Lecture 2: Probability notions

A **probability space** consists of a set Ω called **laboratory** a set of subsets of Ω called **events** and a function P from events to the interval $[0,1]$ so that probabilities of disjoint sets add up and such that the entire laboratory has probability 1. Every point in Ω is an **experiment**. Think of an event A as a collection of experiments and of $P[A]$ as the likelihood that A occurs.

Examples:

1 We turn a wheel of fortune and assume it is fair in the sense that every angle range $[a,b]$ appears with probability $(b-a)/2\pi$. What is the chance that the wheel stops with an angle between 30 and 90 degrees?
Answer: The laboratory here is the circle $[0,2\pi)$. Every point in this circle is a possible experiment. The event that the wheel stops between 30 and 90 degrees is the interval $[\pi/6, \pi/2]$. Assuming that all angles are equally probable, the answer is $1/6$.

2 This example is called **Bertrand's paradox**. Assume we throw randomly a line into the unit disc. What is the probability that its length is larger than the length of the inscribed triangle?
Answer: Interestingly, the answer depends as we will see in the lecture.

3 Lets look at the DowJonesIndustrial average DJI from the start. What is the probability that the index will double in the next 50 years?
Answer: This is a strange question because we have **only one** data set. How can we talk about probability in this situation? One way is to see this graph as a sample of a larger probability space. A simple model would be to fit the data with some polynomial, then add random noise to it. The real DJI graph now looks very similar to a typical graph of those.

4 Lets look at the digits of π. What is the probability that the digit 5 appears? **Answer:** Also this is a strange example since the digits are not randomly generated. They are given by nature. There is no randomness involved. Still, one observes that the digits behave like a random number and that the number is "normal": every digit appears with the same frequency. This is independent of the base.

Here is a more precise list of conditions which need to be satisfied for events.

1. The entire laboratory Ω and the empty set $\emptyset$ are events.
2. If A_j is a sequence of events, then $\bigcup_j A_j$ and $\bigcap_j A_j$ are events.

It follows that also the complement of an event is an event.

Here are the conditions which need to be satisfied for the **probability function** P:

1. $0 \leq P[A] \leq 1$ and $P[\Omega] = 1$.
2. A_j are disjoint events, then $P[\bigcup_{j=1}^{\infty} A_j] = \sum_{j=1}^{\infty} P[A_j]$.

$P[\Omega \setminus A] = 1 - P[A]$.
$P[A \cup B] = P[A] + P[B] - P[A \cap B]$.

An important class of probability spaces are **finite probability spaces**, where every subset can be an event. The most natural choice is to assign them the probability $P[A] = |A|/|\Omega|$ where $|A|$ is the number of elements in A. This reads the "number of good cases" divided by the "number of all cases".

$$P[A] = \frac{|A|}{|\Omega|}$$

It is important that in any situation, we first find out what the laboratory is. This is often the hardest task. Once the setup is fixed, one has a combinatorics or counting problem.

5 We throw a dice twice. What is the probability that the sum is larger than 5?
Answer: We can enumerate all possible cases in a matrix and get Let

$$\Omega = \begin{bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{bmatrix} .$$

be the possible cases, then there are only 8 cases where the sum is smaller or equal to 8.

6 Lets look at all 2×2 matrices for which the entries are either 0 or 1. What is the probability that such a matrix has a nonzero determinant $det(A) = ad - bc$?
Answer: We have 16 different matrices. Our probability space is finite:

$$\Omega = \{ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \} .$$

Now lets look at the event that the determinant is nonzero. It contains the following matrices:

$$A = \{ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \} .$$

The probability is $P[A] = |A|/|\Omega| = 6/16 = 3/8 = 0.375$.

7 Lets pick 2 cards from a deck of 52 cards. What is the probability that we have 2 kings?
Answer: Our laboratory Ω has $52 * 51$ possible experiments. To count the number of good cases, note that there are $4 * 3 = 12$ possible ordered pairs of two kings. Therefore $12/(52 * 51) = 1/221$ is the probability.

Some notation

Set theory in Ω:
The **intersection** $A \cap B$ contains the elements which are in A **and** B.
The **union** $A \cup B$ contains the elements which are in A **or** B.
The **complement** A^c contains the elements in Ω which are **not** in A.
The **difference** $A \setminus B$ contains the elements which are in A but not in B.
The **symmetric difference** $A \Delta B$ contains what is in A or B but not in both.
The **empty set** $\emptyset$ is the set which does not contain any elements.

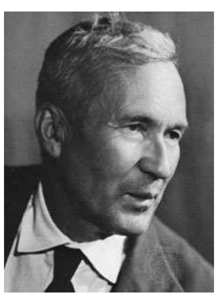
The algebra $\mathcal{A}$ of events:
If Ω is the laboratory, the set $\mathcal{A}$ of events is **σ-algebra**. It is a set of subsets of Ω in which one can perform countably many set theoretical operations and which contains Ω and $\emptyset$. In this set one can perform **countable unions** $\bigcup_j A_j$ for the union of a sequence of sets $A_1, A_2, \dots$ or **countable intersections** $\bigcap_j A_j$.

The probability measure P:
The probability function P from $\mathcal{A}$ to $[0,1]$ is assumed to be normalized that $P[\Omega] = 1$ and that $P[\bigcup_i A_i] = \sum_i P[A_i]$ if A_i are all disjoint events. The later property is called **σ-additivity**. One gets immediately that $P[A^c] = 1 - P[A]$, $P[\emptyset] = 0$ and that if $A \subset B$ then $P[A] < P[B]$.

The Kolmogorov axioms:
A **probability space** $(\Omega, \mathcal{A}, P)$ consists of a laboratory set Ω, a σ-algebra $\mathcal{A}$ on Ω and a probability measure P. The number $P[A]$ is the **probability** of an event A. The elements in Ω are called **experiments**. The elements in $\mathcal{A}$ are called **events**.

Some remarks:

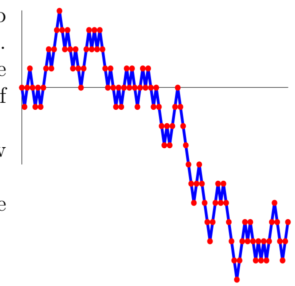
1) In a σ-algebra the operation $A \Delta B$ behaves like **addition** and $A \cap B$ like **multiplication**. We can "compute" with sets like $A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C)$. It is therefore an algebra. One calls it a **Boolean algebra**. Beside the just mentioned **distributivity**, one has **commutativity**, and **associativity**. The **"zero"** is played by $\emptyset$ because $\emptyset \Delta A = A$. The **"one"** is the set Ω because $\Omega \cap A = A$. The algebra is rather special because $A \cap A = A$ and $A \Delta A = \emptyset$. The "square" of each set is the set itself and adding a set to itself gives the zero set.



2) The Kolmogorov axioms form a solid foundation of probability theory. This has only been achieved in the 20th century (1931). Before that probability theory was a bit hazy. For infinite probability spaces it is necessary to restrict the set of all possible events. One can not take all subsets of an interval for example. There is no probability measure P which would work with all sets. There are just too many.

Homework due February 2, 2011

1 You walk 100 steps and chose in each step randomly one step forward or backward. You flip a coin. What is the chance to be back at your starting point 0 at the end of your walk?
a) Set up the probability space Ω. How many elements does it have?
b) Which subset A of Ω is the event to be back at 0 at time 100?
c) Find the probability $P[A]$.
d) What formula do you get for n steps.



2 Do problem 5) in Chapter 2 of the text but with 100 instead of 1000. You choose a random number from $\{1, \dots, 100\}$, where each of the numbers have the same probability. Let A denote the event that the number is divisible by 3 and B the event that the number is divisible by 5. What is the probability $P[A]$ of A, the probability $P[B]$ of B and the probability of $P[A \cap B]$? Compute also the probability $P[A \cap B]/P[B]$ which we will call the **conditional probability** next time. It is the probability that the number is divisible by 3 under the condition that the number is divisible by 5.

3 You choose randomly 6 cards from 52 and do not put the cards back. What is the probability that you got all aces? Make sure you describe the probability space Ω and the event A that we have all aces.