Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

English

Contact during exam:
Håkon Tjelmeland 73 59 35 20
Oddvar Kristian Husby 73 59 35 20
Odd Kolbjørnsen 73 59 35 20

EXAM IN COURSE SIF5060/SIF5505 STATISTICS

Thursday November 30th 2000
Time: 09:00–14:00

Permitted aids: Approved calculator with empty memory.
Tabeller og formler i statistikk, Tapir forlag.
K. Rottman: Matematisk formelsamling.
English-Norwegian vocabulary handed out in class.

Problem 1 Fabrizio Frizzi

An Italian TV station sends a Sunday show where the program director, Fabrizio Frizzi, sits next to a large safe containing 250 000 000 Italian lire. The safe has a secret code consisting of four digits. The viewers are calling Frizzi by phone and guess on codes. The first one who guess the code correctly wins the amount in the safe. So each of the digits in the code is one of the numbers $0, 1, 2, 3, 4, 5, 6, 7, 8$ or 9. Assume that none of the viewers are able to remember any of the previously suggested (wrong) codes, so that each guess on the secret code can be considered a random guess for a four digit code.

a) What is the probability that a random viewer guesses the code correctly?

Let X denote number of calls until a person wins the money. Explain briefly why X has a geometric distribution.

What is the probability that caller number 300 is the first who guesses the code correctly?

Now assume that in the start of the show, the viewers are informed that the digit 7 occurs exactly two times in the secret code (this is the variant actually used in Italian TV). Let m denote the number of codes that is now possible.

b) Show that $m = 486$ (i.e. show how this figure is computed).

If Frizzi speaks sufficiently fast to receive two calls a minute, what is the expected time until a caller wins the 250 000 000 lire?

Now assume that viewers write down previously guessed codes, so that no callers will ever guess on a previously suggested code. Let Y denote the number of calls until a caller guess the correct code, for this situation. Still consider the variant where the viewers are informed that the digit 7 occurs exactly two times in the secret code.

c) Give the answers of the following questions as a function of m:

What is the sample space of Y?

Derive the probability distribution for Y.

What is now the expected time until a caller wins the 250 000 000 lire (still assuming Frizzi to reach two callers a minute)?

Problem 2 Exit polls

At elections and referendums it is today common to perform so called "exit polls". Then a random sample of the voters are asked, when they leave after having voted, on what candidate they have voted. In this exercise we are going to analyse this situation for an election for two candidates only, which we are going to denote G and B, respectively. We will ignore the possibility that voters may give a blank vote and the votes may be rejected. We will also ignore the possibility that voters may deny to answer or give an untrue answer when being asked about their voting.

Let N denote the total number of persons having voted and let p be the fraction that votes for G. Furthermore, let n denote the number of voters that are asked about their voting and let X be the number of those n voters that vote for G.

a) Which condition(s) must be fulfilled for X to be approximately binomially distributed?

Given that this condition is fulfilled, compute the following probabilities for $n = 20$ and $p = 0.50$:

$$\mathrm{P}(X = 9) \quad , \quad \mathrm{P}(X > 9) \quad \text{and} \quad \mathrm{P}(X > 9 | X \leq 12)$$

In the remainder of this exercise we assume that X is (approximately) binomially distributed.

b) Derive the maximum likelihood estimator (MLE) for p.

Find the expectation and variance of this estimator.

A binomial distribution can be approximated with a normal distribution if n is sufficiently large. In the remainder of this exercise we will assume this condition to hold, so that

$$\frac{X - np}{\sqrt{np(1-p)}}$$

has (approximately) a standard normal distribution.

Assume that a TV station is doing the exit poll. If a sufficiently large majority of the persons asked about their voting, have voted for one of the candidates, the TV station will declare that candidate as winner of the election.

c) Formulate this as a hypothesis test. Specify the null hypothesis and the alternative hypothesis and determine the criterion for rejecting the null hypothesis. Use significance level α.

What is the conclusion of the hypothesis test if 5 000 voters are questioned about their voting, 2562 of those voted for G, and you use significants level $\alpha = 0.10$?

Problem 3 SAR-measurements

SAR (Synthetic Aperture Radar) is a method for surveying the earth surface from satellite. Measurements are taken by sending out radar beams from the satellite and observing how much of this that is reflected back to the satellite. The degree of reflection depends on properties of the earth surface in the location surveyed and thereby one is able to differentiate between different types of surfaces. One observation is really done by taking several measurements (so called "looks") and summing the result. From physics of radar beams it is known that an observations, X, is gamma distributed with parameters a and $b = r/a$. Thus, the probability density is given by

$$f(x) = \frac{a^a}{r^a \Gamma(a)} x^{a-1} \exp\left\{-\frac{ax}{r}\right\},$$

where a is number of "looks" used and the reflection parameter r depends on the surface properties. From known formulas for expectation and variance in a gamma distribution, we then know the expectation and variance of X to be

$$\mathrm{E}(X) = r \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{r^2}{a}.$$

We are going to assume that n observations is taken from a homogeneous area (i.e. the parameter r is the same for all n observations). Let $X_1, X_2, \ldots, X_n$ denote the n observations and assume them to be independent. From these observations we want to estimate r. The number of "looks", a, is known. We use the estimator

$$\widehat{r} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

a) Show that $\widehat{r}$ is unbiased and find the variance of the estimator.

Explain briefly the general result of the central limit theorem. Thereafter, explain how we in our situation can deduce from this theorem that

$$\sqrt{na}\,\frac{\widehat{r} - r}{r} \tag{1}$$

has an approximate standard normal distribution when n is large.

b) Use the result in **a**) to derive an approximate $(1-\alpha)\cdot 100\%$ confidence interval for r when n is large. Give the answer as a function of α, a, n and $\overline{X}$.

Compute also numerical values for the interval when $\alpha = 0.05$, $a = 5$, $n = 20$ and observed values for $X_i; i = 1, \ldots, 20$ are

7.98	10.82	15.88	17.00	24.22	12.20	8.17	16.53	7.46	14.31
34.55	19.46	20.21	13.58	10.98	4.42	24.92	30.29	23.45	23.36

It is given that $\overline{x} = 16.99$.

Now assume that we have observations from two different areas, area A and area B. From area A we have n observations, $X_1, X_2, \ldots, X_n$, all assumed to have the same reflection parameter r_A. From area B we also have n observations, $Y_1, Y_2, \ldots, Y_n$, these all have reflection parameter r_B. We are going to assume that all $2n$ observations are independent and that they are all taken with the same number of "looks". From the $2n$ observations we want to test if the two reflection parameters differ, i.e. null hypothesis and alternative hypothesis are

$$\mathrm{H}_0 : r_A = r_B \qquad \text{against} \qquad \mathrm{H}_1 : r_A \neq r_B.$$

c) With n still assumed to be large, find a suitable test statistic to test the hypotheses given above and determine the criterion for rejection so that the test gets (approximately) an α level of significance. If necessary, you may do more approximations, but if so you have to give reasons for their validity.

Problem 4

Assume X to have a standard normal distribution and let $Y = X^2$. Show that then Y has a χ^2 distribution with 1 degree of freedom.