

# Distribution Function of a Bivariate Random Variable

The cumulative distribution function of  $(X, Y)$  is defined by:

$$F_{(X,Y)}(x, y) = \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) = \mathbb{P}(X \leq x, Y \leq y).$$

This *bivariate distribution function* has properties analogous to those of the univariate distribution function. The most important of these is the following:

$$\begin{aligned} & \mathbb{P}(a < X \leq b, c < Y \leq d) \\ &= F_{(X,Y)}(b, d) - F_{(X,Y)}(a, d) - F_{(X,Y)}(b, c) + F_{(X,Y)}(a, c) \end{aligned}$$

which we will derive by drawing a picture of course.

From the definition of the distribution function:

$$F_{(X,Y)}(x, \infty) = F_X(x) \quad \text{and} \quad F_{(X,Y)}(\infty, y) = F_Y(y)$$

so that the distributions of  $X$  and  $Y$  can be obtained from the distribution of  $(X, Y)$ .

In general the converse is not true, because the distributions of  $X$  and  $Y$  are not sufficient to specify the distribution of  $(X, Y)$ . We also need to know something about the relationship between  $X$  and  $Y$ .

However in principle we can imagine choosing a random point by choosing its  $X$  and  $Y$  coordinates completely 'independently'. Then the individual distributions of  $X$  and  $Y$  *should* determine the joint distribution.

If  $X$  and  $Y$  are 'independent' then the events  $\{X \leq x\}$  and  $\{Y \leq y\}$  should also be 'independent' for all  $x$  and  $y$  and so

$$\begin{aligned}F_{(X,Y)}(x,y) &= \mathbb{P}(X \leq x, Y \leq y) \\ &= \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) \\ &= F_X(x)F_Y(y).\end{aligned}$$

In this special case the two univariate distributions do determine the joint distribution. Factorisation of the bivariate distribution function in this way can be taken as one definition of the independence of random variables, which we will define more carefully later.

## Joint and marginal pmf's

If  $S_{(X,Y)}$  is countable then we refer to  $(X, Y)$  as *bivariate discrete random variable* and define the *joint pmf* in the obvious way:

$$\begin{aligned} p_{(X,Y)}(x, y) &= \mathbb{P}(\{X = x\} \cap \{Y = y\}) \\ &= \mathbb{P}(X = x, Y = y). \end{aligned}$$

The pmf is such that

- $p_{(X,Y)}(x,y) \geq 0$ ,
- $\sum_x \sum_y p_{(X,Y)}(x,y) = 1$ .

Just as we did for 'univariate' or one dimensional discrete random variables, we can again interpret this *pmf* as assigning discrete probability masses to particular points in the plane.

## Marginal pmf's

The formula for the marginal distributions follows from the observation that the events  $\{Y = y\}$  are disjoint and exhaustive as  $y$  ranges over the set  $S_Y = \{y : (x, y) \in S_{(X, Y)}\}$ . Then,

$$\mathbb{P}(X = x) = \sum_{y \in S_Y} \mathbb{P}(X = x, Y = y)$$

that is

$$p_X(x) = \sum_{y \in S_Y} p_{(X, Y)}(x, y)$$

and similarly

$$p_Y(y) = \sum_{x \in S_X} p_{(X, Y)}(x, y).$$

## Coin example

If in the coin tossing example the coin is fair then the pmf is given by

$p_{(X,Y)}(x,y)$	$x = 0$	$x = 1$	$x = 2$
$y = 0$	$\frac{1}{8}$	$\frac{1}{8}$	0
$y = 1$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
$y = 2$	0	$\frac{1}{8}$	$\frac{1}{8}$

We can also find the marginal pmf's for  $X$  and  $Y$  from here.

## CDF for discrete rv's

For discrete  $(X, Y)$ ,

$$\begin{aligned}F_{(X,Y)}(x, y) &= \mathbb{P}(X \leq x, Y \leq y) \\&= \sum_{u \leq x \text{ and } v \leq y} \mathbb{P}(X = u, Y = v) \\&= \sum_{u \leq x \text{ and } v \leq y} p_{(X,Y)}(u, v).\end{aligned}$$

However, obtaining the pmf from cdf for a pair of rv's requires more care!



## Example

Suppose the cumulative distribution function of the pair of rv's  $(X, Y)$  (each with values  $0, 1, 2, \dots$ ) is given by the table:

$y \backslash x$	1	2	3	4
1	0	0.1	0.2	0.3
2	0.1	0.2	0.35	0.6
3	0.2	0.3	0.6	0.95
4	0.25	0.35	0.65	1

- 1 Find  $\mathbb{P}(X = 4, Y = 3)$ .
- 2 Find the complete probability mass function.

## Pmf from cdf

For a pair of rv's  $(X, Y)$  with values in  $0, 1, 2, \dots$ ,

$$\begin{aligned} & \mathbb{P}(X = x, Y = y) \\ &= \mathbb{P}(X = x, Y \leq y) - \mathbb{P}(X = x, Y \leq y - 1) \\ &= \mathbb{P}(X \leq x, Y \leq y) - \mathbb{P}(X \leq x - 1, Y \leq y) \\ &\quad - \mathbb{P}(X \leq x, Y \leq y - 1) + \mathbb{P}(X \leq x - 1, Y \leq y - 1) \\ &= F_{(X,Y)}(x, y) - F_{(X,Y)}(x - 1, y) \\ &\quad - F_{(X,Y)}(x, y - 1) + F_{(X,Y)}(x - 1, y - 1). \end{aligned}$$

pmf

$y \backslash x$	1	2	3	4
1	0	0.1	0.1	0.1
2	0.1	0	0.05	0.15
3	0.1	0	0.15	0.1
4	0.05	0	0	0

## Probabilities from cdf - in general

For a pair of rv's  $(X, Y)$  and numbers  $a, b, c, d$ :

$$\begin{aligned} & \mathbb{P}(\{a < X \leq b\} \cap \{c < Y \leq d\}) \\ &= \mathbb{P}(\{a < X \leq b\} \cap \{Y \leq d\}) - \mathbb{P}(\{a < X \leq b\} \cap \{Y \leq c\}) \\ &= \mathbb{P}(\{X \leq b\} \cap \{Y \leq d\}) - \mathbb{P}(\{X \leq a\} \cap \{Y \leq d\}) \\ &\quad - \mathbb{P}(\{X \leq b\} \cap \{Y \leq c\}) + \mathbb{P}(\{X \leq a\} \cap \{Y \leq c\}) \\ &= F_{(X,Y)}(b, d) - F_{(X,Y)}(a, d) \\ &\quad - F_{(X,Y)}(b, c) + F_{(X,Y)}(a, c). \end{aligned}$$

## Example

Suppose Jeff has 4 quokkas which are ranked 1 to 4 in cuteness. If a random sample of size 2 is taken, find the pmf and cdf for the pair of random variables  $(X, Y)$  if  $X$  is the sum of ranks in the sample and  $Y$  is the maximum of ranks in the sample.



## Joint and marginal pdfs

What is the analogy for cts rv's?

A pair of rv's  $(X, Y)$  is said to have a continuous distribution if the cdf  $F_{(X,Y)}$  can be written as an integral of a two variable, non-negative function  $f_{(X,Y)}$ , called the probability density function (pdf):

$$\begin{aligned} F_{(X,Y)}(x, y) &= \mathbb{P}(\{X \leq x\} \cap \{Y \leq y\}) \\ &= \iint_{\{(u,v): u \leq x \text{ and } v \leq y\}} f_{(X,Y)}(u, v) du dv. \end{aligned}$$

## Double integrals

Let  $f(u, v) \geq 0$  and

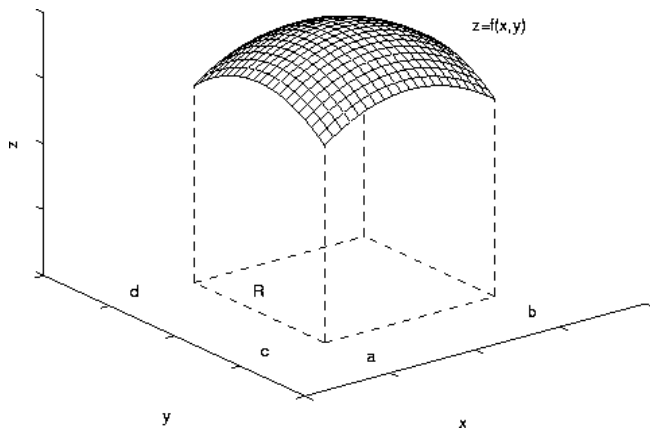
$$V = \int_a^b \int_c^d f(u, v) dv du = \lim \sum_i f(u_i, v_i) \Delta_i$$

where the limit is over all regular partitions of the rectangle  $(a, b) \times (c, d)$  into the small areas  $\Delta_i$ . If we interpret  $f(u, v)$  as a surface above the  $xy$  plane then  $V$  can be interpreted as the volume below that surface.



## Double integrals

Conceptually, the double integral is the volume under the surface given by  $f(x,y)$  in the region defined by  $a, b, c, d$ .



## Double integrals

The double integrals can be evaluated by treating them as two iterated single integrals:

$$\int_a^b \int_c^d f(u, v) dv du = \int_a^b \left( \int_c^d f(u, v) dv \right) du.$$

In the inner integral here we think of  $u$  as fixed. Consider intersecting the hill by the plane through this fixed  $u$  - the resulting cross-sectional area is given by the inner integral. So the whole iterated integral just amounts to saying that the total volume is the sum of cross-sectional 'slices' parallel to the  $x$  axis of volume  $\left( \int_c^d f(u, v) dv \right) du$ . You can reverse the order to get the sum of cross-sectional slices parallel to the  $y$  axis.

## pdf's

- Just as the pdf for a single rv must be non-negative and have 1 as total area under the curve it defines, the pdf for a pair of rv's must be non-negative and have 1 as total volume under the surface that it defines.
- If such a function exists, then it is unique.
- If a *pdf* exists then for 'almost all'  $(x, y)$  values we also have

$$\frac{\partial^2}{\partial x \partial y} F_{(X, Y)}(x, y) = f_{(X, Y)}(x, y).$$

To understand the connection between the distribution function and pdf consider:

$$\begin{aligned}f_{(X,Y)}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y) \\&\approx \frac{\partial}{\partial x} \left( \frac{1}{h} \left( F(x, y + \frac{h}{2}) - F(x, y - \frac{h}{2}) \right) \right) \\&\approx \frac{1}{h} \left( \frac{F(x + \frac{h}{2}, y + \frac{h}{2}) - F(x + \frac{h}{2}, y - \frac{h}{2})}{h} \right. \\&\quad \left. - \frac{F(x - \frac{h}{2}, y + \frac{h}{2}) - F(x - \frac{h}{2}, y - \frac{h}{2})}{h} \right) \\&\approx \frac{1}{h^2} \mathbb{P}(x - \frac{h}{2} < X \leq x + \frac{h}{2}, y - \frac{h}{2} < Y \leq y + \frac{h}{2}).\end{aligned}$$

So

$$\mathbb{P}((X, Y) \in \text{small rectangle area } h^2 \text{ near}(x, y)) \approx h^2 f_{(X, Y)}(x, y)$$

justifying our interpretation of the pdf as measuring the probability density around the point  $(x, y)$ .

For continuous univariate random variables, probabilities are assigned to intervals using the area under the pdf curve  $f_X(x)$ . For continuous bivariate random variables probabilities are assigned to regions in the plane using volumes under the pdf 'hill' or surface  $f_{(X, Y)}(x, y)$ .

The probability that the random point  $(X, Y)$  lies in any rectangle in the plane is given by the volume below the two dimensional surface represented by  $f_{(X,Y)}(x, y)$  over that rectangle:

$$\mathbb{P}(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f_{(X,Y)}(x, y) dy dx$$

As we did with discrete bivariate random variables, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy$$

and similarly

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dx.$$

## pdf example

Suppose the pdf of the bivariate rv  $(X, Y)$  is

$$f_{(X,Y)}(x,y) = \begin{cases} cxy, & \text{if } 0 < x < 1, 0 < y < 1, 0 < x + y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find  $c$  and  $\mathbb{P}(\{X \leq 0.5\} \cap \{Y \leq 0.7\})$ .



## A surprise?

- 1  $f_{(X,Y)}(x,y) = 2x + 2y - 4xy$  for  $0 < x < 1$ ,  $0 < y < 1$ , find the marginal pdf's.
- 2  $f_{(X,Y)}(x,y) = 2 - 2x - 2y + 4xy$  for  $0 < x < 1$ ,  $0 < y < 1$ , find the marginal pdf's.
- 3 What can we see from here?

## Conditional pmfs

It seems natural to ask about the distribution of one component of our bivariate distribution (say  $X$ ) when we have information about the value of the other component ( $Y$ ).

By analogy with the way that we defined conditional probability for events, this is called a *conditional distribution* and for discrete random variables the expression for the conditional distribution follows directly from the conditional probability formula.

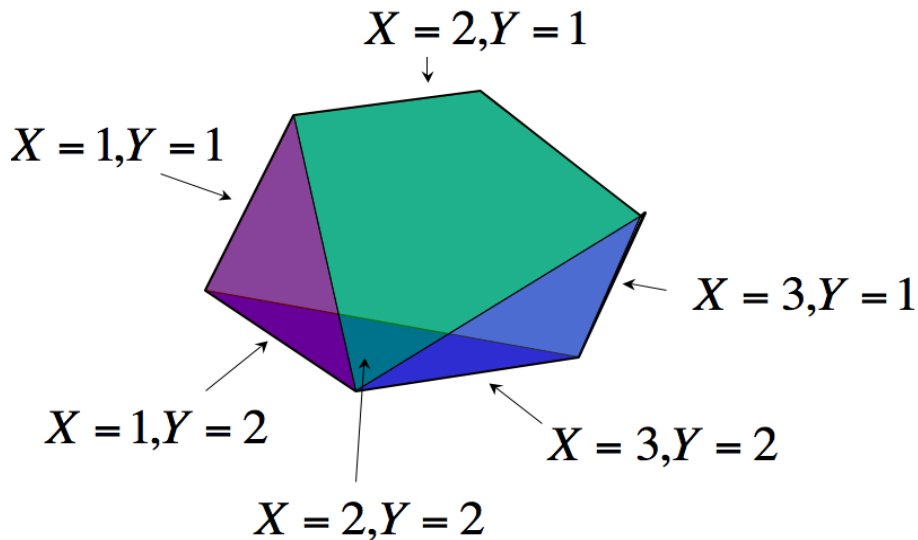
We define the conditional distribution of  $X$  given  $Y$  as:

$$\begin{aligned} p_{X|Y}(x|y) &= \mathbb{P}(X = x | Y = y) \\ &= \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \\ &= \frac{p_{(X,Y)}(x,y)}{p_Y(y)} \end{aligned}$$

Note in this formula we think of  $x$  as a variable and of  $y$  as fixed with  $y \in S_Y$  (as previously defined) and  $p_Y(y) \neq 0$ .

**Remark** The role of  $y$  is to refer to the corresponding partition set.

## Conceptually



## Conditional pdfs

The pdf of  $X$  *conditional on*  $Y = y$  is given by

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}.$$

This can be derived by letting  $\delta y \rightarrow 0$  in the equation

$$\mathbb{P}(X \approx x | Y \approx y) = \frac{\mathbb{P}(X \approx x, Y \approx y)}{\mathbb{P}(Y \approx y)}.$$

## Example

A particular type of rock is analysed. Let  $X$  and  $Y$  denote the proportions of minerals  $A$  and  $B$  respectively found in a sample of the rock. Assume that the pdf of  $(X, Y)$  is given by

$$f_{(X,Y)}(x, y) = 2 \quad (x + y \leq 1, x \geq 0, y \geq 0).$$

Find the marginal distributions, and the conditional distribution of  $X|Y$ .

Are  $X$  and  $Y$  independent?

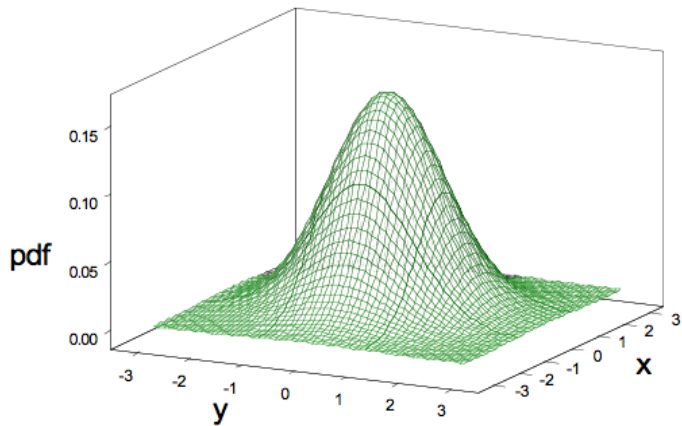
## Bivariate normal distribution

If the pdf of  $(X, Y)$  is given by

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

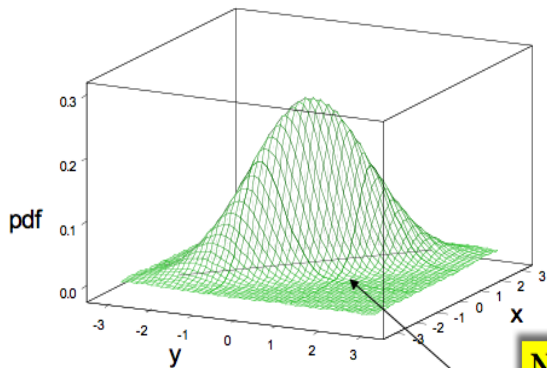
where  $\rho \in [-1, 1]$ , then we say that  $(X, Y)$  has the *standard bivariate normal distribution with parameter  $\rho$* , and we write  $(X, Y) \stackrel{d}{=} N_2(\rho)$ .

$$\rho = 0$$





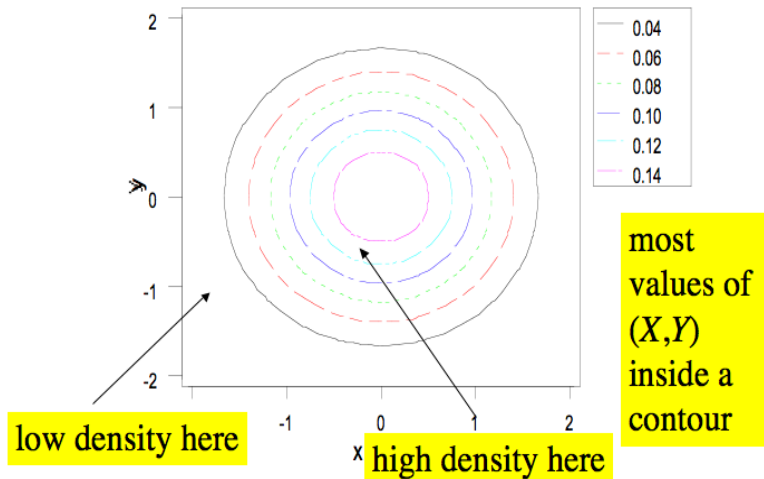
$$\rho = 0.8$$

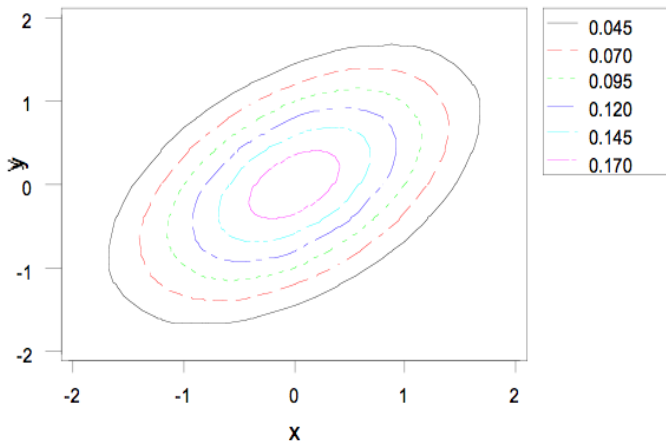


Note: pdf concentrated around line

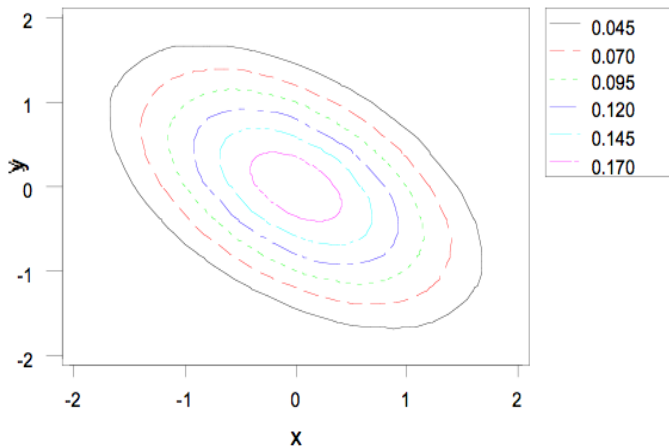
□

$$\rho = 0$$

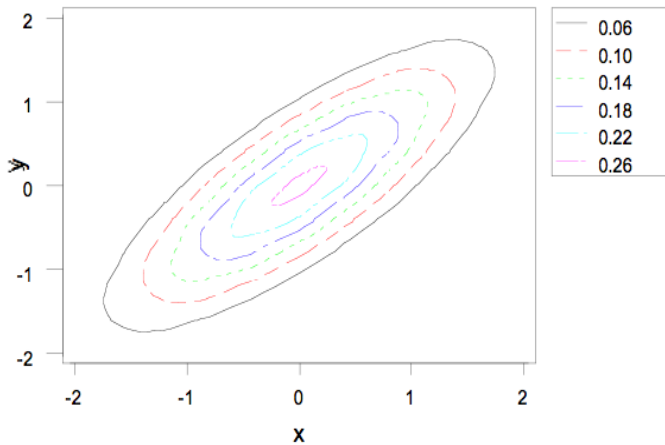




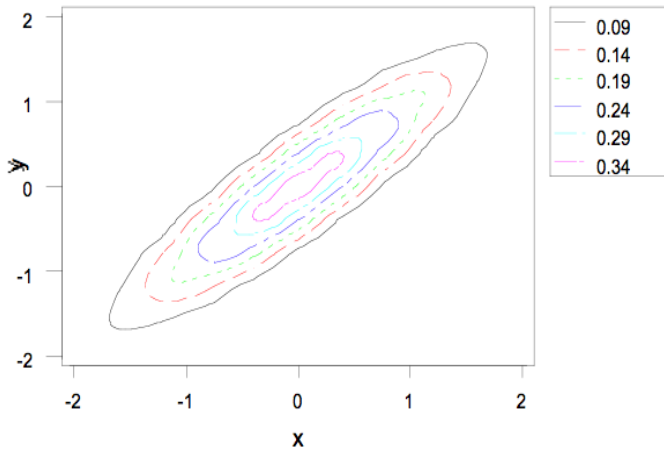
$\rho = 0.5$



$\rho = -0.5$



$\rho = 0.8$



$\rho = 0.9$

The contours of  $f_{(X,Y)}(x,y)$  are ellipses with axes inclined at an angle of  $\pi/4$  to the  $x$ - and  $y$ -axes.

If  $\rho > 0$  then the major axis lies along  $y = x$  and the minor axis lies along  $y = -x$ . This means that  $X$  and  $Y$  tend to be large together, and small together. In this case we say that  $X$  and  $Y$  are *positively-related*.

On the other hand, if  $\rho < 0$  then the axes are reversed and  $X$  tends to be large when  $Y$  is small and vice versa. In this case  $X$  and  $Y$  are said to be *negatively-related*.

The contours of  $f_{(X,Y)}(x,y)$  are ellipses with axes inclined at an angle of  $\pi/4$  to the  $x$ - and  $y$ -axes.

If  $\rho > 0$  then the major axis lies along  $y = x$  and the minor axis lies along  $y = -x$ . This means that  $X$  and  $Y$  tend to be large together, and small together. In this case we say that  $X$  and  $Y$  are *positively-related*.

On the other hand, if  $\rho < 0$  then the axes are reversed and  $X$  tends to be large when  $Y$  is small and vice versa. In this case  $X$  and  $Y$  are said to be *negatively-related*.



As  $|\rho|$  gets close to one, the ellipses become 'skinnier' in whatever orientation that they have and the relationship between  $X$  and  $Y$  becomes stronger.

We see that the parameter  $\rho$  tells us a good deal about the relationship between  $X$  and  $Y$ .

The pdf of  $Y$  is given by

$$\begin{aligned}f_Y(y) &= \int_{-\infty}^{\infty} f_{(X,Y)}(x,y) dx \\&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}\right) dx \\&= \dots = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\end{aligned}$$

Therefore  $Y \stackrel{d}{=} N(0, 1)$ . Similarly, we can show that  $X \stackrel{d}{=} N(0, 1)$ .

Note that  $X \stackrel{d}{=} N(0, 1)$  and  $Y \stackrel{d}{=} N(0, 1)$  does not imply that  $(X, Y) \stackrel{d}{=} N_2(\rho)$ .

For example if  $X \stackrel{d}{=} N(0, 1)$  and we define

$$Y = \begin{cases} X & \text{with probability } 1/2 \\ -X & \text{with probability } 1/2 \end{cases}$$

then  $Y \stackrel{d}{=} N(0, 1)$  also, but  $(X, Y)$  is not bivariate normal. Why?

The conditional pdf of  $X$  given  $Y = y$  is given by

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + \rho^2 y^2)\right) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x - \rho y)^2\right) \end{aligned}$$

hence  $(X|Y = y) \stackrel{d}{=} N(\rho y, 1 - \rho^2)$ . It follows that  $\mathbb{E}[X|Y = y] = \rho y$ .

## Example

Suppose that  $(X, Y) \stackrel{d}{=} N_2(\rho = 0.5)$ . Find  $\mathbb{P}(X > 1)$  and  $\mathbb{P}(X > 1 | Y = 1)$ .

### Solution

Well, what is the marginal distribution of  $X$ ?

We know that  $(X|Y = 1) \stackrel{d}{=} N(0.5, 1 - 0.5^2)$ .

Then we just standardise using techniques we already know.

We see that  $\mathbb{P}(X > 1|Y = 1)$  is greater than  $\mathbb{P}(X > 1)$ . Since  $\rho > 0$ ,  $X$  and  $Y$  are positively related. It follows that we expect  $Y$  to be larger (greater than its mean) when  $X$  is larger.

If

$$\left( \frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y} \right) \stackrel{d}{=} N_2(\rho),$$

then  $(X, Y)$  has a bivariate normal distribution with parameters  $\mu_X, \mu_Y, \sigma_X, \sigma_Y$  and  $\rho$ , and we write

$$(X, Y) \stackrel{d}{=} N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho).$$

It has density

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2(1-\rho^2)} \left( \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{s_x s_y} \right) \right]$$



An alternative notation is

$$\begin{bmatrix} X \\ Y \end{bmatrix} \stackrel{d}{=} N_2 \left( \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \right).$$

Results for the general bivariate normal distribution can be derived from those for the standard case (try this yourself)

$$X \stackrel{d}{=} N(\mu_X, \sigma_X^2)$$

$$Y \stackrel{d}{=} N(\mu_Y, \sigma_Y^2)$$

$$(X|Y = y) \stackrel{d}{=} N\left(\mu_X + \rho\sigma_X \frac{(y - \mu_Y)}{\sigma_Y}, \sigma_X^2(1 - \rho^2)\right).$$

## Example

Suppose that  $(X, Y) \stackrel{d}{=} N_2(50, 50; 10^2, 5^2; -0.7)$ . Find  $\mathbb{P}(X > 55)$  and  $\mathbb{P}(X > 55 | Y = 60)$ .

## Example

Suppose the study scores of students in the subjects of Kangaroo caring and Possum hunting can be considered as bivariate normal with  $\rho = 0.7$ ,  $\mu$ 's = 30 and  $\sigma$ 's = 7. What is the conditional probability that the student gets more than 30 in Kangaroo care given the students score in Possum hunting is (a) 30 (b) 35 (c) 40?

## Sol

Let  $X_s = \frac{X-30}{7}$  and  $Y_s = \frac{Y-30}{7}$ , then  $(Y_s|X_s = z) = N(0.7z, 1 - 0.7^2)$  and

$$\begin{aligned} & \mathbb{P}(Y > 30|X = x) \\ &= \mathbb{P}\left(Y_s > 0 \mid X_s = \frac{x-30}{7}\right) \\ &= \mathbb{P}\left(Z > \frac{-0.7 \cdot \frac{x-30}{7}}{\sqrt{1-0.7^2}}\right) = \mathbb{P}\left(Z < \frac{0.7 \cdot \frac{x-30}{7}}{\sqrt{1-0.7^2}}\right) \end{aligned}$$

- $\mathbb{P}(Y > 30|X = 30) = \frac{1}{2}$
- $\mathbb{P}(Y > 30|X = 35) = 0.758080$
- $\mathbb{P}(Y > 30|X = 40) = 0.919285$

## Independence of random variables

We previously had the notion of independence of events. We now translate this idea to independence of random variables:

Two random variables  $X$  and  $Y$  are *independent* if

$$\mathbb{P}(X \in M, Y \in N) = \mathbb{P}(X \in M)\mathbb{P}(Y \in N)$$

for any (appropriate) subsets  $M, N \in \mathbb{R}$ .

## Independence by cdf

$X$ ,  $Y$  independent rv's if and only if

$$F_{(X,Y)}(x,y) = F_X(x)F_Y(y) \text{ for all } x \text{ and } y.$$

## For discrete rv's

Often more convenient to think of pmf.

Rv's are independent if and only if the joint pmf is the product of the marginals:

$X, Y$  independent



$$\left\{ \begin{array}{l} p_{(X,Y)}(x,y) = p_X(x)p_Y(y) \quad \forall x,y \\ p_X(x) = p_{X|Y}(x|Y=y), p_Y(y) = p_{Y|X}(y|X=x) \quad \forall x,y \end{array} \right\}$$



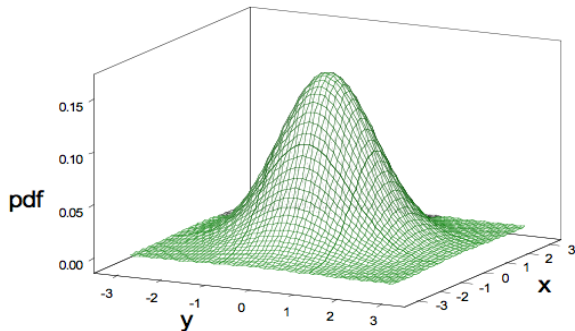
## For continuous rv's

Often more convenient to think of pdf.

Rv's are independent if and only if the pdf is the product of the marginals:

$$\begin{array}{c} X, Y \text{ independent} \\ \Updownarrow \\ \left\{ \begin{array}{l} f_{(X,Y)}(x,y) = f_X(x)f_Y(y) \quad \forall x,y \\ f_X(x) = f_{X|Y}(x|Y=y), f_Y(y) = f_{Y|X}(y|X=x) \quad \forall x,y \end{array} \right\} \end{array}$$

## Pair of indept $N(0,1)$



## Pdf example

Suppose the pdf for the bivariate rv  $(X, Y)$  is

$$f_{(X,Y)}(x,y) = \begin{cases} c, & \text{if } 0 < x < 1, 0 < y < 1, 0 < x + y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Are  $X$  and  $Y$  independent?

## Symmetry

If the density is symmetric in  $x$  and  $y$  (ie the value does not change if  $x$  and  $y$  are swapped), the marginal pdf's are the same:  $X$  and  $Y$  are identically distributed.

## Summary: three criteria for independence

Rv's are independent if and only if

- cdf is product of marginals
- pmf or pdf is product of marginals
- conditional pmf or pdf is same as marginal

## Example

Suppose that  $(X, Y) \stackrel{d}{=} N_2(0)$ , that is,  $X$  and  $Y$  are independent standard normal. Then  $(X, Y)$  defines a point in  $\mathbb{R}^2$ . What is the distribution of  $(R, \Theta)$ , the polar coordinates of this point?

$$\mathbb{P}(X \approx x, Y \approx y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \delta x \delta y.$$

In polar coordinates,  $r = \sqrt{x^2 + y^2}$  and so, for  $(r, \theta) \in [0, \infty) \times [0, 2\pi)$ , we have

$$\mathbb{P}(R \approx r, \Theta \approx \theta) = \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r \delta r \delta \theta.$$

Hence, the pdf of  $(R, \Theta)$  is given by

$$f_{(R, \Theta)}(r, \theta) = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2} \quad (r > 0, 0 < \theta < 2\pi).$$

It follows that

$$f_R(r) = re^{-\frac{1}{2}r^2} \quad (r > 0) \quad \text{and} \quad f_\Theta(\theta) = \frac{1}{2\pi} \quad (0 < \theta < 2\pi).$$

Why?

Since  $f_{(R,\Theta)}(r, \theta) = f_R(r)f_\Theta(\theta)$ , the random variables  $R$  and  $\Theta$  are independent.



The transformation described in the above example can be used to derive a proof of the result that

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi},$$

which we have used to derive the constant  $1/(\sqrt{2\pi})$  in the formula for the pdf of the normal distribution.

Assume that

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx = c.$$

Then

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy = c,$$

and, multiplying them together,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx dy = c^2.$$

Now, transforming to polar coordinates,

$$\int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}r^2} r dr d\theta = c^2.$$

This is equivalent to

$$2\pi \int_0^{\infty} e^{-\frac{1}{2}r^2} r dr = c^2$$

or

$$2\pi \left[ -e^{-\frac{1}{2}r^2} \right]_0^{\infty} = c^2.$$

This implies  $c = \sqrt{2\pi}$ .

## The pull back trick

**The pull back trick:**  $\mathbb{E}(Z) = \sum_{\text{all } \omega} Z(\omega)\mathbb{P}(\omega)$

Hence, for  $\psi : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\mathbb{E}(\psi(Z)) = \begin{cases} \sum_{z \in \mathcal{S}_Z} \psi(z) p_Z(z) & Z \text{ discrete} \\ \int_{-\infty}^{\infty} \psi(z) f_Z(z) dz & Z \text{ continuous} \end{cases}$$

# Expectation of function

## Theorem

If  $(X, Y)$  is a discrete bivariate random variable with set of possible values  $S_{(X,Y)}$  and probability mass function  $p_{(X,Y)}(x, y)$ , then, for any real-valued function  $\psi$ ,

$$\mathbb{E}[\psi(X, Y)] = \sum_{(x,y) \in S_{(X,Y)}} \psi(x, y) p_{(X,Y)}(x, y)$$

provided the sum converges absolutely.