Chapter 2: Random variables and probability distributions

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 $\{0, 1, 2, \ldots\}$;
- uncountable.

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

Random variables (cont.)

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 $\mathbb{R}$, that is

$$Y : S \longrightarrow \mathbb{R}$$

$$s \longmapsto y$$

We are essentially saying that

$$Y(s) = y \qquad \text{for } s \in S.$$

The mapping Y is termed a *random variable*.

Random variables (cont.)

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 (cont.)

Example (Two coins)

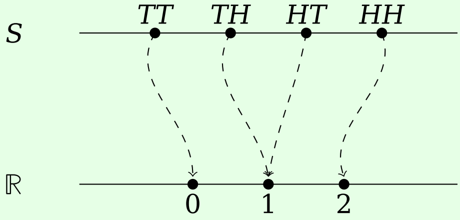
In the original probability specification for events in S,

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

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

Random variables (cont.)

Example (Two coins)



Random variables (cont.)

Example (Two coins)

Then

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

and $Y(.)$ 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 $P(Y(s) = y) = 0$ for all other real values y.

Random variables (cont.)

Notes

- We typically consider the image of S under the mapping $Y(.)$ first
  - that is, we identify the possible values that Y can take.
- In general, $Y(.)$ is a many-to-one mapping.
- We usually suppress the dependence of $Y(.)$ on its argument, and merely write

$$P(Y = y).$$

Notes

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 real values that the random variable takes.

Discrete Random variables

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

$$y_1, y_2, \ldots, y_n, \ldots$$

Could be a

- finite set: eg $\{1, 2, \ldots, 100\}$
- countably infinite set: eg $\mathbb{Z}$.

I will denote the set of values that carry positive probability using the notation $\mathcal{Y}$.

Discrete Random variables (cont.)

In this case, the probabilities assigned to the individual points in the set $\mathcal{Y}$ in must carry all the probability:

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

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

Probability mass function

The *probability mass function* (pmf) for Y, $p(.)$, is the mathematical function that records how probability is distributed across points in $\mathbb{R}$.

That is,

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

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

Probability mass function (cont.)

Note

The notation $f(y)$ is also commonly used, and also it is sometimes helpful to write

$$p_Y(y) \quad \text{or} \quad f_Y(y)$$

to indicate that we are dealing with the pmf for Y.

Probability mass function (cont.)

The pmf could be defined pointwise: for example

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

Probability mass function (cont.)

Alternatively it could be defined using some functional form:

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

Probability mass function (cont.)

The pmf **must** be specified to satisfy two conditions:

(i) The function must specify probabilities: that is

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

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

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

Probability mass function (cont.)

Example (Combinatorial pmf)

A bag contains 7 red balls and 5 black balls. Four balls are selected, with all sequences of selections being equally likely.

Let Y be the random variable recording the number of red balls selected.

What is

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

for $y \in \mathbb{R}$?.

Probability mass function (cont.)

Example (Combinatorial pmf)

First consider $\mathcal{Y}$; we have that

$$\mathcal{Y} = \{0, 1, 2, 3, 4\}$$

- Y must be a non-negative integer;
- there are only four balls in the selection;
- it is possible to select four black balls.

Example (Combinatorial pmf)

What is $P(Y = y)$ for $y \in \{0, 1, 2, 3, 4\}$?

$$(Y = y) \equiv A = \{y \text{ red balls selected out of four}\}$$

That is,

$$(Y = y) \equiv A = \{y \text{ red balls, and } 4 - y \text{ black balls selected}\}$$

Probability mass function (cont.)

Example (Combinatorial pmf)

We compute $P(A)$ as follows: for $y \in \{0, 1, 2, 3, 4\}$

$$n_A = \binom{7}{y} \times \binom{5}{4-y}$$

$\binom{7}{y}$: number of ways of selecting the y red balls from the 7 available.

$\binom{5}{4-y}$: number of ways of selecting the $4 - y$ black balls from the 5 available.

Probability mass function (cont.)

Example (Combinatorial pmf)

Also

$$n_S = \binom{12}{4}$$

which is the number of ways of selecting the 4 balls.

Therefore

$$P(Y = y) = P(A) = \frac{n_A}{n_S} = \frac{\binom{7}{y}\binom{5}{4-y}}{\binom{12}{4}} \quad y = 0, 1, 2, 3, 4$$

and zero for all other values of y.

Probability mass function (cont.)

Example (Combinatorial pmf)

y	0	1	2	3	4
Pr.	$\frac{\binom{7}{0}\binom{5}{4}}{\binom{12}{4}}$	$\frac{\binom{7}{1}\binom{5}{3}}{\binom{12}{4}}$	$\frac{\binom{7}{2}\binom{5}{2}}{\binom{12}{4}}$	$\frac{\binom{7}{3}\binom{5}{1}}{\binom{12}{4}}$	$\frac{\binom{7}{4}\binom{5}{0}}{\binom{12}{4}}$
	0.01010	0.14141	0.42424	0.35354	0.07071

Probability mass function (cont.)

Note

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

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

In this case we would require the infinite sum

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

be equal to 1.

Probability mass function (cont.)

Note

All pmfs take the form

$$\text{constant} \times \text{function of } y$$

or

$$cg(y)$$

say, where $g(y)$ also contains information on the range of values, $\mathcal{Y}$, for which $p(y) > 0$. We have

$$\sum_{y \in \mathcal{Y}} p(y) = 1 \qquad \Longrightarrow \qquad \sum_{y \in \mathcal{Y}} g(y) = \frac{1}{c}$$

Expectations for discrete random variables

The *expectation* (or *expected value*) of a discrete random variable Y, denoted $\mathbb{E}[Y]$, is defined as

$$\mathbb{E}[Y] = \sum_{y \in \mathcal{Y}} y \, p(y)$$

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

Expectations for discrete random variables (cont.)

Notes

- $\mathbb{E}[Y]$ is the centre of mass of the probability distribution.
- If $\mathcal{Y} = \{y_1, y_2, \ldots, y_n, \ldots\}$, we have that

$$\mathbb{E}[Y] = y_1 p(y_1) + y_2 p(y_2) + \cdots + y_n p(y_n) + \cdots$$

- It follows that $\mathbb{E}[Y]$ is finite if

$$\sum_{y \in \mathcal{Y}} |y| \, p(y) < \infty$$

Example (Expectation and Mean)

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

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

and zero otherwise. Then

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

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

Expectations for discrete random variables (cont.)

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

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

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

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

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

The expectation of a function

Suppose $g(.)$ is a real-valued function, and consider $g(Y)$. For example, if Y takes the values $\{-2, -1, 0, 1, 2\}$, and

$$g(t) = t^2$$

then $g(Y)$ takes the values $\{0, 1, 4\}$. If $p(y)$ denotes the pmf of Y, then

$$P(g(Y) = t) = \begin{cases} p(0) & t = 0 \\ p(-1) + p(1) & t = 1 \\ p(-2) + p(2) & t = 4 \\ 0 & \text{otherwise} \end{cases}$$

The expectation of a function (cont.)

Suppose that we denote the set of possible values of $g(Y)$ by

$$\mathcal{G} = \{g_i, i = 1, \ldots, m : g_i = g(y), \text{some } y \in \mathcal{Y}\}.$$

We can define the expectation of $g(Y)$ as

$$\mathbb{E}[g(Y)] = \sum_{y \in \mathcal{Y}} g(y)p(y).$$

The expectation of a function (cont.)

We see that this is the correct formula by considering the random variable $g(Y)$; this is a discrete random variable itself, with pmf given by

$$P(g(Y) = g_i) = \sum_{\substack{y \in \mathcal{Y}: \\ g(y)=g_i}} p(y) = p^*(g)$$

say.

The expectation of a function (cont.)

Then

$$
\begin{aligned}
E[g(Y)] &= \sum_{i=1}^{m} g_i p^*(g_i) \\
&= \sum_{i=1}^{m} g_i \left\{ \sum_{\substack{y \in \mathcal{Y}: \\ g(y)=g_i}} p(y) \right\} \\
&= \sum_{y \in \mathcal{Y}} g(y) p(y).
\end{aligned}
$$

Variance

The *variance* of random variable Y is denoted $\mathbb{V}[Y]$, and is defined by

$$\mathbb{V}[Y] = \mathbb{E}[(Y - \mu)^2]$$

where $\mu = \mathbb{E}[Y]$. The *standard deviation* of Y is

$$\sqrt{\mathbb{V}[Y]}.$$

Note that it is possible that $\mathbb{V}[Y]$ is not finite, even if $\mathbb{E}[Y]$ is finite.

Variance (cont.)

The variance is a measure of how "dispersed" the probability distribution is.

Note that

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

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

In fact, $\mathbb{V}[Y] > 0$ unless the distribution of Y is *degenerate*, that is, for some constant c

$$p(c) = 1$$

in which case $\mathbb{V}[Y] = 0$.

Properties of expectations

For discrete random variable Y with expectation $\mathbb{E}[Y]$, and real constants a and b, we have that

$$\mathbb{E}[aY + b] = a\mathbb{E}[Y] + b.$$

This result follows from the previous result for $g(Y)$ with

$$g(t) = at + b.$$

Properties of expectations (cont.)

This follows as

$$\begin{aligned}\mathbb{E}[aY+b] &= \sum_y (ay+b)p(y) \\ &= a\sum_y yp(y) + b\sum_y p(y) \\ &= a\mathbb{E}[Y]+b\end{aligned}$$

as

$$\sum_y p(y) = 1.$$

Properties of expectations (cont.)

Now suppose we have

$$g(Y) = ag_1(Y) + bg_2(Y)$$

for two separate functions $g_1(.)$ and $g_2(.)$. Then

$$\begin{aligned}
\mathbb{E}[g(Y)] &= \mathbb{E}[ag_1(Y) + bg_2(Y)] \\
&= \sum_y (ag_1(y) + bg_2(y))p(y) \\
&= a\sum_y g_1(y)p(y) + b\sum_y g_2(y)p(y) \\
&= a\mathbb{E}[g_1(Y)] + b\mathbb{E}[g_2(Y)].
\end{aligned}$$

Properties of expectations (cont.)

Note

These results merely follow standard mathematical concepts of the *linearity* of summations:

$$\sum_i (a_i + b_i) = \sum_i a_i + \sum_i b_i.$$

Properties of expectations (cont.)

For the variance, if we denote $\mathbb{E}[Y] = \mu$, then

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

In this formula

$$\mathbb{E}[Y^2] = \sum_y y^2 \; p(y).$$

Properties of expectations (cont.)

From this we can deduce that

$$\mathbb{V}[aY + b] = a^2 \mathbb{V}[Y].$$

We have that

$$\mathbb{E}[aY + b] = a\mathbb{E}[Y] + b = a\mu + b$$

so

$$\mathbb{V}[aY + b] = \mathbb{E}[(aY + b - (a\mu + b))^2]$$

which reduces to

$$\mathbb{E}[(aY - a\mu)^2] = \mathbb{E}[a^2(Y - \mu)^2] = a^2\mathbb{E}[(Y - \mu)^2].$$

The binomial distribution

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

On 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

$$Y = 1 \text{ if outcome is ‘success’} \qquad Y = 0 \text{ if outcome is ‘failure’}$$

and then consider

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

The binomial distribution (cont.)

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 (cont.)

We may then write

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

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 (cont.)

This model covers many experimental settings:

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

etc.

The binomial distribution (cont.)

We have for this model that

$$\mu \equiv \mathbb{E}[Y] = \sum_{i=0}^{1} yp(y) = (0 \times p(0) + 1 \times p(1))$$

$$= (0 \times q + 1 \times p) = p$$

The binomial distribution (cont.)

Similarly

$$\mathbb{E}[Y^2] = \sum_{i=0}^{1} y^2 p(y) = (0^2 \times p(0) + 1^2 \times p(1))$$

$$= (0 \times q + 1 \times p) = p$$

so therefore

$$\mathbb{V}[Y] = \mathbb{E}[Y^2] - \mu^2 = p - p^2 = p(1-p).$$

We now consider a version of the experiment where a sequence of n identical binary experiments (or *'trials'*) 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

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

The binomial distribution (cont.)

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

$$(Y = y) \quad \Longleftrightarrow \quad \text{“}y \text{ successes in } n \text{ trials”}$$

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

$$\begin{array}{c} 10101000 \\ 01010001 \\ 11100000 \end{array}$$

etc.

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

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

that is, in the examples above

$$\begin{gathered} pqpqpqqq \\ qpqpqqqp \\ pppqqqqq \end{gathered}$$

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

The binomial distribution (cont.)

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 (cont.)

Therefore we have that

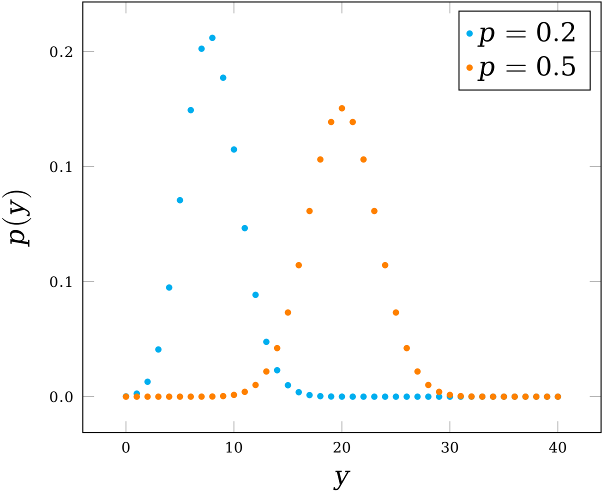
$$p(y) \equiv P(Y = y) = \binom{n}{y} p^y q^{n-y} \qquad y \in \{0, 1, 2, \ldots, n\}$$

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

This is the pmf for the *Binomial distribution*.

The binomial distribution (cont.)

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



The binomial distribution (cont.)

Expectation : For the expectation of a binomial distribution, we may compute directly as follows:

$$\begin{aligned}
\mathbb{E}[Y] = \sum_{y=0}^{n} y p(y) &= \sum_{y=0}^{n} y \binom{n}{y} p^y q^{n-y} \\
&= \sum_{y=1}^{n} y \frac{n!}{y!(n-y)!} p^y q^{n-y} \\
&= n \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!(n-y)!} p^y q^{n-y}
\end{aligned}$$

The binomial distribution (cont.)

Now we may write

$$(n-y) = (n-1) - (y-1)$$

so therefore the sum can be written

$$\sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!((n-1)-(y-1))!} p^y q^{(n-1)-(y-1)}.$$

The binomial distribution (cont.)

Rewriting the terms in the sum by setting $j = y - 1$, this equals

$$\sum_{j=0}^{n-1} \frac{(n-1)!}{j!((n-1)-j)!} p^{j+1} q^{(n-1)-j}.$$

or equivalently, by the binomial theorem, this equals

$$p \sum_{j=0}^{n-1} \frac{(n-1)!}{j!((n-1)-j)!} p^{j} q^{(n-1)-j} = p(p+q)^{n-1}.$$

But $p + q = 1$, so therefore

$$\mathbb{E}[Y] = np.$$

The binomial distribution (cont.)

Variance : For the variance, we have by the previous result that

$$\mathbb{E}[Y^2] = \mathbb{E}[Y(Y-1)] + \mathbb{E}[Y].$$

The binomial distribution (cont.)

Now, we have

$$\begin{aligned}
\mathbb{E}[Y(Y-1)] &= \sum_{y=0}^{n} y(y-1)p(y) \\
&= \sum_{y=0}^{n} y(y-1)\binom{n}{y} p^y q^{n-y} \\
&= \sum_{y=2}^{n} y(y-1)\frac{n!}{y!(n-y)!} p^y q^{n-y} \\
&= n(n-1)\sum_{y=2}^{n} \frac{(n-2)!}{(y-2)!(n-y)!} p^y q^{n-y}
\end{aligned}$$

The binomial distribution (cont.)

Using the same trick as for the expectation, and writing

$$(n-y) = (n-2) - (y-2)$$

we see that we can rewrite the sum

$$n(n-1)\sum_{y=2}^{n} \frac{(n-2)!}{(y-2)!((n-2)-(y-2))!} p^y q^{(n-2)-(y-2)}$$

$$= n(n-1)\sum_{j=0}^{n-2} \frac{(n-2)!}{j!((n-2)-j)!} p^{j+2} q^{(n-2)-j}$$

$$= n(n-1)p^2\sum_{j=0}^{n-2} \frac{(n-2)!}{j!((n-2)-j)!} p^{j} q^{(n-2)-j}$$

$$= n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2.$$

The binomial distribution (cont.)

Therefore

$$\mathbb{E}[Y(Y-1)] = n(n-1)p^2$$

and so

$$\mathbb{E}[Y^2] = \mathbb{E}[Y(Y-1)+Y] = \mathbb{E}[Y(Y-1)]+\mathbb{E}[Y] = n(n-1)p^2+np.$$

Hence

$$\begin{aligned}\mathbb{V}[Y] = \mathbb{E}[Y^2] - \mu^2 &= n(n-1)p^2 + np - (np)^2 \\ &= -np^2 + np \\ &= np(1-p) \\ &= npq.\end{aligned}$$

The binomial distribution (cont.)

Note

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

$$\mathbb{E}[Y] = p \qquad \mathbb{V}[Y] = pq$$

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

$$\mathbb{E}[Y] = np \qquad \mathbb{V}[Y] = npq.$$

The binomial distribution (cont.)

If $n = 1$, we have that

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

and $p(y) = 0$ for all other y. This special case is known as the *Bernoulli* distribution with parameter p.

In this case we consider the number of successes in single binary trial. The expectation is p and the variance is pq.

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 (cont.)

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

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

and zero for other values of y. This follows as

$$Y = y \iff \text{“}y - 1 \text{ failures, followed by a success"}$$

that is, for example, we observe a sequence 00001 which 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 (cont.)

Notes

- Note that

$$\sum_{y=1}^{\infty} q^{y-1} p = p \sum_{j=0}^{\infty} q^j = p \frac{1}{1-q} = p \frac{1}{(1-(1-p))} = 1.$$

- The geometric distribution is an example of a waiting-time distribution.

The geometric distribution (cont.)

Notes

- Sometimes the geometric distribution is written

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

– 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

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

.

Expectation:

$$\mathbb{E}[Y] = \sum_{y=1}^{\infty} y p(y) = \sum_{y=1}^{\infty} y q^{y-1} p = p \sum_{y=1}^{\infty} y q^{y-1}.$$

We have that

$$\sum_{y=1}^{\infty} y q^{y-1} = \frac{1}{(1-q)^2} = \frac{1}{(1-(1-p))^2} = \frac{1}{p^2}.$$

Therefore

$$\mathbb{E}[Y] = p \frac{1}{p^2} = \frac{1}{p}.$$

The geometric distribution (cont.)

Variance: Using the same approach as for the binomial distribution, we have

$$\mathbb{E}[Y(Y-1)] = \sum_{y=1}^{\infty} y(y-1)p(y) = pq\sum_{y=2}^{\infty} y(y-1)q^{y-2}.$$

Now, we have the following calculus results for a convergent sum

$$\frac{d}{dt}\left\{\sum_{j=0}^{\infty} t^j\right\} = \sum_{j=1}^{\infty} jt^{j-1}$$

$$\frac{d^2}{dt^2}\left\{\sum_{j=0}^{\infty} t^j\right\} = \sum_{j=2}^{\infty} j(j-1)t^{j-2}.$$

The geometric distribution (cont.)

But

$$\sum_{j=0}^{\infty} t^j = \frac{1}{1-t}$$

so therefore we must have that

$$\frac{d}{dt}\left\{\sum_{j=0}^{\infty} t^j\right\} = \frac{d}{dt}\left\{\frac{1}{1-t}\right\} = \frac{1}{(1-t)^2}$$

$$\frac{d^2}{dt^2}\left\{\sum_{j=0}^{\infty} t^j\right\} = \frac{d^2}{dt^2}\left\{\frac{1}{1-t}\right\} = \frac{2}{(1-t)^3}.$$

Thus

$$\sum_{y=2}^{\infty} y(y-1)q^{y-2} = \frac{2}{(1-q)^3} = \frac{2}{p^3}$$

and hence

$$\mathbb{E}[Y(Y-1)] = \sum_{y=1}^{\infty} y(y-1)pq^{y-1} = pq\frac{2}{p^3} = \frac{2(1-p)}{p^2}.$$

The geometric distribution (cont.)

Therefore

$$\begin{aligned}\mathbb{V}[Y] &= \mathbb{E}[Y(Y-1)] + \mathbb{E}[Y] - \{\mathbb{E}[Y]\}^2 \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{(1-p)}{p^2}.\end{aligned}$$

The geometric distribution (cont.)

Now consider the probability

$$P(Y > y)$$

for $y \in \{1, 2, 3, \ldots\}$. Using a partitioning argument, we see that

$$P(Y > y) = \sum_{j=y+1}^{\infty} P(Y = j).$$

That is

$$P(Y > y) = \sum_{j=y+1}^{\infty} q^{j-1} p = pq^y \sum_{j=y+1}^{\infty} q^{j-y-1} = pq^y \frac{1}{1-q}.$$

The geometric distribution (cont.)

Therefore

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

and consequently

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

The geometric distribution (cont.)

We can define the function

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

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

The geometric distribution (cont.)

We can define all such ‘interval’ probabilities such as

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

etc. via $F(y)$. This is because for **all** discrete distributions and sets $\mathcal{X}$

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

The geometric distribution (cont.)

Note

We can compute

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

for any real value y. However, we must remember that

$$P(Y \leq y) = P(Y \leq \lfloor y \rfloor)$$

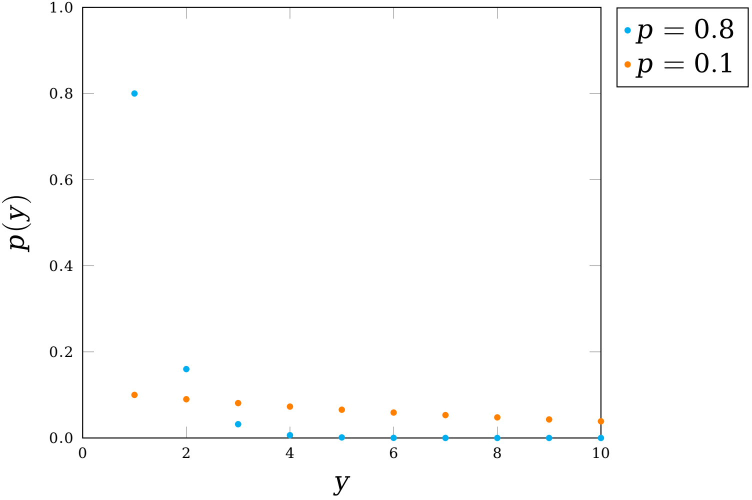
where $\lfloor y \rfloor$ is the largest integer that is no greater than y. For example,

$$F(1.234) = F(1) \qquad F(4.789) = F(4)$$

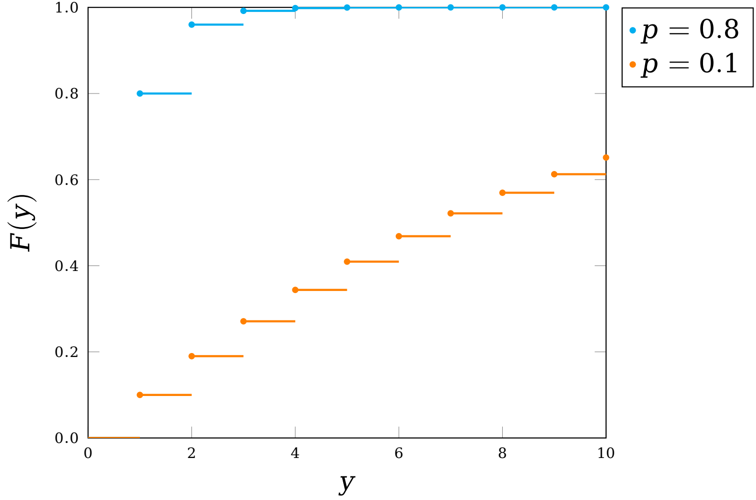
etc.

The geometric distribution (cont.)

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



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



The negative binomial distribution

As an extension to the geometric distribution setting, we now consider a sequence of identical and independent binary trials, but we continue until we observe the rth success, where $r = \{1, 2, \ldots, \}$ is a fixed number. That is, if $r = 3$, we consider sequences such as

00010101
0111
010000011
10101
0000010101

etc.

The negative binomial distribution (cont.)

If random variable Y records the number of trials carried out up to an including the rth success, then we see that

$$(Y = y) \equiv A \cap B$$

where

$$A = \text{“}(r-1) \text{ successes in the first } (y-1) \text{ trials”}$$
$$B = \text{“a success on the } y\text{th trial”}$$

with A and B independent events.

The negative binomial distribution (cont.)

We have that Y can take values on the set

$$r, r+1, \ldots$$

Now, we have from the binomial distribution that

$$P(A) = \binom{y-1}{r-1} p^{r-1} q^{(y-1)-(r-1)} \qquad y = \{r, r+1, r+2, \ldots\}$$

and

$$P(B) = p$$

The negative binomial distribution (cont.)

Therefore, as

$$(y-1)-(r-1)=y-r$$

we have that

$$p(y)=P(Y=y)=\binom{y-1}{r-1}p^r q^{y-r} \qquad y=r,r+1,r+2,\ldots$$

and zero for other values of y.

This is the pmf of the *negative binomial distribution*.

The negative binomial distribution (cont.)

Notes

- We can show (Exercise)

$$\sum_{y=r}^{\infty} p(y) = 1.$$

- The negative binomial distribution is another example of a waiting-time distribution.

The negative binomial distribution (cont.)

Notes

- Sometimes the negative binomial distribution is written

$$p(y) = \binom{y+r-1}{r-1} p^r q^y \qquad y = 0, 1, 2, \ldots$$

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

$$Y^* = Y - r.$$

- Note that by definition

$$\binom{y+r-1}{r-1} = \binom{y+r-1}{y}$$

The negative binomial distribution (cont.)

For random variable Y, we have the following results:

Expectation:

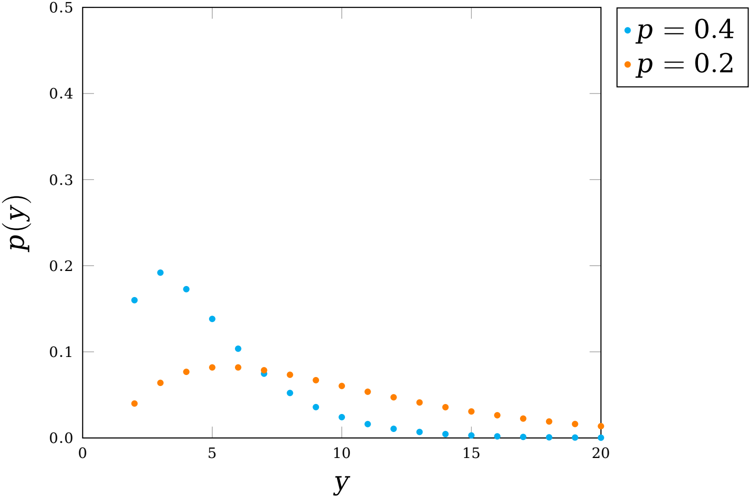
$$\mathbb{E}[Y] = \frac{r}{p}$$

Variance:

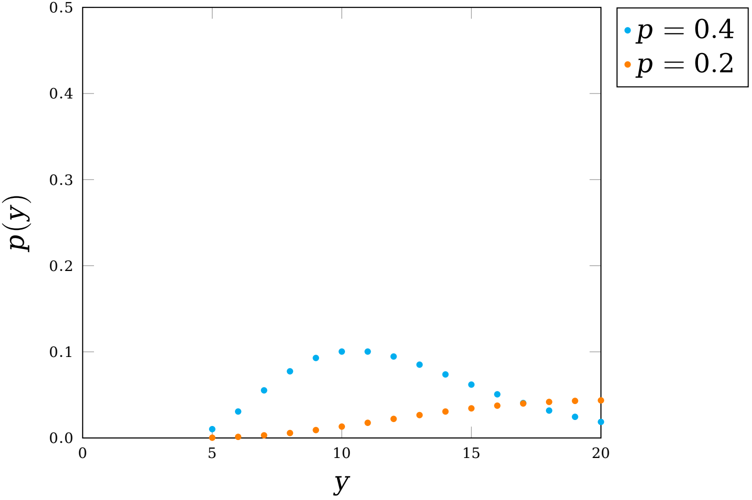
$$\mathbb{V}[Y] = \frac{r(1-p)}{p^2}$$

The negative binomial distribution $r = 2$

Negative binomial $p(y)$: with $p = 0.4$ and $p = 0.2$.



Negative binomial $p(y)$: with $p = 0.4$ and $p = 0.2$.



The hypergeometric distribution

The discrete distribution known as the *hypergeometric distribution* relates to the earlier combinatorial exercises.

Suppose we have a finite 'population' (set) of N items which comprise

- r Type I items,
- $N - r$ Type II objects.

The experiment consists of selecting n items from the population, without replacement, such that all such selections are equally likely.

The hypergeometric distribution (cont.)

Define the random variable Y such that

$$(Y = y) \equiv A = \text{“}y \text{ Type I items in the selection”}$$

First, consider the feasible values for Y; we know that

$$0 \leq y \leq r$$

but also that

$$0 \leq n - y \leq N - r$$

The hypergeometric distribution (cont.)

Putting these two facts together, we must have that

$$\max\{0, n - N + r\} \leq y \leq \min\{n, r\}$$

so

$$\mathcal{Y} = \{\max\{0, n - N + r\}, \ldots, \min\{n, r\}\}.$$

Secondly, we compute $P(A)$ as follows: for $y \in \mathcal{Y}$

$$n_A = \binom{r}{y} \times \binom{N-r}{n-y}$$

$\binom{r}{y}$: number of ways of selecting y Type I items from the r available.

$\binom{N-r}{n-y}$: number of ways of selecting $n-y$ Type II items from the $N-r$ available.

The hypergeometric distribution (cont.)

Also

$$n_S = \binom{N}{n}$$

which is the number of ways of selecting the n items.

Therefore

$$p(y) = P(Y = y) = P(A) = \frac{n_A}{n_S} = \frac{\binom{r}{y} \times \binom{N-r}{n-y}}{\binom{N}{n}} \quad y \in \mathcal{Y}$$

and zero for all other values of y.

This is the pmf of the *hypergeometric distribution*.

The hypergeometric distribution (cont.)

Note

An equivalent expression for $p(y)$ is

$$p(y) = \frac{\binom{n}{y} \times \binom{N-n}{r-y}}{\binom{N}{r}} \quad y \in \mathcal{Y}$$

which we can obtain by simple manipulation of the combinatorial terms, or justify by considering

- the ways of arranging the y Type I items in the sample of size n,
- the $r - y$ Type I items in the $N - n$ 'non-sampled' items,
- the number of ways of arranging the Type I and Type II items.

The hypergeometric distribution (cont.)

Note

When N and r are very large, $p(y)$ is well approximated by a binomial pmf

$$\binom{n}{y}\left(\frac{r}{N}\right)^{y}\left(1-\frac{r}{N}\right)^{n-y}$$

We now examine a limiting case of the binomial experimental setting; recall that binomial distribution arises when we consider n identical and independent trials each with success probability p.

$$10001000101010101010 1$$

etc. We have

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

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

The Poisson distribution (cont.)

Suppose we allow n to get larger, and p to get smaller, but in such a way that

$$n \times p$$

remains constant at the value λ say. That is, we may write

$$p = \frac{\lambda}{n}$$

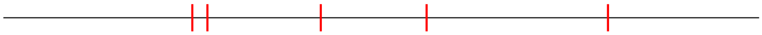
and consider n becoming large.

The Poisson distribution (cont.)

Note

The parameter λ corresponds to the ‘rate’ at which successes occur in a continuous approximation to the experiment.

The Poisson distribution (cont.)



The Poisson distribution (cont.)

We then have that

$$p^y(1-p)^{n-y} = \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y}$$

and

$$\binom{n}{y} = \frac{n!}{y!(n-y)!} = \frac{1}{y!}\frac{n!}{(n-y)!}$$

so we may write the binomial pmf

$$\frac{\lambda^y}{y!}\left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\ldots(n-y+1)}{n^y}\left(1 - \frac{\lambda}{n}\right)^{-y}$$

The Poisson distribution (cont.)

Now we may rewrite

$$\frac{n(n-1)\dots(n-y+1)}{n^y} = \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-y+1}{n}\right)$$

and deduce from this that for fixed y as $n \longrightarrow \infty$

$$\frac{n(n-1)\dots(n-y+1)}{n^y}\left(1-\frac{\lambda}{n}\right)^{-y} \longrightarrow 1$$

as each term in the product converges to 1.

The Poisson distribution (cont.)

Finally, by a standard representation

$$\lim_{n \longrightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Recall that

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

The Poisson distribution (cont.)

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

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

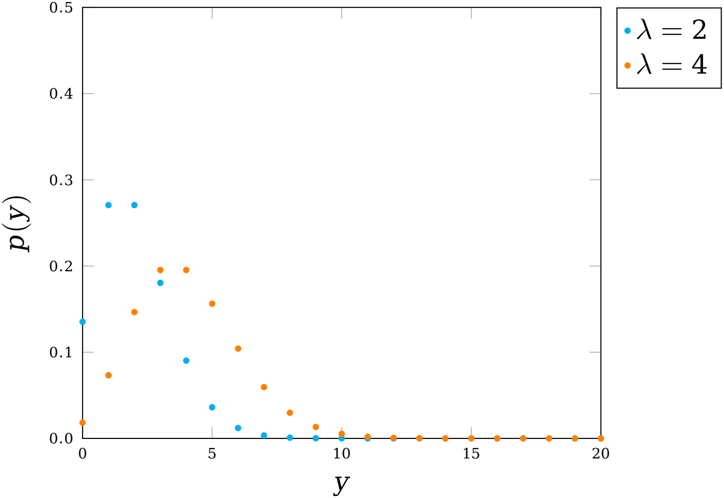
with $p(y) = 0$ for other values of y.

This is the pmf for the *Poisson distribution*.

Note that

$$\sum_{y=0}^{\infty} \frac{\lambda^y}{y!}e^{-\lambda} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda}e^{\lambda} = 1.$$

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



Expectation : For the expectation of a Poisson distribution:

$$
\begin{aligned}
\mathbb{E}[Y] = \sum_{y=0}^{\infty} y p(y) &= \sum_{y=0}^{\infty} y \frac{\lambda^y}{y!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{y=1}^{\infty} \frac{\lambda^y}{(y-1)!} \\
&= e^{-\lambda} \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} \\
&= e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \\
&= e^{-\lambda} \lambda e^{\lambda} = \lambda.
\end{aligned}
$$

The Poisson distribution (cont.)

Variance : For the variance of a Poisson distribution: using the usual trick

$$\mathbb{E}[Y(Y-1)] = \sum_{y=0}^{\infty} y(y-1)p(y) = \sum_{y=2}^{\infty} y(y-1)\frac{\lambda^y}{y!}e^{-\lambda}$$

$$= e^{-\lambda}\lambda^2 \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!}$$

$$= e^{-\lambda}\lambda^2 \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

$$= e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2.$$

The Poisson distribution (cont.)

Therefore

$$\mathbb{V}[Y] = \mathbb{E}[Y(Y-1)] + \mathbb{E}[Y] - \{\mathbb{E}[Y]\}^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Thus for the Poisson distribution

$$\mathbb{E}[Y] = \mathbb{V}[Y] = \lambda.$$

The Poisson distribution (cont.)

Note

The Poisson distribution is widely used to model the distribution of count data.

It corresponds to a model where incidents (‘successes’) occur independently in continuous time at a constant rate.

The Poisson process

The Poisson distribution is widely used to model the distribution of count data.

- We consider the limiting case of a sequence of binary trials with success probability $\lambda\delta$ carried out independently in disjoint time intervals of the form

$$[(k-1)\delta, k\delta) \qquad k = 1, 2, 3, \ldots$$

  as $\delta \longrightarrow 0$.

- This corresponds to a model where 'successes' occur independently in continuous time at a constant rate λ per unit time.

The Poisson process (cont.)

Consider the discrete random variables $X([t_0, t_1))$ that counts the number of "events" (ie successes) that are observed in the interval $[t_0, t_1)$ for any $t_0 < t_1$ according to the probabilities

$$P(X([t, t+\delta t)) = x) = \begin{cases} 1 - \lambda\delta t + \mathrm{o}(\delta t) & x = 0 \\ \lambda\delta t + \mathrm{o}(\delta t) & x = 1 \\ \mathrm{o}(\delta t) & x = 2, 3, \ldots \end{cases}$$

where $\mathrm{o}(\delta t)$ is a constant that converges to zero as $\delta t \longrightarrow 0$ such that the random variables

$$X([a, b)) \qquad X([c, d))$$

are independent for any pair of disjoint intervals $[a, b)$ and $[c, d)$.

The Poisson process (cont.)

Let $X(t) \equiv X([0,t))$ be the discrete random variable counting the number of events that occur in the time interval $[0,t)$. Then

$$X(t) \sim Poisson(\lambda t)$$

so that

$$P(X(t) = x) = e^{-\lambda t}\frac{(\lambda t)^x}{x!} \qquad x = 0, 1, 2 \ldots$$

with $P(X(t) = x) = 0$ for other values of x.

To see this, let

$$p_k(t) = P(X(t) = k) \qquad k = 0, 1, 2, \ldots .$$

Then for $k = 0$

$$\begin{aligned} p_0(t + \delta t) &= P(X(t + \delta t) = 0) \\ &= P(X([0, t + \delta t)) = 0) \\ &= P(X([0, t)) = 0 \cap X([t, t + \delta t)) = 0)) \\ &= P(X([0, t)) = 0) \times P(X([t, t + \delta t)) = 0)) \\ &= p_0(t)(1 - \lambda \delta t + \mathrm{o}(\delta t)) \end{aligned}$$

where the fourth line follows by the independence assumption.

Then

$$\frac{p_0(t+\delta t) - p_0(t)}{\delta t} = -\lambda p_0(t) + \frac{\mathrm{o}(\delta t)}{\delta t}$$

so taking the limit as $\delta t \longrightarrow 0$, we obtain that

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t)$$

which we may solve to obtain that

$$p_0(t) = e^{-\lambda t}$$

For $k > 0$, we have

$$\begin{aligned} p_k(t+\delta t) &= P(X(t+\delta t) = k) \\ &= P(X([0, t+\delta t)) = k) \\ &= \sum_{x=0}^{\infty} P(X([0, t+\delta t)) = k | X(t) = x) P(X(t) = x) \end{aligned}$$

by the Theorem of Total Probability.

Now

$$P(X([0, t+\delta t)) = k | X(t) = x) = P(X[t, t+\delta t) = k - x)$$

and

$$P(X[t, t+\delta t) = k - x) = \begin{cases} 1 - \lambda\delta t + \mathrm{o}(\delta t) & k - x = 0 \\ \lambda\delta t + \mathrm{o}(\delta t) & k - x = 1 \\ \mathrm{o}(\delta t) & k - x = 2, 3, \ldots \end{cases}$$

The Poisson process (cont.)

Thus

$$
\begin{aligned}
p_k(t+\delta t) &= \sum_{x=0}^{\infty} P(X([0,t+\delta t)) = k | X(t) = x) P(X(t) = x) \\
&= \sum_{x=0}^{k} P(X([0,t+\delta t)) = k | X(t) = x) p_x(t) \\
&= \mathrm{o}(\delta t) \sum_{x=0}^{k-2} p_x(t) + (\lambda \delta t + \mathrm{o}(\delta t)) p_{k-1}(t) \\
&\qquad\qquad + (1 - \lambda \delta t + \mathrm{o}(\delta t)) p_k(t) \\
&= (1 - \lambda \delta t) p_k(t) + (\lambda \delta t) p_{k-1}(t) + \mathrm{o}(\delta t)
\end{aligned}
$$

as we may collect all the terms involving $\mathrm{o}(\delta t)$ together.

The Poisson process (cont.)

Hence, rearranging we obtain

$$\frac{p_k(t+\delta t) - p_k(t)}{\delta t} = -\lambda p_k(t) + \lambda p_{k-1}(t) + \frac{\mathrm{o}(\delta t)}{\delta t}$$

so taking the limit as $\delta t \longrightarrow 0$, we obtain that

$$\frac{dp_k(t)}{dt} = -\lambda p_k(t) + \lambda p_{k-1}(t).$$

The Poisson process (cont.)

This equation holds if and only if

$$e^{\lambda t}\frac{dp_k(t)}{dt} = -\lambda e^{\lambda t}p_k(t) + \lambda e^{\lambda t}p_{k-1}(t).$$

which may be rewritten

$$\frac{d}{dt}\left\{e^{\lambda t}p_k(t)\right\} = \lambda e^{\lambda t}p_{k-1}(t)$$

or, writing $g_k(t) = e^{\lambda t}p_k(t)$,

$$\frac{d}{dt}\{g_k(t)\} = \lambda g_{k-1}(t).$$

The Poisson process (cont.)

A solution to this equation is

$$g_k(t) = \frac{(\lambda t)^k}{k!}$$

so therefore we have that

$$p_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \qquad k = 1, 2, 3, \ldots$$

Hence

$$p_k(t) = P(X(t) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \qquad k = 1, 2, 3, \ldots$$

so $X(t) \sim Poisson(\lambda t)$.

The Poisson process (cont.)

Example (Accidents)

Accidents at a road junction occur according to a Poisson process with parameter $\lambda = 0.5$ per week. Then

- the number of accidents in a given week is a Poisson random variable with parameter 0.5;
- the number of accidents in two successive weeks are independent Poisson random variables each with parameter 0.5;
- the number of accidents in any two week period is a Poisson random variable with parameter $0.5 \times 2 = 1$;
- the number of accidents in a one year period is a Poisson random variable with parameter $0.5 \times 52 = 26$.

Moments

We may generalize the concept of expectation and variance by considering the following quantities: if $\mu = \mathbb{E}[Y]$,

$$k^{\text{th}} \text{ moment: } \quad \mu_k' = \mathbb{E}[Y^k] = \sum_y y^k p(y)$$

$$k^{\text{th}} \text{ central moment: } \quad \mu_k = \mathbb{E}[(Y-\mu)^k] = \sum_y (y-\mu)^k p(y)$$

These quantities help to characterize $p(y)$.

Moment-generating functions

In mathematics, the *generating function*, $g(t)$, for the real-valued sequence $\{a_j\}$ is the function defined by the sum

$$g(t) = \sum_j a_j t^j.$$

For example,

$$g(t) = (1+t)^n = \sum_{j=0}^{n} \binom{n}{j} t^j$$

is the generating function for the binomial coefficients

$$a_j = \binom{n}{j} \qquad j = 0, 1, \ldots, n.$$

Moment-generating functions (cont.)

We can also define an (exponential) generating function as

$$g(t) = \sum_j a_j \frac{t^j}{j!}.$$

Moment-generating functions (cont.)

For pmf $p(y)$, the *moment-generating function* (mgf), $m(t)$, is the (exponential) generating function for the moments of $p(y)$:

$$m(t) = \sum_{j=0}^{\infty} \mu_j' \frac{t^j}{j!}$$

whenever this sum is convergent.

More specifically, we require this sum to be finite at least in the neighbourhood of $t = 0$, for $|t| \leq b$, say.

Moment-generating functions (cont.)

Provided the sum is finite, we have for fixed t,

$$
\begin{aligned}
m(t) &= \sum_{j=0}^{\infty} \mu'_j \frac{t^j}{j!} \\
&= \sum_{j=0}^{\infty} \left\{ \sum_{y} y^j p(y) \right\} \frac{t^j}{j!} \\
&= \sum_{y} \left\{ \sum_{j=0}^{\infty} \frac{(ty)^j}{j!} \right\} p(y)
\end{aligned}
$$

Moment-generating functions (cont.)

But

$$\sum_{j=0}^{\infty} \frac{(ty)^j}{j!} = e^{ty}$$

so therefore

$$m(t) = \sum_{y} e^{ty} p(y) = \mathbb{E}\left[e^{tY}\right]$$

Note also that if we write

$$m^{(k)}(t) = \frac{d^k}{dt^k}\{m(t)\}$$

for the kth derivative of $m(t)$ with respect to t, we have

$$\begin{aligned}\frac{d^k}{dt^k}\{m(t)\} &= \frac{d^k}{dt^k}\left\{1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \mu_3' \frac{t^3}{3!} + \cdots\right\} \\ &= \mu_k' + \mu_{k+1}' t + \mu_{k+2}' \frac{t^2}{2!} + \cdots\end{aligned}$$

by properties of the exponential series.

Moment-generating functions (cont.)

Therefore if we evaluate at $t = 0$, we have

$$m^{(k)}(0) = \mu'_k \qquad k = 1, 2, \ldots.$$

This result indicates that we can use the moment-generating function, if it is given, to compute the moments of $p(y)$.

This is the first main use of moment generating function.

Moment-generating functions (cont.)

The second main use is to *identify* distributions.

Suppose $m_1(t)$ and $m_2(t)$ are two moment-generating functions for pmfs $p_1(y)$ and $p_2(y)$. Then

$$m_1(t) = m_2(t) \text{ for all } |t| \leq b$$

for some b if and only if

$$p_1(y) = p_2(y) \qquad \text{for all } y.$$

Example (Poisson mgf)

We have for the Poisson distribution with parameter λ

$$\begin{aligned} m(t) = \sum_{y=0}^{\infty} e^{ty} p(y) &= \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y}{y!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!} \\ &= e^{-\lambda} \exp\{\lambda e^t\} \\ &= \exp\{\lambda(e^t - 1)\} \end{aligned}$$

Note that this holds for all t.

Example (Binomial mgf)

We have for the Binomial distribution with parameters n and p

$$
\begin{aligned}
m(t) = \sum_{y=0}^{n} e^{ty} p(y) &= \sum_{y=0}^{n} e^{ty} \binom{n}{y} p^y q^{n-y} \\
&= \sum_{y=0}^{n} \binom{n}{y} (pe^t)^y q^{n-y} \\
&= (pe^t + q)^n.
\end{aligned}
$$

Moment-generating functions (cont.)

Example (Binomial mgf)

If $p = \lambda/n$, then

$$\begin{aligned}(pe^t + q)^n &= \left(\frac{\lambda}{n}e^t + \left(1 - \frac{\lambda}{n}\right)\right)^n \\ &= \left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n \\ &\longrightarrow \exp\{\lambda(e^t - 1)\}\end{aligned}$$

as $n \longrightarrow \infty$, by definition of the exponential function.

Moment-generating functions (cont.)

Example (Geometric mgf)

We have for the geometric distribution with parameter p

$$
\begin{aligned}
m(t) = \sum_{y=1}^{\infty} e^{ty} p(y) &= \sum_{y=1}^{\infty} e^{ty} q^{y-1} p \\
&= pe^t \sum_{y=1}^{\infty} (qe^t)^{y-1} \\
&= pe^t \sum_{j=0}^{\infty} (qe^t)^{j} \\
&= pe^t \frac{1}{1-qe^t} = \frac{pe^t}{1-qe^t}
\end{aligned}
$$

Moment-generating functions (cont.)

Example

Suppose we are told that the mgf of a distribution takes the form

$$m(t) = \exp\{3(e^t - 1\};$$

then we may deduce that the corresponding distribution is a Poisson distribution with parameter $\lambda = 3$.

The Negative Binomial and the Poisson

Suppose $Y \sim NegBinomial(r, p)$; recall that Y is the random variable recording the number of independent $Bernoulli(p)$ trials required to obtain the rth success. The pmf for Y is

$$p_Y(y) = \binom{y-1}{r-1} p^r q^{y-r} \qquad y = r, r+1, r+2, \ldots$$

with $p_Y(y) = 0$ otherwise, where $q = 1 - p$. The mgf for Y is

$$m_Y(t) = \left(\frac{pe^t}{1 - qe^t}\right)^r = \left(\frac{(1-q)e^t}{1 - qe^t}\right)^r$$

defined when

$$|qe^t| < 1 \qquad \text{that is} \qquad -\ln q > t.$$

Alternate form: If $X = Y - r$, then

$$p_X(x) = \binom{x+r-1}{r-1} p^r q^x \qquad x = 0, 1, 2, \ldots$$

with $p_X(x) = 0$ otherwise. The mgf for X is

$$m_X(t) = e^{-rt} m_Y(t) = \left(\frac{1-q}{1-qe^t} \right)^r = \left(1 + \frac{q(e^t - 1)}{1 - qe^t} \right)^r.$$

The Negative Binomial and the Poisson (cont.)

Let $\lambda = rq$, so that $q = \lambda/r$. Then

$$m_X(t) = \left(1 + \frac{\lambda}{r}\frac{(e^t - 1)}{1 - (\lambda/r)e^t}\right)^r \qquad \ln r - \ln \lambda > t$$

and in the limit as $r \longrightarrow \infty$ with λ fixed,

$$m_X(t) \longrightarrow \exp\{\lambda(e^t - 1)\} \qquad t \in \mathbb{R}$$

Hence the limiting distribution of X is $\mathit{Poisson}(\lambda)$.

Probability-generating functions

A related function is the *probability-generating function* (pgf); this can be defined for a random variable Y that takes values on

$$\{0, 1, 2, \ldots\}$$

where

$$P(Y = i) = p_i.$$

The pgf, $G(t)$, is defined by

$$G(t) = \mathbb{E}\left[t^Y\right] = p_0 + p_1 t + p_2 t^2 + \cdots = \sum_{i=0}^{\infty} p_i t^i$$

whenever this sum is finite. Note that $G(1) = 1$, so we typically focus on evaluating $G(t)$ in a neighbourhood of 1, say $(1 - b, 1 + b)$ for some $b > 0$.

Probability-generating functions (cont.)

Note that

$$G^{(k)}(t) = \frac{d^k}{dt^k}\{G(t)\} = \frac{d^k}{dt^k}\left\{\sum_{i=0}^{\infty} p_i t^i\right\}$$

$$= \sum_{i=k}^{\infty} i(i-1)\dots(i-k+1)p_i t^{i-k}$$

so therefore

$$G^{(k)}(1) = \sum_{i=k}^{\infty} i(i-1)\dots(i-k+1)p_i$$

$$= \mathbb{E}[Y(Y-1)(Y-2)\dots(Y-k+1)].$$

Probability-generating functions (cont.)

The quantity

$$\mu_{[k]} = \mathbb{E}[Y(Y-1)(Y-2)\ldots(Y-k+1)]$$

is termed the k^{th} factorial moment.

Recall that we previously computed

$$\mu_{[2]} = \mathbb{E}[Y(Y-1)]$$

to calculate the variance of a distribution.

Probability-generating functions (cont.)

Note

As we have the identity

$$t = \exp\{\ln t\}$$

it follows that

$$G(t) \equiv \mathbb{E}\left[t^Y\right] = \mathbb{E}\left[\exp\{Y \ln t\}\right] = m(\ln t)$$

so once we have computed $m(t)$ we may immediately deduce $G(t)$.

Note

	$m(t)$	$G(t)$
Binomial	$(pe^t + q)^n$	$(pt + q)^n$
Poisson	$\exp\{\lambda(e^t - 1)\}$	$\exp\{\lambda(t - 1)\}$
Geometric	$\dfrac{pe^t}{1 - qe^t}$	$\dfrac{pt}{1 - qt}$

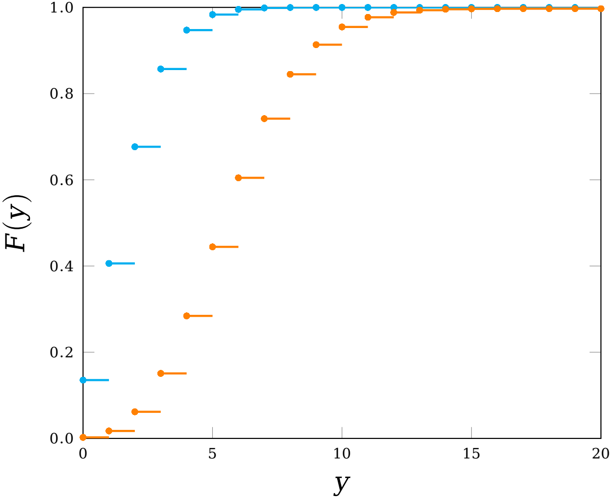
Continuous Random variables

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

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



Continuous Random variables (cont.)

Directly from the probability axioms, we can deduce three properties of the function $F(y)$:

1. "Starts at zero":

$$F(-\infty) \equiv \lim_{y \longrightarrow -\infty} F(y) = 0.$$

2. "Ends at 1":

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

3. "Non-decreasing in between": if $y_1 < y_2$, then

$$F(y_1) \leq F(y_2).$$

Continuous Random variables (cont.)

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

- height/weight,
- temperature,
- time,

and so on.

In this case it seems that the probability that Y lies in an interval $(y_1, y_2]$ for $y_1 < y_2$ must be positive, but should converge to zero the closer y_1 moves to y_2. That is

$$P(Y \in (y_1, y_2]) = P(Y \leq y_2) - P(Y \leq y_1) = F(y_2) - F(y_1)$$

should be positive, but decrease to zero as $y_1 \longrightarrow y_2$.

Continuous Random variables (cont.)

This is not the case for a discrete random variable: suppose

$$p(0) = \frac{1}{2} \qquad p(1) = \frac{1}{4} \qquad p(2) = \frac{1}{4}$$

and $p(y) = 0$ for all other values of y. Then for $0 < \epsilon < 1$

$$F(0) = P(Y \leq 0) = \frac{1}{2}$$

$$F(1 - \epsilon) = \frac{1}{2}$$

$$F(1) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Continuous Random variables (cont.)

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

That is, for any y,

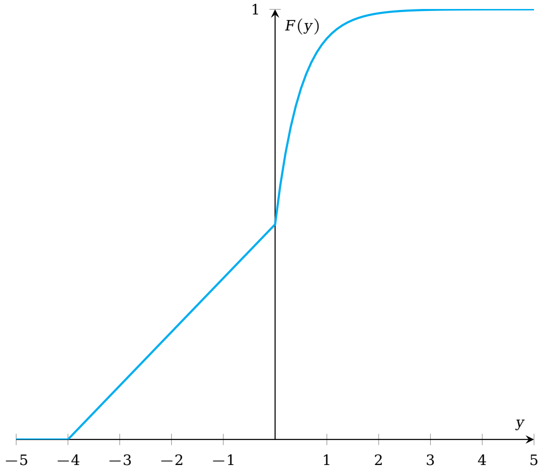
$$\lim_{\epsilon \longrightarrow 0+} F(y - \epsilon) = \lim_{\epsilon \longrightarrow 0+} F(y + \epsilon) = F(y).$$

Continuous Random variables (cont.)

We often also assume that $F(y)$ has a continuous first derivative for all $y \in \mathbb{R}$, except perhaps at a finite number of points.

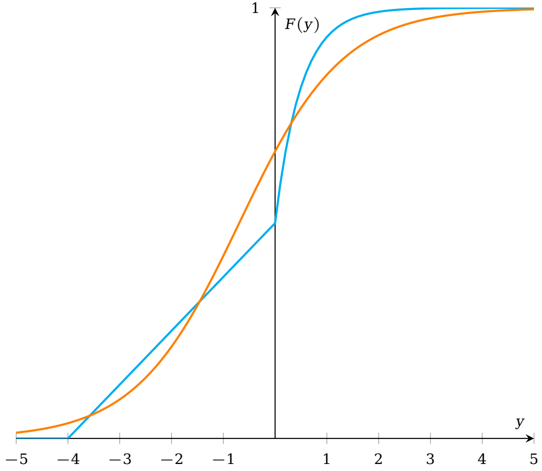
Continuous Random variables (cont.)

$F(y)$ continuous but not differentiable everywhere.



Continuous Random variables (cont.)

$F(y)$ continuous and differentiable.



Continuous Random variables (cont.)

Note

Even if Y is discrete, $F(y)$ is right-continuous.

That is, for any y,

$$\lim_{\epsilon \longrightarrow 0+} F(y+\epsilon) = F(y).$$

Continuous Random variables (cont.)

Important Note

If $F(y)$ is continuous, then

$$P(Y = y) = 0$$

for all $y \in \mathbb{R}$. If there were to exist a y_0 such that

$$P(Y = y_0) = p_0 > 0$$

then $F(y)$ would have a step of size p_0 at y_0, and would be discontinuous.

Probability density functions

For a continuous cdf $F(y)$, denote by $f(y)$ the first derivative of $F(.)$ at y

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

wherever this derivative exists.

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

Probability density functions (cont.)

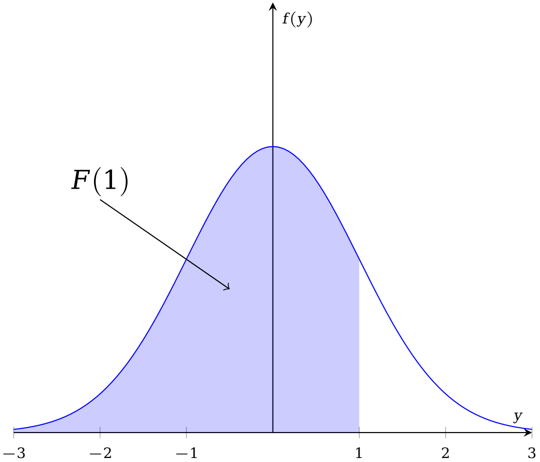
It follows that

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

Probability density functions (cont.)

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

Probability density functions (cont.)



Properties of density functions

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

1. Non-negative:

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

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

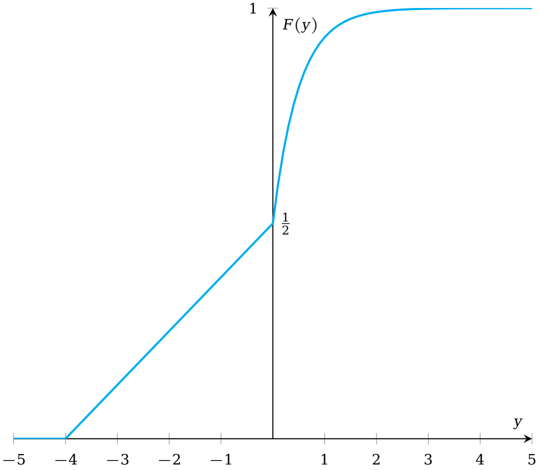
2. Integrates to 1:

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

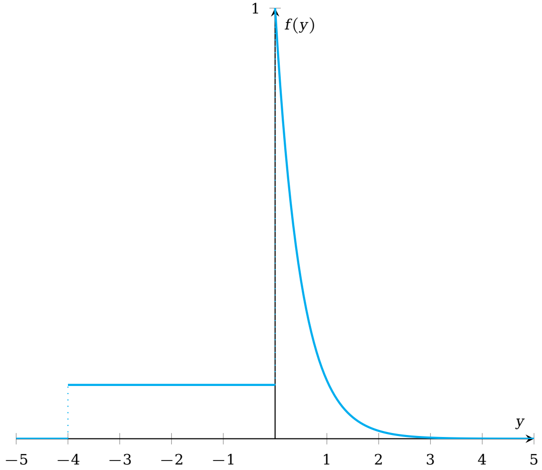
   – this follows as

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

Properties of density functions (cont.)



Properties of density functions (cont.)



Properties of density functions (cont.)

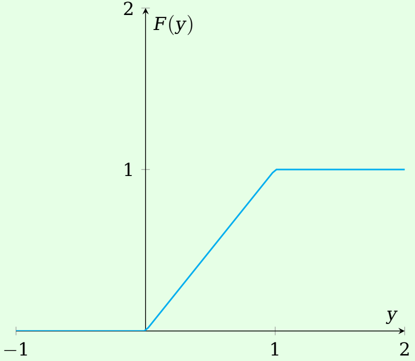
Example

Suppose

$$F(y) = \begin{cases} 0 & y \leq 0 \\ y & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

Properties of density functions (cont.)

Example



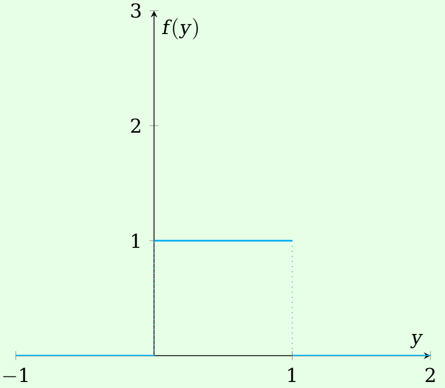
Properties of density functions (cont.)

Example

Then

$$f(y) = \begin{cases} 0 & y < 0 \\ 1 & 0 < y < 1 \\ 0 & y > 1 \end{cases}$$

Example



Properties of density functions (cont.)

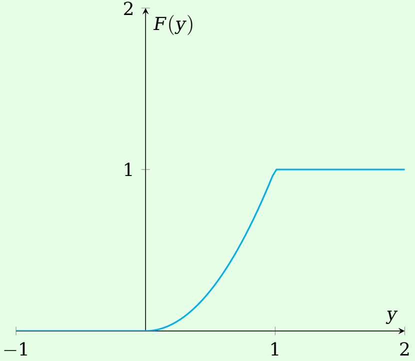
Example

Suppose

$$F(y) = \begin{cases} 0 & y \leq 0 \\ y^2 & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

Properties of density functions (cont.)

Example



Properties of density functions (cont.)

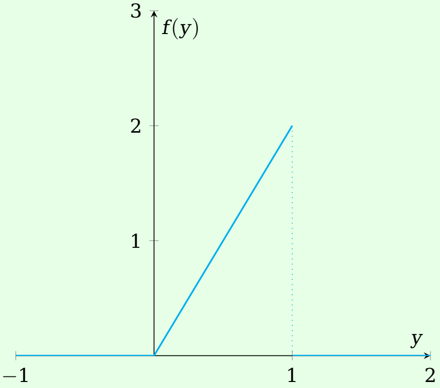
Example

Then

$$f(y) = \begin{cases} 0 & y < 0 \\ 2y & 0 < y < 1 \\ 0 & y > 1 \end{cases}$$

Properties of density functions (cont.)

Example



Properties of density functions (cont.)

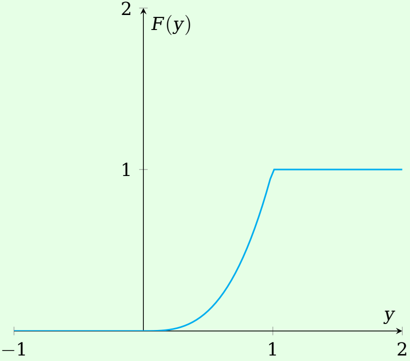
Example

Suppose

$$F(y) = \begin{cases} 0 & y \leq 0 \\ y^3 & 0 \leq y \leq 1 \\ 1 & y > 1 \end{cases}$$

Properties of density functions (cont.)

Example



Properties of density functions (cont.)

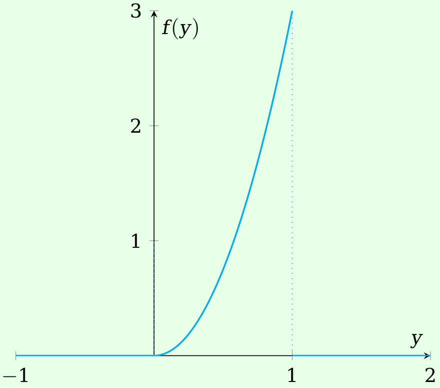
Example

Then

$$f(y) = \begin{cases} 0 & y < 0 \\ 3y^2 & 0 < y < 1 \\ 0 & y > 1 \end{cases}$$

Properties of density functions (cont.)

Example



Interval probabilities

To compute

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

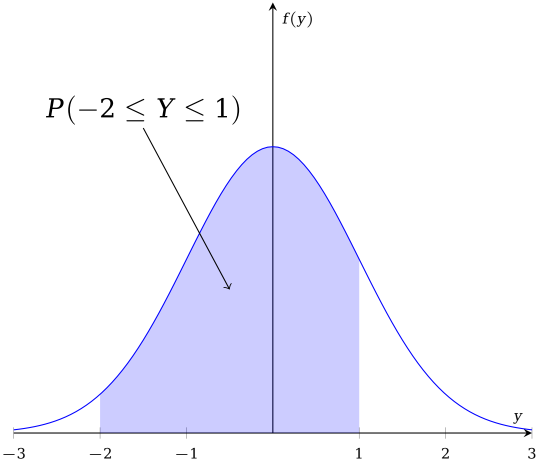
we integrate $f(y)$ over the interval (a, b) as

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

Interval probabilities (cont.)

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

Interval probabilities (cont.)



Interval probabilities (cont.)

Example

Suppose

$$f(y) = \begin{cases} 0 & y \leq 0 \\ cy^2 & 0 \leq y \leq 2 \\ 0 & y > 2 \end{cases}$$

for some $c > 0$.

Example

Then

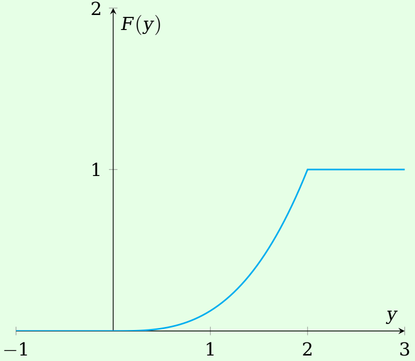
$$F(y) = \begin{cases} 0 & y \leq 0 \\ c\dfrac{y^3}{3} & 0 \leq y \leq 2 \\ c\dfrac{8}{3} & y \geq 2 \end{cases}$$

Therefore, as we require $F(\infty) = 1$, we must have

$$c = \frac{3}{8}.$$

Interval probabilities (cont.)

Example



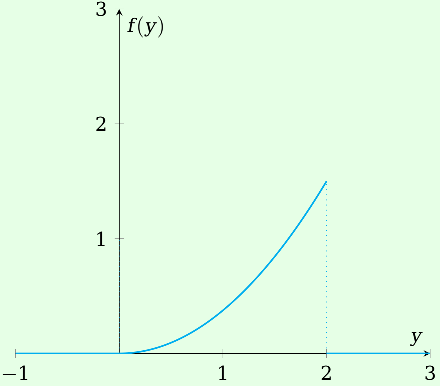
Example

Then

$$f(y) = \begin{cases} 0 & y < 0 \\ \dfrac{3}{8}y^2 & 0 < y < 2 \\ 0 & y > 2 \end{cases}$$

Interval probabilities (cont.)

Example



Interval probabilities (cont.)

To compute

$$P(1/2 \leq Y \leq 3/2)$$

we use the cdf to show that this is equal to

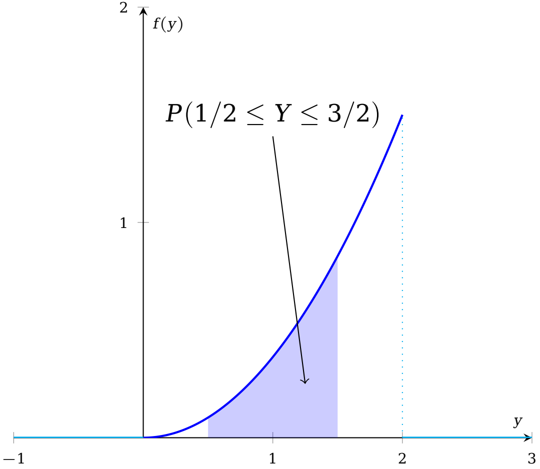
$$F(3/2) - F(1/2) = \frac{3}{8}\frac{1}{3}\left(\frac{3}{2}\right)^3 - \frac{3}{8}\frac{1}{3}\left(\frac{1}{2}\right)^3 = 0.40625$$

Interval probabilities (cont.)

This calculation is merely

$$P(1/2 \leq Y \leq 3/2) = \int_{1/2}^{3/2} \frac{3}{8} y^2 \, dy.$$

Interval probabilities (cont.)



Interval probabilities (cont.)

Note

For a *continuous* random variable

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

Interval probabilities (cont.)

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

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

Interval probabilities (cont.)

Note

All pdfs take the form

$$\text{constant} \times \text{function of } y$$

that is

$$f(y) = cg(y)$$

say, where $g(y)$ also contains information on the range of values, $\mathcal{Y}$, for which $f(y) > 0$. We have

$$\int_{-\infty}^{\infty} f(y)\, dy = 1 \qquad \Longrightarrow \qquad \int_{-\infty}^{\infty} g(y)\, dy = \frac{1}{c}.$$

Expectations

The expectation of continuous random variable Y is defined by

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

say, provided this integral *exists*.

If

$$\int_{-\infty}^{\infty} |y| f(y)\, dy < \infty$$

then it follows that $\mathbb{E}[Y]$ exists and is finite.

Expectations (cont.)

The expectation of a function of Y, $g(Y)$ say, is defined by

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)\,dy$$

if this integral is finite.

Properties of expectations

The key properties of expectations for discrete random variables also hold for continuous random variables. Specifically, if $g_1(.)$ and $g_2(.)$ are two functions, and a and b are constants then

$$\mathbb{E}[ag_1(Y) + bg_2(Y)] = a\mathbb{E}[g_1(Y)] + b\mathbb{E}[g_2(Y)]$$

provided $\mathbb{E}[g_1(Y)]$ and $\mathbb{E}[g_2(Y)]$ exist.

Properties of expectations (cont.)

Note

This result follows by the linearity properties of integrals.

$$\int (h_1(y) + h_2(y))\, dy = \int h_1(y)\, dy + \int h_2(y)\, dy$$

Properties of expectations (cont.)

We have identical definitions to the discrete case of the following quantities:

- Variance:

$$\mathbb{V}[X] = \mathbb{E}\left[(Y-\mu)^2\right] = \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\, dy$$

Properties of expectations (cont.)

- Moments: for $k = 1, 2, \dots$

$$\mu'_k = \mathbb{E}\left[Y^k\right] = \int_{-\infty}^{\infty} y^k f(y)\, dy$$

- Central moments: $k = 1, 2, \dots$

$$\mu_k = \mathbb{E}\left[(Y - \mu)^k\right] = \int_{-\infty}^{\infty} (y - \mu)^k f(y)\, dy$$

Properties of expectations (cont.)

- Factorial moments: $k = 1, 2, \ldots$

$$\begin{aligned}\mu_{[k]} &= \mathbb{E}\left[Y(Y-1)(Y-2)\cdots(Y-k+1)\right] \\ &= \int_{-\infty}^{\infty} y(y-1)(y-2)\cdots(y-k+1)f(y)\,dy\end{aligned}$$

Properties of expectations (cont.)

- Moment-generating function: for $|t| \leq b$, some $b > 0$

$$m(t) = \mathbb{E}\left[e^{tY}\right] = \int_{-\infty}^{\infty} e^{ty} f(y)\, dy$$

In addition, if

$$m^{(k)}(t) = \frac{d^k m(t)}{dt^k}$$

then as in the discrete case

$$m^{(k)}(0) = \mu'_k.$$

Properties of expectations (cont.)

- *Factorial moment-generating function:* for $1 - b \leq t \leq 1 + b$, some $b > 0$

$$G(t) = \mathbb{E}\left[t^Y\right] = \int_{-\infty}^{\infty} t^y f(y)\, dy$$

In addition, if

$$G^{(k)}(t) = \frac{d^k G(t)}{dt^k}$$

then as in the discrete case

$$G^{(k)}(1) = \mu_{[k]}.$$

Note: We do not use the term 'probability-generating function' in the continuous case.

Example

Suppose for some $\alpha > 0$, we have

$$f(y) = \begin{cases} 0 & y \leq 0 \\ \dfrac{c}{(1+y)^{\alpha+1}} & y > 0 \end{cases}$$

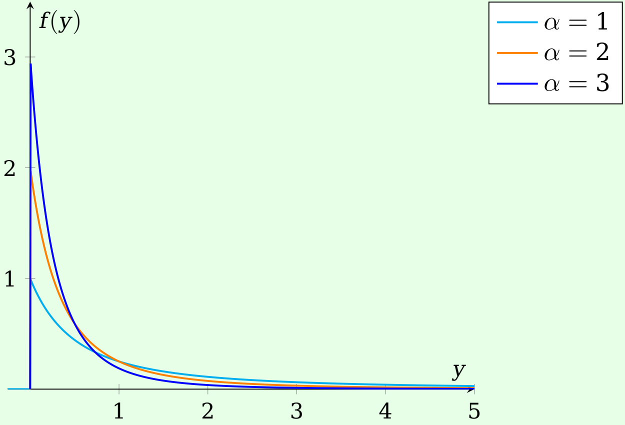
Example

Then

$$F(y) = \begin{cases} 0 & y \leq 0 \\ \\ 1 - \dfrac{c}{(1+y)^{\alpha}} & y > 0 \end{cases}$$

and as we need $F(\infty) = 1$, we must have $c = \alpha$.

Example



Properties of expectations (cont.)

Example

Then

$$\mu'_k = \mathbb{E}\left[Y^k\right] = \int_{-\infty}^{\infty} y^k f(y)\, dy$$

$$= \int_{0}^{\infty} y^k \frac{\alpha}{(1+y)^{\alpha+1}}\, dy$$

Properties of expectations (cont.)

Example

We inspect this integral and note that it is only finite if

$$\alpha > k.$$

If this holds, we can find μ'_k using integration by parts. For example

$$\mu'_1 = \mathbb{E}[Y] = \int_0^\infty y^k \frac{\alpha}{(1+y)^{\alpha+1}}\,dy = \frac{1}{\alpha - 1}$$

provided $\alpha > 1$.

Properties of expectations (cont.)

The quantity α in this pdf is termed a *parameter* of the model. It determines the properties of the model such as

- $f(y)$ and $F(y)$;
- $\mathbb{E}[Y]$ and other moments

The continuous uniform distribution

The continuous uniform distribution has a pdf that is constant on some interval (θ_1, θ_2) say (with $\theta_1 < \theta_2$):

$$f(y) = \frac{1}{\theta_2 - \theta_1} \qquad \theta_1 \leq y \leq \theta_2$$

with $f(y) = 0$ otherwise.

The continuous uniform distribution (cont.)

It is straightforward to show that

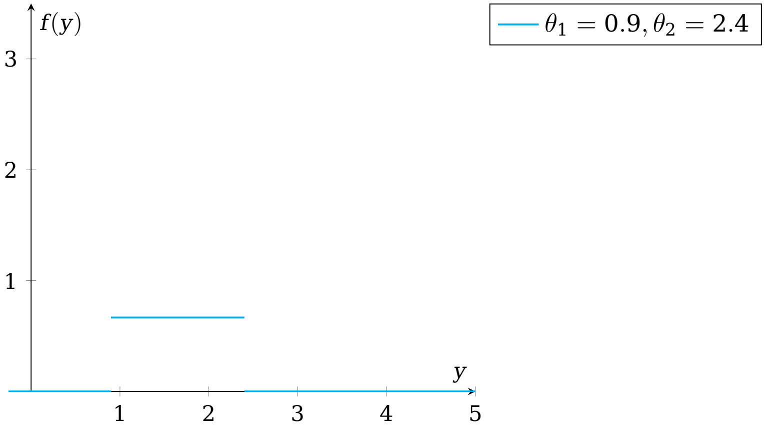
$$F(y) = \begin{cases} 0 & y \leq \theta_1 \\ \dfrac{y - \theta_1}{\theta_2 - \theta_1} & \theta_1 \leq y \leq \theta_2 \\ 1 & y \geq \theta_2 \end{cases} .$$

The continuous uniform distribution (cont.)

By direct calculation, we can compute that

$$\mathbb{E}[Y] = \frac{\theta_1 + \theta_2}{2} \qquad \mathbb{V}[Y] = \frac{(\theta_2 - \theta_1)^2}{12}.$$

The continuous uniform distribution



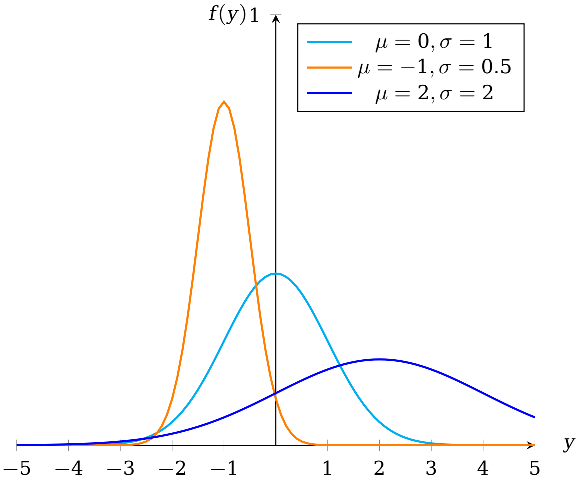
The Normal distribution

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

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

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

The Normal distribution (cont.)



The Normal distribution (cont.)

Notes

- 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

$$\mathbb{E}[Y] = \mu \qquad \mathbb{V}[Y] = \sigma^2.$$

The case where $\mu = 0$ and $\sigma = 1$ yields the standard Normal pmf

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

The Normal distribution (cont.)

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.

The Normal distribution (cont.)

However, 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 we see that for real value z

$$\begin{aligned} F_Z(z) = P(Z \leq z) &= P\left(\sigma Z + \mu \leq \sigma z + \mu\right) \\ &= P\left(Y \leq \sigma z + \mu\right) = F_Y\left(\sigma z + \mu\right) \end{aligned}$$

The quantity on the right-hand side is

$$\int_{-\infty}^{\sigma z+\mu} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}\, dt$$

but if we make the substitution $u = (t-\mu)/\sigma$ in the integral, we see that this becomes

$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\}\, du$$

The Normal distribution (cont.)

This is the integral of the standard Normal pdf, so we may conclude that the random variable Z

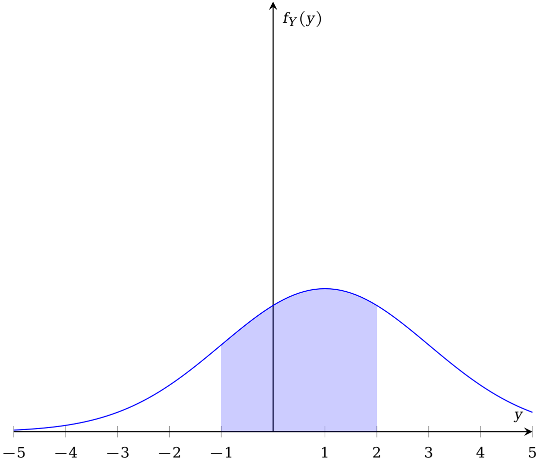
$$Z = \frac{Y - \mu}{\sigma}$$

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.

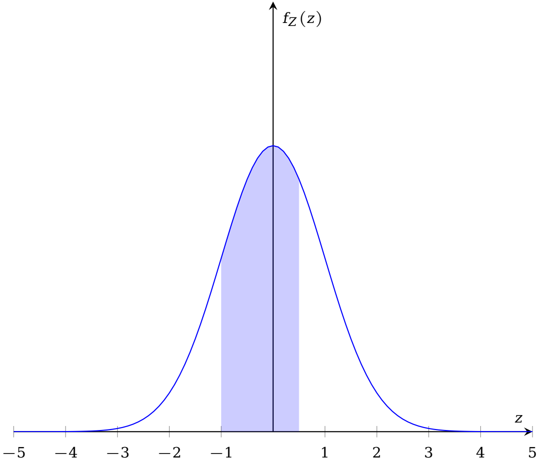
The Normal distribution (cont.)

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



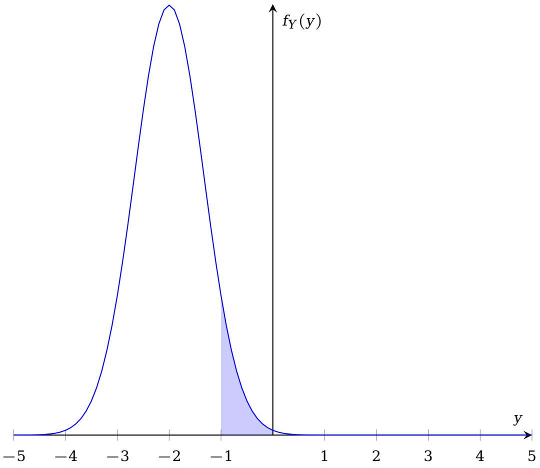
The Normal distribution (cont.)

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

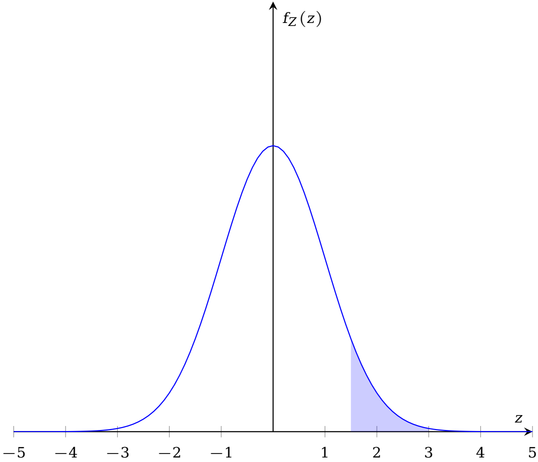


The Normal distribution (cont.)

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



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



The Normal distribution (cont.)

Note: Change of variable

We have by the change of variable rule that

$$\int_a^b g(t)dt = \int_c^d g(t(u))\frac{dt(u)}{du}du$$

where $t \longrightarrow u$ in the integral.

Moment-generating function

First consider the *standard Normal* mgf: we have for $t \in \mathbb{R}$ that

$$\begin{aligned} m(t) = \mathbb{E}\left[e^{tY}\right] &= \int_{-\infty}^{\infty} e^{ty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{ty - \frac{1}{2}y^2\right\}\, dy \\ &= \exp\left\{\frac{t^2}{2}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-t)^2}{2}\right\}\, dy \\ &= \exp\left\{\frac{t^2}{2}\right\} \end{aligned}$$

as the integral is the integral of a Normal pdf where the expectation is $\mu = t$.

The Gamma distribution has pdf defined on the whole real line by

$$f(y) = \begin{cases} 0 & y < 0 \\ \dfrac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} & 0 \leq y < \infty \end{cases}$$

where $\alpha, \beta > 0$ are parameters.

The Gamma distribution (cont.)

In this expression, $\Gamma(\alpha)$ is the *Gamma function*

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} \, dt.$$

This expression cannot be computed analytically, unless α is a positive integer. However, by integration by parts, we have that

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1).$$

Hence, if $n = 2, 3, 4, \ldots$, we have that

$$\Gamma(n) = (n - 1)!$$

with $\Gamma(1) = 1$ by direct calculation.

The Gamma distribution (cont.)

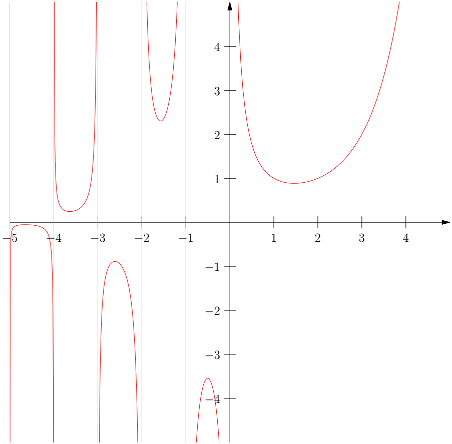
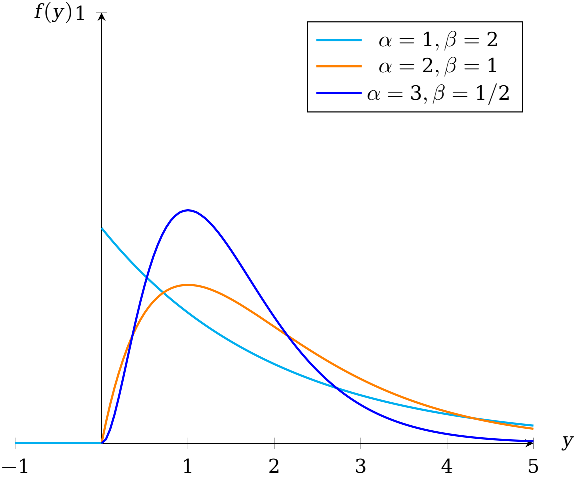


Figure: $\Gamma(\alpha)$ for $-5 \leq \alpha < 5$

The Gamma distribution (cont.)



The Gamma distribution (cont.)

Notes

- α is the shape parameter of the distribution;
- β is the scale parameter of the distribution;
- It can be shown that

$$\mathbb{E}[Y] = \alpha\beta \qquad \mathbb{V}[Y] = \alpha\beta^2.$$

For the calculation of the k^{th} moment of the distribution, we have to compute

$$\mathbb{E}[Y^k] = \int_{-\infty}^{\infty} y^k f(y)\, dy = \int_0^{\infty} y^k \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}\, dy.$$

Now, as we require $f(y)$ integrate to 1, this implies that

$$\int_0^\infty \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}\, dy = 1$$

for any $\alpha > 0$, so

$$\int_0^\infty y^{\alpha-1} e^{-y/\beta}\, dy = \beta^\alpha \Gamma(\alpha).$$

Returning to the moment calculation, we therefore deduce that

$$\begin{aligned}\int_0^\infty y^k \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta}\, dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^{k+\alpha-1} e^{-y/\beta}\, dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \beta^{k+\alpha} \Gamma(\alpha + k)\end{aligned}$$

from the previous result.

Furthermore

$$\begin{aligned}\Gamma(\alpha + k) &= (\alpha + k - 1)\Gamma(\alpha + k - 1) \\ &= (\alpha + k - 1)(\alpha + k - 2)\Gamma(\alpha + k - 2) \\ &= \cdots \\ &= (\alpha + k - 1)(\alpha + k - 2)\cdots(\alpha + 1)\alpha\Gamma(\alpha).\end{aligned}$$

The Gamma distribution (cont.)

Therefore

$$E[Y^k] = \frac{1}{\beta^\alpha \Gamma(\alpha)} \beta^{k+\alpha} (\alpha + k - 1)(\alpha + k - 2) \cdots (\alpha + 1) \alpha \Gamma(\alpha)$$

$$= \beta^k (\alpha + k - 1)(\alpha + k - 2) \cdots (\alpha + 1) \alpha$$

so that

$$\mathbb{E}[Y] = \alpha\beta \qquad \mathbb{E}[Y^2] = \alpha(\alpha + 1)\beta^2$$

and so on.

The Gamma distribution (cont.)

Note

Sometimes the Gamma distribution is written

$$\frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \quad 0 \leq y < \infty$$

that is, using a different parameterization.

The Gamma distribution (cont.)

Special Cases

1. $\alpha = \nu/2$, $\beta = 2$, where $\nu = 1, 2, \ldots$ is a positive integer. This is the *chi-square distribution* with ν degrees of freedom.

2. $\alpha = 1$. This is the *exponential distribution*

$$f(y) = \frac{1}{\beta} e^{-y/\beta} \quad 0 \leq y < \infty.$$

as $\Gamma(1) = 1$.

The Gamma distribution (cont.)

For the moment-generating function of the Gamma distribution:

$$
\begin{aligned}
m(t) = \mathbb{E}\left[e^{tY}\right] &= \int_0^\infty e^{ty} \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} e^{-y/\beta} \, dy \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y(1/\beta - t)} \, dy \\
&= \frac{1}{\beta^\alpha \Gamma(\alpha)} \Gamma(\alpha) \left(\frac{1}{\beta} - t\right)^{-\alpha} = \left(\frac{1}{1-\beta t}\right)^{\alpha}
\end{aligned}
$$

as the integral is proportional to a Gamma pdf with parameters α and

$$\left(\frac{1}{\beta} - t\right)^{-1}.$$

Connections

- *Geometric/Exponential:* Suppose

$$Y \sim Geometric(p)$$

and write $p = \lambda/n$. Let $X = Y/n$. Then as $n \longrightarrow \infty$

$$X \longrightarrow Exponential(1/\lambda)$$

- *Negative Binomial/Gamma:* Suppose

$$Y \sim NegativeBinomial(r, p),$$

and write $p = \lambda/n$. Let $X = Y/n$. Then as $n \longrightarrow \infty$

$$X \longrightarrow Gamma(r, 1/\lambda)$$

Connections (cont.)

- *Poisson process:* If $X(t)$ counts the number of events of a Poisson process with rate λ that occur in the interval of length t, then we have that

$$X(t) \sim Poisson(\lambda t).$$

  - If Y_1 is a continuous random variable that records the time until the *first* event, then

$$Y_1 \sim Exponential(1/\lambda).$$

  - If Y_k is a continuous random variable that records the time until the k^{th} event, then

$$Y_k \sim Gamma(k, \lambda) \qquad k = 1, 2, 3, \ldots$$

see Appendix.

The Beta distribution

The *Beta distribution* is suitable for a random variable defined on the range $[0, 1]$; its pdf takes the form

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \qquad 0 \leq y \leq 1$$

and $f(y) = 0$ for all other values of y. The constant

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

ensures that the pdf integrates to 1.

The Beta distribution (cont.)

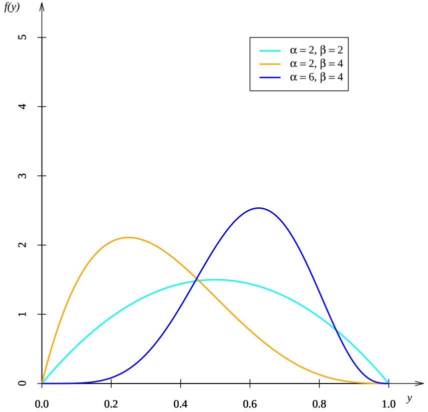
The *Beta function* is defined

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\, dy.$$

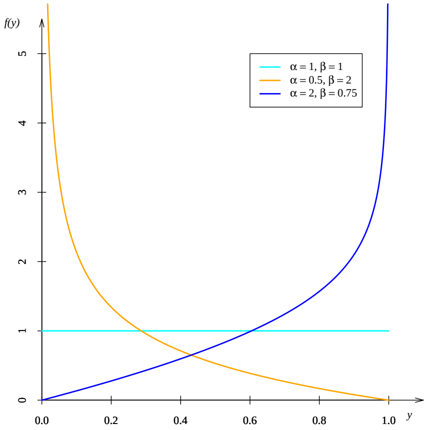
so we may write for the Beta pdf

$$f(y) = \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1} \qquad 0 \le y \le 1.$$

The Beta distribution (cont.)



The Beta distribution (cont.)



Note

The case $\alpha = \beta = 1$ has

$$B(\alpha, \beta) = \frac{\Gamma(1)\Gamma(1)}{\Gamma(2)} = \frac{1 \times 1}{1} = 1$$

and the pdf is

$$f(y) = 1 \qquad 0 \leq y \leq 1$$

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

This is the continuous uniform distribution with parameters $\theta_1 = 0$ and $\theta_2 = 1$.

The Beta distribution (cont.)

Note that for $k = 1, 2, \ldots$

$$\mathbb{E}\left[Y^k\right] = \int_{-\infty}^{\infty} y^k f(y)\, dy = \int_0^1 y^k \frac{1}{B(\alpha, \beta)} y^{\alpha-1}(1-y)^{\beta-1}\, dy$$

$$= \frac{1}{B(\alpha, \beta)} \int_0^1 y^{\alpha+k-1}(1-y)^{\beta-1}\, dy$$

$$= \frac{1}{B(\alpha, \beta)} B(\alpha + k, \beta)$$