Elements of Probability Theory and Combinatorics

Notes of Alessandra Micheletti

Graph Theory and Discrete Mathematics
Data Science and Economics

Probability space

Let Ω be any set and let Σ be some "appropriate" class of subsets of Ω.

Elements of Σ are called events.

For $A \subseteq \Omega$ we write A^C for the complement of A in Ω, i.e.

$$A^C = \{s \in \Omega : s \notin A\}.$$

Definition

A probability measure on Ω is a function $P : \Sigma \to [0, 1]$, satisfying

1. $P(\emptyset) = 0$.
2. $P(A^C) = 1 - P(A)$ for any event A.
3. If A and B are disjoint events (that is if $A \cap B = \emptyset$), then $P(A \cup B) = P(A) + P(B)$. More generally, if $A_1, A_2, \ldots$ is a countable sequence of disjoint events ($A_i \cap A_j = \emptyset$, for any $i \neq j$), then $P(\bigcup_{t=1}^{\infty} A_t) = \sum_{t=1}^{\infty} P(A_t)$.

Note that the first two conditions imply that $P(\Omega) = 1$.

The triple (Ω, Σ, P) is called probability space.

Conditional probability

Definition

If A and B are events, and $P(B) > 0$, we define the conditional probability of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Interpretation: $P(A|B)$ is how likely we consider that A happens, knowing that B happened

Example

$A =$ Tomorrow here will rain
$B =$ Today a storm occurred 100 Km on the west of my position

If I don't know anything about weather forecast or conditions in the surrounding (and I don't know if B occurred) I can only guess that $P(A) = P(\text{tomorrow here will rain}) = \frac{1}{2}$.

But if I know that B happened, it becomes more likely that tomorrow here will rain, thus
$P(A|B) > \frac{1}{2}$.

Independence

Definition

Two events A and B are said to be independent if

$$P(A \cap B) = P(A)P(B).$$

More in general

Definition

The events $A_1, \dots, A_k$ are said to be independent if for any $l \leq k$ and any $i_1, \dots, i_l \in \{1, \dots, k\}$ with $i_1 < i_2 < \dots < i_l$ we have

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_l}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_l}).$$

Note that if A and B are independent, then, since $P(A|B) = \frac{P(A \cap B)}{P(B)}$, we have

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

Then

$$P(A|B) = P(A)$$

Example.
$A =$ Tomorrow here will rain
$B =$ Today I make a cake
A is not influenced by B and viceversa, thus they are independent and $P(A|B) = P(A)$.

In practice, in particular if the space Ω is finite, we compute the probability of an event A as

$$P(A) = \frac{\#\ cases\ in\ favor\ of\ A}{\#\ possible\ cases}$$

The correct counting of cases is the subject of combinatorics.

Combinatorics: counting problems

[C.M.Grinstead, J.L.Snell, Introduction to Probability, AMS publisher, 1997 - Chapter 3]

Consider an experiment that takes place in several stages and is such that the number of outcomes m at the nth stage is independent of the outcomes of the previous stages.
The number m may be different for different stages.

We want to count the number of ways that the entire experiment can be carried out.

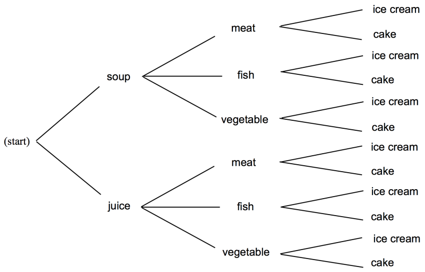
Example 1. You are eating at Emile's restaurant and the waiter informs you that you have

(a) two choices for appetizers: soup or juice;

(b) three for the main course: a meat, fish, or vegetable dish;

(c) two for dessert: ice cream or cake.

How many possible choices do you have for your complete meal?



Your menu is decided in three stage. At each stage the number of possible choices does not depend on what is chosen in the previous stages: two choices at the first stage, three at the second, and two at the third.

From the tree diagram we see that the total number of choices is the product of the number of choices at each stage.

In this example we have $2 \cdot 3 \cdot 2 = 12$ possible menus.

Our menu example is an example of the following general counting technique:

Counting technique. A task is to be carried out in a sequence of r stages. There are n_1 ways to carry out the first stage; for each of these n_1 ways, there are n_2 ways to carry out the second stage; for each of these n_2 ways, there are n_3 ways to carry out the third stage, and so forth. Then the total number of ways in which the entire task can be accomplished is given by the product

$$N = n_1 \cdot n_2 \cdot \cdots \cdot n_r$$

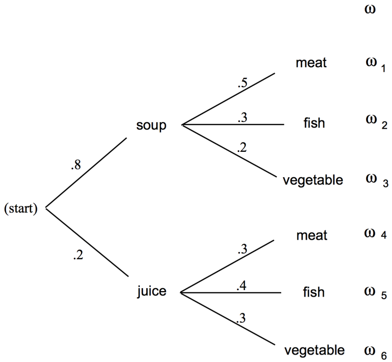
.

Tree diagrams

It will often be useful to use a tree diagram when studying probabilities of events relating to experiments that take place in stages and for which we are given the probabilities for the outcomes at each stage.

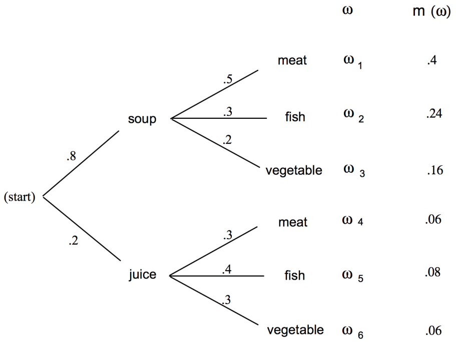
Example 1: consider only appetizers and main course, and assume that the owner of Emile's restaurant has observed that 80% of his customers choose the soup for an appetizer and 20% choose juice. Of those who choose soup, 50% choose meat, 30% choose fish, and 20% choose the vegetable dish.
Of those who choose juice for an appetizer, 30% choose meat, 40% choose fish, and 30% choose the vegetable dish.

We represent these probabilities on the tree diagram.



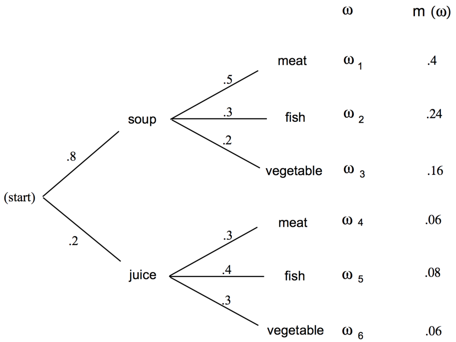
We choose for our sample space the set $\Omega = \{\omega_1, \ldots, \omega_k\}$ of all possible paths along the tree.

Question: what is the probability that a customer chooses first soup and then meat?



Answer: $P([1st\ soup]) \cdot P([2nd\ meat|1st\ soup]) = 0.8 \cdot 0.5 = 0.4$

This suggests choosing our probability distribution for each path through the tree to be the product of the probabilities at each of the stages along the path.



Note that $\sum_{i=1}^{6} m(\omega) = 1$. And if we want to know the probability that a customer eats meat, whatever appetizer he/she gets, we sum the probabilities of having meat:

$$P(meat) = 0.4 + 0.06 = 0.46$$