Probability problems for a 1st course
by Richard Cowan

A. Counting and simple probabilities based on equally-likely outcomes

1. 15 students enter the lift of a building at the ground floor. (a) How many different exit patterns are there on the higher floors? (b) If 9 are female and 6 male, how many different patterns are there if we distinguish between the males and females?

2. Ignoring the white ball, the game of Snooker has balls of 7 colours: red, yellow, green, brown, blue, pink and black. There are 15 indistinguishable red balls and 1 each of the other 6 colours. At the end of the game, all 21 balls are in the 6 distinguishable pockets around the table. How many distinguishable ways are there of arranging the 21 balls in the pockets?

3. A car is parked amongst n cars in a row, not at either end. On his return, the owner finds that exactly r of the n places are still occupied. What is the probability that both neighbouring places are empty?

4. Four Chan's and four Thompson's go to the cinema and sit 'randomly' in a block of 8 seats arranged in a line. (a) What is the probability that no one sits next to someone from his own family? (b) Afterwards they have dinner at a circular table for 8. What is the chance now? Part (b) raises the question of the definition of 'randomly'. In (a), it is clear that we mean that all distinct arrangements of 4C's and 4T's are equally likely. In (b) it depends upon our idea of 'distinct'. Solve (b) for two different definitions of 'distinct'. Firstly consider that different patterns which are equivalent under rotation are indeed distinct. Secondly consider that rotationally equivalent patterns are *not* distinct.

B. Practice with non-equally-likely outcomes

1. If n dice are thrown find the probability of the event A that the product of scores is divisible by 4.

2. The following is a simplified model representing the spread of measles within a family of three susceptible children $\alpha,\ \beta,\ \gamma$. If α catches measles there is a period during which β may get the disease from him, and the probability that β does in fact catch measles is p. Similarly γ may or may not get the disease in this period, <u>independently</u> of β and with the same probabilities p and $1-p$. Thus in the first stage none, one or two of the other children may get the disease. If the first or last event occurs, this completes the matter but, if for instance β gets the disease but not γ there is a further period during which γ may, with probability p, get the disease from β. Show that, given that one child gets measles from outside the family, the probabilities of 0, 1, 2 of the other children getting it are $(1-p)^2,\ 2p(1-p)^2,\ p^2(3-2p)$.

3. In a town of $n+1$ inhabitants, a persons tells a rumour to a second person, who in turn repeats it to a third person, etc. At each step the recipient of the rumour is

chosen at random from the n people available. Find the probability that the rumour will be told r times without: (a) returning to the originator, (b) being repeated to any person. Do the same problem when at each step the rumour is told by one person to a gathering of k randomly chosen people. (The first question is the special case $k = 1$.)

4. Suppose that an airplane engine will fail, when in flight, with probability $1-p$ independently from engine to engine. Suppose that the plane will make a successful flight if at least 50% if its engines remain operative. For what values of p is a 4-engine plane safer than a 2-engine plane?

5. Independent trials that result in a success with probability p and a failure with probability $1-p$ are called Bernoulli trials. Let P_n denote the probability that n Bernoulli trials result in an even number of successes (0 being considered as an even number). Show that

$$P_n = p(1 - P_{n-1}) + (1-p)P_{n-1} \qquad n \geq 1$$

and use this to prove (by induction) that

$$P_n = \frac{1 + (1-2p)^n}{2}.$$

6. Let Q_n denote the probability that in n tosses of a fair coin no run of 3 consecutive heads appears. Show that

$$\begin{aligned} Q_n =& \frac{1}{2}Q_{n-1} + \frac{1}{4}Q_{n-2} + \frac{1}{8}Q_{n-3} \quad n \geq 3 \\ Q_0 =& Q_1 = Q_2 = 1. \end{aligned}$$

Find Q_8.

7. In a given society, all married couples prefer families with at least one child of each sex. If a family continues to have children until there is one of each sex, find the probability that x children result. Assume that the probability of a boy is p, and that birth outcomes are independent. Ignore the chance of twins or other multiple births. What is the expected family size? For what value of p is this expectation a minimum?

8. Two people each toss a true coin n times. Prove that the probability of them getting the same number of heads is equal to the probability that the total number of heads that they get between them is n.
(*Hint:* Use the identity $\sum_{r=0}^{n} \binom{M}{r}\binom{n}{n-r} = \binom{M+N}{n}$.)

9. How many times would you expect to roll a fair die before all 6 sides have appeared at least once?

10. Independent trials are conducted. At each trial, there are 3 possible outcomes α, β or γ. These occur with probabilities p, q and r $(p+q+r=1)$. Let X be the number

of trials until the first α and Y be the number of trials until the first β. Find the joint probability mass function of X and Y.

11. There is a simple game between a banker and n gamblers. Each of the $n+1$ people independently sample a 'score' which has the rectangular distribution on the integers $\{1, 2, \cdots, m\}$. The banker compares his score with each gambler. If the banker has a score which is $\geq$ a particular gambler's score, then that gambler pays the banker \$1. If not, then the banker pays that gambler \$1.

    (a) What is the probability that the banker receives \$1 from a particular gambler?

    (b) What is the probability that the banker wins $\$n$ after one play of the game?

    Hint: Condition on the banker's score b. If your answer involves the expression

    $$\sum_{b=1}^{m} b^n,$$

    you may leave it without further effort to simplify the sum.

12. Two Aa individuals from generation 0 are crossed to produce many offspring of each sex. What is the probability that a randomly selected offspring will be (a) AA (b) Aa (c) aa? One male and one female offspring are selected randomly and mated to produce the 2nd generation. What is (d) the probability that the selected male is Aa and the selected female AA? (e) Fill in the following table of 'selection probabilities'.

    Female

		AA	Aa	aa
	AA			
Male	Aa			
	aa			

    Table: chances that the generation 1 breeding pair is as shown.

    What is the probability that a randomly selected member of generation 2 is (f) AA (g) Aa (h) aa?

13. The breeding scheme in problem 12 is continued. At each generation a breeding pair is randomly selected to create the next generation. What is the probability that a randomly selected individual in generation n is (a) AA; (b) Aa; (c) aa. Assume that at least 1 male and 1 female offspring result from each mating, so that one will always be able to select a breeding pair.

14. Problem 12 concerns an artifical breeding experiment; not a natural population. Yet it demonstrates a phenomenon seen in natural populations, *allele fixation.* It is possible that one allele, say A, is totally absent in the pair chosen. In generation 1 this event has probability $\frac{1}{16}$. If this event occurs for the *first* time in generation n, then allele A will be extinct from generation $n+1$ onward. We say that allele

a has become *fixed*. Let p_n = the probability that allele A is absent for the *first* time in the mating pair of generation n. Show that $p_2 = \frac{1}{25}$. From symmetry, the probability that a is absent for the first time in generation 2 is also $\frac{1}{25}$.

15. It is now known that *twins* are of two types, *fraternal*, resulting from the independent fertilisation of two eggs, or *identical* resulting from the splitting of one fertilised egg. This phenomenon was first suggested by Galton around 1875. Galton's theory of the fraternal/identical mechanism suggests a probability model of the situation. Let m be the probability that an egg produced receives a male sperm and let s be the probability that a fertilised egg splits into 2 in a manner that creates 2 'identical' offspring. Furthermore, let d be the probability that the female ovulation produces two eggs (otherwise one). Assuming that the egg(s) of an ovulation are fertilised, what is (a) the probability that a twin birth occurs? Given that a twin occurs, what is the probability that they are (b) of mixed sex, (c) 'identical'. The model allows for higher birth multiplicity too, triplets, quads etc. (d) what is the probability of > 2 in a birth. If $m = 0.52$, $d = 0.04$ and $s = 0.01$, evaluate the answers.

C. Continuous random variables

1. The presence of a certain disease is generally accompanied by a decrease in the concentration of a particular chemical in the blood. The variability in the concentration X of the chemical is described by the density function

$$f_1(x) = 4e^{-4x} \qquad (x \geq 0)$$

for sufferers from the disease, and by

$$f_2(x) = xe^{-x} \qquad (x \geq 0)$$

for normal persons. Accurate diagnosis of the disease is elaborate and costly, so the following preliminary screening process has been suggested. A person with $X \leq b$ is to be subjected to the further elaborate tests, but a person with $X > b$ is to be classified as normal. As measures of the limitations of such screening, calculate

(a) the probability that a sufferer from the disease will be misclassified as normal;

(b) the probability that a normal person will be unnecessarily subjected to further tests.

What should the "cut-off" value b be if the probability of misclassifying a sufferer as normal is to be set at 0.0001? With such a cut-off value, what is the probability that a normal person will be subjected to further tests?

2. The joint probability density function of X and Y is given by

$$f(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) \qquad 0 < x < 1,\ 0 < y < 2$$

(a) Verify that this is indeed a joint density function.

(b) Compute the distribution function of X.

(c) Find $P\{X > Y\}$.

(d) Find $P\{Y > \frac{1}{2} \mid X < \frac{1}{2}\}$.

3. Two points are selected randomly on a line of length L so as to be on opposite sides of the midpoint of the line. [In other words, the two points X and Y are independent random variables such that X is uniformly distributed over $(0, L/2)$ while Y is uniformly distributed over $(L/2, L)$.] Find the probability that the distance between the two points is greater than $L/3$.

4. In Problem 3 find the probability that the 3 line segments from 0 to X, from X to Y, and from Y to L could be made to form the three sides of a triangle. (Note that three line segments can be made to form a triangle if the length of each of them is less than the sum of the lengths of the others.)

5. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed continuous random variables. We say that a "record" occurs at time j $(j \leq n)$ if X_j is the highest of the variables up to that time. Let Z = number of "records" which occur up until time n. Show that $E(Z) = \sum_{j=1}^{n} 1/j$.

D. Estimation

1. A random sample of size n is drawn from a population. Each observation X_i $(i = 1, 2, \cdots, n)$ has pdf.

$$
\begin{aligned}
f(x) &= \lambda\, x^{\lambda-1} \qquad && 0 \leq x \leq 1 \\
&= 0 && \text{otherwise.}
\end{aligned}
$$

(a) What is the maximum likelihood estimate of λ?

(b) What is the moment estimate of λ?

2. An experimenter thinks that he is sampling from a Poisson (λ) distribution but he is told later that, due to a technical measurement error, no observation of zero is possible. That is, his random sample $X_1, X_2, \cdots, X_n$ is Poisson *conditioned* by the fact that $X_i \neq 0$. What is the pmf of the X_i? Write down the log likelihood for his data $x_1, x_2, \cdots, x_n$. Find an equation for the ML estimator of λ. Solve this equation (by numerical iteration) if $n = 20$ and $\Sigma x_i = 42$.

3. $X_1, X_2, \cdots, X_n$ is a random sample from a continuous distribution with p.d.f.

$$
\begin{aligned}
f(x) &= \frac{1}{\lambda}\sqrt{\frac{2}{\pi}}\, e^{-\frac{x^2}{2\lambda}} \qquad && x > 0 \\
&= 0 && \text{otherwise.}
\end{aligned}
$$

(a) Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Show that $\sqrt{\frac{\pi}{2}}\bar{X}$ is an unbiassed estimator of λ.

(b) Find the maximum likelihood estimator of λ.
Hint: It is not the same as the estimator in part (a).