**RANDOM VARIABLES**

**Defn:** A random variable is a function that associates a real number with each element in the sample space.

**Notation:** capital letter, say $X$, denotes a random variable and its corresponding small letter, $x$, in this case, denotes one of its values.

**Example 1:** $X$: Number of defective products when 3 products are tested.

<table>
<thead>
<tr>
<th>Outcomes in sample space</th>
<th>$x$: value of $X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D: defective</td>
<td>DDD 3</td>
</tr>
<tr>
<td>N: non-defective</td>
<td>DND 2, DNN 1, NND 1, NNN 0</td>
</tr>
</tbody>
</table>

If the sample space contains a finite number of possibilities such as this case, it is called a **discrete sample space**.

**Example 2:** $Y$: Number of products tested before a defective product is found.

<table>
<thead>
<tr>
<th>Sample Space</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>1</td>
</tr>
<tr>
<td>ND</td>
<td>2</td>
</tr>
<tr>
<td>NND</td>
<td>3</td>
</tr>
<tr>
<td>NNNND</td>
<td>4</td>
</tr>
<tr>
<td>NNNNNND</td>
<td>5</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

If the sample space contains an unending sequence with as many elements as there are integers (whole numbers), it is still called a discrete sample space.

This is an example of such a case.
Example 1 and 2 are discrete random variables.

Example 3: Z: Height in inches of a random person. We can measure Z to any degree of accuracy:

71 inches
71.2 inches
71.24
71.243

Continuous sample space
Continuous random variable

If a sample space contains an infinite number of outcomes equal to the number of points on a line segment, it is called a continuous sample space.

Example 4: X: Power consumption of my house measured in kilowatts (kw) at a random time.

This is another example of a continuous random variable.

Example: Let X be the random variable: number of heads in 3 tosses of a fair coin.

<table>
<thead>
<tr>
<th>Sample space</th>
<th>x</th>
<th>P(X = x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>0</td>
<td>P(X=0) = 1/8</td>
</tr>
<tr>
<td>TTH</td>
<td>1</td>
<td>P(X=1) = 3/8</td>
</tr>
<tr>
<td>THT</td>
<td>1</td>
<td>P(X=1) = 3/8</td>
</tr>
<tr>
<td>THH</td>
<td>2</td>
<td>P(X=2) = 3/8</td>
</tr>
<tr>
<td>HTH</td>
<td>2</td>
<td>P(X=3) = 1/8</td>
</tr>
<tr>
<td>HHT</td>
<td>2</td>
<td>P(X=3) = 1/8</td>
</tr>
<tr>
<td>HHH</td>
<td>3</td>
<td>P(X=3) = 1/8</td>
</tr>
</tbody>
</table>

Terminology for discrete random variables

Probability function
Probability mass function
Probability distribution
Defn. \( f(x) \) is a probability mass function for the discrete random variable \( X \) if, for each possible outcome

1. \( f(x) \geq 0 \)
2. \( \sum_{x} f(x) = 1 \)
3. \( P(X = x) = f(x) \)

Example: If the random variable \( X \) has \( n \) possible outcomes \( x_1, x_2, \ldots, x_n \), then the above rules become

1. \( f(x_i) \geq 0 \) for \( i = 1 \) to \( n \)
2. \( \sum_{i=1}^{n} f(x_i) = 1 \)
3. \( P(X = x_i) = f(x_i) \)

Example: \( X \) : Number of bits equal to 1 in an 8-bit random binary number.

The sample space has \( 2^8 = 256 \) possible binary numbers. Each are equally likely. The random variable can have values between 0 and 8. Let's compute the probabilities

\[
P(X = 0) = \frac{1}{256} \quad \text{only 1 way to have all bits OFF}
\]

\[
P(X = 1) = \frac{8}{256} \quad 8 \text{C} 1 = \frac{8!}{7! 1!} = 8 \text{ ways}
\]

\[
P(X = k) = \frac{8 \text{C} k}{256}
\]

\[
P(X = 8) = \frac{1}{256}
\]

\[
f(x) = \frac{8 \text{C} x}{256}
\]

Notice: 1. \( f(x) \geq 0 \) for 0 \( \leq x \leq 8 \)

2. \( \sum_{x} f(x) = \frac{8}{2} \times \frac{8 \text{C} k}{256} = \frac{2^8}{256} = 1 \)

3. \( P(X = x) = f(x) \)
Example continued.

What is the probability that there are less than or equal to 3 bits on?

\[
P(X \leq 3) = \sum_{k=0}^{3} P(X = k) = \sum_{k=0}^{3} \frac{3}{2^5} C_k
\]

\[
\overline{\text{This is like an event}}
\]

\[
\overline{\text{These are the probabilities of outcomes in the event.}}
\]

Definition: The cumulative distribution function \( F(x) \) of a discrete random variable \( X \) with a probability mass function \( f(x) \) is

\[
F(x) = P(X \leq x) = \sum_{t \leq x} f(t)
\]

Example: Number of heads in 3 tosses of a fair coin.

\[
f(0) = \frac{1}{8} \quad f(1) = \frac{3}{8} \quad f(2) = \frac{3}{8} \quad f(3) = \frac{1}{8}
\]

\[
F(0) = \frac{1}{8} \quad F(1) = \frac{4}{8} \quad F(2) = \frac{7}{8} \quad F(3) = \frac{8}{8}
\]

* \( F(x) \) is monotonically increasing since \( f(x) > 0 \)
* \( F(-\infty) = 0 \)
* \( F(\infty) = 1 \)
Continuous Probability Distributions

A continuous random variable has a probability of zero of assuming exactly any one of its values!

\[ P(X = 164 \text{ cm}) = 0 \]

Why? Between any two values, say 163.5 cm and 164.5 cm, or even 163.999 and 164.001, there are an infinite number of possible height (x) values. The probability of \[ x = 164.00000 \ldots \] (to arbitrary precision) is 0.

What to do? With continuous random variables we talk about the probability of \( x \) being in some interval, like \( P(a < x < b) \), rather than \( x \) assuming a precise value like \( P(X = a) \).

Note: \[ P(a < X \leq b) = P(a < x < b) + P(X = b) \]

This holds only for continuous random variables.

Defn. The function \( f(x) \) is a probability density function for the continuous random variable \( X \) if:

1. \( f(x) \geq 0 \) for all \( x \in \mathbb{R} \) (all real numbers)
2. \( \int_{-\infty}^{\infty} f(x)dx = 1 \)
3. \( P(a < X < b) = \int_{a}^{b} f(x)dx \)

Defn. The cumulative distribution function \( F(x) \) of a continuous random variable \( X \) with density function \( f(x) \) is \( F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t) dt \).
Example: Height of a random person over 21

\[ f(x) \]

\[ f(170) \]

This is not a probability value! It is a density value.

\[ \text{Shaded area is } \int_{170}^{180} f(x) \, dx \]

\[ F(x) \]

\[ F(\infty) = 1 \]

\[ P(X < 180) \]

\[ P(X < 170) \]

\[ P(170 < X < 180) = \int_{170}^{180} f(x) \, dx = F(180) - F(170) = \frac{1}{180} \int_{-\infty}^{180} f(x) \, dx - \frac{1}{180} \int_{-\infty}^{170} f(x) \, dx \]

* \[ P(a < X < b) = F(b) - F(a) \]

* From the definition of \( F(x) \) we have

\[ f(x) = \frac{dF(x)}{dx} \]

\( f(x) \) is the derivative of \( F(x) \)

* \[ P(X = a) = \int_{a}^{a} f(x) \, dx = F(a) - F(a) = 0 \]
Example: A random variable $X$ has the probability density function

$$f(x) = \begin{cases} 
\frac{x^2}{3}, & -1 < x < 2 \\
0, & \text{elsewhere}
\end{cases}$$

a) Let's sketch $f(x)$

Starting (at first) fact:

$$f(2) = \frac{4}{3} > 1.$$ 

This is not ok for a density function.

Not ok. If this were a probability mass function of a discrete random variable.

$$\int_{-\infty}^{\infty} f(x) = \int_{-1}^{2} \frac{x^2}{3} \, dx = \frac{x^3}{9} \bigg|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$$

Also $f(x) \geq 0$ for all $x \in \mathbb{R}$ so this is a valid probability density function.

c) What is the probability that $X$ is between 0 and 1?

Directly from density function:

$$P(0 < X < 1) = \int_{0}^{1} \frac{x^2}{3} \, dx = \frac{x^3}{9} \bigg|_{0}^{1} = \frac{1}{9}$$
Via the cumulative distribution function $F(x)$:

$$F(x) = \int_{-\infty}^{x} f(t) \, dt$$

**Case 1** $x < -1$

$$F(x) = \int_{-\infty}^{x} 0 \, dt = 0$$

**Case 2** $-1 \leq x \leq 2$

$$F(x) = \int_{-\infty}^{-1} 0 \, dt + \int_{-1}^{x} \frac{t^2}{3} \, dt$$

$$= \frac{t^3}{9} \bigg|_{-1}^{x} = \frac{x^3 + 1}{9}$$

**Case 3** $x > 2$

$$F(x) = \int_{-\infty}^{-1} 0 \, dt + \int_{-1}^{2} \frac{t^2}{3} \, dt + \int_{2}^{x} 0 \, dt$$

$$= \frac{t^3}{9} \bigg|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$$

So,

$$F(x) = \begin{cases} 
0, & x < -1 \\
\frac{x^3 + 1}{9}, & -1 \leq x \leq 2 \\
1, & x > 2
\end{cases}$$

* $F(x)$ is monotonically increasing always.

**Example 3.21** From textbook 3.5
Exercise 3.5 (textbook)

a) \( f(x) = c \left( x^2 + 4 \right) \) \( x = 0, 1, 2, 3 \)

Find the value of \( c \) that makes \( f(x) \) a valid probability mass function (discrete).

Solution: We require \( \sum_{x=0}^{3} c(x^2 + 4) = 1 \)

\[
4c + 5c + 8c + 13c = 1
\]

\[
30c = 1 \implies c = \frac{1}{30}
\]

b) \( f(x) = c \left( \frac{2}{x} \right) \left( \frac{3}{3-x} \right) \) for \( x = 0, 1, 2 \)

Again, \( \sum_{x=0}^{2} f^2(x) = 1 \)

\[
c \left[ \left( \frac{2}{0} \right) \left( \frac{3}{3} \right) + \left( \frac{2}{1} \right) \left( \frac{3}{2} \right) + \left( \frac{2}{2} \right) \left( \frac{3}{1} \right) \right] = 1
\]

\[
c \left[ 1 + 6 + 3 \right] = 1
\]

\[
10c = 1 \implies c = \frac{1}{10}
\]
Exercise 3.6 (textbook)

\[ f(x) = \begin{cases} \frac{20000}{(x+100)^3} & x > 0 \\ 0 & \text{elsewhere} \end{cases} \]

Shelf life of medicine
Continuous r.v.

a) \[ P(X > 200) = ? \]

\[ P(X > 200) = \int_{200}^{\infty} \frac{20000}{(x+100)^3} \, dx = \left. -\frac{10000}{(x+100)^2} \right|_{200}^{\infty} = \frac{10000}{300^2} = \frac{1}{9} \]

b) \[ P(80 < X < 200) = \int_{80}^{200} \frac{20000}{(x+100)^3} \, dx \]

\[ = \left. -\frac{10000}{(x+100)^2} \right|_{80}^{120} = 0.102 \]

* \[ P(X \leq 200) = ? \]

\[ P(X \leq 200) = 1 - P(X > 200) = 1 - \frac{1}{9} = \frac{8}{9} \]
Exercise 3.12 (textbook)

\[ F(t) = \begin{cases} 
0, & t < 1 \\
1/4, & 1 \leq t < 3 \\
1/2, & 3 \leq t < 5 \\
3/4, & 5 \leq t < 7 \\
1, & t \geq 7 
\end{cases} \]

Notice a) this is a discrete random variable because \( t \) is an integer number of years and b) the cumulative distribution function is given.

a) \( P(T=5) = F(5) - F(4) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2} \)

b) \( P(T > 3) = 1 - P(T \leq 3) = 1 - F(3) = 1 - \frac{1}{2} = \frac{1}{2} \)

c) \( P(1.4 < T < 6) = F(6) - F(1.4) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2} \)

\( F(1.4) = 1/4 \) because \( F(t) \) is like a step function. For all \( t \) values between 1 and 3, it takes on the value \( 1/4 \).
Exercise 3.21 (textbook)

\[ f(x) = \begin{cases} 
  k \sqrt{x}, & 0 < x < 1 \\
  0, & \text{elsewhere} 
\end{cases} \]

Continuous prob. density function

a) Evaluate k.

\[
\int_{-\infty}^{\infty} f(x) \, dx = 1 \quad \text{since } f(x) \text{ is a probability density function}
\]

\[
\int_{-\infty}^{\infty} f(x) \, dx = \int_{0}^{\infty} k \sqrt{x} \, dx = \frac{2k}{3} x^{3/2} \bigg|_{0}^{\infty}
\]

\[
= \frac{2k}{3} = 1
\]

Therefore \( k = \frac{3}{2} \)

b) \( F(x) = \int_{-\infty}^{x} f(t) \, dt \)

for \( x < 0 \) \( F(x) = 0 \)

for \( x > 1 \) \( F(x) = 1 \)

for \( 0 < x < 1 \) \( F(x) = \int_{0}^{x} \frac{1}{3} \sqrt{t} \, dt = \left[ \frac{2}{9} t^{3/2} \right]_{0}^{x} = \frac{2}{9} x^{3/2} \)

\[ F(x) = \begin{cases} 
  0, & x < 0 \\
  \frac{2}{9} x^{3/2}, & 0 \leq x \leq 1 \\
  1, & x \geq 1 
\end{cases} \]

\( P(0.3 < X < 0.6) = F(0.6) - F(0.3) = 0.3004 \)