Finite Probability Spaces

When a situation has some finite number of possible outcomes (*samples*), each its own chance of occurrence (*probability*), the mathematical model we use is that of a *finite probability space*.

- A ***finite probability space*** $(\Omega, \mathbb{P})$ consists of:
  - A set Ω, called the ***sample space***; and
  - A function $\mathbb{P} : \Omega \to \mathbb{R}$, called a ***probability distribution*** on Ω, with the following properties:
    - $\forall\, s \in \Omega,\ \mathbb{P}(s) \geq 0$; and [No sample can have negative probability.]
    - $\displaystyle\sum_{s\in\Omega} \mathbb{P}(s) = 1.$ [The sum of all samples' probabilities is 1.]

    The function $\mathbb{P}$ literally "distributes" a total probability of 1 among the elements of the sample space Ω.
- If the function $\mathbb{P}$ is *constant* (i.e., the probability of every sample is the same), we call the probability ***uniformly distributed***.
  - In this case (but only in this case!), $\forall\, s \in \Omega,\ \mathbb{P}(s) = \dfrac{1}{|\Omega|}$.
- The set $\mathscr{P}(\Omega)$ of all *subsets* of Ω is called the ***event space*** of the probability space $(\Omega, \mathbb{P})$
  - The probability distribution $\mathbb{P}$ lets us define probabilities for *events* $E \in \mathscr{P}(\Omega)$ via $\mathbb{P}(E) = \displaystyle\sum_{s\in E} \mathbb{P}(s)$,

    i.e., the probability of an *event* is the sum of the probabilities of all *samples* in the event.
    - In the case of a *uniform* distribution, we have $\mathbb{P}(E) = \dfrac{|E|}{|\Omega|}$.

  The events space effectively gives us all possible ***events*** (i.e., combinations of samples) that we could ask about in the probability space.

  Be very careful to distinguish the *samples* of the *sample space* Ω from the *events* of the *event space* $\mathscr{P}(\Omega)$!

Product / Joint Probability Spaces

We can "join" two probability spaces $(X, \mathbb{P}_X)$ and $(Y, \mathbb{P}_Y)$, considering a sample $x \in X$ to be taken and, independently, a sample $y \in Y$ being taken.

- The ***product***, or ***joint probability space***, of $(X, \mathbb{P}_X)$ and $(Y, \mathbb{P}_Y)$ is $(\Omega, \mathbb{P})$ with:
  - $\Omega = X \times Y$ [The set of all ordered pairs of a sample in X followed by a sample in Y.]
  - $\mathbb{P} : X \times Y \to \mathbb{R}$, defined by $\mathbb{P}(x, y) = \mathbb{P}_X(x) \cdot \mathbb{P}_Y(y)$.

    [The probability of x then y is the product of their probabilities.]
    - Note that the two probability spaces need have nothing to do with one another: X could be flips of a coin, and Y could be rolls of a die, and $X \times Y$ would consist of all possible pairs of a coin flip and a die roll.
- We can iterate this construction to form a probability space from $X^n = \{(x_1, x_2, \ldots, x_n)\ :\ x_i \in X\}$.

  [e.g., flipping a coin 10 times, or rolling 3 dice.]

Subspaces

Given any event B in a probability space $(\Omega, \mathbb{P})$ with $\mathbb{P}(B) > 0$ (i.e., an event that is possible!), we can build a new probability space for just the samples in the set B.

- If $B \subset \Omega$ has $\mathbb{P}(B) > 0$ is any event of positive probability, the ***subspace*** $(B, \mathbb{P}_B)$ is defined as follows:
  - B is just the set of samples in our event of choice; and
  - for each $A \subset B$, we define $\mathscr{P}_B(A) = \dfrac{\mathbb{P}(A)}{\mathbb{P}(B)}$. [We have to scale the probabilities to get a total of 1.]
    - Note that we could define $\mathbb{P}_B(E)$ for *any* $E \subset \Omega$ by first intersecting E with B: $\mathbb{P}_B(E) = \dfrac{P(E \cap B)}{\mathbb{P}(B)}$.

Mutual Exclusivity, Independence, and Conditional Probability

Suppose that $A, B \in \mathscr{P}(\Omega)$ are events in a probability space $(\Omega, \mathbb{P})$.

- We define the events A and B to be:
  - ***mutually exclusive*** when $\mathbb{P}(A \cap B) = 0$, and/or [i.e., it is impossible for both events to happen]
  - ***independent*** when $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
    [i.e., the chance of both happening simultaneously is just the product of each one happening individually].
    - If $\mathbb{P}(A \cap B) > \mathbb{P}(A) \cdot \mathbb{P}(B)$, then the occurence of one event *promotes* occurrence of the other.
    - If $\mathbb{P}(A \cap B) < \mathbb{P}(A) \cdot \mathbb{P}(B)$, then the occurence of one event *inhibits* occurrence of the other.
- Suppose that $\mathbb{P}(B) > 0$.
  The ***conditional probability*** $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ represents the probability that A happens *if* we know that B happens.
  - Note that this is nothing more than $\mathbb{P}_B(A)$, the probability of A within the subspace B!
  - In the case that A and B are *independent*, this gives $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$.
    - That is, if the events are independent, knowing that B happens doesn't affect the probability of A happening; this justifies our use of the word *independent* above!

Random Variables

Suppose that $(\Omega, \mathbb{P})$ is a probability space.

- A ***random variable*** X on Ω attaches a *numeric value* to each sample in Ω.
  - This simply amounts to a *function* $X : \Omega \to \mathbb{R}$!
  - These "variables" can be combined via our usual arithmetic and other operations ($+$, $-$, $\times$, $\min$, $\max$, etc.) simply by combining their values at each sample: e.g., $X + Y$ is given by $(X + Y)(s) = X(s) + Y(s)$.
  - A random variable X on $(\Omega, \mathbb{P})$ defines a new probability space:
    - samples are the values in the *range* of the function X; and
    - each value v is assigned its probability $\mathbb{P}(X = v)$ (i.e., $\sum_{X(s)=v} \mathbb{P}(s)$) of occurring.
- The ***expected value*** of a random variable X gives a weighted average of its values: $E(X) = \sum_{s \in \Omega} X(s) \cdot \mathbb{P}(s)$.
  - This is literally the average value one would *expect* if they sampled Ω a large number of times.
  - Expected value is ***linear***, i.e., for any random variables X, Y on $(\Omega, \mathbb{P})$:
    - for any constant $c \in \mathbb{R}$, $E(c \cdot X) = c \cdot E(X)$; and
    - $E(X + Y) = E(X) + E(Y)$.
- If $A \in \mathscr{P}(\Omega)$ is any event, then the ***indicator variable*** $\mathbb{1}_A$ is defined by $\mathbb{1}_A(s) = \begin{cases} 1 \text{ if } s \in A \\ 0 \text{ if } s \notin A \end{cases}$
  - Note that $E(\mathbb{1}_A) = \mathbb{P}(A)$.
  - This allows us to convert *events* into *random variables*, which are *linear*!
    Regardless of whether events A and B are independent or not, it is true that $E(\mathbb{1}_A + \mathbb{1}_B) = E(\mathbb{1}_A) + E(\mathbb{1}_B)$.