Norwegian University of Science and Technology
Department of Mathematical Sciences

English

Contact during exam:
Håkon Tjelmeland 73 59 35 38

EXAM IN TMA4240 STATISTICS

Wednesday December 2nd 2009
Time: 09:00–13:00

Aids: Calculator HP30S or Citizen SR-270X with empty memory.
Statistiske tabeller og formler, Tapir forlag.
K. Rottman: Matematisk formelsamling.
One yellow paper (A5 with stamp) with own hand written formulas and notes

Grading finished: December 23rd 2009.

Problem 1

Let X be a continuously distributed random variable with probability density

$$f(x) = \begin{cases} k\,(4x+1) & \text{for } x \in [0,1], \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

a) Find what value the constant k must have. Sketch $f(x)$.
Find the cumulative distribution function $F(x) = \mathrm{P}(X \leq x)$.
Find the probabilities

$$\mathrm{P}\left(X > \frac{1}{4}\right) \quad \text{and} \quad \mathrm{P}\left(X > \frac{1}{2} \middle| X > \frac{1}{4}\right).$$

Let also Y be a continuously distributed random variable, and let the conditional density of Y given $X = x$ be given as

$$f(y|x) = \begin{cases} \frac{4x+2y}{4x+1} & \text{for } y \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

b) Find the joint density function of X and Y.

Find the marginal density function of Y.

Find the probability

$$\mathrm{P}\left(Y \leq X\right).$$

Problem 2 Telephone problems

A group of n students one day discusses problems they have experienced when trying to contact a certain firm by telephone. The firm has no telephone queue system, so when someone is calling the firm he or she either gets in contact with the firm or gets a busy signal. Several of the students have experienced that they apparently always bet the busy signal. The group of students decides to do some experiments to better understand the situation. Each of the students will several times call the firm and each time register whether he or she gets in contact with the firm or gets the busy signal. Each of the students will continue to call the firm until he or she has got in contact with the firm k times. Then each student will report to the group the total number of times he or she has telephoned the firm.

Number the students from 1 to n and let X_i denote the number of telephone calls student number i reported to the group. Let p denote the probability of getting in contact with the firm if doing one try.

The students, who are not very trained in statistics, quickly conclude that $X_1, X_2, \ldots, X_n$ is a random sample from a negative binomial distribution with parameters k and p, i.e.

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \;\; x = k, k+1, \ldots.$$

a) What premises, in addition to the assumptions specified above, must be true for the conclusion of the students about $X_1, X_2, \ldots, X_n$ being a random sample from a negative binomial distribution to be correct?

Now assume these premises to be true, and assume $k = 2$ and $p = 0.1$. Then find the probabilities

$$\mathrm{P}(X_1 = 2) \text{ and } \mathrm{P}(X_1 \leq 4 | X_1 > 2).$$

In the rest of this problem we assume that the premises you specified in item **a**) is fulfilled, so that the conclusion of the students is correct. Moreover we assume k to be a known number, whereas p is unknown and should be estimated.

b) Derive the maximum likelihood estimator (MLE) for p based on the data $X_1, X_2, \ldots, X_n$ and show that it can be written as

$$\widehat{p} = \frac{nk}{\sum_{i=1}^{n} X_i}.$$

Some time later the students see a TV program where a representative of the firm is interviewed about exactly the situation discussed by the students. The representative admitted that it might be somewhat difficult to get in contact with the firm by telephone, but claimed that definitely $p \geq 0.1$. The students decide to test whether their data gives reason to conclude that the claim of the firm is wrong.

c) Formulate the problem of the students as a hypothesis testing problem. Thus, choose the null hypothesis and the alternative, choose a test statistic and decide a rejection criterion.

Find the (approximate) p-value for the hypothesis test when $k = 2$, $n = 50$ and the observations of the students gave $\sum_{i=1}^{n} x_i = 779$. What will you conclude about the claim of the firm? (Give reason for your answer)

Problem 3 Melting point

A metallurgist has participated in the development of a new alloy and wants to present various properties of the alloy to his colleagues. Here we will consider the melting point of the alloy.

To measure the melting point of the alloy the metallurgist has two methods. Here we call these methods for method A and method B, respectively. Because of measuring error repeated measurements of the melting point by method A can be considered as independent realisations of a normally distributed random variable with mean value μ and variance σ_A^2. The metallurgist wants to estimate the mean value μ. Correspondingly, repeated measurements of the melting point by method B are independent realisations of a normally distributed random variable with mean μ and variance σ_B^2.

In this problem you should consider the variances σ_A^2 and σ_B^2 to be known values, whereas the common mean value μ is unknown and should be estimated. For this the metallurgist do n measurements with method A and m measurements with method B. Let $X_1, X_2, \ldots, X_n$ denote the results of the measurements with method A and let $Y_1, Y_2, \ldots, Y_m$ denote the results of the measurements with method B. We will also assume that $X_1, X_2, \ldots, X_n$ are independent of $Y_1, Y_2, \ldots, Y_m$.

As an estimator for μ the metallurgist uses

$$\widehat{\mu} = a\bar{X} + b\bar{Y} = \frac{a}{n}\sum_{i=1}^{n} X_i + \frac{b}{m}\sum_{i=1}^{m} Y_i,$$

where a and b are two constants.

a) Use computational rules for the mean value and the variance to show that

$$\mathrm{E}\left(\widehat{\mu}\right) = (a+b)\mu \quad \text{and} \quad \mathrm{Var}\left(\widehat{\mu}\right) = \frac{a^2\sigma_A^2}{n} + \frac{b^2\sigma_B^2}{m}.$$

b) Find values for the constants a and b so that $\widehat{\mu}$ becomes a best possible estimator for μ.

In the rest of this problem you should still use $\widehat{\mu} = a\bar{X} + b\bar{Y}$, i.e. you should **not** insert the optimal values for a and b found in item **b**).

c) What type of probability distribution has $\widehat{\mu}$? Give reason for you answer.

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

d) Find how the constants a and b should be chosen to make the confidence interval you found in item **c**) as short as possible. Compare your results with what you found in item **b**) and make a comment.