Discrete Structures. CSCI-150.

Probability.

Introduction.

In any probabilistic problem, you have to identify the source of randomness and uncertainty. Find, what are the events that can occur, and quantitatively describe them.

First, you need to identify two things:

(1) The set of all possible outcomes, called **sample space**, denoted by Ω (or S in Rosen).

(2) The probability of each individual outcome.

Example. Tossing a coin:

$$\Omega = \{\text{Heads}, \text{Tails}\}$$

$$P(\text{Heads}) = 1/2, \quad P(\text{Tails}) = 1/2.$$

Another example. Rolling a die once:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$P(i) = 1/6, \qquad \text{for } i = 1, \dots, 6$$

Example. What is Ω for tossing a coin twice? Answer: $\Omega = \{H, T\}^2$, the set of all pairs: (H, H), (H, T), (T, H), (T, T).

There are a few requirements. The **probability** of each outcome must be non-negative, and the sum of the probabilities of all outcomes must be equal to 1:

$$\sum_{u \in \Omega} P(u) = 1.$$

An **event** is a subset of Ω.

Consider an event A that happens when you roll a die and get an odd number: $A = \{1, 3, 5\}$. For example, subsets $\varnothing$, $\{1\}$, $\{2\}$, $\{4, 5, 6\}$, $\{1, 2, 3, 4, 5, 6\}$ are valid events too. Can you describe these events in English?

The **probability of an event** is the sum of the probabilities of the outcomes it contains:

$$P(A) = \sum_{u \in A} P(u)$$

$$P(\{1, 3, 5\}) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$$

So, the probability that the result of a die roll is odd is one half.

Four possible events in coin tossing:

$$P(\varnothing) = 0, \quad P(\{H\}) = 1/2, \quad P(\{T\}) = 1/2, \quad P(\{H, T\}) = 1$$

Example. Given a set of letters $L = \{a, b, c, d, e, f, g, h, i, j\}$. A program generates a string of length 3, so that each string is equally likely to be generated. What's the probability that a generated string starts with an 'a'?

It makes sense to choose the sample space to be the set of all generated strings, $\Omega = L^3$. $|\Omega| = |L|^3 = 10^3 = 1000$. What's the probability to generate a string $s \in \Omega$? Because all strings are equally likely, then for any $s \in \Omega$, their probability is the same, $P(s) = p$. Now, how to find this p?

Because $\sum_{s\in\Omega} P(s) = 1$ by the definition of probability, we can find p. Consider the sum

$$\sum_{s\in\Omega} P(s) = \sum_{s\in\Omega} p = |\Omega| \cdot p = 1000p = 1.$$

Thus $p = 0.001$.

We are interested in the probaiblity that the string starts with an 'a'. What does it mean? It means that we want to compute the probability $P(A)$, where the event A is the set of all strings from Ω that start with an 'a'.

$$P(A) = \sum_{s\in A} P(s) = \sum_{s\in A} p = |A| \cdot p = 0.001 \cdot |A|.$$

The question reduces to simply counting the number of strings in A. By the product rule, if the first letter is 'a', there are 2 letters left, so there are $10^2 = 100$ strings of length 3 that start with an 'a', $|A| = 100$, therefore, $P(A) = 0.001 \cdot 100 = 0.1$.

Example. The same problem as in the previous example, but the program generates a string of length 3, 4, or 5. What's the probability that a generated string starts with 'a', 'b', 'c', or 'd'?

$\Omega = L^3 \cup L^4 \cup L^5$. For any $s \in \Omega$: $P(s) = p = 1/|\Omega| = 1/111000$.
$P(A) = (4 \cdot 10^2 + 4 \cdot 10^3 + 4 \cdot 10^4) \cdot p = 0.4$.

Conditional probability.

Conditional probability of A, given B. What is the probability that one event, A, happens, given that some other event, B, definitely happens?

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

Example 1. We roll a die and get x. Given that $x \le 3$, what is the probability that x is odd?

$$\begin{aligned} A:&\quad x \text{ is odd} \\ B:&\quad x \le 3 \end{aligned}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{1,3\})}{P(\{1,2,3\})} = \frac{1/6 + 1/6}{1/6 + 1/6 + 1/6} = 2/3$$

Example 2. A bit string of length four is generated at random so that each of the 16 bit strings of length four is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?

Note that if you cannot simply list all the elements from the sets $A \cap B$ and B because the sets are too large, apply counting techniques.

Example 3.

In many team sports, players' weight and height are usually correlated. Assume a certain team has the distribution shown in the table below.

The height and the weight are given in abstract units, and the numbers in the table tell you how many players in the squad have certain weight and height.

Height						
5					2	1
4	1		2	2	1	
3		1	3	1	1	
2	1	1	1			
1		1			1	
Weight	1	2	3	4	5	6

Assuming we pick one player at random. What's the probability that the player is heavy? What's the probability that the player is tall?

It's natural to assume that Ω is the set of players. Then, define two events:

$$\begin{aligned} Heavy &= \{x \in \Omega \mid \mathit{Weight}(x) \geq 4\}, \\ Tall &= \{x \in \Omega \mid \mathit{Height}(x) \geq 4\}. \end{aligned}$$

By simply counting the numbers in the table, we can establish that $|\Omega| = 20$,

$P(Heavy) = 9/20$ and $P(Tall \cap Heavy) = 6/20$.

Let's find the probability that a player is tall given that he is heavy.

$$P(Tall \mid Heavy) = \frac{P(Tall \cap Heavy)}{P(Heavy)} = \frac{6}{9} = \frac{2}{3} \approx 66\%.$$

On the other hand, if we did not know that the player is heavy, the chance that he is tall is less than one half: $P(Tall) = 9/20 = 45\%$. So, on average the players are actually short. But with the additional knowledge about the player (we knew that he is heavy), we could tell that he is more likely to be tall rather than short.

Other example of conditional probability questions:

- What is the probability that two rolled dice sum to 10, given that both are odd?
- What is the probability that it will rain this afternoon, given that it is cloudy this morning? (Think about how to formally define the problem, what additional information you may need to compute this probability?)
- Very good example in LL: 19.1, p.246: The Halting problem (about MIT EECS hockey team).

Independent events and Bernoulli trials.

A box with four types of coins (or with four compartments).

	Dime	Nickel
US	4	12
Canada	8	24

Take one coin from the box (all choices are equally likely). $P(\text{Dime} \mid \text{US}) = \dfrac{1}{4} = P(\text{Dime})$.

Two events A and B are **independent** if $P(A \mid B) = P(A)$. The occurence of A does not depend on B in any way.

Since $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, A and B are independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Example. Consider a **biased coin**. When tossing it, with probability p, we get 1. Otherwise, with probability $1 - p$, we get 0. For a **fair coin**, $p = 1/2$.

When a coin is flipped, the possible outcomes are heads and tails. Each performance of an experiment with two possible outcomes is called a Bernoulli trial, after James Bernoulli. In general, a possible outcome of a Bernoulli trial is called a success or a failure. Bernoulli trials are mutually independent if the conditional probability of success on any given trial is p, given any information whatsoever about the outcomes of the other trials. We can use 1 to denote a success, and 0 to denote a failure. Therefore, the result of n trials is a bit-string of length n.

Let x_i be the result if the i^{th} trial. A sequence $x_1 \dots x_n$ is a string of 0s and 1s, produced by a sequence of n trials. The trials are independent, so the result of each particular trial does not depend on the other trials. We can write it down in the following way:

$$P(x_i = 1 \mid x_1 \dots x_{i-1} x_{i+1} \dots x_n) = P(x_i = 1) = p$$
$$P(x_i = 0 \mid x_1 \dots x_{i-1} x_{i+1} \dots x_n) = P(x_i = 0) = 1 - p$$

Consider a particular resulting string of 1s and 0s:

$$10011101$$

The probability to generate this string is

$$p(1-p)(1-p)ppp(1-p)p = p^5(1-p)^3$$

For shorter notation, it's common to denote $q = 1 - p$.

In general, when there are k successes and $n - k$ failures in a given sequence, the probability to generate the sequence is

$$p^k(1-p)^{n-k}$$

Consider the sample space of n Bernoulli trials: $\Omega = \{1, 0\}^n$. Each outcome is a bit string of length n, The probability of each particular outcome that has k successes (and $n - k$ failures) is

$$P(x_1 \dots x_n) = p^k(1-p)^{n-k}$$

Observe that this is not the probability of getting k successes in n trials, this is the probability of getting one particular sequence that has k successes. Can we compute the probability of getting any sequence like that?

The probability of exactly k successes in n trials. Binomial distribution.

What is the probability that in a sequnce of n Bernoulli trials with probability of success p you get exactly k successes? (An alternative formulation: after flipping a coin n times, what is the probability to to get heads exactly k times?)

The answer is not $p^k(1-p)^{n-k}$, because there is more than one sequence of n trials that results in k successes. Let A be the set of all such sequences.

$$P(A) = \sum_{(x_1 \ldots x_n) \in A} P(x_1 \ldots x_n) = \sum_{(x_1 \ldots x_n) \in A} p^k(1-p)^{n-k} = |A| \cdot p^k(1-p)^{n-k}$$

(Note that the last equality is correct, because all sequences from A have the same probability $p^k(1-p)^{n-k}$.)

We know that there are $\binom{n}{k}$ ways to select k elements from a set of n elelments. Therefore, the probability is

$$\sum_{u \in A} p^k(1-p)^{n-k} = \binom{n}{k} p^k(1-p)^{n-k}$$

Probability of a complement

If we know $P(A)$, then

$$P(\overline{A}) \equiv P(\Omega \setminus A) = 1 - P(A)$$

Very useful sometimes, e.g. when the event A is defined by a predicate Q, i.e. $A = \{u \in \Omega \mid Q(u)\}$, and then its complement $\overline{A} = \{u \in \Omega \mid \neg Q(u)\}$.

Random variable

What if we are interested not in subsets of Ω, but is some numerical property of the outcomes. E.g., Ω is a set of people in the room, and we are interested in the height of a randomly selected person $x \in \Omega$.

A **random variable** is a function from the sample space Ω to $\mathbb{R}$. That is, a random variable assigns a real number to each possible outcome

$$X : \Omega \to \mathbb{R}$$

Examples:
1. The height of a randomly selected person.

2. The number of successes in n Bernoulli trials.

3. You are playing a game such that on every stage you either win or lose $1. It continues indefinitely, until you run out of money. Let N_0 be the amount of money you have before the game started. Then the r.v. N_i describes the amount of money you have after the i^{th} stage of the game. Another r.v., Z, is the stage number, when you lose the last dollar.

(Remember that all these random variables are just functions, they map every particular sequence of wins and losses to a number.)

Probability distribution

The **probability distribution** (probability mass function) of a random rariable X assigns the probability to each possible value of X:

$$P(X = r)$$

where $r \in X(\Omega)$, i.e. the probability is defined for each r is in the image of the function X – remember that a random variable is nothing more than a real-valued function.

Example:
Let r.v. X be the sum of the numbers that appear when a pair of dice is rolled. What is the probability distribution of X?

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2 \qquad \text{(the set of pairs } (k, m))$$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Consider one particular value $X = 7$. What is the probability that $P(X = 7)$?

Each pair (k, m) is equally likely, with probability $P((k, m)) = (1/6)^2 = 1/36$, since the rolls are independent, and there are six outcomes with $k + m = 7$:

$$P(X = 7) = \frac{6}{36} = \frac{1}{6}$$

Alternatively,
Let's assume that in the first roll we got k. What's the chance that in the second roll we will have enough to get 7 in total?

(Let R_1 and R_2 be the r.v. corresponding to the individual results of the 1st and 2nd roll respectively.)

Notice that regardless of the outcome of the first roll R_1, there is always a chance (with probability 1/6) that in the second roll you get the exactly matching value $R_2 = 7 - R_1$, so

$$P(X = 7) = \ldots = P(R_2 = (7 - k) \mid R_1 = k) \stackrel{indep.}{=} P(R_2 = (7 - k)) = \frac{1}{6}$$

So, in other words, the first roll determines the column in the table. And in the second roll, one of the six possibilities results in having $X = 7$ in total.

More detailed (notation is being rather lengthy)

$$\begin{aligned} P(X = 7) &= \sum_{k=1}^{6} P(\underbrace{R_1 = k,\ R_2 = (7 - k)}_{\{u|R_1(u)=k\}\cap\{u|R_2(u)=7-k\}}) = \sum_{k=1}^{6} P(R_2 = (7 - k) \mid R_1 = k) \cdot P(R_1 = k) \\ &= \sum_{k=1}^{6} P(R_2 = (7 - k) \mid R_1 = k) \cdot \frac{1}{6} = \\ &= P(R_2 = (7 - k) \mid R_1 = k) \\ &= P(R_2 = (7 - k)) = \frac{1}{6} \end{aligned}$$

Notice the notation: for a joint event $A \cap B$, it's common to write $P(A, B)$, which is equivalent to $P(A \cap B)$.

Binomial distribution
The probability to get k successes in a sequence of n independent Bernoulli trials:

$$b(k; n, p) = P(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Expected value

The **expectation** or mean, of the random variable X on the sample space Ω is

$$E(X) = \sum_{u \in \Omega} P(u) \cdot X(u)$$

The expected value of a die roll:

$$\sum_{k=1}^{6} \frac{k}{6} = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \left(\frac{6 \cdot 7}{2}\right) \cdot \frac{1}{6} = \frac{7}{2} = 3.5$$

Alternatively, we can cumpute the expected value by enumerating all values of the random variable, instead of enumerating the individual outcomes:

$$E(X) = \sum_{k \in X(\Omega)} P(X = k) \cdot k$$

Geometric distribution

We flip a biased coin, $P(T) = p$, $P(H) = 1 - p$. (Alternatively, we could say that this is a series of independent Bernoulli trials).

What is the expected number of flips until the coin comes up tails?

$$\begin{aligned}
P(X = 1) &= p \\
P(X = 2) &= (1-p)p \\
P(X = 3) &= (1-p)^2 p \\
&\dots \\
P(X = k) &= (1-p)^{k-1} p
\end{aligned}$$

Thus

$$E(X) = \sum_{k=1}^{\infty} P(X = k) \cdot k = \sum_{k=1}^{\infty} (1-p)^{k-1} p \cdot k$$

Do we know how to compute such a sum? Essentially, we need a formula for

$$\sum_{k=1}^{\infty} q^{k-1} k$$

We know the sum of an infinite geometric series

$$\sum_{k=1}^{\infty} q^k = q \sum_{k=0}^{\infty} q^k = \frac{q}{1-q} \qquad \text{(when } |q| < 1\text{)}$$

Take the derivative with respect to q:

$$\sum_{k=1}^{\infty} q^{k-1} \cdot k = \frac{1(1-q) - (-1)q}{(1-q)^2} = \frac{1 - q + q}{(1-q)^2} = \frac{1}{(1-q)^2} \qquad \text{(when } |q| < 1\text{)}$$

Thus

$$E(X) = \sum_{k=1}^{\infty} (1-p)^{k-1} p \cdot k = \Big(\sum_{k=1}^{\infty} (1-p)^{k-1} \cdot k \Big) p = \frac{1}{(1-(1-p))^2} \cdot p = \frac{1}{p^2} \cdot p = \frac{1}{p}$$

If $p = 1/2$ (a fair coin), then we expect to wait only two flips. If $p = 1/10$, we expect to wait 10 flips, etc.

We say that this random variable X has a **geometric distribution** with parameter p,

$$P(X = k) = (1-p)^{k-1} p$$

The expected value of such r.v. is

$$E(X) = \frac{1}{p}$$

Bayes' Therem

Suppose that A and B are events from a sample space Ω such that $p(A) \neq 0$ and $p(B) \neq 0$.

By the definition of conditional probability, $P(B \cap A) = P(B \mid A)P(A)$, and $P(B \cap A) = P(A \mid B)P(B)$. Therefore, the right-hand sides of these two equations are equal and

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

Since $P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})$, we can get an extended version of the same formula

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

Example

Suppose that one person in 100,000 has a particular rare disease for which there is a fairly accurate diagnostic test. This test is correct 99.0% of the time when given to a person selected at random who has the disease; it is correct 99.5% of the time when given to a person selected at random who does not have the disease. Given this information can we find

(a) the probability that a person who tests positive for the disease has the disease?

(b) the probability that a person who tests negative for the disease does not have the disease?

Should a person who tests positive be very concerned that he or she has the disease?

To solve, choose A to be the event that a person selected at random has the disease, and B to be the event that a person selected at random is tested positively.

Problem 1

Given six cards:

$$A\spadesuit, J\spadesuit, 2\spadesuit, A\heartsuit, 2\heartsuit, 2\diamondsuit,$$

you pick one card at random.

Consider two events:

$$\begin{aligned} A:&\quad \text{the chosen card is an ace} \\ S:&\quad \text{the chosen card is a spade} \end{aligned}$$

(a) What is the sample space Ω?

(b) Compute the probabilities $P(A)$ and $P(S)$.

(c) Are the events A and S independent?

(d) Can you find any (other?) pair of independent events for the given set of cards?

Problem 2.

We randomly select a permutation of $\{1, 2, 3, 4, 5, 6\}$ (all permutations are equally likely). What is the probability that 1 precedes 2?

Problem 3.

What is the conditional probability that exactly four heads appear when a fair coin is flipped five times, given that the first flip came up heads?

Problem 4.

What is the probability that in a series of five independent Bernoulli trials with probability of success p:

(a) all are successes?

(b) all are failures?

(c) only the first is a success?

(d) there are at least 3 successes?

Problem 5 (Birthdays).

What is the minimum number of people who need to be in a room so that the probability that at least two of them have the same birthday is greater than $1/2$?

Assume that there are $n = 366$ days in a year, and all birthdays are independent and equally likely.

Problem 6.

Assume that when you play a game, you win with probability $p = 0.55$, and lose with probability $1 - p = 0.45$. What is the probability that after playing 10 games the number of your victories is the same as the number of your losses?

Problem 7.

Search and read about "non-transitive dice". Roughly speaking, they are a randomized analogue of Rock-Paper-Scissors.

Problem 8.

Find each of the following probabilities when n independent Bernoulli trials are carried out with probability of success p.

(a) the probability of no successes

(b) the probability of at least one success

(c) the probability of at most one success

(d) the probability of at least two successes

(e) the probability of no failures

(f) the probability of at least one failure

(g) the probability of at most one failure

(h) the probability of at least two failures