Probabilities

Please attach a cover page (-10%).

Exercises 2.3–2.5 from Bulmer, Dover (1979), p. 26 **(75%).**

Problem 2.5 from Bulmer, Dover (1979), p. 28 **(25%).**

Traveling Salesman Problem (Extra Credit 100%):

Please provide your solutions in a Mathematica notebook, i.e., .nb file:

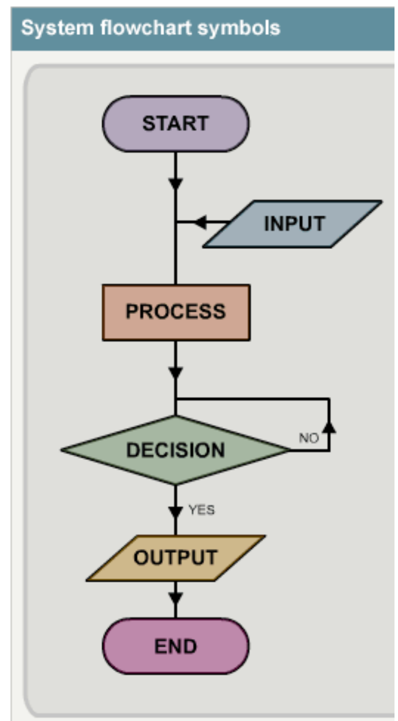
a) How many different possible ways are there for a traveler to visit once and only once each one of the four locations A, B, C and D? Write these down. Use Mathematica to write these down.

b) Assuming these four locations are Arches, Monument Valley, Salt Lake City and Zion, what is the driving distance that each route entails (according to the dataset Utah_Driving_Distances.txt)?

c) How many routes are at least twice as longer as the shortest route?

d) What is the probability for a route to be at least twice as longer as the shortest route?

e) Use these flowchart symbols to draw a flowchart for the brute-force solution of the traveling salesman problem, that is, for an input of a driving distances chart among n locations, output the shortest route or routes among all routes that visit each location once and only once.



Exercises

2.1. In Table 3 on p. 18 find the conditional proportion of (*a*) 3 on red die given 6 on white die, (*b*) 5 on white die given 4 on red die, and compare them with the corresponding marginal proportions.

2.2. Find the probability that the sum of the numbers on two unbiased dice will be even (*a*) by considering the probabilities that the individual dice will show an even number, (*b*) by considering the probabilities in Table 4 on p. 20.

2.3. Find the probabilities of throwing a sum of 3, 4, ..., 18 with three unbiased dice. (This problem was considered by Galileo.)

2.4. In 1654 the Chevalier de Méré, a well-known gambler and an amateur mathematician, put a problem to Pascal which gave rise to a famous correspondence between the latter and Fermat. De Méré's problem was that he had found by calculation that the probability of throwing a six in 4 throws of a die is slightly greater than $\frac{1}{2}$, while the probability of throwing double sixes in 24 throws of two dice is slightly less than $\frac{1}{2}$; it seemed self-evident to de Méré that these probabilities should be the same and he called the result "a great scandal which made him say haughtily that the theorems were not consistent and that the arithmetic was demented" (Smith, 1929). Repeat de Méré's calculations. (It is stated in many books that de Méré had noticed this very slight inequality in the chances in actual play. There is no basis for this improbable story.)

2.5. Three men meet by chance. What are the probabilities that (*a*) none of them, (*b*) two of them, (*c*) all of them, have the same birthday? [P. P. P. Hilary, 1965].

2.6. In a certain survey of the work of chemical research workers, it was found, on the basis of extensive data, that on average each man required no fume cupboard for 60 per cent of his time, one cupboard for 30 per cent and two cupboards for 10 per cent; three or more were never required. If a group of four chemists worked independently of one another, how many fume cupboards should be available in order to provide adequate facilities for at least 95 per cent of the time? [Certificate, 1959].

2.7. A yarborough at whist or bridge is a hand of 13 cards containing no card higher than 9. It is said to be so called from an Earl of Yarborough who used to bet 1000 to 1 against its occurrence. Did he have a good bet?

Problems

2.1. For an arbitrary number of events, $E_1, E_2, \ldots, E_n$, the general law of addition states that

$$P(E_1 \text{ or } E_2 \text{ or} \ldots \text{or } E_n) = \sum_i P(E_i) - \sum_{i<j} P(E_i \& E_j) + \sum_{i<j<k} P(E_i \& E_j \& E_k) - \ldots (-)^{n-1} P(E_1 \& E_2 \& \ldots \& E_n).$$

Prove this formula by induction. (The "or" is inclusive not exclusive.)

2.2. $E_1, E_2, \ldots, E_n$ are events; P is the probability that at least one of them occurs. Prove that

$$\sum_i P(E_i) - \sum_{i<j} P(E_i \;\&\; E_j) \leqslant P \leqslant \sum_i P(E_i).$$

An unbiased roulette wheel has 37 cups. What is the approximate probability that in 400 spins of the wheel there will be at least one cup which the ball never occupies? [Diploma, 1962]

2.3. An arbitrary number of events are said to be *pairwise independent* if

$$P(E_i \& E_j) = P(E_i) \,.\, P(E_j) \text{ for all } i \neq j.$$

They are said to be *mutually independent* if, in addition,

$$P(E_i \& E_j \& E_k) = P(E_i) \,.\, P(E_j) \,.\, P(E_k) \text{ for all } i \neq j \neq k$$
$$\vdots$$
$$P(E_1 \& E_2 \ldots \& E_n) = P(E_1) \,.\, P(E_2) \ldots P(E_n).$$

Show that pairwise independence does not imply mutual independence by supposing that two unbiased dice are thrown and considering the three events, odd number on first die, odd number on second die, odd sum on two dice.

2.4. Suppose that a genetic character is determined by a single pair of genes A and B so that there are three genotypes AA, AB and BB whose frequencies are x, y and z. Write down the expected frequencies of the different mating types, AA × AA, AA × AB, and so on, under random mating and hence show that the expected frequencies of these genotypes in the next generation under random mating are p^2, $2pq$ and q^2, where $p = x + \frac{1}{2}y$ and $q = \frac{1}{2}y + z$ are the frequencies of the A and B genes in the population. Thus the above equilibrium frequencies are attained after one generation of random mating. This law, which is fundamental in population genetics, was discovered independently by the English mathematician Hardy and the German physician Weinberg in 1908.

If the three genotypes are all distinguishable find the probabilities that (*a*) a pair of brothers, (*b*) a pair of unrelated individuals in a randomly mating population will appear the same when $p = q = \frac{1}{2}$.

2.5. The results of Table 7 on p. 25 can be explained on the assumption that the genes for flower colour and pollen shape are on the same chromosome but that there is a probability π that one of the genes will be exchanged for the corresponding gene on the other chromosome. If we denote the genes for purple or red flowers by P and p, and the genes for long and round pollen by L and l, then the hybrids from the cross considered will all be of the genotype PL/pl, the notation indicating that the P and L genes are on one chrosomome and the p and l genes on the other. When these hybrids are allowed to self-fertilise, there is a chance π that the L and l genes will interchange in one parent, giving Pl/pL; there are therefore really three mating types, $PL/pl \times PL/pl$, $Pl/pL \times PL/pl$ and $Pl/pL \times Pl/pL$, which occur with probabilities $(1-\pi)^2$, $2\pi(1-\pi)$ and π^2 respectively. Find the probabilities of the four possible phenotypes resulting from the experiment in terms of $\theta = (1-\pi)^2$.

2.6. Five players enter a competition in which each plays a game against each of the others. The two players in each game have an equal chance of winning it; a game must end in a win for one of the players who then scores a point. The competition is won by the player or players scoring most points. What is the probability that a particular player will (*a*) win the competition outright; (*b*) share in winning with other players? [Certificate, 1959]

2.7. In a game of bridge the declarer finds that he and his partner hold 9 spades between them when dummy is laid down. What are the chances that the remaining spades are distributed (*a*) 4-0, (*b*) 3-1, (*c*) 2-2 among his opponents?

2.8. A poker hand contains 5 cards. A flush is a hand all of the same suit, a straight is a hand in rank order (Aces counting high or low), and a straight flush is a hand all of the same suit in rank order; these categories are exclusive so that a straight flush does not count as either a flush or a straight. What are the chances of dealing (*a*) a flush, (*b*) a straight, (*c*) a straight flush?

CHAPTER 3

RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

So far we have been considering the probabilities of quite arbitrary events. Very often, however, the events in which we are interested are numerical. Such a numerical variable which takes different values with different probabilities is called a *random variable*. There are two types of random variable: *discrete* variables, such as the number of petals on a flower, which arise from counting and which can in consequence only take the integral values 0, 1, 2, ...; and *continuous* variables, such as the length of a petal or the weight of a man, which result from measuring something and can therefore take any value within a certain range. We will now discuss them in turn.

DISCRETE RANDOM VARIABLES

As an example of a discrete random variable, consider the data in Table 8 on the sizes of 815 consecutive litters of rats. We can imagine that if more and more litters of the same species were counted under the same conditions the relative frequencies in the last row would tend to stable limiting values or probabilities. Litter size may thus be thought of as taking different values with different probabilities, that is to say as a random variable.

Table 8 is said to represent the *frequency distribution* of litter size since it shows with what frequencies litters of different sizes are distributed over the possible values of litter size.

TABLE 8

Frequency distribution of litter size in rats (King, 1924)

Litter size	1	2	3	4	5	6	7	8	9	10	11	12	Total
No. of litters	7	33	58	116	125	126	121	107	56	37	25	4	815
Relative frequency	·01	·04	·07	·14	·15	·15	·15	·13	·07	·05	·03	·01	1·00