Norges teknisk–
naturvitenskapelige universitet
Mathematical Sciences

Corresponding teachers:
Turid Follestad 98066880
Jo Eidsvik 90127472
Mette Langaas 98847649

EXAM IN TMA4245 STATISTICS

Friday 19 May, 2006

Time: 09:00–13:00

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

ENGLISH

Results of exam: 14 June, 2006.

Problem 1 Failure times on mobile service

Let X be the number of weeks between two consecutive failures on a mobile service network. The probability density function of X equals

$$f(x) = \beta x^{-\beta-1}, \quad x > 1,\ \beta > 1.$$

a) Show that the cumulative distribution, $F(x)$, of X is $F(x) = 1 - x^{-\beta}$ for $x > 1$.

Let for the rest of this point $\beta = 3$.

What is the probability that a failure occurs more than 2 weeks after the previous one?

Assume that it has been more than 2 weeks since the previous failure, what is then the probability of a failure taking place before 3.5 weeks since the previous failure?

We next want to estimate the parameter β based on data of failures on this network. Let $X_i,\ i = 1, \ldots, n$ be n times between consecutive failures (measured in weeks). Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with probability density function $f(x)$ given above.

Three alternative estimators for β are

$$\hat{\beta}_1 = \frac{n}{\sum_{i=1}^n \ln(X_i)}, \quad \hat{\beta}_2 = \frac{n-1}{\sum_{i=1}^n \ln(X_i)} \quad \text{and} \quad \hat{\beta}_3 = \frac{\sum_{i=1}^n \ln(X_i)}{n},$$

where ln is the natural logarithm.

b) Which of the estimators above is the maximum likelihood estimator (MLE)? Calculate the estimate using the data below:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1.23	2.04	1.27	1.79	1.10	1.29	2.74	1.15	1.10	1.06

In summary $\sum_{i=1}^{10} \ln(x_i) = 3.39$.

c) Show that $2\beta \ln(X_i)$ is chi-square distributed with 2 degrees of freedom, and that $2\beta \sum_{i=1}^n \ln(X_i)$ is chi-square distributed with $2n$ degrees of freedom.

Construct a 95% confidence interval for β. What is this interval when the failure time data are as above?

Problem 2 Transport of waste

An contractor has hired a transport firm to transport waste from a building site to the waste deposit site. There are two possible routes from the building site to the waste deposit site, either through the city center or around the city center. Due to environmental issues the transport firm is instructed that the transport route around the city center should be used.

When the work is ended the transport firm has made 1000 transports, and they inform the contractor that of the 1000 transports a total of 5 transports have been driven through the city center and 995 around the city center.

During the transport period the contractor checked the choice of route for 5 randomly chosen transports. Let X denote the number of times, for the 5 transports that were checked, the waste was driven through the city center.

What is the distribution of X? Justify your answer.

Which value of X has the highest probability?

For the 5 transports checked, the contractor found that all 5 transports were driven through the city center. What is the probability that this could happen, i.e. $P(X = 5)$?

Problem 3 Compressive strength of clay bricks

A company produces a specific type of clay bricks, and we will study the compressive strength, Y, of the clay bricks. Compressive strength is defined as the maximal pressure that can be applied to the clay bricks, not causing damage. Assume that Y is normally distributed with expected value $\mu = \mathrm{E}(Y)$, given in MPa (10^6 Pascal), and standard deviation $\sigma = \mathrm{SD}(Y) = 0.21$ MPa.

a) Assume that $\mu = 2.10$ MPa.

What is the probability that a randomly selected clay brick has a compressive strength larger than 1.83 MPa, i.e. $P(Y > 1.83)$?

What is the probability that a randomly selected clay brick has a compressive strength deviating less than 0.3 MPa from the expected value $\mu = 2.10$ MPa?

We measure compressive strength of $n = 24$ randomly chosen clay bricks. What is the probability that the smallest measurement of compressive strength is less than 1.83 MPa?

The company is producing a new type of clay bricks, and they believe that the expected compressive strength of the new bricks is $\mu = 2.40$ MPa. Assume that it is known that the standard deviation of the compressive strength of the new bricks is $\sigma = 0.21$ MPa. The company wants to investigate whether there is reason to believe that the expected compressive strength for the new bricks is less than 2.40 MPa.

b) Formulate this as an hypothesis test, defining a null and alternative hypothesis.

Select a test statistic and find the critical region for the test, when a significance level of $\alpha = 0.05$ is chosen.

What is the conclusion of the test when we have observed the compressive strength for $n = 24$ clay bricks, and the average compressive strength was 2.30 MPa?

Calculate the power of the test for the alternative hypothesis $H_1 : \mu = 2.30$ MPa using significance level 0.05 and $n = 24$.

Problem 4 Hubble

An important scientific breakthrough took place in 1929 when Edwin Hubble discovered that the universe is expanding. One of Hubble's datasets consisted of; x_i = distance to galaxy i (measured in millions of light years), and y_i = velocity of galaxy i (measured in thousands of km/s). These data are as follows:

Name	Distance, x_i	Velocity, y_i
Virgo	22	1.2
Pegasus	68	3.8
Perseus	108	5.1
Coma Berenices	137	7.5
Ursa Major 1	255	14.9
Leo	315	19.2
Corona Borealis	390	21.4
Gemini	405	23.0
Bootes	685	39.2
Ursa Major 2	700	41.6
Hydra	1100	60.8

In summary we have that $\sum_{i=1}^{11} x_i = 4185$, $\sum_{i=1}^{11} y_i = 237.7$, $\sum_{i=1}^{11} x_i^2 = 2685141$ and $\sum_{i=1}^{11} x_i y_i = 152220$.

Hubble suggested $y = \beta x$ as a model for velocity as a function of distance, where β later has become known as Hubble's constant. A statistical version of this equation is given by:

$$Y_i = \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, 11, \tag{1}$$

where ε_i, $i = 1, \ldots, 11$, are independent and identically distributed Gaussian random variables with mean 0 and variance σ^2.

a) We first want an estimate of β.

Use the method of least squares to estimate β from equation (1), and show that this estimator for β equals $\hat{\beta} = \frac{\sum_{i=1}^{11} x_i Y_i}{\sum_{i=1}^{11} x_i^2}$. Calculate the estimate based on the data above.

Find the mean and variance of $\hat{\beta}$.

b) Assume that another galaxy is $x_0 = 900$ million light years away.

Predict the velocity, $\hat{y}_0$, of this galaxy.

Construct a 95% prediction interval for an observation of the velocity for this galaxy. We have that $\sum_{i=1}^{11} (y_i - \hat{y}_i)^2 = 9.87$, where $\hat{y}_i = \hat{\beta} x_i$.