Chapter 7

Probability.

7.1 Probability Spaces.

Definition 7.1.1. *A probability space defined by a tuple* (Ω, P) *such that:*

1. Ω *is a set, called the sample space. Any element* $\omega \in \Omega$ *is named 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* $A \subset \Omega$, $P(A) = \sum_{\omega \in A} P(w)$.

   (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, each of the faces, indexed by* $1, 2, 3, 4, 5, 6$ *has the same chances to rolled out. Thus, our atomic events associated with the rolling result, and* P *defined to be* $P(\omega) = \frac{1}{6}$ *for any such atomic event.*

Claim 7.1.1. *Probability function satisfies the following properties:*

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

2. *Monotonic, 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 disjointness events. If* $A \cap B = \emptyset$ *then* $P(A \cup B) = P(A) + P(B)$.

5. *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. *[COMMENT] Add dice roll, as an example.*

Definition 7.1.3. *We will say that two random variable $X, Y : \Omega \to \mathbb{R}$ are independent if for any $x \in \operatorname{Im} X$ and $y \in \operatorname{Im} Y$:*

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

7.2 [COMMENT] Throw Keys to Cells.

[COMMENT] Add the description of throwing keys to cells. Define the random variable X_i^j.

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 $X \le Y$ then :

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

2. Linearity,

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

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 7.2.1. *[COMMENT] Expectation of indicators and their multiplication. 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:*

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

Therefore,

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

let B[0 : n − 1] be a new array
for $i \leftarrow [0, n{-}1]$ **do**
  make B_i an empty list
end
for $i \leftarrow [1, n]$ **do**
  insert A_i into list $B_{\lfloor nA_i \rfloor}$
end
for $i \leftarrow [0, n{-}1]$ **do**
  sort list B_i
end
concatenate the lists $B_0, B_1, .., B_{n-1}$ together and
return the concatenated lists

Algorithm 1: bucket-sort(A, n)

Example 7.2.2. *[COMMENT] How many keys trowed into the same cell as the first key thrown to?*

Denote by $X_i : [n] \to [n]$ then random variable that counts the number of elements fallen in the ith bucket. The Expectation of the sorting running time is:

$$\begin{aligned}
\mathbf{E}\left[T\right] &= \mathbf{E}\left[\text{Inserting into buckets } + \sum_i \text{Sorting } i\text{th bucket}\right] \\
&= \mathbf{E}\left[\Theta(n) + \sum_i X_i^2\right] = \Theta(n) + \sum_i \mathbf{E}\left[X_i^2\right] \\
\mathbf{E}\left[X_i^2\right] &= \mathbf{E}\left[\left(\sum X_i^j\right)^2\right] = \mathbf{E}\left[\sum_{j,j'} X_i^j X_i^{j'}\right] = \sum_{j,j'} \mathbf{E}\left[X_i^j X_i^{j'}\right] \\
&= \sum_{j \neq j'} \mathbf{E}\left[X_i^j X_i^{j'}\right] + \sum_j \mathbf{E}\left[X_i^j X_i^j\right] \\
&= \sum_{j \neq j'} \mathbf{E}\left[X_i^j X_i^{j'}\right] + \sum_j \mathbf{E}\left[X_i^j\right] \\
&= 2\binom{n}{2}\left(\frac{1}{n}\right)^2 + n \cdot \frac{1}{n} \\
&= \frac{n-1}{n} + 1 = 2 - \frac{1}{n} \Rightarrow \mathbf{E}\left[T\right] = \Theta(n) + n\left(2 - \frac{1}{n}\right) = \Theta(n)
\end{aligned}$$