is a martingale, but that does not help because this expectation equal to zero does not help us get to $E(S_T)$ because we do not know $E(T)$. Instead let $Z = \left(\frac{q}{p}\right)^{S_n}$, $q = 1 - p$. Claim: $\{Z_n\}_{n \geq 1}$ is a martingale
 - Proof: $E(Z_{n+1}|X_1, \dots, X_n) = E\left(\left(\frac{q}{p}\right)^{S_{n+1}}|X_1, \dots, X_n\right)$. Now $S_{n+1} = S_n + X_{n+1}$. So given X_1, \dots, X_n , $\left(\frac{q}{p}\right)^{S_n}$ is a constant. So $= \left(\frac{q}{p}\right)^{S_n} E\left(\left(\frac{q}{p}\right)^{X_{n+1}}|X_1, \dots, X_n\right)$. Then since $X_{n+1} \perp X_1, \dots, X_n$, say $= \left(\frac{q}{p}\right)^{S_n} E\left(\left(\frac{q}{p}\right)^{X_{n+1}}\right) = \left(\frac{q}{p}\right)^{S_n} \left(\left(\frac{q}{p}\right) P(X_{n+1} = 1) + \left(\frac{q}{p}\right)^{-1} P(X_{n+1} = -1)\right) = Z_n$.
 - As before, let $T_n = \min\{T, n\}$ where $T = \min\{n : S_n = b, 0\}$ Then T_n is a bounded stopping time and $E(Z_{T_n}) = E(Z_0) = (\frac{q}{p})^a$. We also know that $P(T < \infty) = 1$ (we proved this when $p = 1/2$, proof here is similar). Thus $Z_{T_n} \rightarrow Z_T$ as $n \rightarrow \infty$. Also since $T_n \leq T$, $0 \leq Z_{T_n} \leq \max\{(\frac{q}{p})^0, (\frac{q}{p})^1, \dots, (\frac{q}{p})^b\} = \max\{1, (\frac{q}{p})^b\} < \infty$ So by the bounded convergence theorem, $E(Z_{T_n}) \rightarrow E(Z_T)$ and $E(Z_T) = (\frac{q}{p})^a$. As before Z_T can only take two values, namely 1 and $(\frac{q}{p})^b$ (Since S_T can only take on 0 or b). Therefore $E(Z_T) = P(Z_T = 1)1 + P(Z_T = (\frac{q}{p})^b))(\frac{q}{p})^b = 1 - P(S_T = b) + P(S_T = b)(\frac{q}{p})^b$
 - Therefore $1 - P(S_T = b) + P(S_T = b)(\frac{q}{p})^b = (\frac{q}{p})^a \implies P(S_T = b) = \frac{(\frac{q}{p})^a - 1}{(\frac{q}{p})^b - 1}$ for $q \neq p$. For example if $a = 900$, $b = 1000$ and $p = 18/38$, then $P(S_T = b) \approx 0.000027$ - the gambler essentially has no chance of winning now that p is slightly off of $1/2$. In the beginning, the situation looks similar to $p=1/2$ with wins and losses, but later the fluctuations have a slow negative drift, and since it took so long to reach 1000 with $p=1/2$, the losses will slowly increase and 1000 will be further out of reach.
 - If the gambler is allowed to bet larger amounts than a dollar, there are more optimal strategies. If you have more than 500, bet the remaining amount of money to goal, if less than 500 bet everything you have. Has 88% chance of success with a fair coin. With these roulette odds, it still performs very well.

Martingale Varieties

- A sequence Z_1, \dots is called a **submartingale** if $Z_n \leq E(Z_{n+1} | Z_1, \dots, Z_n)$ (equality holds if martingale - martingales are submartingales but not conversely).
 - Example: if $\{Z_n\}$ is a martingale then $\{Z_n^2\}$ is a submartingale.
 - Proof: $E(Z_{n+1}^2 | Z_1, \dots, Z_n) = E((Z_{n+1} - Z_n)^2 + 2Z_n(Z_{n+1} - Z_n) + Z_n^2 | Z_1, \dots, Z_n) = E((Z_{n+1} - Z_n)^2 | Z_1, \dots, Z_n) + 2Z_n E(Z_{n+1} - Z_n | Z_1, \dots, Z_n) + Z_n^2 | Z_1, \dots, Z_n)$
 - Note that given up to Z_n the RV Z_n becomes a constant so $E(Z_n(Z_{n+1} - Z_n) | Z_1, \dots, Z_n) = Z_n E(Z_{n+1} | Z_1, \dots, Z_n) - Z_n^2 = Z_n^2 - Z_n^2 = 0$. Therefore $E(Z_{n+1}^2 | Z_1, \dots, Z_n) = E((Z_{n+1} - Z_n)^2 | Z_1, \dots, Z_n) + Z_n^2 \geq Z_n^2$.
- **Submartingale Stopping Theorem:** Let $\{X_n\}$ be a submartingale and T be a bounded stopping time (ie. there is a constant n st $T \leq n$ always). Then $E(X_n) \geq E(X_T) \geq E(X_1)$.
 - Proof: Exactly the same as for the MST, replacing equalities with inequalities.
- **Submartingale Maximal Inequality:** Let $\{X_n\}$ be a non-negative submartingale, then for any n and any $a > 0$, $P(\max\{X_1, \dots, X_n\} \geq a) \leq \frac{E(X_n)}{a}$
 - Proof: Fix n . Let $N = \begin{cases} \text{first } i \leq n \text{ st } X_i > a & \text{if such } i \text{ exists} \\ n & \text{otherwise} \end{cases}$ Can check N is a stopping time, moreover $N \leq n$, thus by the submartingale stopping theorem $E(X_N) \leq E(X_n)$. But $\max\{X_1, \dots, X_n\} > a \implies X_N > a$, since there is some i st $X_i > a$ (Note that $\max\{X_1, \dots, X_n\} > a$ then there is some $i \leq n$ st $X_i > a$ and so N , being the minimum such i , also satisfies $X_N > a$). Thus $P(\max\{X_1, \dots, X_n\} \geq a) \leq P(X_N > a) \leq \frac{E(X_N)}{a} \leq \frac{E(X_n)}{a}$. First inequality bounded by Markov inequality and second by SST.

Cauchy Martingales

- Recall Polya's Urn model: An urn contains one white ball and one black ball. Pick a ball at random. Replace it back to the urn together with an extra ball of the same color. Repeat. Let X_n be the proportion of white balls at time n (start at time 1). **Claim:** $\{X_n\}_{n \geq 1}$ is Cauchy WP1 and hence convergent.
 - Note that we are not claiming anything about the limit; in fact the limit is random.
 - Recall: a sequence of real numbers x_1, x_2, \dots is called Cauchy if $\forall \epsilon > 0, \exists n_0 \geq 1$ st $\forall m, n \geq n_0, |x_m - x_n| \leq \epsilon$. After a certain stage, all of the numbers in the sequence are within some epsilon of each other. Important fact: a sequence of real numbers is convergent iff it is Cauchy.
 - This follows from the martingale convergence theorem: Let $\{Z_n\}$ be a martingale with $\sup_{n \geq 1} E(Z_n) < \infty$
 - many martingales do not satisfy this, such as random walk. For example, the X_n for the Polya Urn model do satisfy this condition, since $0 \leq X_n \leq 1$ and also recall that X_n is a martingale. Then with probability 1, $\{Z_n\}$ is a Cauchy sequence and hence convergent.
 - Will work towards proof: (under slightly stronger assumption that $\sup_{n \geq 1} E(Z_n^2) < \infty$, note this also holds for Polya).
 - We will now show that if $\{Z_n\}$ is a Martingale such that $\sup_{n \geq 1} E(Z_n^2) < \infty$ then WP1 $\{Z_n\}$ is a Cauchy sequence.

- Step 1: $E(Z_n^2)$ is an increasing sequence and being bounded has a limit. Proof: Let's prove this with $Z_0 = 0$. We will show that for any $m \leq n$, $E(Z_n - Z_m)^2 = E(Z_n^2) - E(Z_m^2)$ (1). This will imply that $E(Z_n^2) - E(Z_1^2) = \sum_{i=2}^n [E(Z_i^2) - E(Z_{i-1}^2)] = \sum_{i=2}^n E[(Z_i - Z_{i-1})^2]$ (Telescoping sum). This clearly shows that $E(Z_n^2)$ is an increasing sequence.

- Let $\mu = \lim_{n \rightarrow \infty} E(Z_n^2)$. Proving eqn (1):

$$\begin{aligned} E(Z_n - Z_m)^2 &= E((Z_n - Z_{n-1} + Z_{n-1} - Z_m)^2) \\ &= E((Z_n - Z_{n-1})^2 + 2(Z_n - Z_{n-1})(Z_{n-1} - Z_m) + (Z_{n-1} - Z_m)^2) \end{aligned}$$

- Claim: expected value of the middle term is zero. Then

$$\begin{aligned} E(2(Z_n - Z_{n-1})(Z_{n-1} - Z_m) | Z_1, \dots, Z_{n-1}) &= 2(Z_{n-1} - Z_m)E(Z_n - Z_{n-1} | Z_1, \dots, Z_{n-1}) \\ &= 2(Z_{n-1} - Z_n)(E(Z_n | Z_1, \dots, Z_{n-1}) - Z_{n-1}) = 0 \end{aligned}$$

- Fact: $E(E(X|Y_1, \dots, Y_n)) = E(X)$. Suppose the RV are discrete, then

$$\begin{aligned} E(E(Z|Y_1, \dots, Y_n)) &= \sum_y E(X|Y_1 = y_1, \dots, Y_n = y_n)P(Y_1 = y_1, \dots, Y_n = y_n) \text{ Then} \\ &= \sum_y (\sum_x P(X = x | Y_1 = y_1, \dots, Y_n = y_n))P(Y_1 = y_1, \dots, Y_n = y_n), \\ &= \sum_y \sum_x X P(X = x, Y_1 = y_1, \dots, Y_n = y_n) \text{ and finally} \\ &= \sum_x X \sum_y P(X = x, Y_1 = y_1, \dots, Y_n = y_n) = \sum_x X P(X = x) = E(X) \end{aligned}$$

- Since middle term is 0, we can take expectation of both sides to get

$$E(2(Z_n - Z_{n-1})(Z_{n-1} - Z_m) | \dots) = 0. Thus E(Z_n^2) - E(Z_m^2) = E((Z_n - Z_{n-1})^2 + (Z_{n-1} - Z_m)^2).$$

- Continue this way to get $E(Z_n - Z_m)^2 = \sum_{i=m+1}^n E(Z_i - Z_{i-1})^2$ In particular

$$E(Z_n^2) = E(Z_n - Z_0)^2 = \sum_{i=1}^n E(Z_i - Z_{i-1})^2 \text{ and}$$

$$E(Z_m^2) = E(Z_m - Z_0)^2 = \sum_{i=1}^m E(Z_i - Z_{i-1})^2. Thus E(Z_n^2) - E(Z_m^2) = E(Z_n - Z_m)^2.$$

- Step 2: Take any $m, n \geq 1$. Note that the sequence M_0, \dots where $M_i = Z_{m+i} - Z_m$ is also a martingale wrt the sequence $((Z_0, \dots, Z_m), Z_{m+1}, Z_{m+2}, \dots)$. Why?

$$\begin{aligned} E(M_{i+1} | Z_0, \dots, Z_{m+i}) &= E(Z_{m+i+1} - Z_m | Z_0, \dots, Z_{m+i}) = Z_{m+i} - Z_m = M_i. So \{M_i^2\} \text{ is a} \\ &\text{submartingale, so by the maximal inequality, } P(\max_{0 \leq i \leq n} |Z_{m+i} - Z_m| > \epsilon) = P(\max_{0 \leq i \leq n} M_i^2 > \epsilon^2) \\ &\text{bounded by } \leq \frac{E(M_n^2)}{\epsilon^2} = \frac{E(Z_{m+n} - Z_n)^2}{\epsilon^2} = \frac{E(Z_{m+n})^2 - E(Z_n)^2}{\epsilon^2}. If A_n = \max_{0 \leq i \leq n} |Z_{m+i} - Z_m| > \epsilon \text{ then} \end{aligned}$$

$A_0 \subset A_1 \subset A_2 \dots$ - increasing events. Therefore $P(\cup_{n=0}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$, as a consequence of

the countable additivity of P. But $\cup_{n=0}^{\infty} A_n = \{|Z_{m+i} - Z_m| > \epsilon\}$ for some $i \geq 0$. Thus

$$P(|Z_{m+i} - Z_m| > \epsilon) \leq \lim_{n \rightarrow \infty} \frac{E(Z_{m+n})^2 - E(Z_n)^2}{\epsilon^2} = \frac{\mu - E(Z_m^2)}{\epsilon^2}.$$

- So if we consider the complementary event $B_n = \{|Z_{m+i} - Z_m| \leq \epsilon\}$ has probability

$$\geq 1 - \frac{\mu - E(Z_m^2)}{\epsilon^2}. Let B = \cup_{m=0}^{\infty} B_m \text{ then } P(B) \geq P(B_m) \text{ for all m}$$

$$\implies P(B) \geq \limsup_{m \rightarrow \infty} P(B_m) \geq \limsup_{m \rightarrow \infty} (1 - \frac{\mu - E(Z_m^2)}{\epsilon^2}). But E(Z_m^2) \rightarrow \mu \text{ as } \mu \rightarrow \infty \text{ so } P(B) \geq 1$$

but that means $P(B) = 1$. So $B = \{\exists m \ st |Z_{m+i} - Z_m| \leq \epsilon \forall i \geq 0\}$. But

$B \implies \{\exists m \ st |Z_k - Z_m| \leq \epsilon \text{ and } |Z_l - Z_m| \leq \epsilon\}$. Then

$|Z_k - Z_l| \leq |Z_k - Z_m| + |Z_l - Z_m| \leq \epsilon$. Thus $P(C_{\epsilon}) = 1$. Let

$C = \cap_{n=1}^{\infty} C_{1/n} = \{\forall n \geq 1, \exists m \geq 1 \ st |Z_k - Z_l| \leq 1/n \forall k, l \geq m\}$ Since $P(C_{1/n}) = 1$, $\forall n$ we get

$P(C) = 1$. $C \implies$ the sequence $\{Z_n\}$ is Cauchy and therefore convergent.

- General Case: Take any martingale Z_n st $\sup_{n \geq 1} E(Z_n^2) < \infty$. Let $a = E(Z_1)$. We proved that $E(Z_n) = E(Z_1)$.

Define $M_n = Z_n - a$, $n \geq 1$. Let $M_0 = 0$. Then $\{M_n\}$ is also a martingale because

$E(M_{n+1} | Z_1, \dots, Z_n) = E(Z_{n+1} - a | Z_1, \dots, Z_n) = Z_n - a = M_n$. We also have

$E(M_1 | M_0) = E(M_1) = E(Z_1) - a = 0 = M_0$. Note also that $E(M_n^2) = E(Z_n - a)^2$, then using

$(x+y)^2 \leq 2x^2 + 2y^2$ we get $\leq 2E(Z_n^2) + 2a^2 \implies \sup_{n \geq 1} E(M_n^2) < \infty$. Thus $\{M_n\}$ is Cauchy WP1 and

therefore so is $\{Z_n\}$.

- Example: If X_n is a fraction of white balls at time n in Polya Urn, then X_n converges WP1 as $n \rightarrow \infty$. The limit is random. Actually the limit is a uniform $[0, 1]$. How could we prove this? Proof by induction that for any n (process starts at time 1 with 1 white and 1 black), $X_n \sim \text{unif}(\{\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}\})$

Concentration Inequalities

- **Azuma - Hoeffding Inequality:** Let Z_0, Z_1, \dots be a martingale and let $\mu = E(Z_0)$. Suppose also that there is a non-random, fixed constant c st $|Z_i - Z_{i-1}| \leq c \forall i$. Then for any n and any $t \geq 0$ the

$$P(|Z_n - \mu| \geq t) \leq 2e^{-\frac{t^2}{2nc^2}}.$$

- A bounded difference martingale allows us to say something about the fluctuation of the nth martingale.
- We will prove that $P(Z_n - \mu \geq t) \leq e^{-\frac{t^2}{2nc^2}}$. Same argument applied $-Z_n$ will show that $P(Z_n - \mu \leq -t) \leq e^{-\frac{t^2}{2nc^2}}$. Combining we get $P(|Z_n - \mu| \geq t) \leq 2e^{-\frac{t^2}{2nc^2}} \leq P(Z_n - \mu \geq t) + P(Z_n - \mu \leq -t) \leq 2e^{-\frac{t^2}{2nc^2}}$.
- First step: show that for any $\theta > 0$, $E(e^{\theta(Z_n - \mu)}) \leq e^{nc^2\theta^2/2}$. Proof:
 $E(e^{\theta(Z_n - \mu)}) = E(E(e^{\theta(Z_n - \mu)} | Z_1, \dots, Z_{n-1}))$. Now
 $E(e^{\theta(Z_n - \mu)} | Z_1, \dots, Z_{n-1}) = E(e^{\theta(Z_n - Z_{n-1})} e^{\theta(Z_{n-1} - \mu)} | Z_1, \dots, Z_{n-1}) = e^{\theta(Z_{n-1} - \mu)} E(e^{\theta(Z_n - Z_{n-1})} | Z_1, \dots, Z_{n-1})$. Given Z_1, \dots, Z_{n-1} the conditional expectation of $Z_n - Z_{n-1}$ is 0. Also $|Z_n - Z_{n-1}| \leq c$. So by the following lemma, we get $E(e^{\theta(Z_n - Z_{n-1})} | Z_1, \dots, Z_{n-1}) \leq e^{\theta^2 c^2/2}$.
- **Lemma:** If X is a RV st $E(X) = 0$ and $|X| \leq c$ then $E(e^{\theta X}) \leq e^{\theta^2 c^2/2}$
 - Proof: The function $f(x) = e^{\theta X}$ is convex on \mathbb{R} meaning that $f''(x) \geq 0$ everywhere. This means that $f'(x)$ is an increasing function of x. So f is constantly bending upwards. Such functions have the property that for any x,y and for any $t \in [0, 1]$, $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$, ie a midpoint between x-axis values x and y, the value of the point on f(x) is bounded above by the line connecting f(y) and f(x) (sensible since curve sags below straight lines).
 - Take any $x \in [-c, c]$, write $x = t(-c) + (1-t)c$ where $t = \frac{c-x}{2c} \in [0, 1]$. So $e^{\theta x} = f(t(-c) + (1-t)c) \leq tf(-c) + (1-t)f(c) = \frac{c-x}{2c}e^{-c\theta} + \frac{c+x}{2c}e^{c\theta}$. This holds if we put X in place of x so $E(e^{\theta X}) \leq E(\frac{c-X}{2c}e^{-c\theta} + \frac{c+X}{2c}e^{c\theta})$. But $E(X) = 0$. So $E(e^{\theta X}) \leq \frac{1}{2}e^{-c\theta} + \frac{1}{2}e^{c\theta} = \cosh(c\theta)$. (Fact: $\cosh(x) \leq e^{x^2/2}$ using series expansion.)
 - Thus $E(e^{\theta(Z_n - Z_{n-1})} | Z_1, \dots, Z_{n-1}) \leq e^{\theta^2 c^2/2}$ which gives us $E(e^{\theta(Z_n - \mu)} | Z_1, \dots, Z_n) \leq e^{\theta(Z_{n-1} - \mu)} e^{\theta^2 c^2/2}$. Taking expectations, $E(e^{\theta(Z_n - \mu)}) \leq E(e^{\theta(Z_{n-1} - \mu)}) e^{\theta^2 c^2/2}$ and continue by backward induction.
 - Thus we end up proving $E(e^{\theta(Z_n - \mu)}) \leq e^{n\theta^2 c^2/2}$. For any $t > 0$, $P(Z_n - \mu \geq t) = P(e^{\theta(Z_n - \mu)} > e^{\theta t}) \leq e^{-\theta t} E(e^{\theta(Z_n - \mu)})$ by markov inequality. Then $\leq e^{-\theta t + n\theta^2 c^2/2}$. Given t find optimal θ , $\hat{\theta} = \frac{t}{nc^2}$. Pluggin this in we get $P(Z_n - \mu \geq t) \leq e^{-t^2/(2nc^2)}$

- **Bounded Difference Inequality:** Let X_1, \dots, X_n be independent RV (or even random vectors of varying dimensions). Let $y = f(X_1, \dots, X_n)$ be some function of X_1, \dots, X_n with the following property:
 $|f(x_1, \dots, x_n) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \leq c$ for any x_1, \dots, x_n any i, any x'_i . Then $\forall t > 0$, $P(|y - E(y)| \geq t) \leq 2e^{-t^2/2nc^2}$.

- If you change one coordinate, the value of the function can change by at most c, then this inequality applies
- Let $Z_i = E(Y | X_1, \dots, X_i)$ for $1 \leq i \leq n$, $Z_0 = E(Y)$. Then $\{Z_i\}$ is a martingale sequence since $E(Z_i | X_1, \dots, X_{i-1}) = Z_{i-1}$.
- Note $Z_n = E(Y | X_1, \dots, X_n) = Y$. Claim: $|Z_i - Z_{i-1}|$ is always bounded by c for any i. Here we use independence of X_1, \dots, X_n .

- Proof: Let $f_i(x_1, \dots, x_i) = E(Y|X_1, \dots, X_i)$. Since $Y = f(X_1, \dots, X_n)$ and X 's are independent, $E(Y|X_1 = x_1, \dots, X_i = x_i) = E(f(x_1, \dots, x_i, X_{i+1}, \dots, X_n))$ (Substituting in specific values does not affect the future RVs due to independence. Why?

$$E(Y|X_1 = x_1, \dots, X_i = x_i) = E(f(X_1, \dots, X_n)|X_1 = x_1, \dots, X_i = x_i)$$
 and expanding
 $= \sum_{x_{i+1}, \dots, X_n} f(x_1, \dots, x_n) P(X_{i+1} = x_{i+1}, \dots, X_n = x_n | X_1 = x_1, \dots, X_i = x_i)$ by independence.)
 $= \sum f(x_1, \dots, x_n) P(X_{i+1} = x_{i+1}, \dots, X_n = x_n)$
- Claim: for any X_1, \dots, X_{i+1} the $|f_i(X_1, \dots, X_i) - f_{i+1}(X_1, \dots, X_{i+1})| \leq c$. Proof:
 $|f_i(X_1, \dots, X_i) - f_{i+1}(X_1, \dots, X_{i+1})|$
 $= |E(f_i(x_1, \dots, x_i, X_{i+1}, \dots, X_n)) - E(f_i(x_1, \dots, x_i, x_{i+1}, X_{i+2}, \dots, X_n))|$
- Then $= |E(f_i(x_1, \dots, x_i, X_{i+1}, \dots, X_n)) - f_i(x_1, \dots, x_i, x_{i+1}, X_{i+2}, \dots, X_n))|$ which is
 $\leq E|(f_i(x_1, \dots, x_i, X_{i+1}, \dots, X_n)) - f_i(x_1, \dots, x_i, x_{i+1}, X_{i+2}, \dots, X_n))|$. The vectors $(x_1, \dots, x_i, X_{i+1}, \dots, X_n), (x_1, \dots, x_i, x_{i+1}, X_{i+2}, \dots, X_n)$ differ only in coordinate $i+1$. Thus the above quantity is bounded by c .
- Therefore by the A-H inequality, $P(|Z_n - E(Z_n)| \geq t) \leq 2e^{-t^2/2nc^2}$ but $Z_n = Y$.

Applications

- Bin packing: You have an infinite number of bins, each of capacity 1. You have m items with iid weights X_1, \dots, X_m , with each weight between 0 and 1. Let L be the minimum number of bins required to pack all the items. We do not know the distribution, just that each item can fit into an empty bin. Note: L is random. Find L .
 - Consider L as a function of the weights X_1, \dots, X_m . If you change exactly one weight, keeping all else constant, you may need at most 1 extra bin to pack all the weights. So if L' is the new value of L , then $L' \leq L + 1$. But by the same logic $L \leq L' + 1$; therefore $|L - L'| \leq 1$. Therefore here $c = 1$, and so by the bounded difference inequality, $P(|L - E(L)| \geq t) \leq 2e^{-t^2/2m}$ since m is the number of independent RVs. This becomes friendlier if we replace t by $x\sqrt{m}$ which gives
 $P(|L - E(L)| \geq x\sqrt{m}) \leq 2e^{-x^2/2}$. Thus L fluctuates around its EV by at most $O(\sqrt{m})$. Eg.
 $Var(L) = E((L - E(L))^2) = \int_0^\infty P(|L - E(L)|^2 \geq t) dt \leq 4m$.
 - We can see why these are called concentration inequalities - RV is concentrated around its EV. Note that $L \geq \sum X_i$ so if $E(X_i) \neq 0$, then $E(L) \geq mE(X_1)$ so the concentration inequality shows that $\frac{L}{E(L)} \xrightarrow{P} 1$ in probability as $m \rightarrow \infty$ (say for $L = L_m$).
 - In our example, as $m \rightarrow \infty$, $\forall \epsilon > 0$,

$$E(L) \geq mE(X_1) \implies P\left(\left|\frac{L}{mE(X_1)} - 1\right| > \epsilon\right)$$

 $= P(|L - E(L)| > \epsilon E(L)) \leq 2e^{\epsilon^2 E(L)^2} 2m \leq 2e^{\epsilon^2 m^2 E(X_1)^2} 2m = 2e^{\epsilon^2 mE(X_1)^2} 2 \rightarrow 0$
- Chromatic Number of a Random Graph: A simple graph consists of a set of vertices, some of which are connected to each other by edges. The chromatic number of such a graph is the minimum number of colors required to color the vertices such that no two vertices that are connected by an edge receive the same color.
 - Consider a random graph on n vertices, defined as follows: for any two vertices, put an edge between them with probability $1/2$ independently of each other. This is equivalent to picking a graph uniformly at random from the set of all graphs on these n vertices. Why? There are $2^{\binom{n}{2}}$ graphs since $\binom{n}{2}$ possible edges and each be present or absent (binary variable on or off). The random graph defined above has probability $= 2^{-\binom{n}{2}}$ of being any of these graphs (certain combination of switching every possible edge on or off, and they are present independently). Let G_n denote this random graph. This is actually an instance of the Erdos - Renyi $G(n, 1/2)$ random graph. Let χ_n be the chromatic number of G_n . What is the behavior of χ_n when n is large?
 - One can show that $E(\chi_n)$ behaves like $\frac{n}{2\log_2 n}$ when n is large. What is the magnitude of $|\chi_n - E(\chi_n)|$?

Let X_1 = set of edges from vertex 1 to vertices 2,...,n. X_2 = set of edges from vertex 2 to vertices 3,...,n, etc. Finally X_{n-1} = set of edges from vertex n-1 to vertex n. Then X_1, X_2, \dots, X_{n-1} are independent (not iid), since we are never counting the same edge twice. Now χ_n is a function of X_1, \dots, X_{n-1} . If one X_i is changed, then it changes the edges coming out of exactly one vertex. So $\chi'_n \leq \chi_n + 1$ - at most you have to change the color of this one vertex. Similarly, $\chi_n \leq \chi'_n + 1 \implies |\chi_n - \chi'_n| \leq 1$ so by the bounded difference inequality $P(|\chi_n - E(\chi_n)| \geq t) \leq 2e^{-t^2/2(n-1)}$. Again this shows that $\chi_n - E(\chi_n)$ is typically of order \sqrt{n} . In particular, $\frac{\chi_n}{E(\chi_n)} \rightarrow 1$ in probability as $n \rightarrow \infty$.

Brownian Motion

- Foundation of continuous time stochastic processes.
- Standard Brownian Motion is a random continuous function B from $[0, \infty) \rightarrow \mathbb{R}$ satisfying the following properties:
 - $B(0) = 0$
 - For any $0 \leq s \leq t$ $B(t) - B(s) \sim N(0, t - s)$
 - For any $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ (any n), the increments $B(t_1), B(t_2) - B(t_1), \dots, B(t_n) - B(t_{n-1})$ are independent
- The existence / uniqueness of the probability distribution on the function space will have these properties, those that share these properties are the same distribution
- Interpret $B(t)$ as the location of a randomly moving particle at time t when it starts from 0.
- Incidentally, if you have any random process in continuous time that moves continuously (not like PP that moves at discrete jumps) and has independent increments, then the increments must necessarily be normal RVs. This is a consequence of CLT. Thus the normality assumption is relatively weak.
- The variance of the normal is proportional to the length of the interval because we are dealing with homogeneous jumps - a linear process.
- **Theorem:** WP1 B has no point of differentiability
- **Theorem:** WP1 there is no interval in which B is increasing or decreasing.
- **Theorem:** The set of local maxima/minima of B is countable and dense in $[0, \infty)$.
- On a computer, cannot really simulate a continuous function, so discretize the real line and apply the 2nd and 3rd properties. Generate independent normal RVs for the increments and then connect them together
- Brownian motion is a scaling limit of random walk. More precisely, let X_1, X_2, \dots be iid RVs with mean 0 and variance 1. For each n, let $S_n = \sum_{i=1}^n X_i$, $S_0 = 0$. Take any n, define a random continuous function from $[0, \infty) \rightarrow \mathbb{R}$ as follows: if $t = \frac{k}{n}$ for some integer k, let $B_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^k X_i = \frac{S_k}{\sqrt{n}}$ (random walk scaled by root n). If t is between k/n and $(k+1)/n$, use linear interpolation. Then **Dansker's Theorem** says that the sequence of random continuous functions $\{B_n\}$ converges in distribution to standard brownian motion B as $n \rightarrow \infty$. This means that for any continuous $f : C[0, \infty) \rightarrow \mathbb{R}$, the sequence of RVs $f(B_n)$ converges in distribution to $f(B)$ as $n \rightarrow \infty$.
- Example: define a map $f : C[0, \infty] \rightarrow \mathbb{R}$ for $C[0, \infty] = \text{set of all continuous functions from } [0, \infty] \text{ to } \mathbb{R}$. $f(\phi) = \max_{0 \leq t \leq 1} \phi(t)$. Take any continuous function, take its max on the domain, that is a continuous map from $0, \text{infty}$ to \mathbb{R} . Can show that f is continuous. Now note that $f(B_n) = \frac{1}{\sqrt{n}} \max_{0 \leq t \leq 1} B(t)$. So Dansker tells us that $\frac{1}{\sqrt{n}} \max_{0 \leq t \leq 1} S_i$ converges in distribution to $\max_{0 \leq t \leq 1} B(t)$. Separately we can show that $\max_{0 \leq t \leq 1} B(t)$ has the same distribution as $|Z|$ where $Z \sim N(0, 1)$. Thus, for any sequence of iid RVs X_1, X_2, \dots mean 0, var 1; if we

define $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$ for all n , then as $n \rightarrow \infty$, $\frac{1}{\sqrt{n}} \max_{0 \leq t \leq 1} S_t \xrightarrow{d} |Z|$. This is a weird result, taking the maximum prevents the ability to use CLT, so this is the only way to prove this.

- **Markov Property of Brownian Motion:** Let B be standard Brownian motion. Take any $s \geq 0$. Let $W(t) = B(s+t) - B(s)$, $\forall t \geq 0$. Then W is also a standard BM and is independent of $(B_t)_{0 \leq t \leq s}$
 - Taking BM up to time s , looking at the future, subtracting off the history of time s , is independent of what happened up to time s .
 - Reason: for any $0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq s$ and any $0 \leq t_1 \leq \dots \leq t_m$ one can prove using the independent increments property that the random vectors $(B(s_1), \dots, B(s_n))$ and $(W(t_1), \dots, W(t_n))$ are independent. Increments are independent so we can obtain the B sequence and W using different sets of increments and this leads to the vector independence. Since this holds for all choices of $0 \leq s_1 \leq \dots \leq s_n \leq s$ and $0 \leq t_1 \leq \dots \leq t_m < \infty$ we can then use measure theory to deduce that the processes $(B(t))_{0 \leq t \leq s}$, $(W(t))_{t \geq 0}$ are independent.
- Other Properties:
 - **Scale Invariance:** For any $c > 0$ if we define $W(t) = c^{-1/2} B(ct)$ then W is again standard BM
 - If B is a standard BM, so is $-B$
 - **Time Inversion:** If B is standard BM and we define W as $W(0) = 0$ and $W(t) = tB(1/t)$ for $t > 0$, that is again Std BM.
 - Proof: Note $W(0) = 0$ by defn. Take any $0 < s \leq t$ then $W(t) - W(s) = tB(1/t) - sB(1/s)$. Then note that for any $u, v \in [0, \infty]$, $u \leq v$, $Cov(B(u), B(v)) = Cov(B(u), B(v) - B(u) + B(u)) = Cov(B(u), B(v) - B(u)) + Var(B(u))$. The first cov is zero since $B(u)$ and $B(v) - B(u)$ are independent by indep of increments. Also $Var(B(u)) = u$ since $B(u) = B(u) - B(0) \sim N(0, u)$. So if $u \leq v$, $Cov(B(u), B(v)) = u$.
 - Conversely if a process (X_t) is such that any $(X(t_1), \dots, X(t_n))$ is jointly normal and $Cov(X(s), X(t)) = s$ whenever $s \leq t$ one can show by calculating covariances that the process has the independence of increments property. As long as you have Gaussians, cov = 0 implies RV independence.
 - Thus $W(t) - W(s) \sim N(0, var = ?)$. So $Var(W(t) - W(s)) = Var(tB(1/t) - sB(1/s)) = t^2 Var(B(1/t)) + s^2 Var(B(1/s)) - 2tsCov(B(1/t), B(1/s)) = t^2(1/t) + s^2(1/s) - 2ts(1/t) = t - s$
 - The independent increments of W follow similarly by calculating covariances.
- We did not prove W is continuous - why? $W(T) = tB(1/t)$, $t > 0$, $W(0) = 0$. W is obviously continuous at every $t > 0$ - continuous transform of continuous function. What about continuity at 0? We have to show that $tB(1/t) \rightarrow 0$, $t \rightarrow 0$. This is the same as showing $\frac{B(s)}{s} \rightarrow 0$, $s \rightarrow \infty$ - this is called the **Law of Large Numbers for BM**.
- Note that $B(1), B(2) - B(1), \dots$ are iid $N(0, 1)$ RVs so $\frac{B(n)}{n} = \frac{1}{n} \sum_{i=1}^n (B(i) - B(i-1)) \xrightarrow{a.s.} 0$, $n \rightarrow \infty$ by LLN but more work is needed to show that $\frac{B(s)}{s} \rightarrow 0$, $s \rightarrow \infty$ since in this case we are not going through the integers.
- **Stopping Time:** A RV T taking value in $[0, \infty]$ is called a stopping time for BM if for any t , the occurrence or non-occurrence of the event $\{T \leq t\}$ is determined by the values of $(B(s); s \in [0, t])$. If you know the value of the BM up to time t , you can tell the occurrence of T .
 - Example: Let $T = \inf\{t : B(t) = a\}$ ($a > 0$ is given). Then T is a stopping time. Looking at the BM up to time t , you can determine if a has been crossed.
 - Let $T = \inf\{t : B(t) \notin (a, b)\}$ for $(a < 0 < b)$ then T is a stopping time.

- Let $T = \sup\{t \in [0, 1] : B(t) = 0\}$. Then T is not a stopping time - this requires looking at the future, though proving it is not a stopping time is difficult.
- Let $T = \sup\{t \geq 0 : B(t) = t\}$ then T is not a stopping time. There will be a last time that B(t) hits t, since B(t) goes to 0 by LLN. But it requires future knowledge
- Adaptive Process:** A stochastic process (X_t) is said to be adapted to a BM B if $\forall t, X(t)$ is a function of $(B(s))_{s \leq t}$ - ie it is BM up to the given time. An adapted process is called a martingale if $\forall s \leq t, E(X_t | (B(u)_{u \leq s})) = X(s)$ and $E|X_t| < \infty, \forall t$. This is just like the discrete case; looking at the BM up to time s, you just get the value of the process X at time s.
 - Example: B itself is a martingale adapted to B
 - Proof: $s \leq t$,

$$E(B(t) | (B(u))_{u \leq s}) = E(B(t) - B(s) + B(s) | (B(u))_{u \leq s}) = E(B(t) - B(s) | (B(u))_{u \leq s}) + B(s)$$
 but by the Markov property $B(t) - B(s) \perp (B(u))_{u \leq s}$ with independent increments. So

$$E(B(t) - B(s) | (B(u))_{u \leq s}) = E(B(t) - B(s)) = 0$$
 Since $B(t) - B(s) \sim N(0, t - s)$. Thus

$$E(B(t) | (B(u))_{u \leq s}) = B(s)$$
 - Example: Let $X(t) = B(t)^2 - t$ then $(X(t))_{t \geq 0}$ is a martingale adapted to B.
 - Proof: Take $s \leq t, E(B(t)^2 - t | (B(u))_{u \leq s}) = E(B(t) - B(s) + B(s)^2 | (B(u))_{u \leq s}) - t$. Then

$$E((B(t) - B(s))^2 + 2B(s)(B(t) - B(s)) + B(s)^2 | (B(u))_{u \leq s}) - t$$
 Evaluating separately,

$$B(t) - B(s) \perp (B(u))_{u \leq s}, \sim N(0, t - s)$$
. So

$$E((B(t) - B(s))^2 | (B(u))_{u \leq s}) = E((B(t) - B(s))^2) = t - s$$
 Next given $(B(u))_{u \leq s}$, B(s) is just a constant. So $E(2B(s)(B(t) - B(s)) | (B(u))_{u \leq s}) = 2B(s)E((B(t) - B(s)) | (B(u))_{u \leq s}) = 0$.
 Finally, $E(B(s)^2 | (B(u))_{u \leq s}) = B(s)^2$ (since just a constant). So

$$E(B(t)^2 - t | (B(u))_{u \leq s}) = t - s + 0 + B(s)^2$$
 and so

$$E(X(t) | (B(u))_{u \leq s}) = t - s + B(s)^2 - t = B(s)^2 - s = X(s) \implies \text{martingale.}$$
 - In fact you can find a polynomial of B(t) of any degree that is a martingale.
 - Example; Exponential family of Martingales: For any $\lambda \in \mathbb{R}$ the process $X(t) = e^{\lambda B(t) - \frac{\lambda^2}{2}t}$ is a martingale adapted to B.

MST for Continuous Time

- MST for Continuous Time:** Let $\{X_t\}$ be a continuous martingale wrt Brownian motion $(B(t))$ (that is $\forall s \leq t, E(X_t) < \infty, E(X_t | (B(u))_{u \leq s}) = X_s$). For any bounded stoppping time T, then $E(X_T) = E(X_0)$.
- Example: Take any 2 numbers, $a < 0 < b$. Let $T = \inf\{t : B(t) = a \text{ or } b\}$ - first time BM hits a or b starting from 0. Then B(T) is either a or b; what are the probabilities of B(T) being a and b?
 - Using martingales, take any t. Let $T \wedge t = \min\{T, t\}$. Then $T \wedge t$ is again a stopping time. Why? To show: the event $\{T \wedge t \leq s\}$ is determined by $(B(u))_{u \leq s}$ for any s. If $s \geq t$, then $T \wedge t \leq t \leq s$ and so $\{T \wedge t \leq s\}$ is always true. If $s < t$ then $T \wedge t \leq s \iff T \leq s$. But since T is a stopping time, then event $\{T \leq s\}$ is determined by $(B(u))_{u \leq s}$. So in either case, the occurrence or non-occurrence of $\{T \wedge t \leq s\}$ is determined by $(B(u))_{u \leq s}$. Thus $T \wedge t$ is a stopping time.
 - So by MST, $E(B(T \wedge t)) = E(B(0)) = 0$. We will show later that $P(T < \infty) = 1$ (T finite WP1). If this is true, then $\lim_{t \rightarrow \infty} B(T \wedge t) = B(T)$. Why? Think of RVs as functions from the sample space Ω into the appropriate space.
 - Eg. $T : \Omega \rightarrow \mathbb{R}, B : \Omega \rightarrow C[0, \infty]$ (the space of contiunous functions from $[0, \infty] \rightarrow \mathbb{R}$). So $P(T < \infty) = 1$ means $P(\{\omega : T(\omega) < \infty\}) = 1$. Take any ω st $T(\omega) < \infty$. Let $X_t = B(T \wedge t), X = B(T)$, This means $X_T(\omega) = B(\omega)(\min\{T(\omega), t\})$. Since $T(\omega) < \infty, \min\{T(\omega), t\} = T(\omega)$ for all large t ($t > T(\omega)$). So for this ω ,

$\lim_{t \rightarrow \infty} X_t(\omega) = B(\omega)(T(\omega)) = X(\omega)$. Thus $P(\lim_{t \rightarrow \infty} X_t = X) = 1$.

- So we have (a) $E(B(T \wedge t)) = 0$, $\forall t$ (b) $B(T \wedge t) \rightarrow B(T)$, $t \rightarrow \infty$ (c) since $T \wedge t \leq T$, $B(T \wedge t) \in [a, b]$ for any t .
- So by the bounded convergence theorem, $E(B(T)) = \lim_{t \rightarrow \infty} E(B(T \wedge t)) = 0$. But $E(B(T)) = aP(B(T) = a) + bP(B(T) = b) = aP(B(T) = a) + b(1 - P(B(T) = a))$. Thus we get $P(B(T) = a) = \frac{b}{b-a}$, $P(B(T) = b) = \frac{-a}{b-a}$. So this is just like gambler's ruin but for Brownian motion.
- Next: Find $E(T)$ using the martingale $B(t)^2 - T$
 - Again, by the MST $E(B(T \wedge t)^2 - T \wedge t) = E(B(0)^2 - 0) = 0$ for any t . Thus $E(T \wedge t) = E(B(T \wedge t)^2)$. As we observed above $\{B(T \wedge t)\}_{t \geq 0}$ is a bounded sequence of RVs converging to $B(T)$, $t \rightarrow \infty$. Thus again by the bounded convergence theorem $E(B(T)^2) = \lim_{t \rightarrow \infty} E(B(T \wedge t)^2)$.
 - Also $\{T \wedge t\}$ is a sequence of non-negative RVs increasing to T . As long as $t < T$, this is equal to t . Once we hit T , its value remains at T and is just flat. Therefore by the monotone convergence theorem, $E(T) = \lim_{t \rightarrow \infty} E(T \wedge t)$. Combining the above, we get $E(T) = E(B(T)^2) = a^2 P(B(T) = a) + b^2 P(B(T) = b) = \frac{a^2 b - b^2 a}{b-a} = -ab$. So $E(T) = -ab$.

Brownian Motion with Drift

- **Brownian Motion with Drift:** Let $(B(t))$ be standard BM. Take any $\mu \in \mathbb{R}$ then the process $B(t) + \mu t$ is called BM with drift μ .
 - Fact: if B is standard BM, then $\limsup_{t \rightarrow \infty} B(t) = \infty$, $\liminf_{t \rightarrow \infty} B(t) = -\infty$.
 - Note that $B(t) + \mu t = t(\frac{B(t)}{t} + \mu)$. We know that $\frac{B(t)}{t} \rightarrow 0$, $t \rightarrow \infty$. This shows that $B(t) + \mu t \rightarrow \begin{cases} \infty & \text{if } \mu > 0 \\ -\infty & \text{if } \mu < 0 \end{cases}$ (oscillates for $\mu = 0$).
- Example: Suppose $Z_t = B(t) - \mu t$ is BM with negative drift $-\mu$. Let $M = \max Z_t$. What is the distribution of M ?
 - We will use the MST to compute the distribution of M . We know that $e^{\lambda B(t) - \frac{1}{2}\lambda^2 t}$ is a martingale for any $\lambda \in \mathbb{R}$. Take $\lambda = 2\mu$. Let $X_t = e^{\lambda B(t) - \frac{1}{2}\lambda^2 t}$. Take any $x > 0$, let $T = \inf\{t : B(t) = \mu t + x\}$ - first time BM crosses the line $y = \mu t + x$ with slope μ and intercept x . T might be infinite; ie. we never hit the line. In fact $\{T = \infty\}$ is the same as saying $M < x$. So $P(M \geq x) = P(T < \infty)$.
 - Note $M \geq 0$ since $M = \sup(B(t) - \mu t) \geq B(0) - \mu 0 = 0$. Also $M < \infty$ WP1 since $M = \infty \implies B(t_n) - \mu t_n \rightarrow \infty$ along a sequence t_n . This sequence cannot be bounded since B is continuous and hence bounded on closed intervals. So \exists sequence $t_n \rightarrow \infty$ st $B(t_n) - \mu t_n \rightarrow \infty$ but this is impossible, since $B(t_n) - \mu t_n = t_n(\frac{B(t_n)}{t_n} - \mu) \rightarrow -\infty$. So M is finite.
 - Recall we have the martingale $X_t = e^{\lambda B(t) - \frac{1}{2}\lambda^2 t}$, $\lambda = 2\mu$. By MST $E(X_{T \wedge t}) = E(X_0) = 1$. What is the $\lim_{t \rightarrow \infty} X_{T \wedge t}$? If $T < \infty$ then arguing as before, we get $\lim_{t \rightarrow \infty} X_{T \wedge t} = X_T = e^{\lambda B(T) - \frac{1}{2}\lambda^2 T}$. But $T = \inf\{t : B(t) = \mu t + x\}$, therefore $B(T) = \mu T + x$, so $X_T = e^{\lambda(\mu T + x) - \frac{1}{2}\lambda^2 T}$.
 - But $\lambda = 2\mu \implies \lambda\mu - (1/2)\lambda^2 = 0$. So $X_T = e^{\lambda x}$ if $T < \infty$. Therefore if $T < \infty$, then $\lim_{t \rightarrow \infty} X_{T \wedge t} = e^{2\mu x}$. OTOH, if $T = \infty$ then $T \wedge t = t \forall t$. Thus $X_{T \wedge t} = X_t = e^{\lambda B(t) - \frac{1}{2}\lambda^2 t} \forall t$. But $\lambda B(t) - \frac{1}{2}\lambda^2 t = \lambda t[\frac{B(t)}{t} - \frac{1}{2}\lambda] \rightarrow -\infty$, $t \rightarrow \infty$. Thus $X_{T \wedge t} \rightarrow 0$, $t \rightarrow \infty$ if $T = \infty$. Therefore $\lim_{t \rightarrow \infty} X_{T \wedge t} = e^{2\mu x} \mathbb{I}\{T < \infty\}$.
 - Finally, observe that for any t , $0 \leq X_{T \wedge t} = e^{\lambda B(T \wedge t) - \frac{1}{2}\lambda^2 (T \wedge t)}$. But $T \wedge t \leq T \implies B(T \wedge t) \leq \mu(T \wedge t) + x$. This in turn implies $X_{T \wedge t} \leq e^{\lambda(\mu(T \wedge t) + x) - \frac{1}{2}\lambda^2 T \wedge t} = e^{\lambda x}$. So

we can apply the bounded convergence theorem and get

$$\lim_{t \rightarrow \infty} E(X_{T \wedge t}) = E(\lim_{t \rightarrow \infty} X_{T \wedge t}) = E(e^{2\mu x} \mathbb{I}\{T < \infty\}) = e^{2\mu x} P(T < \infty). \text{ So } P(T < \infty) = e^{-2\mu x}. \text{ But}$$

$$P(T < \infty) = P(M \geq x) \text{ so } P(M \geq x) = e^{-2\mu x} \text{ and } M \sim \exp(2\mu).$$

Reflections and Maxima

- Let $M(t) = \max_{0 \leq s \leq t} B(s)$. What is the distribution of $M(t)$?
- **Strong Markov Property of BM:** Let T be a stopping time. Define $W(t) = B(T+t) - B(T)$, $t \geq 0$. Just like the markov property but starting at some stopping time. Then W is a standard BM and is independent of $(B(t))_{t \leq T}$.
 - Proof sketch: If T can take only discretely many values t_1, t_2, t_3, \dots then conditional on $T = t_i$ and $(B(s))_{s \leq t_i}$ we can show that W is a StdBM. The result follows from this in the discrete case, then one can pass to the continuous case by taking a limit.
- **Reflection Principle for BM:** Take $a > 0$. Let $T = \inf\{t : B(t) = a\}$ - first time BM hits a. Let $W(t) = B(T+t) - B(T)$, $t \geq 0$. Then W is Std BM, independent of $B^* = (B(t))_{t \leq T}$ by the strong markov property. Note that B can be produced as a function of B^* , T and W as follows:

$$B(t) = \begin{cases} B(t) & \text{if } t \leq T \\ B(T) + W(t-T) & \text{if } t > T \end{cases}$$
 Note that since T is a function of B^* , we can reconstruct B using B^* and W . But $-W$ is also a standard BM and also independent of B^* ; if we apply the same function to B^* and $-W$, the new process obtained in this manner will have the same distribution as B . (Using idea that if X, Y are indep RVs and X, Y' are also indep RVs and $Y' \stackrel{d}{=} Y$ then $f(X, Y') \stackrel{d}{=} f(X, Y)$).
 - The new process is $\tilde{B}(t) = \begin{cases} B(t) & \text{if } t \leq T \\ B(T) + (-W(t-T)) & \text{if } t > T \end{cases}$. Then $B(T) + (-W(t-T)) = B(T) + (-(B(t) - B(T))) = 2B(T) - B(t)$ then \tilde{B} is also a std BM.
 - Our conclusion is we have a new path after T (B) and its reflection (\tilde{B}) about a line $y=a$ and they are both standard BM.
- **Distribution of the Maximum:** Let $M(t) = \max_{0 \leq s \leq t} B(s)$. What is the distribution of $M(t)$?
 - Obviously $0 \leq M(t) < \infty$, since $B(0) = 0$ and B is continuous.
 - $P(M(t) \geq a)$? Let $T = \inf\{t \geq 0 : B(t) = a\}$ - first hitting time of level a. Let B' be B reflected across line $y = a$ after time T . $B'(t) = \begin{cases} B(t) & \text{if } t \leq T \\ 2B(T) - B(t) & \text{if } t > T \end{cases}$. We know that B' is also std BM.
 - Claim: $\{M(t) \geq a, B(t) \leq a\} = \{B'(t) \geq a\}$. Proof: Suppose $M(t) \geq a$ and $B(t) \leq a$. Then the process B hits level a within time t , since the maximum is a or more. That is $T \leq t$. So $B'(t) = 2B(T) - B(t)$ because $t > T$ - the reflection happened before time T . So $= 2a - B(t)$ but $B(t) \leq a$ so $B'(t) = 2a - B(t) \geq 2a - a = a$ Conversely, suppose that $B'(t) \geq a$ then consider 2 possibilities
 - (1) $T > t$. This means B does not hit level a by time t , which implies that $B(t) = B'(t)$ and $B(t) < a$; if you haven't hit level a by time t , then the reflective process is the same as the original. This contradicts the assumption that $B'(t) \geq a$, therefore this case is impossible.
 - (2) $T \leq t$. Then B hits level a by time t which means $M(t) \geq a$ and also $B'(t) = 2B(T) - B(t) = 2a - B(t)$. Thus $B(t) = 2a - B'(t) \leq 2a - a = a$.
 - We have shown both sides of the claim, therefore the claim $\{M(t) \geq a, B(t) \leq a\} = \{B'(t) \geq a\}$ has been shown.
 - Claim: $\{M(t) \geq a, B(t) > a\}$ is the same as $\{B(t) > a\}$. Proof: Suppose $M(t) \geq a, B(t) > a$ then obviously $B(t) > a$. Conversely, if $B(t) > a$ then $M(t) \geq B(t) \geq a$.
 - $\{M(t) \geq a\}$ is the union of the disjoint events $\{M(t) \geq a, B(t) > a\}$ and $\{M(t) \geq a, B(t) \leq a\}$. So

$P(M(t) \geq a) = P(M(t) \geq a, B(t) > a) + P(M(t) \geq a, B(t) \leq a)$ But from the two claims
 $= P(B(t) > a) = P(B'(t) \geq a)$ But B, B' are both BM and so B(t), B'(t) are both $N(0, t)$ RV. Therefore
 $P(M(t) \geq a) = 2P(Z \geq a) = P(|Z| \geq a)$ where $Z \sim N(0, t)$.

- Thus $M(t) \sim |N(0, t)|$
- By Donsker's theorem, one can show that this implies that for any iid RVs X_1, X_2, \dots with mean 0 and variance 1, if set $S_n = \sum_{i=1}^n X_i, S_0 = 0$ then $\frac{1}{\sqrt{n}} \max_{0 \leq i \leq n} S_i \xrightarrow{d} |N(0, 1)|$
- Example: One sided boundary crossing. Take any $a > 0$. Let $T = \inf\{t : B(t) = a\}$. What is the distribution of T?
 - $P(T > t) = P(M(t) < a)$ If you haven't hit a by time t, then the max values must be less than a.
 $= P(|Z| < a), Z \sim N(0, t) = \int_0^a \frac{1}{\sqrt{2\pi t}} 2e^{-x^2/2t} dx$. We have to differentiate wrt t and make negative (since we are looking at 1 - CDF) to get the density. $-\frac{d}{dt} P(T > t) = -\frac{d}{dt} \int_0^a \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} dx$. In the integral, sub $y = x/\sqrt{t}$, then $dy = \frac{1}{\sqrt{t}} dx$. The integral becomes $\int_0^{a/\sqrt{t}} \sqrt{\frac{2}{\pi t}} e^{-y^2/2} \sqrt{t} dy$. If $\phi(b) = \int_0^b \sqrt{\frac{2}{\pi}} e^{-y^2/2} dy$ then $\phi'(b) = \sqrt{\frac{2}{\pi}} e^{-b^2/2}$ But our integral is $\phi(a/\sqrt{t})$ so its derivative wrt t is $\phi'(a/\sqrt{t}) \frac{d}{dt} (a/\sqrt{t}) = \sqrt{\frac{2}{\pi}} e^{-a^2/2t} (-a/2t^{3/2})$. Thus the pdf of T at the point t is $\frac{a}{\sqrt{2\pi t^{3/2}}} e^{-a^2/2t}$
 - Note that this behaves like a constant multiple of $t^{-3/2}$ for large t. In particular, $E(T) = \infty$ and even $E\sqrt{T} = \infty$ - it is a very heavy tailed distribution.

Gaussian Processes

- A stochastic process $(X_t)_{t \geq 0}$ is called Gaussian if for all $t_1, \dots, t_n \geq 0$ and any n, $(X_{t_1}, \dots, X_{t_n})$ is jointly normal. Brownian motion is a Gaussian process, but there are many others. Since $B(t_1), \dots, B(t_n)$ are linear combinations of the increments of B which are independent normal RVs.
- We know that means and covariances uniquely determine the joint distribution of jointly normal RVs. Similarly, if $(X_t)_{t \geq 0}$ is a Gaussian process and we know $E(X_t) \forall t$ and $Cov(X_t, X_s) \forall s, t$ then that uniquely determines the distribution of the process. That is, these numbers uniquely determine the probability of any event related to the process and the expected value of any RV related to the process. Any other process is Gaussian with same mean and same covariances, then it is the same process.
- For BM, $E(B(t)) = 0$ for all t and $Cov(B(t), B(s)) = \min\{t, s\} \forall t, s$
- **Brownian Bridge:** another important Gaussian process. A BB starts at 0 and ends at 0 at time 1.
- Briefly, BB $(X_t)_{0 \leq t \leq 1}$ is just $(B(t))_{0 \leq t \leq 1}$ conditioned to have $B(1) = 0$. That is, it's the limit of conditional distribution given $|B(1)| < \epsilon, \epsilon \rightarrow 0$, since $B(1) = 0$ alone is a 0 probability event.
- If X_1, X_2 are jointly normal with $E(X_1) = \mu_1, E(X_2) = \mu_2, Var(X_1) = \sigma_1^2, Var(X_2) = \sigma_2^2, Cor(X_1, X_2) = \rho$ then the conditional distribution of X_1 given $X_2 = x_2$ is $N(\mu_1 + \frac{\sigma_1}{\sigma_2} \rho(\mu_2 - \mu_1), \sigma_1^2(1 - \rho^2))$.
- Let $(X_t)_{0 \leq t \leq 1}$ be the process $(B(t))$ conditioned on $B(1) = 0$. Since conditional distributions of multivariate normal RVs are again normal, we see that for any $t_1, \dots, t_n \in [0, 1]$ $(X_{t_1}, \dots, X_{t_n})$ must be jointly normal so (X_t) is a Gaussian process. In fact the joint distribution $(X_{t_1}, \dots, X_{t_n})$ is just the distribution of $(B(t_1), \dots, B(t_n))$ given $B(1) = 0$.
- So we only have to calculate $E(X_t) = E(B(t)|B(1) = 0)$ and $Cov(X_t, X_s)$. Starting with expectation, note that $(B(t), B(1))$ is jointly normal with $E(B(t)) = E(B(1)) = 0$ and $Var(B(t)) = t, Var(B(1)) = 1, Cov(B(t), B(1)) = \frac{Cov(B(t), B(1))}{\sqrt{Var(B(t))Var(B(1))}} = \sqrt{t}$. So plugging into our result above and with zero means $E(B(t)|B(1) = 0) = \mu_1 + \rho \frac{\sigma_1}{\sigma_2} (\mu_2 - \mu_1) = 0$
- The $Cov(X_t, X_s) = E(X_t X_s) - E(X_t)E(X_s) = E(X_t X_s)$ since $E(X_t) = E(X_s) = 0$ And $E(X_t X_s) = E(B(t)B(s)|B(1) = 0)$. Say $0 \leq s \leq t \leq 1$, we will compute $f(x) = E(B(s)|B(t) = x, B(1) = 0)$ then use the tower property of conditional expectation

- $E(B(s)B(t)|B(1)=0) = \int_{-\infty}^{\infty} E(B(s)B(t)|B(t)=x, B(1)=0)g(x)dx$ where g is the conditional pdf of $B(t)|B(1)=0$ (just using law of total probability). Then
 $= \int_{-\infty}^{\infty} xE(B(s)|B(t)=x, B(1)=0)g(x)dx = \int_{-\infty}^{\infty} xf(x)g(x)dx$ for defining
 $f(x) = E(B(s)|B(t)=x, B(1)=0)$. Then finally $= E(B(t)f(B(t))|B(1)=0)$.
- $f(x) = E(B(s)|B(t)=x, B(1)=0) = E(B(s)|B(t)=x, B(1)-B(t)=-x)$. However, $B(1)-B(t)$ is independent of $(B(u))_{u\leq t}$ by the Markov property of BM. So $= E(B(s)|B(t)=x)$ Applying our formula, $B(s), B(t)$ is jointly normal with $\mu_1 = E(B(s)) = 0$, $\mu_2 = E(B(t)) = 0$ and $\sigma_1^2 = Var(B(s)) = s$, $\sigma_2^2 = Var(B(t)) = t$ and $\rho = Cor(B(s), B(t)) = \frac{Cov(B(t), B(s))}{\sqrt{Var(B(t))Var(B(s))}} = \sqrt{\frac{s}{t}}$. So
 $f(x) = E(B(s)|B(t)=x) = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x - \mu_1) = \sqrt{\frac{s}{t}}\sqrt{\frac{s}{t}}x = \frac{s}{t}x$ - we just end up scaling x down.
 - Therefore, $Cov(X_s, X_t) = E(B(t)f(B(t))|B(1)=0) = \frac{s}{t}E(B(t)^2|B(1)=0)$. We have seen from the previous computation that $E(B(t)|B(1)=0) = 0$ so $E(B(t)^2|B(1)=0) = Var(B(t)|B(1)=0)$. But we also know that the conditional variance of $B(t)$ given $B(1)=0$ is $\sigma_1^2(1-\rho^2)$ where $\sigma_1^2 = Var(B(t)) = t$ and $\rho = Cor(B(t), B(1)) = \sqrt{t}$. So $Var(B(t)|B(1)=0) = t(1-t)$ Thus $Cov(X_t, X_s) = \frac{s}{t}t(1-t) = s(1-t)$ for $0 \leq s \leq t \leq 1$ (take the smaller value times 1 minus the larger value).
 - Easier description of BB:** Let B be std BM for $0 \leq t \leq 1$, let $Z_t = B(t) - tB(1)$. Then by linearity $(Z_t)_{0 \leq t \leq 1}$ is a Gaussian process. $E(Z_t) = E(B(t)) - tE(B(1)) = 0$ for all t. And for $0 \leq s \leq t \leq 1$, $Cov(Z_s, Z_t) = Cov(B(s) - sB(1), B(t) - tB(1))$ and by linearity of covariance $= Cov(B(s), B(t)) - sCov(B(1), B(t)) - tCov(B(s), B(1)) + stVar(B(1))$. We know these values $= s - st - ts + st = s(1-t)$, exactly what we got before. Thus this process is a BB.
 - BB is just the difference between BM and a line connected $B(0)=0$ and $B(1)$, not necessarily 0.

Stochastic Calculus

Ito integral

- Try to give meaning to integrals like $\int_0^t X(s)dB(s)$ where B is a standard BM and $\{X(s)\}_{s \geq 0}$ is a stochastic process st for each s, $X(s)$ is a function of $(B(t))_{t \leq s}$. Note: the function may depend on s.
- Examples:
 - $X(t) = f(B(t))$ for some $f : \mathbb{R} \rightarrow \mathbb{R}$.
 - $X(t) = f(t, B(t))$ for some $f : \mathbb{R}^2 \rightarrow \mathbb{R}$.
 - $X(t) = f(B(t)) - \int_0^t g(B(s))ds$
- Such processes are called adapted to B.
- Ito integrals look like R-S integrals wrt B. However, if we write down $\sum_{i=0}^{n-1} X(t_i)(B(t_{i+1}) - B(t_i))$ where $0 = t_0 < t_1 < \dots < t_n = t$ and take finer and finer partitions such that $\max(t_{i+1} - t_i) \rightarrow 0$ these sums do not converge. We cannot use the FTC to calculate.
- However, they do converge in L^2 . Meaning: we say that a sequence of RV $\{X_n\}_{n \geq 1}$ converges to a limit RV X in L^2 if $\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0$.
 - Note $X_n \xrightarrow{L^2} X \implies X_n \xrightarrow{P} X$.
- We see values around a limit, but occasional big variations in values.
- The convergence of the approximating sums in L^2 can be proved for a very general class of adapted processes. For simplicity, we will prove it only when X is uniformly bounded by a constant and is continuous. We will then assume that it is true for all X.
 - Proof: Assume that X is a continuous adapted process and there exists some constant M such that $|X(t)| \leq M \forall t$. Take a partition $0 = t_0 < t_1 < \dots < t_n = t$ of $[0, t]$ and take a refinement

$0 = s_0 < s_1 < \dots < s_m = t$ of this partition, meaning that each t_i is s_j for some j . Let $Y_j = X(s_j)$ and let $Z_j = X(t_i)$ where i is the unique index such that $t_i \leq s_j < t_{i+1}$. Note that Y_j, Z_j are both functions of $(B(s))_{s \leq s_j}$. Note also that $I_1 = \sum_{j=0}^{m-1} X(s_j)(B(s_{j+1}) - B(s_j)) = \sum_{j=0}^{m-1} Y_j(B(s_{j+1}) - B(s_j))$ and $I_2 = \sum_{j=0}^{m-1} Z_j(B(s_{j+1}) - B(s_j)) = \sum_{j=0}^{m-1} X(t_i)(B(t_{i+1}) - B(t_i))$ since for all i , $Z_j = X(t_i)$ for every $t_i \leq s_j < t_{i+1}$. We want to show I_1, I_2 are close in norm, meaning the refinement is close to the original in size.

- Let $\Delta_j = B(s_{j+1}) - B(s_j)$.
 $E(I_1 - I_2)^2 = E((\sum_{j=0}^{m-1} (y_j - Z_j)\Delta_j)^2) = \sum_{j,k} E((y_j - Z_j)\Delta_j(Y_k - Z_k)\Delta_k)$. Suppose $j < k$, then $Y_j, Z_j, \Delta_j, Y_k, Z_k$ are all function of $B(s)$. OTOH, $\Delta_k = B(s_{k+1}) - B(s_k)$ is independent of $B(s)$ by the markov property and $E(\Delta_k) = 0$. Thus
 $E((y_j - Z_j)\Delta_j(Y_k - Z_k)\Delta_k) = E((y_j - Z_j)\Delta_j(Y_k - Z_k))E(\Delta_k) = 0$. Similarly this vanishes when $j > k$. Thus $E(I_1 - I_2)^2 = \sum_{j=0}^{m-1} E((Y_j - Z_j)^2\Delta_j^2)$. But again, Y_j, Z_j are functions of $B(s)_{s \leq s_j}$ and $\Delta_j = B(s_{j+1}) - B(s_j)$ so $(Y_j - Z_j)^2, \Delta_j$ are independent. Moreover, $E(\Delta_j^2) = s_{j+1} - s_j$ since $\Delta_j \sim N(0, s_{j+1} - s_j)$.
- Thus
 $E(I_1 - I_2)^2 = \sum_{j=0}^{m-1} E((Y_j - Z_j)^2\Delta_j^2) = \sum_{j=0}^{m-1} E((Y_j - Z_j)^2)E(\Delta_j^2) = \sum_{j=0}^{m-1} E((Y_j - Z_j)^2)(s_{j+1} - s_j)$. Then this becomes $E[\sum_{j=0}^{m-1} (Y_j - Z_j)^2(s_{j+1} - s_j)]$ and inside the expectation, the expression becomes $\int_0^t (f(s) - g(s))^2 dt$ where $f(t) = X(s_j)$ for $s_j \leq t < s_{j+1}$ and $g(t) = X(t_i)$ for $t_i \leq t < t_{i+1}$, ie they are both constant on these intervals. This gives an exact value for the difference. Now if the mesh sizes of the two partitions are sent to zero then f and g both tend pointwise to X since X is continuous. From this observation and bounded convergence theorem, it follows that
 $E(\int_0^t (f(s) - g(s))^2 ds) \rightarrow 0$.
- Now take two arbitrary partitions and let I_1, I_2 be the two sums. We can always find a common refinement of the two partitions, say by union of points. Let I_3 be the sum for this common refinement. The above argument shows that $E(I_1 - I_3)^2, E(I_2 - I_3)^2$ both tend to 0 as the mesh sizes go to 0. Therefore $E(I_1 - I_2)^2$ also goes to 0, since $(a + b)^2 \leq 2a^2 + 2b^2$. In particular, if we take any sequence of partitions with mesh size going to zero and we let I_n be the sum corresponding to the nth partition, then $\{I_n\}_{n \geq 1}$ is a cauchy sequence in L^2 meaning that $\forall \epsilon > 0$ there exists n_0 st for all $m, n \leq n_0$, $E(I_m - I_n)^2 < \epsilon$. Fact: any sequence that is Cauchy in L^2 converges to a limit in L^2 . This limit is denoted by $\int_0^t X(s) dB(s)$, so does not depend on the choice of partitions.
- This integral is random, it is the limit of a sequence of random variables. Nothing we have from regular calculus still works because B is non-differentiable in any way

Ito's Formula

- This is the fundamental theorem of stochastic calculus.
- **Ito's Formula:** For any $f \in C^2$ (twice continuously differentiable functions, satisfying $\int_0^t E(f'(B(s))^2)ds < \infty$) $f(B(t)) - f(B(0)) = \int_0^t f'(B(s))dB(s) + \frac{1}{2} \int_0^t f''(B(s))ds$.
 - We would get the first integral in ordinary calculus, just using the first derivative. But here we get a stochastic integral plus and ordinary integral of the second derivative.
 - If B was differentiable, we would only have the first term on the RHS using the FTC.
 - Proof: Let's assume f''', f'' exists and is bounded by some C for simplicity. Take partition $0 = t_0 < t_1 < \dots < t_n = t$, then $f(B(t)) - f(B(0))$ taking telescoping sum of Taylor expansion
 $= \sum_{i=0}^{n-1} (f(B(t_{i+1})) - f(B(t_i)))$,
 $= \sum_{i=0}^{n-1} f'(B(t_i))(f(B(t_{i+1})) - f(B(t_i))) + \frac{1}{2} \sum_{i=0}^{n-1} f''(B(t_i))(f(B(t_{i+1})) - f(B(t_i)))^2 + R$ where remainder $R |R| \leq \frac{C}{6} \sum_{i=0}^{n-1} |B(t_{i+1}) - B(t_i)|^3$. B is not smooth, so this second derivative sum will not simply go to 0 with a finer mesh. As the mesh size $\max(t_{i+1} - t_i) \rightarrow 0$ then

$\sum_{i=0}^{n-1} f'(B(t_i))(f(B(t_{i+1})) - f(B(t_i))) \rightarrow \int_0^t f'(B(s))dB(s) \in L^2$ as discussed before. The expected value of $|R|$ is bounded by some constant times $|B(t_{i+1}) - f(B(t_i))|^3$ so

$B(t_{i+1}) - f(B(t_i)) \sim N(0, t_{i+1} - t_i) \stackrel{d}{=} \sqrt{t_{i+1} - t_i} Z$ for std normal Z . So

$E(|B(t_{i+1}) - f(B(t_i))|^3) = \sqrt{t_{i+1} - t_i}^3 E|Z|^3$ So

$E|R| \leq C \sum_{i=0}^{n-1} |t_{i+1} - t_i|^{3/2} \leq C \sum_{i=0}^{n-1} \max(t_{i+1} - t_i)^{1/2} \sum_{i=0}^{n-1} (t_{i+1} - t_i) = Ct\sqrt{\text{mesh size}}$, which goes to 0 as mesh size goes to 0. So $P(|R| > \epsilon) \leq \frac{E|R|}{\epsilon} \rightarrow 0$ as mesh size goes to 0. So of our 3 terms for $f(B(t)) - f(B(0))$, the first goes to the stochastic integral, the third term goes to 0, and we just need to determine the second term.

- Claim: $\frac{1}{2} \sum_{i=0}^{n-1} f''(B(t_i))(f(B(t_{i+1})) - f(B(t_i)))^2 \rightarrow \int_0^t f''(B(s))ds \in L^2$ as mesh size goes to 0. If we can prove this claim, then $f(B(t)) - f(B(0))$ is the sum of 3 terms: first tending to $\int_0^t f'(B(s))dB(s)$, second to $\int_0^t f''(B(s))ds$ and third to 0. But $f(B(t)) - f(B(0))$ itself has no dependence on the partition - the difference is always the sum of these three things. Thus we can fix this and send mesh size to 0 and get $f(B(t)) - f(B(0)) = \int_0^t f'(B(s))dB(s) + \frac{1}{2} \int_0^t f''(B(s))ds$.
- Proof: Note that by ordinary Riemann integration, $\sum_{i=0}^{n-1} f''(B(t_i))(t_{i+1} - t_i) \rightarrow \int_0^t f''(B(s))ds$ as mesh size goes to 0. So it suffices to prove that the difference between this integral and the one we are interested in tends to 0: $\sum_{i=0}^{n-1} X_i \Delta_i \rightarrow 0 \in L^2$ for $X_i = f''(B(t_i))$, $\Delta_i = (B(t_{i+1}) - B(t_i))^2 - (t_{i+1} - t_i)$. Now $E((\sum \Delta_i X_i)^2) = \sum_{i,j} E(X_i \Delta_i X_j \Delta_j)$ Now take $i < j$, then X_i, Δ_i, X_j are functions of $(B(s))_{s \leq t_j}$ and $\Delta_j = (B(t_{j+1}) - B(t_j))^2 - (t_{j+1} - t_j) \perp (B(s))_{s \leq t_j}$ by the Markov property. $E(\Delta_j) = 0$, therefore $E(X_i \Delta_i X_j \Delta_j) = 0$. Similarly this is zero if $i > j$. Thus $E(\sum \Delta_i X_i)^2 = \sum_{i=0}^{n-1} E(\Delta_i^2 X_i^2) \stackrel{\perp}{=} \sum_{i=0}^{n-1} E(\Delta_i^2) E(X_i^2)$. Note $X_i = f''(B(t_i))$, f'' bounded, so $E(X_i^2) \leq \text{constant}$. And $B(t_{i+1}) - B(t_i) \sim N(0, t_{i+1} - t_i)$, so $\Delta_i = (\sqrt{t_{i+1} - t_i} Z)^2 - (t_{i+1} - t_i) = (t_{i+1} - t_i)(Z^2 - 1)$. Thus $E(\Delta_i^2) = E((Z^2 - 1)^2)(t_{i+1} - t_i)^2$. So $\sum E(\Delta_i^2) E(X_i^2) \rightarrow 0$ using the same argument as before as mesh size goes to 0.

Applications of Ito's Formula

- Evaluation of **stochastic integrals**:
 - Suppose we want to evaluate $\int_0^t B(s)dB(s)$. Looking at Ito's Formula, it is clear that we need to apply the formula with a function f such that $f'(x) = x$. Take $f(x) = x^2/2$, then Ito gives $f(B(t)) - f(B(0)) = \int_0^t f'(B(s))dB(s) + \frac{1}{2} \int_0^t f''(B(s))ds$. So $\frac{B(t)^2}{2} = \int_0^t B(s)dB(s) + \frac{1}{2} \int_0^t ds$. This implies that $\int_0^t B(s)dB(s) = \frac{1}{2}(B(t)^2 - t)$.
 - Notice stochastic integrals are martingales.
 - Important note: When approximating $\int_0^t X(s)dB(s)$ by $\sum_{i=1}^t X(s_i)(B(s_{i+1}) - B(s_i))$ it is crucial that we use $X(s_i)$, that is the left endpoint of the interval. If we use eg. $X(\frac{s_i+s_{i+1}}{2})$ the limit will be completely different. This is quite unlike Riemann integrals. If we use $(X(s_i) + X(s_{i+1}))/2$, the limit is again different and is called the Stratonovich integral and is denoted by $\int_0^t X(s) \circ dB(s)$. Though generally the Ito integral is preferred as it is easier to work with.
- A more general Ito's Formula: Let $f(t, x) : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$. Suppose f is once continuously differentiable in t , twice in x . Also assume that $\forall t, \int_0^t E[(\frac{\partial f}{\partial x}(s, B(s))^2]ds < \infty$ then for all t $f(t, B(t)) - f(0, B(0)) = \int_0^t \frac{\partial f}{\partial t}(s, B(s))ds + \int_0^t \frac{\partial f}{\partial x}(s, B(s))dB(s) + \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, B(s))ds$. Then $= \int_0^t (\frac{\partial f}{\partial t}(s, B(s)) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(s, B(s)))ds + \int_0^t \frac{\partial f}{\partial x}(s, B(s))dB(s)$
- **Stochastic differential equations**
 - Let $X(t) = e^{B(t)} = f(B(t))$. Then by Ito, $X(t) - X(0) = \int_0^t e^{B(s)}dB(s) + \frac{1}{2} \int_0^t e^{B(s)}ds = \int_0^t X(s)dB(s) + \frac{1}{2} \int_0^t X(s)ds$. Therefore the process X

satisfies a "stochastic integral equation"

- The corresponding diff equation would be $\frac{dX(t)}{dt} = X(t) \frac{dB(t)}{dt} + \frac{1}{2}X(t)$. However B is not differentiable. Therefore the diff eq does not make sense; only the integral equation makes sense. Still, we continue to call this a stochastic diff eq; the shorthand for writing the integral equation is $dX(t) = X(t)dB(t) + \frac{1}{2}X(t)dt$, imagining it makes sense and multiply through by dt.
- Just like in discrete time we have the first difference equation $Z_t - Z_0 = \sum_{j=1}^t \epsilon_{t-j}$ to describe a random walk, we now have our Brownian motion $Z_{t+\Delta} - Z_t \sim N(0, \Delta)$ with normally distributed increments. Our first difference operator, the continuous *epsilon* is dz_t - the forward distance at time t $dz_t = \lim_{\Delta \rightarrow 0} (Z_{t+\Delta} - Z_t)$.
- Since the variance $\sigma^2(Z_{t+\Delta} - Z_t) = \Delta$, the SD is $\sigma(Z_{t+\Delta} - Z_t) = \sqrt{\Delta}$. That means the typical size of $dz_t \sim \sqrt{dt}$ - the infinitesimal change in t.

- Black-Scholes**

- Suppose that $X(t)$ is the price of a stock at time t. The assumption is that in time interval Δt , the stock price changes by a small proportion which is modeled as $N(0, \sigma^2 \Delta t)$ where $\sigma^2 > 0$ is some constant. That is $X(t + \Delta t)/X(t) \approx 1 + N(0, \sigma^2 \Delta t)$.
- Variance being proportional to the length of the interval is a reasonable assumption because if you consider two successive intervals of lengths $\Delta t_1, \Delta t_2$ then the total percentage change will be approximately the sum of the two changes, since $(1 + \alpha)(1 + \beta) \approx 1 + \alpha + \beta$ for small α, β . And if the successive changes are independent then the variances should add up.
- So we can model $X(t + \Delta t)/X(t) = 1 + \sigma(B(t + \Delta t) - B(t))$ where B is a standard BM. Independent increments with the correct variance, since $\sigma(B(t + \Delta t) - B(t)) \sim N(0, \sigma^2 \Delta t)$. This implies that $X(t + \Delta t) = X(t) + \sigma X(t)(B(t + \Delta t) - B(t)) \implies X(t + \Delta t) - X(t) = \sigma X(t)(B(t + \Delta t) - B(t))$. Summing both sides over $0 = t_0 < t_1 < \dots < t_n = t$ where $\Delta t_i = t_{i+1} - t_i$, then summing both sides of $X(t_{i+1}) - X(t_i) = \sigma X(t_i)(B(t_{i+1}) - B(t_i))$. In the limit as $\max(t_{i+1} - t_i) \rightarrow 0$ we get $X(t) - X(0) = \int_0^t \sigma X(s) dB(s)$ so the process $X(t)$ must satisfy the stochastic diff eq $dX(t) = \sigma X(t) dB(t)$, where σ is an unknown constant that we can estimate from data.
- For BM every step is independent of the past and the jumps are homogenous. Here the size of the jump is proportional to the value, which is more realistic to stock prices, ie. taking the log of the value is close to BM.
- To solve the diff eq, assume $X(t) = f(t, B(t))$ for some function $f(t, x)$. We do not yet know if there is a solution of this form, but we will try it. By Ito,
$$X(t) - X(0) = \int_0^t \frac{\partial f}{\partial x}(s, B(s)) dB(s) + \int_0^t (\frac{\partial f}{\partial t}(s, B(s)) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(s, B(s))) ds \text{ and we want}$$

$$X(t) - X(0) = \int_0^t \sigma X(s) dB(s). \text{ So we need (1) } \frac{\partial f}{\partial x} = \sigma f \text{ for the first part, and (2) } \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} = 0$$

(Note $f(0, 0) = X(0)$ is given).

- In (1), treat t like a constant. Solving (1) gives $f(t, x) = A(t) \exp(\sigma x)$ for some $A(t)$ constant that depends only on t. In (2), $\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} = A'(t) \exp(\sigma x) + \frac{\sigma^2}{2} A(t) \exp(\sigma x)$. This $= 0 \implies A'(t) + \frac{\sigma^2}{2} A(t) = 0 \implies A(t) = C \exp(-\frac{1}{2} \sigma^2 t)$. Therefore $f(t, x) = C \exp(\sigma x - \frac{1}{2} \sigma^2 t)$ and $X(t) = C \exp(\sigma B(t) - \frac{1}{2} \sigma^2 t)$ and to evaluate C, we take $t = 0$ since $B(0) = 0$ we get $C = X(0)$. So the solution to the equation is $X(t) = X(0) \exp(\sigma B(t) - \frac{1}{2} \sigma^2 t)$
- A more **general version**: Assume that you are investing in a fund which satisfies $X(t + \Delta t)/X(t) = 1 + N(0, \sigma^2 \Delta t) + \mu \Delta t$ where there is a drift. We have some μ , a guaranteed rate of average growth. We can model this as $X(t + \Delta t)/X(t) = 1 + \sigma(B(t + \Delta t) - B(t)) + \mu \Delta t$. Rearranging get $X(t + \Delta t) = X(t) + \sigma X(t)(B(t + \Delta t) - B(t)) + \mu X(t) \Delta t$ which implies $X(t + \Delta t) - X(t) = \sigma X(t)(B(t + \Delta t) - B(t)) + \mu X(t) \Delta t$. The limiting stochastic diff eq as $\Delta t \rightarrow 0$

given by $dX(t) = \sigma X(t)dB(t) + \mu X(t)dt$.

- μ is called the drift coefficient and σ is the volatility coefficient. There is a guaranteed returns and a random fluctuation controlled by sigma.
- Solving by Ito: Assume $X(t) = f(t, B(t))$. We get (1) $\frac{\partial f}{\partial x} = \sigma f$ and (2) $\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} = \mu f$. Then (1) implies $f(t, x) = A(t)\exp(\sigma x)$ and (2) implies $A'(t)\exp(\sigma x) + \frac{1}{2}\sigma^2 A(t)\exp(\sigma x) = \mu A(t)\exp(\sigma x)$, which implies $A'(t) = (\mu - \frac{1}{2}\sigma^2)A(t)$, $A(t) = C\exp((\mu - \frac{1}{2}\sigma^2)t)$. Then $f(t, x) = C\exp(\sigma x + (\mu - \frac{1}{2}\sigma^2)t)$ and $X(t) = X(0)\exp(\sigma B(t) + (\mu - \frac{1}{2}\sigma^2)t)$
- What is the behavior of $X(t)$ as $t \rightarrow \infty$. Note $X(t) \geq 0$ for all t and $X(0) > 0$. Is it converging, diverging or 0? Noting that $\sigma B(t) + (\mu - \frac{1}{2}\sigma^2)t = t[\sigma \frac{B(t)}{t} + \mu - \frac{1}{2}\sigma^2]$ and $\frac{B(t)}{t} \rightarrow 0$ as $t \rightarrow \infty$. Thus there are three possibilities.
 - (1) If $\mu > \frac{1}{2}\sigma^2$ then $X(t) \rightarrow \infty$.
 - (2) If $\mu < \frac{1}{2}\sigma^2$ then $X(t) \rightarrow 0$
 - (3) If $\mu = \frac{1}{2}\sigma^2$ then $\limsup X(t) = \infty$ and $\liminf X(t) = 0$
- It is not enough that we have a positive drift.. We have to beat a threshold around the volatility. In reality we won't hit the third limit case. If we didn't have any stochastic term, our fund would blow up exponentially. But the volatility allows the fund to go to 0 even with the positive drift.
- **Change of Variable formula:** Suppose that $X(t)$ satisfies $dX(t) = a(t)dt + b(t)dB(t)$ where $a(t), b(t)$ are stochastic processes. Let $Y(t) = f(t, X(t))$. Then

$$dY(t) = \frac{\partial f}{\partial x}(t, X(t))b(t)dB(t) + \left[\frac{\partial f}{\partial t}(t, X(t)) + \frac{\partial f}{\partial x}(t, X(t))a(t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, X(t))b(t)^2 \right] dt$$
 - While the formula is complicated, there are some rules to get here.
 - $(dB(t))^2 = dt$
 - $dB(t)dt = 0$
 - $(dt)^2 = 0$
 - for $Y(t) = f(t, X(t))$ then $dY(t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dX(t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(dX(t))^2$ where $(dX(t))^2$ is computed using the above rules. Looks like Ito / Taylor expansion
 - This gives us $(dX(t))^2 = b(t)^2dt$ and $dX(t) = a(t)dt + b(t)dB(t)$
 - This gives us another way to solve things. $dX(t) = \sigma X(t)dB(t)$ Let $Y(t) = \log X(t)$ then by the change of variable formula, we get

$$dY(t) = \frac{1}{X(t)}dX(t) + (-\frac{1}{2X(t)^2})(dX(t))^2 = \sigma dB(t) - \frac{1}{2X(t)^2}\sigma^2 X(t)^2 dt = \sigma dB(t) - \frac{1}{2}\sigma^2 dt$$
 Integrating both sides: $Y(t) = Y(0) + \sigma B(t) - \frac{1}{2}\sigma^2 t \implies X(t) = X(0)\exp(\sigma B(t) - \frac{1}{2}\sigma^2 t)$
- **Ito Isometry:** Suppose $\{X(t)\}$ is an adapted process, let $Y(t) = \int_0^t X(s)dB(s)$ then $\{Y(t)\}$ is a martingale and $E(Y(t)^2) = \int_0^t E(X(s)^2)ds$
 - Proof: start from simple process, discretization, and pass to the limit.

Markov Chains

Markov Chains on Finite State Spaces

- Let S be a finite or countable set, then a sequence of S -valued random variables X_0, X_1, \dots is called a Markov Chain if for any $n \geq 1$, $(x_0, \dots, x_n) \in S$,
- $P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1})$. Given the whole past, the next step is determined completely by the immediate past

- Example: Simple random walk. Let Y_1, Y_2, \dots iid ± 1 with equal prob. Then let $S_0 = 0$, $S_n = \sum_{i=1}^n Y_i$. What is $P(S_n = s_n | S_0 = s_0, \dots, S_{n-1} = s_{n-1}) = P(Y_n = s_n - s_{n-1} | S_1 = s_1 s_0, \dots, Y_{n-1} = s_{n-1} - s_{n-2}) = P(Y_n = s_n - s_{n-1})$ by independence of Y_1, Y_2, \dots
 - This equals $P(S_n = s_n | S_{n-1} = s_{n-1})$ because $P(S_n = s_n | S_{n-1} = s_{n-1}) = P(Y_n = s_n - s_{n-1} | S_{n-1} = s_{n-1}) = P(Y_n = s_n - s_{n-1})$ by independence. Thus $\{S_n\}$ is a MC on \mathbb{Z} .
- A MC $\{X_0, X_1, \dots\}$ as above is called time homogeneous if for any $x, y \in S$, $P(X_n = y | X_{n-1} = x)$ does not depend on n.
- Now let us assume that S is finite. If S has N elements, we may index the elements as 1,2,3,...N and take $S = \{1, 2, 3, \dots, N\}$. Also assume that the MC is time homogeneous. Then for any $1 \leq i, j \leq N$, $p_{ij} = P(X_n = j | X_{n-1} = i)$ does not depend on n. The $(N \times N)$ matrix $P = (p_{ij})_{1 \leq i, j \leq N}$ is called the transition matrix of the MC. Any finite state space MC that is time homogeneous has a transition matrix.
- Example: two state chain. Let $S = \{1, 2\}$. When the chain is at 1 it goes to 1 wp 1/2 and to 2 wp 1/2. When is at 2, it goes to 1 always. Then we get transition matrix $P = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$
- Note that the row sums are always 1. This is because $\sum_{j=1}^N p_{ij} = \sum_{j=1}^N P(X_1 = j | X_0 = i) = P(\cup_{j \in S} \{X_1 = j\} | X_0 = i) = 1$ - it must take on some value WP1. Any square matrix with non-negative entries which sum to 1 on each row defines the transition matrix of a MC. Such matrices are called stochastic matrices.
- **Chapman-Kolmogorov Equation:** Let X_0, X_1, \dots be a time homogeneous MC on a finite state space S with transition matrix P. Let $p_{ij}^{(n)} = P(X_n = j | X_0 = i)$ and $P^{(n)} = (p_{ij}^{(n)})_{i,j \in S}$ - the n-step transition matrix.
 - C-K Equation: For any n, $P^{(n)} = P^n$ - the n step transition matrix is the nth power of the 1 step transition matrix.
 - Proof:

$$P(X_0 = i_0, \dots, X_n = i_n) = P(X_0 = i_0)P(X_1 = i_1 | X_0 = i_0) \dots P(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1})$$
. The collapses as the denominator of each conditional ratio cancels with the numerator of the next. This equals $P(X_0 = i_0)P_{i_0 i_1} \dots P_{i_{n-1} i_n}$. Thus

$$P(X_n = j | X_0 = i) = \frac{P(X_n = j, X_0 = i)}{P(X_0 = i)} = \frac{1}{P(X_0 = i)} \sum_i P(X_0 = i, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j)$$
. This sum by our previous step $= \frac{1}{P(X_0 = i)} \sum_i P(X_0 = i)P_{ii_1} \dots P_{i_{n-1} j} = \sum_i P_{ii_1} \dots P_{i_{n-1} j} = (i, j)$ th entry of P^n .
- Example: $P = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$, then $P^2 = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix}$, $P^3 = \begin{bmatrix} 5/8 & 3/8 \\ 3/4 & 1/4 \end{bmatrix}$, etc.

Long Run Behavior

- What is the behavior of P^n as $n \rightarrow \infty$?
- Invariant distributions: Let X_0, X_1, \dots be a time homogeneous MC on a finite state space S with matrix P. A probability distribution on S is simply a collection on non-negative numbers $(\mu_i)_{i \in S}$ st $\sum_i \mu_i = 1$. That is μ_i should be interpreted as the probability of i. In the MC world, such distributions are denoted by row vectors. $\mu = [\mu_1, \mu_2, \dots, \mu_N]$ for $S = \{1, \dots, N\}$.
- Let μ be a probability distribution on S. We say that μ is an invariant / stationary / equilibrium distribution for the MC X_0, \dots is $X_0 \sim \mu \implies X_1 \sim \mu$. That is, if $P(X_0 = j) = \mu_j \forall j$ then $P(X_1 = j) = \mu_j \forall i$. Note that $P(X_1 = j) = \sum_{i \in S} P(X_1 = j, X_0 = i) = \sum_{i \in S} P(X_1 = j | X_0 = i)P(X_0 = i)$. So we need that for all j in S, $\mu_j = \sum_{i \in S} \mu_i p_{ij}$. In matrix notation: $\mu P = \mu$ (for row vector μ).

- Remark: let μ be a stationary distribution. Let $x_0 \sim \mu$ then $X_n \sim \mu \forall n$. Why?
 $P(X_n = j) = \sum_{i \in S} P(X_n = j | X_0 = i)P(X_0 = i) = \sum_{i \in S} \mu_i p_{ij}^{(n)}$ = jth element of μP^n by the C-K equation.
But $\mu P = \mu \implies \mu P^n = \mu \forall n$ since μ is the left eigenvector of P . Thus $X_n \sim \mu$.
- Example: finding stationary distribution of $P = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$.
 - $\mu P = \mu \implies (\mu_1 \mu_2) \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix} = (\mu_1 \mu_2)$. So $\mu_1 = \frac{1}{2}\mu_1 + \mu_2$, $\mu_2 = \frac{1}{2}\mu_1 \rightarrow$ both equations are the same, but we have another constraint to solve the system - μ is a probability distribution so $\mu_1 + \mu_2 = 1$.
 - Solving: $\mu_1 = \frac{2}{3}$, $\mu_2 = \frac{1}{3}$. In this case, this is clearly the only stationary distribution.
- **Theorem: Existence of stationary distribution.** Let X_0, X_1, \dots be a T-H MC on a finite state space S with matrix P . Then there is at least one stationary distribution for this chain. (ie, there exists at least one probability distribution μ in S solving $\mu P = \mu$)
 - Proof: Take any probability distribution μ_0 on S . Then note that $\mu_0 P^n$ is a probability distribution on S for any n . Let $\mu^{(n)} = \frac{1}{n+1}(\mu_0 + \mu_0 P + \dots + \mu_0 P^n)$ then $\mu^{(n)}$ is also a probability distribution on S . Note that $\mu^{(n)} - \mu^{(n)} P = \frac{1}{n+1}(\mu_0 + \mu_0 P + \dots + \mu_0 P^n) - \frac{1}{n+1}(\mu_0 + \mu_0 P + \dots + \mu_0 P^n)P$. Then $= \frac{1}{n+1}(\mu_0 - \mu_0 P^{n+1}) \rightarrow 0, n \rightarrow \infty$ since $\mu_0, \mu_0 P^{n+1}$ are both prob distributions and therefore have bounded entries. Note that each $\mu^{(n)}$ belongs to the set $\{(x_1, \dots, x_N) \in \mathbb{R}^N : x_i \geq 0, \sum x_i = 1\}$. This set is closed and bounded so by Bolzano-Weierstrass theorem, any sequence in this set has a convergent subsequence converging to some element in this set.
 - Let $\mu^{(n_k)}$ converge to some prob distribution μ where $n_1 < n_2 < \dots$ But we know that $\mu^{(n)} - \mu^{(n)} P \rightarrow 0, n \rightarrow \infty \implies \mu^{(n_k)} - \mu^{(n_k)} P \rightarrow 0$ but $\mu^{(n_u)} - \mu^{(n_u)} P \rightarrow \mu - \mu P$. Thus $\mu = \mu P$
- Note: stationary distribution may not be unique. For example, $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ then both $(1 0), (0 1)$ are both stationary distributions.
- **Doeblin Condition:** Let X, S, P be as before with stationary distribution μ . We say that P satisfies the Doeblin condition if for some $k \geq 1$ some $\epsilon \in (0, 1)$ and the square matrix M whose rows are all equal to μ , we have $P^k \geq \epsilon M$ where \geq means that the (i,j) th element of P^k is \geq the (i,j) th element ϵM for all i, j .
 - Each entry of P^k is bigger than each entry of ϵM - dominates this matrix. For all i, j $p_{ik}^{(k)} \geq \epsilon \mu_j$. There is some k, ϵ so that this happens.
- For a matrix $A = (a_{ij})_{i,j \in S}$, let $\|A\| = \max_{i \in S} \sum_{j \in S} |a_{ij}|$ - sum abs value of entries along row, take the max row. Then $\|A + B\| \leq \|A\| + \|B\|$ and $\|AB\| \leq \|A\| \|B\|$
- **Doeblin's Theorem:** Suppose that P satisfies Doeblin's condition for some $k \geq 1, \epsilon \in (0, 1)$. Then for all n , $\|P^n - M\| \leq 2(1 - \epsilon)^{\lfloor n/k \rfloor}$ where $[x]$ is the integer part of x .
 - Aside: $C = AB$ then $c_{ij} = \sum a_{ij} b_{kj}$ and therefore $\sum_j c_{ij} \leq \sum_j \sum_k |a_{ij} b_{kj}|$ by triangle inequality.
Then $= \sum_k |a_{jk}| (\sum_j |b_{kj}|) \leq \|B\| \|A\|$
 - Proof: Define $Q = \frac{P^k - \epsilon M}{1 - \epsilon}$. Claim: Q is also a stochastic matrix. First note that the entries of Q are all non-negative since $P^k \geq \epsilon M$. Note that any row sum of $Q = \frac{1}{1-\epsilon}(\text{row sum of } P^k - \epsilon \text{ row sum of } M)$, but the row sums are both 1 since stochastic matrices. So get $\frac{1}{1-\epsilon}(1 - \epsilon) = 1$. Thus Q is a transition matrix
 - Note $P^k = (1 - \epsilon)Q + \epsilon M$, as a rewrite of the formula above. We know that $\mu P = \mu$. Since each row of M is μ , this shows that $MP = M$ and since each column of M is constant and P is a stochastic matrix $PM = M$ - taking row of P times column of M , but the row sum of P is 1 so we just get M back. Thus $P^k M = MP^k = M$. Also $M^2 = M$ since μ is a probability distribution. Thus $QM = M = MQ$. Therefore any product of multiple copies of M and Q equals M as soon as at least one M is present, eg.

$$Q^3 M^2 Q M^5 Q^7 = M$$

- Therefore for any $m \geq 1$, $P^{km} = ((1 - \epsilon)Q + \epsilon M)^m$ = expand... and get
 $= (1 - \epsilon)Q^k + \sum_{k=1}^m \binom{m}{k} (1 - \epsilon)^{m-k} \epsilon^k M$ then by binomial theorem $= (1 - \epsilon)Q^k + (1 - (1 - \epsilon)^m)M$.
Now take any $n \geq 1$. Let $m = [n/k]$ so that $n = km + r$ for some $r \leq k - 1$. Then
 $P^n - M = P^r(P^{km} - M) = (1 - \epsilon)^m P^r(Q^m - M)$ this implies that
 $\|P^n - M\| \leq (1 - \epsilon)^m \|P^r\| \|Q^m - M\| \leq (1 - \epsilon)^m \|P^r\| (\|Q^m\| + \|M\|)$. Now any stochastic matrix has norm = 1. So $\|P^n - M\| \leq 2(1 - \epsilon)^m$
- Intuition behind theorem: If P^k dominates ϵM then if you run the chain up to k steps, equivalent to choosing from the stationary distribution with probability ϵ and with probability $1 - \epsilon$ choosing from new distribution Q . Each time tossing a coin, choosing from one distribution or another, taking a step. After some time, eventually choose from the stationary distribution for sure, and there you remain. Once you choose from the stationary distribution, you will continue to follow the stationary distribution.

- Example: $P = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$ we saw $\mu = (2/3 \ 1/3)$ so $M = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}$. And we saw $P^2 = \begin{bmatrix} 3/4 & 1/4 \\ 1/2 & 1/2 \end{bmatrix}$.
Now $P^2 \geq \frac{3}{4} \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} = \frac{3}{4}M$. So the Doeblin condition is satisfied with $k = 2$, $\epsilon = 3/4$. Thus
 $\|P^n - M\| \leq 2(\frac{3}{4})^{[n/2]}$. This proves convergence to this matrix M .
- Definition: The **period** of a state i in S is the GCD of $\{n \geq 1 : P_{ii}^{(n)} > 0\}$. If no such n exists, then the period = ∞
 - For example SRW on \mathbb{Z} , period of any state is 2. Start from some state, cannot return in an odd number of steps, but for any even number there is a positive probability.
- Example: $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, then period of both states = 2.
- Example: $P = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 0 \end{bmatrix}$, claim both states have period 1. Such states are called **aperiodic**. Clearly $P_{11}^{(1)} = 1/2 \implies \{n \geq 1 : P_{ii}^{(n)} > 0\}$ contains 1 so GCD is 1. To show that state 2 is aperiodic, note that from P^2, P^3 , we can return in 2 or 3 steps \rightarrow thus the GCD is 1.
- The MC is called **aperiodic** if all states are aperiodic.
- Irreducible:** State j is said to be accessible from a state i if $P_{ij}^{(n)} > 0$ for some n - there is some chance of going from j to i in some number of steps. The MC is called irreducible if every state is accessible from every other.
- Lemma:** Take any $k \geq 1$ and $i_0, \dots, i_k \in S$ and $n_1, \dots, n_k \geq 1$. Suppose that every i_l is accessible from i_{l-1} in n_l steps for $l = 1, \dots, k$. Let $n = n_1 + \dots + n_k$. Then i_k is accessible from i_0 in n steps.
 - Proving one state is accessible from another, can find a chain between them to show accessibility.
 - Proof: By C-K eqns.
- Theorem:** Let X_0, \dots be a time homogeneous, irreducible, aperiodic MC on a finite state space S . Then the transition matrix P has a unique invariant distribution μ and if M is the matrix with all rows = μ then $P^n \rightarrow M$ as $n \rightarrow \infty$.
 - For any i, j in S $\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \mu_j$
 - If it is periodic, then P^n will always have some 0 entries that are changing with each n , so we cannot converge to a given matrix.
 - If not irreducible, then it is not unique and we would not get convergence. So the conditions are quite important.
 - Proof: Show that any irreducible and aperiodic chain on a finite state space satisfies Doeblin's condition.

- o Lemma: For any i in S , $P_{ii}^{(n)} > 0$ for all sufficiently large n .
 - Proof: By irreducibility, $P_{ii}^{(n)} > 0$ for some $n \geq 1$. Let A be the set of all such n . By aperiodicity, $\text{GCD}(A) = 1$. Claim: There exists some k , some $a_1, \dots, a_k \in A$ and some $u_1, \dots, u_k \in \mathbb{Z}$ such that $u_1, a_1 + \dots + u_k a_k = 1$. Proof: Let d be the minimum positive value of $u_1 a_1 + \dots + u_k a_k$ over all k , all $a_1, \dots, a_k \in A$, $u_1, \dots, u_k \in \mathbb{Z}$. Take any $a \in A$, let r be remainder when a is divided by d (note that $d \leq a$). But then r is again a linear combination of elements of A with integer coefficients and $r < d$. Thus r must be 0, since it is non-negative. Thus d divides every element of A also $d > 0$. Since $\text{gcd}(A) = 1$, this shows $d=1$.
 - Fix some k , $a_1, \dots, a_k \in A$, $u_1, \dots, u_k \in \mathbb{Z}$. Let $m = a_1 + \dots + a_k$. Take any $n \geq 1$. Let $n = qm + r$ where $q = \text{quotient}$, $r = \text{remainder}$. So $n = qm + r = qm + r(n_1 a_1 + \dots + u_k a_k)$. Since $r \leq m - 1$ and $q \rightarrow \infty$, $n \rightarrow \infty$, we see that the coeffs $(q + ru_1), \dots, q + ru_k$ are positive when n is large enough. In other words, any large enough n can be expressed as a linear combination of a_1, \dots, a_k with positive integer coefficients. But $P_{ii}^{(a_1)} > 0, \dots, P_{ii}^{(a_k)} > 0$, so by the lemma $P_{ii}^{(n)} > 0$ for all sufficiently large n .
- o Lemma: For any i, j in S , $P_{ij}^{(n)} > 0$ for all sufficiently large n .
 - Proof: Take any i and j . By the previous lemma there exists n st $\forall n \geq m$, $P_{ii}^{(n)} > 0$. Also by irreducibility, there exists k such that $P_{ij}^{(k)} > 0$. Let $m' = m + k$. Claim: $P_{ij}^{(n)} > 0$ when $n \geq m'$

Proof: $P_{ij}^{(n)} \geq P_{ii}^{(n-k)} P_{ij}^{(k)}$. Note $n - k \geq m' - k = m \implies P_{ii}^{(n-k)} > 0$ and $P_{ij}^{(k)} > 0 \implies P_{ij}^{(n)} > 0$
- o Corollary: There exists n sufficiently large so that all entries of P^n are strictly positive.
- o Theorem Proof: By the above corr. we can find some k st the entries of P^k are all strictly positive. By a previous result, we know that P has at least one invariant distribution μ . Let M be the matrix whose rows are all equal to μ . Since the entries of P^k are all strictly positive, we can find some small enough $\epsilon > 0$ st $\epsilon M \leq P^k$. Thus P satisfies Doeblin's condition and so $P^n \rightarrow M$, $n \rightarrow \infty$. Moreover if ν is another invariant distribution, we can similarly show that $P^n \rightarrow N$ where N is the matrix with all rows = ν . But P^n can have only one limit, so $\mu = \nu$, proving the uniqueness.
- Example: Top-to-random shuffle. Take a deck on n cards. State space of the chain is the set of all permutations of the n cards = S_n . Note $|S_n| = n!$. Move: Take the top card and insert it uniformly at random into the deck (it may just again be the top card wp 1/n). This defines a MC on S_n . This chain is time homogeneous.
 - o Note: This chain is aperiodic since starting from one state there is probability 1/n of staying at that state.
 - o The chain is irreducible, since it is possible to go from any state to any other state by a sequence of moves that take the top card and places it somewhere else. All of these moves have positive probability.
 - o Therefore there exists a unique invariant distribution and this is the limiting distribution of the chain starting from any state. We will show this is the uniform distribution on S_n
 - o We have to show that if $\mu_\sigma = \frac{1}{n!}$ for all $\sigma \in S_n$ then for any $\sigma \in S_n$, $\sum_{a \in S_n} \mu_a P_{a\sigma} = \mu_\sigma$ ie. $\frac{1}{n!} \sum_a P_{a\sigma} = \frac{1}{n!}$, ie. $\sum_a P_{a\sigma} = 1$
 - o Given a state $\sigma \in S_n$, what are all the states a such that σ is accessible from a in one step? (ie. $p_{a\sigma} > 0$). A state a has this property if and only if a can be obtained by taking out a card from the σ arrangement and placing it on the top of the deck. If a comes before σ , then if we put a card on top, σ reverses this action
 - o So there are exactly n such "a". For each such "a" $P_{a\sigma} = \frac{1}{n}$. If you take the top card from "a", the chance that you insert it exactly at the right place to get σ is $\frac{1}{n}$. Thus $\sum_a p_{a\sigma} = 1$. Therefore the uniform distribution on S_n is indeed the unique stationary distribution of this MC.

- Example: Random walk on a graph. Let V be a finite set. An edge on V is an unordered pair of elements. Picture the elements of V as points on a piece of paper, and an edge as a line segment connecting two points. Let E be some set of edges. The pair $G = (V, E)$ is called an undirected graph.
 - SRW X_0, X_1, \dots on G is a MC on V that evolves according to the following rule: If the chain is at a vertex v it chooses one of its neighbors uniformly at random and moves there. This covers a lot of examples, such as the random transpositions walk on S_n where you start with some arrangement of a deck of cards and choose a pair of cards and swap them. Here you can put a graph structure on S_n by putting an edge between two permutations iff one can be obtained from the other by swapping a single pair of cards.
 - Q: on a general graph, what is the long-term behavior of this walk? Note that in general, the walk may not be aperiodic. Consider $V = \mathbb{Z}$, $\{i, j\} \in E \iff |i - j| = 1$. This is a SRW on \mathbb{Z} , and is not aperiodic because $p_{ii}^{(n)} > 0$ iff n is even. The same is true for any bipartite graph.
 - Easy fix: Lazy SRW - make the chain stay where it is WP $p \in (0, 1)$ and jump to a uniformly chosen neighbor WP $1-p$. Moves like the SRW just has different tempo. This is aperiodic since we can return to the same state in 1 step.
 - If the graph is disconnected, meaning that $V = V_1 \cup V_2$ where V_1, V_2 disjoint and there are no edges between vertices in V_1 and those in V_2 , then the SRW is not irreducible. But if the graph is connected then the SRW and Lazy SRW are irreducible - for any vertices v, w there is a path connecting v to w and there is a positive probability that the walk takes this path. Therefore the lazy SRW on a finite connected graph is irreducible and aperiodic and has a unique stationary distribution.
 - For any vertex v , let d_v be the number of neighbors of v , the degree of the vertex. The stationary distribution for the lazy SRW is given by $\mu_v = \frac{d_v}{\sum_{u \in V} d_u}$, where the denominator is simply normalizing to get a sum to 1 for a probability distribution. Note this has no dependence on the holding probability p .
 - Proof: Clearly μ is a probability distribution. Let P_{uv} denote the transition probability from u to v .
$$P_{uv} = \begin{cases} p & \text{if } u = v \\ \frac{1-p}{d_u} & \text{if } v \text{ is a neighbor of } u \\ 0 & \text{else} \end{cases}$$

Let $N_v = \text{set of neighbors of } v$. Then

$$\sum_{u \in N_v} \mu_u P_{uv} = p\mu_v + (1-p) \sum_{u \in N_v} \frac{\mu_u}{d_u}$$

$$= p\mu_v + (1-p) \sum_{u \in N_v} \frac{1}{d_u} \left(\frac{d_u}{\sum_{w \in V} d_w} \right) = \mu_v$$

Thus μ is the stationary distribution.

- **Reversible Markov Chains:** A time homogeneous MC on a finite state space S with transition matrix P is said to be reversible wrt a probability distribution μ on S if $\mu_u P_{uv} = \mu_v P_{vu}$, $\forall u, v \in S$.
- Remark: if this condition holds, then μ is a stationary distribution.
 - Proof: $\sum_{u \in S} \mu_u P_{uv} = \sum_{u \in S} \mu_u P_{vu} = \mu_v \sum_{u \in S} P_{uu} = \mu_v$
 - Consequently, if the MC has the property that for all u, v either $P_{uv} = P_{vu} = 0$ or both nonzero and $\frac{P_{uv}}{P_{vu}} = \frac{f_v}{f_u}$ for some set of positive numbers $(f_u)_{u \in S}$ then $\mu_u = \frac{f_u}{\sum_{v \in S} f_v}$ satisfies that the MC is reversible wrt μ .
 - In our example $P_{uv} \neq 0 \iff P_{vu} \neq 0$. In this case $\frac{P_{uv}}{P_{vu}} = \frac{(1-p)/d_u}{(1-p)/d_v} = \frac{d_v}{d_u}$.
 - Reversible MCs have many nice properties for example, the transition matrix is self adjoint under a suitable inner product and has a spectral decomposition.

Markov Chains on Countable State Spaces

- Example: SRW on \mathbb{Z} and on \mathbb{Z}^d .

- These chains typically don't have stationary distributions. For example, if μ is a stationary distribution for SRW on Z , then μ must satisfy $\mu_i = \frac{\mu_{i-1} + \mu_{i+1}}{2}$ $\forall i$. Claim, there cannot exist a prob distribution on Z with this property.
 - Proof: If there is such a μ , it cannot be constant everywhere since $\sum_{i=-\infty}^{\infty} \mu_i = 1$. For the same reason it cannot be non-decreasing or non-increasing everywhere on Z . There must exist at least one strict local maximum for μ , ie $\mu_{i-1} < \mu_i, \mu_{i+1} < \mu_i$. But this violates $\mu_i = \frac{\mu_{i-1} + \mu_{i+1}}{2}$.
- **Recurrence and Transience:** A state i is called recurrent if $P(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$ and transient if this is < 1 .
- Fact: a state is recurrent iff the chain returns to the state infinitely many times.

Polya's Theorem

- **Polya's Theorem:** SRW on \mathbb{Z}^d is recurrent if $d = 1, 2$ and transient if $d \geq 3$
- Theorem: Let X_0, X_1, \dots be a TH MC on a countable state space S . Then a state $i \in S$ is recurrent iff $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$ and transient if $\sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty$.
- Proof: Let $N = \text{number of visits to state } i \text{ starting at time 1}$. $N = \sum_{n=1}^{\infty} 1\{X_n = i\}$. By the monotone convergence theorem,

$$E(N|X_0 = i) = \sum_{n=1}^{\infty} E(1\{X_n = i\}|X_0 = i) = \sum_{n=1}^{\infty} P(X_n = i|X_0 = i) = \sum_{n=1}^{\infty} P_{ii}^{(n)}. ($$

$$E(\sum X_i) = \sum E(X_i) \text{ if } X_i \text{ are non-neg RVs.}$$
- Let $p = P(\text{Ever returning to } i | X_0 = i) = P(\cup_{n=1}^{\infty} \{X_n = i\} | X_0 = i)$ for a fixed i . Claim: $N \sim \text{Geom}(1 - p)$, ie. N has the same distribution as the number of tails before the first head if you keep tossing a coin with probability $1-p$ of head. Let's say the MC starts at i . Then it either returns to i at some time or never returns to i . Probability of never returning = p , probability of never returning = $1 - p$.
- Now consider the first time it returns to i (if it does). Since this is a MC, the future behavior will not depend on the past so the chance of returning again after the first time is also = p . Suppose it returns again, then again you can say that the chance of returning is p . Thus each return is like a coin coming up tails where the chance of tails = p .
- Consequence: $P(N \geq k | X_0 = i) = p^k$ for $k=1,2,\dots$ The chance you return at least k times is p^k .
 - Rigorous Proof: We will prove by induction on k . $P(N \geq 1 | X_0 = i) = p$ by definition of p , so this holds for $k = 1$. Suppose it holds for $k-1$, for any $n \geq 1$ and states $j_1, \dots, j_n \in S$ and $j_n = i$ and exactly $k-2$ of the other j 's equal to i . Let A_{j_1, \dots, j_n} be the event $\{X_1 = j_1, \dots, X_n = j_n\}$. Then $\{N \geq k-1\}$ is the union of these disjoint events over all n and all j_1, \dots, j_n as above. Since $\{N \geq k\} \subset \{N \geq k-1\}$ we get $P(N \geq k | X_0 = i) = P(\{N \geq k\} \cap \{N \geq k-1\} | X_0 = i) = \sum P(\{N \geq k\} \cap A_{j_1, \dots, j_n} | X_0 = i)$. But

$$= \sum P(\{N \geq k\} | A_{j_1, \dots, j_n} \cap X_0 = i) P(A_{j_1, \dots, j_n} | X_0 = i)$$

$$P(N \geq k | A_{j_1, \dots, j_n} \cap X_0 = i) = P(N \geq k | X_0 = i, \dots, X_n = j_n) = P(\cup_{m=n+1}^{\infty} \{X_m = i\} | \dots)$$

$$\text{MC, } = P(\cup_{m=n+1}^{\infty} \{X_m = i\} | X_n = i) = p \text{ since } j_n = i. \text{ Thus}$$

$$P(N \geq k | X_0 = i) = p \sum P(A_{j_1, \dots, j_n} | X_0 = i) = p P(N \geq k-1 | X_0 = i) = pp^{k-1} = p^k$$
 - Thus $E(N | X_0 = i) = \sum_{k=1}^{\infty} P(N \geq k | X_0 = i) = \sum_{k=1}^{\infty} p^k$. Therefore recurrence iff $p = 1 \iff \sum p^k = \infty \iff E(N | X_0 = i) = \infty \iff \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ for $p \in [0, 1]$. Similarly transience iff $p < 1 \iff \sum p^k < \infty \iff E(N | X_0 = i) < \infty \iff \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$. From this we also observe recurrent \iff number of returns = ∞ and transience iff this is finite. Since number of returns follow $\text{Geom}(1-p)$.
- Multinomial Coefficients: If have n balls and k boxes, the number of ways of distributing the balls so that box i receives n_i balls is $\frac{n!}{n_1! \dots n_k!}$, where $n_1 + \dots + n_k = n$. This generalizes the binomial coefficient which says the same thing for $k = 2$.

- Stirling's Formula: $\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n^{n+1/2}} e^{-n}} = 1$
 - Proof: the PDF of $Gamma(n+1, 1) = \frac{x^n e^{-x}}{n!}$. Thus $\int_0^\infty \frac{x^n e^{-x}}{n!} dx = 1$ which implies $n! = \int_0^\infty x^n e^{-x} dx$. This implies $\frac{n!}{n^n e^{-n}} = \int_0^\infty (x/n)^n e^{-(x-n)} dx$. Can sub $y = (x-n)/\sqrt{n}$ to get $= \int_{-\sqrt{n}}^\infty (1 + \frac{y}{\sqrt{n}})^n e^{-\sqrt{n}y} \sqrt{n} dy$. Thus $\frac{n!}{n^{n+1/2} e^{-n}} = \int_{-\sqrt{n}}^\infty (1 + \frac{y}{\sqrt{n}})^n e^{-\sqrt{n}y} dy$.
 - Then can show this goes to $\sqrt{2\pi}$ as $n \rightarrow \infty$. Note $\lim_{x \rightarrow 0} \frac{\log(1+x)-x}{x^2} = -1/2$ by using L'Hopital's rule twice. Then $(1 + y/\sqrt{n})^n e^{-\sqrt{n}y} = \exp(n \log(1 + y/\sqrt{n}) - \sqrt{n}y) = \exp(\frac{\log(1+y/\sqrt{n})-y/\sqrt{n}}{1/n}) = \dots$ Thus $\lim_{x \rightarrow 0} (1 + y/\sqrt{n})^n \exp(-\sqrt{n}y) = \exp(-y^2/2)$.
 - Using this and dominated convergence theorem, one can now argue that $\lim_{x \rightarrow 0} \int_{-\sqrt{n}}^\infty (1 + \frac{y}{\sqrt{n}})^n e^{-\sqrt{n}y} dy = \int_{-\infty}^\infty \exp(-y^2/2) dy = \sqrt{2\pi}$ since a normal density.
- Polya's Proof: Remember claim: Let X_0, X_1, \dots be a TH MC on a countable state space S . Then a state $i \in S$ is recurrent iff $\sum_{n=1}^\infty P_{ii}^{(n)} = \infty$ and transient if $\sum_{n=1}^\infty P_{ii}^{(n)} < \infty$.
 - We will prove it for initial state = (0,0,...0). If the chain is irreducible then either every state is recurrent or every state is transient. Let r_n = n-step transition probability from the origin to the origin. Clearly, $r_n = 0$ if n is odd. The origin is rec if $\sum r_n = \infty$ transient if $\sum r_n < \infty$.
 - Take some even n . To come back to the origin in n steps, the walk has to do the following. The number of steps in the positive direction along any axis must equal the number of steps in the negative direction. The n steps can be divided into n_i steps along axis i , $i = 1, \dots, d$ in $\frac{n!}{n_1! \dots n_d!}$ was. Here n_1, \dots, n_d are even numbers such that $\sum n_i = n$. We are trying to count the number of paths that start from the origin and return in d dimensions.
 - Once we have divided the steps into these d categories, the individual steps in category i can be assigned to be positive or negative in $\binom{n_i}{n_{i/2}}$ ways. The total number of paths of length n that start and end at the origin is $N_n = \sum_{n_1, \dots, n_d \text{ even}} \frac{n!}{n_1! \dots n_d!} \prod_{i=1}^d \binom{n_i}{n_{i/2}}$ - this is the exact number of paths. Each of these paths has probability $(2d)^{-n}$ of being taken by the walk. Thus $r_n = N_n (2d)^{-n}$ when n is even.
 - Case $d=1$: Here $r_n = \binom{n}{n/2} 2^{-n}$ for n even. By Stirling's formula, take any $n \geq 1$ then $r_{2n} = \binom{2n}{n} 2^{-2n} = \frac{(2n)!}{(n!)^2} 2^{-2n}$ Let $a_n \sim b_n$ mean $\frac{a_n}{b_n} \rightarrow 1, n \rightarrow \infty$ - "similar". Then $n! \sim \sqrt{2\pi n^{n+1/2}} e^{-n}$ and $r_{2n} \sim \frac{\sqrt{2\pi}(2n)^{2n+1/2} e^{-2n}}{(\sqrt{2\pi}(n)^{n+1/2} e^{-n})^2} 2^{-2n} = \dots = \frac{1}{\sqrt{\pi n}}$. Everything cancels and we are left with $r_{2n} \sim \frac{1}{\sqrt{\pi n}}$. By the ratio test for summability, this shows that $\sum r_{2n} = \infty$ and hence $\sum_{n=1}^\infty r_n = \infty$. Thus 0 is recurrent in $d = 1$.
 - Case $d = 2$: can go through the same process.