# Multivariate Random Variables

# Multivariate Random Variables

Often, a single random variable is not sufficient to describe a physical problem. This may, for example, be the case when we are interested in the effect of one random quantity on another. In such a case we consider two (or more) random variables together.

Formally, we then define a "vector" of which each component is itself a ("scalar") random variable.

We call such a vector a ***random vector*** or a ***multi-variate random variable*** or an ***$n$-dimensional random variable***. The components can be discrete or continuous random variables, and even mixtures of the two.

In this section we will for the most part focus on bivariate (two-dimensional) random variables where either both components are discrete or both components are continuous random variables.

# Discrete Multivariate Random Variables

8.1. Definition. Let $S$ be a sample space and $\Omega$ a countable subset of $\mathbb{R}^n$. A ***discrete multivariate random variable*** is a map

$$\boldsymbol{X} : S \to \Omega$$

together with a function $f_{\boldsymbol{X}} : \Omega \to \mathbb{R}$ with the properties that

(i) $f_{\boldsymbol{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \Omega$ and

(ii) $\sum\limits_{x \in \Omega} f_{\boldsymbol{X}}(x) = 1$.

The function $f_{\boldsymbol{X}}$ is called the ***joint density function*** of the random variable $\boldsymbol{X}$.

# Discrete Multivariate Random Variables

We consider the multivariate random variable $\boldsymbol{X}$ to have $n$ components, i.e.,

$$\boldsymbol{X} = (X_1, \ldots, X_n).$$

We often write

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n) = f_{X_1 \cdots X_n}(x_1, \ldots, x_n)$$

The joint density function $f_{\boldsymbol{X}}$ gives the probability that the tuple $(X_1, \ldots, X_n)$ assumes a given value $x \in \mathbb{R}^n$, i.e.,

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n) = P[X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \ldots \text{ and } X_n = x_n].$$

Given two random variables, we may write $(X, Y)$ instead of $(X_1, X_2)$ and use similar notation for three or larger numbers of components.

# Discrete Bivariate Random Variables

8.2. Example. Suppose we roll two six-sided dice, obtaining results $(i, j)$ with $, j = 1, \ldots, 6$. Let us define

$$X := \boxed{i + j \bmod 5} \qquad Y = i - j \bmod 5.$$

(i+j)÷5 的余数

Then we can find the values of the probability density function by Cardano's rule. The number of outcomes leading to each event $(X, Y)$ is

(4, 6)

$X = 4+6 \bmod 5$
$= 0$

$Y = 4-6 \bmod 5$
$= -2 \bmod 5$
$= 3$

| $x/y$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 1 | 1 |
| 1 | 1 | 2 | 1 | 2 | 1 |
| 2 | 2 | 1 | 1 | 1 | 2 |
| 3 | 2 | 1 | 1 | 1 | 2 |
| 4 | 1 | 2 | 1 | 2 | 1 |

$P = \frac{2}{36} = \frac{1}{18}$

so each number in the table must be divided by 36 to obtain the corresponding probability. For example, $P[X = 1 \text{ and } Y = 1] = 1/18$.

# Marginal Density

While each element of the table gives us $36 \cdot P[X = x \text{ and } Y = y]$, we can find the probability of the event $X = x$ by adding up all relevant probabilities:

$$P[X = x] = \sum_{y=0}^{4} P[X = x \text{ and } Y = y]$$

| $x/y$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 1 | 1 |
| 1 | 1 | 2 | 1 | 2 | 1 |
| 2 | 2 | 1 | 1 | 1 | 2 |
| 3 | 2 | 1 | 1 | 1 | 2 |
| 4 | 1 | 2 | 1 | 2 | 1 |

For example,

$$P[X = 0] = (1 + 1 + 4 + 1 + 1)/36 = 8/36.$$

This procedure can be justified by considering the corresponding en=vent in the sample space.

By summing in this way, we can determine $P[X = x]$ for all $x$. This is called the ***marginal density*** for $X$.

# Marginal Density of a Discrete Random Variable

**8.3. Definition.** Let $(\boldsymbol{X}, f_{\boldsymbol{X}})$ be a discrete multivariate random variable. We define the ***marginal density*** $f_{x_k}$ for $X_k$, $k = 1, \ldots, n$, by

$$f_{X_k}(x_k) = \sum_{x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_n} f_{\boldsymbol{X}}(x_1, \ldots, x_n).$$

**8.4. Example.**

| $x/y$ | 0 | 1 | 2 | 3 | 4 | $f_X(x)$ |
|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 1 | 1 | $8/36$ |
| 1 | 1 | 2 | 1 | 2 | 1 | $7/36$ |
| 2 | 2 | 1 | 1 | 1 | 2 | $7/36$ |
| 3 | 2 | 1 | 1 | 1 | 2 | $7/36$ |
| 4 | 1 | 2 | 1 | 2 | 1 | $7/36$ |
| $f_Y(y)$ | $7/36$ | $7/36$ | $8/36$ | $7/36$ | $7/36$ | 1 |

# Independence of two Random Variables

Question. Considering the table:

| $x/y$ | 0 | 1 | 2 | 3 | 4 | $f_X(x)$ |
|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 1 | 1 | 8/36 |
| 1 | 1 | 2 | 1 | 2 | 1 | 7/36 |
| 2 | 2 | 1 | 1 | 1 | 2 | 7/36 |
| 3 | 2 | 1 | 1 | 1 | 2 | 7/36 |
| 4 | 1 | 2 | 1 | 2 | 1 | 7/36 |
| $f_Y(y)$ | 7/36 | 7/36 | 8/36 | 7/36 | 7/36 | 1 |

Do you think that $X$ and $Y$ are independent?

- Yes
- No ✓ 如果 independent, 这个概率应当为 $\frac{8}{36} \times \frac{7}{36}$
- It's not possible to tell from the table.

# Independence of Random Variables

If $(\boldsymbol{X}, f_{\boldsymbol{X}})$ is a discrete ***bivariate*** random variable, i.e., $\boldsymbol{X} = (X_1, X_2)$, we say that $X_1$ and $X_2$ are ***independent*** if

$$P[X_1 = x_1 \text{ and } X_2 = x_2] = P[X_1 = x_1] \cdot P[X_2 = x_2].$$

In other words, if

$$f_{\boldsymbol{X}}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$

(The joint density is the product of the marginal densities.)

It is possible to generalize this in the obvious (but notationally cumbersome) way to $n$-variate random variables.

We will mostly be interested in cases where $\boldsymbol{X} = (X_1, \dots, X_n)$ and all the components are independent, i.e.,

$$f_{\boldsymbol{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

# Independence of two Random Variables

## 8.5. Example.

| $x/y$ | 0 | 1 | 2 | 3 | 4 | $f_X(x)$ |
|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 1 | 1 | $8/36$ |
| 1 | 1 | 2 | 1 | 2 | 1 | $7/36$ |
| 2 | 2 | 1 | 1 | 1 | 2 | $7/36$ |
| 3 | 2 | 1 | 1 | 1 | 2 | $7/36$ |
| 4 | 1 | 2 | 1 | 2 | 1 | $7/36$ |
| $f_Y(y)$ | $7/36$ | $7/36$ | $8/36$ | $7/36$ | $7/36$ | 1 |

The variables $X$ and $Y$ are not independent since, for example,

$$P[X = 1 \text{ and } Y = 1] = 1/18$$

but

$$P[X = 1] \cdot P[Y = 1] = \frac{7}{36} \cdot \frac{7}{36}$$

and the two expressions are not equal.

# Conditional Density

Suppose that $(\boldsymbol{X}, f_{\boldsymbol{X}})$ is a discrete bivariate random variable, i.e., $\boldsymbol{X} = (X_1, X_2)$, and that $X_2$ is known to have taken on a certain value.

Then, applying elementary probability laws,

$$P[X_1 = x_1 \mid X_2 = x_2] = \frac{P[X_1 = x_1 \text{ and } X_2 = x_2]}{P[X_2 = x_2]} = \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

We hence define the ***conditional density***

$$f_{X_1|x_2}(x_1) := \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad \text{whenever } f_{X_2}(x_2) > 0,$$

where $f_{X_2}$ is the marginal density of $X_2$.

# Continuous Random Variables

8.6. Definition. Let $S$ be a sample space. A ***continuous multivariate random variable*** is a map

$$\boldsymbol{X} : S \to \mathbb{R}^n$$

together with a function $f_{\boldsymbol{X}} : \mathbb{R}^n \to \mathbb{R}$ with the properties that

(i) $f_{\boldsymbol{X}}(x) \geq 0$ for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and

(ii) $\int_{\mathbb{R}^n} f_{\boldsymbol{X}}(x)\, dx = 1$.

The function $f_{\boldsymbol{X}}$ is called the ***joint density function*** of the random variable $\boldsymbol{X}$.

# Continuous Random Variables

The integral of $f_{\boldsymbol{X}}$ is interpreted as the probability that $\boldsymbol{X}$ assumes values in a given domain $\Omega \subset \mathbb{R}^n$,

$$P[\boldsymbol{X} \in \Omega] = \int_{\Omega} f_{\boldsymbol{X}}(x)\, dx.$$

For example, if $\boldsymbol{X} = (X_1, X_2)$,

$$P[a \le X_1 \le b \text{ and } c \le X_2 \le d] = \int_a^b \int_c^d f_{X_1 X_2}(x_1, x_2)\, dx_1\, dx_2$$

for $a \le b$, $c \le d$.

But of course non-rectangular domains can be considered as well.

We now make definitions for continuous random variables that are completely analogous to those for the discrete case.

# Continuous Multivariate Random Variables

We define the ***marginal density*** of $X_k$, $k = 1, \dots, n$, by

$$f_{X_k}(x_k) = \int_{\mathbb{R}^{n-1}} f_{\boldsymbol{X}}(x)\, dx_1 \dots dx_{k-1}\, dx_{k+1} \dots dx_n.$$

We say that two continuous random variables are ***independent*** if

$$f_{\boldsymbol{X}}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$

and we are often interested in the case where a full set of $n$ components of a multivariate random variable is independent:

$$f_{\boldsymbol{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

The ***conditional density*** for continuous bivariate random variables is similarly

$$f_{X_1|x_2}(x_1) := \frac{f_{X_1 X_2}(x_1, x_2)}{f_{X_2}(x_2)} \qquad \text{whenever } f_{X_2}(x_2) > 0.$$

# Expectation

We define the ***expected value*** or ***expectation*** for $\boldsymbol{X}$ as the vector

$$\mathrm{E}[\boldsymbol{X}] = \begin{pmatrix} \mathrm{E}[X_1] \\ \vdots \\ \mathrm{E}[X_n] \end{pmatrix}$$

where $\mathrm{E}[X_k]$ is calculated using the marginal density of $X_k$, $k = 1, \dots, n$,

$$\mathrm{E}[X_k] = \sum_{x_k} x_k f_{X_k}(x_k) = \sum_{x \in \Omega} x_k f_{\boldsymbol{X}}(x)$$

and

marginal density

joint density

$$\mathrm{E}[X_k] = \int_{\mathbb{R}} x_k f_{X_k}(x_k)\, dx_k = \int_{\mathbb{R}^n} x_k f_{\boldsymbol{X}}(x)\, dx$$

for discrete and continuous random variables, respectively.

# Expectation for Discrete Bivariate Random Variables

## 8.7. Example.

| $x/y$ | 0 | 1 | 2 | 3 | 4 | $f_X(x)$ |
|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 1 | 1 | $8/36$ |
| 1 | 1 | 2 | 1 | 2 | 1 | $7/36$ |
| 2 | 2 | 1 | 1 | 1 | 2 | $7/36$ |
| 3 | 2 | 1 | 1 | 1 | 2 | $7/36$ |
| 4 | 1 | 2 | 1 | 2 | 1 | $7/36$ |
| $f_Y(y)$ | $7/36$ | $7/36$ | $8/36$ | $7/36$ | $7/36$ | 1 |

$$\mathrm{E}[X] = \sum_{(x,y)\in\Omega} x \cdot f_{XY}(x,y) = \sum_{x=0}^{4} x \cdot f_X(x) = \frac{70}{36}$$

$$\mathrm{E}[Y] = \sum_{(x,y)\in\Omega} y \cdot f_{XY}(x,y) = \sum_{y=0}^{4} y \cdot f_Y(y) = 2$$

# Expectation for Functions of Random Vectors

Suppose $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ is a continuous function. Then

$$\varphi \circ \boldsymbol{X} \colon S \to \mathbb{R}$$

defines a scalar random variable. It is possible to prove that in this case,

$$\mathrm{E}[\varphi \circ \boldsymbol{X}] = \sum_{x \in \Omega} \varphi(x) f_{\boldsymbol{X}}(x), \qquad \text{or} \qquad \mathrm{E}[\varphi \circ \boldsymbol{X}] = \int_{\mathbb{R}^n} \varphi(x) f_{\boldsymbol{X}}(x)\, dx.$$

For $\varphi(x_1, \dots, x_n) = x_k$ we regain the definition of $\mathrm{E}[X_k]$.

# Expectation for the Sum of Two Random Variables

8.8. Remark. If $(X, Y)$ is a discrete bivariate random variable and $\varphi(x, y) = x + y$, we have

$$\begin{aligned}
\mathrm{E}[X+Y] &= \sum_{(x,y)\in\Omega} (x+y)\cdot f_{XY}(x,y) \\
&= \sum_{(x,y)\in\Omega} x\cdot f_{XY}(x,y) + \sum_{(x,y)\in\Omega} y\cdot f_{XY}(x,y) \\
&= \mathrm{E}[X] + \mathrm{E}[Y].
\end{aligned}$$

This establishes the addition property of the expectation taht we introduced earlier.

An analogous calculation may be used for continuous random variables.

# Variance and Covariance for Bivariate Random Variables

Let us calculate the variance of the sum of two random variables:

$$\begin{aligned}
\operatorname{Var}[X+Y] &= \mathrm{E}\Big[\big((X+Y)-\mathrm{E}[X+Y]\big)^2\Big] \\
&= \mathrm{E}\Big[\big((X-\mathrm{E}[X])+(Y-\mathrm{E}[Y])\big)^2\Big] \\
&= \mathrm{E}\Big[(X-\mathrm{E}[X])^2+(Y-\mathrm{E}[Y])^2+2(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\Big] \\
&= \operatorname{Var}[X]+\operatorname{Var}[Y]+2\,\mathrm{E}\Big[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])\Big] \qquad (8.1)
\end{aligned}$$

In general,

$$\operatorname{Var}[X+Y] \neq \operatorname{Var}[X]+\operatorname{Var}[Y].$$

We define the ***covariance of*** $(X, Y)$,

$$\operatorname{Cov}[X,Y] = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)],$$

where we have used $\mu$ to denote the expectations. Note that

$$\operatorname{Cov}[X,Y] = \operatorname{Cov}[Y,X] \qquad \text{and} \qquad \operatorname{Cov}[X,X] = \operatorname{Var}[X].$$

# The Covariance Matrix

For a multivariate random variable $\boldsymbol{X}$ we define the ***covariance matrix***

$$\operatorname{Var}[\boldsymbol{X}] = \begin{pmatrix} \operatorname{Var}[X_1] & \operatorname{Cov}[X_1, X_2] & \dots & \operatorname{Cov}[X_1, X_n] \\ \operatorname{Cov}[X_1, X_2] & \operatorname{Var}[X_2] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \operatorname{Cov}[X_{n-1}, X_n] \\ \operatorname{Cov}[X_1, X_n] & \dots & \operatorname{Cov}[X_{n-1}, X_n] & \operatorname{Var}[X_n] \end{pmatrix}.$$

It is possible to prove (through tedious calculation) that

$$\operatorname{Var}[C\boldsymbol{X}] = C \operatorname{Var}[\boldsymbol{X}] C^T$$

where $C \in \operatorname{Mat}(n \times n; \mathbb{R})$ is a constant $n \times n$ matrix with real coefficients.

$\operatorname{Var}[cX] = c^2 \operatorname{Var}[X]$ (c为常数)

# Covariance and Independence

Just as for the variance, a direct calculation yields

$$\operatorname{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].$$

Furthermore, if two continuous random variables $X$ and $Y$ are independent, then $f_{XY}(x, y) = f_X(x) f_Y(y)$ and

$$\begin{aligned}
\mathrm{E}[XY] &= \iint_{\mathbb{R}^2} xy \cdot f_{XY}(x, y)\, dx\, dy \\
&= \iint_{\mathbb{R}^2} xy \cdot f_X(x) f_Y(y)\, dx\, dy \\
&= \left( \int_{\mathbb{R}} x \cdot f_X(x)\, dx \right) \left( \int_{\mathbb{R}} y \cdot f_Y(y)\, dy \right) \\
&= \mathrm{E}[X]\,\mathrm{E}[Y]
\end{aligned}$$

# Covariance and Independence

An analogous calculation works for discrete random variables. We have hence proved:

- If $X$ and $Y$ are independent, then $\text{Cov}[X, Y] = 0$.

However, the converse is not true:

- If $\text{Cov}[X, Y] = 0$, then $X$ and $Y$ are ***not necessarily independent***.

We note that we have also established that

$$\begin{aligned} \text{Var}[X - Y] &= \text{Var}[X + (-Y)] \\ &= \text{Var}[X] + \text{Var}[-Y] \\ &= \text{Var}[X] + \text{Var}[(-1) \cdot Y] \\ &= \text{Var}[X] + (-1)^2 \text{Var}[Y] \\ &= \text{Var}[X] + \text{Var}[Y] \end{aligned}$$

$$E[X - Y] = E[X] - E[Y]$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

$$\text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y]$$

***if the random variables are independent***.

The covariance is hence related to the independence of $X$ and $Y$. However, it is not a measure for dependence, since two dependent variables can still have a vanishing covariance.

So we ask: what does the covariance actually measure?

# Standardizing Random Variables

We note that the covariance scales with $X$ and $Y$, i.e., if $X$ and $Y$ take on numerically large values, then the covariance will be large, while if if $X$ and $Y$ take on small values, the covariance will be small. Therefore, by itself it does not serve very well as a measure of any fundamental properties of $X$ and $Y$.

The solution is to standardize the random variables,

$$\widetilde{X} := \frac{X - \mu_X}{\sigma_X}$$

is the standardized variable for $X$ (assuming that both mean and variance of $X$ exist and $\sigma_X \neq 0$).

Recall that

$$\mathrm{E}[\widetilde{X}] = 0, \qquad\qquad \mathrm{Var}[\widetilde{X}] = 1.$$

# The Pearson Correlation Coefficient

***Karl Pearson (1857-1936) in 1912***. File:Karl Pearson, 1912.jpg. (2018, January 17). Wikimedia Commons, the free media repository.

Instead of $\operatorname{Cov}[X, Y]$ we now consider

$$\operatorname{Cov}[\widetilde{X}, \widetilde{Y}] = \operatorname{E}[\widetilde{X}\widetilde{Y}] - \operatorname{E}[\widetilde{X}]\operatorname{E}[\widetilde{Y}]$$
$$= \frac{\operatorname{Cov}[X, Y]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}}$$

The right-hand side is now scale-independent and unit-less (if $X$ and $Y$ have units).

This quotient is known as the ***Pearson coefficient of correlation*** of $(X, Y)$ and denoted

$$\rho_{XY} := \frac{\operatorname{Cov}[X, Y]}{\sqrt{\operatorname{Var}[X]\operatorname{Var}[Y]}}.$$

$$\cos(\angle(\vec{x}, \vec{y})) = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \cdot \|\vec{y}\|} \in [-1, 1]$$

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$$

# Properties of the Correlation Coefficient

It can be shown that $\rho_{XY}$ has the following properties

(i) $-1 \leq \rho_{XY} \leq 1$,

(ii) $|\rho_{XY}| = 1$ if and only if there exist numbers $\beta_0, \beta_1 \in \mathbb{R}$, $\beta_1 \neq 0$, such that

$$Y = \beta_0 + \beta_1 X$$ 线性相关

depedence

almost surely.

The proof is best performed in a vector-space setting, which we omit here.

The above properties give us a clue as to how the correlation coefficient might be interpreted: if it has modulus one, then $X$ and $Y$ are in a deterministically linear relationship. Let us therefore start from that angle.

# Measuring Linearity of $X$ and $Y$

Suppose that $X$ and $Y$ are related in a linear fashion, say

$$Y = \beta_0 + \beta_1 X, \tag{8.2}$$

with $\beta_1 \neq 0$. Then

$$\mu_Y = \beta_0 + \beta_1 \mu_X$$

and $\operatorname{Var}[Y] = \beta_1^2 \operatorname{Var}[X]$, so

$$\sigma_Y = |\beta_1| \sigma_X .$$

# Measuring Linearity of $X$ and $Y$

Using the standardized variables, we find that

$$\begin{aligned}
\widetilde{Y} &= \frac{Y - \mu_Y}{\sigma_Y} \\
&= \frac{\beta_0 + \beta_1 X - (\beta_0 + \beta_1 \mu_X)}{|\beta_1| \sigma_X} \\
&= \frac{\beta_1}{|\beta_1|} \frac{X - \mu_X}{\sigma_X} \\
&= \frac{\beta_1}{|\beta_1|} \widetilde{X}.
\end{aligned}$$

We conclude that $X$ and $Y$ are in a linear relationship if and only if the standardized variables are either equal or the negative of each other.

# Measuring Linearity of $X$ and $Y$

We now know that $X$ and $Y$ are deterministically linearly related if and only if

$$\widetilde{X} + \widetilde{Y} = 0 \qquad \text{or} \qquad \widetilde{X} - \widetilde{Y} = 0.$$

In order to measure in how far $X$ and $Y$ are not linearly related, it makes sense to consider the standard deviation of $\widetilde{X} + \widetilde{Y}$ and $\widetilde{X} - \widetilde{Y}$. If either of these were zero, the relationship would be deterministically linear.

We calculate $E[\widetilde{X} \pm \widetilde{Y}] = 0$

$$\operatorname{Var}[\widetilde{X} + \widetilde{Y}] = \operatorname{Var}[\widetilde{X}] + \operatorname{Var}[\widetilde{Y}] + 2\operatorname{Cov}[\widetilde{X}, \widetilde{Y}] = 2 + 2\varrho_{XY},$$
$$\operatorname{Var}[\widetilde{X} - \widetilde{Y}] = \operatorname{Var}[\widetilde{X}] + \operatorname{Var}[\widetilde{Y}] - 2\operatorname{Cov}[\widetilde{X}, \widetilde{Y}] = 2 - 2\varrho_{XY}.$$

If either of these two variances is small, then $\widetilde{X}$ and $\widetilde{Y}$ are "nearly proportional" and so $X$ and $Y$ are "nearly linearly" related.

# The Fisher Transformation

In order to capture both of these positive quantities in a single manner, let us consider their quotient,

$$\sqrt{\frac{\operatorname{Var}[\widetilde{X}+\widetilde{Y}]}{\operatorname{Var}[\widetilde{X}-\widetilde{Y}]}} = \sqrt{\frac{1+\rho_{XY}}{1-\rho_{XY}}} \in (0,\infty)$$

If $X$ and $Y$ are linearly related, then this quotient will be either very small or very large.

It is "mathematically nicer" to take the logarithm:

Ronald Fisher (1890-1962) in 1913. File:Youngronaldfisher2.JPG. (2018, July 7). Wikimedia Commons, the free media repository.

$$\ln\left(\sqrt{\frac{\operatorname{Var}[\widetilde{X}+\widetilde{Y}]}{\operatorname{Var}[\widetilde{X}-\widetilde{Y}]}}\right) = \frac{1}{2}\ln\left(\frac{1+\rho_{XY}}{1-\rho_{XY}}\right) = \operatorname{Artanh}(\rho_{XY}) \in \mathbb{R}.$$

This is known as the ***Fisher transformation*** of $\rho_{XY}$.

# Positive and Negative Correlation

It follows that

$$\rho_{XY} = \tanh\left(\ln\left(\frac{\sigma_{\widetilde{X}+\widetilde{Y}}}{\sigma_{\widetilde{X}-\widetilde{Y}}}\right)\right).$$

- If $\rho_{XY} > 0$, then $\mathrm{Var}[\widetilde{X} + \widetilde{Y}] > \mathrm{Var}[\widetilde{X} - \widetilde{Y}]$, which implies that the relationship between $X$ and $Y$ is closer to $\widetilde{X} = \widetilde{Y}$ than to $\widetilde{X} = -\widetilde{Y}$. Hence, if $X$ is large, $Y$ tends to be large also.

  We say that $X$ and $Y$ are ***positively correlated***.

- If $\rho_{XY} < 0$, then $\mathrm{Var}[\widetilde{X} + \widetilde{Y}] < \mathrm{Var}[\widetilde{X} - \widetilde{Y}]$ and the situation is reversed. If $X$ is large, $Y$ tends to be small.

  We say that $X$ and $Y$ are ***negatively correlated***.

Since $X$ and $Y$ are still *random* variables, a large value of $X$ only indicates a tendency for $Y$ to be large/small but doesn't guarantee this. The closer $\rho_{XY}$ is to $\pm 1$, the more pronounced these effects are.

# The Bivariate Normal Distribution

8.9. Example. Suppose two random variables $X$ and $Y$ should each follow a (marginal) normal distribution, but are not independent.

The most common model is the so-called ***bivariate normal distribution***, with density function

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}}\, e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

where $-1 < \varrho < 1$.

The marginal distributions can be shown to be normal, $\mu_X = \mathrm{E}[X]$, $\sigma_X^2 = \operatorname{Var} X$ (and similarly for $Y$) and $\varrho = \rho_{XY}$ is indeed the correlation coefficient of $X$ and $Y$.

Furthermore, $X$ and $Y$ are independent if and only if $\varrho = 0$.

This distribution will be discussed in detail in the assignments.