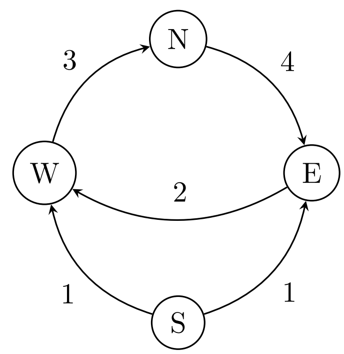
UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 9
Fall 2021

1. **Jump Chain Stationary Distribution**

Use properties of transient states and the jump chain to find the stationary distribution of this CTMC.



Solution: $\pi(S) = 0$ since it's transient. Considering just N, E, and W, we know that the jump chain has stationary distribution $\frac{1}{3}$ in every state. We can use the formula to convert the jump chain stationary distribution to CTMC stationary distribution.

$$\pi_{\text{CTMC}}(x) = \frac{\frac{1}{Q(x)}\pi_{\text{Jump Chain}}(x)}{\sum_y \frac{1}{Q(y)}\pi_{\text{Jump Chain}}(y)}$$

Where $Q(x)$ is the sum of the rates leaving state x. What the formula tells us intuitively is that since we're spending equal time at N, E, and W from a jump standpoint, the true fraction of time we spend in a state should just be proportional $\frac{1}{Q(x)}$. So

$$\begin{aligned}\begin{bmatrix}\pi(N) & \pi(W) & \pi(S) & \pi(E)\end{bmatrix} &= \frac{1}{1/4 + 1/3 + 1/2}\begin{bmatrix}1/4 & 1/3 & 0 & 1/2\end{bmatrix} \\ &= \begin{bmatrix}3/13 & 4/13 & 0 & 6/13\end{bmatrix}\end{aligned}$$

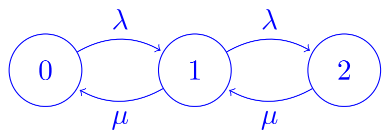
2. **Two-Server System**

A company has two servers (the second server is a backup in case the first server fails, so only one server is ever used at a time). When a server is running, the time until it breaks down is exponentially distributed with rate μ. When a server is broken, it is taken to the repair shop. The repair shop can only fix one server at a time, and its repair time is exponentially distributed with rate λ. Find the long-run probability that no servers are operational.

Solution:

The idea is to model the number of operational servers as a continuous-time Markov chain on the state space $\{0, 1, 2\}$. By thinking about the infinitesimal transition probabilities (which are simply the rates of the exponential distributions), we have the following matrix:

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 \\ \mu & -(\lambda+\mu) & \lambda \\ 0 & \mu & -\mu \end{bmatrix}.$$



Now, we write down the balance equations.

$$\begin{aligned} \lambda\pi(0) &= \mu\pi(1) \\ (\lambda+\mu)\pi(1) &= \lambda\pi(0) + \mu\pi(2) \\ \mu\pi(2) &= \lambda\pi(1) \\ 1 &= \pi(0) + \pi(1) + \pi(2) \end{aligned}$$

We eliminate $\pi(2)$ with $\pi(2) = (\lambda/\mu)\pi(1)$. Plugging this into the second and fourth equations, we have

$$\begin{aligned} \mu\pi(1) &= \lambda\pi(0), \\ 1 &= \pi(0) + \Big(1 + \frac{\lambda}{\mu}\Big)\pi(1). \end{aligned}$$

We next eliminate $\pi(1)$ with $\pi(1) = (\lambda/\mu)\pi(0)$. Plugging this into the second equation above, we have

$$\pi(0) = \frac{1}{1 + \lambda/\mu + (\lambda/\mu)^2}.$$

This is the long-run probability that we will be in state 0, i.e. there are no operational servers.

3. Gaussians and the MSE

Suppose you draw n i.i.d. data points $(x_1, y_1), \ldots, (x_n, y_n)$, where n is a positive integer and the true relationship is $Y = WX + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. (That is, Y has a linear dependence on X, with additive Gaussian noise.) Show that finding the MLE estimate of W given the data points $\{(x_i, y_i) : i = 1, \ldots, n\}$ is equivalent to minimizing the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2$$

Solution:

The likelihood for the data is

$$\mathcal{L}\big((x_1, y_1), \ldots, (x_n, y_n) \mid W = w\big) = \prod_{i=1}^{n} \mathcal{L}\big((x_i, y_i) \mid W = w\big)$$

(the data points are conditionally independent given W)

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-(y_i - wx_i)^2/(2\sigma^2)}$$

(here we say that the likelihood of (x_i, y_i) given W is the density of ε_i, which is $\mathcal{N}(0, \sigma^2)$, evaluated at $y_i - wx_i$)

$$\propto \prod_{i=1}^{n} \mathrm{e}^{-(y_i - wx_i)^2/(2\sigma^2)}$$

(again, we throw out constant factors that do not depend on the data points or w).

We wish to maximize this expression w.r.t. w, but we will find it more convenient to take the log-likelihood instead.

$$\ell\big((x_1, y_1), \ldots, (x_n, y_n) \bigm| w\big) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - wx_i)^2$$

Since we want to *maximize* the log-likelihood, this is equivalent to *minimizing* the cost function

$$J(w) = \sum_{i=1}^{n} (y_i - wx_i)^2$$