1   Chapter 1: Simulation of Discrete Probabilities (Continued)

We will continue our discussion of this chapter.

1.1   Events

Often, in probability, we are interested in the probability that a certain event will occur. For example, we might be interested in whether or not it will rain today, or whether or not we will win the lottery, or so on.

Mathematically, an **event** is a subset of the sample space $E \subset \Omega$.

(Example.) If we roll a die, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. The event,

$$E = \text{``Roll an even number''},$$

is the subset $E = \{2, 4, 6\}$.

(Example.) If we roll a *pair* of dice, then,

$$\Omega = \{(a, b) \mid a, b \in [1, 6] \subset \mathbb{Z}\},$$

and the event that we roll a pair is given by

$$E = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

Note that we do not care about the order, so 12 and 21 are the same outcome.

Recall that each outcome $\omega \in \Omega$ is given a probability mass $m(\omega) \geq 0$ in such a way that

$$\sum_{\omega \in \Omega} m(\omega) = 1.$$

The probability of an event E occurring is given by

$$P(E) = \sum_{\omega \in \Omega} m(\omega).$$

Likewise, for a random variable X, we have

$$P(X = x) = \sum_{\omega : X(\omega) = x} m(\omega).$$

Note that $X = x$ inside the $P(X = x)$ is the shorthand notation for the event

$$\{\omega : X(\omega) = x\} \subset \Omega.$$

1.2   Probability Distribution

We now given a formal definition for a probability distribution.

Definition 1.1: Probability Distribution

Let Ω be a discrete (finite or countably infinite) set. Then, the function

$$P : \Omega \mapsto [0, 1]$$

is called a **probability distribution** on Ω if the following hold:

1. $P(\omega) \geq 0$ for all $\omega \in \Omega$.

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

Now, recall that our basic setup in probability theory is a random experiment, a set of possible outcomes Ω, and a probability distribution m on Ω, where $m(\omega)$ is the probability that $\omega \in \Omega$ occurs.

Now, if X is a random value on Ω, then note that $P(X = x)$ is a probability distribution on the set

$$\Omega_X = \{X(\omega) \mid \omega \in \Omega\}$$

(i.e. the image of the function X) of all possible values that X can take.

(Example.) If we roll a fair die, then our original setup is

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

and $m(i) = \frac{1}{6}$ for all $i \in \Omega$. This is an example of probability distribution.

(Example.) If X is the RV that takes the value 1 if the roll is 1 or 2 and the value 0 otherwise, then:

- $P(X = 1) = \frac{1}{3}$ since the possible ω values are 1 and 2, each with probaiblity $\frac{1}{6}$. Adding them up yields $\frac{1}{3}$.

- $P(X = 0) = \frac{2}{3}$ since the possible ω values are 3, 4, 5, and 6, each with probability $\frac{1}{6}$.

$P(X = 1)$ and $P(X = 0)$ are the probability distributions of X (on the set $\Omega_X = \{0, 1\}$). This is because:

- All of the masses are at least 0 since $\frac{2}{3} > \frac{1}{3} > 0$.

- $\frac{1}{3} + \frac{2}{3} = 1$.

The above example is an example of the **probability mass function** (PMF) of X.

1.3   Set Theory Review

Recall that events are subsets. In particular, let $A, B \subset \Omega$ be two events. Then:

- <u>Intersection:</u> $A \cap B = \{\omega \mid \omega \in A \text{ and } \omega \in B\}$.

- <u>Union:</u> $A \cup B = \{\omega \mid \omega \in A \text{ or } \omega \in B\}$.

- <u>Difference:</u> $A \setminus B = \{\omega \mid \omega \in A \text{ and } \omega \notin B\}$.

- <u>Complement:</u> $A^C = \Omega \setminus A$.

Two events A and B are said to be **disjoint** if $A \cap B = \emptyset$. If two events are disjoint, then it is impossible for them to both occur at the same time.

1.4   More on Probability

We now introduce our first theorem of this class.

Theorem 1.1

Suppose that P is a probability distribution on a discrete set Ω. Then,

1. $P(E) \geq 0$ for all events $E \subset \Omega$.

2. $P(\Omega) = 1$.

3. If $E \subset F \subset \Omega$, then $P(E) \leq P(F)$.

4. If $A \cap B = \emptyset$ are disjoint, then $P(A \cup B) = P(A) + P(B)$.

5. $P(A^C) = 1 - P(A)$ for all events $A \subset \Omega$.

Proof. We'll prove each of the following statements in this theorem.

1. We know that E is a subset of Ω. Take some $\omega \in E$. Then, we know that its mass, $m(\omega) \geq 0$. Thus, it follows that
$$\sum_{\omega \in E} m(\omega) \geq 0.$$

2. Recall that Ω is the set of all possible outcomes. Therefore, it follows that $P(\Omega) = 1$; if this is false, then this implies that there is at least one outcome that isn't in Ω.

3. Using (1) as a baseline, suppose that $P(F) = f$. Since E can only have elements from F, it follows that $P(E) \leq f$; if this statement is false, this implies that E has elements that aren't in F, which cannot be the case.

4. Since $A \cap B = \emptyset$, it follows that
$$P(A \cap B) = \sum_{\omega \in A \cap B} P(\omega) = \sum_{\omega \in A} P(\omega) + \sum_{\omega \in B} P(\omega) = P(A) + P(B).$$

   The key here is that we are not double-counting anything.

5. Recall that $A \cup A^C = \Omega$ and $A \cap A^C = \emptyset$. In particular, since $A \cap A^C = \emptyset$, then
$$P(A \cup A^C) = P(A) + P(A^C).$$

   But, since $A \cup A^C = \Omega$ and $P(\Omega) = 1$, we know that
$$1 = P(A) + P(A^C).$$

   Therefore, it follows that
$$P(A^C) = 1 - P(A^C).$$

This concludes the proof. $\qquad\square$

Looking at #4 in the previous theorem, we can actually generalize this.

Theorem 1.2

If $A_1, \ldots, A_n$ are pairwise disjoint (i.e. $\bigcap_{i \in [1,n] \subset \mathbb{Z}} A_i = \emptyset$), then

$$P\left(\bigcup_{n=1}^{n} A_i \right) = \sum_{i=1}^{n} P(A_i).$$

Proof. One way we can go about this is to take advantage of the fact that $A_1, \ldots, A_n$ are pairwise disjoint. We will use induction on n.

- <u>Base Case:</u> Suppose $n = 1$. Trivially, A_1 is pairwise disjoint since it's the only set and so $P(A_1) = P(A_1)$. Likewise, $n = 2$ is satisfied by the previous theorem.

- <u>Inductive Step:</u> Suppose that this holds for n. We now want to show that this holds for $n + 1$. To do so, we note that
$$A_1 \cap \cdots \cap A_n = \emptyset$$
and
$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n).$$
So, we can define $A = A_1 \cup \cdots \cup A_n$. We now introduce the set A_{n+1}; suppose that $A_{n+1} \cap A = \emptyset$. Then, it follows that
$$P(A \cup A_{n+1}) = P(A) + P(A_{n+1}) = P(A_1) + \cdots + P(A_n) + P(A_{n+1})$$

This concludes the proof. $\square$

Remarks:

- The following consequence of the previous theorem is an extremely useful tool for calculating the probability of events.

- Often, it is difficult to find $P(E)$ directly, and it is easier to "split the job up" into doable subtasks.

1.5   Law of Total Probability & Applications

This brings us to the *Law of Total Probability*.

Theorem 1.3: Law of Total Probability (LoTP)

Let $E \subset \Omega$ be an event, and let $A_1, \ldots, A_n$ be a partition of Ω (that is, a pairwise disjoint collection of sets that "cover" the sample space $\bigcup_{i=1}^{n} A_i = \Omega$). Then, we have that
$$P(E) = \sum_{i=1}^{n} P(E \cap A_i)$$

Proof. Note that E is the pairwise disjoint union of the sets $E \cap A_1, \ldots, E \cap A_n$. Thus, we can just apply the previous theorem. $\square$

Remark: While it might be difficult to find $P(E)$ directly, if you pick the A_i's wisely, it can become easy to find each of the $P(E \cap A_i)$'s.

Corollary 1.1

For any two events A and B,
$$P(A) = P(A \cap B) + P(A \cap B^C)$$

Remark: This holds since B, B^C is a parition of Ω.

> **Theorem 1.4: I**
>
> A and B are subsets of Ω, then
>
> $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

> *Proof.* Recall that
> $$P(A \cup B) = \sum_{\omega \in A \cup B} m(\omega).$$
>
> Now, if ω is in exactly one of the two sets, then it is only counted once (hence the first and second term on the right-hand side). However, if ω is in both A and B, then we would be double-counting since it's counted for in $P(A)$ and $P(B)$. So, we need to subtract it (hence, the last term on the right-hand side). $\qquad\square$

Remark: If $A \cap B = \emptyset$, then $P(A \cap B) = 0$.

1.6   Uniform Distribution

The **uniform distribution** on a <u>finite</u> set is a simple, but very important, example of a probability distribution.

> **Definition 1.2: Uniform Distribution**
>
> The *uniform distribution* on a <u>finite</u> sample space Ω containing n elements is the function m defined by
> $$m(\omega) = \frac{1}{n}$$
> for every outcome $\omega \in \Omega$.

For example, when flipping a fair coin, there is only two possiblities: heads or tails. So,

$$m(\text{Heads}) = \frac{1}{2}.$$

A nice property of the uniform distribution is that, for all events $E \subset \Omega$, we simply have that

$$P(E) = \frac{|E|}{|\Omega|}.$$

1.7   Infinite Sample Sizes

The same definitions apply. However, notice that the rule

$$\sum_{i=1}^{\infty} P(\omega_i) = 1$$

means that this infinite series *converges*, and it converges to 1 (it is not just a usual sum). Now, when Ω is countably infinite, we further assume, in the definition of probability distribution, that

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

for all (possibly countably infinite) collections of pairwise disjoint sets $\{E_i \mid i \in I\}$.

1.7.1   Geometric Distribution

Consider the **geomtric distribution**

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

for $k = 1, 2, \ldots$ and $P(X = x) = 0$ for all other x.

(Example: Geometric Distribution.) To see this, suppose a coin flips "Tails" with probability p. Then, the random variable, X is the number of flips until we flip "Tails" for the first time, has this distribution. For example, if we flip "Heads" twice and get "Tails" on the third attempt, then $X = 3$. Indeed, for this to happen on flip k, we need all of the previous $k-1$ flips to be "Heads," and then the next flip to be "Tails."

To check that this is a bonafide probability distribution, we note that

$$\sum_{k=1}^{\infty} P(X = k)$$

is equal to

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=0}^{\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1.$$

Note that the second-to-last step is from the geometric series

$$\sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}$$

for $|\alpha| < 1$.