Inclusion-Exclusion Formula

For two events \( A \) and \( B \) we have

\[
\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B].
\]

For three events \( A, B, \) and \( C \), we have

\[
\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[A \cap B] - \Pr[B \cap C] - \Pr[A \cap C] + \Pr[A \cap B \cap C].
\]

For events \( A_1, A_2, \ldots, A_n \) in some probability space, let 

\[
S_1 = \{(i_1) | 1 \leq i_1 \leq n\},
\]

\[
S_2 = \{(i_1, i_2) | 1 \leq i_1 \leq i_2 \leq n\},
\]

and more generally let 

\[
S_p = \{(i_1, i_2, \ldots, i_p) | 1 \leq i_1 \leq i_2 \leq \ldots \leq i_p \leq n\}.
\]

Then we have

\[
\Pr[\bigcup_{i=1}^n A_i] = \sum_{i \in S_1} \Pr[A_i] - \sum_{(i_1, i_2) \in S_2} \Pr[A_{i_1} \cap A_{i_2}] + \sum_{(i_1, i_2, i_3) \in S_3} \Pr[A_{i_1} \cap A_{i_2} \cap A_{i_3}] - \cdots + (-1)^{n-1} \Pr[\bigcap_{x=1}^n A_x]
\]

Note that there are \( 2^n - 1 \) non-empty subsets of a set of \( n \) events. To compute the probability of the intersection of every such subset is not possible when \( n \) is large. In such cases we have to approximate the probability of a union of \( n \) events. The successive terms of the above formula actually give an overestimate and underestimate respectively of the actual probability. In many situations the upper-bound given by the first term itself is quite useful. It is called the union-bound and is given by

\[
\Pr[\bigcup_{i=1}^n A_i] \leq \sum_{i=1}^n \Pr[A_i]
\]

Note that when the events are pairwise disjoint, the inequality in the above expression becomes an equality.

**Example.** Consider three flips of a fair coin. What is the probability that result is heads on the first flip or the third flip?

**Solution.** Let \( H_1 \) and \( H_2 \) denote the events that the first flip results in heads and the third flip results in heads respectively. By the inclusion-exclusion formula, we have

\[
\Pr[H_1 \cup H_2] = \Pr[H_1] + \Pr[H_2] - \Pr[H_1 \cap H_2]
\]

\[
= \frac{1}{2} + \frac{1}{2} - \frac{1}{4}
\]

\[
= \frac{3}{4}
\]
Example. When three dice are rolled what is the probability that one of the dice results in 4?

Solution. Let $F_i, i \in \{1, 2, 3\}$ be the event that the $i$th dice results in a 4. We are interested in $\Pr[F_1 \cup F_2 \cup F_3]$. By inclusion-exclusion formula we have

$$\Pr[F_1 \cup F_2 \cup F_3] = \Pr[F_1] + \Pr[F_2] + \Pr[F_3] - \Pr[F_1 \cap F_2] - \Pr[F_1 \cap F_3] - \Pr[F_2 \cap F_3] + \Pr[F_1 \cap F_2 \cap F_3]$$

Since the events $F_1, F_2, F_3$ are mutually independent we can rewrite the above expression as

$$\Pr[F_1 \cup F_2 \cup F_3] = \Pr[F_1] + \Pr[F_2] + \Pr[F_3] - \Pr[F_1] \Pr[F_2] - \Pr[F_1] \Pr[F_3] - \Pr[F_2] \Pr[F_3] + \Pr[F_1] \Pr[F_2] \Pr[F_3]$$

An easier way to solve this is as follows. Let $\overline{F_i}$ be the complement of event $F_i, i = 1, 2, 3$.

$$\Pr[F_1 \cup F_2 \cup F_3] = 1 - \Pr[\overline{F_1} \cap \overline{F_2} \cap \overline{F_3}] = 1 - (5/6)^3 = \frac{91}{216}$$

Random Variables

In an experiment we are often interested in some value associated with an event as opposed to the actual event itself. For example, consider an experiment that involves tossing a coin three times. We may not be interested in the actual head-tail sequence that results but be more interested in the number of heads that occur. These quantities of interest are called random variables.

Definition. A random variable $X$ on a sample space $\Omega$ is a real-valued function that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

In this course we will study discrete random variables which are random variables that take on only a finite or countably infinite number of values.

For a discrete random variable $X$ and a real value $a$, the event “$X=a$” is the set of outcomes in $\Omega$ for which the random variable assumes the value $a$, i.e., $X = a \equiv \{ \omega \in \Omega | X(\omega) = a \}$. The probability of this event is denoted by

$$\Pr[X = a] = \sum_{\omega \in \Omega : X(\omega) = a} \Pr[\omega]$$

Definition. The distribution or the probability mass function (PMF) of a random variable $X$ gives the probabilities for the different possible values of $X$. Thus, if $x$ is a value that $X$ can assume then $p_X(x)$ is the probability mass of $X$ and is given by

$$p_X(x) = \Pr[X = x]$$
Observe that \( \sum_x p_X(x) = \sum_x \Pr[X = x] = 1 \). This is because the events \( X = x \) are disjoint and hence partition the sample space \( \Omega \).

Consider the experiment of tossing three fair coins. Let \( X \) be the random variable that denotes the number of heads that result. The PMF or the distribution of \( X \) is given below.

\[
p_X(x) = \begin{cases} 
1/8 & \text{if } x = 0 \text{ or } x = 3 \\
3/8 & \text{otherwise}
\end{cases}
\]

The definition of independence that we developed for events extends to random variables.

**Definition.** Two random variables \( X \) and \( Y \) are independent if and only if

\[
\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \times \Pr[Y = y]
\]

for all values \( x \) and \( y \). Similarly, random variables \( X_1, X_2, \ldots, X_k \) are mutually independent if and only if, for any subset \( I \subseteq [1, k] \) and any values \( x_i, i \in I \),

\[
\Pr[\bigcap_{i \in I} X_i = x_i] = \prod_{i \in I} \Pr[X_i = x_i]
\]

**Expectation**

The PMF of a random variable, \( X \), provides us with many numbers, the probabilities of all possible values of \( X \). It would be desirable to summarize this distribution into a representative number that is also easy to compute. This is accomplished by the *expectation* of a random variable which is the weighted average (proportional to the probabilities) of the possible values of \( X \).

**Definition.** The *expectation* of a discrete random variable \( X \), denoted by \( \mathbb{E}[X] \), is given by

\[
\mathbb{E}[X] = \sum_i ip_X(i) = \sum_i i \Pr[X = i]
\]

In our running example, in expectation the number of heads is given by

\[
\mathbb{E}[X] = 0 \times \frac{1}{8} + 3 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} = \frac{3}{2}
\]

As seen from the example, the expectation of a random variable may not be a valid value of the random variable.

**Example.** When we roll a die what is the result in expectation?

**Solution.** Let \( X \) be the random variable that denotes the result of a single roll of dice. The PMF for \( X \) is given by

\[
p_X(x) = \frac{1}{6}, x = 1, 2, 3, 4, 5, 6.
\]
The expectation of $X$ is given by

$$E[X] = \sum_{x=1}^{6} p_x(x) \cdot x = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

**Example.** When we roll two dice what is the expected value of the sum?

**Solution.** Let $S$ be the random variable denoting the sum. The PMF for $S$ is given by

$$p_S(x) = \begin{cases} \frac{1}{36}, & x = 2, 12 \\ \frac{2}{36}, & x = 3, 11 \\ \frac{3}{36}, & x = 4, 10 \\ \frac{4}{36}, & x = 5, 9 \\ \frac{5}{36}, & x = 6, 8 \\ \frac{6}{36}, & x = 7 \end{cases}$$

The expectation of $S$ is given by

$$E[S] = \sum_{x=2}^{12} p_S(x) \cdot x$$

$$= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 4 + \frac{5}{36} \times 6 + \frac{6}{36} \times 7 +$$

$$\frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12$$

$$= \frac{252}{36} = 7$$

**Theorem:** Let $Y$ be a random variable that takes on non-negative integer values. Then

$$E[Y] = \sum_{i=1}^{\infty} \Pr[Y \geq i]$$

**Proof.**

$$E[Y] = \sum_{j=1}^{\infty} j \Pr[Y = j]$$

$$= \sum_{j=1}^{\infty} \sum_{i=1}^{j} \Pr[Y = j]$$

$$= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \Pr[Y = j]$$

$$= \sum_{i=1}^{\infty} \Pr[Y \geq i]$$
Linearity of Expectation

One of the most important properties of expectation that simplifies its computation is the *linearity of expectation*. By this property, the expectation of the sum of random variables equals the sum of their expectations. This is given formally in the following theorem. I didn’t cover the proof in the class but I am including it here for anyone who is interested.

**Theorem.** For any finite collection of random variables $X_1, X_2, \ldots, X_n$,

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

**Proof.** We will prove the statement for two random variables $X$ and $Y$. The general claim can be proven using induction.

$$
\mathbb{E}[X + Y] = \sum_{i} \sum_{j} (i + j) \Pr[X = i \cap Y = j] = \sum_{i} \sum_{j} (i \Pr[X = i \cap Y = j] + j \Pr[X = i \cap Y = j]) = \sum_{i} \sum_{j} i \Pr[X = i \cap Y = j] + \sum_{i} \sum_{j} j \Pr[X = i \cap Y = j] = \sum_{i} i \Pr[X = i] + \sum_{j} j \Pr[Y = j] = \mathbb{E}[X] + \mathbb{E}[Y]
$$

It is important to note that no assumptions have been made about the random variables while proving the above theorem. For example, the random variables do not have to be independent for linearity of expectation to be true.

**Lemma.** For any constant $c$ and discrete random variable $X$,

$$
\mathbb{E}[cX] = c \mathbb{E}[X]
$$

**Proof.** The lemma clearly holds for $c = 0$. For $c \neq 0$

$$
\mathbb{E}[cX] = \sum_{j} j \Pr[cX = j] = c \sum_{j} (j/c) \Pr[X = j/c] = c \sum_{k} k \Pr[X = k] = c \mathbb{E}[X]
$$
Example. Using linearity of expectation calculate the expected value of the sum of the numbers obtained when two dice are rolled.

Solution. Let $X_1$ and $X_2$ denote the random variables that denote the result when die 1 and die 2 are rolled respectively. We want to calculate $E[X_1 + X_2]$. By linearity of expectation

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

$$= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) + \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6)$$

$$= 3.5 + 3.5$$

$$= 7$$