NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4245 Statistikk**

Academic contact during examination: Arvid Næss[a], Torstein Fjeldstad[b]

Phone: [a] 99 53 83 50 , [b] 96 20 97 10

Examination date: 10. August 2017

Examination time (from–to): 09.00-13.00

Permitted examination support material:
Tabeller og formler i statistikk, Akademika,
K. Rottmann: *Matematisk formelsamling*,
Calculator Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College eller HP30S,
Yellow stamped A5-sheet with personal hand written notes.

Other information:
All your answers should be justified and the hand-in material should contain calculations leading to your answer.

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1 Probability

Assume that the stochastic variable X has the probability density

$$f(x;\theta) = c\exp\{-(x-\theta)\}\,,\;\; x \geq \theta,$$

while $f(x;\theta) = 0$ for $x < \theta$. θ denotes a constant parameter and c is a positive constant.

a) Determine the constant c, and find the probability that $X > \theta + 1$.

b) Derive an expression for the maximum likelihood estimator $\hat{\Theta}$ for the parameter θ based on n independent observations $x_1, \ldots, x_n$, av X. Hint: Make a sketch of the likelihood function.

c) Let $X_1, X_2, \ldots, X_{10}$ denote 10 independent stochastic variables which have the same probability density function as X. Determine the probability density function of the stochastic variable $W = \min\{X_1, X_2, \ldots, X_{10}\}$. Find the probability that $W > \theta + 1$.

Problem 2 Fatigue

The brackets for the rotor blades on a helikopter are subjected to fracture due to material fatigue. To improve fatigue resilience one may carry out a surface hardening procedure of the brackets by using one of the two methods A and B described below.

Method A, also called ”shot-peening”, consists of subjecting the material to a kind of sandblasting with fairly coarse grains.

Method B (nitration) consists of heating the surface to approximately 500 degrees Celsius and adding ammonia. This results in a thin, durable surface layer.

One is now interested in finding out if method A provides better fatigue resistance than method B. At the factory, 13 brackets are chosen randomly for a test. The first 7 brackets are treated with method A, the remaining 6 are treated with method B. The 13 brackets are mounted in a fatigue testing rig and are subjeted to cyclic loading with tension range 700 MPa (MegaPascal). Experience has shown that the natural logarithm of the number of cycles until fatigue failure is approximately a normally distributed stochastic variable.

The natural logarithm of the number of cycles to fatigue failure for the brackets treated by method A, are $X_1, \ldots, X_7$. For the brackets treated by method B, the corresponding variables are $Y_1, \ldots, Y_6$. All the 13 observations are assumed to be independent, normally distributed stochastic variables with unknown expectation values and variances:

$$E(X_i) = \mu_A\,,\ \mathrm{Var}(X_i) = \sigma_A^2\,,\ i = 1, \ldots, 7.$$
$$E(Y_j) = \mu_B\,,\ \mathrm{Var}(Y_j) = \sigma_B^2\,,\ j = 1, \ldots, 6.$$

Assume in the next two points that $\mu_A = \mu_B = \mu$ and $\sigma_A = \sigma_B = \sigma$, but with unknown common values.

a) Show that

$$\hat{\mu} = \frac{1}{2}\left(\bar{X} + \bar{Y}\right)$$

is an unbiased estimator for μ, and find its variance expressed in terms of σ. Here $\bar{X} = \sum_{i=1}^{7} X_i/7$ and $\bar{Y} = \sum_{j=1}^{6} Y_j/6$.

b) Propose an unbiased estimator μ^* for μ which has smaller variance than $\hat{\mu}$. Show this by calculation.

For the rest of Problem 2 it is no longer assumed that $\mu_A = \mu_B$ and $\sigma_A = \sigma_B$. The following formula is taken directly from the textbook:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

It is used for calculating approximately the number of degrees of freedom ν for estimators that are commonly used for estimation and hypohesis testing as in the rest of this problem.

c) Let $\delta = \mu_A - \mu_B$. Derive a $100(1 - \alpha)\%$ confidence interval for δ. What is this interval numerically for $\alpha = 0.1$ with the empirical results given at the end of this problem?

d) Do the results from the test indicate that method A gives better fatigue resistance than method B? Formulate this as a hypothesis testing problem and justify the choice of null hypothesis and alternative hypothesis. Choose significance level 0.05 and carry out the test with the empirical results given at the end of this problem.

$\bar{x} = 15.22$

$\sum_{i=1}^{7}(x_i - \bar{x})^2 = 0.32$

$\bar{y} = 14.56$

$\sum_{j=1}^{6}(y_j - \bar{y})^2 = 0.47$

$\nu \approx 9$

Problem 3 Quality control

During the production of a type of medicine in powder form, which are sold in sealed canisters packed in cardboard boxes containing k canisters in each box, it is important to make sure that the powder is not contaminated by fungal spores. This is controlled by a laboratory test. The probability that a randomly chosen canister gives a positive test result (i.e. the canister is contaminated), is assumed to be equal to p, and the test results for different canisters are assumed to be independent.

The testing procedure is as follows: Tests are carried out with the k (> 1) canisters in a box at a time. These canisters are referred to as a k-group. Half of the content in each canister of the k-group are mixed together and the mixture is analyzed. If the mixture gives a positive test result, which means that at least one of the canisters was contaminated, new tests are carried out individually with the remaining powder in the k canisters. On the other hand, if the mixture gives a negative test result, none of the canisters in the k-group are assumed contaminated. If there are fungal spores in a canister, they may be assumed to be approximately evenly distributed in the canister.

a) Justify why the number of contaminated canisters in a k-group are binomially distributed, and show that the probability that the mixture of the content in k canisters gives a positive test result, is

$$1 - (1 - p)^k.$$

b) How is the conditional probability of an event A given another event B defined? If a canister is in a k-group that has given a positive test result, what is the probability that the canister is contaminated?

c) Assume that a quality control is carried out by selecting m cardboard boxes with k canisters in each box for testing. Let X denote the number of tests that have to be analyzed before all the mk canisters have been tested for contamination. Show that

$$E(X) = m + mk\Big(1 - (1-p)^k\Big).$$

If $k = 4$, for which values of p is the adopted test procedure to be preferred instead of testing the canisters individually at once?