PROBABILITY

Brailey Sims

Department of Mathematics
The University of Newcastle

THEY USE STATISTICS LIKE A DRUNKEN MAN USES A LAMP-POST --- FOR SUPPORT NOT ILLUMINATION   Andrew Lang 1844–1912

PROBABILITY

Brailey Sims

Department of Mathematics
The University of Newcastle

2/3 unit syllabus: (1) *Random experiments,*
equally likely outcomes,
probability of a given outcome.

(2) *Sum and product results.*

(3) *Experiments involving successive outcomes*
- tree diagrams.

3 unit syllabus: (4) *Systematic enumeration in finite sample spaces*
- permutations and combinations
- binomial probabilities and the binomial distribution.

BASIC FORMALISM

$$\text{EXPERIMENT} \longrightarrow \text{OUTCOME}$$

$$\textit{RANDOM}\ \text{EXPERIMENT} \longrightarrow \text{OUTCOME, } \textit{which we cannot be certain of in}$$
$$\textit{advance.}$$

The possible outcomes form a set S known as the **sample space**.

Example. Experiment: Out of two people *Betty* and *Jim* who will be alive in the year 2001?

Sample space (= set of all possible outcomes) is

$$S = \{ BJ, B'J, BJ', B'J' \},$$

where B denotes *Betty* is alive, B' denotes *Betty* is dead, etc.

A subset **E** of the sample space is termed an **event**.

Example: $E = \{ B'J, BJ' \}$ corresponds to the event that only one of our two people is alive in 2001.

An event is said to occur whenever the outcome of our experiment is one of its elements.

Two events E and F are **mutually exclusive** if they cannot occur concomitantly. That is, if as subsets they are disjoint; $E \cap F = \emptyset$.

1

Example: The event $\mathbf{E} = \{\, B'J,\ BJ' \,\}$, that only one of our two friends are alive in 2001, and the event $\mathbf{F} = \{\, B'J' \,\}$, that both are deceased by 2001, are mutually exclusive.

The **union** of two events, $\mathbf{E} \cup \mathbf{F}$, corresponds to the event that either the event $\mathbf{E}$ **or** the event $\mathbf{F}$ occurred.

The **intersection** of two events, $\mathbf{E} \cap \mathbf{F}$, corresponds to the event that the event $\mathbf{E}$ **and** the event $\mathbf{F}$ have occurred together.

Henceforth we will consider only experiments for which the number of possible outcomes is finite. That is, the number of elements in our sample space, $\#\mathbf{S}$, is a finite number n.

With each event $\mathbf{E}$ ($=$ subset of $\mathbf{S}$) we wish to associate a number, $p(\mathbf{E})$, the *probability* of $\mathbf{E}$, which measures the likelihood of that event occurring.
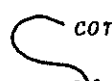
We require:

- $0 \leq p(\mathbf{E})$

- $p(\mathbf{S}) = 1.$

- If $\mathbf{E}$ and $\mathbf{F}$ are mutually exclusive events; that is, $\mathbf{E} \cap \mathbf{F} = \emptyset$, then
$$p(\mathbf{E} \cup \mathbf{F}) = p(\mathbf{E}) + p(\mathbf{F}).$$
$\curvearrowright$ *probability of* $\mathbf{E}$ *or* $\mathbf{F}$

From this we immediately deduce the useful result:

complement of $\mathbf{E}$

$$p(\overline{\mathbf{E}}) = 1 - p(\mathbf{E}).$$

probability of **not** $\mathbf{E}$

We also see that all events have probabilities between 0 and 1. Intuitively, the larger the probability of an event the more likely that event is to occur. In our context, $p(\mathbf{E}) = 0$ means $\mathbf{E}$ cannot occur, while $p(\mathbf{E}) = 1$ means that $\mathbf{E}$ is certain to occur.

Example: If each outcome of the experiment is **equally likely** to occur, then *on average* each outcome would occur one n'th of the time; that is, with probability $1/n$, and so in this case for any event $\mathbf{E}$ we would have
$$p(\mathbf{E}) = \frac{\#\mathbf{E}}{\#\mathbf{S}},$$
and we can readily confirm that for p calculated in this way the three requirements listed above are satisfied.

Note: For many experiments we would expect outcomes to be equally likely, for example the number appearing on the upper most face of a dice after it has been rolled, however in other situations, like our example of who is alive in 2001 (or whether a dropped brick will crack, or remain unbroken), we would not expect the outcomes to be equally likely.
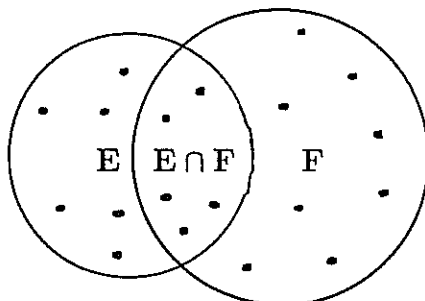
For two events **E** and **F** we have

$$p(\text{F given E}) \; = \; \frac{p(\text{F} \cap \text{E})}{p(\text{E})},$$

where $p(\text{F given E})$ represents the probability of **F** knowing that **E** has occurred [or, the *conditional probability* of **F** given **E**].

In the case of equally likely outcomes this becomes;

$$\begin{aligned} p(\text{F given E}) &= \frac{p(\text{F} \cap \text{E})}{p(\text{E})} \\ &= \frac{\#(\text{F} \cap \text{E})/\#\text{S}}{\#\text{E}/\#\text{S}} \\ &= \frac{\#(\text{F} \cap \text{E})}{\#\text{E}}, \end{aligned}$$

as expected.

Strictly speaking the identity given above serves to define $p(\text{F given E})$, but in reality it is often possible to assign values to $p(\text{E})$ and $p(\text{F given E})$ and hence calculate $p(\text{E} \cap \text{F})$.

Example: Michele buys 5 tickets in a raffle in which 20 tickets are sold. If one ticket is drawn out for first prize and then another for second prize, what is the chance that Michele wins both prizes.

Let **W1** be the event that *Michele wins first prize*, and let **W2** be the event that *Michele wins second prize*, then

$$p(\text{W1}) \; = \; 5/20$$

and

$$p(\text{W2 given W1}) \; = \; 4/19,$$

since there are 19 tickets remaining in the draw for second prize 4 of which are Michele's.

So the probability that Michele wins both first and second prize is

$$p(W1 \cap W2) = p(W2 \text{ given } W1)p(W1) = \frac{4}{19} \cdot \frac{5}{20} = \frac{1}{19}.$$

Tree diagrams are a useful tool in analysing these type of situations, where events are decomposed into mutually exclusive 'subevents'.

Example: If we denote by W that *Michele wins* and by W' that *Michele losses* then a tree diagram for our previous example of a raffle is:

first prize   second prize
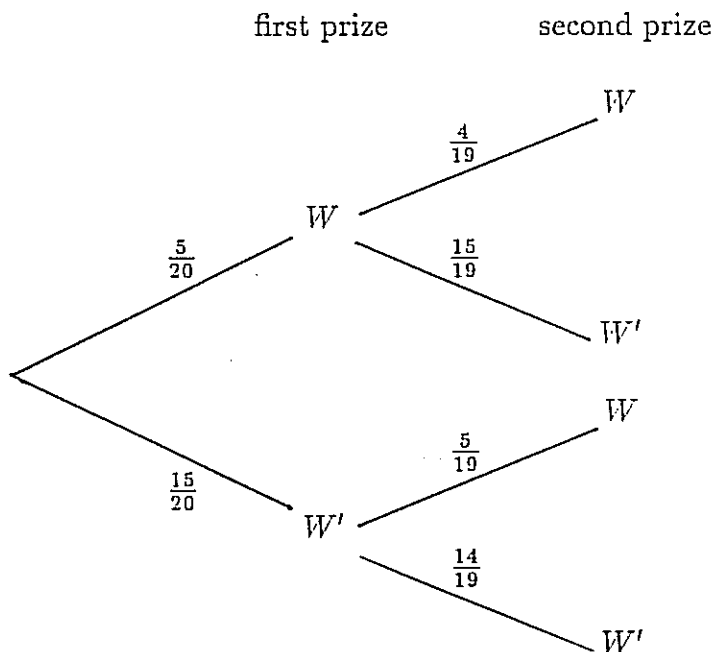
Using this we can readily determine the probability that:

(i) Michele does not win a prize.

(ii) Michele wins at least one prize.

(iii) Michele wins exactly one prize.

Each 'stem' of the tree represents an event, and to determine the likelihood of that event we simply multiply the probabilities on each of its 'stages', thus

(i) the event *Michele does not win a prize* corresponds to the bottom stem and so
$$p(\textit{Michele does not win a prize}) = \frac{15}{20} \cdot \frac{14}{19},$$

then

4

(ii)

$$p(Michele\ wins\ at\ least\ one\ prize) = 1 - p(Michele\ does\ not\ win\ a\ prize)$$
$$= 1 - \frac{15}{20} \cdot \frac{14}{19},$$

and

(iii)

$p(Michele\ wins\ exactly\ one\ prize)$
$= p(Michele\ wins\ first\ prize\ only) + p(Michele\ wins\ second\ prize\ only)$
$= p(W1)p(\overline{W2}\ given\ W1) + p(\overline{W1})p(W2\ given\ \overline{W1})$
$= \dfrac{5}{20} \cdot \dfrac{15}{19} + \dfrac{15}{20} \cdot \dfrac{5}{19} = \dfrac{15}{38}.$

Note: This example is a simplified form of a question in the 1987 2/3 unit paper, which in turn is similar a question in the 1992 paper.

Two events E and F are independent if

$$p(F\ given\ E) = p(F),$$

or equivalently, if

$$p(F \cap E) = p(F)p(E).$$

That is, if the likelihood of F occurring is not affected by knowing that E has occurred.

Exercise: Although it is not apparent from the definition, show that independence is symmetric in E and F. That is, show that if $p(F\ given\ E) = p(F)$. then $p(E\ given\ F) = p(E)$*.

Example: Suppose we draw one card from a normal deck of playing cards.
Let

A be the event that an ace is drawn,

and let

R be the event that a red card (heart, or diamond) is drawn.

Assuming each of the 52 cards is equally likely to be drawn, we have

$$p(A) = 4/52 = 1/13,$$

while

$$p(A\ given\ R) = 2/26 = 1/13.$$

So drawing an ace and drawing a red card are independent events.

---

* While optional this exercise is probably the only *real* mathematics in the section.

Note: This seems to conflict with our intuitive idea that two events are independent if they are not causally connected, although the latter is of course often the reason for our assuming events to be independent.

Example: If statistics indicate that 20% of people in Betty and Jim's age group will have died by 2001 then, in the absence of any other information, it is reasonable to suppose that

$$\overline{B} = \text{the event that } Betty \text{ is dead by } 2001 = \{B'J, B'J'\}$$

has a 1 in 5 (20%) chance of occuring. That is,

$$p(\overline{B}) = \frac{1}{5}.$$

Similarly

$$p(\overline{J}) = p(\{BJ', B'J'\}) = \frac{1}{5}.$$

and if we assume Betty's and Jim's deaths are independent events*, then the probabilities for our earlier example would be;

$$p(B'J') = p(\overline{B} \cap \overline{J}) = \frac{1}{5} \cdot \frac{1}{5} = \frac{1}{25},$$

$$p(B'J) = \frac{1}{5} \cdot \frac{4}{5} = \frac{4}{25}, \qquad [\, p(J) = 1 - p(\overline{J}) \,]$$

$$p(BJ') = \frac{4}{5} \cdot \frac{1}{5} = \frac{4}{25},$$

$$p(BJ) = \frac{4}{5} \cdot \frac{4}{5} = \frac{16}{25}.$$

From this we deduce, for example, that;

probability only one will be alive in 2001

$$= p(\{B'J, BJ'\})$$
$$= p(\{B'J\} \cup \{BJ'\})$$
$$= p(B'J) + p(BJ')$$
$$= \frac{4}{25} + \frac{4}{25} = \frac{8}{25}.$$

Similarly,

probability at least one will be alive in 2001

$$= p(\{B'J, BJ', BJ\})$$
$$= p(B'J) + p(BJ') + p(BJ)$$
$$= \frac{4}{25} + \frac{4}{25} + \frac{16}{25} = \frac{24}{25}.$$

---

* Can you think of circumstances under which this might not be an appropriate assumption?

6

Note: This last probability could have been calculated more efficiently using;

probability at least one will be alive in 2001

$$= 1 - (probability\ neither\ are\ alive\ by\ 2001)$$
$$= 1 - p(B'J')$$
$$= 1 - \frac{1}{25} = \frac{24}{25}.$$

Example: (1990 2/3 unit paper) Two dice are tossed, and the maximum of the two uppermost faces is recorded as a score.

   (i) Find the probability that a score of 1 is recorded in a single throw of the two dice.

   (ii) Find the probability that the scores 1, 1, 1 are recorded in three tosses of the two dice.

   (iii) Find the probability that a score of 6 is recorded in a single throw of the two dice.

The *experiment* consists of tossing two dice and the set of possible outcomes (*sample space*) is

$$\{(1,1),\ (1,2),\ \cdots,\ (1,6),$$
$$(2,1),\ (2,2),\ \cdots,\ (2,6),$$
$$\cdots$$
$$(6,1),\ (6,2),\ \cdots,\ (6,6)\}.$$

We suppose that all 36 outcomes are equally likely (justified if we assume that in a toss of the two dice the number appearing on the uppermost face of one dice is independent of that appearing on the other, and that any one of the six faces of a dice is equally likely to be uppermost; that is, the dice are 'fair').

   (i) Let **S1** be the event that *a score of 1 is recorded in a single throw of the two dice.* Then

$$S1 = \{(1,1)\}$$

and

$$p(S1) = 1/36.$$

   (ii) Assuming the outcome of one toss is independent of the result of other tosses we have

p(scores 1, 1, 1 are recorded in three tosses of the two dice) $= (1/36)^3.$

7

Strictly speaking we are here considering a new *experiment*; three tosses of the two dice, and new *sample space* consisting of ordered triplets of ordered pairs of the numbers 1, 2, 3, 4, 5, 6; for example, $((1, 3), (2, 5), (5, 5))$, of which there are $(36)^3$ in all. Our event, *scores 1, 1, 1 are recorded in three tosses of the two dice*, corresponds to just one point, $((1, 1), (1, 1), (1, 1))$, of the sample space.

(iii) Let S6 be the event that *a score of 6 is recorded in a single throw of the two dice*. Then

$$S6 = \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5)$$
$$(6, 6), (5, 6), (4, 6), (3, 6), (2, 6), (1, 6)\}$$

and so

$$p(S6) = 11/36.$$

In general we may attempt to calculate probabilities in this way; by decomposing our event into suitable subevents and applying the rules found above. This frequently requires us to 'count' (*enumerate*) the elements in large indirectly prescribed sets (events).

ENUMERATIONS

We examine selecting r objects from n, and the application of these ideas to probability.

The 'selection' of r objects from n may be:

with or without replacement — depending on whether or not an object once selected is available for reselection.

ordered or unordered — depending on whether or not the order (*first, second, third, ···*) in which the objects are selected is important.

This leads to four 'types' of selection. Deciding which appropriately describes a given situation is often the most difficult task.

Example: The number of teams of 11 players that can be chosen from a pool of 25 cricketers equals the number of ways of making an *unordered selection of 11 objects from 25 without replacement*.

The number of three digit numbers containing only odd digits equals the number of *ordered selections with replacement of 3 objects from 5*.

8

The following table summarizes the number of ways r objects can be selected from n in each of the 'four' cases.

	with replacement	without replacement
ordered	n^r	^nP_r
unordered	*	$\binom{n}{r}$†

Example: The number of different cricket teams which can be formed from a pool of 25 players is $\binom{25}{11} = 4,457,400$, a truly uninteresting number.

There are $5^3 = 125$ three digit numbers containing only odd digits.

Example: (1990 3/4 unit paper) Three identical blue marbles and four identical yellow marbles are arranged in a row.

(i) How many different arrangements are possible?

(ii) How many different arrangements of just five of these marbles are possible?

(i) Each arrangement corresponds to a selection of three distinct (that is, without replacement) positions, to be occupied by the blue marbles, out of the seven positions in the row. Thus there are $\binom{7}{3} = 35$ possible arrangements.

(ii) Our arrangement may be any one of the following mutually exclusive possibilities.

It may contain 3 blue marbles; and, by reasoning similar to that for (i), there are $\binom{5}{3}$ such arrangements, or

it may contain 2 blue marbles; there are $\binom{5}{2}$ such arrangements, or

it may contain 1 blue marbles; there are $\binom{5}{1}$ such arrangements.

Thus in this case the number of different arrangements is

$$\binom{5}{3} + \binom{5}{2} + \binom{5}{1} = 5 \cdot 4 \cdot 3 / 3 \cdot 2 \cdot 1 + 5 \cdot 4 / 2 \cdot 1 + 5 = 25.$$

---

* while not part of the syllabus, the more adventurous of you might like to contemplate why the answer here is $\binom{n+r-1}{r}$.

† The alternative notation ^nC_r was once common, though it should now be considered obsolete.

Example: (1992 3/4 unit paper) Five players are selected at random from four sporting teams each consisting of ten players numbered 1 to 10.

    (i) What is the probability that of the five players selected three are numbered 6 and two are numbered 8?

    (ii) What is the probability that of the five players selected at least four are from the same team?

The total number of selections of five players from the 40 is $\binom{40}{5}$.
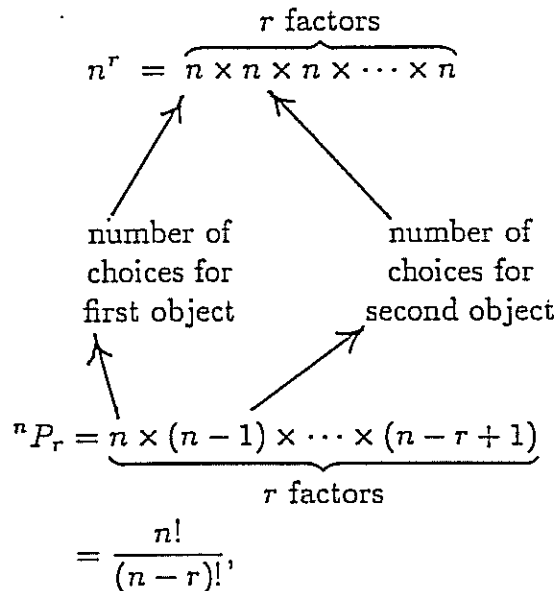
(i) The number of selections with three players numbered 6 and two players numbered 8 is $\binom{4}{3} \times \binom{4}{2}$, the number of ways the three number 6's can be selected from among the four number 6 players (one in each team) multiplied by the number of ways the two number 8's can be selected. So, the probability of such a selection is

$$\frac{\binom{4}{3}\binom{4}{2}}{\binom{40}{5}} = \frac{1}{27417} \doteqdot 0.000036.$$

(ii) The number of selections with at least four players from the same team is the number of ways a team can be selected multiplied by the sum of the number of selections with four from that team and one from one of the other three and the number of selections with all five from the selected team. So the probability of such a selection being made is

$$\frac{\binom{4}{1}\left(\binom{10}{4}\binom{30}{1} + \binom{10}{5}\binom{30}{0}\right)}{\binom{40}{5}} = \frac{28}{703} \doteqdot 0.04.$$

To see why the results tabulated above are correct note that:

$$n^r = \overbrace{n \times n \times n \times \cdots \times n}^{r \text{ factors}}$$

number of choices for first object   number of choices for second object $\cdots$

$$^nP_r = \underbrace{n \times (n-1) \times \cdots \times (n-r+1)}_{r \text{ factors}}$$

$$= \frac{n!}{(n-r)!},$$

10

where $m!$ is m factorial equal to $m \times (m-1) \times (m-2) \times \cdots \times 3 \times 2 \times 1$.

While each unordered choice of r distinct (as will be the case for selection without replacement) objects from n can be ordered in $^{r}P_r = r!$ ways, so $^{n}P_r = \binom{n}{r} \times r!$ or

$$
\begin{aligned}
\binom{n}{r} &= \frac{^{n}P_r}{r!} \\
&= \frac{n!}{(n-r)!\,r!} \\
&= \frac{n(n-1)(n-2)\cdots(n-r+1)}{r(r-1)(r-2)\cdots 3 \cdot 2 \cdot 1}
\end{aligned}
$$

Historical note: The 'P' in $^{n}P_r$ comes from an old meaning of the word *permutations* as the ways of selecting from a larger group. Similarly the 'C' in the archaic notation $^{n}C_r$ for $\binom{n}{r}$ comes from the old meaning of *combination* as a group chosen from a larger one without regard for order.
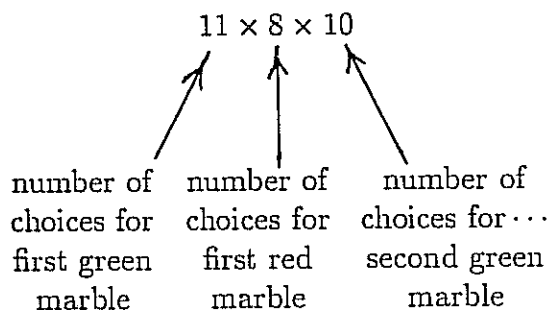
The ideas behind these derivations underlie the solution of many other problems.

Example: (1990 2/3 unit paper) A box contains 8 red and 11 green marbles. Three marbles are randomly selected one at a time and without replacement. What is the probability that the selection is

$$green, \ red, \ green$$

in that order?

There are $^{19}P_3$ equally likely ways of making an ordered selection of 3 objects (marbles) from 19 ($= 8 + 11$) without replacement, of which

$$11 \times 8 \times 10$$

number of	number of	number of
choices for	choices for	choices for $\cdots$
first green	first red	second green
marble	marble	marble

correspond to a selection of *green, red, green* in that order.

Thus the required probability is

$$\frac{11 \times 8 \times 10}{^{19}P_3} = \frac{11 \times 8 \times 10}{19 \times 18 \times 17} \doteqdot 0.15.$$

Problems involving the 'ordered' arrangements of n objects around a circle can be reduced to the arrangements of $n-1$ objects in $n-1$ places by fixing one of the original n objects. Thus there are $^{n-1}P_{n-1} = (n-1)!$ such arrangements.

Example: (**1986 4 unit paper**) A committee of 4 women and 3 men are to be seated at random around a circular table with 7 seats. What is the probability that all the women will be seated together.

To count the number of arrangements with the women seated together, consider the women as forming a 'block'. There are $^4P_4 = 4!$ arrangements of the 4 women within the block and 3! arrangements of the 4 objects; 3 men and the block of women, around the table. Thus there are $4! \cdot 3!$ arrangements with the women seated together. Altogether there are 6! arrangements of the 7 people around the table. Hence, the probability of the women sitting together is

$$\frac{4! \cdot 3!}{6!} = \frac{1}{5}.$$

Selecting r distinct objects out of n without regard to order is the same as choosing the $n-r$ objects to be left out of the selection, so we have

$$\binom{n}{r} = \binom{n}{n-r},$$

a fact which is easily confirmed algebraically.

Another useful fact comes from the observation that making an unordered selection of r objects out of n without replacement can be broken into two mutually exclusive events by first choosing a (favorite) object and then considering those selections which include the chosen object and those which don't. In this way we obtain **Pascal's identity**

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r},$$

number of ways if
chosen object is
in the selection

number of ways if
chosen object isn't $\cdots$
in the selection

an identity which provides an effective way of tabulating the numbers $\binom{n}{r}$ known as Pascal's triangle.

n \ r	0	1	2	3	4
0	1				
1	1	1			
2	1	2	1		
3	1	3	3	1	
4	1	4	6	4	1

12

In the expansion of the *binomial* $(a + b)^n$ the coefficient of $a^{n-r}b^r$ is $\binom{n}{r}$, since when multiplying out we get an $a^{n-r}b^r$ by selecting r of the n factors to contribute a b to the product and letting the remaining $n - r$ factors contribute an a, and this selection can be done in $\binom{n}{r}$ ways. Thus we have the **binomial theorem**:

$$(a + b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + b^n.$$

For this reason the number $\binom{n}{r}$ is sometimes referred to as a **binomial coefficient**.

Repeated trials (Bernoulli trials).

Consider n independent repetitions (trials) of a random experiment which has two possible outcomes, nominally 'success' and 'failure'. Suppose the probability of *success* is p and that of *failure* is $q = 1 - p$. We ask, what is the probability that precisely r of our n trials are *successful*?

There are $\binom{n}{r}$ ways the r 'successful' trials may be selected, each of which corresponds to an event mutually exclusive of the others, and the probability of each such event occurring is, by the independence of the trials, $p^r q^{n-r}$, since r have an outcome of *success* and $n - r$ must have an outcome of *failure*. Thus altogether

$$p(\text{precisely } r \text{ successes out of } n \text{ trials}) = \binom{n}{r}p^r q^{n-r}.$$

A result known for obvious reasons as a **binomial probability**.

Example: **(1990 3/4 unit paper)** A multiple choice examination has 10 questions, each with four possible answers only one of which is correct. What is the probability of answering exactly six questions correctly by chance alone?

Here we take our experiment to be answering a question and deem its outcome to be *success* if our answer is correct, which by chance alone has a probability of $p = \frac{1}{4}$. Thus, the probability of answering exactly six questions correctly; that is, of having precisely 6 'successes' out of 10 trials of our experiment, is

$$\binom{10}{6}\left(\frac{1}{4}\right)^6\left(1 - \frac{1}{4}\right)^{10-6} = \binom{10}{4}\frac{3^4}{4^{10}} \doteq 0.016.$$

Exercise: In our current context what if anything is the significance of the identity,

$$1 = (1 - p + p)^n = (q + p)^n = q^n + \binom{n}{1}pq^{n-1} + \binom{n}{2}p^2 q^{n-2} + \cdots + p^n?$$

13

I will conclude with a few challenges.

Exercise: From among nine mathematicians, five of whom are analysts, a team of five is to be chosen to go on an Antarctic expedition. What is the probability that the team will contain at least two analysts?

[Answer: 121/126?]

Exercise: In 10 tosses of a fair coin what is the probability of a run of five or more heads occurring?

[Hint: Consider the mutually exclusive cases; the run starts on the first toss, on the second toss, on the third toss, $\cdots$.

Answer: 7/64?]

Exercise: Here is one to think about (and inflict on your friends), allegedly due to Erwin Schrödinger, one of the founders of quantum mechanics.

An 'infinite' deck of cards has;

2 cards with the number 1 on one side and 2 on the other side.

4 cards with the number 2 on one side and 3 on the other side.

8 cards with the number 3 on one side and 4 on the other side.

&c.

A card is drawn at random and held up so that player A can see only one side and player B can see only the other side.

A player wins if the number on his side is smaller than the number on the other side.

If A sees a 1 he is sure to win. If he sees $n > 1$, of the $2^{n-1} + 2^n$ cards with an n on them 2^n have a larger number on the other side. Thus, A's chances of winning are at least $2^n/(2^{n-1} + 2^n) = 2/3$. But, by exactly the same reasoning this is also B's chances of winning ?

Exercise: (The car–and–goats fiasco) The following question has caused considerable consternation ever since it was asked and answered(?) in 1990 by Marilyn vos Savant in *Parade*.

> You are the contestant in a popular TV game show. One of three doors hides a car (all three are equally likely) and the other two hide goats. You choose Door 1. The host, who knows where the car is, then opens one of the other two doors to reveal a goat, and asks whether you wish to switch your choice. Say the host opened Door 3; should you switch to Door 2?

Marilyn said yes, arguing that if the car is actually behind Door 1 (probability 1/3) then when you switch you lose; but if it is at Door 2 or 3 (probability 2/3) then the host's revelation of a goat shows you how to switch and win. So the chance you win by switching is 2/3!

Many others have argued that with Door 3 now eliminated the other two are equally likely, so there is no particular advantage in switching.

What do YOU think? [Be careful, it's not even clear that there is a right answer.]