ST3236 Stochastic Processes 1 Cheat Sheet

Law of Total Probability:
Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that $\cup_{i=1}^n F_i = S$(Exactly 1 of F_i will occur). Then, $E = \cup_{i=1}^n E \cap F_i$ &

$$P(E) = P(\cup_{i=1}^n E \cap F_i) = \sum_{i=1}^n P(EF_i) = cup_{i=1}^n E \cap F_i) = \sum_{i=1}^n P(E|F_i)P(F_i)$$

Boole's Inequality: $P(\cup_{i=1}^n A_i) \le \sum_{i=1}^n P(A_i)$

Tail-Sum Formula(Expectation):

$$\mathbb{E}[X] = \sum_{k=1}^\infty kP(X=k) = \sum_{k=1}^\infty \sum_{i=1}^k P(X=k) = \sum_{i=1}^\infty \sum_{k=1}^\infty P(X=k) = \sum_{i=1}^\infty P(X \ge i)$$

Conditional Expectation:

$$\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}[X|Y]] = \sum_y \mathbb{E}[X|Y=y]P(Y=y)$$

$$\mathbb{E}[\sum_{i=1}^N X_1] = \mathbb{E}[\mathbb{E}[\sum_{i=1}^N X_1|N]] = N\mathbb{E}[X]$$ **(Expectation of Random Sum)**

$$\mathbb{E}[XY] = \mathbb{E}[X\mathbb{E}[Y|X]] \qquad Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Chapman-Kolmogorov Equation: Probability of getting from state i to state j, is equivalent to probability of going from state i to state k in n steps, and from state k to state j in m steps

$$p_{i,j}^{n+m} = \sum_{k=0}^\infty p_{i,k}^n p_{k,j}^m \ge p_{i,k}^n p_{k,j}^m \leftrightarrow P^{n+m} = P^n P^m$$

First Step Analysis:
We define the following:

- $u_{ik} = \Pr\{X_T = k|X_0 = i\}$ as the *probability of* ***absorption*** into state k, starting from state k
- $v_i = \mathbb{E}[T|X_0 = i]$ as the *expected duration* of the Markov Chain, starting from state i

Given a Markov Chain that contains first r non-absorbing states and next $n-r$ absorbing states, we have that:

$$u_{ik} = \sum_{j=1}^r \Pr\{X_T = k|X_0 = i, X_1 = j\} = p_{i,k} + \sum_{j=1}^r p_{i,j}u_j$$

$$v_i = \sum_{j=1}^r p_{i,j}(1+v_j) + \sum_{j=r+1}^n p_{i,j}(1) = 1 + \sum_{j=1}^r p_{i,j}v_j$$

Given the transition matrix P for the Markov Chain, we formulate the following:

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

- Q represents transition probabilities from non-absorbing state i to non-absorbing state j
- R represents transition probabilities from non-absorbing state i to absorbing state k

$$U = [u_{ik}] = R + QU \to U = [I-Q]^{-1}R$$
$$V = [v_i] = \vec{1} + QV \to V = [I-Q]^{-1}\vec{1}$$

Are matrices representing absorption probabilities of state i into state k, and the expected duration of the MC starting from state i

Classification of States:
State j is said to be accessible from state i if $P_{i,j}^n > 0$ for $n \ge 0$

- Starting from state i, it is possible that the MC enters state j

States i, j are said to communicate if state i is accessible from j, and state j is accessible from state i, denoted as $i \leftrightarrow j$. Then, the following properties hold:

- (Reflexivity) $i \leftrightarrow i$: State i communicates with itself $\to p_{i,i}^n > 0$
- (Symmetry) If $i \leftrightarrow j$, then $j \leftrightarrow i$
- (Transitivity) if for states i, j, k, $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$

The ***equivalence class*** $[i] = \{j \in S : j \leftrightarrow i\}$ is the collection of states that communicate with i

- The equivalence relation partitions the Markov Chain into equivalence classes → Each equivalence class also a MC
- A Markov Chain is said to be ***irreducible*** if it consists of only one equivalence class → all states communicate

Denote $f_{i,i}^n = \Pr\{X_n = 1, X_v \neq i : i = 1,2,...,n-1|X_0 = 1\}$ as the probability that the *first return time* is n if the MC starts at i. Then, the probability that starting at i, the state ever returns is

$$f_i = \sum_{n=1}^\infty f_{i,i}^n = \lim_{N\to\infty} \sum_{n=1}^N f_{i,i}^n$$

- A state is said to be ***recurrent*** if the probability that it ever returns to state $i \to f_i = 1$
  Alternatively, the *expected number of periods that the Markov Chain stays in state i is infinite:*

  $$\mathbb{E}[\sum_{n=1}^\infty I\{X_n = i|X_0 = i\}] = \sum_{n=1}^n P_{i,i}^n = \infty$$

  - Let C be a communication class of recurrent states. Then, $p_{i,j}^n = 0, \forall n, i \in C, j \notin C \to$ Once in C, stays inside C
  - A recurrent class is also known as a *closed* class → *once the MC is in a recurrent class, it cannot leave the class*
- Similarly, a state i is ***transient*** if the probability of return to the state is finite $\to f_i < \infty$ and:

  $$\mathbb{E}[\sum_{n=1}^\infty I\{X_n = i|X_0 = i\}] = \sum_{n=1}^n P_{i,i}^n = \infty$$

  - $1 - f_i$ is the positive *probability of escape* that the MC will never return to state i again. That is the probability that the MC will return to state i for exactly n times is $f_i^{n-1}(1-f_i)$ decays as n increases

Corollary: If for states $i \leftrightarrow j$ and state i is recurrent, then state j is also recurrent. Consequently, if state i is transient, so is state j
The ***periodicity*** of a state i is denoted by:

$$d(i) = GCD(\{n : p_{i,i}^n > 0, n \ge 1\})$$

Is the lowest common divisor interval of all periods n such that state i can return to i $\to \forall n, p_{i,i}^n > 0 \to n = \alpha d(i)$

- Periodicity is a *class property* → For states i, j, $i \leftrightarrow j \to d(i) = d(j)$
- If a state i has period $d(i)$, there exists integer N such that $\forall n \ge N, p_{i,i}^{nd(i)} > 0$
- If state i has period $d(i)$, and $p_{j,i}^m > 0$, then $p_{j,i}^{m+\alpha d(i)} > 0$
- A Markov Chain that has period 1 is known as *aperiodic*

Basic Limiting Theorem of Markov Chains:
The ***stationary distribution*** $\vec{\pi} = \{\pi_1, \ldots, \pi_n\}$ of a Markov Chain has the following properties:

$$\pi_j = \sum \pi_i p_{i,j} \to \vec{\pi} = \vec{\pi}P$$
$$\sum_i \pi_i = 1$$

And π_i represents the unconditional distribution of the Markov Chain being in state i at any given point in time:

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^n I[X_m = j] = \lim_{n\to\infty} \frac{N_n(j)}{n} \to \lim_{n\to\infty} \Pr\{X_m = j\}$$

Denote $T_i = \min\{n \ge 1 : X_n = 1\}$ as the ***first return time*** of the MC to state i. Then,

$$f_{i,i}^n = \Pr\{T_i = n|X_0 = i\} \text{ and}$$
$$m_i = \mathbb{E}[T_i|X_0 = i] = \sum_{n\to\infty} nf_{i,i}^n$$

Where m_i denotes the *mean duration* between visits to state i.
Theorem 4.1 Consider an *aperiodic, irreducible and recurrent Markov Chain*. Then,

$$\lim_{n\to\infty} P_{j,i}^n = \lim_{n\to\infty} P_{i,i}^n = \frac{1}{m_i}$$

- In the long run, the probability of being in state i is *independent of the starting state j*
- ***Positive Recurrent:*** $m_i = \mathbb{E}[T_i|X_0 = i] < \infty \to \lim P_{i,i}^n > 0$
- ***Null Recurrent:*** $m_i = \mathbb{E}[T_i|X_0 = i] = \infty \to \lim P_{i,i}^n \to 0$
- ***Transient:*** $\lim_{n\to\infty} P_{i,i}^n = 0$

Theorem 4.2 In a positive recurrent, irreducible and aperiodic Markov Chain communication class,

$$\lim_{n\to\infty} P_{j,i}^n = \lim_{n\to\infty} P_{i,i}^n = \pi_i = \frac{1}{m_i} \text{ and}$$

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \vec{\pi} \\ \vec{\pi} \\ \cdots \\ \vec{\pi} \end{bmatrix} \to \begin{bmatrix} \vec{\pi} \\ \vec{\pi} \\ \cdots \\ \vec{\pi} \end{bmatrix} = \lim_{n\to\infty} P^{n+1} = \lim_{n\to\infty} P^n P = \begin{bmatrix} \vec{\pi} \\ \vec{\pi} \\ \cdots \\ \vec{\pi} \end{bmatrix} P$$

- Hence, we can recover m_i from π_i
- Consequently, if a MC is periodic($d_i > 1$), then *the limiting distribution does not exist for this equivalence class*

Regular Transition Matrix:

A Markov Chain and it's transition matrix is called ***regular*** if there exists a $k \geq 1$ such that P^k have all positive entries.

- All states can communicate in k steps
- If a MC contains *absorbing states, then it cannot be regular → MC in absorbing states cannot reach any other states*

Every transition matrix that satisfies the following is regular:

- For every pair of states i, j, there is a path $k_1, \ldots, k_r$ for which their probabilities are strictly positive
- There is at least one state where $P_{i,i} > 0$

Regular implies irreducibility and aperiodic → Markov Chain is positive recurrent

Doubly Stochastic Matrix:

A transition probability matrix is called *double stochastic* if the column entries sum to 1 in addition to the properties of stochastic matrices:

$$P_{i,j} \geq 0, \ \sum_k P_{i,k} = \sum_k P_{k,j} = 1, \ \forall i, j$$

- If an $N \times N$ double stochastic matrix is regular, then the unique limiting distribution is: $\pi_j = \{\frac{1}{N}, \ldots, \frac{1}{N}\}^T$ → equally likely to be in any state after a long period of time

Let P, Q be double stochastic matrices, $0 \leq \lambda \leq 1$

- $A = \lambda P + (1 - \lambda)Q$ is also a double stochastic matrix
- $B = PQ$ is also a double stochastic matrix

Reducible Markov Chains:

If the Markov Chain is reducible, we cannot directly determine the period, class type and limiting distribution(if any). To find $\lim_{n\to\infty} P^n$, we need to decompose the MC into:

Absorbing States

- Once the MC is in an absorbing state i, the probability $p_{i,j}^{(n)}$ from transiting to another state is 0. As a result, row entries of absorbing states do not change as $n \to \infty$

Recurrent Classes

- If the class is positive recurrent and aperiodic, then the limiting distribution exists.

Transient Classes

- For states i that are in transient classes, we know that $\lim_{n\to\infty} P_{j,i}^n = \lim_{n\to\infty} P_{i,i}^n = 0$. Row entries in $\lim_{n\to\infty} P^n$ corresponding to transient state classes = 0
- For stationary distribution probabilities to recurrent/absorbing states j from transient class states, we compute probability of absorption into class $[j]$, then the probability that once in class $[j]$, MC stays in state j:

$$u_{i,[j]} = \sum_{j' \in [j]} p_{i,j'} + p_{i,i} u_{i,[j]}$$

$$\pi_{i,j} = u_{i,[j]} \times \pi_j$$

- This also means that the stationary distribution probabilities from a state i in one transient class to another j is:

$$\pi_{i,j} = u_{i,[j]} \times \pi_j = u_{i,[j]} \times \lim_{n\to\infty} P_{j,j}^n = 0$$

While limiting distributions do not always exists for all equivalence classes such as periodic recurrent classes, *time average distribution exists*: $\lim_{n\to\infty} \left[\frac{1}{n} \sum_{k=1}^{n} P^k \right]$

Time Reversible Markov Chains:

Let $Q_{i,j}$ denote the transition probabilities such that starting at time n, we trace the sequence of states going back in time $X_n, X_{n-1}, \ldots$. Then, the transition probabilities $Q_{i,j}$ is defined by:

$$Q_{i,j} = P\{X_m = j \,|\, X_{m+1} = i\} = \frac{P\{X_m = j, X_{m+1} = i\}}{P\{X_{m+1} = i\}} = \frac{P\{X_m = j\}P\{X_{m+1} = i \,|\, X_m = j\}}{P\{X_{m+1} = i\}} = \frac{\pi_j P_{j,i}}{\pi_i}$$

A Markov Chain is said to be *time reversible* if $P_{i,j} = Q_{i,j}$. That is, it needs to satisfy the **detailed-balance equation**:

$$\pi_j P_{j,i} = \pi_i Q_{i,j} = \pi_i P_{i,j}$$

- This means that the *rate* of forward chain is equals to the *rate* of the backward chain

Theorem 4.2(Kolmogorov's criteria for time reversibility): A regular/ergodic Markov Chain for which $P_{i,j} = 0$ whenever $P_{j,i} = 0$ is time reversible if and only if starting in state i, any path back to state i has the same probability as the reversed path.

$$P_{i,i_1} P_{i_1,i_2} \ldots P_{i_k,i} = P_{i,i_k} P_{i_k,i_{k-1}} \ldots P_{i_1,i}$$

Proposition 4.6: Consider an irreducible Markov Chain with transition probabilities $P_{i,j}$. If we can find positive numbers $\pi_i, i \geq 0$, and a transition probability matrix $\mathbf{Q} = [Q_{i,j}]$ such that:

$$\pi_i P_{i,j} = \pi_j Q_{j,i} \qquad \sum_i \pi_i = 1$$

Then, $Q_{i,j}$ are the transition probabilities of the reversed chain, and π_i are stationary probabilities for both the forward and backwards chain.

Branching Processes:

Let ξ be a non-negative random variable denote the number of offspring that is produced by a single entity in a branching process. Denote the probability distribution of ξ as:

$$\Pr\{\xi = k\} = p_k \quad k = 0,1,\ldots$$

Probability Generating Functions:

Using $s = e^t$, the *probability generating function* of ξ is given as:

$$\phi(s) = \mathbb{E}[e^{t\xi}] = \mathbb{E}[s^\xi] = \sum_{k=0}^{\infty} p_k s^k \quad 0 \leq s \leq 1$$

- Knowing the probability generating function allows us to derive the probability mass functions:

$$p_k = \frac{1}{k!} \frac{d^k \phi(s)}{ds^k} \bigg|_{s=0}$$

- **Uniqueness of PGF:** if random variables X and Y have the same probability generating function, their probability mass functions are identical: $P(X = k) = P(Y = k)$

PGF of sum of Random Variables:

Let ξ and η be independent non-negative integer valued random variables having the PGFs

$$\phi(s) = \mathbb{E}[s^\xi] \qquad \varphi(s) = \mathbb{E}[s^\eta]$$

Then, the PGF of the sum of random variables $\xi + \eta$ is

$$\mathbb{E}[s^{\xi+\eta}] = \mathbb{E}[s^\xi s^\eta] = \mathbb{E}[s^\xi]\mathbb{E}[s^\eta] = \phi(s)\varphi(s)$$

In general, let $\xi_1, \xi_2, \ldots, \xi_m$ be independent random variables with PGF $\phi(s) = \mathbb{E}[s^{\xi_i}]$. Then, the PGF of the sum $\xi_1 + \xi_2 + \ldots + \xi_m$ is:

$$\mathbb{E}[s^{\xi_1 + \xi_2 + \ldots + \xi_m}] = [\phi(s)]^m$$

Let $X = \xi_1 + \xi_2 + \ldots + \xi_N$ be a random sum of independent variables of ξ_i with PGF $\phi(s)$. Then, the PGF of X is given by:

$$h_X(s) = \mathbb{E}[s^X] = \mathbb{E}_N[\phi(s)^N] = g_N(\phi(s))$$

Theorem: Let ϕ be the PGF of a non-negative integer-valued random variable X. Then $\mathbb{E}[X] = \phi'(1)$. More generally

$$\mathbb{E}[X(X-1)\ldots(X-k+1)] = \phi^{(k)}(1), k \geq 1$$

Mean, Variance and Extinction Probability:

Let X_n denote the number of individuals in the population at time n, and $\xi_i^{(n)}$ with $\mathbb{E}[\xi] = \mu$, $\mathrm{Var}(\xi) = \sigma^2$ denote the number of offspring produced by individual i at time n. Then,

$$X_{n+1} = \xi_1^{(n)} + \xi_2^{(n)} + \ldots + \xi_{X_n}^{(n)}$$

With mean:

$$\mathbb{E}[X_{n+1}] = M(n+1) = M(n)\mu = \mu^{n+1}\mathbb{E}[X_0]$$

And variance:

$$\mathrm{Var}(X_{n+1}) = V(n+1) = \sigma^2 M(n) + \mu^2 V(n)$$

$$V(n) = \sigma^2 \mu^{n-1} \times \begin{cases} n & \text{if } \mu = 1 \\ \frac{1-\mu^n}{1-\mu} & \text{if } \mu \neq 1 \end{cases}$$

- Variance of population increases linearly if $\mu = 1$, and geometrically if $\mu > 1$

Let $T = \{\min n \geq 1 : X_n = 0\}$ denote the time of extinction, when $X_0 = 1$. Then we denote:

$$u_n = \Pr\{T \leq n\} = \Pr\{X_n = 0\}$$

As the probability of the first time of extinction of the branching process at time n. We have that:

$$u_n = \sum_{k=0}^{\infty} p_k (u_{n-1})^k = \mathbb{E}[(u_{n-1})^k] = \phi(u_{n-1})$$

- u_n is monotonically increasing $\to u_{n+1} \geq u_n$
- Probability of extinction *exactly* at the n^{th} generation is $u_n - u_{n-1}$

Let the $u_\infty = \Pr\{T \leq \infty\}$ be the probability that the population ever dies out. Then, we have that u_∞ fulfils:

$$u_\infty = \lim_{n\to\infty} u_n = \lim_{n\to\infty} \phi(u_{n-1}) = \phi(\lim_{n\to\infty} u_{n-1}) = \phi(u_\infty)$$

And u_∞ is the *smallest non-negative root to the equation $u = \phi(u)$*

- If $\mathbb{E}[\xi] = \mu < 1$, then $u_\infty = 1$
- If $\mathbb{E}[\xi] = \mu = 1$, then $u_\infty = \begin{cases} 1 & \text{if } p_0 > 0 \\ 0 & \text{if } p_0 = 0 \end{cases}$
- If $\mathbb{E}[\xi] = \mu > 1$, then $0 < u_\infty < 1$