# CHAPTER 4: PROBABILITY

THE NOTION OF **RANDOM**NESS IS QUITE CLEAR, IT RELATES TO A SITUATION WHICH CAN RESULT IN ONE OF SEVERAL POTENTIAL OUTCOMES, BUT IT'S VIRTUALLY IMPOSSIBLE TO PREDICT WHICH ONE (WILL I BE HIRED FOR A JOB I HAVE APPLIED FOR).

A **RANDOM EXPERIMENT** IS A (CAREFULLY DESIGNED) PROCEDURE, HAVING A RANDOM OUTCOME AND (USUALLY) SOME SPECIFIC AIM.
THE MOST IMPORTANT EXAMPLE (TO US) IS RANDOM SAMPLING, BUT IN THIS CHAPTER WE WILL ALSO STUDY SIMPLE EXPERIMENTS SUCH AS ROLLING DICE, FLIPPING COINS, DRAWING MARBLES FROM A BOX, DEALING CARDS.

A COMPLETE <u>LIST</u> OF POSSIBLE OUTCOMES OF A SPECIFIC RANDOM EXPERIMENT IS CALLED ITS **SAMPLE SPACE** (E.G. 1, 2, 3, 4, 5, 6 FOR A SINGLE ROLL OF A DIE).

ANY <u>SUBSET</u> OF THIS IS CALLED AN **EVENT** (GETTING AN <u>EVEN</u> NUMBER OF DOTS).

IN SOME (SYMMETRIC, REGULAR) SITUATIONS, ALL POSSIBLE OUTCOMES ARE **EQUALLY LIKELY** (IN OUR EXAMPLE, EACH POSSIBLE OUTCOME HAS THE PROBABILITY OF 1/6).

SIMILARLY, TWO SIDES OF A COIN MUST HAVE THE SAME PROBABILITY OF ½ EACH.

FINALLY, IN RANDOM INDEPENDENT SAMPLING FROM A POPULATION (OF PEOPLE), EACH PERSON HAS THE SAME CHANCE OF BEING SELECTED - THUS ALL SAMPLES OF SIZE $n$ ARE ALSO EQUALLY LIKELY.

OTHERWISE (NO SUCH SYMMETRY, E.G. FLIPPING A TACK), WE NEED TO PERFORM THE CORRESPONDING EXPERIMENT, REPEATEDLY AND **INDEPENDENTLY** (PAST OUTCOMES CANNOT INFLUENCE THE NEXT ONE) MANY TIMES, KEEPING TRACK OF WHETHER THE PARTICULAR EVENT DID OR DID NOT OCCUR.

THE **RELATIVE FREQUENCY** (PERCENTAGE) OF THE EVENT'S OCCURRENCE IS A (RANDOM) NUMBER WHICH YIELDS THE PROBABILITY OF THE EVENT <u>IN THE LIMIT</u> OF <u>INFINITELY</u> MANY **TRIAL**S (LAW OF LARGE NUMBERS).

THIS MEANS THAT, IN MOST SITUATIONS (NO SYMMETRY), WE WILL NEVER KNOW THE EXACT ANSWER, BUT ONLY AN APPROXIMATE **ESTIMATE** OF IT.

---

PROBABILITY OF AN EVENT IS THE SUM OF THE PROBABILITIES OF THE INDIVIDUAL OUTCOMES THE EVENT CONSISTS OF. FOR AN EXPERIMENT WITH EQUALLY LIKELY OUTCOMES, THIS LEADS TO THE FOLLOWING SIMPLE RULE: PROBABILITY OF AN EVENT EQUALS

$$\frac{\#\ OF\ OUTCOMES\ THE\ EVENT\ CONSISTS\ OF}{TOTAL\ (SAMPLE\ SPACE)\ \#\ OF\ OUTCOMES}$$

IN A SINGLE EXPERIMENT, WE CAN CLEARLY DEFINE MORE THAT ONE EVENT, E.G. $A$ IS GETTING AN EVEN NUMBER, $B$ IS GETTING A NUMBER BIGGER THAN 3, WHEN ROLLING A DIE.

BASED ON TWO EVENTS ($A$ AND $B$), WE CAN DEFINE A FEW DERIVED (COMPOUND) EVENTS, NAMELY:

*NOT A*: THE **COMPLEMENT** OF *A* (ANYTHING BUT *A)*.

*A AND B*: THE 'OVERLAP' (**INTERSECTION**) OF *A* AND *B* (THE SUBSET OF NUMBERS <u>BOTH</u> EVEN AND BIGGER THAN 3, I.E. 4, 6).

*A OR B*: THE **UNION** (JOINING THEM TOGETHER) OF THE TWO (I.E. NUMBERS WHICH ARE EITHER EVEN OR BIGGER THAN 3, OR <u>BOTH</u>: 2, 4, 5, 6).

THERE ARE SOME SIMPLE PROBABILITY RULES RELATING TO THESE:

$$P(\textit{NOT } A) = 1 - P(A)$$

$$P(A \textit{ AND } B) = P(A) @ P(B)$$

BUT ONLY WHEN *A* AND *B* ARE <u>INDEPENDENT</u> (IN OUR EXAMPLE, THEY WERE <u>NOT</u>), OR (THE GENERAL SITUATION):

$$P(A \textit{ AND } B) = P(A) @ P(B \textit{ GIVEN } A)$$

WHERE P(*B GIVEN A*) IS A **CONDITIONAL** PROBABILITY (E.G. ASSUMING THAT AN EVEN NUMBER, I.E. ONE OF 2, 4, 6, HAS RESULTED, THE PROBABILITY THAT IT IS BIGGER THAN 3 IS 2/3).

SIMILARLY:

$$P(A \text{ OR } B) = P(A) + P(B)$$

WHEN $A$ AND $B$ ARE (MUTUALLY) **EXCLUSIVE**, MEANING THAT $A$ AND $B$ HAVE AN EMPTY OVERLAP (THEY CANNOT HAPPEN TOGETHER), AND

$$P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$$

IN GENERAL (*AND* BEING USUALLY EASIER TO DEAL WITH THAN *OR*).

EXAMPLE: THE EXPERIMENT CONSISTS OF ROLLING A DIE TWICE. WE DEFINE $A$ $(B)$ TO MEAN GETTING A SIX IN THE FIRST (SECOND) ROLL (CLEARLY INDEPENDENT).

$P(A \text{ AND } B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 2.78\%$ (PROBABILITY OF 2 SIXES).

PROBABILITY THAT WE DON'T GET TWO SIXES (CLEARLY THE COMPLEMENT OF THE PREVIOUS EVENT) THUS EQUALS $1 - \frac{1}{36} = 97.22\%$.

PROBABILITY OF AT LEAST ONE SIX:

$$P(A \text{ OR } B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{6} \times \frac{1}{6} = \frac{11}{36} = 30.56\%$$

ANOTHER EXAMPLE (RELATING TO CONDITIONAL PROBABILITIES): STUDENTS IN A CERTAIN CLASS HAVE BEEN ASKED WHETHER THEY SMOKED OR NOT. THESE ARE THE RESULTS:

| | NO | YES |
|---|---|---|
| MALE | 36 | 14 |
| FEMALE | 23 | 15 |

THE EXPERIMENT CONSISTS OF SELECTING ONE OF THESE STUDENTS AT RANDOM. THE PROBABILITY THAT IT IS:

- A MALE STUDENT: 50 / 88 = 56.82%
- A STUDENT WHO SMOKES: 29 / 88 = 32.95%

- GIVEN THAT THE STUDENT IS A NON-SMOKER, WHAT IS THE PROBABILITY THAT IT IS A FEMALE: 23 / 59 = 38.98%
- GIVEN IT IS A MALE STUDENT, WHAT IS THE PROBABILITY THAT HE SMOKES: 14 / 50 = 28.00%
- ARE 'SELECTING A MALE' AND 'SELECTING A SMOKER' INDEPENDENT EVENT? $(50/88)\times(29/88) \neq 14/88$ NO

---

SOME OF THE PREVIOUS RULES CAN BE EXTENDED TO MORE THAN TWO EVENTS, IN PARTICULAR:

P(*A* AND *B* AND *C*) = P(*A*) @ P(*B*) @ P(*C*)

WHEN *A*, *B* AND *C* ARE MUTUALLY INDEPENDENT, AND

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$

WHEN *A*, *B* AND *C* ARE MUTUALLY EXCLUSIVE.

# (PROBABILITY) TREE DIAGRAM

IS QUITE OFTEN THE EASIEST WAY TO KEEP TRACK OF OUTCOMES OF AN EXPERIMENT, E.G. HAVING 4 RED AND 7 BLUE MARBLES IN A BOX, AND SELECTING RANDOMLY AND WITHOUT REPLACEMENT, THREE OF THESE.

IT IS A GRAPHICAL REPRESENTATION OF THE SAMPLE SPACE - EACH COMPLETE PATH CORRESPONDS TO A SIMPLE EVENT (ONE COMPLETE OUTCOME).

NOTE THAT THE 'BRANCH' PROBABILITIES ARE OF THE CONDITIONAL TYPE.

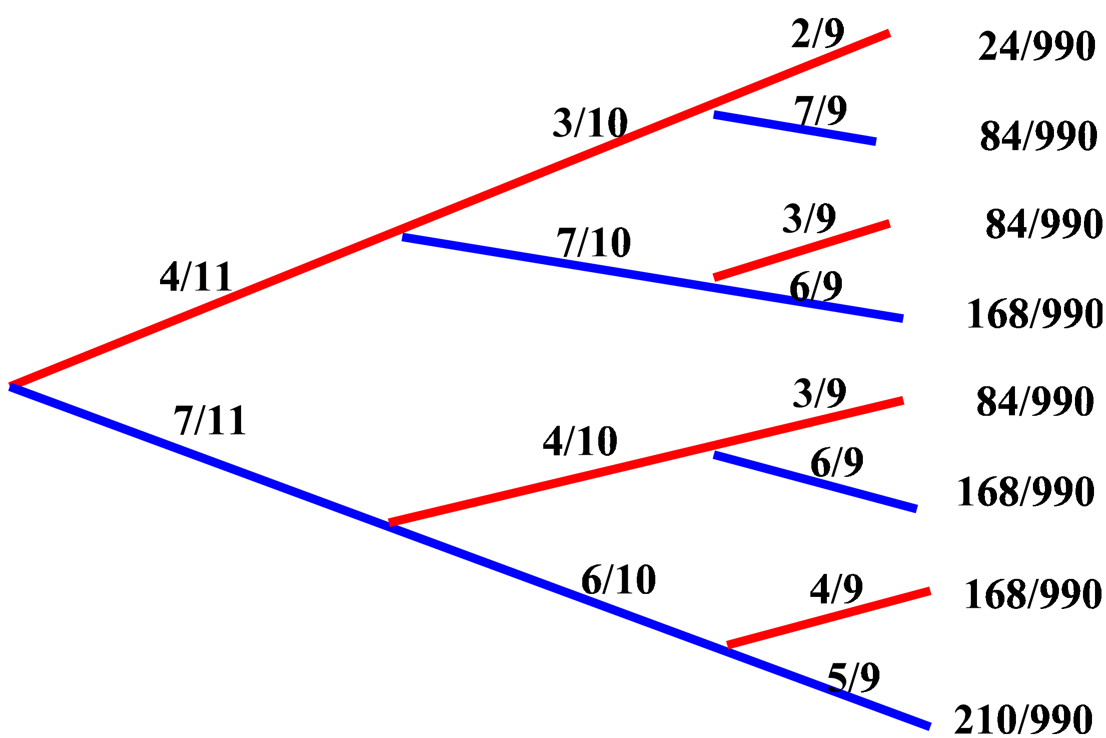
WHEN DOING SAMPLING WITH REPLACEMENT, THE PROBABILITY TREE REMAINS THE SAME, BUT THE INDIVIDUAL PROBABILITIES CHANGE (NOTE THAT THE OUTCOMES OF DRAW 1, 2 AND 3 ARE NOW INDEPENDENT).

## TWO SAMPLE-SPACE EXAMPLES:

ROLLING TWO DICE

| | | | | | |
|---|---|---|---|---|---|
| 1 1 | 1 2 | 1 3 | 1 4 | 1 5 | 1 6 |
| 2 1 | 2 2 | 2 3 | 2 4 | 2 5 | 2 6 |
| 3 1 | 3 2 | 3 3 | 3 4 | 3 5 | 3 6 |
| 4 1 | 4 2 | 4 3 | 4 4 | 4 5 | 4 6 |
| 5 1 | 5 2 | 5 3 | 5 4 | 5 5 | 5 6 |
| 6 1 | 6 2 | 6 3 | 6 4 | 6 5 | 6 6 |

DRAWING THREE MARBLES (TREE DIAGRAM)




PROBABILITY OF GETTING 2 RED MARBLES (IN ANY ORDER) IS THUS 84 / 990 + 84 / 990 + 84 / 990 = 25.45 %

TO UNDERSTAND FORMULAS OF THE NEXT CHAPTER, WE MUST NOW DISCUSS THE ISSUE OF COUNTING, I.E. IN <u>HOW MANY WAYS</u> CAN WE:

- ARRANGE $n$ PEOPLE IN A ROW (PLACING THEM IN EMPTY CHAIRS).

WE CLEARLY HAVE $n$ CHOICES TO FILL THE FIRST CHAIR, $n$-1 CHOICES TO FILL THE SECOND CHAIR (THIS GIVES US $n$@($n$-1) CHOICES FOR FILLING THE FIRST TWO CHAIRS - NOTE THAT INTERCHANGING ANY TWO PEOPLE COUNTS AS DIFFERENT), ETC. SO, AFTER $n$ STEPS, WE GET THE FOLLOWING ANSWER:

$n$ @ ($n$-1) @ ($n$-2) @ ... @ 2 @ 1 / $n$! ($n$ **FACTORIAL**)

E.G. WHEN $n = 3$, WE HAVE: 123, 132, 213, 231, 312, 321 (SIX POSSIBILITIES).

< REPEAT THIS, ASSUMING THAT THERE ARE $r < n$ CHAIRS (SOME PEOPLE WILL REMAIN STANDING):

$n$ @ $(n$-1) @ $(n$-2) @ ... @ $(n$-$r$ +1) = $\frac{n!}{(n-r)!}$ / $P_{n,r}$

(= ${}_nP_r$ = $P^n_r$) - NUMBER OF PERMUTATIONS.

E.G. $n$ = 8 AND $r$ = 3 YIELDS: 8 @ 7 @ 6 = 336.

< IN HOW MANY WAYS CAN WE SELECT $r$ PEOPLE (OUT OF $n$) WHO WILL BE SEATED (NO MATTER WHERE).

VISUALIZE THE PREVIOUS SITUATION, DONE IN TWO STAGES: FIRST, WE SELECT $r$ PEOPLE. SECOND, WE ARRANGE THEM IN THE $r$ CHAIRS. WE HAVE $P_{n,r}$ CHOICES ALTOGETHER, $r$! WAYS TO CARRY OUT THE SECOND STAGE. DUE TO THE **MULTIPLICATION PRINCIPLE**, THE NUMBER OF WAYS TO SELECT $r$ PEOPLE (STAGE ONE) MUST BE:

$$P_{n,r} \div r! = \frac{n!}{(n-r)! \cdot r!} \;/\; C_{n,r} \;/\; \binom{n}{r}$$

(NUMBER OF COMBINATIONS, $n$ CHOOSE $r$).

E.G. $n$ = 8 AND $r$ = 3 YIELDS $\frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1}$ = 56.

NOTE THAT THE FORMULA IS SYMMETRIC IN THE FOLLOWING SENSE: $C_{n,r} = C_{n,n-r}$ E.G. $C_{8,5} = C_{8,3} = 56$ (SELECTING 5 OUT OF 8 IS THE SAME AS SELECTING 3 OUT OF 8 - THE PEOPLE WHO WILL HAVE TO STAND).

WE NOW KNOW THE ANSWER TO ONE OF OUR PREVIOUS QUESTIONS: THERE ARE $C_{N,n}$ RANDOM SAMPLES OF SIZE $n$, IF THE POPULATION HAS $N$ MEMBERS. THEY ARE ALL EQUALLY LIKELY TO BE SELECTED.

<u>EXAMPLE</u>: ONE OF THE MOST COMMON 'SAMPLINGS' OF THIS TYPE IS OF COURSE RANDOMLY SELECTING A 5 CARD HAND, OUT OF A DECK OF 52 CARDS. THERE ARE ALTOGETHER $C_{52,5} = \dfrac{52\cdot 51\cdot 50\cdot 49\cdot 48}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 2{,}598{,}960$ SUCH HANDS, ALL EQUALLY LIKELY (OR SHOULD WE SAY UNLIKELY).

SINCE $C_{13,3} \times C_{39,2} = 286 \times 741 = 211{,}926$ OF THESE WILL CONTAIN EXACTLY 3 SPADES, THE PROBABILITY OF THIS HAPPENING IS $\dfrac{211{,}926}{2{,}598{,}960} = 8.15\%$.

SIMILARLY, THE PROBABILITY OF GETTING (EXACTLY) 2 ACES IS $\dfrac{C_{4,2} \times C_{48,3}}{2{,}598{,}960} = 3.99\%$.

ANOTHER EXAMPLE: FORM A BOX WITH 15 RED AND 27 BLUE MARBLES, WE RANDOMLY SELECT 9 MARBLES (WITHOUT REPLACEMENT). WHAT IS THE PROBABILITY OF GETTING (EXACTLY) FOUR RED MARBLES IN OUR 'SAMPLE'?

ANSWER: QUITE ROUTINELY NOW

$$\frac{C_{15,4} \times C_{27,5}}{C_{42,9}} = \frac{1365 \times 80{,}730}{445{,}891{,}810} = 24.71\%$$

THE FORMULA WE HAVE BEEN USING IS CALLED **HYPERGEOMETRIC**.

IN THE NEXT CHAPTER, WE DEVELOP ANOTHER GENERAL FORMULA, TO SIMILARLY DEAL WITH SAMPLING WITH REPLACEMENT.