

# SMP Exam E-2018

Richard Brooks

**Note:** Please state your answers in the attached notebook (“SMP\_exam”). If you hand in any exercise by hand, please state so in the notebook under the respective assignment (e.g. state “# on paper” in the code block). Please make sure that you hand in in .pdf format (you can save the notebook as .html and then convert/print this to .pdf). Before handing in, change the name of the file to your name. Please make sure that your line of reasoning is apparent either by means of explanation or in terms of code or in terms of both.

## Assignment 1 (15%)

Let  $X$  denote a continuous stochastic variable with the following density function

$$f(x) = \begin{cases} c(1 - x^2) & \text{for } -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- Determine the value of  $c$  and state the cumulative distribution function of  $X$ .
- Determine  $P(X \leq \frac{1}{2})$  and  $P(X \leq -\frac{1}{4})$
- Determine the expected value and the variance of  $X$ .

## Assignment 2 (15%):

In each of the following cases, determine which probability distribution to use and calculate the stated probability:

- The worldwide average number of airplane crashes of commercial airlines is 3.5 crashes per month. Find the probability that there will be at least two crashes within the next month.
- Suppose a batch of 100 items contains 6 that are defective. If you randomly select 10 items, what is the probability that you will select more than two defectives?
- It is known that USB drives produced by a certain company will be defective with probability .01, independent of each other. The company sells the USBs in packages of size 10 and offers a money-back guarantee that at most 1 of the 10 USBs in the package will be defective. If someone buys 3 packages, what is the probability that exactly 1 of the packages will contain a defective? What is the probability that one or more of the packages will contain a defective?

**Assignment 3 (5%):**

It is believed that the number of years a certain laptop functions is exponentially distributed with parameter  $\lambda = \frac{1}{8}$ . If a person buys a used laptop, what is the probability that it will be working after an additional 8 years, i.e. will be working more than 8 years after the person buys it used? Does the exponential distribution seem as a plausible distribution to model durability of laptops? Explain why/why not.

**Assignment 4 (25%)**

A recent study made by the department of education here in Denmark asked students in three different areas what was the most important thing in school: making good grades ('Grades'), having a high learning outcome ('Learning'), or having a good social life ('Social'). Students from rural, suburban, and urban areas were surveyed. A total of 478 students participated in the survey and 149 were from rural areas and of these 50 answered 'Learning'. A total of 247 students answered 'Grades' and of these 103 lived in urban areas. 42 of the 141 who answered 'Learning' lived in a suburban area. 178 of the respondents lived in an urban area. There were 30 more respondents living in a suburban area who answered 'Grades' than there were respondents living in rural areas who answered 'Grades'.

- a. Based on the above information, create a 3 x 3 contingency table and a corresponding joint probability table. Please place area of living on the horizontal axis.
- b. What is the probability that a randomly chosen student was from a suburban area *and* thought having a good social life was most important?
- c. What is the probability that a randomly chosen student was from a suburban area, given the student thought having a good social life was most important?
- d. What is the probability that a randomly chosen student thought having a good social life was most important, given the student was from a suburban area?
- e. Using conditional probability, determine whether 'Rural' and 'Grades' are independent, whether 'Suburban' and 'Grades' are independent, and whether 'Urban' and 'Grades' are independent.
- f. Setup an appropriate hypothesis test to determine whether attitude towards importance in school is independent from area of living.
- g. From the test in (f), state at which level of significance we would have failed to reject and explain the meaning of this value, perhaps even as a conditional probability.

### Assignment 5 (25%)

The data in “Crime Study.xlsx” displays the occurrence of crimes in different states before and after a large U.S. government expenditure to reduce crime.

- a. Setup appropriate tests to test whether the expenditure has had an effect on reducing crimes within each of the three areas (Murder, Rape, and Robbery) as well as on the overall crime rate (the columns with Crime Before/After). State each of the p-values.
- b. For each of the four differences in (a), setup 95% confidence intervals and accompany the intervals with plots that display the rejection region.

### Assignment 6 (15%)

The data in “Salary.xlsx” show the (monthly) salary along with years of experience of 31 software developers.

- a. Create a **complete** regression analysis of the data mentioned above. Your analysis must include a plot of the data, considerations about outliers, estimates for the regression parameters and confidence intervals for these, considerations about the assumptions of the model, as well as an assessment of the adequacy of the model.
- b. According to the model, what salary can a newly graduated software developer with no experience expect?
- c. Assuming the developer starts his/her career at 27 and retires when he/she is 67, what will be the salary of the developer when he/she retires? Does this sound plausible?