Elementary Probability

- **Conditional Probability:** Probability of event A occuring given that event B occured $P(A|B) = \frac{P(A\cap B)}{P(B)}$ provided $P(B) > 0$. Note we get the useful property $P(A \cap B) = P(A|B)P(B)$.
- **Chain Rule:** $P(A_1 \cap A_2 \cap \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$
- **Independence:** Events A and B are independent iff $P(A \cap B) = P(A)P(B)$. RVs X and Y independent iff for all $a, b \in \mathbb{R}$, $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$ iff $p(x, y) = p_X(x)p_Y(y)$ for all $x, y \in \mathbb{R}$.
- **Law of Total Probability:** If $\{B_i\}_{i=1}^n$ partitions the sample space Ω (i.e., $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\cup_{i=1}^n B_i = \Omega$), then $P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i)P(B_i)$.
- **Bayes' Formula:** If $\{B_i\}_{i=1}^n$ partitions the sample space Ω, then $P(B_j|A) = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}$.
- **Expectation:** For $g : \mathbb{R} \to \mathbb{R}$ and an RV X, then $\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x)p(x) & \text{if } X \text{ is a discrete RV} \\ \int_{-\infty}^{\infty} g(x)p(x)\,\mathrm{d}x & \text{if } X \text{ is a continuous RV} \end{cases}$.

  Note special cases include

  (a) $\mathbb{E}[X^n]$ is the nth moment of X.

  (b) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$, i.e., expectation is linear.

  (c) If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

  (d) $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ is the variance of X.

  (e) $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ is the covariance of X and Y. Note $\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$. Note also $\mathrm{Cov}(X, Y) = 0$ if X and Y are independent.

  (f) $\phi_X(t) = \mathbb{E}[e^{tX}]$ is the moment generating function of X. In the joint case, $\phi_{X,Y}(s, t) = \mathbb{E}[e^{sX+tY}]$ is the mgf of (X, Y). Note that the nth derivative of ϕ_X satisfies $\phi_X^{(n)}(0) = \mathbb{E}[X^n]$ and the (m, n)th derivative of $\phi_{X,Y}$ satisfies $\phi_{X,Y}^{(m,n)}(0, 0) = \frac{\partial^{m+n}}{\partial s^m \partial t^n}\phi_{X,Y}(s, t)\Big|_{s=t=0} = \mathbb{E}[X^mY^n]$. The MGF also uniquely characterizes the corresponding probability distribution.
- **Marginal Distributions:** Where X and Y are RVs with joint pdf $f(x, y)$, the marginal distribution of the single RV X is given by $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y$.
- **cdf → pdf:** for an RV X, $f(x) = \lim_{\Delta\downarrow 0} \frac{P(x \leq X \leq x+\Delta)}{\Delta}$.
- **mgf of sum:** If $X_1, X_2, \ldots, X_n$ are independent RVs and $T = \sum_{i=1}^n X_i$, then the mgf of T is the product of the mgfs of X_is: $\phi_T(t) = \prod_{i=1}^n \phi_{X_i}(t)$. In particular, if $X_1, \ldots, X_n$ are iid, then $\phi_T(t) = (\phi_{X_1}(t))^n$.
- **Strong Law of Large Numbers:** If $X_1, X_2, \ldots, X_n$ are iid RVs with common mean μ and $\mathbb{E}[|X_1|] < \infty$, then $\bar{X}_n = \frac{X_1+X_2+\cdots+X_n}{n} \to \mu$ as $n \to \infty$.

Conditional Distributions

- **Conditional Distribution:** The conditional pdf of $X|(Y = y)$ is $f_{X|Y}(x|y) = P(X|Y = y) = \frac{f(x,y)}{f_Y(y)}$. The conditional cdf is $P(a \leq X \leq b|Y = y) = \int_a^b f_{X|Y}(x|y)dx = \frac{\int_a^b f(x,y)dx}{f_Y(y)} = \frac{\int_a^b f(x,y)dx}{\int_{-\infty}^{\infty} f(x,y)dx}$
- **Conditional Expectation:** $\mathbb{E}[g(X, Y)|Y = y] = \begin{cases} \sum_x g(x, y)p_{X|Y}(x|y) & \text{if } X \text{ is a discrete RV} \\ \int_{-\infty}^{\infty} g(x, y)f_{X|Y}(x|y)\,\mathrm{d}x & \text{if } X \text{ is a continuous RV} \end{cases}$.

  Note we also see that $\mathbb{E}[aX + bY|Z = z] = a\mathbb{E}[X|Z = z] + b\mathbb{E}[Y|Z = z]$. It's also worth noting that $\mathbb{E}[g(X, Y)|Y = y]$ is a function of the conditioning value y. Note also that $\mathbb{E}[g(X|Y)] = \mathbb{E}[g(X)|Y]$.

- **Conditional Variance:** $\mathrm{Var}(X|Y=y) = \mathbb{E}[(X-\mathbb{E}[X|Y=y])^2] = \mathbb{E}[X^2|Y=y] - \mathbb{E}[X|Y=y]^2$.
- **Law of Total Expectation:** $\mathbb{E}[g(X)] = \mathbb{E}_Y[\mathbb{E}_X[g(X)|Y]] = \mathbb{E}_Z[\mathbb{E}_Y[\mathbb{E}_X[g(X)|Y,Z]|Y] = \cdots$.
- **Variance Alternate Formula:** $\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$.
- $f_X(x) = \begin{cases} \sum_y f_{X|Y}(x|Y=y) f_Y(Y=y) & \text{if } Y \text{ if is a discrete RV} \\ \int_{-\infty}^{\infty} f_{X|Y}(x|Y=y) f_Y(Y=y)\,\mathrm{d}y & \text{if } Y \text{ if is a continuous RV} \end{cases}$
- $P(X<Y) = \int P(X<y) f_Y(y)\,\mathrm{d}y = \int F_X(y) f_Y(y)\,\mathrm{d}y$

Discrete-time Markov Chains

- **Discrete-time Markov Chain:** A stochastic process $\{X_n : n \in \mathbb{N}\}$ is a DTMC if (1) X_n is a discrete RV for all $n \in \mathbb{N}$ and (2) for all $n \in \mathbb{N}$, the Markov property holds:

$$P(X_{n+1} = x_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x_0) = P(X_{n+1} = x_{n+1}|X_n = x_n).$$

  The Markov property indicates that we only care about the most up-to-date information.

- **Transition Probability Matrix:** The TPM defined by $P_{i,j} = P(X_1 = j|X_0 = i)$ is an $S \times S$ matrix where $S \in \mathbb{N} \cup \{\infty\}$ is the number of states in the DTMC. Note since the entries are probabilities, we see that the rows must sum to 1. We define $P_{i,j}^{(n)} = P(X_{m+n} = j|X_m = i) = (P^n)_{i,j}$ for any $m, n \in \mathbb{N}$.
- **Stationarity of DTMC:** We only consider stationary DTMCs where the distribution of X_m and X_n are identical for all $m, n \in \mathbb{N}$. As a result, $P_{i,j}^{(n)} = P(X_n = j|X_0 = i)$ and so the one-tep TPM completely characterizes its DTMC.
- **Chapman-Kolmogorov Equation:** For any $m, n \in \mathbb{N}$, $P_{i,j}^{(n)} = \sum_{k \in \mathcal{S}} P_{i,k}^{(m)} P_{k,j}^{(n-m)}$. As a result, $P^{(n)} = P^n$
- **Initial Conditions:** For $n \in \mathbb{N}$, let $\underline{\alpha}_n = (\alpha_{n,0}, \alpha_{n,1}, \ldots, \alpha_{n,k}, \ldots)$ be the row vector $\alpha_{n,k} = P(X_n = k)$ (i.e., $\underline{\alpha}_n$ is the marginal pmf of X_n). Then $\underline{\alpha}_0$ is the initial conditions of the DTMC and $\alpha_{n,k} = \sum_{i=0}^{\infty} \alpha_{m,i} P_{i,k}^{(n-m)}$ for all $m, n \in \mathbb{N}$. In particular, $\underline{\alpha}_n = \underline{\alpha}_0 P^n$.
- **Communication:** State j is *accessible* from state i (denoted $i \to j$) if $\exists n \in \mathbb{N}$ such that $P_{i,j}^{(n)} > 0$. If $i \to j$ and $j \to i$, then states i and j *communicate* denoted by $i \leftrightarrow j$. Communication defines an equivalence relation (i.e., reflexivity $i \leftrightarrow i$, symmetry $i \leftrightarrow j \iff j \leftrightarrow i$, and transitivity $i \leftrightarrow j, j \leftrightarrow k \implies i \leftrightarrow k$) and so we usually separate DTMCs into the sets of states which communicate with each other (called *communication classes*). If a DTMC has only one communication class, then it is said to be *irreducible.*
- **Period:** The period of state i is $d(i) = \gcd\{n \in \mathbb{Z}^+ : P_{i,i}^{(n)} > 0\}$. If $d(i) = 1$ then state i is *aperiodic.* Note if $P_{i,i} > 0$, then $d(i) = 1$. By convention, if $P_{i,i}^{(n)} = 0$ for all n, then $d(i) = \infty$. The period is shared by communication classes, in particular, if $i \leftrightarrow j$ then $d(i) = d(j)$.
- **First Visit:** We denote $f_{i,j}^{(n)} = P(X_n = j, X_{n-1} \neq j, \ldots, X_2 \neq j, X_1 \neq j|X_0 = i)$ as the probability that the first visit to state j when starting in state i i happens after n steps. Note $f_{i,j}^{(n)} = P_{i,j}^{(n)} - \sum_{k=1}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)}$. We also denote $f_{i,j} = P(\text{DTMC ever visits state } j|X_0 = i)$. Note $f_{i,j} = \sum_{k=1}^{\infty} f_{i,j}^{(k)} \leq 1$.
- **Recurrence:** State i is *recurrent* if $f_{i,i} = 1$, if $f_{i,i} < 1$ then it is *transient.* If a state is recurrent, it will be visited infinitely often, otherwise it will be visited only finitely often. Recurrence is also shared by communication classes, in particular, if $i \leftrightarrow j$ then state i is recurrent if and only if state j is recurrent.

- Let M_i count the number of number of visits to state i (ever). Note then that $M_i \sim \text{GEO}_f(1 - f_{i,i})$ and so

$$\mathbb{E}[M_i | X_0 = i] = \frac{f_{i,i}}{1 - f_{i,i}} = \sum_{n=1}^{\infty} P_{i,i}^{(n)} = \begin{cases} < \infty & \text{if state } i \text{ is transient} \\ \infty & \text{if state } i \text{ is recurrent} \end{cases}$$

- If $i \leftrightarrow j$ and state i is recurrent then $f_{i,j} = 1$.

- **Test for Transience:** If state i is recurrent and state i does not communicate with state j, then $P_{i,j}^{(k)} = 0$ for all $k \in \mathbb{Z}^+$. This provides a test for transience, in that if states i and j do not communicate but $\exists k$ such that $P_{i,j}^{(k)} > 0$, then i is transient.

- **Limiting Behaviour of Transient States:** For any state i and transient state j, $\lim_{n \to \infty} P_{i,j}^{(n)} = 0$.

- **Mean Recurrent Time:** For a recurrent state i, let $N_i = \min\{n \in \mathbb{Z}^+ : X_n = i\}$ denote the number of steps before visiting (or returning to) X_i for the first time. Note then that $P(N_i = n | X_0 = i) = f_{i,i}^{(n)}$ and the *mean recurrent time* of state i is $\mathbb{E}[N_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{i,i}^{(n)}$. We see then that m_i is the average number of steps between successive visits to state i. If $m_i < \infty$ then we say state i is *positive recurrent*, and if $m_i = \infty$ then we say state i is *null recurrent*. Note positive recurrence is also shared by communication classes, in particular, if $i \leftrightarrow j$ and state i is positive recurrent, then j is positive recurrent.

- **Ergodicity:** If a state i is both positive recurrent and aperiodic, then it is called *ergodic*.

- **Recurrence in Finite-State DTMCs:** A finite-state DTMC has at least one recurrent state. As a result, all states from a finite-state irreducible DTMC are recurrent. In a finite-state DTMC, there are no null recurrent states, so any recurrent state is necessarily positive recurrent.

- **Stationary Distribution:** A distribution $\{p_i\}_{i=0}^{\infty}$ is a stationary distribution (of a DTMC) if $\sum_{i=0}^{\infty} p_i = 1$ and $p_j = \sum_{i=0}^{\infty} p_i P_{i,j}$ for all states j. In matrix form, stationarity is the property $\underline{p} = \underline{p}P$. In particular, if $\underline{\alpha}_0 = \underline{p}$, then $X_0, X_1, X_2, \ldots$ are all identically distributed. A stationary distribution exists iff there is at least one positive recurrent state. Note that stationary distributions are not necessarily unique.

- **Basic Limit Theorem:** For an irreducible, recurrent, and aperiodic DTMC, $\lim_{n \to \infty} P_{i,j}^{(n)} = \pi_j = \frac{1}{m_j}$ exists for all i, j. If the DTMC is also positive recurrent, then $\{\pi_j\}_{j=0}^{\infty}$ is the unique stationary distribution and is the unique positive solution to the system of linear equations given by $\begin{cases} \pi_j = \sum_{i=0}^{\infty} \pi_i P_{i,j} & \forall j \in \mathbb{N} \\ \sum_{j=0}^{\infty} \pi_j = 1 \end{cases}$ or written in matrix form, $\underline{\pi} = \underline{\pi}$ and $\underline{\pi}\underline{e}' = 1$. Note that when the DTMC has finitely many states, this system is overspecified, and so any one equation can be dropped. Note that if state j is null recurrent of transient, then $\pi_j = 0$.

- **Doubly Stochastic:** The TPM of a DTMC is doubly stochastic if all row sums of P are 1 (necessary to be a TPM) and all column sums of P are also 1. An irreducible, aperiodic DTMC with $N < \infty$ states and a doubly stochastic TPM has $\pi_j = \frac{1}{N}$ for all $j \in \mathbb{N}$.

- **Interpretation of Limiting Behaviour:** Assuming the conditions of the BLT, after running the DTMC for a "long" time, the probability of finding the process in state j is π_j. However, π_j also represent the long-run fraction of time that the process spends in state j, i.e., the fraction of time steps with $X_n = j$.

- **Shared Properties of Communication Classes:** The following properties are the same for all states i, j from the same communication class
  - $i \leftrightarrow j$ and so $\exists n,\ P_{i,j}^{(n)} > 0$ and $\exists m,\ P_{j,i}^{(m)} > 0$.

- $d(i) = d(j)$.
- i and j are either both recurrent or both transient.
- if i and j are both recurrent, then they are either both positive recurrent or both null recurrent.

- **Galton-Watson Branching Process:** This process models the size of a population when each person has a certain probability of having m offspring in the next generation. Formally, let X_n denote the population of the nth generation of process, where α_m is the probability an individual has m offspring. Note that since $P_{0,0} = 1$ and $P_{i,0} > 0$, we know that state 0 is recurrent and all other states are transient. Let $Z_i^{(j)}$ denote the number of offspring produced by individual i in the jth generation, then $X_n = \sum_{i=1}^{X_{n-1}} Z_i^{(n-1)}$, so X_n is a DTMC. Let $\mu = \mathbb{E}[Z_i^{(j)}]$ and $\sigma^2 = \mathrm{Var}(Z_i^{(j)})$ denote the (common) mean and variance of the number of offspring respectively. Then $\mathbb{E}[X_n] = \mu\mathbb{E}[X_{n-1}]$ and $\mathrm{Var}(X_n) = \sigma^2\mathbb{E}[X_{n-1}] + \mu^2\,\mathrm{Var}(X_{n-1})$. As a result, assuming $X_0 = 1$, then
$$\mathbb{E}[X_n] = \mu^n \qquad \text{and} \qquad \mathrm{Var}(X_n) = \sigma^2\mu^{n-1}\sum_{i=0}^{n-1}\mu^i = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \sigma^2\mu^{n-1}\left(\frac{1-\mu^n}{1-\mu}\right) & \text{if } \mu \neq 1 \end{cases}$$
Let $\pi_0 = \lim_{n\to\infty} P(X_n = 0)$ denote the probability that the population dies out. Then π_0 is the unique solution in $[0, 1)$ to the equation $z = \sum_{j=0}^{\infty}\alpha_j z^j$ (note that $z = 1$ is always a solution to the equation). Note when $\mu \leq 1$ then $\pi_0 = 1$ (i.e., the population is guaranteed to die out). The value $1 - \pi_0$ is the probability that the population will keep growing infinitely. In the general case where $X_0 = n$ then the extinction probability is π_0^n (for π_0 computed in the $X_0 = 1$ case). Note when $\alpha_0 + \alpha = 1$, X_n will stay at 1 for some number of generations (according to a geometric distribution) and then eventually die out.

- **Gambler's Ruin Problem:** The process models the sum of money a gambler has before he either goes bankrupt or wins the jackpot. Formally, let X_n denote the number of units of money the gambler has at time n, we assume that the gambler either wins one unit with probability $p \in (0, 1)$ and loses one unit with probability $q = 1 - p$ each step, until he either goes bankrupt (reaching $X_n = 0$) or wins the jackpot (reaching $X_n = N$ for some $N < \infty$). Note that states 0 and N are recurrent, and the states $\{1, 2, \ldots, N-1\}$ form a transient communication class. Let $G(i)$ denote the probability that starting with $X_0 = i$ the gambler eventually reaches the jackpot N. Consider the TPM and its limiting behaviour:
$$P = \begin{array}{c c} & \begin{array}{cccccccc} 0 & 1 & 2 & 3 & & N-2 & N-1 & N \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad \lim_{n\to\infty} P^{(n)} = \begin{array}{c c} & \begin{array}{ccccccc} 0 & 1 & 2 & & N-1 & N \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \\ N \end{array} & \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1-G(1) & 0 & 0 & \cdots & 0 & G(1) \\ 1-G(2) & 0 & 0 & \cdots & 0 & G(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-G(N-1) & 0 & 0 & \cdots & 0 & G(N-1) \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \end{array}$$
Notice that
$$(p+q)G(i) = pG(i+1) + qG(i-1) \qquad \Longrightarrow \qquad G(i+1) - G(i) = \tfrac{q}{p}(G(i) - G(i-1)) = (\tfrac{q}{p})^i G(1)$$
Then, for any $k = 1, \ldots, N$, adding together the above equations for each $i = 1, 2, \ldots, k-1$ we get a telescoping sum resulting in
$$G(k) - G(1) = \sum_{i=1}^{k-1} G(1)(\tfrac{q}{p})^i \qquad \Longrightarrow \qquad G(k) = G(1)\sum_{i=1}^{k-1}(\tfrac{q}{p})^i = \begin{cases} G(1)\left(\frac{1-(\frac{q}{p})^k}{1-\frac{q}{p}}\right) & \text{if } p \neq q \\ kG(1) & \text{if } p = q \end{cases}$$
Using the above formula and noting $G(N) = 1$, we can find a formula for $G(1)$. Using this newfound formula, we find $G(k) = \begin{cases} \frac{1-(\frac{q}{p})^k}{1-(\frac{q}{p})^N} & \text{if } p \neq q \\ \frac{k}{N} & \text{if } p = q \end{cases}$ for $k = 0, 1, \ldots, N$. Note when $p \leq \frac{1}{2}$, $\lim_{N\to\infty} G(i) = 0$ and when $p > \frac{1}{2}$, $\lim_{N\to\infty} G(i) = 1 - (\frac{q}{p})^i$.

- **Absorbing DTMCs:** Consider an N state DTMC where states $0, 1, \dots, M-1$ are transient and states $M, M+1, \dots, N-1$ are recurrent. The TPM for this DTMC can be expressed as

$$P = \begin{array}{c} \\ 0 \\ M-1 \\ M \\ N-1 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & M-1 & M & N-1 \end{array} \\ \left[\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array}\right] \end{array} = \begin{array}{c} \\ 0 \\ \\ M-1 \\ M \\ \\ N-1 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & & M-1 & M & & N-1 \end{array} \\ \left[\begin{array}{ccc|ccc} Q_{0,0} & \cdots & Q_{0,M-1} & R_{0,M} & \cdots & R_{0,N-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ Q_{M-1,0} & \cdots & Q_{M-1,M-1} & R_{M-1,M} & \cdots & R_{M-1,N-1} \\ \hline 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{array}\right] \end{array}$$

  Such states $M \le i \le N-1$ are called *absorbing*, since once entered the DTMC will never leave them. Note that the tools we will develop can also be applied where there are recurrent communication classes, replacing them with a single absorbing state and then solving their limiting distributions.

- **Absorption Time:** The absorption time of the DTMC is defined as $T = \min\{n \in \mathbb{Z}^+ : M \le X_n \le N-1\}$. Note if T_j denotes the number of remaining steps until absorption given that the current state is j, then $T|(X_0 = i) \sim T_i$ and $T|(X_1 = j, X_0 = i) \sim (1 + T_j)|X_1 = j$.

- **Absorption Probability:** The absorption probability of transient state $0 \le i \le M-1$ into recurrent state $M \le k \le N-1$ is defined as $U_{i,k} = P(X_T = k | X_0 = i) = R_{i,k} + \sum_{j=0}^{M-1} Q_{i,j} U_{j,k}$. Letting $U = [U_{i,k}]$ be a matrix, it satisfies the expression $U = R + QU$ or equivalently $U = (I-Q)^{-1}R$. Notice this also yields the limiting distribution $\lim_{n\to\infty} P^{(n)} = \begin{bmatrix} \mathbf{0} & U \\ \mathbf{0} & I \end{bmatrix}$

- **Mean Absorption Time:** The mean absorption time from state i is defined as $v_i = \mathbb{E}[T | X_0 = i] = 1 + \sum_{j=0}^{M-1} Q_{i,j} v_j$. Letting $\underline{v} = [v_i]$ be a column vector, it satisfies the expression $\underline{v} = \underline{e}' + Q\underline{v}$ or equivalently $\underline{v} = (1 - Q)^{-1}\underline{e}'$.

- **Mean Number of Visits:** Let ℓ be a transient state and define the indicator variable $A_n = \mathbf{1}[X_n = \ell]$. Then the mean number of visits made to state ℓ (including time 0) before absorption given that $X_0 = i$ is defined as $W_{i,\ell} = \mathbb{E}[\sum_{n=0}^{T-1} A_n | X_0 = i] = \delta_{i,\ell} + \sum_{j=0}^{M-1} Q_{i,j} W_{j,\ell}$. Letting $W = [W_{i,\ell}]$ be a matrix, it satisfies the expression $W = I + QW$ or equivalently $W = (I-Q)^{-1}$. Note we also get the following formula for the probability of ever making a future visit to state ℓ: $f_{i,\ell} = \frac{W_{i,\ell} - \delta_{i,\ell}}{W_{\ell,\ell}}$. Combining this with our prior knowledge, we know

$$f_{i,j} = \begin{cases} \frac{W_{i,j} - \delta_{i,j}}{W_{j,j}} & \text{if } i \text{ and } j \text{ are both transient} \\ U_{i,j} & \text{if } i \text{ is transient and } j \text{ is absorbing} \\ \delta_{i,j} & \text{if } i \text{ is absorbing} \end{cases}$$

Poisson Processes

- **Properties of Exponential:** If $X \sim \text{EXP}(\lambda)$, then X has tpf $P(X > x) = e^{-\lambda x}$ and mgf $\phi_X(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.

- **Properties of Erlang:** If $X_1, \dots, X_n$ are iid $\text{EXP}(\lambda)$ RVs, then $Y = \sum_{i=1}^n X_i \sim \text{Erlang}(n, \lambda)$. So Y has tpf $P(Y > y) = e^{-\lambda y} \sum_{j=0}^{n-1} \frac{(\lambda y)^j}{j!}$ and mgf $\phi_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$ for $t < \lambda$.

- **Minimum of Independent Exponentials:** If $\{X_i\}_{i=1}^n$ is a sequence of independent RVs with $X_i \sim \text{EXP}(\lambda_i)$, then $Y = \min\{X_1, \dots, X_n\}$ has distribution $Y \sim \text{EXP}(\sum_{i=1}^n \lambda_i)$. In particular, if $X_1, \dots, X_n$ are iid $\text{EXP}(\lambda)$, then $Y \sim \text{EXP}(n\lambda)$. Note that $P(X_i = \min\{X_1, \dots, X_n\}) = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_n}$. Also worth noting is

$$P(X_1 < X_2 < \cdots < X_n) = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_n} \cdot \frac{\lambda_2}{\lambda_2 + \cdots + \lambda_n} \cdots \frac{\lambda_{n-1}}{\lambda_{n-1} + \lambda_n} \cdot \frac{\lambda_n}{\lambda_n} = \prod_{i=1}^n P(X_i = \min\{X_i, X_{i+1}, \dots, X_n\})$$

  From this we see that $X_1 | (X_1 < X_2 < \cdots < X_n) \sim Y$.

- **Memoryless Property:** An RV X is memoryless iff $P(X > y+z|X > y) = P(X > z)$ iff $P(X > y+z) = P(X > y)P(X > z)$ iff $P(X > Y + Z|X > Y) = P(X > Z)$ for all $y, z \in \mathbb{R}$ and RVs Y and Z. The exponential distribution is the unique memoryless continuous distribution and the geometric distribution is the unique memoryless discrete distribution. As a result of this, we see that $(X - Y)|(X > Y) \sim X$

- **Counting Process:** A stochastic process $\{N(t)\}_{t\geq 0}$ where $N(t)$ represents the number of events that occurred by (continuous) time t. Some properties of a counting process are (1) $N(0) = 0$, (2) $N(t) \in \mathbb{N}$, (3) if $s < t$ then $N(s) \leq N(t)$, and (4) if $s < t$ then $N(t) - N(s)$ counts the number of events that occurred in the interval $(s, t]$. A counting process is said to have *independent increments* if $N(t_1) - N(s_1)$ and $N(t_2) - N(s_2)$ are independent whenever $(s_1, t_1] \cap (s_2, t_2] = \emptyset$. A counting process is said to have *stationary increments* if $N(s + t) - N(s)$ has the same distribution as $N(t)$ for all $S < t$.

- **Big-$\mathcal{O}$ Notation:** $f(x) \in o(h)$ if $\lim_{h\to 0} \frac{f(h)}{h} = 0$, i.e., $f(h)$ goes to 0 faster than h. This is the only big-oh notation we need.

- **Poisson Process:** A counting process $\{N(t)\}_{t\geq 0}$ is said to be a Poisson process at rate λ if (1) it possesses both independent and stationary increments, (2) $P(N(h) = 1) = \lambda h + o(h)$, and (3) $P(N(h) \geq 2) = o(h)$. Intuitively, conditions 2 and 3 imply that the probability of an event occurring in an interval is proportional to the length of the interval and the probability of more than one event occurring in an interval vanishes as the interval becomes increasingly small. Any such Poisson process at rate λ satisfies $N(t) \sim \text{POI}(\lambda t)$.

- **Interarrival Times:** Let T_i be time elapsed between the $(i - 1)$th event and ith event of a Poisson process, then $\{T_i\}_{i=1}^{\infty}$ is a sequence of iid $\text{EXP}(\lambda)$ RVs. We can also reverse this, if $\{X_i\}_{i=1}^{\infty}$ if a sequence of iid $\text{EXP}(\lambda)$ RVs, then $N(t) = \max\{n \in \mathbb{N} : \sum_{i=1}^{n} X_i \leq t\}$ defines a Poisson process at rate λ. Let S_n denote the time elapsed before the nth event occurs, so that $S_n = \sum_{i=1}^{n}$, then clearly $S_n \sim \text{Erlang}(n, \lambda)$. If $N_1(t)$ and $N_2(t)$ define separate Poisson processes at rates λ_1, λ_2 respectively and $S_m^{(1)}, S_n^{(2)}$ denote the time elapsed before the mth event of the first process and nth event of the second process occurring respectively, then $P(S_m^{(1)} < S_n^{(2)}) = \sum_{j=1}^{n-1} \binom{m+j-1}{m-1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^m \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^j$.

- **Splitting Poisson Processes:** If $N(t)$ is a Poisson process at rate λ, and an event of this process is of type i with probability p_i for k total possible types, then we can define k Poisson processes $N_i(t)$ and with the property $N(t) = \sum_{i=1}^{k} N_i(t)$. Moreover, $N_i(t)$ is a Poisson process at rate λp_i. This also works in reverse, if $N_1(t), \ldots, N_k(t)$ are Poisson processes with associated rates $\lambda_1, \ldots, \lambda_k$, then $N(t) = \sum_{i=1}^{k} N_i(t)$ is a Poisson process at rate $\sum_{i=1}^{k} \lambda_i$.

- **Conditional Poisson Process:** Let $N(t)$ be a Poisson process at rate λ. Then $N(s)|(N(t) = n) \sim \text{Bin}(n, \frac{s}{t})$. Suppose $N(t) = n$, then the conditional distribution of the arrival times is the same as the order statistics of a uniform distribution. In particular, $(S_1, \ldots, S_n)|(N(t) = n) \sim (Y_{(1)}, \ldots, Y_{(n)})$ where $Y_1, \ldots, Y_n$ are iid $U(0, t)$ RVs. Note the joint pdf of the ordered uniform RVs is $f(y_1, \ldots, y_n) = \frac{n!}{t^n}$ for $0 < y_1 < y_2 < \cdots < y_n < t$.

- **Non-homogeneous Poisson Process:** A counting process $\{N(t)\}_{t\geq 0}$ is a *non-homogeneous* Poisson process with rate function $\lambda(t)$ if (1) it has independent increments, (2) $P(N(t + h) - N(t) = 1) = h\lambda(t) + o(h)$, and (3) $P(N(t + h) - N(t) \geq 2) = o(h)$. This generalization allows the rate at which events happen to vary as a function of time, however, we lose stationary increments. If $N(t)$ is a non-homogeneous Poisson process with rate function $\lambda(t)$, then $N(t+s) - N(s) \sim \text{POI}(m(t+s) - m(s))$ where $m(t) = \int_0^t \lambda(\tau)d\tau$ is the mean value function for $t \geq 0$.

- **Compound Poisson Process:** Let $\{Y_i\}_{i=1}^{\infty}$ be an iid sequence of RVs and let $N(t)$ be a Poisson process at rate λ, independent of each Y_i. Then $X(t) = \sum_{i=1}^{N(t)} Y_i$ is a *compound* Poisson process and has both independent and stationary increments. While finding $X(t)$'s distribution is in general intractable, $\mathbb{E}[X(t)] = \lambda t\mathbb{E}[Y_1]$ and $\text{Var}(X(t)) = \lambda t\mathbb{E}[Y_1^2] = \lambda t(\text{Var}(Y_1) + \mathbb{E}[Y_1]^2)$.