Distribution	Probability Function	Mean	Variance	MGF
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geometric	$p(y) = p(1-p)^{y-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hypergeometric	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	No closed form
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$	λ	λ	$e^{\lambda(e^t-1)}$
Negative binomial	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$
Uniform	$f(y) = \frac{1}{\theta_2-\theta_1}$	$\frac{\theta_1+\theta_2}{2}$	$\frac{(\theta_2-\theta_1)^2}{12}$	$\frac{e^{t\theta_2}-e^{t\theta_1}}{t(\theta_2-\theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2}$	μ	σ^2	$e^{\mu t + \frac{t^2\sigma^2}{2}}$
Exponential	$f(y) = \frac{1}{\beta} e^{-\frac{y}{\beta}}$	β	β^2	$(1-\beta t)^{-1}$
Gamma	$f(y) = \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right) y^{\alpha-1} e^{-\frac{y}{\beta}}$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$
Chi-square	$f(y) = \frac{y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}$	ν	2ν	$(1-2t)^{-\frac{\nu}{2}}$
Beta	$f(y) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) y^{\alpha-1}(1-y)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	No closed form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{(ad)-(bc)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

MATH447 Crib Sheet

Julian Lore

Basic Probability

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

$$Pr(A) = \sum_k Pr(A \cap B_k)$$

where $B_k \cap B_j = \emptyset \ \forall k \neq j, Pr(\cup_j B_j) = 1$

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{Pr(A|B)P(B) + Pr(A|\overline{B})Pr(\overline{B})}$$

$$Pr(B_i|A) = \frac{Pr(A|B_i)Pr(B_i)}{\sum_j Pr(A|B_j)Pr(B_j)}$$

$$E(Y) = \begin{cases} \sum_y yPr(Y=y) & \text{discrete} \\ \int_{-\infty}^{\infty} y\ f_y(y)\ dy & \text{continuous} \end{cases}$$

$$E(Y) = \sum_{i=1}^{k} E(Y|A_i)P(A_i)$$

$$E(Y|X=x) = \begin{cases} \sum_y yPr(Y=y|X=x) & \text{discrete} \\ \int_{-\infty}^{\infty} y\ f_y(y|x)\ dy & \text{continuous} \end{cases}$$

$$E(aY+b|X=x) = aE(Y|X=x) + b$$

$$E(g(y)|X=x) = \begin{cases} \sum_y g(y)Pr(Y=y|X=x) & \text{discrete} \\ \int_{-\infty}^{\infty} g(y)\ f_y(y|x)\ dy & \text{continuous} \end{cases}$$

$$E(Y|X=x) = E(Y) \text{ if } X \text{ indep } Y$$

$$E(Y) = E(E(Y|X))$$

$$Y = g(x) \implies E(Y|X=x) = g(x)$$

$$V(Y) = E(Y^2) - (E(Y))^2 = E[(Y-E(Y))^2]$$

$$V(Y|X=x) = E((Y-\mu_x)^2|X=x)$$

$$= \begin{cases} \sum_y (y-\mu_x)^2 Pr(Y=y|X=x) & \text{discrete} \\ \int_{-\infty}^{\infty} (y-\mu_x)^2 f_{Y|X}(y|x)\ dy & \text{cont} \end{cases}$$

$$V(Y) = E(V(Y|X)) + V(E(Y|X))$$

Law of Iterated Expectation

$$E_Y(Y) = E_X[E_{Y|X}(Y|X)]$$

Ex: # of coins to flip according to Poi(λ), coins have prob p heads, $1-p$ tails. Expected number of heads, $T = \sum_{i=1}^{n} x_i$ (indicator). N = # flips $\sim Poi(\lambda)$.

$$E(T) = E_N(E_{T|N}(T|N)) = E_N(NP) = pE_n(N) = p\lambda$$

$E(N\mu_x) = \mu_x\mu_N$ if x_i i.i.d and $\mu(N)$ indep of N

Law of Iterated Variance

$$V_Y(Y) = V_X(E_{Y|X}[Y|X]) + E_X(V_{Y|X}(Y|X))$$

Markov Chains

Stochastic process $\{X_t : t \in T\}$ where $Pr(x_t|x_{t-1},\ldots,x_0) = Pr(x_t|x_{t-1})$.

Markov Property $X_n \perp X_0,\ldots,x_{n-2}|x_{n-1}$

Stochastic Matrix Matrix with min vals ≥ 0, max vals ≤ 1, rows sum to 1. **State space** is set of vals for x_t.

Transition prob matrix: $\mathbf{P} = \begin{array}{cc} & \begin{array}{cc} 0 & 1]x_{t+1} \end{array} \\ \underbrace{\begin{array}{c} 0 \\ 1 \end{array}}_{x_t} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}$ $\mathbf{P}_{ij}$ = prob of going from $i \to j$ in one step. $\mathbf{P}^n_{ij}$ = prob of going from $i \to j$ in n steps.

Limiting Distribution $\{X_t\}$ has a limiting distribution if $\lim_{n\to\infty}(P^n)_{ij} = \lambda_j$ for all i and j (not guaranteed).

$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ doesn't have a limiting distribution.

If $\vec{\lambda}$ is limiting distrib of $\mathbf{P}$, then $\vec{\lambda}\mathbf{P} = \vec{\lambda}$ (not a bi-implication). Still have $\vec{\pi}\mathbf{P} = \vec{\pi}$ for stationary distribution.

Positive TPM

TPM whose entries are $> 0, \mathbf{P} > 0$. If not positive but $\exists n$ s.t. $\mathbf{P}^n > 0$ then $\mathbf{P}$ is a regular matrix. If $\mathbf{P}$ is regular, then $\exists$ **unique** $\vec{\pi}$ that is a stationary distrib for $\mathbf{P}$ and will also be a limiting distribution. $\mathbf{P}$ will not be regular if $\mathbf{P}^n$ and $\mathbf{P}^{n+1}$ have entries that are zero in both matrices.

How to find Stationary Dis

General: $\vec{\pi}\mathbf{P} = \vec{\pi}$, do the mult:

$$(\pi_1, \pi_2)\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = (\pi_1, \pi_2)$$
$$\implies \pi_1 P_{11} + \pi_2 P_{21} = \pi_1$$
$$\pi_1 P_{12} + \pi_2 P_{22} = \pi_2$$

Also have $\pi_1 + \pi_2 = 1$. Solve linear system with 1 redundant eq. If **stationary unique**, then can use:

$$\vec{x} = (1, x_2, \ldots, x_k)$$
$$\vec{x}\mathbf{P} = \vec{x} \text{ (solve)}$$
$$\vec{\pi} = \left(\frac{1}{1 + x_2 + \ldots + x_k}\right)\vec{x}$$

Let W be $k \times k$ matrix. If $W\vec{v} = \lambda\vec{v}$ then $\vec{v}$ is a right eigenvector of W with eigenvalue λ. If we can construct matrix of eigenvectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_u}(V)$ where λ_i is the eigenvalue associated with $\vec{v_i}$ then $W = V\Lambda V \iff WV = V\Lambda$ (eigenvalue decomposition), where Λ is a matrix of 0s with $\lambda_1, \lambda_2, \ldots, \lambda_u$ on the diagonal. $\vec{\pi}\mathbf{P} = \vec{\pi} = 1 \cdot \vec{\pi} \implies \mathbf{P}^T\vec{\pi}^T = 1 \cdot \vec{\pi}^T$. Here $\vec{\pi}^T$ is an eigenvector of $\mathbf{P}^T$ corresponding to eigenvalue of 1.
If we can get from i to j, then j is accessible from i. For time-homogeneous $\mathbf{P}$, if $\exists n \geq 0$ s.t. $\mathbf{P}^n_{ij} > 0$, then j accessible from i.
If j accessible from i and i accessible from j, then i and j communicate. Communication is symmetric, reflexive ($\mathbf{P}^0_{ii} = 1$), transitive.
If all states of chain communicate with each other, then the chain is irreducible.
First hitting time: $T_j = \min\{n > 0 : X_n = j \; if \; x_0 = j\}$
$f_j = Pr(T_j < \infty \mid X_0 = j) = 1 \iff$ state j is a recurrent state. If recurrent, expected number of returns is infinity (will become infinite sum of 1)
$f_j = Pr(T_j < \infty \mid X_0 = j) < 1 \iff$ state j is a transient state. If transient, expected returns is geometric $(1 - f_j)$.
Communication class: set of states who **all communicate with each other** and no one else. **State by itself looping** is a communication class. All states in a communication class are either **all recurrent or all transient**
Closed comm classes: C is closed $\iff \mathbf{P}_{ij} = 0 \; \forall i \in C, j \notin C$. States is just a union of classes of transient and classes of recurrent states.
Canonical decomposition of markov chain:

$$\begin{array}{c} \\ T \\ R_1 \\ R_2 \end{array}\begin{array}{c} \begin{array}{ccc} T & R_1 & R_2 \end{array} \\ \begin{pmatrix} 0 & * & * \\ 0 & [P_1] & 0 \\ 0 & 0 & [P_2] \end{pmatrix} \end{array}$$

Where T is the class of all transient states (may or may not communicate) and R_i are recurrent communication classes. P_i are irreducible mini tpm, mini markov chain on reduced state space.
How to **find expected return time** $\mu_j = E[T_j | x_0 = j]$ (for finite irreducible markov chains, positive stationary distrib is unique)

1. Use stationary distrib. $\{x_0, \ldots, x_n\}$ finite (# of states) irreducible chain. Then $\mu_j < \infty$ and $\exists\vec{\pi}$ s.t. $\pi_j = \frac{1}{\mu_j} \; \forall j$. We have $\pi_j = \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{P}^m_{ij}$ (limit of avg). This is NOT the same as $\lim_{n\to\infty} \mathbf{P}_{ij} = \lambda_j$, as this will not converge if there is no limiting distribution.
2. First step analysis. Find $e_x = E(T_a \mid x_0 = x)$ by finding e_k for all relevant states. Ex. $\mathbf{P} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$. Then $e_a = \frac{1}{2}(1) + 0 \cdot (1 + e_b) + \frac{1}{2}(1 + e_c), e_b = 1(1 + e_c), e_c = \frac{1}{4} \cdot 1 + \frac{1}{4}(1 + e_b) + \frac{1}{2}(1 + e_c)$. Get $e_a = \frac{7}{2} = \mu_a \implies \pi_a = \frac{2}{7}$

Positive recurrent, recurrent j s.t. $E(T_j | x_0 = j) < \infty$
Null recurrent, recurrent j s.t. $E(T_j | x_0 = j) = \infty$
Periodicity, **period of state** i, $d(i)$ is the gcd of the set of possible return times to i. $d(i) = gcd\{n > 0 : \mathbf{P}^n_{ii} > 0\}$. If $d(i) = 1$, then i is aperiodic. If there is no return to i, $d(i) = \infty$. All states in a communication class have the same period. Markov chain is periodic if it is irreducible and all states have period > 1. Otherwise, if irreducible and all states have period 1, then the chain is aperiodic. Chain is ergodic if irreducible, aperiodic and all states have finite return times. If chain is ergodic, then $\exists$ **unique** positive stationary distribution for the chain ($\pi_j = \lim_{n\to\infty}(\mathbf{P}^n)_{ij} \; \forall i, j$). Chain ergodic $\iff$ tpm regular.
A chain is time-reversible if $\pi_i \mathbf{P}_{ij} = \pi_j \mathbf{P}_{ji} \; \forall i, j$ (can't tell if I'm going forward or backwards) for stationary $\vec{\pi}$. These equations are called the detailed balance equations.
$\pi_i \mathbf{P}_{ij} = Pr(x_0 = i)Pr(x_1 = j \mid x_0 = i) = Pr(x_0 = i, x_1 = j) \overset{\text{by time revers}}{=} Pr(x_0 = k, x_i = i) = Pr(x_0 = j)Pr(x_1 = i \mid x_0 = j) = \pi_j \mathbf{P}_{ji}$. Additionally $Pr(x_0 = i_0, x_1 = i_1, \ldots, x_n = i_n) = Pr(x_0 = i_n, x_1 = i_{n-1}, \ldots, x_n = i_0)$
State i is an absorbing state if $\mathbf{P}_{ii} = 1$. Markov chain is an absorbing chain if there is ≥ 1 absorbing state. For an absorbing chain with t transient states and k singleton absorbing states we have the canonical decomp:
$\mathbf{P} = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$, where Q is $t \times t$ matrix for transient states, 0 is $k \times t$ 0 matrix (can't go back to transient), R is $t \times k$ and I is $k \times k$ for absorbing.
$\mathbf{P}^n = \begin{pmatrix} Q^n & (Q^{n-1} + Q^{n-2} + \ldots + Q + I)R \\ 0 & I \end{pmatrix}$
Lemma: square matrix A s.t. $\lim_{n\to\infty} A^n = 0$ then $\sum_{n=0}^{\infty} A^n = (I - A)^{-1}$. For absorbing above, $\lim_{n\to\infty} Q^n = 0 \implies \lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} 0 & (I - Q)^{-1}R \\ 0 & I \end{pmatrix}$, where $F = (I - Q)^{-1}$ is the fundamental matrix of the absorbing chain (not a tpm). F_{ij} contains expected # of visits to j starting in i, where i and j are transient. **Expected time to absorption** from $i = a_i = \sum_{j\in T} F_{ij}$ or $\vec{a} = F\vec{1}$, where T is the set of transient states. i.e. expected time to absorption starting from state 1 is sum of row 1. Absorption time from $i = (F1)_i$. Absorption probability: prob that from transient i, chain is absorbed in j is $(FR)_{ij}$.