Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

English

Contact during exam:

Odd Kolbjørnsen	73 59 35 27
Håkon Tjelmeland	73 59 35 38
Arvid Næss	73 59 70 53

SIF5060/SIF5505 Statistics

Thursday November 29th 2001
Time: 09:00–14:00

Permitted aids: Specific simple calculator
Statistiske tabeller og formler, Tapir forlag.
K. Rottman: Matematisk formelsamling.

Final grading : January 13th 2002.

Problem 1 Lotto gambling

In this problem we analyse two different aspects of Lotto gambling.

In the game Lotto a single sequence is played by choosing 7 out of 34 numbers. The gambler can also play a system. When a gambler plays a system, he chooses m out of 34 numbers, where m denotes the count of numbers in the system. When a gambler plays a system having the size m, the count of single sequences being played is equal to the count of possible combinations of 7 numbers out of the m numbers in the system.

a) How many single sequences is played in a system containing 8 numbers?

How many single sequences is played in a system containing m numbers?

Each single sequence costs NOK 3,- to play. How much does it cost to play a system consisting of 12 numbers? (This is the maximum size of a system allowed by Norsk Tipping.)

As a total there is 5 379 616 possible single sequences that can be played in Lotto. Only one of these sequences is a winning sequence, hence the probability that a single sequence selected at random will be a winning one is $p = (5\,379\,616)^{-1} \approx 1.86 \cdot 10^{-7}$. In the game of Lotto played November 17th $n = 21\,481\,335$ single sequences were played. In the current problem we will assume that all of these n single sequences played this day were drawn independently at random among all possible single sequences. Let X denote the count of sequences that coincide with the winning sequence, that is the count of persons to win the premium prize. By the assumptions above X has a binomial distribution, $b(x; n, p)$.

b) Which conditions must be fulfilled to approximate a binomial distribution with a Poisson distribution?

Compute the Poisson approximation to the distribution of X.

Compute the probability masses for the events $X = 0$ and $X = 1$, both for the exact binomial distribution and for the Poisson approximation. Is the Poisson approximation reasonable in this case?

Problem 2 The first to score wins ?

In todays society many people have a passionate interest in soccer and self claimed experts frequently have opinions regarding soccer matches. In the current problem we only consider matches between two particular teams. We denote the teams R and L. An expert commentator from television made the following claim regarding matches between R and L: “the team that is the first to score in a match, will usually also win the match. In the current problem we use this claim as a motivation for our computations.

For a soccer match between R and L, define the events:
R: Team R wins the match
F: Team R is the first to score
I : There is no goal in the game, that is 0-0.

a) Assume for this point that $\mathrm{P}(R) = 0.4$, $\mathrm{P}(F) = 0.5$, $\mathrm{P}(R \cap F) = 0.3$ and $\mathrm{P}(I) = 0.05$.

Sketch the events R, F and I in a Venn diagram.

What is the probability that team R wins the match given that team R scores the first goal, i.e. what is $\mathrm{P}(R \mid F)$?

What is the probability that team R wins the match given that there is at least one goal in the match, i.e. what is $\mathrm{P}(R \mid I')$ (where I' denotes the complement of I)?

In what follows we only analyse matches between R and L that ends with at least one goal. Let p denote the probability that the team that is the first to score also wins the match. We assume that this probability does not depend on which of the teams that play at their home field. We estimate p based on the last n matches between R and L in the national league (only those matches that contain at least one goal are included). Among these n matches, let X denote the number of matches where the team that got the first goal also won. We assume X to be binomially distributed with parameters n and p, and we use the estimator

$$\widehat{p} = \frac{X}{n}.$$

b) In general, which conditions must be fulfilled in order for a random variable to have a binomial distribution. Is it reasonable to assume X to be binomially distributed?(give a reason for your answer)

Explain briefly the general result of the central limit theorem.

Use the central limit theorem to argue that

$$\frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is approximately a standard normal variable when n is large.

When the expert commentator mentioned above was asked to quantify his claim that in matches between R and L the team that is the first to score will usually also win the match, he said that p is not smaller than 0.80. We want to investigate whether our observed value for X can falsify the claim of the expert.

c) Formulate this as a test of statistical hypothesis. Use 5% level of significance and derive a rule for when to reject H_0.

What is the conclusion of the test when $n = 24$ and $x = 17$? (This is the result of matches between Rosenborg and Lillestrøm in the period 1990-2001. All of the matches in this period contained at least one goal.)

d) Assume the rejection rule derived in **c)** to be used, but that the true value of p is 0.7. How many match observations must be considered to make the probability of detecting the false claim of the expert to be at least 0.9.

Problem 3 The parking ticket

A car owner that does not pay a parking fee, gets a parking ticket of NOK 300,- if a parking guard detects the offence. If parking guards detect the same offence more than once the size of the parking ticket remains the same. In this problem we analyse *statistical* aspects of this situation.

We assume the parking guards to arrive at a specific parking lot according to a Poisson process with parameter λ. Let T denote the time from a car is parked until the first parking guard arrives. Under the assumption of a Poisson process it is known that T has an exponential distribution, so the probability density of T is

$$f(t;\lambda) = \begin{cases} \lambda \exp\{-\lambda t\}, & \text{if} \quad t > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In this problem, you can without proof use that whenever $T_1, T_2, ..., T_n$ are independent and exponentially distributed then

$$\mathrm{E}\left\{\frac{1}{\sum_{i=1}^{n} T_i}\right\} = \frac{\lambda}{n-1}.$$

a) For this point only, assume $\lambda = 1/5$ per hour.

Katrine forgets to pay her parking fee and is away from her car for 2 hours. What is the probability that she has received a parking ticket when she returns?

Assume Katrine forgets to pay her parking fee and is away from the car for t hours. If a parking guard discover that she has not payed the parking fee, she gets a ticket of NOK 300,-. Show that the expected cost for her parking becomes NOK

$$300(1 - \exp\{-\lambda t\}).$$

The regular parking fee is NOK 30,- per hour. Will it pay off for Katrine not to pay the parking fee if her car is parked at the same lot for 8 hours ?

The owners of the delivery firm Cheat & Drive Ltd gets to know that it might pay off not to pay the parking fee. They see this as an opportunity to reduce costs. To get more information about the situation, they want to estimate λ. To estimate λ they let a new employee observe a parking lot. The new employee reports time intervals, $T_1, T_2, ..., T_n$, between arrivals of parking guards. Under the assumption of a Poisson process, these times have an exponential distribution.

The $n = 20$ observed times are

0.56	2.79	2.62	5.95	0.92	3.23	0.04	3.27	5.13	1.82
1.79	1.01	3.01	2.27	0.70	0.05	3.84	0.35	0.81	2.35

It is given that $\sum_{i=1}^{20} t_i = 42.51$.

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

Is the estimator unbiased? If not, propose an unbiased estimator by an adjustment of the maximum likelihood estimator.

Using the unbiased estimator, what is the estimate of λ with the data set given above?

c) Show by the use of moment generating functions that $V = 2\lambda \sum_{i=1}^{n} T_i$ have a χ^2 distribution with $2n$ degrees of freedom.

d) Use the result in **c)** to derive a 95% confidence interval for λ.

In **a)** it was shown that the expected cost for a car being parked for t hours without the parking fee payed, was $\gamma = 300(1 - \exp\{-\lambda t\})$. Use the confidence interval found for λ to derive a 95% confidence interval for γ.

Compute the confidence interval for γ numerically for $t = 8$ hours and with the data given above.