NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4245 Statistikk**

Academic contact during examination: Håkon Tjelmeland, Arvid Næss

Phone: 48221896, 99538350

Examination date: 5.juni 2015

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

Permitted examination support material: C:
– Tabeller og formler i statistikk, Tapir forlag,
– K.Rottman. Matematisk formelsamling,
– Yellow A-5 sheet with your own handwritten notes (stamped by the Department of Mathematical Sciences)
– Calculator: HP30S, Citizen SR-270X, Citizen SR-270X College or Casio fx-82ES PLUS.

Other information:
All answers should be justified, and relevant calculations provided.
The results from the exam are due by June 26, 2015

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1 Data plot

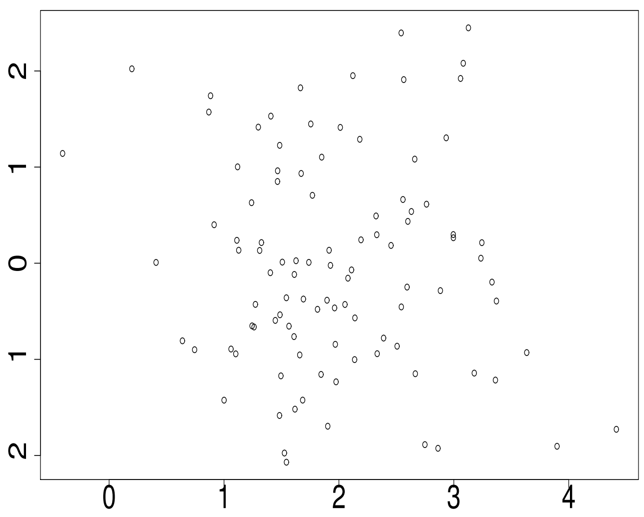

Figure 1: Pairwise data from a joint distribution.

In Figure 1 are shown pairwise outcomes of two stochastic variables X (horizontal) and Y (vertical).

a) What is your estimate of the correlation between X and Y? Provide a short justification for your answer.

The expected value and standard deviation for each of the two variables have integer values. By a visual inspection of Figure 1, give your best estimates of these values.

Problem 2

Assume that A and C are independent events, and that B and C are also independent events.

a) Is it possible for $A \cap B$ and C to be dependent? Can $A \cup B$ and C be dependent? (Hint: Look at e.g. the situation $A \cap B \cap C = \emptyset$.)

Can $A \cup B$ and C be dependent if A and B are disjoint?

Problem 3 Quality control

A farmaceutical company is producing a medicine in liquid form that are sold on bottles containing a specified amount of the medicine. The medicine contains an important ingredient which requires a consistent and running quality control. This is achieved by extracting a random sample from each production series, and by analyzing the content of each of the bottles in the random sample. Each time a random sample is analyzed, a control analysis is also carried out on a sample with known concentration 0.10 mg/l of the ingredient. This is done to ensure that the analysis method is calibrated correctly. Because the analysis method is not completely accurate, the measured concentration will vary. The outcome of the analysis of the control sample can be modelled as a normally distributed stochastic variable X with expected value μ and variance σ^2, where σ^2 denotes the variance of the measurement error inherent in the analysis method, and where μ under normal circumstances is equal to 0.10 mg/l.

An *alarm event* A for the analysis method is defined if the measured value X in the control sample deviates more than one standard deviation from the concentration 0.10 mg/l. That is, $|X - 0.1| > \sigma$ (equivalently, $X < 0.1 - \sigma$ or $X > 0.1 + \sigma$). An *action event* B is defined if the measured value X in the control sample deviates more than two standard deviations from 0.10 mg/l, that is, $|X - 0.1| > 2\sigma$.

a) Assume (only for this subproblem) that $\sigma = 0.01$ mg/l.

Calculate $P(B)$ and $P(B \mid A)$ when $\mu = 0.10$ mg/l.

Then assume that a contaminant has entered the control sample, such that $\mu = 0.11$ mg/l. For this situation, calculate $P(B)$.

Assume for the remaining part of Problem 3 that $\mu = 0.10$ mg/l. To estimate σ^2, the measurement results $x_1, x_2, \ldots, x_n$, from n independent analyses of the control sample will be used. For this purpose, $x_1, x_2, \ldots, x_n$ may be considered as outcomes from n independent stochastic variables $X_1, X_2, \ldots, X_n$, where each X_i, $i = 1, \ldots, n$, has the same distribution as the stochastic variable X.

b) Show that the maximum likelihood estimator (MLE) for σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Even if μ is known, one may also use the estimator

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

where $\overline{X} = \sum_{i=1}^{n} X_i / n$.

In the next subproblems you may use without proof that

$$\sum_{i=1}^{n} (X_i - \mu)^2 / \sigma^2 \ \text{ og } \ \sum_{i=1}^{n} (X_i - \overline{X})^2 / \sigma^2$$

are χ^2-distributed (chi-square distributed) with n and $n-1$ degrees of freedom, respectively.

c) Show that both $\widehat{\sigma}^2$ and S^2 are unbiased.

Which of the two estimators do you prefer? Justify your answer.

The results measured in mg/l of 20 analyses of the control sample have given $\sum_{i=1}^{20} x_i = 1.9240$ and $\sum_{i=1}^{20} x_i^2 = 0.1866$.

d) Derive a 90 % confidence interval for σ^2 by using the estimator you recommended in (c).

Problem 4

Let $X_1, ..., X_n$ denote a random sample (independent, identically distributed stochastic variables) from an exponential distribution with unknown parameter μ, that is, from a distribution with probability density function

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \ ; \ x \geq 0, \ \mu > 0.$$

The null hypothesis $H_0 : \mu = 1$ is to be tested against the alternative $H_1 : \mu < 1$. First the test statistic

$$T = \min\{X_1, ..., X_n\},$$

is used.

a) Show that the probability density function of T is

$$f_T(t) = \frac{n}{\mu} e^{-\frac{nt}{\mu}} \ ; \ t \geq 0.$$

Determine the expected value and the variance of T.

The test that is based on using T, rejects H_0 for small values of T. This can be formulated as:

Test1: If $T < c_1$, reject H_0.

The significance level of the tests in this problem is α.

b) Show that

$$c_1 = \frac{1}{n} \ln \frac{1}{1-\alpha}.$$

The MLE for μ is equal to the sample mean $\overline{X} = \Sigma_{i=1}^{n} X_i/n$.

c) Formulate a Test2 based on $\overline{X}$. If you can assume that $n \geq 30$, write down an approximate probability distribution of $\overline{X}$, and calculate (approximatly) the critical region for Test2.

d) Find an expression for the power of the test for each of the two tests ($n \geq 30$). (The power of the test is the probability of rejecting H_0 when H_0 is not correct; the power of the test depends on a specified alternative.)

Which test is to be preferred? Why? You may base your conclusion on results calculated for $\alpha = 0.05$, $n = 30$, and specified alternative hypothesis $H_1' : \mu = 0.8$.

Why is Test1 a bad test?