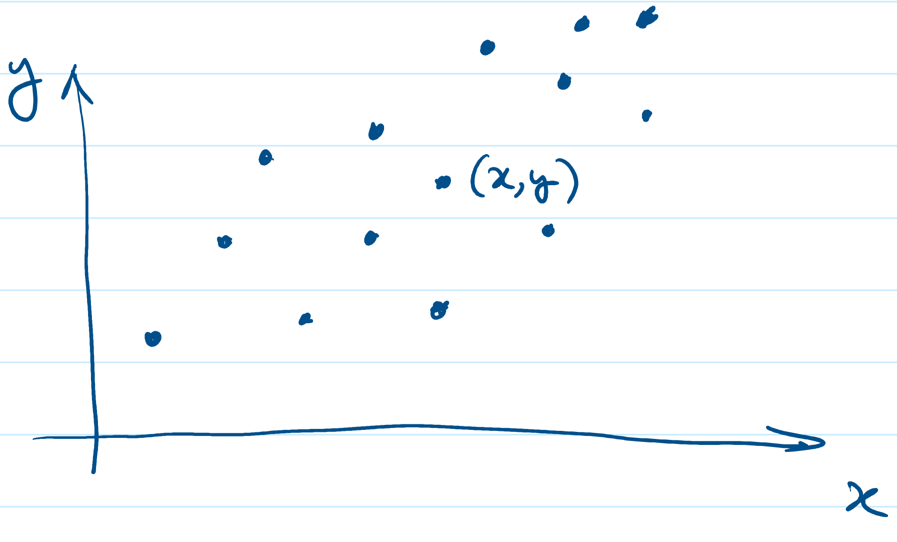
4.5 Covariance and correlation

Let (X,Y) be a bivariate vector $((X,Y) \in \mathbb{R}^2)$

e.g $(X,Y) = $ (height, weight) of a person

want to express the observations than
"if $X \uparrow$ then Y tends to $\uparrow$, and vice versa"



we say X and Y are possitively correlated.

Another example
For $(X,Y) = $ (time on screen, time on reading)
X and Y may be negatively correlated.

Def. Covariance of Random variables X and Y is

$$\mathrm{Cov}(X,Y) := \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)$$

Remark.

① if $X=Y$ then $\mathrm{Cov}(X,X) \equiv \mathrm{Var}(X)$.

② if $\mathbb{E}X = \mathbb{E}Y = 0$ then $\mathrm{Cov}(X,Y) = \mathbb{E}XY$

③ Correlation of X and Y

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}}$$

in other words, $\mathrm{Corr}(X,Y)$ is the covariance of the standardized versions of X and Y.

indeed,

let $X' = \dfrac{X - \mathbb{E}X}{\sqrt{\operatorname{var} X}}$, $Y' = \dfrac{Y - \mathbb{E}Y}{\sqrt{\operatorname{var} Y}}$

Then $\begin{cases} \mathbb{E}X' = \mathbb{E}Y' = 0 \\ \operatorname{var} X' = \operatorname{var} Y' = 1. \end{cases}$

By definition,

$$\begin{aligned}
\operatorname{corr}(X', Y') &= \operatorname{cov}(X', Y') \\
&= \mathbb{E}(X'Y') \\
&= \mathbb{E}\, \frac{(X - \mathbb{E}X)}{\sqrt{\operatorname{var} X}} \frac{(Y - \mathbb{E}Y)}{\sqrt{\operatorname{var} Y}} \\
&= \frac{\mathbb{E}\,(X - \mathbb{E}X)(Y - \mathbb{E}Y)}{\sqrt{\operatorname{var} X}\sqrt{\operatorname{var} Y}} \\
&= \operatorname{corr}(X, Y). \qquad \square
\end{aligned}$$

Theorem

(i) $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$.

(ii) if $X \perp\!\!\!\perp Y$ then $\mathrm{Cov}(X,Y) = \mathrm{corr}(X,Y) = 0$.

Remark.

(1) generalizes the fact that $\mathrm{var}\,X = \mathbb{E}X^2 - (\mathbb{E}X)^2$

(2) Conversely, $\mathrm{cov}(X,Y) = 0$ does <u>NOT</u> imply $X \perp\!\!\!\perp Y$ (unless (X,Y) is bivariate normal, to be seen)

Theorem $\forall a,b \in \mathbb{R}$

$$\mathrm{var}(aX + bY) = a^2\,\mathrm{var}\,X + b^2\,\mathrm{var}\,Y + 2ab\,\mathrm{Cov}(X,Y).$$

Remarks

A nice consequence of this theorem is the following.

Fix $b=1$ then

$$a^2 \operatorname{var} X + 2a \operatorname{Cov}(X,Y) + \operatorname{var} Y \geq 0 \quad \forall a.$$

Suppose $\sqrt{\operatorname{var} X} > 0$. Minimizing LHS with respect to a gives

$$a = -\sqrt{\operatorname{var} Y} / \sqrt{\operatorname{var} X}, \quad \text{for which}$$

$$\text{LHS} = \left(a\sqrt{\operatorname{var} X} + \sqrt{\operatorname{var} Y}\right)^2 + 2a\left(\operatorname{Cov}(X,Y) - \sqrt{\operatorname{var} X \cdot \operatorname{var} Y}\right)$$

$$= 2a\left(\operatorname{Cov}(X,Y) - \sqrt{\operatorname{var} X \operatorname{var} Y}\right) \geq 0.$$

So $$\operatorname{Cov}(X,Y) - \sqrt{\operatorname{var} X}\sqrt{\operatorname{var} Y} \leq 0.$$

if moreover, $\operatorname{var} X \operatorname{var} Y \neq 0$ then by dividing by $\sqrt{\operatorname{var} X}\sqrt{\operatorname{var} Y}$ we obtain

Theorem. if $\operatorname{var} X \operatorname{var} Y \neq 0$ then

$$\boxed{-1 \leq \operatorname{Corr}(X,Y) \leq 1.}$$

Proof.

We already proved $\operatorname{Corr}(X,Y) \le 1$.

But also, $\operatorname{Corr}(X,-Y) \le 1$

But $\operatorname{Corr}(X,Y) = -\operatorname{Corr}(X,-Y)$

So $\operatorname{Corr}(X,Y) \ge -1$ $\square$.

Suppose $\operatorname{Corr}(X,Y) = 1$?

From above proof,

$$
\begin{aligned}
\operatorname{Var}(aX+Y) &= a^2 \operatorname{Var}X + 2a \operatorname{Cov}(X,Y) + \operatorname{Var}Y \\
&= a^2 \operatorname{Var}X + 2a\sqrt{\operatorname{Var}X \operatorname{Var}Y} + \operatorname{Var}Y \\
&= \left(a\sqrt{\operatorname{Var}X} + \sqrt{\operatorname{Var}Y}\right)^2
\end{aligned}
$$

So for $a = -\sqrt{\operatorname{Var}Y} / \sqrt{\operatorname{Var}X}$, RHS $= 0$

$$\Rightarrow \quad \mathrm{var}(aX+Y) = 0 \quad \Rightarrow \quad aX+Y = \text{Const.}$$

$$Y - (-a)X = \text{Const.}$$

"with probability 1".

Theorem Suppose $\mathrm{var}X \neq 0$, $\mathrm{var}Y \neq 0$.

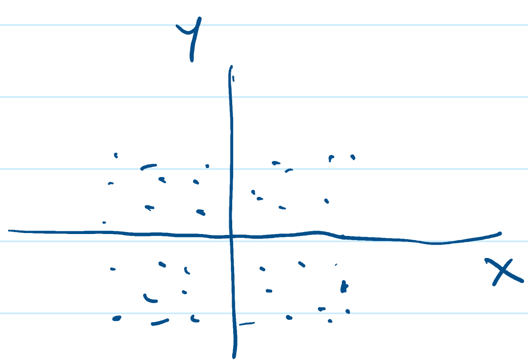
(i) if $\mathrm{Corr}(X,Y) = 1$ then there are constants $b>0$ and c such that

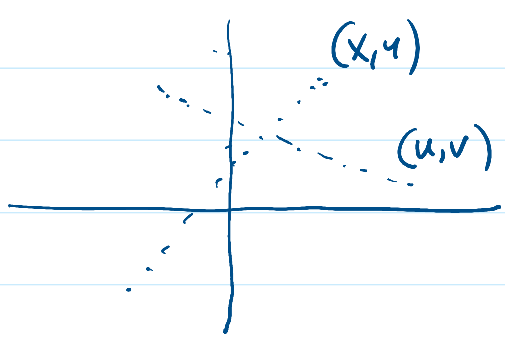
$$Y = bX + c \quad \text{with prob. } 1.$$

(ii) if $\mathrm{Corr}(X,Y) = -1$
then $Y = bX + c$ with prob 1.
for some constants $b<0$, c.

Remarks

The Role of Correlation $-1 \leq \mathrm{Corr}(X,Y) \leq 1$





if $X \perp\!\!\!\perp Y$ then
$\mathrm{Corr}(X,Y) = 0$

if $\mathrm{Corr}(X,Y) = 1 \Rightarrow Y = bX + c$, $b>0$.

if $\mathrm{Corr}(X,Y) = 1 \rightarrow Y = bX + c$, $b<0$.

Example: if $\mathrm{Corr}(X,Y) \approx 1 \iff Y \approx bX + c,\ b>0$

$\mathrm{Corr}(X,Y) \approx -1 \iff Y \approx bX + c,\ b<0.$

Let $X \sim \mathrm{Unif}(0,1)$

$Z \sim \mathrm{Unif}(0, 1/10)$, $Z \perp\!\!\!\perp X$

$Y = X + Z.$

draw a pic of the joint pdf!

Then
$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}XY - \mathbb{E}X\,\mathbb{E}Y \\
&= \mathbb{E}X(X+Z) - \mathbb{E}X(\mathbb{E}X + \mathbb{E}Z) \\
&= \mathbb{E}X^2 + \mathbb{E}XZ - (\mathbb{E}X)^2 - (\mathbb{E}X)(\mathbb{E}Z) \\
&= \mathrm{var}\,X \\
&= 1/12
\end{aligned}$$

Hence
$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{var}X\ \mathrm{var}Y}} = \frac{1/12}{\sqrt{1/12}\sqrt{1/12 + 1/1200}} = $$

$$= \frac{1}{\sqrt{1+\frac{1}{100}}} = \left(\frac{100}{101}\right)^{1/2} \approx 1$$

Remark $\mathrm{CORR}(X,Y) = 0$ does not imply $X \perp\!\!\!\perp Y$

- in fact it is possible that Y and X are strongly dependent on one another
- if that is the case, then the relationship between Y and X must be nonlinear!

Example.

Let $X \sim \mathrm{Unif}(-1,1)$
$Z \sim \mathrm{Unif}(0, 1/10)$, $Z \perp\!\!\!\perp X$.
$Y = X^2 + Z$.

draw a pic of the joint pdf!

Then
$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y) \\ &= \mathbb{E}\,X(X^2+Z) - (\mathbb{E}X)(\mathbb{E}X^2 + \mathbb{E}Z) \\ &= \mathbb{E}X^3 + \mathbb{E}X\,\mathbb{E}Z - (\mathbb{E}X)(\mathbb{E}X^2) - (\mathbb{E}X)(\mathbb{E}Z) \\ &= \mathbb{E}X^3 - \mathbb{E}X\,\mathbb{E}X^2 = 0 - 0 = 0. \end{aligned}$$

Bivariate Normal distribution

Def. A bivariate random vector $(X,Y) \in \mathbb{R}^2$ is said to have bivariate normal distribution and denoted by

$$Z = (X,Y) \sim N(\mu, \Sigma)$$

$\mu \in \mathbb{R}^2$

Σ positive definite, symmetric 2×2 matrix

if its joint pdf, for $z = (x,y)$

$$\boxed{\begin{aligned} f_{XY}(x,y) &:= f_Z(z) \\ &:= \frac{1}{\left((2\pi)^2 |\Sigma|\right)^{1/2}} \exp\left\{-\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)\right\}. \end{aligned}} \quad (*)$$

Remark.

Write $\mu = (\mu_1, \mu_2)^T$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.

then μ and Σ have the meaning as the mean vector and covariance matrix of Z.

$$\begin{cases} \mathbb{E} Z = \mu \\ \operatorname{cov} Z = \Sigma. \end{cases}$$

(see next page)

$\mathbb{E}Z = \mu$ means $\mathbb{E}\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \Leftrightarrow \begin{cases} \mathbb{E}X = \mu_1 \\ \mathbb{E}Y = \mu_2 \end{cases}$

$\operatorname{cov} Z = \Sigma$ means

$$\mathbb{E}[Z - \mathbb{E}Z][Z - \mathbb{E}Z]^T = \Sigma$$

$$\Leftrightarrow \mathbb{E}\underbrace{\begin{bmatrix} X - \mathbb{E}X \\ Y - \mathbb{E}Y \end{bmatrix}}_{2\times 1} \underbrace{\begin{bmatrix} X - \mathbb{E}X \\ Y - \mathbb{E}Y \end{bmatrix}^T}_{1\times 2} = \Sigma$$

$$\Leftrightarrow \mathbb{E}\begin{bmatrix} (X-\mathbb{E}X)^2 & (X-\mathbb{E}X)(Y-\mathbb{E}Y) \\ (Y-\mathbb{E}Y)(X-\mathbb{E}X) & (Y-\mathbb{E}Y)^2 \end{bmatrix}_{2\times 2} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} \Sigma_{11} = \operatorname{var} X \\ \Sigma_{22} = \operatorname{var} Y \\ \Sigma_{12} = \Sigma_{21} = \operatorname{cov}(X, Y). \end{cases}$$

This is the remarkable property of the normal dist.

Due to the identities:

if

$$f_{XY}(x,y) = f_Z(z) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp -\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)$$

then

$$\boxed{\begin{aligned} \iint_{\mathbb{R}\mathbb{R}} f_{XY}(x,y)\, dx\, dy &= 1 \\ \iint x\, f_{XY}(x,y)\, dx\, dy &= \mu_1 \\ \iint y\, f_{XY}(x,y)\, dx\, dy &= \mu_2 \\ \iint (z-\mu)(z-\mu)^T f_Z(z)\, dz &= \Sigma. \end{aligned}}$$

These identities follow from the identities for (univariate) normal distribution (see Sec 3.2b)

Translating the definition (*) into univariate forms:
(which you don't need to remember!)

$$|\Sigma| = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix} = \Sigma_{11}\Sigma_{22} - \Sigma_{21}\Sigma_{22}$$

$$\equiv \sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2$$

$$= \sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)$$

where $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} = \mathrm{CORR}(X,Y)$.

$$\mu = (\mu_1, \mu_2)^T \equiv (\mu_X, \mu_Y)^T.$$

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} \Sigma_{22} & -\Sigma_{21} \\ -\Sigma_{12} & \Sigma_{22} \end{pmatrix}$$

$$= \frac{1}{\sigma_X^2 \sigma_Y^2 (1 - \rho_{XY}^2)} \begin{pmatrix} \sigma_Y^2 & -\sigma_{XY} \\ -\sigma_{XY} & \sigma_X^2 \end{pmatrix}$$

$$= \frac{1}{1 - \rho_{XY}^2} \begin{pmatrix} \dfrac{1}{\sigma_X^2} & -\dfrac{\sigma_{XY}}{\sigma_X^2 \sigma_Y^2} \\ -\dfrac{\sigma_{XY}}{\sigma_X^2 \sigma_Y^2} & \dfrac{1}{\sigma_Y^2} \end{pmatrix}$$

$$(z-\mu)^T \Sigma^{-1} (z-\mu) = -\frac{1}{2(1-\rho_{XY}^2)} \left\{ \frac{1}{\sigma_X^2}(x-\mu_X)^2 - \frac{2\sigma_{XY}}{\sigma_X^2\sigma_Y^2}(x-\mu_X)(y-\mu_Y) + \frac{1}{\sigma_Y^2}(y-\mu_Y)^2 \right\}$$

$$2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)$$

Hence,

$$\boxed{f_{XY}(x,y) = \frac{1}{2\pi\,\sigma_X\sigma_Y(1-\rho_{XY}^2)^{1/2}} \times \exp\left\{ -\frac{1}{2(1-\rho_{XY}^2)} \left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \right] \right\}.}$$

Theorem if $\operatorname{Cov}(X,Y) = 0$ and $(X,Y) \sim$ bivariate normal then $X \perp\!\!\!\perp Y$.

Proof. $\operatorname{Cov}(X,Y) = 0 \Rightarrow \rho_{XY} = 0$

From (*), $f_{XY}(x,y)$ factorizes into a product of $g(x)\,h(y)$, so $X \perp\!\!\!\perp Y$ $\square$

Other important and useful facts of bivariate normal.

if $Z = (X,Y)^T \sim N(\mu, \Sigma)$, where $\begin{cases} \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \\ \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{cases}$

then

- the marginal of X are normal too:

$$X \sim N(\mu_1, \Sigma_{11}) \qquad (\text{i.e. } \mu_1 = \mu_X,\ \Sigma_{11} = \sigma_X^2$$
$$Y \sim N(\mu_2, \Sigma_{22}) \qquad \mu_2 = \mu_Y,\ \Sigma_{22} = \sigma_Y^2\ )$$

and $\operatorname{cov}(X,Y) = \Sigma_{12} = \Sigma_{21}$ $\left(\rho_{XY} = \rho_{YX} = \dfrac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}\right)$

- the conditional distributions are also normal:

$$Y \mid X = x \sim N\left(\mathbb{E}[Y \mid X=x], \operatorname{var}[Y \mid X=x]\right), \text{ where}$$

$$\mathbb{E}[Y \mid X=x] = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1) = \mu_Y + \rho_{XY}\frac{\sigma_Y}{\sigma_X}(x-\mu_X)$$

$$\operatorname{var}[Y \mid X=x] = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \Sigma_{22}\left(1 - \frac{\Sigma_{21}\Sigma_{12}}{\Sigma_{11}\Sigma_{22}}\right)$$
$$= \sigma_Y^2(1-\rho_{XY}^2).$$