In general, suppose that $X_1, \ldots, X_n$ are random variables measured on the same random experiment. For arbitrary constants $b_0, b_1, \ldots, b_n$, we have

$$\boxed{\mathbb{E}\left[b_0 + b_1 X_1 + \cdots b_n X_n\right] = b_0 + b_1 \mathbb{E}(X_1) + \cdots + b_n \mathbb{E}(X_n)}$$

Important: Linearity of expectations holds for any collection of random variables measured on the same random experiment.

Suppose that $X_1, \ldots, X_n$ are *independent* random variables measured on the same random experiment. We have

$$\mathbb{E}\left[X_1 X_2 \cdots X_n\right] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdot \cdots \cdot \mathbb{E}[X_n]$$

and

$$\operatorname{Var}(X_1 + X_2 + \cdots + X_n) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \cdots + \operatorname{Var}(X_n)$$

Example. Let Z_1 and Z_2 be two independent standard normal random variables and define $X = Z_1 + Z_2$ and $Y = Z_1 - Z_2$. Compute $N(0,1)$

$$\mathbb{E}\,X = \mathbb{E}[Z_1 + Z_2] = \mathbb{E}[Z_1] + \mathbb{E}[Z_2] = 0 + 0 = 0$$

$$\mathbb{E}\,Y = \mathbb{E}[Z_1 - Z_2] = \mathbb{E}[Z_1] - \mathbb{E}[Z_2] = 0 - 0 = 0.$$

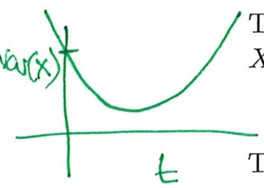
$$\operatorname{Var}(X) = \operatorname{Var}(Z_1 + Z_2) = \operatorname{Var}(Z_1) + \operatorname{Var}(Z_2) = 1 + 1 = 2.$$

$$\operatorname{Var}(Y) = \operatorname{Var}(Z_1 - Z_2) = \operatorname{Var}(Z_1) + \operatorname{Var}(-Z_2) = \operatorname{Var}(Z_1) + (-1)^2 \operatorname{Var}(Z_2) = 1 + 1 = 2$$

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}(X) \cdot \mathbb{E}(Y) \\ &= \mathbb{E}[(Z_1 + Z_2)(Z_1 - Z_2)] \\ &= \mathbb{E}[Z_1^2 + \cancel{Z_2 Z_1} - \cancel{Z_1 Z_2} - Z_2^2] = \mathbb{E}[Z_1^2] - \mathbb{E}[Z_2^2] = 1 - 1 = 0\end{aligned}$$

We know $\operatorname{Var}(Z_1) = 1$ because $Z_1 \sim N(0,1)$

$$\begin{aligned}\operatorname{Var}(Z_1) &= \mathbb{E}[Z_1^2] - (\mathbb{E}[Z_1])^2 \\ &= \mathbb{E}[Z_1^2] - (0)^2 \\ &= 1\end{aligned}$$



The size of the covariance between X and Y is constrain by the respective variances of X and Y. To see this, we note that for any $t \in \mathbb{R}$,

$$0 \leqslant \operatorname{Var}(X + tY) = \operatorname{Var}(X) + 2t\operatorname{Cov}(X,Y) + t^2 \operatorname{Var}(Y).$$

The above quadratic in t must be non-negative for all $t \in \mathbb{R}$. This leads to the inequality

$$|\operatorname{Cov}(X,Y)| \leqslant \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}.$$

Cauchy-Schwarz inequality

Definition. The **correlation** (or **correlation coefficient**) of X and Y is defined by

$$\varrho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

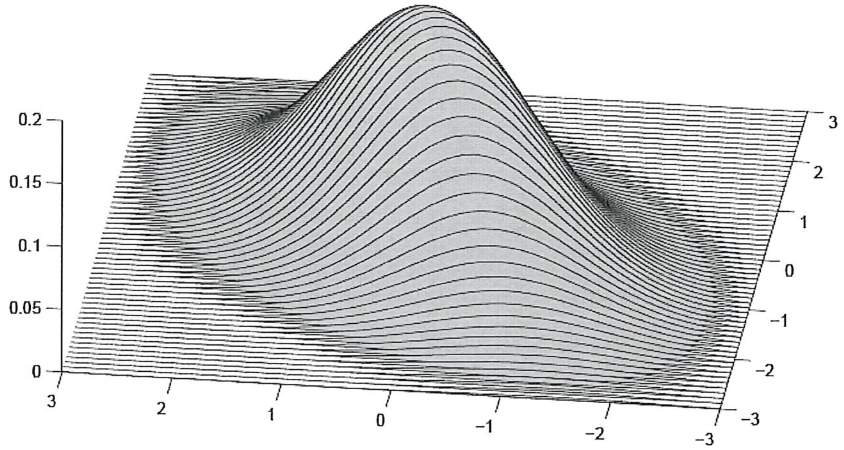
The random variables X and Y are said to be *positively correlated* or *negatively correlated* according as $\varrho(X,Y) > 0$ or $\varrho(X,Y) < 0$; otherwise, they are *uncorrelated.* The larger the value of $|\varrho(X,Y)|$ the more strongly correlated are X and Y.

Multivariate normal (Gaussian) distribution

Let (X,Y) be a pair of random variables with joint probability density function

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

where $\varrho \in (-1,1)$. Below is a plot of this joint pdf with $\varrho = 0.5$.



This is an important model; X and Y are said to have a (standard) *bivariate normal distribution.* To determine the marginal pdfs of X and Y, we first write

$$x^2 - 2\varrho xy + y^2 = (1-\varrho^2)x^2 + (y - \varrho x)^2.$$

$$= x^2 - \cancel{\rho^2 x^2} + y^2 - 2\rho y x + \cancel{\rho^2 x^2}$$

Then

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}\left(x^2 - 2\varrho xy + y^2\right)\right) dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}\left((1-\varrho^2)x^2 + (y-\varrho x)^2\right)\right) dy \\
&= \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)} \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \cdot \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right) dy}_{\text{pdf of } N(\rho x,\, 1-\rho^2)}
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

pdf of $N(\mu, \sigma^2)$

$N(0,1)$

So marginally X has a $\boxed{\text{standard normal dist.}}$. Similarly, the marginal distribution of Y is $\boxed{\text{standard normal}}$. From this we can also see that X and Y are independent (that is, $f_{X,Y}$ factorises as $f_X f_Y$) if and only if $\boxed{\rho = 0.}$.

The mean and variance of X and Y are

$$\mathbb{E}X = 0 \qquad \mathrm{Var}(X) = 1$$
$$\mathbb{E}Y = 0 \qquad \mathrm{Var}(Y) = 1$$

We can also evaluate the correlation between X and Y using the same trick that we used to evaluate the marginal pdfs of X and Y.

$$\begin{aligned}
\mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y) = \mathbb{E}(XY) \qquad & \text{[since the marginals of } X \text{ and } Y \text{ are standard normal]} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\frac{1}{2\pi\sqrt{1-\varrho^2}}\exp\left(-\frac{1}{2(1-\varrho^2)}\left((1-\varrho^2)x^2+(y-\varrho x)^2\right)\right)\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\left[\int_{-\infty}^{\infty} y \underbrace{\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\left(-\frac{(y-\rho x)^2}{2(1-\rho^2)}\right)}_{\text{pdf of } N(\rho x, 1-\rho^2)} dy\right]dx \\
&= \int_{-\infty}^{\infty} \rho x \cdot x \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right)dx \\
&= \rho\int_{-\infty}^{\infty} x^2\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)dx = \rho\times 1 = \rho
\end{aligned}$$

The general bivariate normal distribution is only very slightly more complicated: X and Y are said to have a bivariate normal distribution if its joint pdf $f_{X,Y}(x,y)$ has the form

$$\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\varrho^2}}\exp\left(-\frac{1}{2(1-\varrho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\varrho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right).$$

The marginal distributions are both normal: $X \sim \mathsf{Normal}(\mu_X, \sigma_X^2)$ and $Y \sim \mathsf{Normal}(\mu_Y, \sigma_Y^2)$. Also ϱ is is the correlation of (X,Y), and X and Y are independent if and only if $\varrho = 0$.

A random vectors $\mathbf{X} := (X_1, \ldots, X_n)$ has a multivariate Normal distribution if the joint pdf has the form

Σ covariance matrix

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\det(\Sigma)}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

$\Sigma_{ij} = \mathrm{cov}(X_i, X_j)$

$\Sigma_{ii} = \mathrm{cov}(X_i, X_i) = \mathrm{var}(X_i)$

μ mean vector

$\mu_i = \mathbb{E}(X_i)$

Σ $(n\times n)$ matrix

μ $(n\times 1)$ vector

$\underline{x}$ $(n\times 1)$ vector

$\Sigma_{ij} = \Sigma_{ji}$ since $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$

where

$$\mathbb{E}(X_i) = \mu_i, \qquad \text{and} \qquad \text{Cov}(X_i, X_j) = \Sigma_{ij}$$

In particular, if Σ is diagonal, then the $X_1, \ldots, X_n$ are independent random variables with $X_i \sim \mathsf{Normal}(\mu_i, \Sigma_{ii})$.

For us, the most important property of the multivariate Normal distribution is its behaviour under linear transformations.

very important

> Suppose $\mathbf{X} := (X_1, \ldots, X_n)'$ has a multivariate Normal distribution. Let $\boldsymbol{a} \in \mathbb{R}^m$ and B is an $(m \times n)$ matrix (with $m \leqslant n$). If $\boldsymbol{X} \sim \mathsf{Normal}(\boldsymbol{\mu}, \Sigma)$, then the random vector $Y := \boldsymbol{a} + B\boldsymbol{X}$ has a $\mathsf{Normal}(\boldsymbol{a} + B\boldsymbol{\mu}, B\Sigma B^T)$.

Example: Suppose that $X_1 \sim \mathsf{Normal}(-1, 2)$ and $X_2 \sim \mathsf{Normal}(1, 3)$ are independent. What is the distribution of $Y = 3 + 2X_1 - X_2$?

Observe

$$Y = \overset{a}{3} + \overset{B}{\begin{bmatrix} 2 & -1 \end{bmatrix}} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

and

$$\mathbf{X} \sim \text{Normal}\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \right)$$

so

$$Y \sim \text{Normal}\left(\overset{a}{3} + \overset{B}{\begin{bmatrix} 2 & -1 \end{bmatrix}} \overset{\mu}{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}, \overset{B}{\begin{bmatrix} 2 & -1 \end{bmatrix}} \overset{\Sigma}{\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}} \overset{B^T}{\begin{bmatrix} 2 \\ -1 \end{bmatrix}} \right)$$

$$\begin{bmatrix} 4 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$Y \sim \text{Normal}(0, 11).$$

Conditional probability density functions and conditional expectation

Recall the definition of conditional probability mass function for discrete random variables;

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

provided $f_Y(y) = \mathbb{P}(Y = y) > 0$.

For continuous random variables (X, Y) we can similarly define the **conditional probability density function** of X given $\{Y = y\}$, denoted by $f_{X|Y}(x|y)$, ;

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

when $f_Y(y) > 0$.

Note that if X and Y are independent, then

> $$f_{X|Y}(x|y) =$$
>
> when $f_Y(y) > 0$.