Solutions of the first exam:
Stochastic Processes and Simulation II
May 22nd 2024

Exercise 1: Poisson processes

(i) Give the definition of a continuous-time Markov chain.

Solution: A stochastic process $\{X(t), t \geqslant 0\}$ is a continuous-time Markov chain if for all $s, t \geqslant 0$ and nonnegative integers $i, j, x(u), 0 \leqslant u < s$,

$$\mathbb{P}(X(t+s) = j \,|\, X(s) = i, X(u) = x(u), 0 \leqslant u < s) = \mathbb{P}(X(t+s) = j \,|\, X(s) = i).$$

In other words, the process has the Markovian property that the conditional distribution of the future given the present and the past depends only on the present and is independent of the past.

The Amazon rainforest has mainly two seasons, wet and dry, providing the perfect environment for plants and wildlife to thrive. While heavy rainfalls are dominant during the wet season, during the dry season we can observe an alternation in the weather conditions, from sunny periods (S) to light rainfalls (L) to heavy rainfalls (H). During the dry season, assume that a sunny period lasts for an amount of time that can be described by an exponential random variable with parameter λ_S, after which light rainfalls occur with probability 3/4 and heavy rainfalls with probability 1/4. The length of a period with light rain can be described by an exponential random variable with parameter λ_L, while the length of a period with heavy rain by an exponential random variable with parameter λ_H. After light rainfalls, it becomes sunny again with probability 2/3 while heavy rainfalls begin with probability 1/3. After heavy rainfalls, there is an equal probability of light rainfalls and sun. Assume that these exponential random variables are all independent of each other.

(ii) Draw a Markov chain whose states represent the weather conditions during the dry season. Write the associated transition probability matrix P and transition rate matrix Q. What is the relation between the entries of the two matrices?

Solution: Consider the three states in the order S, L, H. We then have that

$$P = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

and

$$Q = \begin{pmatrix} -\lambda_S & \frac{3}{4}\lambda_S & \frac{1}{4}\lambda_S \\ \frac{2}{3}\lambda_L & -\lambda_L & \frac{1}{3}\lambda_L \\ \frac{1}{2}\lambda_H & \frac{1}{2}\lambda_H & -\lambda_H \end{pmatrix}.$$

For any pair of different states we have $q_{ij} = \lambda_i P_{ij}$ with $i, j = S, L, H$.

(iii) Write down the balance equations and calculate the proportion of time in which it rains.

Solution: By solving the balance equations

$$\begin{aligned} \lambda_S P_S &= \frac{2}{3}\lambda_L P_L + \frac{1}{2}\lambda_H P_H \\ \lambda_L P_L &= \frac{3}{4}\lambda_S P_S + \frac{1}{2}\lambda_H P_H \\ \lambda_H P_H &= \frac{1}{4}\lambda_S P_S + \frac{1}{3}\lambda_L P_L \end{aligned}$$

we obtain $P_L = \frac{21}{20}\frac{\lambda_S}{\lambda_L} P_S$ and $P_H = \frac{3}{5}\frac{\lambda_S}{\lambda_H} P_S$. By using the fact that $P_S + P_L + P_H = 1$, we get that $P_S = \frac{1}{1+\frac{21}{20}\frac{\lambda_S}{\lambda_L}+\frac{3}{5}\frac{\lambda_S}{\lambda_H}}$. The proportion of time in which it rains is then $1 - P_S$.

Exercise 2: Renewal theory

(i) State and prove the inspection paradox.

Solution: Consider a renewal process $\{N(t), t \geqslant 0\}$ with interarrival distribution F. Let $X_{N(t)+1} = S_{N(t)+1} - S_{N(t)}$ denote the length of the renewal interval containing the time point t. The inspection paradox states that

$$\mathbb{P}(X_{N(t)+1} > x) \geqslant 1 - F(x).$$

In other words, the length of the renewal interval containing the point t tends to be larger than an ordinary renewal interval.
To prove that, condition on the time of the last renewal prior to or at t to write

$$\mathbb{P}(X_{N(t)+1} > x) = \mathbb{E}[\mathbb{P}(X_{N(t)+1} > x \,|\, S_{N(t)} = t - s)].$$

If $s > x$, since there are no renewals between $t - s$ and t, then

$$\mathbb{P}(X_{N(t)+1} > x \,|\, S_{N(t)} = t - s) = 1.$$

If $s \leqslant x$, no renewals should occur for an additional time $x - s$, hence

$$\mathbb{P}(X_{N(t)+1} > x \,|\, S_{N(t)} = t - s) = \mathbb{P}(X > x \,|\, X > s) = \frac{\mathbb{P}(X > x)}{\mathbb{P}(X > s)} = \frac{1 - F(x)}{1 - F(s)} \geqslant 1 - F(x).$$

Hence, $\mathbb{P}(X_{N(t)+1} > x \mid S_{N(t)} = t - s) \geqslant 1 - F(x)$ for all s, and we conclude by taking expectations.

Consider one of the cultural centers in the Amazon rainforest, where tourists arrive to visit its museum about indigenous culture and/or to explore the unique nature by a river cruise.

(ii) Assume that, during the opening hours of the cultural center, tourists arrive according to a renewal process, where the average time between two arrivals is 10 minutes and its variance is 5 minutes. Calculate, for a given time, the average time passed from the last arrival and the average time until the next arrival. Explain how these results are related to the inspection paradox.

<u>Solution</u>: We need to calculate the average age $\mathbb{E}[A(t)]$ and the average excess $\mathbb{E}[Y(t)]$ of a renewal process. We have seen through renewal reward theory that

$$\mathbb{E}[A(t)] = \mathbb{E}[Y(t)] = \frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]},$$

where X is distributed as the interarrival times. Hence $\mathbb{E}[X] = 10$ and $\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2 = 5 + 100 = 105$, so we obtain

$$\frac{\mathbb{E}[X^2]}{2\mathbb{E}[X]} = \frac{105}{20} = 5.25.$$

The results are paradoxical in the sense that the average length of the given interval $\mathbb{E}[X(t)] = \mathbb{E}[A(t)] + \mathbb{E}[Y(t)] = 10.5$ is larger than the average length of a renewal interval.

(iii) Consider now the boats that are used for the river cruise and assume that each boat has a lifetime described by an exponential random variable with parameter λ, independently of the other boats. The cultural center has a policy for which they replace a boat as soon as it breaks or reaches the age of K years. The cost of a new boat is C_1 and an additional cost of C_2 has to be paid when a boat breaks. Assume that the cultural center owns 2 boats. What is the long-run average cost?

<u>Solution</u>: Using reward renewal theory, we can define a cycle to end whenever a boat is replaced. The long-run average cost is then given by the expected cost in a cycle divided by the expected length of a cycle.
Let A be the event that both boats reach age K years without breaking. If this occurs then the cost of a cycle is C_1. We have that $\mathbb{P}(A) = e^{-2\lambda K}$, since we are asking for the minimum of two exponential random variables with the same parameter λ to be larger than K. Denote by A^C the complementary event that one of the boats breaks before time K. If this occurs then the cost of a cycle is $C_1 + C_2$. We have that $\mathbb{P}(A^C) = 1 - e^{-2\lambda K}$. The expected cost in

a cycle is then

$$\begin{aligned}
\mathbb{E}[\text{cost in a cycle}] &= C_1\mathbb{P}(A) + (C_1 + C_2)P(A^C) \\
&= C_1 e^{-2\lambda K} + (C_1 + C_2)(1 - e^{-2\lambda K}) \\
&= C_1 + C_2(1 - e^{-2\lambda K}).
\end{aligned}$$

The length of a cycle is x if one of the boats breaks at time $x < K$, otherwise it is K. Note that the density of the minimum of two exponential random variables with parameter λ is $f(x) = 2\lambda e^{-2\lambda x}$. Hence we get

$$\begin{aligned}
\mathbb{E}[\text{length of a cycle}] &= \int_0^K x f(x)\, dx + \int_K^\infty K f(x)\, dx \\
&= 2\lambda \int_0^K x e^{-2\lambda x}{,}dx + K e^{-2\lambda K} \\
&= -K e^{-2\lambda K} - \frac{1}{2\lambda} e^{-2\lambda K} + \frac{1}{2\lambda} + K e^{-2\lambda K} \\
&= \frac{1}{2\lambda}(1 - e^{-2\lambda K}).
\end{aligned}$$

The long-run average cost is then

$$\frac{C_1 + C_2(1 - e^{-2\lambda K}).}{\frac{1}{2\lambda}(1 - e^{-2\lambda K})}.$$

Exercise 3: Queueing theory

The cultural center also has a restaurant where tourists can have a break eating some local food. Assume that the restaurant has a total of 10 sitting indoor spots that can be used. Tourists arrive according to a Poisson process with rate λ and they are allowed to order and consume as long as there are free sitting spots, otherwise they wait in a queue. Assume also that the total time it takes to order and consume the food is exponentially distributed with rate μ, independently of everything else.

(i) What type of queueing model best describes the situation at the restaurant? What condition, if any, must λ and μ satisfy? Write down the balance equations.

Solution: The situation at the restaurant can be described as an $M/M/10$ queueing model, where the arrival times are i.i.d. $\text{Exp}(\lambda)$ and the service times are i.i.d. $\text{Exp}(\mu)$. The condition to impose in order for the queue size not to increase indefinitely is that $\lambda < 10\mu$. The balance equations are

$$\begin{aligned}
\lambda P_0 &= \mu P_1, & \\
(\lambda + n\mu)P_n &= \lambda P_{n-1} + (n+1)\mu P_{n+1}, & n < 10, \\
(\lambda + 10\mu)P_n &= \lambda P_{n-1} + 10\mu P_{n+1}, & n \geqslant 10.
\end{aligned}$$

Next, assume that next to the restaurant there is a very small bar that serves tropical drinks to go. During the wet season, tourists arrive according to a Poisson process with rate λ, but they are allowed to enter the bar only as long as there are no more than 5 tourists already inside and if they find the bar full then they leave. Assume also that the time it takes to be served in the bar is exponentially distributed with rate μ, independently of everything else.

(ii) What type of queueing model best describes the situation at the bar? What conditions, if any, must now λ and μ satisfy? Write down the balance equations and show how they can be solved to compute the limiting probability P_0 that there are no tourists inside the bar.

<u>Solution</u>: The situation at the bar can be described as an $M/M/1$ queueing model with capacity 6, where the arrival times are i.i.d. Exp(λ) and the service times are i.i.d. Exp(μ). In this case, there is no need to impose any condition on λ and μ because the queue size is bounded and cannot increase indefinitely. The balance equations are

$$\begin{aligned} \lambda P_0 &= \mu P_1, \\ (\lambda+\mu)P_n &= \lambda P_{n-1} + \mu P_{n+1}, \qquad 1 \leqslant n \leqslant 5, \\ \mu P_6 &= \lambda P_5. \end{aligned}$$

Rewriting in terms of P_0, we obtain $P_1 = \frac{\lambda}{\mu} P_0$ and $P_n = \left(\frac{\lambda}{\mu}\right)^n P_0$. Using $\sum_{n=0}^{6} P_n = 1$, we get that

$$P_0 = \frac{\left(1 - \frac{\lambda}{\mu}\right)}{1 - \left(\frac{\lambda}{\mu}\right)^7}.$$

Assume now that, during the dry season, the bar opens an outdoor counter and removes the restriction that allows no more than 5 tourists inside, so that now there is enough space for everyone and every tourist that arrives can join the queue (if there is any) and order. Assume that tourists still arrive according to a Poisson process with rate λ. However, the time takes to be served is not anymore exponentially distributed, but is described by a continuous random variable S with mean μ and variance σ^2, independently of everything else.

(ii) What condition, if any, must now λ and μ satisfy? What is the limiting probability P_0 that there are no tourists at the bar.

<u>Solution</u>: The situation at the bar can now be described as an $M/G/1$ queueing model, where the service times are not anymore exponential random variables, but they have mean μ and variance σ^2. The condition to be satisfied is $\lambda < \frac{1}{\mathbb{E}[S]} = \frac{1}{\mu}$. The average number of people in service is $\lambda \mathbb{E}[S] = \lambda\mu$ from the cost equation, and it is also $0P_0 + 1(1-P_0) = 1 - P_0$. Hence we obtain

$$P_0 = 1 - \lambda\mu.$$

(iv) Bonus (2 points): State and prove the Pollaczek-Khintchine formula to show that

$$W_Q = \frac{\lambda(\sigma^2 + \mu^2)}{2(1 - \lambda\mu)},$$

where W_Q is the average time a tourist spends in the queue at the bar.

<u>Solution</u>: Recall that the work in the bar at any time t is defined as the sum of the remaining service times of all tourists in the bar at time t. Let V denote the average work in the bar. Assume that each tourist pays at a rate of y per unit time when his/her remaining service time is y. More precisely, a tourist pays at rate S per time unit while he is in queue and at rate $S - x$ after being in service for time x. Under this cost rule, the rate at which the system earns is the work in the bar, and from the basic cost equation we also have that

$$\mathbb{E}[\text{rate at which the system earns}] = \lambda\mathbb{E}[\text{amount paid by a tourist}].$$

Hence,

$$V = \lambda\mathbb{E}\left[SW_Q^* + \int_0^S (S - x)\,dx\right] = \lambda\mathbb{E}[S]W_Q + \frac{\lambda\mathbb{E}[S^2]}{2},$$

where W_Q^* is the time a given tourist spends in queue. Since the time a tourist waits in the queue equals the work that he/she sees in the bar when he/she arrives, by taking expectations we obtain that W_Q equals the average work seen by an arrival, which for the PASTA principle equals the average work in the bar. Hence, $W_Q = V$, from which we get the Pollaczek-Khintchine formula

$$W_Q = \frac{\lambda\mathbb{E}[S^2]}{2(1 - \lambda\mathbb{E}[S])}.$$

Since $\mathbb{E}[S] = \mu$ and $\mathrm{Var}(S) = \sigma^2$, we get

$$W_Q = \frac{\lambda(\sigma^2 + \mu^2)}{2(1 - \lambda\mu)}.$$

Exercise 4: Simulation

(i) Describe and prove the rejection method.

<u>Solution</u>: Suppose that we can simulate a random variable with density $g(x)$. The rejection method says, that, if $\frac{f(x)}{g(x)} \leqslant c$ for all x, then we can simulate a continuous random variable X with density $f(x)$ in the following way:

(1) Simulate Y with density g and $U \sim U(0, 1)$;

(2) If $U \leqslant \frac{f(Y)}{cg(Y)}$, set $X = Y$, otherwise return to step 1.

To prove that, let $K = \mathbb{P}\big(U \leqslant \frac{f(Y)}{cg(Y)}\big)$ and write

$$\mathbb{P}(X \leqslant x) = \mathbb{P}\left(Y \leqslant x \,\middle|\, U \leqslant \frac{f(Y)}{cg(Y)}\right) = \frac{\mathbb{P}\left(Y \leqslant x, U \leqslant \frac{f(Y)}{cg(Y)}\right)}{K}$$
$$= \frac{\int_{-\infty}^{x} \mathbb{P}\big(U \leqslant \frac{f(y)}{cg(y)}\big) g(y)\, dy}{K} = \frac{\int_{-\infty}^{x} \frac{f(y)}{c}\, dy}{K}.$$

Letting $x \to \infty$ shows that $K = \frac{1}{c}$, hence we obtain that

$$\mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(y)\, dy,$$

which means that the simulated random variable X has density $f(x)$ as desired.

(ii) Describe in detail how we can simulate the time of the n-th arrival of a tourist at the bar of the cultural center during the dry season. To do so, use the rejection method with trial density g being the density of an exponential random variable with parameter λ/n.

<u>Solution</u>: The arrival time of the n-th tourist S_n has a Gamma distribution with parameters n and λ, so its density is $f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$. The density g of an exponential random variable with parameter λ/n is $g(x) = \frac{\lambda}{n} e^{-\frac{\lambda}{n} x}$. We then have that

$$\frac{f(x)}{g(x)} = \frac{n\lambda^{n-1} x^{n-1}}{(n-1)!} e^{-\lambda\left(1-\frac{1}{n}\right)x}$$

and its derivative

$$\frac{d}{dx}\frac{f(x)}{g(x)} = (n - \lambda x)\frac{\lambda^{n-1} x^{n-2}}{(n-2)!} e^{-\lambda\left(1-\frac{1}{n}\right)x}$$

is 0 when $x = 0$ or $x = \frac{n}{\lambda}$. We get that $\frac{f(0)}{g(0)} = 0$ and $\frac{f\left(\frac{n}{\lambda}\right)}{g\left(\frac{n}{\lambda}\right)} = \frac{n^n}{(n-1)!} e^{-(n-1)}$, so that $\frac{f(x)}{g(x)}$ attains its maximum value $c = \frac{n^n}{(n-1)!} e^{-(n-1)}$ in $x = \frac{n}{\lambda}$. We can now use the rejection method: simulate Y with density $g(x)$ (for example using the inverse transformation method), simulate $U \sim U(0,1)$, and set $S_n = Y$ if $U \leqslant \frac{f(Y)}{cg(Y)}$, otherwise repeat the procedure.

(iii) What is the average number of iterations in the most efficient case? What distribution does the number of iterations follow?

<u>Solution</u>: The most efficient case is when $c = \max_{x \geqslant 0} \frac{f(x)}{g(x)}$. The number of iterations follows a geometric distribution with mean exactly $c = \frac{n^n}{(n-1)!} e^{-(n-1)}$.

Exercise 5: Brownian motion

(i) Given a Brownian motion $\{X(t), t \geqslant 0\}$, define the hitting time T_a of barrier $a > 0$ and derive its distribution. Prove that the maximum $M(t)$ of a Brownian motion in the interval $[0, t]$ has the same distribution of $|X(t)|$.

Solution: For $a > 0$, the hitting time of a barrier a is defined as $T_a = \inf\{t \geqslant 0 : X(t) \geqslant a\}$. To find its distribution, write

$$\mathbb{P}(X(t) \geqslant a) = \mathbb{P}(X(t) \geqslant a \,|\, T_a \leqslant t)\mathbb{P}(T_a \leqslant t) + \mathbb{P}(X(t) \geqslant a \,|\, T_a > t)\mathbb{P}(T_a > t) = \frac{1}{2}\mathbb{P}(T_a \leqslant t),$$

which gives

$$\mathbb{P}(T_a \leqslant t) = 2\mathbb{P}(X(t) \geqslant a) = \frac{2}{\sigma\sqrt{2\pi t}} \int_a^\infty e^{-x^2/(2\sigma^2 t)}\, dx = \frac{2}{\sqrt{2\pi}} \int_{\frac{a}{\sigma\sqrt{t}}}^\infty e^{-y^2/2}\, dy.$$

To answer the secon part, note that $\mathbb{P}(M(t) \geqslant a) = \mathbb{P}(T_a \leqslant t)$.. We can then write

$$\mathbb{P}(M(t) \geqslant a) = 2\mathbb{P}(X(t) \geqslant a) = \mathbb{P}(X(t) \geqslant a) + \mathbb{P}(X(t) \leqslant -a) = \mathbb{P}(|X(t)| \geqslant a),$$

which proves the statement.

In the 1970s the total surface of the Brazilian Amazon rainforest was around 4 million km^2, while at the end of 2023 it was estimated to be around 3 250 000 km^2. Current data show that the average rate of deforestation in the decade from 2009 to 2018 was approximately 5 000 km^2 per year, while in the past five years, from 2019 to 2023, it was 10 000 km^2 per year. Assume that we can model the deforestation process as a Brownian motion $\{X(t), t \geqslant 0\}$ with negative drift and variance parameter σ^2, where $X(t)$ is the size of the rainforest at time t. (We are not interested in $X(0)$, so we can ignore the initial condition and start looking at the process from when we have data).

(ii) Lately, due to the crucial role deforestation plays in the current climate crisis, more attention has been given to this matter. Assume that from this year, 2024, we are able to reduce the deforestation rate of 50% and bring it back to 5 000 km^2 per year. In which year should we expect to reach the threshold of 3 million km^2 for the total size of the Brazilian Amazon rainforest?

Solution: The total size of the rainforest is expected to decrease by 5 000 km^2 per year. To decrease by 250 000 km^2 (from 3 250 000 km^2 to 3 000 000 km^2) it will take on average 50 years starting from 2024, so should expect to reach the threshold in 2073.

Assume now that governments will be able to find an agreement to reduce the deforestation rate by another 50% from the year 2029, and eventually bring it to zero from the year 2039.

(iii) If in 2039 the size of the rainforest will be exactly at its mean value, what is the probability that it will not reach the threshold of 3 million km^2 before the year 2100? (You can leave the answer in the form of an integral, without calculating it).

Solution: In the next 5 years (2024-2028) the size of the rainforest is expected to decrease by $5 \cdot 5000 = 25\,000$ km^2 and reach the value of 3 225 000 km^2 in 2028. In the following 10 years (2029-2038) it is expected to decrease by $10 \cdot 2500 = 25\,000$ km^2 and reach the value of 3 200 000 km^2 in 2038. It will then be described by a Brownian motion without drift and variance parameter σ^2 starting from the year 2039. The probability of not reaching the threshold of 3 million km^2 before 2100 is then equal to the probability that a Brownian motion $\{X(t), t \geqslant 0\}$ with variance parameter σ^2 starting at 0 does not decrease by 200 000 km^2 within 61 years. Since by symmetry for $a < 0$ the distribution of the hitting time T_a is the same as that of T_{-a}, we have that

$$\mathbb{P}(T_a \leqslant t) = \frac{2}{\sqrt{2\pi}} \int_{\frac{|a|}{\sigma\sqrt{t}}}^{\infty} e^{-y^2/2}\, dy.$$

Hence the requested probability is

$$\mathbb{P}(T_{-200000} > 61) = 1 - \mathbb{P}(T_{-200000} \leqslant 61) = 1 - \frac{2}{\sqrt{2\pi}} \int_{\frac{200000}{\sigma\sqrt{61}}}^{\infty} e^{-y^2/2}\, dy.$$