Every random variable has an associated probability space.

What is a probability space?
Let $S$ be an arbitrary set. Let $\mathcal{F}$ be a set of subsets of $S$ satisfying
1. $\emptyset \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
3. If $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup A_i \in \mathcal{F}$

Any set of subsets of $S$ satisfying (i), (ii) and (iii) is called a $\sigma$-field.

A probability measure is any function $P$ defined on $\mathcal{F}$ that takes values in $[0,1]$, and satisfies the axioms of probability.

**Axioms of Probability**
1. $P(A) \geq 0$ for all $A \in \mathcal{F}$
2. $P(S) = 1$
3. If $A_1, A_2, \ldots \in \mathcal{F}$ is a sequence of mutually disjoint subsets of $S$ (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then
   \[ P\left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i) \]
   This is called countable additivity.

If $P$ is a probability measure on $S$, the pair $(S, P)$ is called a probability space.

**Random Variables**
A random variable is a function from $S$ to $\mathbb{R}$.

**Example** Suppose you conduct an experiment where you flip a coin infinitely often. The outcome is an infinite sequence of heads and tails. Let $S$ be the set of all possible sequences of heads and tails. For $s \in S$, let $X(s)$ be the long-run proportion of heads. Then $X(s) \in [0,1] \subset \mathbb{R}$, so $X$ is a random variable.
The distribution of a random variable $X$ is the probability measure on $\mathbb{R}$ induced by $P$. If this distribution is denoted by $P_X$, then for $A \subseteq \mathbb{R}$, $P_X(A) = P(\{s \in S : X(s) \in A\})$.

This subset of $S$ is called the pre-image of $A$ under the mapping $X$.

We often do not make explicit reference to the underlying probability measure $P$ or explicitly give the probability measure $P_X$ (but they are always there). The usual ways that we specify the distribution of a random variable $X$ is through:

1. a distribution function (df) $F_X$
2. a probability mass function (pmf) $p_X$, if $X$ is discrete
3. a probability density function (pdf) $f_X$, if $X$ is continuous.

All of these functions are functions from $\mathbb{R}$ to $[0, \infty)$. This is in contrast to $P_X$, which is a function defined on subsets of $\mathbb{R}$.
Remark on Notation

Last time we denoted the distribution of a random variable $X$ by $P_X$, which is a probability measure on $\mathbb{R}$, and given by

$$P_X(A) = P(\{s \in S : X(s) \in A\})$$

where $P$ is the underlying probability measure on $S$. Notationally, it is simpler to not refer to the probability measure $P_X$ but rather just refer to the probability measure $P$, when writing probabilities involving $X$. Also, for simplicity we will drop the argument $s$ in the notation. So we will write $X \in A$ instead of $\{s \in S : X(s) \in A\}$ (we will also drop the curly brackets). So we will write $D(\{X \in A\})$ to mean $P(\{s \in S : X(s) \in A\})$. For specific forms of the subset $A$ we will also write:

If $A = (-\infty, x]$ we write $P(X \leq x)$ to mean $P(X \in (-\infty, x])$

If $A = (-\infty, x)$ we write $P(X < x)$

If $A = [x, \infty)$ we write $P(X \geq x)$

If $A = (x, \infty)$ we write $P(X > x)$

If $A = (a, b)$ we write $P(a < X < b)$

If $A = (a, b]$ we write $P(a < X \leq b)$

If $A = \{a\}$ we write $P(X = a)$

etc.

Review of Distribution Functions

For a random variable $X$, its distribution function (df), denoted by $F_X$, is a function from $\mathbb{R}$ to $[0, 1]$ defined by

$$F_X(x) = P(X \leq x)$$

for $x \in \mathbb{R}$.

Distribution functions have the following properties.

- Any distribution function is a nondecreasing function on $\mathbb{R}$.
  
  That is, if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$.
  
  To see this, note that if $x_1 \leq x_2$ then $(-\infty, x_1] \subseteq (-\infty, x_2]$.
  
  So we can write $(-\infty, x_2) = (-\infty, x_1] \cup (x_1, x_2]$. The 2 sets
\(-\infty, x_1\) and \((x_1, x_2]\) are disjoint, so
\[
P(\{X \in (-\infty, x_2]\}) = P(\{X \in (-\infty, x_1]\} \cup \{X \in (x_1, x_2]\})
\]
\[
F_x(x_2) = P(X \in (-\infty, x_1]) + P(X \in (x_1, x_2])
\]
\[
\geq P(X \in (-\infty, x_1])
\]
\[
F_x(x_1)
\]

So \(F_x(x_1) \leq F_x(x_2)\)

2. Any distribution function is continuous from the right. To say that \(F_X\) is continuous from the right at a point \(x\) means that if \(x_1 \geq x_2 \geq x_3 \geq \ldots \) and \(X_n \downarrow x\) as \(n \to \infty\) then
\[
\lim_{n \to \infty} F_X(x_n) = F_X(x).
\]
To show this consider the sets
\[
A = \bigcap_{n=1}^{\infty}\{X \leq x_n\} \quad \text{and} \quad B = \{X \leq x\}.
\]
First, note that \(A = B\). To show this one shows that \(A \subseteq B\) and \(B \subseteq A\). To show that \(A \subseteq B\), let \(s \in A\). Then \(s \in \{X \leq x_n\}\) for every \(n\).
That is, \(X(s) \leq x_n\) for every \(n\). But since \(X_n \downarrow x\) we must have \(X(s) \leq x\). But then \(s \in B\). So \(A \subseteq B\). Conversely, if \(s \in B\) then \(X(s) \leq x\). But \(x \leq x_n\) for every \(n\), so \(X(s) \leq x_n\) for every \(n\). So \(s \in \bigcap_{n=1}^{\infty}\{X \leq x_n\} = A\).

So \(B \subseteq A\). Therefore, \(A = B\). Also, note that
\[
\{X \leq x_n\}_{n=1}^{\infty}
\]
is a decreasing sequence of sets. That is, if \(n_1 \leq n_2\) then \(\{X \leq x_{n_2}\} \subseteq \{X \leq x_{n_1}\}\). Now we can use the continuity property of probabilities. By the continuity of probability (Theorem 1.8) for a decreasing sequence of sets,
\[
P\left(\bigcap_{n=1}^{\infty}\{X \leq x_n\}\right) = \lim_{n \to \infty} P(X \leq x_n)
\]

\[
\text{this set is} \quad \{X \leq x\}
\]

So then we have
\[
F_x(x) = P(X \leq x) = P\left(\bigcap_{n=1}^{\infty}\{X \leq x_n\}\right) = \lim_{n \to \infty} P(X \leq x_n) = \lim_{n \to \infty} F_x(x_n)
\]
So if there is a discontinuity in $F_x$ at the point $x_1$, the picture looks like

\[ F_x(x) \]

\[ x \]

\[ x_1 \]

3. The distribution function $F_x$ of a random variable $X$ determines the distribution of $X$. Note that $F_x$ gives the probability of sets of the form \{ $X \leq x$ \} for $x \in \mathbb{R}$ but does not directly give the probabilities \{ $X \in A$ \} for arbitrary subset $A$ of $\mathbb{R}$. It is a theorem from advanced probability that the collection of probabilities \{ $P(X \leq x)$ : $x \in \mathbb{R}$ \} determine all the probabilities in the collection \{ $P(X \leq A)$ : $A \in \mathcal{F}$ \}, where $\mathcal{F}$ is the $\sigma$-field of subsets of $\mathbb{R}$ that is of interest.
Probability Mass Functions (pmf)

If $X$ is a discrete random variable then we can use its pmf to specify its distribution. We say that $X$ is discrete if there is a finite or countable set $S_X \subseteq \mathbb{R}$ such that $P(X \in S_X) = 1$. The set $S_X$ is called a support of the distribution of $X$. The probability mass function (pmf) of $X$ is the function $p_X : \mathbb{R} \to \mathbb{R}$ given by

$$p_X(x) = P(X = x), \quad x \in \mathbb{R}.$$  

$p_X(x)$ is defined for every $x \in \mathbb{R}$ and $p_X(x) > 0$ only if $x \in S_X$. The pmf determines the distribution of $X$: if $A \subseteq \mathbb{R}$ then $P(X \in A) = \sum_{x \in A} p_X(x)$. The df of $X$ when $X$ is discrete is always a step function. If we order the points in $S_X$ as $x_1 < x_2 < x_3 < \ldots$

$$F_X(x) = P(X \leq x)$$

Probability Density Functions (pdf)

If $X$ is a continuous random variable then we can use its pdf to specify its distribution. Unlike the discrete case, a continuous random variable may not have a pdf. We say that a random variable $X$ is continuous if its df $F_X(x)$ is continuous at all $x \in \mathbb{R}$. If there exists a function $f_X : \mathbb{R} \to [0, \infty)$ satisfying

$$P(X \in A) = \int_A f_X(x) \, dx$$

for $A \subseteq \mathbb{R}$, then $f_X(x)$ is called a probability density function (pdf) of $X$. Remark: Unlike the df of $X$ or the pmf of a discrete $X$,
Remark: If \( A = \mathbb{R} \), then if \( f_X \) is a pdf for \( X \) then 
\[
P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(x) \, dx.
\]
But \( P(X \in \mathbb{R}) = 1 \) for any random variable \( X \). So any pdf \( f_X(x) \) must satisfy \( \int f_X(x) \, dx = 1 \). Any nonnegative function \( f_X(x) \) defined on \( -\infty, \infty \) that integrates to 1 is the pdf of some random variable.

Remark: You can change the value of any given pdf at any finite or countable set of \( x \) values, and it will still be a pdf for the same random variable. This is because doing so will not change the value of the integral \( \int f_X(x) \, dx \) for any \( A \subseteq \mathbb{R} \). So the pdf of a random variable \( X \) is not unique. Normally, we use a standard version of any pdf that we use in practice, typically a continuous function.

Remarks for both pmfs and pdfs:

1. Both pmfs and pdfs are defined for all \( x \in \mathbb{R} \), and when specifying a pmf or pdf it should be specified for all values of \( x \in \mathbb{R} \) (including at points \( x \) where \( P_X(x) = 0 \) or \( f_X(x) = 0 \)).

2. In most practical situations, the distribution of a random variable \( X \) is specified or given only by giving the pmf of \( X \) (if \( X \) is discrete) or the pdf of \( X \) (if \( X \) is continuous), or the df, without any specification of the underlying probability measure \( P \) on \( S \), or \( X \) as a function. For any given pmf or pdf, there can be many (usually infinitely many) domains \( S \), underlying probability measures \( P \), and functions \( X \), that produce the distribution on \( \mathbb{R} \) specified by the given pmf or pdf.