Chapter 7

Probability.

7.1 Probability Spaces.

Definition 7.1.1. A probability space is defined by a tuple (Ω, P), where:

1. Ω is a set, called the sample space. Any element $\omega \in \Omega$ is an atomic event. Conceptually, we think of atomic events as possible outcomes of our experiment. Any subset $A \subset \Omega$ is an event.

2. P, called the probability function, is a function that assigns a number in $[0, 1]$ to any event, denoted as $P : 2^{\Omega} \to [0, 1]$, and satisfies:

   (a) For any event $A \subset \Omega$, $P(A) = \sum_{\omega \in A} P(\omega)$.

   (b) Normalization over the atomic events to 1, which means $\sum_{\omega \in \Omega} P(\omega) = 1$.

Example 7.1.1. Consider a dice rolling, where each of the faces is indexed by $1, 2, 3, 4, 5, 6$ and has an equal chance of being rolled. Therefore, our atomic events are associated with the rolling result, and P is defined as $P(\omega) = \frac{1}{6}$ for any such atomic event. An example of an event can be $A =$ "the dice falls on an even number". The probability of this outcome is:

$$P(A) = \sum_{\omega \in A} P(\omega) = P(\{2\}) + P(\{4\}) + P(\{6\}) = 3 \cdot \frac{1}{6} = \frac{1}{2}$$

Claim 7.1.1. *The probability function satisfies the following properties:*

1. $P(\emptyset) = 0$.

2. *Monotonicity: If* $A \subset B \subset \Omega$, *then* $P(A) \leq P(B)$.

3. *Union Bound:* $P(A \cup B) \leq P(A) + P(B)$.

4. *Additivity for disjoint events: If* $A \cap B = \emptyset$, *then* $P(A \cup B) = P(A) + P(B)$.

5. *Complementarity: Denote by* $\bar{A}$ *the complementary event of* A, *which means* $A \cup \bar{A} = \Omega$. *Then,* $P(\bar{A}) = 1 - P(A)$.

Example 7.1.2. Let's proof the additivity of disjointness property. Let A, B disjointness events, so $A \cap B = \emptyset$ then

$$\begin{aligned} P(A \cup B) &= \sum_{w \in A \cup B} P(w) \\ &= \overbrace{\sum_{w \in A, w \notin B} P(w)}^{P(A)} + \overbrace{\sum_{w \in B, w \notin A} P(w)}^{P(B)} + \overbrace{\sum_{w \in A, w \in B} P(w)}^{0} \\ &= P(A) + P(B) \end{aligned}$$

Definition 7.1.2. Let (Ω, P) be a probability space. A random variable X on (Ω, P) is a function $X : \Omega \to \mathbb{R}$. An indicator, is a random variable defined by an event $A \subset \Omega$ as follows

$$X(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Sometimes, we will use the notation $\{X = x\}$ to denote the event A such:

$$A = \{\omega : X(\omega) = x\} := \{X = x\}$$

Example 7.1.3. Consider rolling a pair of dice. Denote by $X : [6] \times [6] \to [6]$ the random variable that is set to be the result of the first roll. Let Y be defined in almost the same way, but setting the result of the second die. Namely, if we denote by $\{(i, j)\}$ the atomic event associated with sample i on the first die and j on the second die, then:

$$\begin{aligned} X(\{i, j\}) &= i \\ Y(\{i, j\}) &= j \end{aligned}$$

In addition, one can define the random variable z as the sum, $Z = X + Y$. Since the sum is also a function from Ω to $\mathbb{R}$, Z is also a random variable. An example of an indicator could be W, which gets 1 if $Z \in \{2, 7, 8\}$.

Example 7.1.4. Let X be an indicator of event A. Then $1 - X$ is the indicator of $\bar{A}$.

$$1 - X(\omega) = \begin{cases} 0 & \omega \in A \Leftrightarrow \omega \notin \bar{A} \\ 1 & \omega \notin A \Leftrightarrow \omega \in \bar{A} \end{cases}$$

Definition 7.1.3. We will say that two events A, B are independent if:

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

Similarly we will say that random variables $X, Y : \Omega \to \mathbb{R}$ are independent if for any $x \in \text{Im } X$ and $y \in \text{Im } Y$:

$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$$

Example 7.1.5. X, Y defined in Example 7.1.3 are independent.

$$\begin{aligned} P(\{X = i\} \cap \{Y = j\}) &= \sum_{i' = i \text{ and } j' = j} P(\{(i', j')\}) = P(\{(i, j)\}) \\ &= \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = P(X = i) P(Y = j) \end{aligned}$$

Example 7.1.6. Let A and B be independent events. Then, $\bar{A}$ and B are also independent events, since:

$$\begin{aligned} P(B) &= P(B \cap \Omega) = P(B \cap (A \cup \bar{A})) = P((B \cap A) \cup (B \cap \bar{A})) \\ &= P(B \cap A) + P(B \cap \bar{A}) = P(B)P(A) + P(B \cap \bar{A}) \\ \Rightarrow P(B \cap \bar{A}) &= P(B)(1 - P(A)) = P(B)P(\bar{A}) \end{aligned}$$

Example 7.1.7. Let X and Y be indicators of independent events A and B. Then $P(X \cdot Y = 1) = P(X = 1) \cdot P(Y = 1)$. The proof is left as an exercise.

7.2 Throwing Keys to Cells.

Example 7.2.1. Imagine that following experiment, we have m cells and n keys (balls, numbers, or your favorite object type). We throw each of the keys independently into the cells. The cells are identical, so the probability of hitting any of them is the same, $1/m$. We would like to analyze how the capacity of the cells is distributed.

1. What is the probability that the first and the second keys will be thrown to the first cell? What is the probability that the first and the second keys will be thrown to the same cell?
2. What is the probability that in the first cell there is exactly one key?

Let us define the indicator X_i^j which indicate that the jth key fallen into the ith cell.

1. So first we been asked whether $X_1^1 \cdot X_1^2 = 1$, Since this happens only if both $X_1^1 = 1$, $X_1^2 = 1$ then by independently we have that:

$$\begin{aligned} P(X_1^1 \cdot X_1^2 = 1) &= P(X_1^1 = 1 \cap X_1^2 = 1) \\ &= P(X_1^1 = 1) \cdot P(X_1^2 = 1) = \frac{1}{m^2} \end{aligned}$$

   Now, to answer if the first and second keys fall into the same cell, we need to check if there exists an i such that $X_i^1 \cdot X_i^2 = 1$. Observes that for any different i and i', the X_i^j and $X_{i'}^j$ are indicators of disjoint events. This is because j cannot be in both the i and i' cells. Therefore, $X_i^1 \cdot X_i^2$ and $X_{i'}^1 \cdot X_{i'}^2$ are also indicators of disjoint events. Thus:

$$\begin{aligned} P(\exists i : X_i^1 \cdot X_i^2 = 1) &= P(\bigcup_i X_i^1 \cdot X_i^2 = 1) \\ &= \sum_i P(X_i^1 \cdot X_i^2 = 1) = m \cdot \frac{1}{m^2} = \frac{1}{m} \end{aligned}$$

   We are basically done. However, we want to present the same calculation in a different notation that will be useful for computing expectations later on. Note that the random variable that counts "how many" cells both the first and the second fall into is $\sum_i X_i^1 \cdot X_i^2$. In other words, the sum can be either 0 if the keys fall into different cells, or 1 if they both fall into the same cell.

2. The event that only the jth key falls into the first cell matches to

$$\left\{ X_1^j \prod_{j \neq j'} \left(1 - X_1^{j'}\right) = 1 \right\}$$

Therefore, due to the disjointness of $1 - X_1^{j'}$ and $X_1^{j'}$, the indicator for the first cell containing exactly one key is:

$$\left\{ \sum_j X_1^j \prod_{j \neq j'} \left(1 - X_1^{j'}\right) = 1 \right\}$$

Since the terms in the sum are disjoint and the products are products of independent indicators, we have:

$$P\left(\sum_j X_1^j \prod_{j \neq j'} \left(1 - X_1^{j'}\right) = 1 \right) = \sum_j P\left(X_1^j \prod_{j \neq j'} \left(1 - X_1^{j'}\right) = 1 \right)$$
$$= m \cdot \frac{1}{m} \left(1 - \frac{1}{m}\right)^{n-1} = \left(1 - \frac{1}{m}\right)^{n-1}$$

Definition 7.2.1. Let $X : \Omega \to \mathbb{R}$ be a random variable, the expectation of X is

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{x \in \operatorname{Im} X} x P(X = x)$$

Observes that if P is distributed uniformly, then the expectation of X is just the arithmetic mean:

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega)$$

Claim 7.2.1. *The expectation satisfies the following properties:*

1. *Monotonic, If $X \leq Y$ (for any $\omega \in \Omega$) then $\mathbf{E}[X] \leq \mathbf{E}[Y]$.*
2. *Linearity, for $a, b \in \mathbb{R}$ it holds that $\mathbf{E}[aX + by] = a\mathbf{E}[X] + b\mathbf{E}[Y]$.*
3. *Independently, if X, Y are independent, then $\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$.*
4. *For any constant $a \in \mathbb{R}$ we have that $\mathbf{E}[a] = a$.*

Proof. 1. Monotonic, if $X \leq Y$ then :

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega) \leq \sum_{\omega \in \Omega} Y(\omega) P(\omega) = \mathbf{E}[Y]$$

2. Linearity,

$$\mathbf{E}[aX + bY] = \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega)) P(\omega)$$
$$= a \sum_{\omega \in \Omega} X(\omega) P(\omega) + b \sum_{\omega \in \Omega} Y(\omega) P(\omega)$$

3. Independently,

$$\begin{aligned}
\mathbf{E}[XY] &= \sum_{x,y\in \operatorname{Im} X\times \operatorname{Im} Y} xyP(X=x\cap Y=y)\\
&= \sum_{x,y\in \operatorname{Im} X\times \operatorname{Im} Y} xyP(X=x)P(Y=y)\\
&= \sum_{x\in \operatorname{Im} X}\sum_{y\in \operatorname{Im} Y} xyP(X=x)P(Y=y)\\
&= \sum_{x\in \operatorname{Im} X} xP(X=x)\sum_{y\in \operatorname{Im} Y} yP(Y=y)\\
&= \sum_{x\in \operatorname{Im} X} xP(X=x)\mathbf{E}[Y]\\
&= \mathbf{E}[X]\,\mathbf{E}[Y]
\end{aligned}$$

4. Let X be the random variable which is also the constant function $X(\omega)=a$ for any $\omega\in\Omega$. Then we have that

$$\begin{aligned}
\mathbf{E}[X] &= \sum_{\omega\in\Omega} X(\omega)P(\omega)\\
&= \sum_{\omega\in\Omega} aP(\omega) = a\cdot 1 = a
\end{aligned}$$

□

Example 7.2.2. Let X be an indicator of event A, what are $\mathbf{E}[X]$ and $\mathbf{E}\left[X^2\right]$? Recall that $X(\omega)=1$ only if $\omega\in A$ and 0 otherwise, thus:

$$X^k(\omega)=\begin{cases}1^k=1 & \omega\in A\\ 0^k=0 & \text{else}\end{cases}\Rightarrow X^k(\omega)=X(\omega)$$

Therefore,

$$\mathbf{E}\left[X^k\right]=\sum_{\omega\in\Omega}X^k(\omega)P(\omega)=\sum_{\omega\in\Omega}X(\omega)P(\omega)=\mathbf{E}[X]$$

Example 7.2.3. Consider the experiment of throwing keys into cells again. What is the expected number of keys that fell into the same cell as the first key? The indicator of the event j and 1 falling into the same cell is given by $\sum_i X_i^1X_i^j$ and it remains to sum over all the j's. So:

$$\begin{aligned}
\mathbf{E}\left[\sum_i\sum_j X_i^1X_i^j\right] &= \sum_i\sum_j \mathbf{E}\left[X_i^1X_i^j\right]\\
&= \sum_i\sum_{j\neq 1}\mathbf{E}\left[X_i^1\right]\mathbf{E}\left[X_i^j\right]+\overbrace{\sum_i \mathbf{E}\left[X_i^1\right]}^{j=1}\\
&= m\cdot(n-1)\frac{1}{m^2}+m\cdot\frac{1}{m}=\frac{n-1}{m}+1
\end{aligned}$$

Note 1: Despite the ease of computing the expectation, calculating the exact probability of $\sum_i\sum_j X_i^1X_i^j=L$ for some arbitrary L is a difficult task.

7.3 Running Time as a Random Variable.

Randomness might appears in algoritmic context in two main cases, in the first the algorithm might behave randomly, means that it flips coins and decided what to do conditioning on the outcomes. In the second case, the input might distributed according to probability function. In both cases the result and running time of the algorithm are random variable. And it's interesting to ask what is the expected running time.

Let us introduce an example for the first case. We are given an array $A_1, A_2, ..., A_n$ and are asked to find the k smallest elements in it. Here, we are going to suggest a random algorithm that is expected to return in linear time, even if we do not make any assumptions about the input, particularly how it is distributed.

Result: returns the k smallest element in $A_1...A_n \in \mathbb{R}^n$
1 **if** *left* = *right* **then**
2 | **return** A_{left}
3 **end**
4 pivot ← select random pivot in [left, right]
5 pivot ← partition(A, left, right, pivot)
6 **if** k = *pivot* **then**
7 | **return** A_k
8 **end**
9 **else if** $k <$ *pivot* **then**
10 | right ← pivot - 1
11 **end**
12 **else**
13 | left ← pivot + 1
14 | $k \leftarrow k-$ pivot
15 **end**
16 **return** call select(A, left, right, k)

Algorithm 1: select(A, left, right, k)

Consider the first call to 'select' and let X_m be the indicator for selecting the index of the mth smallest number on line (4). Notice that X_m and the running time of the recursive calls are independent random variables. Additionally, we will assume in induction that $T(n', k) \leq 2cn'$ for any $n' < n$. Therefore, the

expected running time is:

$$\begin{aligned}
T(n,k) &= c\cdot n + \sum_{m<k} X_m \cdot T(n-m, k-m) \\
&\quad + X_k \cdot 1 + \sum_{m>k} X_m \cdot T(m-1,k) \\
\Rightarrow \mathbf{E}\left[T(n,k)\right] &\leq cn + 2c + \sum_{m<k} \mathbf{E}\left[X_m \cdot T(n-m,k-m)\right] \\
&\quad + \sum_{m>k} \mathbf{E}\left[X_m \cdot T(m-1,k)\right] \\
&\leq c\cdot n + 2c + 2c\sum_{m<k} \frac{n-m}{n} + 2c\sum_{m>k}\frac{m-1}{n} \\
&\leq c\cdot n + 2c + 2c\sum_{m<k} \frac{n-m}{n} + 2c\sum_{m>k}\frac{m}{n}
\end{aligned}$$

Claim 7.3.1. *The sum above is maximal when $k = \lfloor n/2 \rfloor$.*

Proof. We will prove that for $k = i+1$, the sum is greater than $k = i$ if $i < \lfloor n/2 \rfloor$. Denote $S_i = \sum_{m<i} \frac{n-m}{n} + \sum_{m>i} \frac{m}{n}$. Then, the substitution of $S_{i+1} - S_i$ becomes:

$$S_{i+1} - S_i = \frac{n-i-1}{n} - \frac{i}{n} = \frac{n-2i-1}{n}$$

And that quantity is positive for any $i < \lfloor n/2 \rfloor$. By symmetry, we obtain that for any $i > \lceil n/2 \rceil + 1$, the quantity $S_i - S_{i+1}$ is positive. Hence, $\lfloor n/2 \rfloor$ is a global maximum. □

Therefore, the expectation is bounded by:

$$\begin{aligned}
&\leq c\cdot n + 2c + 2c \sum_{m<\lfloor n/2 \rfloor} \frac{n-m}{n} + 2c \sum_{m>\lfloor n/2 \rfloor} \frac{m}{n} \\
&= c\cdot n + 2c + 2\cdot 2c \sum_{m>\lfloor n/2 \rfloor} \frac{m}{n} \\
&= c\cdot n + 2c + 2\cdot 2c \cdot \frac{(n/2)\cdot(n/2-1)}{2n} \\
&\leq cn + 2c + n/2 \cdot 2c \leq 2c\cdot n
\end{aligned}$$