MATH 556: MATHEMATICAL STATISTICS I

MULTIVARIATE PROBABILITY DISTRIBUTIONS: EXAMPLES

Discrete bivariate distributions: We consider two variables X_1 and X_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

$$f_{X_1,X_2}(x_1, x_2)$$

that distributes probability across the possible values of the vector (X_1, X_2) so that

$$f_{X_1,X_2}(x_1, x_2) = P_{X_1,X_2}((X_1 = x_1) \cap (X_2 = x_2)) \equiv P_{X_1,X_2}(X_1 = x_1, X_2 = x_2)$$

for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$. The function $f_{X_1,X_2}(x_1, x_2)$ is the *joint probability mass function*: it has two basic properties

- "specifies probabilities"
$$0 \leq f_{X_1,X_2}(x_1, x_2) \leq 1 \quad \text{for all } x_1, x_2$$
- "sums to one"
$$\sum_{x_1=-\infty}^{\infty} \sum_{x_2=-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) = 1.$$
although $f_{X_1,X_2}(x_1, x_2)$ may be zero for some arguments.

We can think of $f_{X_1,X_2}(x_1, x_2)$ as specifying the values in a probability table.

		X_2		
		1	2	3
	1	$f_{X_1,X_2}(1,1)$	$f_{X_1,X_2}(1,2)$	$f_{X_1,X_2}(1,3)$
	2	$f_{X_1,X_2}(2,1)$	$f_{X_1,X_2}(2,2)$	$f_{X_1,X_2}(2,3)$
X_1	3	$f_{X_1,X_2}(3,1)$	$f_{X_1,X_2}(3,2)$	$f_{X_1,X_2}(3,3)$
	4	$f_{X_1,X_2}(4,1)$	$f_{X_1,X_2}(4,2)$	$f_{X_1,X_2}(4,3)$
	5	$f_{X_1,X_2}(5,1)$	$f_{X_1,X_2}(5,2)$	$f_{X_1,X_2}(5,3)$

Example: For $1 \leq x_1 \leq 5, 1 \leq x_2 \leq 3$

$$f_{X_1,X_2}(x_1, x_2) = \frac{(x_1 + x_2)}{75}$$

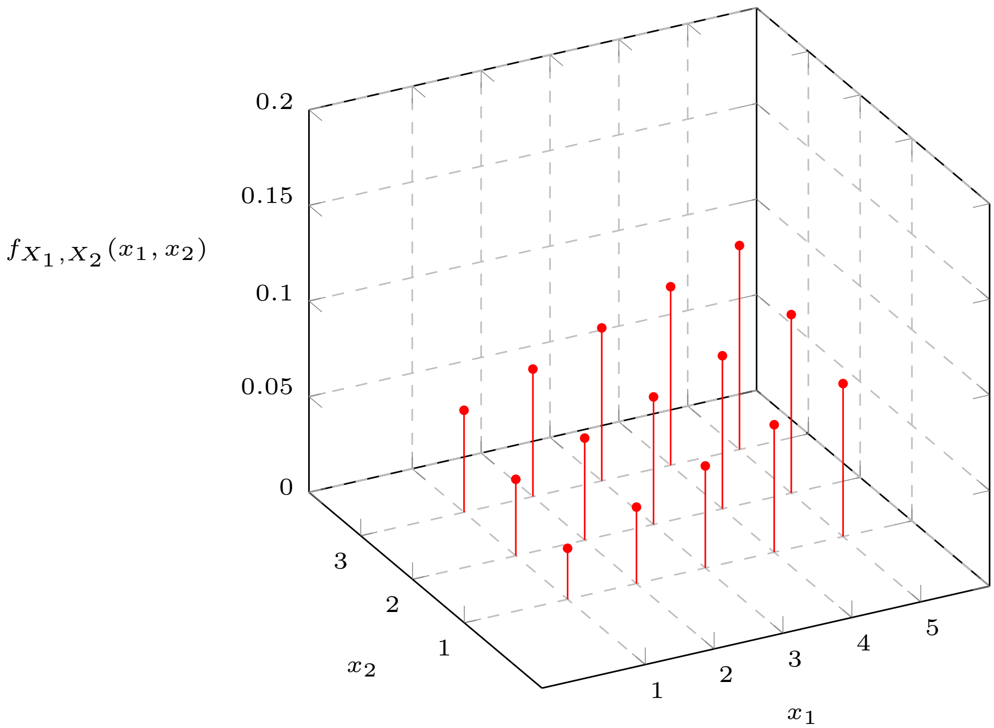
		X_2		
		1	2	3
	1	2/75	3/75	4/75
	2	3/75	4/75	5/75
X_1	3	4/75	5/75	6/75
	4	5/75	6/75	7/75
	5	6/75	7/75	8/75

In the above example,

$$\sum_{x_1=1}^{5} \sum_{x_2=1}^{3} \frac{(x_1 + x_2)}{75} = \frac{1}{75} \sum_{x_1=1}^{5} \sum_{x_2=1}^{3} (x_1 + x_2) = \frac{1}{75} \left[3 \sum_{x_1=1}^{5} x_1 + 5 \sum_{x_2=1}^{3} x_2 \right]$$

$$= \frac{1}{75} \left[3 \frac{5 \times 6}{2} + 5 \frac{3 \times 4}{2} \right]$$

$$= \frac{1}{75} [45 + 30] = 1.$$

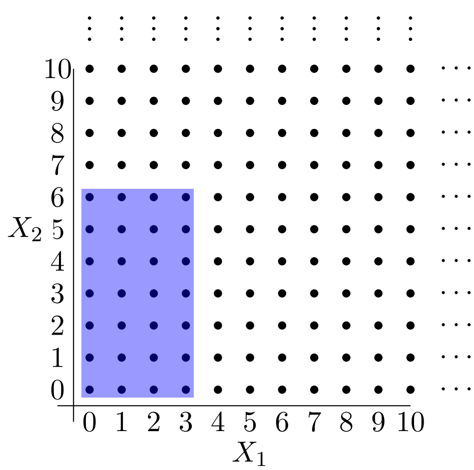


We define the *joint cumulative distribution function* $F_{X_1,X_2}(x_1, x_2)$ by

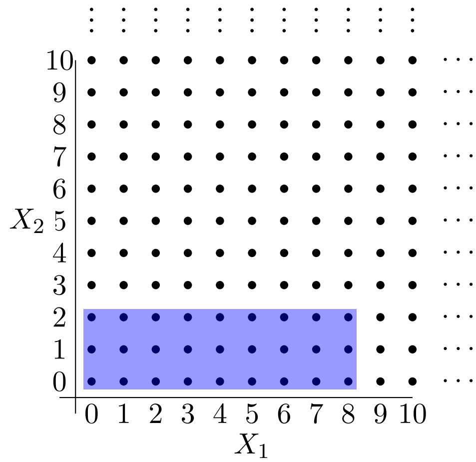
$$F_{X_1,X_2}(x_1, x_2) = P_{X_1,X_2}(X_1 \leq x_1, X_2 \leq x_2) = \sum_{t_1=-\infty}^{x_1} \sum_{t_2=-\infty}^{x_2} f_{X_1,X_2}(t_1, t_2)$$

that is, by summing probabilities in the joint pmf over a range of values up to and including (x_1, x_2)

$$F_{X_1,X_2}(3, 6) = P_{X_1,X_2}(X_1 \leq 3, X_2 \leq 6) = \sum_{t_1=0}^{3} \sum_{t_2=0}^{6} f_{X_1,X_2}(t_1, t_2)$$



$$F_{X_1,X_2}(8,2) = P_{X_1,X_2}(X_1 \le 8, X_2 \le 2) = \sum_{t_1=0}^{8}\sum_{t_2=0}^{2} f_{X_1,X_2}(t_1,t_2)$$



Example 1 A bag contains ten balls:

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

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

- X_1 denote the number of red balls selected;
- X_2 denote the number of yellow balls selected.

Then using combinatorial arguments, we see that the joint pmf of X_1 and X_2 is given by

$$f_{X_1,X_2}(x_1,x_2) = \frac{\binom{5}{x_1}\binom{3}{x_2}\binom{2}{4-x_1-x_2}}{\binom{10}{4}}$$

for (x_1,x_2) such that the combinatorial terms are defined, and zero when the terms are not. We need (x_1,x_2) simultaneously to satisfy

$$0 \le x_1 \le 5 \qquad 0 \le x_2 \le 3 \qquad 0 \le 4 - x_1 - x_2 \le 2$$

in order to have a non-zero probability. Total number of selections: $\binom{10}{4} = 210$.

Red (x_1)	Yellow (x_2)	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

$f_{X_1,X_2}(x_1,x_2)$

		X_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
X_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

The marginal mass function: Suppose that the joint pmf for X_1 and X_2 is denoted $f_{X_1,X_2}(.,.)$. Then the marginal pmf for X_1, $f_{X_1}(.)$ is given by

$$f_{X_1}(x_1) = P_{X_1}(X_1 = x_1) = \sum_{x_2=-\infty}^{\infty} P_{X_1,X_2}(X_1 = x_1, X_2 = x_2)$$

that is

$$f_{X_1}(x_1) = \sum_{x_2=-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)$$

This result uses a partitioning argument:

$$(X_1 = x_1) = \bigcup_{x_2=-\infty}^{\infty} (X_1 = x_1) \cap (X_2 = x_2)$$

For example

$$P_{X_1}(X_1 = 2) = P_{X_1,X_2}(X_1 = 2, X_2 = 1) + P_{X_1,X_2}(X_1 = 2, X_2 = 2) + P_{X_1,X_2}(X_1 = 2, X_2 = 3).$$

If $f_{X_1,X_2}(x_1, x_2)$ specifies the values in a probability table, we compute the marginal pmf

- for X_1 by summing *across the rows* of the table;
- for X_2 by summing *down the columns* of the table.

		X_2			
		1	2	3	$f_{X_1}(.)$
	1	$f_{X_1,X_2}(1,1)$	$f_{X_1,X_2}(1,2)$	$f_{X_1,X_2}(1,3)$	$f_{X_1}(1)$
	2	$f_{X_1,X_2}(2,1)$	$f_{X_1,X_2}(2,2)$	$f_{X_1,X_2}(2,3)$	$f_{X_1}(2)$
X_1	3	$f_{X_1,X_2}(3,1)$	$f_{X_1,X_2}(3,2)$	$f_{X_1,X_2}(3,3)$	$f_{X_1}(3)$
	4	$f_{X_1,X_2}(4,1)$	$f_{X_1,X_2}(4,2)$	$f_{X_1,X_2}(4,3)$	$f_{X_1}(4)$
	5	$f_{X_1,X_2}(5,1)$	$f_{X_1,X_2}(5,2)$	$f_{X_1,X_2}(5,3)$	$f_{X_1}(5)$
	$f_{X_2}(.)$	$f_{X_2}(1)$	$f_{X_2}(2)$	$f_{X_2}(3)$	1

Example 2 [Previous example]
Four balls selected from 10.

- X_1 denote the number of red balls selected;
- X_2 denote the number of yellow balls selected.

The joint pmf of X_1 and X_2 is given by

$$f_{X_1,X_2}(x_1, x_2) = \frac{\binom{5}{x_1}\binom{3}{x_2}\binom{2}{4 - x_1 - x_2}}{\binom{10}{4}}$$

for (x_1, x_2) such that the combinatorial terms are defined, and zero when the terms are not.

We can compute the marginal pmf for X_1 by summing probabilities in the joint probability table.

$$f_{X_1}(0) = \frac{\binom{5}{0}\binom{3}{2}\binom{2}{2} + \binom{5}{0}\binom{3}{3}\binom{2}{1}}{\binom{10}{4}} = \frac{3+2}{210} = \frac{5}{210}$$

$$f_{X_1}(1) = \frac{\binom{5}{1}\binom{3}{1}\binom{2}{2} + \binom{5}{1}\binom{3}{2}\binom{2}{1} + \binom{5}{1}\binom{3}{3}\binom{2}{0}}{\binom{10}{4}} = \frac{15+30+5}{210} = \frac{50}{210}$$

$$f_{X_1}(2) = \frac{\binom{5}{2}\binom{3}{0}\binom{2}{2} + \binom{5}{2}\binom{3}{1}\binom{2}{1} + \binom{5}{2}\binom{3}{0}\binom{2}{2}}{\binom{10}{4}} = \frac{10+60+30}{210} = \frac{100}{210}$$

$$f_{X_1}(3) = \frac{\binom{5}{3}\binom{3}{0}\binom{2}{1} + \binom{5}{3}\binom{3}{1}\binom{2}{0}}{\binom{10}{4}} = \frac{20+30}{210} = \frac{50}{210}$$

$$f_{X_1}(4) = \frac{\binom{5}{4}\binom{3}{0}\binom{2}{0}}{\binom{10}{4}} = \frac{5}{210}$$

Note: In this example we can compute $f_{X_1}(.)$ *directly* using the hypergeometric formula

$$f_{X_1}(x_1) = \frac{\binom{5}{x_1}\binom{5}{4-x_1}}{\binom{10}{4}}$$

for $0 \le x_1 \le 5$ and $0 \le 4 - x_1 \le 5$.

x_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 pmfs $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ have all the properties of single variable pmfs; specifically

$$\sum_{x_1=-\infty}^{\infty} f_{X_1}(x_1) = \sum_{x_1=-\infty}^{\infty}\sum_{x_2=-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2) = 1.$$

The conditional mass function: Once we have the joint pmf

$$f_{X_1,X_2}(x_1,x_2) = P_{X_1,X_2}(X_1 = x_1, X_2 = x_2)$$

and the marginal pmf

$$f_{X_1}(x_1) = P_{X_1}(X_1 = x_1)$$

we can consider *conditional* pmfs. We have that if $P_{X_1}(X_1 = x_1) > 0$, then the conditional probability that $X_2 = x_2$, given that $X_1 = x_1$, is

$$P(X_2 = x_2 | X_1 = x_1) = \frac{P_{X_1,X_2}(X_1 = x_1, X_2 = x_2)}{P_{X_1}(X_1 = x_1)}$$

For a **fixed** value of x_1, we can consider how this conditional probability varies as argument x_2 varies. The *conditional probability mass function* for X_2, given that $X_1 = x_1$, is denoted

$$f_{X_2|X_1}(x_2|x_1)$$

and defined by

$$f_{X_2|X_1}(x_2|x_1) = P(X_2 = x_2 | X_1 = x_1)$$

whenever $P_{X_1}(X_1 = x_1) > 0$.

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

$$f_{X_2|X_1}(x_2|x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} = \frac{f_{X_1,X_2}(x_1,x_2)}{\sum\limits_{t_2=-\infty}^{\infty} f_{X_1,X_2}(x_1,t_2)}.$$

Given $X_1 = 2$, we define the conditional pmf $f_{X_2|X_1}(x_2|2)$ by examining the second row of the table.

		X_2 1	2	3	$f_{X_1}(.)$
	1	$f_{X_1,X_2}(1,1)$	$f_{X_1,X_2}(1,2)$	$f_{X_1,X_2}(1,3)$	$f_{X_1}(1)$
	2	$f_{X_1,X_2}(2,1)$	$f_{X_1,X_2}(2,2)$	$f_{X_1,X_2}(2,3)$	$f_{X_1}(2)$
X_1	3	$f_{X_1,X_2}(3,1)$	$f_{X_1,X_2}(3,2)$	$f_{X_1,X_2}(3,3)$	$f_{X_1}(3)$
	4	$f_{X_1,X_2}(4,1)$	$f_{X_1,X_2}(4,2)$	$f_{X_1,X_2}(4,3)$	$f_{X_1}(4)$
	5	$f_{X_1,X_2}(5,1)$	$f_{X_1,X_2}(5,2)$	$f_{X_1,X_2}(5,3)$	$f_{X_1}(5)$
	$f_{X_2}(.)$	$f_{X_2}(1)$	$f_{X_2}(2)$	$f_{X_2}(3)$	1

Given $X_2 = 3$, we define the conditional pmf $f_{X_1|X_2}(x_1|3)$ by examining the third column of the table.

		X_2 1	2	3	$f_{X_1}(.)$
	1	$f_{X_1,X_2}(1,1)$	$f_{X_1,X_2}(1,2)$	$f_{X_1,X_2}(1,3)$	$f_{X_1}(1)$
	2	$f_{X_1,X_2}(2,1)$	$f_{X_1,X_2}(2,2)$	$f_{X_1,X_2}(2,3)$	$f_{X_1}(2)$
X_1	3	$f_{X_1,X_2}(3,1)$	$f_{X_1,X_2}(3,2)$	$f_{X_1,X_2}(3,3)$	$f_{X_1}(3)$
	4	$f_{X_1,X_2}(4,1)$	$f_{X_1,X_2}(4,2)$	$f_{X_1,X_2}(4,3)$	$f_{X_1}(4)$
	5	$f_{X_1,X_2}(5,1)$	$f_{X_1,X_2}(5,2)$	$f_{X_1,X_2}(5,3)$	$f_{X_1}(5)$
	$f_{X_2}(.)$	$f_{X_2}(1)$	$f_{X_2}(2)$	$f_{X_2}(3)$	1

Note: We have the fundamental relationship

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1)$$

whenever $f_{X_1}(x_1) > 0$.

Continuous case: joint density function: If X_1 and X_2 are two *continuous* random variables, then we can still consider statements of the form

$$P_{X_1,X_2}((X_1 \leq x_1) \cap (X_2 \leq x_2))$$

and hence define the *joint cumulative distribution function* cdf

$$F_{X_1,X_2}(x_1, x_2) = P((X_1 \leq x_1) \cap (X_2 \leq x_2))$$

for any pair of real numbers (x_1, x_2).

The joint cdf has the following properties:

- "starts at zero"
$$\lim_{x_1 \longrightarrow -\infty} \lim_{x_2 \longrightarrow -\infty} F_{X_1,X_2}(x_1, x_2) = 0$$
- "ends at one"
$$\lim_{x_1 \longrightarrow \infty} \lim_{x_2 \longrightarrow \infty} F_{X_1,X_2}(x_1, x_2) = 1$$
- "non-decreasing in x_1 and x_2 in between"
$$F_{X_1,X_2}(x_1, x_2) \leq F_{X_1,X_2}(x_1 + h, x_2)$$
$$F_{X_1,X_2}(x_1, x_2) \leq F_{X_1,X_2}(x_1, x_2 + h)$$
for all x_1, x_2, and any $h > 0$.

Furthermore, we have that

$$\lim_{x_1 \longrightarrow \infty} F_{X_1,X_2}(x_1, x_2) = P_{X_1,X_2}(X_1 < \infty, X_2 \leq x_2) = P_{X_1}(X_2 \leq x_2) = F_{X_2}(x_2)$$

and similarly

$$\lim_{x_2 \longrightarrow \infty} F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1).$$

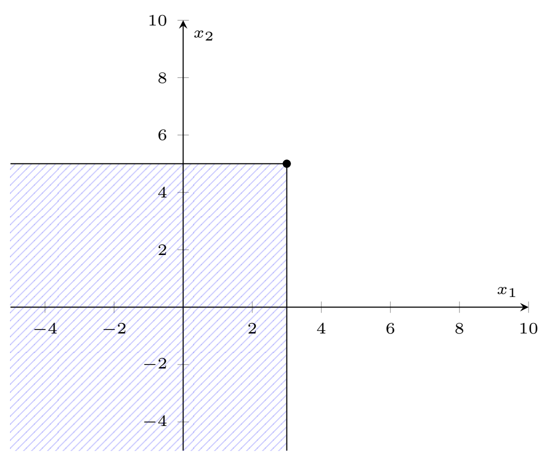
Regions of integration: to compute the joint cdf, we accumulate probability over the shaded region, the rectangle

$$(-\infty, x_1] \times (-\infty, x_2],$$

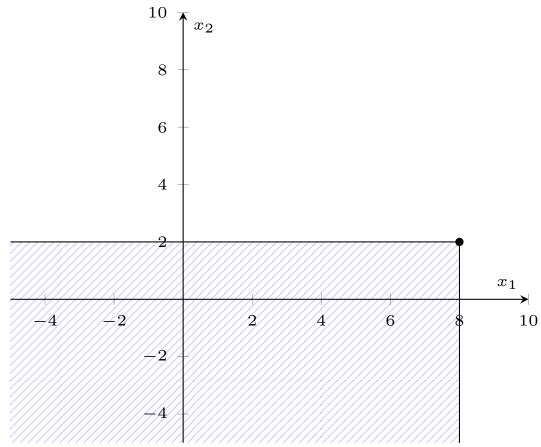
to compute $F_{X_1,X_2}(x_1, x_2)$. However, as in the single variable case, we must have

$$P_{X_1,X_2}(X_1 = x_1, X_2 = x_2) = 0$$

for all x_1 and x_2.

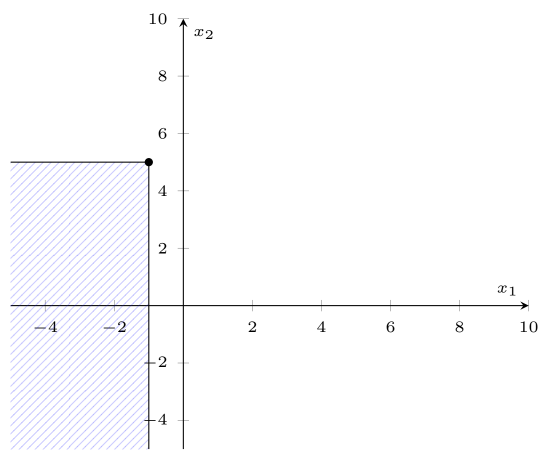


$$F_{X_1,X_2}(3, 5) = P_{X_1,X_2}(X_1 \leq 3, X_2 \leq 5)$$



$$F_{X_1,X_2}(8,2) = P_{X_1,X_2}(X_1 \leq 8, X_2 \leq 2)$$

$F_{X_1,X_2}(-1,5) = P_{X_1,X_2}(X_1 \leq -1, X_2 \leq 5)$



$$F_{X_1,X_2}(-1,5) = P_{X_1,X_2}(X_1 \leq -1, X_2 \leq 5)$$

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

$$f_{X_1,X_2}(x_1,x_2)$$

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

$$F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1,X_2}(t_1,t_2)\,dt_2\,dt_1$$

that is, to compute $F_{X_1,X_2}(x_1,x_2)$ we *integrate* $f_{X_1,X_2}(x_1,x_2)$ over the rectangle

$$(-\infty, x_1] \times (-\infty, x_2].$$

We compute the double integral as follows: writing

$$\int_{-\infty}^{x_1}\left\{\int_{-\infty}^{x_2} f_{X_1,X_2}(t_1,t_2)\,dt_2\right\}dt_1$$

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

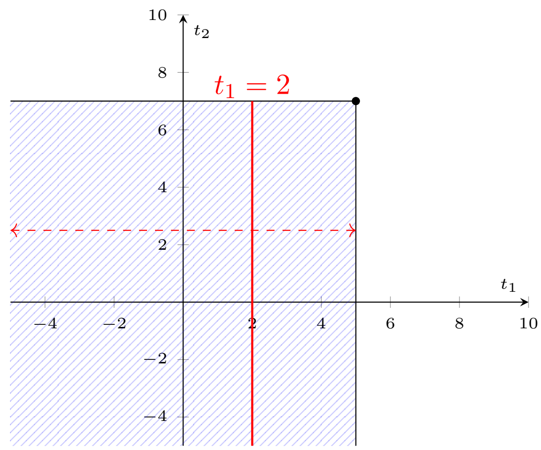
$$\int_{-\infty}^{x_2} f_{X_1,X_2}(t_1,t_2)\, dt_2$$

  in the 'strip' at t_1 to obtain a function $g(t_1, x_2)$, say;

- perform the second (outer) integration

$$\int_{-\infty}^{x_1} g(t_1,x_2)\, dt_1.$$

  to obtain the joint cdf.



$$F_{X_1,X_2}(5,7) = P_{X_1,X_2}(X_1 \leq 5, X_2 \leq 7)$$

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*

$$f_{X_1,X_2}(x_1,x_2) \geq 0 \qquad -\infty < x_1 < \infty, -\infty < x_2 < \infty$$

  (as the joint cdf is non-decreasing in both x_1 and x_2);

- the joint pdf *integrates to 1*

$$\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2)\, dx_2 \right\} dx_1 = 1$$

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

Example 3 Suppose X_1 and X_2 are continuous with joint pdf

$$f_{X_1,X_2}(x_1,x_2) = c(x_1+x_2) \qquad 0 \le x_1 \le 1, 0 \le x_2 \le 1$$

with $f_{X_1,X_2}(x_1,x_2) = 0$ otherwise. Then for $0 \le x_1 \le 1, 0 \le x_2 \le 1$,

$$\begin{aligned}
F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1} \left\{ \int_{-\infty}^{x_2} f_{X_1,X_2}(t_1,t_2)\, dt_2 \right\} dt_1 &= \int_0^{x_1} \left\{ \int_0^{x_2} c(t_1+t_2)\, dt_2 \right\} dt_1 \\
&= \int_0^{x_1} \left[c\left(t_1 t_2 + \frac{1}{2} t_2^2 \right) \right]_0^{x_2} dt_1 \\
&= \int_0^{x_1} c\left(t_1 x_2 + \frac{1}{2} x_2^2 \right) dt_1 \\
&= \left[c\left(\frac{1}{2} t_1^2 x_2 + \frac{1}{2} x_2^2 t_1 \right) \right]_0^{x_1} \\
&= \frac{c}{2}\left(x_1^2 x_2 + x_1 x_2^2 \right)
\end{aligned}$$

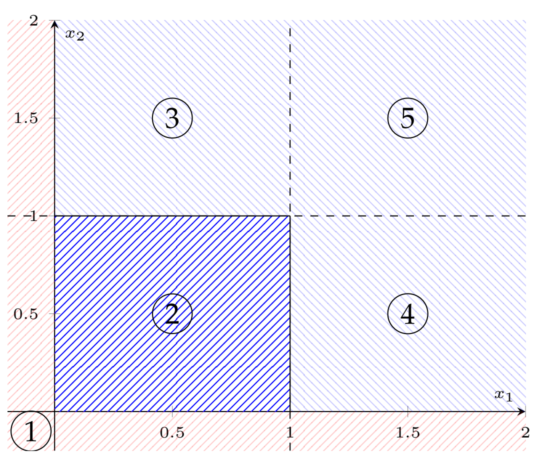
We require that $F_{X_1,X_2}(1,1) = 1$, so we must have $c = 1$. That is

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} (x_1+x_2) & 0 \le x_1 \le 1, 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(no probability outside of the unit square).

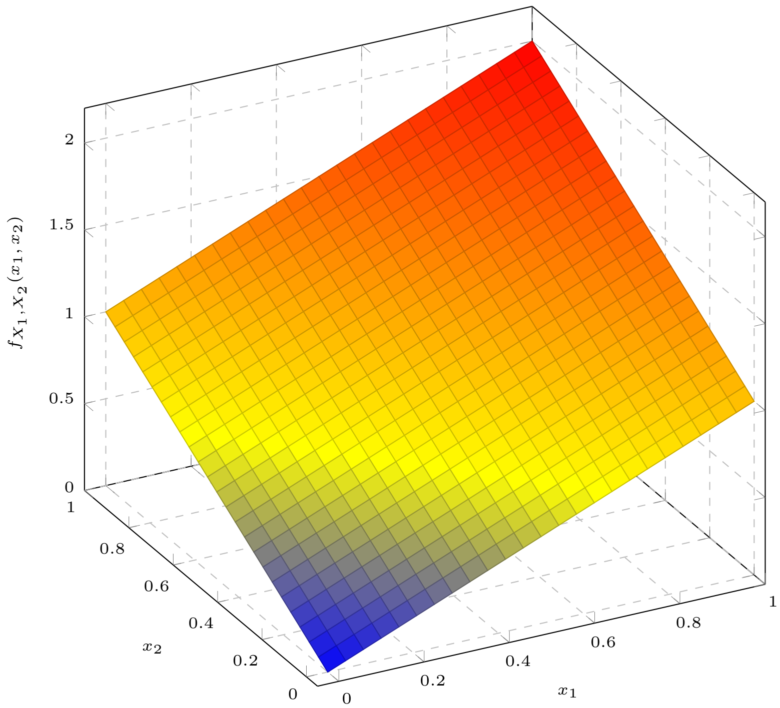
$$F_{X_1,X_2}(x_1,x_2) = \begin{cases} 0 & x_1 < 0 \text{ or } x_2 < 0 & (1) \\ \left(x_1^2 x_2 + x_1 x_2^2\right)/2 & 0 \le x_1, x_2 \le 1 & (2) \\ \left(x_1^2 + x_1\right)/2 & 0 \le x_1 \le 1, x_2 > 1 & (3) \\ \left(x_2 + x_2^2\right)/2 & 0 \le x_2 \le 1, x_1 > 1 & (4) \\ 1 & x_1 > 1 \text{ and } x_2 > 1 & (5) \end{cases}$$

Note: regions for $F_{X_1,X_2}(x_1,x_2)$

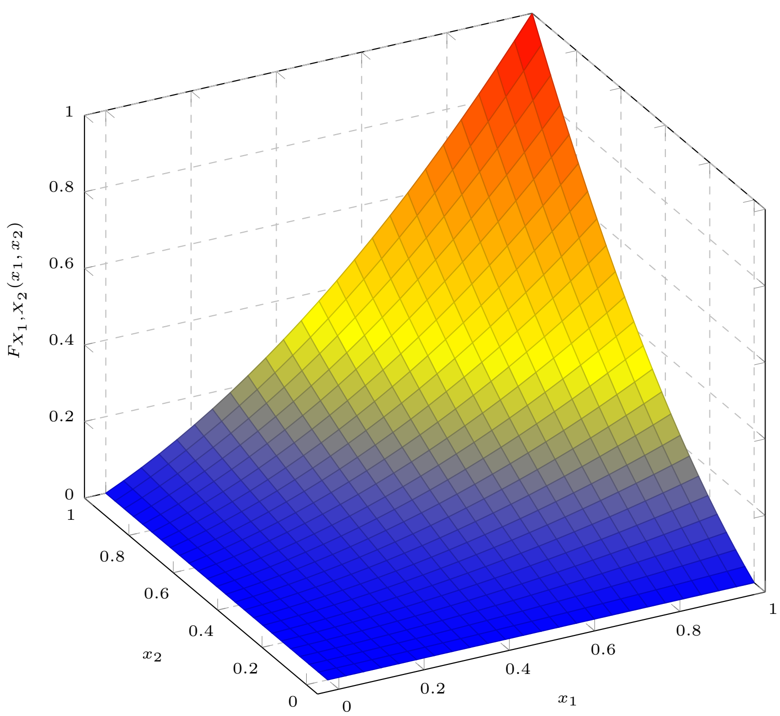


Note: To compute $F_{X_1,X_2}(x_1,x_2)$ we always integrate the joint pdf *below and to the left* of (x_1,x_2).

$$f_{X_1,X_2}(x_1, x_2) = (x_1 + x_2)$$



$$F_{X_1,X_2}(x_1, x_2) = (x_1^2 x_2 + x_1 x_2^2)/2$$



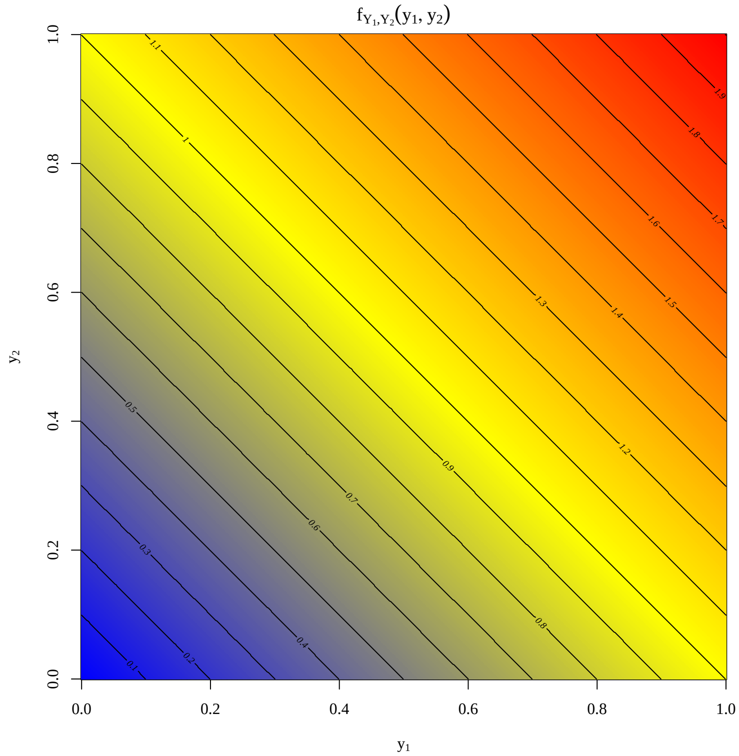


Figure 1: $f_{X_1,X_2}(x_1, x_2)$: image and contour plot

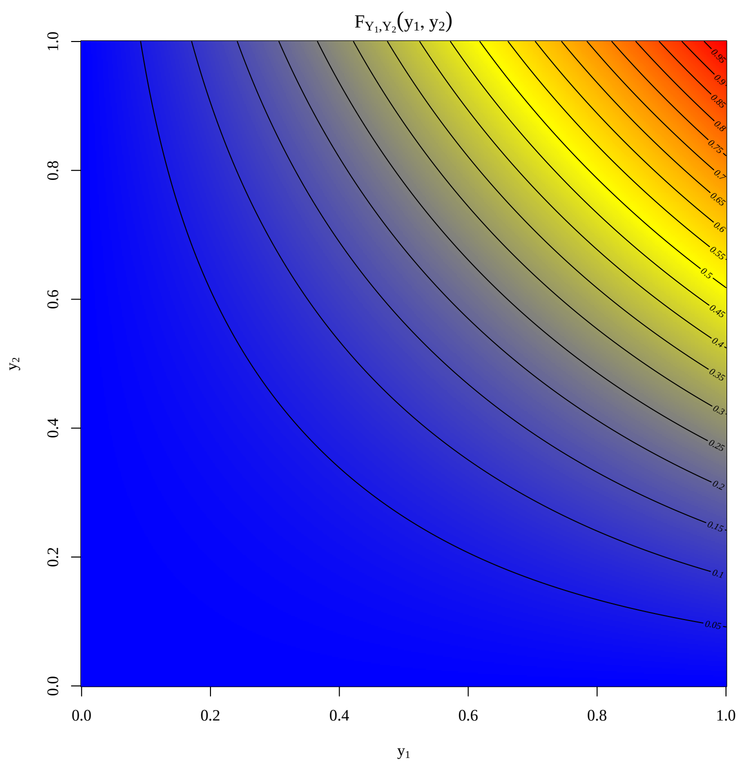


Figure 2: $F_{X_1,X_2}(x_1, x_2)$: image and contour plot

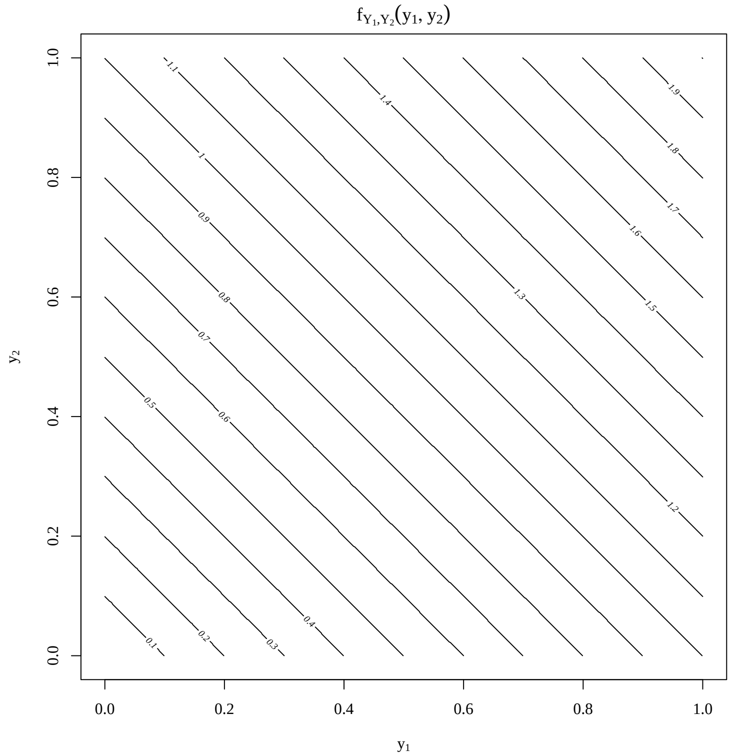


Figure 3: $f_{X_1,X_2}(x_1, x_2)$: contour plot

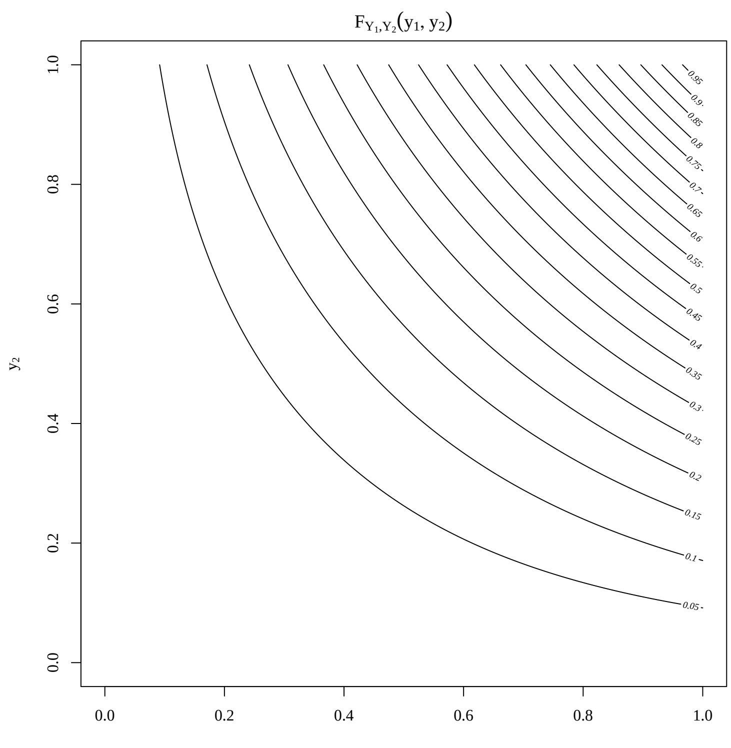


Figure 4: $F_{X_1,X_2}(x_1, x_2)$: contour plot

We compute $f_{X_1,X_2}(x_1, x_2)$ from $F_{X_1,X_2}(x_1, x_2)$ using *partial differentiation*:

$$f_{X_1,X_2}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} \{F_{X_1,X_2}(x_1, x_2)\}$$

- **Step 1:** differentiate $F_{X_1,X_2}(x_1, x_2)$ with respect to x_2 while holding x_1 constant;
- **Step 2:** take the result of **Step 1**, and differentiate it with respect to x_1.

We can regard the calculation as

$$f_{X_1,X_2}(x_1, x_2) = \frac{\partial}{\partial x_1} \left\{ \frac{\partial F_{X_1,X_2}(x_1, x_2)}{\partial x_2} \right\}$$

Example 4 [Previous example] For $0 \leq x_1, x_2 \leq 1$,

$$F_{X_1,X_2}(x_1, x_2) = \frac{1}{2}\left(x_1^2 x_2 + x_1 x_2^2\right)$$

- **Step 1:**

$$\frac{\partial F_{X_1,X_2}(x_1, x_2)}{\partial x_2} = \frac{1}{2}\left(x_1^2 + 2x_1 x_2\right)$$

- **Step 2:**

$$\frac{\partial \left\{ \frac{1}{2}\left(x_1^2 + 2x_1 x_2\right) \right\}}{\partial x_2} = \frac{1}{2}(2x_1 + 2x_2) = (x_1 + x_2)$$

Note: In the partial differentiation, we could choose to differentiate with respect to x_1 in Step 1, and then differentiate the result with respect to x_2: we will get the same answer

$$\frac{\partial^2}{\partial x_1 \partial x_2} \{F_{X_1,X_2}(x_1, x_2)\} = \frac{\partial^2}{\partial x_2 \partial x_1} \{F_{X_1,X_2}(x_1, x_2)\}.$$

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

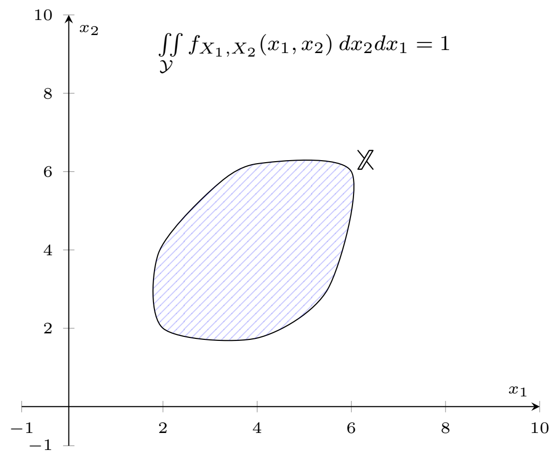
Constant pdfs: If the joint pdf is *constant* over a finite region $\mathcal{Y} \in \mathbb{R}^2$

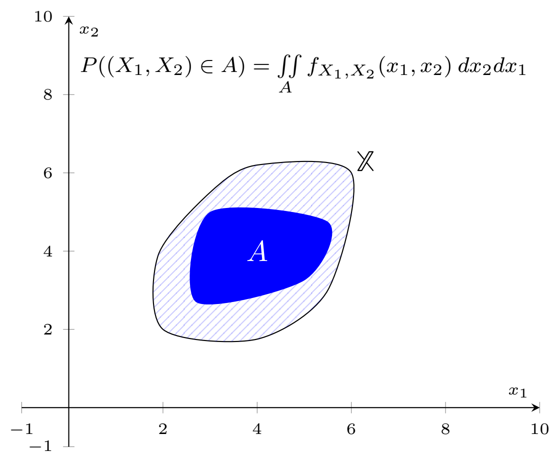
$$f_{X_1,X_2}(x_1, x_2) = c \qquad (x_1, x_2) \in \mathcal{Y}$$

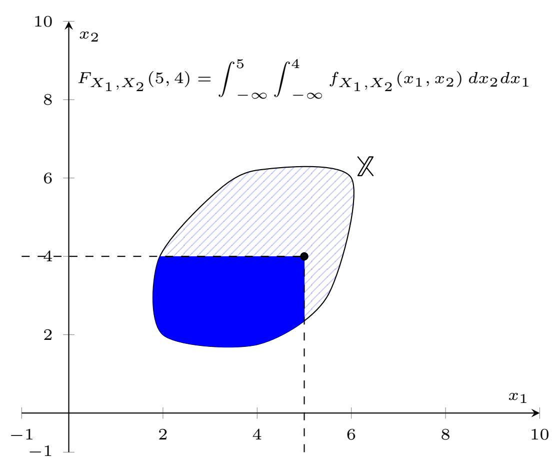
with $f_{X_1,X_2}(x_1, x_2)$ zero otherwise, then to compute probabilities associated with this pdf, for example

$$P_{X_1,X_2}((X_1, X_2) \in A) = \iint\limits_A f_{X_1,X_2}(x_1, x_2)\, dx_2 dx_1$$

we must compute the *area* of A and divide it by the *area* of $\mathbb{X}$.







The marginal density function: Once the joint pdf $f_{X_1,X_2}(x_1, x_2)$ is specified, we define the *marginal probability density functions* (marginal pdfs) analogously to the discrete case.

We have for X_1 the marginal pdf $f_{X_1}(x_1)$ for each fixed x_1 by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2$$

and for X_2 the marginal pdf $f_{X_2}(x_2)$ for each fixed x_2 by

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_1$$

Example 5 [Distribution on the unit square]
For

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} (x_1 + x_2) & 0 \le x_1 \le 1, 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

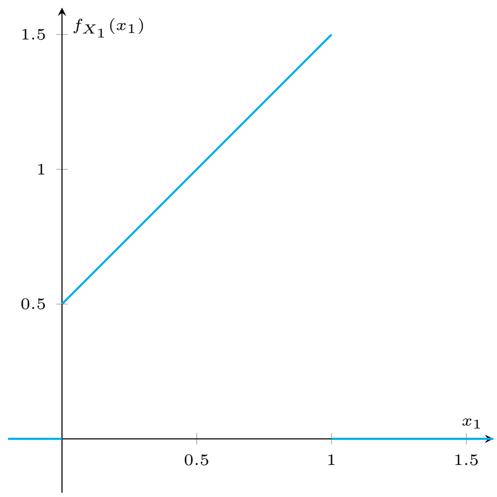
We have for $0 \le x_1 \le 1$

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2 \\ &= \int_0^1 (x_1 + x_2)\, dx_2 \\ &= \left[x_1 x_2 + \frac{x_2^2}{2} \right]_0^1 \\ &= x_1 + \frac{1}{2}. \end{aligned}$$

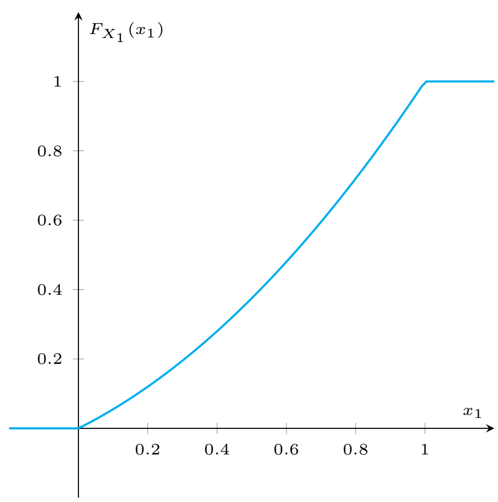
with $f_{X_1}(x_1) = 0$ otherwise. Also for $0 \le x_1 \le 1$

$$\begin{aligned} F_{X_1}(x_1) &= \int_{-\infty}^{x_1} f_{X_1}(t_1)\, dt_1 \\ &= \int_0^{x_1} \left(t_1 + \frac{1}{2} \right)\, dt_2 \\ &= \left[\frac{t_1^2}{2} + \frac{t_1}{2} \right]_0^{x_1} \\ &= \frac{x_1^2}{2} + \frac{x_1}{2}. \end{aligned}$$

with $F_{X_1}(x_1) = 0$ for $x_1 < 0$ and $F_{X_1}(x_1) = 1$ for $x_1 > 1$.



$f_{X_1}(x_1) = x_1 + 1/2$ for $0 \leq x_1 \leq 1$



$F_{X_1}(x_1) = (x_1^2 + x_1)/2$ for $0 \leq x_1 \leq 1$

We can perform a similar calculation and obtain

$$f_{X_2}(x_2) = x_2 + \frac{1}{2} \qquad 0 \leq x_2 \leq 1$$

with $f_{X_2}(x_2) = 0$ otherwise, and

$$F_{X_2}(x_2) = \frac{(x_2^2 + x_2)}{2} \qquad 0 \leq x_2 \leq 1$$

with $F_{X_2}(x_2) = 0$ for $x_2 < 0$ and $F_{X_2}(x_2) = 1$ for $x_2 > 1$.

The conditional density function: Once we have the joint pdf

$$f_{X_1,X_2}(x_1,x_2)$$

and the marginal pdf

$$f_{X_1}(x_1)$$

we can consider *conditional* pdfs.

The *conditional probability density function* for X_2, given that $X_1 = x_1$, is denoted

$$f_{X_2|X_1}(x_2|x_1)$$

and defined by

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}$$

whenever $f_{X_1}(x_1) > 0$. The function $f_{X_2|X_1}(x_2|x_1)$ is a pdf in x_2 for every fixed x_1. That is, for every fixed x_1 where $f_{X_1}(x_1) > 0$,

- the conditional pdf is *non-negative*

$$f_{X_2|X_1}(x_2|x_1) \geq 0 \qquad -\infty < x_2 < \infty;$$

- the conditional pdf is *integrates to 1 over x_2*

$$\int_{-\infty}^{\infty} f_{X_2|X_1}(x_2|x_1)\, dx_2 = 1.$$

We can define $f_{X_1|X_2}(x_1|x_2)$ in an identical fashion. Conditional pdfs are obtained by taking a 'slice' through the joint pdf at $X_1 = x_1$, and then standardizing the slice so that the density integrates to one. As in the discrete case, we have the *chain rule factorization*

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1)$$

whenever $f_{X_1}(x_1) > 0$.

As ever, we can exchange the roles of X_1 and X_2 to obtain that

$$f_{X_1,X_2}(x_1,x_2) = f_{X_2}(x_2) f_{X_1|X_2}(x_1|x_2).$$

whenever $f_{X_1}(x_1) > 0$.

Example 6 [Distribution on the unit square]
For

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} (x_1 + x_2) & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

we can consider conditioning on values of x_1 or x_2 in the interval $[0,1]$. As

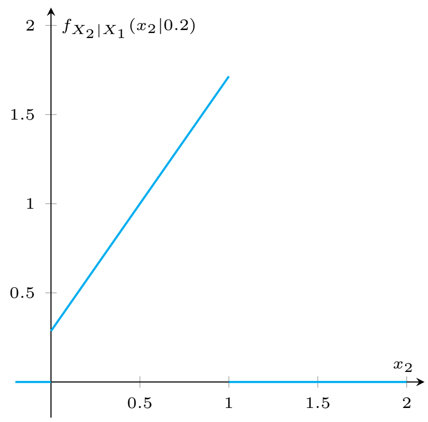
$$f_{X_1}(x_1) = x_1 + \frac{1}{2} \qquad 0 \leq x_1 \leq 1.$$

we have for each fixed x_1 in this range

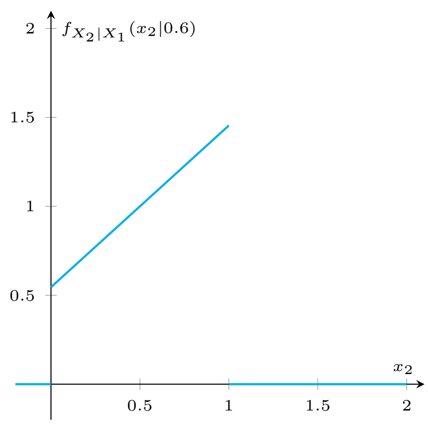
$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} = \frac{(x_1 + x_2)}{x_1 + 1/2}$$

for $0 \leq x_2 \leq 1$, with the function zero for other values of x_2.

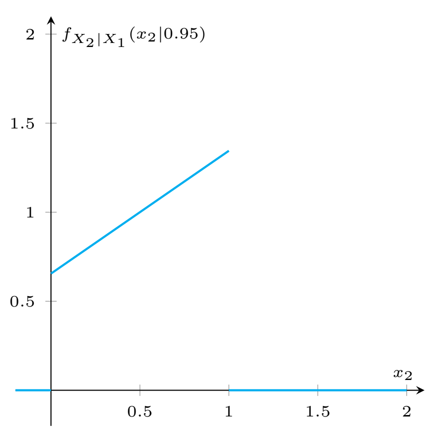
$f_{X_2|X_1}(x_2|x_1) = \dfrac{(x_1 + x_2)}{x_1 + 1/2}$ for $0 \leq x_2 \leq 1$, with $x_1 = 0.2$



$$f_{X_2|X_1}(x_2|x_1) = \frac{(x_1 + x_2)}{x_1 + 1/2} \text{ for } 0 \leq x_2 \leq 1\text{, with } x_1 = 0.2$$



$$f_{X_2|X_1}(x_2|x_1) = \frac{(x_1 + x_2)}{x_1 + 1/2} \text{ for } 0 \leq x_2 \leq 1\text{, with } x_1 = 0.6$$



$$f_{X_2|X_1}(x_2|x_1) = \frac{(x_1 + x_2)}{x_1 + 1/2} \text{ for } 0 \leq x_2 \leq 1\text{, with } x_1 = 0.95$$

General multivariate distributions: All of the above ideas extend naturally to more than two variables. If $(X_1, \ldots, X_d)$ form an n-dimensional random vector, we can consider the *joint pdf*

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$$

and *joint cdf*

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d).$$

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

$$f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)\, dx_3 dx_4$$

or

$$f_{X_2,X_4}(x_2, x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)\, dx_1 dx_3$$

and so on. This construction works with any d and any k. We can also define conditional pdfs for example

$$f_{X_1,X_3|X_2,X_4}(x_1, x_3|x_2, x_4) = \frac{f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)}{f_{X_2,X_4}(x_2, x_4)}$$

provided the denominator $f_{X_2,X_4}(x_2, x_4) > 0$.

Independence

Continuous random variables $(X_1, \ldots, X_d)$ are *independent* if

$$f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \prod_{i=1}^{d} f_{X_i}(x_i)$$

for **all** vectors $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Equivalently they are independent if

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \prod_{i=1}^{d} F_{X_i}(x_i)$$

for **all** vectors $(x_1, \ldots, x_d) \in \mathbb{R}^d$. Random variables that are not independent are termed *dependent*.

In the bivariate case, X_1 and X_2 are independent if

$$f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2)$$

for all (x_1, x_2) where the conditional pdf is defined, or equivalently if

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1)$$

for all (x_1, x_2).

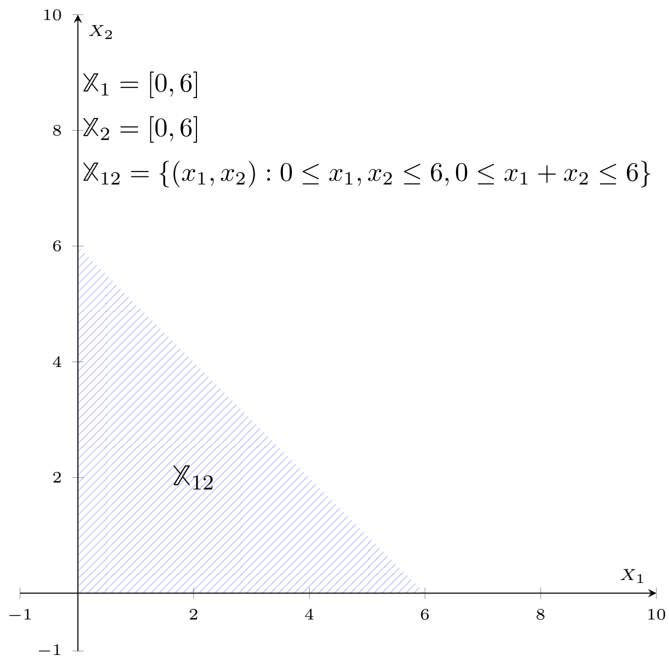
Note: A straightforward way to exclude the possibility of two variables being independent is to assess whether the set of values, $\mathbb{X}_{12}$ (called the *support* of the joint pdf), for which

$$f_{X_1,X_2}(x_1, x_2) > 0$$

is identical to the Cartesian product of the two sets, $\mathbb{X}_1$ and $\mathbb{X}_2$ for which the marginal densities satisfy $f_{X_1}(x_1) > 0$ and $f_{X_2}(x_2) > 0$. That is, if

$$\mathbb{X}_{12} \neq \mathbb{X}_1 \times \mathbb{X}_2$$

then X_1 and X_2 are **not** independent.



We can find points where

- $f_{X_1}(x_1) > 0$ and $f_{X_2}(x_2) > 0$, but
- $f_{X_1,X_2}(x_1, x_2) = 0$

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

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

for **all** $(x_1, x_2) \in \mathbb{R}^2$.