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

Discrete Random Variables and Distributions

1. A random experiment consists of tossing n (distinct) coins and recording the sequence of scores $X_1, X_2, ..., X_n$, where 1 denotes head and 0 denotes tail. Let Y denote the number of heads.

   (a) What is a suitable sample space Ω? How many elements contains Ω?

   (b) Express Y as a function on the sample space Ω.

   (c) Show that $|\{Y = k\}| = \binom{n}{k}$ for $k \in \{0, 1, ..., n\}$.

   (d) With n=5, explicitly list the elements in the event $\{Y = 3\}$.

2. Suppose that two fair, standard dice are tossed and the sequence of scores (X_1, X_2) recorded. Let $Y = X_1 + X_2$, denote the sum of the scores, $U = \min(X_1, X_2)$, the minimum score, and $V = \max(X_1, X_2)$ the maximum score.

   (a) Find the probability density function of (X_1, X_2).

   (b) Find the probability density function of Y.

   (c) Find the probability density function of U.

   (d) Find the probability density function of V.

   (e) Find the probability density function of (U, V).

3. R offers for a large number of probability distributions functions. The commands for each distribution are prepended with a letter to indicate the functionality:

   - "d" returns the height of the probability density function
   - "p" returns the cumulative density function
   - "q" returns the inverse cumulative density function (quantiles)
   - "r" returns randomly generated numbers

   Consider an urn with 100 balls, where are 30 balls of them are red. 20 balls are randomly drawn and let X be the number of red drawn balls.

(a) Determine the distribution of X if the balls are drawn with resp. without replacement.

(b) Plot the density of X.

(c) Generate a sample of size 20 of values of X.

(d) Compute $P(5 < X < 15)$.

(e) Determine the 25% quantile, the median and the 75% quantile of X.

4. In a game a player can bet 1$ on any of the numbers 1, 2, 3, 4, 5 and 6. Three dice are rolled. If the players number appears k times, where $k \geq 1$, the player gets k$ back plus the original stack of 1$. Over the long run, how many cents per game a player expects to win or lose playing this game?

5. Consider a lottery of 20 tickets. Among the tickets there is are 1 first prize, 4 second prizes and 15 rivets. 5 tickets are drawn from the lottery drum. Determine the probability that

   (a) 2 rivets have been drawn.

   (b) 2 rivets, 2 second prizes and the first prize were drawn.

   (c) the 5th ticket drawn is the first ticket which is not a rivet.

   Calculate the probabilities if the tickets are drawn with replacement.

6. compare Heumann, Schomaker p 176, Exercise 8.6
   A company organizes a raffle at an end-of-year function. There are 4000 raffle tickets to be sold, of which 500 win a prize. The price of each ticket is 1.5 Euro. The value of the prizes varies between 80 Euro and 250 Euro with an average of 142 Euro.

   (a) An employee wants to have a 99% guarantee of receiving three prizes. How much money does he need to spend? Use R to solve the question.

   (b) Use R to plot the function which describes the relationship between the number of tickets bought and the probability of winning at least three prizes.

   (c) Given the value of the prizes and the costs of the tickets, it is worth taking part in raffle?

Continous Random Variables and Distributions

1. The time T (in minutes) required to perform a certain job is uniformly distributed over the interval $[15, 60]$.

   (a) Find the probability that the job requires more than 30 minutes.

   (b) Given that the job is not finished after 30 minutes, find the probability that the job will require more than 15 additional minutes.

2. A continous random variable T over the positive real numbers is exponentially distributed with parameter λ ($T \sim E(\lambda)$) if
   $P(T \leq t) = 1 - e^{-\lambda t}$ for $t \geq 0$.
   The exponential distribution is one of the widely used continuous distributions. It is often used to model the time elapsed between events.

   (a) Show that exponentially distributed random variables are memoryless, i.e.

   $$P(T \leq t_2 \,|\, T > t_1) = P(T \leq t_2 - t_1) \quad 0 \leq t_1 \leq t_2$$

   (b) Find the value of λ if $E(T) = 100$ and calculate

   - $P(T = 100)$
   - $P(90 < T \leq 110)$
   - $P(T = 100 \,|\, T > 50)$
   - $P(90 < T \leq 110 \,|\, T > 50)$

Normal Distributions

1. R offers a number of functions for calculating with normal distributions. Call them up in the RStudio Help area with the keyword Normal and familiarize yourself with them.
   R has four in built functions to generate normal distribution.

   (a) The function qnorm() gives height of the probability distribution at each point for a given mean and standard deviation. Apply this function to create a plot the density of the normal distribution with mean 2.5 and standard deviation 1.5.

   (b) pnorm() gives the probability of a normally distributed random number to be less that the value of a given number (cumulative distribution function). Apply this function to create a plot of the normal distribution function with mean 2.5 and standard deviation 1.5.

(c) qnorm() takes the probability value and gives a number whose cumulative value matches the probability value (quantile). Apply this function to plot the quantiles of the normal distribution with mean 2.5 and standard deviation 1.5.

(d) rnorm() is used to generate random numbers whose distribution is normal. It takes the sample size as input and generates that many random numbers. Draw a histogram to show the distribution of the generated numbers which are normally distributed with mean 2.5 and standard deviation 1.5 .

2. If scores are normally distributed with a mean of 35 and a standard deviation of 10, what percent of the scores is:

   (a) greater than 34?

   (b) smaller than 42?

   (c) between 28 and 34?

3. Assume a normal distribution with a mean of 70 and a standard deviation of 12. What limits would include the middle 65% of the cases?

4. Suppose that weights of bags of potato chips coming from a factory follow a normal distribution with mean 12.8 ounces and standard deviation 0.6 ounces. If the manufacturer wants to keep the mean at 12.8 ounces but adjust the standard deviation so that only 1% of the bags weigh less than 12 ounces, how small does he need to make that standard deviation?

5. In a silk spinning mill, raw fibers from silk cocoons are prepared to silk threads. It can be assumed that the useful silk thread length per cocoon is a normally distributed variable with expectation 800 m and variance 6400 m^2.

   (a) Calculate the probability that the useful silk thread length from a randomly selected cocoon is at least 750 m. Also calculate the probability that the useful silk thread length from a randomly selected cocoon exceeds 1000 m.

   (b) Use appropriate assumptions and calculate the lower boundary $\underline{c}$ and the higher boundary $\bar{c}$ for the total length of the useful silk thread for 10000 cocoons. These boundaries should at the same time be guaranteed with a probability of 95%.
   The boundaries should be selected in such a way that the probability for exceeding $\bar{c}$ and going below $\underline{c}$ should be equally high.

(c) Assume that the variance still is the same as before. How high must the expetation of the useful silk thread length at least be, if we would like the total useful silk length to be at least 750 m with a probability of 0.90.

(d) 10 cocoons are randomly chosen. With which probability is at most for one of these cocoons the useful silk thread length less than 750 m?

6. Peter and Paul agree to meet at a restaurant at noon. Peter arrives at a time normally distributed with mean 12:00 and standard deviation 5 minutes. Paul arrives at a time normally distributed with mean 12:02 and standard deviation 2 minutes. Assuming the two arrivals are independent, find the probabilty that

   (a) Peter arrives before Paul

   (b) both men arrive within 3 minutes of noon

   (c) the two men arrive within 3 minutes of each other

7. The weight of a melon, X, in kg is $N(\mu = 1.2, \sigma^2 = 0.3^2)$, i.e. normally distributed with expectation 1.2 kg and standard deviation 0.3 kg. The weight Y for a pineapple is in kg $N(\mu = 0.6, \sigma^2 = 0.2^2)$. We assume that a melon and a pineapple are chosen independently of each other.

   (a) Which distribution has the total weight of the two fruits?

   (b) Calculate the probability that the total weight of the two fruits does not exceed 2.0 kg.

   (c) The melon costs 2 euro per kg and the pineapple 4 euro per kg. Give an expression for the total price Z using X and Y. What is the distribution of Z?

   (d) Calculate the probability that the price Z is higher than 4 euro.

Central Limit Theorem

1. A machine consists of the three modules A, B and C. The machine works only if all three modules are working and if no error occured during the construction phase. The probabilities that the modules A, B and C are defect are 1%, 1% and 5%. The probability for an error during the construction phase 2%. The four kinds of errors occur independently of each other.

(a) Calculate the expectation and the variation of the number of defect machines in a lot of 1000 randomly chosen machines.

(b) The producer is thinking about guaranteeing that not more than 110 machines are defect i such a lot. With which approximate probability can this guarantee promise be kept?

(c) Each defect machine provokes an extra cost of 100 euro. The producer considers to buy a better module C (at a higher price) but with an error rate of is 1%.
How high can the additional cost for each machine for module C be, in order to say that it is (according to the expectation) profitable to buy the more expensive module C?

2. An airline knows that over the long run, 90% of passengers who reserve seats show up for their flight. On a particular flight with 300 seats, the airline accepts 324 reservations.

   (a) Assuming that passengers show up independently of each other, what is the chance that a passenger with a reservation do not get a seat?

   (b) How many reservations can be given, if the airline will accept an overbooking probability of 1%?

3. As a new residential area with 1000 domestic homes is going to be built, the number of required parking lots is calculated in the following way: We assume that there is no relation between the number of cars in different homes. Furthermore, we assume that a domestic home has no car with probability 0.2, one car with probability 0.7 and two cars with probability 0.1. The number of parking lots should be planned in such way that the probability that each car gets a parking lot is 0.99.

   How many parking lots should be built?

4. Starting in the origin, a particle is moving along the integer axes in this way:
   At each time point $1, 2, 3, \ldots$, the particle is moving either one step to the left or one step to the right with the same probability. There is also a third alternative: The particle stays constant, without moving. This alternative has the probability p and $0 < p < 1$. The movement of the particle is independent of earlier movements.

   How should p be chosen, if we want that the probability is 1% that the particle at time point 100 is located to the right of point 15?