# Lecture 5

Andrei Antonenko

Feb 02, 2005

## 1 Conditional independence

As we saw in previous lectures, conditional probabilities form a probability law. We can generalize a notion of independence of events to the case of conditional probabilities.

**Definition 1.1.** *Given an event $C$, two events $A$ and $B$ are called* ***conditionally independent****, if*

$$P(A \cap B|C) = P(A|C)P(B|C). \tag{1}$$

This definitional is parallel with the definition of the conditional probability given before: $P(A \cap B) = P(A)P(B)$. From the definition of conditional probability and multiplication rule, we get:

$$\begin{aligned} P(A \cap B|C) &= \frac{P(A \cap B \cap C)}{P(C)} = \\ &= \frac{P(C)P(B|C)P(A|B \cap C)}{P(C)} = \\ &= P(B|C)P(A|B \cap C). \end{aligned}$$

Combining the last equality with the definition of conditionally independent events, we obtain:

$$P(A|C) = P(A|B \cap C). \tag{2}$$

In words, this relation states that if $C$ is known to have occurred, the additional knowledge whether $B$ occurred or not, does not alter the probability of $A$.

The following examples will show, that independence of two events $A$ and $B$ (unconditional) does not imply conditional independence, and vice versa.

**Example 1.2.** *Consider an experiment, consisting of 2 tosses of a fair coin. The events we are interested in are:*

$$H_1 = \{\textit{1st toss resulted in } H\} = \{HH, HT\} \qquad P(H_1) = \frac{1}{2};$$

$$H_2 = \{\textit{2nd toss resulted in } H\} = \{HH, TH\} \qquad P(H_2) = \frac{1}{2};$$

$$D = \{\textit{1st and 2nd tosses had different outcomes}\} = \{HT, TH\} \qquad P(D) = \frac{1}{2}.$$

*Obviously, the events $H_1$ and $H_2$ are independent. Let's check whether they are conditionally independent.*

$$\begin{aligned}
P(H_1|D) &= \frac{P(H_1 \cap D)}{P(D)} = \frac{1/4}{1/2} = 1/2; \\
P(H_2|D) &= \frac{P(H_2 \cap D)}{P(D)} = \frac{1/4}{1/2} = 1/2; \\
P(H_1 \cap H_2|D) &= \frac{P(H_1 \cap H_2 \cap D)}{P(D)} = \frac{0}{1/2} = 0.
\end{aligned}$$

*Therefore, since $P(H_1|D)P(H_2|D) \neq P(H_1 \cap H_2|D)$, events $H_1$ and $H_2$ are not conditionally independent.*

**Example 1.3.** *There are two coins: Red and Blue. At the beginning we choose a random coin with probability $1/2$, and than make 2 tosses. The coins are biased: the Red coin results in $H$ with probability $0.99$, and the Blue coin results in $H$ with probability $0.01$. Let events $H_1$ and $H_2$ be the same as in previous example. Let $B$ be the event of selecting Blue coin, and $R$ be the event of selecting Red coin.*

*Given the choice of a coin, the events $H_1$ and $H_2$ are independent, because tosses are independent. So,*

$$P(H_1 \cap H_2|R) = P(H_1|R)P(H_2|R).$$

*Thus, the events $H_1$ and $H_2$ are conditionally independent, given choice of the coin.*

*Now, let's check whether these events are (unconditionally) independent. By law of total probability, we can obtain:*

$$\begin{aligned}
P(H_1) &= P(H_1|R)P(R) + P(H_1|B)P(B) = \\
&= 0.99 \cdot \frac{1}{2} + 0.01 \cdot \frac{1}{2} = \frac{1}{2}; \\
P(H_2) &= P(H_2|R)P(R) + P(H_2|B)P(B) = \\
&= 0.99 \cdot \frac{1}{2} + 0.01 \cdot \frac{1}{2} = \frac{1}{2}; \\
P(H_1 \cap H_2) &= P(H_1 \cap H_2|R)P(R) + P(H_1 \cap H_2|B)P(B) = \\
&= 0.99 \cdot 0.99 \cdot \frac{1}{2} + 0.01 \cdot 0.01 \cdot \frac{1}{2} \approx \frac{1}{2}.
\end{aligned}$$

*So, we see that $P(H_1 \cap H_2) \neq P(H_1)P(H_2)$, and thus events $H_1$ and $H_2$ are not independent.*

# 2 Counting

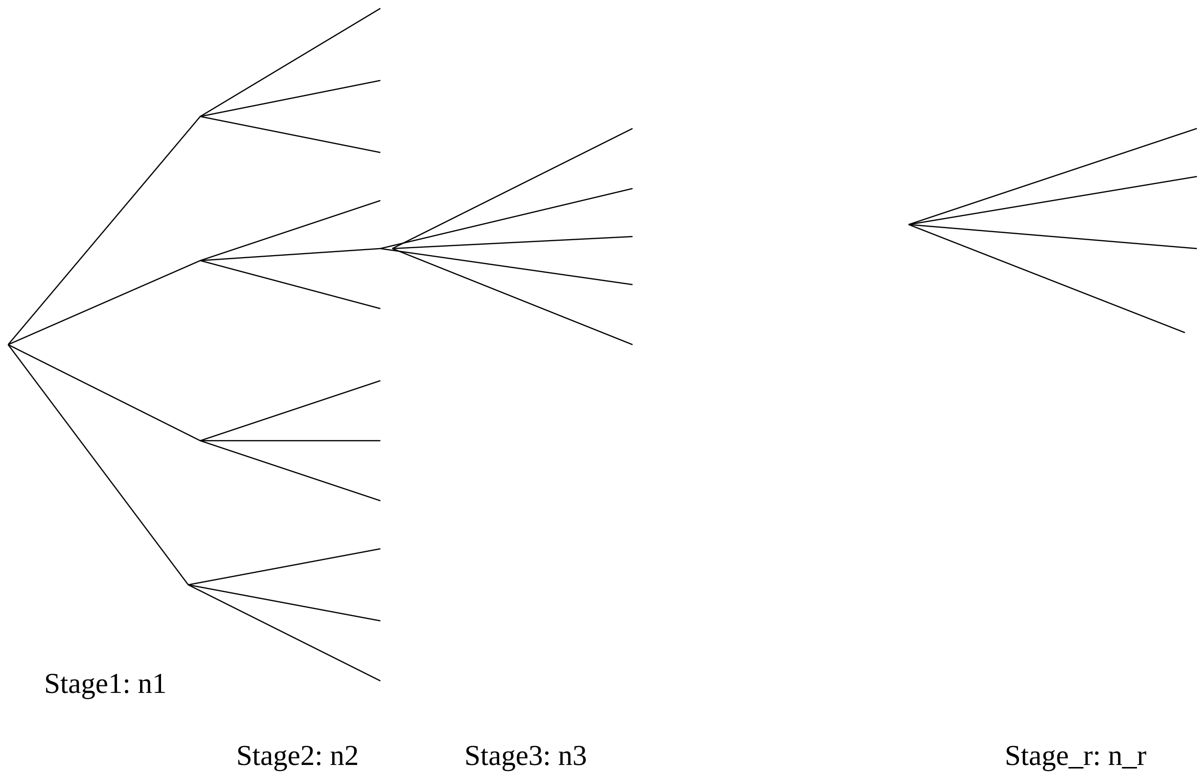
Assume the sample space of the experiment $\Omega$ is finite, and all the outcomes are equally likely to happen. In this case, for any event $A$,

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega}. \tag{3}$$

The problem of finding the number of elements in $A$ and in $\Omega$ is not necessarily an easy one. In this section we will discuss ways of counting, which can help in solving several problems.

## 2.1 Counting principle

Consider a multistage process with $r$ stages. Assume, there are $n_1$ outcomes of the 1st stage. For each of the outcomes of the 1st stage, there are $n_2$ possible outcomes of the 2nd stage. For any possible outcomes of the first two stages there are $n_3$ possible outcomes of the 3rd stage, etc, and for any possible outcomes of the first $r-1$ stages there are $n_r$ possible outcomes of the $r$th stage. Schematically,




In this case the total number of possible outcomes of such a process is equal to

$$n_1 \cdot n_2 \cdot n_3 \dots n_r.$$

**Example 2.1.** *What is the total number of phone numbers in a given area code? The first digit can be chosen in 8 different ways (0 and 1 can not be first digits of the phone number), for any choice of the first digit, the second digit can be chosen in 10 different ways, etc. Thus the total number of phone numbers is*

$$8 \cdot \underbrace{10 \cdot 10 \dots 10}_{6 \text{ times}}$$

**Example 2.2.** *Assume a set $A$ has $n$ elements. What is the total number of subsets of the set $A$?*

*For the first element of $A$ we have 2 choices: include it into the subset or not, the same happens for the second element, etc. Thus, the total number of subsets is*

$$\underbrace{2 \cdot 2 \dots 2}_{n \text{ times}} = 2^n.$$

## 2.2 Permutations

Assumer there are $n$ objects, and assume $k \leq n$. What is the total number of different $k$-object sequences? These sequences are called **$k$-permutations**.

We have $n$ choices for the 1st elements of the sequence, $n-1$ choices for the 2nd element of the sequence, etc., and $n-k+1$ choices for the $k$th element of the sequence. Thus, the total number of sequences is

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!},$$

where $n!$ is defined as

$$n! = 1 \cdot 2 \cdot 3 \dots (n-1) \cdot n. \tag{4}$$

**Example 2.3.** *What is the number of words, consisting of 4 different letters?*

*The English alphabet has 26 letters, and we need to count all 4 letter sequences. This number of sequences is equal to*

$$\frac{26!}{(26-4)!} = \frac{26!}{22!} = 26 \cdot 25 \cdot 24 \cdot 23.$$

**Example 2.4.** *Assume you have a CD collection with $n_1$ CDs with classical music, $n_2$ CDs with rock music, and $n_3$ CDs with pop music. You want to arrange your collection on the shelf, putting all CDs of the same genre together. How many ways do you have to do it?*

*CDs with classical music can be arranged in $n_1!$ different ways, CDs with rock music can be arranged in $n_2!$ different ways, and CDs with pop music can be arranged in $n_3!$ different ways. The three groups on the shelf can be arranged in 3! different ways. Thus, the total number of arrangements is equal to*

$$3! \cdot n_1! \cdot n_2! \cdot n_3!$$

## 2.3 Combinations

Assume the set $A$ consists of $n$ elements. What is the total number of $k$-element subsets of $A$?

The difference between combinations and permutations is that in the combinations there is no ordering of its elements. For example, if $A = \{a, b, c, d\}$, then the 2-letter permutations are

$$ab, ac, ad, ba, ca, da, bc, bd, cb, db, cd, dc,$$

and combinations are

$$ab, ac, ad, bc, bd, cd.$$

Let's note that selecting a permutation is the same as selecting a combination and ordering its elements. If $k$ is the size of permutation/combination, then there are $k!$ orderings, and thus

$$\# \text{ of permutations} = k! \cdot \# \text{ of combinations}.$$

Therefore, the number of combinations of the size $k$ is given by the following formula:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}. \tag{5}$$

These numbers are called **binomial coefficients**, and are read as "$n$ choose $k$."

There are several interesting identities about binomial coefficients worth mentioning.

The total number of subsets of the $n$-element set is (as we saw before) $2^n$. In another way, the total number of subsets is equal to total number of 0 element subsets, total number of 1-element subsets, total number of 2-element subsets, etc. Thus, we get the following identity:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n,$$

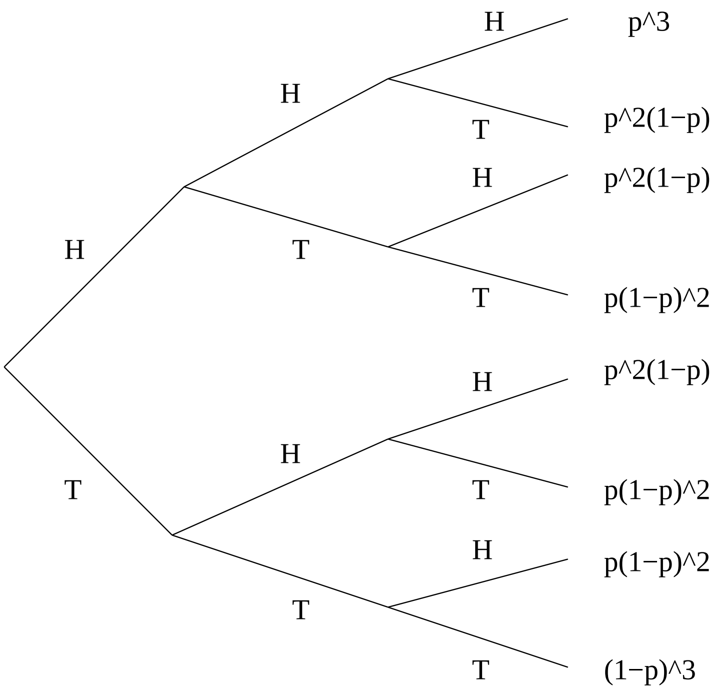
or

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n. \tag{6}$$

## 2.4 Bernoulli trials

Now we will look at the following experiment. Assume we toss a coin $n$ times. The coin has probability $p$ of a Head and probability $(1-p)$ of a Tail. That's an $n$-stage experiment, with each stage having exactly 2 outcomes: $H$ or $T$. Thus, the total number of outcomes in $2^n$. They are sequences of $T$'s and $H$'s of the length $n$.

Schematically, the experiment is represented on the following picture (for the case $n = 3$).




What is the probability of each outcome of this event? Assume, the sequence has $k$ heads and $n-k$ tails. Then, the probability is $p^k(1-p)^{n-k}$.

Assume we would like to compute the probability of getting exactly $k$ heads. This probability is equal to:

$$\begin{aligned} P(k \text{ heads}) &= \#\text{ of sequences with } k \text{ heads and } (n-k) \text{ tails} \cdot \text{probability of each such sequence} = \\ &= \#\text{ of sequences with } k \text{ heads and } (n-k) \text{ tails} \cdot p^k(1-p)^{n-k}. \end{aligned}$$

Now, we need to know the total number of sequences with $n$ heads and $(n-k)$ tails. This can be counted using the formula for number of combinations. We have $k$ out of $n$ slots available for

$H$s. These $k$ slots can be chosen in $\binom{n}{k}$ different ways. Thus, the total number of the sequences of the length $n$ with exactly $k$ heads is equal to $\binom{n}{k}$. Therefore, we obtain the following formula:

$$P(\text{getting exactly } k \text{ heads out of } n \text{ tosses}) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{7}$$

Such **trials**, or experiments with exactly two outcomes (like coin tosses) are called **Bernoulli trials**. The experiment described above is called a **sequence of independent Bernoulli trials**. The probabilities from the equation (7) are called **binomial probabilities**.

Since all binomial probabilities should add up to 1, we have the following identity:

$$\binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} + \binom{n}{2} p^2 (1-p)^{n-2} + \cdots + \binom{n}{n-1} p^{n-1} (1-p)^1 + \binom{n}{n} p^n = 1,$$

or

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1. \tag{8}$$