Continuous Time Markov Chains (CTMCs)

For most systems of interest, events may occur at any point in time. This leads us to consider continuous time Markov chains. A *continuous time Markov chain* (*CTMC*) has the following property:

$$P\left[X(t+\tau)=j \middle| X(t)=i, X(t-t_1)=k_1, X(t-t_2)=k_2, \ldots, X(t-t_n)=k_n\right]$$
$$= P\left[X(t+\tau)=j \middle| X(t)=i\right],$$
$$= P_{ij}(\tau)$$

for all $\tau > 0, 0 < t_1 < t_2 < \ldots < t_n$

A CTMC is completely described by the initial probability distribution $\pi(0)$ and the transition probability matrix $P(t) = [p_{ij}(t)]$. Then we can compute $\pi(t) = \pi(0)P(t)$.

The problem is that $p_{ij}(t)$ is generally very difficult to compute.

CTMC Properties

This definition of a CTMC is not very useful until we understand some of the properties.

First, notice that $p_{ij}(\tau)$ is independent of how long the CTMC has previously been in state i, that is,

$$
\begin{aligned}
&P\left[X(t+\tau)=j \middle| X(u)=i \text{ for } u\in[0,t]\right] \\
&\quad = P\left[X(t+\tau)=j \middle| X(t)=i\right] \\
&\quad = p_{ij}(\tau)
\end{aligned}
$$

There is only one random variable that has this property: the exponential random variable. This indicates that CTMCs have something to do with exponential random variables. First, we examine the exponential r.v. in some detail.

Exponential Random Variables

Recall the property of the exponential random variable. An exponential random variable X with parameter λ has the CDF

$$P[X \leq t] = F_x(t) = \begin{cases} 0 & t \leq 0 \\ 1 - e^{-\lambda t} & t > 0 \end{cases}.$$

The distribution function is given by $f_x(t) = \dfrac{d}{dt} F_x(t)$;

$$f_x(t) = \begin{cases} 0 & t \leq 0 \\ \lambda e^{-\lambda t} & t > 0 \end{cases}$$

The exponential random variable is the only random variable that is "memoryless."

To see this, let X be an exponential random variable representing the time that an event occurs (e.g., a fault arrival).

We will show that $P\left[X > t + s \middle| X > s\right] = P[X > t]$.

Memoryless Property

Proof of the memoryless property:

$$P[X > t + s | X > s] = \frac{P[X > t + s, X > s]}{P[X > s]}$$

$$= \frac{P[X > t + s]}{P[X > s]}$$

$$= \frac{1 - F_X(t + s)}{1 - F_X(s)}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}$$

$$= \frac{e^{-\lambda t} e^{-\lambda s}}{e^{-\lambda s}}$$

$$= e^{-\lambda t}$$

$$= P[X > t]$$

Event Rate

The fact that the exponential random variable has the memoryless property indicates that the "rate" at which events occur is constant, i.e., it does not change over time.

Often, the event associated with a random variable X is a failure, so the "event rate" is often called the *failure rate* or the *hazard rate*.

The *event rate* of random variable X is defined as the "time-averaged probability" that the event associated with X occurs within the small interval $[t, t + \Delta t]$, given that the event has not occurred by time t, per the interval size Δt:

$$\frac{P\left[t < X \le t + \Delta t \mid X > t\right]}{\Delta t}.$$

This can be thought of as looking at X at time t, observing that the event has not occurred, and measuring the expected number of events (probability of the event) that occur per unit of time at time t.

Observe that:

$$\frac{P\left[t < X \le t + \Delta t \,\middle|\, X > t\right]}{\Delta t} = \frac{P\left[t < X \le t + \Delta t, X > t\right]}{P\left[X > t\right] \cdot \Delta t}$$

$$= \frac{P\left[t < X \le t + \Delta t\right]}{P\left[X > t\right] \cdot \Delta t}$$

$$= \frac{F_X(t + \Delta t) - F_X(t)}{(1 - F_X(t))\Delta t}$$

$$= \frac{F_X(t + \Delta t) - F_X(t)}{\Delta t} \cdot \frac{1}{1 - F_X(t)}$$

$$= \frac{f_X(t)}{1 - F_X(t)} \quad \text{in general.}$$

In the exponential case,

$$\frac{f_X(t)}{1 - F_X(t)} = \frac{\lambda e^{-\lambda t}}{1 - (1 - e^{-\lambda t})} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

This is why we often say a random variable X is "exponential with rate λ."

Minimum of Two Independent Exponentials

Another interesting property of exponential random variables is that the minimum of two independent exponential random variables is also an exponential random variable.

Let A and B be independent exponential random variables with rates α and β respectively. Let us define $X = \min\{A,B\}$. What is $F_X(t)$?

$$
\begin{aligned}
F_X(t) &= P[X \le t] \\
&= P[\min\{A,B\} \le t] \\
&= P[A \le t \text{ OR } B \le t] \\
&= 1 - P[A > t \text{ AND } B > t] \qquad \text{- see lecture 2} \\
&= 1 - P[A > t]\, P[B > t] \\
&= 1 - (1 - P[A \le t])(1 - P[B \le t]) \\
&= 1 - (1 - F_A(t))(1 - F_B(t)) \\
&= 1 - (1 - [1 - e^{-\alpha t}])(1 - [1 - e^{-\beta t}]) \\
&= 1 - e^{-\alpha t} e^{-\beta \mathrm{t}} \\
&= 1 - e^{-(\alpha + \beta)t}
\end{aligned}
$$

Thus, X is exponential with rate $\alpha + \beta$.

Competition of Two Independent Exponentials

If A and B are independent and exponential with rate α and β respectively, and A and B are competing, then we know that the time until one of them wins is exponentially distributed time (with rate $\alpha + \beta$). But what is the probability that A wins?

$$P[A < B] = \int_0^\infty P[A < B \mid A = x] P[A = x] dx$$

$$= \int_0^\infty P[A < B \mid A = x] f_A(x) dx$$

$$= \int_0^\infty P[A < B \mid A = x] \alpha e^{-\alpha x} dx$$

$$= \int_0^\infty P[x < B] \alpha e^{-\alpha x} dx$$

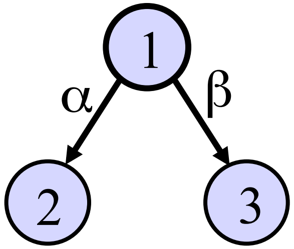
$$= \int_0^\infty (1 - P[B \le x]) \alpha e^{-\alpha x} dx$$

$$= \int_0^\infty \left(1 - \left[1 - e^{-\beta x}\right]\right) \alpha e^{-\alpha x} dx$$

$$= \int_0^\infty e^{-\beta x} \alpha e^{-\alpha x} dx$$

$$= \alpha \int_0^\infty e^{-(\alpha+\beta)x} dx = \frac{\alpha}{\alpha + \beta}$$

Competing Exponentials in CTMCs



$X(0) = 1$
$P[X(0) = 1] = 1$

Imagine a random process X with state space $S = \{1,2,3\}$. $X(0) = 1$. X goes to state 2 (takes on a value of 2) with an exponentially distributed time with parameter α. Independently, X goes to state 3 with an exponentially distributed time with parameter β. These state transitions are like competing random variables.

We say that from state 1, X goes to state 2 with rate α and to state 3 with rate β.

X remains in state 1 for an exponentially distributed time with rate $\alpha + \beta$. This is called the *holding time* in state 1. Thus, the expected holding time in state 1 is $\frac{1}{\alpha+\beta}$.

The probability that X goes to state 2 is $\frac{\alpha}{\alpha+\beta}$. The probability X goes to state 3 is $\frac{\beta}{\alpha+\beta}$.

This is a simple continuous-time Markov chain.

Competing Exponentials vs. a Single Exponential With Choice

Consider the following two scenarios:

1. Event *A* will occur after an exponentially distributed time with rate α. Event *B* will occur after an independent exponential time with rate β. One of these events occurs first.
2. After waiting an exponential time with rate $\alpha + \beta$, an event occurs. *A* occurs with probability $\frac{\alpha}{\alpha+\beta}$, and event *B* occurs with probability $\frac{\beta}{\alpha+\beta}$.

These two scenarios are indistinguishable. In fact, we frequently interchange the two scenarios rather freely when analyzing a system modeled as a CTMC.

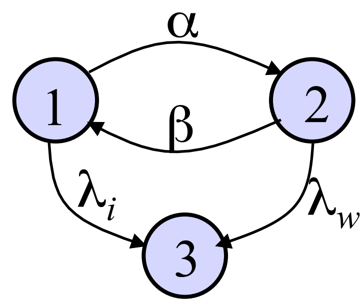
State-Transition-Rate Matrix

A CTMC can be completely described by an initial distribution $\pi(0)$ and a state-transition-rate matrix. A *state-transition-rate* matrix $Q = [q_{ij}]$ is defined as follows:

$$
q_{ij} = \begin{cases} \text{rate of going from state } i \text{ to state } j & i \neq j, \\ -\sum_{k \neq i} q_{ik} & i = j. \end{cases}
$$

Example: A computer is idle, working, or failed. When the computer is idle, jobs arrive with rate α, and they are completed with rate β. When the computer is working, it fails with rate λ_w, and with rate λ_i when it is idle.

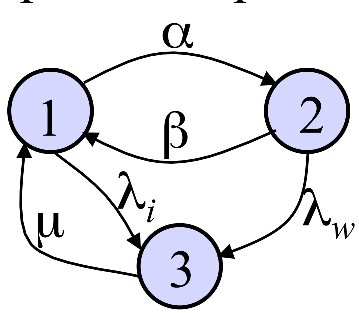
"Simple Computer" CTMC



Let $X = 1$ represent "the system is idle," $X = 2$ "the system is working," and $X = 3$ a failure.

$$Q = \begin{bmatrix} -(\alpha + \lambda_i) & \alpha & \lambda_i \\ \beta & -(\beta + \lambda_w) & \lambda_w \\ 0 & 0 & 0 \end{bmatrix}$$

If the computer is repaired with rate μ, the new CTMC looks like



$$Q = \begin{bmatrix} -(\alpha + \lambda_i) & \alpha & \lambda_i \\ \beta & -(\beta + \lambda_w) & \lambda_w \\ \mu & 0 & -\mu \end{bmatrix}$$

Analysis of “Simple Computer” Model

Some questions that this model can be used to answer:

- What is the availability at time *t*?
- What is the steady-state availability?
- What is the expected time to failure?
- What is the expected number of jobs lost due to failure in [0,*t*]?
- What is the expected number of jobs served before failure?
- What is the throughput of the system (jobs per unit time), taking into account failures and repairs?