Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Contacts during exam:
Mette Langaas 98847649
Eirik Mo 73593541/41106633

EXAM IN COURSE TMA4245 STATISTICS

Saturday June 5th 2004

Time: 09:00–14:00

Permitted aids:
Yellow A5 sheet with your own handwritten notes.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Calculator: HP30S.

ENGLISH

The examination results are due June 28th 2004.

Problem 1 Atle, you are lying!

Participants in programmes on TV are often asked personal questions. We will divide the types of questions asked into three categories, and define the following three disjoint events:

A_1="a question of non-sensitive nature is asked, e.g. what is your name?"
A_2="a question of partly sensitive nature is asked, e.g. how old are you?"
A_3="a question of sensitive nature is asked, e.g. have you been unfaithful to your partner?".

In addition we define the event:
L="the participant is lying."

The following probabilities are given:
$P(A_1) = 0.1$, $P(A_2) = 0.4$, $P(A_3) = 0.5$, $P(L|A_1) = 0.05$, $P(L|A_2) = 0.2$, $P(L|A_3) = 0.6$.

a) Draw the four events in a Venn diagram.

Given that a participant is asked a question of type A_2, what is the probability that the participant is not lying, $P(L'|A_2)$?

What is the probability that a randomly selected participant is lying, $P(L)$?

One of the questions that is of a partly sensitive nature is "how old are you?". A total of n individuals were asked this question and the answers were recorded and compared to information in public registers. Let X be a random variable denoting the number of individuals lying among the n individuals asked, and let p be the probability that a randomly selected person is lying.

b) Which conditions must be fulfilled in order for X to have a binomial distribution?

Assume that $p = 0.2$, and that we ask $n = 20$ individuals. What is then $P(X = 4)$? What is $P[(X \leq 2) \cup (X > 5)]$?

We now assume that p is unknown. The following two estimators for p have been suggested,

$$\hat{p} = \frac{X}{n} \quad \text{and} \quad p^* = \frac{X}{n-1}.$$

c) Find the expected value and the variance of each of the two estimators $\hat{p}$ and p^*.

Which two properties describe a good estimator? Which of the two estimators, $\hat{p}$ and p^*, will you prefer to use. Justify you answer.

We assume that n is large and choose to use the estimator $\hat{p}$. Then,

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately standard normally distributed.

Based on previous experience we assume that $p = 0.2$. We would like to investigate if the collected data give us reason to believe that p is larger than 0.2.

d) Formulate a hypothesis test by defining a null hypothesis and an alternative hypothesis.

Use the fact that Z approximately follows a standard normal distribution to find a rejection region for the null hypothesis when we use $\alpha = 0.01$ as our significance level.

What is the conclusion to the hypothesis test when we have asked $n = 200$ individuals and $x = 55$ of them lied?

Will the p-value (significance probability) for the hypothesis test be less than or larger than 0.01? Justify you answer. (You do not need to calculate the precise value of the p-value.)

Problem 2 The stock exchange

The company Agderfrukt is listed on the stock exchange. We assume that the change X in the value of one share of Agderfrukt during one day is normally distributed with mean $\mu_X = 0.15$ kroner and standard deviation $\sigma_X = 0.60$ kroner. If you have one share of Agderfrukt, then $X > 0$ will give you profit, while $X < 0$ is loss.

a) What is the probability of loosing money on Agderfrukt during one day, i.e. $P(X < 0)$?

What is $P(0 \leq X \leq 0.15)$?

If you buy 10 shares of Agderfrukt today and sell the shares tomorrow, what is then the expected profit of your trade? Also find the variance of the profit.

The company Trønderfrukt is also listed on the stock exchange. We denote the change in one share of Trønderfrukt during one day by Y, where Y is normally distributed with mean $\mu_Y = 0.15$ kroner and standard deviation $\sigma_Y = 0.80$ kroner.

b) We examine the change in the value of shares for Agderfrukt and Trønderfrukt on the same day, and we assume at this point that the changes in the two shares, X and Y, are independent.

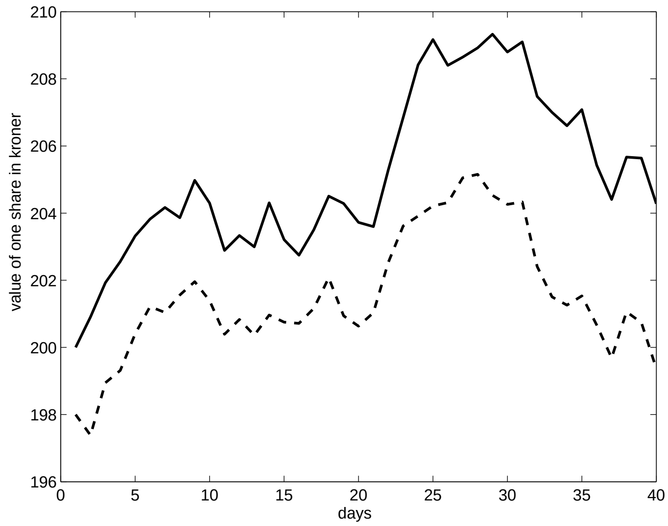
Today the value of one share of Agderfrukt is the same as the value of one share of Trønderfrukt. We would like to investigate the following three possible strategies for shares purchase, where we buy today and sell tomorrow.

i) Buy two shares of Agderfrukt.

ii) Buy one share of Agderfrukt and one share of Trønderfrukt.

iii) Buy two shares of Trønderfrukt.

If you want to minimize the risk in your investment, which of the three strategies will you choose? Justify your answer.



In the figure we see the development of the exchange rate for one share of Agderfrukt (dashed) together with one share of Trønderfrukt (solid).

The change in the value of one share of Agderfrukt on day i is denoted X_i, and we assume that X_i is normally distributed with mean $\mu_X = 0.15$ kroner and standard deviation $\sigma_X = 0.60$ kroner.

The change in the value of one share of Trønderfrukt on day i is denoted Y_i, and we assume that Y_i is normally distributed with mean $\mu_Y = 0.15$ kroner and standard deviation $\sigma_Y = 0.80$ kroner.

We assume that the changes in values of shares on different days are independent.

We compare the two companies by calculating the differences between the daily changes in the shares for the two companies, $D_i = X_i - Y_i$, and then taking the average. We look at this difference over 10 days, $\overline{D} = \frac{1}{10}\sum_{i=1}^{10} D_i = \frac{1}{10}\sum_{i=1}^{10}(X_i - Y_i)$.

c) Does the figure give you reason to believe that the changes in the shares of Agderfrukt, X_i, and Trønderfrukt, Y_i, on the same day are mutually independent?

The correlation between X_i and Y_i, $\rho(X_i, Y_i)$, is either -0.5, 0.0 or 0.5. Which of these values are most reasonable when you look at the figure? Give a short explanation of your choice.

What is the expected value and the variance of $\overline{D}$? Use the value for the correlation, $\rho(X_i, Y_i)$, that you have chosen above.

Problem 3 Wave heights

In design of offshore structures, e.g. when selecting the deck elevation for a fixed offshore platform, it is important to know the wave conditions in the area where the structure will be located. A wave height gauge is therefore placed in the area.

Let X denote the largest wave height measured on a randomly chosen day. We assume that the probability density function of X is given as

$$f(x;\theta) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, \qquad x \geq 0, \qquad \theta > 0.$$

In this problem you can use without proof that $\mathrm{E}[X^2] = \theta$ and $E[X^4] = 2\theta^2$.

a) Show that the cumulative distribution of X, $F(x) = P(X \leq x)$, is $F(x) = 1 - e^{-\frac{x^2}{\theta}}$. (Use substitution with $u = x^2$.)

Given that the largest wave height is larger than 10 meter, find the probability that the largest wave height is larger than 15 meter when $\theta = 25$, i.e. $P(X > 15 | X > 10)$?

In the rest of this problem we assume that θ is unknown.

We have observed the largest wave height for n days. Let X_i be a random variable denoting the largest wave height on day i. We assume that $X_1, ..., X_n$ are mutually independent and identically distributed with probability density function $f(x;\theta)$.

b) Derive the maximum likelihood estimator (MLE) $\hat{\theta}$ of θ.

Is the estimator $\hat{\theta}$ unbiased?

Find the variance of $\hat{\theta}$.

c) Use the central limit theorem to explain why

$$Z = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \theta}{\sqrt{\frac{\theta^2}{n}}}$$

is approximately standard normally distributed.

Use Z to find an approximate 95% confidence interval for θ.

The probability that the largest wave height a randomly chosen day is larger than 10 meter is $P(X > 10) = e^{-\frac{100}{\theta}}$. Use the approximate confidence interval that you found to find an approximate 95% confidence interval for $e^{-\frac{100}{\theta}}$.