Engineering Mathematics 2B

Module 15: Introduction to Probability

Nick Polydorides

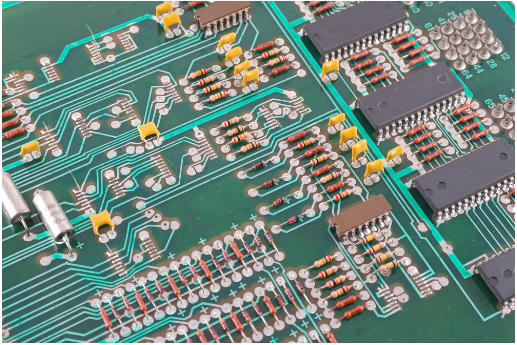
School of Engineering

THE UNIVERSITY of EDINBURGH

Module 15 contents

Independent random variables

Electrical engineers rely on independent random variables to model the variability in the behaviour of individual electronic components. These models are essential for designing reliable and efficient circuits, ensuring that they can withstand fluctuations in temperature, voltage, and other environmental factors.



Joint normal variables

Random variables exist in isolation or as a group.

Two or more random variables are said to be joint if they occur simultaneously. Their relation may be causal or not!

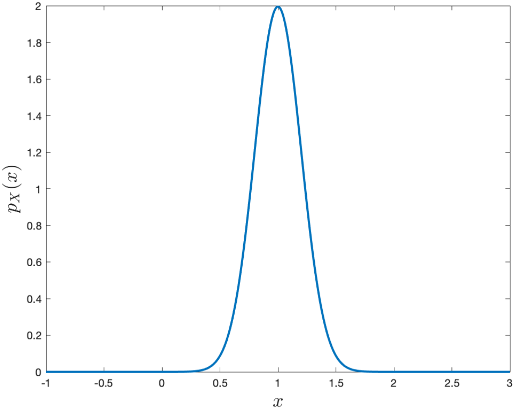
"rolling 2 dice together",

"daily study and Tik-Tok times",

"daily exercise time and calorific intake",...

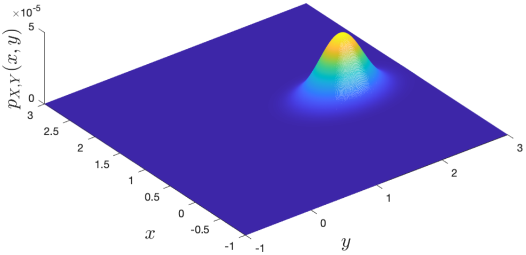
SOS: We will discuss joint normal (Gaussian) random variables but the ideas extend to other continuous and discrete types.

Gaussian siblings



$p_X(x)$ is $\mathcal{N}(\mu_X, \sigma_X^2)$ with $\mu_X = 1$ and $\sigma_X = 0.2$.

Gaussian siblings



$$(X,Y) \sim \mathcal{N}\left(\begin{bmatrix}\mu_X \\ \mu_Y\end{bmatrix}, \begin{bmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2\end{bmatrix}\right)$$

with $\mu_X = 1$, $\mu_Y = 2$, $\sigma_X = 0.2$, $\sigma_Y = 0.3$ and $\rho = 0$.

Joint Density Functions

For X and Y be continuous random variables, $p_{X,Y}$ is the **joint density function** of X and Y used to compute probabilities

$$\mathbb{P}\big((X,Y) \in R\big) = \iint_R p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

where R is a region on the xy plane.

The joint cumulative distribution function of X and Y is defined as

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x \cap Y \leq y) = \int_{-\infty}^{y}\int_{-\infty}^{x} p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

As with single variable PDFs we require

$$p_{X,Y}(x,y) \geq 0, \quad -\infty < x, y < +\infty$$

and

$$\iint_{-\infty}^{+\infty} p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = 1$$

Joint random variables

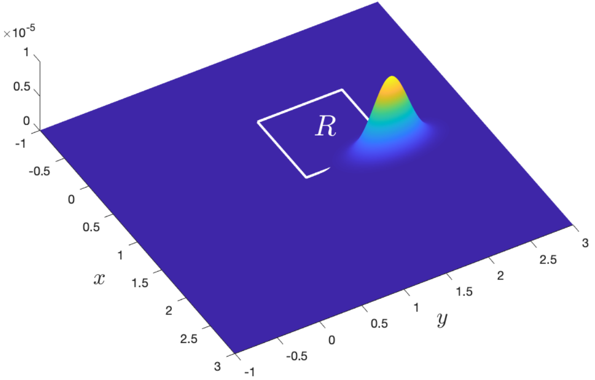
If X, Y have joint PDF $p_{X,Y}(x, y)$ what is the probability of $0 \leq X \leq 1$ and $1 \leq Y \leq 2$?

Recall that for any pair of inputs (x, y) anywhere on the xy plane, the density $p_{X,Y}(x, y)$ is a scalar function that yields a non-negative likelihood value at that point.

$$\begin{aligned}\mathbb{P}(0 \leq X \leq 1 \cap 1 \leq Y \leq 2) &= \iint_R p_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x \\ &= \int_0^1 \int_1^2 p_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x\end{aligned}$$

Recall: Double integrals!

Integral under the joint Gaussian



$$\mathbb{P}(0 \leq X \leq 1 \cap 1 \leq Y \leq 2) = \iint_R p_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y$$

Matrix algebra

To understand the joint density of bivariate normal variables involves some matrix algebra skills.

A symmetric matrix with real values whose diagonal is non-zero has an inverse. The 2×2 covariance matrix Σ is symmetric with correlation ρ as $-1 < \rho < 1$, and σ_X^2, σ_Y^2 the individual variances

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2 \end{bmatrix} \text{ then } \Sigma^{-1} = \frac{1}{|\Sigma|} \begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_Y\sigma_X & \sigma_X^2 \end{bmatrix}$$

where the determinant is $|\Sigma| = \sigma_X^2 \sigma_Y^2 (1 - \rho^2)$.

For a column vector $\boldsymbol{\mu} = \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}$ the transpose $\boldsymbol{\mu}^\top$ is the same vector as a row vector.

Alternative expressions for the bivariate normal

Let X, Y be joint normal $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2 \end{bmatrix}$. A 'tidy' matrix expression for the joint is

$$p_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left\{ -\frac{1}{2} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right)^\top \Sigma^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right) \right\}$$

It can be shown that this can also be expressed as

$$\begin{aligned} p_{X,Y}(x, y) = & \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \\ & \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right.\right. \\ & \qquad \left.\left. - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right\} \end{aligned}$$

Covariance of joint variables

Although for $X \sim p_X(x)$, the variance $\mathrm{Var}(X) = \sigma_X^2$ tells us how X varies about its mean (expected) value, in joint variables that's not enough.

The way the pair (X, Y) varies about its mean $\boldsymbol{\mu}$ is captured by the 2×2 covariance matrix Σ

$$\Sigma = \begin{bmatrix} \mathrm{Cov}(X,X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Cov}(Y,Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{bmatrix}$$

where

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] := \sigma_{XY}. \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] := \sigma_{YX} = \mathrm{Cov}(Y,X) \end{aligned}$$

Notice:

$$\mathrm{Cov}(X,X) = \mathrm{Var}(X) = \sigma_X^2, \quad \mathrm{Cov}(Y,Y) = \mathrm{Var}(Y) = \sigma_Y^2$$

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X,Y)$$

Covariance of joint normal variables

By comparison with the bivariate **normal** template

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2 \end{bmatrix}$$

it is easy to see that

$$\sigma_{XY} = \rho\sigma_X\sigma_Y = \sigma_{YX},$$

where $-1 \leq \rho \leq 1$.

Correlation of joint variables

For (X, Y) joint normal with $\text{Var}(X), \text{Var}(Y) > 0$ the **correlation coefficient** of X and Y is

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \, \sigma_Y}, \text{ with } -1 \leq \rho(X, Y) \leq 1.$$

The value of ρ tells us how close to being linearly dependent X and Y are, i.e. think of realisations (x_i, y_i) being points on the xy axes.

In the extreme cases $\rho = \pm 1$ "knowing the one implies knowing the other". (degenerative joint case - (x_i, y_i) form a line)

In the case where $\rho = 0$, "knowing the one tells us nothing about the other", as this is the case where the two don't vary together, $\text{Cov}(X, Y) = 0$. ((x_i, y_i) do not fit a line)

The marginals

The marginal(s) of a joint distribution allows us to "say something about one of the variables, regardless of the other(s)".

For (X, Y) joint normal with means μ_X, μ_Y, and covariance Σ then

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\mathrm{d}y, \quad \text{and} \quad p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\mathrm{d}x,$$

so the marginal for one variable is found by integrating the 2D joint along the other axis.

From these we can get to the individual expectations as

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x) p_X(x)\mathrm{d}x,$$

and similarly for $\mathbb{E}[f(Y)]$.

The marginals of the bivariate normal

If

$$
\begin{aligned}
p_{X,Y}(x,y) = & \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \\
& \exp\Big\{-\frac{1}{2(1-\rho^2)}\Big[\Big(\frac{x-\mu_X}{\sigma_X}\Big)^2 + \Big(\frac{y-\mu_Y}{\sigma_Y}\Big)^2 \\
& - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\Big]\Big\}
\end{aligned}
$$

then, no matter of what the value of ρ is, (try to show that)

$$
p_X(x) = \int_{-\infty}^{+\infty} p_{X,Y}(x,y)\mathrm{d}y = \underbrace{\frac{1}{\sqrt{2\pi}\sigma_X}\exp\Big\{-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2}\Big\}}_{\mathcal{N}(\mu_X,\sigma_X^2)}
$$

$$
p_Y(y) = \int_{-\infty}^{+\infty} p_{X,Y}(x,y)\mathrm{d}x = \underbrace{\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\Big\{-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}\Big\}}_{\mathcal{N}(\mu_Y,\sigma_Y^2)}
$$

Conditional distributions of joint continuous variables

Let X and Y be continuous with joint $p_{X,Y}(x, y)$.

The **conditional** PDF of X given $Y = y$ for a fixed y and $p_Y(y) > 0$ is

$$p_{X|Y}(x|y) = \mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x \cap Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)},$$

where $p_Y(y)$ can be obtained by marginalising the joint.

Graphically, the conditional $p_{X|Y}(x|y)$ cuts through the joint $p_{X,Y}(x, y)$ at the line $Y = y$, dividing by a normalisation constant that makes the integral along that line equal to 1.

Conditional distributions

$p_{X|Y}(x|y)$ gives the likelihood of X taking a single value conditioned on Y taking a single value. This leads to the probability that X falls inside a one-dimensional interval $[x_1, x_2]$ if $Y = y$ as

$$\mathbb{P}(x_1 \leq X \leq x_2 \,|\, Y = y) = \int_{x_1}^{x_2} p_{X|Y}(x|y)\mathrm{d}x$$

Similarly, one can define a conditional for Y conditioned on a specific outcome for X,

$$p_{Y|X}(y|x) = \mathbb{P}(Y = y \,|\, X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}, \quad p_X(x) > 0.$$

is the conditional PDF of Y given $X = x$ for a fixed x.

Conditional expectation

Let X and Y be random variables. The conditional expectation (or conditional mean) of Y given $X = x$ is denoted by $\mathbb{E}[Y|X = x]$.

$\mathbb{E}[Y|X = x]$ is defined to be the expectation of the conditional distribution of Y given $X = x$.

If Y has a continuous conditional distribution given $X = x$ with conditional PDF $p_{Y|X}(y|x)$, then

$$\mathbb{E}[Y|X = x] = \int_{-\infty}^{+\infty} y\, p_{Y|X}(y|x) \mathrm{d}y$$

Notice: $\mathbb{E}[Y|X]$ is a random variable whose realisation when $X = x$ is $\mathbb{E}[Y|X = x]$.

Independent random variables

Joint random variables can be independent or not.

Let X and Y have joint PDF $p_{X,Y}(x, y)$. X and Y are said to be **independent** if and only if their joint is the product of their marginals

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad \forall x, y$$

or

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

If X and Y are independent then $f(X)$ and $g(Y)$ are independent, no matter what the functions f and g are. Effectively, if f and g are functions of **independent** variables X and Y then

$$\mathbb{E}[f(X)\, g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

Independent random variables

When $n > 2$ random variables $X_1, \ldots, X_n$ are (mutually) independent, then any pair of them are independent (the reverse doesn't always hold).

Under mutual independence, the following *nice* properties hold:

- $\mathbb{E}[X_1 X_2 \ldots X_n] = \mathbb{E}[X_1]\mathbb{E}[X_2] \ldots \mathbb{E}[X_n]$
- $\mathrm{Var}(X_1 \pm \ldots \pm X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$

Irrespective of independence

$$\mathbb{E}[X_1 \pm \ldots \pm X_n] = \mathbb{E}[X_1] \pm \ldots \pm \mathbb{E}[X_n]$$

Independent random variables

A set of random variables are said to be **independent and identically distributed** (iid) if each random variable has the same probability distribution as the others and are all mutually independent. For normal variables for example

$$X_1, X_2, \ldots, X_n \text{ are iid } \Rightarrow X_i \sim \mathcal{N}(\mu, \sigma^2),\ i = 1, \ldots, n$$

and

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] = \mu^2$$

If joint (X_1, X_2) are independent, then their covariance and correlation is zero

$$\sigma_{X_1 X_2} = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0 \ \Rightarrow\ \rho(X_1, X_2) = 0$$

Bivariate independent normal

The PDF $p_{X,Y}(x,y)$ of independent X and Y can be found by substituting $\rho = 0 = \sigma_{XY}$ into the template for the joint normal

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{ -\frac{1}{2}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right)^\top \Sigma^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right) \right\}$$

where $\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$, while $\Sigma = \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{bmatrix}$ and $\Sigma^{-1} = \begin{bmatrix} \sigma_X^{-2} & 0 \\ 0 & \sigma_Y^{-2} \end{bmatrix}$ are diagonal matrices.

Alternatively,

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{ -\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\},$$

and by the properties of the exponential $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

Dependence versus Correlation

Recall that we have defined the **covariance** between two random variables X and Y

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

If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$, but the reverse does not always hold, i.e. "if they are uncorrelated that doesn't make them independent".

If $\mathrm{Cov}(X,Y) = 0$ then X and Y are **uncorrelated** (otherwise they are called correlated), but not necessarily independent, i.e. "they may have a nonlinear correlation".

We show this using an example on the next slide.

Example

Let $X \sim \mathcal{N}(0,1)$ and define $Y = X^2$. We will show that in this case $\mathrm{Cov}(X,Y) = 0$ although X and Y are not independent.

Indeed, $\mathbb{E}[X] = 0 \implies \mathbb{E}[X]\mathbb{E}[Y] = 0$ and thus

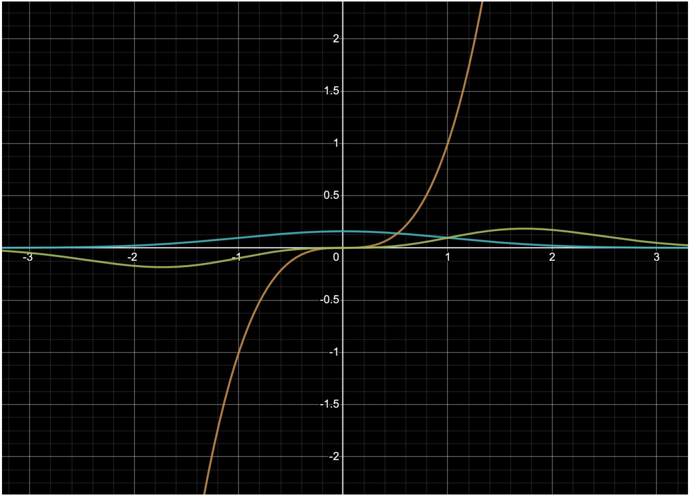
$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] = \mathbb{E}[X^3],$$

but

$$\begin{aligned} \mathbb{E}[X^3] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^3 \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} x^3 \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x + \int_{0}^{+\infty} x^3 \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x \right) = 0 \end{aligned}$$

as the two integrals cancel out due to their integrands being skew symmetric. From this: "X and Y are uncorrelated".

Example cont



Graphs of $\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$, x^3 and their product, justifying why the integral over the x axis equals zero.

Example cont

Recall that by definition X and Y are independent if

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

hence checking for

$$\mathbb{P}(X \leq -1, Y \leq 1) = \mathbb{P}(X \leq -1, X^2 \leq 1) = 0$$

since any number that is less than -1 has a square that's more than 1.

At the same time

$$\mathbb{P}(X \leq -1)\mathbb{P}(Y \leq 1) = \mathbb{P}(X \leq -1)\mathbb{P}(X^2 \leq 1) > 0.$$

hence the criterion for independence doesn't hold, and thus X and Y are not independent despite being uncorrelated. (Recall correlation is a linear relation)

Formulas

General form of bivariate Gaussian. Set $\rho = 0$ if X and Y are independent. $(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_Y\sigma_X & \sigma_Y^2 \end{bmatrix}$

$$p_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left\{ -\frac{1}{2}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right)^{\top} \Sigma^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \boldsymbol{\mu} \right) \right\}$$

- $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ ($= 0$ iff independent)
- $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X, Y)$
- $\rho(X, Y) = \mathrm{Cov}(X, Y) / \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}$
- $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \quad p_Y > 0$
- $\mathbb{E}[Y|X = x] = \int_{-\infty}^{+\infty} y\, p_{X|Y}(y|x) \mathrm{d}y$
- $\mathbb{E}[X_1 X_2 \ldots X_n] = \mathbb{E}[X_1]\mathbb{E}[X_2] \ldots \mathbb{E}[X_n]$ (iff independent)
- $\mathbb{E}[X_1 \pm \ldots \pm X_n] = \mathbb{E}[X_1] \pm \ldots \pm \mathbb{E}[X_n]$
- $\mathrm{Var}(X_1 \pm \ldots \pm X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$ (iff uncorrelated or if independent)

Main outcomes of module 15

You **MUST** know:

1. What is meant for two random variables to be joint.
2. To compute the expectation, covariance and correlation of joint normal variables.
3. The marginals of joint normal variables.
4. The conditional distribution.
5. The concept of independence in random variables.