Probability Spaces

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Summary

Probability spaces encode common sense

Key idea

Model uncertain phenomena as experiments that can be repeated over and over

Sample space

Set of all possible outcomes of the experiment

Usually denoted by Ω

Rolling a 6-sided die

What is the sample space?

$\Omega := \{1, 2, 3, 4, 5, 6\}$

Rolling a die until it lands on 6

What is the sample space?

It depends on how we model!

1. $\Omega := \{1, 2, 3, \ldots\}$ (number of rolls)
2. $\Omega := \{6, 1 \to 6, \ldots, 5 \to 6, 1 \to 1 \to 6, \ldots\}$

Weather tomorrow in NYC

What is the sample space?

Depends *a lot* on modeling choices

Just interested in temperature at a given time: $\Omega := \mathbb{R}$

Overview of probability space

1. Model phenomenon of interest as experiment with a sample space of mutually exclusive outcomes

2. Group outcomes in sets called events

3. Assign a probability to each event

Rolling a six-sided die

Examples of events:

$A := \{1, 3, 5\}$

$B := \{4\}$

$C := \{1, 2, 3, 4, 5, 6\}$

If we roll a 4, which of these events have occurred?

Rolling a die until it lands on 6

How many outcomes does the event *Rolling twice* contain?

It depends!

1. $\Omega_1 := \{1, 2, 3, \ldots\}$ (number of rolls)
2. $\Omega_2 := \{6, 1 \to 6, \ldots, 5 \to 6, 1 \to 1 \to 6, \ldots\}$

Weather in NYC tomorrow $\Omega = \mathbb{R}$

Examples of events:

$A := [30, \infty)$

$B := \{35\}$

$C := \mathbb{R}$

If temperature is 40 degrees, which events have occurred?

Probability measure

The probability of an event quantifies how likely it is

(A function of a set is called a measure)

Intuitive definition: If we repeat the experiment many times

$$\mathrm{P}(\text{event}) = \frac{\text{number of times event occurs}}{\text{total repetitions}}$$

Events and probability measures

Key question:

What events should we assign probabilities to?

Sample space

This is the event that any outcome occurs

From our intuitive definition:

$$
\begin{aligned}
P(\Omega) &= \frac{\text{outcomes in } \Omega}{\text{total}} \\
&= \frac{\text{total}}{\text{total}} \\
&= 1
\end{aligned}
$$

Union and intersection of events

If we assign probabilities to A and B, we should assign a probability to

- $A \cup B$ (A or B happen)
- $A \cap B$ (A and B happen)

Union of disjoint events

Disjoint events don't share any common outcomes

From our intuitive definition:

$$
\begin{aligned}
P\left(D_1 \cup D_2\right) &= \frac{\text{outcomes in } D_1 \text{ or } D_2}{\text{total}} \\
&= \frac{\text{outcomes in } D_1 + \text{outcomes in } D_2}{\text{total}} \\
&= \frac{\text{outcomes in } D_1}{\text{total}} + \frac{\text{outcomes in } D_2}{\text{total}} \\
&= \mathrm{P}\left(D_1\right) + \mathrm{P}\left(D_2\right)
\end{aligned}
$$

Union of non-disjoint events

Non-disjoint events share some common outcomes

From our intuitive definition:

$$
\begin{aligned}
P(E_1 \cup E_2) &= \frac{\text{outcomes in } E_1 \text{ or } E_2}{\text{total}} \\
&= \frac{\text{outcomes in } E_1 + \text{outcomes in } E_2 - \text{outcomes in } E_1 \text{ and } E_2}{\text{total}} \\
&= \frac{\text{outcomes in } E_1}{\text{total}} + \frac{\text{outcomes in } E_2}{\text{total}} - \frac{\text{outcomes in } E_1 \text{ and } E_2}{\text{total}} \\
&= \mathrm{P}(E_1) + \mathrm{P}(E_2) - \mathrm{P}(E_1 \cap E_2)
\end{aligned}
$$

$$\mathrm{P}(E_1 \cap E_2) = \mathrm{P}(E_1) + \mathrm{P}(E_2) - P(E_1 \cup E_2)$$

Complement of an event

If we assign a probability to A, we should also assign a probability to its complement A^c

This is the probability that A does not happen

What should $\mathrm{P}(A^c)$ equal as a function of $\mathrm{P}(A)$?

(*Hint:* $A \cup A^c = \Omega$)

$1 = \mathrm{P}(\Omega) = \mathrm{P}\left(A \cup A^c\right) = \mathrm{P}(A) + \mathrm{P}(A^c)$

so $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$

This is enough!

We are ready to define probability spaces mathematically

Collection of events

Given a sample space Ω, we assign probabilities to a collection $\mathcal{C}$ of events that satisfies:

1. $\Omega \in \mathcal{C}$
2. If an event $A \in \mathcal{C}$ then $A^c \in \mathcal{C}$
3. If the events $A, B \in \mathcal{C}$, then $A \cup B \in \mathcal{C}$

   If a countably infinite sequence $A_1, A_2, A_3, \ldots \in \mathcal{C}$
   then $\cup_{i=1}^{\infty} A_i \in \mathcal{C}$

Collections with these properties are called σ-algebras

Collection of events

If A and B are in the collection, what about $A \cap B$?

De Morgan's law: $A \cap B = (A^c \cup B^c)^c$

Rolling a six-sided die

$\Omega := \{1, 2, 3, 4, 5, 6\}$

Collection of events if we want to include $\{1\}$, $\{2\}$, ..., $\{6\}$?

All possible subsets of Ω ($2^6 = 64$ events)

Smaller collections can also be valid

Smallest collection that contains $\{2\}$?

$\{\Omega, \emptyset, \{2\}, \{1, 3, 4, 5, 6\}\}$

Probability measure

Function mapping events in the collection to probabilities

1. $\mathrm{P}(A) \geq 0$ for any event $A \in \mathcal{C}$
2. $\mathrm{P}(\Omega) = 1$
3. If $A, B \in \mathcal{C}$ are disjoint then

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$$

   For countably infinite sequences of disjoint sets: $A_1, A_2, A_3, \ldots \in \mathcal{C}$

$$\mathrm{P}\left(\lim_{n\to\infty} \cup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{P}(A_i)$$

Consequences of definition

$$\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$$

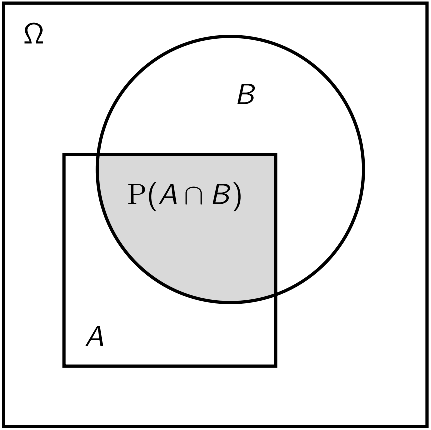
$$\mathrm{P}\left(E_1 \cup E_2\right) = \mathrm{P}\left(E_1\right) + \mathrm{P}\left(E_2\right) - \mathrm{P}\left(E_1 \cap E_2\right)$$

Analogy with other measures

Mass, length, area or volume satisfy similar properties

We can use Venn diagrams to represent events

Venn diagram



Rolling a six-sided die

$\Omega := \{1, 2, 3, 4, 5, 6\}$

Collection: All possible subsets of Ω

Probability measure?

We need to assign consistent probabilities to all events...

Idea: Divide Ω into smallest possible *components* and assign probabilities to them

Partition

$A_1, A_2, \ldots \in \mathcal{C}$ is a partition of Ω if

- A_i and A_j are disjoint for $i \neq j$
- $\Omega = \cup_i A_i$

Rolling a six-sided die

We assign probabilities to the partition $\{1\}$, $\{2\}$, $\ldots$, $\{6\}$

$P(\{i\}) = \theta_i$ for $1 \leq i \leq 6$

$P(\{i\}) = \theta_i$ for $1 \leq i \leq 6$

What conditions should θ_1, θ_2,. . . , θ_6 satisfy?

Nonnegative and

$$\begin{aligned}\sum_{i=1}^{6} \theta_i &= \mathrm{P}(\cup_{i=1}^{6} \{i\}) \\ &= \mathrm{P}(\Omega) \\ &= 1\end{aligned}$$

$P(\{i\}) = \theta_i$ for $1 \leq i \leq 6$

What about the rest of events in the collection?

$$\begin{aligned} \mathrm{P}\left(\{2,4,6\}\right) &= \mathrm{P}\left(\{2\}\right) + \mathrm{P}\left(\{4\}\right) + \mathrm{P}\left(\{6\}\right) \\ &= \theta_2 + \theta_4 + \theta_6 \end{aligned}$$

What have we learned?

A probability space consists of

- A sample space Ω containing all possible outcomes
- A collection of events $\mathcal{C}$
- A probability measure P that assigns probabilities to the events in the collection

This sounds very complicated...

We never do this, instead we use random variables