Solutions of the second exam:
Stochastic Processes and Simulation II
August 20th 2024

Exercise 1: Poisson processes

(i) Give the definition of a counting process and explain what it means for the process to have independent increments and stationary increments.

Solution: A stochastic process $\{N(t), t \geqslant 0\}$ is said to be a counting process if $N(t)$ represents the total number of events that occur by time t. A counting process must satisfy the following:

(i) N(0) = 0;

(ii) N(t) is integer-valued;

(iii) if $s < t$, then $N(s) \leqslant N(t)$;

(iv) for $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

A counting process has independent increments if the number of events that occur in disjoint time intervals are independent, while it has stationary increments if the distribution of the number of events that occur in any time interval depends only on the length of the interval.

(ii) Give the definition of a mixed Poisson process and explain why it is not a Poisson process.

Solution: For a positive random variable L, the counting process $\{N(t), t \geqslant 0\}$ is said to be a mixed Poisson process if, conditional on $L = \lambda$, it is a Poisson process with rate λ. In general, a mixed Poisson process does not have independent increments: knowing how many events occur in an interval gives information about the possible value of L, which affects the distribution of the number of events in any other interval. Hence, it is not a Poisson process.

The Amazon rainforest is suffering a major deforestation problem which is drastically reducing its size year after year. The rate of deforestation has changed over the years ranging from an annual forest loss of around 25 000 km^2 in 2003 to an annual forest loss of around 9 000 km^2 in 2023. In the Brazilian Amazon rainforest, it has been observed that the leading political party has deeply influenced the rate of deforestation over the last decades. Denote by $\{N(t), t \geqslant 0\}$ a Poisson process describing the amount of land reduction per month, where, for $k \geqslant 1$, the k-th event occurs when $100 \cdot k$ km^2 of total land have been deforested. For example, if after 1 month has been deforested 750 km^2 of land, then N(1) = 7, and if after 2 months have been deforested 1 600 km^2 of land then N(2) = 16. Assume that the process starts at the beginning of 2024, so that time units correspond to calendar months. The rate at which events occur, i.e., the rate of deforestation, depends on the leading political party and is described by a positive random variable L with mean μ and variance σ^2.

(iii) Calculate the expected value and variance of the total land reduction of the year 2024.

Solution: Since it is a mixed Poisson process, we have that $\mathbb{E}[N(12)] = 12\,\mathbb{E}[L] = 12\mu$ and $\mathrm{Var}(N(12)) = 12\,\mathbb{E}[L] + 144\mathrm{Var}(L) = 12\mu + 144\sigma^2$.

Exercise 2: Renewal theory

(i) Let $\{N(t), t \geqslant 0\}$ be a renewal process with renewal function $m(t) = \mathbb{E}[N(t)]$. Assuming that the interarrival distribution F is continuous with density function f, derive the renewal equation for $m(t)$. For $t \leqslant 1$, what is the solution of the renewal equation when F is uniform on $(0, 1)$?

Solution: The renewal equation is

$$\begin{aligned} m(t) &= \mathbb{E}[N(t)] \\ &= \int_0^\infty \mathbb{E}(N(t) \mid X_1 = x) f(x)\, dx \\ &= \int_0^t (1 + \mathbb{E}[N(t-x)]) f(x)\, dx \\ &= F(t) + \int_0^t m(t-x) f(x)\, dx, \end{aligned}$$

where X_1 is the first interarrival time. When F is uniform on $(0, 1)$ we have $m(t) = e^t - 1$ for $t \in [0, 1]$.

Consider now a new type of tree-cutting machine for deforestation that has recently started being produced. Assume that every day such machines are introduced in the Amazon rainforest one by one at random time intervals $U_1, U_2, \dots,$ that are distributed uniformly on $(0, 1)$ independently of each other, where a unit time indicates a day. Assume also that every day the process restarts, independently of the previous day.

(ii) How many new machines are introduced on average in a day? How many in a month (30 days)?

Solution: We can describe the introduction of the machines in a single day via a renewal process $\{N(t), t \in [0, 1]\}$ with interarrival distribution $F \sim U(0, 1)$. By using the solution of the renewal equation from point (i), we have that the average number of machines introduced in a day is then $m(1) = e - 1$. Moreover, the average number of machines introduced in a month is $30\, m(1) = 30(e - 1)$. Note that we cannot write $m(30) = e^{30} - 1$, because the process regenerates every day and the solution of the renewal equation from point (i) is valid only for $t \in [0, 1]$.

Consider now one of the above machines and assume that a workers association has just bought it. The machine breaks down according to a renewal process with interarrival times uniformly distributed between 0 and 4 years. Assume that the cost of a new machine is 2 million SEK, while the expected cost of a new repair increases with the number of earlier repairs in such a way that the expected cost of the k-th repair is $100\,000 \cdot k$ SEK. Assume that at the 5-th breakdown the association decides to replace the machine and buy a new one instead of repairing it.

(iii) (6 points) What is the expected cost the association has per year?

Solution: Let a new renewal cycle begin as soon as a new machine is bought. Using renewal reward theory, we know that

$$\mathbb{E}[\text{cost per time unit}] = \frac{\mathbb{E}[\text{cost per cycle}]}{\mathbb{E}[\text{length of a cycle}]}.$$

The expected cost per cycle is given by

$$2\,000\,000 + \sum_{k=1}^{4} 100\,000 \cdot k = 2\,000\,000 + 1\,000\,000 = 3\,000\,000 \text{ SEK}.$$

Moreover, breakdowns are distributed uniformly as $U(0,4)$, hence the average time between breakdowns is 2 years. Since the machine is replaced at the 5-th breakdown, we obtain that the expected age of a machine when replaced, i.e., the expected length of a cycle, is 10 years. We conclude that the expected cost per year is $\frac{3000000}{10} = 300\,000$ SEK.

Exercise 3: Queueing theory

(i) Write the basic cost equation and describe how it can be used to derive Little's formula.

Solution: The basic cost equation is

$$\text{average rate at which the system earns} = \lambda_a \times \text{average amount an entering customer pays},$$

where λ_a is the average arrival rate of entering customers. Little's formulas can be obtained by supposing that each customer pays 1 SEK per unit time while in the system. Under this cost rule, the rate at which the system earns is just the number of customers in the system L, and the amount a customer pays is just equal to its time in the system W. Hence we get $L = \lambda_a W$.

In order to contrast the Amazon deforestation, lots of initiatives are arising to re-plant trees in the forest. The Brasilian Ministry of Environment in 2017 sponsored a project aiming at planting 73 million new trees in the following 6 years. On a smaller scale, many associations allow people to donate money online to plant trees and help with the cause. To maintain biodiversity, a lot of different species need to be planted, among which are the andiroba and the Roucouyer tree, known for their medicinal properties. Consider an association specializing in medicinal plants and planting andiroba and Roucouyer trees in different areas of the Amazon forest. Assume that anyone can donate money and place an order to plant a tree through the association website and can even specify which type of tree to plant. Assume that the association receives orders for andiroba trees according to a Poisson process with rate λ_a, while it receives orders for Roucouyer trees according to a Poisson process with rate λ_r, independently of all the other orders. When an order arrives, it is normally picked up by the people in the association who process it and make sure the tree is planted. The time it takes from when the order is picked up to when the tree is planted is exponentially distributed with mean μ, independently of the other orders and the type of tree. However, the association consists of a limited number of people, hence they can only pick up and process k orders at the time. If there are more than k orders in the system, the remaining ones are placed in a queue according to their arrival time.

(ii) What type of queueing model best describes the above situation? What is the total arrival rate λ of orders in the system? What conditions, if any, must λ and μ satisfy? Write down the balance equations.

Solution: The situation can be described as an $M/M/k$ queueing model where the arrival times are i.i.d. $\text{Exp}(\lambda)$ with $\lambda = \lambda_a + \lambda_r$ and the service times are i.i.d. $\text{Exp}(1/\mu)$. Note that the combination of two independent Poisson process with rates λ_a and λ_r is a Poisson process with rate $\lambda_a + \lambda_r$. The condition to be satisfied is $\lambda < k/\mu$. The balance equations are

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

(iii) Assuming that you know the average number of orders in the system, what is the average time between the placing of an order and the planting of its tree?

Solution: It is a simple application of Little's formula where we want W knowing L, hence $W = L/\lambda$.

Exercise 4: Simulation

(i) Describe how to use the inverse transformation method to simulate an exponential random variable with parameter λ.

Solution: The inverse transformation method says that, when F^{-1} is computable, we can simulate a random variable X from a continuous distribution F by simulating $U \sim U(0,1)$ and then setting $X = F^{-1}(U)$. Since exponential random variables have distribution $F(x) = 1 - e^{-\lambda x}$, we just need to simulate $U \sim U(0,1)$ and we obtain that $\frac{-\log(U)}{\lambda} \sim \text{Exp}(\lambda)$.

(ii) Consider the tree-planting association of the previous exercise. Describe how we can simulate the arrival times of the andiroba tree orders by simulating only standard uniform random variables.

Solution: We can simulate the standard Poisson process with rate λ_a by simulating the sequence of exponentially distributed interarrival times $X_1, X_2, \ldots$. We can simulate standard uniform random variables $U_1, U_2, \ldots$, and use the inverse transformation method setting $X_i = -\frac{\log(U_i)}{\lambda_a}$. The arrival times are then given by $S_i = \sum_{j=1}^{i} X_j$, for $i \geqslant 1$.

(iii) Let T_n denote the time of the n-th order placed on the website of the association. What is its distribution? How can we simulate T_n in such a way that we can tell if it is for an andiroba tree or a Roucouyer tree?

Solution: The arrival times of the orders follow a Poisson process with rate $\lambda = \lambda_a + \lambda_r$, hence $T_n \sim \Gamma(n, \lambda)$ has a gamma distribution with parameters n and λ. We can simulate the first n orders placed for andiroba trees by simulating a Poisson process with rate λ_a as in point (ii). We label the arrival times with the letter a. Analogously, we can simulate the first n orders placed for Roucouyer trees by simulating a Poisson process with rate λ_r as in point (ii). We label the arrival times with the letter r. We then merge the two processes, order the arrival times through a list and select the n-th smallest one. Its label indicates the type of tree.

Exercise 5: Brownian motion

(i) (3 points) Give the definition of a standard Brownian motion $\{X(t), t \geqslant 0\}$. Give some intuition on why $X(t)$ is a continuous function of t and it is nowhere differentiable.

Solution: A stochastic process $\{B(t), t \geqslant 0\}$ is said to be a standard Brownian motion if:

(i) $B(0) = 0$;

(ii) $\{B(t), t \geqslant 0\}$ has independent and stationary increments;

(iii) for every $t > 0$, $B(t) \sim \mathcal{N}(0, t)$.

To show that $X(t)$ is a continuous function of t, we must show that $\lim_{h\to 0}(X(t+h)-X(t)) = 0$ almost surely. Note that the random variable $X(t+h) - X(t) \sim \mathcal{N}(0, h)$ has mean 0 and variance h, and so it would seem to converge to a random variable with mean 0 and variance 0 as $h \to 0$. Moreover, note that $\frac{X(t+h)-X(t)}{h} \sim h^{-1}\mathcal{N}(0, h) \sim \mathcal{N}(0, h^{-1})$ has mean 0 and variance $1/h$, which converges to ∞ if $h \to 0$. Hence, $X(t)$ is not differentiable.

The Amazon rainforest represents over half of Earth's remaining rainforests and comprises the largest and most biodiverse tract of tropical rainforest in the world, with around 390 billion trees in about 16 000 species. Assume that we can model the number of trees in the Amazon rainforest as a standard Brownian motion $\{X(t), t \geqslant 0\}$ with drift, where $X(t)$ is the number of trees at time t. We can assume as a starting point the initial condition

$X(0) = 390\ 000\ 000\ 000$ at the beginning of the year 2024. The drift is given by the estimate that deforestation is responsible for a loss of trees at an average rate of 1.5 billion per year.

(ii) (3 points) In how many years should we expect to reach the threshold of 300 billion trees in total in the Amazon rainforest?

Solution: The deforestation rate can be described by a negative drift $\mu_1 < 0$ of 1.5 billion trees per year, so that to go from 390 billion trees to 300 billion trees we must decrease by 90 billion trees. This occurs in 60 years.

Consider now also the action of re-planting trees and assume that various initiatives from governments and associations manage to re-plant trees at an average rate of 500 million (half billion) per year.

(iii) (3 points) In how many years should we expect to reach the threshold of 300 billion trees in total in the Amazon rainforest?

Solution: The re-planting rate can be described by a positive drift $\mu_2 > 0$ of 0.5 billions trees per year. Together with the deforestation drift μ_1, we obtain that the number of trees evolves according to a standard Brownian motion with drift $\mu_1 + \mu_2$, hence with a negative drift of 1 billion trees a year. It takes then 90 years to decrease by 90 billion trees and reach the threshold.

Assume that the average deforestation rate will decrease linearly in the next 4 years and then it will suddenly drop to match the average re-planting rate. More precisely, it will still be 1.5 billion trees in the year 2024, then 1.4 billion trees in 2025, 1.3 billion trees in 2026, 1.2 billion trees in 2027, 1.1 billion in 2028, and then drop to 500 million trees starting from the year 2029.

(iv) (3 points) If the number of trees in 2029 will be exactly at its mean value, what is the probability that it will reach the threshold of 350 billion trees before the year 2065? (You can leave the answer in the form of an integral, without calculating it).

Solution: The deforestation rate is decreasing linearly every year, hence the drift associated with deforestation is changing every year. In 2024 the sum of the drifts of deforestation and re-planting is still negative 1 billion trees, in 2025 it is negative 0.9 billion trees, in 2026 negative 0.8 billion trees, in 2027 negative 0.7 billion trees and in 2028 negative 0.6 billion trees. Since the number of trees in 2029 is at its mean value, it means that it has decreased by a total of $1+0.9+0.8+0.7+0.6 = 4$ billion trees, hence it is at the value of 386 billion trees in 2029. From this year on, the drifts from deforestation and re-planting sum up to zero, hence the number of trees evolves as a standard Brownian motion without drift. The probability of reaching the threshold of 350 million trees before 2065 is then equal to the probability that a standard Brownian motion decreases by 36 billion within 36 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

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

Hence the requested probability is

$$\mathbb{P}(T_{-36000000000} \leqslant 36) = \frac{2}{\sqrt{2\pi}} \int_{\frac{36000000000}{\sqrt{36}}}^{\infty} e^{-y^2/2}\, dy = \frac{2}{\sqrt{2\pi}} \int_{6000000000}^{\infty} e^{-y^2/2}\, dy \approx 0.$$