Department of mathematics and statistics, University of Helsinki

Johdatus tilastolliseen päättelyyn (Arjas), kurssikuulustelu 9.5.06
Introduction to statistical inference (Arjas), final exam 9.5.06

1. Pekka has a number of white and black balls, which are similar apart from their color. He places them one-by-one into a bowl (hiding this process from you), always deciding on the color of an individual ball on the basis of a coin toss: choosing a white ball if the coin lands Heads, and black if it lands Tails. Pekka places in this way two balls into the bowl, but does not tell you what colors they were. He asks you to take one of them "blind", without looking into the bowl, and it turns out to be white. (a) What is the probability that also the other ball still in the bowl is a white one? Justify your answer by using Bayes' formula. (b) You put the ball, which you just picked, back into the bowl, after which Pekka shakes it properly. You are allowed to again pick a ball "blind". What is the probability, given the result of the first pick, that also the second pick turns out white?

2. Let X_1, X_2 be a simple random sample from the uniform distribution Unif(0, θ), i.e., from the distribution with density function equal to $1/\theta$ on the interval $(0, \theta)$ and 0 outside, with θ being the parameter of this distribution. What is the expression of the likelihood function corresponding to the observed value (x_1, x_2)? Find also the resulting maximum likelihood estimate of θ. (*Hint*: Begin by considering only a single observation.)

3. In a poll people were asked to answer the question "Do you happen to jog?" There were 400 respondents. Suppose that in reality 20% of the Finnish adult population are joggers and that the poll was based on simple random sampling from this population.
   (a) Denote by $\overline{X}_{400}$ the proportion of joggers in the sample. What is its expected value and standard deviation?
   (b) By applying the normal approximation, find the probability that $\overline{X}_{400}$ is between 0.18 and 0.22.
   (c) What sample size would be required to reduce the standard deviation of $\overline{X}_{400}$ to one-half the value you found in (a)?

4. The developer of a new filter of filter-tipped cigarettes claims that it leaves less nicotine in the smoke than does the current filter, thus absorbing more. Because cigarette brands differ in a number of ways, he tests each filter on one cigarette of each of nine brands and records the difference between the nicotine content for the new filter and the current filter. The mean difference is $\bar{x}_9 = 1.32$ mg and the standard deviation $s = 2.35$ mg. Formulate this problem in the form of a statistical test (state the null hypothesis H and its alternative A) and consider the outcome in terms of the resulting p-value.

Matematiikan ja tilastotieteen laitos, Helsingin yliopisto

Johdatus tilastolliseen päättelyyn (Arjas), final exam 9.8.2006 (English text)

1. Suppose that the variables $X1, X2, \ldots, Xn$ have been sampled according to simple random sampling from a normal distribution $N(\mu, \sigma^2)$, where the variance σ^2 has a known value. Show that the maximum likelihood estimator of μ is given by the arithmetic mean of the observations.

2. The balls in a bowl, of which K are white and the rest are black. However, you don't know the value of K. Therefore you think it will be reasonable to describe your uncertainty about its value, initially, by saying that each of the three balls is equally likely – and independently of the colours of the other balls - to be either black or white. After this you are asked to take "blind" two balls from the bowl, without putting either of them back. They both turn out white. What is then the (posterior) probability of that also the third ball, still in the bowl, is white?

3. A study of working couples measures the income X of the husband and the income Y of the wife in a large number of couples in which both partners are employed. Suppose that you knew the means and the variances of both variables in the population. (a) Is it reasonable to take the mean of the total income of both partners to be the sum of these two means? Explain why you think that this is, or is not, reasonable. (b) Is it reasonable to take the variance of the total income to be the sum of the two variances? Again, explain your reasoning.

4. A shopkeeper regularly receives shipments of potatoes from a producer. They have been packed into sacks with a nominal weight of 25 kg, but in practice there are minor differences in their true weights. We assume that the producer has actually decided on some target weight μ, and that this variability from one sack to another can be reasonably well be described by independent sampling from the normal distribution $N(\mu, (0.2)^2)$. The shopkeeper does not know the true value of μ, however, and he would like to verify that it is not below the nominal weight 25 kg. He therefore decides to weigh a number of sacks with a pair of scales which he has in his shop. Suppose that the scales are not biased either way, but that they are not very accurate. More precisely, suppose that the measurement errors, each time when a sack is being weighed, can be treated as independent random variables distributed according to $N(0, (0.3)^2)$. Based on this, what is the distribution of the individual weighing results? Suppose that the shopkeeper actually has five sacks in his shop, and that he then gets the readings 25.5, 25.0, 24.9, 25.3 ja 25.3 kg when weighing them. Based on this, determine the level 0.95 confidence interval for the nominal weight μ.

5. The shopkeeper who was mentioned above has attended the lectures of the course "Johdatus tilastolliseen päättelyyn" and then decides to try out what kind of a result he would get by applying statistical hypothesis testing to his problem. He would like to be able to demonstrate a deviation of 0.2 kg downwards from the nominal weight of 25 kg with a probability at least equal to 0.90 (i.e., the power of the test at 24.8 kg should be at least 0.90). On the other hand, he would rather not make a formal complaint to the producer if there was no real reason for making one, and therefore he chooses the significance level of the test to be $\alpha = 0.01$. Design a statistical test (determine sample size and critical value) such that the sample size (the number of sacks that need to be weighed) is just enough to satisfy these two desiderata. For simplicity, we assume here, just as above, that the true weights of the sacks vary according to the distribution $N(\mu, (0.2\text{ kg})^2)$, but now for simplicity that the scales which the shopkeeper uses are so good that there is no need to take into account their measurement errors.

THE STATISTICAL TABLES THAT ARE NEEDED WILL BE MADE AVAILABLE IN THE EXAM. "CLOSED BOOK" EXAM: NO OTHER SOURCE MATERIAL IS ALLOWED!