CHAPTER 5: RANDOM VARIABLES, BINOMIAL AND POISSON DISTRIBUTIONS

DEFINITION: IF AN OUTCOME OF A RANDOM EXPERIMENT IS CONVERTED TO A SINGLE (RANDOM) NUMBER (E.G. THE TOTAL NUMBER OF DOTS WHEN ROLLING TWO DICE), THIS IS CALLED A **RANDOM VARIABLE** (NOT TO BE CONFUSED WITH THE OLD 'EVENT').

A TABLE WHICH LISTS ALL POSSIBLE VALUES OF A RANDOM VARIABLE, TOGETHER WITH THE CORRESPONDING PROBABILITIES, IS CALLED THE RANDOM VARIABLE'S **DISTRIBUTION**.

E.G. ROLLING TWO DICE AND COUNTING THE TOTAL NUMBER OF DOTS (LET US CALL THIS RANDOM VARIABLE X) YIELDS:

$X =$	2	3	4	5	6	7	8	9	10	11	12
Pr:	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{35}$	$\frac{2}{36}$	$\frac{1}{36}$

WE CAN ALSO PLOT THE CORRESPONDING HISTOGRAM.

THE **MEAN** (EXPECTED VALUE) OF A DISTRIBUTION IS COMPUTED BY

$$\mu = \Sigma x \cdot P(x)$$

I.E. MULTIPLYING EACH VALUE BY THE CORRESPONDING PROBABILITY AND SUMMING UP (OVER THE WHOLE TABLE).
E.G. (USING THE PREVIOUS EXAMPLE):

$$\frac{2\times 1+3\times 2+4\times 3+...+11\times 2+12\times 1}{36}=\frac{252}{36}=7$$

IF THE EXPERIMENT IS PERFORMED, REPEATEDLY, THE CORRESPONDING <u>SAMPLE MEAN</u> WILL <u>APPROACH</u> THIS (THEORETICAL) MEAN, MORE AND MORE CLOSELY AS WE INCREASE THE SAMPLE SIZE (**LAW OF LARGE NUMBERS**).

STANDARD ('TYPICAL') **DEVIATION** OF THE RANDOM VARIABLE (THE CORRESPONDING DISTRIBUTION) IS:

$$\sigma = \sqrt{\Sigma(x-\mu)^2 \cdot P(x)}$$

$$= \sqrt{\left(\Sigma x^2 \cdot P(x)\right) - \mu^2}$$

THE SECOND EXPRESSION BEING EASIER TO EVALUATE, E.G. (OUR EXAMPLE):

$$\sqrt{\frac{1\times 2^2 + 2\times 3^2 + \ldots + 1\times 12^2}{36} - 7^2} = \sqrt{\frac{35}{6}} = 2.4152$$

SOME DISTRIBUTIONS (AND TYPES OF EXPERIMENTS) ARE SO COMMON THAT THEY HAVE THEIR OWN NAME. ONE OF THE MOST IMPORTANT EXAMPLES IS THE SO CALLED **BINOMIAL** DISTRIBUTION.

IT ARISES FROM THE FOLLOWING SITUATION:

WE PERFORM, INDEPENDENTLY, A FIXED NUMBER (DENOTED *n*) **TRIALS** (E.G. FLIPPING A COIN, ROLLING A DIE, VACCINATING A PATIENT, DRAWING A MARBLE FROM A BOX, ETC.), EACH HAVING ONLY TWO POSSIBLE OUTCOMES CALLED **SUCCESS** AND **FAILURE** (GETTING HEADS, A SIX, AN ADVERSE REACTION, A RED MARBLE, ETC.).

THE PROBABILITY OF EACH SUCCESS IS p, OF A FAILURE IS q / 1- p (THESE MUST NOT CHANGE FROM TRIAL TO TRIAL - SAMPLING WITH REPLACEMENT).

THE RANDOM VARIABLE OF INTEREST IS THE TOTAL NUMBER OF SUCCESSES OBTAINED (WE WILL CALL IT X). WHAT IS ITS DISTRIBUTION?

THE POSSIBLE VALUES ARE CLEARLY 0, 1, 2, ... n-1, n, WE ALSO NEED TO KNOW THE PROBABILITY OF EACH OF THESE.

LET'S FIRST TRY THIS WITH $n = 5$. X WILL HAVE A VALUE OF 0 ONLY WHEN ALL TRIALS RESULT IN A FAILURE (*FFFFF*).

DUE TO INDEPENDENCE, THIS OUTCOME HAS THE PROBABILITY OF $q \cdot q \cdot q \cdot q \cdot q = q^5$.

THE VALUE OF 1 COMES FROM 5 OUTCOMES (*SFFFF, FSFFF, FFSFF, FFFSF, FFFFS*). LUCKILY, THEY ALL HAVE THE SAME PROBABILITY OF $p \cdot q^4$, AND ARE MUTUALLY <u>EXCLUSIVE</u>. THIS MEANS THAT THE PROBABILITY OF $X = 1$ EQUALS TO $5pq^4$.

HOW ABOUT $X = 2$? THE CONTRIBUTING OUTCOMES ARE: *SSFFF, SFSFF, SFFSF, ... FFFSS* - HOW MANY OF THESE ARE THERE? WELL, WE HAVE TO CHOOSE 2 OUT OF 5 POSITIONS (FOR THE TWO *S*'S), THE ANSWER IS THUS $\binom{5}{2} = 10$.

PUTTING IT TOGETHER: $P(X = 2) = \binom{5}{2} p^2 q^3$.

ANY GUESS AS TO $P(X = 3)$?

$\binom{5}{3} p^3 q^2$, OF COURSE!

AND $P(X = i)$, WHERE i IS ANY INTEGER

BETWEEN 0 AND n (INCLUSIVE)?

$$\binom{5}{i} \cdot p^i \cdot q^{5-i}$$

THE GENERAL FORMULA (FOR ANY n) IS

THUS $$P(X=i) = \binom{n}{i} \cdot p^i \cdot q^{n-i}$$

ONE CAN SHOW THAT THIS DISTRIBUTION HAS THE MEAN OF $n \cdot p$ AND THE STANDARD DEVIATION OF $\sqrt{n \cdot p \cdot q}$.

EXAMPLE: ROLL A DIE 4 TIMES, CALL GETTING 5 OR 6 A SUCCESS. FIND THE DISTRIBUTION OF THE TOTAL NUMBER OF SUCCESSES, AND VERIFY THE : $= np$ AND F $= \sqrt{npq}$ FORMULAS.

$X=$	0	1	2	3	4
Pr =	$\left(\frac{2}{3}\right)^4$	$4 \times \left(\frac{2}{3}\right)^3 \times \frac{1}{3}$	$6 \times \left(\frac{2}{3}\right)^2 \times \left(\frac{1}{3}\right)^2$	$4 \times \frac{2}{3} \times \left(\frac{1}{3}\right)^3$	$\left(\frac{1}{3}\right)^4$
	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{24}{81}$	$\frac{8}{81}$	$\frac{1}{81}$

THIS MEANS THAT THE PROBABILITY OF, SAY, (EXACTLY) 2 SUCCESSES EQUALS 24/81 = 29.63%.

THE MEAN : IS COMPUTED BY

$$\frac{0\times 16+1\times 32+2\times 24+3\times 8+4\times 1}{81}=\frac{108}{81}=\frac{4}{3}$$ (CHECK)

SIMILARLY, TO COMPUTE THE VARIANCE, WE FIRST DO

$$\frac{0^2\times 16+1^2\times 32+2^2\times 24+3^2\times 8+4^2\times 1}{81}=\frac{216}{81}=\frac{8}{3}$$ THEN

SUBTRACT $:^2$, RESULTING IN 8 / 9 (CHECK).

OUR SECOND (AND FINAL) ‘COMMON’ DISTRIBUTION OF THIS (INTEGER) TYPE IS CALLED POISSON:

THIS RELATES TO AN EXPERIMENT IN WHICH WE ARE COUNTING THE RANDOM NUMBER OF:

< ARRIVING CUSTOMERS (A STORE)

< ACCIDENTS (AN INTERSECTION)

< PHONE CALLS (AN OFFICE)

< FISHES CAUGHT (A LAKE)

< ......

DURING SOME FIXED TIME INTERVAL, ASSUMING THAT THE THEORETICAL RATE (OF ARRIVALS, ACCIDENTS, ETC.) IS CONSTANT IN TIME (NO PEAK OR SLACK PERIODS), AND ANY TWO (ARRIVALS, ACCIDENTS, ETC.) CAN HAPPEN ARBITRARILY CLOSE (IN TIME) TO EACH OTHER.

WITH THESE ASSUMPTIONS, THE RANDOM VARIABLE (TOTAL NUMBER OF ARRIVALS BETWEEN 9:00 A.M. AND 10:00 A.M.) CAN HAVE (AT LEAST THEORETICALLY) ANY (NON-NEGATIVE) INTEGER VALUE. THE CORRESPONDING PROBABILITIES ARE GIVEN BY THE FOLLOWING FORMULA:

$$P(i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda}$$

WHERE e IS THE BASE OF NATURAL LOGARITHMS (–2.7183), AND 8 IS THE RATE (OF ARRIVALS, ETC.) MULTIPLIED BY THE LENGTH OF THE TIME INTERVAL SELECTED.

ONE CAN SHOW THE 8 IS ALSO THE MEAN OF THIS DISTRIBUTION. THE STANDARD DEVIATION EQUALS $\sqrt{\lambda}$.

EXAMPLE: CUSTOMERS ARRIVE AT AN AVERAGE RATE OF 17.2 PER HOUR. FIND THE PROBABILITY OF MORE THAN THREE ARRIVALS DURING THE NEXT 5 MINUTES.

SOLUTION: FIRST, WE COMPUTE 8 $= \frac{17.2}{12} = 1.4\overline{3}$. UTILIZING THE COMPLEMENT, WE EVALUATE

$$1 - \left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6}\right) \cdot e^{-\lambda} = 0.05756 = 5.756\%$$

ADDITIONAL EXAMPLES

LET'S RE-DO THE PROBLEM OF THE PREVIOUS LECTURE (15 RED, 27 OTHER MARBLES), ASSUMING THE SAMPLING IS NOW DONE WITH REPLACEMENT:

$$C_{9,4} \times \left(\frac{15}{42}\right)^4 \times \left(\frac{27}{42}\right)^5 = 22.51\%$$

(WITHOUT REPLACEMENT RESULTED IN 24.71%).

IN 21 ROLLS OF A DIE, WHAT IS THE PROBABILITY OF GETTING MORE THAN 3 SIXES?
RATHER THAN ADDING ALL PROBABILITIES FROM 4 TO 21, IT IS EASIER TO DEAL WITH THE COMPLEMENT:

$$1 - \left[\left(\frac{5}{6}\right)^{21} + 21 \times \left(\frac{5}{6}\right)^{20}\left(\frac{1}{6}\right) + C_{21,2}\left(\frac{5}{6}\right)^{19}\left(\frac{1}{6}\right)^2 + C_{21,3}\left(\frac{5}{6}\right)^{18}\left(\frac{1}{6}\right)^3\right]$$

= 47.31 %

A FISHERMAN KNOWS (FROM PAST EXPERIENCE) THAT HE CAN CATCH FISH AT THE AVERAGE RATE OF 3.2 PER HOUR. IF HE GOES FISHING BETWEEN 4:30 AND 6:00 PM, WHAT IS THE PROBABILITY THAT HE WILL CATCH:

- (EXACTLY) 5 FISHES
- FEWER THAN 4
- BETWEEN 3 TO 6 (INCLUSIVE).

SOLUTION: FIRST WE FIND THE VALUE OF 8 = 3.2 × 1.5 = 4.8 (THIS IS THE MEAN OF THE CORRESPONDING DISTRIBUTION). THEN,

$$\frac{4.8^5}{5!}\times e^{-4.8}=17.47\%$$

$$\left(1+4.8+\frac{4.8^2}{2}+\frac{4.8^3}{6}\right)\times e^{-4.8}=29.42\%$$

$$\left(\frac{4.8^3}{6}+\frac{4.8^4}{24}+\frac{4.8^5}{120}+\frac{4.8^6}{720}\right)\times e^{-4.8} = 64.83\%$$

ON MOST COUNTY ROADS POTHOLES OCCUR AT THE RATE OF 3.7 PER KILOMETRE. WE HAVE 1.5 KM LEFT TO REACH HOME.
WHAT IS THE PROBABILITY OF ENCOUNTERING FEWER THAN 3 POTHOLES?

$$\lambda=3.7\times 1.5=5.55$$

$$\left(1+\lambda+\frac{\lambda^2}{2}\right)\times e^{-\lambda} = 8.53\%$$

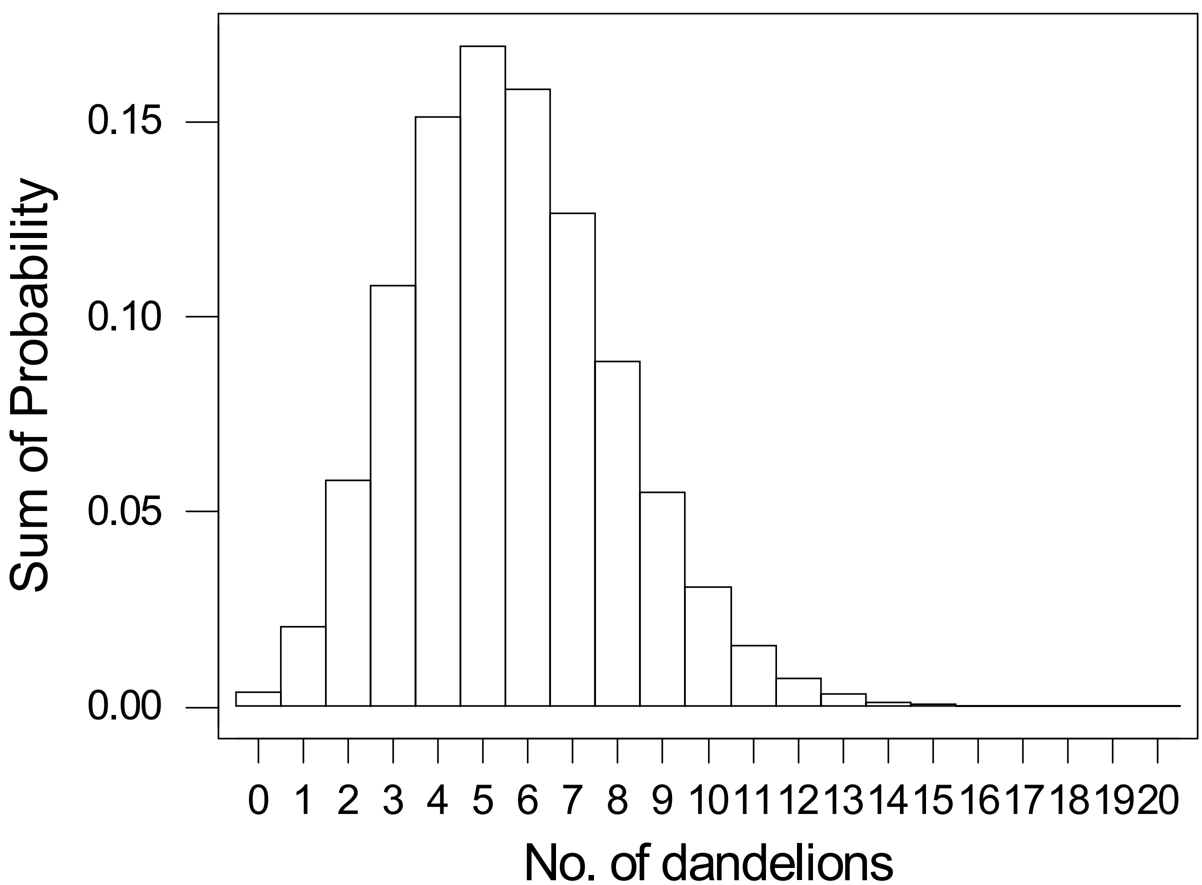
IN A CERTAIN TYPE OF LAWN, THE AVERAGE DENSITY OF DANDELIONS IS 0.56 PER SQUARE METER.
WHAT IS THE EXPECTED VALUE AND STANDARD DEVIATION OF THE NUMBER OF DANDELIONS IN AN AREA MEASURING 10 SQUARE METERS?
FIND THE PROBABILITY THAT THIS (10 m^2) AREA WILL CONTAIN EXACTLY 5 DANDELIONS? MORE THAN 5 DANDELIONS?

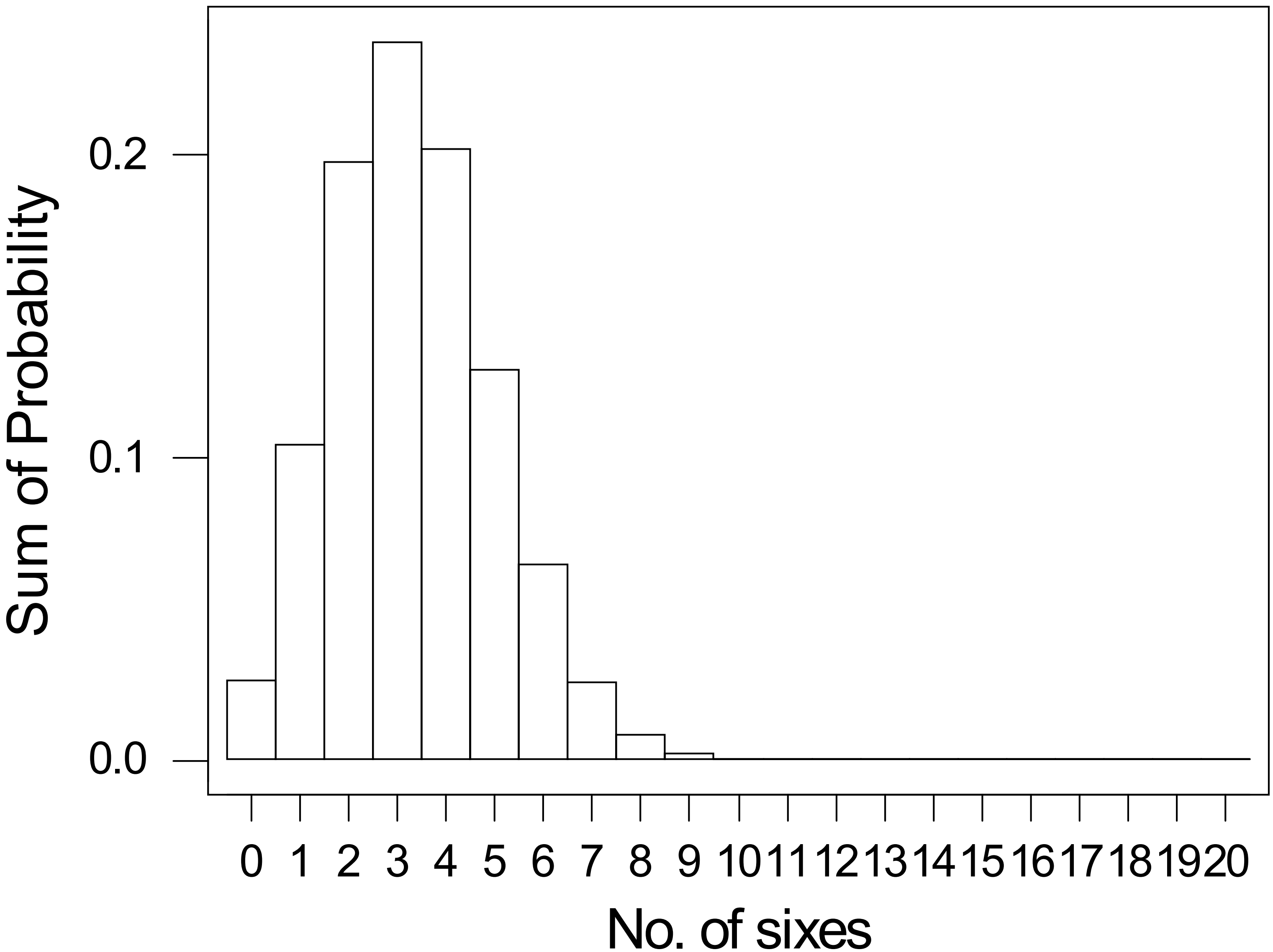
8 = 0.56 × 10 = 5.6 THIS IS ALSO THE CORRESPONDING MEAN (EXPECTED) VALUE

THE STANDARD DEVIATION IS $\sqrt{\lambda}$ = 2.366

$$P(5) = \frac{5.6^5}{5!} \times e^{-5.6} = 16.97\%$$

$$P(X>5) = 1 - \left(1 + 5.6 + \ldots + \frac{5.6^5}{5!}\right) \times e^{-5.6} = 48.81\%$$




Sum of Probability
0.0
0.1
0.2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
No. of sixes