37161 Probability and Random Variables

Lecture 4

Probability Mass Functions

- In Lecture 3, we saw about defining the probability mass function of a discrete random variable.

- This was written as a list of all possible values of the random variable and the probability that the variable takes each of those values.

- For example, if we are rolling one regular fair six-sided die and defining the random variable X to be the number shown, then the probability mass function of X is

$$P(X = k) = \begin{cases} 1/6 & k = 1 \\ 1/6 & k = 2 \\ 1/6 & k = 3 \\ 1/6 & k = 4 \\ 1/6 & k = 5 \\ 1/6 & k = 6 \\ 0 & \text{otherwise} \end{cases}$$

“Standard” Probability Distributions

- Many times our random variable of interest might be a “standard” type of variable.

- That is, whatever the context of the experiment from which it is obtained, there are a number of random variables which commonly arise and are well studied.

- For example, flipping a single fair coin once and seeing how many times (0 or 1) it lands Tails gives rise to exactly the same random variable as rolling a fair regular six-side die once and counting how many even numbers are obtained.

The Simplest Common Random Variable

- Consider two independent random experiments.

- Let X be the number of Hearts cards selected when picking one card at random from a standard deck of 52 with all cards equally likely to be chosen.
- Let Y be the number of tickets ending in a 7 selected when one raffle ticket is selected at random from a bucket containing tickets numbered with integers 1-100 with all tickets equally likely to be chosen.
- Clearly $P(X = k) = \begin{cases} 0.75 & k = 0 \\ 0.25 & k = 1 \\ 0 & \text{otherwise} \end{cases}$ and $P(Y = k) = \begin{cases} 0.9 & k = 0 \\ 0.1 & k = 1 \\ 0 & \text{otherwise} \end{cases}$

The Simplest Common Random Variable

- Any random variable whose probability mass function can

be written as $P(X=k)=\begin{cases} 1-p & k=0 \\ p & k=1 \\ 0 & \text{otherwise} \end{cases}$

is known as a **Bernoulli** variable.

- In this case, we write $X \sim Bern(p)$.

- This distribution has range $\{0,1\}$.

Jacob Bernoulli

(1655-1705)

Bernoulli Distribution

- The distribution depends on one parameter, *p*, which gives the probability of obtaining a 1, rather than a zero.

- For example, the number of Tails from a single fair coin flip $\sim Bern(0.5)$ or, when selecting one person at random, the number of selected people born on a Saturday $\sim Bern\left(\frac{1}{7}\right)$

- The expectation and variance of $X \sim Bern(p)$ can easily be calculated.

- $E(X) = \sum k \times p(X = k) \;= [0 \times (1-p)] + [1 \times p] \;= p$

Jacob Bernoulli

(1655-1705)

The Simplest Common Random Variable

- Similarly, $E(X^2) = \sum k^2 \times p(X = k) \; = \left[0^2 \times (1-p)\right] + \left[1^2 \times p\right] = p$

- This therefore gives $Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$.

- These values are perhaps intuitive. If we expect half of our experiments to give a 1 then, on average, each experiment gives the value 0.5.

- The variance is zero if $p = 0$ or $p = 1$. This is because there is no variability between realisations of this experiment – we already knew the outcome would either certainly happen (1) or certainly not happen (0).

Jacob Bernoulli

(1655-1705)

Binomial Distribution

- A generalisation of Bernoulli variables gives rise to another commonly seen variable.

- Adding the outcomes of *n* identical independent Bernoulli variables gives a **Binomial** variable.

- If $X_1 \sim Bern(p), X_2 \sim Bern(p), \ldots, X_n \sim Bern(p)$ then $\left[X_1 + X_2 + \ldots + X_n\right] \sim Bin(n,p)$.

- Clearly $Y \sim Bern(p)$ and $Y \sim Bin(1,p)$ mean exactly the same thing.
- A binomial random variable requires two parameters:

- *n*: The number of independent Bernoulli variables
- *p*: The probability of a 1 from each Bernoulli variable

Binomial Distribution

- Even before we obtain the probability mass function of $X \sim Bin(n,p)$, we can calculate its range, expectation and variance.

- Since each individual Bernoulli variable takes the value 0 or 1 and we are adding n independent outcomes of these, the range of a $Bin(n,p)$ variable is $\{0,1,2,\ldots,n\}$.

- We have already seen that that if $X_1 \sim Bern(p)$ then $E(X_1) = p$.
- Because, for any random variables A and B, $E(A+B) = E(A) + E(B)$, we know that

$[X_1 + X_2 + \ldots + X_n] \sim Bin(n,p)$ has expectation $E(X_1) + E(X_2) + \ldots + E(X_n) = p + p + \ldots + p = np$

- Similarly, if $X \sim Bin(n,p)$ then $Var(X) = np(1-p)$.

Binomial Distribution

- Consider flipping a (possibly biased) coin which lands Heads on each flip with probability p. Let X be the total number of Heads obtained in 4 flips. What is the probability mass function of $X \sim Bin(4,p)$?

- $P(X=0)=(1-p)\times(1-p)\times(1-p)\times(1-p)=(1-p)^4$ since this only occurs if each independent flip is Tails, each of which happens independently with probability $(1-p)$.

- $$P(X=1)=\begin{cases} P(TTTH)+ \\ P(TTHT)+ \\ P(THTT)+ \\ P(HTTT)+ \end{cases} = \begin{cases} (1-p)\times(1-p)\times(1-p)\times p+ \\ (1-p)\times(1-p)\times p\times(1-p)+ \\ (1-p)\times p\times(1-p)\times(1-p)+ \\ p\times(1-p)\times(1-p)\times(1-p) \end{cases} = 4p(1-p)^3$$

Binomial Distribution

- Similarly, $P(X=2)=6p^2(1-p)^2$ since the outcome could arise from *HHTT, HTHT, HTTH, TTHH, THTH* or *THHT* – six different ways.
- We also have $P(X=3)=4p^3(1-p)$ and $P(X=4)=p^4$.
- $$P(X=k)=\begin{cases}(1-p)^4 & k=0\\ 4p(1-p)^3 & k=1\\ 6p^2(1-p)^2 & k=2\\ 4p^3(1-p) & k=3\\ p^4 & k=4\\ 0 & \text{otherwise}\end{cases}$$

Binomial Distribution: Combinations

- To obtain the probability mass function of $X \sim Bin(n,p)$, we need to know how many ways X can take each possible value.

- For example, if we want to know the probability of flipping a (possibly biased) coin 10 times and getting 3 Heads, we need to know how many ways this could happen. For example, we could have *HHHTTTTTTT*, *TTTTTTTHHH*, *TTTHHHTTTT* etc.

- In other words, we need to know how many ways we could write a string of 7 Tails and 3 Heads.

Binomial Distribution: Combinations

- Call the three coins that land Heads H_1, H_2, H_3 . There are ten places in the string of possible outputs which could be H_1 . Once this is placed, there are nine places in the string which could be H_2 etc.

- The total number of strings containing H_1, H_2, H_3 is therefore $10 \times 9 \times 8 = 720$.

- However, we are only counting the total number of Heads in that string, so $TTTTTTTH_1H_2H_3$ and $TTTTTTTH_1H_3H_2$ are equivalent.

- For three Heads, there are $3 \times 2 \times 1 = 6$ orderings of these.
- Since each of the 720 orderings of the 10 outputs corresponds to each combination 6 times, we therefore have 720/6 = 120 combinations of three Heads and seven Tails.

Binomial Distribution: Combinations

- When looking for the number of ways that k outcomes can be ordered in a string of n trials, we have $\binom{n}{k} = {}^{n}C_{k} = \dfrac{n!}{(n-k)!k!}$ ways, where $n! = n(n-1)(n-2)...2\times 1$
- As with the example on the previous slide, we have $\binom{10}{3} = {}^{10}C_{3} = \dfrac{10!}{7!3!} = 120$.
- The probability of getting k Heads out of n flips of a coin which lands Heads with probability p is therefore $P(X=k) = \binom{n}{k} p^{k}(1-p)^{n-k}$.
- For example, when flipping a fair coin 12 times, the probability of obtaining exactly 7 Heads is $P(X=7) = \binom{12}{7} 0.5^{7}(0.5)^{5} = 792 \times (0.5)^{12} \approx 0.193$

Binomial Distribution

- For $X \sim Bin(n,p)$, calculation of $E(X)$ or $Var(X)$ directly from the probability mass function $P(X=k)=\frac{n!}{(n-k)!k!}p^k(1-p)^{n-k}$ is not a simple task and requires combinatorics beyond the scope of this subject.

- Similarly, even verifying that $\sum_{k=0}^{n} P(X=k)=1$ is not trivial.

- We have already seen, though, how these van be easily obtained via understanding that adding independent identical Bernoulli variables gives rise to a Binomial variable.

Binomial Distribution: Example

- In North-West Europe, it is observed that around 6% of the population has red hair.
- Selecting 5 people at random, what is the probability that exactly two of them have red hair?
- $X \sim Bin(5, 0.06)$ so $P(X=2) = \binom{5}{2}(0.06)^2(0.94)^3 \approx 0.0299$
- What is the probability that three siblings all have red hair?
- This **cannot** be calculated by a binomial distribution, since the Bernoulli trials (i.e. does each sibling individually have red hair) are not independent, as hair colour is a genetic trait.

Geometric Distribution

- One other common random variable which can arise from independent Bernoulli trials is a **Geometric** variable.

- We write $X \sim Geo(p)$ if X is the number of successive independent identical Bernoulli variables until the first 1 is obtained.

- For example, when flipping a fair coin repeatedly, the number of flips until the first Heads $\sim Geo(0.5)$.
- The range of $X \sim Geo(p)$ is easily seen to be $\{1, 2, 3, \ldots\}$.

Geometric Distribution

- When considering a number of independent $Bern(p)$ variables, we obtain the first 1 on the kth variable if and only if the first $(k-1)$ are 0s and the kth is a 1.

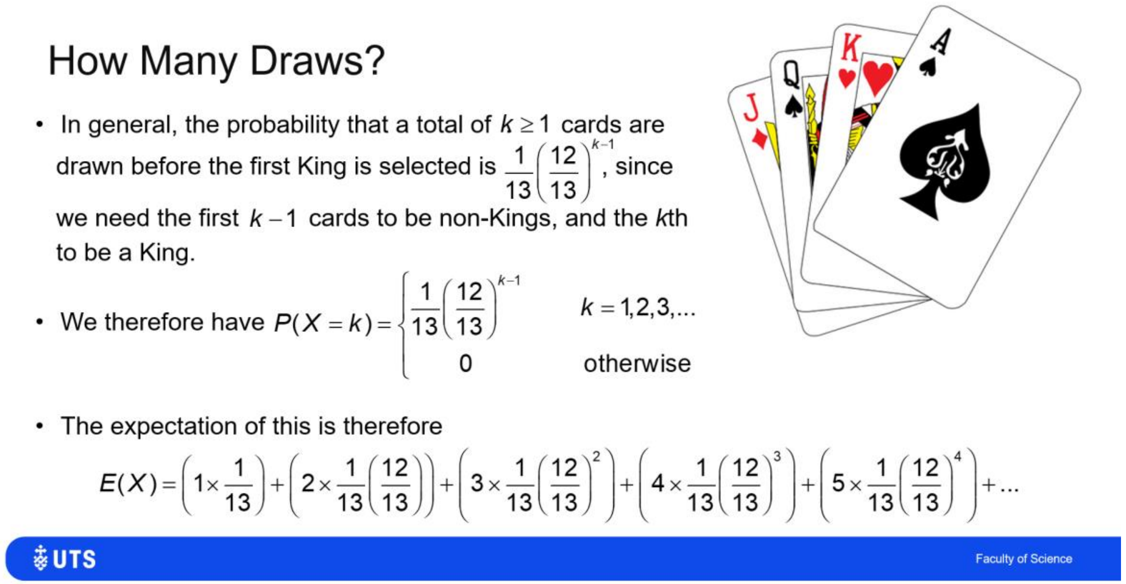
- That is, $P(X=k)=\begin{cases}(1-p)^{k-1}p & k=1,2,3,\ldots \\ 0 & \text{otherwise}\end{cases}$

- We can verify that this is a valid probability mass function since

$$\sum_{k=1}^{\infty}P(X=k)=\sum_{k=1}^{\infty}(1-p)^{k-1}p \quad =p+(1-p)p+(1-p)^2p+(1-p)^3p+(1-p)^4p+$$

- This is a geometric series, first term p, common ratio $(1-p)$.
- The infinite sum is therefore $\sum_{k=1}^{\infty}P(X=k)=\dfrac{p}{1-(1-p)}=1.$

Geometric Distribution



- We already saw last lecture (via a geometric series of geometric series) that the expectation of $X \sim Geo(p)$ is $E(X)=\frac{1}{p}$.
- This is again perhaps intuitive, since if, on average one trial out of every ten is a 1 then, on average, we would have to look at around ten outcomes before expecting to see a 1.

- Note, though, that a geometric variable arises only if the Bernoulli trials are independent. For example, if we are selecting a card from a deck without replacement, then the number of cards needed until the first King is drawn is not geometrically distributed.
- (Sampling without replacement gives rise to a hypergeometric distribution – beyond the scope of 37161.)