Data Science 1
Probability
Probability Distributions

Edward L. Boone

Introduction

The last video introduced the idea of a *Random Variable*.

- A random variable turns and outcome into a number.
- Formally it is a function X that goes from $\mathcal{S}$ to $\mathbb{R}$.
- Easy to think of it as "How do we measure our outcome?"
- Not unique.
- Random since we don't know the measurement before the experiment.

Combinations

A probability density(mass) function, f is a function that goes from $\mathbb{R}$ to $[0, 1]$ such that the probability of an event $E \in \mathcal{S}$ is the same as the $f(X(E)) = P(X = x) = P(X(E) = x) = P(E)$.

- Basically $f(x)$ needs to match the probability of the event that was mapped to x.
- The idea seems a bit difficult, but it doesn't need to be.
- Looking a few examples should help you understand the concept.

Single Die Example

If we are looking at a simple *fair* six sided die, then the sample space is:

$$\mathcal{S} = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\}$$

Let X be the random variable that corresponds to the number of dots on the up facing side.

$$X\left(\epsdice{1}\right) = 1, X\left(\epsdice{2}\right) = 2, X\left(\epsdice{3}\right) = 3,$$

$$X\left(\epsdice{4}\right) = 4, X\left(\epsdice{5}\right) = 5, X\left(\epsdice{6}\right) = 6$$

Single Die Example

If we are looking at a simple *fair* six sided die, then the sample space is:

$$\mathcal{S} = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\}$$

Let X be the random variable that corresponds to the number of dots on the up facing side.

$$X\left(\text{⚀}\right) = 1, X\left(\text{⚁}\right) = 2, X\left(\text{⚂}\right) = 3,$$

$$X\left(\text{⚃}\right) = 4, X\left(\text{⚄}\right) = 5, X\left(\text{⚅}\right) = 6$$

Hence the probability mass function is:

$$f(1) = P(X = 1) = P\left(\text{⚀}\right) = 1/6$$

Single Die Example

If we are looking at a simple *fair* six sided die, then the sample space is:

$$\mathcal{S} = \{\text{⚀}, \text{⚁}, \text{⚂}, \text{⚃}, \text{⚄}, \text{⚅}\}$$

Let X be the random variable that corresponds to the number of dots on the up facing side.

$$X\left(\text{⚀}\right) = 1, X\left(\text{⚁}\right) = 2, X\left(\text{⚂}\right) = 3,$$

$$X\left(\text{⚃}\right) = 4, X\left(\text{⚄}\right) = 5, X\left(\text{⚅}\right) = 6$$

Hence the probability mass function is:

$$f(1) = P(X = 1) = P\left(\text{⚀}\right) = 1/6$$

$$f(2) = P(X = 2) = P\left(\text{⚁}\right) = 1/6$$

Single Die Example

If we are looking at a simple *fair* six sided die, then the sample space is:

$$\mathcal{S} = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\}$$

Hence the probability mass function is:

$$f(1) = P(X = 1) = P\left(\epsdice{1}\right) = 1/6$$

$$f(2) = P(X = 2) = P\left(\epsdice{2}\right) = 1/6$$

$$f(3) = P(X = 3) = P\left(\epsdice{3}\right) = 1/6$$

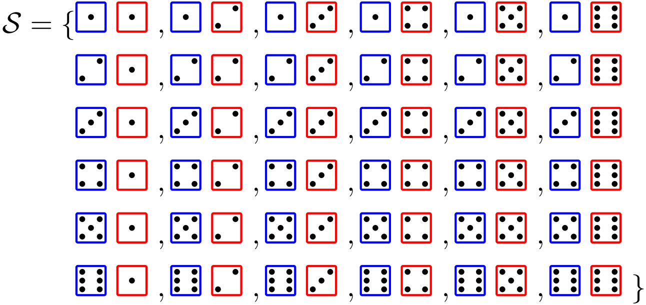
$$f(4) = P(X = 4) = P\left(\epsdice{4}\right) = 1/6$$

$$f(5) = P(X = 5) = P\left(\epsdice{5}\right) = 1/6$$

$$f(6) = P(X = 6) = P\left(\epsdice{6}\right) = 1/6$$

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red. Then the Sample Space is:



There are $|\mathcal{S}| = 36$ possible outcomes! But the random variable, X, that is the sum of all the up facing dots only has 11 outcomes!

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red.

$$f(1) = 0$$

Since there are no outcomes that have 1 as the total of the up facing dots.

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red.

$$f(1) = 0$$

Since there are no outcomes that have 1 as the total of the up facing dots.

$$f(2) = P(X = 2) = P\left(\boxdot \;\; \boxdot\right) = 1/36$$

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red.

$$f(2) = P(X = 2) = P\left(\color{blue}{⚀}\ \color{red}{⚀}\right) = 1/36$$

$$f(3) = P(X = 3) = P\left(\color{blue}{⚀}\ \color{red}{⚁}, \color{blue}{⚁}\ \color{red}{⚀}\right) = 2/36$$

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red.

$$f(2) = P(X = 2) = P\left({\color{blue}\text{⚀}}\ {\color{red}\text{⚀}} \right) = 1/36$$

$$f(3) = P(X = 3) = P\left({\color{blue}\text{⚀}}\ {\color{red}\text{⚁}}, {\color{blue}\text{⚁}}\ {\color{red}\text{⚀}} \right) = 2/36$$

$$f(4) = P(X = 4) = P\left({\color{blue}\text{⚀}}\ {\color{red}\text{⚂}}, {\color{blue}\text{⚂}}\ {\color{red}\text{⚀}}, {\color{blue}\text{⚁}}\ {\color{red}\text{⚁}} \right) = 3/36$$

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red.

$$f(2) = P(X = 2) = P\left(\text{⚀ ⚀}\right) = 1/36$$

$$f(3) = P(X = 3) = P\left(\text{⚀ ⚁}, \text{⚁ ⚀}\right) = 2/36$$

$$f(4) = P(X = 4) = P\left(\text{⚀ ⚂}, \text{⚂ ⚀}, \text{⚁ ⚁}\right) = 3/36$$

$$f(5) = P(X = 5) = P\left(\text{⚀ ⚃}, \text{⚃ ⚀}, \text{⚁ ⚂}, \text{⚂ ⚁}\right) = 4/36$$

Two Dice

Things get considerably more complicated when playing a game with two dice. Since at each roll two faces are facing up we need to record two faces at a time. Suppose one die is blue and the other is red.

If you keep the process up you end up with the following probability mass function:

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Properties

For discrete (countable) outcome experiments f has the following properties:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$.
- $\sum_{x \in \mathbb{R}} f(x) = 1$.

For continuous (uncountable) outcome experiments f has the following properties:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$.
- $\int_{x \in \mathbb{R}} f(x)dx = 1$.