

Chapter 3

Random variables and probability distributions

Random variables and distributions: big idea

We seek a way to

- efficiently describe the sample outcomes of experiments in numerical form
- describe how probability is spread around the sample outcomes using mathematical functions.

Random variables

Up to this point we have dealt with probability assignments for events defined on a general sample space S .

We now focus on the specific case where S is a subset of the real numbers, $\mathbb{R}$. We could have S

- finite or countable $\{0, 1, 2, \dots\}$;
- uncountable.

Events in a sample space that is a subset of $\mathbb{R}$ can be termed **numerical** events.

Random variables

For most experiments defined on an arbitrary space S , we can usually consider a simple mapping, Y , defined on S that maps the sample outcomes onto $\mathbb{R}$.

The mapping Y is termed a **random variable**; it is just a way of recording the outcome in a **numerical** fashion.

Random variables

Example: Two coins

Two coins are tossed with the outcomes being independent. The sample space is the list of sequences of possible outcomes

$$S = \{HH, HT, TH, TT\}.$$

with all sequences equally likely.

Define the random variable Y by

$$Y(s) = \text{“the number of heads in sequence } s\text{”}$$

Random variables

Example: Two coins

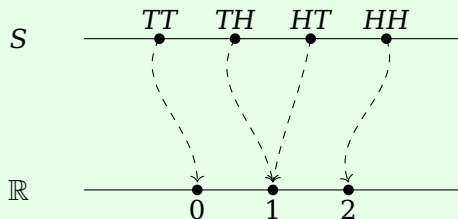
In the original probability specification for events in S ,

$$P(\{HH\}) = P(\{HT\}) = P(\{TH\}) = P(\{TT\}) = \frac{1}{4}$$

Each $s \in S$ carries its probability with it through the mapping via $Y(\cdot)$.

Random variables

Example: Two coins



Random variables

Example: Two coins

Then

$$Y(HH) = 2 \quad Y(HT) = Y(TH) = 1 \quad Y(TT) = 0$$

and $Y(\cdot)$ maps S onto the set $\{0, 1, 2\}$. We have that

$$P(Y = 0) \equiv P(s \in \{TT\}) = \frac{1}{4}$$

$$P(Y = 1) \equiv P(s \in \{HT, TH\}) = \frac{2}{4}$$

$$P(Y = 2) \equiv P(s \in \{HH\}) = \frac{1}{4}$$

with $P(Y = y) = 0$ for all other values of y .

Random variables

Note

- $Y(\cdot)$ can be a **many-to-one** mapping.
- We must consider the possible values that Y can take.
- We usually simply write

$$P(Y = y)$$

Random variables

Note

We typically use **upper case** letters

$$X, Y, Z$$

etc. to denote random variables, and **lower case** letters

$$x, y, z$$

etc. to denote the numerical values that the random variable can take.

Discrete Random variables

A random variable Y is termed **discrete** if the possible values that Y can take is a **countable** set:

$$y_1, y_2, \dots, y_n, \dots$$

i.e. we can make a list of the possible values.

Could be

- a finite set: e. g. $\{1, 2, \dots, 100\}$
- an infinite set: e.g the integers

Write the set of values that carry **positive** probability $\mathcal{Y}$.

Discrete Random variables

The probabilities assigned to the individual points in the set $\mathcal{Y}$ must carry all the probability:

$$\sum_{y \in \mathcal{Y}} P(Y = y) = 1.$$

This follows by the probability rules, as all the probability from S has been mapped into $\mathcal{Y}$.

Probability mass function

The **probability mass function** (pmf) for Y , $p(\cdot)$, is the mathematical function that records how probability is distributed.

That is,

$$p(y) = P(Y = y)$$

- If $y \in \mathcal{Y}$, $p(y) > 0$,
- otherwise $p(y) = 0$.

Probability mass function

The pmf could be defined **pointwise**: for example

$$p(y) = \begin{cases} 0.2 & y = -2 \\ 0.1 & y = 0 \\ 0.2 & y = 3 \\ 0.5 & y = 5 \\ 0 & \text{otherwise} \end{cases}$$

e.g. The probability that the random variable takes the value 5 is 0.5 etc.

Probability mass function

Alternatively it could be defined using some **functional form**:

$$p(y) = \begin{cases} \frac{y}{15} & y \in \{1, 2, 3, 4, 5\} \\ 0 & \text{otherwise} \end{cases}$$

Probability mass function

The pmf **must** satisfy two conditions: it

(i) must specify probabilities: that is

$$0 \leq p(y) \leq 1 \quad \text{for all } y$$

(ii) must sum to 1 when evaluated at points in $\mathcal{Y}$:

$$\sum_{y \in \mathcal{Y}} P(Y = y) = 1.$$

Probability mass function

Note

The collection of y values for which $p(y) > 0$ may be a countably infinite set, say

$$\mathcal{Y} = \{y_1, y_2, \dots\}.$$

In this case we would require the infinite sum

$$\sum_{i=1}^{\infty} p(y_i)$$

be equal to 1.

Expectations for discrete random variables

The **expectation** (or **expected value**) of a discrete random variable Y , denoted $E(Y)$, is defined as

$$E(Y) = \sum_{y \in \mathcal{Y}} y p(y)$$

whenever this sum is finite; if it is not finite, we say that **the expectation does not exist**.

Expectations for discrete random variables

Note

- $E(Y)$ is the **centre of mass** of the probability distribution.
- $E(Y)$ is also termed the **mean** of the distribution.
- If $\mathcal{Y} = \{y_1, y_2, \dots, y_n, \dots\}$, we have that

$$E(Y) = y_1p(y_1) + y_2p(y_2) + \cdots + y_np(y_n) + \cdots$$

Expectations for discrete random variables

Example: Expectation and Mean

Suppose that $\mathcal{Y} = \{y_1, y_2, \dots, y_n\}$, and

$$p(y) = \frac{1}{n} \quad y \in \mathcal{Y}$$

and zero otherwise. Then

$$\begin{aligned} E(Y) &= y_1 p(y_1) + y_2 p(y_2) + \cdots + y_n p(y_n) \\ &= \frac{1}{n} (y_1 + y_2 + \cdots + y_n) = \bar{y}. \end{aligned}$$

This is an example of a **discrete uniform distribution**.

Expectations for discrete random variables

The discrete uniform distribution is a distribution with **constant** $p(y)$, that is,

$$p(y) = c \quad y \in \mathcal{Y}$$

for some **finite** set $\mathcal{Y}$. It follows that

$$\frac{1}{c}$$

equals the number of elements in $\mathcal{Y}$.

Variance

We can also compute

$$E(Y^2) = \sum_{y \in \mathcal{Y}} y^2 p(y)$$

and interpret it as the expectation of Y^2 . That is

$$E(Y^2) = y_1^2 p(y_1) + y_2^2 p(y_2) + \cdots$$

the sum of the squares of the possible values, weighted by their probabilities.

Variance

The **variance** of random variable Y is denoted $\text{Var}(Y)$, and is defined by

$$\text{Var}(Y) = E[(Y - \mu)^2]$$

where $\mu = E(Y)$. The **standard deviation** of Y is

$$\text{sd}(Y) = \sqrt{\text{Var}(Y)}.$$

Variance

The variance is a measure of how “dispersed” the probability distribution is. Note that

$$\text{Var}(Y) = E[(Y - \mu)^2] = \sum_y (y - \mu)^2 p(y) \geq 0$$

as $p(y) \geq 0$ and $(y - \mu)^2 \geq 0$ for all y .

In fact, $\text{Var}(Y) > 0$ unless the distribution of Y places **all** its probability at a single value, in which case $\text{Var}(Y) = 0$.

Variance

Now

$$\begin{aligned} E[(Y - \mu)^2] &= E(Y^2 - 2\mu Y + \mu^2) \\ &= E(Y^2) - 2\mu E(Y) + \mu^2 && \text{by linearity} \\ &= E(Y^2) - 2\mu^2 + \mu^2 \\ &= E(Y^2) - \mu^2. \end{aligned}$$

so

$$\text{Var}(Y) = E(Y^2) - \mu^2$$

which gives an alternative method for calculation.

The binomial distribution

We can now consider specific pmfs that correspond to particular experimental settings.

One simple pmf is obtained if we consider an experiment with two possible outcomes, which we can refer to as 'success' and 'failure' respectively. We can define the random variable Y by

$$Y = 1 \text{ if outcome is 'success'} \quad Y = 0 \text{ if outcome is 'failure'}$$

and then consider

$$P(Y = 0) \text{ and } P(Y = 1).$$

The binomial distribution

Note however that

$$P(Y = 0) + P(Y = 1) = 1$$

so if we specify

$$P(Y = 1) = p$$

then automatically

$$P(Y = 0) = 1 - p = q$$

say.

The binomial distribution

We may then write

$$p(y) = p^y(1 - p)^{1-y} = p^y q^{1-y} \quad y \in \{0, 1\}$$

with $p(y) = 0$ for all other values of y .

This distribution is known as the **Bernoulli** distribution with parameter p .

Note

To avoid the trivial case, we typically assume that $0 < p < 1$.

The binomial distribution

This model covers many experimental settings:

- coin tosses;
- thumbtack tosses;
- pass/fail;
- cure/not cure

etc.

The binomial distribution

We have for this model that

$$\begin{aligned}\mu \equiv E(Y) &= \sum_{y=0}^1 yp(y) = (0 \times p(0) + 1 \times p(1)) \\ &= (0 \times q + 1 \times p) = p\end{aligned}$$

The binomial distribution

Similarly

$$\begin{aligned} E(Y^2) &= \sum_{i=0}^1 y^2 p(y) = (0^2 \times p(0) + 1^2 \times p(1)) \\ &= (0 \times q + 1 \times p) = p \end{aligned}$$

so therefore

$$\text{Var}(Y) = E(Y^2) - \mu^2 = p - p^2 = p(1 - p).$$

The binomial distribution

We now consider a version of the experiment where a sequence of n **identical** binary experiments with success probability p whose outcomes are **independent** is carried out.

0101011101

We consider the **total** number of successes in the sequence.

Provided $0 < p < 1$, we have that

$$\mathcal{Y} = \{0, 1, 2, \dots, n\}$$

The binomial distribution

Now if Y is the random variable that records the number of successes in the sequence, we have that

$(Y = y)$ means “ y successes in n trials”

Consider a sequence that results in y successes and $n - y$ failures: for example, if $n = 8$ and $y = 3$, we might have

10101000

01010001

11100000

etc.

The binomial distribution

All such sequences are **equally likely**, under the ‘identical and independent’ assumption. The probability of any such sequence is, for $y \in \{0, 1, 2, \dots, n\}$,

$$p^y q^{n-y}$$

that is, in the examples above

pqpqpqqq

qpqpqqqp

pppqqqqq

etc. all yield $p^3 q^5$.

The binomial distribution

However for the total number of successes, we do not care about the order of the outcomes; we must count the number of all such sequences, as the individual sequences **partition** the event of interest.

By the previous combinatorial arguments, we see that the number of such sequences is

$$C_y^n = \binom{n}{y} = \frac{n!}{y!(n-y)!}.$$

The binomial distribution

Therefore we have that

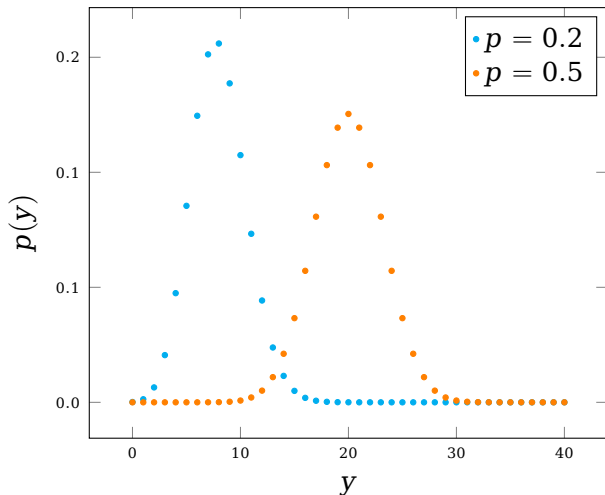
$$p(y) \equiv P(Y = y) = \binom{n}{y} p^y q^{n-y} \quad y \in \{0, 1, 2, \dots, n\}$$

and $p(y) = 0$ for all other y .

This is the pmf for the **Binomial distribution**.

The binomial distribution

Two examples: $n = 40$, with $p = 0.2$ and $p = 0.5$.



The binomial distribution

For the binomial distribution, we have that if

$$Y \sim \textit{Binomial}(n, p)$$

then

$$\text{Expectation: } E(Y) = np$$

$$\text{Variance: } \text{Var}(Y) = npq$$

The binomial distribution

Note

Recall that for a **single** binary experiment, for the total number of successes Y we have that

$$E(Y) = p \quad \text{Var}(Y) = pq$$

whereas for **n** identical and independent binary experiments, for the total number of successes Y we have that

$$E(Y) = np \quad \text{Var}(Y) = npq.$$

The binomial distribution

If $n = 1$, we have that

$$p(y) = p^y q^{1-y} \quad y \in \{0, 1\}$$

and $p(y) = 0$ for all other y , which coincides with **Bernoulli** distribution with parameter p .

The geometric distribution

We consider again identical and independent binary trials, but rather than counting the **total** number of successes, we count the number of trials carried out **until the first success**.

Using the 0/1 notation, we consider sequences such as

0001

01

000000001

1

000001

etc.

The geometric distribution

If random variable Y records the number of trials carried out up to and including the first success, then

$$p(y) = P(Y = y) = q^{y-1}p \quad y \in \{1, 2, 3, \dots\}$$

and $p(y) = 0$ for other values of y .

- $y - 1$ failures, followed by a success.
- e.g. the sequence 00001 occurs with probability

$$q \times q \times q \times q \times p = q^4 p.$$

This is the pmf of the **geometric distribution**.

The geometric distribution

Note

- Sometimes the geometric distribution is written

$$p(y) = q^y p \quad y \in \{0, 1, 2, \dots\}$$

– this is the same distribution, but for the random variable that records the **number of failures** observed up to the first success, that is, for the random variable

$$Y^* = Y - 1.$$

The geometric distribution

For the geometric distribution, we have that if

$$Y \sim \textit{Geometric}(p)$$

then

$$\text{Expectation: } E(Y) = 1/p$$

$$\text{Variance: } \text{Var}(Y) = (1 - p)/p^2$$

Y takes values on the whole number set $\{1, 2, 3, \dots\}$; this distribution does not place probability anywhere else.

Cumulative probability

We now study another way to describe how probability is distributed. Rather than

$$P(Y = y)$$

for a single value y , we consider

$$P(Y \in A)$$

for some set of values A .

Cumulative probability

For example if $A = \{2, 3, 4, 5\}$, we have

$$\begin{aligned}P(Y \in A) &= P(Y = 2 \text{ or } Y = 3 \text{ or } Y = 4 \text{ or } Y = 5) \\&= P(2 \leq Y \leq 5)\end{aligned}$$

which we can compute as

$$P(Y \in A) = P(Y = 2) + P(Y = 3) + P(Y = 4) + P(Y = 5).$$

Such probabilities are called **cumulative** probabilities.

Cumulative probability

For the geometric distribution, consider the probability

$$P(Y > y)$$

for $y \in \{1, 2, 3, \dots\}$. We see that

$$\begin{aligned} P(Y > y) &= P(Y = y + 1) + P(Y = y + 2) + \cdots \\ &= \sum_{j=y+1}^{\infty} P(Y = j). \end{aligned}$$

Cumulative probability

Using the geometric pmf

$$\begin{aligned}P(Y > y) &= \sum_{j=y+1}^{\infty} q^{j-1} p \\&= pq^y \sum_{j=y+1}^{\infty} q^{j-y-1} \quad \text{a geometric progression} \\&= pq^y \frac{1}{1-q}.\end{aligned}$$

Cumulative probability

Therefore

$$P(Y > y) = q^y \quad y \in \{1, 2, 3, \dots\}$$

and consequently, as

$$P(Y \leq y) = 1 - P(Y > y)$$

we have that

$$P(Y \leq y) = 1 - q^y = 1 - (1 - p)^y \quad y \in \{1, 2, 3, \dots\}.$$

Cumulative probability

We can define the function

$$F(y) = P(Y \leq y) = 1 - (1 - p)^y \quad y \in \{1, 2, 3, \dots\}.$$

$F(y)$ is the **cumulative distribution function** (cdf).

Cumulative probability

We can define all probabilities such as

$$P(Y \leq a) \quad P(Y \geq b) \quad P(c \leq Y \leq d)$$

etc. via $F(y)$. This is because for discrete distributions

$$P(Y \in A) = \sum_{y \in A} p(y).$$

for set A .

Cumulative probability

Note

We can compute

$$F(y) = P(Y \leq y)$$

for **any** y . However, we must remember that, e.g.

$$P(Y \leq 2.635) = P(Y = 1) + P(Y = 2) = P(Y \leq 2)$$

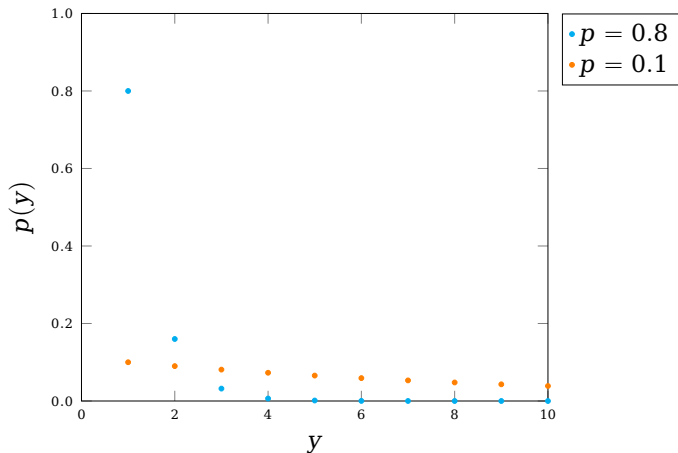
as the probability is only assigned to **whole number** values.
Therefore

$$F(2.635) = F(2) \quad F(1.234) = F(1) \quad F(4.789) = F(4)$$

etc.

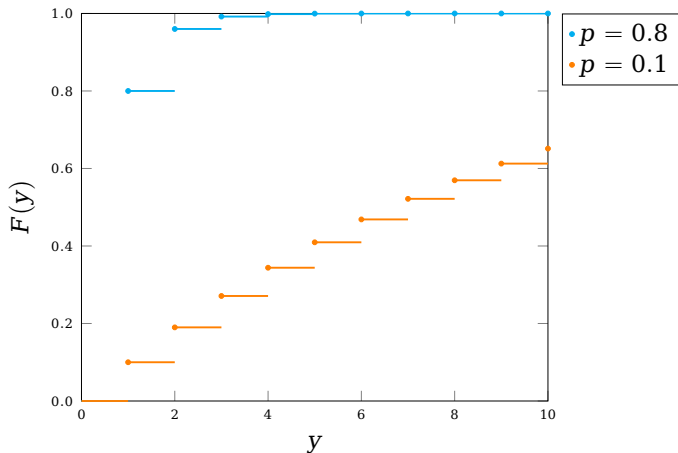
Cumulative probability

Geometric $p(y)$: with $p = 0.8$ and $p = 0.1$.



Cumulative probability

Geometric $F(y)$: with $p = 0.8$ and $p = 0.1$.



Cumulative probability

We can define the $F(y)$ function for **any** discrete distribution; for example, for the binomial distribution

$$F(y) = P(Y \leq y) = \sum_{j=0}^y \binom{n}{j} p^j q^{n-j} \quad y \in \{0, 1, 2, 3, \dots, n\}$$

but there is no simple formula for this.

Cumulative probability

Note that

$$\begin{aligned}p(y) &= P(Y = y) \\&= P(Y \leq y) - P(Y \leq y - 1) \\&= F(y) - F(y - 1)\end{aligned}$$

so that we can compute $p(y)$ from $F(y)$ and vice versa.

Cumulative probability

$F(y)$ must have the following properties:

1. 'Starts at 0'

$$F(-\infty) = P(Y \leq -\infty) = 0$$

2. 'Ends at 1'

$$F(\infty) = P(Y \leq \infty) = 1$$

3. 'Non-decreasing'

$$\text{If } y_1 < y_2 \text{ then } F(y_1) \leq F(y_2)$$

4. 'Lies in the range $[0, 1]$ '

$$0 \leq F(y) \leq 1 \quad \text{for all } y.$$

The Poisson distribution

The binomial distribution arises when we consider n identical and independent trials each with success probability p :

100010001010101010101

etc. We have

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad y \in \{0, 1, 2, \dots, n\}$$

with $p(y) = 0$ for all other values of y .

The Poisson distribution

If $n \rightarrow \infty$ with $p = \lambda/n$ for fixed λ , we have that the binomial pmf converges to

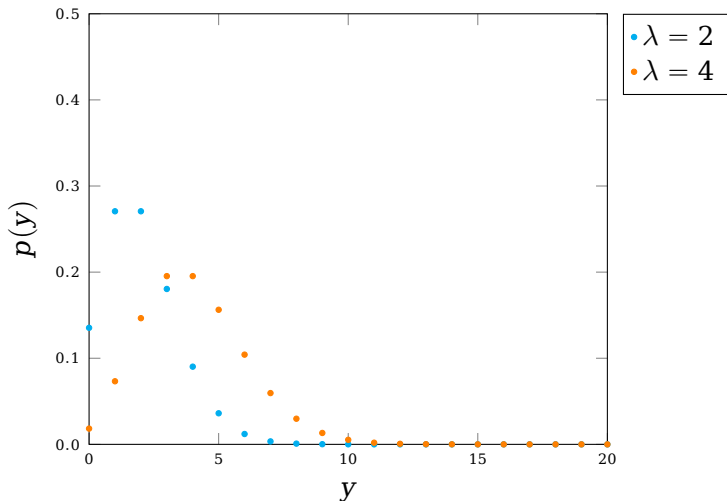
$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda} \quad y \in \{0, 1, 2, \dots\}$$

with $p(y) = 0$ for other values of y .

This is the pmf for the **Poisson distribution**.

The Poisson distribution

Poisson $p(y)$: with $\lambda = 2$ and $\lambda = 4$.



The Poisson distribution

For the Poisson distribution, we have that if

$$Y \sim \text{Poisson}(\lambda)$$

then

$$\text{Expectation: } E(Y) = \lambda$$

$$\text{Variance: } \text{Var}(Y) = \lambda$$

Y takes values on the set $\{0, 1, 2, 3, \dots\}$; this distribution does not place probability anywhere else.

The Poisson distribution

Note

The Poisson distribution is widely used to model the distribution of count data, and acts as an approximation to the binomial when

- n is large,
- p is small.

e.g.

$$\textit{Binomial}(10000, 0.002) \approx \textit{Poisson}(20)$$

Continuous Random variables

Recall the previous definition of the **cumulative distribution function** (or cdf), $F(y)$, for a random variable:

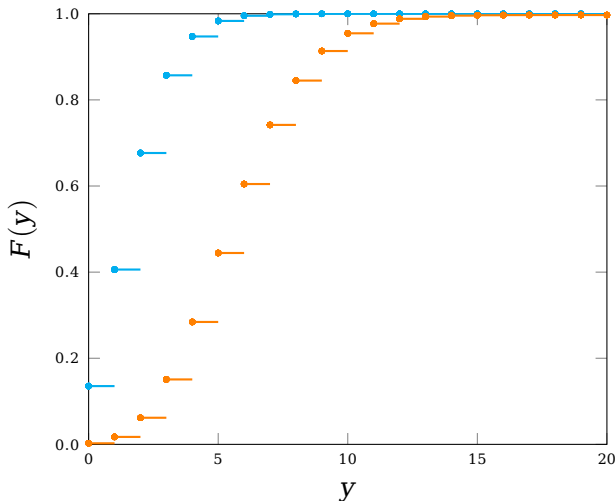
$$F(y) = P(Y \leq y)$$

which can be considered for **any** value y .

For a discrete random variable, this function is termed a **step** function, with steps at only a countable number of points.

Continuous Random variables

Poisson $F(y)$: with $\lambda = 2$ and $\lambda = 6$.



Continuous Random variables

Now suppose that the random variable Y is most accurately thought of as varying **continuously**, that is, can potentially take any value in an interval

- height/weight,
- temperature,
- time,

and so on.

Continuous Random variables

A random variable Y with distribution function $F(y)$ is called **continuous** if $F(y)$ is continuous for all y .

- we forgo the formal mathematical definition here;
- we can interpret to mean ‘has no discontinuities’

The probability accumulates in a continuous fashion.

Probability density functions

Let $f(y)$ be the derivative of $F(\cdot)$ at y

$$f(y) = \frac{dF(y)}{dy}$$

wherever this derivative exists.

$f(y)$ is known as the **probability density function** (pdf).

Probability density functions

It follows that

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

that is, $f(y)$ is the function you need to integrate to get $F(y)$.

i.e. you compute probability by performing an “area under the pdf” calculation.

Probability density functions

The pdf is equivalent to the pmf for discrete distributions

$$F(1) = P(Y \leq 1) = \int_{-\infty}^1 f(y) dy.$$

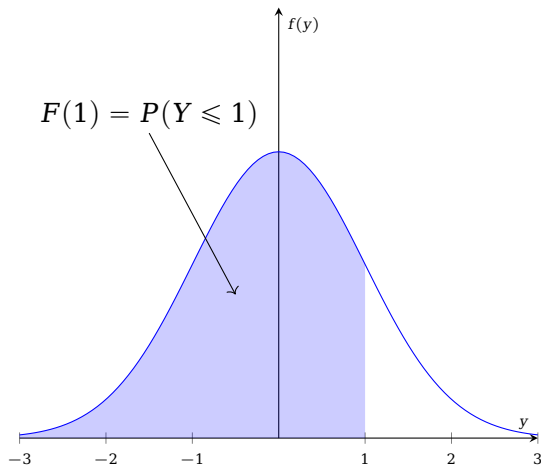
means that to compute $P(Y \leq 1)$, we “add up” all the contributions given by $f(y)$ for all the y lying in the range $(-\infty, 1]$.

Similarly

$$P(Y \in (3, 5)) = \int_3^5 f(y) dy$$

and so on.

Probability density functions



Probability density functions

We compute probabilities as “area under $f(y)$ ” calculations:

- the pdf measures ‘density’, **not probability**, that is

$$f(y) \neq P(Y = y)$$

- you can think of $f(y)$ as similar to what is measured in the bars in a histogram
- $f(y)$ displays the “shape” of the distribution.

Properties of density functions

From the definition, we have two properties of the pdf $f(y)$:

1. Non-negative:

$$f(y) \geq 0 \quad -\infty < y < \infty$$

– this follows as $F(y)$ is non-decreasing.

2. Integrates to 1:

$$\int_{-\infty}^{\infty} f(y) dy = 1$$

– this follows as

$$F(\infty) \equiv \lim_{y \rightarrow \infty} F(y) = 1.$$

Interval probabilities

To compute

$$P(a < Y \leq b)$$

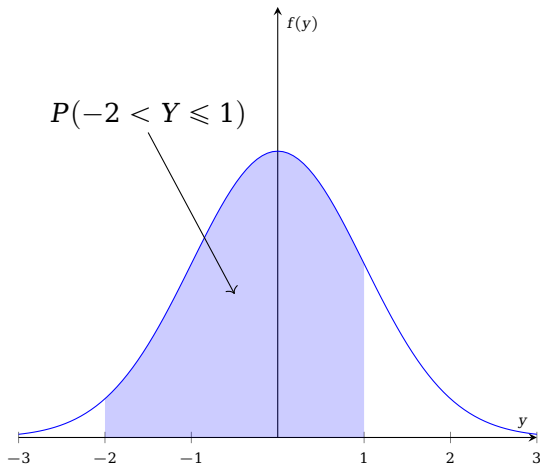
we integrate $f(y)$ over the interval (a, b) as

$$\begin{aligned} P(a < Y \leq b) &= P(Y \leq b) - P(Y \leq a) \\ &= \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt \\ &= \int_a^b f(t) dt. \end{aligned}$$

e.g.

$$P(-2 < Y \leq 1) = \int_{-2}^1 f(y) dy = F(1) - F(-2).$$

Interval probabilities



Interval probabilities

Note

For a **continuous** random variable

$$\begin{aligned}P(a < Y \leq b) &= P(a \leq Y \leq b) \\&= P(a \leq Y < b) \\&= P(a < Y < b)\end{aligned}$$

Interval probabilities

Note

We often have that for some $y \in \mathbb{R}$

$$f(y) = 0$$

and so we again denote by $\mathcal{Y}$ the values for which $f(y) > 0$, that is

$$\mathcal{Y} = \{y : f(y) > 0\}$$

Expectation and Variance

The expectation of continuous random variable Y is

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy = \int_{\mathcal{Y}} y f(y) dy = \mu$$

say, provided this integral **exists**.

For the variance

$$\text{Var}(Y) = E[(Y - \mu)^2] = \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy$$

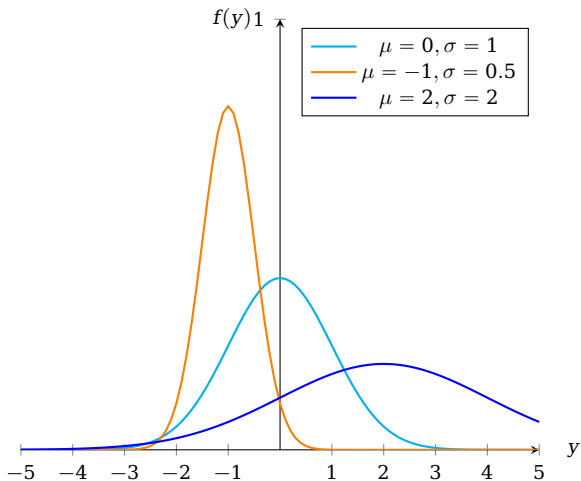
The Normal distribution

The Normal (or Gaussian) distribution has pdf defined on the whole real line by

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \mu)^2 \right\} \quad -\infty < y < \infty$$

where μ and $\sigma > 0$ are parameters.

The Normal distribution



The Normal distribution

Note

- μ is the **centre** of the distribution;
 - ▶ $f(y)$ is **symmetric** about μ .
- σ measures the **spread** of the distribution;
 - ▶ as σ decreases, $f(y)$ becomes more concentrated around μ .
- It can be shown that

$$E(Y) = \mu \quad \text{Var}(Y) = \sigma^2.$$

The Normal distribution

If $\mu = 0$ and $\sigma = 1$ then we have the **standard Normal** pdf

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2}{2} \right\} \quad -\infty < y < \infty$$

The Normal distribution

The cdf, $F(y)$, in the general case is given by

$$F(y) = \int_{-\infty}^y \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(t - \mu)^2\right\} dt$$

but this cannot be computed analytically.

Often, this cdf is written using the special notation $\Phi(y)$.

The Normal distribution

If Y has a Normal distribution with parameters μ and σ , and we consider the random variable Z where

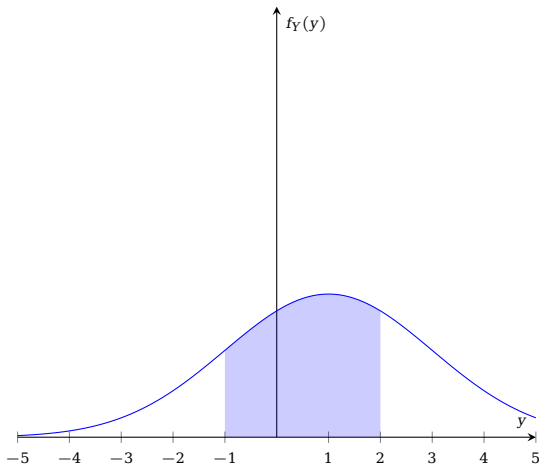
$$Z = \frac{Y - \mu}{\sigma}$$

then Z has a **standard** Normal distribution.

This result tells us that we can always transform a general Normal random variable Y into a standard Normal random variable Z .

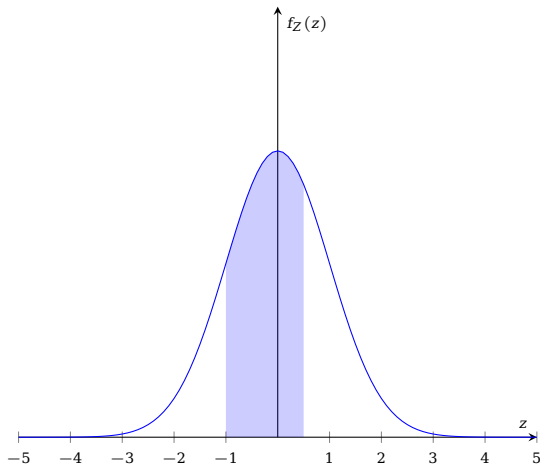
The Normal distribution

$$Y : \mu = 1, \sigma = 2 : P(-1 \leq Y \leq 2)$$



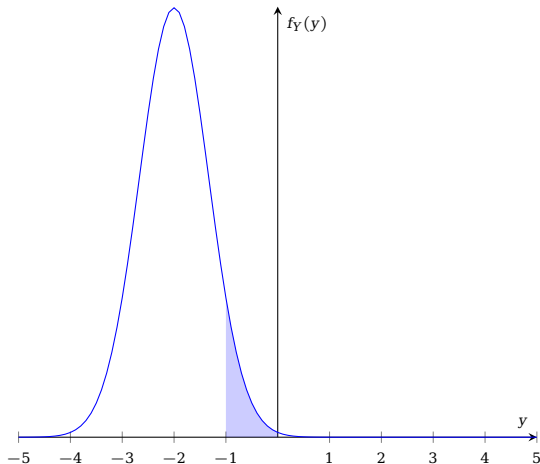
The Normal distribution

$$Z : \mu = 0, \sigma = 1 : P(-1 \leq Z \leq 1/2)$$



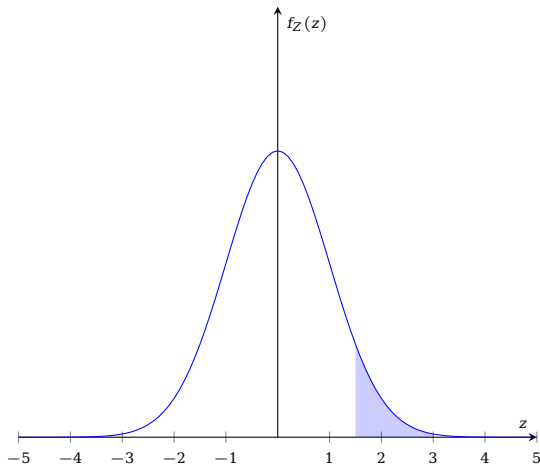
The Normal distribution

$$Y : \mu = -2, \sigma = 2/3 : P(-1 \leq Y \leq 0)$$



The Normal distribution

$$Z : \mu = 0, \sigma = 1 : P(3/2 \leq Z \leq 3)$$



Continuous distributions

There is a large number of continuous distributions that can be used to model data. The details of the

- $f(y)$
- $F(y)$

functions change, but the principles of all the calculations remain the same.

In this course, we will meet only two continuous distributions

- Normal distribution
- t -distribution.

Distributions in R

There are a lot of functions in R for computing with probability distributions.

All the standard distributions are available, specifically

- Binomial, Poisson, Geometric
- Normal (and related distributions)

There are four types of function

1. pmf/pdf functions
2. cdf functions
3. inverse cdf functions
4. random number generating functions

Example: Normal distribution

For the Normal distribution:

1. pdf function: `dnorm`
2. cdf function: `pnorm`
3. inverse cdf function: `qnorm`
4. random number generating function: `rnorm`.

Distributions in R

Example: Normal distribution

Suppose $X \sim \text{Normal}(3.2, 1.5^2)$: to compute

- $f(2.7)$
- $f(-1.2)$

```
dnorm(2.7,3.2,1.5)
```

```
+ [1] 0.2515888
```

```
dnorm(-1.2,3.2,1.5)
```

```
+ [1] 0.003600704
```

The second and third arguments to `dnorm` are $\mu = 3.2$ and $\sigma = 1.5$.

Distributions in R

Example: Normal distribution

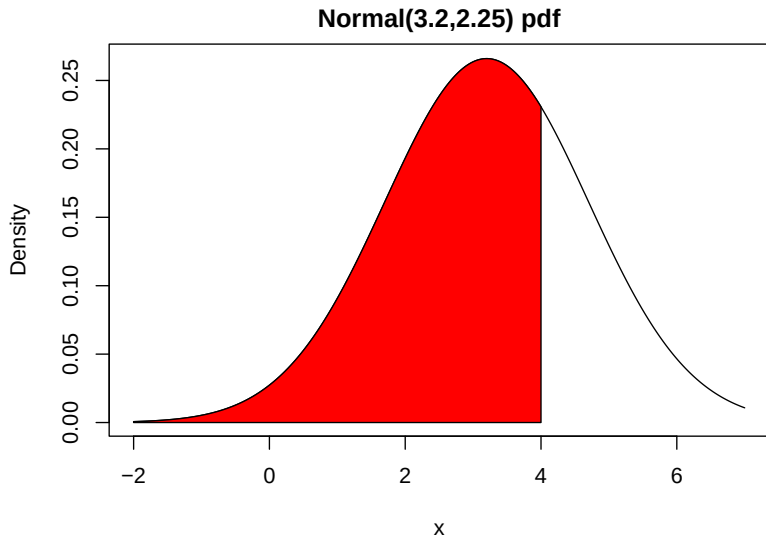
To compute $P(X \leq 4.0)$, we use `pnorm`:

```
pnorm(4.0, 3.2, 1.5)
```

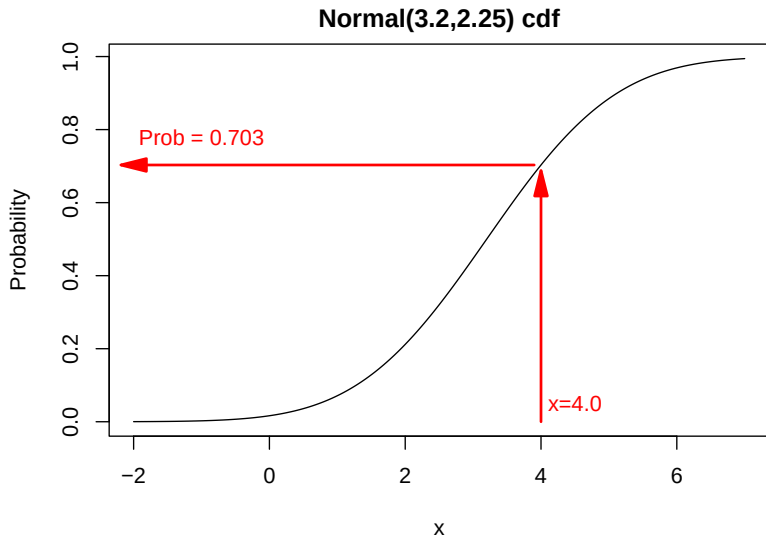
```
+ [1] 0.7030986
```

Recall that this is the area under the Normal pdf from negative infinity up to 4.0.

Distributions in R



Distributions in R



Distributions in R

Example: Normal distribution

We can also compute the x value that solves

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

for some specified p .

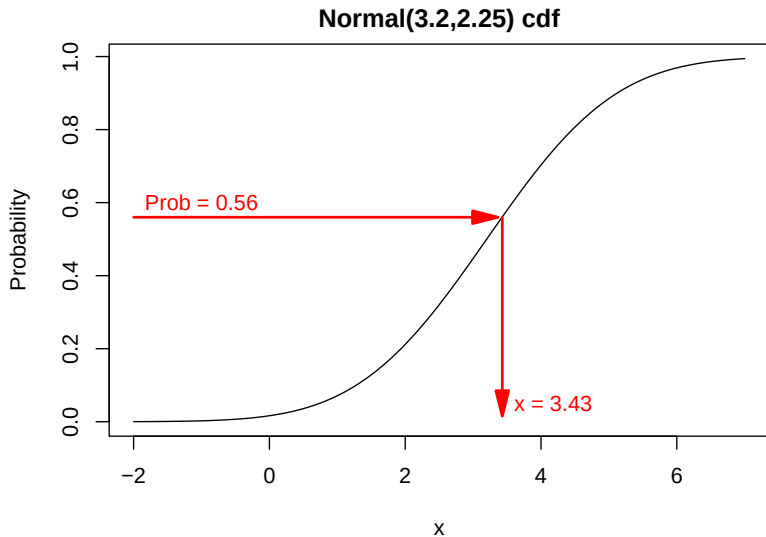
For example, if $p = 0.56$, we use the function `qnorm`.

```
qnorm(0.56, 3.2, 1.5)
```

```
+ [1] 3.426454
```

This value represents how far we need to go to get an area under the curve to be 0.56.

Distributions in R



Distributions in R

Example: Normal distribution

The function `qnorm` computes the inverse cdf. It can be used to compute probability tables.

```
pvec<-c(0.8,0.9,0.95,0.99,0.999)
qvec<-qnorm(pvec)
cbind(pvec,qvec)
```

```
+      pvec      qvec
+ [1,] 0.800 0.8416212
+ [2,] 0.900 1.2815516
+ [3,] 0.950 1.6448536
+ [4,] 0.990 2.3263479
+ [5,] 0.999 3.0902323
```

Distributions in R

Example: Normal distribution

The function `rnorm` simulates values of a Normal random variable

```
n<-10  
rnorm(n)
```

```
+ [1] -0.7204912 -0.7894890 -1.1909317 -1.4324848  0.6493333  1.8838497  
+ [7] -0.9629458  1.1761699 -0.3092402  1.0042386
```

Example: Binomial distribution

For the Binomial distribution, we have

- pmf: `dbinom`
- cdf: `pbinom`
- inverse cdf: `qbinom`
- simulated values: `rbinom`

Distributions in R

Example: Binomial distribution

If $X \sim \text{Binomial}(300, 0.25)$: for $x = 75, 76, \dots, 85$

```
n<-300
p<-0.25
x<-c(75:85)
prob<-dbinom(x,n,p)
rbind(x,round(prob,3))
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]
+ x	75.000	76.000	77.000	78.000	79.000	80.000	81.000	82.000	83.000	84.000	85.000
+	0.053	0.052	0.051	0.048	0.045	0.042	0.038	0.034	0.029	0.025	0.022

Distributions in R

Example: Binomial distribution

To compute


$$P(X > 70) = \sum_{j=71}^{300} \left(\frac{1}{4}\right)^j \left(1 - \frac{1}{4}\right)^{300-j} = 1 - P(X \leq 70)$$

```
1-pbinom(70,n,p)
```

```
+ [1] 0.7233254
```

```
sum(dbinom(71:n,n,p))
```

```
+ [1] 0.7233254
```

typo: binomial coefficient omitted
here, numerical calculation is
correct.

Distributions in R

Example: Binomial distribution

For the special case of $Bernoulli(p) \equiv Binomial(1, p)$:

```
rbinom(20, 1, p)
```

```
+ [1] 0 1 0 0 0 1 0 0 0 0 1 1 1 1 0 0 0 0 0 0
```