Probability for Data Science - Formula Sheet 3

tbimbato

Discrete Markov Chains

-
  - Markov prop. :$P(X_{n+1}|X_n, X_{n-1}, \ldots, X_0) = P(X_{n+1}|X_n)$
  - $P(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1}, \ldots, X_0) =$ $P(X_{n+1} = x_{n+1} \mid X_n = x_n)$ (1-step memory)
  - $P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) =$ $P(X_n \mid X_{n-1})P(X_{n-1} \mid X_{n-2}) \cdots P(X_1 \mid X_0)P(X_0)$ (follow the path)
- **n-step P**: $\mathbb{P}^n$

  $\sum row = 1$, always
- **Classification of States:**
  - ⋆ **Irreducible**: (MC composed by one class)
  - ⋆ **Aperiodic**: gcd(cycle lengths) $= 1 within in class$

  Recurrent: Visited infinitely often (when i enter a recurrent class I never leave)

  Transient: Visited finitely often (when i enter a transient class I can leave it)

  - ⋆ **Positive Recurrent / ergodic**: finte mean return time or finite number of states
- **First Step Analysis**: $P_i =$

  Prob. of reaching a state from i, solving $P_i = \sum_j P_{ij} P_j$ with boundary conditions ($P_{goal} = 1, P_{abs.} = 0$).
- **Stationary dist / prop. of time**: $\lim_{n \to +\infty} P_{ij}^n = \pi_j$ for all $i, j \in S$

$$\begin{cases} \pi = \pi \cdot P \\ \sum_{i \in \mathbb{S}} \pi_i = 1 \end{cases}$$

Poisson Process

- $(N(t))_{t \geq 0}$ is a Poisson Process with rate λ if:
  - no. of arrivals: $N(t) \sim Pois(\lambda t)$ Inter-arrival times: $T_i \sim Exp(\lambda)$
- **Properties:**
  - $P(N(t+s) - N(s) = k) = P(N(t) = k)$ $P(T > t + s | T > s) = P(T > t)$
  - $N(t) - N(s) \sim \text{Pois}(\lambda(t-s))$
  - The no. of arrivals in **disjoint** intervals are independent.
  - if $N_1(t) \sim \text{Pois}(\lambda_1 t)$ and $N_2(t) \sim \text{Pois}(\lambda_2 t)$ $\rightarrow$ $N_1(t) + N_2(t) \sim \text{Pois}((\lambda_1 + \lambda_2)t)$
  - Filtered PP: e.g., 'special' calls occur with $P = p \rightarrow \lambda_{\text{special}} = (p \cdot \lambda)$
- **Probability Distributions:**

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k \geq 0$$

- **Interarrival Distribution:**

$$P(T > t) = e^{-\lambda t}, \quad T \sim Exp(\lambda)$$

Continuous Markov Chains

- **Birth-Death Process:** Special case with rates:

  $q_{i,i+1} = \lambda_i, \quad q_{i,i-1} = \mu_i, \quad q_{ii} = -(\lambda_i + \mu_i)$
- **Stationary Distribution** (Birth-Death Process):

$$\pi_0 = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} + \ldots}$$

$$\pi_n = \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \pi_0$$

- **Holding time params.** $v_0, v_1 .. \sim \ min(exp(\lambda_1), exp(\lambda_2)..)$

  Mean time in state i $= \frac{1}{v_i}$
- Jump Chain $= P_{i,j} \rightarrow P(T_a < T_b)$
- Generator matrix :

  $q_{ii} = -v_i$ diagonals

  $q_{ij} = v_i P_{i,j}$ for $j \neq i$