Homework 7 - Continuous-time Markov chains

1) Machine repair times. The time T required to repair a machine is an exponentially-distributed random variable (RV) with mean $1/2$ (hours). What is the probability that a repair time exceeds $1/2$ hour? What is the probability that a repair takes at least 12 and $1/2$ hours given that its duration exceeds 12 hours?

2) The post office. Consider a post office that is run by two clerks. Suppose that when Mr. Smith enters the system he discovers that Mr. Jones is being served by one of the clerks, and Mr. Brown by the other. Mr. Smith is told that his service will begin as soon as either Jones or Brown leaves. Suppose that the time that clerk $i = 1, 2$ spends with a customer is exponentially distributed with rate λ_i. Show that

$$\mathrm{P}\left[\text{Smith is not the last to leave the office}\right] = \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^2 + \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^2.$$

3) Conditional expectation of exponential random variables. Suppose T is an exponentially-distributed RV with parameter λ. Use the identity

$$\mathbb{E}\left[T\right] = \mathbb{E}\left[T \,\middle|\, T < t\right] \mathrm{P}\left[T < t\right] + \mathbb{E}\left[T \,\middle|\, T > t\right] \mathrm{P}\left[T > t\right]$$

to calculate $\mathbb{E}\left[T \,\middle|\, T < t\right]$, $t > 0$.

4) Sum of i.i.d. exponential random variables. Suppose that $T_1, \ldots, T_n$ are i.i.d. exponential RVs with parameter λ. In this problem, we will show that the sum $T = \sum_{i=1}^{n} T_i$ has the gamma distribution with parameters n and λ; i.e. T has pdf

$$f_T(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t \geq 0.$$

To that end, we will argue by mathematical induction.

A) Base case: Show that the claim holds true for $n = 1$.

B) Inductive step: Supposing the claim is true for $n-1$, i.e., $T^{(n-1)} := T_1 + \ldots + T_{n-1}$ has the gamma distribution with parameters $n-1$ and λ; show it also holds for n. (Hint: Notice that $T = T^{(n-1)} + T_n$, and recall that for independent, non-negative, continuous RVs Y and Z, the pdf $f_X(x)$ of $X = Y + Z$ is given by the convolution

$$f_X(x) = \int_0^x f_Y(y) f_Z(x-y) dy$$

of the pdfs $f_Y(x)$ and $f_Z(x)$ of Y and Z.)

5) Sum of a random number of exponential random variables. Suppose $T_{\mathbb{N}} = T_1, T_2, \ldots$ is an i.i.d. sequence of exponential RVs, each with parameter λ. Suppose N is a geometric RV with parameter p. Assume that N and $T_{\mathbb{N}}$ are independent, and let

$$X = \sum_{n=1}^{N} T_n.$$

Show that X has the exponential distribution, and determine its parameter.

6) Poisson process. Let $N(t)$ be a Poisson process with rate λ, and let S_i denote the i-th arrival time. Determine $\mathbb{E}\left[S_4\right]$, $\mathbb{E}\left[S_4 \,\middle|\, N(1) = 2\right]$, and $\mathbb{E}\left[N(4) - N(2) \,\middle|\, N(1) = 3\right]$.

7) Arrivals to a subway station. Consider the operation of a subway station during a time interval T. Our goal is to study the arrival of public transportation users to the station. For that purpose divide the time interval in n subintervals of duration h. Notice that the i-th subinterval, for $i = 1, \ldots, n$ consists of times $((i-1)h, ih]$ and that it must be $T = nh$. The total number of potential customers is very large but only a small random fraction arrive at the station during time interval T. The probability of one customer arriving in a small time interval of duration h is λh, and we assume that potential customers make independent decisions as to when to arrive to the station. For sufficiently small time interval h the probability of having more than one arrival can be ignored leading to the approximation

$$\mathrm{P}\left[N_i(h) = 1\right] = \lambda h, \qquad \mathrm{P}\left[N_i(h) = k\right] = 0, \text{ for all } k > 1 \tag{1}$$

where we have introduced $N_i(h)$ as the number of customers that arrive in the i-th time interval of duration h.

In principle the arrival of a customer in the i-th time interval, is *not* independent of the arrival of a customer in the j-th time interval. However, if the number of potential customers is very large, it is a reasonable assumption to suppose that the arrival of a customer in the i-th time interval is independent of the corresponding arrival in the j-th interval, allowing us to write

$$\mathrm{P}\left[N_i(h) = 1/0, N_j(h) = 1/0\right] = \mathrm{P}\left[N_i(h) = 1/0\right]\mathrm{P}\left[N_j(h) = 1/0\right]. \tag{2}$$

The expressions in (1) and (2), are accepted as the definition of the random process describing the arrival of passengers to the subway station. Implicit in this definition is the assumption that T is sufficiently small to allow the assumption that the probability of a customer arriving is independent of time.

The two quantities of interest in this problem are the number $N(t) = \sum_{i=1}^{t/h} N_i(h)$ of customers arriving by time t, with $N(t) = 0$; and the time S_1 elapsed until the first customer arrival at the train station. According to (1), $N(t)$ must satisfy

$$\mathrm{P}\left[N(t+h) - N(t) = 1 \,\middle|\, N(t)\right] = \lambda h. \tag{3}$$

To obtain time S_1 from $N(t)$ we must notice that S_1 is the time at which $N(t)$ transitions from 0 to 1. Thus,

$$S_1 = \min_t (N(t) \geq 1). \tag{4}$$

We will see in this exercise that for sufficiently small h, the first arrival time S_1 is exponentially distributed with parameter λ, and that the number of arrivals $N(t)$ for *any* time t is Poisson with parameter λt.

A) Write a Matlab function to simulate an arrival process. Use $T = 10$ minutes, $\lambda = 1$ customer per minute, and $n = 10^3$. Compare the histogram of $N(T)$ obtained from 10^4 experiments with the Poisson pmf with parameter λT. Compare also the histogram of $N(T)/2$ with the Poisson pmf with parameter $\lambda T/2$.

B) The comparisons in Part *A* should have yielded accurate fits. Use the Poisson approximation of the binomial distribution to justify why this is true. Argue that this implies that the pmf of $N(t)$ is Poisson with parameter λt for all t, i.e.,

$$\mathrm{P}\left[N(t) = k\right] = e^{-\lambda t}\frac{(\lambda t)^k}{k!}. \tag{5}$$

C) Compute a cumulative histogram of the first arrival time S_1 to estimate the cdf of S_1. Compare with the cdf of an exponential RV with parameter λ. A good fit should be observed.

D) Use the fact that the probability of having no arrivals by time t is approximately given by $e^{-\lambda t}$ [cf. (5)], to explain the good fit observed in Part *C*. Notice that we have $S_1 > t$ if and only if there are no arrivals by time t.