HIROSHIMA UNIVERSITY

# Fundamental Data Science (30104001)

## Lecture 12 — Bivariate probability distributions

Jorge N. Tendeiro

Hiroshima University

# Review

> *Consider a fair coin (i.e., with 50% chance of landing heads up).*
> *Throw the coin twice. Count the number of heads.*

Let $X =$ *number of heads*.

Then we may assume that

$$X \sim B\left(n = 2, p = \frac{1}{2}\right),$$

where

- $n =$ number of trials
- $p =$ probability of a 'success' (in this case, flipping heads).

# Today

Today we will learn how to consider two random variables simultaneously!

For example:

> *Consider a fair coin (i.e., with 50% chance of landing heads up).*
> *Throw the coin twice. Count the number of heads and the number of tails.*

We can set two random variables:

- $X =$ *number of heads.*
- $Y =$ *number of tails.*

And we can entertain probabilities of events occurring at the same time, for example:

- $P(X = 1, Y = 2)$
- $P(X = 0, Y = 1)$

# Today

Why consider two random variables simultaneously?

Because...

- We want to be able to find the probability of two phenomena occurring at the same time, like $P(X = 1, Y = 2)$ and $P(X = 0, Y = 1)$.

  We will learn about joint probability distributions for this.

- We want to learn about the relationship between two phenomena.
  For example, when $X$ increases, what can we expect about $Y$ (increases, decreases, or stays the same)?

  To answer such questions, we will learn about the covariance and correlation coefficients for pairs of random variables.

# Joint probability distribution

# Joint probability distribution (discrete)

> *Consider a fair coin (i.e., with 50% chance of landing heads up).*
> *Throw the coin twice. Count the number of heads and the number of tails.*

Set two random, and in this case discrete, variables:

- $X =$ *number of heads.*
- $Y =$ *number of tails.*

**Joint probability distribution:**

> *Probability distribution which takes two random variables into account at the same time.*

When $X$ and $Y$ are discrete variables we use the notation

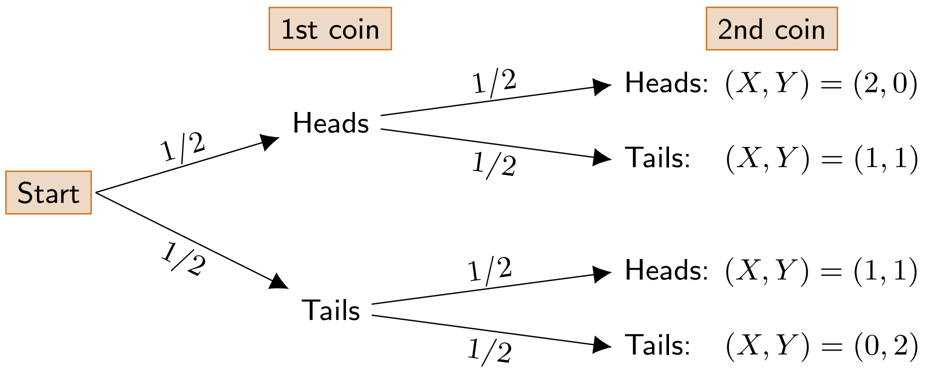
$$P(X = k, Y = l)$$

to denote the joint probability of $X$ being equal to $k$ and $Y$ being equal to $l$.

# Joint probability distribution (discrete)

- $X$ = *number of heads*
- $Y$ = *number of tails*

| $X \backslash Y$ | 0 | 1 | 2 | Total |
|---|---|---|---|---|
| 0 | 0 | 0 | $^1/_4$ | |
| 1 | 0 | $^2/_4$ | 0 | |
| 2 | $^1/_4$ | 0 | 0 | |
| Total | | | | 1 |



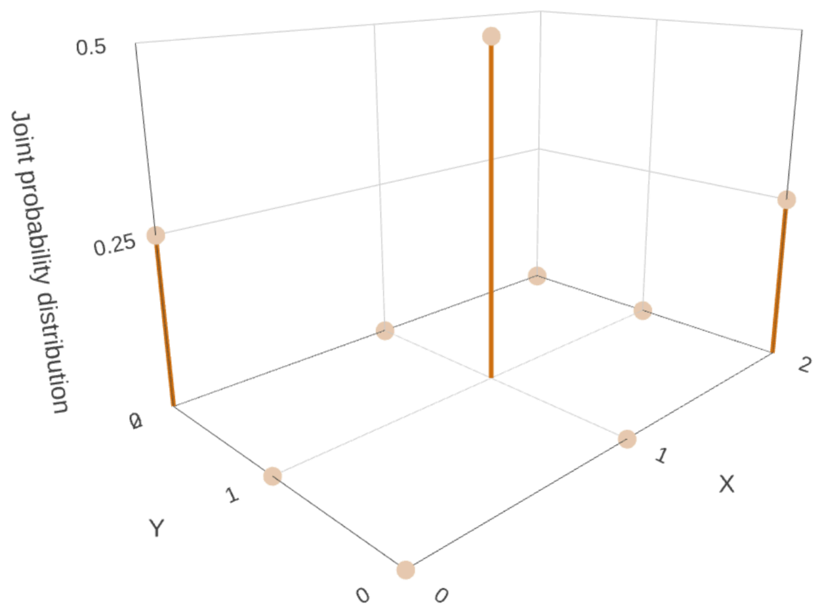

Joint probabilities = probabilities of combinations of values of $X$ and $Y$:

$$P(X = k, Y = l)$$

The joint probabilities add up to $1$.

# Joint probability distribution (discrete)

Plot of the joint probability distribution of variables $X =$ *number of heads* and $Y =$ *number of tails*:




Notice that:

- To plot two random variables we need a 3-dimensional plot.
- The joint probability distribution of discrete random variables is represented with points.

# Marginal probability distribution (discrete)

| $X \backslash Y$ | 0 | 1 | 2 | Total |
|---|---|---|---|---|
| 0 | 0 | 0 | $1/4$ | $1/4$ |
| 1 | 0 | $2/4$ | 0 | $2/4$ |
| 2 | $1/4$ | 0 | 0 | $1/4$ |
| Total | | | | 1 |

The rightmost column is the marginal probability distribution of $X$.
Each row's marginal probability is equal to the sum of the probabilities of the corresponding row.

For example:

$$
\begin{aligned}
P(X=1) &= P(X=1, Y=0) + P(X=1, Y=1) + P(X=1, Y=2) \\
&= 0 + \frac{2}{4} + 0 = \frac{2}{4}
\end{aligned}
$$

# Marginal probability distribution (discrete)

| $X \backslash Y$ | 0 | 1 | 2 | Total |
|---|---|---|---|---|
| 0 | 0 | 0 | $^1/_4$ | |
| 1 | 0 | $^2/_4$ | 0 | |
| 2 | $^1/_4$ | 0 | 0 | |
| Total | $^1/_4$ | $^2/_4$ | $^1/_4$ | 1 |

The bottom row is the marginal probability distribution of $Y$.
Each column's marginal probability is equal to the sum of the probabilities of the corresponding column.

For example:

$$\begin{aligned} P(Y=2) &= P(X=0, Y=2) + P(X=1, Y=2) + P(X=2, Y=2) \\ &= \frac{1}{4} + 0 + 0 = \frac{1}{4} \end{aligned}$$

# Summary (discrete random variables)

Let $X$ and $Y$ denote two discrete random variables.

Joint probability distribution of $X$ and $Y$:

> *A distribution providing probabilities for pairs of values of $X$ and $Y$:*
>
> $$f_{X,Y}(k,l) = P(X = k, Y = l), \qquad \text{for } k, l = 0, 1, 2, \ldots$$
>
> $f_{X,Y}$ is called the joint probability mass function of $X$ and $Y$.

Marginal probability distribution of $X$ (and similarly for $Y$):

> *A distribution providing probabilities for the values of $X$, across all values of $Y$:*
>
> $$f_X(k) = P(X = k) = f_{X,Y}(k,0) + f_{X,Y}(k,1) + \cdots$$
>
> *$f_X$ is called the marginal probability function of $X$.*

# Exercise (1)

Consider a fair coin, that is, a coin with a $^1/_2$ chance of landing heads up.
You toss the coin twice.

Here, we define random variables $X$ and $Y$ as follows:

$$X = \begin{cases} 1, & \text{if 1st throw is heads} \\ 0, & \text{if 1st throw is tails} \end{cases} \qquad Y = \begin{cases} 1, & \text{if 2nd throw is heads} \\ 0, & \text{if 2nd throw is tails} \end{cases}.$$

Find the joint probability distribution of $X$ and $Y$, as well as both marginal probability distributions of $X$ and of $Y$.

| $X \backslash Y$ | 0 | 1 | Total |
|---|---|---|---|
| 0 | | | |
| 1 | | | |
| Total | | | 1 |

# Exercise (1) — ANSWER

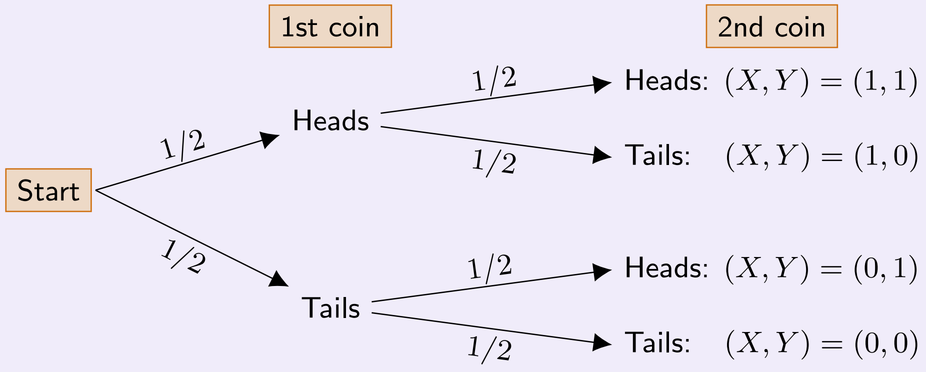
Consider a fair coin, that is, a coin with a $^1/_2$ chance of landing heads up.
You toss the coin twice.

Here, we define random variables $X$ and $Y$ as follows:

$$X = \begin{cases} 1, & \text{if 1st throw is heads} \\ 0, & \text{if 1st throw is tails} \end{cases} \qquad Y = \begin{cases} 1, & \text{if 2nd throw is heads} \\ 0, & \text{if 2nd throw is tails} \end{cases}.$$

Find the joint probability distribution of $X$ and $Y$, as well as both marginal probability distributions of $X$ and of $Y$.

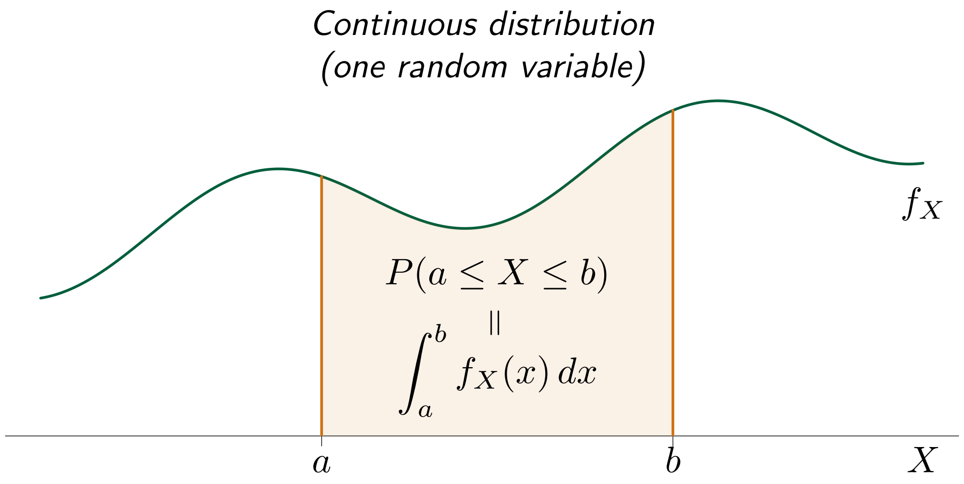
| $X$ \ $Y$ | 0 | 1 | Total |
|---|---|---|---|
| 0 | ¼ | ¼ | ½ |
| 1 | ¼ | ¼ | ½ |
| Total | ½ | ½ | 1 |



# Joint probability distribution (continuous)

Examples of continuous random variables:

- Random variable $X$ = height in cm.
- Random variable $Y$ = weight in kg.




Joint probability distribution of $X$ and $Y$:

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

$f_{X,Y}$ is called the joint probability density function of $X$ and $Y$.

# Joint probability distribution (continuous)

$f_{X,Y}$ represents a two-dimensional surface.

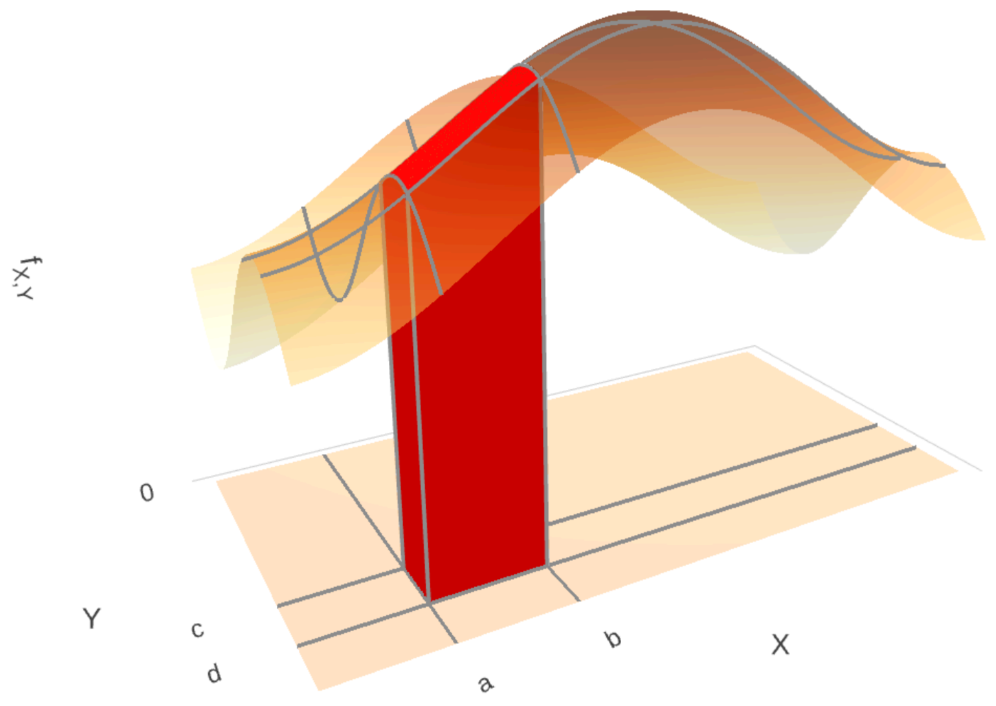
Now, instead of

probability = *area* under the *curve*
(for one variable)

we have that:

probability = volume under the surface
(two variables).

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



# Covariance, correlation

# Relationships between $X$ and $Y$

> *Throw a fair coin twice.*

Consider the following two random variables:

- $X =$ number of coins landing heads up.
- $Y =$ number of coins landing tails up.

Clearly,

> *The more coins landing heads up, the less coins landing tails up.*

In other words,

> *The bigger the value of $X$, the smaller the value of $Y$.*

It's important to study the relationship between $X$ and $Y$ via the joint probability distribution.
In particular, we should consider the covariance and correlation of a pair of random variables.

# Covariance

The covariance of random variables $X$ and $Y$ is

$$\mathrm{Cov}[X, Y] = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big].$$

Using this formula requires knowing the joint probability distribution of $X$ and $Y$.

This formula looks like the sample covariance between variables $x$ and $y$ for data (recall from Lecture 06):

$$s_{xy} = \text{mean of } (x_i - \overline{x})(y_i - \overline{y}).$$

But there are some **differences** between the two concepts:

| | $s_{xy}$ | Cov$[X, Y]$ |
|---|---|---|
| *Applies to...* | Observed data | Random variables |
| *Based on...* | Sample mean | Expected value |
| *Describes...* | Characteristics of the data | Characteristics of a joint probability distribution modeling a probabilistic phenomenon |

# Correlation

The correlation coefficient of random variables $X$ and $Y$ is

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{V[X]}\sqrt{V[Y]}}.$$

Similarly, this looks like the sample correlation between variables $x$ and $y$ for data (recall from Lecture 06):

$$r_{xy} = \frac{s_{xy}}{s_x s_y}.$$

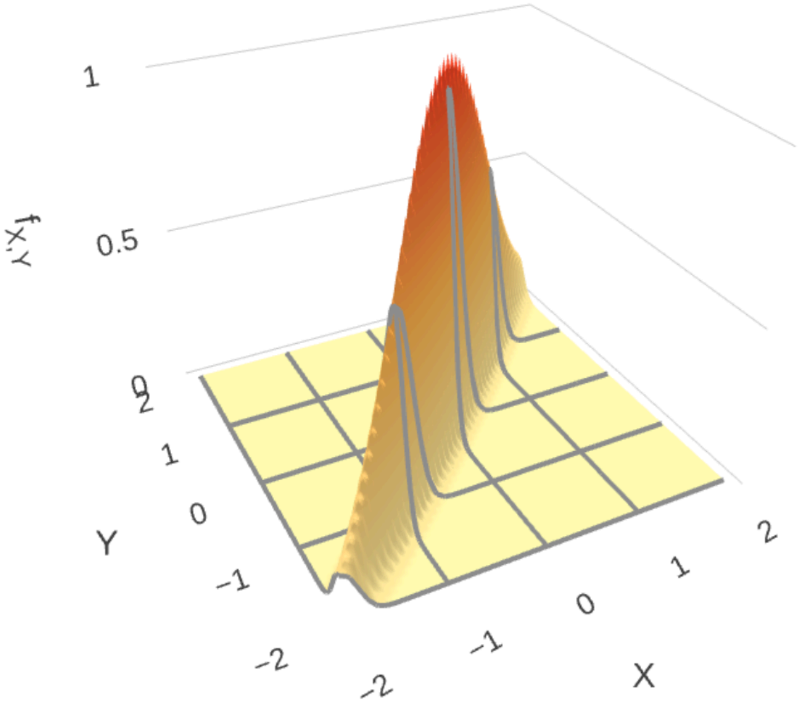
**Characteristics of $\rho_{X,Y}$:**

- $-1 \le \rho_{X,Y} \le 1$
- $\rho_{X,Y}$ indicates the strength of the linear relationship between $X$ and $Y$:
  - $\rho_{X,Y} > 0 \longrightarrow$ positive correlation
  - $\rho_{X,Y} < 0 \longrightarrow$ negative correlation
  - $\rho_{X,Y} = 0 \longrightarrow$ no correlation
- When $V[X] = 0$ or $V[Y] = 0$ we define $\rho_{X,Y} = 0$.

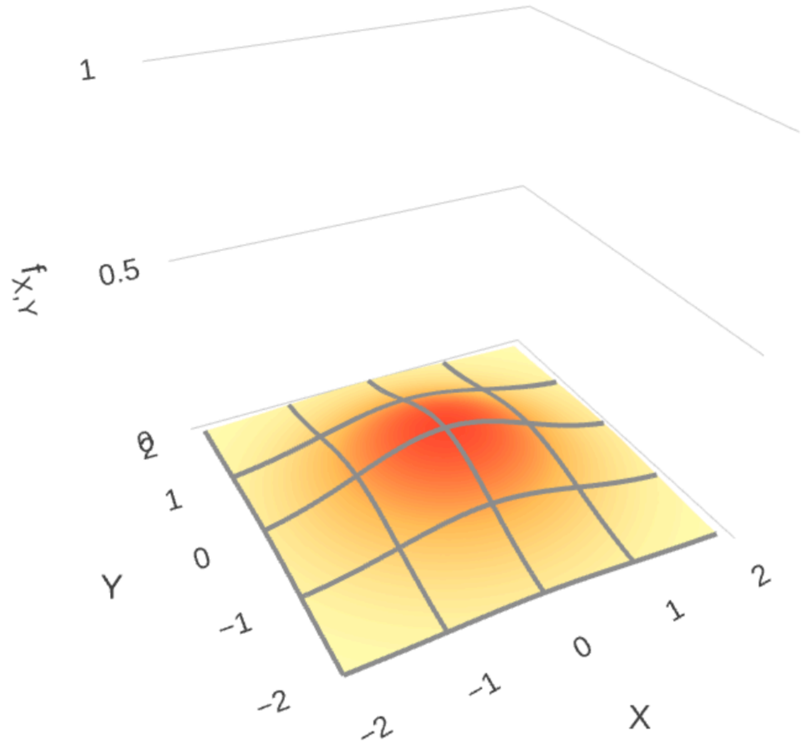
| $\rho_{X,Y}$ | **Interpretation** |
|---|---|
| $-1 \le \rho_{X,Y} \le -0.7$ | Strong negative correlation |
| $-0.7 \le \rho_{X,Y} \le -0.4$ | Somewhat negative correlation |
| $-0.4 \le \rho_{X,Y} \le -0.2$ | Weak negative correlation |
| $-0.2 \le \rho_{X,Y} \le 0.2$ | Almost uncorrelated |
| $0.2 \le \rho_{X,Y} \le 0.4$ | Weak positive correlation |
| $0.4 \le \rho_{X,Y} \le 0.7$ | Somewhat positive correlation |
| $0.7 \le \rho_{X,Y} \le 1$ | Strong positive correlation |

# Visualizing correlation coefficients

$\boxed{\rho_{X,Y} = 0.99}$

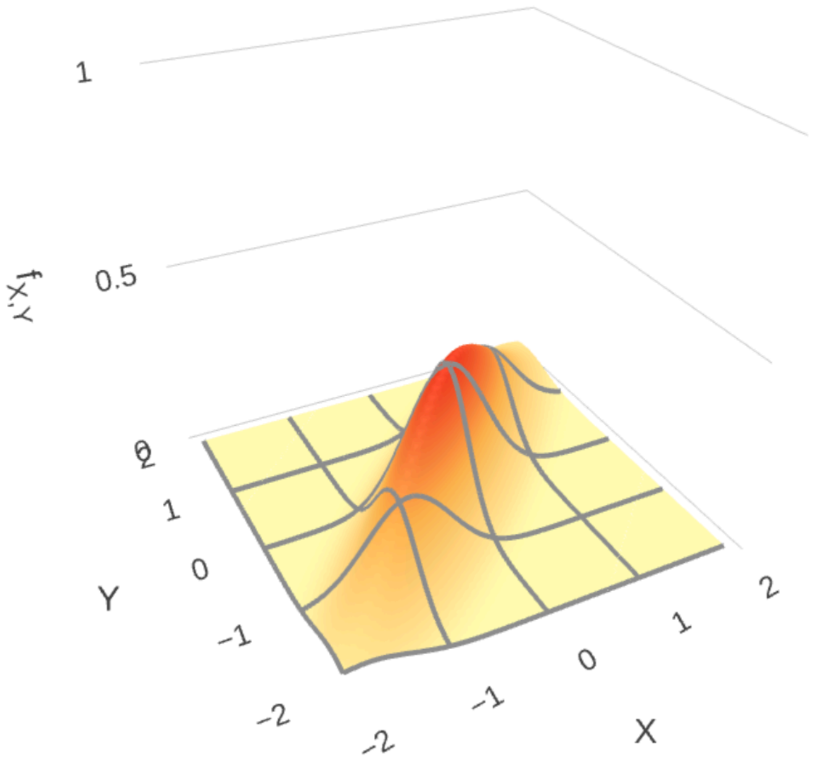



$\boxed{\rho_{X,Y} = 0}$

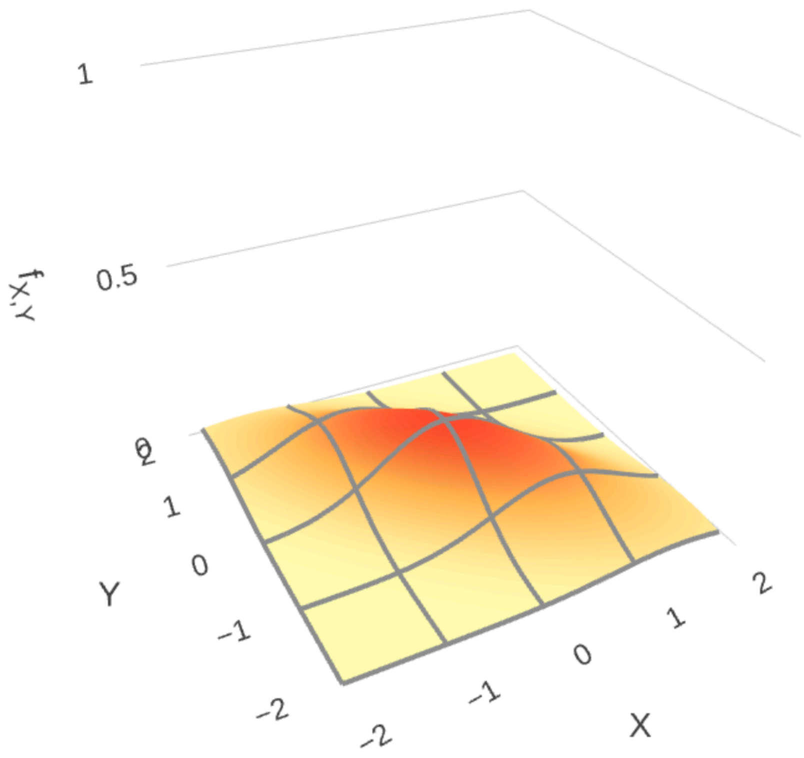


# Visualizing correlation coefficients

$\boxed{\rho_{X,Y} = 0.9}$




$\boxed{\rho_{X,Y} = -0.9}$



# Example: Tossing coins

> *Throw a fair coin twice.*

Consider the following two random variables:

- $X$ = number of coins landing heads up.
- $Y$ = number of coins landing tails up.

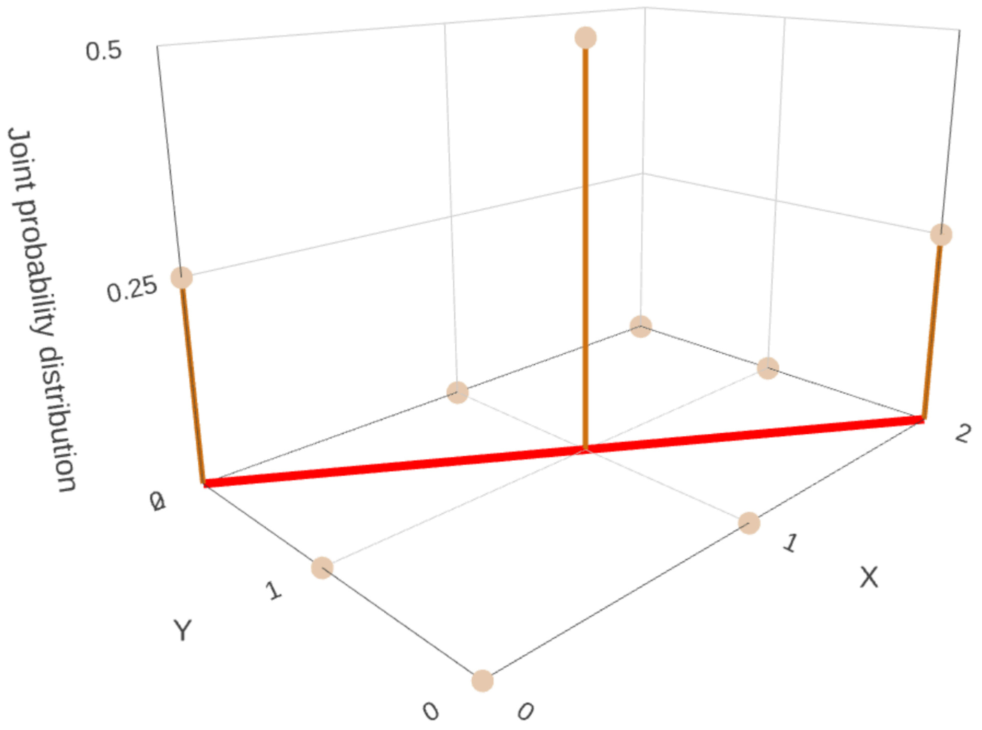
| $X \backslash Y$ | 0 | 1 | 2 | Total |
|---|---|---|---|---|
| 0 | 0 | 0 | $1/4$ | $1/4$ |
| 1 | 0 | $2/4$ | 0 | $2/4$ |
| 2 | $1/4$ | 0 | 0 | $1/4$ |
| Total | $1/4$ | $2/4$ | $1/4$ | 1 |

In this case $\rho_{X,Y} = -1$ (the strongest negative correlation possible):

> *The values of $X$ and $Y$ are perfect inverses of each other ( $X = 2 - Y$ ).*

# Example: Tossing coins

Red line: $Y = 2 - X$




By means of the correlation coefficient $\rho_{X,Y}$ we can quantify the strength of the linear relationship between $X$ and $Y$ which is displayed by the joint probability distribution.

# Correlation and independence

# Example: Tossing coins

> *Toss a fair coin twice.*

Consider random variables $X$ and $Y$ as follows:

$$X = \begin{cases} 1, & \text{if 1st throw is heads} \\ 0, & \text{if 1st throw is tails} \end{cases} \qquad Y = \begin{cases} 1, & \text{if 2nd throw is heads} \\ 0, & \text{if 2nd throw is tails} \end{cases}.$$

In this case, $\rho_{X,Y} = 0$:

> *$X$ and $Y$ are uncorrelated.*
> *This means that there is no linear relationship between $X$ and $Y$.*

| $X$ \ $Y$ | 0 | 1 | Total |
|---|---|---|---|
| 0 | $1/4$ | $1/4$ | $1/2$ |
| 1 | $1/4$ | $1/4$ | $1/2$ |
| Total | $1/2$ | $1/2$ | 1 |

$X$ and $Y$ are also independent:

> *There is **no** relationship between $X$ and $Y$.*

In general:

$$\boxed{\text{independent} \implies \text{uncorrelated}}$$

# Uncorrelated but not independent

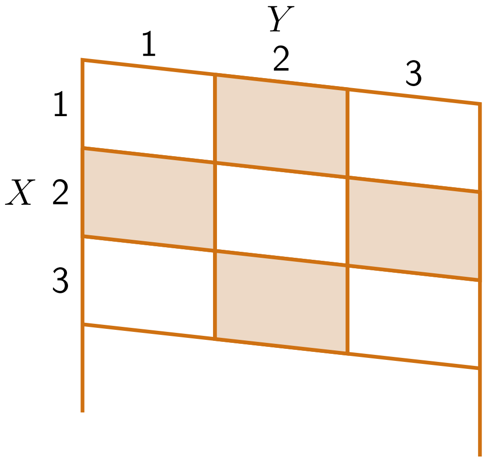
Notice that

$$\boxed{\text{uncorrelated} \not\Rightarrow \text{independent}}$$

**Example**

> *There is a pitcher who always throws a ball in one of the four areas , equally randomly.*

$X =$ number of row which was hit
$Y =$ number of column which was hit



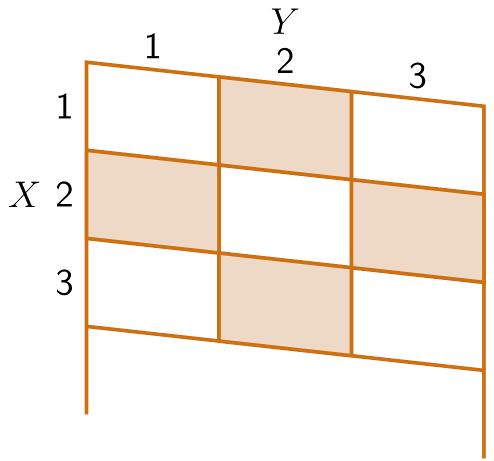

> $X$ *and* $Y$ *are* ***not*** *independent.*

For example, knowing that $X = 2$ does give us
some information about $Y$ (namely, $Y$ **cannot** be 2).

# Uncorrelated but not independent

However, $X$ and $Y$ are uncorrelated.
To see this, consider the joint probability distribution of $X$ and $Y$:

| $X$ \ $Y$ | 1 | 2 | 3 | Total |
|---|---|---|---|---|
| 1 | 0 | $^1/_4$ | 0 | $^1/_4$ |
| 2 | $^1/_4$ | 0 | $^1/_4$ | $^2/_4$ |
| 3 | 0 | $^1/_4$ | 0 | $^1/_4$ |
| Total | $^1/_4$ | $^2/_4$ | $^1/_4$ | 1 |




$\rho_{X,Y} = 0$:

> *There is **no linear relationship** between $X$ and $Y$.*

Thus, two variables may not be linearly related ( $X$ and $Y$ are uncorrelated), but they may still be associated ( $X$ and $Y$ are not independent)!

# Correlation and independence — Summary

$X$ and $Y$ are independent:

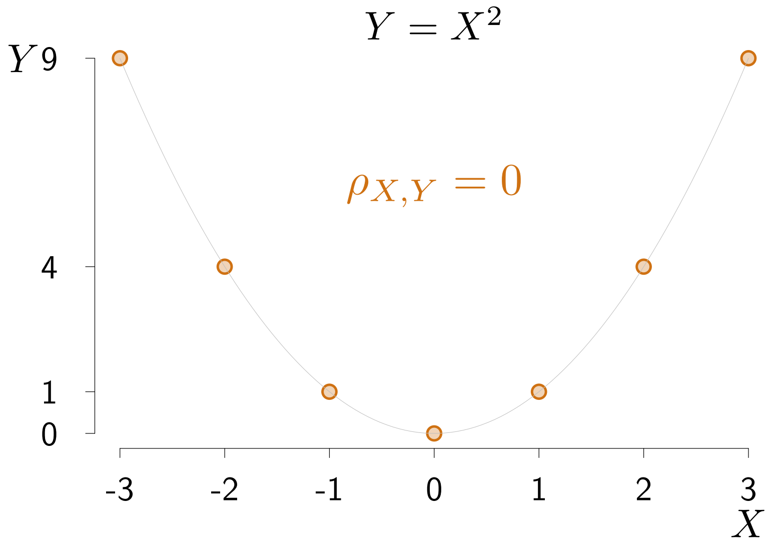
> *There is no relationship between $X$ and $Y$.*

Independent variables are **always** uncorrelated.

$X$ and $Y$ uncorrelated:

> *X and Y are not in a linear relationship.*

Uncorrelated variables are **not necessarily** independent.



# Trinomial distribution

# Trinomial distribution

Recall that the binomial distribution can be used to model counts of events when there are two possible outcomes.

**Example:**

> *Number of heads $X$ when tossing a coin $n$ times.*

Here, the two possible outcomes are 'heads' or 'tails'.

The trinomial distribution is an extension of the *binomial* distribution.
The trinomial distribution can be used to model counts of events when there are three possible outcomes.

# Trinomial distribution

**Example:**

> *Play a match $n$ times, under the following conditions:*
>
> - *Each match ends in one of three possible outcomes: winning/losing/drawing, with probability $p$, $q$, and $(1-p-q)$, respectively.*
> - *The matches are independent of each other.*

Let

- $X =$ number of wins
- $Y =$ number of losses

We say that $(X, Y)$ follows a trinomial distribution.

# Trinomial distribution

The joint probability distribution of the trinomial distribution is given as follows:

$$f_{X,Y}(k,l) = \frac{n!}{k!l!(n-k-l)!} p^k q^l (1-p-q)^{n-k-l},$$

for $0 \leq k+l \leq n, 0 \leq p \leq 1, 0 \leq q \leq 1.$

Here,

- $n =$ total number of events.
- $p =$ probability of one of the three possible outcomes.
- $q =$ probability of another of the three possible outcomes.

# Trinomial distribution

$$\boxed{f_{X,Y}(k,l) = \frac{n!}{k!l!(n-k-l)!}p^k q^l (1-p-q)^{n-k-l}}$$

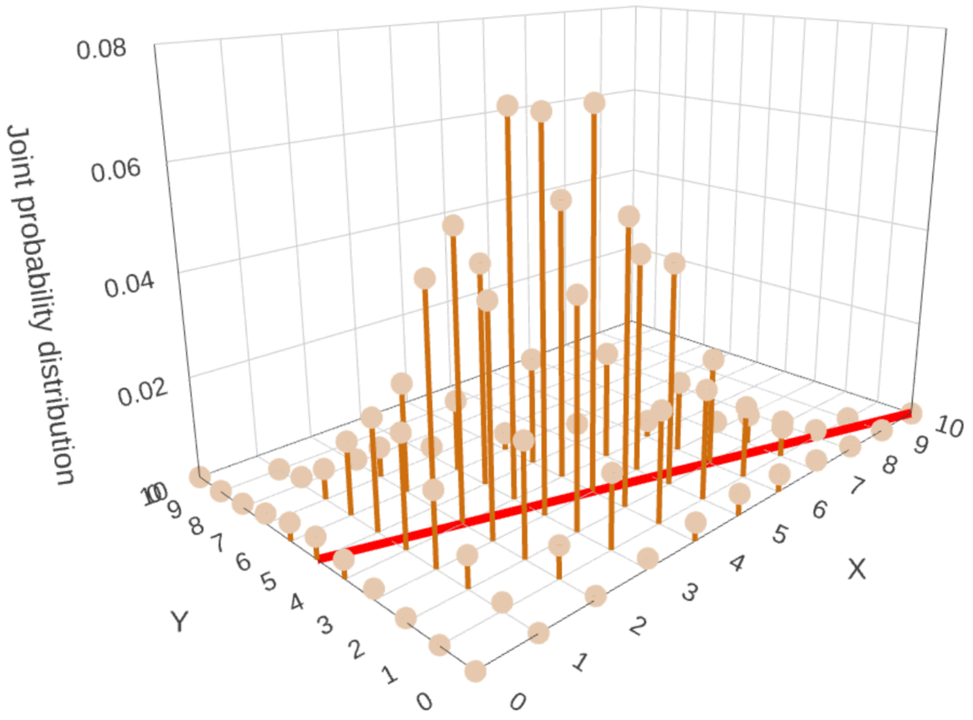
We write $(X,Y) \sim M_3(n,p,q)$ to denote that the joint distribution of $X$ and $Y$ is a trinomial distribution with parameters $n$, $p$, and $q$.

**Properties:**

$$\begin{aligned} V[X] &= p(1-p) \\ V[Y] &= q(1-q) \\ \mathrm{Cov}[X,Y] &= -npq \\ \rho_{X,Y} &= -\sqrt{\frac{pq}{(1-p)(1-q)}} \text{ when } p \neq 1 \text{ and } q \neq 1 \end{aligned}$$

# Trinomial distribution

Joint probability function of $(X,Y) \sim M_3(10, 1/3, 1/3)$:




$\rho_{X,Y} = -0.5$:
There is a negative correlation between $X$ and $Y$.

This *does* make sense:

> *As $X$ increases, $Y$ tends to decrease since $0 \leq X + Y \leq 10$.*

# Bivariate normal distribution

# Bivariate normal distribution

The bivariate normal distribution is an extension of the binomial distribution.
The joint probability distribution of the bivariate normal distribution is given as follows:

$$f_{X,Y}(x,y) = (\text{constant}) \times e^{\text{quadratic function of } X \text{ and } Y}$$

(the exact formula is omitted here).

We write $(X,Y) \sim N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{X,Y})$ to denote that the joint distribution of $X$ and $Y$ is a bivariate normal distribution with parameters $\mu_X$, $\mu_Y$, $\sigma_X^2$, $\sigma_Y^2$, and $\rho_{X,Y}$.

**Properties:**

$$\mathbb{E}[X] = \mu_X, \quad \mathbb{E}[Y] = \mu_Y, \quad V[X] = \sigma_X^2, \quad V[Y] = \sigma_Y^2, \quad \text{Cor}[X,Y] = \rho_{X,Y}$$

# Bivariate normal distribution — Location

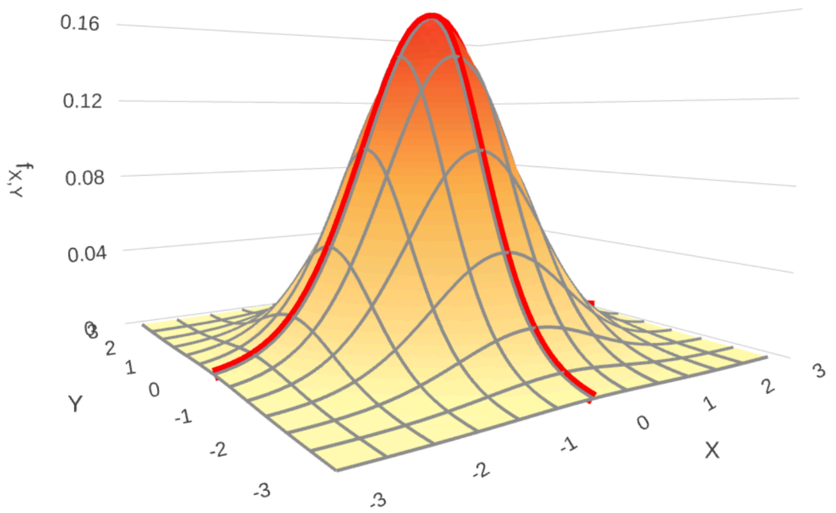
Parameters $\mu_X$ and $\mu_Y$ determine the central location of the $x$ and $y$ axes, respectively:

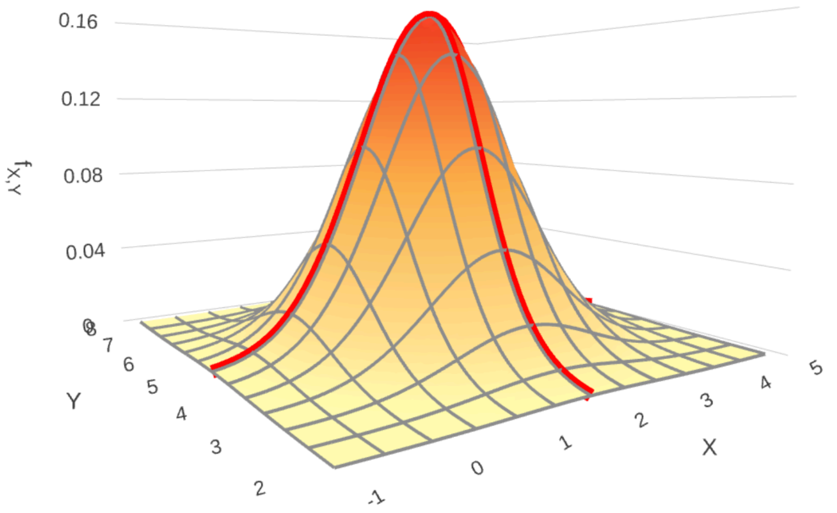
> *Changing the values of $\mu_X$ and $\mu_Y$ preserves the shape of the distribution, but moves it across the $XY$-plane.*

# Bivariate normal distribution — Location

$\boxed{\mu_X = 0, \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1, \rho_{X,Y} = 0}$

$\boxed{\mu_X = 2, \mu_Y = 5, \sigma_X^2 = 1, \sigma_Y^2 = 1, \rho_{X,Y} = 0}$



# Bivariate normal distribution — Spread

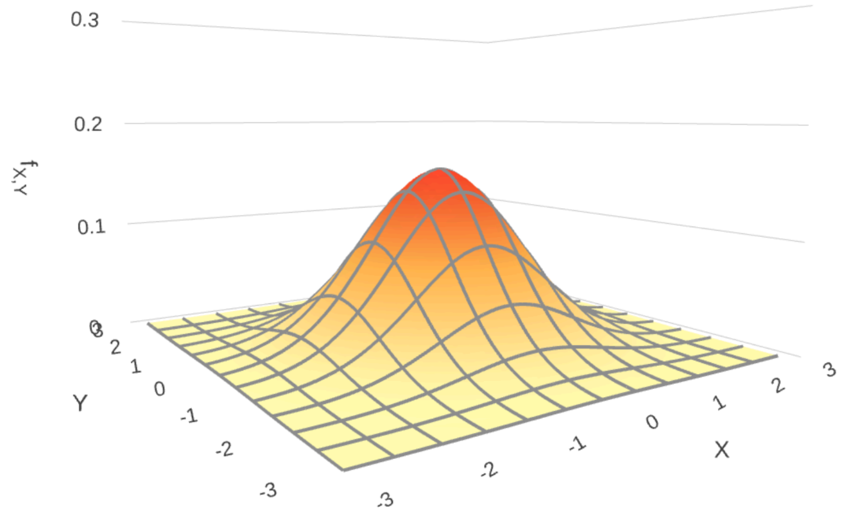
Parameters $\sigma_X^2$ and $\sigma_Y^2$ determine the spread of random variable $X$ and $Y$, respectively:

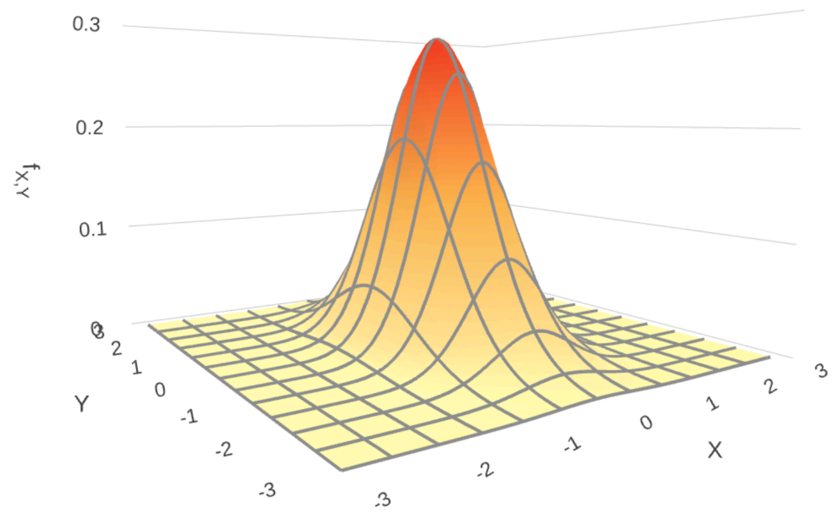
> *Changing the values of $\sigma_X^2$ and $\sigma_Y^2$ changes the shape of the distribution.*

# Bivariate normal distribution — Spread

$\boxed{\mu_X = 0, \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1, \rho_{X,Y} = 0}$

$\boxed{\mu_X = 0, \mu_Y = 0, \sigma_X^2 = 0.1, \sigma_Y^2 = 1, \rho_{X,Y} = 0}$



# Bivariate normal distribution — Linear relationship

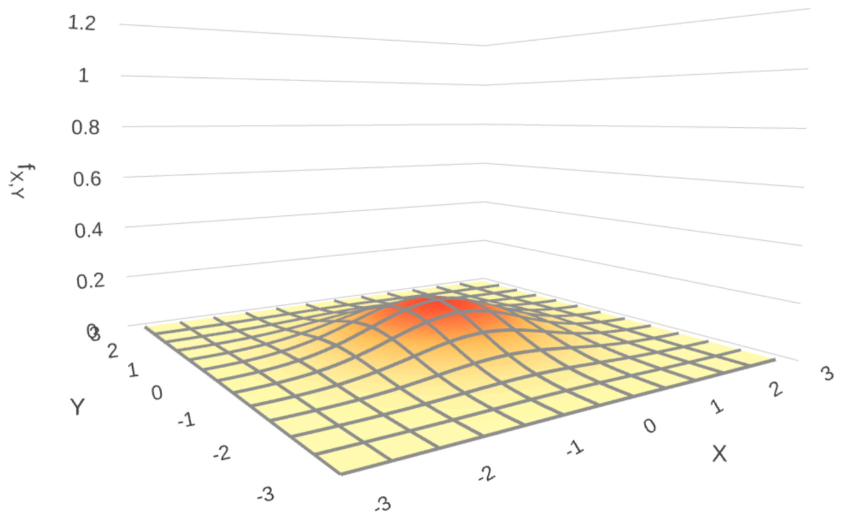
Parameter $\rho_{X,Y}$ indicates the strength of the linear relationship between $X$ and $Y$:

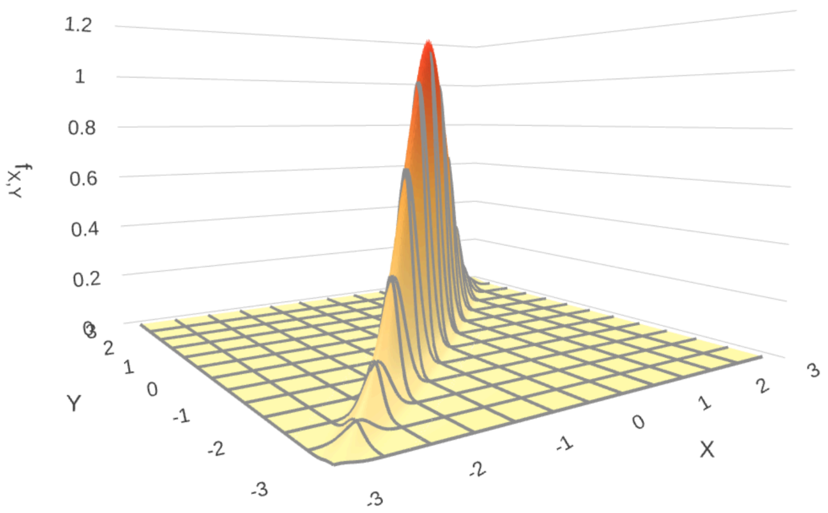
> *The larger $\rho_{X,Y}$ (in absolute value), the stronger the linear relationship between $X$ and $Y$.*

# Bivariate normal distribution — Linear relationship

$\mu_X = 0, \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1, \rho_{X,Y} = 0$

$\mu_X = 0, \mu_Y = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1, \rho_{X,Y} = 0.99$



# Summary

We learned how to handle two random variables $X$ and $Y$ at the same time!

- We can learn about the probability of two events occuring simultaneously from a joint probability distribution.

**Example:** If $X$ and $Y$ are discrete then we can compute

$$P(X = 1, Y = 2),$$

and if $X$ and $Y$ are continuous then we can compute

$$P(1 \leq X \leq 2, -2 \leq Y \leq 0).$$

- The correlation coefficient can tell us about the strength of the linear relationship between $X$ and $Y$.

When two events seem to have a relationship and you are interested to study them jointly, use a bivariate probability distribution.