Probability for Computer Science

Spring 2021

Lecture 5

Prof. Claire Monteleoni

Boulder

Today

- Conditional Probability (continued)
- Multiplication Rule
- The Total Probability Theorem
  - Divide and Conquer method
- Bayes' Rule
- Independence

If time:

- Counting
- Independent trials and Binomial probabilities

Conditional Probability

A technique to reason about the outcome of an experiment, given partial information.

E.g.

How likely is it that a person has a particular disease, given that the medical test for it turned out negative?

If a word starts with the letter t, what is the probability that its second letter is h?

Properties of Conditional Probability

- The conditional probability of an event A, given an event B with $\mathbf{P}(B) > 0$, is defined by

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)},$$

and specifies a new (conditional) probability law on the same sample space Ω. In particular, all properties of probability laws remain valid for conditional probability laws.

- Conditional probabilities can also be viewed as a probability law on a new universe B, because all of the conditional probability is concentrated on B.

- If the possible outcomes are finitely many and equally likely, then

$$\mathbf{P}(A \mid B) = \frac{\text{number of elements of } A \cap B}{\text{number of elements of } B}.$$

Credit: Bertsekas & Tsitsiklis, 2008

Multiplication Rule

Assuming that all of the conditioning events have positive probability, we have

$$\mathbf{P}\big(\cap_{i=1}^{n} A_i\big) = \mathbf{P}(A_1)\mathbf{P}(A_2 \,|\, A_1)\mathbf{P}(A_3 \,|\, A_1 \cap A_2) \cdots \mathbf{P}\big(A_n \,|\, \cap_{i=1}^{n-1} A_i\big).$$

The multiplication rule can be verified by writing

$$\mathbf{P}\big(\cap_{i=1}^{n} A_i\big) = \mathbf{P}(A_1) \cdot \frac{\mathbf{P}(A_1 \cap A_2)}{\mathbf{P}(A_1)} \cdot \frac{\mathbf{P}(A_1 \cap A_2 \cap A_3)}{\mathbf{P}(A_1 \cap A_2)} \cdots \frac{\mathbf{P}\big(\cap_{i=1}^{n} A_i\big)}{\mathbf{P}\big(\cap_{i=1}^{n-1} A_i\big)}$$

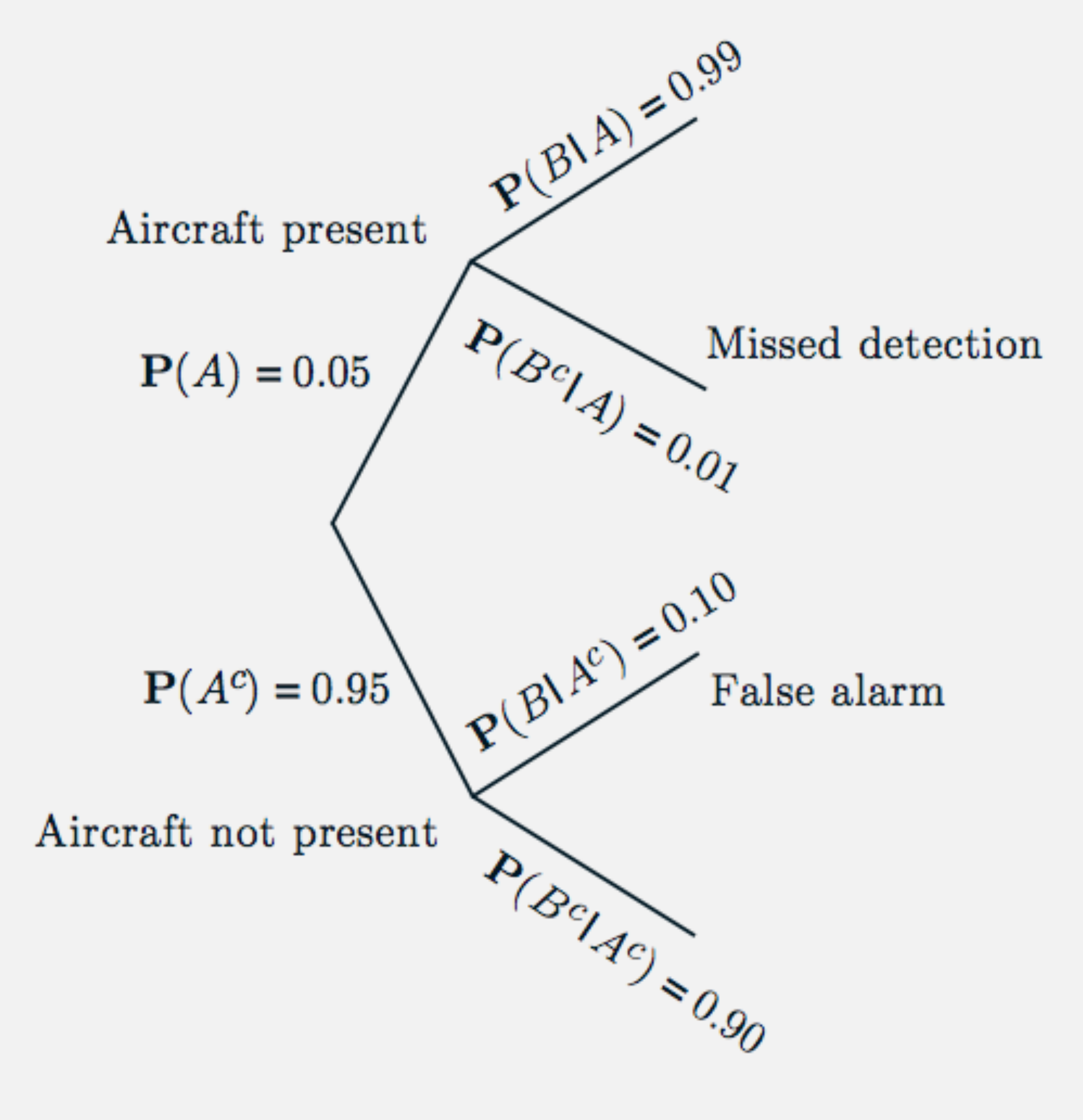
Credit: Bertsekas & Tsitsiklis, 2008

Example

- If there's an aircraft in a certain region, the radar system generates an alarm with probability 0.99.
- If there's no aircraft in the region, the radar system generates an alarm with probability 0.1.
- The probability of an aircraft being in the region is 0.05.

Q: What is the probability that an aircraft is in the region, and the radar system does not generate an alarm?

Credit: Bertsekas & Tsitsiklis, 2008



Credit: Bertsekas & Tsitsiklis, 2008

The Sequential Method

(a) We set up the tree so that an event of interest is associated with a leaf. We view the occurrence of the event as a sequence of steps, namely, the traversals of the branches along the path from the root to the leaf.

(b) We record the conditional probabilities associated with the branches of the tree.

(c) We obtain the probability of a leaf by multiplying the probabilities recorded along the corresponding path of the tree.

Given enough practice, you can apply the Multiplication Rule without drawing a tree.

Credit: Bertsekas & Tsitsiklis, 2008

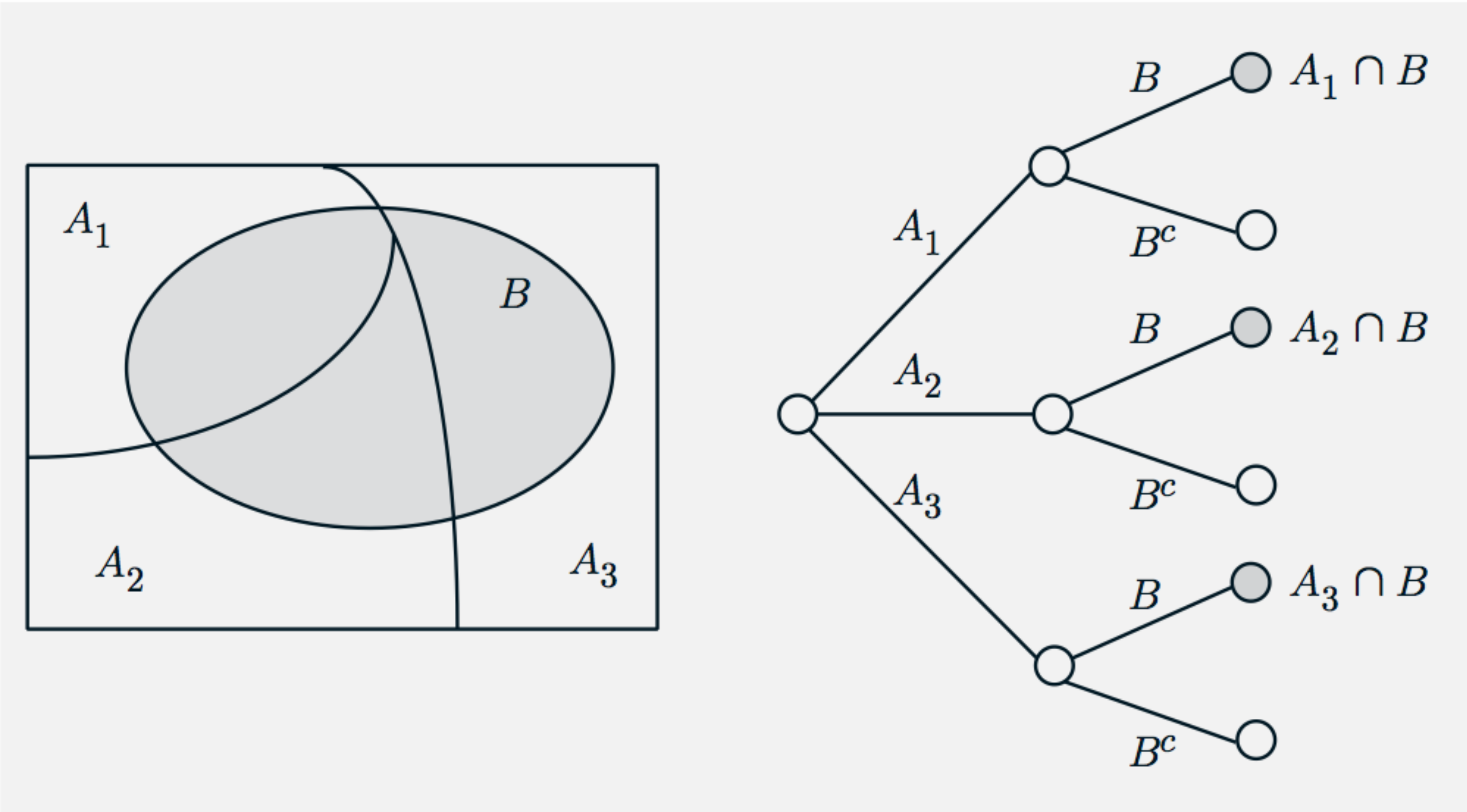
Total Probability Theorem

Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in exactly one of the events $A_1, \ldots, A_n$) and assume that $\mathbf{P}(A_i) > 0$, for all i. Then, for any event B, we have

$$\begin{aligned}\mathbf{P}(B) &= \mathbf{P}(A_1 \cap B) + \cdots + \mathbf{P}(A_n \cap B)\\ &= \mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n).\end{aligned}$$

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

Credit: Bertsekas & Tsitsiklis, 2008



Credit: Bertsekas & Tsitsiklis, 2008

Divide and Conquer Method

- Choose a set of events A_1 , … , A_n that partition Ω and have known probabilities, $P(A_i)$.

- Compute $P(B \mid A_i)$ for each $i : 1 \leq i \leq n$.

- Solve for P(B) using Total Probability Theorem:

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

Bayes' Rule

Bayes' Rule

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $\mathbf{P}(A_i) > 0$, for all i. Then, for any event B such that $\mathbf{P}(B) > 0$, we have

$$\begin{aligned}\mathbf{P}(A_i \mid B) &= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(B)} \\ &= \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\mathbf{P}(A_1)\mathbf{P}(B \mid A_1) + \cdots + \mathbf{P}(A_n)\mathbf{P}(B \mid A_n)}.\end{aligned}$$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

Credit: Bertsekas & Tsitsiklis, 2008

Independence

- Two events A and B are said to be **independent** if

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

  If in addition, $\mathbf{P}(B) > 0$, independence is equivalent to the condition

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

- If A and B are independent, so are A and B^c.

Credit: Bertsekas & Tsitsiklis, 2008

Definition of Independence of Several Events

We say that the events $A_1, A_2, \ldots, A_n$ are **independent** if

$$\mathbf{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbf{P}(A_i), \qquad \text{for every subset } S \text{ of } \{1, 2, \ldots, n\}.$$

Credit: Bertsekas & Tsitsiklis, 2008

For the case of three events, A_1, A_2, and A_3, independence amounts to satisfying the four conditions

$$
\begin{aligned}
\mathbf{P}(A_1 \cap A_2) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2), \\
\mathbf{P}(A_1 \cap A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_3), \\
\mathbf{P}(A_2 \cap A_3) &= \mathbf{P}(A_2)\,\mathbf{P}(A_3), \\
\mathbf{P}(A_1 \cap A_2 \cap A_3) &= \mathbf{P}(A_1)\,\mathbf{P}(A_2)\,\mathbf{P}(A_3).
\end{aligned}
$$

Credit: Bertsekas & Tsitsiklis, 2008

Example

Experiment: toss a fair coin two times, independently.

H_1 = {The first toss is a head}

H_2 = {The second toss is a head}

D = {The two tosses have different outcomes}

Independence

- Two events A and B are said to be **independent** if

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

If in addition, $\mathbf{P}(B) > 0$, independence is equivalent to the condition

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

- If A and B are independent, so are A and B^c.
- Two events A and B are said to be **conditionally independent**, given another event C with $\mathbf{P}(C) > 0$, if

$$\mathbf{P}(A \cap B \mid C) = \mathbf{P}(A \mid C)\mathbf{P}(B \mid C).$$

If in addition, $\mathbf{P}(B \cap C) > 0$, conditional independence is equivalent to the condition

$$\mathbf{P}(A \mid B \cap C) = \mathbf{P}(A \mid C).$$

- Independence does not imply conditional independence, and vice versa.

Credit: Bertsekas & Tsitsiklis, 2008

Equally likely outcomes

Recall the Discrete Uniform Probability Law: If the sample space has a finite number of equally likely outcomes, then the probability of any event A is:

$$P(A) = \frac{|A|}{|\Omega|}$$

If we know the probability, $p = \frac{1}{|\Omega|}$, of each outcome, then we can compute the the probability of any event A as:

$$P(A) = p \cdot |A|$$

The Counting Principle

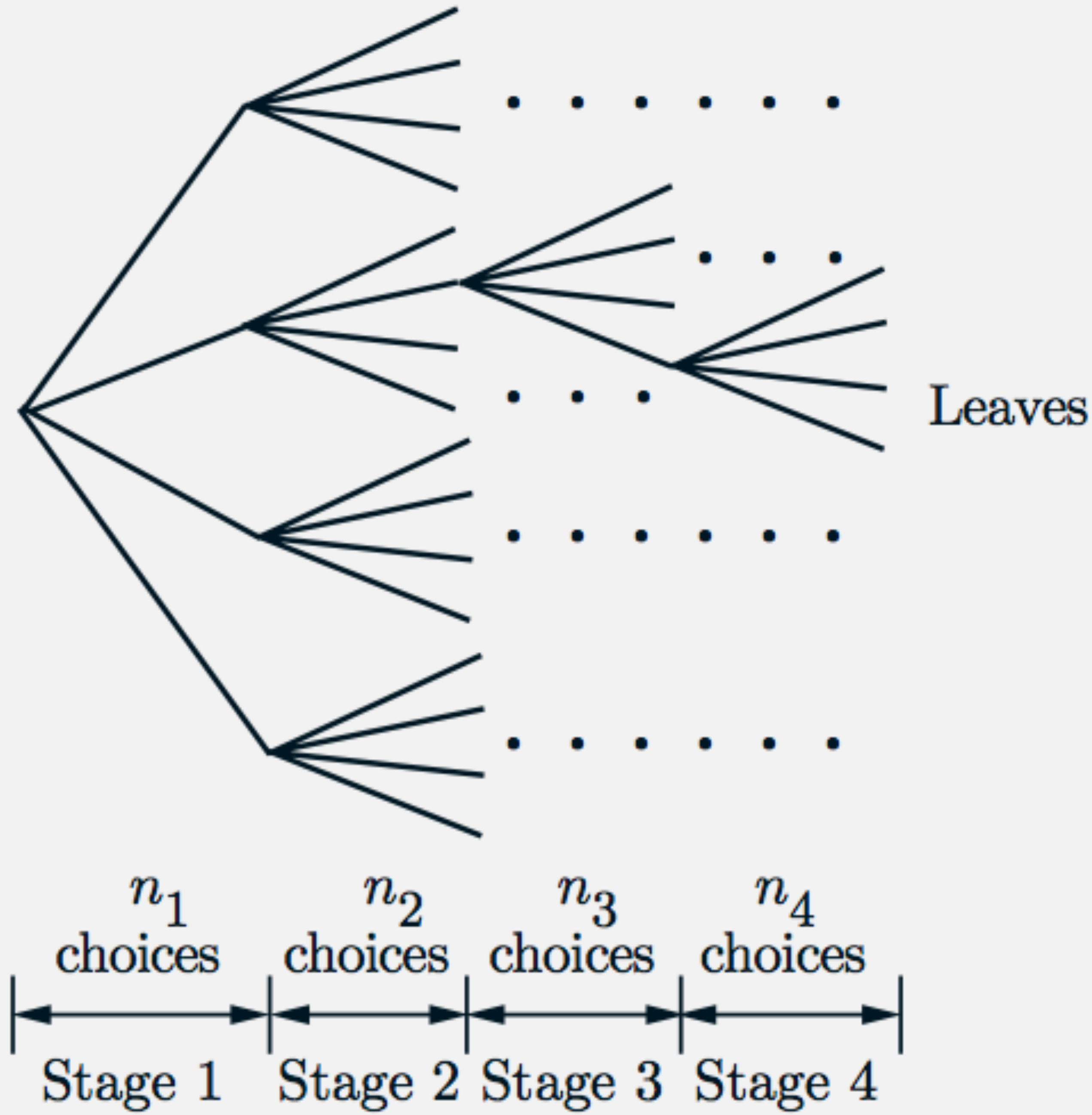
Consider an experiment consisting of r stages, such that:

- There are n_1 possible results at the first stage.
- For any result at the first stage, there are n_2 possible results at the second stage.
- For any result at stage i-1, there are n_i possible results at stage i.

Then, the total number of possible outcomes of the r-stage experiment is:

$$\boxed{|\Omega| = n_1 n_2 \cdots n_r}$$

Product, over r stages, of number of results possible at each stage.



Credit: Bertsekas & Tsitsiklis, 2008

Examples

1. Count the number of possible telephone numbers of the following form:

   7 digit sequence that does not start with 0 or 1.

2. Count the number of subsets of a set containing n elements, $S = \{s_1, s_2, \dots, s_n\}$.

3. Count the number of ways to order a deck of 52 cards.

k-permutations

Given n distinct objects, how many distinct k-object sequences are possible? ($0 < k \leq n$)

- There are n choices for the first object in the sequence.
- There are n - 1 choices for the second, since one object has already been placed in the sequence.
- There are n - (i - 1) = n - i + 1 choices for the i-th, since i - 1 objects have already been placed in the sequence.
- There are n - k + 1 choices for the k-th. This is the last object in the sequence.

Thus, the total number of distinct k-object sequences is:

$$\boxed{n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}}$$

Example

How many ways are there to order your books on a shelf, such that your Computer Science books are all together, your novels are all together, and your math books are all together?

- You have c Computer Science books
- You have n novels
- You have m Math books
- Shelf can hold c + n + m books

Combinations

Given a set containing n elements, how many subsets of size k are there? (0 < k ≤ n)

- Notice: unlike permutations, the order of the k objects does not matter.

The answer is $\binom{n}{k}$ which is called "n choose k" and is defined as:

$$\boxed{\binom{n}{k} = \frac{n!}{k!(n-k)!}}$$

Example

Count the number of distinct pairs of (2) cards, in a deck of 52 cards, where the order within the pair does not matter.

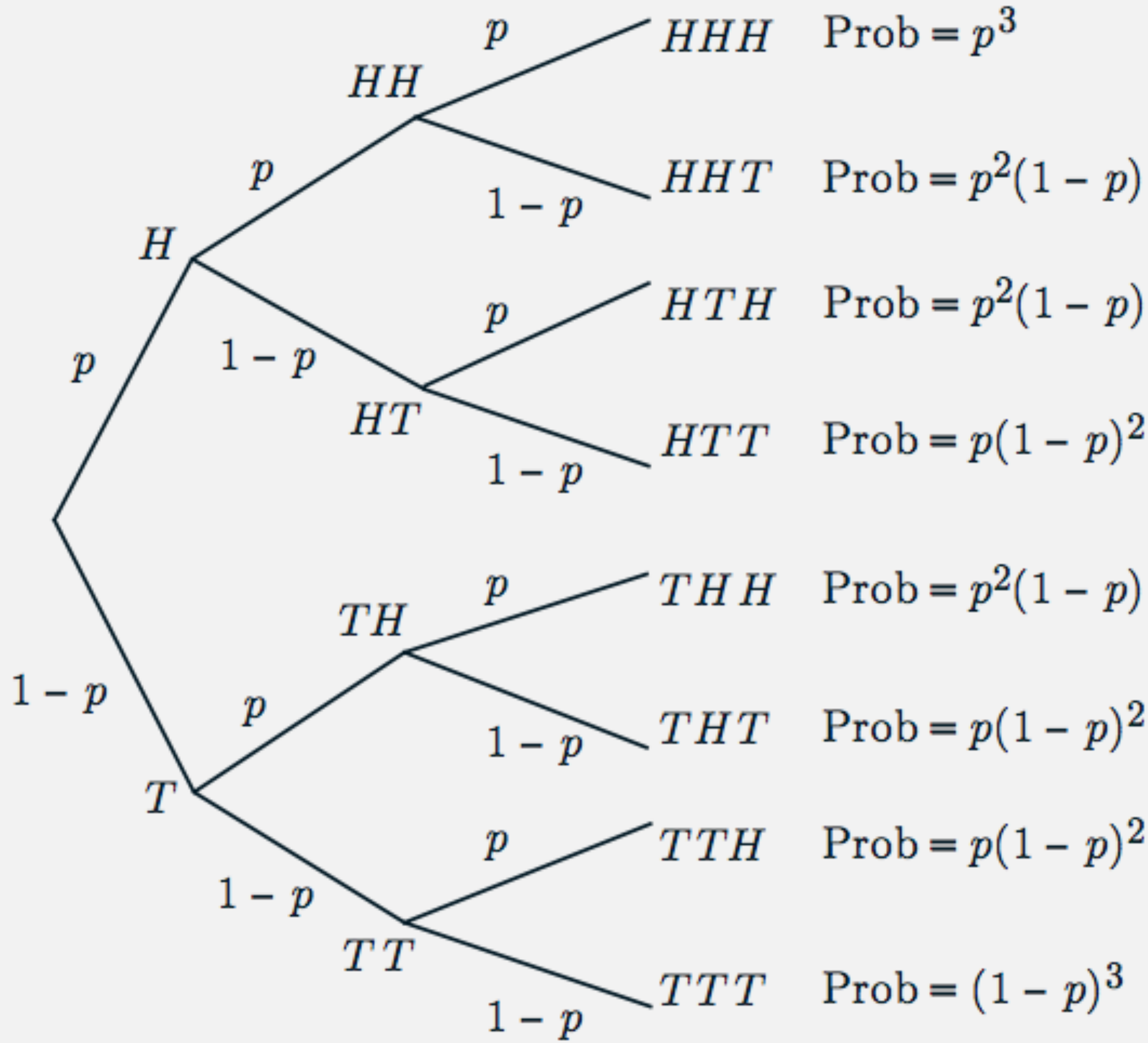
Independent Trials

An experiment that consists of a sequence of independent but identical stages is called a sequence of independent trials. E.g.

- Repeatedly flipping the same coin
- Repeatedly rolling the same die

In the special case where there are only 2 possible outcomes this is called a sequence of independent Bernoulli trials. E.g.

- Repeatedly flipping the same coin
- Repeatedly receiving emails that are either spam or not spam



Credit: Bertsekas & Tsitsiklis, 2008

Binomial probabilities

What is the probability that exactly k heads come up in a sequence of n independent coin tosses?

- Define the event A = {the sequence contains exactly k heads}.
- Define the coin's probability of heads, P[H] = p.
- The probability of each outcome in A:

  For any particular sequence containing exactly k heads, its probability is: $p^k(1-p)^{n-k}$

- |A| = the number of sequences containing exactly k heads = $\binom{n}{k}$
- P(A) = |A|× (Probability of each outcome in A)

The answer is therefore: $P(A) = \boxed{p(k) = \binom{n}{k} p^k (1-p)^{n-k}}$

Example

- Assume a cell phone provider can handle up to d data requests at once.
- Assume that every minute, each of its n customers makes a data request with probability p, independent of the other customers.

1. What is the probability that exactly c customers will make a data request during a particular minute?
2. What is the probability that more than d customers will make a data request during a particular minute?