BBM 205 Discrete Mathematics
Hacettepe University
http://web.cs.hacettepe.edu.tr/∼bbm205

Lecture 9: Introduction to Discrete Probability

Lecturer: Lale Özkahya

Resources:
Kenneth Rosen, "Discrete Mathematics and App."
inf.ed.ac.uk/teaching/courses/dmmr

The “sample space” of a probabilistic experiment

Consider the following probabilistic (random) experiment:

“Flip a fair coin 7 times in a row, and see what happens”

Question: What are the possible outcomes of this experiment?

Answer: The possible outcomes are all the sequences of “Heads” and “Tails”, of length 7. In other words, they are the set of strings $\Omega = \{H, T\}^7$.

The set $\Omega = \{H, T\}^7$ of possible outcomes is called the sample space associated with this probabilistic experiment.

Sample Spaces

For any probabilistic experiment or process, the set Ω of all its possible outcomes is called its **sample space**.

In general, sample spaces need not be finite, and they need not even be countable. In "Discrete Probability", we focus on finite and countable sample spaces. This simplifies the axiomatic treatment needed to do probability theory. We only consider discrete probability (and mainly finite sample spaces).

Question: What is the sample space, Ω, for the following probabilistic experiment:

"Flip a fair coin repeatedly until it comes up heads."

Sample Spaces

For any probabilistic experiment or process, the set Ω of all its possible outcomes is called its **sample space**.

In general, sample spaces need not be finite, and they need not even be countable. In "Discrete Probability", we focus on finite and countable sample spaces. This simplifies the axiomatic treatment needed to do probability theory. We only consider discrete probability (and mainly finite sample spaces).

Question: What is the sample space, Ω, for the following probabilistic experiment:

"Flip a fair coin repeatedly until it comes up heads."

Answer: $\Omega = \{H, TH, TTH, TTTH, TTTTH, \ldots\} = T^*H$.

Note: This set is not finite. So, even for simple random experiments we do have to consider countable sample spaces.

Probability distributions

A probability distribution over a finite or countable set Ω, is a function:

$$P : \Omega \to [0, 1]$$

such that $\sum_{s \in \Omega} P(s) = 1$.

In other words, to each outcome $s \in \Omega$, $P(s)$ assigns a probability, such that $0 \leq P(s) \leq 1$, and of course such that the probabilities of all outcomes sum to 1, so $\sum_{s \in \Omega} P(s) = 1$.

Simple examples of probability distributions

Example 1: Suppose a fair coin is tossed 7 times consecutively. This random experiment defines a probability distribution

Simple examples of probability distributions

Example 1: Suppose a fair coin is tossed 7 times consecutively. This random experiment defines a probability distribution $P : \Omega \to [0, 1]$, on $\Omega = \{H, T\}^7$, where, for all $s \in \Omega$, $P(s) = 1/2^7$. and $|\Omega| = 2^7$, so $\sum_{s \in \Omega} P(s) = 2^7 \cdot (1/2^7) = 1$.

Simple examples of probability distributions

Example 1: Suppose a fair coin is tossed 7 times consecutively. This random experiment defines a probability distribution $P : \Omega \to [0,1]$, on $\Omega = \{H,T\}^7$, where, for all $s \in \Omega$, $P(s) = 1/2^7$. and $|\Omega| = 2^7$, so $\sum_{s\in\Omega} P(s) = 2^7 \cdot (1/2^7) = 1$.

Example 2: Suppose a fair coin is tossed repeatedly until it lands heads. This random experiment defines a probability distribution $P : \Omega \to [0,1]$, on $\Omega = T^*H$,

Simple examples of probability distributions

Example 1: Suppose a fair coin is tossed 7 times consecutively. This random experiment defines a probability distribution $P : \Omega \to [0, 1]$, on $\Omega = \{H, T\}^7$, where, for all $s \in \Omega$, $P(s) = 1/2^7$. and $|\Omega| = 2^7$, so $\sum_{s\in\Omega} P(s) = 2^7 \cdot (1/2^7) = 1$.

Example 2: Suppose a fair coin is tossed repeatedly until it lands heads. This random experiment defines a probability distribution $P : \Omega \to [0, 1]$, on $\Omega = T^*H$, such that, for all $k \geq 0$,

$$P(T^k H) = \frac{1}{2^{k+1}}$$

Note that
$\sum_{s\in\Omega} P(s) = P(H) + P(TH) + P(TTH) + \ldots = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1$.

Events

For a countable sample space Ω, an **event**, E, is simply a subset $E \subseteq \Omega$ of the set of possible outcomes.
Given a probability distribution $P : \Omega \to [0, 1]$, we define the probability of the event $E \subseteq \Omega$ to be $P(E) \doteq \sum_{s \in E} P(s)$.

Example: For $\Omega = \{H, T\}^7$, the following are events:

- "The third coin toss came up heads".

Events

For a countable sample space Ω, an **event**, E, is simply a subset $E \subseteq \Omega$ of the set of possible outcomes.
Given a probability distribution $P : \Omega \rightarrow [0, 1]$, we define the probability of the event $E \subseteq \Omega$ to be $P(E) \doteq \sum_{s \in E} P(s)$.

Example: For $\Omega = \{H, T\}^7$, the following are events:

- "The third coin toss came up heads".
  This is event $E_1 = \{H, T\}^2 H \{H, T\}^4$; $P(E_1) = (1/2)$.

Events

For a countable sample space Ω, an **event**, E, is simply a subset $E \subseteq \Omega$ of the set of possible outcomes.
Given a probability distribution $P : \Omega \to [0, 1]$, we define the probability of the event $E \subseteq \Omega$ to be $P(E) \doteq \sum_{s \in E} P(s)$.

Example: For $\Omega = \{H, T\}^7$, the following are events:

- "The third coin toss came up heads".
  This is event $E_1 = \{H, T\}^2 H \{H, T\}^4$; $P(E_1) = (1/2)$.
- "The fourth and fifth coin tosses did not both come up tails".
  This is $E_2 = \Omega - \{H, T\}^3 TT \{H, T\}^2$; $P(E_2) = 1 - 1/4 = 3/4$.

Events

For a countable sample space Ω, an **event**, E, is simply a subset $E \subseteq \Omega$ of the set of possible outcomes.
Given a probability distribution $P : \Omega \to [0, 1]$, we define the probability of the event $E \subseteq \Omega$ to be $P(E) \doteq \sum_{s \in E} P(s)$.

Example: For $\Omega = \{H, T\}^7$, the following are events:

- "The third coin toss came up heads".
  This is event $E_1 = \{H, T\}^2 H \{H, T\}^4$; $P(E_1) = (1/2)$.
- "The fourth and fifth coin tosses did not both come up tails".
  This is $E_2 = \Omega - \{H, T\}^3 TT \{H, T\}^2$; $P(E_2) = 1 - 1/4 = 3/4$.

Example: For $\Omega = T^* H$, the following is an event:

- "The first time the coin comes up heads is after an even number of coin tosses."

Events

For a countable sample space Ω, an **event**, E, is simply a subset $E \subseteq \Omega$ of the set of possible outcomes.
Given a probability distribution $P : \Omega \to [0, 1]$, we define the probability of the event $E \subseteq \Omega$ to be $P(E) \doteq \sum_{s \in E} P(s)$.

Example: For $\Omega = \{H, T\}^7$, the following are events:

- "The third coin toss came up heads".
  This is event $E_1 = \{H, T\}^2 H \{H, T\}^4$; $P(E_1) = (1/2)$.
- "The fourth and fifth coin tosses did not both come up tails".
  This is $E_2 = \Omega - \{H, T\}^3 TT \{H, T\}^2$; $P(E_2) = 1 - 1/4 = 3/4$.

Example: For $\Omega = T^* H$, the following is an event:

- "The first time the coin comes up heads is after an even number of coin tosses."
  This is $E_3 = \{T^k H \mid k \text{ is odd}\}$; $P(E_3) = \sum_{k=1}^{\infty} (1/2^{2k}) = 1/3$.

Basic facts about probabilities of events

For event $E \subseteq \Omega$, define the complement event to be $\overline{E} \doteq \Omega - E$.

Theorem: Suppose $E_0, E_1, E_2, \ldots$ are a (finite or countable) sequence of pairwise disjoint events from the sample space Ω. In other words, $E_i \in \Omega$, and $E_i \cap E_j = \emptyset$ for all $i, j \in \mathbb{N}$. Then

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

Furthermore, for each event $E \subseteq \Omega$, $P(\overline{E}) = 1 - P(E)$.

Proof: Follows easily from definitions:

Basic facts about probabilities of events

For event $E \subseteq \Omega$, define the complement event to be $\overline{E} \doteq \Omega - E$.

Theorem: Suppose $E_0, E_1, E_2, \ldots$ are a (finite or countable) sequence of pairwise disjoint events from the sample space Ω. In other words, $E_i \in \Omega$, and $E_i \cap E_j = \emptyset$ for all $i, j \in \mathbb{N}$. Then

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

Furthermore, for each event $E \subseteq \Omega$, $P(\overline{E}) = 1 - P(E)$.

Proof: Follows easily from definitions:
for each E_i, $P(E_i) = \sum_{s \in E_i} P(s)$, thus, since the sets E_i are disjoint, $P(\bigcup_i E_i) = \sum_{s \in \bigcup_i E_i} P(s) = \sum_i \sum_{s \in E_i} P(s) = \sum_i P(E_i)$.

Likewise, since $P(\Omega) = \sum_{s \in \Omega} P(s) = 1$, $P(\overline{E}) = P(\Omega - E) = \sum_{s \in \Omega - E} P(s) = \sum_{s \in \Omega} P(s) - \sum_{s \in E} P(s) = 1 - P(E)$. □

Brief comment about non-discrete probability theory

In general (non-discrete) probability theory, with uncountable sample space Ω, the conditions of the prior theorem are actually taken as axioms about a "probability measure", P, that maps events to probabilities, and events are not arbitrary subsets of Ω. Rather, the axioms say: Ω is an event; If $E_0, E_1, \ldots,$ are events, then so is $\bigcup_i E_i$; and If E is an event, then so is $\overline{E} = \Omega - E$.

A set of events $\mathcal{F} \subseteq 2^\Omega$ with these properties is called a σ-algebra. General probability theory studies probability spaces consisting of a triple $(\Omega, \mathcal{F}, P)$, where Ω is a set, $\mathcal{F} \subseteq 2^\Omega$ is a σ-algebra of events over Ω, and $P : \mathcal{F} \rightarrow [0, 1]$ is a probability measure, defined to have the properties in the prior theorem.

We only discuss discrete probabability, and will not assume you know definitions for general (non-discrete) probability.

Conditional probability

Definition: Let $P : \Omega \to [0, 1]$ be a probability distribution, and let $E, F \subseteq \Omega$ be two events, such that $P(F) > 0$.

The **conditional probability** of E given F, denoted $P(E \mid F)$, is defined by:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Example: A fair coin is flipped three times. Suppose we know that the event $F =$ "heads came up exactly once" occurs. what is the probability then of the event $E =$ "the first coin flip came up heads" occurs?

Conditional probability

Definition: Let $P : \Omega \rightarrow [0, 1]$ be a probability distribution, and let $E, F \subseteq \Omega$ be two events, such that $P(F) > 0$.

The **conditional probability** of E given F, denoted $P(E \mid F)$, is defined by:

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Example: A fair coin is flipped three times. Suppose we know that the event $F =$ "heads came up exactly once" occurs. what is the probability then of the event $E =$ "the first coin flip came up heads" occurs?

Answer: There are 8 flip sequences $\{H, T\}^3$, all with probability $1/8$. The event that "heads came up exactly once" is $F = \{HTT, THT, TTH\}$. The event $E \cap F = \{HTT\}$.
So, $P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{1/8}{3/8} = \frac{1}{3}$.

Independence of two events

Intuitively, two events are *independent* if knowing whether one occurred does not alter the probability of the other. Formally:

Definition: Events A and B are called **independent** if $P(A \cap B) = P(A)P(B)$.

Note that if $P(B) > 0$ then A and B are independent if and only if

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

Thus, the probability of A is not altered by knowing B occurs.

Example: A fair coin is flipped three times. Are the events $A =$ "the first coin toss came up heads" and $B =$ "an even number of coin tosses came up head", independent?

Independence of two events

Intuitively, two events are *independent* if knowing whether one occurred does not alter the probability of the other. Formally:

Definition: Events A and B are called **independent** if $P(A \cap B) = P(A)P(B)$.

Note that if $P(B) > 0$ then A and B are independent if and only if

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

Thus, the probability of A is not altered by knowing B occurs.

Example: A fair coin is flipped three times. Are the events $A =$ "the first coin toss came up heads" and $B =$ "an even number of coin tosses came up head", independent?

Answer: Yes. $P(A \cap B) = 1/4$, $P(A) = 1/2$, and $P(B) = 1/2$, so $P(A \cap B) = P(A)P(B)$.

Pairwise and mutual independence

What if we have more than two events: $E_1, E_2, \ldots, E_n$.
When should we consider them "independent"?

Pairwise and mutual independence

What if we have more than two events: $E_1, E_2, \ldots, E_n$.
When should we consider them "independent"?

Definition: Events $E_1, \ldots, E_n$ are called **pairwise independent**, if for every pair $i, j \in \{1, \ldots, n\}$, $i \neq j$, E_i and E_j are independent (i.e., $P(E_i \cap E_j) = P(E_i)P(E_j)$).

Events $E_1, \ldots, E_n$ are called **mutually independent**, if for every subset $J \subseteq \{1, \ldots, n\}$,
$$P(\bigcap_{j \in J} E_j) = \prod_{j \in J} P(E_j).$$

Clearly, mutual independence implies pairwise independent.
But... **Warning:** pairwise independence does not imply mutual independence.
Typically, when we refer to > 2 events as "independent", we mean they are "mutually independent".

Biased coins and Bernoulli trials

In probability theory there are a number of fundamental probability distributions that one should study and understand in detail.
One of these distributions arises from (repeatedly) flipping a biased coin.
A **Bernoulli trial** is a probabilistic experiment that has two outcomes: success or failure (e.g., heads or tails).
We suppose that p is the probability of success, and $q = (1 - p)$ is the probability of failure.

We can of course have repeated Bernoulli trials. We typically assume the different trials are mutually independent.
Question: A biased coin, which comes up heads with probability $p = 2/3$, is flipped 7 times consecutively. What is the probability that it comes up heads exactly 4 times?

The Binomial Distribution

Theorem: The probability of exactly k successes in n (mutually) independent Bernoulli trials, with probability p of success and $q = (1 - p)$ of failure in each trial, is

$$\binom{n}{k} p^k q^{n-k}$$

The Binomial Distribution

Theorem: The probability of exactly k successes in n (mutually) independent Bernoulli trials, with probability p of success and $q = (1 - p)$ of failure in each trial, is

$$\binom{n}{k} p^k q^{n-k}$$

Proof: We can associate n Bernoulli trials with outcomes $\Omega = \{H, T\}^n$. Each sequence $s = (s_1, \ldots, s_n)$ with exactly k heads and $n - k$ tails occurs with probability $p^k q^{n-k}$. There are $\binom{n}{k}$ such sequences with exactly k heads. $\square$

Definition: The **binomial distribution**, with parameters n and p, denoted $b(k; n, p)$, defines a probability distribution on $k \in \{0, \ldots, n\}$, given by

$$b(k; n, p) \doteq \binom{n}{k} \cdot p^k q^{n-k}$$

Random variables

Definition: A **random variable**, is a function $X : \Omega \to \mathbb{R}$, that assigns a real value to each outcome in a sample space Ω.

Example: Suppose a biased coin is flipped n times. The sample space is $\Omega = \{H, T\}^n$. The function $X : \Omega \to \mathbb{N}$ that assigns to each outcome $s \in \Omega$ the number $X(s) \in \mathbb{N}$ of coin tosses that came up heads is one random variable.

For a random variable $X : \Omega \to \mathbb{R}$, we write $P(X = r)$ as shorthand for the probability $P(\{s \in \Omega \mid X(s) = r\})$. The **distribution** of a random variable X is given by the set of pairs $\{(r, P(X = r)) \mid r \text{ is in the range of } X\}$.

Note: These definitions of a random variable and its distribution are only adequate in the context of discrete probability distributions. For general probability theory we need more elaborate definitions.

Biased coins and the Geometric Distribution

Question: Suppose a biased coin, comes up heads with probability p, $0 < p < 1$, each time it is tossed. Suppose we repeatedly flip this coin until it comes up heads.
What is the probability that we flip the coin k times, for $k \geq 1$?

Biased coins and the Geometric Distribution

Question: Suppose a biased coin, comes up heads with probability p, $0 < p < 1$, each time it is tossed. Suppose we repeatedly flip this coin until it comes up heads.
What is the probability that we flip the coin k times, for $k \geq 1$?

Answer: The sample space is $\Omega = \{H, TH, TTH, \ldots\}$.
Assuming mutual independence of coin flips, the probability of $T^{k-1}H$ is $(1-p)^{k-1}p$. Note: this does define a probability distribution on $k \geq 1$, because
$\sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=0}^{\infty}(1-p)^k = p(1/p) = 1$. □

A random variable $X : \Omega \to \mathbb{N}$, is said to have a **geometric distribution** with parameter p, $0 \leq p \leq 1$, if for all positive integers $k \geq 1$, $P(X = k) = (1-p)^{k-1}p$.