Chapter 4: Multivariate distributions

Multivariate probability distributions

We now consider probability specifications for more than one variable at a time.

- We have already used the straightforward concept of *independent* random variables, that is, variables associated with independent events;
- for example, the random variables associated with the outcomes of two independent rolls of dice.

Multivariate probability distributions (cont.)

However, there are many instances where we wish to consider random variables where there is an absence of independence;

- for example where the outcome of one part of the experiment depends on the outcome of the other.

In general we can consider a *vector* of random variables (or *vector random variable*)

$$(Y_1, Y_2, \ldots, Y_n).$$

Multivariate probability distributions (cont.)

Note

The critical thing to remember is that when we distribute probability amongst the possible values of the random variables, the total probability we assign must still equal one.

Discrete bivariate distributions

We consider two variables Y_1 and Y_2 that are both discrete. We can suppose that both variables take values on the integers, $\mathbb{Z}$.

A *discrete bivariate probability mass function* is a function of two arguments

$$p(y_1, y_2)$$

that distributes probability across the possible values of the vector (Y_1, Y_2) so that

$$p(y_1, y_2) = P((Y_1 = y_1) \cap (Y_2 = y_2))$$

for $-\infty < y_1 < \infty$ and $-\infty < y_2 < \infty$.

Discrete bivariate distributions (cont.)

For shorthand, we write

$$P((Y_1 = y_1) \cap (Y_2 = y_2)) \equiv P(Y_1 = y_1, Y_2 = y_2)$$

Discrete bivariate distributions (cont.)

We can think of $p(y_1, y_2)$ as specifying the values in a probability table.

		Y_2		
		1	2	3
	1	$p(1,1)$	$p(1,2)$	$p(1,3)$
	2	$p(2,1)$	$p(2,2)$	$p(2,3)$
Y_1	3	$p(3,1)$	$p(3,2)$	$p(3,3)$
	4	$p(4,1)$	$p(4,2)$	$p(4,3)$
	5	$p(5,1)$	$p(5,2)$	$p(5,3)$

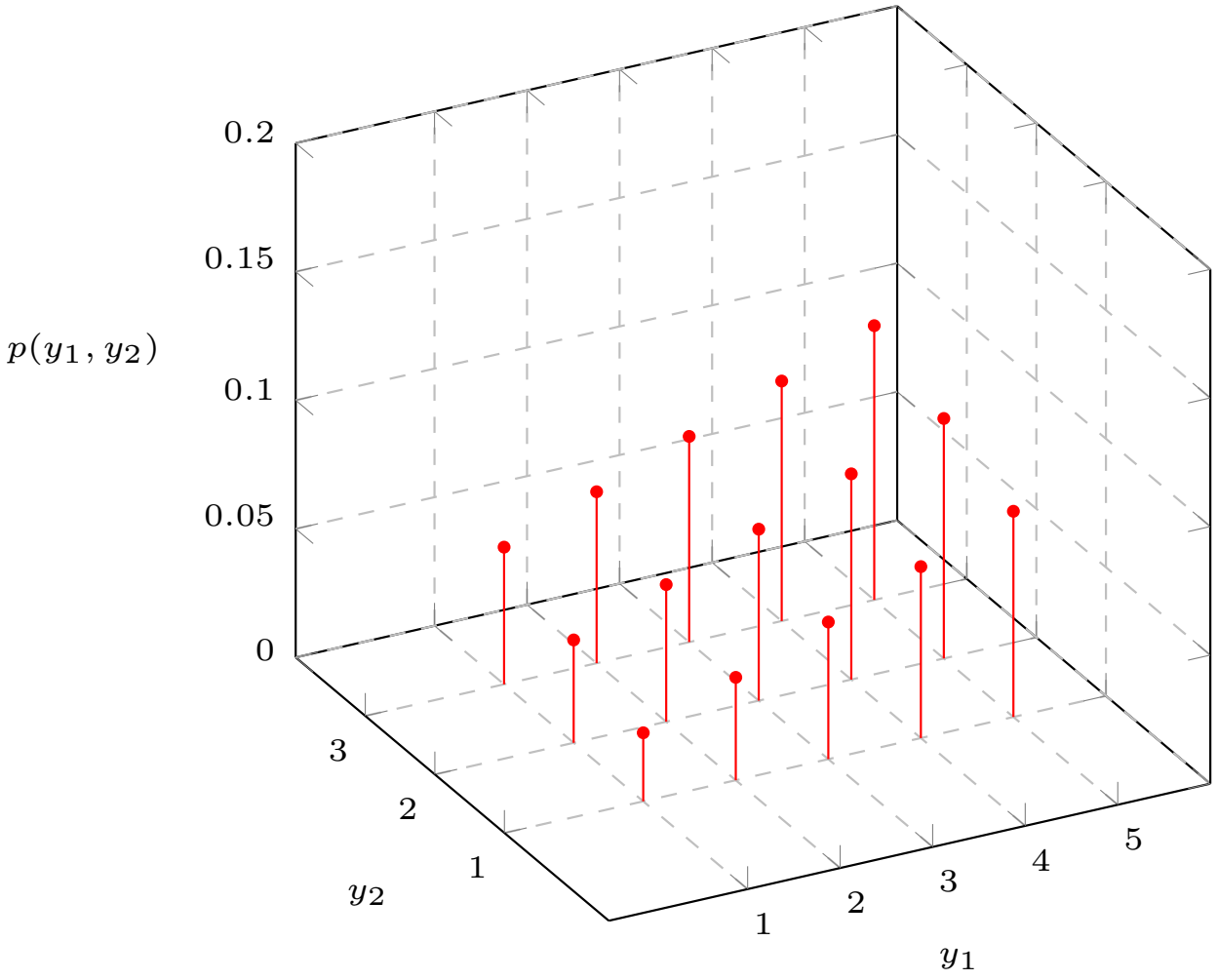
Discrete bivariate distributions (cont.)

For $1 \leq y_1 \leq 5, 1 \leq y_2 \leq 3$

$$p(y_1, y_2) = \frac{(y_1 + y_2)}{75}$$

		Y_2		
		1	2	3
	1	2/75	3/75	4/75
	2	3/75	4/75	5/75
Y_1	3	4/75	5/75	6/75
	4	5/75	6/75	7/75
	5	6/75	7/75	8/75

Discrete bivariate distributions (cont.)



Discrete bivariate distributions (cont.)

The function $p(y_1, y_2)$ is the *joint probability mass function*: it has two basic properties

- "specifies probabilities"

$$0 \leq p(y_1, y_2) \leq 1 \quad \text{for all } y_1, y_2$$

- "sums to one"

$$\sum_{y_1=-\infty}^{\infty} \sum_{y_2=-\infty}^{\infty} p(y_1, y_2) = 1.$$

although $p(y_1, y_2)$ may be zero for some arguments.

Discrete bivariate distributions (cont.)

In the above example,

$$\begin{aligned}
\sum_{y_1=1}^{5}\sum_{y_2=1}^{3}\frac{(y_1+y_2)}{75} &= \frac{1}{75}\sum_{y_1=1}^{5}\sum_{y_2=1}^{3}(y_1+y_2) \\
&= \frac{1}{75}\left[3\sum_{y_1=1}^{5}y_1 + 5\sum_{y_2=1}^{3}y_2\right] \\
&= \frac{1}{75}\left[3\frac{5\times 6}{2} + 5\frac{3\times 4}{2}\right] \\
&= \frac{1}{75}\left[45+30\right] \\
&= 1.
\end{aligned}$$

Discrete bivariate distributions (cont.)

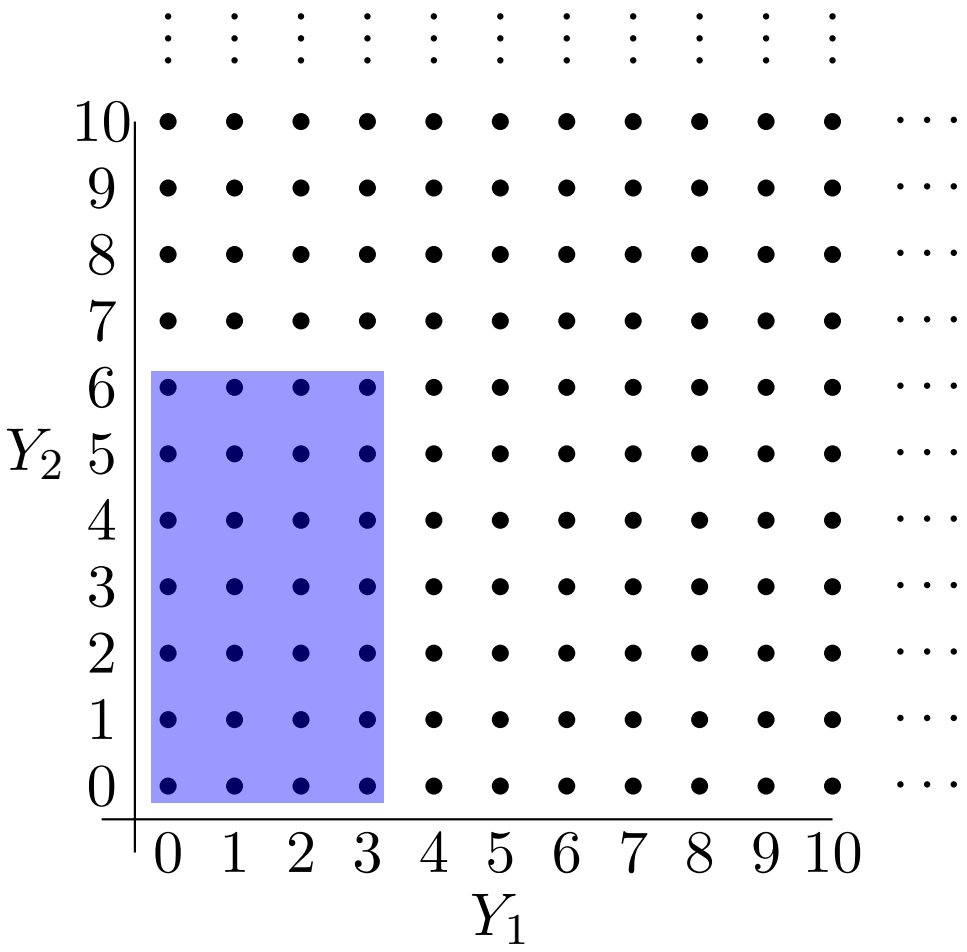
We define the *joint cumulative distribution function* $F(y_1, y_2)$ by

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2) = \sum_{t_1=-\infty}^{y_1} \sum_{t_2=-\infty}^{y_2} p(t_1, t_2)$$

that is, by accumulating the probabilities in the joint pmf over a range of values up to and including (y_1, y_2)

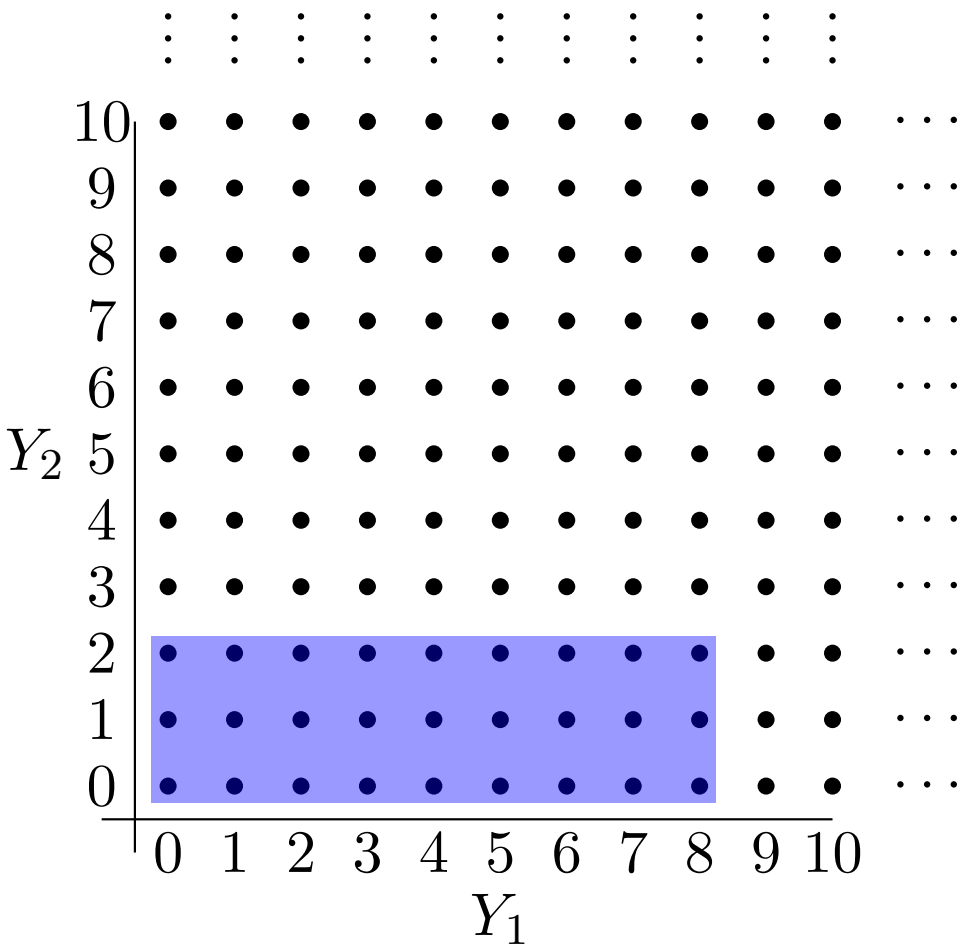
Discrete bivariate distributions (cont.)

$$F(3,6) = P(Y_1 \leq 3, Y_2 \leq 6) = \sum_{t_1=0}^{3} \sum_{t_2=0}^{6} p(t_1, t_2)$$



Discrete bivariate distributions (cont.)

$$F(8,2) = P(Y_1 \leq 8, Y_2 \leq 2) = \sum_{t_1=0}^{8} \sum_{t_2=0}^{2} p(t_1, t_2)$$



Discrete bivariate distributions (cont.)

Example

A bag contains ten balls:

- five red;
- three yellow;
- two white;

Four balls are selected, with all such selections being equally likely.

Discrete bivariate distributions (cont.)

Example

Let

- Y_1 denote the number of red balls selected;
- Y_2 denote the number of yellow balls selected.

Then using combinatorial arguments, we see that the joint pmf of Y_1 and Y_2 is given by

$$p_{Y_1,Y_2}(y_1, y_2) = \frac{\binom{5}{y_1}\binom{3}{y_2}\binom{2}{4 - y_1 - y_2}}{\binom{10}{4}}$$

for (y_1, y_2) such that the combinatorial terms are defined, and zero when the terms are not.

Discrete bivariate distributions (cont.)

Example

We need (y_1, y_2) simultaneously to satisfy

$$0 \leq y_1 \leq 5 \qquad 0 \leq y_2 \leq 3 \qquad 0 \leq 4 - y_1 - y_2 \leq 2$$

in order to have a non-zero probability.

Total number of selections: $\binom{10}{4} = 210$.

Discrete bivariate distributions (cont.)

Example

		Y_2			
		0	1	2	3
	0	0	0		
	1	0			
	2				0
Y_1	3			0	0
	4		0	0	0
	5	0	0	0	0

Red	Yellow	White
0	2	2
0	3	1
1	1	2
1	2	1
1	3	0
2	0	2
2	1	1
2	2	0
3	0	1
3	1	0
4	0	0

Discrete bivariate distributions (cont.)

Example

Red	Yellow	White	Count
0	2	2	3
0	3	1	2
1	1	2	15
1	2	1	30
1	3	0	5
2	0	2	10
2	1	1	60
2	2	0	30
3	0	1	20
3	1	0	30
4	0	0	5

Discrete bivariate distributions (cont.)

Example

		Y_2			
		0	1	2	3
	0	0.0000	0.0000	0.0143	0.0095
	1	0.0000	0.0714	0.1429	0.0238
	2	0.0476	0.2857	0.1429	0.0000
Y_1	3	0.0952	0.1429	0.0000	0.0000
	4	0.0238	0.0000	0.0000	0.0000
	5	0.0000	0.0000	0.0000	0.0000

Discrete bivariate distributions (cont.)

For more than two variables, we can define a joint pmf in the same way:

$$p(y_1, y_2, \ldots, y_n) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

to represent an n-way table.

The marginal mass function

Once we have specified a joint pmf, we can easily compute the probability distribution for one of the variables alone.

The *marginal probability mass function* for Y_1 is merely the pmf of Y_1, which is computed via the Theorem of Total Probability.

The marginal mass function (cont.)

Suppose that the joint pmf for Y_1 and Y_2 is denoted $p_{Y_1,Y_2}(.,.)$. The the marginal pmf for Y_1, $p_{Y_1}(.)$ is given by

$$p_{Y_1}(y_1) = P(Y_1 = y_1) = \sum_{y_2=-\infty}^{\infty} P(Y_1 = y_1, Y_2 = y_2)$$

that is

$$p_{Y_1}(y_1) = \sum_{y_2=-\infty}^{\infty} p_{Y_1,Y_2}(y_1, y_2)$$

The marginal mass function (cont.)

This result uses a partitioning argument:

$$(Y_1 = y_1) = \bigcup_{y_2=-\infty}^{\infty} (Y_1 = y_1) \cap (Y_2 = y_2)$$

The marginal mass function (cont.)

For example

$$P(Y_1 = 2) = P(Y_1 = 2, Y_2 = 1) + P(Y_1 = 2, Y_2 = 2) + P(Y_1 = 2, Y_2 = 3).$$

The marginal mass function (cont.)

If $p_{Y_1,Y_2}(y_1, y_2)$ specifies the values in a probability table, we compute the marginal pmf

- for Y_1 by summing across the rows of the table;
- for Y_2 by summing down the columns of the table.

		Y_2			
		1	2	3	$p_{Y_1}(.)$
Y_1	1	$p_{Y_1,Y_2}(1,1)$	$p_{Y_1,Y_2}(1,2)$	$p_{Y_1,Y_2}(1,3)$	$p_{Y_1}(1)$
	2	$p_{Y_1,Y_2}(2,1)$	$p_{Y_1,Y_2}(2,2)$	$p_{Y_1,Y_2}(2,3)$	$p_{Y_1}(2)$
	3	$p_{Y_1,Y_2}(3,1)$	$p_{Y_1,Y_2}(3,2)$	$p_{Y_1,Y_2}(3,3)$	$p_{Y_1}(3)$
	4	$p_{Y_1,Y_2}(4,1)$	$p_{Y_1,Y_2}(4,2)$	$p_{Y_1,Y_2}(4,3)$	$p_{Y_1}(4)$
	5	$p_{Y_1,Y_2}(5,1)$	$p_{Y_1,Y_2}(5,2)$	$p_{Y_1,Y_2}(5,3)$	$p_{Y_1}(5)$
	$p_{Y_2}(.)$	$p_{Y_2}(1)$	$p_{Y_2}(2)$	$p_{Y_2}(3)$	1

The marginal mass function (cont.)

Example (Previous example)

Four balls selected from 10.

- Y_1 denote the number of red balls selected;
- Y_2 denote the number of yellow balls selected.

The joint pmf of Y_1 and Y_2 is given by

$$p_{Y_1,Y_2}(y_1, y_2) = \frac{\binom{5}{y_1}\binom{3}{y_2}\binom{2}{4 - y_1 - y_2}}{\binom{10}{4}}$$

for (y_1, y_2) such that the combinatorial terms are defined, and zero when the terms are not.

The marginal mass function (cont.)

Example (Previous example)

We can compute the marginal pmf for Y_1 by summing probabilities in the joint probability table.

$$p_{Y_1}(0) = \frac{\binom{5}{0}\binom{3}{2}\binom{2}{2} + \binom{5}{0}\binom{3}{3}\binom{2}{1}}{\binom{10}{4}} = \frac{3+2}{210} = \frac{5}{210}$$

The marginal mass function (cont.)

Example (Previous example)

$$p_{Y_1}(1) = \frac{\binom{5}{1}\binom{3}{1}\binom{2}{2} + \binom{5}{1}\binom{3}{2}\binom{2}{1} + \binom{5}{1}\binom{3}{3}\binom{2}{0}}{\binom{10}{4}}$$

$$= \frac{15 + 30 + 5}{210} = \frac{50}{210}$$

The marginal mass function (cont.)

Example (Previous example)

$$p_{Y_1}(2) = \frac{\binom{5}{2}\binom{3}{0}\binom{2}{2} + \binom{5}{2}\binom{3}{1}\binom{2}{1} + \binom{5}{2}\binom{3}{0}\binom{2}{2}}{\binom{10}{4}}$$

$$= \frac{10 + 60 + 30}{210} = \frac{100}{210}$$

The marginal mass function (cont.)

Example (Previous example)

$$p_{Y_1}(3) = \frac{\binom{5}{3}\binom{3}{0}\binom{2}{1} + \binom{5}{3}\binom{3}{1}\binom{2}{0}}{\binom{10}{4}}$$

$$= \frac{20+30}{210} = \frac{50}{210}$$

The marginal mass function (cont.)

Example (Previous example)

$$p_{Y_1}(4) = \frac{\binom{5}{4}\binom{3}{0}\binom{2}{0}}{\binom{10}{4}} = \frac{5}{210}$$

The marginal mass function (cont.)

Note

In this example we can compute $p_{Y_1}(.)$ directly using the hypergeometric formula

$$p_{Y_1}(y_1) = \frac{\binom{5}{y_1}\binom{5}{4-y_1}}{\binom{10}{4}}$$

for $0 \leq y_1 \leq 5$ and $0 \leq 4 - y_1 \leq 5$.

The marginal mass function (cont.)

Note

y_1	Numerator
0	$\binom{5}{0}\binom{5}{4} = 1 \times 5 = 5$
1	$\binom{5}{1}\binom{5}{2} = 5 \times 10 = 50$
2	$\binom{5}{2}\binom{5}{3} = 10 \times 10 = 100$
3	$\binom{5}{3}\binom{5}{1} = 10 \times 5 = 50$
4	$\binom{5}{4}\binom{5}{0} = 5 \times 1 = 5$

The marginal mass function (cont.)

Note

The marginal pmfs $p_{Y_1}(y_1)$ and $p_{Y_2}(y_2)$ have all the properties of single variable pmfs; specifically

$$\sum_{y_1=-\infty}^{\infty} p_{Y_1}(y_1) = \sum_{y_1=-\infty}^{\infty} \sum_{y_2=-\infty}^{\infty} p_{Y_1,Y_2}(y_1, y_2) = 1.$$

The conditional mass function

Once we have the joint pmf

$$p_{Y_1,Y_2}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$$

and the marginal pmf

$$p_{Y_1}(y_1) = P(Y_1 = y_1)$$

we can consider *conditional* pmfs.

The conditional mass function (cont.)

We have that if $P(Y_1 = y_1) > 0$, then the conditional probability that $Y_2 = y_2$, given that $Y_1 = y_1$, is

$$P(Y_2 = y_2 | Y_1 = y_1) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_1 = y_1)}$$

For a **fixed** value of y_1, we can consider how this conditional probability varies as argument y_2 varies.

The conditional mass function (cont.)

The *conditional probability mass function* for Y_2, given that $Y_1 = y_1$, is denoted

$$p_{Y_2|Y_1}(y_2|y_1)$$

and defined by

$$p_{Y_2|Y_1}(y_2|y_1) = P(Y_2 = y_2|Y_1 = y_1)$$

whenever $P(Y_1 = y_1) > 0$.

The conditional mass function (cont.)

The conditional pmfs are obtained by taking 'slices' through the joint pmf, and then standardizing the slice so that the probabilities sum to one.

Recall that

$$p_{Y_2|Y_1}(y_2|y_1) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_1 = y_1)} = \frac{p_{Y_1,Y_2}(y_1, y_2)}{p_{Y_1}(y_1)}$$

$$= \frac{p_{Y_1,Y_2}(y_1, y_2)}{\sum_{t_2=-\infty}^{\infty} p_{Y_1,Y_2}(y_1, t_2)}.$$

The conditional mass function (cont.)

Given $Y_1 = 2$, we define the conditional pmf $p_{Y_2|Y_1}(y_2|2)$ by examining the second row of the table.

		Y_2			
		1	2	3	$p_{Y_1}(.)$
	1	$p_{Y_1,Y_2}(1,1)$	$p_{Y_1,Y_2}(1,2)$	$p_{Y_1,Y_2}(1,3)$	$p_{Y_1}(1)$
	2	$p_{Y_1,Y_2}(2,1)$	$p_{Y_1,Y_2}(2,2)$	$p_{Y_1,Y_2}(2,3)$	$p_{Y_1}(2)$
Y_1	3	$p_{Y_1,Y_2}(3,1)$	$p_{Y_1,Y_2}(3,2)$	$p_{Y_1,Y_2}(3,3)$	$p_{Y_1}(3)$
	4	$p_{Y_1,Y_2}(4,1)$	$p_{Y_1,Y_2}(4,2)$	$p_{Y_1,Y_2}(4,3)$	$p_{Y_1}(4)$
	5	$p_{Y_1,Y_2}(5,1)$	$p_{Y_1,Y_2}(5,2)$	$p_{Y_1,Y_2}(5,3)$	$p_{Y_1}(5)$
	$p_{Y_2}(.)$	$p_{Y_2}(1)$	$p_{Y_2}(2)$	$p_{Y_2}(3)$	1

The conditional mass function (cont.)

Given $Y_2 = 3$, we define the conditional pmf $p_{Y_1|Y_2}(y_1|3)$ by examining the third column of the table.

		Y_2			
		1	2	3	$p_{Y_1}(.)$
Y_1	1	$p_{Y_1,Y_2}(1,1)$	$p_{Y_1,Y_2}(1,2)$	$p_{Y_1,Y_2}(1,3)$	$p_{Y_1}(1)$
	2	$p_{Y_1,Y_2}(2,1)$	$p_{Y_1,Y_2}(2,2)$	$p_{Y_1,Y_2}(2,3)$	$p_{Y_1}(2)$
	3	$p_{Y_1,Y_2}(3,1)$	$p_{Y_1,Y_2}(3,2)$	$p_{Y_1,Y_2}(3,3)$	$p_{Y_1}(3)$
	4	$p_{Y_1,Y_2}(4,1)$	$p_{Y_1,Y_2}(4,2)$	$p_{Y_1,Y_2}(4,3)$	$p_{Y_1}(4)$
	5	$p_{Y_1,Y_2}(5,1)$	$p_{Y_1,Y_2}(5,2)$	$p_{Y_1,Y_2}(5,3)$	$p_{Y_1}(5)$
	$p_{Y_2}(.)$	$p_{Y_2}(1)$	$p_{Y_2}(2)$	$p_{Y_2}(3)$	1

The conditional mass function (cont.)

Note

We have the fundamental relationship

$$p_{Y_1,Y_2}(y_1, y_2) = p_{Y_1}(y_1)p_{Y_2|Y_1}(y_2|y_1)$$

whenever $p_{Y_1}(y_1) > 0$.

Independence

Two discrete random variables Y_1 and Y_2 are called *independent* if

$$p_{Y_1,Y_2}(y_1, y_2) = p_{Y_1}(y_1)p_{Y_2}(y_2)$$

for **all** values y_1, y_2. Equivalently, they are independent if

$$p_{Y_1|Y_2}(y_1|y_2) = p_{Y_1}(y_1) \qquad p_{Y_2|Y_1}(y_2|y_1) = p_{Y_2}(y_2)$$

for **all** values y_1, y_2 where the conditional pmfs are defined.

If Y_1 and Y_2 are not independent, they are *dependent*.

Independence (cont.)

This definition extends to multiple variables: $Y_1, \ldots, Y_n$ are independent if

$$p_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \prod_{i=1}^{n} p_{Y_i}(y_i)$$

for **all** vectors $(y_1, \ldots, y_n)$.

Note that we could make equivalent statements in terms of the joint cdf

$$F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = P\left(\bigcap_{i=1}^{n}(Y_i \leq y_i)\right) = \prod_{i=1}^{n} F_{Y_i}(y_i)$$

The joint density function

If Y_1 and Y_2 are two *continuous* random variables, then we can still consider statements of the form

$$P((Y_1 \leq y_1) \cap (Y_2 \leq y_2))$$

and hence define the *joint cumulative distribution function* cdf

$$F_{Y_1,Y_2}(y_1, y_2) = P((Y_1 \leq y_1) \cap (Y_2 \leq y_2))$$

for any pair of real numbers (y_1, y_2).

The joint density function (cont.)

The joint cdf has the following properties:

- "starts at zero"

$$\lim_{y_1 \longrightarrow -\infty} \lim_{y_2 \longrightarrow -\infty} F_{Y_1,Y_2}(y_1, y_2) = 0$$

- "ends at one"

$$\lim_{y_1 \longrightarrow \infty} \lim_{y_2 \longrightarrow \infty} F_{Y_1,Y_2}(y_1, y_2) = 1$$

- "non-decreasing in y_1 and y_2 in between"

$$F_{Y_1,Y_2}(y_1, y_2) \leq F_{Y_1,Y_2}(y_1 + c, y_2)$$

$$F_{Y_1,Y_2}(y_1, y_2) \leq F_{Y_1,Y_2}(y_1, y_2 + c)$$

for all y_1, y_2, and any $c > 0$.

The joint density function (cont.)

Furthermore, we have that

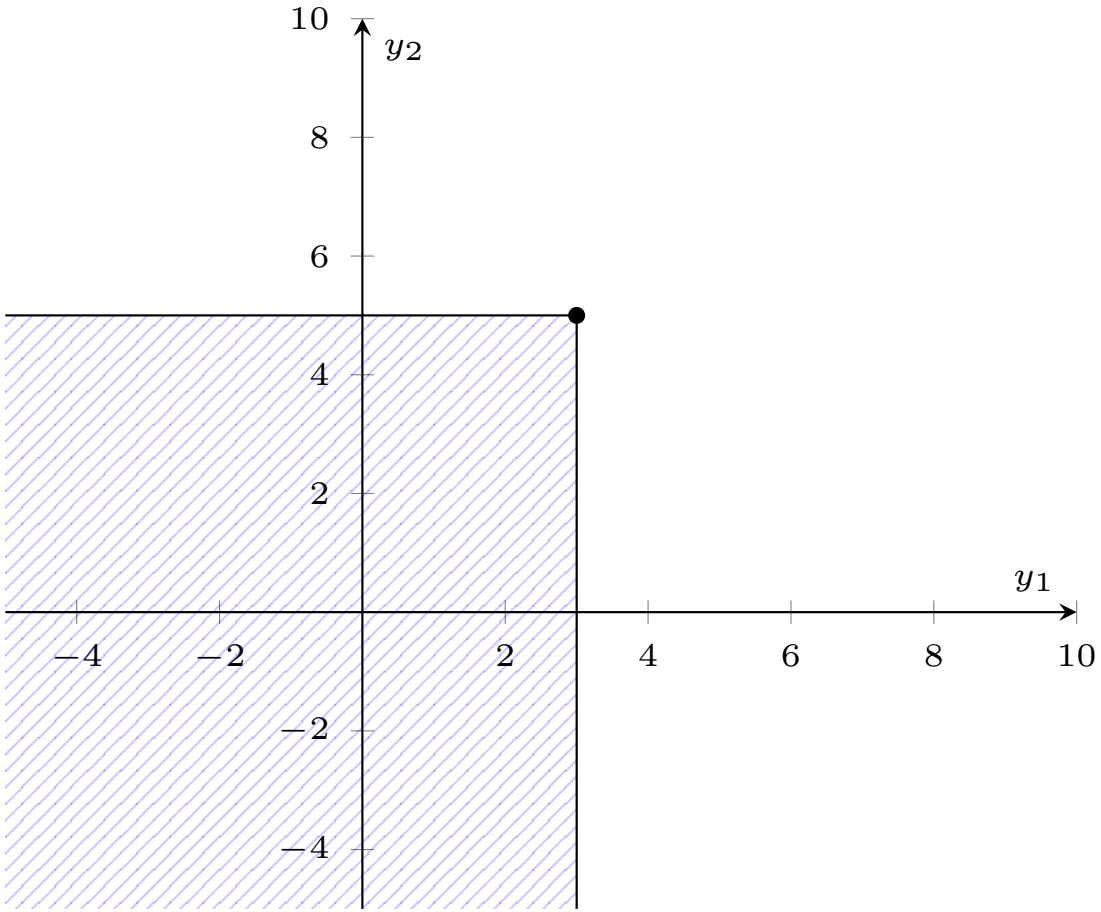
$$\begin{aligned}\lim_{y_1 \longrightarrow \infty} F_{Y_1,Y_2}(y_1, y_2) &= P(Y_1 < \infty, Y_2 \leq y_2) \\ &= P(Y_2 \leq y_2) \\ &= F_{Y_2}(y_2)\end{aligned}$$

and similarly

$$\lim_{y_2 \longrightarrow \infty} F_{Y_1,Y_2}(y_1, y_2) = F_{Y_1}(y_1).$$

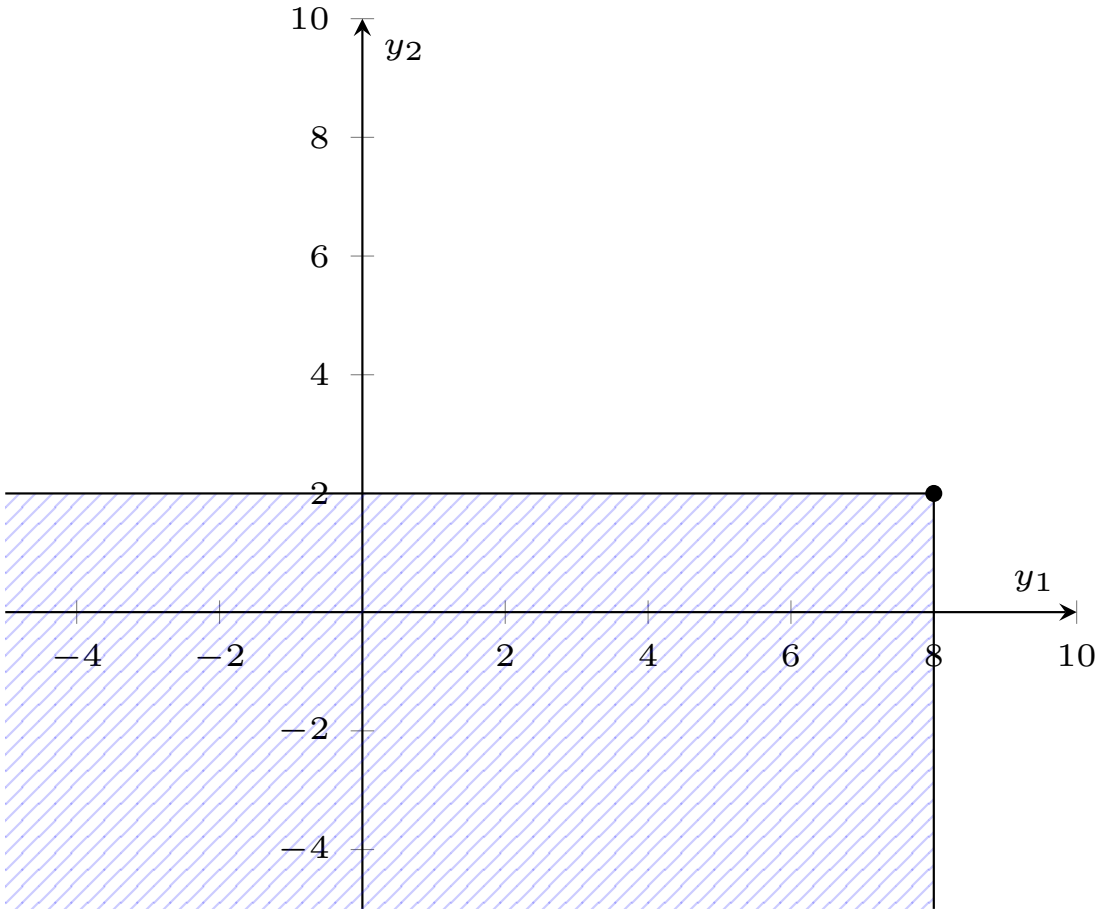
The joint density function (cont.)

$$F_{Y_1,Y_2}(3,5) = P(Y_1 \leq 3, Y_2 \leq 5)$$



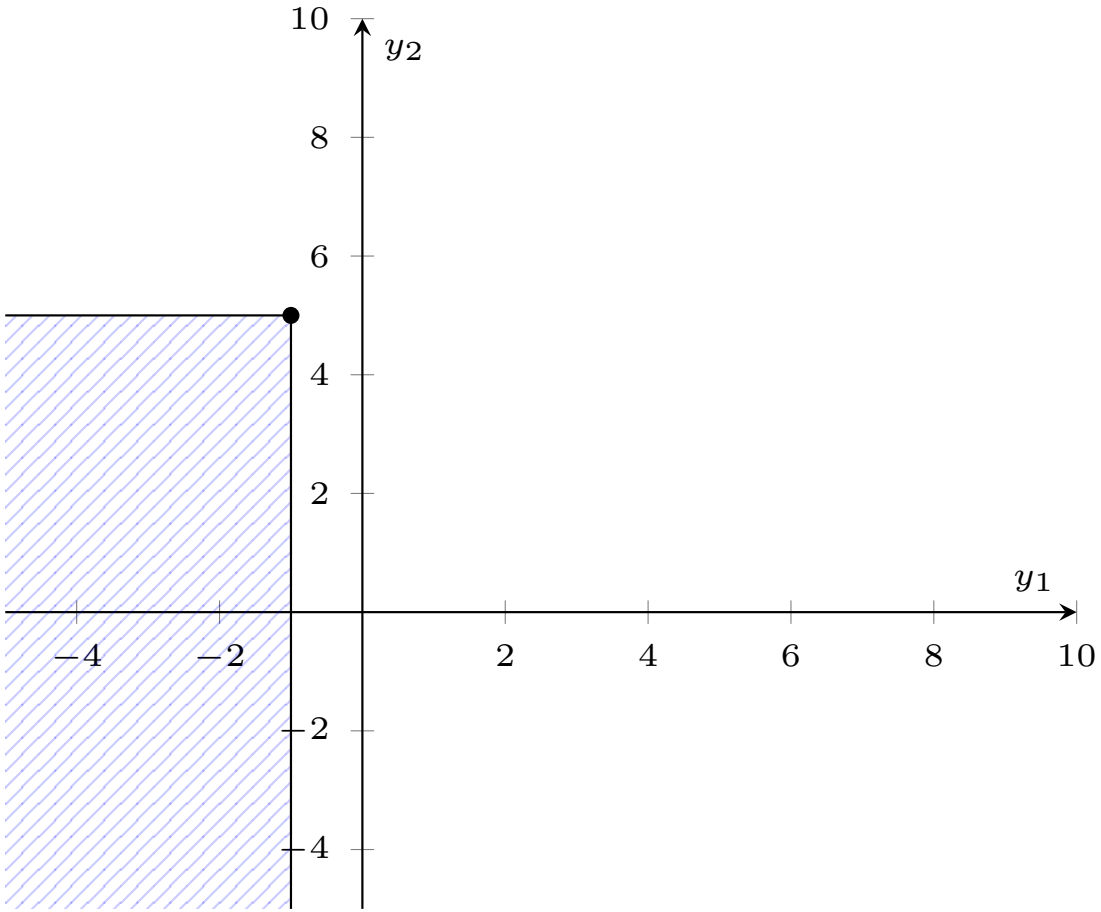
The joint density function (cont.)

$$F_{Y_1,Y_2}(8,2) = P(Y_1 \leq 8, Y_2 \leq 2)$$



The joint density function (cont.)

$$F_{Y_1,Y_2}(-1,5) = P(Y_1 \leq -1, Y_2 \leq 5)$$



The joint density function (cont.)

We accumulate probability over the shaded region, the rectangle

$$(-\infty, y_1] \times (-\infty, y_2],$$

to compute $F_{Y_1,Y_2}(y_1, y_2)$.

However, as in the single variable case, we must have

$$P(Y_1 = y_1, Y_2 = y_2) = 0$$

for all y_1 and y_2.

The joint density function (cont.)

As in the single variable case, we introduce the *joint probability density function* (joint pdf)

$$f_{Y_1,Y_2}(y_1, y_2)$$

to describe how probability is spread around the possible values, where

$$F_{Y_1,Y_2}(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2 dt_1$$

that is, to compute $F_{Y_1,Y_2}(y_1, y_2)$ we integrate $f_{Y_1,Y_2}(y_1, y_2)$ over the rectangle

$$(-\infty, y_1] \times (-\infty, y_2].$$

The joint density function (cont.)

We can compute the double integral as follows: writing

$$\int_{-\infty}^{y_1} \left\{ \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2 \right\} dt_1$$

- fix t_1, and perform the first (inner) integration

$$\int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2$$

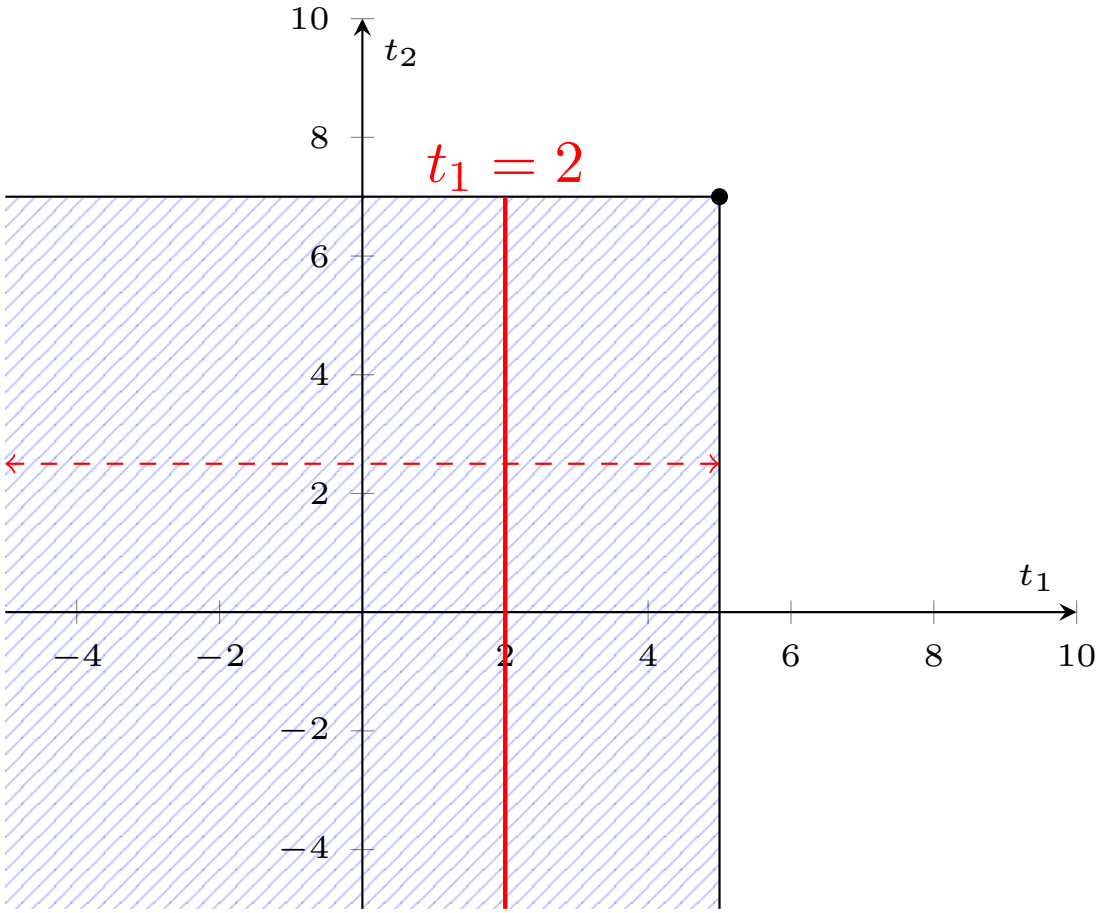
  in the 'strip' at t_1 to obtain a function $g(t_1, y_2)$, say;
- perform the second (outer) integration

$$\int_{-\infty}^{y_1} g(t_1, y_2)\, dt_1.$$

  to obtain the joint cdf.

The joint density function (cont.)

$$F_{Y_1,Y_2}(5,7) = P(Y_1 \leq 5, Y_2 \leq 7)$$



The joint density function (cont.)

The joint pdf describes how the probability is spread 'point-by-point' across the real plane. By the probability axioms, we must have that

- the joint pdf is non-negative

$$f_{Y_1,Y_2}(y_1, y_2) \geq 0 \qquad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

  (as the joint cdf is non-decreasing in both y_1 and y_2);

- the joint pdf integrates to 1

$$\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 \right\} dy_1 = 1$$

  (as the probability must accumulate to 1 over the real plane).

The joint density function (cont.)

Example (Distribution on the unit square)

Suppose Y_1 and Y_2 are continuous with joint pdf

$$f_{Y_1,Y_2}(y_1, y_2) = c(y_1 + y_2) \qquad 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1$$

with $f_{Y_1,Y_2}(y_1, y_2) = 0$ otherwise.

The joint density function (cont.)

Example (Distribution on the unit square)

Then for $0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1$,

$$\begin{aligned}
F_{Y_1,Y_2}(y_1, y_2) &= \int_{-\infty}^{y_1} \left\{ \int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2 \right\} dt_1 \\
&= \int_{0}^{y_1} \left\{ \int_{0}^{y_2} c(t_1 + t_2)\, dt_2 \right\} dt_1 \\
&= \int_{0}^{y_1} \left[c\left(t_1 t_2 + \frac{1}{2} t_2^2 \right) \right]_0^{y_2} dt_1 \\
&= \int_{0}^{y_1} c\left(t_1 y_2 + \frac{1}{2} y_2^2 \right) dt_1
\end{aligned}$$

The joint density function (cont.)

Example (Distribution on the unit square)

$$= \left[c \left(\frac{1}{2} t_1^2 y_2 + \frac{1}{2} y_2^2 t_1 \right) \right]_0^{y_1}$$

$$= \frac{c}{2} \left(y_1^2 y_2 + y_1 y_2^2 \right)$$

We require that $F_{Y_1,Y_2}(1,1) = 1$, so we must have $c = 1$.

The joint density function (cont.)

Example (Distribution on the unit square)

That is

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} (y_1+y_2) & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(no probability outside of the unit square).

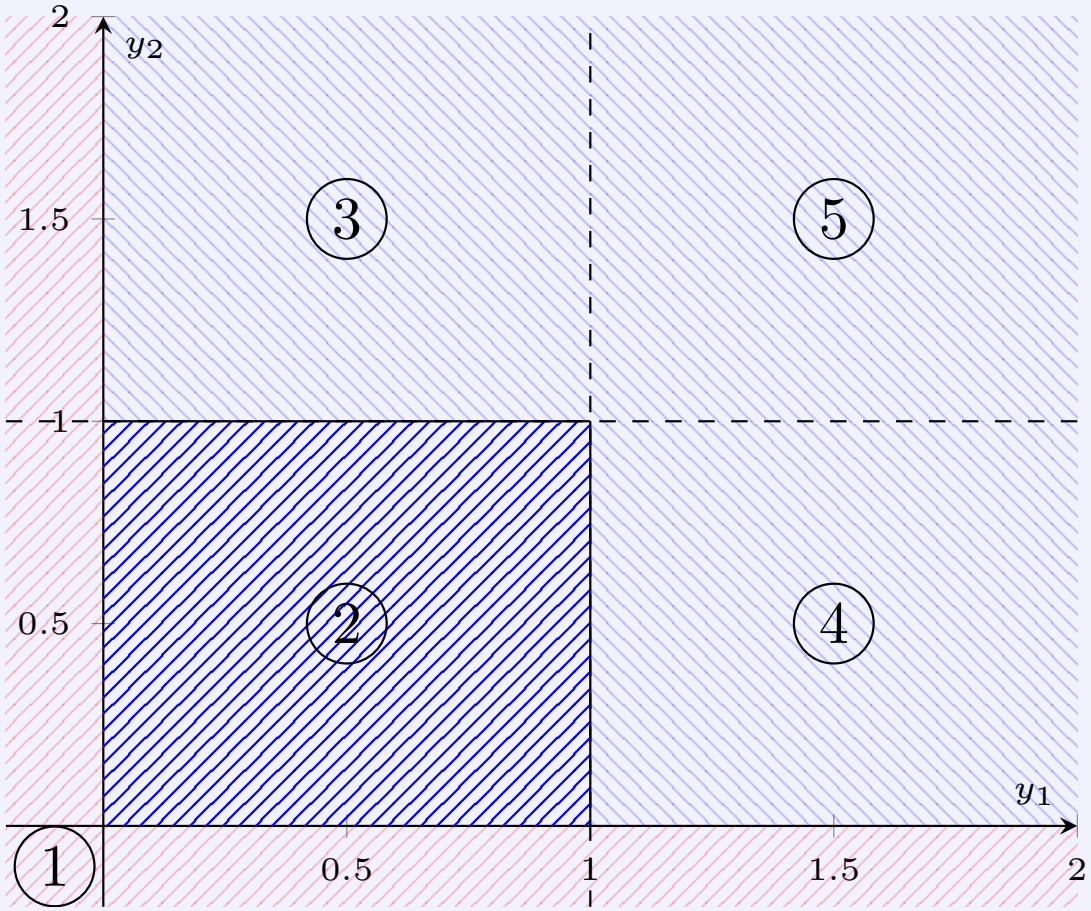
The joint density function (cont.)

Example (Distribution on the unit square)

$$F_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 0 & y_1 < 0 \text{ or } y_2 < 0 \quad (1) \\ \left(y_1^2 y_2 + y_1 y_2^2\right)/2 & 0 \le y_1, y_2 \le 1 \quad (2) \\ \left(y_1^2 + y_1\right)/2 & 0 \le y_1 \le 1, y_2 > 1 \quad (3) \\ \left(y_2 + y_2^2\right)/2 & 0 \le y_2 \le 1, y_1 > 1 \quad (4) \\ 1 & y_1 > 1 \text{ and } y_2 > 1 \quad (5) \end{cases}$$

The joint density function (cont.)

Note: regions for $F_{Y_1,Y_2}(y_1, y_2)$



The joint density function (cont.)

Note

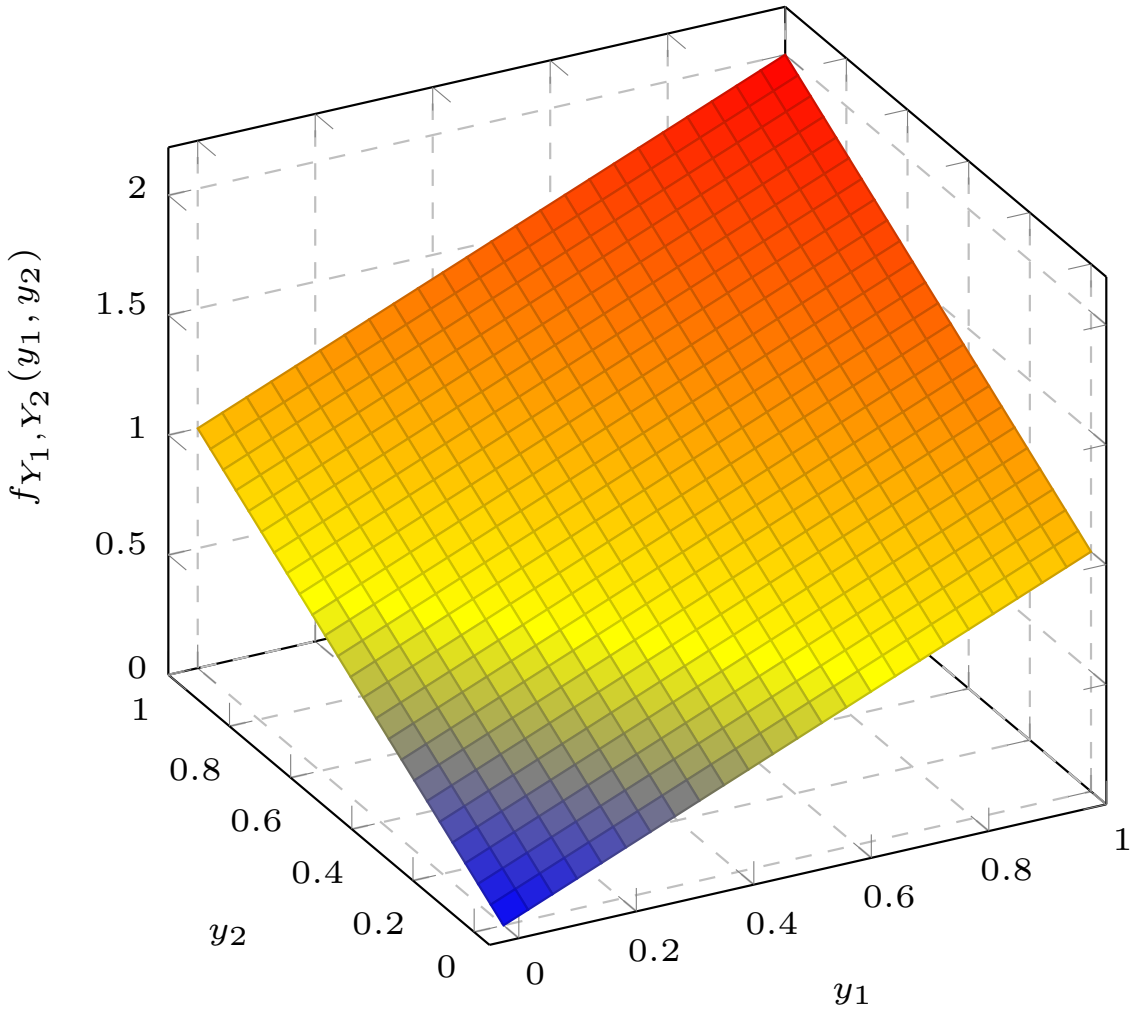
To compute $F_{Y_1,Y_2}(y_1, y_2)$ we always integrate the joint pdf

below and to the left

of (y_1, y_2).

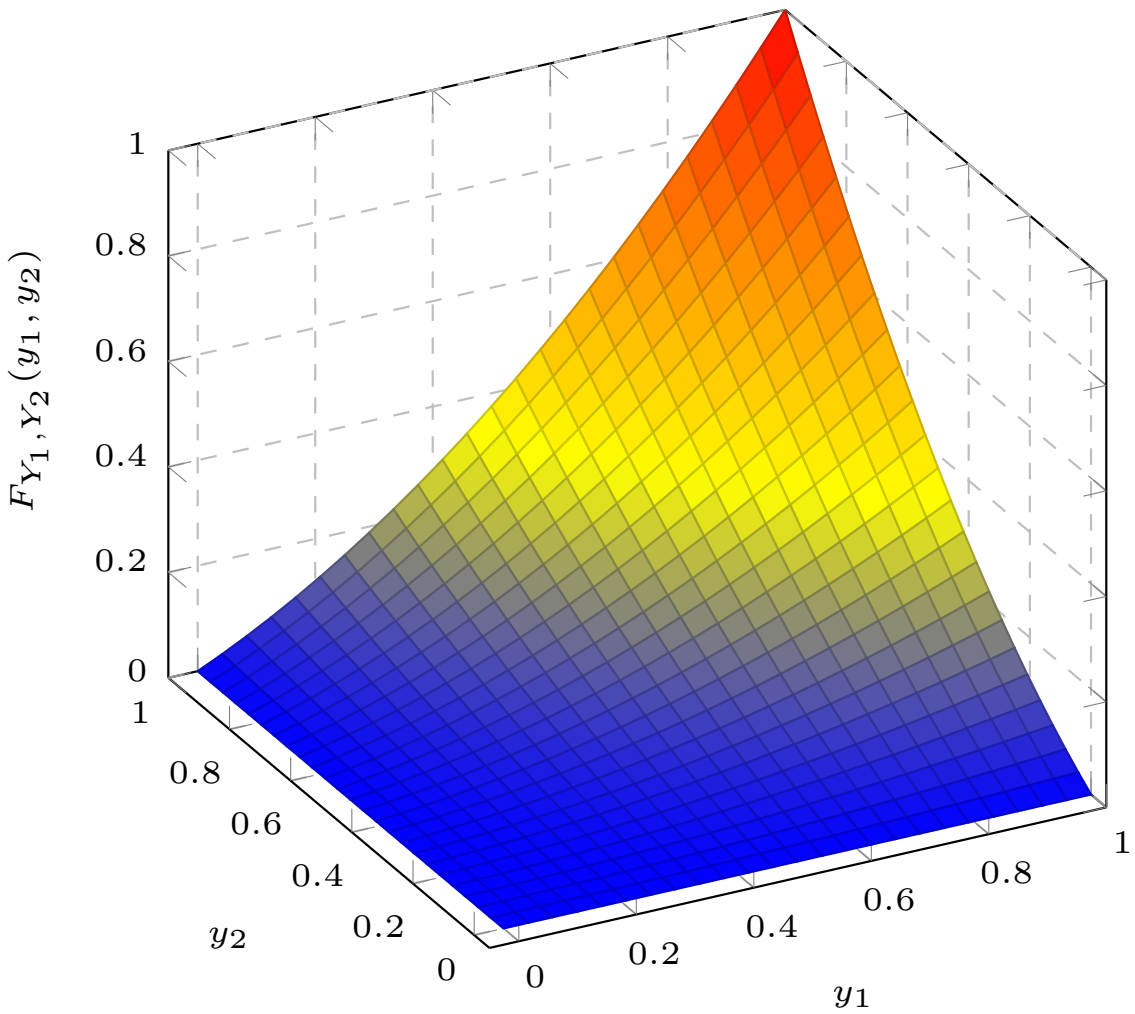
The joint density function (cont.)

$$f_{Y_1,Y_2}(y_1, y_2) = (y_1 + y_2)$$



The joint density function (cont.)

$$F_{Y_1,Y_2}(y_1, y_2) = (y_1^2 y_2 + y_1 y_2^2)/2$$



The joint density function (cont.)

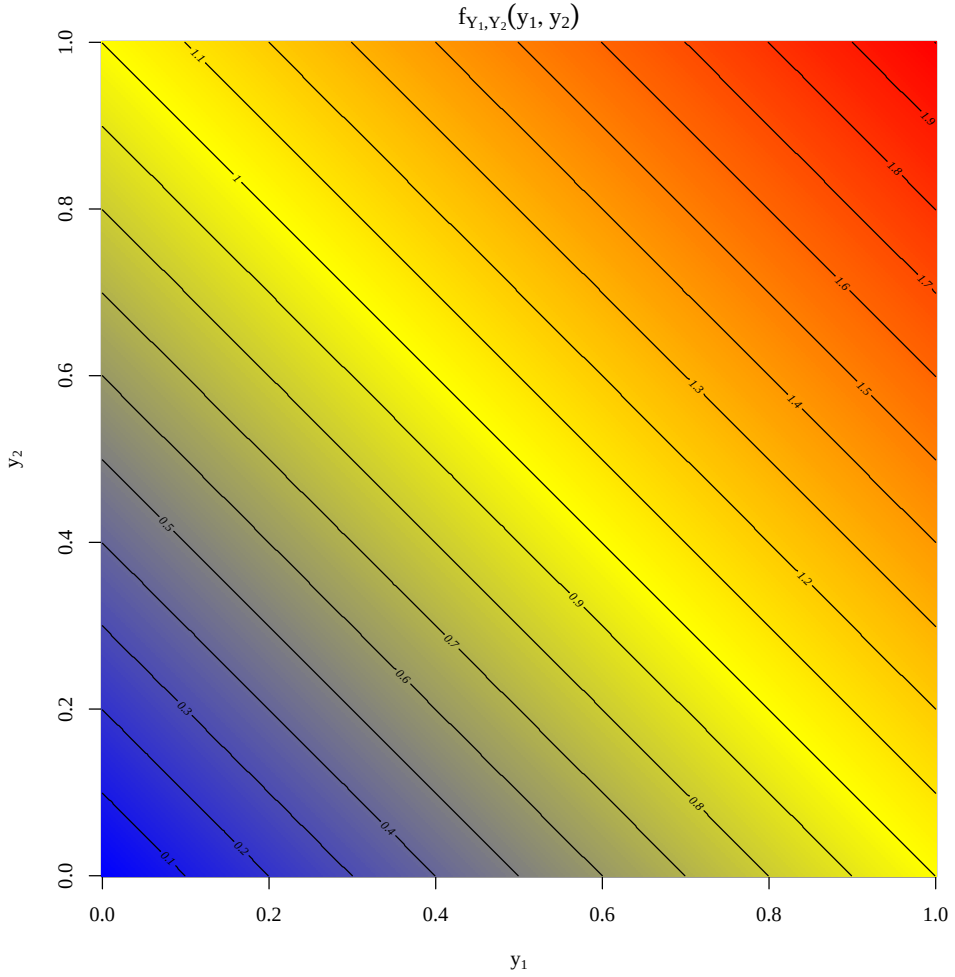


Figure: $f_{Y_1,Y_2}(y_1, y_2)$: image and contour plot

The joint density function (cont.)

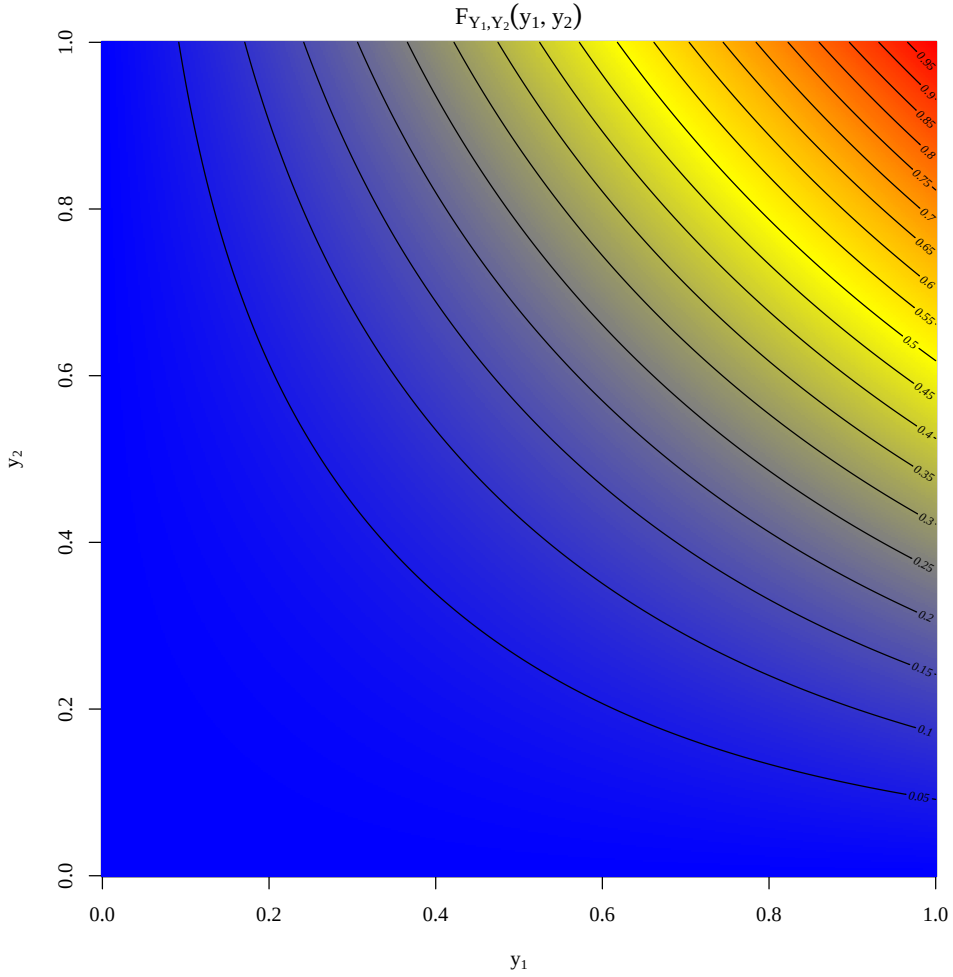


Figure: $F_{Y_1,Y_2}(y_1, y_2)$: image and contour plot

The joint density function (cont.)

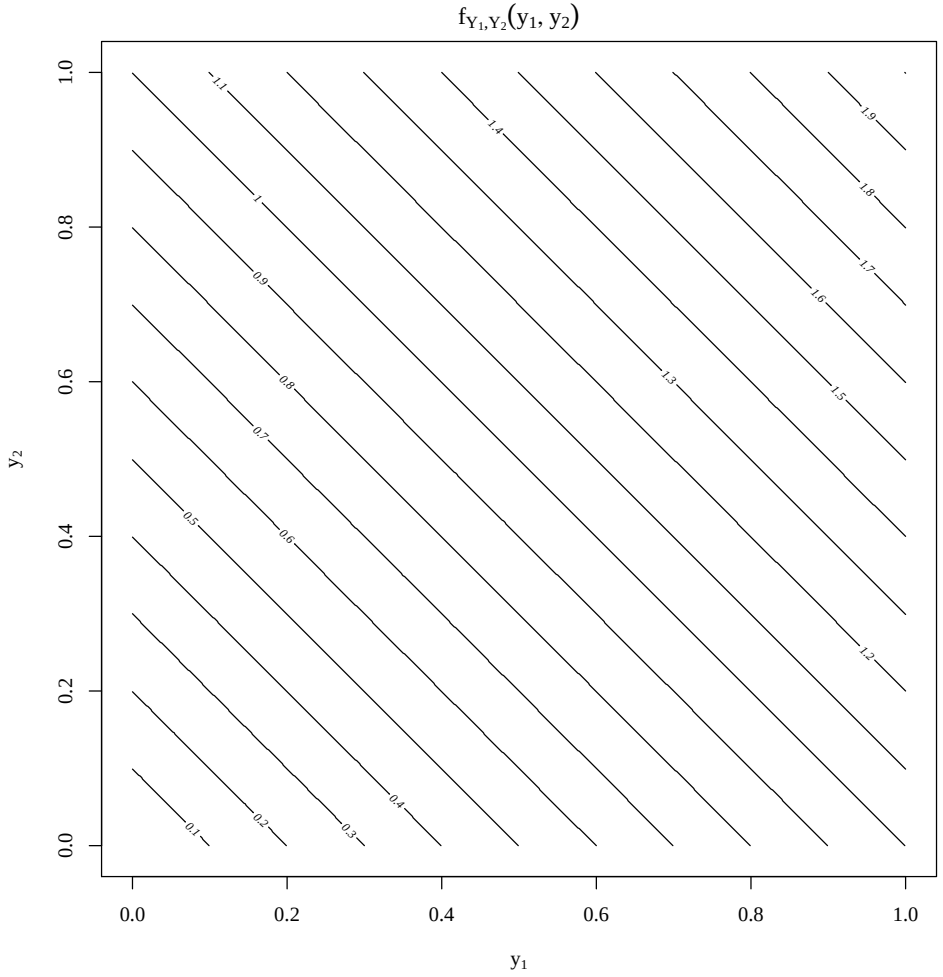


Figure: $f_{Y_1,Y_2}(y_1, y_2)$: contour plot

The joint density function (cont.)

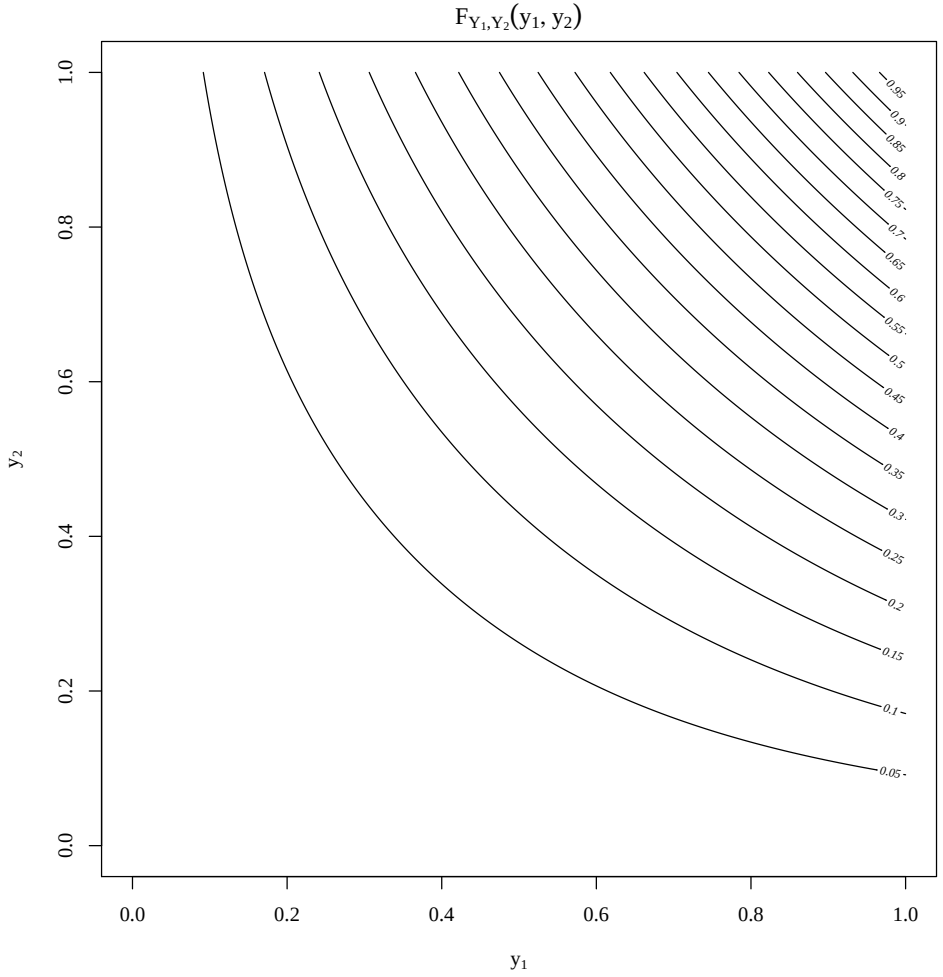


Figure: $F_{Y_1,Y_2}(y_1, y_2)$: contour plot

The joint density function (cont.)

We compute $f_{Y_1,Y_2}(y_1, y_2)$ from $F_{Y_1,Y_2}(y_1, y_2)$ using partial differentiation:

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{\partial^2}{\partial y_1 \partial y_2} \left\{ F_{Y_1,Y_2}(y_1, y_2) \right\}$$

- **Step 1:** differentiate $F_{Y_1,Y_2}(y_1, y_2)$ with respect to y_2 while holding y_1 constant;
- **Step 2:** take the result of **Step 1**, and differentiate it with respect to y_1.

The joint density function (cont.)

We can regard the calculation as

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{\partial}{\partial y_1}\left\{\frac{\partial F_{Y_1,Y_2}(y_1, y_2)}{\partial y_2}\right\}$$

The joint density function (cont.)

Example

For $0 \leq y_1, y_2 \leq 1$,

$$F_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2}\left(y_1^2 y_2 + y_1 y_2^2\right)$$

- **Step 1:**

$$\frac{\partial F_{Y_1,Y_2}(y_1, y_2)}{\partial y_2} = \frac{1}{2}\left(y_1^2 + 2y_1 y_2\right)$$

- **Step 2:**

$$\frac{\partial\left\{\frac{1}{2}\left(y_1^2 + 2y_1 y_2\right)\right\}}{\partial y_2} = \frac{1}{2}\left(2y_1 + 2y_2\right) = \left(y_1 + y_2\right)$$

The joint density function (cont.)

Note

In the partial differentiation, we could choose to differentiate with respect to y_1 in Step 1, and then differentiate the result with respect to y_2: we will get the same answer

$$\frac{\partial^2}{\partial y_1 \partial y_2}\left\{F_{Y_1,Y_2}(y_1,y_2)\right\} = \frac{\partial^2}{\partial y_2 \partial y_1}\left\{F_{Y_1,Y_2}(y_1,y_2)\right\}.$$

This is a general property of partial derivatives that holds provided the result of each side is continuous.

Constant pdfs

If the joint pdf is constant over a finite region $\mathcal{Y} \in \mathbb{R}^2$

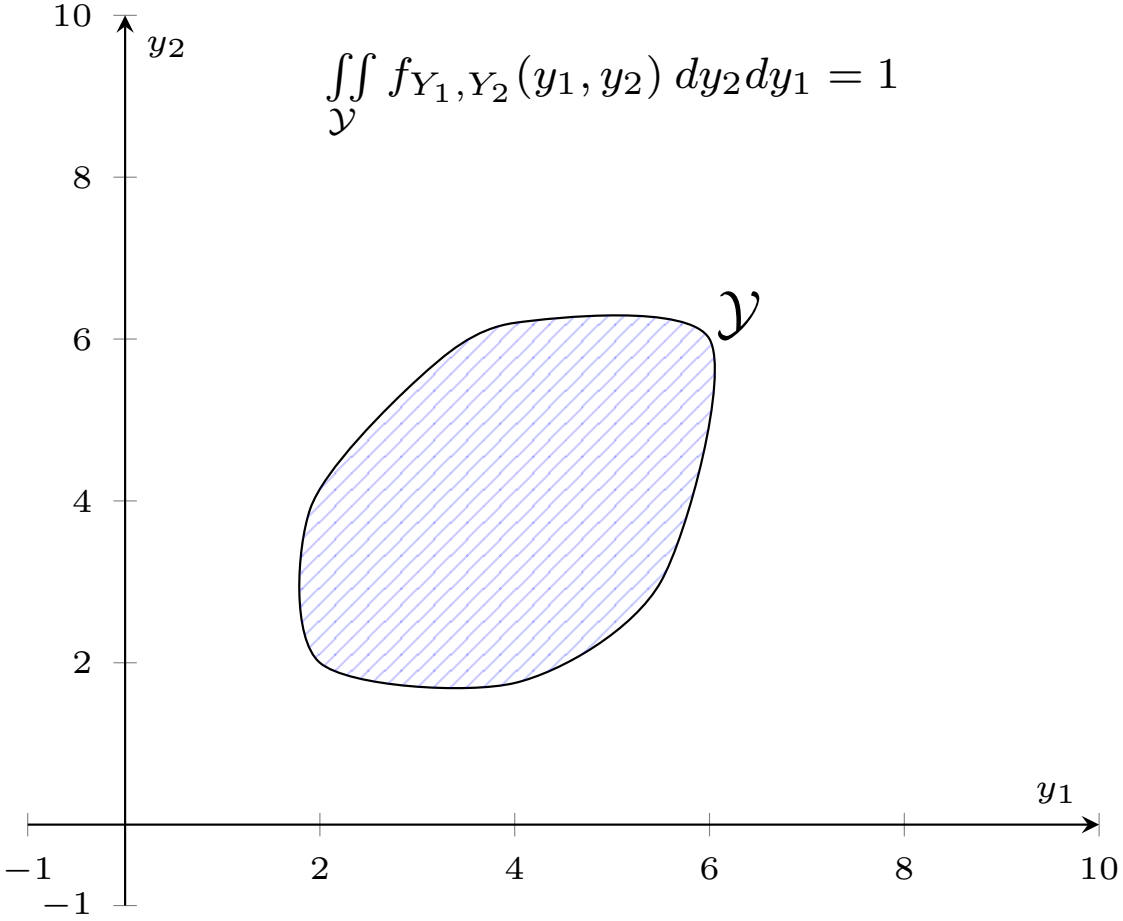
$$f_{Y_1,Y_2}(y_1, y_2) = c \qquad (y_1, y_2) \in \mathcal{Y}$$

with $f_{Y_1,Y_2}(y_1, y_2)$ zero otherwise, then to compute probabilities associated with this pdf, for example

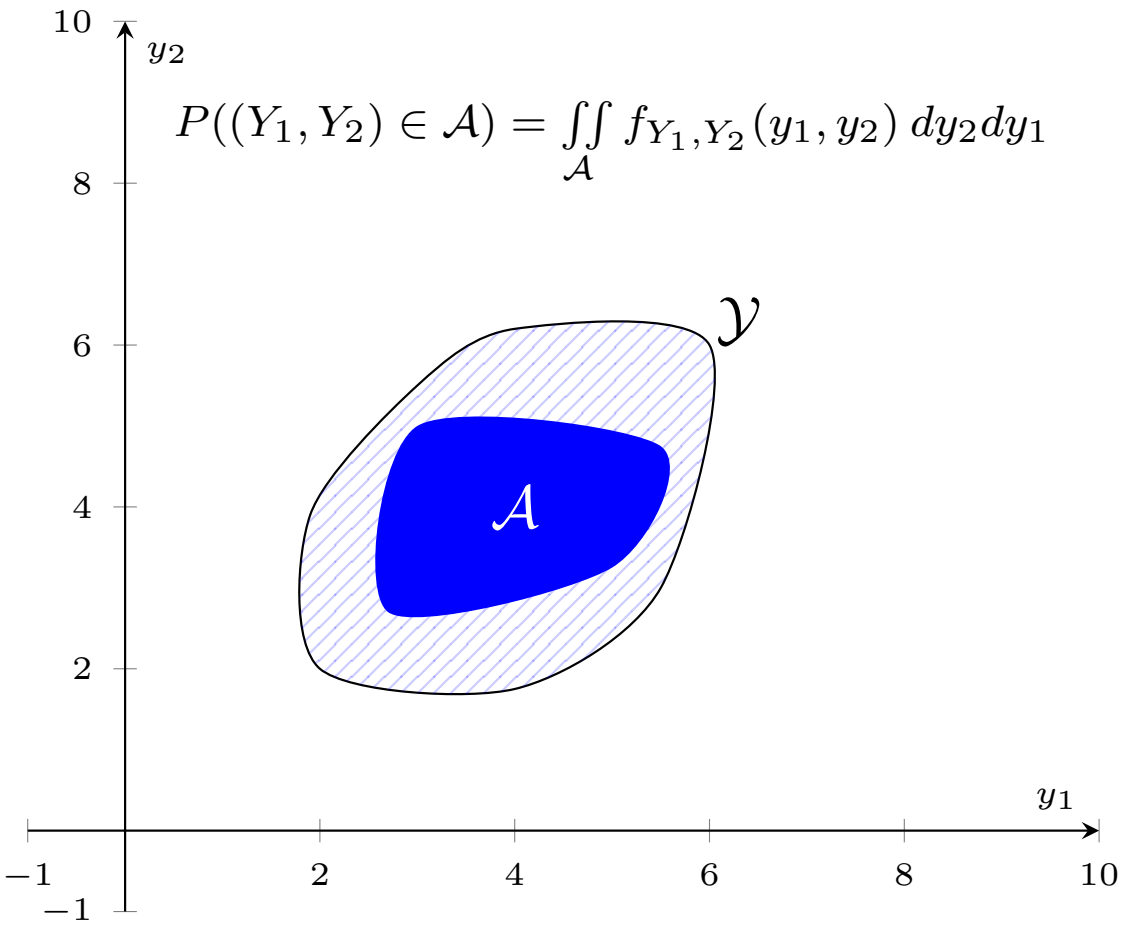
$$P((Y_1, Y_2) \in \mathcal{A}) = \iint_{\mathcal{A}} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 dy_1$$

we must compute the area of $\mathcal{A}$ and divide it by the area of $\mathcal{Y}$.

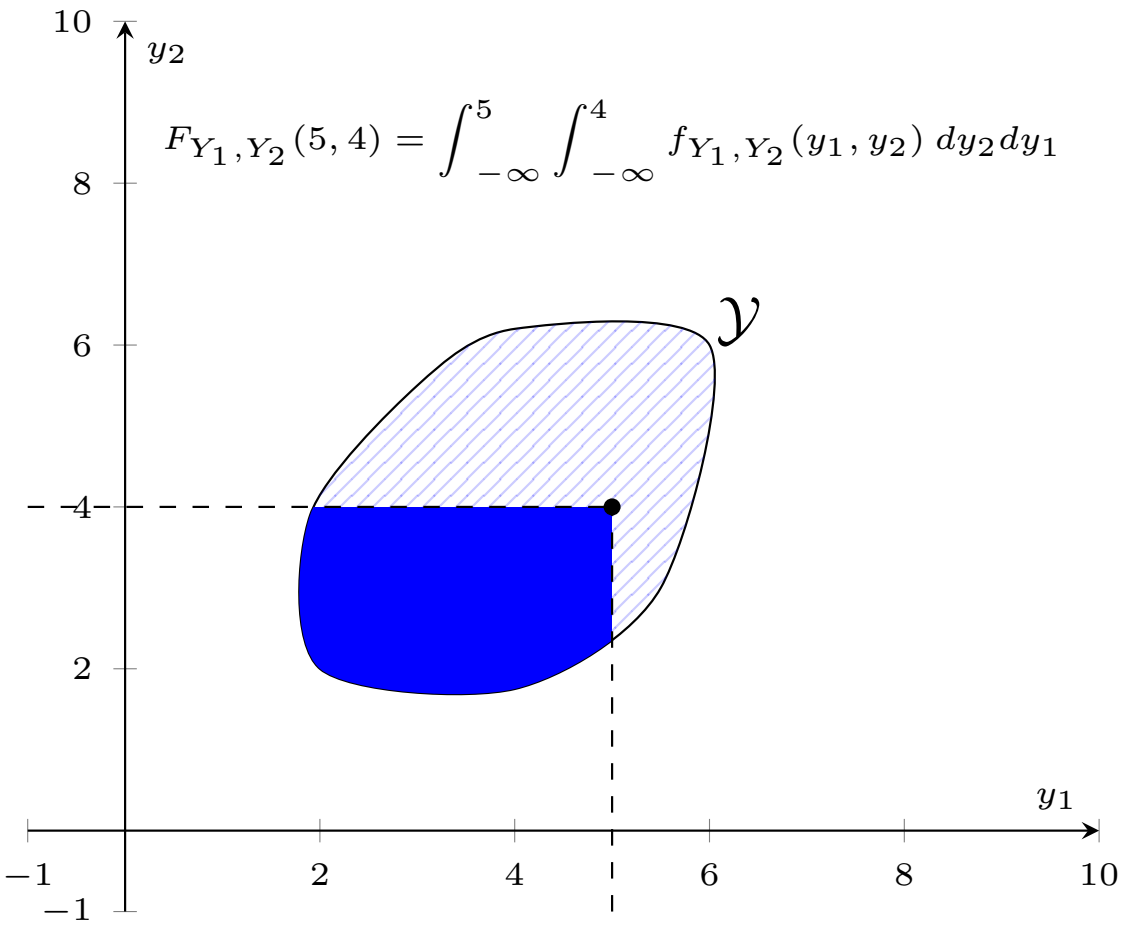
Constant pdfs (cont.)



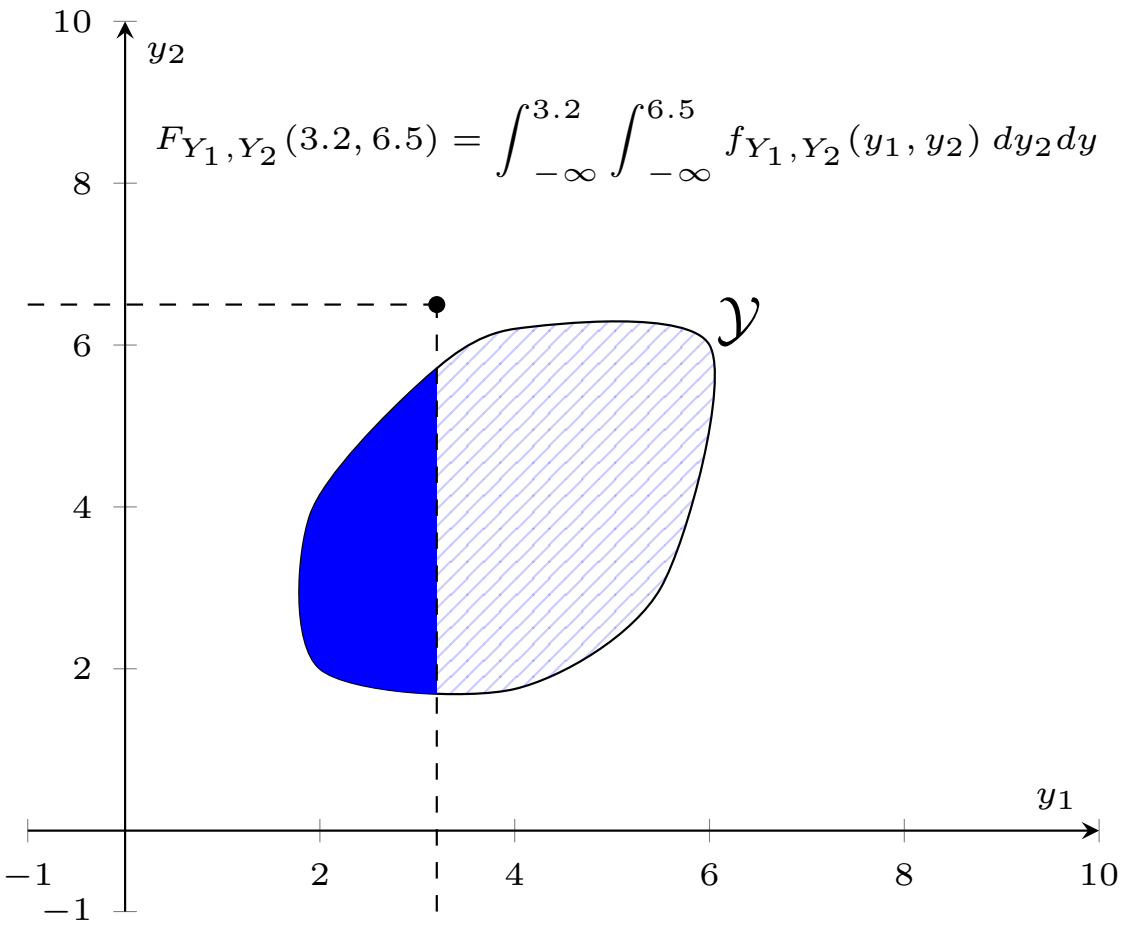
Constant pdfs (cont.)



Constant pdfs (cont.)



Constant pdfs (cont.)



The marginal density function

Once the joint pdf $f_{Y_1,Y_2}(y_1, y_2)$ is specified, we define the *marginal probability density functions* (marginal pdfs) analogously to the discrete case.

We have for Y_1 the marginal pdf $f_{Y_1}(y_1)$ for each fixed y_1 by

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2$$

and for Y_2 the marginal pdf $f_{Y_2}(y_2)$ for each fixed y_2 by

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_1$$

The marginal density function (cont.)

Example (Distribution on the unit square)

For

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} (y_1+y_2) & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density function (cont.)

Example (Distribution on the unit square)

We have for $0 \leq y_1 \leq 1$

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 \\ &= \int_0^1 (y_1 + y_2)\, dy_2 \\ &= \left[y_1 y_2 + \frac{y_2^2}{2} \right]_0^1 \\ &= y_1 + \frac{1}{2}. \end{aligned}$$

with $f_{Y_1}(y_1) = 0$ otherwise.

The marginal density function (cont.)

Example (Distribution on the unit square)

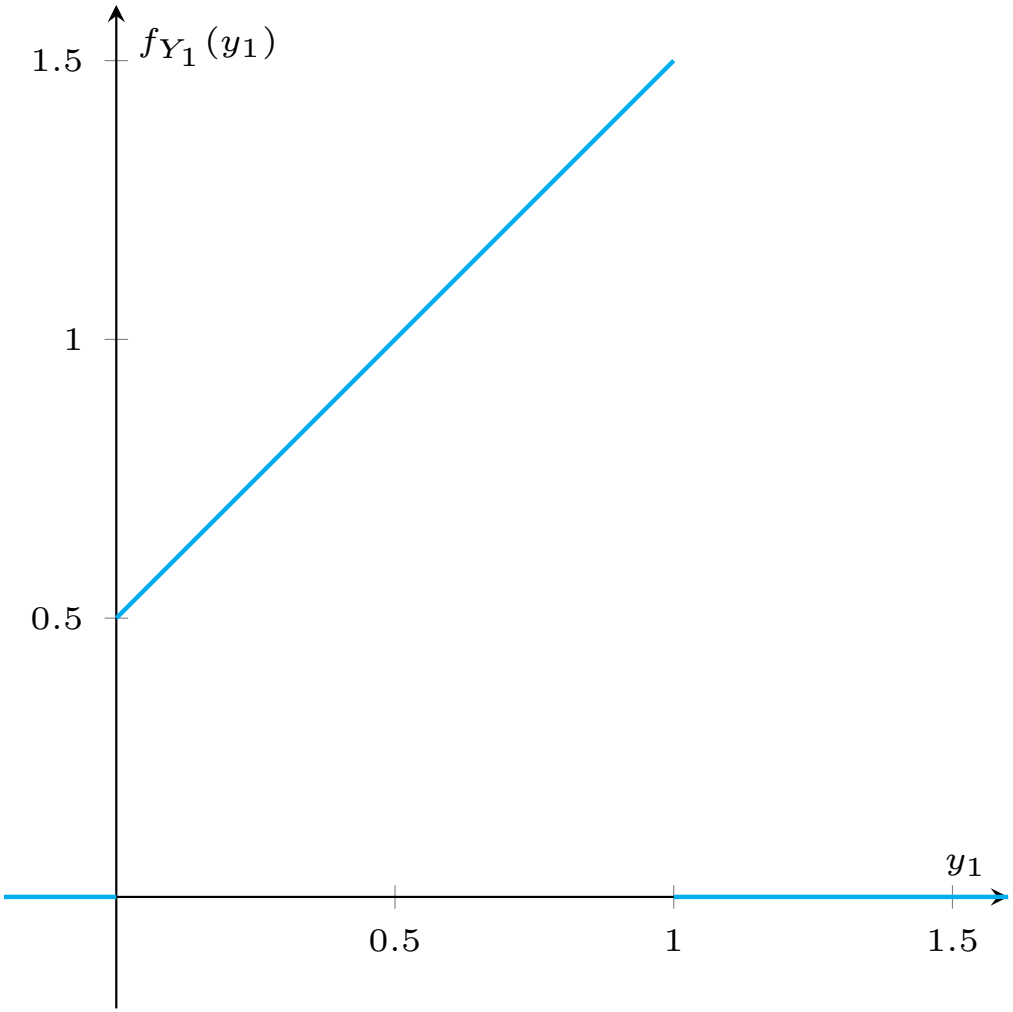
Also for $0 \leq y_1 \leq 1$

$$\begin{aligned} F_{Y_1}(y_1) &= \int_{-\infty}^{y_1} f_{Y_1}(t_1)\, dt_1 \\ &= \int_0^{y_1} \left(t_1 + \frac{1}{2}\right)\, dt_2 \\ &= \left[\frac{t_1^2}{2} + \frac{t_1}{2}\right]_0^{y_1} \\ &= \frac{y_1^2}{2} + \frac{y_1}{2}. \end{aligned}$$

with $F_{Y_1}(y_1) = 0$ for $y_1 < 0$ and $F_{Y_1}(y_1) = 1$ for $y_1 > 1$.

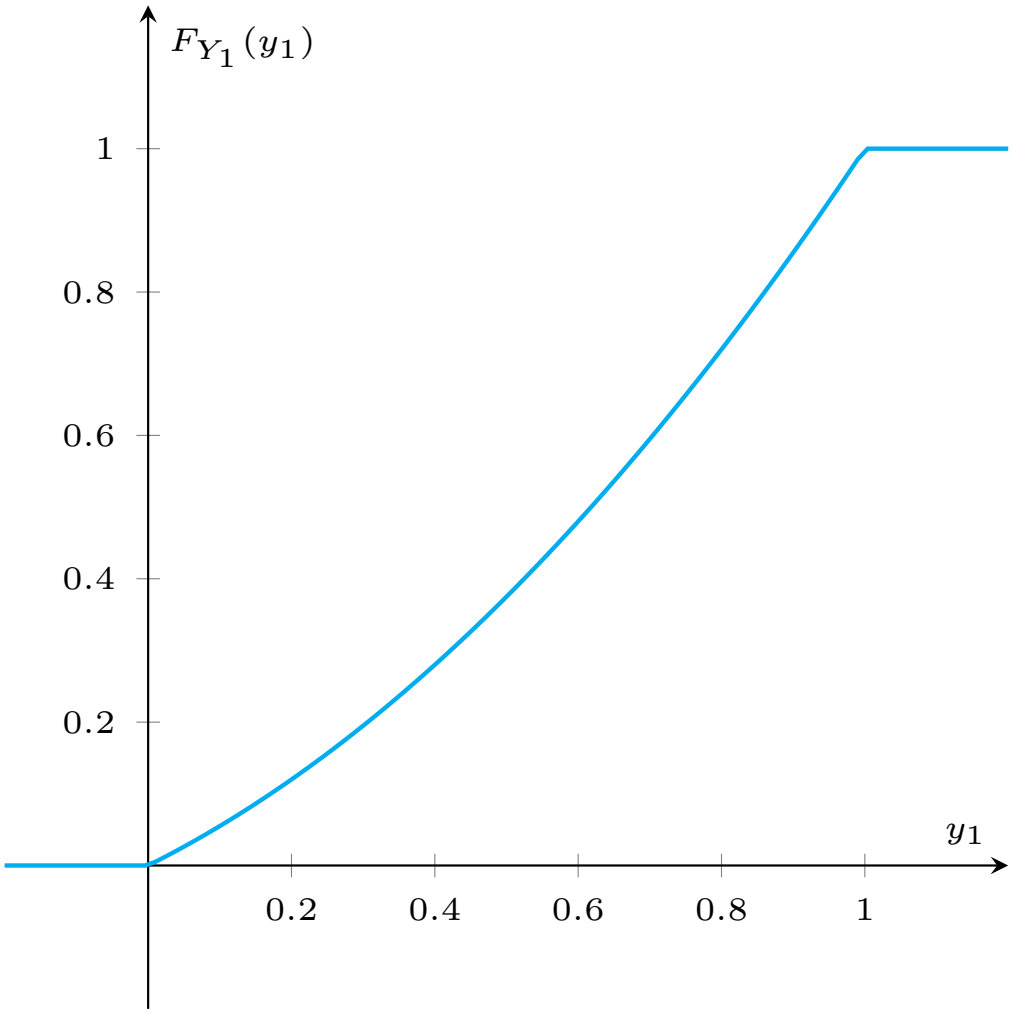
The marginal density function (cont.)

$f_{Y_1}(y_1) = y_1 + 1/2$ for $0 \leq y_1 \leq 1$.



The marginal density function (cont.)

$F_{Y_1}(y_1) = (y_1^2 + y_1)/2$ for $0 \leq y_1 \leq 1$.



The marginal density function (cont.)

Example (Distribution on the unit square)

We can perform a similar calculation and obtain

$$f_{Y_2}(y_2) = y_2 + \frac{1}{2} \qquad 0 \leq y_2 \leq 1$$

with $f_{Y_2}(y_2) = 0$ otherwise, and

$$F_{Y_2}(y_2) = \frac{(y_2^2 + y_2)}{2} \qquad 0 \leq y_2 \leq 1$$

with $F_{Y_2}(y_2) = 0$ for $y_2 < 0$ and $F_{Y_2}(y_2) = 1$ for $y_2 > 1$.

The conditional density function

Once we have the joint pdf

$$f_{Y_1,Y_2}(y_1, y_2)$$

and the marginal pdf

$$f_{Y_1}(y_1)$$

we can consider *conditional* pdfs.

The conditional density function (cont.)

The *conditional probability density function* for Y_2, given that $Y_1 = y_1$, is denoted

$$f_{Y_2|Y_1}(y_2|y_1)$$

and defined by

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_1}(y_1)}$$

whenever $f_{Y_1}(y_1) > 0$.

The conditional density function (cont.)

The function $f_{Y_2|Y_1}(y_2|y_1)$ is a pdf in y_2 for every fixed y_1. That is, for every fixed y_1 where $f_{Y_1}(y_1) > 0$,

- the conditional pdf is non-negative

$$f_{Y_2|Y_1}(y_2|y_1) \geq 0 \qquad -\infty < y_2 < \infty;$$

- the conditional pdf is integrates to 1 over y_2

$$\int_{-\infty}^{\infty} f_{Y_2|Y_1}(y_2|y_1)\, dy_2 = 1.$$

We can define $f_{Y_1|Y_2}(y_1|y_2)$ in an identical fashion.

The conditional density function (cont.)

Conditional pdfs are obtained by taking a 'slice' through the joint pdf at $Y_1 = y_1$, and then standardizing the slice so that the density integrates to one.

The conditional density function (cont.)

Example (Distribution on the unit square)

For

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} (y_1 + y_2) & 0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

we can consider conditioning on values of y_1 or y_2 in the interval $[0, 1]$.

The conditional density function (cont.)

Example (Distribution on the unit square)

As

$$f_{Y_1}(y_1) = y_1 + \frac{1}{2} \qquad 0 \leq y_1 \leq 1.$$

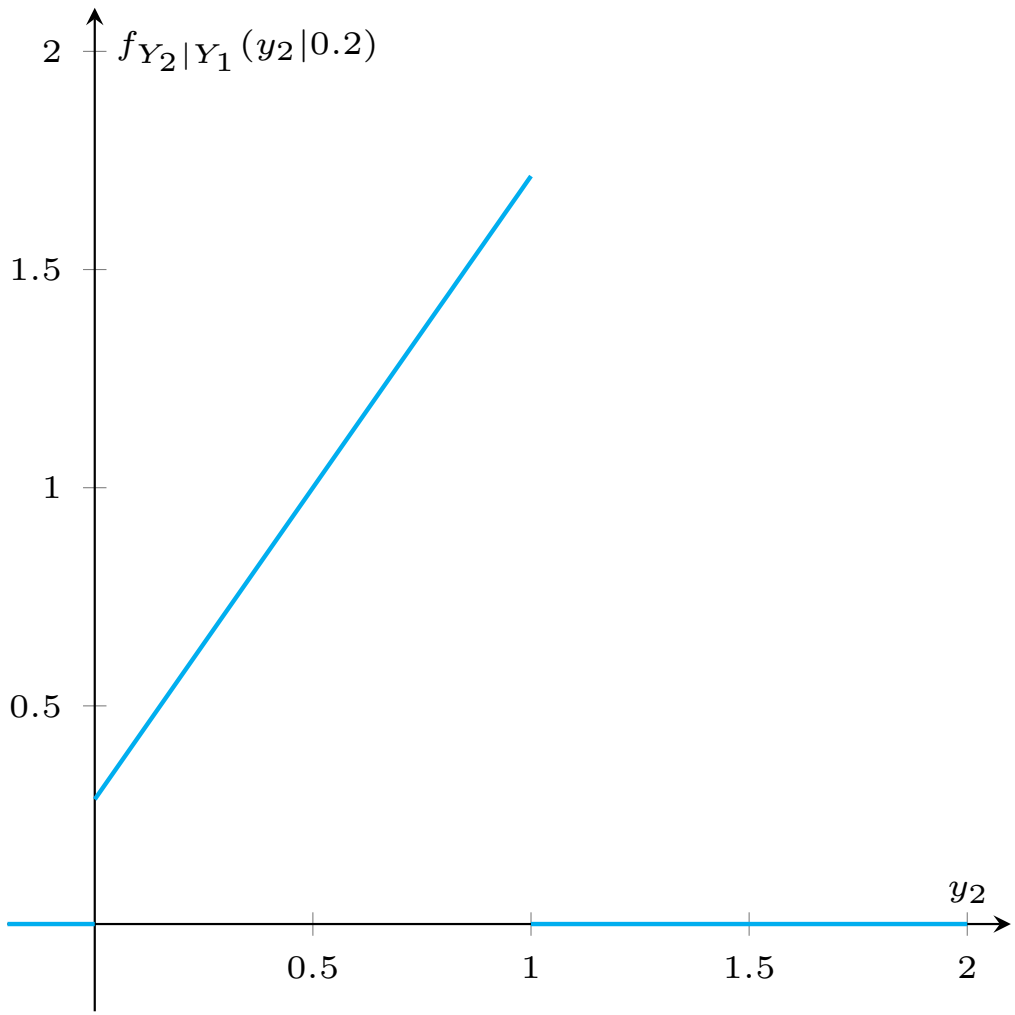
we have for each fixed y_1 in this range

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{f_{Y_1,Y_2}(y_1, y_2)}{f_{Y_1}(y_1)} = \frac{(y_1 + y_2)}{y_1 + 1/2}$$

for $0 \leq y_2 \leq 1$, with the function zero for other values of y_2.

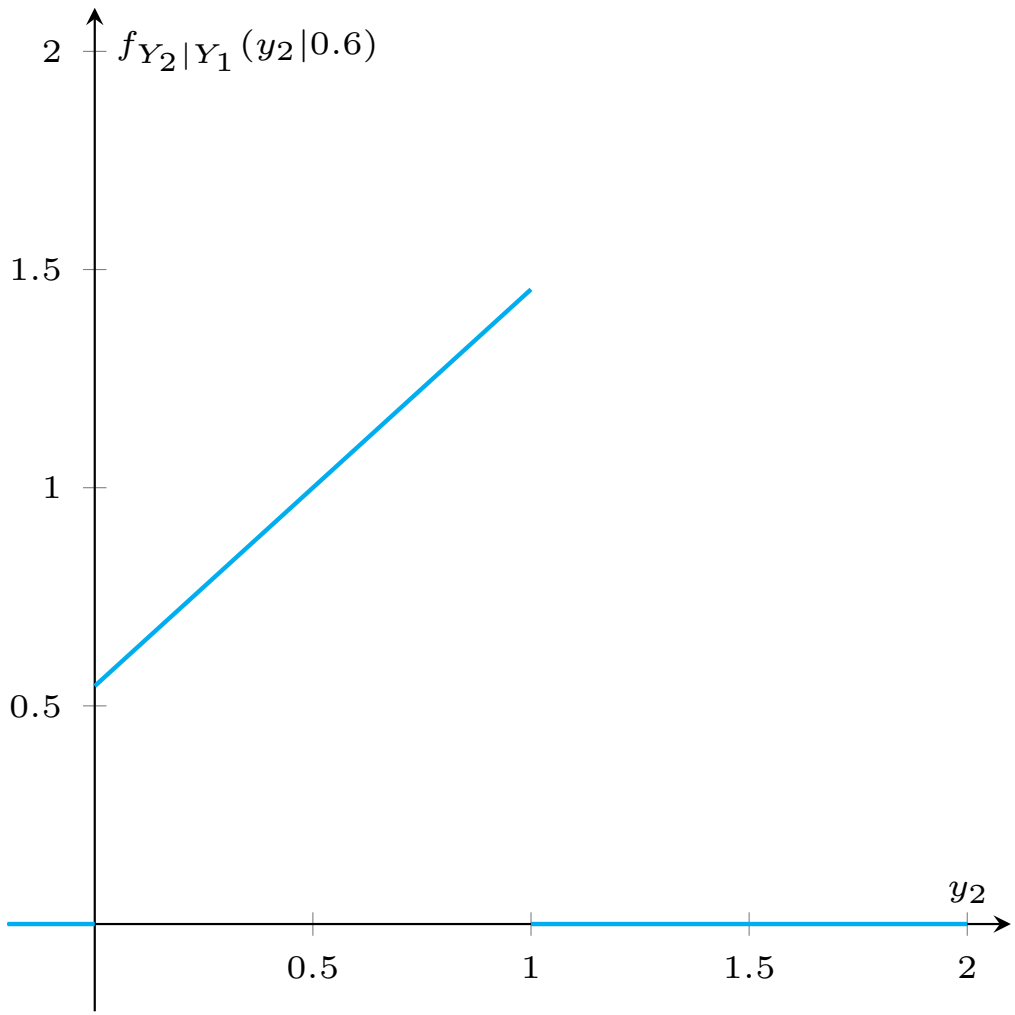
The conditional density function (cont.)

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{(y_1 + y_2)}{y_1 + 1/2} \text{ for } 0 \leq y_2 \leq 1, \text{ with } y_1 = 0.2$$



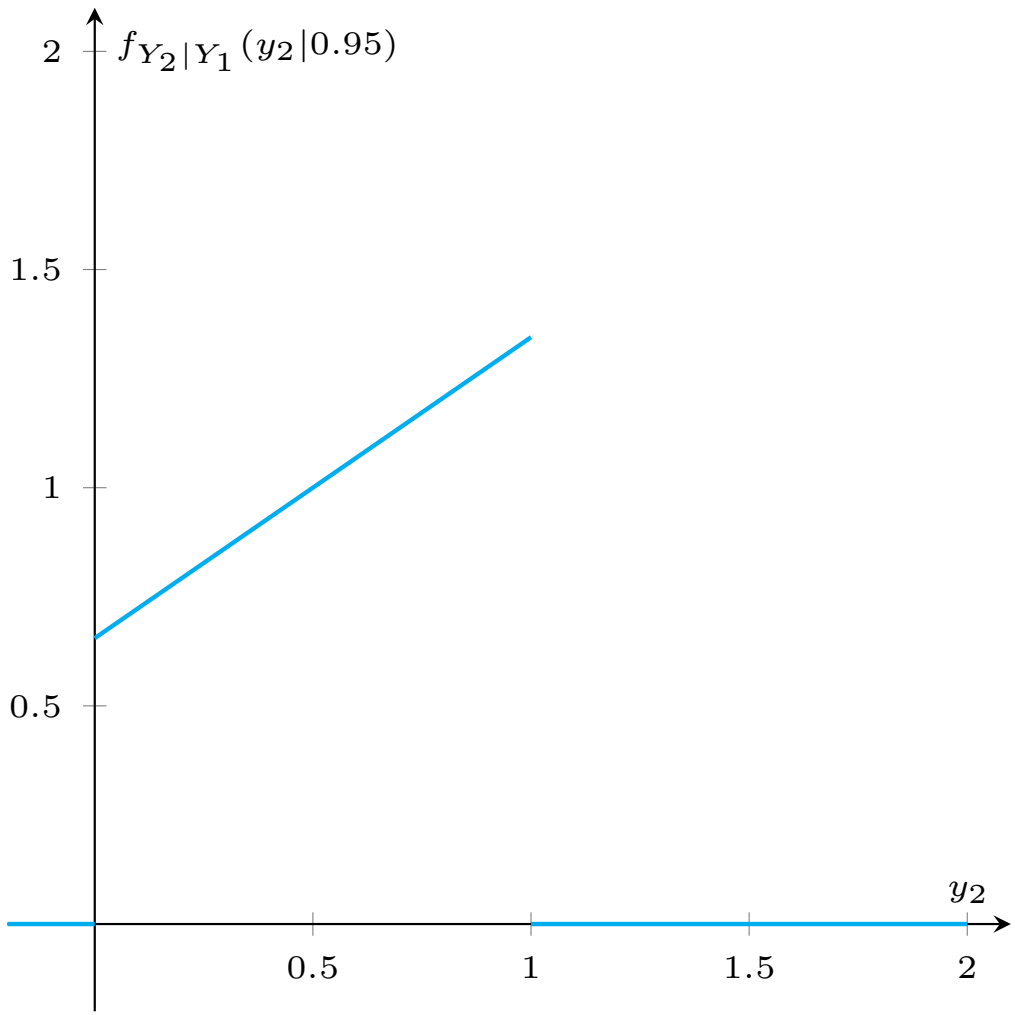
The conditional density function (cont.)

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{(y_1 + y_2)}{y_1 + 1/2} \text{ for } 0 \leq y_2 \leq 1, \text{ with } y_1 = 0.6$$



The conditional density function (cont.)

$$f_{Y_2|Y_1}(y_2|y_1) = \frac{(y_1 + y_2)}{y_1 + 1/2} \text{ for } 0 \leq y_2 \leq 1, \text{ with } y_1 = 0.95$$



The conditional density function (cont.)

As in the discrete case, we have the *chain rule factorization*

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2|Y_1}(y_2|y_1)$$

whenever $f_{Y_1}(y_1) > 0$.

As ever, we can exchange the roles of Y_1 and Y_2 to obtain that

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_2}(y_2) f_{Y_1|Y_2}(y_1|y_2).$$

whenever $f_{Y_1}(y_1) > 0$.

Multivariate distributions

All of the above ideas extend naturally to more than two variables. If $(Y_1, \ldots, Y_n)$ form an n-dimensional random vector, we can consider the *joint pdf*

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$$

and *joint cdf*

$$F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n).$$

Multivariate distributions (cont.)

We can consider *marginalization* by integrating out $n - d < n$ of the dimensions to leave a d-dimensional probability distribution: for example if $n = 4$ and $d = 2$, we have

$$f_{Y_1,Y_2}(y_1, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Y_1,Y_2,Y_3,Y_4}(y_1, y_2, y_3, y_4)\, dy_3 dy_4$$

or

$$f_{Y_2,Y_4}(y_2, y_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Y_1,Y_2,Y_3,Y_4}(y_1, y_2, y_3, y_4)\, dy_1 dy_3$$

and so on. This construction works with any n and any d.

Independence

Continuous random variables $(Y_1, \ldots, Y_n)$ are *independent* if

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$

for **all** vectors $(y_1, \ldots, y_n) \in \mathbb{R}^n$. Equivalently they are independent if

$$F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \prod_{i=1}^{n} F_{Y_i}(y_i)$$

for **all** vectors $(y_1, \ldots, y_n) \in \mathbb{R}^n$.

Random variables that are not independent are termed *dependent*.

Independence (cont.)

In the bivariate case, Y_1 and Y_2 are independent if

$$f_{Y_2|Y_1}(y_2|y_1) = f_{Y_2}(y_2)$$

for all (y_1, y_2) where the conditional pdf is defined, or equivalently if

$$f_{Y_1|Y_2}(y_1|y_2) = f_{Y_1}(y_1)$$

for all (y_1, y_2).

Independence (cont.)

Note

A straightforward way to exclude the possibility of two variables being independent is to assess whether the set of values, $\mathcal{Y}_{12}$ (called the *support* of the joint pdf), for which

$$f_{Y_1,Y_2}(y_1, y_2) > 0$$

is identical to the Cartesian product of the two sets, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ for which the marginal densities satisfy $f_{Y_1}(y_1) > 0$ and $f_{Y_2}(y_2) > 0$.

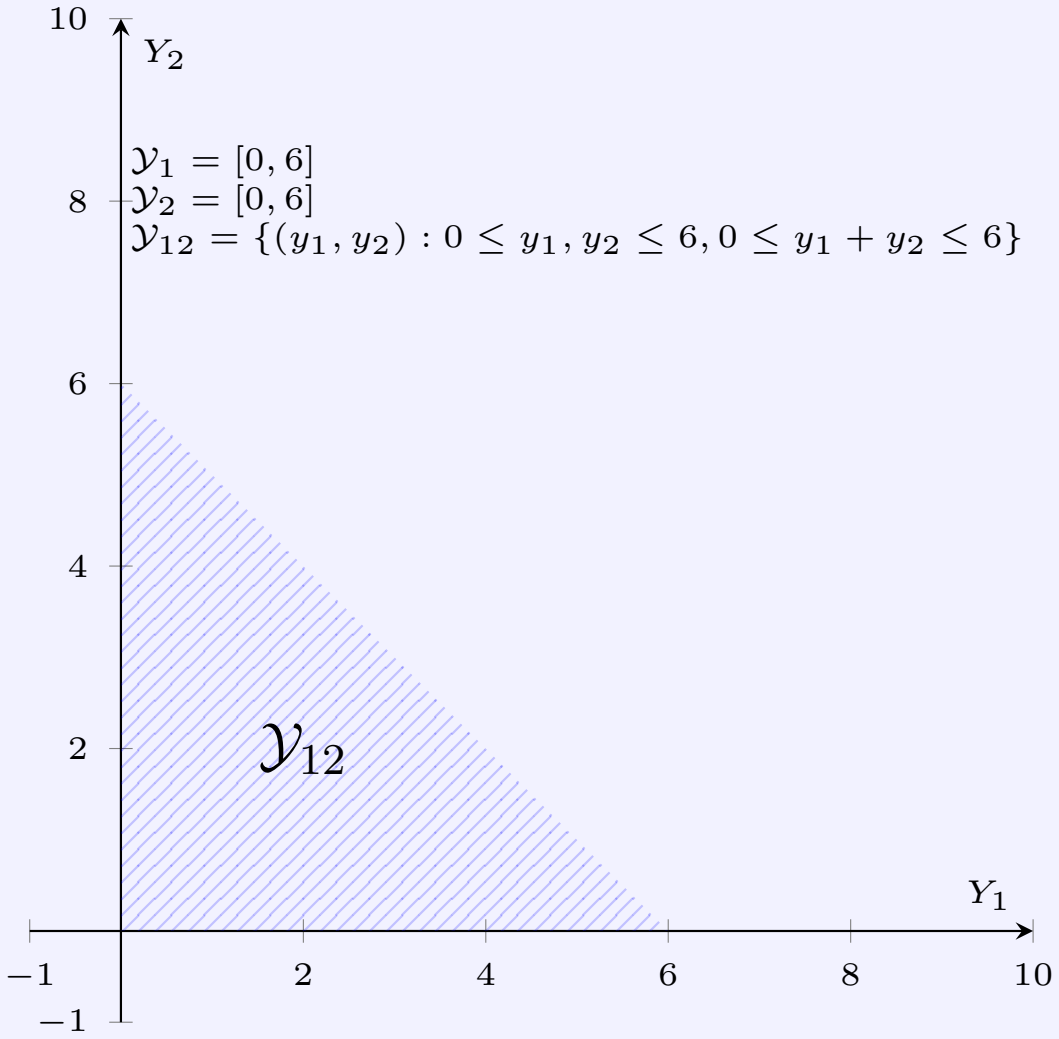
That is, if

$$\mathcal{Y}_{12} \neq \mathcal{Y}_1 \times \mathcal{Y}_2$$

then Y_1 and Y_2 are **not** independent.

Independence (cont.)

Note



Independence (cont.)

Note

We can find points where

- $f_{Y_1}(y_1) > 0$ and $f_{Y_2}(y_2) > 0$, but
- $f_{Y_1,Y_2}(y_1, y_2) = 0$

(for example, the point $(5, 5)$), so we do not meet the requirement for independence that

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$$

for **all** $(y_1, y_2) \in \mathbb{R}^2$.

Multivariate expectations

We can extend the idea of expectation to two or more dimensions.

We consider a function $g(y_1, \ldots, y_n)$ of n random variables. The expectation of this function is defined in the discrete case by

$$\mathbb{E}[g(Y_1, \ldots, Y_n)] =$$

$$\sum_{y_1=-\infty}^{\infty} \cdots \sum_{y_n=-\infty}^{\infty} g(y_1, \ldots, y_n) f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n).$$

Multivariate expectations (cont.)

In the continuous case we define

$$\mathbb{E}[g(Y_1, \ldots, Y_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, \ldots, y_n) f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)\, dy_1 \ldots dy_n.$$

These definitions hold when the sum/integral is finite.

Multivariate expectations (cont.)

For two variables, this becomes

$$\mathbb{E}[g(Y_1, Y_2)] = \sum_{y_1=-\infty}^{\infty} \sum_{y_2=-\infty}^{\infty} g(y_1, y_2) f_{Y_1,Y_2}(y_1, y_2).$$

or

$$\mathbb{E}[g(Y_1, Y_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2) f_{Y_1,Y_2}(y_1, y_2)\, dy_1 dy_2.$$

Multivariate expectations (cont.)

The properties of expectations for more than one variable match those in the single variable case.

In particular, if a_1, a_2 are constants, and $g_1(.,.)$ and $g_2(.,.)$ are two functions, then

$$\mathbb{E}[a_1 g_1(Y_1, Y_2) + a_2 g_2(Y_1, Y_2)] = a_1 \mathbb{E}[g_1(Y_1, Y_2)] + a_2 \mathbb{E}[g_2(Y_1, Y_2)]$$

provided all the expectations are finite.

Covariance and correlation

For two random variables Y_1 and Y_2, with expectations μ_1 and μ_2 respectively, the *covariance* between Y_1 and Y_2, $\mathrm{Cov}[Y_1, Y_2]$, is defined by

$$\mathrm{Cov}[Y_1, Y_2] = \mathbb{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

Covariance and correlation (cont.)

Note that

$$\begin{aligned}
\mathrm{Cov}[Y_1, Y_2] &= \mathbb{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)] \\
&= \mathbb{E}[Y_1 Y_2 - \mu_1 Y_2 - Y_1 \mu_2 + \mu_1 \mu_2] \\
&= \mathbb{E}[Y_1 Y_2] - \mu_1 \mathbb{E}[Y_2] - \mu_2 \mathbb{E}[Y_1] + \mu_1 \mu_2 \\
&= \mathbb{E}[Y_1 Y_2] - \mu_1 \mu_2 - \mu_2 \mu_1 + \mu_1 \mu_2 \\
&= \mathbb{E}[Y_1 Y_2] - \mu_1 \mu_2.
\end{aligned}$$

Covariance and correlation (cont.)

The *correlation* between Y_1 and Y_2, $\mathrm{Corr}(Y_1, Y_2)$, is defined by

$$\mathrm{Corr}[Y_1, Y_2] = \frac{\mathrm{Cov}[Y_1, Y_2]}{\sqrt{\mathbb{V}[Y_1]\mathbb{V}[Y_2]}}$$

Covariance and correlation (cont.)

Both covariance and correlation are *measures of dependence.* We have that for any pair of random variables,

$$-1 \leq \mathrm{Corr}[Y_1, Y_2] \leq 1$$

Covariance and correlation (cont.)

If

$$\mathrm{Cov}[Y_1, Y_2] = \mathrm{Corr}[Y_1, Y_2] = 0$$

we say that Y_1 and Y_2 are *uncorrelated*.

Independent random variables

If Y_1 and Y_2 are independent then (in the continuous case – discrete case works in the same way), we have that

$$\begin{aligned}
\mathbb{E}[Y_1 Y_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1,Y_2}(y_1, y_2)\, dy_1 dy_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f_{Y_1}(y_1) f_{Y_2}(y_2)\, dy_1 dy_2 \\
&= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) \left\{ \int_{-\infty}^{\infty} y_1 f_{Y_1}(y_1)\, dy_1 \right\} dy_2 \\
&= \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) \mathbb{E}[Y_1] dy_2 \\
&= \mathbb{E}[Y_1] \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) dy_2 = \mathbb{E}[Y_1]\mathbb{E}[Y_2]
\end{aligned}$$

Independent random variables (cont.)

More generally, if $g_1(Y_1)$ and $g_2(Y_2)$ are two functions of the independent random variables, then

$$\mathbb{E}[g_1(Y_1)g_2(Y_2)] = \mathbb{E}[g_1(Y_1)]\mathbb{E}[g_2(Y_2)]$$

Independent random variables (cont.)

Therefore if two random variables Y_1 and Y_2 are independent, then

$$\begin{aligned}\mathrm{Cov}[Y_1, Y_2] &= \mathbb{E}[Y_1 Y_2] - \mu_1\mu_2 \\ &= \mathbb{E}[Y_1]\mathbb{E}[Y_2] - \mu_1\mu_2 \\ &= \mu_1\mu_2 - \mu_1\mu_2 = 0\end{aligned}$$

and Y_1 and Y_2 are uncorrelated.

Linear combinations of random variables

Suppose that $Y_1, \ldots, Y_n$ and $X_1, \ldots, X_m$ be random variables with

$$\mathbb{E}[Y_i] = \mu_i \qquad \mathbb{E}[X_j] = \xi_j.$$

Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \qquad U_2 = \sum_{j=1}^{m} b_j X_j$$

for real constants $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$.

Linear combinations of random variables (cont.)

(a) Expectations:

$$\mathbb{E}[U_1] = \sum_{i=1}^{n} a_i \mathbb{E}[Y_i] = \sum_{i=1}^{n} a_i \mu_i$$

(b) Variances:

$$\mathbb{V}[U_1] = \sum_{i=1}^{n} a_i^2 \mathbb{V}[Y_i] + 2 \sum_{j=2}^{n} \sum_{i=1}^{j-1} a_i a_j \mathrm{Cov}[Y_i, Y_j]$$

(c) Covariances:

$$\mathrm{Cov}[U_1, U_2] = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathrm{Cov}[Y_i, X_j]$$

Linear combinations of random variables (cont.)

For (b):

$$\mathbb{V}[U_1] = \mathbb{E}[U_1^2] - \{\mathbb{E}[U_1]\}^2.$$

But

$$U_1^2 = \left(\sum_{i=1}^{n} a_i Y_i\right)^2 = \sum_{i=1}^{n} a_i^2 Y_i^2 + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j Y_i Y_j$$

so therefore

$$\mathbb{E}[U_1^2] = \sum_{i=1}^{n} a_i^2 \mathbb{E}[Y_i^2] + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j \mathbb{E}[Y_i Y_j]$$

Linear combinations of random variables (cont.)

Similarly

$$\{\mathbb{E}[U_1]\}^2 = \left\{\sum_{i=1}^{n} a_i \mu_i\right\}^2 = \sum_{i=1}^{n} a_i^2 \mu_i^2 + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j \mu_i \mu_j$$

Therefore

$$\begin{aligned}\mathbb{E}[U_1^2] - \{\mathbb{E}[U_1]\}^2 = &\sum_{i=1}^{n} a_i^2 (\mathbb{E}[Y_i^2] - \mu_i^2) \\ &+ 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} a_i a_j (\mathbb{E}[Y_i Y_j] - \mu_i \mu_j)\end{aligned}$$

and the result follows.

Linear combinations of random variables (cont.)

For (c): consider the simple case

$$U_1 = a_1 Y_1 + a_2 Y_2 \qquad U_2 = b_1 X_1 + b_2 X_2$$

Then

$$\begin{aligned}
\mathbb{E}[U_1 U_2] &= \mathbb{E}[(a_1 Y_1 + a_2 Y_2)(b_1 X_1 + b_2 X_2)] \\
&= \mathbb{E}[a_1 b_1 Y_1 X_1 + a_1 b_2 Y_1 X_2 + a_1 b_2 Y_1 X_2 + a_2 b_2 Y_2 X_2] \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} a_i b_j \mathbb{E}[Y_i X_j]
\end{aligned}$$

Linear combinations of random variables (cont.)

Using the same logic

$$\begin{aligned}\mathbb{E}[U_1]\mathbb{E}[U_2] &= \left(\sum_{i=1}^{2} a_i \mathbb{E}[Y_i]\right)\left(\sum_{j=1}^{2} b_j \mathbb{E}[X_j]\right)\\ &= \left(\sum_{i=1}^{2} a_i \mu_i\right)\left(\sum_{j=1}^{2} b_j \xi_j\right)\\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} a_i b_j \mu_i \xi_j\end{aligned}$$

Linear combinations of random variables (cont.)

and therefore

$$\mathbb{E}[U_1 U_2] - \mathbb{E}[U_1]\mathbb{E}[U_2] = \sum_{i=1}^{2}\sum_{j=1}^{2} a_i b_j (\mathbb{E}[Y_i X_j] - \mu_i \xi_j)$$

and the result follows.

Linear combinations of random variables (cont.)

Example (Sums of independent random variables)

If $Y_1, \ldots, Y_n$ are independent and identically distributed with expectation μ and variance σ^2, then

$$U = \sum_{i=1}^{n} Y_i$$

then

$$\mathbb{E}[U] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} \mu = n\mu$$

and

$$\mathbb{V}[U] = \sum_{i=1}^{n} \mathbb{V}[Y_i] = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$$

as the covariance terms are all zero.

Sums of random variables (revisited)

For two random variables Y_1 and Y_2 consider the distribution of

$$Y = Y_1 + Y_2$$

which we studied previously in the context using the convolution calculation in Chapter 3.

Sums of random variables (revisited) (cont.)

We now see that the convolution formula can be viewed in the context of a joint distribution calculation:

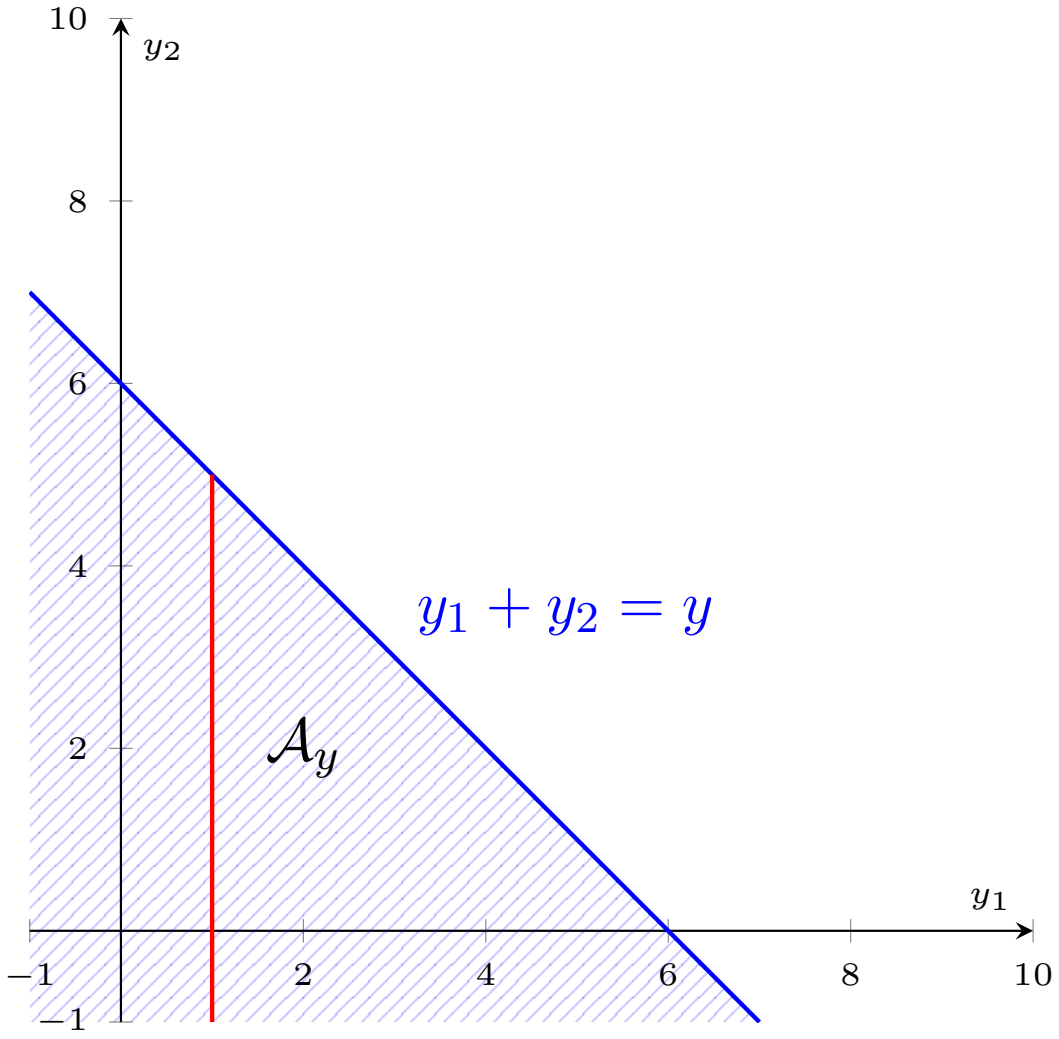
$$F_Y(y) = P(Y \leq y) = P(Y_1 + Y_2 \leq y) = \iint_{\mathcal{A}_y} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 dy_1$$

where

$$\mathcal{A}_y = \left\{ (y_1, y_2) \in \mathbb{R}^2 : y_1 + y_2 \leq y \right\}.$$

Sums of random variables (revisited) (cont.)

Need to integrate over the region below the line $y_1 + y_2 = y$.



Sums of random variables (revisited) (cont.)

Therefore we can write

$$F_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y-y_1} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 dy_1.$$

By differentiating with respect to y under the integral sign, using the Fundamental Theorem of Calculus, we see that

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y - y_1)\, dy_1.$$

Sums of random variables (revisited) (cont.)

We now consider an alternate method of calculation: recall the definition of the moment generating function:

$$m_Y(t) = \mathbb{E}\left[e^{tY}\right]$$

for $-b \leq t \leq b$, some $b > 0$.

We can define a similar quantity for the random variable U which is a transformation of Y, $U = g(Y)$ say, as

$$m_U(t) = \mathbb{E}\left[e^{tU}\right] \equiv \mathbb{E}\left[e^{tg(Y)}\right].$$

Sums of random variables (revisited) (cont.)

For two random variables Y_1 and Y_2, with

$$Y = a_1 g_1(Y_1) + a_2 g_2(Y_2),$$

we have

$$m_Y(t) = \mathbb{E}\left[e^{tY}\right] = \mathbb{E}\left[e^{t\{a_1 g_2(Y_1) + a_2 g_2(Y_2)\}}\right]$$

This is the mgf for Y. We can write

$$m_Y(t) = \mathbb{E}\left[e^{tY}\right] = \mathbb{E}\left[e^{ta_1 g_1(Y_1)} e^{ta_2 g_2(Y_2)}\right].$$

Sums of random variables (revisited) (cont.)

If Y_1 and Y_2 are independent we can further write

$$m_Y(t) = \mathbb{E}\left[e^{ta_1 g_1(Y_1)}\right] \mathbb{E}\left[e^{ta_2 g_2(Y_2)}\right].$$

– this follows by the earlier result on expectations.

Sums of random variables (revisited) (cont.)

Now consider the special case $Y = Y_1 + Y_2$:

$$m_Y(t) = \mathbb{E}\left[e^{t\{Y_1+Y_2\}}\right].$$

It follows immediately that if Y_1 and Y_2 are independent,

$$m_Y(t) = \mathbb{E}\left[e^{tY_1}\right]\left[e^{tY_2}\right] = m_{Y_1}(t)m_{Y_2}(t).$$

This gives us a way to compute the distribution of sums of independent random variables.

Sums of random variables (revisited) (cont.)

Example (Sum of Binomial random variables)

If Y_1 and Y_2 are independent Binomial random variables with parameters (n_1, p) and (n_2, p), we have

$$m_{Y_1}(t) = (pe^t + q)^{n_1} \qquad m_{Y_2}(t) = (pe^t + q)^{n_2}$$

then if $Y = Y_1 + Y_2$, we have that

$$m_Y(t) = m_{Y_1}(t) m_{Y_2}(t) = (pe^t + q)^{n_1} (pe^t + q)^{n_2} = (pe^t + q)^{n_1 + n_2}$$

and we recognize this as the mgf of the Binomial distribution with parameters $(n_1 + n_2, p)$.

Sums of random variables (revisited) (cont.)

Example (Sum of Poisson random variables)

If Y_1 and Y_2 are independent Poisson random variables with parameters λ_1 and λ_2, we have

$$m_{Y_1}(t) = \exp\{\lambda_1(e^t - 1)\} \qquad m_{Y_2}(t) = \exp\{\lambda_2(e^t - 1)\}$$

then if $Y = Y_1 + Y_2$, we have that

$$\begin{aligned} m_Y(t) = m_{Y_1}(t)m_{Y_2}(t) &= \exp\{\lambda_1(e^t - 1)\}\exp\{\lambda_2(e^t - 1)\} \\ &= \exp\{(\lambda_1 + \lambda_2)(e^t - 1)\} \end{aligned}$$

and we recognize this as the mgf of the Poisson distribution with parameter $\lambda_1 + \lambda_2$.

Sums of random variables (revisited) (cont.)

Example (Sum of Normal random variables)

If Y_1 and Y_2 are independent Normal random variables with parameters (μ_1, σ_1^2) and (μ_2, σ_2^2), we have

$$m_{Y_1}(t) = \exp\{\mu_1 t + \sigma_1^2 t^2/2\} \qquad m_{Y_2}(t) = \exp\{\mu_2 t + \sigma_2^2 t^2/2\}$$

then if $Y = Y_1 + Y_2$, we have that

$$\begin{aligned} m_Y(t) = m_{Y_1}(t) m_{Y_2}(t) &= \exp\{\mu_1 t + \sigma_1^2 t^2/2\} \exp\{\mu_2 t + \sigma_2^2 t^2/2\} \\ &= \exp\{(\mu_1 + \mu_2)t + (\sigma_1^2 + \sigma_2^2)t^2/2\} \end{aligned}$$

and we recognize this as the mgf of the Normal distribution with parameters $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Multivariate transformations

It is possible to consider general multivariate transformations of variables. For example if Y_1 and Y_2 are two variables with joint pdf $f_{Y_1,Y_2}(y_1, y_2)$, we can consider

$$U_1 = h_1(Y_1, Y_2) \qquad U_2 = h_2(Y_1, Y_2)$$

and attempt to compute

$$f_{U_1,U_2}(u_1, u_2).$$

Multivariate transformations (cont.)

For example, we might have

$$U_1 = Y_1 + Y_2 \qquad U_2 = Y_1 - Y_2$$

or

$$U_1 = Y_1 Y_2 \qquad U_2 = Y_1 / Y_2$$

etc.

Multivariate transformations (cont.)

We can approach this from a 'change of variable' perspective. If the transformation

$$(Y_1, Y_2) \longrightarrow (U_1, U_2)$$

is a one-to-one transformation, then the inverse transformation

$$(U_1, U_2) \longrightarrow (Y_1, Y_2)$$

is well-defined. Suppose we write

$$Y_1 = h_1^{-1}(U_1, U_2) \qquad Y_2 = h_2^{-1}(U_1, U_2).$$

Multivariate transformations (cont.)

Then we may derive that for $(u_1, u_2) \in \mathbb{R}^2$,

$$f_{U_1,U_2}(u_1, u_2) = f_{Y_1,Y_2}(h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2))|J(u_1, u_2)|$$

where $J(u_1, u_2)$ is the *Jacobian* of the transformation, that is

$$J(u_1, u_2) = \det \begin{bmatrix} \dfrac{\partial h_1^{-1}(u_1, u_2)}{\partial u_1} & \dfrac{\partial h_1^{-1}(u_1, u_2)}{\partial u_2} \\ \dfrac{\partial h_2^{-1}(u_1, u_2)}{\partial u_1} & \dfrac{\partial h_2^{-1}(u_1, u_2)}{\partial u_2} \end{bmatrix}$$

where det[.] denotes determinant, and |.| denotes absolute value.

Multivariate transformations (cont.)

This result holds provided that all the partial derivatives exist and are continuous, and that

$$|J(u_1, u_2)| > 0.$$

The result extends beyond two dimensions.