Warm-up sessions

Copernicus Master on Digital Earth

Probabilities and Statistics for data science

Prof. Nicolas Courty
ncourty@irisa.fr

Definitions

Definition of probability

Event space Ω All possible events together from a given experience.

Definition of $\mathrm{P}\left(\mathrm{A}\right)$ Let A be a set of events included in Ω,

$$\mathrm{P}\left(\mathrm{A}\right) = \lim_{\mathrm{n}\to\infty} \frac{\mathrm{n}(\mathrm{A})}{\mathrm{n}}$$

with

- n the number of experiments performed,
- $\mathrm{n}(\mathrm{A})$ the number of experiments where A was performed.

Example, 6-sided dice

- $\Omega = \mathrm{faces}{:}1, 2, 3, 4, 5, 6$
- If dice not pipped, then $\mathrm{P}\left(\mathrm{k}\right) = 1/6, \quad \forall \mathrm{k} \in 1, \ldots, 6$.

Axioms of probabilities

- First axiom If $\mathrm{A} \in \Omega$ then

$$0 \leq \mathrm{P}(\mathrm{A}) \leq 1$$

- Second axiom

$$\mathrm{P}(\Omega) = 1 \quad \mathrm{P}(\mathrm{O}) = 0$$

  with O the empty set

- Union and intersection If $\mathrm{A} \in \Omega, \ \mathrm{B} \in \Omega$, then

$$\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B}) - \mathrm{P}(\mathrm{A} \cap \mathrm{B})$$

- If $\mathrm{A} \cap \mathrm{B} = \mathrm{O}$ then

$$\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B})$$

Random Variable

Definition

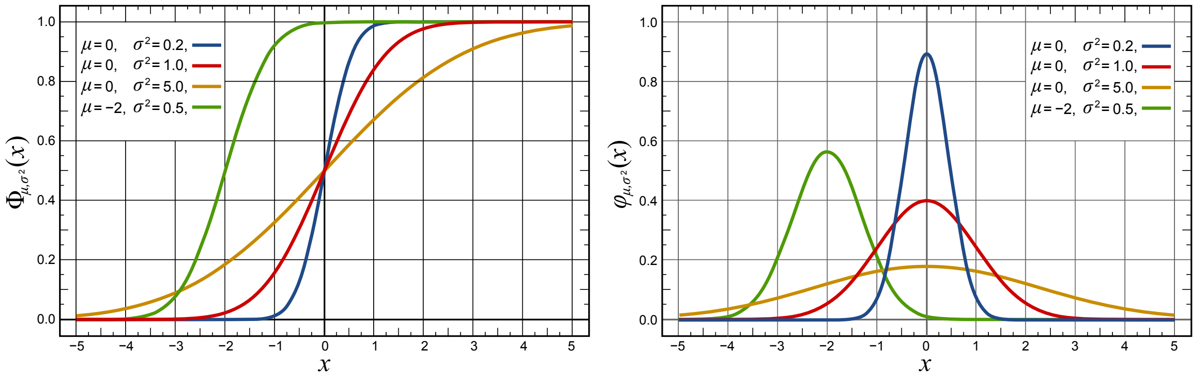
A random variable is a (real) X_ω number whose value is is determined by the ω result of a randomized experiment.

Example: 6-sided dice

- The random event ω is the appearance of a face.
- An integer 1 to 6 is associated to each face.

Distribution function

Distribution function and derivative



- Distribution function (a.k.a. Cumulative distribution function) The $\mathrm{F_X}$ distribution function of a random variable (r.v.) X is defined as the probability that X is less than or equal to x,

$$\mathrm{F_X(x) = P(X \leq x)}$$

- Probability density function. It is defined as the derivative of the distribution function,

$$\mathrm{p(x) = \frac{dF(x)}{dx}}$$

Properties

- Properties of the distribution function

$$F_X(-\infty) = 0, \quad F_X(\infty) = 1$$
$$0 \leq F_X(x) \leq 1$$
$$P(x_1 \leq x \leq x_2) = F_X(x_2) - F_X(x_1)$$

- Properties of the probability density

$$p(x) \geq 0 \quad \int_{-\infty}^{+\infty} p(x)dx = 1$$
$$P(x \leq x_1) = F_X(x_1) = \int_{-\infty}^{x_1} p(x)dx$$
$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} p(x)dx$$

Moments of a random variable

Definition of a moment. The moment $g(x)$ of one random variable is given by its expectation

$$E(g(x)) = \int_{-\infty}^{+\infty} g(x)p(x)dx$$

Whenever $g(x) = x^m$, we refer to this quantity as the moment of order m,

$$\text{Moment of order 1} \qquad m_X = E(X) = \int_{-\infty}^{+\infty} xp(x)dx$$

$$\text{Moment of order 2} \qquad m_X^{(2)} = E(X^2) = \int_{-\infty}^{+\infty} x^2p(x)dx$$

The 1st order moment is also often called average (or mean).

- Property linearity of expectation

$$E(X+Y) = E(X) + E(Y), \qquad E(kX) = kE(X)$$

For k a constant.

Moments of a random variable

Definition of the variance. The variance is the expectation of the square of the deviations from mean value $m_X = E(X)$,

$$
\begin{aligned}
\sigma_X^2 &= E\left((X - m_X)^2\right) = \int_{-\infty}^{+\infty} (x - m_X)^2 p(x) dx \,, \\
\sigma_X^2 &= E\left(X^2\right) - E\left(X\right)^2
\end{aligned}
$$

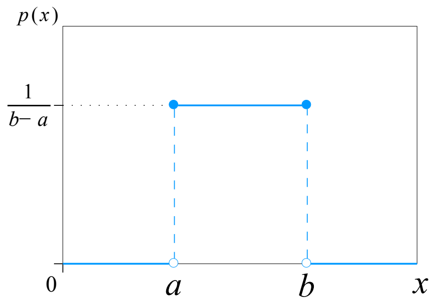
The notion of standard deviation σ is also often used,

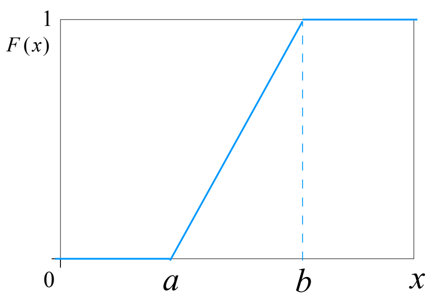
$$
\sigma_X = \sqrt{\sigma_X^2} \,.
$$

- Incomplete characterization. Incomplete characterization of a random variable by its mean and variance.

Examples of laws

Uniform Law $\mathcal{U}(a, b)$





Probability density

$$p(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{elsewhere} \end{cases}$$

Distribution function

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & \text{if } x \in [a, b], \\ 1 & x' > b \end{cases}$$

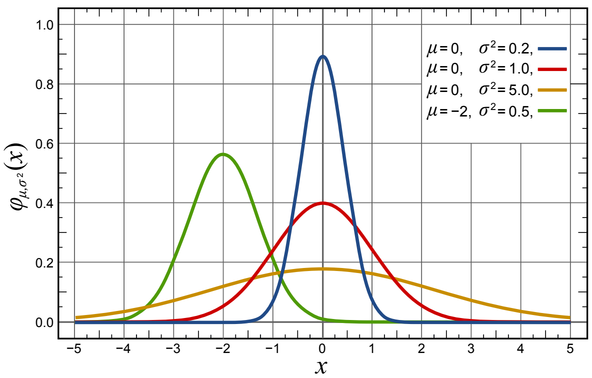
Expectation:

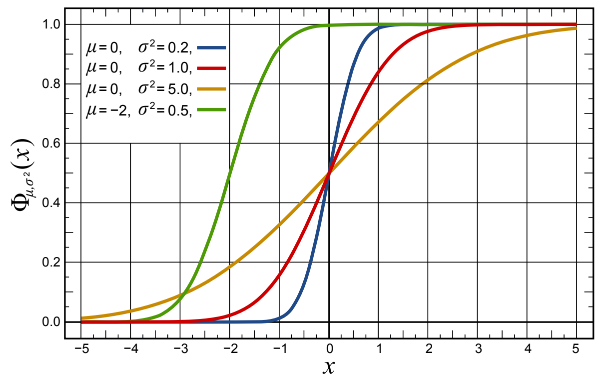
$$m_X = E(X) = \frac{b+a}{2}$$

Variance:

$$\mathrm{Var}(X) = \frac{1}{12}(b-a)^2$$

Normal Law $\mathcal{N}(\mu, \sigma^2)$





Probability density

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Distribution function

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

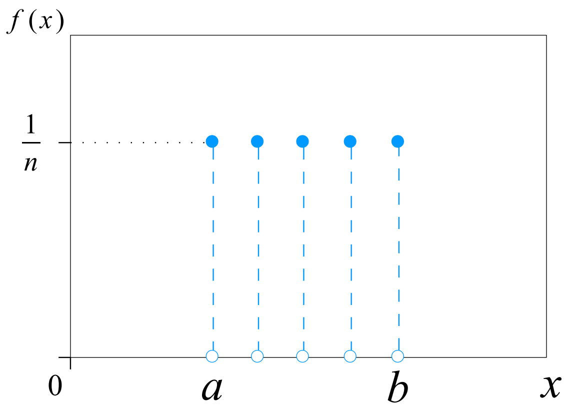
Expectation:

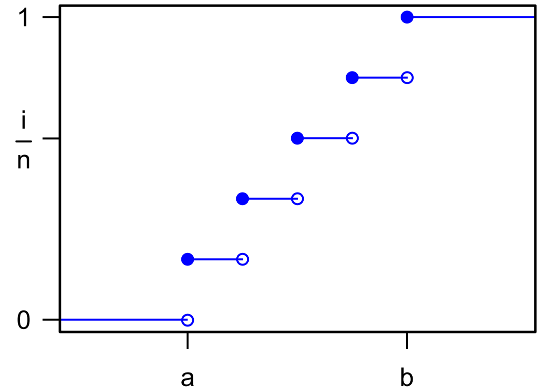
$$m_X = E(X) = \mu$$

Variance:

$$\operatorname{Var}(X) = \sigma^2$$

Empirical Law $\mathcal{U}(x_1, \ldots, x_n)$





Probability density

$$p(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i)$$

Expectation:

$$m_X = E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Distribution function

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} \ _{x\geq x_i}$$

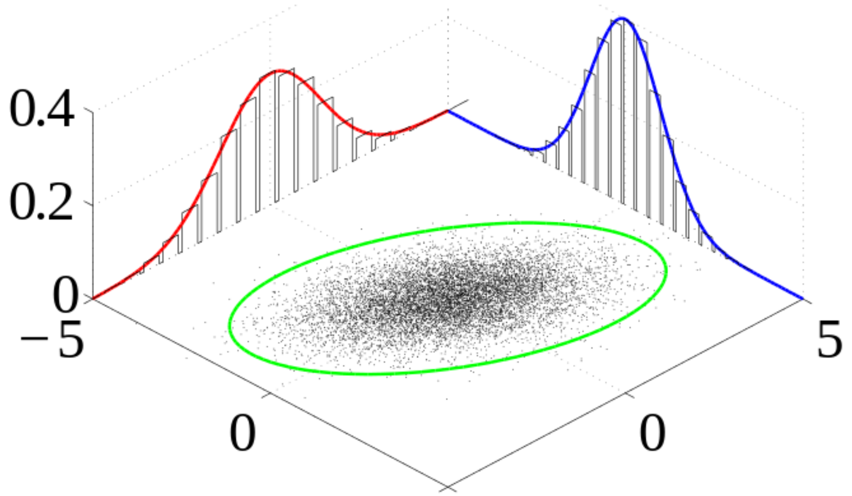
Variance:

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_X)^2$$

System of random variables

- We consider the case where several random variables are available simultaneously.

- We need to model jointly those variables



- When those variables $X_1, X_2, \ldots, X_d$ are given, we can model them by a vector of random variables $\mathbf{X} \in \mathbb{R}^d$

Joint probability density

Cumulative distribution function. Let X and Y be two r.v. then,

$$\mathrm{F}(\mathrm{x},\mathrm{y}) = \mathrm{P}(\mathrm{X} \leq \mathrm{x}, \mathrm{Y} \leq \mathrm{y})$$

Properties

$$0 \leq \mathrm{F}(\mathrm{x},\mathrm{y}) \leq 1$$

$$\mathrm{F}(-\infty,-\infty) = 0$$

$$\mathrm{F}(\infty,\infty) = 1$$

Joint probability density. Let X and Y be two r.v. then,

$$\mathrm{p}(\mathrm{x},\mathrm{y}) = \frac{\partial^2 \mathrm{F}(\mathrm{x},\mathrm{y})}{\partial \mathrm{x} \partial \mathrm{y}}$$

$$\mathrm{p}(\mathrm{x}) = \int \mathrm{p}(\mathrm{x},\mathrm{y}) \mathrm{d}\mathrm{y}$$

$$\mathrm{p}(\mathrm{y}) = \int \mathrm{p}(\mathrm{x},\mathrm{y}) \mathrm{d}\mathrm{x}$$

$\mathrm{p}(\mathrm{x})$ and $\mathrm{p}(\mathrm{y})$ are called marginal laws.

Properties

$$\mathrm{p}(\mathrm{x},\mathrm{y}) \geq 0$$

$$\int \mathrm{p}(\mathrm{x},\mathrm{y}) \mathrm{d}\mathrm{x}\mathrm{d}\mathrm{y} = 1$$

$$\mathrm{p}(\mathrm{A},\mathrm{B}) = \mathrm{P}(\mathrm{x} \in \mathrm{A}, \mathrm{y} \in \mathrm{B}) = \int_{\mathrm{A}} \int_{\mathrm{B}} \mathrm{p}(\mathrm{x},\mathrm{y}) \mathrm{d}\mathrm{x}\mathrm{d}\mathrm{y}$$

Joint probability density

Conditional probability

- Joint law $p(x, y)$.
- Probability of one of the variables knowing the value of the second.
- Notation: $p(x|y)$.

Bayes' theorem

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

$$p(y|x) = \frac{p(x, y)}{p(x)}$$

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y)$$

Covariance and correlation

- Moments of a joint law

$$E(g(x,y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)p(x,y)dxdy$$

- Correlation

$$R_{XY} = E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyp(x,y)dxdy$$

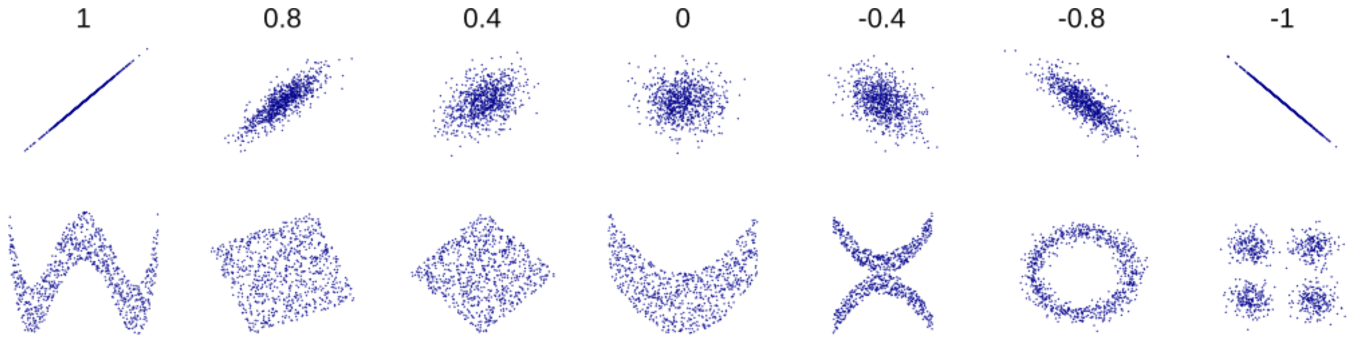
- Covariance

$$C_{XY} = \sigma_{XY} = E\left((X - m_X)(Y - m_Y)\right)$$
$$C_{XY} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - m_X)(y - m_Y)p(x,y)dxdy$$

- Correlation coefficient

$$r_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y}$$

Independence and correlation



Covariance and Correlation

$$R_{XY} = E(XY) = C_{XY} + m_X m_Y$$

Independence

- Two r.v. X and Y are independent if

$$p(x, y) = p(x)p(y)$$

.

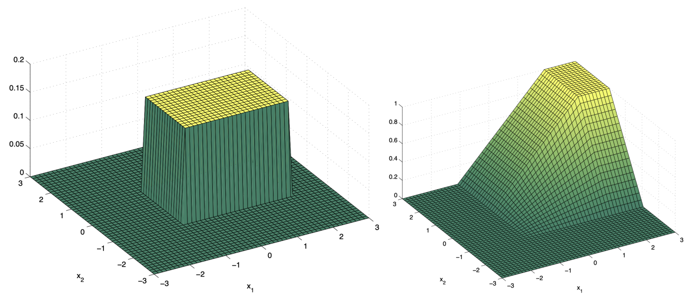
- If the variables are independent then

$$R_{XY} = m_X m_Y \qquad \text{and} \qquad C_{XY} = 0.$$

Examples of multivariate distribution function

Multivariate Uniform Law

- $X \sim U(a_x, b_x)$ and $Y \sim U(a_y, b_y)$
- $\mathbf{X} = [X, Y]^\top$



Probability density

$$p(x,y) = \begin{cases} \frac{1}{S} & \text{if } x \in [a_x, b_x] \text{ and } y \in [a_y, b_y] \\ 0 & \text{else} \end{cases}$$

Surface $S = (b_x - a_x)(b_y - a_y)$

Expectation:

$$\mathbf{m}_X = E(\mathbf{X}) = \begin{bmatrix} \frac{b_x + a_x}{2} \\ \frac{b_y + a_y}{2} \end{bmatrix}$$

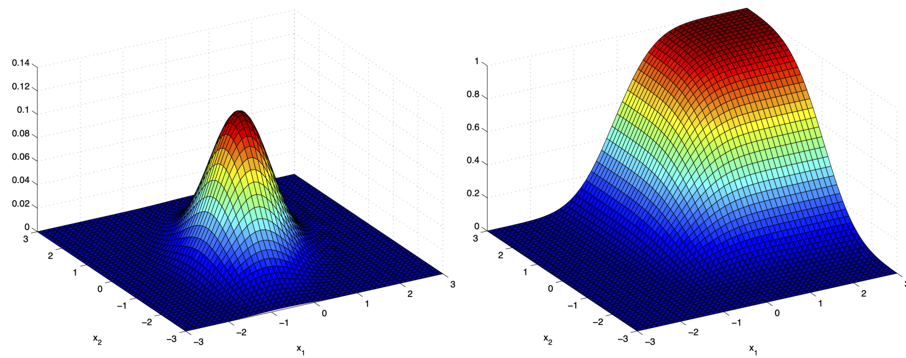
Variance:

$$\mathrm{Cov}(\mathbf{X}) = E((\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X) = \begin{bmatrix} \mathrm{Var}(X) & 0 \\ 0 & \mathrm{Var}(Y) \end{bmatrix}$$

Examples of multivariate distribution function

Multivariate Gaussian Law

- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$



Probability density

$$\mathrm{p}(\mathrm{x}, \mathrm{y}) = \mathrm{K} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Coefficient $\mathrm{K} = \dfrac{1}{(2\pi)^{\mathrm{N}/2} |\boldsymbol{\Sigma}|^{1/2}}$

Expectation:

$$\mathbf{m}_{\mathrm{X}} = \mathrm{E}(\mathbf{X}) = \boldsymbol{\mu}$$

Covariance:

$$\begin{aligned} \mathrm{Cov}(\mathbf{X}) &= \mathrm{E}((\mathbf{X} - \mathbf{m}_{\mathrm{X}})(\mathbf{X} - \mathbf{m}_{\mathrm{X}}) \\ &= \boldsymbol{\Sigma} \end{aligned}$$

Practical session

Let's practice now:

- Numpy, scipy and Pandas for manipulating dataframes
- Explore basic statistics computation using python

The end.