# 1 5.1 Random Variables

### 5.1 Random Variables

Roll a pair of dice. What is the probability of getting more than 3 but less than 10?

Carry out a census of Ireland. What is the probability that a household chosen at random has more than 2 children?

Test a 100W light bulb. What is the probability that it will last for less than six months?

A convenient way to answer questions like these is to use probability densities of random variables.

### 5.2 Examples of random variables.

$X =$ sum of numbers on pair of dice.

$X =$ number of children in a given household.

$X =$ lifetime of a light bulb.

In each case, $X$ has these two important properties:

(1) $X$ can take a number of different values; the value that $X$ does take is determined by doing the experiment.

(2) Probability that $X$ has a particular value may be determined.

Look at the first experiment in detail. The sample space is

$$S = \{(1,1), (1,2), \ldots, (3,5), \ldots, (6,6)\}.$$

Consider the random variable $X =$ sum of the two numbers rolled.

Look carefully at how $X$ is calculated:

| $INPUT$ | | | | $OUTPUT$ |
|---|---|---|---|---|
| $(2,3)$ | $\rightarrow$ | $X =$ sum of numbers | $\rightarrow$ | 5 |
| $(6,1)$ | $\rightarrow$ | $X =$ sum of numbers | $\rightarrow$ | 7 |
| $(2,5)$ | $\rightarrow$ | $X =$ sum of numbers | $\rightarrow$ | |

Notice that we can determine

$P(X = 4)$ = "probability that $X = 4$" = $\frac{3}{36}$.
$P(X = 7)$ = "probability that $X = 7$" = $\frac{6}{36}$.
$P(X = 19)$ = "probability that $X = 19$" = $\frac{0}{36}$.

**5.3 Random variable $X$.**

$X$ is a function with inputs from the sample space of an experiment and outputs which are real numbers.

For any number $a$, we can calculate $P(X = a)$ = "Probability that $X = a$".

Another example: $Y =$ product of two numbers on a pair of dice.

**Question**
Suppose that

$$\begin{array}{rcl} A = & \{1,2,3,4,5\} & \text{Finite} \\ B = & \{1,2,3,4,\dots\} & \text{Countably infinite} \\ C = & \{2,4,6,8,\dots\} & \text{Countably infinite} \\ D = & \{x \in \mathbb{R} : 1 \leq x \leq 5\} & \text{Uncountably infinite} \end{array}$$

Discuss the ranking of these sets in terms of the number of elements that they each contain.
Count

**Finite and infinite sets.**

The set $A$ is *finite* if it contains a finite number of elements.

$A$ is *countably infinite* if its elements can be put in a $1-1$ correspondence with the set $\mathbb{N} = \{1, 2, 3, \dots\}$, i.e. if the elements of $A$ can be counted.

$A$ is *uncountably infinite* if its elements cannot be counted.

Example1

**5.4 Discrete random variables**
A random variable $X$ is called *discrete* if the number of different values of $X$ (i.e. the *range* $R_X$ of $X$) is either finite or countably infinite. Examples are

$X_1 =$ product of numbers on a pair of dice. $X_2 =$ number of drops of rain falling in Dublin today.

In these cases, we can list the possible output values of the random variable:

$$R_{X_1} = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 14, ..., 30, 36\}$$

$$R_{X_2} = \{0(?), 1, 2, 3, \dots\}$$

The most convenient way of dealing with probabilities like $P(X_1 = 16)$ is using probability distributions.

**5.5 Probability distributions.**
Let $X$ be a discrete r.v. and $R_X$ its range. With each $x_i \in R_X$ we associate a number

$$p(x_i) = P(X = x_i)$$

called the probability of $x_i$. The numbers $p(x_i)$ satisfy

(i) $p(x_i) \geq 0$ for all $i$.

(ii) $\sum_{i=1}^{\infty} p(x_i) = p(x_1) + p(x_2) + \cdots = 1$.

This function $p$ is called the *probability distribution function* of the discrete random variable $X$. It is sometimes useful to draw the graph of $p(x)$.

**5.6 Example**
Toss two coins. The sample space is

$$S = \{(H, H), (H, T), (T, H), (T, T)\}.$$

Let $X =$ number of tails obtained. Then

$$\begin{array}{llll} X((H,H)) & = 0 & X((H,T)) = 1 \\ X((T,H)) & = 1 & X((T,T)) = 2 \end{array}$$

Thus the range of $X$ is

$$R_X = \{0, 1, 2\}.$$

We can write down the probabilities:

$$\begin{array}{lll} p(0) & = & P(X = 0) = P(\text{no tails}) = \frac{1}{4} \\ p(1) & = & P(X = 1) = P(\text{one tail}) = \frac{1}{2} \\ p(2) & = & P(X = 2) = P(\text{two tails}) = \frac{1}{4} \end{array}$$

Can check that this $p(x)$ satisfied the requirements of the definition.

**5.7 Continuous random variables**
A continuous random variable $X$ is a r.v. whose range $R_X$ is uncountably infinite. For example, $X =$ lifetime of a bulb has

$$R_X = \{x : x \geq 0\}.$$

Every continuous random variable has associated with it a *probability density function* (pdf) $f$ satisfying

(i) $f(x) \geq 0$ for all $x$.

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

**5.8 Cumulative distribution functions (cdf)**
Given a random variable $X$ (discrete or continuous) we define the cumulative distribution function $F$ of $X$ by

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

This has the useful property that

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a). \end{aligned}$$

**Two important facts:**

For a discrete random variable $X$,

$$F(x) = \sum_{x_j \leq x} p(x_j).$$

For a continuous random variable with pdf $f$,

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

**5.9 Examples**

(i) Consider the random variable $X$ with range $R_X = \{0, 1, 2\}$ and with

| $x_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $p(x_i)$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{2}$ |

Is $X$ discrete or continuous?

Verify that $p(x)$ is a valid probability distribution and sketch its graph.

Find the cdf of $X$.

(ii) Given the (candidate) pdf

$$f(x) = \begin{cases} 2x, & \text{if } 0 < x < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

Is $X$ discrete or continuous?

Verify that $f$ is a valid pdf and sketch its graph.

Calculate the cdf and sketch its graph.

Calculate $P(X \leq \frac{1}{2})$.

(iii) Given the (candidate) pdf

$$f(x) = \begin{cases} \dfrac{9 - x^2}{k}, & \text{if } -3 < x < 3; \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the value of $k$ for which $f$ is a valid pdf.

(b) Calculate the cdf and sketch its graph.

(a) If $f$ is a pdf,then

$$\begin{aligned} 1 = & \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{-3} f(x)dx + \int_{-3}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx \\ & = \int_{-\infty}^{-3} 0\,dx + \int_{-3}^{3} \frac{9-x^2}{k}dx + \int_{3}^{\infty} 0\,dx \\ & = \frac{1}{k}\left[9x - \frac{x^3}{3}\right]_{-3}^{3} = \frac{1}{k}(27 - 9 - (-27+9)) = \frac{36}{k} \\ \Longrightarrow & \ k = 36. \end{aligned}$$

(b)

If $x \leq -3$, then

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

If $-3 < x < 3$, then

$$\begin{aligned} F(x) & = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-3} 0\,dt + \frac{1}{36}\int_{-3}^{x} 9 - x^2\,dt \\ & = \frac{1}{36}\left[9t - \frac{t^3}{3}\right]_{-3}^{x} = \frac{x}{4} - \frac{x^3}{108} + \frac{1}{2}. \end{aligned}$$

If $3 \leq x$, then

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-3} 0\,dt + \frac{1}{36}\int_{-3}^{3} 9 - x^2\,dt + \int_{3}^{x} 0\,dt = 1.$$

Therefore

$$F(x) = \begin{cases} 0, & \text{if } x \leq -3; \\ \dfrac{x}{4} - \dfrac{x^3}{108} + \dfrac{1}{2}, & \text{if } -3 < x < 3; \\ 1, & \text{if } 3 \leq x \end{cases}$$

# 2 5.2 Characteristics of Random Variables

### 5.10 Characteristics of Random Variables

Consider these random variables...

$X$ = sum of numbers on two dice;

$X$ = total number of correct guesses in a multiple choice test of 10 questions each with 5 suggested answers;

$X$ = lifetime of a light bulb.

...and these related questions.

What is the average value of each $X$?

How spread out around this average are the other values of $X$?

The answer to these questions for a given random variable gives us a lot of useful information. We use objects called the *expectation* and the *variance* to get a rigourously defined notion of what we mean by 'average' and 'spread' respectively. (What *should* the definitions be? Consider a 3-sided dice...)

**5.11 Expectation**

Suppose that $X$ is a discrete random variable with range $R_X = \{x_1, x_2, \ldots, x_n\}$ and probability distribution $p$. Then the expectation (or expected value or mean or average) of $X$ is defined to be

$$E(X) = \sum_{i=1}^{n} x_i p(x_i) = x_1 p(x_1) + x_2 p(x_2) + \cdots + x_n p(x_n).$$

For a continuous random variable $X$ with probability density function $f$, the expectation is

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

**5.12 Examples**

(i) Calculate $E(X)$ for the random variable $X$ with range $R_X = \{0, 1, 2\}$ and with

| $x_i$ | 0 | 1 | 2 |
|---|---|---|---|
| $p(x_i)$ | $\frac{1}{3}$ | $\frac{1}{6}$ | $\frac{1}{2}$ |

(ii) Calculate $E(X)$ for the random variable $X$ with pdf

$$f(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Var

**5.13 Properties of $E(X)$**

(i) If $X = c$ constant, then $E(X) = c$.

(ii) If $X$ is a random variable and $c$ is a number, then $E(cX) = cE(X)$.

(iii) If $X$ is a random variable and $a, b$ are numbers, then $E(aX + b) = aE(X) + b$.

(iv) If $X$ is a random variable and $g$ is a function from $\mathbb{R}$ to $\mathbb{R}$, then $g(X)$ is a random variable. The expectation $E(g(X))$ is calculated as follows. If $X$ is discrete with probability distribution $p$, then

$$E(g(X)) = \sum_{i=1}^{\infty} g(x_i)p(x_i).$$

If $X$ is continuous with probability density function $f$, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx.$$

(v) If $X$ is a random variable, $g_1, \ldots, g_m$ are functions and $c_1, \ldots, c_m$ are numbers, then

$$E(c_1 g_1(X) + \cdots + c_m g_m(X)) = \\ c_1 E(g_1(X)) + \cdots + c_m E(g_m(X)).$$

**5.14 Variance**

Want to measure something like: "the average distance of the values of $X$ from the average value $E(X)$".

$X - E(X)$: random variable which measures displacement of $X$ from the expectation (average) $E(X)$.

$(X - E(X))^2$: r.v. which measures *distance*...

$E[(X - E(X))^2]$: expectation (i.e. average) of distance from the expectation.

So we define the *variance* $V(X)$ of the random variable $X$ by

$$V(X) = E[(X - E(X))^2].$$

**Important and useful fact**

$$V(X) = E(X^2) - [E(X)]^2.$$

This gives us a simple way of calculating the variance:
If $X$ is a discrete random variable, then

$$V(X) = \sum x_i^2 p(x_i) - \left[\sum x_i p(x_i)\right]^2.$$

If $X$ is a continuous random variable, then

$$V(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \left[\int_{-\infty}^{\infty} x f(x)\, dx\right]^2.$$

**5.15 Examples**
Calculate the variance for the random variables in the previous examples.
Example

**Properties of $V(X)$**

(i) If $X = c$ constant, then $V(X) = 0$.

(ii) If $X$ is a random variable and $c$ is a number, then $V(cX) = c^2 V(X)$.

(iii) If $X$ is a random variable and $a, b$ are numbers, then $V(aX + b) = a^2 V(X)$.