Probability Preliminaries

- *Set operations*

  Note the following basic property of unions and intersections of sets $\left[\bigcup_n A_n\right]^c = \bigcap_n A_n^c$ Or equivalently: $\bigcup_n A_n^c = \left[\bigcap_n A_n\right]^c$

  - ***Proposition***: Let $\{A_n\}$ be a sequence of (measurable) events such that $\mathbb{P}(A_n) = 1$ for all n. Then $\mathbb{P}\left(\bigcap_n A_n\right) = 1$.

    Proof: First consider that $\left[\bigcap_n A_n\right]^c = \bigcup_n A_n^c$. Then consider that, using the union bound $\mathbb{P}\left(\bigcup_n A_n^c\right) \le \sum_n \mathbb{P}\left(A_n^c\right) \quad = \sum_n 1 - \mathbb{P}\left(A_n\right) = \sum_n 0 = 0$. The union bound, used in the first line, simply states that if we add the probability of events without subtracting their intersections, the result will be greater than when considering the union (which leaves out intersections). ■

    Definition (limsup and liminf): Let $\{a_n\}$ be a real-value sequence. Then we write $\limsup_n a_n := \inf_m \sup_{n\ge m} a_n$. This is the least upper bound on that sequence. Similarly $\liminf_n a_n := \sup_m \inf_{n \ge m} a_n$. Note that these limits need not be finite. ***Remarks***:

    - If $\limsup_n a_n < a$, then $a_n < a$ *eventually* (ie: for sufficiently large n). If $\limsup_n a_n > a$, then $a_n > a$ infinitely often – in other words, there are infinitely many elements in the sequence above a.
    - Similarly, $\liminf_n a_n > a \Rightarrow a_n > a$ ev. and $\liminf a_n < a \Rightarrow a_n < a$ i.o.

  - ***Definition (limsup and liminf for sets)***: Let $\{A_n\}$ be a sequence of measurable sets. Then

    $$\limsup_n A_n := \bigcap_m \bigcup_{n\ge m} A_n = \lim_{m\to\infty} \bigcup_{n \ge m} A_n$$
    $$\liminf_n A_n := \bigcup_m \bigcap_{n\ge m} A_n = \lim_{m\to\infty} \bigcap_{n \ge m} A_n$$

    We can interpret these as follows. The lim sup of A_n is the set of events that are contained in infinitely many of the A_n (but not necessarily in all the A_n past a point – the event could "bounce in and out" of the A_n) $\left\{\limsup_n A_n\right\} = \left\{\omega \in \Omega : \omega \in A_n \text{ i.o.}\right\}$. Similarly, the lim inf of A_n is the set of events that eventually appear in an A_n and then in *all* A_n past that point. $\left\{\liminf_n A_n\right\} = \left\{\omega\in\Omega : \omega \in A_n \text{ ev.}\right\}$. Clearly, if an event is in the lim inf, it also appears infinitely often (because it appears in all the A_n past a point, and so $\liminf_n A_n \subseteq \limsup_n A_n$.

    Remark: Take $\left\{A_n, \text{ev.}\right\} = \left\{\liminf_n A_n\right\}$. Then $\left\{A_n, \text{ ev.}\right\}^c = \left[\bigcup_m \bigcap_{n\ge m} A_n\right]^c = \bigcap_m \left\{\left(\bigcap_{n\ge m} A_n\right)^c\right\} = \bigcap_m \bigcup_{n \ge m} A_n^c = \liminf_n A_n^c = \left\{A_n^c, \text{ i.o.}\right\}$

  - ***Proposition***: Let $\{A_n\}$ be a sequence of measurable sets, then $\mathbb{P}\left(\liminf_n A_n\right) \le \liminf_n \mathbb{P}\left(A_n\right)$ and $\mathbb{P}\left(\limsup_n A_n\right) \ge \limsup_n \mathbb{P}\left(A_n\right)$. (The first statement can be thought of as a generalization of Fatou's Lemma for probabilities).

    Proof of (i): Let us define $B_m = \bigcap_{n\ge m} A_n$. Then we know that $\cdots \subseteq B_m \subseteq B_{m+1} \subseteq \cdots$. In other words, the sets B_m increase monotonically to

    $$B = \bigcup_m B_m = \bigcup_m \bigcap_{n \ge m} A_n = \liminf_n A_n$$

    Since the events are increasing, a simple form of monotone convergence gives us that $\mathbb{P}\left(B_n\right) \nearrow \mathbb{P}\left(B\right)$. But we also have that $\mathbb{P}\left(B_m\right) \le \mathbb{P}\left(A_n\right) \qquad n \ge m$ because B_m is the intersection of all events from A_m onwards, so its probability is less than or equal to the probability of any single event. Thus $\mathbb{P}\left(B_m\right) \le \inf_{n\ge m} \mathbb{P}\left(A_n\right) \Rightarrow \sup_m \mathbb{P}\left(B_m\right) \le \sup_m \inf_{n \ge m} \mathbb{P}\left(A_n\right) \Rightarrow \mathbb{P}\left(B\right) \le \liminf_n \mathbb{P}\left(A_n\right)$. And given how we have defined B, $\mathbb{P}\left(\liminf_n A_n\right) \le \liminf_n \mathbb{P}\left(A_n\right)$ As required.

- *Borel-Cantelli Lemmas & Independence*

  - ***Proposition (First Borel-Cantelli Lemma)*** Let $\{A_n\}$ be a sequence of measurable events such that $\sum_n \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}\left(A_n \text{ i.o.}\right) = 0$. In other words, $\mathbb{P}\left(\limsup_n A_n\right) = 0$.[1] We offer two proofs – the first is somewhat mechanical, the second is more intuitive.

    Proof (version 1): Consider that $\limsup_n A_n = \bigcap_m \bigcup_{n\ge m} A_n$. As such $\mathbb{P}\left(\limsup_n A_n\right) \le \mathbb{P}\left(\bigcup_{n\ge m} A_n\right) \le \sum_{n\ge m} \mathbb{P}\left(A_n\right) \to 0$ As required. ■

    The second proof will require a lemma:

    Lemma (Fubini's): If $\{X_n\}$ is a sequence of real-valued random variables with $X_n \ge 0$, then $\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}\left(X_n\right)$ (which could be infinite). This is effectively a condition under which we can exchange a summation and an integration.

    Proof (version 2): Let $X_n(\omega) = \mathbb{I}_{A_n} = \begin{cases} 1 & \text{if } \omega \in A_n \\ 0 & \text{otherwise} \end{cases}$. Note that $\mathbb{E}\left(X_n\right) = \mathbb{P}\left(A_n\right)$. We also have that $\sum_n \mathbb{P}\left(A_n\right) = \sum_n \mathbb{E}\left(\mathbb{I}_{A_n}\right) = \mathbb{E}\left(\sum_n \mathbb{I}_{A_n}\right) < \infty$. This means that the random variable $\sum_n \mathbb{I}_{A_n}$ must be less than or equal to infinity for every outcome in Ω; in other words $\sum_n \mathbb{I}_{A_n} < \infty$ a.s. This means that $\mathbb{I}_{A_n} = 1$ only finitely many times. So, A_n occurs only finitely often

  - ***Definition (independence)***: A sequence of random variables $\{X_1, \cdots, X_n\}$ are independent iff $\mathbb{P}\left(X_1 \le x_1, \cdots, X_n \le x_n\right) = \prod_{i=1}^n \mathbb{P}\left(X_i \le x_i\right)$ for all $x_1, \cdots, x_n \in \mathbb{R}$.

[1] It might not be clear that the statements $\mathbb{P}\left(A_n \text{ i.o.}\right) = 0$ and $\mathbb{P}\left(\limsup_n A_n\right) = 0$ are equivalent. To see this more clearly, note that the first statement is simply a shorthand for $\mathbb{P}\left(\omega \in \Omega : \omega \in A_n \text{ i.o.}\right) = 0$. This is clearly identical, by definition, to $\mathbb{P}\left(\limsup_n A_n\right) = 0$.

  Defininition (independence): $\{A_1, \cdots, A_n\}$ are said to be *independent* if and only if the sequence or random variables $\left\{\mathbb{I}_{A_1}, \cdots, \mathbb{I}_{A_n}\right\}$ are independent.

  - ***Proposition (second Borrel-Cantelli Lemma)***: Let $\{A_n\}$ be a sequence of independent measurable events such that $\sum_n \mathbb{P}\left(A_n\right) = \infty$, then $\mathbb{P}\left(A_n \text{ i.o.}\right) = 1$.

    Proof: We showed in a remark above that $\left\{A_n, \text{i.o}\right\}^c = \left\{A_n^c, \text{ev.}\right\} = \bigcup_m \bigcap_{n\ge m} A_n^c$. Let us fix $m \ge 1$. Then using $1 - x \le e^{-x}$, we get $\mathbb{P}\left(\bigcap_{n\ge m} A_n^c\right) \quad = \prod_{n \ge m} \mathbb{P}\left(A_n^c\right)$ $= \prod_{n\ge m} 1 - \mathbb{P}\left(A_n\right) \le \prod_{n\ge m} \exp\left(-\mathbb{P}(A_n)\right) = \exp\left(-\sum_{n\ge m} \mathbb{P}(A_n)\right) = 0$. But we have that $\mathbb{P}\left(\bigcup_m \bigcap_{n\ge m} A_n^c\right) \le \sum_m \mathbb{P}\left(\bigcap_{n \ge m} A_n^c\right) = 0$. As required. ■

- *Notions of convergence*

  - ***Definition (convergence almost surely)*** $X_n \to X$ almost surely if $\mathbb{P}\left\{\omega \in \Omega : X_n(\omega) \to X(\omega)\right\} = 1$. More generally, $\{X_n\}$ converges almost surely if $\mathbb{P}\left(\limsup_n X_n = \liminf_n X_n\right) = 1$
  - ***Example***: Consider the sequence $X_n = \frac{1}{n} U[0,1]$. We claim that $X_n \to 0$ almost surely.

    Proof: In this case, $\Omega = [0,1]$. For any ω we might drawn, we will find $0 \le X_n(\omega) \le \frac{1}{n} \to 0$. As required. ■ □
  - ***Definition (convergence in probability)***: $X_n \to_p X$ as $n \to \infty$ if, for all $\varepsilon > 0$. $\mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) \to 0 \qquad$ as $n \to \infty$
  - ***Definition (convergence in expectation)*** – also called L_1 convergence. We say that $X_n \to_{L_1} X$ as $n \to \infty$ if $\mathbb{E}\left|X_n - X\right| \to 0 \qquad$ as $n \to \infty$.
  - ***Claim***: $X_n \to_{L_1} X$ implies that $\mathbb{E}\left|X_n\right| \to \mathbb{E}\left|X\right|$ and $\mathbb{E}(X_n) \to \mathbb{E}(X)$.
  - ***Theorem (Markov's Inequality)***: Let $X \ge 0$ have finite expectation. Then for all $a > 0$ $\mathbb{P}\left(X > a\right) \le \frac{\mathbb{E}(X)}{a}$ ***Proof***: $\mathbb{E}(X) = \mathbb{E}\left(X, X > a\right) + \mathbb{E}\left(X, X \le a\right) \ge \mathbb{E}\left(X, X > a\right) \ge a\mathbb{P}\left(X > a\right)$
  - ***Claim***: If $X_n \to_{L_1} X$ then $X_n \to_p X$.

    Proof: Fix $\varepsilon > 0$, use Markov's Inequality $\mathbb{P}\left(\left|X_n - X\right| > \varepsilon\right) \le \mathbb{E}\left(\left|X_n - X\right|\right) / \varepsilon \to 0$
  - ***Example***: Let X_n be a sequence of IID random variables with exponential distribution with parameter 1. We will prove that $\limsup(X_n / \log n) = 1$ a.s.

    Proof: Take $\varepsilon > 0$. Consider $\mathbb{P}\left(X_n > (1+\varepsilon)\log n\right) = e^{-(1+\varepsilon)\log n} = 1/n^{1+\varepsilon}$. Thus $\sum_n \mathbb{P}\left(X_n > (1+\varepsilon)\log n\right) < \infty$. By BC-1, $\mathbb{P}\left(X_n > (1+\varepsilon)\log n, \text{i.o.}\right) = 0$. So, $X_n \le (1+\varepsilon)\log n$ ev. a.s. and so $\boxed{\limsup_n \frac{X_n}{\log n} \le 1 + \varepsilon}$ Similarly $\mathbb{P}\left(X_n > (1-\varepsilon)\log n\right) = \frac{1}{n^{1-\varepsilon}} \Rightarrow \sum_n \mathbb{P}\left(X_n > (1-\varepsilon)\log n\right) = \infty$, so by BC-2 and independence of the X_n, $\mathbb{P}\left(X_n > (1-\varepsilon)\log n, \text{ i.o.}\right) = 1 \Rightarrow \boxed{\limsup_n \frac{X_n}{\log n} > 1 - \varepsilon \quad \text{a.s.}}$

- *Interchange arguments*

  - ***Theorem (BDD – Bounded Convergence)***: Let $\{X_n\}$ be a sequence of random variables such that $\left|X_n\right| \le K \le \infty$ (where K is a deterministic constant) and $X_n \to X$ almost surely, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.
  - ***Lemma (Fatou)***: If $\{X_n\}$ is a sequence of non-negative random variables, then $\mathbb{E}\left(\liminf_n X_n\right) \le \liminf_n \mathbb{E}(X_n)$. (Note: there is no converse for lim sup).
  - ***Theorem (DOM – Dominated Convergence)***: Let $\{X_n\}$ be a sequence of random variables such that $\left|X_n\right| \le Y$ with $\mathbb{E}(Y) < \infty$ and $X_n \to X$ almost surely, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.
  - ***Theorem (MON – Monotone Convergence)***: Let $\{X_n\}$ be a sequence of random variables such that $0 \le X_1 \le X_2 \le \cdots$ a.s., then (1) $X_n \nearrow X$ (poss. ∞) almost surely. (2) $\mathbb{E}(X_n) \nearrow \mathbb{E}(X)$ (poss. ∞).
  - ***Proposition (Fubini I)***: Let $\{X_n\}$ be with $X_n \ge 0$ for all n, then $\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}(X_n)$
  - ***Prop. (Fubini II)***: Let $\{X_n\}$ be a sequence with $\mathbb{E}\left(\sum_n \left|X_n\right|\right) < \infty$. Then $\mathbb{E}\left(\sum_n X_n\right) = \sum_n \mathbb{E}(X_n)$
  - ***Definition (u.i)***: A sequence of random variables $\{X_n\}$ is said to be *uniformly integrable* (u.i.) if for all $\varepsilon > 0$, there exists a $K(\varepsilon) < \infty$ such that $\sup_n \mathbb{E}\left(\left|X_n\right| : \left|X_n\right| > K(\varepsilon)\right) \equiv \sup_n \mathbb{E}\left(\left|X_n\right| \mathbb{I}_{\{|X_n| > K(\varepsilon)\}}\right) \le \varepsilon$
  - ***Proposition***: Let $\{X_n\}$ be a sequence of random variables and suppose $X_n \to X$ almost surely (1) If $\{X_n\}$ are uniformly integrable, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$. In fact, a stronger result applies – that $\mathbb{E}\left(\left|X_n - X\right|\right) \to 0$. (2) If $X_n > 0$ for all n and $\mathbb{E}(X_n) \to \mathbb{E}(X) < \infty$, then $\{X_n\}$ are u.i.
  - ***Prop (Sufficient conditions for u.i.)*** (1) If $\sup_n \mathbb{E}\left|X_n\right|^p < \infty$ for some $p > 1$, then $\{X_n\}$ is uniformly integrable. (2) If $\left|X_n\right| \le Y_n$ and $\{Y_n\}$ is uniformly integrable, then $\{X_n\}$ is u.i. (3) If $\{X_n\}$ and $\{Y_n\}$ are uniformly integrable, then $\{X_n + Y_n\}$ is u.i.

- *Kolmogorov's 3-Series Theorem*

  - ***Theorem (Kolmogorov 3-series)***: Let $\{X_n\}$ be a sequence of *independent* random variables. Then $\sum_n X_n$ converges **if and only if** for some (and therefore every) $K > 0$, the following conditions hold: (1) $\sum_n \mathbb{P}\left(\left|X_n\right| > K\right) < \infty$ (clearly, for example, if the X_n are IID without bounded support, we'll never be able to satisfy this). This is a statement that the "mass in the tails" must decrease with n. (2) $\sum_n \mathbb{E}\left(X_n; \left|X_n\right| \le K\right) < \infty$ (3) $\sum_n \mathrm{Var}\left(X_n; \left|X_n\right| \le K\right) < \infty$. ***Note*** that there are no issues of *existence* of the expected value and the variance in the last two points because these are taken over a *finite* range – namely on $X_n \in (-K, K)$
  - ***Proposition (2nd version of 3-series theorem)***: Let $\{X_n\}$ be independent random variables with $\mathbb{E}(X_n) = 0$ for all n. If $\sum_n \mathrm{Var}(X_n) < \infty$, then $S_n = \sum_{m=1}^n X_m$ converges almost surely.
  - ***Example***: Set $X_n = Y_n / n$, where the Y_n are IID exponential random variables with mean 1. Does the sum $S_n = \sum_n X_n$ converge? The deterministic series $1/n$ does *not* converge, and we wonder whether the exponential variables will be close enough to 0 often enough to "fix" that. In light of the result we derived in the previous lecture (that draws from an exponential will be arbitrarily large infinitely often) one might expect this series *not* to converge. *First condition*: $\mathbb{P}\left(\left|X_n\right| > K\right) = \mathbb{P}\left(\frac{Y_n}{n} > K\right) = \mathbb{P}\left(Y_n > nK\right) = e^{-nK}$. This does indeed sum to a finite number. The first condition is therefore met. *Second condition*: $\mathbb{E}\left(X_n; \left|X_n\right| \le K\right) = \mathbb{E}\left(\frac{Y_n}{n}; Y_n \le nK\right) = \frac{1}{n}\mathbb{E}\left(Y_n; Y_n \le nK\right)$. But we note that $\mathbb{E}\left(Y_1; Y_1 \le s\right) = \int_0^s y e^{-y}\, dy > 0 \qquad \forall s > 0$. It is clear, therefore, that condition 2 is violated, because the sum of $1/n$ diverges. So, as we expected, S_n does not converge almost surely.
  - ***Example***: Set $X_n = Y_n/n$, where the Y_n are IID random variables with $\mathbb{P}(Y_n = 1) = 1 - \mathbb{P}(Y_n = -1) = p$. Once again, we wonder whether $S_n = \sum_{m=1}^n X_m$ converges. The intuition here is that the deterministic series $\sum_n (-1)^n / n$ *does* converge. Our version is this sum, but instead of deterministically flipping between positive and negative, we *randomly* switch between positive and negative. We wonder whether this will "spoil" the convergence. We fix $K > 0$ (say $K = 1$) and test the three conditions: *First condition* $\mathbb{P}\left(\left|X_n\right| > 1\right) = 0$ So clearly, the sum does converge. *Second condition* $\mathbb{E}\left(X_n; \left|X_n\right| \le 1\right) = \frac{1}{n}\mathbb{E}\left(Y_n; \left|Y_n\right| \le n\right) = \frac{1}{n}\mathbb{E}(Y_n) = \frac{2p-1}{n}$. We know, however, that $\sum \frac{k}{n}$ diverges, so the infinite sum of these expectations *only* converges if p = ½, in which case the numerator is 0. We therefore restrict our attention to that case. *Third condition* (*assuming* p = ½) $\mathrm{Var}\left(X_n; \left|X_n\right| \le 1\right) = \frac{1}{n^2}\mathrm{Var}\left(Y_n \mid \left|Y_n\right| \le n\right) = \frac{1}{n^2}\mathrm{Var}(Y_n) = \frac{\mathbb{E}(Y^2) - 0}{n^2}$ The infinite sum of the variances is therefore finite.
  - Example: Set $X_n = Y_n / \sqrt{n}$, where Y_n are IID random variables with probability ½ of being 1 and probably ½ of being –1. Again, the deterministic sum $\sum_n (-1)^n / \sqrt{n}$ converges, and we wonder whether adding randomness will make a difference. Since, in this case, $\mathbb{E}(X_n) = 0$, we can simply apply the second form of the 3-series theorem: $\mathrm{Var}\left(X_n; \left|X_n\right| \le 1\right) = \frac{1}{n}\mathrm{Var}\left(X_n; \left|Y_n\right| \le n\right) = \frac{\mathbb{E}(X^2)}{n}$ The infinite sum of these variances clearly diverges, and so S_n does *not* converge.

- *The Strong Law of Large numbers*

  - ***Theorem (SLLN)***: Let $\{X_n\}$ be IID with $\mathbb{E}\left|X_1\right| < \infty$, then $\frac{1}{n}\sum_{k=1}^n X_k \to \mathbb{E}(X_1) \qquad$ a.s.
  - ***(Cesaro Sum Property)***: $\{a_n\}$ be real valued sequence $a_n \to a_\infty$. Then $\frac{1}{n}\sum_{k=1}^n a_k \to_{n\to\infty} a_\infty$
  - ***(Kronecker's Lemma)***: $\{a_n\}$ real valued sequence. Then $\left(\sum_{k=1}^\infty \frac{a_k}{k} \text{ converges}\right) \Rightarrow \left(\frac{1}{n}\sum_{k=1}^n a_k \to 0\right)$
  - ***Example***: Let $\{X_n\}$ be an IID sequence of random variables with $\mathbb{E}(X_n) = 0$ and $\mathbb{E}\left(X_n^2\right) = \sigma^2 < \infty$. Let $S_n = \sum_{k=1}^n X_n$, and consider S_n / a_n with $a_n = \sqrt{n}\left(\log n\right)^{\frac{1}{2}+\varepsilon}$. log n grows extremely slowly, so this is really very close to a $\sqrt{n}$ growth. Consider that $\mathrm{Var}(X_n / a_n) = \sigma^2 / \sqrt{n}(\log n)^{(1/2)+\varepsilon}$, and this implies that $\sum_n \mathrm{Var}\left(\frac{X_n}{a_n}\right) < \infty$ Thus, by the Kolmogorov 3-series theorem, $\sum_n \frac{X_n}{a_n}$ converges almost surely, and by Kronecker's Lemma, this means that $S_n / a_n \to 0$. Thus, we see that even with such a small-growing a_n, an almost-sure result still holds. However, it is interesting to note that if we take $a_n = \sqrt{n}$ (slightly smaller), then *no almost sure result* holds anymore. The best we can say is $S_n / \sqrt{n} \to \sigma N(0,1)$.

- *Weak convergence & the Central Limit Theorem*

  - ***Definition*** (***Weak convergence***): A sequence of random variables $\{X_n\}$ is said to converge weakly to a random variable X, denoted $X_n \Rightarrow X$ as $n \to \infty$, if and only if $\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right]$ for all bounded continuous functions f. ***Comments***: If $X_n \to X$ almost surely, then $f(X_n) \to f(X)$, because f is continuous, and $\mathbb{E}\left[f(X_n)\right] \to \mathbb{E}\left[f(X)\right]$, by bounded convergence (since the functions are bounded). Thus, almost sure convergence implies weak convergence.
  - ***Definition (Weak convergence)***: A sequence of *real valued* random variables $\{X_n\}$ converges weakly to X if and only if $F_n(x) \to F(x)$ for all continuity points of F, where $X_n \sim F_n$ and $X \sim F$.
  - ***Definition (Tightness)***: A sequence of random variables $\{X_n\}$ with corresponding distribution functions $\{F_n\}$ is said to be *tight* if $\forall\varepsilon > 0$, $\exists K(\varepsilon) < \infty$ s.t. $\sup_n \mathbb{P}\left\{\left|X_n\right| > K(\varepsilon)\right\} \le \varepsilon$. Equivalently $\sup_n \left|1 - F_n\left(K(\varepsilon)\right)\right| \le \varepsilon$. ***Comment***: A sufficient condition for tightness if $\sup_n \mathbb{E}\left|X_n\right| < \infty$. To see why, with $K > 0$, and consider that by Markov's Inequality $\mathbb{P}\left(\left|X_n\right| > K\right) \le \frac{\mathbb{E}\left|X_n\right|}{K} \le \frac{\sup_n \mathbb{E}\left|X_n\right|}{K}$
  - ***Proposition***: Let $\{X_n\}$ be a sequence of random variables (1) If $X_n \Rightarrow X$, then $\{X_n\}$ is tight (2) If $\{X_n\}$ is tight, then there exists a subsequence $\{n_k\}$ such that $\{X_{n_k}\}$ converges weakly. This is somewhat isomorphic to the concept of compactness in real analysis.
  - ***Definition (Characteristic Function)***: A characteristic function (CF) of a random variable X is given by $\varphi_X(\theta) = \mathbb{E}\left[\exp(i\theta X)\right] \qquad \theta \in \mathbb{R}$ ***Remarks*** (1) $\varphi_X(0) = 1$ (2) $\left|\varphi_X(\theta)\right| \le 1$, because by Jensen's inequality, $\left|\varphi_X(\theta)\right| \le \left\{\mathbb{E}\left|\exp(i\theta X)\right|\right\} = 1$ (3) $\varphi_X^{(n)}(0) = i^n \mathbb{E}(X^n)$, provided $\mathbb{E}\left|X\right|^n < \infty$. (4) If X and Y are independent, $\varphi_{X+Y}(\theta) = \varphi_X(\theta)\varphi_Y(\theta)$, though the converse is false.
  - ***Proposition (Levy characterization theorem)***: Let $\{X_n\}$ be a sequence of random variables with distribution functions $\{F_n\}$. If $\varphi_{X_n}(\theta) \to \varphi(\theta)$ for all $\theta \in \mathbb{R}$ and $\varphi(\theta)$ is continuous at 0 [kind of a tightness condition], Then $X_n \Rightarrow X$ and X has characteristic function $\varphi(\theta)$.

- *Odds & Ends*

  - ***Prop. (Skorohod Representation)***: Let $\{X_n\}$ be a sequence of random variables such that $X_n \Rightarrow X$. Then there exists a probability space $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ supporting a sequence $\{X_n'\}$ and X' such that $X_n' \overset{d}{=} X_n$ and $X' \overset{d}{=} X$ And $X_n' \to X'$ almost surely.
  - ***Prop. (Continuous Mapping Theorem – CMT)***: Let $\{X_n\}$ be such that $X_n \Rightarrow X$. Let f be a function with $\mathbb{P}\left(X \in D_f\right) = 0$ with D_f being the set of discontinuity points of f. Then $f(X_n) \Rightarrow f(X)$
  - ***Proposition (Converging Together Lemma*** – CTL): If $\{X_n\}$ is a sequence such that $X_n \Rightarrow X$ and $\{Y_n\}$ is a sequence such that $Y_n \Rightarrow a$ (where a is a deterministic constant), then (1) $X_n + Y_n \Rightarrow X + a$ (2) $X_n Y_n \Rightarrow Xa$ (3) $\frac{X_n}{Y_n} \Rightarrow \frac{1}{a} X$ provided $a \ne 0$.

- *Introduction to Large Deviations*

  - ***Theorem***: Let $X_1, X_2, \cdots$ be IID with mean μ and such that $M(\theta) = \mathbb{E}(e^{\theta X_1})$ exists for all $\theta \in \mathbb{R}$, and assume $\mathbb{P}(X_1 = \mu) < 1$. Then for any $a > \mu$, $\frac{1}{n}\log\mathbb{P}\left(S_n > na\right) \to -I(a)$ Where $I(a) = \sup_{\theta\in\mathbb{R}}\left[\theta a - \log M(\theta)\right] = \sup_{\theta\in\mathbb{R}}\left[\theta a - \varphi(\theta)\right]$
  - ***Example***: Let $X_1, X_2, \cdots$ be IID $N(\mu, \sigma^2)$ variables variables. Fix $a > \mu$. We saw, earlier that $\mathbb{P}\left(S_n > na\right) \le \exp\left(-nI(a)\right) = \exp\left(-n\frac{(a-\mu)^2}{2\sigma^2}\right)$. The central limit theorem allows us to "handwave" $S_n \approx n\mu + \sigma\sqrt{n}N(0,1)$. Using this result, however, we can be more exact – we can find a sequence a_n such that S_n eventually lies below $n\mu + a_n$ a.s. Consider: $\mathbb{P}\left(S_n > n\mu + a_n\right) \le \exp\left(-n\frac{(a_n/n)^2}{2\sigma^2}\right)$. For summability, choose $a_n = \sqrt{2(1+\delta)\sigma^2 n \log n}$. This is very close to $\sqrt{n}$, which we might expect to work given the CLT. $\mathbb{P}\left(S_n > n\mu + a_n\right) \le n^{-(1+\delta)}$. Use BC-1.
  - ***Theorem (Moderate Deviations)***: Under the conditions of Cramer's Theorem, for any sequence α_n such that (1) $\frac{\alpha_n}{\sqrt{n}} \to \infty$ (2) $\frac{\alpha_n}{n} \to 0$, then for any $a > 0$, $\frac{n}{\alpha_n^2}\log\mathbb{P}\left(S_n > n\mu + a\alpha_n\right) \to -\frac{a^2}{2\sigma^2}$

Random Walks & Martingales

- *Random walks*

  - ***Definition*** RW is $S_n = \sum_{i=1}^n X_i$, with the $\{X_n\}$ IID, and independent of S_0.
  - ***Definition*** (Stopping time): Suppose T is a non-negative integer-valued random variable. Then T is said to be a *stopping time* with respect to an underlying sequence $\{X_n\}$ if, $\{T = k\} \in \mathcal{F}_k \qquad \forall k$
  - ***Proposition*** (Wald's First Identity): Let S_n be a random walk $S_n = \sum_{i=1}^n X_i$, with $S_0 = 0$, and let T be a stopping time w.r.t. the sequence $\{\mathcal{F}_n\}$ (1) $X_i \ge 0 \Rightarrow \mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$ (2) If $\mathbb{E}\left|X_i\right| < \infty, \mathbb{E}(T) < \infty \Rightarrow \mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_0)$

    Proof: Let $S_T = \sum_{i=1}^T X_i = \sum_{i=1}^\infty X_i \mathbb{I}_{\{i \le T\}}$. Then $\mathbb{E}(S_T) = \mathbb{E}\left(\sum_{i=1}^\infty X_i \mathbb{I}_{\{i\le T\}}\right)$

    Now, let us do both parts:

    - ***First part***: $X_i \ge 0$, and indicators are always positive, so by Fubini I, we can interchange the expectation and the sum $\mathbb{E}(S_T) = \sum_{i=1}^\infty \mathbb{E}\left(X_i \mathbb{I}_{\{i\le T\}}\right)$. Consider, however, that $\mathbb{I}_{\{T\ge i\}} = 1 - \mathbb{I}_{\{T < i\}} = 1 - \mathbb{I}_{\{T \le i-1\}}$. So $\mathbb{I}_{\{T \ge i\}} \in \mathcal{F}_{i-1}$. Thus $\mathbb{E}(S_T) = \sum_{i=1}^\infty \mathbb{E}\left[\mathbb{E}\left(X_i \mathbb{I}_{\{i\le T\}} \mid \mathcal{F}_{i-1}\right)\right] = \sum_{i=1}^\infty \mathbb{E}\left[\mathbb{I}_{\{i\le T\}}\mathbb{E}\left(X_i \mid \mathcal{F}_{i-1}\right)\right]$ $= \sum_{i=1}^\infty \mathbb{E}\left[\mathbb{I}_{\{i\le T\}}\mathbb{E}\left(X_1\right)\right] = \mathbb{E}(X_1)\sum_{i=1}^\infty \mathbb{E}\left(\mathbb{I}_{\{i\le T\}}\right)$ Using Fubini I again $\mathbb{E}(S_T) = \mathbb{E}(X_1)\mathbb{E}\left(\sum_{i=1}^\infty \mathbb{I}_{\{i\le T\}}\right) = \mathbb{E}(X_1)\mathbb{E}(T)$
    - ***Second part***: Triangle inequality $\left|S_T\right| \le \sum_{i=1}^T \left|X_i\right|$. By part 1, however, $\mathbb{E}\left(\sum_{i=1}^T \left|X_i\right|\right) = \mathbb{E}(T)\mathbb{E}\left|X_1\right| < \infty$. The result then follows by Fubini II.

  - ***Proposition*** (Wald II): Suppose $\{X_i\}$ are IID *bounded* random variables, and $\mathbb{E}(X_1) = 0$, $\mathbb{E}(X_1^2) = \sigma^2$. Let T be a stopping time with respect to $\{\mathcal{F}_n\}$, and such that $\mathbb{E}(T^2) < \infty$. Then $\mathrm{Var}(S_T) = \sigma^2\mathbb{E}(T)$

- *Martingales*

  - Consider that if we let $D_n = M_n - M_{n-1}$, we can write $M_n = \sum_{i=1}^n D_i + M_0$.

- *Optional Stopping Theorem for Martingales*

  $$\boxed{\mathbb{E}\left[\lim_{n\to\infty} M_{T\wedge n}\right] = \lim_{n\to\infty} \mathbb{E}\left[M_{T\wedge n}\right]}$$

  - ***Corollary I***: Let $\left(M_n : n \ge 0\right)$ be a martingale with respect to $\{\mathcal{F}_n\}$ and T be a stopping time such that T is bounded (in other words, there exists a $K < \infty$ such that $\mathbb{P}\left(T < K\right) = 1$), then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.
  - ***Corollary II***: Let $\left|M_{T\wedge n}\right| \le Z$, with $\mathbb{E}(Z) < \infty$. Then by dominated convergence, the interchange holds and $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.
  - ***Corollary III***: Let $\left(M_n : n \ge 0\right)$ be a martingale with respect to $\{\mathcal{F}_n\}$ and T be a stopping time such that $\mathbb{E}(T) < \infty$. Provided the martingale differences are uniformly bounded ($\mathbb{E}\left[\left|D_i\right| \mid \mathcal{F}_{i-1}\right] \le C < \infty$), then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

    Proof: We have $M_{T\wedge n} = \sum_{i=1}^{T\wedge n} D_i + M_0$. M_0 is integrable. Now consider that, $\left|\sum_{i=1}^{T\wedge n} D_i\right| \le \sum_{i=1}^{T\wedge n}\left|D_i\right| = \sum_{i=1}^\infty \left|D_i\right|\mathbb{I}_{\{T\ge i\}}$. Let us now take expectations. By Fubini I, we can then swap the expectation and sum, since the summands are positive

    $$\mathbb{E}\sum_{i=1}^\infty \left|D_i\right|\mathbb{I}_{\{T\ge i\}} = \sum_{i=1}^\infty \mathbb{E}\left[\left|D_i\right|\mathbb{I}_{\{T\ge i\}}\right] = \sum_{i=1}^\infty \mathbb{E}\left[\mathbb{I}_{\{T\ge i\}}\mathbb{E}\left[\left|D_i\right| \mid \mathcal{F}_{n-1}\right]\right]$$

    Since the martingale differences are uniformly bounded bt C, this is $\le C\mathbb{E}(T)$. As such $\left|M_{T\wedge n}\right| \le C\mathbb{E}(T) + \left|M_0\right|$ M_0 is integrable, and by the statement of the theorem, $\mathbb{E}(T) < \infty$. As such, martingale bounded by integrable variable.

- *Martingale Limit Theory & Concentration Inequalities*

  - ***Proposition (Martingale SLLN)***: Let $\{M_n : n \ge 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ such that $\sup_{n\ge1}\mathbb{E}(D_n^2) < \infty$ (this, as we saw above, is a requirement for the increments to be uncorrelated). Then $\frac{M_n}{n} \to 0$ a.s.

    Proof: Put $\tilde{M}_n = \sum_{k=1}^n \frac{D_k}{k}$ (where, as usual $D_k = M_k - M_{k-1}$). Clearly, $\tilde{M}_n \in \mathcal{F}_n$, and $\mathbb{E}\left[\tilde{M}_n \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\sum_{k=1}^{n-1}\frac{D_k}{k} + \frac{D_n}{n} \mid \mathcal{F}_{n-1}\right] = \sum_{k=1}^{n-1}\frac{D_k}{k} + \frac{1}{n}\mathbb{E}\left[D_n \mid \mathcal{F}_{n-1}\right] = \tilde{M}_{n-1}$ and $\mathbb{E}\left|\tilde{M}_n\right| < \infty$ for all n, due to the conditions on the second moments of D_n. Thus, it is a martingale. Now $\mathbb{E}\left|\tilde{M}_n\right|^2 = \mathbb{E}\left[\sum_{k=1}^n \frac{D_k^2}{k^2}\right] + \mathbb{E}\left[\sum_{i\ne j}^n \frac{D_i D_j}{ij}\right]$ The second term is equal to 0, since the D are uncorrelated. $\mathbb{E}\left|\tilde{M}_n\right|^2 \le \sum_{k=1}^n \frac{\sup_{n\ge1}\mathbb{E}(D_k^2)}{k^2} < \infty$. This implies that $\sup_{n\ge1}\mathbb{E}\left|\tilde{M}_n\right|^2 < \infty \Rightarrow \sup_{n\ge1}\mathbb{E}\left|\tilde{M}_n\right| < \infty$. If so, by the Martingale Convergence Theorem, $\tilde{M}_n \to \tilde{M}_\infty$ a.s.. Now, let $A = \left\{\omega : \sum_{k=1}^n \frac{D_k(\omega)}{k} \to M_\infty(\omega)\right\}$, $\mathbb{P}(A) = 1$ $B = \left\{\omega : \frac{1}{n}\sum_{k=1}^n D_k(\omega) \to 0\right\}$. By Kroenecker's Lemma $A \subseteq B$. Thus, $\mathbb{P}(B) = 1$. This proves our Theorem. ■
  - ***Proposition (Central Limit Theorem for Martingales)***: Let $\{M_n : n \ge 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ and put $V_n = \max_{1\le i\le n}\left|D_i\right|$. If (1) $\sup_n \frac{\mathbb{E}(V_n^2)}{n} < \infty$ (2) $\left|V_n\right| / \sqrt{n} \Rightarrow 0$ (3) $\frac{1}{n}\sum_{i=1}^n D_i^2 \Rightarrow \sigma^2$ (deterministic and finite). This makes our martingale "similar" to a random walk. Then $\frac{1}{\sqrt{n}} M_n \Rightarrow \sigma N(0,1)$
  - ***Proposition (Azuma-Hoeffding Inequality)***: Let $\{M_n : n \ge 0\}$ be a martingale with respect to $\{\mathcal{F}_n\}$ such that $\left|M_n - M_{n-1}\right| = \left|D_n\right| \le c_n \ \forall n$. Then, for all $n > 0, \lambda > 0$, $\mathbb{P}\left(\left|M_n - M_0\right| \ge \lambda\right) \le 2\exp\left(-\frac{\lambda^2}{2\sum_{k=1}^n c_k^2}\right)$

    Remark: Suppose that $\left|D_n\right| \le c \ \forall k$, then we can re-write this as $\mathbb{P}\left(\left|M_n - M_0\right| \ge \lambda\right) \le 2\exp\left\{-\frac{\lambda^2}{2c^2 n}\right\}$ or $\mathbb{P}\left(\left|M_n - M_0\right| \ge xc\sqrt{n}\right) \le 2\exp\left\{-\frac{x^2}{2}\right\}$. Or, choosing $\lambda = \sqrt{\kappa n \log n}$, we obtain $\mathbb{P}\left(\left|M_n - M_0\right| \ge \sqrt{\kappa n \log n}\right) \le 2\exp\left(-\frac{\kappa \log n}{2c}\right)$. Choosing, for example, $\kappa = 4c$ produces a summable sequence, which can be used to obtain an almost sure result.

- *Stochastic stability*

  - ***Proposition***: Let $\{X_n : n \ge 0\}$ be an irreducible markov chain on a countable state space $\mathcal{S}^k$. Let $K \subseteq \mathcal{S}$ be a set containing a finite number of states. Then, if there exists a function $g : \mathcal{S} \to \mathbb{R}_+$ such that (recall $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot \mid X_0 = x]$) $\mathbb{E}_x g(X_1) - g(x) \le -\varepsilon \quad \forall x \in \bar{\mathcal{K}}$ and some $\varepsilon > 0$ and $\mathbb{E}_x g(X_1) < \infty \quad \forall x \in \mathcal{K}$ Then $(X_n, n \ge 0)$ is positive recurrent

- "***Application***": Consider the stochastic system $X_{n+1} = \alpha X_n + Z_{n+1}$ with $\{Z_n\}$ iid and $X_0 = x$. We want conditions on α and distributions of Z such that X_n is stable. Use $g(x) = |x|$ and $\mathbb{E}_x g(X_1) = \mathbb{E}_x |\alpha x + Z_1| \le |\alpha||x| + \mathbb{E}|Z_1|$. So if $\mathbb{E}|Z_1| < \infty$ and $|\alpha| < 1$, then we might be able to make this work, because we would then have $\mathbb{E}_x g(X_1) - g(x) \le (|\alpha| - 1)|x| + \mathbb{E}(Z) \le -\varepsilon$ provided $|X| \ge \frac{\mathbb{E}|Z|+\varepsilon}{1-|\alpha|}$

Martingales

- ***Definition***: $\{X_n\}$ is a sub-martingale with respect to $\{\mathcal{F}_n\}$ if (1) $X_n \in \mathcal{F}_n$ (2) $\mathbb{E}(X_n) < \infty$ (it is often convenient to work with the stronger condition $\mathbb{E}|X_n| < \infty$]. (3) $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \ge X_n$ [$\le$ gives a super-martingale, = gives a martingale]. Implies the weaker property $\mathbb{E}[X_{n+1}] \ge \mathbb{E}[X_n]$
- ***Remarks***: A convex function of a martingale is a submartingale. (2)An increasing convex function of a submartingale is a submartingale. ***Proof***: (i) and (ii) are simple. $\mathbb{E}[f(X_{n+1}) \mid \mathcal{F}_n] \ge f(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) \ge f(X_n)$.
- ***Example***: Let $S_n = \sum_{i=1}^n X_i$, where the X_i are IID with $\mathbb{E}(X_i) = 0, \mathbb{E}|X_i| < \infty$
  - S_n is a martingale [$\mathbb{E}|S_n| \le n\mathbb{E}|X_1|$] (the *mean martingale*).
  - If $\mathrm{Var}(X_i) = \sigma^2 < \infty$, $X_n^2 - \sigma^2 n$ is a martingale (the *variance martingale*). □
- ***Example (the exponential martingale)***: Let $\varphi(\theta) = \mathbb{E}(e^{\theta X_1})$. $M_n = e^{\theta S_n} / \varphi^n(\theta)$ is a martingale. For example, if $S_n = \sum_{i=1}^n X_i$ is an asymmetric random walk with $p = \mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = -1)$, then $M_n = \left(\frac{1-p}{p}\right)^{S_n}$ is an exponential martingale, with $e^\theta = \frac{1-p}{p}$ and $\varphi(\theta) = 1$. □
- ***Example***: Suppose an urn starts with one black and one white ball. We pull out balls from the urn, and return them to the urn with *another, new* ball of the same color. Y_n, the proportion of white balls after n draws, is a martingale (mean ½). □
- ***Example***: Let $\{X_n\}$ be a Markov Chain with transition matrix $P(x, y)$ and let $h(x)$ be a bounded function with $h(x) = \sum_y p(x,y)h(y)$. $\{h(X_n)\}$ is then a martingale. □

Optional stopping

- ***Definition (stopping time)***: If T is an integer valued random variable, we say it is a *stopping time* with respect to a filtration $\mathcal{F}_n$ if $\{T = n\} \in \mathcal{F}_n$ for all n (or $\{T \le t\} \in \mathcal{F}_t$ for all t, in continuous time).
  Remark: If T_1 and T_2 are stopping times, so are $T_1 + T_2, T_1 \wedge T_2, T_1 \vee T_2$.
- ***Theorem***: If $\{X_n\}$ is a (sub)martingale and T is a stopping time, $\{X_{T\wedge n}\}$ is also a (sub)martingale. If $\{X_n\}$ is u.i., so it $\{X_{T\wedge n}\}$. ***Proof***: Write $X_{T\wedge n} = \sum_{k=0}^{n-1} X_k \mathbb{I}_{\{T=k\}} + X_n \mathbb{I}_{\{T>n-1\}}$. Clearly, this is $\mathcal{F}_n$ measurable, and $|X_{T\wedge n}| \le \sum_{k=0}^{n-1} |X_k| + |X_n|$ so it is integrable. Conditioning follows. $\boxed{u.i.}$: to show u.i., first note that $\{X_{T\wedge n}^+\}$ is also a submartingale, and so $\mathbb{E}(X_{T\wedge n}^+) \le \mathbb{E}|X_n^+| \le \mathbb{E}|X_n|$. Taking limits, $\sup_n \mathbb{E}(X_{T\wedge n}^+) \le \sup_n \mathbb{E}|X_n| < \infty$ [the last inequality follows by u.i. of $\{X_n\}$]. By the theorem in the martingale convergence section, $X_{T\wedge n}^+ \to_{n\to\infty} X_T$ with $\mathbb{E}|X_T| < \infty$. Finally, consider $\mathbb{E}\left[|X_{T\wedge n}| \mathbb{I}_{X_{T\wedge n} > a}\right]$. Simply split it over $\mathbb{I}_{T \le n}$ and $\mathbb{I}_{T > n}$, drop these indicators and use integrability of X_T and u.i. of $\{X_n\}$. ■
- ***Example***: Consider a gambler's ruin with wealth S_t at time t with $S_0 = i$ and with probably ½ of going each direction at each time step. Let $\begin{array}{c} p = \mathbb{P}(\text{Probability we hit } N > i, \text{ at which point we stop}) \\ 1 - p = \mathbb{P}(\text{Probability we hit 0, at which point we're ruined}) \end{array}$ And let $T = \inf\{n : S_n = N \text{ or } S_n = 0\}$ We can now use the OST $\boxed{\text{On } S_t}$ OST says that $\mathbb{E}(S_T) = \mathbb{E}(S_0) = i$. Logic says that $\mathbb{E}(S_T) = pN + 0$. Together, we obtain $p = i / N$. $\boxed{\text{On } S_t^2 - n}$ OST says that $\mathbb{E}(S_T^2 - T) = \mathbb{E}(S_0^2) = i^2$, and so $\mathbb{E}(T) = \mathbb{E}(S_T^2) - i^2$. Logic says that $\mathbb{E}(S_T^2) = pN^2 + 0$. Together, $\mathbb{E}(T) = i(N - i)$. □
- ***Counterexample***: Consider the example above, but with $T' = \inf\{n : S_n = N > i\}$. This is well defined, in that $\mathbb{P}(T' < \infty) = 1$ (because the random walk is an irreducible Markov chain, which means every state will eventually be visited) but blindly applying the OST gives $\mathbb{E}(S_{T'}) = i$, which implies that $N = i$. Clearly, something has gone awry. □
- …***Theorem***: If $T \le n_0$ a.s. then $\mathbb{E}[X_T] = \mathbb{E}[X_{n_0}]$. ***Proof***: $\mathbb{E}[X_T] = \mathbb{E}[X_{T\wedge n_0}] = \mathbb{E}[X_0]$ ■
- Optional stopping: $\mathbb{E}[X_0] \overbrace{=}^{(1)} \lim_n \mathbb{E}[X_{T\wedge n}] \overbrace{=}^{(2)} \mathbb{E}[\lim_n X_{T\wedge n}] \overbrace{=}^{(3)} \mathbb{E}[X_T]$ [u.i. ok for eg]
- ***Theorem***: $\mathbb{E}\left[|X_{n+1} - X_n| \mid \mathcal{F}_n\right] \le C < \infty$ for $n \le T$ and $\mathbb{E}(T) < \infty$ is enough.
- ***Example (Wald's Identity)***: Let $\{X_i\}$ be IID with $\mathbb{E}|X_i| < \infty$ and $\mathbb{E}(X_i) = \mu$. Let T be a stopping time with $\mathbb{E}(T) < \infty$. Then $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$. ***Proof***: Consider $\{S_n - n\mu\}$. Applying the OST gives the required result, given that $\mathbb{E}(|S_n - S_{n-1}| \mid \mathcal{F}_n) = \mathbb{E}|X_n| < \infty$.
- ***Example***: Consider an asymmetric random walk with $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = -1)$, and let $T = \inf\{n : S_n = -a \text{ or } b\}$. We would like to find $p_b = \mathbb{P}(S_T = b)$. (1) Try the exponential martingale ρ^{S_n} (where $\rho = \frac{1-p}{p}$). Note that $0 \le M_{n\wedge T} = \rho^{S_{n\wedge T}} \le \rho^{a+b} \vee 1$; it is therefore bounded, and we can apply the OST to deduce that $1 = \mathbb{E}(\rho^0) = \mathbb{E}(\rho^T)$. Also, $\mathbb{E}(\rho^T) = \rho^b p_b + \rho^a(1 - p_b)$. This allows us to find p_b. (2) Now use the mean martingale $\{S_n - n(p - q)\}$. OST gives $\mathbb{E}[S_T - T(p - q)] = 0 \Rightarrow \mathbb{E}(S_T) = (p - q)\mathbb{E}(T)$. Using p_b, we can work out $\mathbb{E}(S_T)$ and find $\mathbb{E}(T)$.
- Note that the OST does not necessarily require $\mathbb{P}(T < \infty) = 1$. Indeed, $\mathbb{E}(X_T) = \mathbb{E}(X_T \mathbb{I}_{T<\infty} + X_\infty \mathbb{I}_{T=\infty})$, and if the stopped martingale is u.i., X_∞ must exist.
- It is important to remember that when we invoke the martingale convergence theorem so say that $X_{T\wedge n} \to_{n\to\infty} X_T$, we are implicitly implying that $\mathbb{P}(T < \infty) = 1$.

Martingale Inequalities

- ***Doob's Inequality/The Maximal Inequality***
  - ***Motivation***: Markov's Inequality states that $a\mathbb{P}(|X| > a) \le \mathbb{E}|X|$.
  - ***Theorem***: If $\{X_n\}$ is a submartingale and $A = \{\max_{0\le k\le n} X_k \ge a\}$, then $a\mathbb{P}(A) \le \mathbb{E}[X_n \mathbb{I}_A] \le \mathbb{E}[X_n^+]$
  - If we let $S_0 = 0$ and $S_n = \sum_{i=1}^n X_n$ were the X_n are IID with $\mathbb{E}[X_n] = 0, \mathrm{Var}[X_n] = \sigma^2$, then $\{S_n\}$ is a martingale, and $\{K_n\} = \{S_n^2\}$ is a submartingale. As such, we get *Kolmogorov's Inequality*:
$$\mathbb{P}\left(\max_{0\le k\le n} |S_k| \ge x\right) = \mathbb{P}\left(\max_{0\le k\le n} X_k \ge x^2\right) \le \frac{\mathbb{E}[S_n^2]}{x} = \frac{n\sigma^2}{x^2}$$
  - ***Theorem (Azuma's Inequality)***: Let $\{Z_n\}$ be a zero-mean martingale with bounded MG differences (ie: $-\alpha \le Z_i - Z_{i-1} \le \beta$ for $\alpha, \beta \ge 0$). Then
$$\mathbb{P}\left(\bigcup_{n=m}^\infty |Z_n| > n\varepsilon\right) \le 2\exp\left(-\frac{2m\varepsilon^2}{(\alpha + \beta)^2}\right)$$
    This bound is not as tight as the CLT's, but it requires less.
  - ***Definition (Doob Martingale)***: Let X be a random variable in L_1 and $\mathcal{F}_n$ be a set of filtrations. Then $X_n = \mathbb{E}(X \mid \mathcal{F}_n)$ is a martingale.
    Proof: $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}[\mathbb{E}(X \mid \mathcal{F}_{m+1}) \mid \mathcal{F}_n] = \mathbb{E}(X \mid \mathcal{F}_n) = X_n$. ■
  - Let $\boldsymbol{X} = (X_1, \cdots, X_n)$, where the X_i are independent and with CDF F_i. Define $\mathcal{F}_i = \sigma(X_1, \cdots, X_i)$. Finally, let $h : \mathbb{R}^n \to \mathbb{R}$ such that, if $\boldsymbol{x}$ differs from $\boldsymbol{y}$ in only one component, $|h(\boldsymbol{x}) - h(\boldsymbol{y})| \le \ell$, for some $\ell \ge 0$. Then $S_i = \mathbb{E}[h(\boldsymbol{X}) \mid \mathcal{F}_i]$ is a Doob martingale. Provided we can prove $|S_i - S_{i-1}| \le \ell$, we can apply Azuma's Inequality with $\alpha + \beta = \ell$ to $S_n = h(\boldsymbol{X})$ $\mathbb{P}\left(|h(\boldsymbol{X}) - \mathbb{E}[h(\boldsymbol{X})]| > n\varepsilon\right) \le 2\exp\left(-\frac{n\varepsilon^2}{2\ell^2}\right)$

Martingale Convergence

- ***Theorem***: If $\{X_n\}$ is a submartingale and $\sup_n \mathbb{E}|X_n| < \infty$ (this is a weaker condition than u.i.), then $X_n \to X$ a.s. and $\mathbb{E}|X| < \infty$. ***Remark***: $\sup_n \mathbb{E}|X_n| < \infty$ is equivalent to $\sup_n \mathbb{E}(X_n^+) < \infty$ because $x^+ \le |x| = 2x^+ - x$, and so, for example, $\mathbb{E}|X_n| = 2\mathbb{E}[X_n^+] - \mathbb{E}[X_n] \le 2\mathbb{E}[X_n^+] - \mathbb{E}[X_0]$.
  Proof: Note that $(X_n - a)^+ \le X_n^+ + |a|$ and write the upcrossing inequality as $\mathbb{E}[U_n] \le \frac{\mathbb{E}(X_n^+)+|a|}{b-a} = K < \infty$ [using the fact $|X|$ dominates X^+ and the condition in the theorem). However, U_n is increasing by definition and therefore tends to some U (possibly ∞), but by monotone convergence, $\mathbb{E}(U_n) \to \mathbb{E}(U) < \infty$. Thus, the number of up-crossings must be finite, and so $\mathbb{P}(\liminf X_n < a < b < \limsup X_n) = 0$. This is true for *any a* and *b*, and so $\mathbb{P}(\liminf X_n = \limsup X_n) = 1$. Integrability of the limit follows by Fatou.■
- ***Corollary***: If $\{X_n\}$ is a supermartingale and $X_n \ge 0$, then $X_n \to X$ and $\mathbb{E}[X] \le \mathbb{E}[X_0]$.
- ***Proof***: Let $Y_n = -X_n$ – this is a submartingale with $\mathbb{E}[Y_n^+] = 0$. The condition of the theorem above (see the remark) is therefore satisfied. Note, however, that almost sure converge *does not* imply convergence in the means or variances, as the next two examples illustrate..
- ***Example***: Assume $X_0 = i > 0$, and $X_n \mid X_{n-1} \sim \mathrm{Po}(X_{n-1})$. Clearly, this is a martingale, and once we hit 0, we stay there. Let $T = \inf\{n : X_n = 0 \text{ or } X_n \ge b\}$. By optional stopping, $\mathbb{E}(X_T) = \tilde{b}(1 - p)$, where $\tilde{b} \ge b$ and $p = \mathbb{P}(X_T = 0)$. But $\mathbb{E}(X_T) = i$. As such, $1 - p \to \frac{i}{b} \to 0$ as $b \to \infty$. Thus, $p \to 1$. How do we know stopping time is finite? Note, however, that $\mathbb{E}[X_n^2] = \mathbb{E}[\mathbb{E}(X_n^2 \mid X_{n-1})] = \mathbb{E}(X_{n-1}^2) + i$. As such, $\mathrm{Var}(X_n) = ni$; the variable itself tends to 0, but the variance blows up. □
- ***Example***: Let $X_1 \sim U[0,1]$, and $X_n \mid X_{n-1} \sim U[0, X_{n-1}]$. Let $Y_n = 2^n X_n$. We can write $X_n = U_1 \cdots U_n$ with each U IID $U[0,1]$. This is a martingale, and, by the SLLN, $\frac{1}{n}\log Y_n = \log 2 + \frac{1}{n}\sum_{i=1}^n \log U_i \to \log 2 + \mathbb{E}(\log U_1) < \infty$ a.s. So $Y_n \to 0$. Note, however, that $\mathrm{Var}(X_n) = \left(\frac{4}{3}\right)^n - 1$. Again, the variance blows up. □
- …***Theorem***: When $\{X_n\}$ is a martingale, the following are equivalent (1) X_n is u.i. (and therefore converges almost surely) (2) $X_n \to_{L_1} X$ (3) X_n can be written as a Doob martingale; $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$, with $\mathbb{E}|X| < \infty$. For submartingales, only (i) and (ii) are equivalent.
- ***Example (Polya's Urn)***: Consider the example of Polya's Urn, discussed above. Let X_n be the proportion of red balls in the urn after draw n. By symmetry, $\mathbb{E}(X_n) = \frac{1}{2}$.

Queuing

- ***Sample path methods***
  - $\omega = \{(t_n, S_n) : n \in \mathbb{Z}\}$. S_n is work, t_n is arrival time. Processing at unit rate. $W_t(\omega)$ is work in system at time t. $X_t^s(\omega)$ is work in system at time t if empty at s. $X_t^*(\omega) = \lim_{s\to-\infty} X_t^s(\omega)$ exists because variable is monotonic.
  - Now $W_t(\omega) = \begin{cases} \alpha & t = 0 \\ (\alpha - t)^+ & 0 \le t \le t_1 \\ (\alpha - t)^+ + S_1 & t = t_2 \\ (W_{t_n} - t - t_n)^+ & t \in [t_n, t_{n+1}) \\ \left(W_{t_n} - (t_{n-2} - t_n)\right)^+ + S_{n+1} & t = t_{n+1} \end{cases}$ These are called Lindley's Euqations.
  - ***Proposition***: If $\rho < 1$, then $X_t^* < \infty$ a.s. for all t.
    Proof: Fix $t \in \mathbb{R}$ and $\omega \in \Omega$ and define $T_t^s(\omega) = \sup\{\tau < t : X_\tau^s(\omega) = 0\}$; last empty time before time t. Now
$$X_t^s = \overbrace{X_{T_t^s}^s}^{\substack{\text{Work in system} \\ \text{at } T_t^s}} + \overbrace{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}}^{\substack{\text{Work entering the} \\ \text{system in } [T_t^s, t]}} - \overbrace{(t - T_t^s)}^{\substack{\text{Work processed} \\ \text{in } [T_t^s, t]}} = \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}} - (t - T_t^s) \le \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}$$
    Now clearly, for $s' < s$, $T_t^{s'} \le T_t^s$. Thus $\lim_{s\to-\infty} T_t^s = T_t^{-\infty}$ exists. Let us assume that $T_t^{-\infty} > -\infty$. This implies that there is some finite time in the past at which the system was empty, which implies that the amount of work at t, X_t^*, is finite.

    Show $\rho < 1 \Rightarrow T_t^{-\infty} > -\infty$ a.s. We do this by contradiction; suppose $\rho < 1$ and $T_t^{-\infty} = -\infty$ on a set ω of sample paths. From the second line above, we have, $X_t^s = \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}} - (t - T_t^s)$. Now, however, that $X_t^s \ge 0$ (there can never be negative work in the system), and so, dividing by $t - T_t^s$ throughout, we obtain $\frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}}{t - T_t^s} \ge 1$. Note that if we were to replace T_t^s with s, the LHS of this expression would form a sequence with limit to ρ as $s \to -\infty$. By the definition of a limit, however, this is also true for any *subsequence* of that sequence. But since we have assumed $T_t^s \to -\infty$ as $s \to -\infty$, the LHS above is such a subsequence. Thus, letting $s \to -\infty$, we find $1 \le \frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [T_t^s, t]\}}}{t - T_t^s} \to_{s\to-\infty} \rho$
  - ***Proposition***: If $\rho < 1$, then coupling time $T_0 < \infty$ a.s.
    Proof: WLOG, assume $\tau = 0$ and denote $W_{\tau,t}^\alpha \equiv W_t^\alpha$ [work in system at t given started at 0 with α. Assume that $\alpha > X_\tau^*(\omega)$. Then $T_0 = \inf\{t > 0 : W_t^\alpha = 0\}$. Assume $\rho < 1$ and suppose $T_0 = \infty$ on a set of nonzero probability. This implies that $W_t^\alpha > 0 \; \forall t \ge 0$ and that the server is *constantly* working for all $t \ge 0$. As such $0 \le W_t^\alpha = \overbrace{\alpha}^{\substack{\text{Work at} \\ t=0}} + \overbrace{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}}}^{\text{New work in } [0,t]} - \overbrace{t}^{\substack{\text{Work processed} \\ \text{in } [0,t]}}$. Dividing by t and re-arranging, we obtain $1 \le \frac{\alpha}{t} + \frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}}}{t} \to_{t\to\infty} \rho$. Contradiction.
  - ***Proposition***: If $\rho < 1$ $\sup_{A_1,\cdots,A_n} \left|\mathbb{P}\{W_{t_1+\tau}^\alpha \in A_1, \cdots, W_{t_n+\tau}^\alpha \in A_n\} - \mathbb{P}\{X_{t_1}^* \in A_1, \cdots, X_{t_n}^* \in A_n\}\right| \to_{\tau\to\infty} 0$
    Where $A_1, \cdots, A_n$ are measurable sets and for all $(t_1, \cdots, t_n) \in \mathbb{R}^n$ and $\alpha > 0$. This is called *convergence is total variation* and is stronger than weak convergence.
  - We can compare to what happens in Markov chains. Consider a finite-state, irreducible Markov chain, & consider $\left|\mathbb{P}\{X_n = i \mid X_0 = j\} - \mathbb{P}_\pi\{X_n = i\}\right| \le \max_{i,j} \mathbb{P}_j\{T_i > n\} \le \left(\max_{i,j} \mathbb{E}_j e^{\theta T_i}\right) e^{-\theta n}$
  - ***Proposition***: If $\rho > 1$, then for all $\alpha > 0$, $\liminf_{t\to\infty} \frac{W_t^\alpha}{t} > 0$ almost surely. In other words, the workload increases linearly with time.
    Proof: Denote the cumulative idle time up to time t as I_t. We then have $W_t^\alpha = \alpha + \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}} - (t - I_t) \ge \sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}} - t$. Dividing by t $\frac{W_t^\alpha}{t} \ge \frac{\sum_{n\in\mathbb{Z}} S_n \mathbb{1}_{\{t_n \in [0,t]\}}}{t} - 1$. Letting $t \to \infty$, we get that $\liminf_{t\to\infty} \frac{W_t^\alpha}{t} \ge \rho - 1 > 0$.
  - ***Proposition***: Busy probability of an arbitrary server is ρ. ***Proposition***: By Little's Law, ρ is also the expected number of customers in service [arrival λ, sojourn $\frac{1}{\mu}$]
- ***Little's Law (Conservation Laws)***
  - We now consider a setting in which work $(t_n, S_n : n \in \mathbb{Z})$ enters a system and then leaves the system at a time $(d_n : n \in \mathbb{Z})$. We let θ_n be the *sojourn time* of the n^{th} job in the system, given by $d_n - t_n$. We define the following quantities $A(t) = \sup\{n : t_n \le t\}$ = number of arrivals in $[0, t]$ and $D(t)$ = number of of departures in $[0, t]$ and $N(t) = A(t) - D(t)$ = work in the system at time t
    The only two assumptions we make is that the following are true for every sample path (1) $\lim_{t\to\infty} \frac{A(t)}{t} = \lambda \in (0, \infty)$ (2) $\lim_{n\to\infty} \frac{\sum_{k=1}^n \theta_k}{n} = \bar{\theta} \in (0, \infty)$
  - ***Proposition***: Under these assumptions, $\frac{1}{t}\int_0^t N(s)\,\mathrm{d}s \to \lambda\bar{\theta}$

    Proof: $\int_0^t N(s)\,\mathrm{d}s = \int_0^t A(s) - D(s)\,\mathrm{d}s$. We can write $A(s) = \sum_{n\in\mathbb{Z}} \mathbb{1}_{\{t_n \le s\}}$ and $D(s) = \sum_{n\in\mathbb{Z}} \mathbb{1}_{\{d_n \le s\}}$. As such $N(s) = \sum_{n\in\mathbb{Z}} \mathbb{1}_{\{t_n \ge s\}} - \mathbb{1}_{\{d_n \ge s\}} = \sum_{n\in\mathbb{Z}} \mathbb{1}_{\{t_n \le s \le d_n\}}$ (The last equality follows because any jobs that arrive after s won't be counted at all, and any events that arrive and leave before s will be counted by both indicators and therefore cancel out). Swapping the summation and integration (Fubini)
$$\int_0^t N(s)\,\mathrm{d}s = \sum_{n\in\mathbb{Z}} \int_0^t \mathbb{1}_{\{t_n \le s \le d_n\}}\,\mathrm{d}s = \sum_{n\in\mathbb{Z}} \begin{Bmatrix} \text{Amount of time job } n \text{ was} \\ \text{in the system during } [0,t] \end{Bmatrix}$$
    We can bound this above by considering the sojourn time of all arrivals up to and including time t (though some of them may overrun past t) and lower bound it by considering the sojourn time of all job that depart before time t (even though some jobs that leave after t do spend *some* time in the system before t). This gives $\sum_{i=0}^{D(t)} \theta_{n_i} \le \int_0^t N(s)\,\mathrm{d}s \le \sum_{n=0}^{A(t)} \theta_n$. Where n_i is the index of the i^{th} job to leave the system (since we have no assumed FIFO processing discipline, we cannot assume that $\theta_i = i$). Before we continue, we will need the following claim

    Claim: Under the two assumptions above $\theta_n / t_n \to_{n\to\infty} 0$

    Proof: Consider that $\frac{\theta_n}{n} = \frac{\sum_{k=1}^n \theta_k - \sum_{k=1}^{n-1} \theta_k}{n} = \frac{1}{n}\sum_{k=1}^n \theta_k - \frac{1}{n}\sum_{k=1}^{n-1} \theta_k \to \bar{\theta} - \bar{\theta} \to 0$

    Consider also that $\frac{n}{t_n} = \frac{A(t_n)}{t_n} \to \lambda$ as $t_n \to \infty$ Finally, $\frac{\theta_n}{n}\frac{n}{t_n} \to 0 \cdot \lambda = 0$

    Now by our claim, for all $\varepsilon > 0$, there exists an $N(\varepsilon)$ such that for $n > N(\varepsilon)$ $\frac{\theta_n}{t_n} \le \varepsilon \Rightarrow \frac{d_n - t_n}{t_n} \le \varepsilon$. This implies that $d_n \le (1 + \varepsilon)t_n \; \forall n > N(\varepsilon)$. This means that all jobs after the $N(\varepsilon)^{\text{th}}$ job that arrive in [0, t] will have departed by time $t(1 + \varepsilon)$. Therefore $\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n \le \sum_{i=0}^{D(t(1+\varepsilon))} \theta_{n_i}$. Putting this together with the bounds developed above, we find that
$$\frac{1}{t(1+\varepsilon)}\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n \le \frac{1}{t(1+\varepsilon)}\int_0^{t(1+\varepsilon)} N(s)\,\mathrm{d}s \le \frac{1}{t(1+\varepsilon)}\sum_{n=1}^{A(t(1+\varepsilon))} \theta_n$$
    Let's first consider the upper bound
$$\frac{1}{t(1+\varepsilon)}\sum_{n=1}^{A(t(1+\varepsilon))} \theta_n = \left(\frac{A(t(1+\varepsilon))}{t(1+\varepsilon)}\right)\left(\frac{1}{A(t(1+\varepsilon))}\sum_{n=1}^{A(t(1+\varepsilon))} \theta_n\right) \to \lambda\bar{\theta}$$
    The first term tends to λ, by assumption 1. The second term tends to $\bar{\theta}$ because (1) By assumption 1, $A(t) \to_{t\to\infty} \infty$ (2) By assumption 2, $\frac{\sum \theta_n}{n} \to \bar{\theta}$, which means any subsequence thereof also $\to \bar{\theta}$. Since $A(t(1+\varepsilon)) \to \infty$, the second term above is precisely such a subsequence. This implies that $\limsup_{t\to\infty} \frac{1}{t}\int_0^t N(s)\,\mathrm{d}s \le \lambda\bar{\theta}$. Now the lower bound $\frac{1}{t(1+\varepsilon)}\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n$
$$= \frac{1}{1+\varepsilon}\frac{A(t)}{t}\frac{1}{A(t)}\sum_{n=N(\varepsilon)+1}^{A(t)} \theta_n = \frac{1}{1+\varepsilon}\frac{A(t)}{t}\frac{\sum_{n=0}^{A(t)} \theta_n - \sum_{n=0}^{N(\varepsilon)} \theta_n}{A(t)}$$
$$= \frac{1}{1+\varepsilon}\frac{A(t)}{t}\left(\frac{\sum_{n=0}^{A(t)} \theta_n}{A(t)} - \frac{\sum_{n=0}^{N(\varepsilon)} \theta_n}{A(t)}\right) \to \frac{\lambda\bar{\theta}}{1+\varepsilon}$$
    Again, the first term tends to λ by assumption 1. The second term tends to $\bar{\theta}$ by a similar logic as above, and by noting that since $N(\varepsilon) < \infty$, the second term in brackets tends to 0. As such, we get $\liminf_{t\to\infty} \frac{1}{t}\int_0^t N(\varepsilon)\,\mathrm{d}s \ge \frac{\lambda\bar{\theta}}{1+\varepsilon}$. Since this is true for all $\varepsilon > 0$, this, together with the lim sup above, proves our theorem. ■
- ***Single-server queue; IID case***
  - We now specialize our analysis to a situation in which the workloads and inter-arrival times are IID. Letting $\tau_n = t_n - t_{n-1}$ be the time gap before the n^{th} job arrives, this situation requires the $\{S_n\}$ and $\{\tau_n\}$ to be IID. There is, once again, only one server. We often denote this situation *GI*/*GI*/1.
  - We also denote by w_n the time that the n^{th} job has to wait in queue before it is served. In that respect, $d_n = t_n + w_n + S_n$. Now, consider that $w_{n+1} = \begin{cases} 0 & d_n \le t_{n+1} \\ d_n - t_{n+1} & d_n > t_{n+1} \end{cases} = (d_n - t_{n+1})^+$. As such $\boxed{w_{n+1} = (w_n + S_n - \tau_{n+1})^+}$ Let $Z_n = S_{n-1} - \tau_n$. We then have $w_{n+1} = (w_n + Z_{n+1})^+ = \max(0, w_n + Z_{n+1})$. Since the Z_n are IID random variables, this is a random walk "capped off" at the origin. Letting $\sigma_n = \sum_{j=1}^n Z_j$, we get $w_1 = \max(Z_1, 0) = \max(\sigma_1, 0)$ and then $w_2 = (w_1 + Z_2)^+ = \sigma_2 + \max(0, -\sigma_1, -\sigma_2)$. Generally, $w_n = \sigma_n - \min_{0\le k\le n} \sigma_k$. The second term takes into account the *reflected* random walk, to prevent it going negative.
  - One thing the *GI*/*GI*/1 framework gives us over the *G*/*G*/1 framework is that we can now say something more about the distribution of the waiting times: $w_n = \sigma_n - \min_{k\le n} \sigma_k = \max_{0\le k\le n}\{\sigma_n - \sigma_k\} = \max_{0\le k\le n}\{\sum_{k=h+1}^n Z_j\}$. Since IID, we can change indices on the Z and $=_d \max_{0\le k\le n}\{\sum_{j=1}^k Z_j\} = \max_{0\le k\le n}\{\sigma_k\} = M_n$ Since M_n is a non-decreasing sequence, $M_n \to_{\text{a.s.}} M_\infty = \max_{k\ge 0} \sigma_k$. As such, $w_n \Rightarrow M_\infty$ as $n \to \infty$. [Note: convergence in distribution is the best we can hope for in this case, because M_n has some structure (the fact it's non-decreasing) that w_n does not. However, since this is a Markov chain, a stationary distribution is all we could really want].
  - If the random walk has positive drift – in other words, if $\mathbb{E}(Z_n) = \mathbb{E}(S_{n-1}) - \mathbb{E}(\tau_n) > 0$ – then the random walk drifts to infinity and the waiting times get infinitely large. This is consistent with our findings in the *G*/*G*/1 queue, since $\mathbb{E}(\tau_{n+1}) > \mathbb{E}(S_n) \Leftrightarrow \rho < 1$. On the other hand, if the random walk has negative drift, the chain is stable and the waiting times return to 0 infinitely often almost surely (we motivated this result in homework 2 using a simpler reflected random walk).
- ***The single-server M/M/1 Markovian queue***
  - We now consider the most tractable of all single-server queue models; the *M*/*M*/1 in queue. In that case, we assume the $\{S_n\}$ are IID and exponentially distributed with parameter μ whereas the $\{\tau_n\}$ are IID and exponentially distributed with parameter λ.
  - Consider the process $X(t)$ = Number of jobs in system at time $t \ge 0$ $\{X(t) : t \ge 0\}$ is CTMC.
$$\mathbb{P}\{X(t+h) = j \mid X(t) = i\} = \begin{cases} \lambda h + o(h) & j = j + 1 \\ \mu h + o(h) & j = j - 1 \\ 1 - (\lambda + \mu)h + o(h) & j = i \\ o(h) & \text{otherwise} \end{cases}$$
    Our transition matrix P then takes the form (rows sum to 1)
$$P_h = I + \begin{vmatrix} -\lambda & \lambda & 0 & & \\ \mu & -(\lambda+\mu) & \lambda & 0 & \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 \\ & 0 & \mu & \ddots & \ddots \\ & & 0 & \ddots & \ddots \end{vmatrix} h + o(h)$$
$$P_h = I + Qh + o(h)$$
    Q is the rate matrix, defined by $Q = \lim_{h\downarrow 0} \frac{P_h - I}{h}$. The steady-state equations in this case are $\pi^\top P_h = \pi^\top$, $\pi \ge 0$, $\pi \cdot c = 1$. Feeding our expression for P_h into the first equation, we obtain $\pi^\top + \pi^\top Qh + o(h) = \pi^\top \Rightarrow \pi^\top Q = 0$. For our particular matrix, this give $-\lambda\pi_0 + \mu\pi_1 = 0$ and $\lambda\pi_{n-1} + (\mu + \lambda)\pi_n + \mu\pi_{n+1} = 0$. Letting $\rho = \lambda / \mu$ and assuming $\rho < 1$, we obtain $\pi_0\rho = \pi_1$ and $\rho\pi_{n+1} - (1 + \rho)\pi_n + \pi_{n+1} = 0$. Solving this recurrence relation, we obtain $\pi_n = c\rho^n$ and $\sum \pi_n = 1 \Rightarrow c = 1 - \rho$.So $\boxed{\pi_n = (1 - \rho)\rho^n \qquad n \ge 0}$ This expression passes a "sanity check" at n = 0; the probability the chain is empty is given by $\pi_0 = 1 - \rho = 1 - \frac{\lambda}{\mu}$, which we would expect to be the average amount of time the system is empty.
  - We can use Little's Law to good effect; consider the following two examples; the first is trivial, the second less so
    - ***Consider the server as the "system"***
      Let S be the number of items in the server. The average sojourn time in the server is $\frac{1}{\mu}$. As such, by Little's Law $\mathbb{E}_\pi(S) = \frac{\lambda}{\mu} = \rho$. Note, however, that $S = 1 \cdot 1_{\text{System is busy}} + 0 \cdot \mathbb{1}_{\text{System is idle}}$. As such, $\mathbb{E}_\pi(S) = \mathbb{P}(\text{System is busy}) = \rho$. The result above is therefore consistent with what we would expect.
    - ***Consider the queue as the "system"***
      Let Q be the number of items in the queue. The sojourn time in the queue is simply the waiting time, w_i. As such, by Little's Law $\mathbb{E}_\pi(Q) = \lambda\mathbb{E}_\pi(w)$ (where w is the waiting time for any new item that joins the queue). Now, note that $Q(t) = (X(t) - 1)^+$ (we subtract item currently in the server). Thus $E_\pi(Q) = \sum_{n=1}^\infty (n-1)(1-\rho)\rho^n = \frac{\rho^2}{1-\rho}$. Thus $\rho / \mu(1 - \rho)$.
  - We know consider the more difficult problem of deriving the probability distribution of waiting times, $\mathbb{P}_\pi(w > x)$. Once again, recall w is the wait time experienced by a random job entering the queue.
$$\mathbb{P}_\pi(w > x) = 0 \cdot \mathbb{P}(X = 0) + \sum_{k=1}^\infty \mathbb{P}_\pi(w > x \mid X = k)\mathbb{P}_\pi(X = k)$$
$$= \sum_{k=1}^\infty \mathbb{P}_\pi(w > x \mid X = k)(1 - \rho)\rho^k$$
    Now, note that since the exponential distribution is memoryless, we can write
$$(w \mid X = k) =_d \overbrace{S_k}^{\substack{\text{Remaining time} \\ \text{for job currently} \\ \text{in server}}} + \overbrace{\sum_{j=1}^{k-1} S_j}^{\substack{\text{Processing time} \\ \text{for the } k-1 \text{ other} \\ \text{jobs waiting}}} =_d \text{Erlang}(k, \mu)$$
    (the Erlang distribution is the distribution of the sum of exponential variables; it is a special case of the Gamma distribution). As such
$$f_w(x) = \sum_{k=1}^\infty \frac{e^{-\mu x}(\mu x)^{k-1}\mu}{(k-1)!}\rho^k(1 - \rho) = e^{-\mu x}\mu(1 - \rho)\rho\sum_{k=1}^\infty \frac{(\mu x)^{k-1}}{(k-1)!}\rho^{k-1} = \rho\mu(1 - \rho)e^{-\mu(1-\rho)x}$$
    And so $w \sim \begin{cases} \exp(\mu(1 - \rho)) & \text{with prob } \rho \\ 0 & 1 - \rho \end{cases}$ and $\mathbb{P}(w > x) = \rho e^{-\mu x(1-\rho)}$

Prop 6.11 (M/M/c queue). *$\mu_n = n\mu$ if $n < c$ and $c\mu$ otherwise. $\lambda_n = \lambda$. Then $P(L = n) \equiv p_n = p_0\frac{\lambda^n}{n!\mu^n}$ if $n < c$ and $p_0\frac{\lambda^n}{c^{n-c}c!\mu^n}$ otherwise. Summing the probabilities to 1, $p_0 = \left(\sum_{n=0}^{c-1}\frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)}\right)^{-1}$ where $r = \frac{\lambda}{\mu}$ and $\frac{r}{c} = \rho$. We still need $\rho < 1$. $E_\pi L_q = p_0\frac{r^c\rho}{c!(1-\rho)^2}$. Waiting time distribution is $P(W_q > w) = \frac{p_0 r^c}{c!(1-\rho)}\rho e^{-(c\mu-\lambda)w}$ When $n \ge c$ the system output is Poisson with mean $c\mu$ and the distribution of the time for the $n - c + 1$ completions is Erlang type $n - c + 1$.*

Prop 6.12 (M/M/c/K queue). *$\mu_n = n\mu$ if $n < c$ and $c\mu$ o.w. $\lambda_n = \lambda$ if $n < K$ and 0 o.w. Then $P(L = n) = p_n = p_0\frac{\lambda^n}{n!\mu^n}$ if $n < c$ and $p_0\frac{\lambda^n}{c^{n-c}c!\mu^n}$ if $c \le n \le K$. Since a fraction p_K of arrivals do not join the system, by the PASTA property effective arrival rate seen by the servers $\lambda(1 - p_K) \equiv \lambda_{eff}$. Thus in Little's formula, we use $\lambda_{eff}E_\pi W_q = E_\pi L_q$. In an M/M/1/K model, the limiting distribution is truncated geometric: $p_j = \frac{\rho^j(1-\rho)}{1-\rho^{K+1}}$ if $\rho \ne 1$, and $p_j = \frac{1}{K+1}$ if $\rho = 1$. And $L_q = \frac{\rho}{1-\rho} - \frac{\rho(K\rho^K+1)}{1-\rho^{K+1}}$ if $\rho \ne 1$, and $L_q = \frac{K(K-1)}{2(K+1)}$ if $\rho = 1$. Note $\mu(1 - p_0) = \lambda(1 - p_K)$.*

Prop 6.13. *Consider a M/M/1/K queue. Then, $q_n = \frac{p_n}{1-p_K}$, i.e.: PASTA does not hold.*

Prop 6.14 (M/M/∞ queue). *$\mu_n = n\mu$ and $\lambda_n = \lambda$. Then $P(L = n) \equiv p_n = p_0\frac{r^n}{n!}$ where $p_0 = e^{-r}$ and $r = \frac{\lambda}{\mu}$ (also valid for M/G/∞). Steady state solution exists $\forall\rho$. $L = r$.*

Prop 6.15 (Finite source queue). *$\mu_n = n\mu$ if $n < c$ and $c\mu$ o.w. $\lambda_n = (M - n)\lambda$ if $n < M$ and 0 o.w. Then $P(L = n) \equiv p_n = p_0 r^n\frac{M!}{n!(M-n)!}$ if $n < c$ and $p_0 r^n\frac{M!}{(M-n)!c^{n-c}c!}$ $c \le n \le M$. Effective arrival rate: $\lambda_{eff} = \sum_{n=0}^{M-1}\lambda p_n(M - n) = \lambda(M - L)$. If we have Y spares then $\lambda_n = M\lambda$ if $0 \le n < Y$, $(M - n + Y)\lambda$ if $Y \le n < M + Y$ and 0 o.w.*

Def 6.16 (Busy period). *A busy cycle is the sum of the busy period and the idle period. For all M/G/1 type queues, we have $E(T_{bp}) = \frac{1}{\mu-\lambda}$ and $E(T_{bc}) = \frac{1}{\lambda} + \frac{1}{\mu-\lambda}$.*

- ***The M/G/1 queue***
  - Also require $R(t)$. To sidestep this complication, we will consider the following *embedded Markov chain* X_n. Let T_n be the time at which the n^{th} job concludes processing, and denote $X(T_{n^+}) \equiv X_n \equiv \frac{\text{Number of jobs in the system}}{\text{immediately after the } n^{\text{th}} \text{ job as departed}}$. We then have $X_{n+1} = (X_n - 1)^+ + A_{n+1}$, where A_{n+1} is the number of arrivals during the processing time of the $(n + 1)^{\text{th}}$ job. In this case, by assumption, the A_n are IID, and $(A_n \mid S_n = s) \sim \mathrm{Po}(\lambda s)$, where λ is the rate of arrivals.
  - Now, using the ergodic theorem for DTMCs (and assuming we have stability; ie: $\rho < 1$), we have $\mathbb{P}(X_n = j) \to_{n\to\infty} \pi(j)$. Furthermore, from the *G*/*G*/1 case, we know that for each sample path,

$X(t) = X_t^*$ exists, which implies that $\mathbb{P}\left(X(t) = j\right) \to_{t\to\infty} \hat{\pi}(j)$. The challenge now, however, is to prove that $\pi = \hat{\pi}$. ***Theorem***: $\pi = \hat{\pi}$

- The PK formula $\mathbb{E}_\pi(w) = \dfrac{1+\frac{\sigma_s^2}{\mu_s^2}}{2} \cdot \dfrac{\rho}{1-\rho} \cdot \mathbb{E}(S)$.

Def 6.20 (Departure Point Steady State). *Let the imbedded Markov chain at the departure points be $\mathbf{P} = \{p_{ij}\}$. $p_{ij} = P(X_{n+1} = j|X_n = i) = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^{j-i+1}}{(j-i+1)!} dF(t)$. Letting k_n denote the probability of n arrivals during service time $S = t$, $k_n = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{(n)!} dF(t)$ and thus $p_{ij} = k_{j-i+1}$. Let π_n denote the probability of n in the system at a departure point in the steady state $\pi\mathbf{P} = \pi$ yields $\pi_i = \pi_0 k_i + \sum_{j=1}^{i+1} \pi_j k_{i-j+1}$. Letting $\Pi(z) = \sum_0^\infty \pi_i z^i$ and $K(z) = \sum_0^\infty k_i z^i$, by multiplying the above equation by z^i and summing over i, we obtain $\Pi(z) = \frac{\pi_0(1-z)K(z)}{K(z)-z}$. $\Pi'(1)$: expected system size.*

Thm 6.21 (Erlang loss formula (M/G/k/k) (valid for M/M/k/k)). *The limiting distribution of the number of customers in the system is $P(n) = \frac{(\lambda ES)^n/n!}{\sum_{i=0}^k (\lambda ES)^i/i!}$, and given that there are n in the system, the residual times are iid with the equilibrium distribution of G, given by $G_e(x) = \frac{\int_0^x \bar{G}(u)\,du}{ES}$.*

- ***Addendum on PASTA***
  - The principle of PASTA (Poisson Arrivals see Time Averages) states that for *any* stochastic process $X(t)$ over a state space $\mathcal{S}$, and for any $A \subseteq \mathcal{S}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X(t_k)\in A\}} = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{1}_{\{X(s)\in A\}}\, \mathrm{d}s$$

provided the t_k form a Poisson process – in other words, $t_k - t_{k-1} \sim \exp(\lambda)$.

6.6 G/M/1: General input and exponential service

We now consider the system just before an arrival and X_i will represent the number in the system that the ith arrival sees upon joining the system. Then $X_{n+1} = (X_n + 1 - B_n)^+$ where B_n is the number of customers served during the interarrival time T^n between nth and $n+1$st arrivals. Let us denote its cdf by $A(t)$. Then $P(B = b) = \int_0^\infty P(B = b|T = t)dA(t)$ $= \int_0^\infty \frac{e^{-\mu t}(\mu t)^b}{(b)!} dA(t)$. Define $p_{ij} = P(X_{n+1} = j|X_n = i)$ and then for $j = 0, 1, ..., i$ we have $p_{i,i+1-j} = \int_0^\infty \frac{e^{-\mu t}(\mu t)^j}{(j)!} dA(t)$. The formula for $p_{i,0} = \int_0^\infty \sum_{k=i+1}^\infty \frac{e^{-\mu t}(\mu t)^k}{(k)!} dA(t)$ Let $b_n = P(B = n)$ and q be the corresponding steady state solution that an arrival finds n in the system. Then $qP = q$ and $q\mathbf{1} = 1$ yields $q_i = \sum_{k=0}^\infty q_{i+k-1} b_k$ for $i \geq 1$ and $q_0 = \sum_{j=0}^\infty q_i \left(1 - \sum_{k=0}^j b_k\right)$. The solution is of the form $q_k = c\beta^k$ where $c = 1 - \beta$ and $\beta = \int_0^\infty e^{-\mu t(1-\beta)} dA(t)$.

Renewal & Regenerative Process

- ***Renewal processes***
  - Let $\{X_n\}$ be IID, with $\mathbb{E}(X_1) = \mu < \infty$ and $\mathbb{P}(X_1 = 0) < 1$. Let $S_n = \sum_{i=1}^n X_i$, with $S_0 = 0$. Let $N(t) = \sup\{n \geq 1 : S_n \leq t\}$. $\{N(t) : t \geq 0\}$ is called a *renewal process*.
  - ***Definition (renewal function)***: The *renewal function* is defined as $m(t) = \mathbb{E}[N(t)]$.
  - ***Example***: Let $X_1 \sim \exp(1/\mu)$. $\{N(t) : t \geq 0\}$ is then a Poisson process. Consider, incidentally, that in a Poisson process, the following two facts are true. Much of our work in this section will be concerned with generalizing these results to general renewal processes:
    - $\mathbb{E}(N(t)) = m(t) = \mu t = \mathbb{E}(X_1)t$ (generalizes to the *elementary renewal theorem*).
    - $m'(t) = \mu = \mathbb{E}(X_1)$ (generalizes to the *key renewal theorem*). □
  - ***Proposition***: $m(t) = \sum_{n=1}^\infty F_n(t)$, where $F_n(t) = \mathbb{P}(S_n \leq t)$.

    Proof: Note that $N(t) = \sum_{n=1}^\infty \mathbb{1}_{\{S_n \leq t\}} \Rightarrow \mathbb{E}[N(t)] = \sum_{n=1}^\infty \mathbb{P}(S_n \leq t)$, using Fubini I

    Remark: The CDF of S_n is the n-fold convolution of the CDF of each individual RV X. $F_n(\cdot) = \underbrace{F_1(\cdot) * F_1(\cdot) * \cdots * F_1(\cdot)}_{n \text{ fold}}$. For example, $F_2(t) = \int F(t-n)\,\mathrm{d}F(n)$
  - Note that $N(t) \geq n \Leftrightarrow S_n \leq t$
  - Analogies: $m(t) \approx \lambda t$, $\mathbb{P}\left[N(t+h) - N(t) = 1\right] \approx \lambda h$
  - $\mathbb{P}\left(S_{N(t)} \leq s\right) = \bar{F}(t) + \int_0^s \bar{F}(t-y)\,\mathrm{d}m(y) \qquad 0 \leq s \leq t$
- ***Some Theorems***
  - ***Proposition (SLLN for renewal processes)*** $\frac{N(t)}{t} \to_{t\to\infty} \frac{1}{\mathbb{E}(X_1)}$ a.s.

    Proof: Consider that $S_{N(t)} \leq t \leq S_{N(t)+1}$. As such $\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{S_{N(t)+1}}{N(t)+1}\frac{N(t)+1}{N(t)}$ Last term goes to 1 a.s., and by the SLLN, $\frac{S_n}{n} \to \mathbb{E}(X_1)$. Feeding this into the above proves our theorem.
  - ***Proposition (Elementary/Baby Renewal Theorem)***: $\frac{m(t)}{t} = \frac{\mathbb{E}[N(t)]}{t} \to_{t\to\infty} \frac{1}{\mathbb{E}(X_1)}$

    Proof: Stop at $N(t) + 1$ renewal. Stopping time, because $N(t) + 1 = n \Leftrightarrow N(t) = n-1 \Leftrightarrow X_1 + \cdots + X_{n-1} \leq t$ and $X_1 + \cdots + X_n > t$. Also $\mathbb{E}\left(S_{N(t)+1}\right) = \mu\left[m(t) + 1\right]$
  - ***Proposition (CLT for renewal processes)***: Suppose that $\mathrm{Var}[X_1] = \sigma^2 < \infty$ then $\sqrt{t}\left(\frac{N(t)}{t} - \frac{1}{\mathbb{E}(X_1)}\right) \Rightarrow \sigma \frac{1}{\left[\mathbb{E}(X_1)\right]^{3/2}} N(0,1)$ Or in other words $N(t) \approx \frac{t}{\mathbb{E}(X_1)} + \sqrt{t}\frac{\sigma}{\left[\mathbb{E}(X_1)\right]^{3/2}} N(0,1)$
  - ***Definition:*** A distribution F is said to be of Lattice-type if there exists an $h > 0$ such that F is supported on $\{X_n\} = \{nh : n \in \mathbb{Z}\}$ For example, the Poisson distribution is lattice with $h = 1$.
  - ***Theorem (Blackwell's Theorem)***: If F is *non-lattice*, then $m(t+a) - m(t) \to_{t\to\infty} \frac{1}{\mathbb{E}(X_1)} \cdot a \quad \forall a > 0$

    Remark: *not* implied by $\frac{m(t)}{t} \to \frac{1}{\mathbb{E}(X_1)}$, because this theorem concerns *increments* in m
  - ***Definition***: A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is *directly Riemann integrable (dRi)* if (1) $\int_0^\infty \bar{f}_n(x)\,\mathrm{d}x < \infty$ (2) $\int_0^\infty \bar{f}_n(x)\,\mathrm{d}x - \int_0^\infty \underline{f}_n(x)\,\mathrm{d}x \to_{h\to 0} 0$. Where $\bar{f}_n(x) = \sum_{k=0}^\infty \sup\left\{f(x) : kn \leq x \leq (k+1)n\right\}\mathbb{1}_{\{x\in[kh,(k+1)h]\}}$, $\underline{f}_n(x) = \sum_{k=0}^\infty \inf\left\{f(x) : kn \leq x \leq (k+1)n\right\}\mathbb{1}_{\{x\in[kh,(k+1)h]\}}$

    Remark: If $f : \mathbb{R}_+ \to \mathbb{R}$, we simply require both f^+ and f^- to be dRi for f to be dRi.

    Remark: *Any* of the following conditions are sufficient for $f : \mathbb{R}_+ \to \mathbb{R}$ to be dRi (1) f is continuous with compact support. (2) f is bounded and continuous, and $\int_0^\infty \bar{f}_h(x)\,\mathrm{d}x < \infty$ for some $h > 0$. (3) f is non-increasing and $\int_0^\infty f(x)\,\mathrm{d}x < \infty$
  - ***Theorem (Key Renewal Theorem)***: Providing the usual assumptions on F holds (IID increments with finite mean that are not masses at 0) and that the renewal process is non-lattice, then for any $b : \mathbb{R}_+ \to \mathbb{R}$ that is dRi $a(t) \to_{t\to\infty} \frac{1}{\mu}\int_0^\infty b(s)\,\mathrm{d}s$. Where a is a solution to renewal equation $a = b + a * F$.
- ***Laplace Transforms & Co.***
  - The Laplace Transform of a function of a distribution with CDF $F(x)$ is defined as $\hat{F}(s) = \int_0^\infty e^{-sx}\,\mathrm{d}F(x) = \int_0^\infty e^{-sx} f(x)\,\mathrm{d}x$. A few important results:
    - In CDF form, if $F = A * B$, then $\hat{F}(s) = \hat{A}(s)\hat{B}(s)$
    - Note that if f is a density function, then $\left|\hat{F}(s)\right| = \left|\int_0^\infty e^{-sx}\,\mathrm{d}F(x)\right| < 1$
  - Together, the points above imply that $\hat{m}(s) = \sum_{n=1}^\infty \hat{F}(s)^n = \frac{\hat{F}(s)}{1-\hat{F}(s)}$. Similarly, re-arranging $\hat{F}(s) = \frac{\hat{m}(s)}{1+\hat{m}(s)}$. Thus, there is a 1-to-1 correspondence between F and m.
  - ***Theorem***: If $b(t)$ is bounded on any interval, then the solution to the following *renewal equation* $a = b + a * F$ (or $a(t) = b(t) + \int_0^\infty a(t-s)\,\mathrm{d}F(s)$) is $a = b + b * m$ (or $a(t) = b(t) + \int_0^t b(t-s)\,\mathrm{d}m(s)$)
  - ***Example***: Consider that $\mathbb{E}\left[N(t) \mid X_1 = x\right] = \begin{cases} 0 & x > t \\ 1 + \mathbb{E}[N(t) - N(x) \mid X_1 = x] = 1 - m(t-x) & x \leq t \end{cases}$

    Now, $m(t) = \mathbb{E}[N(t)] = \int_0^\infty \mathbb{E}[N(t) \mid X_1 = k]\,\mathrm{d}F(x) = \int_0^t [1 + m(t-x)]\,\mathrm{d}F(x)$

    $F(t) + \int_0^t m(t-x)\,\mathrm{d}F(x) = F(t) + (m * F)(t)$. This is a renewal equation, with solution $m(t) = F(t) + \int_{s=0}^t F(t-s)\,\mathrm{d}m(s)$.
- ***Applications of the Key Renewal Theorem***
  - ***Alternating Renewal Processes***
    - Consider a system which can be "on" or "off". The distribution of "on" times is H, the distribution of "off" times is G and the distribution of an on-off cycle is F. We also let $P(t) = \mathbb{P}\left(\text{System is "on" at time } t\right)$, and we let the system begin, at $t = 0$, at the start of an "on" period.
    - We can write down the following *renewal equation* $P(t) = \bar{H}(t) + 0 + \int_0^t P(t-s)\,\mathrm{d}F(s)$. This equation reflects the fact that, at t (1) Either the first "on" period hasn't finished yet… (2) …or it has finished and we're in the "off" period… (3) …or it has finished, and so has the subsequent "off" period, and the total cycle took a period $s < t$. In which case, we simply restart with $P(t - s)$.
    - This is an equation of the form above, the solution of which is $P(t) = \bar{H}(t) + \int_0^t \bar{H}(t-s)\,\mathrm{d}m_F(s)$. We now find $\lim_{t\to\infty} P(t)$. To do this, we note that $\bar{H}(\infty) = 0$, and that, by the key renewal theorem, $m_F'(s) \approx 1/\mu_F$. Thus

$$\lim_{t\to\infty} P(t) = \frac{\int_0^\infty \bar{H}(t-s)\,\mathrm{d}s}{\mathbb{E}(F)} = \frac{\mathbb{E}(\text{"On" time per cycle})}{\mathbb{E}(\text{"On" t. per c.}) + \mathbb{E}(\text{"Off" t. per c.})}$$

  - ***The Excess and the Age***
    - We define the following three random variables (1) $\boxed{A(t) = S_{N(t)+1} - t}$ is the *age at* t – the amount of time since the last renewal. (2) $\boxed{Y(t) = t - S_{N(t)}}$ is the *excess/residual life at* t – the amount of time till the next renewal. (3) $\boxed{L(t) = Y(t) + A(t) = S_{N(t)+1} - S_{N(t)}}$ is the length of the time interval between the two events flanking the time point t.

      As usual, let X_i be the arrival time distributions. Use alternating renewal
      - ***Age***: To find $P(t) = \mathbb{P}\left(A(t) \leq x\right)$, let the cycle be "on" for the first x units of time since its last renewal, and "off" thereafter. We then have

$$\lim_{n\to\infty} P(t) = \frac{\mathbb{E}\left(\text{Length of "on" time}\right)}{\mathbb{E}\left(\text{Length of cycle}\right)} = \frac{\mathbb{E}\left(\min[X, x]\right)}{\mu} = \frac{\int_0^\infty \mathbb{P}\left(\min[X, x] > y\right)\,\mathrm{d}y}{\mu} = \frac{\int_0^x \bar{F}(y)\,\mathrm{d}y}{\mathbb{E}(X)}$$

      - ***Residual life***: To find $P(t) = \mathbb{P}\left(Y(t) \leq x\right)$, we consider a cycle to be "on" until the last x units of the renewal cycle. We then have

$$\lim_{n\to\infty} P(t) = \frac{\mathbb{E}\left(\text{Length of "off" time}\right)}{\mathbb{E}\left(\text{Length of cycle}\right)} = \frac{\mathbb{E}\left(\min[X, x]\right)}{\mu} = \frac{\int_0^x \bar{F}(y)\,\mathrm{d}y}{\mathbb{E}(X)}$$

- ***Renewal-reward processes***
  - In a renewal-reward process, a reward R_n is earned at each renewal. The renewal-reward stochastic process is then given by $R(t) = \sum_{n=1}^{N(t)} R_n$. We may assume that R_n depends on X_n, the length of the n^{th} renewal interval. We assume, however, that the pairs (X_n, R_n) are IID.
  - ***Theorem***: With probability 1, $R(t)/t \to \mathbb{E}(R)/\mathbb{E}(X)$. Where X is the length of a cycle (ie: between renewals). This theorem simply states that the average long-term reward is the reward per cycle divided by the length of the cycle. ***Proof***: We write $\frac{R(t)}{t} = \frac{\sum_{n=1}^{N(t)} R_n}{N(t)} \frac{N(t)}{t}$. By the strong law for R_n, the first term tends to $\mathbb{E}(R)$. By the strong law for renewal processes, second term to $1/\mathbb{E}(X)$.
- ***Assorted examples***
  - Let $\mathbb{P}\left(X_n = -1\right) = \mathbb{E}\left(X_n = 1\right) = \frac{1}{2}$, and let $N = \min\left\{n : X_1 + \cdots + X_n = 1\right\}$. N is a stopping time. Wald's Equation leads a contradiction, and $\mathbb{E}(N) = \infty$.
  - A computer includes three parts, with exponentially distributed lives (parameters $\lambda_1, \lambda_2, \lambda_3$). Repairs take a time X_1, X_2, X_3 to come, with arbitrary distributions and means μ_1, μ_2, μ_3.
    - The length of time of an uptime is the minimum of three exponential variables, and so it is $\sim \exp(\lambda_+)$. Similarly, the probability part i fails is λ_i / λ_+.
    - Consider a renewal-reward process in which one cycle is an uptime and a downtime. The long-run proportion of uptime is then $\mathbb{E}(\text{Uptime}) / \mathbb{E}(\text{Uptime} + \text{Downtime})$. And note that $\mathbb{E}(\text{Downtime}) = \sum \mathbb{P}(\text{Part } i \text{ fails})\mathbb{E}(\text{Downtime}|\text{Part } i \text{ failed}) = \sum \frac{\lambda_i}{\lambda_+}\mu_i$. Similarly, if we'd like to find the amount of time part 2 is in suspended animation, all we need is $\mathbb{E}(\text{Part 2 suspended}) / \mathbb{E}(\text{Uptime} + \text{Downtime})$. Where $\mathbb{E}(\text{Part 2 suspended}) = \frac{\lambda_1}{\lambda_+}\mu_1 + \frac{\lambda_3}{\lambda_+}\mu_3$
  - A truck driver goes from A to B at a fixed speed $\sim U(a,b)$. He then goes from B to A at a speed of *either a or b* (with equal probability). He then repeats with no break.
    - By considering this as a renewal-reward process, with a cycle being $A \to B \to A$, the proportion of time spent going from A to B is $\mathbb{E}(A \to B) / \mathbb{E}(A \to B \to A)$. And note that if the distance between the two locations is d: $\mathbb{E}(A \to B) = d\int_a^b \frac{1}{v}\frac{1}{b-a}\,\mathrm{d}v = \frac{d}{b-a}\ln\left(\frac{b}{a}\right)$ and $\mathbb{E}(B \to A) = d\left\{\frac{1}{2}\frac{1}{a} + \frac{1}{2}\frac{1}{b}\right\} = \frac{d(b+a)}{2ab}$
    - Similarly, we can find the proportion of time spent going at 40 miles per hour, by noting that the only time this would ever happen is going back from B to A at that speed. If we assign that event a reward of 1 per unit time, we're good. $\mathbb{E}(\text{Reward per } A \to B \to A \text{ cycle}) = \frac{1}{2} \cdot \frac{d}{40} + 0$.
- ***Regenerative processes***
  - A regenerative process is a stochastic process with time points at which, from a probabilistic point of view, the process restarts itself. A good example is CTMC (for example, corresponding to an $M/M/1$ queue). We consider that the process "resets" each time the queue empties.
  - ***Definition***: Let $\left(X(t) : t \geq 0\right)$ be a stochastic process with a sample path that is right-continuous with left limits (RCLL or CADLAG). Without loss of generality, we let the renewal occur at $X(t) = 0$. Now, define the following quantities:
    - $\tau(n+1) = \inf\left\{t \geq \tau(n) : X(t) = 0, X(t\_) \neq 0\right\}$, the time of the n^{th} renewal.
    - $\tau_{n+1} = \tau(n+1) - \tau(n)$, the inter-renewal time.
    - $\tilde{X}_n(t) = \begin{cases} X\left(\tau(n-1) + t\right) & t \in [0, \tau_n] \\ \Delta & t > \tau_n \end{cases}$, where Δ is outside the state space of X.

    Then X is said to be *regenerative* if
    - $\tau_n < \infty$ a.s. $\forall n$
    - $\tau(n) \to \infty$ as $n \to \infty$
    - $\tilde{X}_0, \tilde{X}_1, \cdots$ are independent
    - $\tilde{X}_1, \tilde{X}_2, \cdots$ are identically distributed (the first renewal might have a different distribution if the system doesn't start "empty").
  - ***Definition (recurrence)***: A regenerative process is *positive-recurrent* if $\mathbb{E}(\tau_1) < \infty$ and *null-recurrent* otherwise.
  - ***Proposition (SLLN for regenerative processes)***: Let $\left(X(t) : t \geq 0\right)$ be a regenerative process over a state space $\mathcal{S}$ with $\mathbb{E}(\tau_1) < \infty$, and let $f : \mathcal{S} \to \mathbb{R}$ be such that $\mathbb{E}\left(\int_{\tau(0)}^{\tau(1)} \left|f\left(X(s)\right)\right|\,\mathrm{d}s\right) < \infty$. Then

$$\frac{1}{t}\int_0^t f\left(X(s)\right)\,\mathrm{d}s \to \frac{\mathbb{E}\left(\int_{\tau(0)}^{\tau(1)} f\left(X(s)\right)\,\mathrm{d}s\right)}{\mathbb{E}(\tau_1)} \text{ a.s. as } t \to \infty$$

  - ***Proposition (CLT for regenerative processes)***: Let $\left(X(t) : t \geq 0\right)$ be a regenerative process on a state space $\mathcal{S}$ with $\mathbb{E}(\tau_1^2) < \infty$. Let $f : \mathcal{S} \to \mathbb{R}$ be a function satisfying $\mathbb{E}\left(Y_1(|f|)^2\right) < \infty$. Then

$$\sqrt{t}\left(\frac{\int_0^t f\left(X(s)\right)\,\mathrm{d}s}{t} - \frac{\mathbb{E}\left(Y_1(t)\right)}{\mathbb{E}(\tau_1)}\right) \Rightarrow \sigma N\left(0,1\right)$$

    With $\sigma^2 = \mathbb{E}\left(Y_1^2(f_c)\right) / \mathbb{E}(\tau_1)$ where $f_c(\cdot) = f(\cdot) - \frac{\mathbb{E}(Y_1(f))}{\mathbb{E}(\tau_1)}$.
  - ***Theorem***: Let $\left(X(t) : t \geq 0\right)$ be a positive recurrent regenerative process on a state space $S \subseteq \mathbb{R}^d$. Suppose that either of the following conditions hold
    - $F(x) = \mathbb{P}\left(\tau_1 = x\right)$ has a density and there is a function $h : \mathbb{R}^d \to \mathbb{R}$ that is bounded
    - $F(x) = \mathbb{P}\left(\tau_1 = x\right)$ is non-lattice and there is a function h that is bounded and continuous

    Then $\mathbb{E}\left[h(X(t))\right] \to \frac{\mathbb{E}\left[\int_{\tau(0)}^{\tau(1)} h(X(s))\,\mathrm{d}s\right]}{\mathbb{E}\left[\tau_1\right]}$ as $t \to \infty$ Then $X(t) \Rightarrow X(\infty)$ as $t \to \infty$ and $\mathbb{P}\left(X(t) \in A\right) = \frac{\mathbb{E}\left[\int_{\tau(0)}^{\tau(1)} \mathbb{1}_{\{X(s)\in A\}}\,\mathrm{d}s\right]}{\mathbb{E}\left[\tau_1\right]}$

    Remark: If $X(0) =_d X(\infty)$, then $\left(X(t) : t \geq 0\right)$ is stationary.

The Poisson Process

- ***Definition***: A stochastic process $\left\{N(t), t \geq 0\right\}$ is said to be a *counting process* if $N(t)$ represents the "number of events that have occurred up to time t". Hence: (1) $N(t) \geq 0$ (2) $N(t)$ is integer valued (3) $s < t \Rightarrow N(s) \leq N(t)$ (4) For $s < t$, $N(t) - N(s) =$ number of events in $(s, t]$
- In addition, the counting process is said to possess
  - *Independent increments* if the number of events in disjoint time intervals are independent.
  - *Stationary increments* if the distribution of events that occur in any interval of time depends only on length of interval. Eg, distribution of $N(t_2) - N(t_1)$ must be the same as that of $N(t_2 + s) - N(t_1 + s)$.
- ***The Poisson Process***
  - ***Definition***: A counting process $\left\{N(t), t \geq 0\right\}$ is said to be a *poisson process* having *rate* $\lambda > 0$ if (1) $N(0) = 0$ (2) The process has independent increments (3) $\mathbb{P}\left(N(s+t) - N(s)\right) \sim \mathrm{Po}\left(\lambda t\right)$. This automatically implies the process has stationary increments, and that $\mathbb{E}\left(N(t)\right) = \lambda t$.
  - ***Definition***: A counting process $\left\{N(t), t \geq 0\right\}$ is said to be a *poisson process* having *rate* $\lambda > 0$ if (1) $N(0) = 0$ (2) The process has stationary and independent increments. (3) $\mathbb{P}\left(N(h) = 1\right) = \lambda h + o(h)$ (4) $\mathbb{P}\left(N(h) \geq 2\right) = o(h)$. Where, if we say f is $o(h)$, we mean that $\lim_{h\to 0} f(h)/h = 0$.
- ***Inter-arrival and waiting time distributions***
  - Let X_1 be the time of the first event, and X_n be the time between the n^{th} and $(n-1)^{\text{th}}$ events. (1) First, note that $\mathbb{P}\left(X_1 > t\right) = \mathbb{P}\left(N(t) = 0\right) = e^{-\lambda t}$ (2)Then, note that $\mathbb{P}\left(X_2 > t \mid X_1 = s\right) = \mathbb{P}\left(0 \text{ events in } (s, s+t] \mid X_1 = s\right) =_{\text{Indep.}} \mathbb{P}\left(0 \text{ events in } (s, s+t]\right) =_{\text{stat.}} e^{-\lambda t}$ Thus, X_2 is also an exponential random variable, independent of X_1.
  - Now, consider the arrival time of the n^{th} events, $S_n = \sum_{i=1}^n X_i$. It can be shown that $S_n \sim \Gamma(\lambda, n)$ in three ways: (1) Using moment generating functions. (2) Noting that $N(t) \geq n \Leftrightarrow S_n \leq t$, which implies that $\mathbb{P}\left(S_n \leq t\right) = \mathbb{P}\left(N(t) \geq n\right) = \sum_{j=n}^\infty e^{-\lambda t}\frac{(\lambda t)^j}{j!}$ Differentiating leads to $f(t)$. (3) Using the independent increment assumptions $\mathbb{P}\left(S_n \in (t, t+\mathrm{d}t\right) = \mathbb{P}\left(N(t) = n-1, N(\mathrm{d}t) = 1\right) + o(\mathrm{d}t)$ $= \frac{e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}\lambda \mathrm{d}t + o(\mathrm{d}t)$ Which, letting $\mathrm{d}t \to 0$, leads f as required.
- ***Conditional distribution of arrivals times***
  - If $Y_1, \cdots, Y_n$ are n random variables, we define the k^{th} order statistic $Y_{(k)} = k^{\text{th}}$ smaller value among the Y. If the Y are IID with density f, then $f\left(Y_{(1)} = y_1, \cdots, Y_{(n)} = y_n\right) = n!\prod_{i=1}^n f(y_i) \qquad y_1 < \cdots < y_n$. This is because given a specific set of y, there are $n!$ ways to permute them, and each have probability $\prod f(y_i)$. Specifically, if the Y are uniformly distributed in $(0, t)$, then $f\left(Y_{(1)} = y_1, \cdots, Y_{(n)} = y_n\right) = \frac{n!}{t^n} \qquad 0 < y_1 < \cdots < y_n < t$
  - ***Theorem***: Given that $N(t) = n$, the n arrival times $S_1, \cdots, S_n$ have the same distribution as the order statistics corresponding to n IID $U(0, t)$ random variables. In other words, each of the S_i are IID $U(0, t)$.
  - ***Proposition***: Consider a situation in which, as events arrive at time s, they get classified into category i with probability $p_i(s)$. Then $N_i(t)$, the number of type-i events that have arrived by time t, is an independent Poisson random variable with mean $\lambda t p_i$, where $p_i = \frac{1}{t}\int_0^t p_i(s)\,\mathrm{d}s$.

    Proof: Consider that

$$\begin{aligned}\mathbb{P}\left(N_i(t) = n_i \ \forall i\right) &= \sum_{k=0}^\infty \mathbb{P}\left(N_i(t) = n_i \ \forall i \mid N(t) = k\right)\mathbb{P}\left(N(t) = k\right) \\ &= \mathbb{P}\left(N_i(t) = n_i \ \forall i \mid N(t) = n_+\right)\mathbb{P}\left(N(t) = n_+\right)\end{aligned}$$

    Now, consider an arbitrary event in the interval $(0, t)$. We know that the time at which that event occurred is uniformly distributed in $(0, t)$, so $p_i = \mathbb{P}\left(\text{It's type } i\right) = \int_0^t \mathbb{P}\left(\text{It's type } i \mid \text{Occured at } s\right)\mathbb{P}\left(\text{Occured at } s\right)\,\mathrm{d}t = \frac{1}{t}\int_0^t p_i(s)\,\mathrm{d}s$

    And so $\mathbb{P}\left(N_i(t) = n_i \ \forall i \mid N(t) = n_+\right)$ is simply multinomial: $\mathbb{P}\left(N_i(t) = n_i \ \forall i \mid N(t) = n_+\right) = \frac{n_+!}{\prod_{\forall i} n_i!}\prod_{\forall i} p_i^{n_i}$ And so (noting that $p_+ = 1$) $\mathbb{P}\left(N_i(t) = n_i \ \forall i\right) = \frac{n_+!}{\prod_{\forall i} n_i!}\prod_{\forall i} p_i^{n_i}\frac{(\lambda t)^{n_+}e^{-\lambda t}}{n_+!} = \prod_{\forall i}\frac{e^{-\lambda t p_i}(\lambda t p_i)^{n_i}}{n_i!}$ Which proves our theorem.
  - This theorem is useful when considering the distribution of customers that are still in service at a time t. For example, in an infinite-server queue with processing time distribution G, let a type-I customer be one that completes its service by time t. Then for a customer that arrives at s, $p_i(s) = G(t-s)$ for $s \leq t$.
- ***Nonhomogeneous Poisson Processes***
  - ***Definition***: A counting process $\left\{N(t), t \geq 0\right\}$ is said to be a *nonhomogeneous poisson process* having *rate* $\lambda > 0$ if (1) $N(0) = 0$ (2) The process has independent increments (3) $\mathbb{P}\left(N(s+t) - N(s)\right) \sim \mathrm{Po}\left(m(t+s) - m(t)\right)$, where $m(t) = \int_0^t \lambda(s)\,\mathrm{d}s$.
  - ***Definition***: A counting process $\left\{N(t), t \geq 0\right\}$ is said to be a *poisson process* having *rate* $\lambda > 0$ if (1) $N(0) = 0$ (2) The process has stationary and independent increments. (3) $\mathbb{P}\left(N(h) = 1\right) = \lambda(t)h + o(h)$ (4) $\mathbb{P}\left(N(h) \geq 2\right) = o(h)$
  - If $\lambda(t) \leq \lambda$ (ie: it is bounded), the process can be obtained by running a larger Poisson process with rate λ, and counting an event at s with probability $\lambda(s)/\lambda$. To see why, consider that

$$\begin{aligned}\mathbb{P}\left(\text{One event counted} \in (t, t+\delta)\right) &= \mathbb{P}\left(\text{One event} \in (t, t+\delta)\right)\frac{\lambda(t)}{\lambda} + o(h) \\ &= h\lambda(t) + o(h)\end{aligned}$$

  - ***Proposition***: Consider a situation in which, as events arrive at time s, they get classified into category i with probability $p_i(s)$. Then $N_i(t)$, the number of type-i events that have arrived by time t, is an independent Poisson random variable with mean $p_i = \int_0^t \lambda(s)p_i(s)\,\mathrm{d}s$.

    Proof: Consider a larger Poisson process with rate $\lambda \geq \lambda(t)$. From our previous proposition, we then have that $p_i = \Lambda\int_0^t \frac{\lambda(s)}{\Lambda}p_i(s)\,\mathrm{d}s = \int_0^t \lambda(s)p_i(s)\,\mathrm{d}s$
- ***Compound Poisson Processes***
  - If $X_1, X_2, \cdots$ are IID with distribution F, independently of $N \sim \mathrm{Po}(\lambda)$, then $W = \sum_{i=1}^N X_i$ is a *compound Poisson variable*.
  - Properties
    - $M_W(s) = \mathbb{E}\left(e^{sW}\right) = \mathbb{E}\left[\mathbb{E}\left(\exp(sX_1)\right)^n\right] = \mathcal{G}\left(M_X(s)\right) = \exp\left(\lambda t[M_X(s) - 1]\right)$
    - $\mathbb{E}(W) = \lambda\mathbb{E}(X)$ and $\mathrm{Var}(W) = \lambda\mathbb{E}(X^2)$
  - ***Important application***: consider a Poisson process with rate α in which an event at time s will contribute an amount X_i whose value is a random variable with distribution F_s. Even though the X_i are neither identical nor independent, the variable $W(t) = \sum_{i=1}^{N(t)} X_i$ is then a compound Poisson random variable with rate $\lambda = \alpha t$ and distribution $F(x) = \frac{1}{t}\int_0^t F_s(x)\,\mathrm{d}s$.

    Proof: Consider a given event i at a time $s \in (0, t)$.

$$\mathbb{P}\left(X_i \leq x \mid s\right) = F_s(x)$$

    But we know that $s \sim U(0, t)$, so

$$\mathbb{P}\left(X_i \leq x\right) = \mathbb{E}\left[\mathbb{P}\left(X_i \leq x \mid s\right)\right] = \frac{1}{t}\int_{s=0}^t F_s(x)\,\mathrm{d}s$$

    Now, consider that

$$M_W(\tau) = \mathbb{E}\left(e^{\tau W}\right) = \mathbb{E}\left[\mathbb{E}\left(e^{\tau X_1}\right)^n\right] = \mathbb{E}\left[M_{X_1}(\tau)^n\right] = \exp\left(\lambda t\left\{M_{X_1}(\tau) - 1\right\}\right)$$

    This is the MFG of a compound Poisson as quoted above.
  - ***Important application II***: If F is discrete, such that $\mathbb{P}(X_i = j) = p_j$, and we define $N_j =$ number of events of type j that occur, we can also write $W = \sum_{\forall j} jN_j \qquad N_j \sim \mathrm{Po}(\lambda p_j)$

    Proof: Consider numbers $n_1, \cdots, n_j$, and note that

$$\begin{aligned}\mathbb{P}\left(N_i = n_i \ \forall i\right) &= \sum_{k=0}^\infty \mathbb{P}\left(N_i = n_i \ \forall i \mid N = k\right)\mathbb{P}\left(N = k\right) \\ &= \mathbb{P}\left(N_i = n_i \ \forall i \mid N = n_+\right)\mathbb{P}\left(N = n_+\right)\end{aligned}$$

    And note that the first probability is multinomial.
- ***Conditional Poisson Processes***

- Conditional Poisson processes are Poisson processes in which the parameter Λ has been mixed over a distribution G. So, for example $\mathbb{P}\left(N(t+s)-N(s)=n\right)=\int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}\,\mathrm{d}G(\lambda)$

  Is a Poisson processes, because $N(t+s)-N(s)=n \mid \Lambda=\lambda \sim \mathrm{Po}(\lambda t)$
- Given $N(t)=n$, the conditional distribution of Λ is

$$\mathbb{P}\left(\Lambda\le x \mid N(t)=n\right)=\frac{\mathbb{P}\left(\Lambda\le x \text{ and } N(t)=n\right)}{\mathbb{P}\left(N(t)=n\right)}=\frac{\int_0^\infty \mathbb{P}\left(\Lambda\le x \text{ and } N(t)=n \mid \Lambda=\xi\right)\mathrm{d}G(\xi)}{\int_0^\infty \mathbb{P}\left(N(t)=n\mid\Lambda=\lambda\right)\,\mathrm{d}G(\lambda)}$$
$$=\frac{\int_0^x \mathbb{P}\left(N(t)=n\mid\Lambda=\xi\right)\mathrm{d}G(\xi)}{\int_0^\infty \mathbb{P}\left(N(t)=n\mid\Lambda=\lambda\right)\,\mathrm{d}G(\lambda)}=\frac{\int_0^x e^{-\lambda t}(\lambda t)^n\,\mathrm{d}G(\xi)}{\int_0^\infty e^{-\lambda t}(\lambda t)^n\,\mathrm{d}G(\lambda)}$$

- ***Assorted examples***
  - If $Y=\min(X_1,X_2)$ and $Z=\max(X_1,X_2)$, finding the covariance between them is simplified by writing $Z = Y + W$, where W is independent of Y. Note that W is a mixture distribution; it either takes the parameter of X_1 or of X_2, with probabilities defined at the start of this section ("Important properties of the exponential").
  - $N(t) = N_1(t) + N_2(t)$. Probability that first event of combined process is from N_1 is $\mathbb{P}\left(N_1(t)=1 \mid N(t)=1\right)=\mathbb{P}\left(N_1(t)=1,N_2(t)=0\right)/\mathbb{P}\left(N(t)=1\right)$.
  - Find the joint distribution of S_1, S_2, S_3. We know the joint distribution of X_1, X_2, X_3, and we know that $\left(S_1,S_2,S_3\right)=\left(s_1,s_2,s_3\right)\Leftrightarrow\left(X_1,X_2,X_3\right)=\left(s_1,s_2-s_1,s_3-s_2\right)$.
  - $N(t)$ is Poisson process with rate λ. Mark first event after s. If haven't stopped by T, lose. If event occurs before T and after mark, lose. $\mathbb{P}(\text{Winning})=\mathbb{P}\left(1\text{ event }\in(s,T)\right)=\lambda(T-s)e^{-\lambda(T-s)}$.
  - Cars arriving with Poisson process; cross if next car will arrive in more than T minutes. Condition on t till next car $\mathbb{E}\left(\text{Wait}\right)=\int_0^\infty \mathbb{E}\left(\text{Wait}\mid\text{Next in }x\right)\lambda e^{-\lambda x}\,\mathrm{d}x$, and note that $\mathbb{E}\left(\text{Wait}\mid\text{Next in }x\right)=\left\{x+\mathbb{E}\left(\text{Wait}\right)\right\}\mathbb{I}_{x\le T}$.
  - Shocks occur with Poisson process. Each might cause failure (independently) with probability p. Can consider two Poisson processes: "failing events", with rate λp, and "non-failing events", with rate $\lambda(1-p)$. Num shocks $\mid T=t\sim\mathrm{Po}\left(\lambda(1-p)t\right)$.
  - Number of trials is $\mathrm{Po}(\lambda)$. Each results in outcome i with p_i. X_j is total number of outcomes that occurs j times. Write $X_j=\sum I_i$, where I_i is 0 if event i occurs j times.
  - $f_{S_i}\left(x\mid N(t)=n\right)$: use the fact that S_i is the i^{th} smallest of n random variables. $i-1$ of these variables must be $>x$ than i, and $n-i$ of them must be $<x$. Thus

$$f_{S_i}(x\mid N(t)=n)=\frac{n!}{(i-1)!(n-i)!}\left[F(x)\right]^{i-1}\left[\bar{F}(x)\right]^{n-1}f(x)$$

    Where f is the distribution of a uniform $(0, t)$.

- $M/G/\infty$ queue. Type I event if departs in $(s, s+t)$. It's a Poisson process, with the right rate. To prove independent, type I if leaves in I_1, type II if leaves in I_2 ($I_1\cap I_2=\varnothing$) and type III otherwise. All independent.
- Note that

$$\mathrm{Cov}\left(N(t_1),N(t_2)\right)=\mathrm{Cov}\left(N(t_1),N(t_1)+N(t_2)-N(t_1)\right)=\mathrm{Cov}\left(N(t_1),N(t_1)\right)$$

Markov Chains

- We assume everything is time homogeneous
- $i\sim j$ means that $\mathbb{P}_j(T_i<\infty)>0$.
- ***Definition (Positive Recurrence)***: $\mathbb{P}_i\left(T_i<\infty\right)=1$ means the state is recurrent. If $\mathbb{E}_i(T_i)<\infty$ means that it's positive recurrent.
- ***Result***: If the chain is *irreducible*, the following are equivalent
  - Some state is positive recurrent
  - All states are positive recurrent (and therefore the chain is positive recurrent)
  - The chain has an invariant distribution π.
- The "ratio" formula for the steady state distribution is

$$\pi_i=\mathbb{P}_\pi\left\{X_n=i\right\}=1/\mathbb{E}_i(T_i)$$

  In the finite state space, it is enough to solve the following set of equations to find the steady state distribution

$$\pi^\top P=\pi^\top \qquad\qquad \pi_+=1$$

  Whereas for a countable state space, we need those two equations *and* either $\pi\ge\boldsymbol{0}$ *or* $\sum\left|\pi\right|<\infty$.
- ***Result***: For a finite state space, suppose we have $P_{ij}^n\to\pi_j$ as $n\to\infty$ for some j, i, then π_j is the invariant measure of the state j.
- ***Definition***: A state i is *aperiodic* if $P_{ii}^n>0$ eventually.
- ***Theorem***: For a irreducible, aperiodic Markov Chain for which there is a stationary distribution π,

$$P_{ij}^n\to\pi_j \qquad \text{as } n\to\infty\ \forall i,j\in\mathcal{S}$$

  In other words, $\mathbb{P}_i\left(X_n=j\right)\to\mathbb{P}_\pi\left(X_0=j\right)$.
- Let X_n: $n\ge0$ be a Markov Chain on a state space S which is irreducible [countable state space] Let λ be any distribution over S. The nfor X_0 ~ \lambda,

$$\frac{N_i(n)}{n}=\frac{\sum_{k=1}^n \mathbb{1}_{\{X_k=i\}}}{n}\to_{a.s.}\frac{1}{\mathbb{E}_i(T_i)}$$

  For all $i\in\mathcal{S}$.
- ***Theorem***: Assume the same setup as previous ergodic theorem and in addition that the chain is positive recurrent and aperiodic. Then for any bounded f for $f:\mathcal{S}\to\mathbb{R}$, then

$$\frac{1}{n}\sum_{i=1}^n f(X_i)\to\mathbb{E}_\pi f(X_0)=\sum_{k\in\mathcal{S}}\pi_k f(k)$$

- ***Introduction & Transition Probabilities***
  - $\mathbb{P}\left(X_{n+1}=j\mid X_n=i,X_{n-1}=i_{n-1},\cdots\right)=P_{ij}$ This is called the *Markovian property*.
  - We let P denote the matrix of one-step transition probabilities P_{ij}. Since the contents of this matrix are probabilities, we have $P_{ij}\ge0 \qquad \sum_{\forall j}P_{ij}=1$
  - The n-step transition probabilities P_{ij}^n give the probability that a process in state i will be in state j after n additional transitions $P_{ij}^n=\mathbb{P}\left(X_{n+m}\mid X_m=i\right) \qquad n\ge0,\ i,j\ge0$

    Chapman-Kolmogorov Equations: The n-step transition probabilities can be found using matrix multiplication $P^{n+m}=P^nP^m \qquad \forall n,m\ge0$

    Proof: Consider that

$$\begin{aligned}P_{ij}^{n+m}&=\mathbb{P}\left(X_{n+m}=j\mid X_0=i\right)=\sum_{\forall k}\mathbb{P}\left(X_{n+m}=j,X_n=k\mid X_0=i\right)\\&=\sum_{\forall k}\mathbb{P}\left(X_{n+m}=j\mid X_n=k,X_0=i\right)\mathbb{P}\left(X_n=k\mid X_0=i\right)\\&=\sum_{\forall k}\mathbb{P}\left(X_{n+m}=j\mid X_n=k,X_0=i\right)\mathbb{P}\left(X_n=k\mid X_0=i\right)=\sum_{\forall k}P_{kj}^mP_{ik}^n\end{aligned}$$

- ***Types of Markov chains***
  - Markov chains come in different types and flavours. The main distinction is
    - *Irreducible/ergodic Markov chains* in which every state can be reached from every other state.
    - *Non-irreducible Markov chains*, in which some states cannot be reached from other states. An example is chains containing one or more absorbing element i, for which $P_{ii}=1$.
  - ***Classes***
    - ***Definition (Communicating states)***: State j is said to be *accessible* from state i if, for some $n\ge0$, $P_{ij}^n>0$. If two states are accessible to each other, we write $i\leftrightarrow j$ and say states i and j communicate. Notes:
      - Communication is an equivalence relation. In other words: (1) $i\leftrightarrow i$ (2) $i\leftrightarrow j\Leftrightarrow j\leftrightarrow i$ (3) $i\leftrightarrow j$ and $j\leftrightarrow k\Rightarrow i\leftrightarrow k$
      - Two states that communicate are said to be in the same *class*. Any two classes are either disjoint or identical, so the set of states is partitioned into classes.
    - ***Definition (irreducible chain)***: If a Markov chain contains only one class, it is clearly irreducible, since every item in the class can be reached from every other item.
  - ***Properties of states***
    - ***Definition (Period)***: State i is said to have period d, denoted $d(i)$, if $P_{ii}^n=0$ whenever n is not divisible by d, and d is the greatest integer with this property. (1) A state with period 1 is said to be *aperiodic*. (2) If $P_{ii}^n=0$ for all $n>0$, then we say the period of i is infinite.

      Theorem: Period is a class property; if $i\leftrightarrow j$, then $d(i)=d(j)$.

      Proof: Suppose $P_{ii}^s>0$; in other words, s is divisible by $d(i)$. Furthermore, choose some m and n such that $P_{ji}^mP_{ij}^n>0$; this is clearly possible since the staes communicate. Then $P_{jj}^{n+m}\ge P_{ji}^nP_{ij}^m>0\Rightarrow P_{jj}^{n+m+s}\ge P_{ji}^nP_{ii}^sP_{jj}^m>0$. To see why these inequalities hold, consider that

      to obtain the RHS, we add an additional constraint to the LHS. For example, the LHS of the first equation talks of getting from j back to j in $n+m$ steps. The RHS talks of going from j to i in n steps, and then back to j in m additional steps.

      These two equations imply that $d(j)$ divides both $n+m$ and $n+m+s$; and therefore also divides their difference, which is equal to s. Thus any s that is divisible by $d(i)$ is also divisible by $d(j)$. A similar argument yields the other direction, giving $d(i)=d(j)$.

      Intuitively, we are simply finding a way of getting from state j back to itself via i, and then realize that we can add s to that number no problem by looping at i an appropriate number of times.
    - ***Definition (Recurrence)***: For any states i and j, define f_{ij}^n to be the probability that, starting in i, the first transition into j occurs at time n.

$$f_{ij}^0=0 \qquad\qquad f_{ij}^n=\mathbb{P}\left(X_n=j,X_k\ne j\ \forall k=1,\cdots,n-1\mid X_0=i\right)$$

      We let $f_{ij}=\sum_{n=1}^\infty f_{ij}^n$. It is the probability of ever making a transition into state j, given the process starts in i. We say state j is *recurrent* if and only if $f_{jj}=1$, and *transient* otherwise.

      Theorem: Sate j is recurrent if and only if $\sum_{n=1}^\infty P_{jj}^n=\infty$

      Proof: State j is recurrent if and only if $\mathbb{E}\left(\text{Number of visits to }j\mid X_0=j\right)=\infty$ (if the state were not recurrent, each time the state was entered, there would be a fixed probability of never returning there, leading to a geometrically distributed number of visits with finite mean). We can then write

$$\mathbb{E}\left(\text{Number of visits to }j\mid X_0=j\right)=\sum_{n=0}^\infty\mathbb{E}\left(\mathbb{I}_{\{X_n=j\}}\mid X_0=j\right)=\sum_{n=0}^\infty P_{jj}^n$$

      Theorem: Recurrence is a class property; if $i\leftrightarrow j$ and i is recurrent, then j is recurrent.

      Proof: Intuitively, if we can get from j to i and back (since $i\leftrightarrow j$) and we know we'll always get back to i if we leave from there, then we can necessarily also always get back to j.

      More formally, find m and n such that $P_{ij}^m>0$ and $P_{ji}^n>0$, we then have $P_{jj}^{m+n+s}\ge P_{ji}^nP_{ii}^sP_{ij}^m>0$. As such $\sum P_{jj}^{m+n+s}\ge P_{ji}^nP_{ij}^m\sum P_{ii}^s=\infty$
    - Another way of defining a recurrent state is to let T_i be the first return time to state i when the system starts at $X_0=i$. An event is transient if and only if $\mathbb{P}\left(T_i=\infty\right)>0$. Even for recurrent events, the *expectation* of T_i need not be finite. Indeed, we define...

      Definition (positive recurrence and null recurrence): Let

$$\mu_{ii}=\mathbb{E}(T_i)=\begin{cases}\infty & \text{if } i \text{ is transient}\\ \sum_{n=1}^\infty nf_{ii}^n & \text{if } i \text{ is recurrent}\end{cases}$$

      We say event i is *positive recurrent* if $\mu_{ii}<\infty$, and *null recurrent* otherwise.

      Theorem: Null recurrence is a class property.
  - ***Types of Markov Chains – more formally***
    - ***Definition (aperiodicity)***: A class is said to be *aperiodic* if it has period 1. A Markov chain is said to be *aperiodic* if every one of its classes has period 1.
    - ***Definition (irreducibility)***: A Markov chain is said to be *irreducible* if it has a single class.
    - ***Definition (ergodicity)***: A state is said to be *ergodic* if it is aperiodic and positive recurrent. A Markov chain is said to be *ergodic* if every one of its states is ergodic.
- ***Irreducible Markov chains***
  - A probability distribution $\boldsymbol{\pi}$ is said to be *stationary* for a Markov chain with probability transition matrix P if $\boldsymbol{\pi}=\boldsymbol{\pi}P$, where $\boldsymbol{\pi}$ is considered to be a *row* vector.
  - ***Theorem***: An irreducible, aperiodic Markov chain belongs to one of the following two classes:
    - *Either* all the states are transient or null recurrent [in which case, $\lim_{n\to\infty}P_{ij}^n=0\ \ \forall i,j$, and there exists no stationary distribution].
    - *Or* all states are positive recurrent, with $\pi_j=\lim_{n\to\infty}P_{ij}^n>0$

      $\boldsymbol{\pi}$ is a stationary distribution, and there exists no other stationary distribution. Notes: (1) In that case, the chain is ergodic. (2) The distribution $\boldsymbol{\pi}$ can be found by solving $\boldsymbol{\pi}=\boldsymbol{\pi}P$ together with $\sum\pi_i=1$. (3) Note that if the Markov chain has a finite number of states, this is the only option. (4) It turns out that $\pi_j=1/\mu_{jj}$; the long-term probability of transitioning into a state is the reciprocal of the expected time till the state is returned to. Thus, if $\pi_j=0$, our state is null-recurrent.
  - If the chain is *periodic*, things change slightly. The stationary vector $\boldsymbol{\pi}$ is still the *long run proportion of time* the Markov chain is in state j, but because of the periodicity, this is now equal to $\lim_{n\to\infty}\frac{P_{ii}^{nd}}{d}=\pi_i$. To see why, consider that P_{ii}^{nd} is the average proportion of time spent in state i *amongst all the other states in the same "periodicity class"*. We need to divide by d to ensure we are accurately finding the proportion of time spent in *all* the states.
- ***Non-irreducible Markov chain***
  - We now consider non-irreducible Markov chain. The interesting case is one in which there are some recurrent classes. We partition the transition matrix as follows: $P=\begin{pmatrix}\iota & 0\\ R & Q\end{pmatrix}$. Where Q only contains transient states (ie: it denotes the probabilities of motion amongst the transient states), and R contains the one-step absorption probabilities. The matrix ι will be equal to I if all recurrent states are absorbing; otherwise, it represents the probabilities of motion amongst the recurrent states).
  - ***Theorem***: If ρ is a recurrent class of states and $i\in\rho$, $j\ne\rho$, then $P_{ij}=0$.

    Proof: i and j are in different classes, and therefore do not communicate. Thus, if we had $P_{ij}>0$, we would have to have $P_{ji}^n=0$ for all n. Hence, if the process starts at i, there is a positive probability it might end up in j and then never return to i; this contradicts the fact i is recurrent.
  - The question arises, therefore, of how many times we visit any given state before being absorbed. Let N_{ij} be the expected number of times spent in transient j, given we started in state i.

    Now, we write $N_{ij}=\sum_{k=0}^\infty\mathbb{P}\left(\text{End up in }j\text{ after }k\text{ moves if started at }i\right)=\mathbb{I}_{\{i=j\}}+Q_{ij}+Q_{ij}^2+Q_{ij}^3+\cdots$

    Since Q is a matrix of transient states, its rows do not sum to 1 and $Q^n\to0$, so the equation above makes sense. In matrix notation, we have $N=I+Q+Q^2+\cdots\Rightarrow(I-Q)N=I\Rightarrow N=(I-Q)^{-1}$

    This is called the *fundamental matrix* of the Markov chain.
  - We can also find the expected number of step until absorption given a start in state i – denote this M_i. Clearly, this is simply the sum of the times we will spend in each transient state. Thus, $\boldsymbol{M}=\boldsymbol{N1}$ Another, equivalent, way of obtaining this quantity is writing each equation manually: $M_i=1+\sum_{\forall\nu}P_{i\nu}P_{\nu\to\text{recurrent state}}$
  - Similarly, let $B_{i\ell}$ be the probability of being absorbed into state ℓ given a start in state i. We can write

$$\begin{aligned}B_{i\ell}&=\sum_{k=0}^\infty\mathbb{P}\left(\text{Absorbed into state }\ell\text{ after step }k\text{ if started at }i\right)\\&=R_{i\ell}+\sum_{\forall\nu}Q_{i\nu}R_{\nu\ell}+\sum_{\forall\nu,\mu}Q_{i\nu}Q_{\nu\mu}R_{\mu\ell}+\cdots\end{aligned}$$

    In matrix form $B=R+QR+Q^2R+\cdots=(I+Q+Q^2+\cdots)R=NR$. Again, in the absence of a computer, we often need to do this manually. This can be done using a recurrence relation

$$B_{i\ell}=\sum_{k\in\text{Transient}}P_{ik}B_{k\ell}+\sum_{\substack{k\in\text{States}\\ \text{communicating}\\ \text{with }\ell}}P_{ik}$$

- ***Time-Reversible Markov Chains***
  - Take a Markov chain whose initial state is chosen to be equal to $\boldsymbol{\pi}$. We can then imagine the chain was started at $t=-\infty$. The "reverse chain" is then also a Markov chain, with transition probabilities

$$P_{ij}^*=\mathbb{P}\left(X_n=j\mid X_{n+1}=i\right)=\frac{\mathbb{P}\left(X_{n+1}=i\mid X_n=j\right)\mathbb{P}\left(X_n=j\right)}{\mathbb{P}\left(X_{n+1}=i\right)}=\frac{P_{ji}\pi_j}{\pi_i}$$

  - We say a Markov chain is *time reversible if* $P=P^*$; in other words, P satisfies the *detailed balance equations* $\pi_iP_{ij}=\pi_jP_{ji}\quad\forall i,j$. These equations imply $\boldsymbol{\pi}=\boldsymbol{\pi}P$ (to see why, simply sum over j); thus, it is enough to check for the second equation. These equations can be interpreted as stating that the rate at which the process goes from i to j is the same as the reverse rate.
  - Consider a graph with a positive number w_{ij} associated with each edge. Suppose a particle currently on vertex i moves to vertex j with probability $P_{ij}=w_{ij}/\sum_{\forall j}w_{ij}$. This is a Markov chain, and if it is irreducible, then it is, in steady state, time-reversible with stationary probabilities given by

$$\pi_i=\sum_{\forall j}w_{ij}/\sum_{\forall i}\sum_{\forall j}w_{ij}$$

    Proof: detailed balance equations reduce to $\pi_iw_{ij}/\sum_kw_{ik}=\pi_jw_{ji}/\sum_kw_{jk}$. We have, that $w_{ij}=w_{ji}$, and so $\pi_i/\sum_kw_{ik}=\pi_j/\sum_kw_{jk}\Rightarrow\pi_i\propto\sum_kw_{ik}$. Since the π must sum to 1, the result is as above.
  - ***Theorem***: If we have a Markov chain with transition probabilities P, and we can find a vector $\boldsymbol{\pi}$ and a matrix P^* such that $\pi_iP_{ij}=\pi_jP_{ji}^*$, then the $\boldsymbol{\pi}$ are stationary probabilities and P^* is the transition matrix of the reversed chain.
- ***Assorted examples***
  - Consider a 3 by 3 grid in which an unfortunate mouse is free to move (but not diagonally). The cells are labeled 1 to 9, left-to-right and then top-down. This is a periodic markov chain with period 2, since we can only ever go from an odd state to an even state and vice-versa. We can work out various limiting probabilities intuitively
    - $\boxed{(P^{2k})_{2,2}\approx\frac{1}{4}}$, because by symmetry, the probability of being in each even state is equal in the long-run. So $P_{2,2}^{2k}\approx P_{2,4}^{2k}\approx P_{2,6}^{2k}\approx P_{2,8}^{2k}$, and the sum is equal to 1.
    - $\boxed{(P^{2k+1})_{1,2}=\frac{1}{4}}$, because the long run probability of being in any even state, starting at any state, is equal. Or in other words, $P_{1,2}^{2k+1}=P_{1,2}P_{2,2}^{2k}+P_{1,4}P_{4,2}^{2k}$.
    - $\boxed{(P^{2k})_{1,1}\approx\frac{1}{6}}$, because $(P^{2k})_{1,1}=P_{1,2}^{2k-1}P_{2,1}+P_{1,4}^{2k-1}P_{4,1}$.
    - $\boxed{(P^{2k})_{5,5}\approx\frac{1}{3}}$, because $P_{5,5}^{2k}=\sum_{i=\{2,4,6,8\}}P_{5,i}^{2k-1}P_{i,5}=4\cdot\frac{1}{4}\cdot\frac{1}{3}$. Alternatively, note that the probability of being in any odd cells adds up to 1, and the probability of being in each of the cells 1, 3, 7 and 9 is $\frac{1}{6}$.
    - $\boxed{(P^{2k})_{1,5}\approx\frac{1}{3}}$, because $P_{1,5}^{2k}=\sum_{i=\{2,4,6,8\}}P_{1,i}^{2k-1}P_{i,5}$.
  - Consider an M/G/1 queue. Let X_n be the number of customers left in the queue *just after the n^{th} customer has departed*. This is an *embedded* Markov chain. Let Y_n denote the number of customers that arrive during the service time of the n^{th} customer: $X_{n+1}=\begin{cases}X_n-1+Y_{n+1} & \text{if } X_n>0\\ Y_{n+1} & \text{if } X_n=0\end{cases}$ Now, clearly, $\mathbb{P}\left(Y_n=j\right)=\int_0^\infty e^{-\lambda x}\frac{(\lambda x)^j}{j!}\mathrm{d}G(x)=a_j$, and $P_{ij}=\begin{cases}\mathbb{P}\left(Y_n=j\right)=a_j & i=0,j\ge0\\ \mathbb{P}\left(Y_n=j-i+1\right)=a_{j-i+1} & i\ge1,j\ge i-1\\ 0 & \text{otherwise}\end{cases}$

    The chain is clearly irreducible. Let's try and check whether it's positive recurrent or transient. We have that, for $j\ge0$: $\begin{aligned}\pi_j&=\sum_{i=0}^\infty\pi_iP_{ij}\\ \pi_j&=\pi_0a_j+\sum_{i=1}^{j+1}\pi_ia_{j-i+1}\end{aligned}$. We introduce the generating functions $\hat{\pi}(s)=\sum_{j=0}^\infty\pi_js^j$ and $\hat{A}(s)=\sum_{j=0}^\infty a_js^j$. We then multiply the above equation by s^j on both sides, and sum over j

$$\begin{aligned}\hat{\pi}(s)&=\pi_0\hat{A}(s)+\sum_{j=0}^\infty\sum_{i=1}^{j+1}\pi_ia_{j-i+1}s^j\\ \hat{\pi}(s)&=\pi_0\hat{A}(s)+\tfrac{1}{s}\sum_{j=1}^\infty\pi_is^i\sum_{j=i-1}^\infty a_{j-i+1}s^{j-i+1}\\ \hat{\pi}(s)&=\pi_0\hat{A}(s)+\tfrac{1}{s}\sum_{j=1}^\infty\pi_is^i\sum_{\alpha=0}^\infty a_\alpha s^\alpha\\ \hat{\pi}(s)&=\pi_0\hat{A}(s)+\frac{\pi(s)-\pi_0}{s}\hat{A}(s)\\ \hat{\pi}(s)&=\frac{(s-1)\pi_0\hat{A}(s)}{s-\hat{A}(s)}\end{aligned}$$

    To find π_0, we let $s\to1$, and note that $\lim_{s\to1}\hat{A}(s)=\sum_{j=0}^\infty a_j=1$, whereas $\lim_{s\to1}\hat{\pi}(s)=\sum_{j=0}^\infty\pi_j$. Using L'Hopital's rule, we find: $\lim_{s\to1}\hat{\pi}(s)=\pi_0\left[1-A'(1)\right]^{-1}$. Consider, however, that

$$A'(1)=\sum_{j=0}^\infty ja_j=\mathbb{E}\left(\text{Arrivals during service time}\right)=\lambda\mathbb{E}\left(\text{Service time}\right)=\rho$$

    Thus $\sum_{j=0}^\infty\pi_j=\frac{\pi_0}{1-\rho}$. For the chain to be recurrent, this sum must be equal to 1. Thus, this can only occur when $\rho<1$, in which case $\pi_0=1-\rho$. In that case, we get

$$\hat{\pi}(s)=\frac{\left(1-\rho\right)\left(s-1\right)\hat{A}(s)}{s-\hat{A}(s)}=\frac{\left(1-\lambda\mathbb{E}(S)\right)\left(s-1\right)\hat{A}(s)}{s-\hat{A}(s)}$$

    Now, note that:
    - We can take the derivative of this, and let $z\to1$. This gives us the expected number of people in the system. The details are horrible, and involve multiple applications of L'Hopital's Rule $\mathbb{E}\left(Q_\infty^D\right)=\rho+\frac{\rho^2}{1-\rho}\frac{\left(c_s^2+1\right)}{2}$. Where $c_s^2=\mathrm{Var}(S)/\mathbb{E}(S)^2$.
  - Let $\mathbb{P}\left(X_i=j\right)=a_j$, where $j=0,\pm1,\pm2,\cdots$. Let $S_0=0 \qquad S_n=\sum_{i=1}^nX_i$. Then S_n is a Markov chain, for which $P_{ij}=a_{j-i}$. Consider the situation in which $a_1=p$ and $a_{-1}=1-p$. It is then clear that the chain is irreducible and periodic with $d=2$; thus, $P_{00}^{2n+1}=0$. However, using $n!\sim n^{n+\frac{1}{2}}e^{-n}\sqrt{2\pi}$, we have $P_{00}^{2n}=\binom{2n}{n}p^n(1-p)^n=\frac{(2n!)}{n!n!}\left[p(1-p)\right]^n\sim\frac{1}{\sqrt{\pi n}}\left[4p(1-p)\right]^n$. The series $\sum P_{00}^{2n}$ therefore only converges if $4p(1-p)\le1$, which only occurs at p = ½. Thus, the symmetric simple random walk is recurrent. Other simple random walks are transient.
  - Suppose a battery fails at period i with probability p_i, where p_i as aperiodic and $\sum_iip_i<\infty$. Let X_n denote the age of the battery in use at period n. This is a Markov chain, with probability $P_{i,1}=1-P_{i,i+1}=p_i \qquad i\ge1$. To find the stationary distribution, solve $\boldsymbol{\pi}=\boldsymbol{\pi}P$. In this case, this reduces to $\pi_1=\sum_i\pi_ip_i\Rightarrow\pi_{i+1}=\pi_i(1-p_i) \qquad i\ge1$. Iterating the second equation, we obtain $\pi_{i+1}=\pi_1\prod_{j=1}^i(1-p_j)=\pi_1\sum_{j=i+1}^\infty p_j=\pi_1\mathbb{P}\left(X\ge i+1\right)$. Using the fact that the sum of all the π has to be 1, we get that $\pi_1=1/\mathbb{E}(X)$. Thus, we find $\pi_i=\mathbb{P}\left(X\ge i\right)/\mathbb{E}(X)$, which can easily be seen to satisfy the first equation above.
  - During each time period, every member of a population dies with p, and $\mathrm{Po}(\lambda)$ new members join. X_n, the population during period n, is a Markov chain. Note that if $X_0=\alpha$, then $X_1\sim\mathrm{Po}\left[\alpha(1-p)+\lambda\right]$. Thus, if we choose α to satisfy $\alpha=\alpha(1-p)+\lambda$ [ie: $\alpha=\lambda/p$], the chain is stationary, with $\pi_j=\mathbb{P}\left[\mathrm{Po}(\frac{\lambda}{p})=j\right]$.
  - Let $P_{ij/k}^n$ be the probability that, starting in state i, the Markov chain will end up in state j without ever passing through state k. We can show that $P_{ij}^n=\sum_{k=0}^nP_{ij/i}^{n-k}P_{ii}^k$. Intuitively, we can split our journey from i to j into two parts – the first part perambulating around and possibly returning to i a number of times, and the last part finally heading from i to j. To prove, let Y be the last time we leave state i. Now $P_{ij}^n=\mathbb{P}(X_n=j\mid X_0=i)=\sum_{k=0}^n\mathbb{P}(X_n=j,Y=k\mid X_0=i)$ $=\sum_{k=0}^n\mathbb{P}(X_n=j,Y=k,X_k=i\mid X_0=i)=\sum_{k=0}^n\mathbb{P}(X_n=j,Y=k\mid X_k=i)\mathbb{P}(X_k=i\mid X_0=i)$ $=\sum_{k=0}^nP_{ij/i}^{n-k}P_{ii}^k$ To go from the second to the third step, the thought-process involves the fact that we're conditioning on the position we reach at time k, so that should somehow be involved...

Continuous-Time Markov Chains

- ***Introduction***
  - $\mathbb{P}\left(X(s+t)=j\mid X(s)=i,X(r)=i_r,r\in A_s\subseteq[0,s)\right)=\mathbb{P}\left(X(s+t)=j\mid X(s)=i\right)$
  - $\mathbb{P}\left(X(s+t)=j\mid X(s)=i\right)$ is a *transition probability*. We will typically assume *stationary* or *homogeneous transition probabilities*: $\mathbb{P}\left(X(s+t)=j\mid X(s)=i\right)=\mathbb{P}\left(X(t)=j\mid X(0)=i\right)=P_{i,j}(t)$
  - We define *limiting probabilities* as follows $\alpha_j\equiv\lim_{t\to\infty}P_{i,j}(t)=\lim_{t\to\infty}\mathbb{P}\left(X(t)=j\mid X(0)=i\right)$. Providing the chain is irreducible, α_j does not depend on the initial state.
- ***Modeling CTMCs***
  - We consider four methods of modeling CTMCs.
  - ***Method 1 – DTMC with Exponential Transition Times***
    - Our first method involves modeling our CTMC as a DTMC in which the time till a transition out of state i is exponentially distributed with rate ν_i.
    - Though this is not necessary, we assume that there are no one-step transitions from any state to itself – in other words, the diagonal elements of the transition probabilities of the DTMC $(P_{i,i})$ are 0.
  - ***Method 2 – Transition Rates and ODEs***
    - We define $Q_{i,j}=P_{i,j}'(0^+)=\lim_{t\downarrow0}\left.\frac{\mathrm{d}P_{i,j}}{\mathrm{d}t}\right|_t$

      It is more usual to define

$$P(0) = I$$
$$P_{i,j}(h) = Q_{i,j}h + o(h) \qquad \text{as } h \downarrow 0 \text{ if } j \neq i$$
$$P_{i,i}(h) - 1 = Q_{i,i}h + o(h) \qquad \text{as } h \downarrow 0$$

For a finite state space, the above implies that $-Q_{i,i} = \sum_{j,j\neq i} Q_{i,j}$. Furthermore, since we have assumed $P_{i,i} = 0$ in the first approach, we have $-Q_{i,i} = \nu_i$.

- This kind of approach is useful when we have a rate diagram that shows the exponential rates of going from one step to another. $Q_{i,j}$ is then equal to the exponential rate of going from one state to another (with negative diagonal elements).

- ***Method 3 – Competing Jobs with Exponential Timers***
  - Method 1 worked because *all exits from state i* were exponentially distributed with the *same* rate ν_i. We generalize this to a case in which exits to different states have different means.
  - For each state i, we associate a clock to each state j to which the process could move, and we let $T_{i,j}$ be the time at which that clock goes off. Whenever a clock goes off, we move to the corresponding state and reset all active clocks at the new state. By the lack-of-memory property of the exponential distribution, resetting running timers is equivalent to not resetting the times.
  - Let's analyze this. Let $T_i = \min_j(T_{i,j})$ $\qquad N_i = \operatorname{argmin}_j(T_{i,j})$. By the properties of the exponential distribution, T_i and N_i are independent, and $T_i \sim \text{Exp}(\nu_i)$ and $\mathbb{P}(N_i = j) = \frac{Q_{i,j}}{\nu_i}$

    Where $\nu_i \equiv -Q_{i,i} = \sum_{j,j\neq i} Q_{i,j}$. As such, we can model our CTMC as a DTMC with transition probabilities $Q_{i,j} / \nu_i$ and waiting time ν_i.
  - We can also go the other way. Given a DTMC with transitions P and inter-transition times ν_i, we can write $Q_{i,j} = \nu_i P_{i,j}$ $\qquad Q_{i,i} = -\sum_{j,j\neq i} Q_{i,j} = \nu_i$
  - A CTMC is uniquely defined by a given Q. It is also uniquely defined by a given $(P,\boldsymbol{\nu})$ provided $P_{i,i} = 0$ (as we have assumed). Otherwise, we can transform our chain into one satisfying this property by defining a new DTMC transition matrix $\hat{P}$. We let $\hat{P}_{i,i} = 0$ and

    $$\hat{P}_{i,j} = \mathbb{P}\left(i \text{ goes to } j \mid i \text{ doesn't go to } i\right) = \frac{P_{i,j}}{1 - P_{i,i}}$$

    We the change the holding time at i to $\hat{\nu}_i > \nu_i$, to reflect the fact we removed "stay at i" transitions $1/\hat{\nu}_i = \frac{1/\nu_i}{1 - P_{i,i}} \Rightarrow \hat{\nu}_i = \nu_i(1 - P_{i,i})$

    Theorem: This leaves the chain unchanged.

    Proof: Note that (1) transition rates determine transition probabilities (2) transition probabilities determine finite-dimensional distributions (3) finite-dimensional distributions are the probability law of the CTMC. So all we need to show is that the transition rates are unchanged. $\hat{Q}_{i,j} = \hat{\nu}_i\hat{P}_{i,j} = \nu_i(1 - P_{i,i})\frac{P_{i,j}}{1 - P_{i,i}} = \nu_i P_{i,j}$

- ***Method 4 – A DTMC with Poisson Transitions***
  - Imagine, in the first approach, that all the ν_i were equal. In that case, we could represent our CTMC as a DTMC with transitions governed by an independent Poisson process. In fact, if the mean transition time is $1/\nu_0$ for all states, $\{Y_n : n \geq 0\}$ is a DTMC with transition matrix P and $\{N(t) : t \geq 0\}$ is a Poisson process with rate ν_0, then $X(t) = Y_{N(t)} \qquad t \geq 0$

    And $P_{i,j}(t) = \mathbb{P}\left(X(t) = j \mid X(0) = i\right) = \sum_k P_{i,j}^k \mathbb{P}\left(N(t) = k\right) = \sum_k P_{i,j}^k \frac{e^{-\nu_0 t}(\nu_0 t)^k}{k!}$
  - Any finite-state CTMC can be represented in this way by *uniformization*, which involves introducing *fictitious transitions* from states to themselves.

    We generate *potential transitons* from a Poisson process with rate $\lambda \geq \nu_i \ \forall i$, and we then independently thin this Poisson process by making the transition a "real transition" with probability ν_i / λ and a "fictitious transition" with probability $1 - (\nu_i / \lambda)$

    Specifically, we define

    $$\tilde{P}_{i,j} = \frac{Q_{i,j}}{\lambda} \qquad\qquad \tilde{P}_{i,i} = 1 - \frac{\nu_i}{\lambda} = 1 - \frac{\sum_{j,j\neq i} Q_{i,j}}{\lambda} = 1 + \frac{Q_{i,i}}{\lambda}$$

    Or, in matrix notation, $\tilde{P} = I + \lambda^{-1}Q$

    Theorem: Performing the uniformization process above leaves the chain unchanged.

- ***Limiting probabilities***
  - Limiting probabilities are not the same as stationary probabilities.
    - A limiting distribution is $\lim_{t\to\infty} X(t)$
    - A stationary distribution is a vector $\boldsymbol{\beta}$ such that $\mathbb{P}\left(X(t) = j\right) = \beta_j$ for all t and j whenever $\mathbb{P}\left(X(0) = j\right) = \beta_j$.
  - ***Theorem***: For irreducible, finite-state CTMC, the story is simpler. There exists a unique stationary vector, which is also a limiting vector.
  - We consider four methods for finding the limiting probabilities.
    - ***Theorem***: Given a CTMC characterized as a DTMC with transition matrix $\tilde{P}$ and transitions according to a Poisson process with rate λ

      $$\alpha_j = \tilde{\pi}_j \quad \forall j$$

      Where $\tilde{\pi}$ is the unique solution to

      $$\tilde{\pi} = \tilde{\pi}\tilde{P} \qquad\qquad \tilde{\pi}\boldsymbol{1} = 1$$
      $$\text{or } \sum_i \tilde{\pi}_i\tilde{P}_{i,j} = \tilde{\pi}_j \qquad\qquad \sum_j \tilde{\pi}_j = 1$$
    - ***Theorem***: Give a CTMC characterized in terms of a DTMC with one-step transition matrix P and exponential transition times with means $1/\nu_i$

      $$\alpha_j = \frac{\pi_j / \nu_j}{\sum_k \pi_k / \nu_k}$$

      Where π is the unique solution is

      $$\pi = \pi P \qquad\qquad \pi\boldsymbol{1} = 1$$
    - ***Theorem***: Given a CTMC characterized by its transition-rate matrix Q, α is the unique solution to

      $$\alpha Q = 0 \qquad\qquad \alpha\boldsymbol{1} = 1$$
      $$\text{or } \sum_i \alpha_i Q_{i,j} = 0 \qquad\qquad \sum_i \alpha_i = 1$$

      These are "global-balance equations" – all probabilities of transition into a state must be equal to transitions out of it.
    - ***Theorem***: Given a CTMC characterized by its transition function $P(t)$, α is the unique solution to

      $$\alpha P(t) = \alpha \text{ for any } t > 0 \qquad\qquad \alpha\boldsymbol{1} = 1$$
      $$\sum_i \alpha_i P_{i,j}(t) = \alpha_j \text{ for all } j \qquad\qquad \sum_i \alpha_i = 1$$

- ***Birth and Death Processes***
  - Birth and death processes are those in which only transitions up and down one state is possible. In such processes, the global-balance equations $\alpha Q = 0$ take the simpler form of *detailed-balance equations*.
  - ***Theorem***: For birth-and-death processes, the limiting probability vector $\boldsymbol{\alpha}$ is the unique solution to the detailed-balance equations $\alpha_j\lambda_j = \alpha_{j+1}\mu_{j+1} \ \forall\ 0 \leq j \leq n-1$. With $\alpha\boldsymbol{1} = 1$. This reflects conservation of flow again; transitions from j to $j+1$ must be equal to transitions from $j+1$ to j.
  - ***Theorem***: For a birth-and-death process with state space $\{0,1,\cdots,n\}$

    $$\alpha_j = \frac{r_j}{\sum_{k=0}^n r_k} \qquad 0 \leq j \leq n \text{ where } r_0 = 1 \qquad r_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}$$

    Proof: We simply note that the detailed balance equations require that $\alpha_j = \frac{\lambda_{j-1}}{\mu_j}\alpha_{j-1}$ and so $\alpha_j = r_j\alpha_0$. This is precisely the above. To get the final form, we use the fact that $\alpha\boldsymbol{1} = 1$.
  - ***Assorted examples and applications***
    - ***Blocking***: Consider a barbershop with two chairs (exponential services times of rate μ) and three waiting spaces. Customers arrive at Poisson λ. Any customers arriving when both waiting chairs are full are lost:
      - This is a birth-and-death process with states 0, 1, 2, 3, 4 and 5.

      - In each case, growths have $Q_{i,i+1} = \lambda$, except for state 5, which has growth 0 – this incorporates blocking.
      - $Q_{1,0} = \mu$, but all the other decreasing Q are $Q_{i,i-1} = 2\mu$ because two customers are being served at the same time.

      The long run proportion of time each barber is busy, for example, can be found as $\frac{1}{2}\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5$.
    - ***Balking***: Consider the situation above, but in which an arriving customer finding both barbers taken will stay with probability β and balk otherwise.
      - We then model this precisely as above, but the Poisson arrival process for $2 \to 3$, $3 \to 4$ and $4 \to 5$ will be a thinned Poisson process, with rate $\beta\lambda$.

      The PASTA (Poisson Arrivals See Time Averages) principle is also useful in calculating some quantities of interest. It states that the proportion of customers that see some state upon arrival coincides with the proportion of times the process spends in that state, provided the arrival process is Poisson.

      Thus, for example, the proportion of customers that balk is $(1-\beta)(\alpha_3 + \alpha_4 + \alpha_5)$. The proportion that are blocked is simply α_5. And the proportion that enter upon arrival is $1 - (1-\beta)(\alpha_2 + \alpha_3 + \alpha_4) - \alpha_5$.
    - ***Abandoning***: Consider the situation above but in which any customer having to wait more than X for service will leave, where $X \sim \text{Exp}(\theta)$.
      - The death rate for $1 \to 0$ is still μ. The death rate for $2 \to 1$ is still 2μ. The death rate for $3 \to 2$, however, is now $2\mu + \theta$, because it is the minimum of *three* variables; loss of patience and either barber finishing service. Similarly, the death rate $4 \to 3$ is $2\mu + 2\theta$, because it is the minimum of *four* variables – either customer losing patience and either barber finishing.

      The rate of customer abandonment can then be calculated as follows:
      - When the process is in 1 or 2, customers are not abandoned.
      - When the process is in 3, abandonment occurs as $\sim \text{Exp}(\theta)$, and so customers are abandoned at rate θ per hour.
      - When the process is in 4, customers are abandoned as $\sim \text{Exp}(2\theta)$, and so customers are abandoned at a rate 2θ per hour.
      - Similarly for 5.

      Thus, the rate of customer abandonment is $\theta\alpha_3 + 2\theta\alpha_4 + 3\theta\alpha_5$. To find the proportion of customers that abandon, we just divide by the arrival rate.

      Similarly, if we wanted to find the long-run proportion of potential customers eventually served, we could simply calculate 1 minus balking, blocking and abandoning. Another way is to note that the service rate is θ in state 1 and 2θ is other states, whereas the arrival rate is λ, so the total proportion served is $\frac{\theta\alpha_1 + 2\theta(\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5)}{\lambda}$. These give the same answer.
    - ***The M/M/1 queue***: Arrivals are rate λ, services times are $\sim \text{Exp}(\mu)$ and there are in infinite number of waiting spaces. The number of customers in the queue is $Q(t)$, which is a birth-and-death process with infinite state space. To guard against pathologies, we require $\rho \equiv \lambda/\mu < 1$. Applying an obvious extension of the birth-and-death process in this case, we get $\alpha_j = \frac{r_j}{\sum_{k=1}^{\infty} r_k}$ $\qquad r_j = \rho^j$. This implies that α is the *geometric distribution* – in other words $\lim_{t\to\infty} \mathbb{P}\left(Q(t) = j \mid Q(0) = i\right) = \alpha_j = (1-\rho)\rho^j$ which has mean $\rho/(1-\rho)$.

      If we had, instead, considered the *M/M/1/r* queuing model (with one server and r waiting spaces, we would have $r + 2$ states with a truncated geometric distribution $\alpha_j = \frac{(1-\rho)}{(1-\rho^{r+2})}\rho^j$ $\qquad 0 \leq j \leq r+1$. This applies without limits on ρ.

- ***Reverse-time CTMC***
  - A stochastic process $\{X(t) : -\infty < t < \infty\}$ is said to be *reversible* if it has the same probability law as $\{X(-t) : -\infty < t < \infty\}$.
  - We define *reverse transition probabilities* as

    $$\overleftarrow{P}_{i,j}(t) \equiv \mathbb{P}\left(X(s) = j \mid X(s+t) = i\right) = \frac{\mathbb{P}\left(X(s) = j\right)P_{j,i}(t)}{\mathbb{P}\left(X(s+t) = i\right)}$$

    Unfortunately, these do not necessarily need to be bona-fide probabilities that sum to 1. If, however, we start our chain with its stationary distribution α, the above becomes $\overleftarrow{P}_{i,j}(t) = \frac{\alpha_j P_{j,i}(t)}{\alpha_i}$. Which, it can be shown, *does* sum to 1. Similarly, the transition-rate matrix for the reverse chain is $\overleftarrow{Q}_{i,j} = \frac{\alpha_j Q_{j,i}}{\alpha_i}$
  - This, however, is not enough for the chain to be time reversible. It can be shown that the extra required condition is for $\overleftarrow{Q} = Q$, or equivalently $\alpha_i Q_{i,j} = \alpha_j Q_{j,i} \qquad \forall i,j$

    Effectively, this requires all steady-state transitions from $i \to j$ to be equal to those from $j \to i$.
  - It can be shown that all birth-and-death processes are reversible.
  - We can also consider *truncated processes*, in which all transitions out of a subset of states $A \subset S$ are disallowed. In other words, we set $Q_{i,j} = 0$ for $i \in A, j \in S \setminus A$, and as usual, we set $Q_{i,i} = -\sum_{j\in A} Q_{i,j}$

    Theorem: The truncated process is also reversible and has stationary probability vector $\alpha_j^{(A)} = \frac{\alpha_j}{\sum_{k\in A}\alpha_k}$ $\qquad \forall j \in A$ where α is the stationary probability vector of the original process.
  - ***Assorted examples***
    - The departures from an *M/M/s* queue (with $\rho \equiv \lambda / s\mu < 1$) in equilibrium is also a Poisson process, with departure rate equal to λ. To see why, simply note the CTMC consisting of the number of people in the queue – $Q(t)$ – is a reversible CTMC. Running it backwards changes arrivals into departures.

Inventory management

- ***Notation***: λ is the demand rate, L is the order lead time, q is the batch size, r is the re-order point, $D[t, u)$ is the demand in that interval, $I(t) = IN(t)^+$ is the inventory on hand, $B(t) = IN(t)^-$ are the backorders outstanding, $IN(t) = I(t) - B(t)$ is the net inventory, $A(t)$ is the stockout indicator $\mathbb{1}(IN(t) \leq 0)$, $IO(t)$ is the inventory on order, $IP(t) = IN(t) + IO(t)$ is the inventory position.
- ***Performance measures***: $\overline{A}$ average stockout frequency. $\overline{B}$ average back-orders. $\overline{I}$ average inventory. $\overline{OF}$ average order frequency. I is the equilibrium inventory, IN is the equilibrium net inventory, A is the equilibrium stockout indicator, B equilibrium back-orders, IP equilibrium inventory position, $D = D[t, t + L]$ leadtime demand (Poisson λL). $\overline{BW}$ average customer back-order waiting time. $\overline{IW}$ time a unit spends in inventory.
- General strategy; find IN (for which we need IP and D) and use it to find I, A and B.
- ***Base-stock policy (q=1)***: [Order 1 batch when IP falls below r] Let $s = r + 1$. $IO(t) = D[t - L, t) =$ Poisson (λL). $IP(t)$ is always kept at s, and so IN(t) = s – IO(t) ; IN = s – IO. Alternatively, IN(t+L) = s – D(t, t+L) and IN = s – D, where $D =_d$ D(t, t+L].
- ***Performance measures***: The g denote the PMF (probability mass function) of D, G^0 its inverse CDF and $G^1(d) = \mathbb{E}[(D-d)^+] = \sum_{j\in[d+1,\infty]} G^0(j) = \lambda L - \sum_{j\in[0,d)} G^0(j)$. We then get $\overline{A} = \mathbb{P}(IN \leq 0) = \mathbb{P}(D \geq s) = G^0(s-1)$. $\overline{B} = \mathbb{E}[IN^-] = \mathbb{E}[(D-s)^+] = G^1(s)$. $\overline{I} = \mathbb{E}[IN^+] = \mathbb{E}[IN + IN^-] = s - \lambda L + G^1(s)$. Of course, $\overline{OF} = \lambda$. For Poisson demands, $\overline{A}$ is also the proportion of demands that ends up having to wait (by PASTA). $\overline{BW} = \overline{B}/\lambda$ is the average customer waiting time and $\overline{IW} = \overline{I}/\lambda$ is the average stocking time (the time a unit spends in inventory).
- ***Notes***: (1) Everything only depends on λL (2) $\overline{B}$ is decreasing and convex as function of s (3) $\overline{I}$ is increasing and convex in s (4) $\overline{A}$ is decreasing in s and convex only when $\lambda L \leq 1$ [not trivial]
- ***General batch sizes (any (r,q))***: (A) $IP \sim U[r+1, r+q]$ (find corresponding DTMC) (B) $IN(t+L) = IP(t) - D(t, t+L]$ (IN at $t + L$ is equal to IN at t plus IO at t; which will get there by $t + L$) (C) IN = IP – D.
- ***Performance measures***: Now, we know [IN | IP = s] is as given in the base-stock policy. But we know the distribution of IP is uniform. So, for eg: $\overline{A}(r,q) = \frac{1}{q}\sum_{s=r+1}^{r+1}\overline{A}(s)$. Same is true for $\overline{B}$ and $\overline{I}$. Writing $G^2(d) = \sum_{j>d}^{\infty} G^1(j) = \frac{1}{2}(\lambda L)^2 - \sum_{j\in[0,d]}^{\infty} G^1(j)$, we also have $\overline{A} = \frac{1}{q}\left[G^1(r) - G^1(r+q)\right]$, $\overline{B} = \frac{1}{q}\left[G^2(r) - G^2(r+q)\right]$ and $\overline{I} = \frac{1}{2}(q+1) - r - \lambda L + \overline{B}$. $\overline{OF} = \lambda / q$ since orders come in at rate λ and orders are made every q demands.
- ***Notes***: (1) As functions of r, they're all as above. $\overline{OF}$ is decreasing and convex as a function of q. (2) $\overline{B} = \lambda\overline{BW}$ and $\overline{I} = \lambda\overline{IW}$; application of Little's Law, and true not just for Poisson demands and constant leadtimes, provided the averages exist.
- ***Theorem***: If BW denotes the limiting distribution of waiting times, and $D(BW)$ denote the demand during BW. Then if $r \geq -1$, $B = D(BW)$. In particular, this gives $\mathbb{E}[B] = \lambda\mathbb{E}[BW]$ and $\mathbb{V}\text{ar}[B] = \lambda\mathbb{E}[BW] + \lambda^2\mathbb{V}\text{ar}[\text{BW}]$. – can be shown that for a given s $\frac{1}{2}\mathbb{E}[B(B-1)] = G^2(s)$. We can weigh that across different s to get the quantity generally, and use that to find the variance of B.

Other

- **Exponential**: $f(x) = \lambda e^{-\lambda x}$; $F(x) = 1 - e^{-\lambda x}$; $\mathbb{E} = \lambda^{-1}$; $\mathbb{V} = \lambda^{-2}$; $M(\theta) = \frac{\lambda}{\lambda - \theta}$.
  - $X \sim \exp(\lambda)$, $\qquad Y \sim \exp(\mu)$: $\qquad \min(X,Y) \sim \exp(\lambda + \mu)$ $\quad$ [$\mathbb{P}(\min \geq z) = \mathbb{P}(X \text{ and } Y \geq z)$] $\max(X,Y) = (X+Y) - \min(X,Y)$; $\mathbb{P}(\min(X,Y) = X) = \lambda/(\lambda+\mu)$ [$= \mathbb{P}(X < Y)$, condition on $Y = y$ and integrate]. $\{\min(X,Y) = X\} \perp \{\min(X,Y) > t\}$ [take $\mathbb{P}$ of both events, condition on $X = x$, integral then goes from $t \to \infty$].
- **Normal**: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$; $\qquad M(\theta) = \exp\left[\mu\theta + \frac{1}{2}\sigma^2\theta^2\right]$; $\qquad \mathbb{E}[Z^3] = 4$, $\qquad \mathbb{E}[Z^4] = 10$; $\frac{\sigma}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-x^2/2\sigma^2} \leq \overline{\Phi}(x) \leq \frac{\sigma}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2\sigma^2}$; $z\phi(z) = -\phi'(z)$.
- **Gamma** (α,λ): $f(x) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}$; $\mathbb{E} = \frac{\alpha}{\lambda}$; $\mathbb{V} = \frac{\alpha}{\lambda^2}$; $M(\theta) = (\frac{\lambda}{\lambda-\theta})^\alpha$; sum of α $\exp(\lambda)$ RV. $\Gamma(n) = (n-1)!$
- **Lognormal**: $f(x) = \frac{1}{x}\phi(\log x)$; $F(x) = \Phi(\frac{1}{\sigma}[\log x - \mu])$; $\mathbb{E} = \exp(\mu + \frac{1}{2}\sigma^2)$
- **Beta**: $x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta)$ mean $\alpha/(\alpha+\beta)$ variance $\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1)$
- **Binomial**: $\mathbb{P}(=k) = {}^nC_k p^k(1-p)^{n-k}$; $\mathbb{E} = np$; $\mathbb{V} = np(1-p)$; $M(\theta) = (1 - p + pe^\theta)^n$
- **Negative binomial** (r,p): Number of **successes** before r failures occur (success probability p). $\mathbb{P}(=k) = {}^{k+r-1}C_k(1-p)^r p^k$ (support 0, 1, …). $\mathbb{E} = \frac{pr}{1-p}$; $\mathbb{V} = \frac{pr}{(1-p)^2}$; $M(\theta) = \left(\frac{1-p}{1-pe^\theta}\right)^r$ $(\theta < -\log t)$. When r = 1, geometric.
- **Poisson**: $\mathbb{P}(=k) = \frac{1}{k!}\lambda^k e^{-\lambda}$ sum of Poisson RVs results in sum of parameters (prove using MGF). Mean and variance λ. MGF $M(\theta) = \exp(\lambda(e^t - 1))$

Other

- $\left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-x^2/2} \leq \int_x^{\infty} e^{-y^2/2}\,\mathrm{d}y \leq \frac{1}{x}e^{-x^2/2}$
- $f(x) \sim g(x)$ if $\lim_{x\to\infty}\frac{f(x)}{g(x)} = 1$. Doesn't imply either function converges.
- Hazard function is $\lambda(t) = f(t)/\overline{F}(t)$
- $\text{Cov}(X,Y) = \mathbb{E}\{(X-\mu_X)(Y-\mu_Y)\} = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- $\mathbb{V}\text{ar}\left(\sum_i X_i\right) = \sum_i \mathbb{V}\text{ar}(X_i) + 2\sum_{i<j}\text{Cov}(X_i, X_j)$
- When dealing with moments, recall that $\sum_{n=0}^{\infty} z^n = (1-z)^{-1}$
- Markov: $\mathbb{P}\left(X \geq a\right) \leq \mathbb{E}\left(X\right)/a$ [$a\mathbb{I}_{X\geq a} \leq X$, take expectations].
- Jensen: $\mathbb{E}\left(f(x)\right) \geq f\left(\mathbb{E}(X)\right)$ for convex f [take taylor series, second derivative +ve]
- Tail-integral formula: For a non-negative RV, $\mathbb{E}(X^n) = \int_0^{\infty} nt^{n-1}\overline{F}(t)\,\mathrm{d}t$. For a non-negative integer value RV, $\mathbb{E}(N)\sum_{i=1}^{\infty}\mathbb{E}(N \geq i)$.
- For conditional var, $\mathbb{V}\text{ar}(X) = \mathbb{E}(S^2) - \mathbb{E}(S)^2$, write in conditional form, replace the first by $\mathbb{V}\text{ar}(S \mid N)$.
- ${}^nC_k = n!/\left[k!(n-k)!\right]$
- MGF: $M(t) = \mathbb{E}[e^{tX}]$ and CF = $M(it)$
- The k^{th} order statistic of n U[0,1] RVs is $B(k, n+1-k)$
- $e^x = 1 + x + x^2/2$ $\qquad$ log(1-x) = - sum_{n=1}^{\infty} x^n / n
- Geometric: $\sum_{k=0}^{n-1} ar^k = a\frac{1-r^n}{1-r}$,
- Law of total variance: $\operatorname{var}(Y) = \operatorname{E}(\operatorname{var}(Y \mid X)) + \operatorname{var}(\operatorname{E}(Y \mid X))$.

Sums, etc…

- a_n and b_n only positive and $a_n / b_n \to L \neq 0$ then if sum b_n converges, so does sum a_n. An example of this is if $a_n \sim b_n$
- If $0 \leq \lim a_{n+1} / a_n < 1$, then sum a_n converges
- Two possibly non real sequences a_n and b_n. If a_n → *0* and b_n → infty and a_nb_n → c then $(1 + a_n)^{b_n}$ → e^c
- If $\lim_{x\to\infty} f(x) = l$ and $\lim_{x\to l} g(x) = L$ then $\lim_{x\to\infty} g(f(x)) = L$, provided g defined and continuous at l.
- $\sum 1/n^2 = \pi^2/6$

Introduction

- ***Definition (Strong Stationarity)***: The process $\{X(t) : t \geq 0\}$ is *stationary* if

  $$\left(X(t_1), X(t_2), \cdots, X(t_n)\right) =_d \left(X(t_1+n), X(t_2+n), \cdots, X(t_n+n)\right)$$

  For all $t_1 < t_2 < \cdots < t_n$ and $n \geq 0$. In other words, a process is stationary if it is "shift invariant".
- ***Definition (Weak Stationarity)***: The process $\{X(t) : t \geq 0\}$ is said to be *covariance stationary* if $\mathbb{E}\left[X(t)\right] = m$ (ie: it has constant mean) and $\text{Cov}\left[X(t), X(s)\right] = \mathbb{E}\left[(X(t)-m)(X(s)-m)\right] = R\left(\left|t-s\right|\right)$, where R is some function of $\left|t-s\right|$ only.
- Note that a Gaussian process is entirely defined by its mean and covariance; as such, for a Gaussian process, stationarity implies covariance stationarity.

Theorems

- ***Theorem (Weak Ergodic Theorem)***: If $\{X_n\}$ is covariance stationary, then $\overline{X}_n \to_{L_2} \overline{X}$, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Note that in general, $\overline{X}$ will be a random variable. Under one of two equivalent conditions, it is a deterministic constant equal to $\mathbb{E}(X_1) = m$:
  - $\mathbb{V}\text{ar}(\overline{X}_n) \to 0$
  - $\mathbb{C}\text{ov}\left(\overline{X}_n, X_1\right) \to 0$
- ***Theorem (Strong Ergodic Theorem)***: If $\{X_n\}$ is stationary and $X_n \in L_1$, then $\overline{X}_n \to \overline{X}$ a.s. and $\overline{X}_n \to_{L_1} \overline{X}$. Again, $\overline{X}$ is generally a random variable. Under *ergodicity* (defined below), it is constant. An intermediate step in proving this is…

  Lemma (Maximal Ergodic Lemma): Let $\{Y_n\}$ be stationary and let $S_{i,n} = Y_i + \cdots + Y_{i+n-1}$ and $M_{i,n} = \max\{0, S_{i,1}, \cdots, S_{i,n}\}$. Then $\mathbb{E}[Y_0 Y \mathbf{1}_{\{M_{0,n}>0\}}] \geq 0 \ \forall n \geq 1$.
- ***Definition (Shift Operator)***: φ is a shift operation. If $x = (x_1, x_2, \cdots)$ then $\varphi(x) = (x_2, x_3, \cdots)$. We say the set A is *shift invariant* if $x \in A \Rightarrow \varphi(x) \in A$. For example, the following two sets are shift invariant (we can replace $\leq$ by $\geq$ and =, and lim sup by lim inf)

  $$A_1 = \left\{x : \lim_{n\to\infty}\tfrac{x_1+\cdots+x_n}{n} \leq a\right\} \qquad\qquad A_2 = \left\{x : \limsup_{n\to\infty}\tfrac{x_1+\cdots+x_n}{n} \leq a\right\}$$
- ***Definition (Ergodicity)***: $X = \left(X_1, \cdots, X_n, \cdots\right)$ is ergodic if $\mathbb{P}\left(X \in A\right) = 0$ or 1 for all shift invariant A.

  Note: We can use this definition of ergodicity to show that $\overline{X}$ is constant.

  $$\mathbb{P}\left(X \in A\right) = \mathbb{P}\left(\lim_{n\to\infty}\overline{X} \leq a\right) =_{a.s.} \mathbb{P}\left(\overline{X} \leq a\right) = 0 \text{ or } 1$$

  Thus, $\overline{X}$ is a constant.
- ***Definition (Mixing)***: $X = \left(X_1, X_2, \cdots, X_n, \cdots\right)$ is *mixing* if

  $$\lim_{n\to\infty}\mathbb{P}\left((X_1, X_2, \cdots) \in A, (X_{n+1}, X_{n+2}, \cdots) \in B\right) = \mathbb{P}\left((X_1, X_2, \cdots) \in A\right) \cdot \mathbb{P}\left((X_1, X_2, \cdots) \in A\right)$$

  Where A and B are sets of infinite sequences (not necessarily shift invariant). We can look at pieces of size k instead $\lim_{n\to\infty}\mathbb{P}\left((X_n, \cdots, X_{n+k}) \in A, (X_{n+m}, \cdots, X_{n+m+k}) \in B\right)$.

  Theorem: If a sequence is mixing, it is ergodic.

  Proof: Let A be any shift invariant set

  $$\begin{aligned} p &= \mathbb{P}\left((X_0, X_1, \cdots) \in A\right) \\ &=_{A \text{ shift invariant}} \mathbb{P}\left((X_0, X_1, \cdots) \in A, (X_n, X_{n+1}, \cdots) \in A\right) \\ &\to_{\text{mixing}} \mathbb{P}\left((X_0, X_1, \cdots) \in A\right) \cdot \mathbb{P}\left((X_0, X_1, \cdots) \in A\right) \\ &= p^2 \end{aligned}$$

  And so p must be equal to 0 or 1. ■
- ***Example***: Let $X_0 = Y$ w.p. p and Z w.p $1-p$, where $Y, Z \in L_1$. Consider two cases:
  - If $X_n \equiv X_0$, $\overline{X} = Y$ w.p p and Z w.p $1-p$
  - If $X_n =_d X_0$, $\overline{X} = \mathbb{E}(Y)$ w.p p and $\mathbb{E}(Z)$ w.p $1-p$

  Neither case is ergodic.