1 Essentials of Probability
Lecture One May 24, 2022
Probability Spaces

The model: \( \Omega \), which is a set, is the set of possible outcomes of an experiment. The set \( \Omega \) is called the sample space.

Denote the number of possible outcomes by \(|\Omega|\) or \(#\Omega\). This is the cardinality of \( \Omega \).

Example: Toss a single coin, let \( \Omega = \{H,T\} \).

Example: Toss a single coin twice, observe the sequence of heads and tails. Then \( \Omega = \{HH,HT,TH,TT\} \).

Example: Toss a coin \( n \) times, then observe the sequence of heads and tails. The \( \Omega \) = the set of length \( n \) sequences of \( H's \) and \( T's \). Note that \(|\Omega| = 2^n\).

An event is ”something that may or may not happen”. We think of an event as a collection of outcomes, i.e. a subset of \( \Omega \).

Example: Toss a single coin twice, observe the sequence of heads and tails, and let \( E \) be the event ”heads at least once”. Then \( E \subset \Omega \) and \( E = \{HH,HT,TH\} \).

Note that \( E = \emptyset \) and \( E = \Omega \) are both events, sometimes called the 'impossible event’ and the ‘certain event’.

What subsets are events?

If \( \Omega \) is finite or countably infinite then there is no problem thinking of every subset of \( \Omega \) as an event. However when \( \Omega \) is uncountably infinite, for example if \( \Omega = \mathbb{R} \) or \( \Omega = [0,1] \), then there will problems if we allow every subset to be an event. We must consider a certain collection of subsets to be events, and this collection must have the following properties.

Definition 1.1. Given a set \( \Omega \), a collection of subsets \( \mathcal{F} \) is called a field if:

1. \( \emptyset \in \mathcal{F} \)
2. If \( A \in \mathcal{F} \) then \( A^c \in \mathcal{F} \)
3. If \( A,B \in \mathcal{F} \) then \( A \cup B \in \mathcal{F} \).

So a field is a collection of subsets close under compliments and unions and containing the empty set. Note that by De Morgan’s law, it is also closed under intersections. Also we have

Definition 1.2. Given a set \( \Omega \), a field of subsets \( \mathcal{F} \) is called a \( \sigma \)-field if:
4. If $A_1, A_2, A_3, \cdots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A probability measure is a way of assigning a probability to each event. A probability is a real number between 0 and 1.

**Definition 1.3.** Given $(\Omega, \mathcal{F})$ where $\mathcal{F}$ is a $\sigma$-field on $\Omega$, a probability measure on $(\Omega, \mathcal{F})$ is a function $P : \mathcal{F} \to [0, 1]$ such that:

1. $P(\emptyset) = 0$, $P(\Omega) = 1$
2. If $A_1, A_2, A_3, \cdots \in \mathcal{F}$ and are disjoint then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

**Definition 1.4.** The triple of objects $(\Omega, \mathcal{F}, P)$ is called a probability space.

**Example** Toss a coin twice, then observe the sequence of heads and tails $\Omega = \{HH, HT, TH, TT\}$ and $\mathcal{F} = 2^\Omega$, the power set of $\Omega$.

Here is an example of an acceptable probability measure:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>4/9</td>
</tr>
<tr>
<td>HT</td>
<td>2/9</td>
</tr>
<tr>
<td>TH</td>
<td>2/9</td>
</tr>
<tr>
<td>TT</td>
<td>1/9</td>
</tr>
</tbody>
</table>

So if $E$ is the event 'heads at least once', then $E = \{HH, HT, TH\}$ and $P(E) = 8/9$.

Here’s the same example, only this time it’s a fair coin:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>1/4</td>
</tr>
<tr>
<td>HT</td>
<td>1/4</td>
</tr>
<tr>
<td>TH</td>
<td>1/4</td>
</tr>
<tr>
<td>TT</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Now if $E$ is the event 'heads at least once', then $P(E) = 3/4$.

There are two approaches to deriving the probability assignment: Theoretical – deducing probability assignment from theory, and Empirical – inferring a probability assignment from data.

**Theoretical**
Example Roll a pair of fair dice. Then we have

\[ \Omega = \{(1, 1), (1, 2), (1, 3), \ldots, (2, 1), (2, 2), (2, 3), \ldots, (6, 6)\}. \]

Because the dice are fair, each of the 36 outcomes is equally likely. So

\[ P(\text{the sum is 7}) = P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{6}{36} = \frac{1}{6}. \]

This can be stated as a proposition.

**Proposition 1.5.** In a probability space, if \( \Omega \) is finite and all outcomes are equally likely, then for any event \( A \),

\[ P(A) = \frac{|A|}{|\Omega|}. \]

Note that the hypotheses are required in this proposition.

**Empirical Example** Suppose we test a drug for efficacy on 2,000 patients, and we find that the drug is effective on 200 patients and ineffective on 1,800. Then we deduce that \( P(\text{Efficacy}) = \frac{1}{10} \).

In other words we interpret the probability of an event as the relative frequency of the event occurring over many trials of the experiment.

Now we’ll do a few more examples.

**Example** Probability of getting a poker hand in a five-card draw. A deck has 52 cards, 13 denominations, 4 suits.

We deal a hand of 5 cards, \( \Omega \) is the set of subsets of the deck with 5 cards in them. We’ll use the formula for a combination. Denote by \( C(n, r) \) the number of ways of choosing \( r \) things from \( n \) things, without regard to order. Here is the formula:

\[ C(n, r) = \frac{n!}{(r!)(n-r)!} \]

So,

\[ | \Omega | = C(52, 5) = \frac{(52!)}{(5!)(47!)} = \frac{(52 \cdot 51 \cdot 50 \cdot 49 \cdot 48)}{(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} = 2,598,960. \]

In other words, there are roughly 2.5 million different 5 card hands that can be dealt from a deck of cards.

Now let’s compute the probability of getting 3 of a kind. This means that 3 of the cards have the same denomination, and the other 2 cards have different denominations, and not the same denomination as each other. (If the other 2 cards has the same denomination as each other, the hand would be called a full house). Denote the 3 of a kind by \( E \).
\[ P(E) = \frac{13 \cdot C(4, 3) \cdot C(12, 2) \cdot 16}{2,598,960} = 0.021128 \]

We also have the following formula for a permutation.

\[ P(n, r) = \frac{n!}{(n-r)!} \]

Let \( F \) be the event full house:

\[ P(F) = \frac{P(13, 2) \cdot C(4, 3) \cdot C(4, 2)}{2,598,960} = .001441 \]

It’s useful to note

\[
(r!)C(n, r) = P(n, r)  \\
C(n, r) = C(n, n-r)  \\
C(n, 1) = n  \\
C(n, n) = 1
\]

Before talking about conditional probability and independence, we should observe a few basic consequences of the definition of a probability space.

**Proposition 1.6.** 1. If \( A \subset B \), then \( P(A) \leq P(B) \).

2. \( P(A^c) = 1 - P(A) \).

3. If \( A, B \) are mutually exclusive, i.e. \( A \cap B = \emptyset \), then \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \).

4. If \( A_1 \subset A_2 \subset A_3 \subset \ldots \) and \( A = \bigcup_{n=1}^{\infty} A_n \) then \( P(A) = \lim_{n \to \infty} P(A_n) \).

5. If \( B_1 \supset B_2 \supset B_3 \supset \ldots \) and \( B = \bigcap_{n=1}^{\infty} B_n \) then \( P(B) = \lim_{n \to \infty} P(B_n) \).