KASDI MERBAH UNIVERSITY, OUARGLA
DEPARTMENT OF COMPUTER SCIENCE AND IT
STATISTICS FOR DATA SCIENCE COURSE
M1 ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

PROBABILITY DISTRIBUTIONS: EXERCISES

Exercise 1. *F*or the driver's license exam, candidates must answer a 20-question questionnaire by choosing one of four proposed answers, only one of which is correct for each question. A completely ignorant candidate decides to try their luck by randomly selecting an answer for each question.

1. What is the probability distribution of the number X of correct answers for the candidate?
2. Calculate $P(X \geq 10)$.

Exercise 2. *I*t is known that the probability that a person is allergic to a certain medication is 0.05. A sample of 20 people is chosen, and we call X the random variable whose value is the number of allergic people in the sample.

1. Determine, with justification, the probability distribution of X.
2. Calculate the probabilities of the following events:
   (a) There are exactly two allergic people in the sample.
   (b) There are at least two allergic people in the sample.
3. Deduce the average number of allergic people in this sample.

Exercise 3. *I*n a group of 30 students, 10 have a personal vehicle. Five students are randomly chosen from this group. We denote by X the variable that counts the number of chosen students with a personal vehicle.

1. What is the probability distribution of X?
2. Calculate the probability that there are 3 students out of the 5 who have personal vehicles.

Exercise 4. *I*n a box, there are n cards numbered from 1 to n. Cards are drawn successively with replacement until the card with the number n is drawn. Let Z be the number of draws performed.

1. Determine the distribution of Z. Calculate the probability that the number of cards drawn is equal to r, for $r \geq 1$.
2. Suppose that $n = 100$, what is the probability that the number of cards drawn is less than or equal to 50?

Exercise 5. *A* hamster is placed in a cage with 5 gates, only one of which allows it to exit the cage. In each unsuccessful attempt, the hamster receives an electric shock and is placed back in its initial location.

1. Assuming that the hamster does not learn and randomly chooses one of the 5 options in each new attempt, determine the probabilities of the following events:
   a) The hamster escapes on the first attempt.
   b) The hamster escapes on the third attempt.
   c) The hamster escapes on the seventh attempt.
2. The hamster now remembers the unsuccessful attempts and chooses equally among the gates he has not tried before. We denote by X the random variable equal to the number of attempts.
   a) What values can X take? Determine its probability distribution.
   b) Determine the expected value $E(X)$.
   c) Determine the variance $V(X)$.

Exercise 6. *T*he average number of accidents on a highway per day is 2.

- Calculate the probability that the number of accidents does not exceed 4 next Monday.
- What is the probability that the number of accidents is greater than 10 for the coming week?

Exercise 7. *A*n airline provides flights between different cities. It is assumed that the total number of flight delays per week follows a Poisson distribution with a parameter of 3. Calculate the probability that the total number of delays is less than 6 for a given month.

Exercise 8. *A* bus stop is served every exact quarter of an hour (e.g., at 8:00, 8:15, 8:30, etc.).

A commuter arrives at the bus stop at a randomly uniformly distributed time between 8:00 and 8:30. We model their arrival time with a random variable X, which is the number of minutes that have elapsed since 8:00 when the commuter arrives at the bus stop. Therefore, $X \sim \mathcal{U}[0, 30]$.

- What is the probability that the commuter waits for the bus for less than 5 minutes?
- What is the probability that the waiting time is between 5 and 10 minutes?

Exercise 9. *A* real number is randomly chosen between -3 and 5.

1. What is the probability of obtaining a number strictly less than 1?
2. What is the probability of obtaining a number greater than or equal to 3?
3. What is the probability that the chosen number is strictly less than 1, given that it is strictly positive?

Exercise 10. *T*he lifetime of a processor (in years) is modeled by an exponential distribution with a rate parameter of $\frac{1}{2}$.

1. What is the average lifetime of this processor?
2. What is the probability that the processor lasts more than 6 months?
3. What is the probability that the processor lasts less than 3 months?
4. What is the probability that the processor lasts between 3 and 6 months?

Exercise 11. *T*he lifetime of an electronic device is a random variable, expressed in hours, that follows an exponential distribution with a rate parameter of 0.00026.

1. What is the probability that the device's lifetime is at most 1000 hours?
2. Deduce the probability that the device's lifetime is at least 1000 hours.
3. Given that the device's lifetime has exceeded 1000 hours, what is the probability that its lifetime exceeds 2000 hours?
4. Given that the device has operated for more than 2000 hours, what is the probability that it fails before 3000 hours?

Exercise 12. *L*et X be the random variable modeling the lifetime of the radioactive component C_{14}. X follows an exponential distribution with a parameter $\lambda > 0$.

The half-life of the random variable X is the duration θ for which $P(X \leq \theta) = 0.5$.

1. Show that $\theta = \frac{\ln 2}{\lambda}$.
2. If it is known that the half-life of C_{14} is estimated at 2000 years, calculate:
   a) The average lifetime of C_{14}.
   b) The probability that its lifetime is at most 1000 years.
   c) The probability that its lifetime exceeds 3000 years, given that it has exceeded 2000 years.
3. Determine t such that $P(X \leq t) = 0.2$.

Exercise 13. *W*e assume that the duration of a phone call, measured in minutes, follows an exponential distribution with a rate parameter of $\frac{1}{10}$. A person arrives at a phone booth, and at that precise moment, another person passes by this person.

1. What is the probability that this person waits for more than ten minutes?
2. What is the probability that they wait between ten and twenty minutes?

Exercise 14. *L*et X be a random variable representing the mathematics grades of a group. It is assumed that X follows a normal distribution $\mathcal{N}(40, 10)$.

We want to divide this group into three groups A, B, C such that $A = \{X \geq \alpha\}$, $B = \{\beta \leq X \leq \alpha\}$, and $C = \{X < \beta\}$.

- Determine the values of α and β if it is known that $P(A) = 2P(B)$ and $P(B) = P(C)$.

Exercise 15. *T*he random variable Z follows a $\mathcal{N}(0,1)$ distribution. What is the probability that:

- $Z < 1.2$;
- $0.8 < Z < 1.2$;
- $-0.4 < Z < 1.5$;
- $|Z| > 0.01$.

Exercise 16. *L*et X be a normal random variable with parameters $m = 4$ and $\sigma = 5$.

1. Calculate the probabilities: $P(|X| \leq 4)$, $P(X > 2)$.
2. Determine a and b such that $P(X > a) \approx 0.04947$ and $P(X < b) \approx 0.15865$.

Exercise 17. Calculate the mathematical expectation and variance of a normal random variable X given that $P(X \leq 2) = 0.5793$ and $P(X > 5) = 0.2119$.

Exercise 18. *I*t is assumed that blood sugar levels in a population are normally distributed with a mean of 1 g/L and a standard deviation of 0.03 g/L. We measure the blood sugar level in a randomly selected individual.

1. Calculate the probability that their blood sugar level is:
   a) Less than 1.06 g/L;
   b) Greater than 0.9985 g/L;
   c) Between 0.94 and 1.08 g/L.
2. We measure the blood sugar levels of 1000 randomly selected individuals. Calculate the average number of individuals whose blood sugar level is greater than 0.99 g/L.

Exercise 19. *I*t is assumed that the heights of 615 students are normally distributed with a mean of 1.75 meters and a standard deviation of 20 cm. Calculate the number of students with heights:

1. Less than or equal to 1.50 meters;
2. Between 1.50 meters and 1.65 meters;
3. Greater than or equal to 2 meters.

Exercise 20. *A* farmer wants to collect statistics on his production of chicken eggs. He knows that out of the two thousand eggs collected in a day, 104 had a weight less than 53 grams, and 130 had a weight greater than 63 grams.

1. Assuming that the random variable representing the weight in grams of an egg follows a normal distribution, provide an estimate of the parameters of this distribution.
2. Deduce the average weight of an egg.
3. How many very large eggs (XL) can he expect to sell in a year (a very large egg weighs more than 71 grams)?

Exercise 21. *The capacity of elevators is determined by the fact that a person's weight follows a normal distribution with a mean of 75 kg and a standard deviation of 5 kg.*

In an elevator of type WH1, the maximum number of people is 9. A light indicator shows that there is an overweight if the total weight exceeds 700 kg, in which case the elevator does not work. Calculate

1. the probability that a person getting into the elevator has a weight greater than 85 kg,
2. the probability of overweight when a group of 9 people enters the elevator.

Exercise 22. *A factory produces electronic components, and 5% of them are defective. A sample of 20 components is considered. Let X be the number of defective components in this sample.*

1. Determine the probability distribution of X and write its probability mass function.
2. What is the probability that at most 2 components are defective?
3. What is the average number of defective components in this sample?

Now, consider a sample of 2000 components. X can be approximated by a random variable Y that follows a normal distribution.

1. Specify the parameters of Y.
2. Calculate approximately the probability that:
   (a) The number of defective components is equal to 120.
   (b) The number of defective components is less than 85.
3. Determine the approximate number α of components for which $P(X \leq \alpha) = 0.67$.