EXAM IN: STA500 INTRODUCTION TO PROBABILITY AND STATISTICS 2

DURATION: 4 HOURS DATE: DESEMBER 1, 2014

PERMITTED AIDS: Approved simple calculator (HP30S, Casio FX82, TI-30, Citizen SR-270X, Texas BA II Plus or HP17bII+).

One yellow A4 size sheet with your own handwritten notes.

THE EXAM CONSISTS OF 3 PROBLEMS ON 4 PAGES, 9 PAGES OF ENCLOSURES.

COURSE RESPONSIBLE: Jan Terje Kvaløy PHONE: 51 83 22 55

Problem 1:

The time until failure X (in years) of a certain type of CPUs (central processing units) has a Weibull distribution with parameters $\alpha = 0.25$ and $\beta = 0.5$. (Look up the information you need about the Weibull distribution in the attached collection of tables and formulas.)

a) Show that $F(x) = 1 - e^{-0.25x^{0.5}}$ and that $P(X > 3) = 0.65$.
Find the hazard rate, and comment briefly what this tells us about the ageing properties of these CPUs.
Find $P(X > 5|X > 2)$.

The CPUs are used in certain types of industrial robots, where there is one CPU in each robot. In a factory they have invested in 52 such robots. Let Y be the number of these robots that experience CPU failure within 3 years. The CPUs in different robots fail independently of each others.

b) Which probability distribution does Y have? Explain why.
Find $\mathrm{E}(Y)$ and $\mathrm{Var}(Y)$.
Calculate $P(Y > 20)$.

Another application of the CPUs is in an autonomous underwater vehicle where they use three CPUs of the considered type. All these three CPUs need to work for the vehicle to function, i.e. the time until the vehicle stop working due to CPU problems is $U = \min\{X_1, X_2, X_3\}$, where X_i is the time until failures for CPU i, $i = 1, 2, 3$. We assume that X_1, X_2 and X_3 are independent.

c) Find the cumulative distribution function of U.
Find $P(U > 3)$.
Find $\mathrm{E}(U)$. (Hint: Which type of distribution does U have?)
Also find $\mathrm{E}(X)$. Compare $\mathrm{E}(X)$ and $\mathrm{E}(U)$ and comment briefly.

Problem 2:

In a recent modelling of oil prices since year 2000 the monthly average oil price (of Europe Brent oil) was modelled as a five state Markov chain. The five states of the Markov chain were defined according to five price intervals (in \$) for the oil price as follows: 0: [0-30), 1: [30-60), 2: [60-90), 3: [90-120), 4: [120-∞). I.e. for instance state 2 means an oil price between 60 to 90, while state 4 means an oil price of at least 120.

The following transition probability matrix was found:

$$P = \begin{pmatrix} 0.93 & 0.07 & 0 & 0 & 0 \\ 0.05 & 0.83 & 0.12 & 0 & 0 \\ 0 & 0.10 & 0.85 & 0.05 & 0 \\ 0 & 0 & 0.03 & 0.87 & 0.1 \\ 0 & 0 & 0 & 0.6 & 0.4 \end{pmatrix}$$

The following matrix is also given:

$$P^{12} = \begin{pmatrix} 0.51 & 0.29 & 0.17 & 0.03 & 0.00 \\ 0.21 & 0.33 & 0.32 & 0.12 & 0.02 \\ 0.10 & 0.27 & 0.34 & 0.25 & 0.04 \\ 0.01 & 0.06 & 0.15 & 0.66 & 0.12 \\ 0.01 & 0.05 & 0.14 & 0.68 & 0.12 \end{pmatrix}$$

a) Find $P(X_2 = 2|X_1 = 1)$, $P(X_5 = 3|X_4 = 2, X_3 = 1, X_2 = 1, X_1 = 1)$ and $P(X_4 = 2, X_3 = 1, X_2 = 1|X_1 = 1)$.
In November this year the price was in state 2. What is the probability that the price will not be in state 2 in December this year?

b) When the price is in state 2, what is the expected number of months until the price is in some other state?
For which state is the expected number of months until the process go to some other state highest?
Will the Markov chain have steady probabilities? Explain why/why not.
Explain briefly what we can use the P^{12} matrix for. Find $P(X_{n+12} > 2|X_n = 2)$ and explain what this means in practice.

In another part of the modelling of oil prices it was only looked at whether the price for each month did go up or down compared to the month before. With the states 0: "down" and 1: "up" the following transition matrix was found:

$$P = \begin{pmatrix} 0.38 & 0.62 \\ 0.42 & 0.58 \end{pmatrix}$$

c) First assume that the up/down process is a Markov chain. Find the steady state probabilities and give a practical interpretation of these probabilities in this situation.

In practice whether the price a month is up or down depends on the state the two previous months. Explain how you still can model the price development as a Markov process by defining new states and sketching the transition probability matrix. In the transition probability matrix, mark with crosses the possible transitions and with zeros the impossible.

Problem 3:

The occurrence of major accidents on floating oil-rigs on the Norwegian shelf is examined. Let λ be the expected number of accidents per thousand man hours. The value of λ is unknown. We look at n oil-rigs, and let X_i denote the number of major accidents on oil-rig number i, $i = 1, \ldots, n$.

Assume that X_i has a Poisson distribution with $\mathrm{E}(X_i) = \lambda t_i$, where t_i is how many thousand man hours that are carried out on oil-rig number i.

The numbers of accidents on different oil-rigs are assumed independent. For each of the n oil-rigs, data has been gathered on the number of man hours and the number of major accidents. Based on this information we want to estimate λ. Two estimators are suggested:

$$\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{t_i} \qquad \text{and} \qquad \hat{\lambda}_2 = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} t_i}$$

It can be shown that $\mathrm{E}(\hat{\lambda}_2) = \lambda$ and $\mathrm{Var}(\hat{\lambda}_2) = \frac{\lambda}{\sum_{i=1}^{n} t_i}$

a) Calculate the expectation and the variance of $\hat{\lambda}_1$.
Calculate numerical values of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ using the data from $n = 5$ oil-rigs given in the table below.
Which one of the two estimators would you prefer in this situation? Explain.

Oil-rig i	t_i	x_i
1	150	1
2	272	4
3	87	0
4	328	2
5	63	1

It is given that $\sum_{i=1}^{5} t_i = 900$, $\sum_{i=1}^{5} t_i^2 = 215606$, $\sum_{i=1}^{5} \frac{1}{t_i} = 0.041$, $\sum_{i=1}^{5} \frac{1}{t_i^2} = 0.00045$ and $\sum_{i=1}^{5} \frac{x_i}{t_i} = 0.043$.

After some time data from another 9 oil-rigs also became available. Summarised for all the $n = 14$ oil-rigs, it was observed $\sum_{i=1}^{14} x_i = 25$ major accidents during a total time (in thousand man hours) of $\sum_{i=1}^{14} t_i = 3200$. These are the data we shall use in the rest of the problem for estimating λ.

b) Show that the likelihood can be written

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} t_i^{x_i}}{\prod_{i=1}^{n} x_i!} e^{-\lambda \sum_{i=1}^{n} t_i}$$

One of the estimators $\hat{\lambda}_1$ or $\hat{\lambda}_2$ is the maximum likelihood estimator (MLE), find out which one by deriving the maximum likelihood estimator for λ.
Find the 90% Wald confidence interval for λ. Can the Wald interval be trusted in this case? Comment briefly.

Experts on risk analysis have information about λ based on other sources than the data (e.g. experience from other areas and knowledge on differences and similarities between these areas and the Norwegian shelf). This knowledge is expressed by a gamma prior distribution with parameters a and b, i.e.

$$p(\lambda) = \frac{1}{b^a \Gamma(a)} \lambda^{a-1} e^{-\lambda/b}$$

For this distribution the expectation is ab and the variance is ab^2.

c) Find the posterior distribution for λ.
Find an expression for the Bayes estimate for λ.
Calculate the estimate based on the data from 14 oil-rigs given above and a gamma prior distribution with expectation 0.01 and variance 0.00001.
Compare the Bayes estimate with the maximum likelihood estimate and the expectation in the prior distribution and comment what this tells us.