Probability Theory

A. Tsamoglou, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 4.2.1

Artificial Intelligence &
Information Analysis Lab

Probability Space

Probability definitions

- $\mathcal{S}$: set of every possible outcomes of an experiment.
- ***Axiomatic probability*** P: it is a measure in a ***Borel field*** $\mathcal{F}$ of every possible subsets of $\mathcal{S}$.
- ***Probability space***: triad $\{\mathcal{S}, \mathcal{F}, P\}$.
- Properties :
  - $\mathcal{A} \in \mathcal{F} \Rightarrow P(\mathcal{A}) \geq 0,$
  - $P(\mathcal{S}) = 1\,,$
  - If $\mathcal{A} \cap \mathcal{B} = \{\emptyset\}$, then $P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$.

Probability Space

- Probability properties:

$$P\{\emptyset\} = 0,$$
$$P(\bar{\mathcal{A}}) = 1 - P(\mathcal{A}),$$

$$P(\mathcal{A} \cup \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B} - \mathcal{A}) \Rightarrow$$
$$P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \cap \mathcal{B}) \leq P(\mathcal{A}) + P(\mathcal{B}),$$

$$\mathcal{A} \subset \mathcal{B} \Rightarrow P(\mathcal{A}) \leq P(B).$$

Conditional Probabilities. Independence

Bayes theorem:

- $P(\mathcal{A}|\mathcal{M})$: Probability of event $\mathcal{A}$ happening under the condition that $\mathcal{M}$ also happens:

$$P(\mathcal{A}|\mathcal{M}) = \frac{P(\mathcal{A} \cap \mathcal{M})}{P(\mathcal{M})} = \frac{P(\mathcal{M}|\mathcal{A})P(\mathcal{A})}{P(\mathcal{M})}.$$

- $P(\mathcal{A}|\mathcal{M})$ is called ***conditional probability***.

Conditional Probabilities. Independence

- Conditional probability properties:
  - $\mathcal{M} \subset \mathcal{A} \Rightarrow P(\mathcal{A}|\mathcal{M}) = 1,$
  - $\mathcal{A} \subset \mathcal{M} \Rightarrow P(\mathcal{A}|\mathcal{M}) = \frac{P(\mathcal{A})}{P(\mathcal{M})}.$

Artificial Intelligence &
Information Analysis Lab

Conditional Probabilities. Independence

- Event $\mathcal{A}, \mathcal{B}$ ***independence***:

$$P(\mathcal{A}\,|\mathcal{B}) = P(\mathcal{A}).$$

Probability independence properties:

$$P(\mathcal{B}\,|\mathcal{A}) = P(\mathcal{B}),$$

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B}).$$

- In general, the events $\{\mathcal{A}_i, i = 1, \dots, N\}$ are independent, if $\mathcal{A}_i, \mathcal{A}_k$ are independent for every $i \neq k$.

Random Variable

Let a Borel field $\mathcal{F}$, $\mathcal{S} = \mathbb{R} = \{z\}$ be used to define an ***event***:

$$\{z : X(z) \leq x\} \in \mathcal{F}.$$

- $X(z)$: ***random variable*** having probability $P\{z: X(z) \leq x\}$ for every real x.

Artificial Intelligence & Information Analysis Lab

Random Variable

Function $F_X(x) = P\{z : X(z) \leq x\}$
is called ***probability distribution*** of X.

- Also called ***cumulative distribution function*** (***cdf***) of X.

Artificial Intelligence & Information Analysis Lab

Random Variable

Probability distribution properties :

$$F_X(-\infty) = P\{X \le -\infty\} = 0,$$
$$F_X(+\infty) = P\{X \le +\infty\} = 1.$$

- Function $F_X(x)$ is non decreasing:

$$x_1 < x_2 \Rightarrow F_X(x_1) \le F_X(x_2),$$
$$P\{x_1 \le X \le x_2\} = F_X(x_2) - F_X(x_1).$$

Random Variable

- $F_X(x)$ is a right-continuous function:

$$F_X\,(x^+) = \lim_{\varepsilon\to 0} F_X(x+\varepsilon), \quad \varepsilon > 0,$$

$$F_X\,(x^-) = \lim_{\varepsilon\to 0} F_X(x-\varepsilon), \varepsilon > 0\,.$$

- Finally:

$$P\{X = x\} = F_X\{x^+\} - F_X\{x^-\} = F_X(x) - F_X(x^-).$$

Propability Density Fuction

The ***probability density function*** (***pdf***) $f_X(x)$ of variable X is the derivative of the probability distribution function $F_X(x)$:

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Pdf properties:

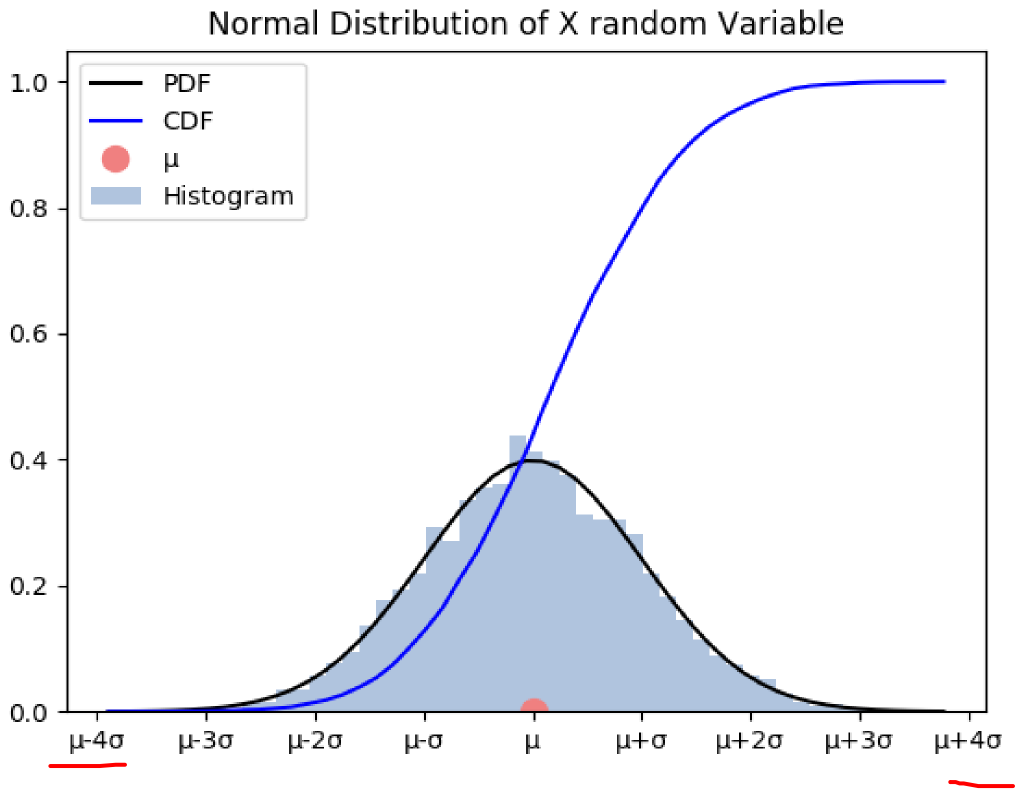
- Since $F_X(x)$ is non-decreasing, then $f_X(x) \geq 0$.

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx,$$

$$\int_{x_1}^{x_2} f_X(x)dx = F_X(x_2) - F_X(x_1).$$

Artificial Intelligence & Information Analysis Lab

Propability Density Fuction



Cumulative probability function and probability density function.

Propability Density Fuction

Normal (Gaussian) distribution $N(m, \sigma)$:

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

Gaussian distribution parameters

- m: position parameter
- σ: scale parameter.

Propability Density Fuction

- Normal (Gaussian) distribution $N(0,1)$:

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

- Uniform distribution $\mathrm{U}(0,1)$:

$$f_X(x) = 1, \qquad x \in [0,1].$$

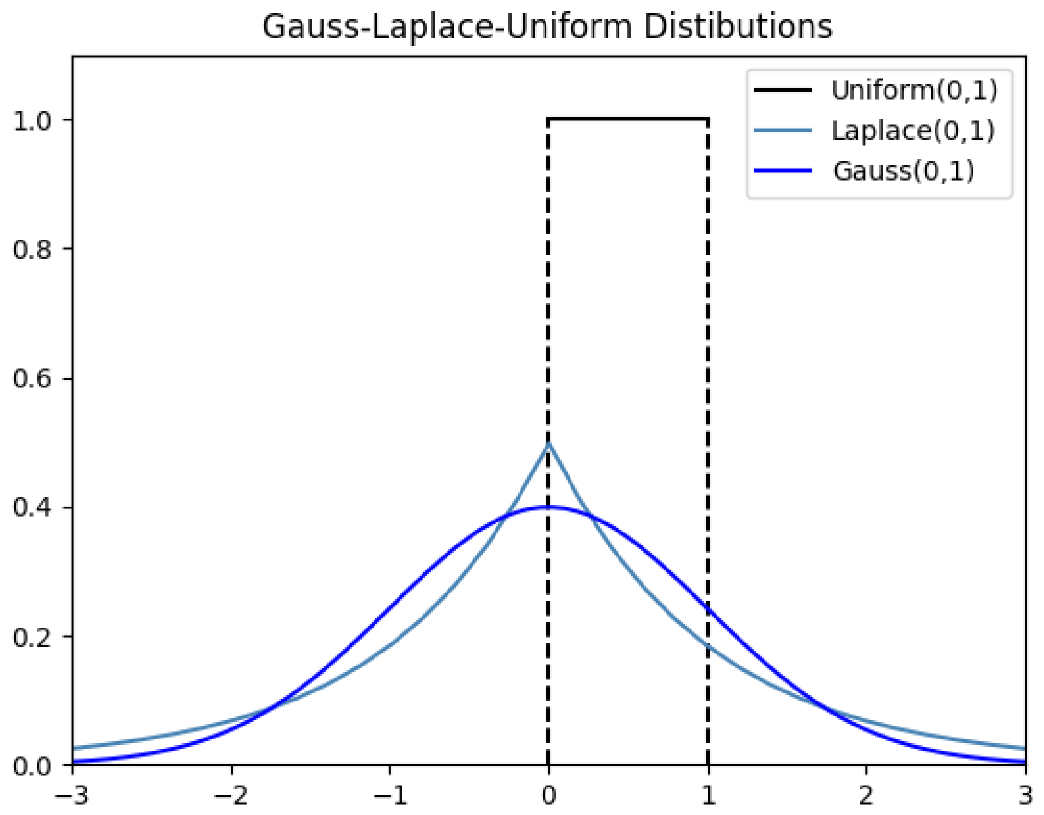
- Laplacian distribution:

$$f_X(x) = \frac{1}{2} e^{-|x|}.$$

Propability Density Fuction

1D probability distributions:

- Uniform (short-tailed).
- Gaussian.
- Laplacian (long-tailed).



Propability Density Fuction

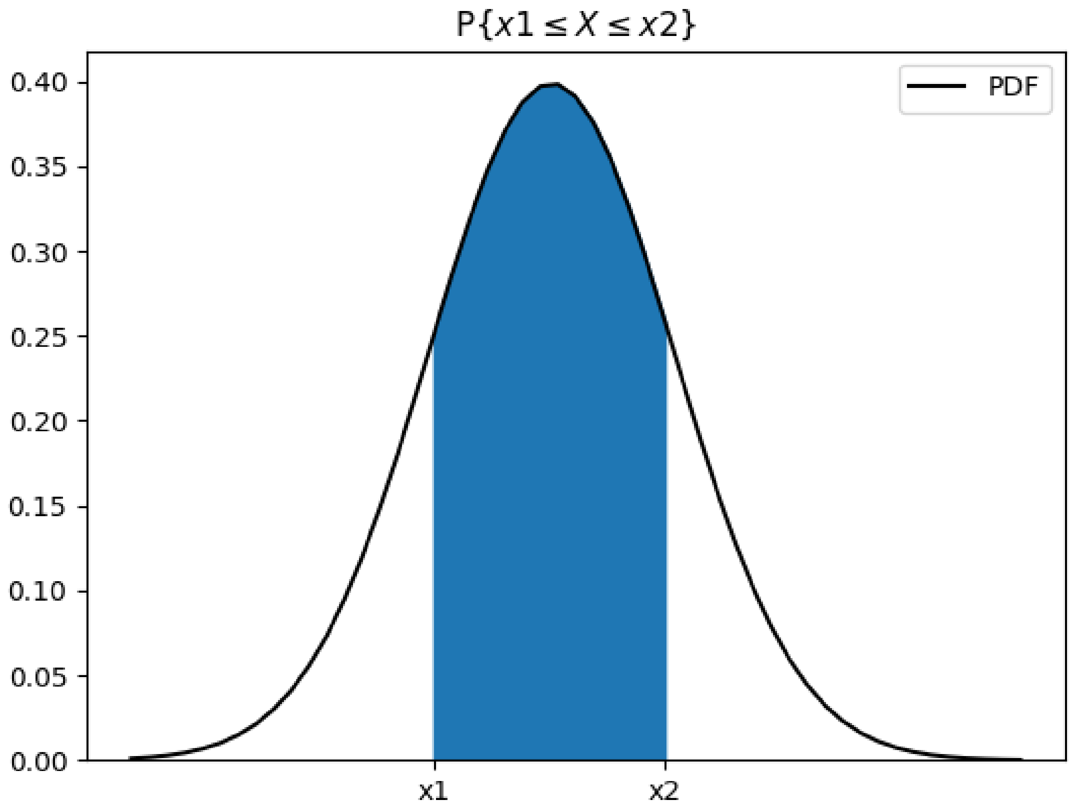
- If $\mathcal{A}$ is a subset of $\mathbb{R}$:

$$P(X \in \mathcal{A}) = \int_{\mathcal{A}} f_X(x)\, dx.$$

- Thus:

$$P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} f_X(x) dx = F_X(x_2) - F_X(x_1).$$

Propability Density Fuction



Probability of a random variable lying in an interval.

Artificial Intelligence & Information Analysis Lab

Discrete Random Variable

Let a discrete-valued random variable X taking values in the set $\mathcal{X} = \{x_1, \dots, x_{|\mathcal{X}|}\}$, with probabilities p_i:

$$\sum_{i=1}^{|\mathcal{X}|} p_i = 1.$$

***Probability density function* (*pdf*)** $p_X(x)$ of X:

$$f_X(x) = \sum_{i=1}^{|\mathcal{X}|} p_i \delta(x - x_i),$$

Artificial Intelligence & Information Analysis Lab

Discrete Random Variable

Probability distribution $F_X(x)$ of X for ordered x_i:

$$F_X(x) = \begin{cases} 0\ , & \text{if } x < x_1 \\ \sum_{i=1}^{k} f_X(x_i), & \text{if } x_k \leq x < x_{k+1}, k = 1,2,\dots,|\mathcal{X}|. \end{cases}$$

Expectation operator

Expectation operator $E\{.\}$ on a real function $g(X)$ of X:

$$E\{g(X)\} = \int g(x)f(x)dx.$$

- Sometimes the notation $E_X\{.\}$ is used to prevent confusion.
- ***Expected value*** m_X of the random variable X :

$$m_X = E\{X\} = \int x f_X(x)dx.$$

- Sometimes also called ***mean value***:
  - It defines the *location* of the distribution.
  - Not to be confused with (sample) ***arithmetic mean***!

Expectation operator

- ***Standard deviation*** σ_X and ***variance*** ${\sigma_X}^2$:

$$ {\sigma_X}^2 = E\{(X - m_X)^2\} = \int (x - m_X)^2 f_X(x)dx \,. $$

- It quantifies the ***dispersion*** of the distribution.
- Not to be confused with ***sample standard deviation***!

Nonlinear random variable transformation

- $Y = g(X)$: real function $g(.)$ of a real random variable X.
- Y is a random variable.
- Expectation operator on random variable Y is given by:

$$E\{g(x)\} = \int y f_Y(y) dy.$$

- $f_Y(y)$ is the probability density function of Y.

Nonlinear random variable transformation

- If $g(x)$ is differentiable and
- x_i are solutions of the equation $g(x) = y$, then:

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|}.$$

Two Random Variables

- The intersection of events $\{X \leq x\} \cap \{Y \leq y\}$ is an event.
- Its probability is called ***joint probability distribution*** of X and Y:

$$F_{XY}(x, y) = P\{X \leq x, Y \leq y\}.$$

- Function $F_{XY}(x, y)$ gives the probabilities of X and Y on the infinite third of quadrant $\{X \leq x, Y \leq y\}$ of plane $\mathbb{R}^2$.
- ***Joint probability density function*** of X, Y:

$$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y).$$

Artificial Intelligence & Information Analysis Lab

Two Random Variables

Joint probability properties:

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)dxdy.$$

$$F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0\,, \qquad F_{XY}(+\infty,+\infty) = 1.$$

- $f_{XY}(x,y) \geq 0$, since $F_{XY}(x,y)$ is non-decreasing.

Two Random Variables

Joint probability properties:

- Integration:

$$P\{(X,Y) \in \mathcal{A}\} = \iint_{\mathcal{A}} f_{XY}(x,y)dxdy,$$

$$P\{x_1 \leq X \leq x_2 \, , y_1 \leq Y \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(x,y)dxdy.$$

Two Random Variables

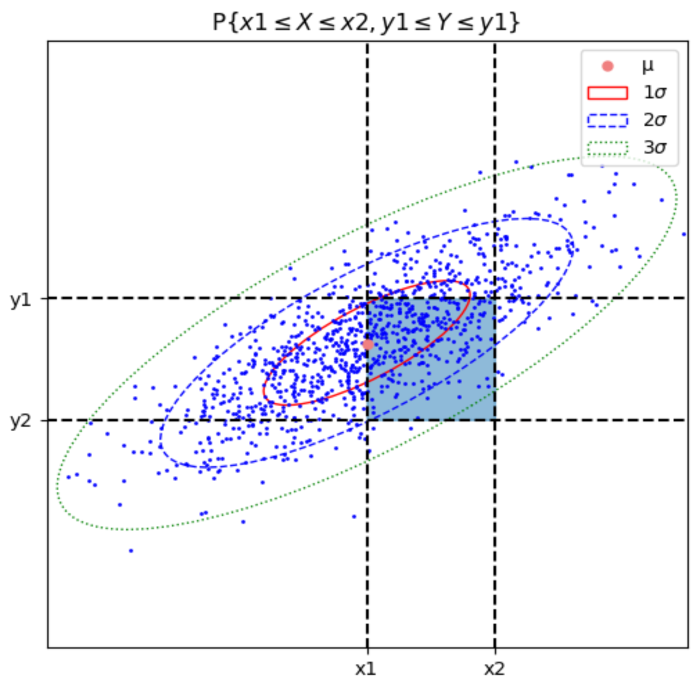
Joint probability properties:

- Integration:

$$P\{x_1 \le X \le x_2 \,, y_1 \le Y \le y_2\} =$$
$$F_{XY}(x_2, y_2\,) + F_{XY}(x_1, y_1) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1)$$

$$P\{x_1 \le X \,, y_1 \le Y\} =$$
$$F_{XY}(+\infty, +\infty) + F_{XY}(x_1, y_1) - F_{XY}(x_1, +\infty) - F_{XY}(+\infty, y_1) =$$
$$1 - F_X(x_1) - F_Y(y_1) + F_{XY}(x_1, y_1).$$

Two Random Variables



Joint pdf integration in a rectangle.

Two Random Variables

Marginal distributions:

- Marginal cdfs:

$$F_X(x) = F_{XY}(x, +\infty)\,, \qquad F_Y(y) = F_{XY}(+\infty, y).$$

- Marginal pdfs:

$$f_X(x) = \int_{-\infty}^{-\infty} f_{XY}(x, y)dy\,, \quad f_Y(y) = \int_{-\infty}^{-\infty} f_{XY}(x, y)dx.$$

Two Random Variables

- Knowledge of $f_{XY}(x, y)$ entails knowledge of $f_X(x), f_Y(y)$.
- The reverse is not generally true.

- Two random variables X and Y are ***independent*** if:

$$f_{XY}(x, y) = f_X(x) f_Y(y).$$

- Then, the marginal probability functions fully determine the joint probability function.

Two Random Variables

- Given independence of X, Y:

$$F_{XY}(x,y) = F_X(x)F_Y(y).$$

$$f_{XY}(x,y) = f_X(x)f_Y(y).$$

- Expected value of the function $g(X,Y)$ of X, Y:

$$E\{g(X,Y)\} = \iint g(x,y)f_{XY}(x,y)dxdy\,.$$

Artificial Intelligence & Information Analysis Lab

Two Random Variables

- ***Correlation*** of X, Y:

$$E\{XY\} = \iint xyf_{XY}(x,y)dxdy.$$

- ***Expected mean vector*** $\mathbf{m} = [m_X, m_Y]^T$:

$$m_X = E\{X\} = \int xf_X(x)dx, \qquad m_Y = E\{Y\} = \int yf_Y(y)dy.$$

- It defines the ***location*** of the distribution.
- Marginal pdfs are used.

Two Random Variables

- ***Standard deviations*** σ_X, σ_Y and ***variances*** ${\sigma_X}^2, {\sigma_Y}^2$:

$$ {\sigma_X}^2 = \int (x - m_X)^2 f_X(x)dx\,, $$

$$ {\sigma_Y}^2 = \int (y - m_Y)^2 f_Y(y)dy\,. $$

- They define the ***dispersion*** of the 2D distribution along axes X, Y.
- Marginal pdfs are used.

Two Random Variables

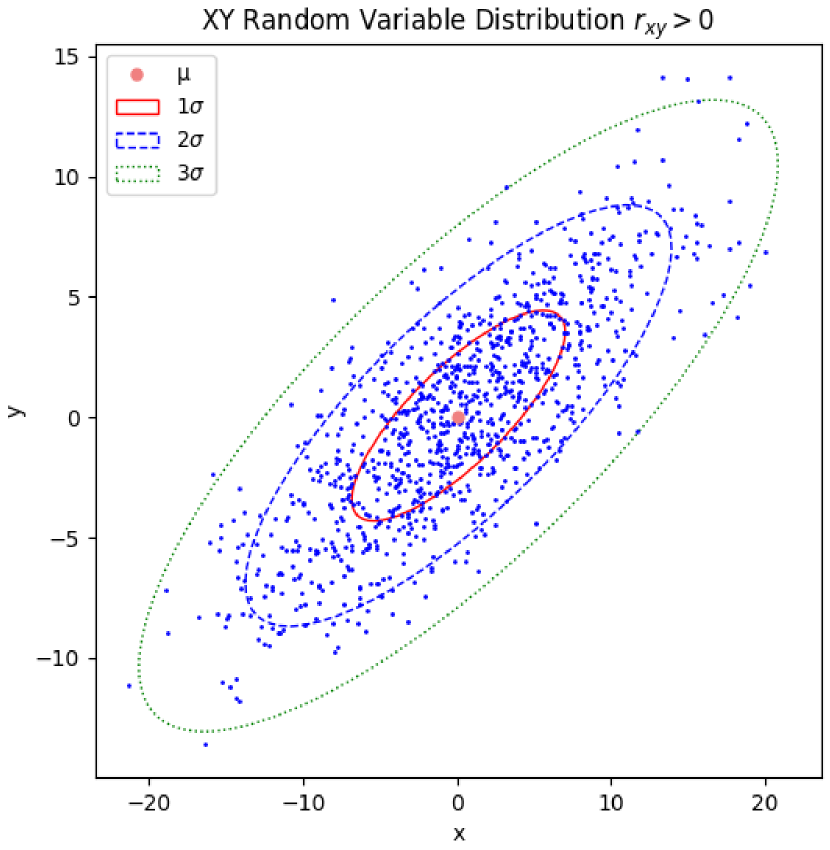
- ***Covariance*** of X, Y:

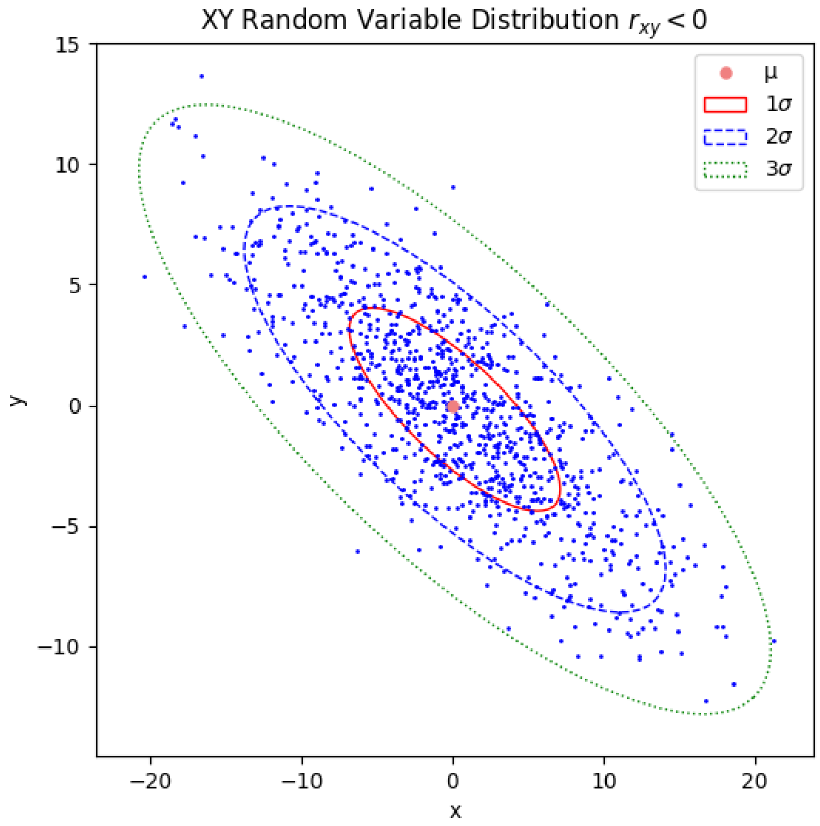
$$E\{(X - m_X)(Y - m_Y)\} = \iint (x - m_X)(y - m_Y) f_{XY}(x, y) dx dy.$$

- ***Correlation coefficient*** of X, Y:

$$r_{XY} = \frac{E\{(X - m_X)(Y - m_Y)\}}{\sigma_X \sigma_Y}.$$

Two Random Variables





Joint Gaussian pdfs with positive and negative r_{XY}.

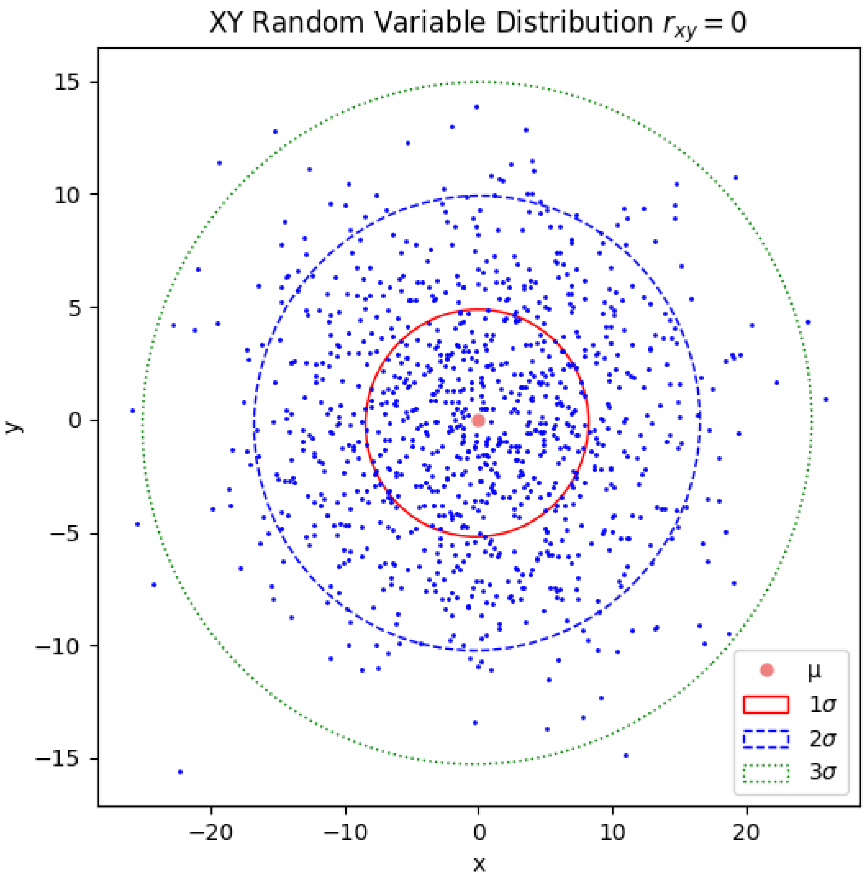
Two Random Variables

- Two random variables X and Y are uncorrelated if:

$$r_{XY} = 0.$$

- Two independent variables are always uncorrelated.
- The opposite is not generally true.

Two Random Variables



Joint Gaussian pdf with $r_{XY} = 0$.

Normal Random Variables

Gaussian (normal) joint pdf $N(m_1, m_2, \sigma_1, \sigma_2, r)$:

- Two random variables are joint normal if:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\{A\},$$

$$A = -\frac{1}{2(1-r^2)}\left(\left(\frac{x-m_1}{\sigma_1}\right)^2 + \left(\frac{y-m_2}{\sigma_2}\right)^2 - \frac{2r(x-m_1)(y-m_2)}{\sigma_1\sigma_2}\right).$$

- r: correlation coefficient of X, Y.

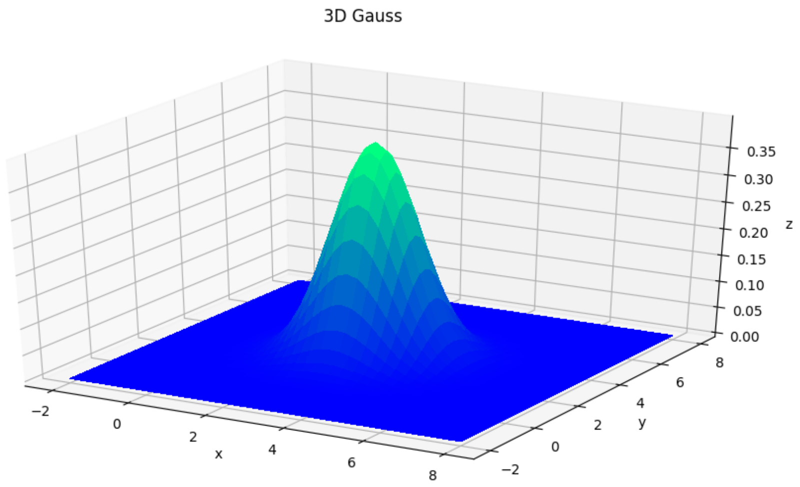
Normal Random Variables

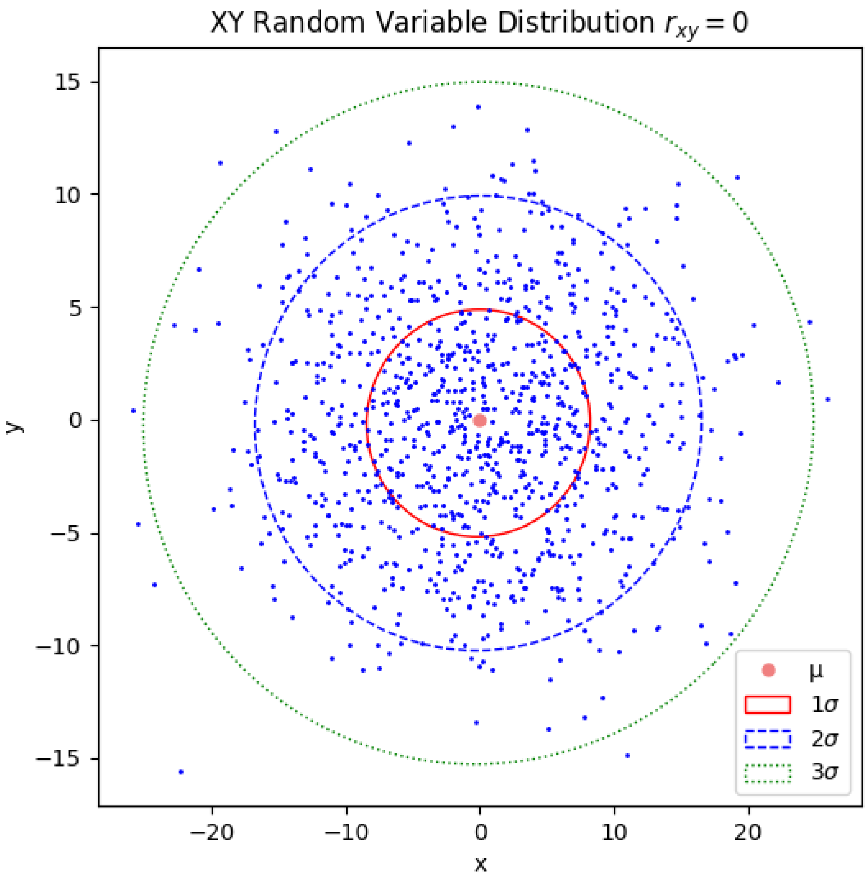
- Locus of the normally distributed points:

$$\left(\left(\frac{x-m_1}{\sigma_1}\right)^2+\left(\frac{y-m_2}{\sigma_2}\right)^2-\frac{2r(x-m_1)(y-m_2)}{\sigma_1\sigma_2}\right)=c.$$

- ***Ellipsis***: 2nd degree curve on $\mathbb{R}^2$.
  - Special case: ***circle*** for $r=0, \sigma_1=\sigma_2=\sigma$.

Normal Random Variables





a) Gaussian pdf; b) Iso-contours.

Normal Random Variables

Properties:

- $m_1, m_2, \sigma_1, \sigma_2, r$ are equal to $m_X, m_Y, \sigma_X, \sigma_Y, r_{XY}$, respectively.
- A joint Gaussian pdf produces Gaussian marginal pdfs:

$$f_X(x) = \int_{-\infty}^{-\infty} f_{XY}(x,y)dy = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x-m_1}{\sigma_1}\right)^2\right\}.$$

$$f_Y(y) = \int_{-\infty}^{-\infty} f_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{y-m_2}{\sigma_2}\right)^2\right\}.$$

Normal Random Variables

Properties:

- If X and Y are joint normal, then X and Y are normal.
- The opposite is not true.

- If X, Y are individually normal and independent, then they are joint normal.
- If X, Y are joint normal and uncorrelated then they are independent.

Functions of Two Random Variables

If $Z = g(X,Y)$, $W = h(X,Y)$ and (x_i, y_i) are the solutions of the system of two ***differentiable functions***:

$$z = g(x_i, y_i), \qquad w = h(x_i, y_i),$$

then :

$$f_{ZW}(z,w) = \sum_i \frac{f_{XY}(x_i,y_i)}{|\det(\mathbf{J}(x_i,y_i))|},$$

$\mathbf{J}(x,y) = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix}$ is the ***Jacobian matrix*** of $g(.), h(.)$.

Functions of Two Random Variables

- Derivation of the pdf of one random function $Z = g(X, Y)$.
- The ***virtual variable trick*** can be used, i.e., $W = X$.

- Example:
  - For the linear system $Z = X + Y, \; W = X$,

    the one and only solution is: $x = w, \; y = z - w$.

$$f_{ZW}(z, w) = \left. \frac{f_{XY}(x,y)}{\begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix}} \right|_{x=w, y=z-w} = f_{XY}(w, z - w).$$

Functions of Two Random Variables

- If X, Y are independent:

$$f_{XY}(x, y) = f_X(x) f_Y(y).$$

- The pdf of the sum $Z = X + Y$ is the convolution of the the marginal pdfs of f_X, f_Y:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{ZW}(z, w) dw = \int_{-\infty}^{+\infty} f_X(w) f_Y(z - w) dw$$
$$= f_X * f_Y.$$

Functions of Two Random Variables

- Properties:
- Invariance of Gaussian distributions under linear transformations.
- If X, Y are jointly normal, their linear transformations Z and W:

$$Z = aX + bY + c, \qquad W = dX + eY + f$$

are jointly normal.

Two Functions of Two Random Variables

- Separable functions and independence:
  - If X and Y are independent, then the random functions

$$Z = g(X), \qquad W = h(Y)$$

produce also independent random variables.

Functions of Two Random Variables

Non-differentiable functions.

- Example:

$$Z = \max\{X, Y\}.$$

Pdf derivation:

$$F_Z(z) = P\{Z \leq z\} = P\{\max\{X, Y\} \leq z\} = P\{X \leq z, Y \leq z\} = F_{XY}(z, z).$$

$$f_Z(z) = \frac{d}{dz} F_{XY}(z, z).$$

Multiple Random Variables

Random vector $\mathbf{X} \in \mathbb{R}^n$ (column of n random variables):

$$\mathbf{X} = [X_1, \dots, X_n]^T.$$

- Joint probability distribution $F_{\mathbf{X}}(\mathbf{x})$ of $X_1, \dots, X_n$:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1 \dots X_n}(x_1, \dots, x_n) = P\{X_1 \leq x_1, \dots, X_N \leq x_n\}.$$

Multiple Random Variables

- Joint probability density distribution $f_{\mathbf{X}}(\mathbf{x})$ of $X_1, \dots, X_n$:

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1 \dots X_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{\mathbf{X}}(\mathbf{x}),$$

$$f_{\mathbf{X}}(\mathbf{x}) \geq 0.$$

- Joint probability density distribution of $X_1, \dots, X_n$ (but X_i):

$$f_{X_1 \dots X_{i-1} X_{i+1} \dots X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = \int_{-\infty}^{+\infty} f_{X_1 \dots X_n}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) dx_i.$$

Multiple Random Variables

- Probability of an event in a subspace $\mathcal{A} \in \mathbb{R}^n$:

$$P\{\mathbf{X} \in \mathcal{A}\} = \int_{\mathcal{A}} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} .$$

- Expectation operator:

$$E\{g(X_1, \dots, X_2)\} = \int \dots \int g(x_1, \dots, x_n) f_{X_1 \dots X_n}(x_1, \dots, x_n) dx_1 \dots dx_n,$$

$$E\{g(\mathbf{X})\} = \int g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Multiple Random Variables

- **Expected vector** of X:

$$\mathbf{m_X} = E\{\mathbf{X}\} = [E\{X_1\} \dots E\{X_n\}]^T,$$

$$m_{X_i} = E\{X_i\} = \int x_i f_{X_i}(x_i) dx_i \,.$$

- Sometimes called mean vector.
- It defines the ***location*** of the distribution.
- Marginal pdfs are used.

Multiple Random Variables

- ***Standard deviations*** σ_{X_i} and ***variances*** $\sigma_{X_i}{}^2, \quad i = 1, \dots, n$:

$$\sigma_{X_i}{}^2 = \int (x_i - m_{X_i})^2 f_{X_i}(x_i) dx_i \,.$$

- They define the ***dispersion*** of the distribution along axes $X_1, \dots, X_n$.
- Marginal pdfs are used.

Multiple Random Variables

- ***Correlation matrix*** of $\mathbf{X}$ ($n \times n$ matrix):

$$\mathbf{R_X} = E\{\mathbf{XX}^T\} = [E\{X_i X_j\}],$$

$$E\{X_i X_j\} = \int\int x_i x_j f_{X_i X_j}(x_i, x_j) dx_i dx_j \,.$$

- ***Covariance matrix*** of $\mathbf{X}$ ($n \times n$ matrix):

$$\mathbf{C_X} = E\{(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T\} = \mathbf{R_X} - \mathbf{m_X}\mathbf{m_X}^T.$$

Multiple Random Variables

- Variables $X_1, \dots, X_n$ are independent, if:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_i f_{X_i}(x_i) \quad \text{or} \quad F_{\mathbf{X}}(\mathbf{X}) = \prod_i F_{X_i}(x_i) \,.$$

- Variables $X_1, \dots, X_n$ are uncorrelated if for every $i \neq j$:

$$E\{X_i X_j\} = E\{X_i\}E\{X_j\} \quad \text{or} \quad E\{\mathbf{X}\mathbf{X}^T\} = \mathbf{m}_{\mathbf{X}}\mathbf{m}_{\mathbf{X}}^T.$$

Multiple Random Variables

Gaussian (normal) random vectors:

- Variables $X_1, \ldots, X_n$ are jointly normal if:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\mathbf{C})^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})\right\}.$$

- $\mathbf{C}$: covariance matrix.
- $\det(\mathbf{C})$: determinant of $\mathbf{C}$.

Multiple Random Variables

Gaussian (normal) pdf properties:

- Its parameters define directly data location and dispersion:

$$\mathbf{m}_{\mathbf{X}} = \mathbf{m},\ \mathbf{C}_{\mathbf{X}} = \mathbf{C}.$$

- If $X_1, \dots, X_n$ are independent, then they are uncorrelated. The opposite is not generally true.
- If $X_1, \dots, X_n$ are joint normal and uncorrelated, they are independent as well.

Multiple Random Variables

Eigenanalysis of covariance matrix $\mathbf{C}$:

$$\mathbf{C}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, \dots, n,$$

$\boldsymbol{\Lambda} = \text{diag}\{\lambda_1\ \lambda_2 \dots\ \lambda_n\}$.

$\mathbf{P} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_n]$.

$\mathbf{P}^T = \mathbf{P}^{-1}$.

Multiple Random Variables

The symmetrical covariance matrix $\mathbf{C}$ can be diagonalized using eigenanalysis:

$$\mathbf{C} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T.$$

- $\mathbf{C}$ is positive definite matrix:
  - All its n eigenvalues are non-negative $\lambda_i \geq 0, \;\; i = 1, \dots, n$.

Multiple Random Variables

- Locus of the normally distributed points $x_1, \dots, x_n$ having equal pdf values:

$$d(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_\mathbf{X})^T \mathbf{C}_\mathbf{X}{}^{-1} (\mathbf{x} - \mathbf{m}_\mathbf{X}) = c.$$

- Iso-hypersurfaces are ***hyperellipsoids.***

Multiple Random Variables

- If $\mathbf{C}$ is diagonal with non-zero elements C_{ii}:

$$d(\mathbf{x}) = \frac{(x_1 - m_1)^2}{C_{11}} + \dots + \frac{(x_n - m_n)^2}{C_{nn}} = c$$

describes an hyperellipse having axes parallel to $X_1, \dots, X_n$.

- Special case: ***hypersphere*** for $\mathbf{C} = \sigma^2 \mathbf{I}$.
- 3D case: ellipsoid, sphere.

Multiple Random Variables

Mahalanobis distance:

$$d(\mathbf{m}_i, \mathbf{m}_j) = \sqrt{(\mathbf{m}_i - \mathbf{m}_j)^T \mathbf{C}^{-1} (\mathbf{m}_i - \mathbf{m}_j)}$$

defines equiprobable points of multivariate distribution $p(\mathbf{x}|\mathcal{C}_k)$:

$$(\mathbf{x} - \mathbf{m}_k)^T \mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}^T (\mathbf{x} - \mathbf{m}_k) = c^2.$$

It describes hyper-ellipsoid ***iso-hypersurfaces***:

- Ellipses in 2D, ellipsoids in 3D space.

Multiple Random Variables

- The linear coordinate transformation $\mathbf{x}' = \mathbf{P}^T\mathbf{x}$ results in data point projection to a new coordinate system based on eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.
- Iso-hypersurface equation becomes:

$$\frac{(x_1' - m_{k_1}')^2}{\lambda_1} + \frac{(x_2' - m_{k_2}')^2}{\lambda_2} + \dots + \frac{(x_n' - m_{k_n}')^2}{\lambda_n} = c^2.$$

It is a hyper-ellipsoid in $\mathbb{R}^n$ having axes parallel to eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

Multiple Functions of Multiple Variables

If random vector $\mathbf{Y} = \boldsymbol{g}(\mathbf{X})$ is a differentiable multivalued function of random vector $\mathbf{X}$:

$$Y_1 = g_1(\mathbf{X}) = g_1(X_1, \ldots, X_n),$$
$$Y_2 = g_2(\mathbf{X}) = g_2(X_1, \ldots, X_n),$$
$$\ldots$$
$$Y_n = g_n(\mathbf{X}) = g_n(X_1, \ldots, X_n),$$

and $\mathbf{x}_i$ are the solutions of $\boldsymbol{g}(\mathbf{x}) = \mathbf{y}$, then:

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_i \frac{f_{\mathbf{X}}(\mathbf{x})}{|\det(\mathbf{J}(\mathbf{x}_i))|},$$

$\mathbf{J}(\mathbf{x}) = \left[\frac{\partial g_i}{\partial x_j}\right]$ is the ***Jacobian matrix*** of this transformation.

Multiple Functions of Multiple Variables

Special cases

- ***Multiple-Input Multiple-Output*** (***MIMO***) linear system:
$$\mathbf{Y} = \mathbf{A}\,\mathbf{X},$$
- $\mathbf{A}$: transformation matrix having $\det(\mathbf{A}) \neq 0$. Then:
$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{y})}{|\det(\mathbf{A})|}.$$
- Neural networks are differentiable MIMO nonlinear systems $\mathbf{Y} = \boldsymbol{f}(\mathbf{X}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a trainable parameter vector.

Random Number Generation

Artificial ***noise generation*** is primarily needed for simulations.

- Additive/multiplicative image noise generators:

$$g(i,j) = f(i,j) + n(i,j),$$

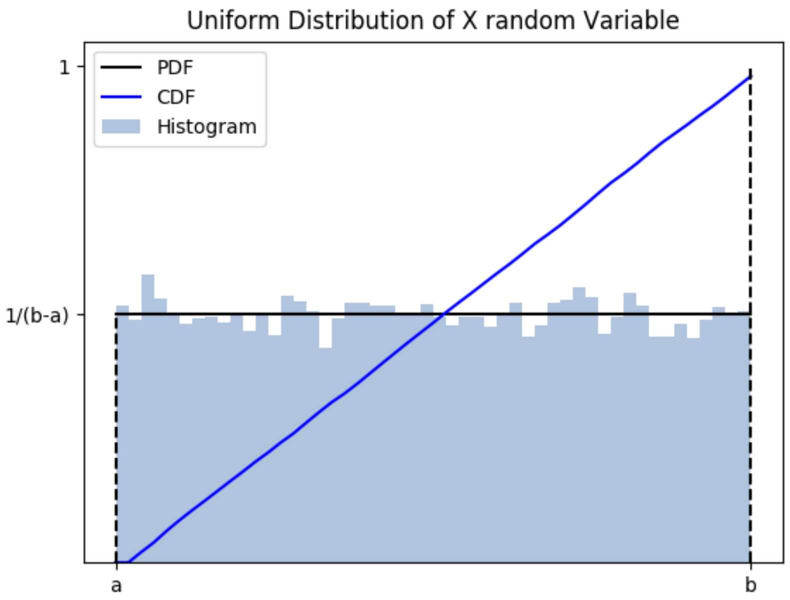
$$g(i,j) = f(i,j)n(i,j).$$

- Random number generators: they produce uniform noise in $[0,1]$.

Random Number Generation

Uniform random number generation in $[0,1]$:

- Use of a math library random number generator (C/C++).

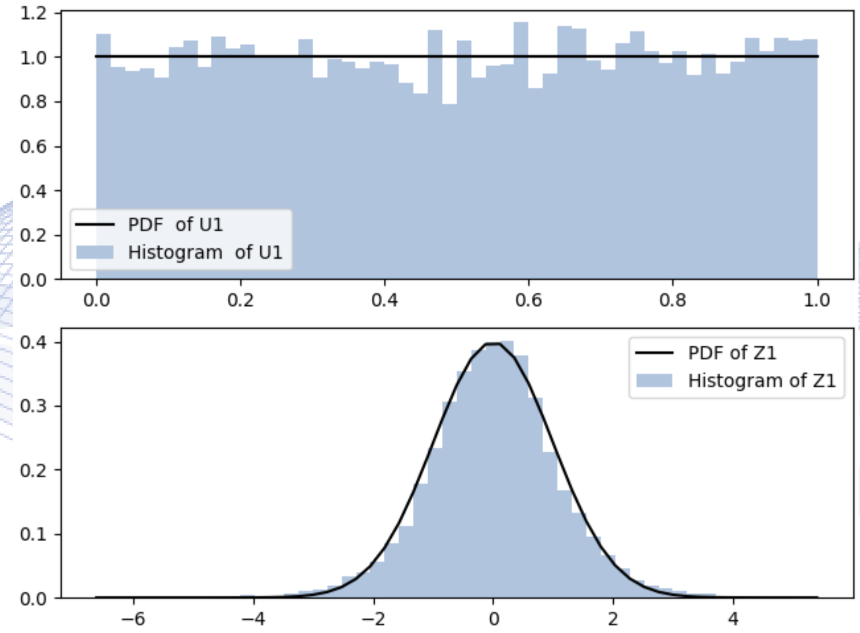


Random Number Generation

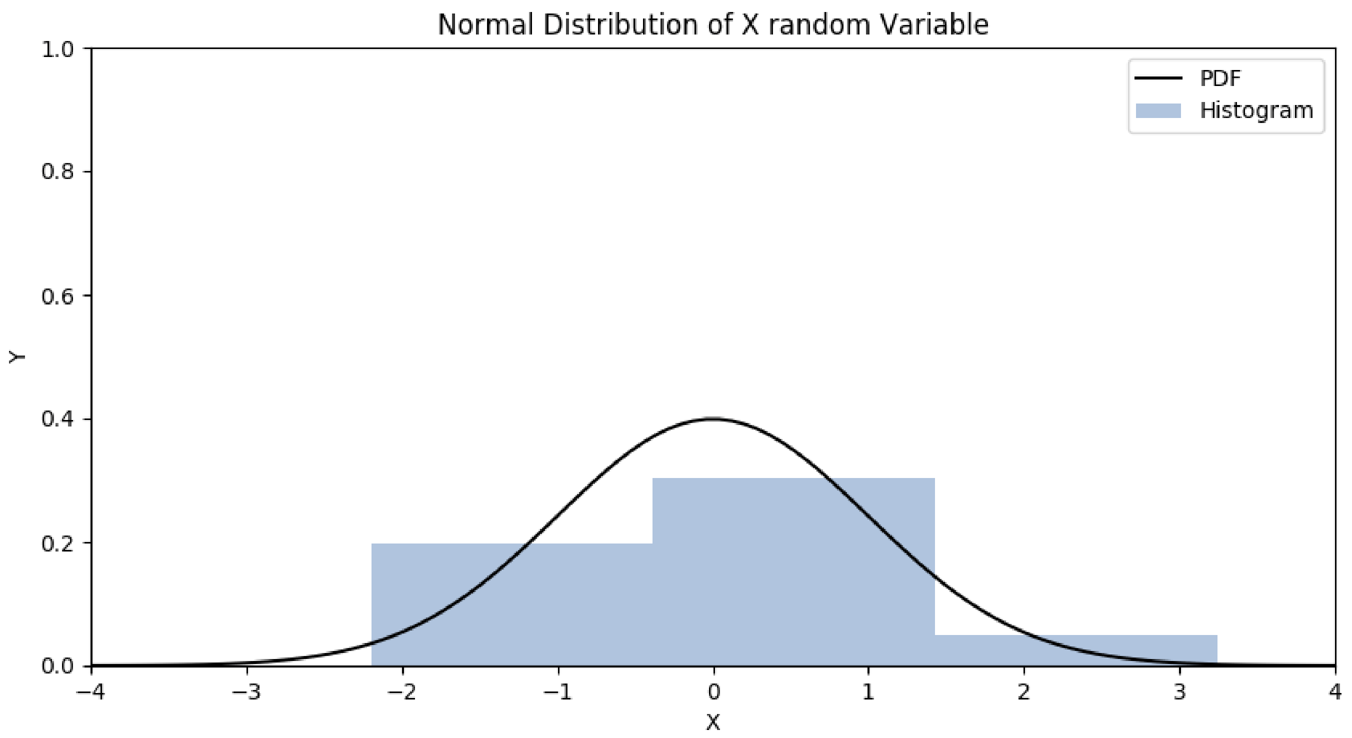
Gaussian $N(0,1)$random number generation:

- Input uniform $\mathrm{U}(0,1)$ distribution.
- Transformation function:

$$Z = -\frac{\ln\left(\frac{1}{U}-1\right)}{1.702}.$$



Random Number Generation



Animation Gaussian random number generator histogram vs the number of generated samples.

Random Number Generation

Laplacian random number generation:

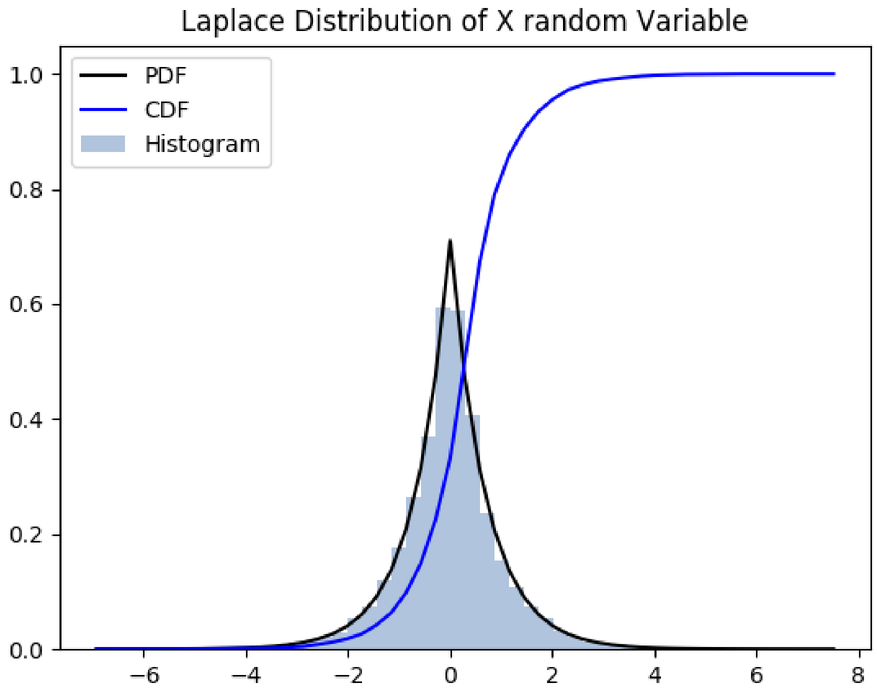
- Laplacian distribution:

$$f_X(x) = \frac{1}{2} e^{-|x|}.$$

- Transformation of uniform noise X distributed in $[0,1]$ to a Laplacian probability distribution Y:

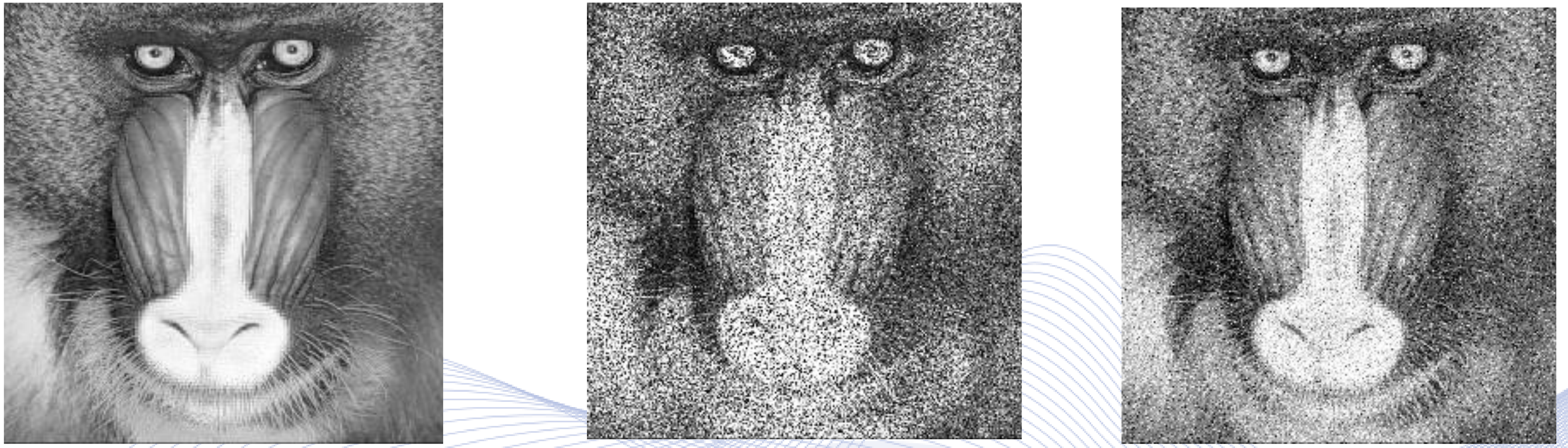
$$Y = \begin{cases} \ln(2X) & 0 \le x \le 1/2. \\ -\ln(2-2X) & 1/2 \le x < 1. \end{cases}$$

Random Number Generation



Laplacian random number generator histogram.

Noise generators for digital image processing



a) Original image; b) Image corrupted by multiplicative Gaussian noise;
c) image corrupted by additive Laplacian noise.

Random Vector Generation

- If the random variables X and Y have two-dimensional normal distribution with known $m_X, m_Y, \sigma_X, \sigma_Y$ and $r_{XY} \neq 0$, then the random variable:

$$Z = Y - r_{XY} \frac{\sigma_X}{\sigma_Y} X,$$

$$m_Z = m_Y - r_{XY} \frac{\sigma_Y}{\sigma_X} m_X, \qquad \sigma_Z = \sigma_Y \sqrt{1 - r_{XY}^2}.$$

has a normal distribution and is independent of X.

Random Vector Generation

Box-Muller transformation:

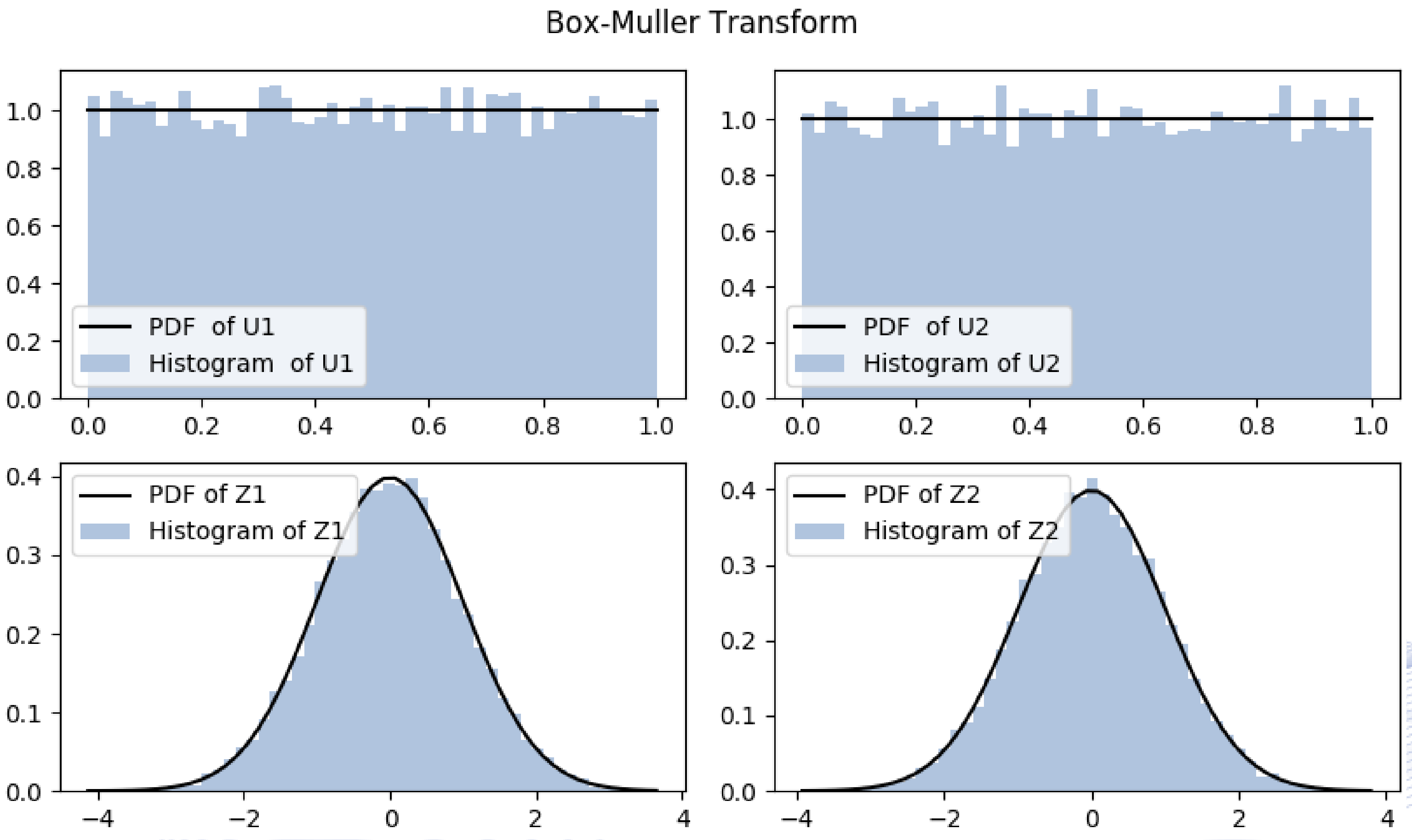
Generation of a random variable pair (vector) $[X,Y]^T$ following 2D Gaussian distribution $\mathrm{N}(m_X, m_Y, \sigma_X, \sigma_Y)$.

- Generate two random variables X_1, X_2 having uniform distribution in $[0,1]$.
- The two random variables Z_1, Z_2:

$$Z_1 = \sqrt{-2\ln(X_1)}\cos(2\pi X_2),\ Z_2 = \sqrt{-2\ln(X_1)}\sin(2\pi X_2).$$

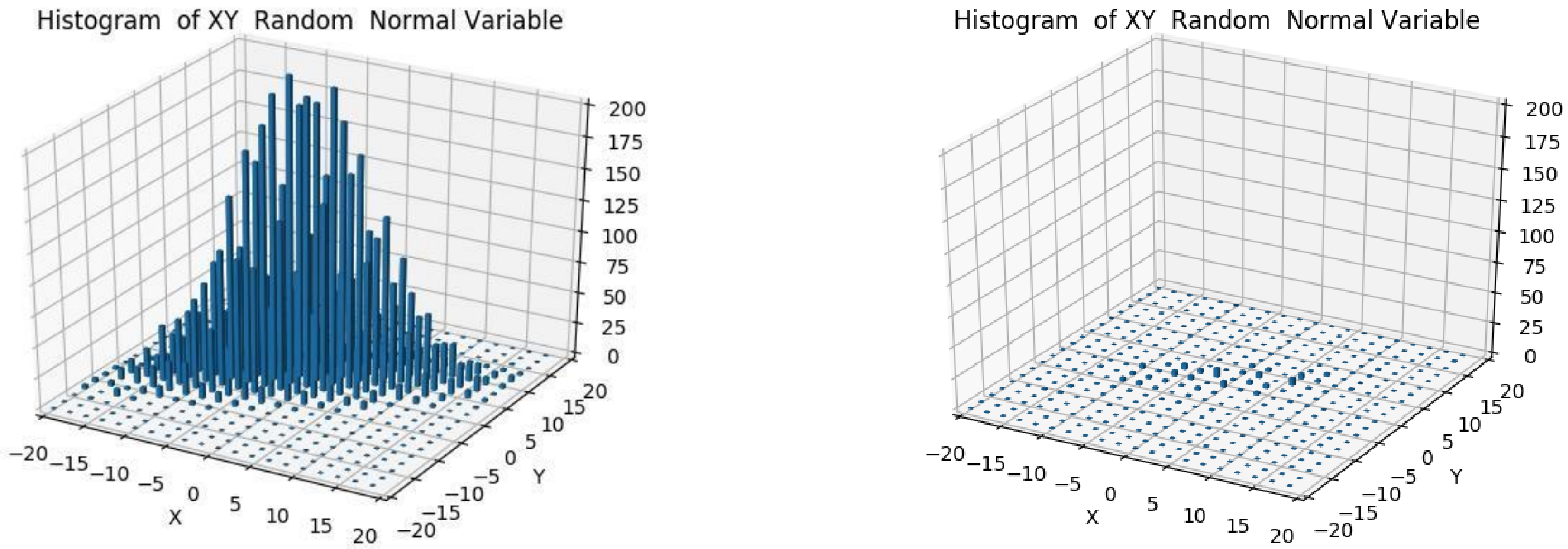
are independent and follow Gaussian distribution $N(0,1)$.

Random Number Generation



Box-Muller transformation.

Generation of two Random Variables



a) 2D Gaussian random number generator histogram; b) Its animation vs the number of generated samples.

BIBLIOGRAPHY

[PAP2002] A. Papoulis, S. U. Pillai, Probability, Random Variables and Stochastic Processes, McGraw-Hill, 2002.

[STR1999] M.G. Strintzis, Pattern Recognition, 1999 (in Greek).

Q & A

Thank you very much for your attention!

Contact: Prof. I. Pitas
pitas@csd.auth.gr