MAS131/231: Introduction to Probability
Exercises

Some of the exercises below are for you to tackle on your own for assessment purposes. Others are for discussion at tutorials. The rest are for practice. Homework will be due each fortnight. The numbers of questions to be assessed, and those to be discussed in tutorials will be emailed to everyone registered for MAS131 and MAS231, and also posted to the local newsgroup `ncl.mas.131`. The hand-in date and the date of the tutorials will also be given. Your completed exercises should be posted in the box in the Foyer by the General Office. Remember to attach a completed cover sheet to your work before posting it into the box.

1. *You should work through the sheet marked* ***Practical 1: Introduction to Minitab*** *before attempting this exercise.*

   In this exercise you will analyse the data collected on members of the class. This question should be tackled during the first Computer Practical (after working through Practical 1), and finished in your own time. You may wish to write your solutions to this exercise in Microsoft Word, with Minitab plots "copied" and "pasted" in.

   A copy of the data is given on an additional sheet, but is also stored on the network ready for reading into Minitab — *do not enter the data by hand into Minitab!* The data has *not* been checked for outliers or obvious errors/inconsistencies; such checking and (if possible) correction should be part of your analysis.

   Using Minitab, and commenting on the results you obtain:

   (a) find the number of males and females, and the number of left- and right-handed people;

   (b) produce an appropriate diagram for the distance from home of class members;

   (c) produce diagrams to compare the hand-span distribution of males and females;

   (d) produce a scatter-plot of foot length against height, distinguishing between males and females;

   (e) produce appropriate summary statistics for numbers of siblings (brothers and sisters) and age;

   (f) for males and females separately, produce appropriate summaries of height, foot length and hand-span.

   (g) Investigate at least one more interesting feature of the data. For example, you could see if there are any special characteristics of left-handers, eldest children, locals, or people who have taken a year-out.

2. *By hand* (using the data on the printed sheet), draw a histogram and a stem and leaf plot of the heights of the female members of the class (females are identified by a 2 in the first column). Using a calculator, find the mean and standard deviation of female height. Again by hand, find the median, lower and upper quartiles of female height, and draw a box-plot. Be sure to show all of your working. Briefly summarise your findings.

3. *By hand* (using the data on the printed sheet), produce a frequency table, a bar chart and a frequency polygon for the number of siblings.

4. If 85% of people have a bowl of cereal for breakfast, 60% of people have some toast for breakfast, and 50% of people have both cereal and toast for breakfast, what proportion of people have neither cereal nor toast for breakfast?

5. If 50 percent of families in a certain city subscribe to the morning newspaper, 65 percent subscribe to the afternoon newspaper, and 85 percent of the families subscribe to at least one of the two newspapers, what proportion of families subscribe to both newspapers?

6. Show *algebraically*, using only the probability axioms and properties, and basic set-theoretic results given in the lectures, that

   (a) $\mathrm{P}(A \cap B^c) = \mathrm{P}(A) - \mathrm{P}(A \cap B)$

   (b) $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

   (c) $\mathrm{P}\big(A \cup \bar{B}\big) = 1 - \mathrm{P}(B) + \mathrm{P}(A \cap B)$

   (d) Use (b) to show that

$$\begin{aligned}\mathrm{P}(A \cup B \cup C) = {} & \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) \\ & - \mathrm{P}(A \cap B) - \mathrm{P}(B \cap C) - \mathrm{P}(A \cap C) + \mathrm{P}(A \cap B \cap C)\end{aligned}$$

7. A pair of six-sided dice are thrown. The sample space consists of all 36 possibilities $\{(1,1), (1,2), \ldots, (6,6)\}$, all equally likely. Let E denote the event that the sum of the faces is odd, and let F denote the event that at least one face is a six. List the points in the following events, and hence write down their probabilities:

   (a) $E \cap F$

   (b) $(E \cup F)^c$

   (c) $E \cup \bar{F}$

   (d) $E^c \cap F$.

8. (a) A student is to answer 7 out of 10 questions in an examination. How many choices has she? How many choices does she have if at least 3 of the first 5 questions have to be answered? *Hint: Split the exam into the first 5 and last 5 questions.*

   (b) A line-up of 10 men is conducted in order that a witness can identify 3 suspects. Suppose that 3 people in the line-up actually committed the crime in question. If the witness does not recognise any of the suspects, but simply chooses three men at random, what is the probability that the three guilty men are selected? What is the probability that the witness selects three innocent men?

9. (a) What is the probability of being dealt two Aces and two Kings in a hand of Poker? In Poker, you are dealt 5 cards from a pack of 52. The 52 cards are made up of 13 cards in each of 4 suits. The "other" card should not be an Ace or a King, as this would give a Full House.

   (b) What is the probability of being dealt Two Pairs in a hand of Poker. Again, be sure not to include Full Houses.

10. (a) What is the probability of getting exactly 4 numbers on the National Lottery?

    (b) What is the probability of getting the Jackpot (all 6 main numbers)?

(c) What is the probability of getting 5 main numbers plus the bonus ball?

(d) What is the probability of winning a prize (getting *at least* 3 numbers)?

(e) If you play one line per week for 2080 weeks (40 years), what is the probability that you win the Jackpot at least once?

11. Capture-recapture is a widely used statistical technique in environmental statistics.

(a) A lake contains N fish. One day n of these are caught, marked with a tag, and returned to the lake. A week later, m fish are caught. What is the probability that k of those m are marked?

(b) If the lake contains 1,000 fish, 100 of which are marked, and you catch 10, what is the probability that *at least* one of the fish you catch is marked?

12. *This question is to be started during the tutorial, then finished in your own time.*

You are provided with 10 drawing pins in a cup. Shake the pins onto the desk and count the number which land point upwards ("successes"). Repeat the procedure 8 times and contribute your data to the group frequency table.

(a) Calculate $\hat{p}$ =proportion of successes for the whole data.

(b) Calculate the probabilities of $0, 1, 2, \ldots, 10$ assuming a $B(10, \hat{p})$ distribution.

(c) Determine the expected frequency, E_i for each value of i using the fact that $E_i = Np_i$, where N is the total number of times that 10 pins were thrown onto the desk and p_i is the binomial probability of i successes.

(d) Use suitable diagrams to compare the observed frequencies with those expected from the binomial model.

(e) Comment on your results.

13. An important discrete probability distribution not covered in the lectures is the *discrete uniform distribution*. If X is a discrete uniform random variable with parameter n, then X takes each of the n possible values $1, 2, \ldots, n$ with equal probability. Therefore, the PMF for X is

$$\mathrm{P}(X = k) = \frac{1}{n}, \quad k = 1, 2, \ldots, n.$$

Show that

$$\mathrm{E}(X) = \frac{n+1}{2}, \quad \text{and} \quad \mathrm{Var}\,(X) = \frac{n^2 - 1}{12}.$$

14. In a squash tournament between three players A, B and C, each player plays the others once (*ie.* A plays B, A plays C and B plays C). Assume the following probabilities:

$$\mathrm{P}(A \text{ beats } B) = 0.6, \quad \mathrm{P}(A \text{ beats } C) = 0.7, \quad \mathrm{P}(B \text{ beats } C) = 0.6.$$

Assuming independence of the match results, calculate the probability that A wins at least as many games as any other player.

15. An insurance company classifies drivers as class X, Y or Z. Experience indicates that the probability a class X driver has at least one accident in any one year is 0.02, while the corresponding probabilities for Y and Z are 0.04 and 0.1 respectively. They have found that of the drivers who apply to them for cover, 30% are class X, 60% are class Y, and 10% are class Z. Assume that for each class of driver, accidents in subsequent years occur independently. However, *for a driver of unknown class, accidents in subsequent years will* ***not*** *occur independently*.

    (a) What the the probability of a new client having an accident in their first year?

    (b) What is the probability of a new client having no accidents in their first 5 years?

    (c) If a new client has no accidents in their first five years, what is the probability that they are a class X driver.

16. A player rolls a die. They stop at the fourth roll or when a six appears, whichever occurs first. Let X be the number of rolls. Find the PMF for X, and hence calculate $\mathrm{E}(X)$ and $\mathrm{Var}\,(X)$.

17. A medical experiment consists of subjecting a patient to a certain stimulus and recording whether or not they respond. The probability of a response is p. A sequence of 5 independent trials is called a series.

    (a) What is the probability that there is *exactly* one response in a series?

    (b) What is the probability that there is *at least* one response in a series?

    (c) What is the probability that, in 8 independent series, *exactly* two consist entirely of responses.

18. Whether or not certain mice are black or brown is genetically determined by a pair of alleles, each of which is either B or b. If both alleles are alike, then the mouse is said to be *homozygous* in that gene, and if they are different, the mouse is said to be *heterozygous*. So, for this gene, BB and bb are homozygous, and Bb and bB are heterozygous. A mouse is only brown if it is homozygous bb. If it is BB, Bb or bB, then it is black. Here, the brown allele b is said to be *recessive*, and the black allele, B, is said to be *dominant*.

    The offspring of a pair of mice receive one allele from each parent, and the allele that they get is randomly chosen from the two possible with equal probability. So, if a parent is heterozygous, it is equally likely to donate a B or a b allele to its offspring. If it is homozygous, the allele donated to the offspring is determined. Thus, a brown mouse, bb is certain to donate a b allele to its offspring. Whether or not the offspring is brown will then depend on the allele donated by the other parent.

    For this question, suppose that two *heterozygous* mice are mated.

    (a) What is the probability that a homozygous BB mouse results?

    (b) What is the probability that a black offspring results?

    In fact, when these two heterozygous mice are mated, a black mouse results. Call this black mouse "Mickey".

    (c) What is the probability that Mickey is homozygous BB?

(d) If Mickey is mated with a brown mouse, what is the probability that its offspring will be brown?

(e) If Mickey is mated with a heterozygous mouse, what is the probability that its offspring will be brown?

Now suppose that Mickey is mated with a brown mouse, resulting in seven offspring, all of which turn out to be black.

(f) Conditional on this information, what is the probability that Mickey is in fact homozygous BB?

19. Use Neave's tables or the analytic CDF (if available) to calculate the following probabilities. Be sure to show all of your working.

(a) For $X \sim B(20, 0.4)$, calculate $\mathrm{P}(X \leq 12)$.

(b) For $X \sim B(20, 0.4)$, calculate $\mathrm{P}(X < 10)$.

(c) For $X \sim B(15, 0.75)$, calculate $\mathrm{P}(X \geq 7)$.

(d) For $X \sim B(15, 0.75)$, calculate $\mathrm{P}(X < 10)$.

(e) For $X \sim Geom(0.1)$, calculate $\mathrm{P}(X \leq 10)$.

(f) For $X \sim Geom(0.2)$, calculate $\mathrm{P}(5 \leq X \leq 10)$.

(g) For $X \sim Geom(0.3)$, calculate $\mathrm{P}(X \geq 5)$.

(h) For $X \sim P(5)$, calculate $\mathrm{P}(X < 6)$.

(i) For $X \sim P(5)$, calculate $\mathrm{P}(3 \leq X \leq 8)$.

(j) For $X \sim P(10)$, calculate $\mathrm{P}(X > 15|X > 10)$.

20. The probability of winning a prize on the National lottery is 0.0186. Suppose that you decide to play one line per week. Let X be the number of weeks until you first win a prize.

(a) What is the distribution of X?

(b) What is $\mathrm{P}(X < 10)$?

(c) What is $\mathrm{P}(X < 20)$?

(d) What is the probability that you *don't win* a prize in the first 52 weeks (one year)?

(e) What is the probability that you *do win* a prize in the first 104 weeks (two years)?

21. Suppose that the number of times during a year that an individual catches a cold can be modelled by a Poisson random variable with an expectation of 4. Further suppose that a new drug based on Vitamin C reduces the expectation to 2 (but is still a Poisson distribution) for 80% of the population, but has no effect on the remaining 20% of the population. Calculate

(a) the probability that an individual taking the drug has 2 colds in a year if they are part of the population which benefits from the drug;

(b) the probability that an individual has 2 colds in a year if they are part of the population which does not benefit from the drug;

(c) the probability that a randomly chosen individual has 2 colds in a year if they take the drug;

(d) the conditional probability that a randomly chosen individual is in that part of the population which benefits from the drug given that they had 2 colds in a year during which they took the drug.

22. Each day a hospital makes available two beds to each of two surgeons (four beds in total). The demand each surgeon has for these beds is assumed to be independently Poisson with expectation 1. Calculate the probability

    (a) that the demand for beds for a particular surgeon exceeds those available;

    (b) that the demand for beds for at least one of the two surgeons exceeds the number available;

    (c) that the demand exceeds the number of available beds if the two surgeons decide to co-operate and pool their resources. *Hint: remember that the sum of independent Poissons is Poisson.*

23. Let X be a continuous random variable with probability density function defined by

$$f_X(x) = \begin{cases} kx, & 0 \le x \le 1, \\ k, & 1 \le x \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

    (a) What value of k is needed in order for this to define a valid probability density?

    (b) Sketch the density.

    (c) Evaluate the cumulative distribution function, $F_X(x)$.

    (d) Find the median, lower quartile and upper quartile of the distribution.

24. Let X be a continuous random variable with probability density function defined by

$$f_X(x) = \begin{cases} k\,x^2, & 0 \le x \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

    (a) What value must k take for this to be a valid density?

    (b) Derive the distribution function, $F_X(x)$.

    (c) Find the median, lower quartile and upper quartile of X.

25. For each of the following, compute the required probability using the specified method.

    (a) For $X \sim U(0, 1)$ use the PDF to calculate $\mathrm{P}(X > 0.6)$.

    (b) For $X \sim U(-3, 10)$ use the PDF to calculate $\mathrm{P}(0 < X < 5)$.

    (c) For $X \sim U(3, 10)$ use the CDF to calculate $\mathrm{P}(0 < X < 5)$.

    (d) For $X \sim Exp(0.1)$ use the PDF to calculate $\mathrm{P}(X < 3)$.

    (e) For $X \sim Exp(0.3)$ use the CDF to calculate $\mathrm{P}(1 < X < 5)$.

    (f) For $X \sim Exp(0.4)$ use the PDF to calculate $\mathrm{P}(X < 2)$.

    (g) For $X \sim Exp(3)$ use the CDF to calculate $\mathrm{P}(X > 1)$.

    (h) For $X \sim N(0, 1)$ use tables to calculate $\mathrm{P}(X > 0.5)$.

(i) For $X \sim N(0,1)$ use tables to calculate $\mathrm{P}(0.2 < X < 0.8)$.

(j) For $X \sim N(3,5^2)$ use tables to calculate $\mathrm{P}(2 < X < 10)$.

(k) For $X \sim N(2,9)$ use tables and linearly interpolate to calculate $\mathrm{P}(X < 4)$.

26. A continuous random quantity, X has PDF

$$f(x) = \max\{0, 1 - |x|\}, \qquad \forall x \in \mathbf{R}.$$

(a) Sketch $f(x)$.

(b) Calculate $\mathrm{P}(|X| > 0.5)$.

(c) Find $\mathrm{E}(X)$ and $\mathrm{Var}\,(X)$.

27. Let the continuous random quantity X have PDF

$$f(x) = \begin{cases} \lambda^2 x e^{-\lambda x}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

for some $\lambda > 0$.

(a) Verify that this does define a PDF.

(b) Find $\mathrm{E}(X)$.

(c) Find $\mathrm{Var}\,(X)$.

Hint: integration by parts may come in handy...

28. In the production of Nylon fibres, the extrusion process is interrupted from time to time by blockages occurring in the extrusion dyes. The time in hours between blockages, T, has an $Exp(0.1)$ distribution. Find the probabilities that

(a) a run continues for at least 20 hours;

(b) a run lasts less than 10 hours;

(c) a run continues for at least 20 hours, given that it has lasted 10 hours.

(d) If S is the time between interruptions measured in days, what is the distribution of S?

29. Leptin is an important appetite controlling substance. The natural logarithm of leptin concentration in men is known to have a normal distribution with expectation 0.7 and standard deviation 0.25.

(a) What is the probability that the log concentration is less than 0.4?

(b) What is the probability that the log concentration lies in the interval $(0.4, 0.8)$?

30. The width of a slot of a Duralumin forging is (in centimeters) normally distributed with $\mu = 0.9$ and $\sigma = 0.003$. All slots must lie in the range 0.9 ± 0.005 otherwise they are considered to be defective.

(a) What percentage of forgings will be defective?

(b) If it were possible to change σ, what is the maximum allowable value of σ that would result in no more than 1% defective forgings?