<table>
<thead>
<tr>
<th>y</th>
<th>( P(Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.237</td>
</tr>
<tr>
<td>1</td>
<td>0.396</td>
</tr>
<tr>
<td>2</td>
<td>0.264</td>
</tr>
<tr>
<td>3</td>
<td>0.088</td>
</tr>
<tr>
<td>4</td>
<td>0.015</td>
</tr>
<tr>
<td>5</td>
<td>0.001</td>
</tr>
</tbody>
</table>

You can see that a rv \( Z \) is completely specified by two things: the values it can take on, and the probability for those values.

(see p. 109) Definition: The (probability) distribution of a random variable \( Y \) is the collection of all probabilities of the form \( P(Y \in A) \) for all sets \( A \) of real numbers in the non-void collection \( \mathcal{C}_R \) of subsets of the real number line \( R \). The rv \( Y \) in the T-s age study has a finite set of possible values —
Definition: A random variable has a discrete distribution, or equivalently, it is a discrete RV, if the set of (distinct) possible values is finite or at most countably infinite; RVs for which the set of possible values is uncountable are called continuous random variables.

Example 1: The RV $X = \begin{cases} 1 & \text{if } Y > 0 \\ 0 & \text{otherwise} \end{cases}$ (with $Y = \# T-s$ holding) is discrete, taking on only the values $\{0, 1\}$ - such RVs are called dichotomous or binary.
Imagine a scale for weighing things that has a dial you can set to specify how many significant figures of precision you want. Buy a "1 pound" package of butter at your favorite market and weigh it.

If there's no conceptual limit to the number of sigfigs you could get, a rv $Y = (\text{the actual (true) weight of the package})$ should be modeled as continuous, having values (e.g.) on $(0, \infty)$, the positive part of $\mathbb{R}$.

Reality check: Infinite precision is impossible in practice.
every measurement you ever make is in actuality discrete, but it's useful to regard rvs that are conceptually continuous (i.e., no limit in principle to the precision of measurement) as continuous. **Definition** Given a (mass) discrete rv $X$, the probability function (pmf or pf) of $X$ is the function that keeps track of the probabilities associated with $X$: $f_X(y) = P(X = y)$. The set \( \{ y : f_X(y) > 0 \} \) is called the support of (the distribution of) $X$.

(As is almost unique in using "pf", nearly everybody talks about the pmf.)
Example 1. In the powerball lottery (see homework 1 problem 2) 5 white balls are drawn at random without replacement from a bin with balls numbered \{1, 2, \ldots, 69\}. Let \( W_i \) denote the number of the drawn ball on the \( i \)-th draw. Clearly \( p(W_i = w_i) = \begin{cases} \frac{1}{69} & \text{if } w_i = 1, 2, \ldots, 69 \\ 0 & \text{otherwise} \end{cases} \)

less clearly (but true) \( W_2, \ldots, W_5 \) follow the same distribution if nothing is known about the previous draws.

Definition 2. For any two integers \( a \leq b \), a rv \( Y \) that is equally likely to be any of the values \( \{a, a+1, \ldots, b\} \) has the uniform distribution \( \text{Uniform} \{a, b\} \). Evidently
A rv $Y$ that only takes on the values $\{0, 1\}$ is said to have a binary rv. It is said to have a Bernoulli distribution with parameter $p$, written $\text{Bernoulli}(p)$, if

\[
\mathbb{P}(Y = y) = \begin{cases} 
  p & \text{for } y = 1 \\
  1-p & \text{for } y = 0 \\
  0 & \text{else}
\end{cases}
\]

Notation

$I$ follows a Bernoulli$(p)$ distribution, is distributed as $I \sim \text{Bernoulli}(p)$ or $(I | p)$
If its pmf is \( f(y) = P(T = y) = \begin{cases} \frac{1}{b - a + 1} & \text{for } y = a, \ldots, b \\ 0 & \text{else} \end{cases} \), then \( Y \sim \text{Uniform} \{a, b\} \) is chosen at random from \( \{a, a+1, \ldots, b\} \).

Given \( n \) trials are performed, with each trial recorded as a success or failure \( F \). If each trial is independent of all the others and the chance \( p \) of success is constant across the trials, then \( Y = \# \text{ of successes} \) has the \text{Binomial distribution}.

\[
f(y) = P(T = y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{for } y = 0, 1, \ldots, n \\ 0 & \text{else} \end{cases}
\]
In shorthand $Y \sim \text{Binomial} (n, p)$.

or $(Y | n, p)$

Let $b_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{failure} \end{cases}$

for $i = 1, \ldots, n$; then under these assumptions $b_i \sim \text{Bernoulli} (p)$ and all the $b_i$ are independent.

Notation $\{ X_i \sim f(x_i) \}$ means that all of the rvs $X_1, X_2, \ldots$ are independent and identically distributed as draws from the distribution with pdf $f(x_i)$. Thus with the success/failure trials, $b_i \sim \text{Bernoulli} (p)$ and $(i=1, \ldots, n)$

$(Y = \sum_{i=1}^{n} b_i) \sim \text{Binomial} (n, p)$. 
This is our first example of the distribution of the sum of a bunch of IID rvs, a topic we'll examine in detail later.
Continuous random variables

Example (round-off error in computer science)

Single-precision floating point decimal numbers carry about 7 sigfis of accuracy,

leading to round off error in the last digit;

it's important to study how these errors accumulate as the number of steps in a calculation increases. Since there's no reason one decimal digit would be favored over another in rounding, the uniform distribution is key to these calculations. Consider first

Uniform \{ 0, 0.1, \ldots, 0.9 \} and then \{ 0, 0.01, 0.02, \ldots \} discrete of \( 0.00\overline{1} \ldots \)
In the limit with more & more right

this should go to \[ \int_0^1 f(u) \, du \], the
continuous uniform distribution
Uniform \((0,1)\) on the unit interval.

The analogue of the discrete pdf in this
continuous case is the smooth function
\[ f(u) = \begin{cases} 1 & \text{for } 0 \leq u \leq 1 \\ 0 & \text{else} \end{cases} \]
Definition
A random variable
(\(X\) has a continuous distribution)
\(\Leftrightarrow (X\) is a continuous rv) if there
exists a continuous non-negative function
\(f_y\) defined on \([-\infty, \infty]\) such that for every
interval \([a, b]\), \( P(a \leq X \leq b) = \int_a^b f_x(y) \, dy \).
In this definition, \( a \) can be \(-\infty\) and \( b \) can be \( \infty \).

**Definition**

If \( Y \) is a continuous rv, the function \( f_{\bar{Y}} \) in the previous definition is called the probability density function (pdf) of \( Y \). The set \( \{ y : f_{\bar{Y}}(y) > 0 \} \) is called the support of \( (\text{the distribution of}) \ Y \).

1. Clearly (a) \( f_{\bar{Y}}(y) \geq 0 \) for all \( y \) and (b) \( \int_{-\infty}^{\infty} f_{\bar{Y}}(y) \, dy = 1 \).

You'll recall from calculus that if \( f_{\bar{Y}} \) is continuous, what about individual points - singletons - \( \{ y \} \) on \( \mathbb{R} \)?
on its support, \( \int_a^b f_\gamma(y) \, dy \) can equally well stand for \( P(a \leq \gamma \leq b) \) or \( P(a < \gamma \leq b) \) or \( P(a \leq \gamma < b) \) or \( P(a < \gamma < b) \), because (e.g.) \( \int_a^a f_\gamma(y) \, dy = 0 \) if \( f_\gamma \) is continuous at \( y = a \). Thus, importantly, \( \mathbb{P}(\gamma = y) = 0 \) for all \( -\infty < y < \infty \).

Weirdly, this doesn't mean that the value \( y \) of \( \gamma \) is impossible, or it does with discrete rv; it just means that singletons have to have 0 probability (otherwise \( \int_{-\infty}^\infty f_\gamma(y) \, dy = \infty \) not 1).
Definition (with \( a \) and \( b \) any two real numbers satisfying \( a < b \),

\[ X \sim \text{Uniform}(a, b) \iff P(\text{of } (a, b)) = \text{the length of the subinterval} \]

\[ f_X(y) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq y \leq b \\ 0 & \text{else} \end{cases} \]

Definition (The indicator function)

The indicator function (true/false) for any proposition \( A \) is \( I(A) = \begin{cases} 1 & \text{if } A \text{ true} \\ 0 & \text{if } A \text{ false} \end{cases} \)

People sometimes also write (with \( A \) set)

\[ I_A(y) = \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{else} \end{cases} \]
with this definition, $Y \sim \text{Uniform}(a, b)$

$\iff f_Y(y) = \frac{1}{b-a} \mathbf{1}(a \leq y \leq b) = \frac{\mathbf{1}_{[a, b]}(y)}{b-a}.$

**Contrast**

$Y \sim \text{Uniform}(a, b)$ continuous and uniform on $(a, b)$ or $[a, b)$

$Y \sim \text{Uniform}\{a, b\}$ for $a, b$ integers with $a < b \iff Y$ discrete and uniform on $\{a, a+1, \ldots, b\}.$

Density and probability are not the same thing.

Density values $f_Y(y)$ are themselves not probabilities; for example, they can easily be $> 1$ and even be too.
As we’ll see later, the density function defines probability: \( P(a \leq Y \leq b) = \int_{a}^{b} f_{Y}(y) \, dy \).

For small \( \varepsilon > 0 \) you can see from this sketch that
\[
P \left(a - \frac{\varepsilon}{2} \leq Y \leq a + \frac{\varepsilon}{2} \right) = \int_{a - \frac{\varepsilon}{2}}^{a + \frac{\varepsilon}{2}} f_{Y}(y) \, dy
\]

\( = \text{area of rectangle} = \varepsilon \cdot f_{Y}(a). \)

Example
(triangular distribution)

\( f_{Y}(y) \) line with slope \( m = \frac{c}{b-a} \)

Can a continuous rv have a pdf that looks like a triangle? Let’s see what, if any, restrictions would be needed.
The line in the sketch has slope \( \frac{c}{b-a} \), and passes through the point \((a, b)\), so the equation of the line is:

\[
y - b = \frac{c}{b-a}(x - a) \quad \Rightarrow \quad y = \frac{c}{b-a}x + \frac{c - ab}{b-a}
\]

Density have to integrate to 1, so

\[
\int_{a}^{b} \frac{c}{b-a} (x-a) \, dx = 1 \quad \Leftrightarrow \quad c = \frac{2}{b-a}
\]

\[\text{Easier way: area of a triangle is } \frac{1}{2} \text{ (base)(height)}, \text{ so }\]

\[
1 = \frac{1}{2} (b-a) c \quad \text{and} \quad c = \frac{2}{b-a}
\]
Thus the triangular distribution that starts at $y=a$ and rises linearly to $y=b$ has density $f(y) = \begin{cases} \frac{2(y-a)}{(b-a)^2} & \text{if } a < y < b \\ 0 & \text{else} \end{cases}$

You can see that calculating probabilities with continuous r.v.s requires you to dust off your integral calculus.

**Example**

With the triangular distribution $f(y)$ above, what's $P(a < y < \frac{b-a}{2})$?

Hard (?): $\int_{a}^{\frac{b-a}{2}} \frac{2(y-a)}{(b-a)^2} \, dy = \frac{(3a-b)^2}{4(b-a)^2}$

Easy (?) $\int_{a}^{\frac{b-a}{2}} \frac{2(y-a)}{(b-a)^2} \, dy = \frac{\text{area of triangle}}{2}$
Sometimes it's mathematically convenient to work with unbounded continuous r.v.s, just as was true in the Poisson case study for discrete r.v.s.

\[ \begin{align*}
\mathbb{E}[Y] &= \int_0^\infty \frac{1}{1+y} \cdot 1 \, dy = \left[ \ln(1+y) \right]_0^\infty = -1(0 - 1) = 1 \checkmark
\end{align*} \]

Example (DS p. 104)

\[ Y = \text{voltage in an electrical system} : \]

in practice, \( Y \) cannot be infinite, but you may not know ahead of time what its maximum practical value is, so model it as unbounded but without much probability for extremely large values. DS give as an example the pdf

\[ f_Y(y) = \frac{1}{(1+y)^2}, \quad y > 0. \]
You can check that \( \int_{100}^{\infty} \frac{1}{1 + x} \, dx = \frac{1}{100} \approx 0.01 \), so the right tail beyond \( x = 1000 \) has almost no probability, matching the correct qualitative behavior. Sometimes a rv will be neither discrete nor continuous; people then say that it has a mixed \( (\text{discrete/continuous}) \) distribution. Example:

In medical clinical trials of people with potentially fatal diseases, the outcome variable \( Y_i \), for person \( i \), is (say) the treatment group might be
$T$ = survival time in days from the beginning of the trial; however, and a good thing too, some patients may still be alive at the time $T_{\text{end}}$ at which the trial finishes. Your model for $T$ would then have a continuous part for $0 \leq T \leq T_{\text{end}}$ and a discrete lump of probability $p$ at $T = T_{\text{end}}$ signifying $(T > T_{\text{end}})$ but we don't know what $T$ would have been if we could have observed it:

$$\int_0^{T_{\text{end}}} f_T(y) dy = (1-p) \quad \text{and} \quad P(T > T_{\text{end}}) = p.$$

(Ex: by example) (see p. 18 of doc. can notes)
The cumulative distribution function, denoted as \( F(y) \), is defined as:

\[
F(y) = P(X \leq y)
\]

For a continuous distribution, it is uniform on \((a, b)\).

For a discrete distribution, it is \( \sum_{y \leq x} f(y) \).

Distribution of \( X \) is either continuous or discrete.

If \( f(x) \) is continuous, then:

- There exists a unique characteristic function \( \phi(t) \).

- Have something continuous \( \phi(t) \).

- Continuous \( \phi(\text{pdf}) \).

- Discrete \( \phi(\text{pdf}) \).

- Continuous \( \phi \).

- Continuous \( \phi \).

- Continuous \( \phi \).