Probability for Computer Science

Spring 2021

Lecture 7

Prof. Claire Monteleoni

Boulder

Today

- Independence (continued)
- Counting
- Independent trials and Binomial probabilities
- Discrete Random Variables

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

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

- 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

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

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

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

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

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

The Counting Principle

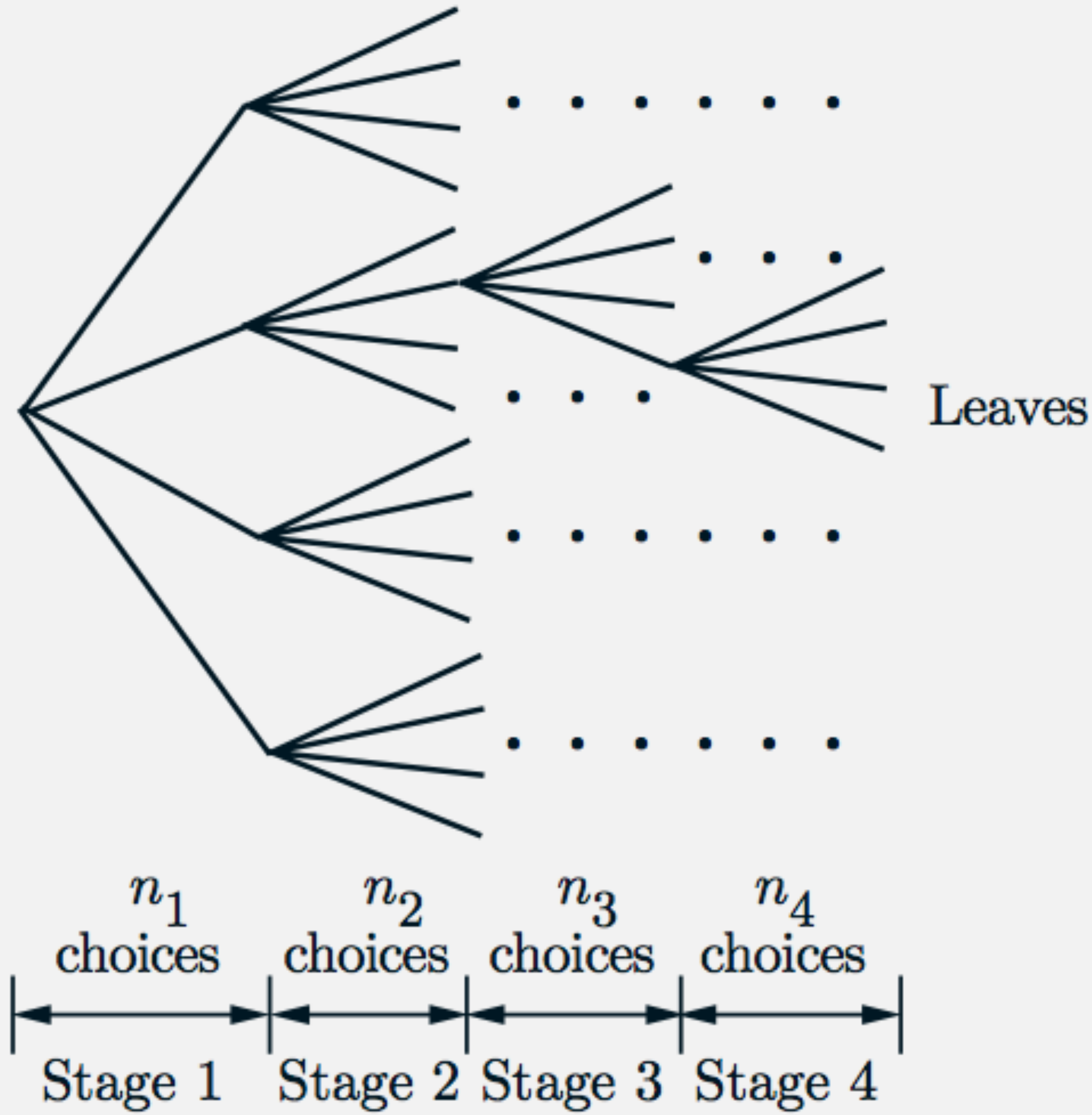
Consider an experiment consisting of r stages, such that:

- 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, $S = \{s_1, s_2, \ldots, s_n\}$.

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? ($0 < k \leq n$)

- 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.

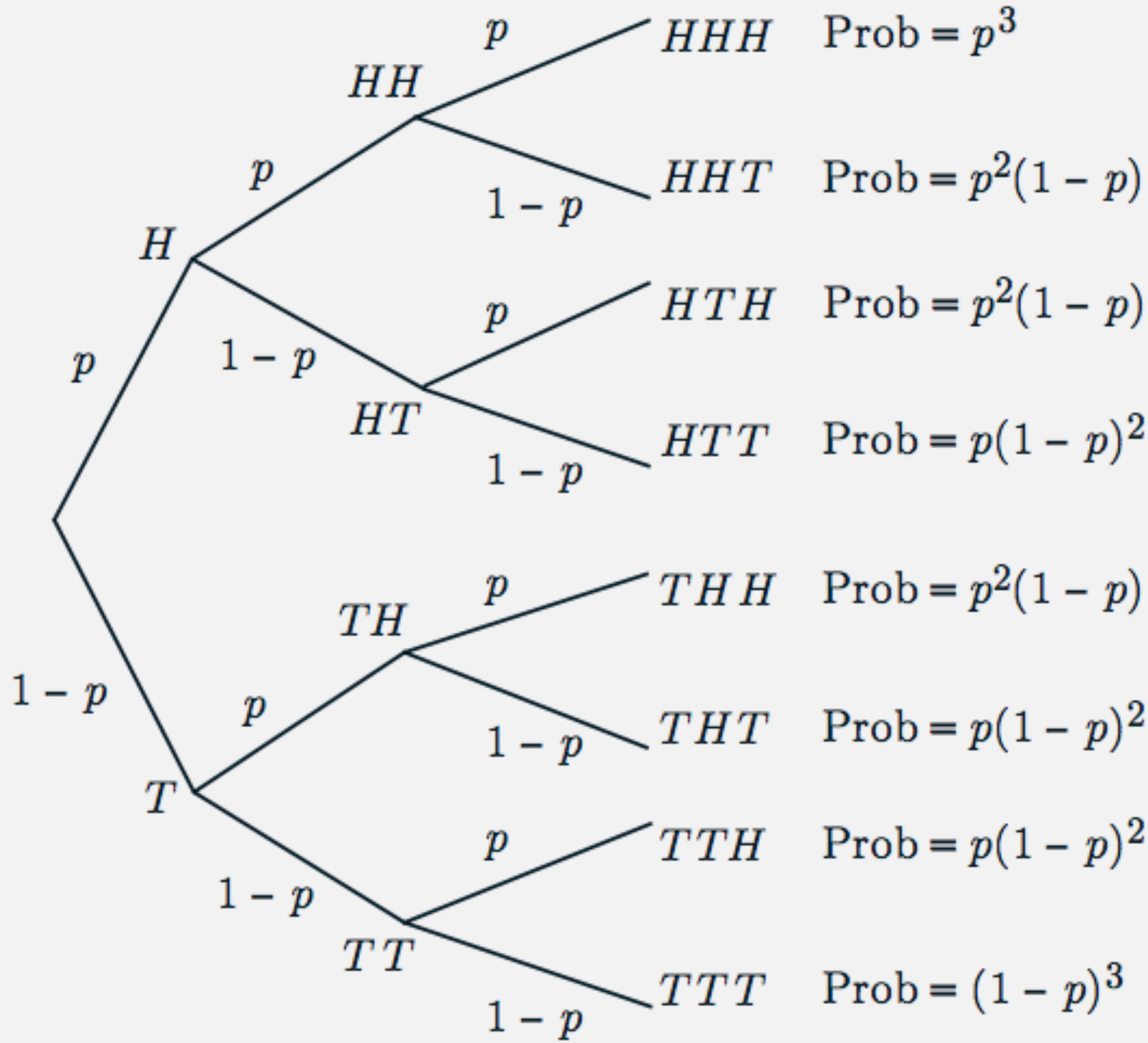
Independent Trials

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?

Random Variables

A random variable (r.v.) is a real-valued function of the outcome of an experiment. $f : \Omega \to \mathbb{R}$

Ex. A) If the experiment is rolling a die twice, here are some r.v.s:

- The sum of the two rolls
- The number of 3's rolled
- (Value of the second roll)^3

Ex. B) If the experiment is transmitting a message, here are some r.v.s:

- The number of bits that are transmitted incorrectly
- The time the message takes to be transmitted on the network

Discrete Random Variables

A random variable is discrete if the set of values it can take is finite, or countably infinite.

Ex. A) all the r.v.s listed are discrete.

Ex. B) if the r.v.s listed here take integer values (bits, milliseconds), then they are discrete.

Ex. C) Experiment is picking a point a in [-1, 1].

- Define r.v. as a^2. Is this a discrete r.v.?
- Define r.v. as sign(a). Is this a discrete r.v.?

Discrete Random Variables

Given a probabilistic model of an experiment:

- A discrete random variable is a real-valued function of the outcome of the experiment, that can take a finite or countably infinite number of values.

- A discrete r.v. has a probability mass function (PMF) which gives the probability of each numerical value that the r.v. can take.

- A function of a discrete random variable is itself a discrete random variable.
  - Its PMF can be computed from the PMF of the original r.v.

Probability Mass Function (PMF)

Each r.v., X, has an associated PMF, defined as follows, for each value x that X can take:

$$p_X(x) = P(\{X = x\})$$

To compute the PMF of a random variable X:

For each possible value x of X:

Collect all the possible outcomes in the event {X = x}

Sum their probabilities to obtain $p_X(x)$

For simplicity, we will use the notation:

$$p_X(x) = P(X = x)$$

Probability Mass Function (PMF)

By additivity and normalization axioms: $\sum_{x} p_X(x) = 1$

The events {X = x} are disjoint and form a partition of the sample space.

For any set S of possible values of X, $P(X \in S) = \sum_{x \in S} p_X(x)$

Common random variables

1. The Bernoulli random variable
   - A binary r.v. with "success" probability p.
   - Takes values 0 and 1.
   - PMF:

$$\boxed{\begin{aligned} p_X(1) &= p \\ p_X(0) &= 1 - p \end{aligned}}$$

Common random variables

2. The Binomial random variable

- The number of “successes” in n independent Bernoulli trials, each with probability of success p.
- Possible values are 0, …, n
- PMF: $$\boxed{p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}$$
- The normalization property is therefore:

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

Common random variables

3. The Geometric random variable

– The number of trials until the first "success," in repeated independent Bernoulli trials, each with success probability $0 < p < 1$.

– Takes as possible values: all the positive integers.

– PMF: $$\boxed{p_X(k) = P(X = k) = (1-p)^{k-1}p}$$

i.e., we have k - 1 trials that don't yield a success and then we get a success on the k-th trial.

Common random variables

4. The Poisson random variable

- The number of “rare” events (p small) in a large number of independent Bernoulli trials (n large).
- Possible values: all the non-negative integers
- PMF: $\boxed{p_X(k) = e^{-\lambda}\left(\frac{\lambda^k}{k!}\right)}$ where λ is a parameter.

Ex. Consider a book with n words. For each word in the book, the probability it is misspelled is p, (independent of whether any other word is misspelled). Let X be the number of misspelled words in the book.

We could use a Binomial r.v. for X. But when n is very large and p is very small, the Poisson r.v. is a good approximation to the Binomial (their PMFs are similar for large n and small p), and is simpler to compute.

Functions of a random variable

If X is a random variable, then any function of X, Y = g(X), is also a random variable.

The PMF of Y can be computed from the PMF of X as follows:

$$\boxed{p_Y(y) = \sum_{\{x \mid g(x)=y\}} p_X(x)}$$