Two

Random Variables

1 Definition of a Random Variable

Often, when an experiment is performed, we are interested in some function (numerical characteristic) of the outcome, rather than the actual outcome itself.

For example,

- in an experiment involving the examination of 100 electronic components, our interest is in the number of defective components.
- in an experiment of flipping a coin 100 times, our interest is in the number of heads obtained, instead of the "H" and "T" sequence.

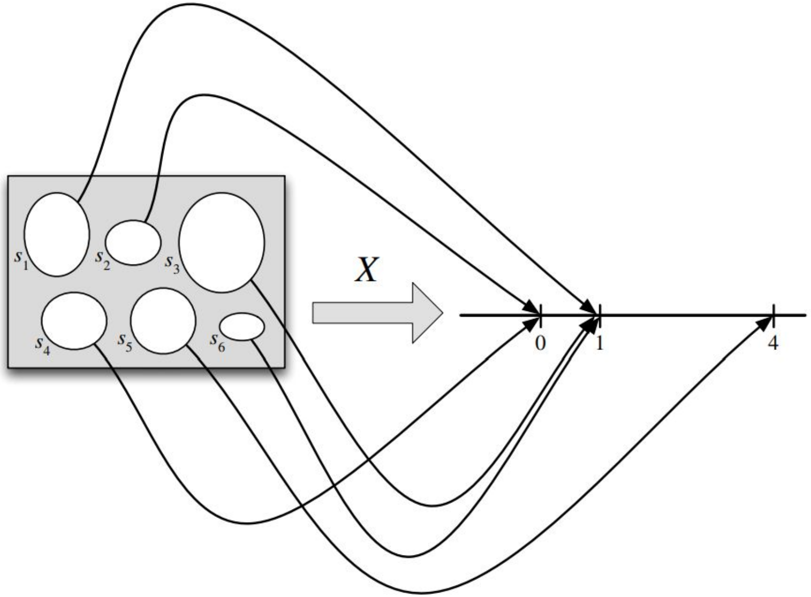
This motivates us to assign numerical values to (possible) outcomes of an experiment.

Definition 1 (Random Variable)
Let S be the sample space of an experiment. A function X, which assigns a real number to every $s \in S$ is called a ***random variable***.

Remark

So a random variable X is a function from S to $\mathbb{R}$:

$$X : S \mapsto \mathbb{R}.$$



A random variable maps the sample space into the real line. The random variable X depicted here is defined on a sample space with 6 elements $\{s_1, s_2, \ldots, s_6\}$, and has possible values 0, 1, and 4.

Definition 2 (Range space)
*The **range space** of X is the set of real numbers*

$$R_X = \{x | x = X(s), s \in S\}.$$

Each possible value x of X corresponds to an event that is a subset or element of the sample space S.

EXAMPLE 2.1

Let $S = \{HH, HT, TH, TT\}$ be the sample space associated with the experiment of flipping two fair coins.

Define the random variable:

$$X = \text{number of heads obtained.}$$

Note that X is a *function* from S to $\mathbb{R}$, the set of real numbers:

$$X(HH) = 2, \quad X(HT) = X(TH) = 1, \quad X(TT) = 0.$$

The range of X is $R_X = \{0, 1, 2\}$.

NOTATIONS

(i) We use upper case letters $X, Y, Z, X_1, X_2, \ldots$ to denote **random variables**.

(ii) We use lower case letters $x, y, z, x_1, x_2, \ldots$ to denote their **observed values** in the experiment.

(iii) The set $\{X = x\} = \{s \in S : X(s) = x\}$ is a subset of S.

(iv) If A is a subset of $\mathbb{R}$, the set $\{X \in A\} = \{s \in S : X(s) \in A\}$ is a subset of S.

(v) With the above expressions, we define $P(X = x)$ and $P(X \in A)$ as

$$\begin{aligned} P(X = x) &= P(\{s \in S : X(s) = x\}); \\ P(X \in A) &= P(\{s \in S : X(s) \in A\}). \end{aligned}$$

EXAMPLE 2.2

We revisit Example 2.1, where $S = \{HH, HT, TH, TT\}$ is the sample space associated with flipping two fair coins, and X is the number of heads obtained.

We then have

$$\{X = 0\} = \{TT\}; \qquad \{X = 1\} = \{HT, TH\};$$
$$\{X = 2\} = \{HH\}; \qquad \{X \geq 1\} = \{HT, TH, HH\}.$$

Thus

$$P(X=0)=P(TT)=1/4; \qquad P(X=1)=P(\{HT,TH\})=2/4;$$
$$P(X=2)=P(HH)=1/4; \qquad P(X\geq 1)=P(\{HT,TH,HH\})=3/4.$$

We can then summarize the probabilities of the random variable X using a table:

x	0	1	2
$P(X=x)$	1/4	1/2	1/4

2 Probability Distribution

There are two main types of random variables used in practice: discrete and continuous.

Let us denote by X the random variable, and it's range by R_X. For a

- **discrete random variable**, the number of values in R_X is **finite** or **countable**. That is, we can write $R_X = \{x_1, x_2, x_3, \ldots\}$.
- **continuous random variable**, R_X is an **interval** or a **collection of intervals**.

Discrete random variable

Consider a discrete random variable X with $R_X = \{x_1, x_2, x_3, \ldots\}$.

For each $x \in R_X$, let $P(X = x)$ be the probability that X takes the value x.

Definition 3 (Probability mass function)
For a discrete random variable X, define

$$f(x) = \begin{cases} P(X = x), & \text{for } x \in R_X; \\ 0, & \text{for } x \notin R_X. \end{cases}$$

Then $f(x)$ is known as the ***probability function (pf)****, or* ***probability mass function (pmf)*** *of X.*

The collection of pairs $(x_i, f(x_i)), i = 1, 2, 3, \ldots,$ is called the ***probability distribution*** *of X.*

Properties of the Probability mass function

The probability mass function $f(x)$ of a discrete random variable **must** satisfy:

(1) $f(x_i) \geq 0$ for all $x_i \in R_X$;

(2) $f(x) = 0$ for all $x \notin R_X$;

(3) $\sum_{i=1}^{\infty} f(x_i) = 1$, or $\sum_{x_i \in R_X} f(x_i) = 1$.

For any set $B \subset \mathbb{R}$, we have

$$P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i).$$

EXAMPLE 2.3

We revisit Examples 2.1 and 2.2.

Recall that random variable X is the number of heads observed when we flip two fair coins.

The probability function of X is given below

x	0	1	2
$f(x)$	1/4	1/2	1/4

Note that $f(x)$ satisfies

(1) $f(x_i) \geq 0$ for $x_i = 0, 1,$ or 2;

(2) $f(x) = 0$ for $x \notin R_X$;

(3) $f(0) + f(1) + f(2) = 1$.

When $B = [1, \infty)$,

$$P(X \in B) = f(1) + f(2) = 3/4.$$

Continuous random variable

For a continuous random variable X, R_X is an interval or a collection of intervals.

We next define the **probability function (pf)**, or **probability density function (pdf)**, to quantify the probability that X is in a certain range.

DEFINITION 4 (PROBABILITY DENSITY FUNCTION) *The **probability density function** of a continuous random variable X, denoted by $f(x)$, is a function that satisfies:*

(1) $f(x) \geq 0$ *for all* $x \in R_X$; *and* $f(x) = 0$ *for* $x \notin R_X$;

(2) $\int_{R_X} f(x)\, \mathrm{dx} = 1$;

(3) *For any* a *and* b *such that* $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\, \mathrm{dx}\,.$$

Remark

- Note that Condition (2) is equivalent to

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

  since $f(x) = 0$ for $x \notin R_X$.

- For any specific value x_0, we have

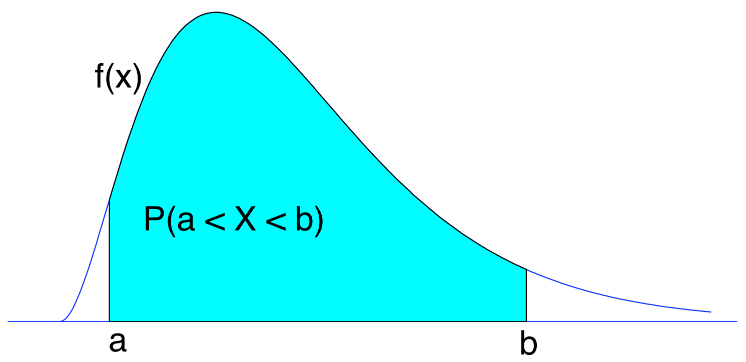
$$P(X = x_0) = \int_{x_0}^{x_0} f(x)\,\mathrm{dx} = 0.$$

  The gives an example of "$P(A) = 0$, but A is not necessarily $\emptyset$."

- Furthermore, we have

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{dx}\,.$$

They all represent the area under the graph of $f(x)$ between $x = a$ and $x = b$.



- To check that a function $f(x)$ is a probability density function, it suffices to check Conditions (1) and (2). Namely,

  (1) $f(x) \geq 0$ for all $x \in R_X$; and $f(x) = 0$ for $x \notin R_X$.

  (2) $\displaystyle\int_{R_X} f(x)\,\mathrm{dx} = 1.$

EXAMPLE 2.4

Let X be a continuous random variable with probability density function given by

$$f(x) = \begin{cases} cx, & \text{for } 0 < x < 1; \\ 0, & \text{otherwise }. \end{cases}$$

(i) Find the value of c;

(ii) Find $P(X \leq 1/2)$.

Solution:

(i) Since

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{dx} = \int_0^1 cx\,\mathrm{dx} = c \cdot \frac{x^2}{2}\bigg|_0^1 = c/2,$$

we set $c/2 = 1$. This results in $c = 2$.

(ii)

$$P(X \leq 1/2) = \int_{-\infty}^{1/2} f(x)\,\mathrm{dx} = \int_0^{1/2} 2x\,\mathrm{dx} = 1/4.$$

3 Cumulative Distribution Function

Another function that describes the distribution of a random variable is the **cumulative distribution function (cdf)**.

Definition 5 (Cumulative Distribution Function)
*For any random variable X, we define its **cumulative distribution function (cdf)** by*

$$F(x) = P(X \leq x).$$

Remark

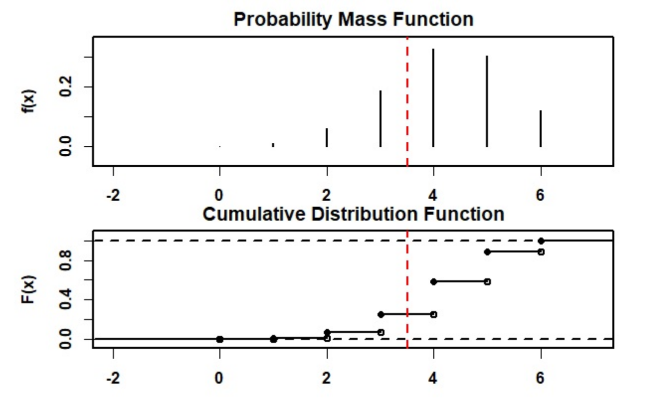
This definition is applicable whether X is a discrete or continuous random variable.

CDF: Discrete random variable

If X is a discrete random variable, we have

$$F(x) = \sum_{t \in R_X; t \leq x} f(t)$$

$$= \sum_{t \in R_X; t \leq x} P(X = t)$$



The cumulative distribution function of a discrete random variable is a step function.

For any two numbers $a < b$, we have

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a-),$$

where "$a-$" represents the "largest value in R_X that is smaller than a". Mathematically,

$$F(a-) = \lim_{x \uparrow a} F(x).$$

EXAMPLE 2.5

We revisit Examples 2.1 and 2.2. The random variable X is the number of heads observed when we flip two fair coins, and has the probability function

x	0	1	2
$f(x)$	1/4	1/2	1/4

We have

$$F(0) = f(0) = 1/4,\ F(1) = f(0) + f(1) = 3/4,\ F(2) = f(0) + f(1) + f(2) = 1.$$

We can therefore obtain the cumulative distribution function:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}.$$

EXAMPLE 2.6

Consider the cumulative distribution function from Example 2.5:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/4, & 0 \leq x < 1 \\ 3/4, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}.$$

Derive the corresponding probability function. [1]

[1] Let us pretend for a while that the cumulative distribution function is the only information available for this distribution.

Solution:
As $F(\cdot)$ only has four possible values, the distribution is a discrete distribution.

We obtain $R_X = \{0, 1, 2\}$, which are the jumping points of $F(\cdot)$. It is also the set where $f(x)$ is non-zero.

We have

$$\begin{aligned} f(0) &= P(X=0) = F(0) - F(0-) = 1/4 - 0 = 1/4; \\ f(1) &= P(X=1) = F(1) - F(1-) = 3/4 - 1/4 = 1/2; \\ f(2) &= P(X=2) = F(2) - F(2-) = 1 - 3/4 = 1/4. \end{aligned}$$

CDF: Continuous random variable

If X is a continuous random variable,

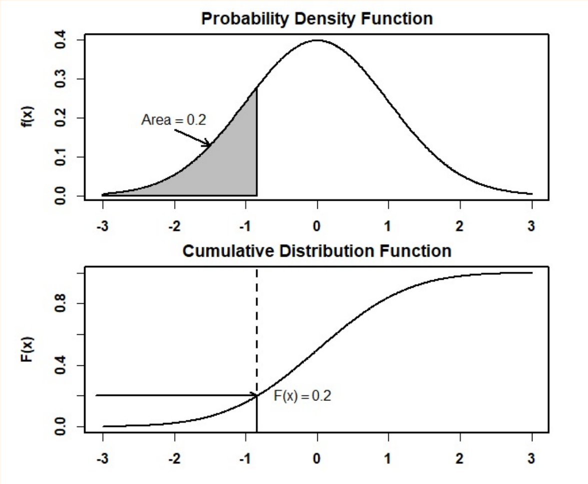
$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

and

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{dx}}.$$

Further

$$P(a \le X \le b) = P(a < X < b) = F(b) - F(a).$$



EXAMPLE 2.7

Suppose the probability density function of a random variable X is

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

The cumulative distribution function of X is then

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x < 1 \\ 1, & 1 \leq x \end{cases}.$$

EXAMPLE 2.8

Consider the cumulative distribution function from Example 2.7:

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x < 1 \\ 1, & 1 \leq x \end{cases} .$$

Derive the corresponding probability function. [2]

[2] Let us pretend for a while that the cumulative distribution function is the only information available for this distribution.

Solution:
$F(x)$ is a cumulative distribution function for a continuous distribution, because when it is not equal to 0 and 1, it assumes different values in the interval $x \in [0, 1)$.

$f(x) = 0$ when $x \notin [0, 1)$ because $\frac{\mathrm{d}}{\mathrm{dx}}(0) = \frac{\mathrm{d}}{\mathrm{dx}}(1) = 0$.

$f(x) = \frac{\mathrm{d}}{\mathrm{dx}}(x^2) = 2x$ when $x \in [0, 1)$.

Remark

(i) No matter if X is discrete or continuous, $F(x)$ is non-decreasing. In the sense that for any $x_1 < x_2$, $F(x_1) \leq F(x_2)$.

(ii) The probability function and cumulative distribution function have a one-to-one correspondence. That is, for any probability function given, the cumulative distribution function is uniquely determined; and vice versa.

(iii) The ranges of $F(x)$ and $f(x)$ satisfy:

- $0 \leq F(x) \leq 1$;
- for discrete distributions, $0 \leq f(x) \leq 1$;
- for continuous distributions, $f(x) \geq 0$, but **not necessary** that $f(x) \leq 1$.

4 Expectation and Variance

For a random variable X, one natural question to ask is: what is the **average value** of X, if the corresponding experiment is repeated many times?

For example, suppose X is the number obtained when we roll a die. We may want to know the average value obtained if we roll the die continuously.

Such an average, over the long run, is called the **mean** or **expectation** of X.

Definition 6 (Expectation: Discrete random variable)
Let X be a discrete random variable with $R_X = \{x_1, x_2, x_3, \ldots.\}$ and probability function $f(x)$. The ***expectation*** *or* ***mean*** *of X is defined by*

$$E(X) = \sum_{x_i \in R_X} x_i f(x_i).$$

By convention, we also denote $\mu_X = E(X)$.

Definition 7 (Expectation: Continuous random variable) *Let X be a continuous random variable with probability function $f(x)$. The* ***expectation*** *or* ***mean*** *of X is defined by*

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

REMARK

The mean of X is *not necessarily* a possible value of the random variable X.

EXAMPLE 2.9

Suppose we toss a fair die and the upper face is recorded as X. We have

$$P(X = k) = 1/6, \quad \text{for } k = 1, 2, \ldots, 6,$$

and

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5.$$

Here we have a random variable whose mean is not a value that X assumes.

EXAMPLE 2.10

The probability density function of weekly gravel sales X is

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

We then have

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{dx} = \int_0^1 \frac{3x}{2}(1-x^2)\,\mathrm{dx} \\ &= \frac{3}{2}\int_0^1 (x-x^3)\,\mathrm{dx} = \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Big|_0^1 = 3/8. \end{aligned}$$

Properties of Expectation

(1) Let X be a random variable, and let a and b be any real numbers. Then

$$E(aX+b) = aE(X)+b.$$

(2) Let X and Y be two random variables. We have

$$E(X+Y) = E(X)+E(Y).$$

(3) Let $g(\cdot)$ be an arbitrary function.

- If X is a **discrete** random variable with probability mass function $f(x)$ and range R_X,
$$E[g(X)] = \sum_{x \in R_X} g(x) f(x).$$

- If X is a **continuous** random variable with probability density function $f(x)$ and range R_X,
$$E[g(X)] = \int_{R_X} g(x) f(x) \, \mathrm{dx}\,.$$

Variance

Let $g(x) = (x - \mu_X)^2$, then $E[g(X)]$ is defined as the **variance** for X.

Definition 8 (Variance)
*Let X be a random variable. The **variance** of X is defined as*

$$\sigma_X^2 = V(X) = E(X - \mu_X)^2.$$

REMARK

- This definition is applicable whether X is discrete or continuous.
- If X is a **discrete** random variable with probability mass function $f(x)$ and range R_X,

$$V(X) = \sum_{x \in R_X} (x - \mu_X)^2 f(x).$$

- If X is a **continuous** random variable with probability density function $f(x)$,

$$V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,\mathrm{dx}\,.$$

- $V(X) \geq 0$ for any X. Equality holds if and only $P(X = E(X)) = 1$, that is, when X is a **constant**.
- Let a and b be any real numbers, then $V(aX + b) = a^2 V(X)$.
- The variance can also be computed by an alternative formula:

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

- The positive square root of the variance is defined as the **standard deviation** of X:

$$\sigma_X = \sqrt{V(X)}.$$

EXAMPLE 2.11

Let the probability function of a random variable X be given by

x	-1	0	1	2
$f(x)$	$1/8$	$2/8$	$1/8$	$4/8$

Find $E(X)$ and $V(X)$.

Solution:

The mean is given as

$$\begin{aligned} E(X) &= \sum_{x \in R_X} x f(x) \\ &= (-1)\left(\frac{1}{8}\right) + 0\left(\frac{2}{8}\right) + 1\left(\frac{1}{8}\right) + 2\left(\frac{4}{8}\right) = 1. \end{aligned}$$

The variance is given as

$$\begin{aligned} V(X) &= \sum_{x \in R_X} [x - E(X)]^2 f(x) = \sum_{x \in R_X} [x-1]^2 f(x) \\ &= (-1-1)^2\left(\frac{1}{8}\right) + (0-1)^2\left(\frac{2}{8}\right) \\ &\quad + (1-1)^2\left(\frac{1}{8}\right) + (2-1)^2\left(\frac{4}{8}\right) = \frac{5}{4}. \end{aligned}$$

EXAMPLE 2.12

Denote by X the amount of time that a book on reserve at the library is checked out by a randomly selected student. Suppose X has the probability density function

$$f(x) = \begin{cases} x/2, & 0 \le x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Compute $E(X), V(X)$, and σ_X.

Solution:
We can compute

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{dx} = \int_0^2 x \cdot x/2\,\mathrm{dx} = \frac{x^3}{6}\bigg|_0^2 = 4/3;$$
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{dx} = \int_0^2 x^2 \cdot x/2\,\mathrm{dx} = \frac{x^4}{8}\bigg|_0^2 = 2.$$

Using $V(X) = E(X^2) - [E(X)]^2$, we obtain

$$V(X) = 2 - (4/3)^2 = 2/9 \quad \text{and} \quad \sigma_X = \sqrt{V(X)} = \sqrt{2}/3.$$