Introduction to Probability Theory and Statistics
Exam, June 2021

Probability

Problem 1
A result of a COVID-19 test can be either positive (indicating that a person is infected) , negative (indicating that a person is not infected) or inconclusive (an error has occurred during test processing and the result can not be determined). Currently in Denmark many people are being tested and 90% of all test results are negative, 7% of all results are positive, and 3% of all results are inconclusive.

However, tests are not perfect and 10% of all tests that show "negative" are wrong and a person is actually infected. Tests with a positive result are wrong in 20% of the cases. If a test result is "inconclusive", in 50% of the cases a person is infected and in 50% not infected.

(a) Given that we take a person who has been tested at random, what is the probability that a person is infected?

(b) Suppose that we know that a person is not infected. What is the probability that the test will show "positive"?

Problem 2
The joint probability density function of a random variable X and a random variable Y equals

$$f(x,y) = \begin{cases} c & 0 \le x \le a,\ 0 \le y \le b \\ 0 & otherwise \end{cases}$$

(a) Find numerical values for a, b, and c, if it is known that $E[X] = 4$ and $E[Y] = 8$.

(b) Find cumulative distribution function for X. Draw a graph for this function.

(c) Find the following probabilities:

(c.1) $P(X > 1.5)$

(c.2) $P(X \ge 1.5)$

(c.3) $P(X < 2)$

(c.4) $P(X = 0.5)$

(d) Find $E[XY]$.

(e) Find $P(-3 < X < 3, -5 < Y < 5)$.

(f) Does distribution $f(x, y)$ have a special name?

Problem 3

A student group has built a robot consisting of 10 components. Some components are inserted to achieve redundancy and the robot is functioning if at least 4 components are working. Each component is functioning independently of the other components.

Each component has been tested individually and it was found that a lifetime of an individual component is following an exponential distribution with a parameter $\lambda = 1/50$. Lifetime is measured in days.

(a) What is an average lifetime of an individual component?

(b) Given that an individual component is already working for 25 days, what is the probability that it will be working for additional 100 days?

(c) Now we consider the whole robot. What is the probability that a robot is working after 100 days?

Statistics

Problem 4
To understand how the human memory works, a group of Psychology students at Aalborg University prepared an experiment to test memory retention. The experiment consists of 13 subjects that were initially asked to memorize a list of disconnected numbers. The subjects were then asked to recall the numbers H hours after the time of the experiment. The proportion of numbers (P) correctly recalled H hours after the list was memorized were recorded, with the data shown in the Table. For example, 30 hours after the experiment, the subjects recalled on average 56% of the numbers.

Time (H)	proportion (P)
1	0.84
5	0.71
15	0.61
30	0.56
60	0.54
120	0.47
240	0.45
480	0.38
720	0.36
1440	0.26
2880	0.20
5760	0.16
10080	0.08

(a) Use a linear regression to model the relationship between the proportion of recalled numbers and the elapsed time. Calculate the parameters A and B of the regression and the coefficient of determination.

(b) The students quickly realized that a logarithmic rule is a better model for this study, so they repeated the regression with this transformation: $P = A + B\log(H)$, where A and B are unknown. Repeat the regression with this model and calculate the coefficient of determination.

(c) Plot the standardized residuals for both regressions (you can use any software package/visualization option for the plot).

(d) In view of the results in (a), (b) and (c), do you agree that the logarithmic dependency is a better model? Why?

Problem 5
The time between phone calls in a grocery shop is given by an exponential random variable $X \sim Exp(\lambda)$, with an unknown rate of arrivals λ. The pdf of this exponential variable is written as

$$f(x, \lambda) = \lambda \exp(-\lambda x) \tag{1}$$

(a) Find the maximum likelihood estimator of λ

(b) The shop attendant has been writing down the time between calls, and so far he has noted the next 10 values, which are given in hours:

$1, 4, 6, 1, 3, 10, 2, 6, 9, 2$

What is the probability that there are no calls for more than 5 hours?

Problem 6

An exam includes a question with four possible responses. The question is designed to be very difficult, with none of the four responses being obviously wrong, yet with only one correct answer. The exam is taken by 100 students. The teacher wants to test whether more people answered the question correctly than would be expected just due to chance (i.e., if every student picked the answer at random).

(a) Set up the hypotheses for the test

(b) Of the 100 students, 33 correctly answer the question. Find the p-value and make a decision about the null hypothesis at a 5% significance level.