

MA3160: Probability and Stochastic Processes
Midterm exam – solutions

1. *We study a medical test, used to detect a disease which occurs in 1% of the population. If a given person has the disease, then the test detects it correctly with probability 0.98, while if this person is healthy, the test wrongly detects the disease with probability 0.03.*

Let us denote by D the event that the person has the disease, so that $\mathbb{P}(D) = 0.01$. We also denote by T the event that the test is positive, i.e. it says that the person is sick. We know that $\mathbb{P}(T|D) = 0.98$, and $\mathbb{P}(T|D^c) = 0.03$.

- (i) *If the test says that a person has the disease, find the probability that this person is indeed sick.*

We want to determine $\mathbb{P}(D|T)$, which can be done by using Bayes' formula:

$$\begin{aligned}\mathbb{P}(D|T) &= \frac{\mathbb{P}(T|D)\mathbb{P}(D)}{\mathbb{P}(T|D)\mathbb{P}(D) + \mathbb{P}(T|D^c)\mathbb{P}(D^c)} \\ &= \frac{(0.98) \cdot (0.01)}{(0.98) \cdot (0.01) + (0.03) \cdot (0.99)} \\ &\simeq 0.2481 \dots\end{aligned}$$

(we used that $\mathbb{P}(D^c) = 1 - \mathbb{P}(D) = 1 - (0.01) = 0.99$).

- (ii) *If the test says that a person is healthy, find the probability that this person does not have the disease.*

We use again Bayes' formula:

$$\begin{aligned}\mathbb{P}(D^c|T^c) &= \frac{\mathbb{P}(T^c|D^c)\mathbb{P}(D^c)}{\mathbb{P}(T^c|D^c)\mathbb{P}(D^c) + \mathbb{P}(T^c|D)\mathbb{P}(D)} \\ &= \frac{(0.97) \cdot (0.99)}{(0.97) \cdot (0.99) + (0.02) \cdot (0.01)} \\ &\simeq 0.9998 \dots\end{aligned}$$

(we used that $\mathbb{P}(T^c|D) = 1 - \mathbb{P}(T|D) = 1 - (0.98) = 0.02$, and $\mathbb{P}(T^c|D^c) = 1 - \mathbb{P}(T|D^c) = 1 - (0.03) = 0.97$).

2. *Let $p \in (0, 1)$. Consider two independent random variables X and Y , each equal to 1 with probability p , and to -1 with probability $1 - p$. Let $Z = X \cdot Y$.*

- (i) *What is the range E_Z of Z ? Determine $\mathbb{P}(Z = z)$ for each $z \in E_Z$.*

There are two possible values that Z can take, 1 and -1 , so its range is $E_Z = \{1, -1\}$.

- We have

$$\begin{aligned}\mathbb{P}(Z = 1) &= \mathbb{P}(\{X = 1, Y = 1\} \cup \{X = -1, Y = -1\}) \\ &= \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = -1, Y = -1).\end{aligned}$$

Since X and Y are assumed to be independent,

$$\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 1) = p \cdot p = p^2,$$

and similarly, $\mathbb{P}(X = -1, Y = -1) = (1 - p)^2$. Hence,

$$\mathbb{P}(Z = 1) = p^2 + (1 - p)^2.$$

- Analogously, we have

$$\begin{aligned}\mathbb{P}(Z = -1) &= \mathbb{P}(X = 1, Y = -1) + \mathbb{P}(X = -1, Y = 1) \\ &= p(1 - p) + (1 - p)p \\ &= 2p(1 - p).\end{aligned}$$

- (ii) Compute $\mathbb{P}((X, Z) = (x, z))$ for all possible values (x, z) that the pair (X, Z) can take.

The set all possible values that the pair (X, Z) can take, i.e. its range, is

$$E_{(X, Z)} = \{1, -1\}^2 = \{(1, 1), (1, -1), (-1, 1), (-1, -1)\}.$$

We have

$$\mathbb{P}((X, Z) = (1, 1)) = \mathbb{P}(X = 1, Z = 1) = \mathbb{P}(X = 1, X \cdot Y = 1) = \mathbb{P}(X = 1, Y = 1).$$

Since X and Y are independent by assumption,

$$\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 1) = p \cdot p = p^2.$$

Hence, $\mathbb{P}((X, Z) = (1, 1)) = p^2$.

Similarly,

- $\mathbb{P}((X, Z) = (1, -1)) = \mathbb{P}(X = 1, Y = -1) = p(1 - p),$
- $\mathbb{P}((X, Z) = (-1, 1)) = \mathbb{P}(X = -1, Y = -1) = (1 - p)^2,$
- and $\mathbb{P}((X, Z) = (-1, -1)) = \mathbb{P}(X = -1, Y = 1) = (1 - p)p.$

- (iii) For which values of $p \in (0, 1)$ are X and Z independent (explain)?

For the independence of X and Z , we need in particular

$$p(1 - p) = \mathbb{P}((X, Z) = (1, -1)) = \mathbb{P}(X = 1, Z = -1) = \mathbb{P}(X = 1) \cdot \mathbb{P}(Z = -1).$$

We already know $\mathbb{P}(Z = -1) = 2p(1 - p)$, and $\mathbb{P}(X = 1) = p$. Hence,

$$p(1 - p) = p \cdot (2p(1 - p)).$$

Since $p \neq 0, 1$ by assumption, we can simplify it to get:

$$1 = 2p,$$

so $p = \frac{1}{2}$. Conversely, it is easy to check that for this particular value of p , X and Z are independent.

3. Let X be a continuous random variable, with density $f_X(x)$. We denote by F_X its cumulative distribution function.

- (i) Let $Y = X^2$. Find its cumulative distribution function and its density, in terms of F_X and f_X .

We have $F_Y(y) = 0$ and $f_Y(y) = 0$ for $y \leq 0$. For $y > 0$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \mathbb{P}(X \leq \sqrt{y}) - \mathbb{P}(X \leq -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Hence,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} (F_X(\sqrt{y}) - F_X(-\sqrt{y})) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

- (ii) Assume now that X has exponential distribution with parameter 2, i.e. it is continuous with density $f_X(x) = 2e^{-2x} \mathbb{1}_{[0,+\infty)}(x)$. Determine the cumulative distribution function and the density of X^2 .

First, $F_{X^2}(y) = f_{X^2}(y) = 0$ for $y \leq 0$. For $y > 0$, we use the result from the previous question with $f_X(x) = 2e^{-2x} \mathbb{1}_{[0,+\infty)}(x)$, and the associated cumulative distribution function $F_X(x) = 1 - e^{-2x}$:

$$F_{X^2}(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = 1 - e^{-2\sqrt{y}}$$

(since $F_X(-\sqrt{y}) = 0$), and

$$f_{X^2}(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) = \frac{1}{2\sqrt{y}} \cdot 2e^{-2\sqrt{y}} = \frac{1}{\sqrt{y}} e^{-2\sqrt{y}}$$

(since $f_X(-\sqrt{y}) = 0$).

4. The arrival times of customers in a bakery can be modeled by a Poisson process $(N_t)_{t \geq 0}$ with rate $\lambda = 4$.

- (i) Let $\mu > 0$, and let X be a random variable having exponential distribution with parameter μ (i.e. it is continuous with density $f_X(x) = \mu e^{-\mu x} \mathbb{1}_{[0,+\infty)}(x)$). Prove that it has expectation $\mathbb{E}[X] = \frac{1}{\mu}$.

This is the formula seen in class for the expectation of a random variable with exponential distribution. We can compute it as follows: first, using the definition of the expectation for a continuous random variable,

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^{+\infty} x \mu e^{-\mu x} \mathbb{1}_{[0,+\infty)}(x) dx = \int_0^{+\infty} x \mu e^{-\mu x} dx.$$

We then use an integration by parts:

$$\mathbb{E}[X] = - \int_0^{+\infty} x d(e^{-\mu x}) = (-x e^{-\mu x}) \Big|_0^{+\infty} + \int_0^{+\infty} e^{-\mu x} dx = -\frac{1}{\mu} e^{-\mu x} \Big|_0^{+\infty} = \frac{1}{\mu}.$$

- (ii) Denote by $(S_n)_{n \geq 1}$ the successive arrival times of customers, and $S_0 = 0$. For each $n \geq 1$, recall what the distribution of $(S_n - S_{n-1})$ is.

We know that in a Poisson process with rate λ , the interarrival times have exponential distribution with parameter λ . Hence in this case, we obtain that $(S_n - S_{n-1})$ has exponential distribution with parameter $\lambda = 4$.

- (iii) For $n \geq 1$, determine the expectation of $(S_n - S_{n-1})$, and then of S_n .

We know from (ii) that $(S_n - S_{n-1})$ has exponential distribution with parameter 4, so we deduce from (i) that it has expectation

$$\mathbb{E}[S_n - S_{n-1}] = \frac{1}{4}.$$

By writing $S_n = \sum_{i=1}^n (S_i - S_{i-1})$ (since $S_0 = 0$), we then obtain

$$\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^n (S_i - S_{i-1})\right] = \sum_{i=1}^n \mathbb{E}[S_i - S_{i-1}] = n \cdot \frac{1}{4}.$$

- (iv) Compute the probability of the following event: $\{\text{the second customer arrives before time 2, two customers arrive during } [4, 10], \text{ and at least one arrives in } [8, 16]\}$.

Denote the event under consideration by A . First, $\{\text{the second customer arrives before time 2}\}$ can be expressed equivalently in terms of the counting process $(N_t)_{t \geq 0}$ as $\{N_2 \geq 2\}$. Hence, the desired probability is equal to

$$\mathbb{P}(A) = \mathbb{P}(N_2 \geq 2, N_{10} - N_4 = 2, N_{16} - N_8 \geq 1).$$

We then make appear independent increments by using the decomposition

$$\begin{aligned} \{N_{10} - N_4 = 2, N_{16} - N_8 \geq 1\} &= \{N_8 - N_4 = 2, N_{10} - N_8 = 0, N_{16} - N_{10} \geq 1\} \\ &\cup \{N_8 - N_4 = 1, N_{10} - N_8 = 1\} \cup \{N_8 - N_4 = 0, N_{10} - N_8 = 2\} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(N_2 \geq 2) (\mathbb{P}(N_8 - N_4 = 2, N_{10} - N_8 = 0, N_{16} - N_{10} \geq 1) \\ &\quad + \mathbb{P}(N_8 - N_4 = 1, N_{10} - N_8 = 1) + \mathbb{P}(N_8 - N_4 = 0, N_{10} - N_8 = 2)) \\ &= (1 - \mathbb{P}(N_2 = 0) - \mathbb{P}(N_2 = 1)) (\mathbb{P}(N_8 - N_4 = 2) \mathbb{P}(N_{10} - N_8 = 0) \mathbb{P}(N_{16} - N_{10} \geq 1) \\ &\quad + \mathbb{P}(N_8 - N_4 = 1) \mathbb{P}(N_{10} - N_8 = 1) + \mathbb{P}(N_8 - N_4 = 0) \mathbb{P}(N_{10} - N_8 = 2)), \end{aligned}$$

using that the increments $N_2 = N_2 - N_0$, $N_8 - N_4$, $N_{10} - N_8$ and $N_{16} - N_{10}$ are independent. Moreover, these increments are all Poisson distributed, with respective parameters $2\lambda = 8$, $4\lambda = 16$, $2\lambda = 8$, and $6\lambda = 24$. We deduce

$$\begin{aligned} \mathbb{P}(A) &= (1 - e^{-8} - e^{-8} \cdot 8) \\ &\quad \cdot \left(e^{-16} \cdot \frac{16^2}{2} \cdot e^{-8} \cdot (1 - e^{-24}) + e^{-16} \cdot 16 \cdot e^{-8} \cdot 8 + e^{-16} \cdot e^{-8} \cdot \frac{8^2}{2} \right) \\ &= (1 - 9e^{-8}) (288e^{-24} - 128e^{-48}). \end{aligned}$$

- (v) We now assume that upon entering the bakery, each customer decides to buy either one bread with probability $\frac{2}{3}$, or one croissant with probability $\frac{1}{3}$ (i.e. the customer chooses exactly one of these two options, it never chooses both), independently of the other customers. What can you say about the corresponding counting processes, denoted (respectively) by $(N_t^B)_{t \geq 0}$ and $(N_t^C)_{t \geq 0}$?

From the theorem seen in class about “decomposing” Poisson processes, $(N_t^B)_{t \geq 0}$ and $(N_t^C)_{t \geq 0}$ are independent Poisson processes, with respective rates $\lambda^B = \lambda \cdot \frac{2}{3} = \frac{8}{3}$ and $\lambda^C = \lambda \cdot \frac{1}{3} = \frac{4}{3}$.

5. The arrival times of customers in two neighboring restaurants (called A and B) can be modeled by two independent Poisson processes $(N_t)_{t \geq 0}$ and $(N'_t)_{t \geq 0}$, with respective rates λ and λ' .

- (i) For $t \geq 0$, compute $\text{Cov}(N_{4t} - N_t, N_{5t} - N_{2t})$.

We make appear independent increments: write $N_{4t} - N_t = (N_{4t} - N_{2t}) + (N_{2t} - N_t)$ and $N_{5t} - N_{2t} = (N_{5t} - N_{4t}) + (N_{4t} - N_{2t})$, so

$$\begin{aligned} \text{Cov}(N_{4t} - N_t, N_{5t} - N_{2t}) &= \text{Cov}(N_{4t} - N_{2t}, N_{5t} - N_{4t}) + \text{Cov}(N_{4t} - N_{2t}, N_{4t} - N_{2t}) \\ &\quad + \text{Cov}(N_{2t} - N_t, N_{5t} - N_{4t}) + \text{Cov}(N_{2t} - N_t, N_{4t} - N_{2t}) \\ &= \text{Cov}(N_{4t} - N_{2t}, N_{4t} - N_{2t}) = \text{Var}(N_{4t} - N_{2t}) \end{aligned}$$

since $\text{Cov}(N_{4t} - N_{2t}, N_{5t} - N_{4t}) = \text{Cov}(N_{2t} - N_t, N_{5t} - N_{4t}) = \text{Cov}(N_{2t} - N_t, N_{4t} - N_{2t}) = 0$ by independence. Since $N_{4t} - N_{2t}$ has Poisson distribution with parameter $\lambda \cdot (4t - 2t) = 2\lambda t$, it has variance $2\lambda t$. Hence,

$$\text{Cov}(N_{4t} - N_t, N_{5t} - N_{2t}) = 2\lambda t.$$

- (ii) In each of the two restaurants, we consider the arrival time of the first customer: it is denoted by T in restaurant A, and T' in restaurant B. Recall what the distributions of T and T' are.

They both have exponential distribution, with respective parameters λ and λ' .

- (iii) Compute the probability that the first customer to arrive enters the first restaurant, i.e. $\mathbb{P}(T < T')$.

We use the distribution of T and T' , recalled in (ii), and also the fact that they are independent: (T, T') is thus jointly continuous with density

$$f_{(T, T')}(t, t') = (\lambda e^{-\lambda t} \mathbb{1}_{[0, +\infty)}(t)) \cdot (\lambda' e^{-\lambda' t'} \mathbb{1}_{[0, +\infty)}(t')).$$

Hence,

$$\mathbb{P}(T < T') = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{(T, T')}(t, t') \mathbb{1}_{t < t'} dt dt' = \int_0^{+\infty} \left(\int_t^{+\infty} \lambda' e^{-\lambda' t'} dt' \right) \lambda e^{-\lambda t} dt.$$

Using now that $\int_t^{+\infty} \lambda' e^{-\lambda' t'} dt' = e^{-\lambda' t}$, we obtain

$$\mathbb{P}(T < T') = \int_0^{+\infty} e^{-\lambda' t} \lambda e^{-\lambda t} dt = \lambda \int_0^{+\infty} e^{-(\lambda + \lambda')t} dt = \frac{\lambda}{\lambda + \lambda'}.$$

- (iv) *What can you say about the counting process $(N_t + N'_t)_{t \geq 0}$ of customers entering one of the two restaurants?*

By the theorem seen in class about “combining” independent Poisson processes, $(N_t + N'_t)_{t \geq 0}$ is still a Poisson process, with rate the sum of the rates, i.e. $\lambda + \lambda'$.