1

- **Recall**: A Bernoulli trial has two outcomes, success (S) and failure (F), with $p(S) = p$.

A binomial random variable, denoted by $X \sim Bin(n, p)$, counts the number of successes in n independent Bernoulli trials.

→ As the trials are independent, the probability of success remains fixed at p for each trial.

Eg. Suppose we flip a fair coin 10 times & consider tails as "success".

Then $X \sim Bin(n = 10, p = 0.5)$ counts the number of tails that result.

Here we have 2^{10} equally-likely outcomes $\{SSSSSSSSSS, \ldots, FFFFFFFFFF\}$, with X having $\{0, 1, 2, \ldots, 10\}$ as its possible values.

As trials are independent, the probability of a particular outcome such as SSSSFSFFSFS is $p^6(1-p)^4$, which is $(\frac{1}{2})^6(1-\frac{1}{2})^4$.

As there are $\binom{10}{6}$ ways of selecting 6 positions in 10, we have $\binom{10}{6}$ outcomes consisting of 6S & 4F.

Thus, the probability of getting exactly 6 successes is

$$P(X=6) = \binom{10}{6} p^6 (1-p)^4$$

$$= \left(\frac{10\times 9\times 8\times 7}{4\times 3\times 2\times 1}\right)\left(\tfrac{1}{2}\right)^6\left(1-\tfrac{1}{2}\right)^4 = \underline{0.205}$$

In general, for $X \sim Bin(n, p)$, we have

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$.

- The <u>expected value</u> (or "expected number" or "expectation") of $X \sim Bin(n,p)$, denoted by $E(X)$, is the weighted average of the values of $X \sim Bin(n,p)$ with respect to their probabilities.

Thus for $X \sim Bin(n,p)$, we have that

$$E(X) = \sum_{k=0}^{n} k \cdot P(X=k)$$

$$= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

Using the Binomial Theorem (i.e. $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$) we can show that

$\underline{E(X) = np}$ for $X \sim Bin(n,p)$.

(3)

The expected value can be considered as the long-run average number ~~number~~ of successes.

For example, suppose we repeated our experiment of flipping 10 fair coins numerous times.

In the long run, we'd expect the average number of successes across all experiments to approach

$$E(X) = np = (10)(0.5) = 5.$$

Here, $E(X) = np$ is a possible value of $X \sim Bin(n, p)$ but that need not be the case in general.

Eg. Let $X \sim Bin(n=5, p=0.5)$

(eg. X counts number of heads, say, in 5 fair flips)

Then X has values $\{0, 1, \ldots, 5\}$ with

$$P(X=k) = \binom{5}{k}\left(\tfrac{1}{2}\right)^k\left(\tfrac{1}{2}\right)^{5-k}.$$

So $P(X=0) = \binom{5}{0}\left(\tfrac{1}{2}\right)^0\left(\tfrac{1}{2}\right)^5 = (1)(1)\left(\tfrac{1}{2}\right)^5 = \frac{1}{32}$

& $P(X=1) = \binom{5}{1}\left(\tfrac{1}{2}\right)^1\left(\tfrac{1}{2}\right)^4 = (5)\left(\tfrac{1}{2}\right)^5 = \frac{5}{32}$, etc.

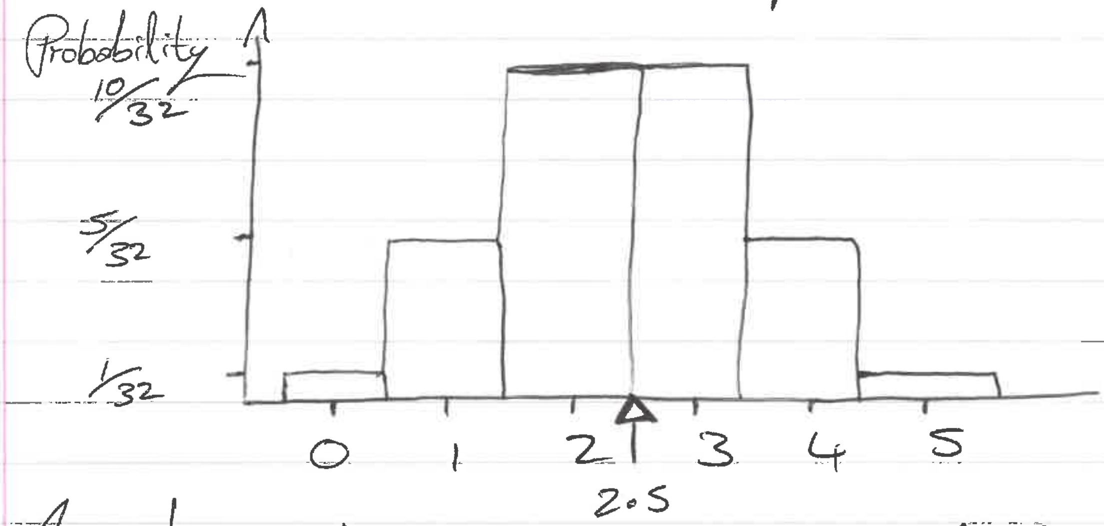
We have:

k	0	1	2	3	4	5
$P(X=k)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Here $E(X) = np = (5)\left(\tfrac{1}{2}\right) = \frac{5}{2}$ which is <u>not</u> a value of X.

Exercise: Confirm that $\sum_{k=0}^{5} k \cdot p(X=k) = 5/2$

We can use the above table of probabilities to graphically display $X \sim Bin(n=5, p=\frac{1}{2})$:



This shape has total area equal to 1 and is balanced at $E(X) = 2.5$

In general, for X a random variable, $E(X)$ is the balance point or "mean" of its distribution.

- For $X \sim Bin(n=5, p=\frac{1}{2})$ as above, we can calculate the probability of any event in the usual way:

Eg. • $p(X \leq 1) = p(X=0) + p(X=1)$
$= \frac{1}{32} + \frac{5}{32} = \frac{6}{32}$,

• $p(X \geq 4) = p(X=4) + p(X=5) = 6/32$,

• $p(2 \leq X \leq 3) = p(X=2) + p(X=3) = \frac{10}{32} + \frac{10}{32}$
$= 20/32$

Note $p(2 \leq X \leq 3) = 1 - p(X \leq 1) - p(X \geq 4)$ by Complement Rule.

5

- As before, when working with $X \sim Bin(n, p)$, we require the independence of the n Bernoulli trials.
  In practice, that's hard to guarantee.

To randomly select 2 cards from a deck of 52 we should shuffle, select one and note it, replace this card, shuffle again and select and note a second card (this is called "sampling with replacement").

In practice, we may instead "sample without replacement": if we discard our first card before drawing the second, the probability of 2 spades is

$$p(\text{Spade on Draw 1}) \times p(\text{Spade on Draw 2} \mid \text{Spade on Draw 1})$$

$$= \left(\frac{1}{4}\right) \times \left(\frac{12}{51}\right)$$

This is not the same as $p(\text{Spade on Draw 1}) \times p(\text{Spade on Draw 2})$, violating our independence assumption.

Aside: What's $p(\text{Spade on Draw 2})$?
By the Law of Total Probability, we have

$$p(\text{Spade on Draw 2}) =$$
$$p(\text{Spade on D2} \mid \text{Spade on D1})\, p(\text{Spade on D1})$$
$$+\ p(\text{Spade on D2} \mid \text{No spade on D1})\, p(\text{No Spade on D1})$$

$$= \left(\frac{12}{51}\right) \times \left(\frac{1}{4}\right) + \left(\frac{13}{51}\right) \times \left(\frac{3}{4}\right) = \frac{51}{(51)4} = \frac{1}{4}$$

6

In practice, if n is large, we can relax our independence assumption slightly and use a Binomial random variable with respect to sampling without replacement

Eg. Suppose 5% of the Irish population are colourblind.
Let X count the number of colourblind people in a randomly sample of 100 people

Thus $X \sim Bin(n, p)$ with

$$X \sim Bin(100, 0.5).$$

Thus, $p(X \geq 2) = 1 - (p(X=0) + p(X=1))$

$$= 1 - \binom{100}{0}(.05)^0(1-.05)^{100} - \binom{100}{1}(.05)^1(1-.05)^{99}$$

$$= 1 - .0059 - .0312$$

$$= .9629$$

- Exam will be similar in format to the 2017-18 papers

Do 3 questions from 4
Main sections in notes
→ Sets and Logic
→ Relations and Functions
→ Counting and Combinatorics
→ Probability
Notes & Tutorial Sheets