Course of Study Bachelor Computer Science	Exercises Statistics WS 2022/23
Sheet V	

Probability Spaces and Basic Rules

1. Consider a random experiment of tossing two dice. Let A denote the event that the first die score is 1 and B the event that the sum of the scores is 7.
   (a) Give the sample space Ω and find $|\Omega|$.
   (b) Explicitly list the elements of the following events:
   $$A, B, A \cup B, A \cap B, A^c \cap B^c$$
2. Suppose that A and B are events in an experiment with $P(A) = 1/3$, $P(B) = 1/4$, $P(A \cap B) = 1/10$. Express each of the following events verbally and find its probability:
   $$A \setminus B, A \cup B, A^c \cup B^c, A^c \cap B^c, A \cup B^c$$
3. Suppose that A, B, and C are events in an experiment with $P(A) = 0.3$, $P(B) = 0.2$, $P(C) = 0.4$, $P(A \cap B) = 0.04$, $P(A \cap C) = 0.1$, $P(B \cap C) = 0.1$, $P(A \cap B \cap C) = 0.01$. Express each of the following events in set notation and find its probability:
   (a) At least one of the three events occurs.
   (b) None of the three events occurs.
   (c) Exactly one of the three events occurs.
   (d) Exactly two of the three events occur.
4. **Law of Large Numbers**
   The Law of Large Numbers says that the average (mean) approaches what it's estimating. As we flip a fair coin over and over, it average eventually converges to the true probability of a head (.5).

(a) To see this in action, create a R function coinPlot, which takes an integer n which is the number of coin tosses that will be simulated. As coinPlot does these coin flips it computes the cumulative sum (assuming heads are 1 and tails 0), but after each toss it divides the cumulative sum by the number of flips performed so far. It then plots this value for each of the k=1...n tosses.
Hint: Use the function sample() to simulate tossing coins.

(b) Call coinPlot several times for n=10, 100 and 1000 and describe what you see.

5. **Urn Models**
A large number of discrete probability spaces can be traced back to so-called urn models. An urn contains n balls, which do not all have to be different. From these urns r balls are drawn with or without replacement. For the result of the drawing, the order or only the quantity of the drawn balls can be of importance.
Here an urn with 10 balls is considered. 5 of them are red, 3 balls are blue and 2 balls are green. 3 balls are drawn. The following 4 cases should be distinguished:

I Drawing with replacement with respect to the order

II Drawing with replacement without observing the order

III Drawing without replacement with respect to the order

IV Drawing without replacement without observing the order

Solve the following tasks.

(a) Load the library gtools and inspect the commands combinations() and permutations(). Consider the bags $b_1 = \{a, b, c\}$ and $b_2 = \{a, a, b, c\}$ = and list all combinations and all permutations of order 2 if duplicated elements in the outpot are allowed or not allowed.

(b) Use the function sample() to determine the result of 10 random draws.

(c) Determine a suitable event space Ω and its size to describe the random experiment.
Note that depending on whether the order of the drawn balls is important or not, the result of a drawing is considered as a r-variation or as a r-combination of the n set of balls. With the

help of the permutations() and combinations() functions of the R-package gtools, the corresponding r-variations or r-combinations can be determined.

(d) Determine the probabilities of all elementary events in Ω using a Laplace model, i.e. as a determination of the ratio of the number of favorable cases by the number of all cases. The probabilities are first determined by counting methods and then by using the R function permutations().
Hint: To determine the probabilities with R, assume that the n balls are numbered consecutively, i.e. they are distinguishable, and that the order is first observed in a drawing. Every r-variation of the numbers 1 to n is equally probable. Determine the set of all these drawings with permutations(). Then map each such drawing to the corresponding elementary event. By dividing the number of drawings belonging to an elementary event and the number of all drawings, you then obtain the corresponding probabilities.

Independence and Conditional Probabilities

1. Suppose that A and B are events in an experiment with $P(A) = 1/3$, $P(B) = 1/4$, $P(A \cap B) = 1/10$. Find each of the following:

$$P(A|B), P(B|A), P(A^c|B), P(B^c|A), P(A^c|B^c)$$

2. In a certain population, 30% of the persons smoke and 8% have a certain type of heart disease. Moreover, 12% of the persons who smoke have the disease.

   (a) What percentage of the population smoke and have the disease?

   (b) What percentage of the population with the disease also smoke?

   (c) Are smoking and the disease positively correlated, negatively correlated, or independent?

3. Suppose that a bag contains 12 coins: 5 are fair, 4 are biased with probability of heads 1/3 and 3 are two-headed. A coin is chosen at random from the bag and tossed.

   (a) Find the probability that the coin shows head.

   (b) Given that the coin shows head, find the conditional probability of each coin type.

4. Suppose we know the accuracy rates of the test for both the positive case (positive result when the patient has HIV) and negative case (result when the patient doesn't have HIV). These are referred to as test sensitivity and specificity, respectively. Let "D" be the event that the patient has HIV, and let "+" indicate a positive test result and "-" a negative.

   (a) Describe test sensitivity and specifity by the above notation.

   (b) Suppose a person gets a positive test result. Express the probability that he really has HIV by the above notation.

   (c) Let the disease prevalence be .001, test sensitivity be 99.7% and test specificity be 98.5%. The probability that a person has the disease given his positive test result, i.e. $P(D|+)$. This quantity is called the positive predictive value. Similarly, $P(D^c|-)$, is called the negative predictive value, the probability that a patient does not have the disease given a negative test result. Apply Bayes Rule to evaluate both values.

   (d) The diagnostic likelihood ratio of a positive test, DLR_+, is the ratio of the two positive conditional probabilities, one given the presence of disease and the other given the absence, i.e. $DLR_+ = \frac{P(+|D)}{P(+|D^c)}$. Do you expect DLR_+ to be large or small? Evaluate DLR_+.

   (e) Similarly, the DLR_- is defined as a ratio. Do you expect DLR_- to be large or small? Evaluate DLR_-.

   (f) Show that $\frac{P(D|+)}{P(D^c|+)}$, i.e. the post-test odds of disease given a positive test result equals the pre-test odds of disease, i.e. $\frac{P(D)}{P(D^c)}$ times DLR_+ and evaluate it.

   (g) Similarly we define $\frac{P(D|-)}{P(D^c|-)}$ the post-test odds of disease given a negative test result. Show that it equals the pre-test odds of disease, i.e. $\frac{P(D)}{P(D^c)}$ times DLR_- and evaluate it.

   (h) Are post-test odds greater than pre-test odds or post-test odds are less than pre-test odds?

5. In a computer science course at an university we have the following data over a long time.
   10% of all students have attended the exercises in statistics regularly. 2% of the students who have failed the statistics exam have attended the exercises regularly. 5% of the students who have attended the exercises regularly have failed the statistics exam.

(a) Find the probability to fail the exam in statistics if the exercises in statistics are not attended regularly.

(b) What is the effect of attending the exercises regularly to passing the exam?