1 Probability Spaces

1.1 Sample Spaces

Definition 1. *A sample space $\mathcal{S}$ is the set of all possible outcomes of an experiment. An event in a sample space $\mathcal{S}$ is a subset $E \subseteq \mathcal{S}$.*

Definition 2. *For an event E in $\mathcal{S}$, define $E^{c} = S \setminus E$ to be the complement of E, i.e, the event that E does not happen.*
    For two events E, F in $\mathcal{S}$,

- *$E \cap F$ is the event that both E and F happen,*
- *$EF = E \cup F$ is the event that E or F happens,*
- *E and F are said to be mutually exclusive if $E \cap F = \emptyset$.*

1.2 Discrete Probability Spaces

Definition 3. *A discrete probability space is defined by a discrete sample space[1] $\mathcal{S}$, and a probability mass function (pmf) $p : \mathcal{S} \to [0, 1]$ such that $\sum_{x \in \mathcal{S}} p(x) = 1$.*
    The probability function *for the probability space is a function P such that*

$$P(E) = \sum_{x \in E} p(x), \text{for every event } E \subseteq \mathcal{S}.$$

Example 1. Consider a fair even toss. Then the sample space is $\mathcal{S} = \{H,'T\}$ where H indicates "head-up" and T indicates "tail-up". The pmf is $p(H) = p(T) = \frac{1}{2}$. The probability function P has $P(\emptyset) = 0, P(\{H\}) = P(\{T\}) = \frac{1}{2}$ and $P(\{H, T\}) = 1$.

Example 2. Consider the experiment where we toss a fair coin repeatedly until we see a head-up. The outcome of the experiment is the number of coin tosses. Then $\mathcal{S} = \mathbb{Z}_{>0} = \{1, 2, 3, ...\}$ and the pmf p satisfies $p(i) = \frac{1}{2^i}$ for every integer $i \geq 1$. $P(\# \text{ coin tosses} \leq 3) = P(\{1, 2, 3\}) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$.

1.3 Continuous Probability Spaces

An example of a continuous probability space is the uniform probability space over the real numbers $[0, 1]$. In this continuous probability space, a pmf becomes meaningless since the probability that the outcome is any real number is 0. In this case, a probability density function is needed to define the probability space.

    A general probability space could be discrete, continuous in some d-dimensional space, or a mixture of discrete and continuous spaces. Thus, to define a general probability space, we need to define the probability function P directly. This P is over a subset Σ of "interesting" events in the samples space $\mathcal{S}$. Σ may not be the family of all events since there are some subsets of $\mathcal{S}$ that are not measurable. So, a general probability space is defined by a tuple $(\mathcal{S}, \Sigma, P)$ that satisfies some conditions, which we omit here. For this course, most of the probability spaces we shall use are discrete; occasionally we shall use the uniform distribution over an real internal and Gaussian distributions. Many lemmas and theorems hold for general probability spaces; however, in some proofs we may only focus on the discrete probability space case for notational convenience. Also, by default, we use $\mathcal{S}, p$ and P to denote the sample space, probability mass function and probability

---

[1]Recall that a set is discrete if it contains finite or countably infinite number of elements.

function of the probability space considered. We also use a *distribution* to refer to a probability space.

1.4 Some Invariants

Let E, F and G be events in a (discrete) probability space. Then,

- $P(E) + P(E^c) = 1$.
- $P(E \cup F) = P(E) + P(F) - P(EF)$.
- $P(E \cup F \cup G) = P(E) + P(F)P(G) - P(EF) - P(EG) - P(FG) + P(EG)$.

So if two events E and F are mutually exclusive then $P(EF) = P(\emptyset) = 0$. In this case, we have $P(E \cup F) = P(E) + P(F)$.

The third equation can be extended to n events as follows.

Lemma 4. *Let $E_1, E_2, E_3, \cdots E_n$ be n events. Then*

$$P(E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n) = \sum_{S \subseteq [n]: S \neq \emptyset} (-1)^{|S|-1} P\left(\bigcap_{i \in S} E_i\right).$$

2 Conditional Probability Spaces and Conditional Probabilities

Definition 5. *Let E, F be events in probability space and assume $P(F) > 0$. The probability of E conditioned on F, denoted as $P(E|F)$, is defined as $P(E|F) = \frac{P(E \cap F)}{P(F)}$.*

Example 3. Consider 2 fair coin tosses in a row. Then, $\mathcal{S} = HH, HT, TH, TT$ and $p(HH) = p(HT) = p(TH) = p(TT) = \frac{1}{4}$. Given that the first result is H, what is the probability that both results are H? Here is the solution:

$$P(\text{both results are } H|\text{first result is } H) = P(\{HH\}|\{HH, HT\})$$
$$= \frac{P(\{HH\} \cap \{HH, HT\})}{P(\{HH, HT\})} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Example 4. Consider the same experiment as above. Given that at least there is one H in 2 coin tosses , what is the probability that both tosses are H?

$$P(\text{both results are } H|\text{at least one } H) = P(\{HH\}|\{HH, HT, TH\})$$
$$= \frac{P(\{HH\} \cap \{HH, HT, TH\})}{P(\{HH, HT\})} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Example 5. Consider the experiment of throwing a fair dice. The value is big if it is $4, 5$ or 6. Then,

$$P(\text{value is even}|\text{value is big}) = P(\{2, 4, 6\}|\{4, 5, 6\})$$
$$= \frac{P(\{2, 4, 6\} \cap \{4, 5, 6\})}{P(\{4, 5, 6\})} = \frac{P(\{4, 6\})}{P(\{4, 5, 6\})} = \frac{2/6}{3/6} = \frac{2}{3}.$$

2.1 Independence of Events

Definition 6. *Two events E and F are said to be independent if the probability that one event occurs in no way affects the probability of the other event occurring. Precisely, E and F are independent if $P(E|F) = P(E)$, or equivalently $P(EF) = P(E)P(F)$.*

Example 6. Toss a fair coin twice. Let E be the event that the first toss is H and F be the event where the second toss is H. Then $P(E) = \frac{1}{2}, P(F) = \frac{1}{2}$ and $P(EF) = \frac{1}{4} = P(E)P(F)$. Thus the events E and F are independent.

Example 7. Example: We toss a fair coin repeatedly until we see a head up and let X be the number of coin tosses we did. Let E be the event that $X \leq 2$ and F be the event X is even. Then,

- $P(E) = p(1) + p(2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.
- $P(F) = p(2) + p(4) + p(6) + p(8) + \cdots = \frac{1}{4} + \frac{1}{16} + \frac{1}{32} + \frac{1}{128} + \cdots = \frac{1/4}{1-1/4} = \frac{1}{3}$.
- $P(EF) = p(2) = \frac{1}{4} = P(E)P(F)$.

Thus, the two events E and F are independent.

Notice that the difference between E and F being mutually exclusive and that E and F being independent. The former happens iff $EF = \emptyset$, which implies $P(EF) = 0$ and the second happens iff $P(EF) = P(E)P(F)$. So, $P(E) > 0$ and $P(F) > 0$, then E and F can not be both mutually exclusive and independent. Also notice the definition of mutual exclusiveness only requires the sample space, while that of independence requires both the sample space and the probability function.

Observation 7. *If E and F are independent, then E^{c} and F are independent.*

2.2 Bayes Formula

Lemma 8 (Bayes Formula)**.** *Let E and F be two events with $P(F) > 0$, then*

$$P(E) = P(F)P(E|F) + (1 - P(F))P(E|F^{\mathrm{c}}).$$

Proof. $P(E) = P(E \cap F) + P(E \cap F^{\mathrm{c}}) = P(F)P(E|F) + P(F^{\mathrm{c}})P(E|F^{\mathrm{c}})$. The lemma follows by observing that $P(F^{\mathrm{c}}) = 1 - P(F)$. $\square$

Example 8. Suppose we have two different types of weather: sunny and rainy. The probability each weather occurs tomorrow and the probability that some bus will be delayed given the weather are as follows:

- $P(\text{sunny}) = 0.7, P(\text{rainy}) = 0.3, P(\text{delayed|sunny}) = 0.1, P(\text{delayed|rainy}) = 0.3$.

Then we have $P(delayed) = 0.7 \times 0.1 + 0.3 \times 0.3 = 0.16$.

3 Expectation of Random Variables

3.1 Functions of Random Variables

We have defined the probability space. A random variable in the probability space is then a variable indicating the outcome of a random experiment for the probability space. That is, the random variable take values in the sample space. However, often we need to consider functions of random variables. A function of a random variable can be also viewed as a random variable. For example, consider the experiment of throwing a fair dice. The sample space is $\{1, 2, 3, 4, 5, 6\}$ and pmf is $p(1) = p(2) = \cdots = p(6) = \frac{1}{6}$. Let X be the random variable denoting the value obtained from the dice throwing. Let Y be a function of X such that $Y = small$ if $X \in \{1, 2, 3\}$ and $Y = big$ if $X \in \{4, 5, 6\}$. Then, for convenience, we also say that Y is a random variable from the probability space.

3.2 Expectations

Definition 9. *Given a real-valued random variable X from some discrete probability space, the expectation of X, or the expected value of X, denoted as $\mathbb{E}[X]$, is defined as*

$$\mathbb{E}[X] = \sum_x P(X = x)x.$$

Example 9. Toss a fair coin and let $X = 0$ if the result is tail-up and $X = 1$ if it is head-up. Then we have $\mathbb{E}[X] = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}$. This suggests that the expectation of a variable may be some value that the variable can never take.

Example 10. We throw a fair dice. Let $Y = -100$ if the value is in $\{1, 2, 3\}$ and 100 if it is in $\{4, 5, 6\}$. Then $\mathbb{E}[Y] = \frac{1}{2} \times (-100) + \frac{1}{2} \times 100 = 0$.

In the above example, we used a function as a random variable. The following lemma holds when computing the expectation of a function of a random variable:

Lemma 10. *Let Y be a real-value function of a random variable X. Then*

$$\mathbb{E}[Y] = \sum_x p(X = x) Y(x).$$

Example 11. Consider a biased coin toss which takes head-up with probability $p \in (0, 1)$ and tail-up with probability $1 - p$. We repeatedly toss the coin until we see a head-up. Let X be the number of coin-tosses we have. Then we have

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{\infty} P(X = i) i = \sum_{i=1}^{\infty} p(1-p)^{i-1} i \\
&= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} p(1-p)^{i-1} = \sum_{j=1}^{\infty} \frac{p(1-p)^{j-1}}{1 - (1-p)} = \sum_{j=1}^{\infty} (1-p)^{j-1} \\
&= \frac{1}{1 - (1-p)} = \frac{1}{p}.
\end{aligned}
$$

As we shall define later, the above distribution is called the geometric distribution with parameter p.

3.3 Linearity of Expectation

Lemma 11 (Linearity of Expectation). *Let $X_1, X_2, \cdots, X_n$ be n random variables. Then,*

$$\mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i].$$

Proof. Let $\mathcal{S}$ and p be the sample space and pmf for the common probability space for all the n random variables. Then, each X_i is a function over elements $x \in \mathcal{S}$.

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{x \in \mathcal{S}} p(x) \sum_{i=1}^{n} X_i(x) = \sum_{i=1}^{n} \sum_{x \in \mathcal{S}} p(x) X_i(x) \sum_{i=1}^{n} \mathbb{E}[X_i]. \qquad \square$$

In particular, for two random variables X and Y we have $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Example 12. Suppose we throw a fair dice to get an integer value between 1 and 6. Let $X = 0$ if the value is in $\{1, 2, 3\}$ and 1 if it is in $\{4, 5, 6\}$. Let Y be the value modular 2. Then, we have

dice value	1	2	3	4	5	6
X	0	0	0	1	1	1
Y	1	0	1	0	1	0
$X + Y$	1	0	1	1	2	1

Then, we have $\mathbb{E}[X] = \frac{1}{2}, \mathbb{E}[Y] = \frac{1}{2}$ and $\mathbb{E}[X + Y] = 1 = \mathbb{E}[X] + \mathbb{E}[Y]$.

3.4 Dependence of Random Variables

Definition 12. *Assume X and Y are 2 random variables from a discrete probability space. Then X and Y are said to be independent if for every x and y, we have*

$$P(X = x, Y = y) = P(X = x) P(Y = y).$$

Example 13. Assume a dice can only land on 1,2,3 or 4 . Assume that X and Y are random variables as follows:

$$X = \begin{cases} 0 & \text{value} \in \{1,2\} \\ 1 & \text{value} \in \{3,4\} \end{cases}, \qquad Y = \begin{cases} 0 & \text{value} \in \{2,4\} \\ 1 & \text{value} \in \{1,3\}. \end{cases}$$

Then X and Y are independent, since for every $x, y \in \{0,1\}$, we have $P(X = x) = \frac{1}{2}, P(Y = y) = \frac{1}{2}$ and $P(X = x, Y = y) = \frac{1}{4}$.

In general, for two real-value variables X and Y, we may not have $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. Consider Example 12. We have $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{1}{2}$ but $\mathbb{E}[XY] = \frac{1}{6}$ since $XY = 1$ if and only if the dice value is 5. However, we the equation holds if X and Y are independent:

Lemma 13. *If X and Y are independent real value variables, we have*

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$