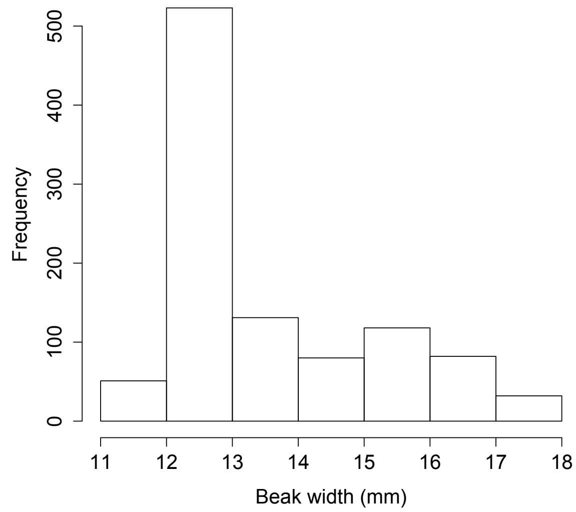
Statistics for the Doctoral School in Biomolecular Sciences
Academic year 2013-2014

Exercises after Lecture 1

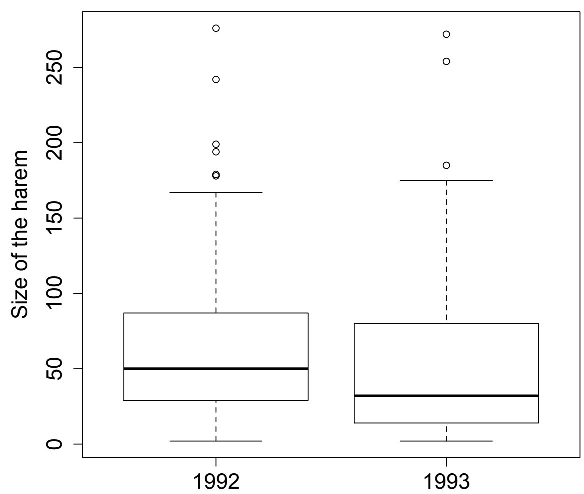
Probability, random variables, empirical distributions.

EXERCISE 1.1 The following diagram represents the distribution of beak width in 1017 individuals of *Pyrenestes ostrinus*, an African bird.



1. Which is the most frequent beak width? Estimate the fraction of birds whose measure lies in the most frequent interval?
2. Estimate approximately the mean of the data.
3. There is a hint of a second peak in the distribution. What would be a good strategy to investigate whether there is actually a second peak? What is the name of a distribution with 2 peaks?

EXERCISE 1.2 A study by Modig (1996) has obtained data on the size of harems of male elephant seals (*Mirounga leonina* in two different years, 1992 and 1993. the data are represented size by size in the following box-plot.



1. In which year was the median larger? What is approximately the size of the central interval that contains 50% of the data?
2. Are the distributions roughly symmetrical?
3. Can you sketch how a histogram of these data may look like? *[It may be appropriate using intervals of different lengths in a histogram]*

EXERCISE 1.3 A population is composed of 6 individuals; the morphogenic indices[1] of each of them are:

$$-1 \qquad 0 \qquad 1 \qquad 1 \qquad 3 \qquad 5.$$

Compute mean and variance of the morphogenic index in this population.

Consider all possible samples of size 2 that can be extracted from this population, and compute sample mean and variance for each sample. What is their relation with the *true*[2] mean and variance?

EXERCISE 1.4 A 'balanced' coin is tossed 6 times. Two possible sequences of results are

1. H T T H T H H = head
2. T T T T T T T = tails.

Choose, among the three following statements, the correct one, explaining why:

(i) sequence 1. is more likely.

(ii) sequence 2. is more likely.

(iii) both sequences are equally likely.

EXERCISE 1.5 In a human population it is estimated that the probability of being deaf is $5 \cdot 10^{-3}$ and that of being blind is $8.5 \cdot 10^{-3}$. The probability of being deaf and blind is $6 \cdot 10^{-4}$. What is the probability of being deaf and/or blind?

EXERCISE 1.6 A urn contains 5 white balls and 2 red ones. We extract 3 balls (without putting back in the urn the extracted ones). What is the probability that all balls are white? What is the probability that at least 2 of them are white?

EXERCISE 1.7 Knowing that in Japan the probability that a child is born female is 48.8%, and assuming that the sex of each child is independent on the others', how many children must the emperor of Japan have to reach a probability above 90% of having a male son?

EXERCISE 1.8 A student gives random answers to 5 multiple-choice questions (for each questions 4 answers are given of which only 1 is correct).

- What is the probability that at least 1 answer is correct?
- What is the probability that exactly 1 answer is correct?

EXERCISE 1.9 A balanced die is tossed 24 times; What is the probability to get 6 at least once? On average, how many times is obtained 6 in 24 tosses? Which is the standard deviation of the number of times that 6 is obtained? Explain an intuitive meaning of this value.

EXERCISE 1.10 A coin is tossed 2 times. You win 2 euros if the two coins show the same face; otherwise you lose 3 euros. Compute mean and standard deviation of the gain *(a loss is counted as a negative gain)*.

[1] a fake measure...
[2] of the population

EXERCISE 1.11 A firm sells ball bearings in chests containing 1.000 pieces. The firm knows that around 1 ball bearing every 2,000 does not fulfill the standards. Let X the number of defective ball bearings in a chest

1. What is the probability that $X = 2$?
2. Compute $\mathbb{E}(X)$ and the standard deviation.
3. What is the probability that a chest contains at least 1 defective ball bearing?
4. Answer to the previous question using a Poisson distribution to approximate X; is the answer very different?

EXERCISE 1.12 The Internet traffic on an old data transmission line is affected by random errors with probability 10^{-4} errors per bit.
Keeping in mind that a character is represented by a sequence of 8 bits, compute the probability that a text 100-characters long is transmitted without errors, and the probability that the number of errors is between 1 and 2 [included.
[Assume that the number of errors is distributed according to a Poisson.]

EXERCISE 1.13 In a bacterial culture there are on average 512 bacteria per cm^3. Assuming that the spatial distribution of bacteria follows a Poisson, which is the standard deviation of the number of bacteria per cm^3?

EXERCISE 1.14 A study has examined the spatial distribution of yeast cells in a haemocytometer. In particular, data have been obtained on the number of cells present in each of 400 quadrats in which the surface of the haemocytometer has been divided; the data are in the following table.

Number of cells per quadrat	Frequency observed	Number of cells per quadrat	Frequency observed
0	75	5	13
1	103	6	2
2	121	7	1
3	54	8	0
4	30	9	1

1. Represent graphically these data.
2. Compute the mean, standard deviation and median of them.
3. Check whether the distribution of yeast cells can be approximated with a Poisson (this is considered the random distribution).

EXERCISE 1.15 A balanced die is tossed 24 times. On average, how many times is obtained 6 in 24 tosses? Which is the standard deviation of the number of times that 6 is obtained? Explain an intuitive meaning of this value.

EXERCISE 1.16 A balanced die is tossed n times, where n is a very large number. What does the law of large numbers tell us about the number of times that 6 will occur out of the n trials? What does the central limit theorem tell us? Explain why these results are not in contradiction with each other.

EXERCISE 1.17 A balanced die is tossed 100 times. Show how one can use the central limit theorem to approximate the probability of obtaining 6 20 time or more. *[You simply have to write the probability in terms of the normal distribution, without actually computing it.]*

EXERCISE 1.18 A researcher tests a drug on a sample of 150 individuals, 50 males and 100 females; of these 24 males and 76 females recover. In the control sample (100 males and 50 females) 50 males and 40 females recover. The researcher concludes that 100 individuals out 150 using the drug have recovered, while only 90 out of 150 in the control sample, so that the drug is effective. Does it seem a correct conclusion?
State the results in the language of probability.

EXERCISE 1.19 You extract 2 cards (without re-immission) from a deck of 52 cards (26 red and 26 black).

(a) Which is the probability that the first card extracted is red?

(b) Which is the probability that the second card extracted is red if the first is red?

(c) Which is the probability that the second card extracted is red if the first is black?

(d) Which is the probability that the second card extracted is red?

(e) Which is the probability that the first card extracted is red if the second is red?

EXERCISE 1.20 An urn coloured in blue contains 3 red balls and 2 white ones. An urn coloured in yellow contains 6 red balls and 1 white one. First an urn is randomly selected, then a ball is extracted by it.

(a) Compute the probability that the yellow urn is chosen and a red ball extracted from it.

(b) Compute the probability that a red ball is extracted.

(c) Knowing that a red ball has been extracted, compute the probability that it comes from the yellow urn.

EXERCISE 1.21 The seeds of a red-flowered plant variety sprout in 60% of the cases, those of the white–flowered variety sprout in 90% of the cases. In a box there are 40 seeds of the red-flowered variety and 80 of the white-flowered variety. Taking a seed at random from the box, with which probability will it sprout?

EXERCISE 1.22 Let p be the probability that a child is born male. In a family with 2 children, which is the probability that both are males knowing that one is male? In a family with 2 children, which is the probability that both are males knowing that the first one is male?