SABS PROBABILITY PRACTICAL

DAVID STEINSALTZ

The first 5 questions are intended for the first practical session. The last question is a more involved computing question, to practice data handling and exploring distributions in `R`.

(1) Blood groups are distributed in the UK as follows (`https://www.blood.co.uk/why-give-blood/blood-types/`):

O	A	B	AB
48%	39%	10%	3%

(a) If two people are picked at random from the population, what is the probability that their blood is of the same type? Of different types?
(b) If four people are picked at random, let p_k be the probability that there are exactly k different blood types among them. Find all values of p_k.

(2) In a genetic experiment, the offspring of a particular cross have 25% chance of being yellow, and 75% chance of being green.
(a) If there are 10 offspring, calculate the probability that at least 8 are green.
(b) Suppose they also have a 25% chance of being short and 75% chance of being tall. Calculate the expected number that are tall and yellow. What assumptions do you need to make?

(3) Continuing an example from the lecture, suppose there is a disease that occurs in three forms: Mild, Severe, and Lethal. There is a gene that is known to occur in two *alleles* (variants), denoted A_1 and A_2, where the A_1 allele provides some protection against the symptoms of the disease, but does not prevent the disease. 75% of the general population has A_1, and 75% of those with the disease has A_1. Of those people with A_1 who have the disease, 90% have the Mild form, and the rest have the Severe form. The A_2 sufferers are evenly split between the Severe and Lethal forms.

Suppose you observe a patient with the Severe form. Calculate the probability that the patient is of type A_1. Do this with Bayes' Rule and with natural frequencies.

(4) I roll a fair die, and then flip a number of fair coins equal to the number that came up on the die.
(a) Calculate the probability that exactly four heads come up on the coins.
(b) Given that four heads come up, calculate the probability that the die roll was 5.

(5) Suppose we have a sequence of independent trials, each with probability p of success (called *Bernoulli trials* with parameter p). Let X be the number of the trial on which you have the first success.
(a) Calculate the probability mass function $\mathbb{P}\{X = k\}$.
(b) Calculate the expectation and variance of X. (This requires the formula for the sum of a geometric series, and some useful variants which you can find in the last section of the Wikipedia article `https://en.wikipedia.org/wiki/Geometric_series`.)

Date: 5 December, 2019.

(6) Stroke patients with aphasic deficits are each given a number of straightforward tasks in a psychometric test. The number of errors made by 123 patients are shown in the table below. Calculate the expectation and variance of the number of errors per patient and comment on these values. Fit a Poisson distribution and comment on how well it fits the observed data.

Number of errors	0	1	2	3	4	5 or more
Number of patients	5	30	56	15	10	7

(7) Let X have uniform distribution on the interval $[a, b]$ defined as the continuous distribution whose density is constant on that interval and 0 outside it.
  (a) What are the density and the cdf of this distribution?
  (b) What are the expectation and variance?

(8) In a certain country the heights of adult males have mean 170cm and standard deviation 10cm, and the heights of adult females have mean 160cm and standard deviation 8cm; for each sex the distribution of heights approximates closely to a normal probability model. On the hypothesis that height is not a factor in selecting a mate, calculate the probability that
  (a) a husband and wife selected at random are both taller than 164cm
  (b) in a randomly selected husband and wife the wife is taller than the husband;
  (c) the average height of a random couple is greater than 168cm.

(9) 100 students each perform an experiment to estimate a parameter μ, and each one independently computes a 99% confidence interval for μ. What is the probability that there will be at least 3 students whose confidence intervals do not include μ? (Hint: Use the binomial distribution or the Poisson distribution.)

(10) (a) Using `R` (or another computer language) compute the following:
    (i) $\mathbb{P}\{X = 112\}$ where X is binomial with $n = 200$, $p = 0.6$.
    (ii) $\mathbb{P}\{X \geq 4\}$ where X is Poisson with parameter 8.
    (iii) $\mathbb{P}\{1 < X < 2\}$ where X is Exponential with parameter 2.
    (iv) $\mathbb{P}\{X < 2\}$ where X is normal with mean 3 and variance 7.

  For all examples but the last one, do the computation two ways, using your knowledge of the probability mass function or density, and using the appropriate distribution function in `R`. For the last one, find z such that the probability is the same as $\mathbb{P}\{Z < z\}$, where Z is standard normal (i.e., mean 0 and variance 1).

  (b) Simulate 1000 groups of k copies of an exponential random variable with parameter 2, where $k = 1, 2, 10, 100$. For each k do the following:
    (i) Compute the mean of each group. So you now have 1000 numbers.
    (ii) Use a normal Q–Q plot to compare it to a normal distribution.
    (iii) Plot a histogram.
    (iv) *[optional]* Draw a normal curve with the same mean and variance on top of the histogram.