Norwegian University of Science and Technology
Department of Mathematical Sciences

Contacts during the exam:
Henning Omre (909 37848)
Mette Langaas (988 47649)

EXAM IN TMA4240 Statistics

December 18, 2010
Time: 9:00–13:00

Number of credits: 7.5

Permitted aids:

- Calculator HP30S or Citizen SR-270X with empty memory.
- *Statistiske tabeller og formler*, Tapir forlag.
- K. Rottman: *Matematisk formelsamling* or *Matematische Formelsammlung*.
- One yellow stamped A5 sheet with own handwritten formulas and notes.

Grading finished: January 18, 2011.

Exam results are annonunced at `http://studweb.ntnu.no/`.

ENGLISH

Problem 1 Hay fever and eczema

In a population of 11 year old children the prevalence of hay fever and eczema has been studied.

Define two events:

- E: a randomly chosen child from the population has eczema.
- H: a randomly chosen child from the population has hay fever.

Assume that in this population we have the following probabilities:

$$\begin{aligned} P(E) &= 0.04 \\ P(H) &= 0.07 \\ P(E \cap H) &= 0.009 \end{aligned}$$

a) Draw a Venn diagram depicturing the two events.
Are the events E and H independent? Justify your answer.
Among the children that do not have eczema in the population, we randomly select one child. What is the probability that this child does not have hay fever?

Problem 2 Newspaper sales

We study the daily sale of the local newspaper at a kiosk, and assume that an unlimited number of copies are available for sale, so that the newspaper will not be sold out.

Let X be the number of copies of the newspaper that is sold during a randomly chosen day, and assume that X follows a Poisson distribution with expected value $\mathrm{E}(X) = \mu$.

a) Assume now that the expected sale is $\mu = 10$ copies, and that we study the sale at a randomly chosen day.

What is the probability that exactly 10 copies are sold on this day?
What is the probability that 13 or more copies are sold on this day?

From previous years we know that the expected number of newspapers sold is approximately the same for every weekday during the autumn season.

Let $X_1, X_2, \ldots, X_n$ be independent Poisson distributed random variables denoting the number of newspaper copies sold at n randomly chosen weekdays in the autumn season. Assume that the expected sale is μ every weekday, but that this parameter is unknown. An estimator for μ is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

b) Find the expected value and the variance of $\bar{X}$.
Use the central limit theorem to construct an approximate 95% confidence interval for μ.
Calculate the confidence interval numerically when $n = 30$ and $\frac{1}{30}\sum_{i=1}^{30} x_i = 10.75$.

The owner of the kiosk needs to decide the number of newspapers to buy from the publisher on a chosen day. Assume that the expected number of newspapers, μ, that can be sold at the kiosk that day is known. Then there are two competing interests to take into consideration:

- If the newspaper is sold out the kiosk owner will lose potential profit.
- If there are unsold copies of the newspaper in the end of the day the kiosk owner loses money because unsold copies returned to the publisher will not be fully compensated.

Let, as previously, X denote the number of copies sold when there is an unlimited number of newspapers available. Assume that $Y = \min(X, a)$ is the sale when the kiosk has a copies of the newspaper available for sale. We assume, as previously, that X follows a Poisson distribution, and that the expected value is $\mu = 10$.

c) Find the probability distribution function of Y, $P(Y = y)$ for $y = 0, 1, \ldots, a$.
Show that $E(Y) = a - \sum_{y=0}^{a-1}(a - y)P(X = y)$.

We may choose a from considerations of total profit for the kiosk. The owner of the kiosk pays 5 kr for each copy of the newspaper and sells each copy for 12 kr. The copies that are not sold can be returned to the publisher, and the owner of the kiosk then gets 3 kr from the publisher for each copy.

Show that the totalt profit is $9Y - 2a$.
How many copies, a, of the newspaper should the owner of the kiosk buy from the publisher so that the *expected* total profit is maximized?

Hint: Call the expected totalt profit $h(a)$. Investigate for which values of a the difference $h(a) - h(a - 1)$ is positive and negative.

Problem 3 Covariance

We consider two random variables X and Y. Let X have expected value $\mathrm{E}(X) = 10$ and variance $\mathrm{Var}(X) = 4$, and Y have expected value $\mathrm{E}(Y) = 8$ and variance $\mathrm{Var}(Y) = 9$. Further, the covariance between X and Y is $\mathrm{Cov}(X, Y) = 5$.

a) Calculate numerical values for the following expressions:
$\mathrm{E}(2X - Y)$
$\mathrm{Var}(2X - Y)$
$\mathrm{E}\left((X - 3)(Y - 5)\right)$

Problem 4 The experimental farm

Experiments on production of biomass are performed at an experimental farm. Assume that the biomass, Y, from a plant is normally (Gaussian) distributed with expected value $\mathrm{E}(Y) = 5$ and variance $\mathrm{Var}(Y) = 4$.

a) Calculate the following three probabilities:
$P(Y > 6)$
$P(4 < Y \leq 6)$
$P(Y > 6 \mid Y > 4)$

Assume that the biomass Y made from a specific plant is dependent on the cultivation period x of the plant. The cultivation period of a plant is defined as the time from the plant is observed above the soil until the time of the biomass measurement.

Further, assume that the dependence between the biomass Y and a given cultivation period x, can be modelled as a linear regression, without intercept and with an error term that is dependent of the cultivation period,

$$Y = \beta x + \varepsilon(x) \text{ for } x > 0,$$

where $\varepsilon(x)$ is normally (Gaussian) distributed with expected value $\mathrm{E}(\varepsilon) = 0$ and variance $\mathrm{Var}(\varepsilon) = \tau^2 x^2$. This means that the standard deviation of the error term is proportional to the cultivation period x. The biomass measurement is only performed for cultivation periods of positive length, $x > 0$.

The model parameters β and τ are assumed to be unknown.

An experiment with $n = 5$ independent measurements at cultivation periods $x_1, x_2, \ldots, x_5$ and corresponding biomasses $Y_1, Y_2, \ldots, Y_5$ gave the following observations:

i	1	2	3	4	5
x_i	3	6	7	10	14
y_i	1.0	5.0	3.0	3.0	10.0

The following sums are given: $\sum_{i=1}^{5} \frac{y_i}{x_i} = 2.61$ and $\sum_{i=1}^{5} \frac{y_i^2}{x_i^2} = 1.59$.

b) What is the probability distribution of Y_i given x_i?
Derive expressions for estimators of β and τ^2, for example by applying the maximum likelihood method.
Use the numerical values from the above table to calculate estimates for β and τ^2.

The following estimator for β is to be used in the rest of the problem:

$$B = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{x_i}$$

c) Assume now that $\tau^2 = 0.04$ is known.

We would like to test the hypothesis

$$H_0\colon \beta = 0.50 \text{ versus } H_1\colon \beta > 0.50$$

Derive a rejection rule at significance level 0.05.
Use the numerical values from the above table to conduct the test.
Derive an expression for the power of the test when $\beta = 0.7$, and give a numerical answer when $n = 5$.