

# Randomized Algorithms and Probabilistic Techniques

Khalid Hourani

September 17, 2021

## 1 08/24

This course will involve only a few techniques with a variety of applications. Many domains, such as Distributed Computing, Networks, etc., all depend on randomized algorithms.

We will focus on a few techniques that lead to understanding.

1. Union Bound
2. Linearity of Expectation
3. Markov's Inequality
4. Chernoff Bounds

These four things prove very useful in the design and understanding of algorithms.

### 1.1 Basic Definitions

#### Definition 1.1

A *Sample Space* is a set  $S$  whose elements consist of *simple events* (also called *elementary events*). When  $S$  is finite or countably infinite, we say it is a *Discrete* sample space.

#### Definition 1.2

An *event* in a sample space  $S$  is a subset of  $S$ .

#### Definition 1.3

A *Probability Distribution* on  $S$  is a function  $\mathbb{P} : 2^S \rightarrow [0, 1]$  that satisfies

1.  $\mathbb{P}(S) = 1$
2. If  $E_1, E_2, \dots$  are pairwise disjoint events (i.e.,  $E_i \cap E_j = \emptyset$  for all pairs  $i, j$ , also called mutually exclusive), ed by some finite or countably infinite set  $I$ , then

$$\mathbb{P}\left(\bigcup_{i \in I} E_i\right) = \sum_{i \in I} \mathbb{P}(E_i)$$

### 1.2 Conditional Probability

The expression  $\mathbb{P}(E_2 \mid E_1)$  is read “the probability of  $E_2$  given  $E_1$ .” For example, if we randomly select a person from Texas, we might write

$E_1$  = person chosen is in Houston

$E_2$  = person chosen is a UH student

in which case  $\mathbb{P}(E_2 \mid E_1)$  is simply the probability that a randomly selected person from Texas is a UH student *given that they are in Houston*.

#### Definition 1.4 ► Conditional Probability

The conditional probability  $\mathbb{P}(E_2 \mid E_1)$  is defined

$$\mathbb{P}(E_2 \mid E_1) = \frac{\mathbb{P}(E_2 \cap E_1)}{\mathbb{P}(E_1)}$$

Intuitively, this can be thought of as taking the probability that  $E_2$  and  $E_1$  occur and “normalizing it” by dividing by the probability that  $E_1$  occurs.

### 1.3 Independence

#### Definition 1.5 ► Independent Events

Two events,  $E_1$  and  $E_2$ , are *independent* if

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \mathbb{P}(E_2)$$

or, equivalently, if

$$\mathbb{P}(E_1 \mid E_2) = \mathbb{P}(E_1)$$

For example, suppose  $C_1$  and  $C_2$  are the outcomes of two fair coin tosses. These are independent, since

$$\mathbb{P}(C_1 = H \cap C_2 = T) = \mathbb{P}(C_1 = H) \mathbb{P}(C_2 = T) = \frac{1}{4}$$

Independence is not always related to physical independence. For example, say we are given a fair die and let

$$E_1 = \text{roll is even}$$

$$E_2 = \text{roll is less than or equal to 4}$$

In this case, we can enumerate the sample space and explicitly determine  $\mathbb{P}(E_1)$ ,  $\mathbb{P}(E_2)$ ,  $\mathbb{P}(E_1 \cap E_2)$ , and  $\mathbb{P}(E_1) \mathbb{P}(E_2)$ , to see if the events are independent:

$$E_1 = \{2, 4, 6\}$$

$$E_2 = \{1, 2, 3, 4\}$$

$$E_1 \cap E_2 = \{2, 4\}$$

Then

$$\mathbb{P}(E_1) = \frac{3}{6} = \frac{1}{2}$$

$$\mathbb{P}(E_2) = \frac{4}{6} = \frac{2}{3}$$

$$\mathbb{P}(E_1 \cap E_2) = \frac{2}{6} = \frac{1}{3}$$

$$\mathbb{P}(E_1) \mathbb{P}(E_2) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

Thus, we see the events are independent. On the other hand, if  $E_2$  is the event that the roll is *strictly less* than 4, we have

$$E_1 = \{2, 4, 6\}$$

$$E_2 = \{1, 2, 3\}$$

$$E_1 \cap E_2 = \{2\}$$

Then

$$\begin{aligned}\mathbb{P}(E_1) &= \frac{3}{6} = \frac{1}{2} \\ \mathbb{P}(E_2) &= \frac{3}{6} = \frac{1}{2} \\ \mathbb{P}(E_1 \cap E_2) &= \frac{1}{6} \\ \mathbb{P}(E_1) \mathbb{P}(E_2) &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\end{aligned}$$

and we see that the events are *not independent*.

## 1.4 The Inclusion-Exclusion Principle

A basic result in set theory is that

$$\begin{aligned}|A \cup B| &= |A| + |B| - |A \cap B| \\ |A \cup B \cup C| &= |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|\end{aligned}$$

which can be generalized to an arbitrary finite union by

$$\bigcup_{i=1}^n A_i = \sum_{k=1}^n (-1)^{k+1} \left( \sum_{1 \leq i_1 < \dots < i_k \leq n} |A_{i_1} \cap \dots \cap A_{i_k}| \right)$$

or equivalently

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \left| \bigcap_{j \in J} A_j \right|$$

This yields the probability formulas

### Theorem 1.1 ► Inclusion-Exclusion Principle

For any events  $E_1, E_2, \dots, E_n$ ,

$$\begin{aligned}\mathbb{P}(E_1 \cup E_2) &= \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2) \\ \mathbb{P}\left(\bigcup_{i=1}^n E_i\right) &= \sum_{\emptyset \neq J \subseteq \{1, \dots, n\}} (-1)^{|J|+1} \mathbb{P}\left(\bigcap_{j \in J} E_j\right)\end{aligned}$$

## 1.5 Union Bound

### Theorem 1.2 ► Union Bound

Let  $E_1, E_2, \dots$ , be any countable set of events. Then

$$\mathbb{P}\left(\bigcup E_i\right) \leq \sum \mathbb{P}(E_i)$$

While this bound is often not very precise, it is useful in many cases where the events  $E_i$  are “bad” and we can bound the likelihood of a single  $E_i$ . This allows us to bound the likelihood of *any*  $E_i$ .

## 1.6 Conditioning on Multiple Events (Chain Rule)

By repeatedly applying the definition of conditional probability, we have

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \dots \cap E_n) &= \prod_{i=1}^n \mathbb{P}(E_i \mid E_1 \cap E_2 \cap \dots \cap E_{i-1}) \\ &= \mathbb{P}(E_1) \mathbb{P}(E_2 \mid E_1) \dots \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \dots \cap E_{n-1})\end{aligned}$$

## 1.7 Birthday Paradox

### Problem 1.1 ► Birthday Paradox

Suppose  $n$  people are in a room and they have birthdays chosen uniformly at random from the 365 calendar days. What is the probability that two people share a birthday?

The probability that *no two* people share a birthday can be calculated by considering the following events:

$E_1$  = person 1 has a birthday

$E_2$  = person 2 has a different birthday than person 1

$\vdots$

$E_i$  = person  $i$  has a different birthday than people 1 through  $i - 1$

Specifically, the probability that no two people share a birthday is

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \dots \cap E_n) &= \mathbb{P}(E_1) \mathbb{P}(E_2 \mid E_1) \dots \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \dots \cap E_{n-1}) \\ &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \dots \times \left(1 - \frac{365 - n + 1}{365}\right)\end{aligned}$$

For what value of  $n$  is the above probability less than  $1/2$ ?

## 2 08/26

### 2.1 Birthday Paradox

The probability that no two people share a birthday is

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \dots \cap E_n) &= \mathbb{P}(E_1) \mathbb{P}(E_2 \mid E_1) \dots \mathbb{P}(E_n \mid E_1 \cap E_2 \cap \dots \cap E_{n-1}) \\ &= 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \dots \times \left(1 - \frac{365 - n + 1}{365}\right)\end{aligned}$$

Applying the inequality  $1 - x \leq e^{-x}$  when  $|x| < 1$

$$\begin{aligned}&\leq 1 \cdot e^{-\frac{1}{365}} \cdot e^{-\frac{2}{365}} \dots e^{-\frac{n-1}{365}} \\ &= e^{-\frac{n(n-1)}{2 \cdot 365}}\end{aligned}$$

Setting

$$e^{-\frac{n(n-1)}{2 \cdot 365}} \leq \frac{1}{2}$$

yields  $n \geq 23$ .

### 2.2 Randomized Algorithms

#### Problem 2.1 ► Toy Problem

Let  $A$  be an array of  $n$  numbers with the property that at least one number occurs at least  $n/2$  times and the rest are distinct. Our goal is to find the element with duplicate entries.

Any deterministic algorithm will require at least  $n/2 + 1$  operations. However, we can construct a randomized algorithm that only requires  $\mathcal{O}(\log n)$  operations.

We do this by *random sampling*. Assume that choosing a single element in  $\{a_1, a_2, \dots, a_n\}$  requires  $\mathcal{O}(1)$  time and that we sample *with replacement*. We will frequently use sampling with replacement because it allows the samples to be independent, and therefore simplifies the analysis. Now, consider the following algorithm:

---

**Algorithm 1** Randomized Algorithm for Toy Problem

---

```
1: Function RAND-DUPLICATE-TOY(arr):  
2:   loop:  
3:     sample1 ← SAMPLE(arr)  
4:     sample2 ← SAMPLE(arr)  
5:     if sample1 ≠ sample2:  
6:       return sample1
```

---

This algorithm simply chooses two elements (not necessarily distinct) at random from the array and repeats until the two elements are identical, after which it outputs the sampled element as the duplicate.

Let us analyze the probability that this algorithm outputs the duplicate value after a single iteration. Suppose the two values chosen are  $a_i$  and  $a_j$ . Then

$$\begin{aligned}\mathbb{P}(a_i = a_j \text{ and } i \neq j) &= \frac{n/2}{n} \cdot \frac{n/2 - 1}{n} \\ &= \frac{1}{2} \left( \frac{1}{2} - \frac{1}{n} \right) \\ &= \frac{1}{4} - \frac{1}{2n} \\ &\geq \frac{3}{20} \text{ for } n \geq 5\end{aligned}$$

What is the probability, then, that our algorithm fails after  $k$  iterations of Algorithm 1? These events are independent, hence the probability is bounded above by

$$\begin{aligned}\mathbb{P}(\text{failure}) &\leq \left( 1 - \frac{3}{20} \right)^k \\ &\leq e^{-\frac{3}{20}k}\end{aligned}$$

Setting  $k = \frac{20}{3} \ln n$ ,

$$\begin{aligned}&= e^{-\ln n} \\ &= \frac{1}{n}\end{aligned}$$

## 2.3 Baye's Law

### Theorem 2.1 ► Baye's Law

For any events  $A$  and  $B$ ,

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \mathbb{P}(A)}{\mathbb{P}(B)}$$

## 2.4 Concepts in Randomized Algorithms

### Definition 2.1

An algorithm is *Monte Carlo* if it has a non-zero probability of outputting an incorrect solution. If an algorithm will output the correct solution with probability 1, it is *Las Vegas*.

### Definition 2.2 ► High Probability

Given an input of size  $n$ , we say an event occurs *with high probability* (whp) if it occurs with probability  $1 - 1/n^c$  for some  $c > 0$ .

### Problem 2.2 ► Modified Toy Problem

Let  $A$  be an array of  $n$  numbers with the property that *either* all elements are distinct *or*  $n/2$  elements are repeated.

Let us define

$E_1$  = event that exactly  $n/2$  elements are distinct

$E_2$  = event that all  $n$  elements are distinct

At first, our “belief” in  $E_1$  and  $E_2$  are approximately the same, i.e., we have no information to determine which event is more likely. However, let us assume we have sampled two elements with replacement and determined that they are not equal (call this event  $F$ ). We can update our “belief” via Baye’s Law.

$$\begin{aligned}\mathbb{P}(E_1 | F) &= \frac{\mathbb{P}(F | E_1) \mathbb{P}(E_1)}{\mathbb{P}(F)} \\ &\leq \frac{(1 - \frac{3}{20}) \cdot \frac{1}{2}}{(1 - \frac{3}{20}) \cdot \frac{1}{2} + \frac{1}{2}} \\ &= \frac{17}{37}\end{aligned}$$

Thus, we see that the probability has shifted from “50-50” to skewing *in favor* of the belief that  $E_2$  has occurred.

## 2.5 Law of Total Probability

### Theorem 2.2 ► Law of Total Probability

For any event  $A$  and any countable partition of the sample space  $B_1, B_2, \dots$ , that is, the events  $B_i$  are pairwise disjoint and their union forms the entire sample space, we have

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A | B_i) \mathbb{P}(B_i)$$

## 2.6 Naive Bayesian Classifier

A Naive Bayesian Classifier is a classic example of how we can apply probabilistic techniques and Baye’s Law in practice. Suppose we have a *feature set*  $e$ , a vector of boolean  $n$  boolean variable. These could be, for example, whether a certain keyword appears in an email in an attempt to detect spam. Then, assuming the independence of the boolean variables (this is why it is referred to as *naive*), we have

$$\begin{aligned}\mathbb{P}(E | e = (b_1, b_2, \dots, b_n)) &= \frac{\mathbb{P}(e = (b_1, b_2, \dots, b_n) | E) \mathbb{P}(E)}{\mathbb{P}(e = (b_1, b_2, \dots, b_n))} \\ &= \frac{\mathbb{P}(b_1 | E) \cdot \mathbb{P}(b_2 | E) \cdots \mathbb{P}(b_n | E) \cdot \mathbb{P}(event)}{\mathbb{P}(e = (b_1, b_2, \dots, b_n))}\end{aligned}$$