Statistics for Data Science -1

Lecture 8.1: Random Variable: Introduction

Usha Mohan

Indian Institute of Technology Madras

1. Define what is a random variable.

- 1. Define what is a random variable.
- 2. Types of random variables: discrete and continuous.

- 1. Define what is a random variable.
- 2. Types of random variables: discrete and continuous.
- 3. Probability mass function, graph, and examples.

- 1. Define what is a random variable.
- 2. Types of random variables: discrete and continuous.
- 3. Probability mass function, graph, and examples.
- 4. Cumulative distribution function, graphs, and examples.

- 1. Define what is a random variable.
- 2. Types of random variables: discrete and continuous.
- 3. Probability mass function, graph, and examples.
- 4. Cumulative distribution function, graphs, and examples.
- 5. Expectation and variance of a random variable.

Example: Rolling a dice twice

Example: Tossing a coin three times

When a probability experiment is performed, often we are not interested in all the details of the experimental result, but rather are interested in the value of some numerical quantity determined by the result.

- When a probability experiment is performed, often we are not interested in all the details of the experimental result, but rather are interested in the value of some numerical quantity determined by the result.
- For example, in rolling a dice twice, often we care about only their sum of outcomes and are not concerned about the values on the individual dice.

- When a probability experiment is performed, often we are not interested in all the details of the experimental result, but rather are interested in the value of some numerical quantity determined by the result.
- For example, in rolling a dice twice, often we care about only their sum of outcomes and are not concerned about the values on the individual dice.
 - That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was (1,6),(2,5),(3,4),(4,3),(5,2), or (6,1).

- When a probability experiment is performed, often we are not interested in all the details of the experimental result, but rather are interested in the value of some numerical quantity determined by the result.
- For example, in rolling a dice twice, often we care about only their sum of outcomes and are not concerned about the values on the individual dice.
 - That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was (1,6),(2,5),(3,4),(4,3),(5,2), or (6,1).
- ► These quantities of interest, or, more formally, these real-valued functions defined on the sample space, are known as random variables.

- When a probability experiment is performed, often we are not interested in all the details of the experimental result, but rather are interested in the value of some numerical quantity determined by the result.
- For example, in rolling a dice twice, often we care about only their sum of outcomes and are not concerned about the values. on the individual dice.
 - That is, we may be interested in knowing that the sum is 7 and may not be concerned over whether the actual outcome was (1,6), (2,5), (3,4), (4,3), (5,2), or (6,1).
- These quantities of interest, or, more formally, these real-valued functions defined on the sample space, are known as random variables.
- Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable., (3) (2) (2) (2) (2) (4/15)

► Experiment: Roll a dice twice

- Experiment: Roll a dice twice
- ► The sample space for this experiment is

$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), \end{array} \right\}$$

- Experiment: Roll a dice twice
- The sample space for this experiment is

The sample space for this experiment is
$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), \end{array} \right\}$$

Consider the probabilities associated with the two questions

- Experiment: Roll a dice twice
- ► The sample space for this experiment is

$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), \end{array} \right\}$$

- ▶ Consider the probabilities associated with the two questions
 - 1. Of the outcomes, how many outcomes will result in a sum of outcomes as 7?

- Experiment: Roll a dice twice
- ► The sample space for this experiment is

$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), \end{array} \right\}$$

- Consider the probabilities associated with the two questions
 - Of the outcomes, how many outcomes will result in a sum of outcomes as 7?
 - 2. Of the outcomes, how many outcomes will have the smaller of the outcomes as 3?

- Experiment: Roll a dice twice
- ► The sample space for this experiment is

$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), \end{array} \right\}$$

- Consider the probabilities associated with the two questions
 - 1. Of the outcomes, how many outcomes will result in a sum of outcomes as 7?
 - 2. Of the outcomes, how many outcomes will have the smaller of the outcomes as 3?
- Notice, the experiment and sample space used to answer both the questions are the same.

Statistics for Data Science -1
Random variable
Example: Rolling a dice twice

Example: Rolling a dice twice

Let *X* denote the sum of outcomes of the two rolls.

- Let X denote the sum of outcomes of the two rolls.
- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the outcome is taken as value of *Y*.

- Let X denote the sum of outcomes of the two rolls.
- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the outcome is taken as value of Y.

Outcome	Χ	Y	Outcome	X	Y	Outcome	X	Y
(1,1)	2	1	(3,1)	4	1	(5,1)	6	1
(1,2)	3	1	(3,2)	5	2	(5,2)	7	2
(1,3)	4	1	(3,3)	6	3	(5,3)	8	3
(1,4)	5	1	(3,4)	7	3	(5,4)	9	4
(1,5)	6	1	(3,5)	8	3	(5,5)	10	5
(1,6)	7	1	(3,6)	9	3	(5,6)	11	5
(2,1)	3	1	(4,1)	5	1	(6,1)	7	1
(2,2)	4	2	(4,2)	6	2	(6,2)	8	2
(2,3)	5	2	(4,3)	7	3	(6,3)	9	3
(2,4)	6	2	(4,4)	8	4	(6,4)	10	4
(2,5)	7	2	(4,5)	9	4	(6,5)	11	5
(2,6)	8	2	(4,6)	10	4	(6,6)	12	6

Let X denote the sum of outcomes of the two rolls.

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

- Let X denote the sum of outcomes of the two rolls.
- X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

- Let X denote the sum of outcomes of the two rolls.
- X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

 Value of $X \mid \text{Relevant event}$

- Let X denote the sum of outcomes of the two rolls.
- ► X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

7 takes the values 2,5, 1,5,6,1,6,5,16,11		
Value of X	Relevant event	
2		

- Let X denote the sum of outcomes of the two rolls.
- ► X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

t tunes the v	aracs 2,5,1,5,6,1,6,5,16,11,	unc
Value of X	Relevant event	
2	$\{(1,1)\}$	

Example: Rolling a dice twice

- Let X denote the sum of outcomes of the two rolls.
- ► X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

1 taite tii taiae				
Value of X	Relevant event			
2	$\{(1,1)\}$			
3				

- Let X denote the sum of outcomes of the two rolls.
- ► X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

71 tantes tile talaes 2,0,1,0,0,1,0,5,10,11, all			
Value of X	Relevant event		
2	$\{(1,1)\}$		
3	$\{(1,2),(2,1)\}$		

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

	Value of X	Relevant event	
	2	$\{(1,1)\}$	
	3	$\{(1,2),(2,1)\}$	
	4		

- Let X denote the sum of outcomes of the two rolls.
- ➤ X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

Value of X	Relevant event			
2	$\{(1,1)\}$			
3	$\{(1,2),(2,1)\}$			
4	$\{(1,3),(2,2),(3,1)\}$			

- Let X denote the sum of outcomes of the two rolls.
- ➤ X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

	- - - - - - -
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

/ takes the v	alucs 2,5,1,5,6,1,6,5,16,11, all
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
:	:

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

/ takes the v	araes 2,0,1,0,0,1,0,5,10,11, arr
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
:	:
9	

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

t takes the v	alucs 2,5,1,5,0,1,0,5,10,11, all
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
:	:
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
-	

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

/ takes the v	alucs 2,5,1,5,0,1,0,5,10,11, all
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
:	:
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

/ takes the v	alucs 2,5,7,5,0,1,0,5,10,11, all
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
•	:
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	$\{(4,6),(5,5),(6,4)\}$

- Let X denote the sum of outcomes of the two rolls.
- ➤ X takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

, t cances time t	a.a.c. 2,0, 1,0,0,1,0,5,10,11, a.i.
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
÷	:
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	$\{(4,6),(5,5),(6,4)\}$
11	

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

/ takes the v	alucs 2,5,1,5,0,1,0,5,10,11, all
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
÷	:
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	$\{(4,6),(5,5),(6,4)\}$
11	{(5,6),(6,5)}

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

/ takes the v	alucs 2,5,1,5,6,1,6,5,16,11, all
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
:	<u>:</u>
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	$\{(4,6),(5,5),(6,4)\}$
11	{(5,6),(6,5)}
12	

- Let X denote the sum of outcomes of the two rolls.
- ► *X* takes the values 2,3,4,5,6,7,8,9,10,11, and 12.

71 tancs the v	<u>ulucs 2,5, 1,5,6,1,6,5,16,11, ull</u>
Value of X	Relevant event
2	$\{(1,1)\}$
3	$\{(1,2),(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
:	:
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	$\{(4,6),(5,5),(6,4)\}$
11	$\{(5,6),(6,5)\}$
12	$\{(6,6)\}$

Statistics for Data Science -1
Random variable
Example: Rolling a dice twice

Example: Rolling a dice twice

Example: Rolling a dice twice

We say X is a random variable taking on one of the values 2,3,4,5,6,7,8,9,10,11, and 12 with respective probabilities

Probability of X Probability of relevant event Probability

_,0,.,0,0,.,0,0,=0,=	, and min respective product	
Probability of X	Probability of relevant event	Probability
$P\{X=2\}$	$P(\{(1,1)\})$	$\frac{1}{36}$

D(V, O) = D(((1, 1)))	
$P\{X=2\} P(\{(1,1)\})$	$\frac{1}{36}$
$P{X = 3}$ $P({(1,2),(2,1)})$	$\frac{2}{36}$

$P\{X = 2\} \qquad P(\{(1,1)\})$ $P\{X = 3\} \qquad P(\{(1,2),(2,1)\})$	1
$P\{X-3\}$ $P(\{(1,2),(2,1)\})$	36
' (\(\frac{1}{2}\), \(\frac{1}{2}\), \(\frac{1}\), \(\frac{1}\), \(\frac{1}{2}\), \(\frac{1}{2}\), \(<u>2</u> 36
$P{X = 4}$ $P({(1,3),(2,2),(3,1)})$	3 36

Probability of X	Probability of relevant event	Probability
$P\{X=2\}$	$P(\{(1,1)\})$	$\frac{1}{36}$
$P{X = 3}$	$P(\{(1,2),(2,1)\})$	$\frac{2}{36}$
$P\{X=4\}$	$P(\{(1,3),(2,2),(3,1)\})$	3 36
$P\{X=5\}$	$P(\{(1,4),(2,3),(3,2),(4,1)\})$	$\frac{4}{36}$
:	÷	:

Probability of X	Probability of relevant event	Probability
$P\{X=2\}$	$P(\{(1,1)\})$	$\frac{1}{36}$
$P\{X=3\}$	$P(\{(1,2),(2,1)\})$	2 36
$P{X = 4}$	$P(\{(1,3),(2,2),(3,1)\})$	3 36
$P\{X=5\}$	$P(\{(1,4),(2,3),(3,2),(4,1)\})$	4 36
i i		:
$P\{X=9\}$	$P(\{(3,6),(4,5),(5,4),(6,3)\})$	<u>4</u> 36

Probability of X	Probability of relevant event	Probability
$P\{X=2\}$	$P(\{(1,1)\})$	$\frac{1}{36}$
$P{X = 3}$	$P(\{(1,2),(2,1)\})$	2 36
$P{X = 4}$	$P(\{(1,3),(2,2),(3,1)\})$	$\frac{3}{36}$
$P\{X=5\}$	$P(\{(1,4),(2,3),(3,2),(4,1)\})$	$\frac{4}{36}$
i:	:	:
$P\{X=9\}$	$P(\{(3,6),(4,5),(5,4),(6,3)\})$	<u>4</u> 36
$P\{X=10\}$	$P(\{(4,6),(5,5),(6,4)\})$	$\frac{3}{36}$

Probability of X	Probability of relevant event	Probability
$P\{X=2\}$	$P(\{(1,1)\})$	$\frac{1}{36}$
$P{X = 3}$	$P(\{(1,2),(2,1)\})$	2 36
$P{X = 4}$	$P(\{(1,3),(2,2),(3,1)\})$	3 36 4
$P\{X=5\}$	$P(\{(1,4),(2,3),(3,2),(4,1)\})$	$\frac{4}{36}$
i:	:	:
$P\{X=9\}$	$P(\{(3,6),(4,5),(5,4),(6,3)\})$	4 36 3
$P\{X = 10\}$	$P(\{(4,6),(5,5),(6,4)\})$	3 36
$P\{X=11\}$	$P(\{(5,6),(6,5)\})$	$\frac{2}{36}$

Probability of X	Probability of relevant event	Probability
$P\{X=2\}$	$P(\{(1,1)\})$	$\frac{1}{36}$
$P{X = 3}$	$P(\{(1,2),(2,1)\})$	2 36
$P{X = 4}$	$P(\{(1,3),(2,2),(3,1)\})$	3 36
$P{X = 5}$	$P(\{(1,4),(2,3),(3,2),(4,1)\})$	$\frac{4}{36}$
i i	:	:
$P\{X=9\}$	$P(\{(3,6),(4,5),(5,4),(6,3)\})$	<u>4</u> 36
$P\{X=10\}$	$P(\{(4,6),(5,5),(6,4)\})$	3 36
$P\{X=11\}$	$P(\{(5,6),(6,5)\})$	$\frac{2}{36}$
$P\{X = 12\}$	$P(\{(6,6)\})$	$\frac{1}{36}$

▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.

- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.
- Y takes the values 1,2,3,4,5, and 6.

- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.
- Y takes the values 1,2,3,4,5, and 6.

Y

- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.
- Y takes the values 1,2,3,4,5, and 6.

```
Y Relevant event
```

- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event	
1		

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

•		,=,0, .,0,
7	Y	Relevant event
	1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$
	2	

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

	21.00 01.0 12.200 =,=,0, 1,0, 21.2 01
Y	Relevant event
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$

- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.
- Y takes the values 1,2,3,4,5, and 6.

,	7 takes the values 1,2,0, 1,0, and 0.		
Y	Relevant event		
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$		
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$		
3			

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$
	$\{(3,3),(3,1),(3,3),(3,3),(3,3),(3,3)\}$

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event	
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$	
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$	
4		

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event	
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$	
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$	
4	$\{(4,4),(4,5),(4,6),(5,4),(6,4)\}$	

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

7 takes the values 1,2,0, 1,0, and 0.		
Y	Relevant event	
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$	
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$	
4	$\{(4,4),(4,5),(4,6),(5,4),(6,4)\}$	
5		

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$
4	$\{(4,4),(4,5),(4,6),(5,4),(6,4)\}$
5	$\{(5,5),(5,6),(6,5)\}$

Lesser of the two values

- ▶ Let Y denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of Y.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event		
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$		
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$		
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$		
4	$\{(4,4),(4,5),(4,6),(5,4),(6,4)\}$		
5	$\{(5,5),(5,6),(6,5)\}$		
6			

Lesser of the two values

- ▶ Let *Y* denote the lesser of the two outcomes. If the outcomes are the same, the value of the out come is taken as value of *Y*.
- Y takes the values 1,2,3,4,5, and 6.

Y	Relevant event		
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$		
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$		
3	$\{(3,3),(3,4),(3,5),(3,6),(4,3),(5,3),(6,3)\}$		
4	$\{(4,4),(4,5),(4,6),(5,4),(6,4)\}$		
5	$\{(5,5),(5,6),(6,5)\}$		
6	{(6,6)}		

Statistics for Data Science -1
Random variable
Example: Rolling a dice twice

Example: Rolling a dice twice

Example: Rolling a dice twice

► We say *Y* is a random variable taking on one of the values 1,2,3,4,5, and 6 with respective probabilities

Y Relevant event Probability

-,			
7	Y	Relevant event	Probability
	1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	11 36

2,2,0, 1,0, and 0 min respective production		
Y	Relevant event	Probability
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	11 36
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$	9 36

-,=,o, .,o, aa ooop ooto p. oodas		
Y	Relevant event	Probability
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	11 36
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$	9 36
3	{(3,3), (3,4), (3,5), (3,6), (4,3), (5,3), (6,3)}	7 36

Y	Relevant event	Probability
1	$\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1),(5,1),(6,1)\}$	11 36
2	$\{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)\}$	9 36
3	{(3,3), (3,4), (3,5), (3,6), (4,3), (5,3), (6,3)}	7/36
4	{(4,4),(4,5),(4,6),(5,4),(6,4)}	<u>5</u> 36

Y	Relevant event	Probability
1	$ \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1), (5,1),(6,1)\} $	11 36
2	{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)}	9/36
3	{(3,3), (3,4), (3,5), (3,6), (4,3), (5,3), (6,3)}	7 36
4	{(4,4),(4,5),(4,6),(5,4),(6,4)}	<u>5</u> 36
5	{(5,5), (5,6), (6,5)}	<u>3</u> <u>36</u>

Y	Relevant event	Probability
1	$ \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(3,1),(4,1), (5,1),(6,1)\} $	11 36
2	{(2,2),(2,3),(2,4),(2,5),(2,6),(3,2),(4,2),(5,2),(6,2)}	9 36
3	{(3,3), (3,4), (3,5), (3,6), (4,3), (5,3), (6,3)}	7 36
4	{(4,4),(4,5),(4,6),(5,4),(6,4)}	<u>5</u> 36
5	{(5,5), (5,6), (6,5)}	3/36
6	{(6,6)}	1/36

Example: Tossing a coin three times

Experiment: Toss a coin three times.

- Experiment: Toss a coin three times.
- ▶ The sample space for this experiment is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- Experiment: Toss a coin three times.
- ► The sample space for this experiment is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Consider the probabilities associated with the two questions:

- Experiment: Toss a coin three times.
- ► The sample space for this experiment is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- ► Consider the probabilities associated with the two questions:
 - 1. Of the three tosses, how many tosses will be heads?

- Experiment: Toss a coin three times.
- ▶ The sample space for this experiment is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- Consider the probabilities associated with the two questions:
 - 1. Of the three tosses, how many tosses will be heads?
 - 2. Of the three tosses, which toss results in a heads first?, i.e first, second or third toss is a head?

- Experiment: Toss a coin three times.
- ► The sample space for this experiment is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- Consider the probabilities associated with the two questions:
 - 1. Of the three tosses, how many tosses will be heads?
 - 2. Of the three tosses, which toss results in a heads first?, i.e first, second or third toss is a head?
- Notice the experiment and sample space used to answer both the questions are the same.

Statistics for Data Science -1
Random variable
Example: Tossing a coin three times

Example: Tossing a coin three times

▶ Let *X* denote the number of heads that appear. Let *Y* denote the toss in which a head appears first.

▶ Let *X* denote the number of heads that appear. Let *Y* denote the toss in which a head appears first.

Outcome	X	Y
HHH	3	1
HHT	2	1
HTH	2	1
HTT	1	1
THH	2	2
THT	1	2
TTH	1	3
TTT	0	NIL

Example: Tossing a coin three times

▶ Let *X* denote the number of heads that appear.

- Let *X* denote the number of heads that appear.
- X takes the values 0,1,2,3

- Let X denote the number of heads that appear.
- \triangleright X takes the values 0,1,2,3

```
Value of X Relevant event
```

- Let *X* denote the number of heads that appear.
- \triangleright X takes the values 0,1,2,3

Value of X	Relevant event
0	

- Let *X* denote the number of heads that appear.
- \triangleright X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}

Example: Tossing a coin three times

- Let *X* denote the number of heads that appear.
- ► X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	

- Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$

- Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	

- Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	$\{(HHT), (HTH), (THH)\}$

- Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	$\{(HHT), (HTH), (THH)\}$
3	

- Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	$\{(HHT), (HTH), (THH)\}$
3	{(HHH)}

- Let X denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	{(HHT), (HTH), (THH)}
3	{(<i>HHH</i>)}

- ▶ Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	$\{(HHT), (HTH), (THH)\}$
3	{(<i>HHH</i>)}

- We say X is a random variable taking on one of the values 0, 1, 2, and 3 with respective probabilities
 - $P\{X=0\}=P\{(TTT)\}=\frac{1}{8}$

Number of tosses that will be heads

- Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	{(HHT), (HTH), (THH)}
3	{(HHH)}

- We say X is a random variable taking on one of the values 0, 1, 2, and 3 with respective probabilities
 - $P\{X=0\}=P\{(TTT)\}=\frac{1}{8}$
 - $P\{X = 1\} = P\{(HTT), (THT), (TTH)\} = \frac{3}{8}$

Number of tosses that will be heads

- Let X denote the number of heads that appear.
- X takes the values 0.1.2.3

Value of X	Relevant event
0	{(<i>TTT</i>)}
1	$\{(HTT), (THT), (TTH)\}$
2	{(HHT), (HTH), (THH)}
3	{(<i>HHH</i>)}

- ▶ We say X is a random variable taking on one of the values 0, 1, 2, and 3 with respective probabilities
 - $P\{X=0\}=P\{(TTT)\}=\frac{1}{9}$

 - ► $P{X = 1} = P{(HTT), (THT), (TTH)} = \frac{3}{8}$ ► $P{X = 2} = P{(HHT), (HTH), (THH)} = \frac{3}{8}$

Number of tosses that will be heads

- ▶ Let *X* denote the number of heads that appear.
- ➤ X takes the values 0,1,2,3

Value of X	Relevant event
0	$\{(TTT)\}$
1	$\{(HTT), (THT), (TTH)\}$
2	{(HHT), (HTH), (THH)}
3	{(HHH)}

- We say X is a random variable taking on one of the values 0,
 - 1, 2, and 3 with respective probabilities
 - $P\{X=0\}=P\{(TTT)\}=\frac{1}{8}$
 - $P\{X=1\} = P\{(HTT), (THT), (TTH)\} = \frac{3}{8}$
 - $P\{X=2\} = P\{(HHT), (HTH), (THH)\} = \frac{3}{8}$
 - $P\{X=3\} = P\{(HHH)\} = \frac{1}{8}$

▶ Let *Y* denote the toss in which a head appears first.

- Let Y denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

- Let Y denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

- Let Y denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

```
Value of Y \mid Relevant event
```

- Let *Y* denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

	Relevant event
1	

- Let Y denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$

- Let Y denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	

- ▶ Let *Y* denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

	· · · ·
Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH), (THT)}

- ▶ Let *Y* denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH),(THT)}
3	

- Let Y denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH),(THT)}
3	{(TTH)}

- Let Y denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH),(THT)}
3	{(<i>TTH</i>)}
NIL	

- Let Y denote the toss in which a head appears first.
- ▶ Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH),(THT)}
3	{(<i>TTH</i>)}
NIL	$\{(TTT)\}$

- Let *Y* denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH),(THT)}
3	{(<i>TTH</i>)}
NIL	{(<i>TTT</i>)}

We say Y is a random variable taking on one of the values 1,
2, 3, and NIL with respective probabilities

- ▶ Let *Y* denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH), (THT)}
3	{(TTH)}
NIL	{(<i>TTT</i>)}

- We say Y is a random variable taking on one of the values 1,
 2, 3, and NIL with respective probabilities
 - ► $P{Y = 1} = P{(HHH), (HHT), (HTH), (HTT)} = \frac{4}{8}$

- Let Y denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	$\{(THH),(THT)\}$
3	{(<i>TTH</i>)}
NIL	{(<i>TTT</i>)}

- We say Y is a random variable taking on one of the values 1, 2, 3, and NIL with respective probabilities
 - $P\{Y=1\}=P\{(HHH),(HHT),(HTH),(HTT)\}=\frac{4}{8}$
 - $P\{Y=2\} = P\{(THH), (THT)\} = \frac{2}{8}$

- Let Y denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	{(THH),(THT)}
3	{(<i>TTH</i>)}
NIL	{(<i>TTT</i>)}

- We say Y is a random variable taking on one of the values 1, 2, 3, and NIL with respective probabilities
 - $P\{Y=1\}=P\{(HHH),(HHT),(HTH),(HTT)\}=\frac{4}{8}$
 - $P\{Y=2\} = P\{(THH), (THT)\} = \frac{2}{8}$
 - $P\{Y=3\} = P\{(TTH)\} = \frac{1}{8}$

- ▶ Let *Y* denote the toss in which a head appears first.
- Y takes the values 1,2,3, and NIL

Value of Y	Relevant event
1	$\{(HHH), (HHT), (HTH), (HTT)\}$
2	$\{(THH),(THT)\}$
3	{(<i>TTH</i>)}
NIL	{(<i>TTT</i>)}

- We say Y is a random variable taking on one of the values 1,
 2, 3, and NIL with respective probabilities
 - $P\{Y=1\} = P\{(HHH), (HHT), (HTH), (HTT)\} = \frac{4}{8}$
 - $P\{Y=2\} = P\{(THH), (THT)\} = \frac{2}{8}$
 - $P{Y = 3} = P{(TTH)} = \frac{1}{8}$
 - ► $P{Y = NIL} = P{(TTT)} = \frac{1}{8}$

Section summary

1. Definition of a random variable and examples