NTNU
Norwegian University of Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4240 Statistics**

Academic contact during examination: Sara Martino[a], Torstein Fjeldstad[b]

Phone: [a]994 03 330 , [b]962 09 710

Examination date: November 28th 2018

Examination time (from–to): 09:00–13:00

Permitted examination support material: C
- Tabeller og formler i statistikk, Akademika,
- A yellow stamped A5-sheet with personal hand written notes.
- A specific basic calculator

Other information:
All your answers should be justified.

The hand-in material should contain calculations leading to your answer.

There are 4 tasks with a total of 10 subtasks which have equal weights in the grading.

Language: English

Number of pages: 5

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1 Let X be a stochastic variable with probability density function:

$$f_X(x) = \begin{cases} \frac{1+x}{2} & x \in (-1, 1) \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X^2$.

a) Find the cumulative distribution function of Y.

Find the probability density function of Y.

Find $E(2Y - Y^2)$.

Problem 2 Assume that, for a particular stretch of road, the cars traveling into town past a particular point follow a Poisson process:

$$X(t) = \{\text{number of cars passing during } t \text{ minutes}\},$$

with parameter λ, that is $X(t) \sim \text{Poisson}(\lambda t)$.

a) Describe the properties of a Poisson process.

Provide an interpretation of the parameter λ for the situation described above.

b) Assume, only for this point, that $\lambda = 1.5$.

What is the probability that exactly 2 cars pass during a 1-minute period?

What is the probability that at least 2 cars pass during a 2-minutes period?

Consider now 10 non-overlapping 1-minute periods. What is the probability that at least for one of the periods there are more than 5 cars passing?

Assume now that the parameter λ is unknown. To estimate λ an inspector from the municipality visits that particular stretch of road n times to count how many cars pass. Each time he stands along the street for t_i minutes and counts X_i cars, $i = 1, \ldots, n$. Assume also that the stochastic variables X_i, $i = 1, \ldots, n$ are mutually independent.

It is given that the maximum likelihood estimator for λ is

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} t_i}$$

The municipality wants to reduce the car traffic in the city center and decides that if there are, on average, more than 1.5 cars per minute traveling into town then it is necessary to install a toll house. The municipality wants to test the following hypotheses

$$H_0 : \lambda = \lambda_0 = 1.5 \qquad \text{against} \qquad H_1 : \lambda > 1.5$$

The inspector visits therefore the stretch of road $n = 10$ times and each time he counts the cars for 10 minutes. We call each 10 minutes visit a "trial". The result of the 10 trials is the following:

Table 1: Number of observed cars in each trial.

t_i	10	10	10	10	10	10	10	10	10	10
x_i	17	14	25	22	18	17	19	27	14	19

where $\sum_{i=1}^{10} x_i = 192$ and $\sum_{i=1}^{10} t_i = 100$.

c) Starting from the MLE estimator $\hat{\lambda}$, formulate a reasonable test statistic and specify its distribution under the null hypotheses. Justify your answer. (Hint: you can use, without proof, that a Poisson distribution with parameter λ can be approximated by a normal distribution if $\lambda > 14$.)

Perform a hypotheses test using a 1 % level of significance. What is the decision of the municipality regarding the construction of the toll house?

An alternative way to test if there are, on average, more than 1.5 cars per minute is to define

$$Z = \{\text{number of trials where the inspector counts more than } \lambda_0 t \text{ cars}\}$$

where λ_0 is the value of car per minutes one wants to test and $t = 10$ is the duration of each trial, and reject if $Z \geq k$ where k is a constant to be defined.

d) Based on Z, determine the smallest value of k such that the significance level of the test is $\alpha \leq 0.01$

Perform the test based on Z with the data provided in Table 1 and determine the municipality's decision.

Problem 3 Oscar is an active athlete who aims at performing at his best. To improve his level of performance he decides to consume an illegal doping substance. Let x be the amount of doping substance that Oscar takes. Assume that it is possible to perfectly control the dose so that x is not a stochastic variable.

One week after taking the drug Oscar checks the concentration of the substance in his blood. Let Y be the measured concentration. Assume the following linear relationship between the amount x that Oscar takes and the concentration level Y that is measured after one week

$$Y = \beta x + \epsilon \tag{1}$$

where ϵ is normally distributed with mean 0 and known variance $\sigma^2 = 4^2$ and β is an unknown constant that describes the relationship between the dose Oscar take and the concentration of the substance in his blood one week after.

a) Assume, only for this point, that $\beta = 0.5$.

In a given day, Oscar takes a dose of $x = 30$. The observed concentration one week after, Y, is then normally distributed with mean $0.5 \cdot 30$ and variance 4^2.

Find the probability that the observed concentration of the doping substance one week after the given day is higher than 20, that is find $P(Y > 20)$.

Find the probability $P(Y < 10 \cup Y > 20)$.

Given that, one week after the given day, the concentration of the doping substance in the blood is higher than 10, find the probability that it is higher than 20, that is find $P(Y > 20 | Y > 10)$.

Tuva uses the same illegal doping substance as Oscar. Together they want to investigate the relationship between the taken dose and the concentration of the substance on the blood after one week, that is they want to estimate β.

It is known that, because of some genetic variation, Tuva has a reference concentration level c_0 of the doping substance in her blood and that such reference concentration is equal to 0 for Oscar. Assume that Tuva and Oscar take an identical dose x of the doping substance. The concentration, Z, that Tuva has in her blood after one week is then:

$$Z = c_0 + \beta x + \tau \tag{2}$$

where τ is normally distributed with mean 0 and variance $\sigma^2 = 4^2$. Assume that c_0 is a known constant that describes the reference concentration level of the doping substance in Tuva's blood and β is the same as in equation (1).

For n times, both take the exact same amount of doping substance x_i, $i = 1, \dots, n$ and one week after measure the corresponding concentrations Y_i and Z_i, $i = 1, \dots, n$. For simplicity assume that they take a new dose only when the previous one has disappeared from their body. That is, Oscar has the random sample $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$ from model (1) and Tuva has the random sample $(x_1, Z_1), (x_2, Z_2), \dots, (x_n, Z_n)$ from model (2). Assume moreover that the two random samples are independent of each other.

b) Show that the maximum likelihood estimator for β based on the random sample $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n), (x_1, Z_1), (x_2, Z_2), \dots, (x_n, Z_n)$ is

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i(Y_i + Z_i) - c_0 \sum_{i=1}^n x_i}{2\sum_{i=1}^n x_i^2}$$

Show that $\hat{\beta}$ is an unbiased estimator of β with variance

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{2\sum_{i=1}^n x_i^2} = \frac{4^2}{2\sum_{i=1}^n x_i^2}$$

c) Derive a $(1 - \alpha) \cdot 100$ % confidence interval for β.

It is given that $n = 10$, $c_0 = 5$, $\sum_{i=1}^{10} x_i = 982$, $\sum_{i=1}^{10} x_i^2 = 97324$, $\sum_{i=1}^{10} x_i y_i = 68586$ and $\sum_{i=1}^{10} x_i z_i = 72398$.

Find a numerical expression for a 90 % confidence interval for β.

We want to perform the following hypotheses test:

$$H_0 : \beta = 0.5 \qquad \text{against} \qquad H_1 : \beta \neq 0.5$$

Would you reject the null hypotheses based on the given observations with a significance level $\alpha = 0.1$? Justify your answer.

Problem 4

Eva works in a fish farm producing salmons. Her job is to check the quality of the fishes that are placed on the marked. The fish farm assumes that the weight of a random salmon X (in kilogram) is normally distributed with unknown mean μ kilogram and unknown standard deviation σ kilogram.

a) Assume, for this point that Eva has captured 10 salmons with weights $X_1, X_2, \ldots, X_{10}$. Assume that the 10 salmons are independent of each other.

Use the random sample $X_1, X_2, \ldots, X_{10}$ to give an expression for a 95 % prediction interval for the weight in kilogram for a new salmon, X_0. Assume that X_0 is independent from $X_1, X_2, \ldots, X_{10}$.

Use that $\sum_{i=1}^{10} x_i = 53.37$ kilogram and $\sqrt{\frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2} = 0.73$ kilogram to find a numerical expression for the prediction interval.

Previously the fish farm has assumed the expected weight to be 5 kilogram, but based on the feedbacks of the shops selling the salmon produced in Eva's fish farm, she suspects that the expected weight is higher than 5 kilogram. Therefore, Eva wants to test the hypotheses

$$H_0 : \mu = 5 \text{ kilogram} \qquad \text{against} \qquad H_1 : \mu > 5 \text{ kilogram}$$

with a significance level $\alpha = 0.05$ based on a random sample $X_1, X_2, \ldots, X_n$.

b) Assume, in this point, that the standard deviation of X is known and equal to 1 kilogram.

Assume that the real expected weight is 5.5 kilogram. Derive an expression for the minimum number of salmon n Eva has to weigh if she requires that the power of the test should be at least 95 % when the true expected value is 5.5 kilogram.

Find a numerical value for the smallest number n which satisfies the requirements.