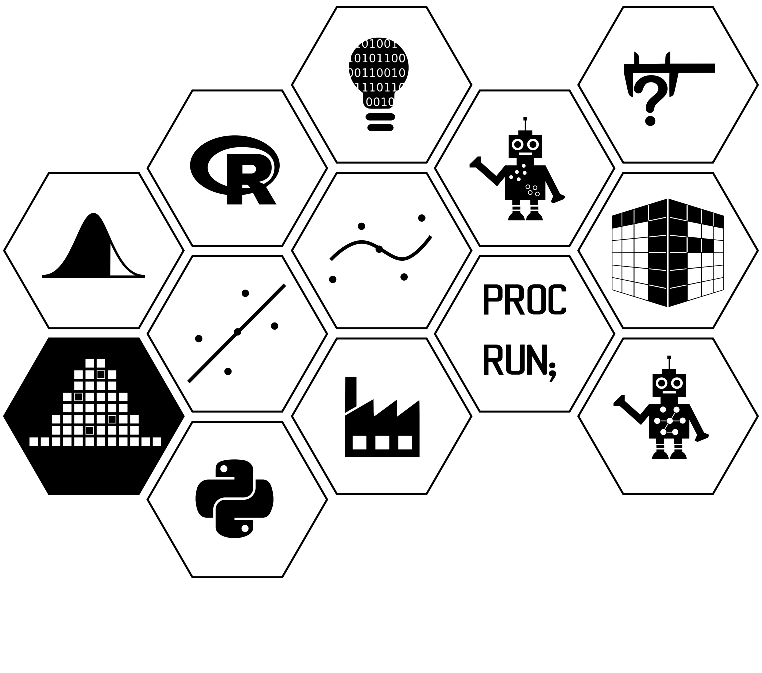
Probability and Sampling Fundamentals

Week 3: Discrete Random Variables



DATA ANALYTICS
GLASGOW

Random Variables

Week 3 learning material aims

The material in week 3 covers:

- Definition of a random variable;
- Describe the range space of a random variable;
- Definition of a discrete random variable;
- Describe the probability mass function of a discrete random variable;
- Calculate the expectation and variance of a discrete random variable;
- Describe the cumulative distribution function of a discrete random variable;
- Identify and describe the properties of the Bernoulli distribution, binomial distribution and Poisson distribution.

This week, we will define random variables, in particular discrete random variables. In order to understand the nature of random variables, we will re-visit some examples we have seen over the last two weeks.

First let's reconsider the example where we rolled a dice three times (from week 1)

Example 1

Rolling a dice three times

Suppose we rolled a dice three times. How many possible ways could we roll a 5 exactly two times?

Recall that we were able to answer this question in two ways depending on how we defined the sample space.

Method 1

In method 1 we considered each of the three rolls separately and within each roll, we noted down whether we rolled a five of not and defined

$$E = \{5\}$$
$$E' = \{1, 2, 3, 4, 6\}.$$

In the first roll, we can either roll a 5 or not. Therefore we will either be in E or E'. Likewise, in the second and third rolls, we can either roll a 5 or not.

So, for example, if we first rolled a four, then a six, and then a five, we would denote this as $E'E'E$, because we did not roll a five in the first two rolls, and then rolled a five in the final roll.

Putting everything together we obtain.

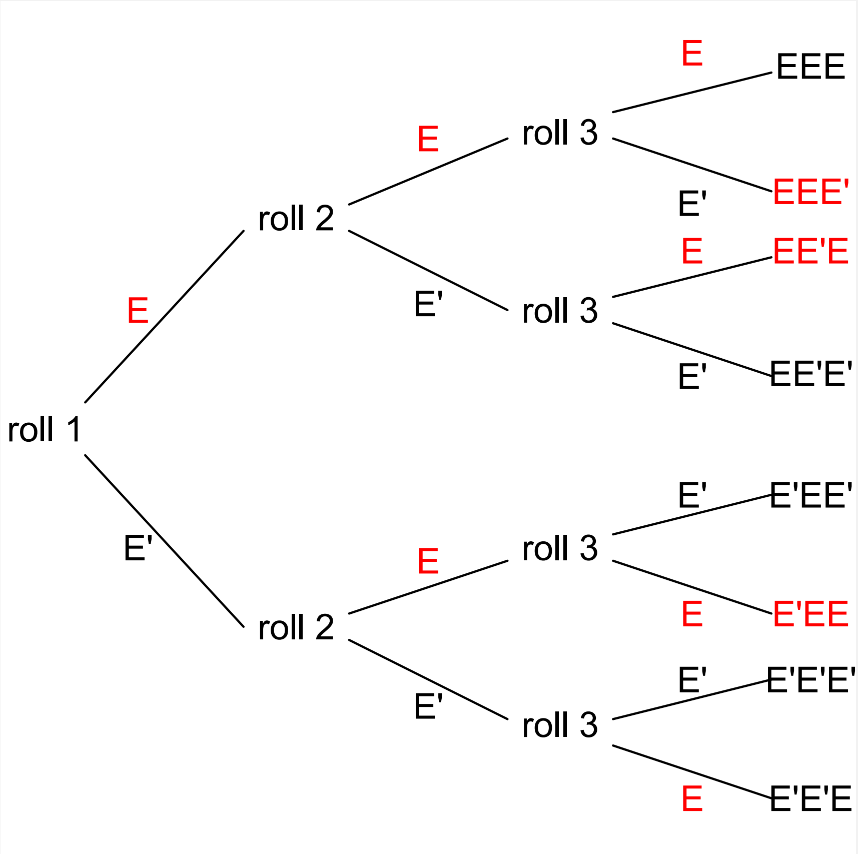


Figure 1

In this approach we can write down the sample space as

$$S_{\text{method 1}} = \{EEE, EEE', EE'E, EE'E', E'EE', E'EE, E'E'E, E'E'E'\}$$

Method 2

If all we are interested in is how many fives we have rolled, then method 1 provides more information that we need.

We could redefine the experiment in terms of just recording the number of times we roll a five out of a total of three rolls. In our example of rolling a four, then a six, and then a five, we would record this as 1, as we have rolled five one time only.

When rolling a dice 3 times, we could roll a five either 0, 1, 2 or 3 times. Thus, we can define our sample space as

$$S_{\text{method 2}} = \{0, 1, 2, 3\}$$

referring now to the total number of times we roll a five.

The two methods we have looked at each provide a way of describing the same experiment and in this example there are even other ways of recording the outcome of the experiment.

If we are interested in calculating the probability of the rolling a five a given number of times, the two methods are equivalent and we obtain the same results irrespective of which method we use. We can see this in the tree diagram above.

We can see this by writing down the event "Rolling a five x times" for each of the two methods.

Event	**Method 1**	**Method 2**
Rolling a five 0 times	$E_0 = \{E'E'E'\}$	$E_0 = \{0\}$
Rolling a five 1 time	$E_1 = \{EE'E', E'EE', E'E'E\}$	$E_1 = \{1\}$
Rolling a five 2 times	$E_2 = \{EEE', EE'E, E'EE\}$	$E_2 = \{2\}$
Rolling a five 3 times	$E_3 = \{EEE\}$	$E_3 = \{3\}$

Method 2 provides a *numerical* summary of the experiment. Method two is actually a special case of the what is called a random variable.

So, let's look at the relationship between the two methods again. We can map each outcome using method 1 (for example the outcome $EE'E$) to exactly one outcome using method 2 (for example the outcome 1). In other words, the outcome in method 2 is a *function* of the outcome in method 1. The figure below illustrates this idea.

We can make this relationship mathematically precise by defining a random variable.

Definition 1

Random Variable

A random variable is a function

$$X : S \to \mathbb{R},$$

which maps each elementary outcome element $s \in S$ to a real number $X(s)$.

Random variables are a key concept in Probability. Because of their nature as being real-valued we can order them and perform arithmetic with them, which means we can do a lot more with them than with general sample spaces. For example, we can look at statements like "What is the probability of rolling a five at least once (i.,e,. $X \geq 1$)?" or calculate the average number of fives we can expect to roll.

We couldn't do this with the outcomes from method 1. It would make so sense to talk about at least rolling $E'EE$, nor would a calculation of the form $(E'EE + EE'E)/2$ make any sense.

Example 2

In our example of rolling a dice three times, we can now define method 2 as a random variable X. Let method 1 define the sample space.

$$S = S_{\text{method 1}} = \{EEE, EEE', EE'E, EE'E', E'EE', E'EE, E'E'E, E'E'E'\}.$$

Then we can write the corresponding outcome of method two as $X(S)$:

$$\begin{aligned} X(E'E'E') &= 0 \\ X(EE'E') &= 1 \\ X(E'EE') &= 1 \\ X(E'E'E) &= 1 \\ X(EEE') &= 2 \\ X(EE'E) &= 2 \\ X(E'EE) &= 2 \\ X(EEE) &= 3 \end{aligned}$$

X is thus a random variable that represents the number of times five appears from three rolls.

In the example, it was very natural to define X since we were counting the number of times we rolled a five. In some cases, elements in a sample space may already be numerical or they may have no natural ordering or clear numerical equivalence.

Example 3

Goals scored

Write down a random variable to describe the total number of goals scored in a football match.

The outcome of this experiment is the total number of goals scored which is numerical. We know the number of goals scored is a positive integer including zero

$$S = \mathbb{Z}_{\geq 0}$$

and we could define a random variable X such that $X(s) = s$ for all $s \in S$. Here X corresponds to the total number of goals scored in football match.

Video

Probability and Sampling Fundamentals - Introduction to discrete random variables

Duration 3:05

Example 4

Student satisfaction

Suppose we surveyed everyone on this course and asked how satisfied they were with this course with possible responses being

$$\text{very satisfied, satisfied, neutral, dissatisfied or very dissatisfied.}$$

Write down a random variable corresponding to this experiment.

The sample space is

$$S = \{\text{very satisfied}, \text{ satisfied}, \text{ neutral}, \text{ dissatisfied}, \text{ very dissatisfied}\}$$

and X a random variable corresponding to the student satisfaction rate in this course.

Although the sample space is not numerical, you may think there is a natural ordering to the elements of S. For example, a score of "very satisfied" is better than "satisfied" which is better than "neutral" etc. We then set

$$\begin{aligned} X(\text{very satisfied}) &= 5 \\ X(\text{satisfied}) &= 4 \\ X(\text{neutral}) &= 3 \\ X(\text{dissatisfied}) &= 2 \\ X(\text{very dissatisfied}) &= 1 \end{aligned}$$

such that "very satisfied" is higher than "satisfied" which is higher than "neutral" etc. However, alternative codes are equally as valid such as

$$\begin{aligned} X(\text{very satisfied}) &= -5 \\ X(\text{satisfied}) &= 10 \\ X(\text{neutral}) &= 0 \\ X(\text{dissatisfied}) &= 3 \\ X(\text{very dissatisfied}) &= 2 \end{aligned}$$

$$\begin{aligned} X(\text{very satisfied}) &= -1 \\ X(\text{satisfied}) &= 3 \\ X(\text{neutral}) &= 7 \\ X(\text{dissatisfied}) &= 2 \\ X(\text{very dissatisfied}) &= -5 \end{aligned}$$

$$\begin{aligned} X(\text{very satisfied}) &= 2 \\ X(\text{satisfied}) &= 1 \\ X(\text{neutral}) &= 0 \\ X(\text{dissatisfied}) &= -1 \\ X(\text{very dissatisfied}) &= -2 \end{aligned}$$

The main point here is that we want to be able to distinguish between the five levels of satisfaction and therefore each level of satisfaction should have a unique numerical value $X(s)$.

Example 5

Time to run a mile

Suppose I asked everyone in this class how long it would take them to run a mile. Write down a random variable to describe the time take to run a mile.

The outcome of this experiment is the time take to run a mile and has to be a positive real number excluding zero

$$S = \mathbb{R}_{>0}.$$

We can define a random variable X such that $X(s) = s$ for all $s \in S$. Here X corresponds to the time taken to run a mile.

In Example 4, the sample space was not naturally numerical and in order to define a random variable we were required to provide numerical codes for each element of the sample space. It is therefore useful to be able to distinguish between the sample space of an experiment and the values $X(s)$ for $s \in S$.

Definition 2

Range space

The range space R_X of a random variable X is the set of all possible realisations $X(s)$ for all outcomes $s \in S$. More mathematically, we can write this as

$$R_X = \{X(s) : \ s \in S\}.$$

A major distinction between these four examples is the type of range space. Recall from **week 1** that a sample space can fall into three categories

- Finite
- Countable
- Uncountable

Range spaces also fall into one of these categories.

- Ref://rolldice1 and Example 4 are examples of finite range spaces.
- Ref://goals is an example of a countable range space.
- Ref://timemile is an example of an uncountable range space (assuming we have an arbitrarily precise clock)

Discrete Random Variables

The type of range space defines the type of random variable. We will see that if we perform calculations we need to treat random variables with an uncountable range space differently from random variables with a finite or countable range space.

Definition 3

Discrete Random Variable

A random variable that has a finite or countable range space is a discrete random variable.

Task 1

Discrete Random Variable

From the four examples rolling a dice three times, goals scored, student satisfaction and time to run a mile, which random variables are discrete?

Dice: DISCRETE
Goals: DISCRETE
Student Satisfaction: DISCRETE
Mile: NOT DISCRETE

The purpose of defining experiments in terms of random variables and range spaces is to answer questions such as

1. What is the probability that a randomly selected student from this class can run a mile in under 8 minutes?
2. What is the expected time taken to run a mile?
3. What is the probability that students found this course satisfactory or better?
4. What is the average rating of the course from students?
5. What is the probability that I roll a 5 exactly two times from three rolls?
6. What is the expected number of times I roll a 5 from three rolls?
7. What is the average number of goals scored in a football match,

In order to answer questions like these we need to work out the probability of values $x \in R_X$. This week, we will focus only on discrete random variables.

Random variables with uncountable range spaces, called continuous random variables, will be covered in week 5.

Probability mass function

Given a discrete random variable X with range space R_X, we often need to be able to work out the probability of each value $x \in R_X$. Every random variable X has a corresponding function that allows us to do so. It is called the probability mass function (pmf).

Definition 4

Probability Mass Function

The probability mass function (pmf) of a discrete random variable X is a function $p : R_x \to (0, 1]$ such that $p_X(x) = P(X = x)$ for all $x \in R_X$.

In other words, the probability mass function gives for each possible realisation x from the range space R_X the probability $P(X = x)$ that we observe exactly that outcome x.

One can show that a probability mass function satisfies the following three properties.

1. $p_X(x)$ is positive for all elements in the range space, i.e. $p_X(x) > 0$ for all $x \in R_X$
2. $p_X(x)$ sums to 1, i.e. $\sum_{x \in R_X} p_X(x) = 1$
3. To find the probability $P(A)$ of any event A, we just sum up the probabilities of all $x \in A$, i.e. $P(X \in A) = \sum_{x \in A} p_X(x)$ for any event $A \subset R_X$.

Any function which satisfies 1. and 2. is a valid probability mass function.

In order to illustrate the concept of a probability mass function we will take the example of flipping a coin.

Example 6

Flipping a coin

Suppose I flipped a fair coin twice and counted the number of heads.

1. Write down the sample space.
2. Define a random variable X associated with this experiment.
3. Write down R_X.
4. Find the probability mass function $p_X(x) = f(x) = P(X = x)$.
5. Check the probability mass function satisfies the three listed criteria.

To begin with we know that if we flip a coin twice then we can either

1. Flip head then tails.
2. Flip head then head.
3. Flip tails then head.
4. Flip tails then tails.

Using the notation H=head and T=tails

$$S = \{HT, HH, TH, TT\}.$$

We can define a random variable $X : S \to \mathbb{R}$ such that X corresponds to the number of heads from two coin flips. Let

$$\begin{aligned} X(HT) &= 1 \\ X(HH) &= 2 \\ X(TH) &= 1 \\ X(TT) &= 0 \end{aligned}$$

And so $R_X = \{0, 1, 2\}$ and we can see that there are four possible outcomes in this experiment.

In order to find the probability mass function associated with this experiments, we need to find $P(X = 0), P(X = 1)$ and $P(X = 2)$. We know that from the four possible outcomes, there is only one way to obtain $X = 0$, there are two ways to obtain $X = 1$ and only one way to obtain $X = 2$. Therefore,

x	0	1	2
P(X=x)	1/4	2/4	1/4

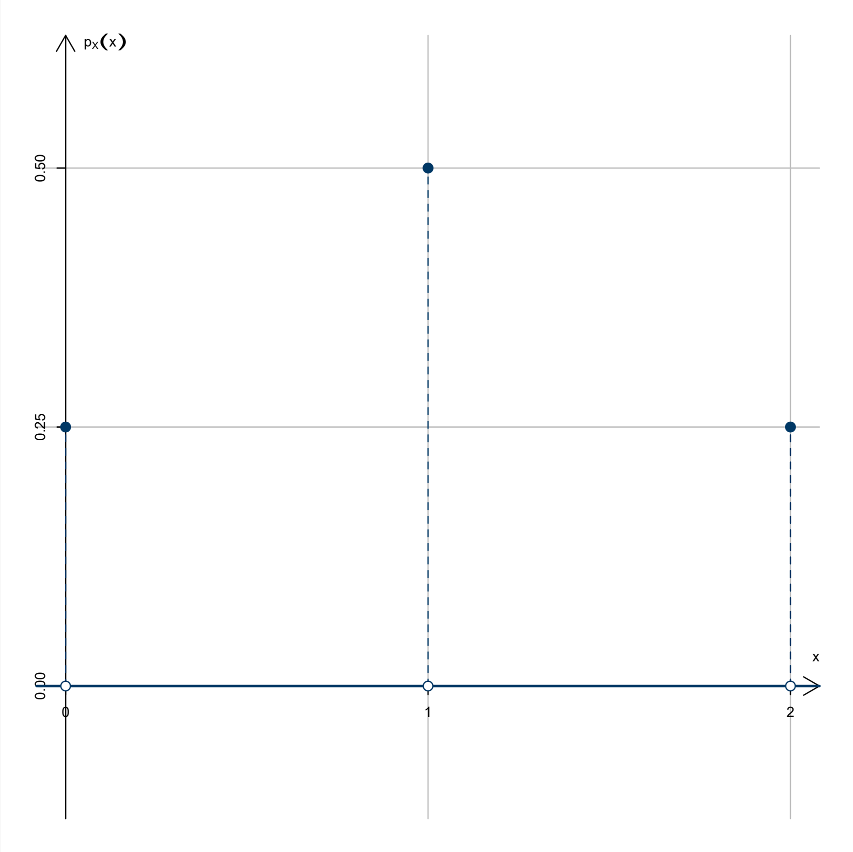


Figure 2

Lastly, we check our probability mass function (pmf) satisfies the necessary conditions.

1. All of $p_X(x) = P(X = x) > 0$ for $x = 0, 1, 2$.

2. $\sum_{x \in \{0,1,2\}} p_X(x) = p_X(0) + p_X(1) + p_X(2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$

3. Let's just consider a few possible examples of events.

   a. What is the probability that we flip at least one head? Elements HT, HH and TH satisfy flipping at least one head and the probability of this event is 3/4. Or, the probability of at least one head corresponds to $P(X \geq 1) = P(X = 1) + P(X = 2) = 3/4$.

   b. What is the probability that we flip less than two heads? HT, TT and TH satisfy flipping less than two heads and the probability of this event is 3/4. Or, the probability of flipping less than two heads corresponds to $P(X < 2) = P(X = 0) + P(X = 1) = 3/4$.

This example illustrates how to 'translate' an experiment into a well defined random variable and probability mass function.

Let's return to the example of rolling a dice three times.

Example 7

Let's return to the example in which we rolled a dice three times and counted the number of times we rolled a 5. What is the probability that we rolled a five exactly two times?

We have already defined the sample space

$$S = \{EEE, EEE', EE'E, EE'E', E'EE', E'EE, E'E'E, E'E'E'\}$$

where $E = \{5\}$ and $E' = \{1, 2, 3, 4, 6\}$. The range space

$$R_X = \{0, 1, 2, 3\},$$

and random variable

$$X = \text{the number of times we roll a five.}$$

What we want to know now is $P(X = 2)$.

We can begin by setting up a table similar to the previous example

x	0	1	2	3
P(X=x)				

In a single roll of a fair dice, we know

$$P(\text{we roll a 5}) = P(E) = 1/6$$
$$P(\text{we do not roll a 5}) = P(E') = 5/6.$$

What is the probability that in an experiment we roll two 5's and another number in $\{1, 2, 3, 4, 6\}$?

Knowing that the three rolls are independent, the multiplication rule introduced in **week 1** helps us answer this question.

$$P(\text{exactly two fives out of three}) = 1/6 \times 1/6 \times 5/6.$$

We also know there are three possible ways to roll a 5, namely

$$EEE', EE'E, E'EE$$

and so

$$
\begin{aligned}
P(X=2) &= P(EEE' \cup EE'E \cup E'EE) \\
&= P(EEE') + P(EE'E) + P(E'EE) \\
&= (1/6 \times 1/6 \times 5/6) + (1/6 \times 1/6 \times 5/6) + (1/6 \times 1/6 \times 5/6) \\
&= 3(1/6 \times 1/6 \times 5/6) \\
&= 0.069
\end{aligned}
$$

Let's add this to the table

x	0	1	2	3
P(X=x)			0.069	

Task 2

Rolling a dice three times.

Following on from the last example, fill in the rest of the table

x	0	1	2	3
P(X=x)	.5787	.347	0.069	.0046

Moments

Moments are numerical summaries of a distribution. In this section we will learn about the first two moments:

- the expected value, which is a measure of average location, and
- the variance, which is a measure of the spread.

Expected value

Example 8

Estimating the average number of fives

Suppose now I asked you how many times could I expected to roll a 5 out of three rolls. One way to answer this question would be to repeat the experiment and count the number of times you roll a 5 from three rolls. Suppose you repeated this 15 times as illustrated below

	outcome 1	outcome 2	outcome 3	count
roll 1	6	4	2	0
roll 2	1	2	2	0
roll 3	2	6	5	1
roll 4	6	2	1	0
roll 5	6	4	1	0
roll 6	3	5	3	1
roll 7	6	6	5	1
roll 8	2	1	1	0
roll 9	3	3	6	0
roll 10	5	4	3	1
roll 11	2	2	4	0
roll 12	4	4	1	0
roll 13	2	4	1	0
roll 14	3	1	4	0
roll 15	6	6	3	0
			average count =	0.27

Figure 4

We can see that the majority of times, you roll a five either zero or one time out of three rolls. How can we summarise this information?

One way to summarise this information is to take the average number of times you roll a five from three rolls. That is, calculate

$$\frac{\text{the number of times you roll five from three rolls in all experiments}}{\text{the number of times you repeat the experiment}} = \frac{4}{15} = 0.27.$$

Let's start by calculating the numerator, i.e. the total number of times we obtained a five across all experiments. We just need to calculate the sum of the right-most column in the above table showing the outcomes of the 15 rolls.

$$0+0+1+0+0+1+1+0+0+1+0+0+0+0+0$$

Collecting the terms this is

$$0 \times 11 + 1 \times 4 + 2 \times 0 + 3 \times 0 = 4$$

Collecting the terms is nothing other than creating a tally of how often we obtained each outcome.

Outcome	Absolute frequency	Relative frequency
$X = 0$	11	11/15=0.73
$X = 1$	4	4/15=0.27
$X = 2$	0	0/15=0
$X = 3$	0	0/15=0

Thus the average number of fives is

$$\frac{0 \times 11 + 1 \times 4 + 2 \times 0 + 3 \times 0}{15} = \frac{4}{15} = 0.27$$

We can rewrite the calculation using relative frequencies. It simply is the weighted average of the possible outcomes, using the relative frequencies as weights.

$$0 \times \frac{11}{15} + 1 \times \frac{4}{15} + 2 \times \frac{0}{15} + 3 \times \frac{0}{15} = \frac{4}{15} = 0.27$$

This value gives us an estimate of the average, or expected, number of times you rolled a five.

In the example, it was easy to simulate the experiment and calculate the average number of times you roll a five. But we cannot do this for every conceivable experiment and we might want to obtain a more precise answer. The *expected value* gives us exactly this information we were looking for.

For a discrete random variable, we can formally define the expected value.

Definition 5

Expected Value of a Discrete Random Variable

The expected value (or expectation) of a discrete random variable X is defined as

$$
\begin{aligned}
E(X) &= \sum_{x \in R_X} x P(X = x) \\
&= \sum_{x \in R_X} x p_X(x).
\end{aligned}
$$

We can re-write the definition without $\sum$-notation if we assume that $R_X = \{x_1, x_2, x_3, \ldots\}$. Then

$$
\begin{aligned}
E(X) &= x_1 \times P(X = x_1) + x_2 \times P(X = x_2) + x_3 \times P(X = x_3) + \ldots \\
&= x_1 \times p_X(x_1) + x_2 \times p_X(x_2) + x_3 \times p_X(x_3) + \ldots
\end{aligned}
$$

In other words, the expected value is the weighted average of the possible outcomes in the range space R_X, using the probabilities as weights. We have done a similar calculation in **Example 8**, where we have used the relative frequencies as weights.

The Greek letter μ or μ_X is often used to denote the expected value $E(X)$.

The expected value of random variable is a measure of the average location of the distribution.

Example 9

Expected number of fives

Using the probability mass function we derived earlier (Task 2) we can now calculate the expected number of fives when rolling a dice three times.

x	0	1	2	3
P(X=x)	0.579	0.347	0.069	0.005

Using the definition of the expectation of a discrete random variable,

$$
\begin{aligned}
E(X) &= \sum_{x \in R_X} x P(X = x) \\
&= \sum_{x \in R_X} x p_X(x) \\
&= 0 \times p_X(0) + 1 \times p_X(1) + 2 \times p_X(2) + 3 \times p_X(3) \\
&= 0 \times 0.579 + 1 \times 0.347 + 2 \times 0.069) + 3 \times 0.005 \\
&= 0.5.
\end{aligned}
$$

So on average we obtain half a five every time we roll the dice three times. Of course, this is a value that can never occur. This is true in general, the expected value does not need to lie in R_X.

Video

Raffle ticket prize

Duration 5:10

Task 3

Rolling a dice three times.

Compare the average value estimated after repeating this experiment 15 times and the expected value calculated using the formal definition.

Should the two values be equal? If so, why are they not?

Because you can never account for every single iteration of this experiment that will take place. As you increase the number of repetitions, the estimated average value converges to the calculated expected value.

A very useful property of expectation is that it can be extended to linear functions of X.

Expectation of linear functions of a random variable

Let X be a discrete random variable with expectation $E(X)$ and let a and $b \in \mathbb{R}$. Then

$$E(aX+b) = \quad = aE(X)+b$$

This is a useful property, as it allows us deal with changes of units of measurement. If the expected value of the height of a child in centimetres is 127cm, then the expected value of the height in inches is 127/2.54=50in.

Note that we can swap the order of evaluating a function and calculating the expectation only if the function is linear, i.e $E(g(X))$ is, in general, not the same as $g(E(X))$. In this was true, then the definition of the variance we will soon see would be pretty pointless, as it always would be 0.

Supplement 1

Proof

$$
\begin{aligned}
E(aX+b) = &= \sum_{x\in R_X}(ax+b)p_X(x)\\
&= a\underbrace{\left(\sum_{x\in R_X} xp_X(x)\right)}_{=E(X)} + b\underbrace{\left(\sum_{x\in R_X} p_X(x)\right)}_{=1}\\
&= aE(X)+b.
\end{aligned}
$$

We could have written the proof also without $\sum$ notation. Assuming $R_X = \{x_1, x_2, x_3, \ldots\}$,

$$
\begin{aligned}
E(aX+b) &= (ax_1+b)p_X(x_1) + (ax_2+b)p_X(x_2) + (ax_3+b)p_X(x_3) + \ldots\\
&= ax_1p_X(x_1) + bp_X(x_1) + ax_2p_X(x_2) + bp_X(x_2) + ax_1p_X(x_3) + bp_X(x_3) + \ldots\\
&= ax_1p_X(x_1) + ax_2p_X(x_2) + ax_1p_X(x_3) + \ldots + bp_X(x_1) + bp_X(x_2) + bp_X(x_3) + \ldots\\
&= a\underbrace{(x_1p_X(x_1) + x_2p_X(x_2) + x_3p_X(x_3) + \ldots)}_{=E(X)} + b\underbrace{(p_X(x_1) + p_X(x_2) + p_X(x_3) + \ldots)}_{=1}\\
&= aE(X)+b
\end{aligned}
$$

Variance and standard deviation

The expected value gives us information about the average value of a distribution. Another important quantity for any random variable is the spread.

Example 10

Goals scored

Suppose we are interested in the number of goals scored in Scottish football matches during the 2018/2019 season.

In total 228 games were played. In order to assess the number of goals scored, we decided to watch 15 matches and noted down the total number of goals in each match and we repeated this 10 times. The results are given below.

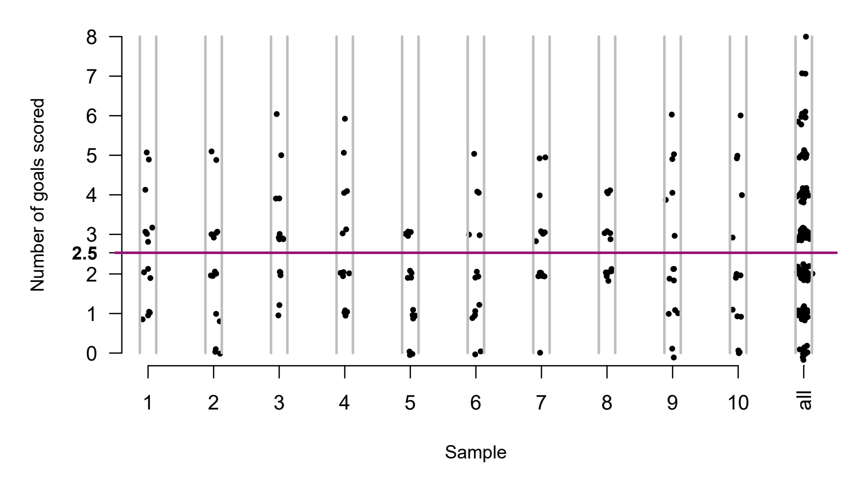


Figure 5

Each point on this plot shows the total number of goals in a single match. Each of the 10 samples, segregated by the grey lines, contains 15 data points. The last column shows all data from the 228 matches.

The average, or expected, number of goals was 2.5 (indicated by the purple horizontal line). Within the total 228 matches, the minimum number of goals was zero and the maximum number of goals was 8.

We can see that within each of the 10 samples, there was a range of total goals scored. However, most of the time, we observe between one and three total goals scored. Although we expected 2.5 goals in each match, we can see that actually there is some spread around this number

Likewise, when we look at the data from the 228 matches, there is some spread around the expected value. Most of the games had a total of 1, 2 or 3 goals but we can see that two games had a total of 7 goals and one game had a total of 8 goals.

We have identified two sources of spread. Firstly, there are differences between the 10 samples. Secondly, there are differences in the number of goals scored when we look at the 228 matches. The total number of goals in some games lie very closed to the expected 2.5 whereas some games lie quite far away.

We want to be able to quantify the latter form of spread.

The variance of a distribution $\mathrm{Var}(X)$ is a measure of the spread. A random variable with a high variance has a more spread-out distribution than one with a lower variance.

Definition 6

Variance of a Random Variable

The variance of a random variable X is defined as

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= E[X^2]-\mu^2\end{aligned}$$

where $\mu = E(X)$.

The variance is the average squared distance between a random draw from X and its mean $\mu = E(X)$.

Given that the variance is defined as the expected value of a squared difference it cannot be negative. One can even show that unless a random variable is constant, the variance must be positive.

The Greek letter σ^2 (or σ_X^2) is often used to denote the variance.

Supplement 2

Equivalence of the two definitions of the variance

We have given two formulae for calculating the variance. We will now show that they are equivalent by deriving the latter formula from the former.

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= E(X^2-2X\mu+\mu^2) \\ &= E(X^2)-2\underbrace{E(X)}_{=\mu}\mu+\mu^2 \\ &= E(X^2)-2\mu^2+\mu^2 \\ &= E(X^2)-\mu^2\end{aligned}$$

Example 11

Returning to our example of rolling a dice three times, we can now calculate the variance of X given $\mu = E(X) = 0.5$ in two ways.

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= (0-0.5)^2(0.579)+(1-0.5)^2(0.347)+(2-0.5)^2(0.069)+(3-0.5)^2(0.005) \\ &= (0.25)(0.579)+(0.25)(0.347)+(2.25)(0.069)+(6.25)(0.005) \\ &= 0.418.\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2]-[E(X)]^2 \\ &= \sum_{x\in R_X} x^2 p_X(x) - (0.5)^2 \\ &= [0^2(0.579)+1^2(0.347)+2^2(0.069)+3^2(0.005)]-0.25 \\ &= [0+0.347+4(0.069)+9(0.005)]-0.25 \\ &= 0.668-0.255 \\ &= 0.418.\end{aligned}$$

CALCULATE VARIANCE

We have looked at the expected value of linear functions of a random variable. We will now look at a formula for the variance of linear functions.

Variance of linear functions of a random variable

Let X be a discrete random variable with variance $\mathrm{Var}(X)$ and let a and $b \in \mathbb{R}$. Then

$$\mathrm{Var}(aX+b) = \; = a^2\mathrm{Var}(X)$$

In other words, if we scale the random variable by a factor of a, then the variance is multiplied by its square. This is because the variance is a squared distance.

We can also see that the additive term b drops out. This makes sense, as the variance is a measure of the spread of a distribution and the additive term b corresponds to shifting the distribution to the left or to the right, which does not change the spread.

Supplement 3

Let's write $Y = aX + b$ and we will use $\mu_Y = E(Y)$. We have seen that $\mu_Y = E(Y) = E(aX+b) = aE(X)+b = a\mu_X + b$.

We can now calculate the variance of $Y = aX + b$.

$$\begin{aligned}\mathrm{Var}(Y) &= E[(Y-\mu_y)^2] \\ &= E[(aX+b-(a\mu_X+b))^2] \\ &= E[a^2(X-\mu_X)^2] \\ &= a^2E[(X-\mu_X)^2] \\ &= a^2\mathrm{Var}(X)\end{aligned}$$

The fact that the variance is an average squared distance, makes its interpretation slightly odd. Suppose we measured the height of people in metres (m) then both the expectation would be measured in metres (m), whereas the variance would be measured in square meters (m^2).

A closely related measurement to variance is the standard deviation, which addresses the problem.

Definition 7

Standard Deviation of a Random Variable

The standard deviation of a random variable X is defined as

$$sd(X) = \sqrt{\mathrm{Var}(X)}$$

The standard deviation is often denoted as σ or σ_X.

The standard deviation is a little more intuitive than the variance, as it is on the same scale as the measurements itself, i.e. has the same units as the values of the random variable.

Example 12

Returning to our example of rolling a dice three times. We can now calculate the standard deviation of X given $\mathrm{Var}(X) = 0.418$.

$$\begin{aligned} sd(X) &= \sqrt{\mathrm{Var}(X)} \\ &= \sqrt{0.418} \\ &= 0.647. \end{aligned}$$

Both the variance and the standard deviation tell us about the spread of the distribution around its expected value.

- A small standard deviation implies the distribution of X is narrowly concentrated around the expected value.
- A large standard deviation implies the distribution of X is widely spread around the expected value.

This of course gives no indication of what is small or large. This depends on the units of measurement.

Cumulative Distribution Function

We have so far described the distribution of a random variable using its probability mass function $p_X(x) = P(X = x)$.

Often we are however interested in probabilities of the form $P(X \leq x)$ or $P(X > x)$. We can work out these probabilities from the probability mass function, but this is quite cumbersome.

$$P(X \leq x) = \sum_{x' \leq x} p_X(x')$$
$$P(X > x) = \sum_{x' > x} p_X(x')$$

Probabilities of the form $P(X \leq x)$ or $P(X > x)$ are more easily obtained from the cumulative distribution function.

Definition 8

Cumulative Distribution Function

Let X be a random variable with range space R_X and pmf $p_X(x) = P(X = x)$. Then the cumulative distribution function (cdf) is defined as

$$F_X(x) = P(X \leq x).$$

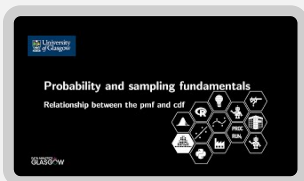
Using the fact that $\{X \leq x\}' = \{X > x\}$ we can use the cdf to express a variety of probabilities.

$$P(X \leq x) = F_X(x)$$
$$P(X > x) = 1 - F_X(x)$$
$$P(a < X \leq b) = F_X(b) - F_X(a)$$

Continuing with the rolling a dice three times example, we can now derive the cumulative distribution function (cdf).

Video

Probability and Sampling Fundamentals - Relationship between the pmf and cdf



Duration 3:05

Example 13

We defined the random variable X= the number of times we roll a 5 from 3 dice rolls and range space $R_X(x) = \{0, 1, 2, 3\}$ such that

x	0	1	2	3
P(X=x)	0.579	0.347	0.069	0.005

We can now construct the cdf.

$$F_X(0) = P(X \leq 0) = P(X = 0) = 0.579$$
$$F_X(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 0.579 + 0.347 = 0.926$$
$$F_X(2) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0.579 + 0.347 + 0.069 = 0.995$$
$$F_X(3) = P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0.579 + 0.347 + 0.069 + 0.005 = 1$$

A plot of the pmf and cdf are shown below.

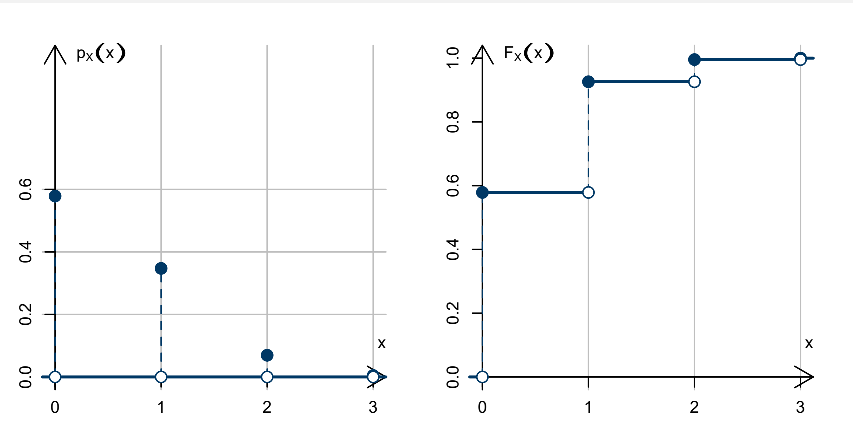


Figure 6

Notice the cumulative distribution function lies between 0 and 1. For values of $x \notin R_X$ $P(X = x) = 0$ therefore $P(X < 0) = 0$ and $P(X > 3) = 0$.

For discrete random variables, the cdf is a step function that starts at 0 and increases to 1 in positive steps.

So, to be more precise, the cdf in the above example should be written as

The cumulative distribution function contains exactly the same information as the probability mass function, just in a different format. We can retrieve the probability mass function from the cumulative distribution function and vice versa.

In general their relationship if given by the equation

$$F_X(x) = \sum_{x' \leq x} p_X(x')$$

However, it is easier to state their relationship if we assume a finite range space $R_X = \{x_1, \ldots, x_n\}$.

$$\begin{aligned} F_X(x_1) &= p_X(x_1) \\ F_X(x_i) &= p_X(x_1) + \ldots + p_X(x_i) \\ F_X(x_n) &= 1 \\ p_X(x_i) &= F_X(x_i) - F_X(x_{i-1}) \qquad \text{for} i = 2, \ldots, n \end{aligned}$$

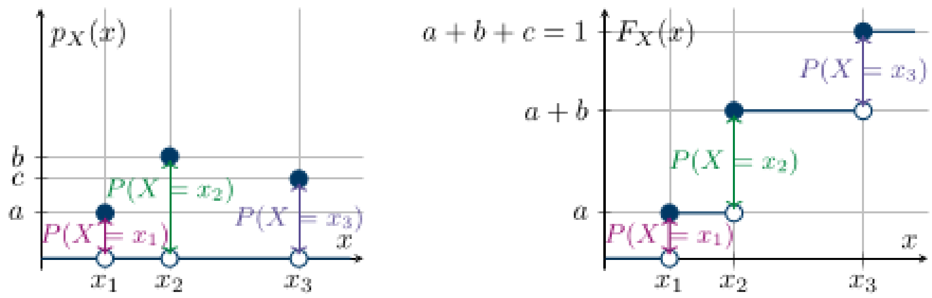


Figure 7

We can now use this to express even more probabilities in terms of the pdf and the cdf.

$$\begin{aligned} P(X \leq x_i) &= p_X(x_1) + \ldots + p_X(x_i) \\ &= F_X(x_i) \\ P(X < x_i) &= p_X(x_1) + \ldots + p_X(x_{i-1}) \\ &= P(X \leq x_i) - P(X = x_i) = F_X(x_i) - p_X(x_i) \\ P(X > x_i) &= p_X(x_{i+1}) + \ldots + p_X(x_n) \\ &= 1 - P(X \leq x) = 1 - F_X(x_i) \\ P(X \geq x_i) &= p_X(x_i) + \ldots + p_X(x_n) \\ &= 1 - P(X < x) = 1 - F_X(x_i) + p_X(x_i) \end{aligned}$$

Example 14

We can now use these properties to derive the pmf from the cdf in the example of counting the number of fives.

$$
\begin{aligned}
P(X=0) &= F_X(0) = 0.579 \\
P(X=1) &= F_X(1) - F_X(0) = 0.926 - 0.579 = 0.347 \\
P(X=2) &= F_X(2) - F_X(1) = 0.995 - 0.926 = 0.069 \\
P(X=3) &= F_X(3) - F_X(2) = 1 - 0.995 = 0.005
\end{aligned}
$$

In other words, for each $x \in R_X$ we are figuring out the lengths of each of the arrows shown in the plot below.

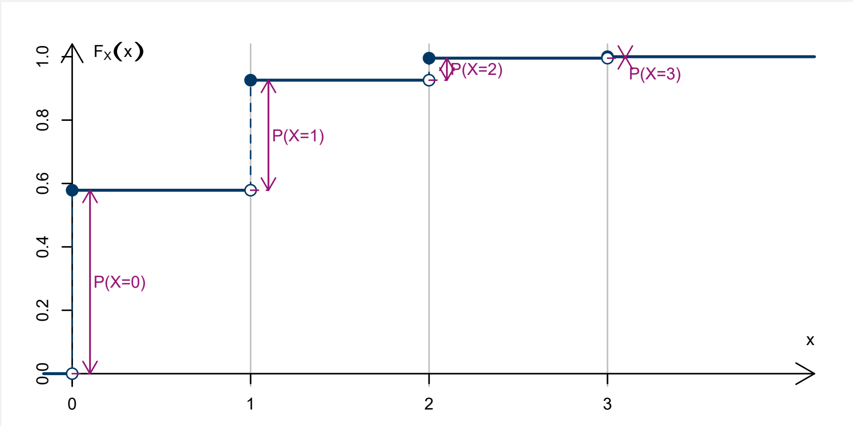


Figure 8

Task 4

Number of languages cdf

Suppose I randomly sampled 1000 school children and asked them how many languages they spoken. The minimum number was 1 and the maximum was 6. Let

$$X = \text{number of languages spoken by a school child}$$

such that

x	1	2	3	4	5	6
P(X=x)	0.387	0.297	0.172	0.101	0.035	.008

Compute the cdf.

Until now, we have explored only generic examples of discrete random variables. There are several standard discrete distributions that you should know, which we will look at next.

The Bernoulli distribution

A Bernoulli random variable has a range space that can take only two values normally denoted by 0 and 1. We assume that $P(X=1)=\theta$ and $P(X=0)=1-\theta$ for some probability θ.

Therefore

$$\begin{aligned} E(X) &= \sum_{x \in R_X} x p_X(x) \\ &= 0(1-\theta)+1(\theta) \\ &= \theta \end{aligned}$$

$$\begin{aligned} \operatorname{Var}(X) &= E[X^2]-[E(X)]^2 \\ &= 0^2(1-\theta)+1^2(\theta)-\theta^2 \\ &= \theta-\theta^2 \\ &= \theta(1-\theta) \end{aligned}$$

It is conventional to think of a Bernoulli random variables as a success or failure experiment with a success coded as 1 and a failure coded as 0.

Example 15

Flipping a coin

Suppose I flipped a coin and noted down if I flipped a head.

A "success" in this example is flipping a head and a failure is flipping a tail. Therefore we should code Head=1 and Tail=0.

We know that we are equally probable to flip a head or a tails and so $\theta=0.5$.

x	0	1
P(X=x)	0.5	0.5

:blockMath[209]

The binomial distribution

The binomial distribution is an extension to the Bernoulli distribution. Instead of completing an experiment once, suppose now we complete the experiment n times with the n experiments being independent.

Suppose I flipped a coin 100 times and noted down if I flipped a head after each flip. The number of heads then has a binomial distribution with $n = 100$. If the coin is fair $\theta = 0.5$.

Suppose I had a population of 200 people, of which 20 are smokers, and I drew a sample of 15 people and asked them whether they smoke. If I sampled with replacement, then every time I select a person, each person is equally likely to be drawn, so the probability of drawing a smoker is $\theta = \frac{20}{200} = 0.1$. I ask $n = 15$ people whether they smoke, thus the number of smokers I find has binomial distribution with $n = 15$ and $\theta = 0.1$. Note that if I had drawn without replacement, so that I don't ask the same person twice, the resulting distribution would not be binomial, as the Bernoulli experiments would be dependent. We will look at such sampling schemes in more detail an the end of the course.

Unlike our other example, writing down the pmf in a table is more complicated with a binomial random variable. It is conventional to instead present the pmf as a formula. In this case

$$P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

This formula is quite intuitive. All we are saying is from a total of n independent repetitions,

- the probability that we succeed x times is θ^x
- the probability that we fail the remaining $n - x$ times is $(1 - \theta)^{n-x}$
- the total number of ways to succeed x times out of n experiments is $\binom{n}{x}$.

We are counting the number of successes, so we can only observe integers and we cannot observe more successes than attempts, so the range space is $R_X = \{0, 1, \ldots, n\}$.

One can show that the expected value and variance of the binomial distribution are

$$\begin{aligned} E(X) &= n\theta \\ \mathrm{Var}(X) &= n\theta(1 - \theta) \end{aligned}$$

Supplement 4

Derivation of the mean and variance of the binomial distribution

We will derive the expected value and variance of the binomial distribution by relating it to the Bernoulli distribution.

The "trick" is that we write the number of successes as a sum of Bernoulli random variables.

$$X = X_1 + X_2 + \ldots + X_n$$

Within each experiment we code a success as 1 and a failure as 0, i.e. $X_i = 1$ if the i-th experiment is a success and $X_i = 0$ otherwise.

Imagine I tossed a coin three times and counted the number of heads. Suppose I obtained first tails, then heads, and finally once more heads. This would correspond to $X_1 = 0$, $X_2 = 1$ and $X_3 = 1$. In total, I obtained $X = X_1 + X_2 + X_3 = 0 + 1 + 1 = 2$ times heads.

For each Bernoulli trial we have seen that $E(X_i) = \theta$ and $\mathrm{Var}(X_i) = \theta(1-\theta)$.

We will see that the expected value of a sum of random variables is the sum of the expected values, thus

$$\begin{aligned} E(X) &= E(X_1 + \ldots + X_n) \\ &= E(X_1) + \ldots + E(X_n) \\ &= \theta + \ldots + \theta \\ &= n\theta \end{aligned}$$

We will also see that the variance of a sum of *independent* random variables is the sum of the variances, thus

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(X_1 + \ldots + X_n) \\ &= \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) \\ &= \theta(1-\theta) + \ldots + \theta(1-\theta) \\ &= n\theta(1-\theta) \end{aligned}$$

Example 16

Actually the example of rolling a dice three times is a great example of a binomial distribution.

Here, $n = 3$ since we rolled the dice three times. A success is rolling a 5 with $\theta = \frac{1}{6}$ and a failure is rolling a number in $\{1, 2, 3, 4, 6\}$ with probability $1 - \theta = \frac{5}{6}$.

Given $R_X = \{0, 1, 2, 3\}$ we can now commute the probability of each $x \in R_X$ using the binomial formula.

$$P(X=0) = \binom{3}{0}\left(\frac{1}{6}\right)^0\left(1-\frac{1}{6}\right)^{3-0} = 1\left(\frac{5}{6}\right)^3 = 0.579$$
$$P(X=1) = \binom{3}{1}\left(\frac{1}{6}\right)^1\left(1-\frac{1}{6}\right)^{3-1} = 3\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = 0.347$$
$$P(X=2) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right)^{3-2} = 3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = 0.069$$
$$P(X=3) = \binom{3}{3}\left(\frac{1}{6}\right)^3\left(1-\frac{1}{6}\right)^{3-3} = 1\left(\frac{1}{6}\right)^3 = 0.005.$$

The Poisson distribution

The Poisson distribution is commonly used to describe the number of events occurring within a given interval. Examples include the number of customers entering a shop within an hour, the number of accidents on one road during one year or the number of typos in these notes.

One key property of the Poisson distribution that is has, in contrast to the binomial distribution, no upper bound. Its range space consist of all (non-negative) integers $\{0, 1, 2, 3, \ldots\}$. Its probability mass function is

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

with λ the expected number of successes or events.

$$E(X) = \lambda$$
$$\mathrm{Var}(X) = \lambda$$

Supplement 5

Relationship to the binomial distribution

Consider again the example of the number typos in these notes. In this case, the Poisson distribution would appear to be a natural choice.

However, if we assumed that each word will not contain more than one typo, we could assume that the number of typos has a binomial distribution with n being the number of words in these notes and θ being the probability that a word is miss-spelt. n is quite large and θ (hopefully) quite small.

In more mathematical language, we can describe this situation as $n \to \infty$ and $\theta \to 0$. Lets assume that the expected number of typos $n\theta$ tends to some constant λ.

Now one can show that in this setting the binomial pmf tends to the Poisson pmf, i.e. there is no difference between the two distributions.

In a nutshell the argument is that (using $\theta = \frac{\lambda}{n}$)

$$
\begin{aligned}
P(X=x) &= \binom{n}{x}\theta^x(1-\theta)^{n-x} \\
&= \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{1}{x!}\underbrace{\frac{n!}{(n-x)!n^x}}_{\to 1}\lambda^x\underbrace{\left(1-\frac{\lambda}{n}\right)^{n-x}}_{\to e^{-\lambda}} \\
&\to \frac{e^{-\lambda}\lambda^x}{x!}.
\end{aligned}
$$

The details are omitted but the interested reader can find more information **here**.

Let's revisit the example of the number of languages spoke by school children.

Example 17

Number of languages

Suppose I randomly sampled 1000 school children and asked them how many languages they spoken. The minimum number was 1 and the maximum was 6. Let

$$X = \text{number of languages spoken by a school child}$$

x	1	2	3	4	5
P(X=x)	0.387	0.297	0.172	0.101	0.008

We previously found

$$
\begin{aligned}
E(X) &= 2.124 \\
&= \lambda
\end{aligned}
$$

Given $R_X = \{1, 2, 3, 4, 5, 6\}$ we can now commute the probability of each $x \in R_X$ using the Poisson pmf formula.

$$
\begin{aligned}
P(X=1) &= \frac{e^{-2.124}2.124^1}{1!} = 0.254 \\
P(X=2) &= \frac{e^{-2.124}2.124^2}{2!} = 0.270 \\
P(X=3) &= \frac{e^{-2.124}2.124^3}{3!} = 0.191 \\
P(X=4) &= \frac{e^{-2.124}2.124^4}{4!} = 0.101 \\
P(X=5) &= \frac{e^{-2.124}2.124^5}{5!} = 0.043 \\
P(X=6) &= \frac{e^{-2.124}2.124^6}{6!} = 0.015
\end{aligned}
$$

First thing to notice here is that the numbers do not add up to 1. This is because we are now assuming that the random variable X follows a Poisson distribution and we are using the theoretical pmf to work out the probabilities of each $x \in R_X$. Although in our experiment we found no children that could speak more than 6 languages, theoretically there is no reason why the number of languages could be greater than 6. In fact,

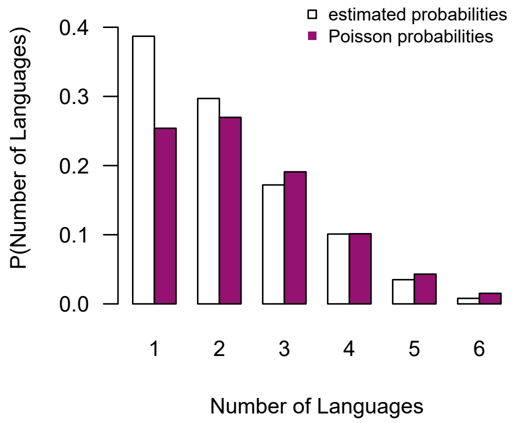
$$
\begin{aligned}
P(X=7) &= \frac{e^{-2.124}2.124^7}{7!} = 0.0046 \\
P(X=8) &= \frac{e^{-2.124}2.124^8}{8!} = 0.0012 \\
P(X=9) &= \frac{e^{-2.124}2.124^9}{9!} = 0.00029 \\
P(X=10) &= \frac{e^{-2.124}2.124^{10}}{10!} = 0.000062
\end{aligned}
$$

Although these probabilities are very small, they are still not equal to zero. As we increase the value of x then $P(X=x) \to 0$.

Secondly, our calculated probabilities are not equal to our estimated probabilities that we derived from the experiment that sampled 1000 school children.

If we repeated this same experiment again with 1000 different school children then we would have different probabilities in this table. In fact, if we did repeat this experiment there is no reason why we would have $R_X = \{1, 2, 3, 4, 5, 6\}$. We could sample a child who spoke 10 languages albeit very unlikely.

What you should be asking yourself is whether our estimated probabilities are 'close' to the calculated probabilities.



This plot shows the estimated probabilities (white) and the Poisson calculated probabilities (purple). How well do you think they match up?

Distribution notation

With some standard distributions such as the Bernoulli, binomial and Poisson distributions, we have some standard notation

1. If a discrete random variable X follows a Bernoulli distribution then we can write

$$X \sim Bern(\theta)$$

   where θ is the probability of success. This distribution is parameterised by θ.

2. If a discrete random variable X follows a binomial distribution then we can write

$$X \sim Bin(n, \theta)$$

   where θ is the probability of success and
   n the number of experiments. This distribution is parameterised by p and
   n.

3. If a discrete random variable X follows a Poisson distribution then we can write

$$X \sim Pois(\lambda)$$

   where λ is the expected value of X. This distribution is parameterised by λ.

##Learning outcomes for week 3

Now you have reached the end of week 3, you should be able to:

- Define a discrete random variable;

- Calculate the expectation and variance of a discrete random variable;

- Derive the pdf of a discrete random variable from the cdf and derive the cdf from the pdf;
- Describe the Bernoulli, binomial and Poisson distributions.

Summary of results from week 3

Random variable

A random variable is a function $X : S \to \mathbb{R}$. For each element $s \in S$, $X(s)$ is a real number (in $\mathbb{R}$).

Range space

The range space R_X of a random variable X is the set of all possible realisation, $X(s)\ \forall s \in S$, of X.

Discrete Random Variable

A random variable that has a finite or countable range space is a discrete random variable.

Probability Mass Function

The probability mass function (pmf) of a discrete random variable X is a function $f : R_x \to (0, 1]$ such that $f(x) = P(X = x) = p_X(x)\ \forall x \in R_X$ such that $f(x) > 0\ \forall x \in R_X$, $\sum_{x \in R_X} f(x) = 1$ and $P(X \in A) = \sum_{x \in A} f(x)$ for some event A.

Expected Value of a Discrete Random Variable

The expected value of a discrete random variable X is defined as $E(X) = \sum_{x \in R_X} xP(X = x) = \sum_{x \in R_X} xp_X(x)$.

Variance of a Random Variable

The variance of a random variable X is defined as $\mathrm{Var}(X) = E[(X - \mu)^2]$ where $\mu = E(X)$. Variance must be a non-negative value $\mathrm{Var}(X) \geq 0$.

Cumulative Distribution Function

Let X be a random variable with range space R_X and pmf $p_X(x) = P(X = x)$. The the cumulative distribution function (cdf) is defined as $F_X(x) = P(X \leq x)$.

The Bernoulli distribution

A Bernoulli random variable X describes an experiment that can take two values, 1 (success) and 0 (failure), with $P(1) = \theta$ and $P(0) = 1 - \theta$. $E(X) = \theta$ and $\text{Var}(X) = \theta(1-\theta)$. We use the notation $X \sim Bern(\theta)$.

The binomial distribution

A binomial random variable X describes n independent and identical experiments that each can take two values, 1 (success) and 0 (failure), with $P(1) = \theta$ and $P(0) = 1 - \theta$.

$$
\begin{aligned}
P(X = x) &= \binom{n}{x} \theta^x (1-\theta)^{n-x} \\
E(X) &= n\theta \\
\text{Var}(X) &= n\theta(1-\theta).
\end{aligned}
$$

We use the notation $X \sim Bin(n, \theta)$.

The Poisson distribution

The Poisson distribution is commonly used to describe the number of events occurring within a given interval. Its range space consist of all (non-negative) integers $\{0, 1, 2, 3, \ldots\}$. For a Poisson random variable X, its probability mass function is

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}.$$

with λ the expected number of successes or events.

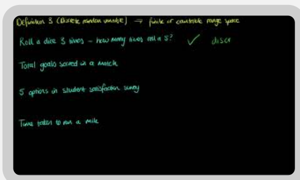
$$
\begin{aligned}
E(X) &= \lambda \\
\text{Var}(X) &= \lambda
\end{aligned}
$$

Answer 1

Rolling a dice three times, goals scored and student satisfaction are discrete.

Video

Video model answer



Duration 1:51

Answer 2

$$P(X=0) = P(E'E'E') = (5/6 \times 5/6 \times 5/6) = 0.579$$
$$P(X=1) = P(EE'E \cap E'EE' \cap E'E'E) = 3(1/6 \times 5/6 \times 5/6) = 0.347$$
$$P(X=3) = P(EEE) = (1/6 \times 1/6 \times 1/6) = 0.005$$

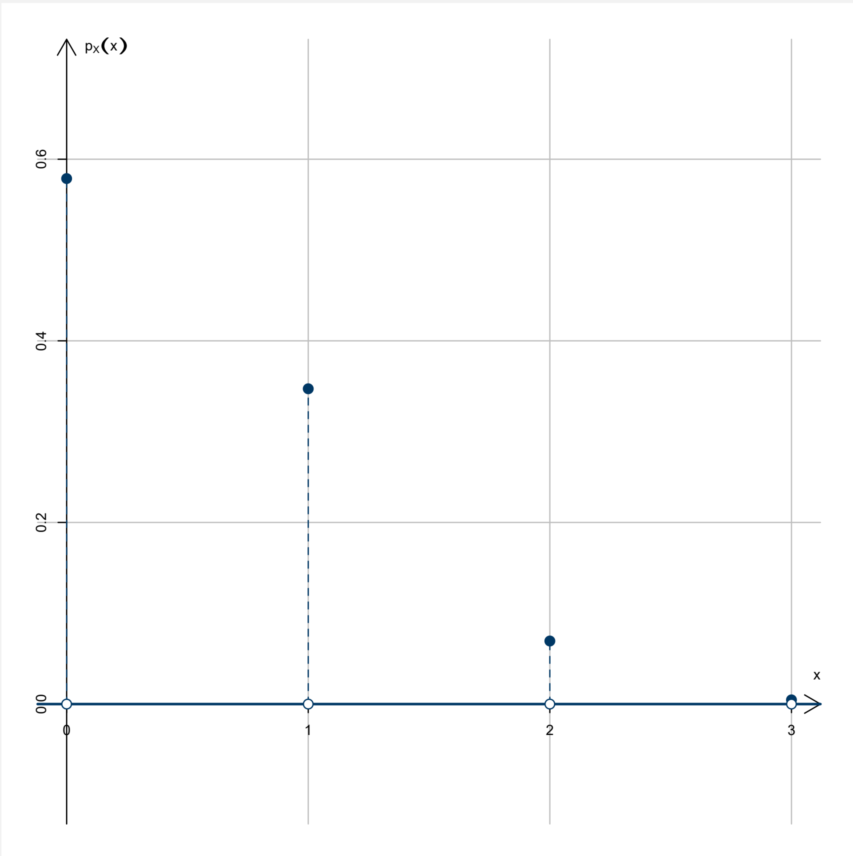
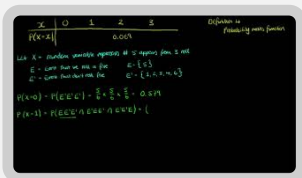


Figure 3

Video



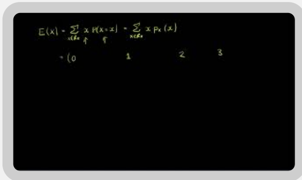
Video model answer

Duration 5:16

Answer 3

We only repeated the experiment 15 times. If you were to repeat this experiment say 100 times, you should find the estimated average value to be closer to the calculated expected value. As you increase the number of repetitions, the estimated average value converges to the calculated expected value. This is called the Law of Large Numbers and we will look at this later in more detail on in this course.

Video



Video model answer

Duration 1:47

Answer 4

In order to construct the cdf, we need to work out

$$F_X(1), F_X(2), F_X(3), F_X(4), F_X(5), F_X(6)$$

$$\begin{aligned}
F_X(1) &= P(X \leq 1) = P(X = 1) = 0.387 \\
F_X(2) &= P(X \leq 2) = P(X = 1) + P(X = 2) = 0.387 + 0.297 = 0.684 \\
F_X(3) &= P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = 0.387 + 0.297 + 0.172 = 0.856 \\
F_X(4) &= P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0.387 + 0.297 + 0.172 + 0.101 = 0.957 \\
F_X(5) &= P(X \leq 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\
&= 0.387 + 0.297 + 0.172 + 0.101 + 0.035 = 0.992 \\
F_X(6) &= P(X \leq 6) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) \\
&= 0.387 + 0.297 + 0.172 + 0.101 + 0.035 + 0.008 = 1
\end{aligned}$$

Therefore

$$F_X(x) = \begin{cases} 0 & \text{for } x < 1 \\ 0.387 & \text{for } 1 \leq x < 2 \\ 0.684 & \text{for } 2 \leq x < 3 \\ 0.856 & \text{for } 3 \leq x < 4 \\ 0.957 & \text{for } 4 \leq x < 5 \\ 0.992 & \text{for } 5 \leq x < 6 \\ 1 & \text{for } x \geq 6 \end{cases}$$

Video

Video model answer

Duration 6:10