Gdańsk University of Technology

Random Processes and Stochastic Control – Part 1 of 2

by

Maciej Niedźwiecki

Gdańsk University of Technology

Departament of Automatic Control

Gdańsk, Poland

Does God play dice ?

Czy Bóg gra w kości ?

“Quantum mechanics is certainly imposing. But an inner voice tells me that it is not yet the real thing. The theory says a lot, but does not really bring us any closer to the secret of the ”old one.” I, at any rate, am convinced that He does not throw dice.”

Albert Einstein

Yes, He does !

to get the modern view watch Richard Feynman’s
The Douglas Robb Memorial Lectures 1979
http://vega.org.uk/video/subseries/8

Feynmans theory states that the probability of an event is determined by summing together all the possible histories of that event. For example, for a particle moving from point A to B we imagine the particle traveling every possible path, curved paths, oscillating paths, squiggly paths, even backward in time and forward in time paths. Each path has an amplitude, and when summed the vast majority of all these amplitudes add up to zero, and all that remains is the comparably few histories that abide by the laws and forces of nature. Sum over histories indicates the direction of our ordinary clock time is simply a path in space which is more probable than the more exotic directions time might have taken otherwise.

Scalar random variables
Skalarne zmienne losowe

random variable / zmienna losowa

A random variable is a function $X(\xi)$ that associates a number with each outcome ξ of a certain experiment. For each outcome $\xi \in \Xi$ the probability $P(\xi)$ is defined.

$$X : \Xi \longrightarrow \Omega_X \subseteq \mathrm{R}, \qquad \{X(\xi), \xi \in \Xi\}$$

where:

Ξ	-	sample space / zbiór zdarzeń elementarnych
ξ	-	outcome / zdarzenie elementarne
Ω_X	-	observation space / zbiór obserwacji

realization of a random variable
realizacja zmiennej losowej

$$x = X(\xi)$$

Ξ is the set of all different possibilities that *could* happen

ξ is one of different possibilities that *did* happen

discrete random variable – when Ω_X takes finite or countably infinite number of values

continuous random variable – when Ω_X takes uncountably infinite number of values

Cumulative distribution function (CDF)
Dystrybuanta

$$F_X(x) = P(X \leq x)$$

$$P(\cdot), \quad 0 \leq P \leq 1$$

probability of an event / prawdopodobieństwo zdarzenia

Properties:

- $F_X(-\infty) = 0, \;\; F_X(\infty) = 1$
- $0 \leq F_X(x) \leq 1$
- if $x_1 < x_2$ then

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \geq 0$$

Probability density function (PDF)
Gęstość rozkładu prawdopodobieństwa

$$p_X(x) = \lim_{\epsilon \to 0} \frac{P(x \leq X < x + \epsilon)}{\epsilon} = \frac{dF_X(x)}{dx}$$

Properties:

- $p_X(x) \geq 0$
- $F_X(x) = \int_{-\infty}^{x} p_X(z)dz$
- $\int_{-\infty}^{\infty} p_X(x)dx = 1$
- $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} p_X(z)dz$

Is it possible that

$p_X(x) > 1$?

$P(X = x) \neq 0$?

Characteristics of random variables
Charakterystyki zmiennych losowych

mean (expected) value / wartość oczekiwana (średnia)

$$m_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x \, dF_X(x) = \int_{-\infty}^{\infty} x \, p_X(x) dx$$

median / mediana

$$\alpha_X = \mathrm{med}[X] : \; P(X \leq \alpha_X) = P(X \geq \alpha_X)$$
$$= \int_{-\infty}^{\alpha_X} p_X(x) dx = \frac{1}{2}$$

variance / wariancja

$$\sigma_X^2 = \mathrm{var}[X] = \mathrm{E}\left[(X - m_X)^2\right]$$
$$= \int_{-\infty}^{\infty} (x - m_X)^2 \, dF_X(x) = \int_{-\infty}^{\infty} (x - m_X)^2 \, p_X(x) dx$$

standard deviation / średnie odchylenie standardowe

$$\sigma_X = \sqrt{\mathrm{var}[X]}$$

Characteristics of random variables

n-th central moment / moment centralny n-tego rzędu

$$\mu_X^n = \mathrm{E}\left[(X - m_X)^n\right] = \int_{-\infty}^{\infty} (x - m_X)^n \, p_X(x) dx$$

coefficient of skewness / współczynnik asymetrii

$$\gamma_X = \frac{\mathrm{E}\left[(X - m_X)^3\right]}{\sigma_X^3} = \frac{\mu_X^3}{\sigma_X^3}$$

for a symmetric distribution $\gamma_X = 0$

coefficient of kurtosis (excess kurtosis) / kurtoza

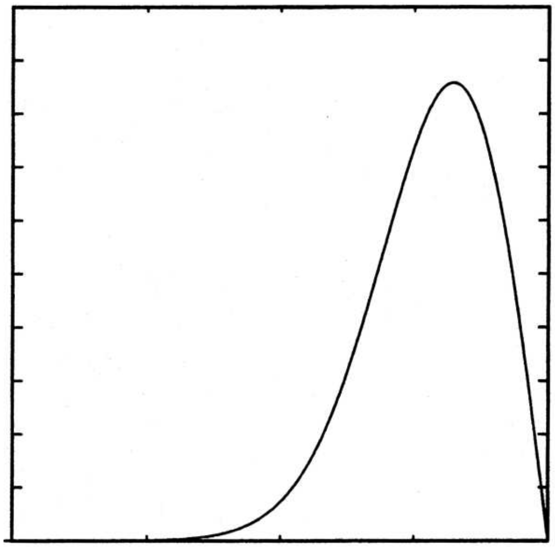
miara spłaszczenia rozkładu (od gr. $kurtos =$ rozdęty)

$$\kappa_X = \frac{\mathrm{E}\left[(X - m_X)^4\right]}{\sigma_X^4} - 3 = \frac{\mu_X^4}{\sigma_X^4} - 3$$

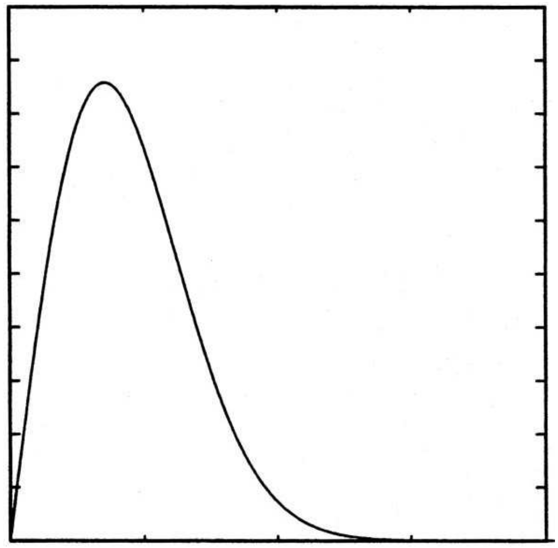
for a Gaussian distribution $\kappa_X = 0$

$\kappa_X > 0$ – leptocurtic distribution / rozkład leptokurtyczny
$\kappa_X = 0$ – mesocurtic distribution / rozkład mezokurtyczny
$\kappa_X < 0$ – platocurtic distribution / rozkład platokurtyczny

Characteristics of random variables

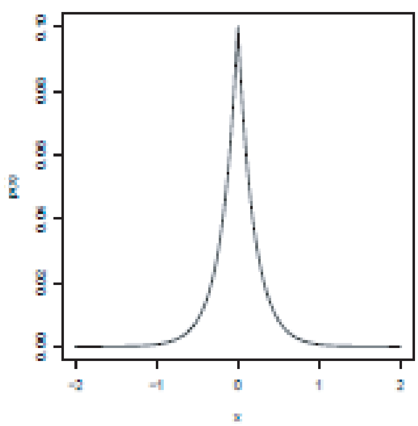


$\gamma_X < 0$

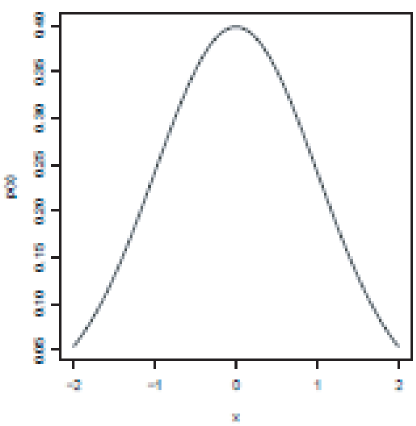


$\gamma_X > 0$

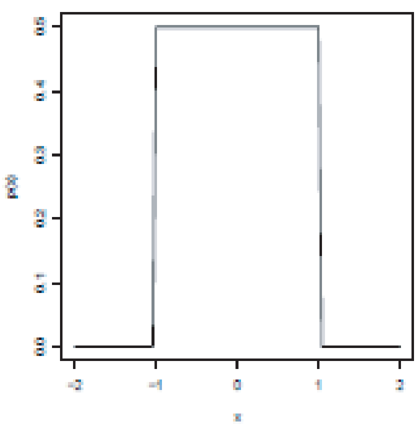
Probability density function with negative skew (left) and positive skew (right).



$\kappa_X = 3$



$\kappa_X = 0$



$\kappa_X = -1.5$

Super-Gaussian (leptocurtic) [left], Gaussian [middle] and sub-Gaussian (platycurtic) [right] probability density functions.

Important random variables
Ważne zmienne losowe

uniform random variable / zmienna losowa o rozkładzie równomiernym (jednostajnym)

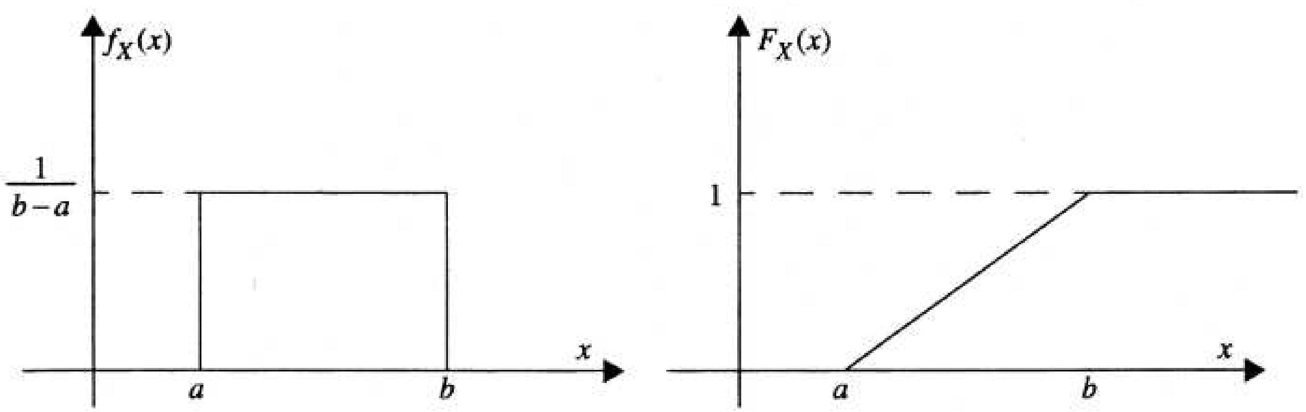
$$X \sim \mathcal{U}(a,b), \quad a < b$$

$$p_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x < b \\ 0 & \text{elsewhere} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

$$m_X = \frac{a+b}{2}, \quad \sigma_X^2 = \frac{(b-a)^2}{12}$$

MATLAB : $\mathcal{U}(0,1) = \texttt{rand}$



Probability density function (left) and cumulative distribution function (right) of a uniform random variable.

Important random variables

Gaussian (normal) random variable / zmienna losowa o rozkładzie gaussowskim (normalnym)

$$X \sim \mathcal{N}(m_X, \sigma_X^2), \quad \sigma_X^2 > 0$$

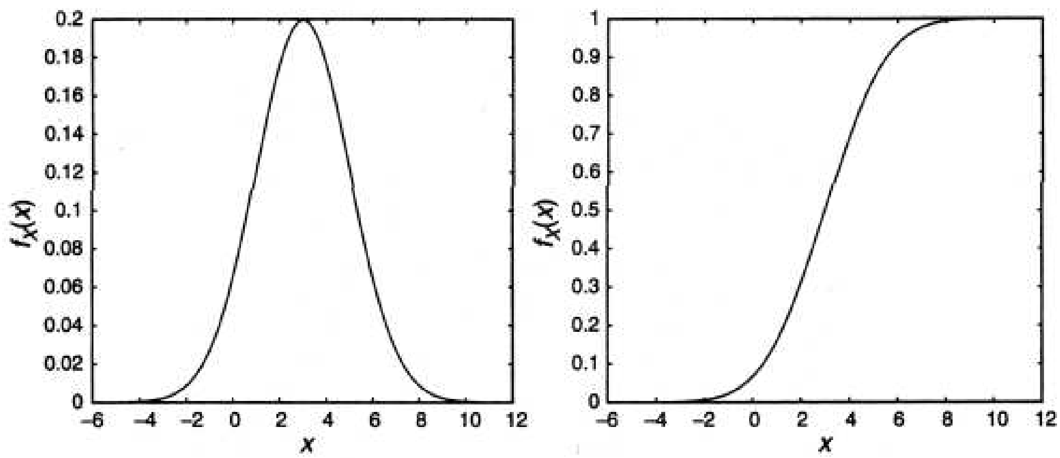
$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x-m_X)^2}{2\sigma_X^2}\right\}$$

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(z-m_X)^2}{2\sigma_X^2}\right\} dz$$

$$= \Phi\left(\frac{x-m_X}{\sigma_X}\right)$$

where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} \exp\{-z^2/2\} dz$

MATLAB : $\mathcal{N}(0,1) = \texttt{randn}$



Probability density function (left) and cumulative distribution function (right) of a Gaussian random variable.

Central limit theorem (CLT)
Centralne twierdzenie graniczne

THEOREM

Let $X_i, i = 1, 2, \ldots$, be a sequence of independent random variables with identical (any!) distribution, mean m_X and variance $\sigma_X^2 < \infty$. Define a new random variable Z_n:

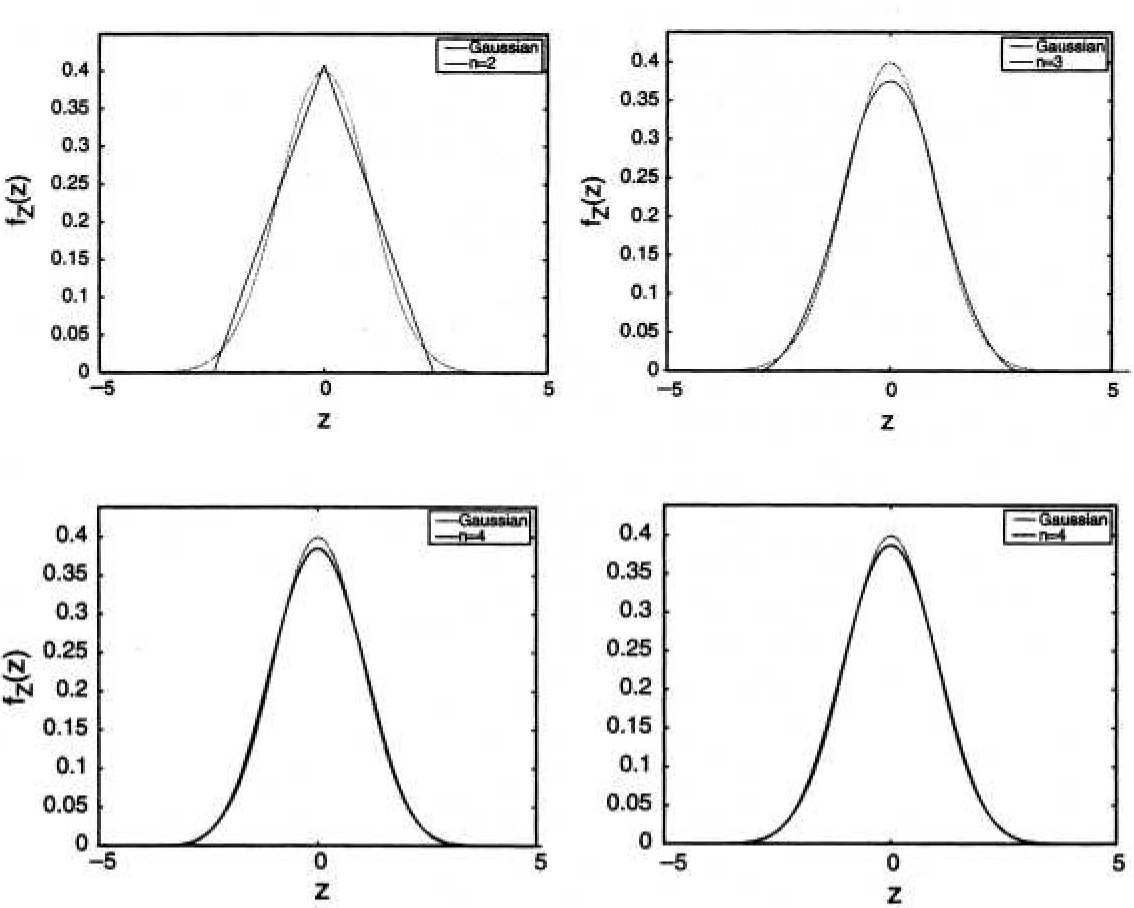
$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m_X}{\sigma_X}$$

In the limits as n approaches infinity, the random variable Z_n converges in distribution to a standard normal random variable $\mathcal{N}(0, 1)$:

$$\lim_{n\to\infty} p_{Z_n}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m_X}{\sigma_X}\right) = \mathcal{N}(0, 1)$$

CLT explains why the Gaussian distribution is of such a great importance, and why it occurs so frequently

The central limit theorem (CLT)



PDF of the sum of independent uniform random variables:
$n = 1, 2, 3, 4.$

Important random variables

Laplace random variable / zmienna losowa o rozkładzie Laplace'a (dwustronnym wykładniczym)

$$X \sim \mathcal{L}(a, b), \quad b > 0$$

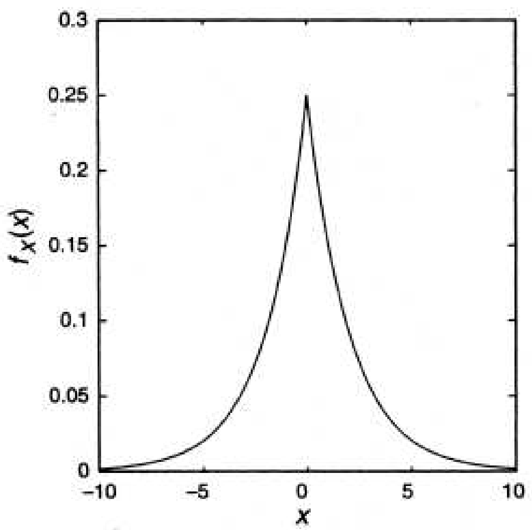
a – location parameter, b – scale parameter

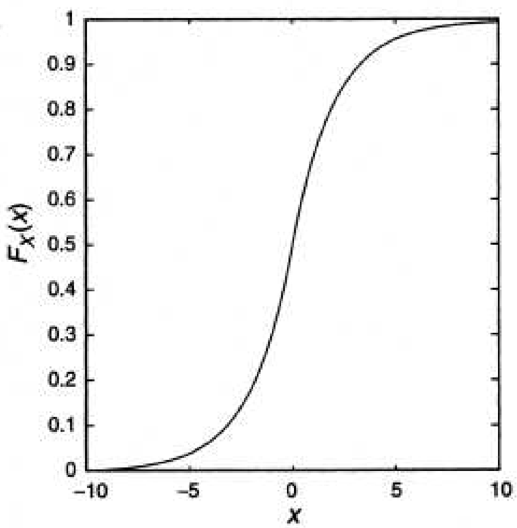
$$p_X(x) = \frac{1}{2b} \exp\left\{ -\frac{|x-a|}{b} \right\}$$

$$F_X(x) = \begin{cases} \frac{1}{2}\exp\left(\frac{x}{b}\right) & x < 0 \\ 1 - \frac{1}{2}\exp\left(\frac{x}{b}\right) & x \geq 0 \end{cases}$$

$$m_X = a\ , \quad \sigma_X^2 = b^2/2$$

$$Y \sim \mathcal{U}(-\tfrac{1}{2}, \tfrac{1}{2}) \Rightarrow X = \operatorname{sgn}[Y] \ln(1 - 2|Y|) \sim \mathcal{L}(0, 1)$$





Probability density function (left) and cumulative distribution function (right) of a Laplace random variable.

Important random variables

Cauchy random variable / zmienna losowa o rozkładzie Cauchy'ego

$$X \sim \mathcal{C}(a, b), \quad b > 0$$

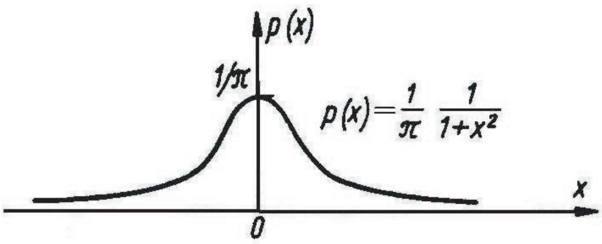
a – location parameter, b – scale parameter

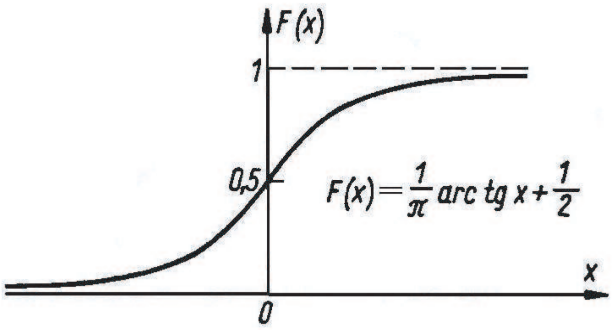
$$p_X(x) = \frac{b/\pi}{b^2 + (x-a)^2}$$

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{x-a}{b}\right)$$

$$m_X = ? \, , \quad \sigma_X^2 = ?$$

$$Y \sim \mathcal{N}(0,1), \, Z \sim \mathcal{N}(0,1) \Rightarrow X = Y/Z \sim \mathcal{C}(0,1)$$





Probability density function (left) and cumulative distribution function (right) of a Cauchy random variable.

Cauchy random variables

The Cauchy distribution is an example of a distribution which has no mean, variance or higher moments defined. Its median is well defined and equal to a.

standard Cauchy distribution / standardowy rozkład Cauchy'ego

$$X \sim \mathcal{C}(0,1): \quad p_X(x) = \frac{1}{\pi(1+x^2)}$$

mean value:

$$\int_{-\infty}^{\infty} x\, p_X(x) dx = \int_{0}^{\infty} x\, p_X(x) dx - \int_{-\infty}^{0} |x|\, p_X(x) dx = \infty - \infty$$

note that due to symmetry of $p_X(x)$

$$\lim_{a\to\infty} \int_{-a}^{a} x\, p_X(x) dx \neq \lim_{a\to\infty} \int_{-2a}^{a} x\, p_X(x) dx$$

variance:

$$\mathrm{E}[X^2] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x^2}{1+x^2}\, dx = \frac{1}{\pi}\int_{-\infty}^{\infty} dx - 1 = \infty$$

Pairs of random variables
Pary zmiennych losowych

$$X(\xi) \longleftrightarrow Y(\xi)$$

joint cumulative distribution function / łączna dystrybuanta

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

Properties:

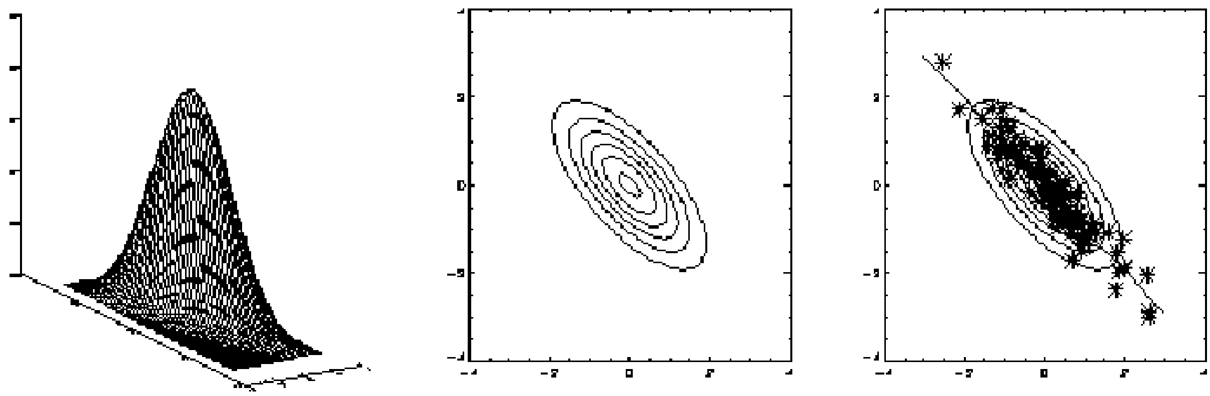
- $F_{XY}(-\infty,-\infty) = 0$
- $F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0$
- $F_{XY}(\infty,\infty) = 1$
- $0 \leq F_{XY}(x,y) \leq 1$
- $F_{XY}(x,\infty) = F_X(x)$
- $F_{XY}(\infty,y) = F_Y(y)$
- if $x_1 < x_2$ and $y_1 < y_2$ then

$$\begin{aligned} &P(x_1 < X \leq x_2, y_1 < Y \leq y_2) \\ &= F_{XY}(x_2,y_2) - F_{XY}(x_1,y_2) \\ &- F_{XY}(x_2,y_1) + F_{XY}(x_1,y_1) \geq 0 \end{aligned}$$

Pairs of random variables

joint probability density function / funkcja łącznej gęstości prawdopodobieństwa

$$p_{XY}(x,y)$$
$$= \lim_{\epsilon_x \to 0, \epsilon_y \to 0} \frac{P(x \le X < x+\epsilon_x, y \le Y < y+\epsilon_y)}{\epsilon_x \epsilon_y}$$
$$= \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$$



Joint probability density function two of random variables.

Properties:

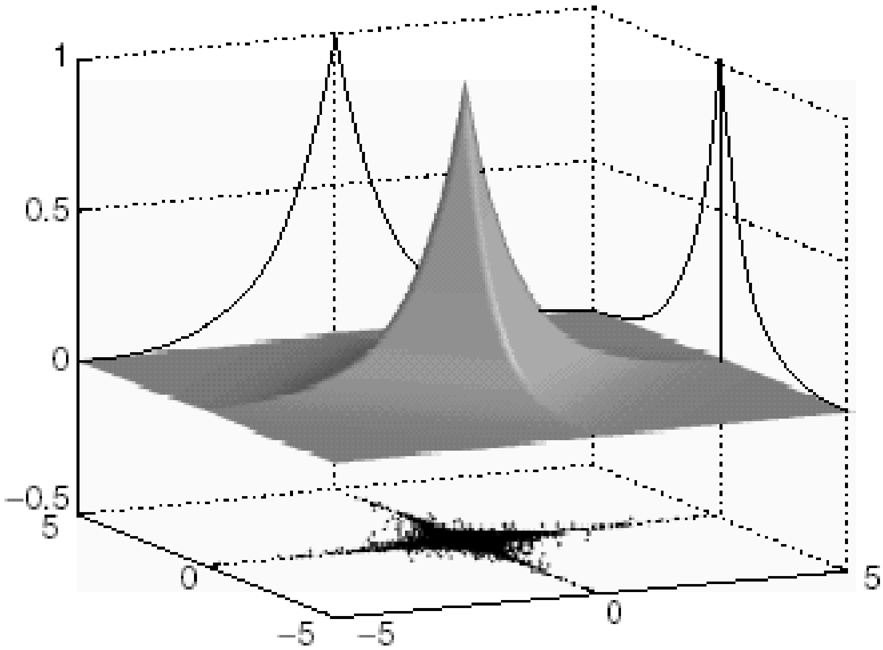
- $p_{XY}(x,y) \ge 0$
- $F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p_{XY}(x,y) dx dy$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{XY}(x,y) dx dy = 1$

- if $x_1 < x_2$ and $y_1 < y_2$ then

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

marginal distributions / rozkłady brzegowe

$$\int_{-\infty}^{\infty} p_{XY}(x,y)dy = p_X(x)$$
$$\int_{-\infty}^{\infty} p_{XY}(x,y)dx = p_Y(y)$$



Joint distribution of two random variables and their marginal distributions

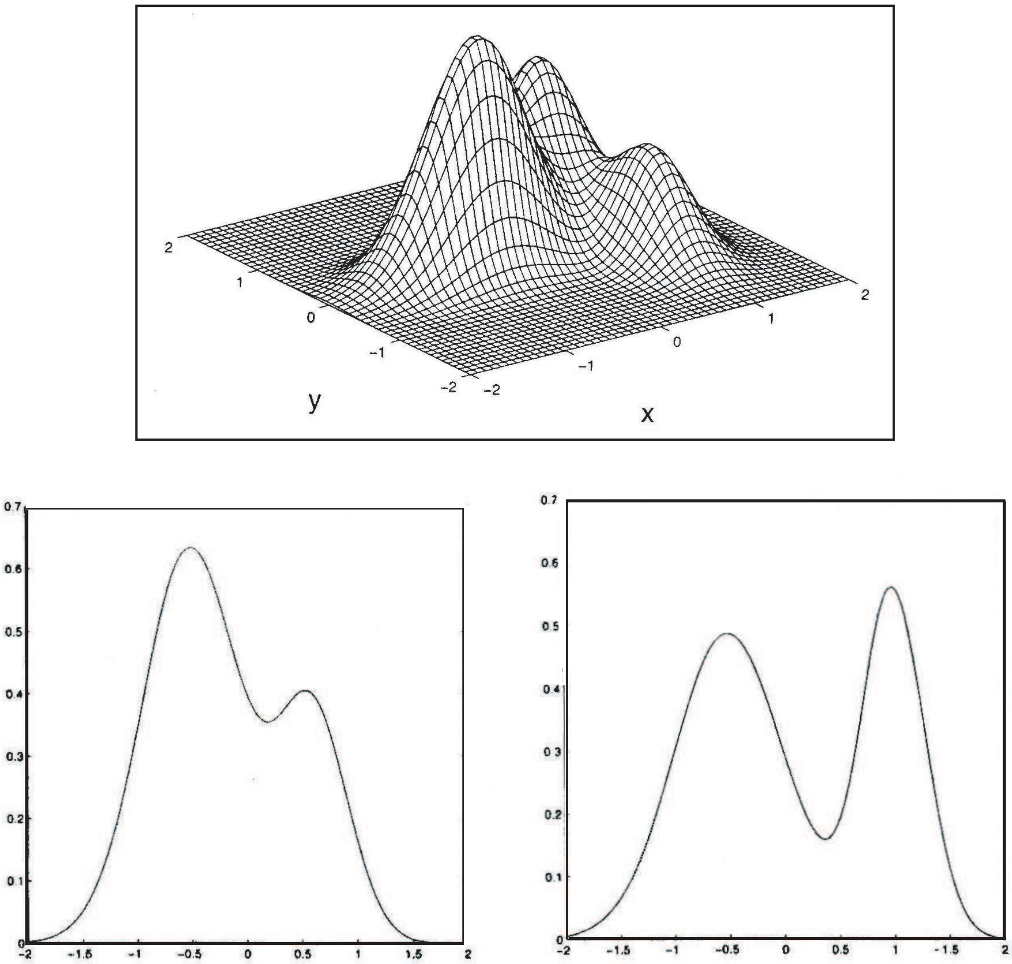
note: there exist many different joint distributions that produce the same marginal distributions

Pairs of random variables

conditional probability density function / warunkowa gęstość rozkładu prawdopodobieństwa

$$X|Y(\xi) = y$$

$$p_{X|Y}(x|y) = \frac{p_{XY}(x,y)}{p_Y(y)}$$



Joint distribution of random variables X and Y (top) and two conditional distributions: $p_{Y|X}(y|x = 1.27)$ (bottom left) and $p_{X|Y}(x|y = -0.37)$ (bottom right).

correlation between random variables / współczynnik korelacji zmiennych losowych

$$r_{XY} = \mathrm{E}[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, p_{XY}(x,y)dxdy$$

covariance between random variables / współczynnik kowariancji zmiennych losowych

$$\begin{aligned} c_{XY} &= \mathrm{E}[(X-m_X)(Y-m_Y)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_X)(y-m_Y)\, p_{XY}(x,y)dxdy \\ &= r_{XY} - m_X m_Y \end{aligned}$$

correlation coefficient of two random variables / unormowany współczynnik kowariancji dwóch zmiennych losowych

normalized random variables / unormowane zmienne losowe

$$\widetilde{X} = \frac{X-m_X}{\sigma_X}\ , \quad \widetilde{Y} = \frac{Y-m_Y}{\sigma_Y}$$

$$m_{\widetilde{X}} = m_{\widetilde{Y}} = 0\ , \quad \sigma^2_{\widetilde{X}} = \sigma^2_{\widetilde{Y}} = 1$$

$$\rho_{XY} = \mathrm{E}[\widetilde{X}\widetilde{Y}] = \frac{c_{XY}}{\sigma_X \sigma_Y}$$

$$-1 \le \rho_{XY} \le 1$$

Pairs of random variables

THEOREM

The correlation coefficient is less than 1 in magnitude.

PROOF

Consider taking the second moment of $X + aY$, where a is a real constant:

$$\mathrm{E}[(X + aY)^2] = \mathrm{E}[X^2] + 2a\mathrm{E}[XY] + a^2\mathrm{E}[Y^2] \geq 0.$$

Since this is true for any a, we can tighten the bound by choosing the value of a that minimizes the left-hand side. This value turns out to be $a = -\mathrm{E}[XY]/\mathrm{E}[Y^2]$. Plugging in this value gives

$$\begin{aligned} \mathrm{E}[X^2] + \frac{(\mathrm{E}[XY])^2}{\mathrm{E}[Y^2]} - 2\,\frac{(\mathrm{E}[XY])^2}{\mathrm{E}[Y^2]} \\ = \mathrm{E}[X^2] - \frac{(\mathrm{E}[XY])^2}{\mathrm{E}[Y^2]} \geq 0. \end{aligned}$$

After replacing X with $X - m_X$ and Y with $Y - m_Y$, one arrives at

$$\rho_{XY}^2 \leq 1$$

which is the desired result.

Pairs of random variables

orthogonal random variables / ortogonalne zmienne losowe

$$r_{XY} = 0$$

uncorrelated random variables / nieskorelowane zmienne losowe

$$c_{XY} = 0$$

independent random variables / niezależne zmienne losowe

$$F_{XY}(x, y) = F_X(x)F_Y(y), \quad p_{XY}(x, y) = p_X(x)p_Y(y)$$

note: independent random variables are always uncorrelated but the converse is not true.

<u>EXAMPLE</u> (independence versus uncorrelatedness)

Case 1

Consider random variables X and Y that are uniformly distributed on the square defined by $0 \leq x, y \leq 1$, i.e.,

$$p_{XY}(x, y) = \begin{cases} 1 & 0 \leq x, y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

The marginal probability density functions of X and Y turn out to be

$$p_X(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases} \qquad p_Y(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

These random variables are statistically independent since $p_{XY}(x, y) = p_X(x)p_Y(y)$.

Pairs of random variables

Case 2

Consider random variables X and Y that are uniformly distributed over the unit circle, i.e.,

$$p_{XY}(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

The marginal probability density function of X can be found as follows

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x,y)dy = \frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy$$

$$= \frac{2}{\pi}\sqrt{1-x^2}, \quad -1 \leq x \leq 1$$

By symmetry, the marginal PDF of Y must take on the same functional form. Hence

$$p_X(x)p_Y(y) = \frac{4}{\pi^2}\sqrt{(1-x^2)(1-y^2)} \neq p_{XY}(x,y)$$

The correlation between X and Y is

$$\mathrm{E}[XY] = \int_{x^2+y^2\leq 1} \frac{xy}{\pi}\, dxdy$$

$$= \frac{1}{\pi}\int_{-1}^{1} x \left[\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} ydy\right] dx = 0$$

Hence X and Y are uncorrelated but not independent !

Jointly Gaussian random variables
Zmienne losowe o łącznym rozkładzie gaussowskim

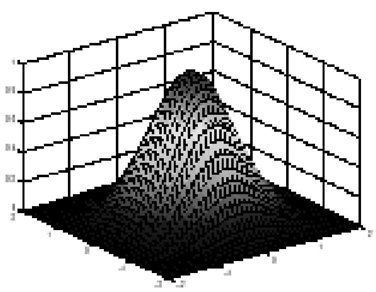
$$p_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \times$$

$$\exp\left\{-\frac{\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-m_X}{\sigma_X}\right)\left(\frac{y-m_Y}{\sigma_Y}\right) + \left(\frac{y-m_Y}{\sigma_Y}\right)^2}{2(1-\rho_{XY}^2)}\right\}$$

marginal distributions of a bivariate Gaussian distribution are also Gaussian

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x-m_X)^2}{2\sigma_X^2}\right\}$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left\{-\frac{(y-m_Y)^2}{2\sigma_Y^2}\right\}$$



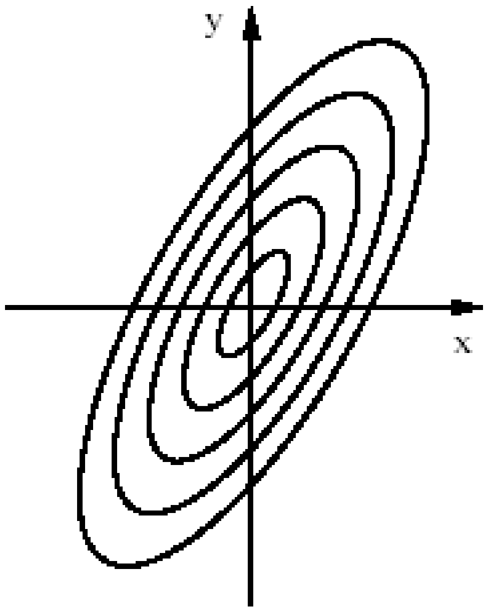
Probability density function of two jointly Gaussian random variables.

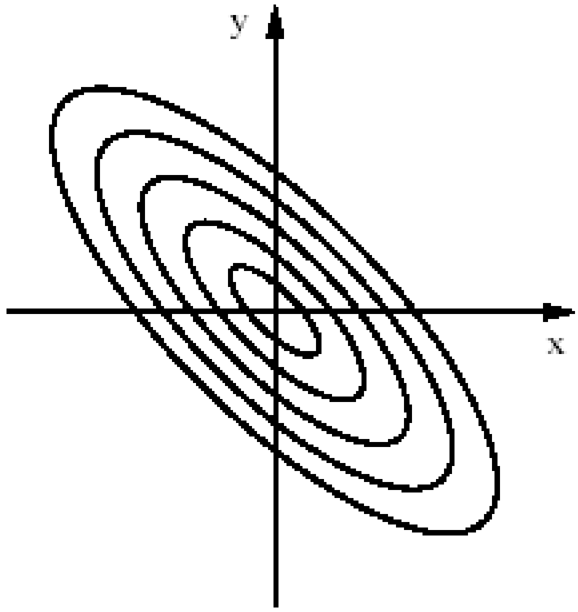
Jointly Gaussian random variables

equidensity contours / izolinie rozkładu

$$\left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x-m_X}{\sigma_X}\right)\left(\frac{y-m_Y}{\sigma_Y}\right) + \left(\frac{y-m_Y}{\sigma_Y}\right)^2 = 1-\rho_{XY}^2$$

covariance ellipse / elipsa kowariancji





Contours of the bivariate Gaussian PDF corresponding to the positive (left) and negative (right) correlation coefficient ρ_{XY}.

note: for $m_X = m_Y = 0$ and $\rho_{XY} = 0$, one obtains

$$\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2} = 1$$

Jointly Gaussian random variables

covariance ellipse is a counterpart of the interval $[m_X - \sigma_X, m_X + \sigma_X]$ for a one-dimensional Gaussian distribution

$$\text{if } X \sim \mathcal{N}(m_X, \sigma_X^2) \text{ then}$$
$$P(X \in [m_X - \sigma_X, m_X + \sigma_X]) = 0.68$$

$$\text{if } (X,Y) \sim \mathcal{N}(m_X, m_Y; \sigma_X^2, \sigma_Y^2, \rho_{XY}) \text{ then}$$
$$P[(X,Y) \in D_{XY}] = 0.68$$

where D_{XY} denotes the two-dimensional region bounded by the covariance ellipse

conditional distribution / rozkład warunkowy

$$[X|Y(\xi) = y] \sim \mathcal{N}(m_{X|Y}, \sigma_{X|Y}^2)$$
$$m_{X|Y} = m_X + \frac{\sigma_X}{\sigma_Y}\,\rho_{XY}(y - m_Y)$$
$$\sigma_{X|Y}^2 = (1 - \rho_{XY}^2)\sigma_X^2$$

Jointly Gaussian random variables

note: two uncorrelated jointly Gaussian variables are independent

$$\text{if } \rho_{XY} = 0 \text{ then}$$

$$\begin{aligned}
p_{XY}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{ -\frac{\left(\frac{x-m_X}{\sigma_X}\right)^2 + \left(\frac{y-m_Y}{\sigma_Y}\right)^2}{2} \right\} \\
&= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{ -\frac{(x-m_X)^2}{2\sigma_X^2} \right\} \\
&\times \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left\{ -\frac{(y-m_Y)^2}{2\sigma_Y^2} \right\} \\
&= p_X(x) p_Y(y)
\end{aligned}$$

note: two normally distributed random variables need not be jointly normal – normally distributed and uncorrelated does not imply independent

<u>EXAMPLE</u>

Let $X \sim \mathcal{N}(0,1)$. Define Y in the form

$$Y = \begin{cases} -X & \text{if } |X| < c \\ X & \text{if } |X| \geq c \end{cases} \qquad c > 0$$

Jointly Gaussian random variables

Note that

1. Y is Gaussian – $Y \sim \mathcal{N}(0,1)$ – because

   $$\begin{aligned}
   &P(Y \le x) = \\
   &P(\{|X| < c \text{ and } -X < x\} \text{ or } \{|X| \ge c \text{ and } X < x\}) \\
   &= P(|X| < c \text{ and } -X < x) + P(|X| \ge c \text{ and } X < x) \\
   &= P(|X| < c \text{ and } X < x) + P(|X| \ge c \text{ and } X < x) \\
   &= P(X < x)
   \end{aligned}$$

   This follows from the symmetry of the distribution of X and the symmetry of the condition that $|X| < c$.

2. Since X completely determines Y, X and Y are not independent.

3. There exists such $c > 0$ for which X and Y are uncorrelated. If c is very small, then the correlation c_{XY} is close to 1. If c is very large, then c_{XY} is close to -1. Since the correlation is a continuous function of c, the intermediate value theorem implies there is some particular value of c that makes the correlation 0. That value is approximately 1.54.

Independent component analysis (ICA)
Analiza składowych niezależnych

Consider two independent and non-Gaussian random variables X_1 and X_2. Suppose we can observe realizations of two other random variables that are mixtures of X_1 and X_2:

$$Y_1 = a_{11}X_1 + a_{12}X_2$$
$$Y_2 = a_{21}X_1 + a_{22}X_2$$

but we do not know the values of the mixing coefficients $a_{11}, a_{12}, a_{21}, a_{22}$.

Given $y_1 = Y_1(\xi)$ and $y_2 = Y_2(\xi)$ can we find out $x_1 = X_1(\xi)$ and $x_2 = X_2(\xi)$?

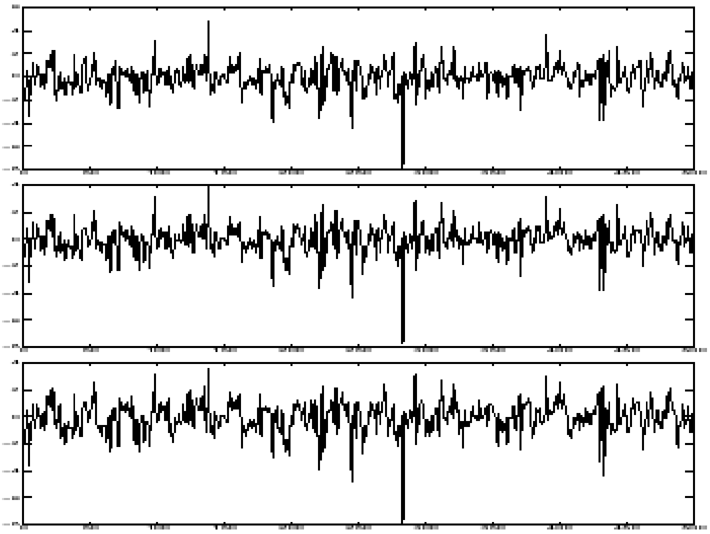
Real-world example: If two people speak at the same time in a room containing two microphones, the output of each microphone is a mixture of two voice signals. Given these two *signal mixtures* can we recover the two original voices or *source signals*?

blind source separation / ślepa separacja sygnałów

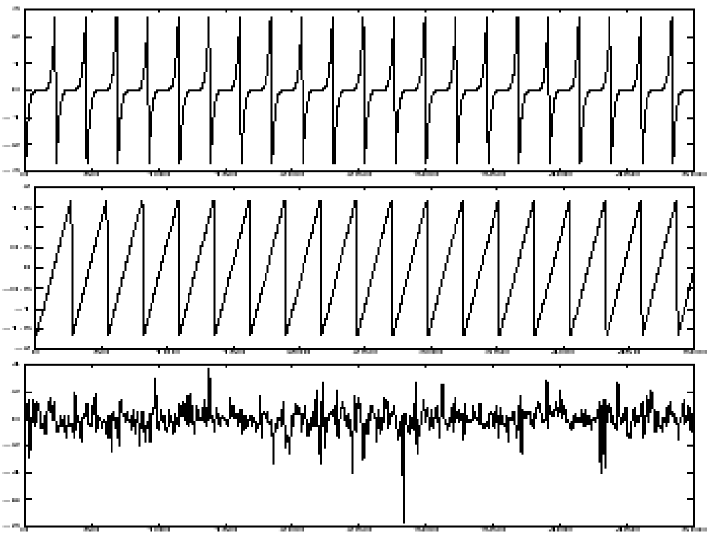
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

But how would we know $a_{11}, a_{12}, a_{21}, a_{22}$?

What results can ICA provide ?



input signals (mixtures)



output signals (sources)

Limitations:

- We cannot determine the variances (energies) of the independent components.

- We cannot determine the order of the independent components.

What makes it possible ?

Jak to możliwe ?

sources:

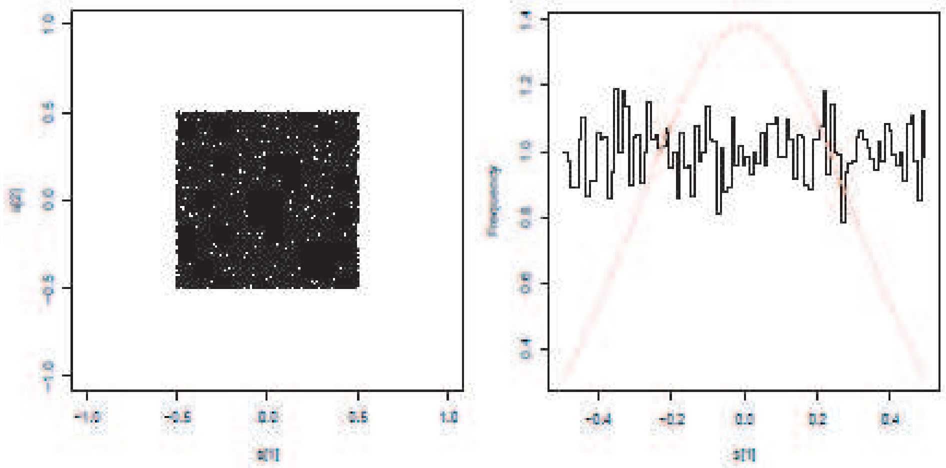
$$X_1 \sim \mathcal{U}(-0.5, 0.5),\ X_2 \sim \mathcal{U}(-0.5, 0.5)$$

$$m_{X_1} = m_{X_2} = 0,\ \sigma^2_{X_1} = \sigma^2_{X_2} = 1/12,\ c_{X_1 X_2} = 0$$

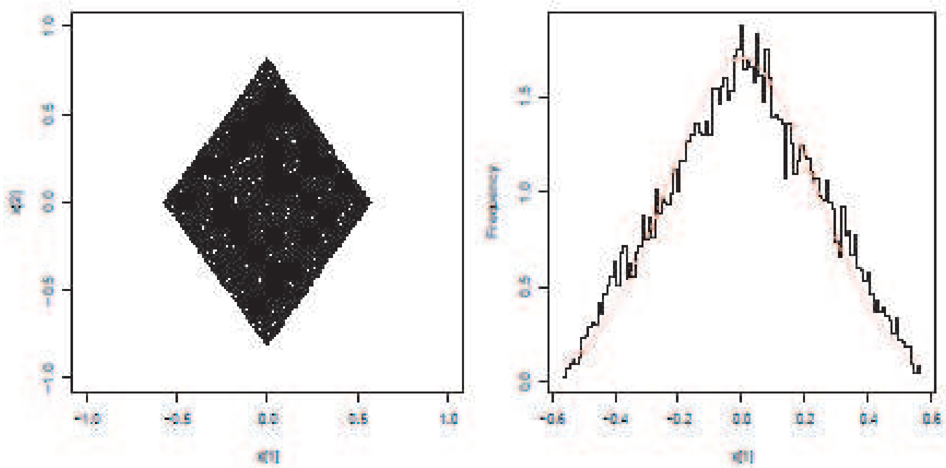
mixtures:

$$Y_1 = \frac{1}{\sqrt{2}}\, X_1 + \frac{1}{\sqrt{2}}\, X_2\ ,\ Y_2 = \frac{1}{\sqrt{2}}\, X_1 - \frac{1}{\sqrt{2}}\, X_2$$

$$m_{Y_1} = m_{Y_2} = 0,\ \sigma^2_{Y_1} = \sigma^2_{Y_2} = 1/12,\ c_{Y_1 Y_2} = 0$$



(a) Independent sources (b) Histogram of the sources



(c) Linear mixture of the sources (d) Histogram of the mixes

One of the possible approaches to ICA

Consider $n > 1$ sources and n mixtures

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j, \quad i = 1, \ldots, n$$

Two-step procedure:

- STEP 1: Prewhite the available measurement data - the new random variables $\widetilde{Y}_1, \ldots, \widetilde{Y}_n$ should be (approximately) mutually uncorrelated

$$\widetilde{y}_i = \textstyle\sum_{j=1}^{n} b_{ij}(y_j - m_{Y_j}),\ \mathrm{var}[\widetilde{Y}_i] = 1,\ i = 1, \ldots, n$$
$$c_{\widetilde{Y}_i \widetilde{Y}_j} = 0,\ i \neq j$$

- STEP 2: Recover source signals using the relationship

$$\widehat{x}_i = \sum_{j=1}^{n} \widehat{c}_{ij} \widetilde{y}_j, \quad i = 1, \ldots, n$$

where $\{\widehat{c}_{i1}, \ldots, \widehat{c}_{in}\}$ are the coefficients chosen (iteratively) so as to maximize non-Gaussianity of $\widehat{X}_i = \sum_{j=1}^{n} \widehat{c}_{ij} \widetilde{Y}_j$:

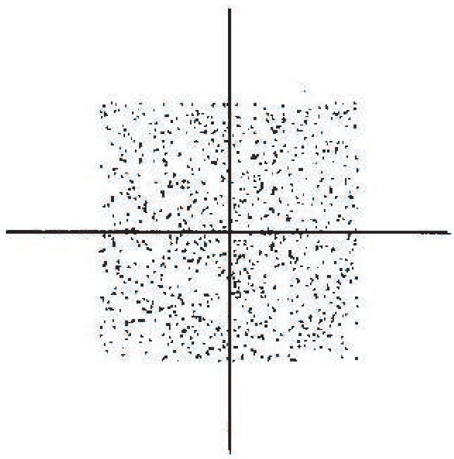
$$\{\widehat{c}_{i1}, \ldots, \widehat{c}_{in}\} = \arg \max_{c_{i1}, \ldots, c_{in}} \left| \kappa_{\widehat{X}_i} \left(\sum_{j=1}^{n} c_{ij} \widetilde{y}_j \right) \right|$$

FastICA algorithm

Picture guide to ICA

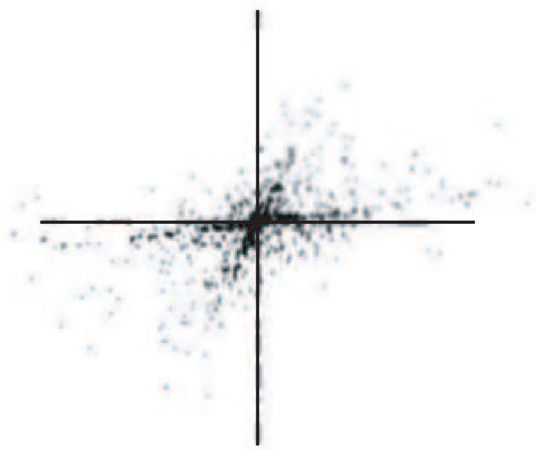
joint distribution of input signals (mixtures)

joint distribution of whitened mixtures

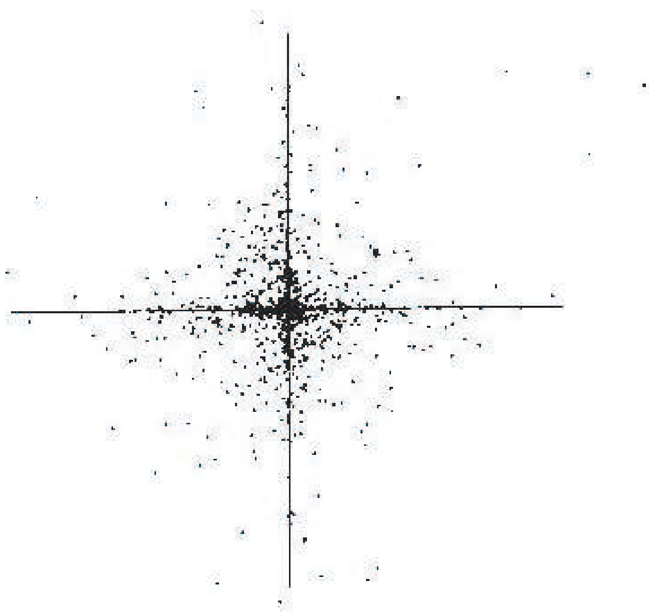


joint distribution of output signals (sources)

Picture guide to ICA

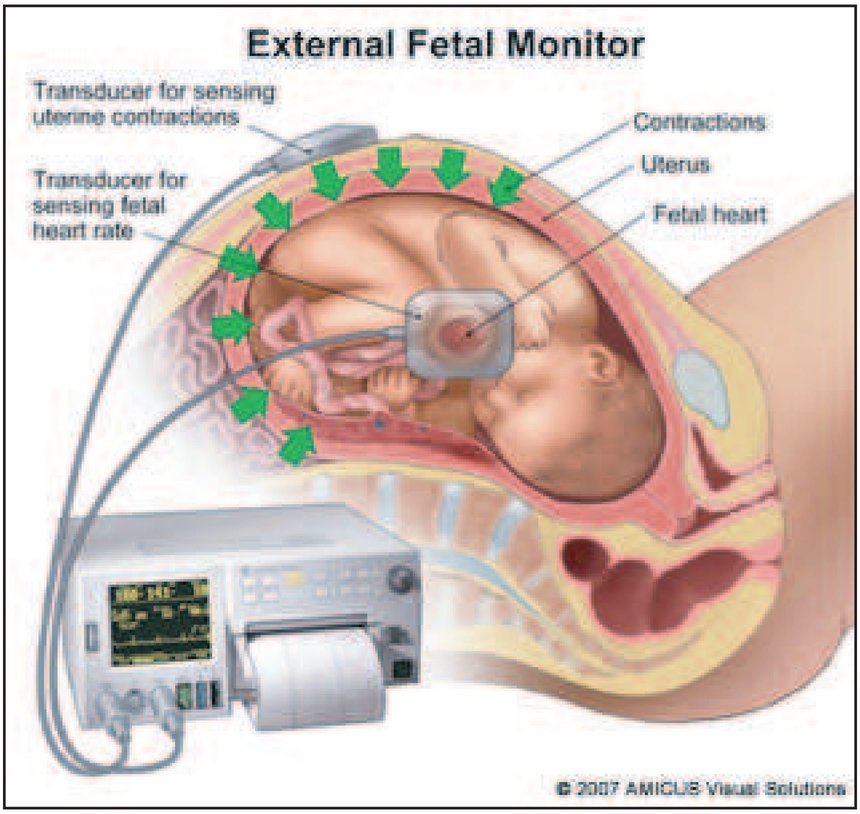


joint distribution of a mixture of Laplace variables



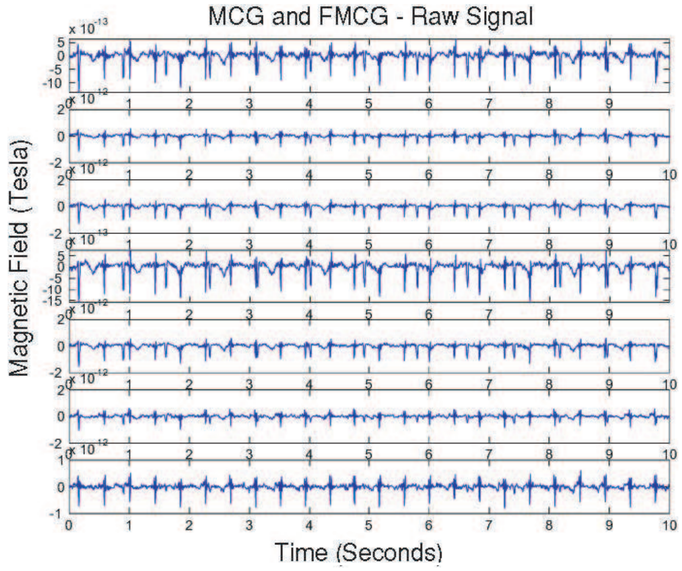
joint distribution of independent Laplace variables

Example of ICA application
- fetal hart monitoring

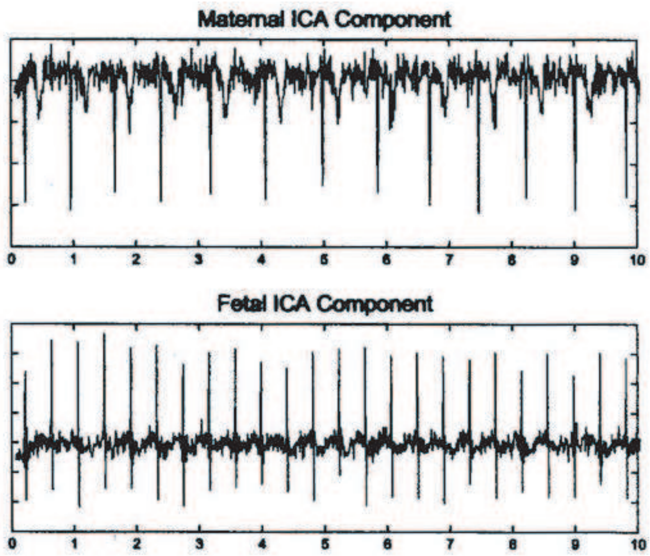


The magnetic field generated by electrical activity in the heart is measured at 37 different locations using fetal megnetocardiography (FMCG). Each measured signal is a different mixture of fetal and maternal cardiac signals.

Example of ICA application
- fetal hart monitoring



Detail of 7 out of 37 measured signals



Separated fetal and maternal cardiac signals

Vector random variables
Wektorowe zmienne losowe

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}_{n\times 1}, \quad \mathbf{x} = \mathbf{X}(\xi)$$

mean (expected) value / wartość oczekiwana (średnia)

$$\mathbf{m}_X = \mathrm{E}[\mathbf{X}] = \begin{bmatrix} m_{X_1} \\ \vdots \\ m_{X_n} \end{bmatrix}_{n\times 1}$$

covariance matrix / covariance matrix

$$\mathbf{\Sigma}_X = \mathrm{cov}[\mathbf{X}] = \mathrm{E}\left[(\mathbf{X}-\mathbf{m}_X)(\mathbf{X}-\mathbf{m}_X)^{\mathrm{T}}\right]$$
$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}_{n\times n}$$

where $\sigma_{ij} = \mathrm{E}[(X_i - m_{X_i})(X_j - m_{X_j})] = c_{X_iX_j}$
and $\sigma_i^2 = \mathrm{var}[X_i] = c_{X_iX_i}$

note: covariance matrix in nonnegative definite

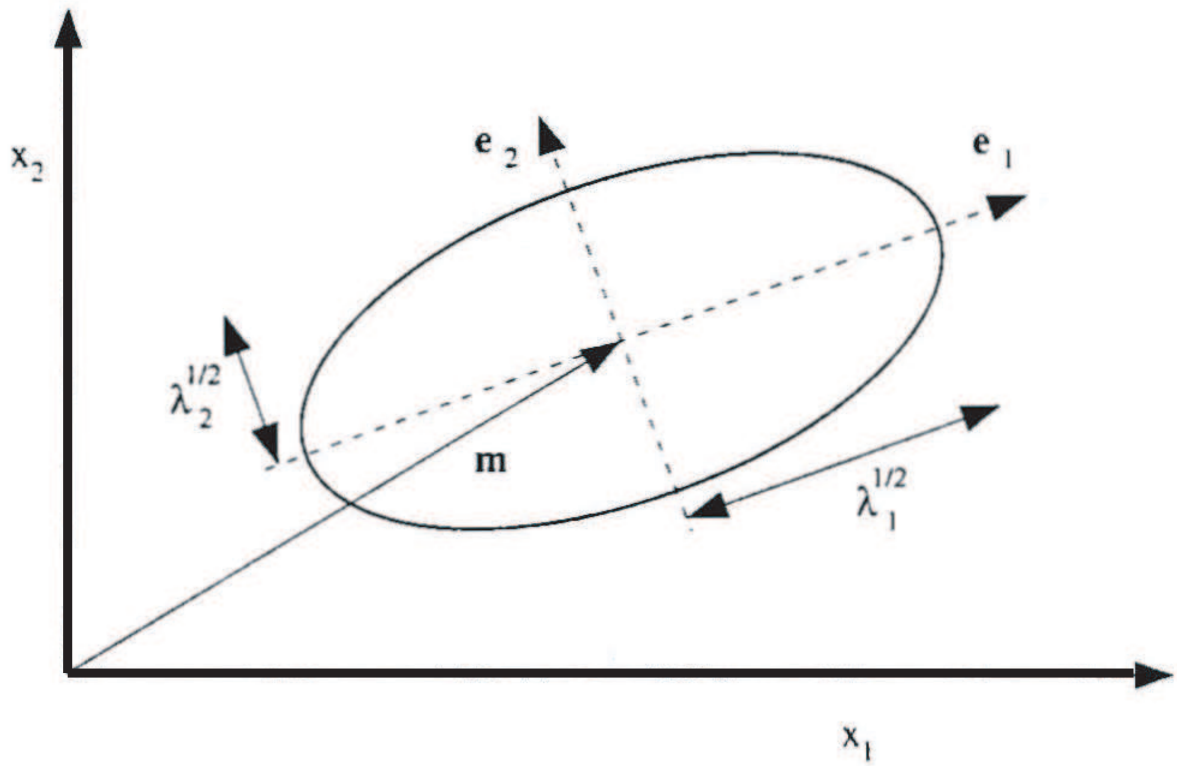
$$\mathbf{\Sigma}_X \geq 0 \iff \mathbf{y}^{\mathrm{T}}\mathbf{\Sigma}_X\mathbf{y} \geq 0, \quad \forall \mathbf{y}$$

Vector random variables

covariance ellipsoid / elipsoida kowariancji

$$(\mathbf{x} - \mathbf{m}_X)^{\mathrm{T}} \mathbf{\Sigma}_X^{-1} (\mathbf{x} - \mathbf{m}_X) = 1$$

- ellipsoid centered at $\mathbf{m}_X$
- the directions of the principal axes are given by the eigenvectors of the covariance matrix $\mathbf{\Sigma}_X$
- the squared lengths of the semi-principal axes are given by the corresponding eigenvalues



Covariance ellipse ($n = 2$).

partial ordering of covariance matrices

$$\boldsymbol{\Sigma}_X \overset{?}{<} \boldsymbol{\Sigma}_Y, \quad \boldsymbol{\Sigma}_X \overset{?}{\leq} \boldsymbol{\Sigma}_Y, \quad \boldsymbol{\Sigma}_X \overset{?}{>} \boldsymbol{\Sigma}_Y, \quad \boldsymbol{\Sigma}_X \overset{?}{\geq} \boldsymbol{\Sigma}_Y$$

$$D_X = \left\{ \mathbf{x} : \ \mathbf{x}^{\mathrm{T}} \boldsymbol{\Sigma}_X^{-1} \mathbf{x} \leq 1 \right\}$$

$$D_Y = \left\{ \mathbf{y} : \ \mathbf{y}^{\mathrm{T}} \boldsymbol{\Sigma}_Y^{-1} \mathbf{y} \leq 1 \right\}$$

$$\boldsymbol{\Sigma}_X \leq \boldsymbol{\Sigma}_Y \iff \boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_X \geq 0 \iff D_X \subseteq D_Y$$

covariance ellipsoid of $\boldsymbol{\Sigma}_X$ is located inside (possibly touching) the covariance ellipsoid of $\boldsymbol{\Sigma}_Y$

singular distributions / rozkłady osobliwe

singular distribution is a probability distribution concentrated on an uncountable set $\Omega_X \subset R^n$ of Lebesque measure 0

- distribution defined on a line in a two-dimensional space
- distribution defined on a circle in a two-dimensional space
- distribution defined on a plane in a three-dimensional space
- distribution defined on a sphere in a three-dimensional space

Vector random variables

multivariate Gaussian (normal) distribution / wielowymiarowy rozkład gaussowski (normalny)

$$\mathbf{X} \sim \mathcal{N}(\mathbf{m}_X, \mathbf{\Sigma}_X), \quad \mathbf{\Sigma}_X > 0$$

$$p_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}_X|}} \times \exp\left\{ -\frac{1}{2}(\mathbf{x} - \mathbf{m}_X)^{\mathrm{T}} \mathbf{\Sigma}_X^{-1} (\mathbf{x} - \mathbf{m}_X) \right\}$$

$|\mathbf{\Sigma}_X| = 0$: singular Gaussian distribution

note: the equidensity contours of a nonsingular multivariate Gaussian distribution are ellipsoids centered at the mean

$$P\left([\mathbf{X} - \mathbf{m}_X]^{\mathrm{T}} \mathbf{\Sigma}_X^{-1} [\mathbf{X} - \mathbf{m}_X] \le 1 \right) = 0.68$$

conditional distribution / rozkład warunkowy

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \begin{matrix} n \\ m \end{matrix} \sim \mathcal{N}(\mathbf{m}_Z, \mathbf{\Sigma}_Z)$$

$$\mathbf{m}_Z = \begin{bmatrix} \mathbf{m}_X \\ \mathbf{m}_Y \end{bmatrix}, \quad \mathbf{\Sigma}_Z = \begin{bmatrix} \mathbf{\Sigma}_X & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_Y \end{bmatrix} > 0$$

Vector random variables

$$\mathbf{\Sigma}_{XY} = \mathrm{E}\left[(\mathbf{X}-\mathbf{m}_X)(\mathbf{Y}-\mathbf{m}_Y)^{\mathrm{T}}\right] = \mathbf{\Sigma}_{YX}^{\mathrm{T}}$$

cross-covariance matrix between random variables $\mathbf{X}$ and $\mathbf{Y}$

$$[\mathbf{X}|\mathbf{Y}(\xi)=\mathbf{y})] \sim \mathcal{N}(\mathbf{m}_{X|Y}, \mathbf{\Sigma}_{X|Y})$$
$$\mathbf{m}_{X|Y} = \mathbf{m}_X + \mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}(\mathbf{y}-\mathbf{m}_Y)$$
$$\mathbf{\Sigma}_{X|Y} = \mathbf{\Sigma}_X - \mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{YX}$$

higher-order moments / momenty wyższych rzędów

higher-order moments of jointly Gaussian variables can be expressed in terms of the first order and second order moments

EXAMPLE

Consider 4 zero-mean and jointly normally distributed random variables $X_1, \ldots, X_4$. It holds that

$$\mathrm{E}[X_1X_2X_3X_4] = \mathrm{E}[X_1X_2]\mathrm{E}[X_3X_4] + \mathrm{E}[X_1X_3]\mathrm{E}[X_2X_4] + \mathrm{E}[X_1X_4]\mathrm{E}[X_2X_3] = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23}$$

Suppose that $\mathbf{W}$ is a symmetric matrix. Show that $\mathrm{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X}\mathbf{X}^{\mathrm{T}}] = \mathbf{\Sigma}_X \mathrm{tr}\{\mathbf{W}\mathbf{\Sigma}_X\} + 2\mathbf{\Sigma}_X\mathbf{W}\mathbf{\Sigma}_X$.

Linear transformations of vector random variables

Liniowe przekształcenia wektorowych zmiennych losowych

$$\mathbf{X}_{n\times 1}, \quad \mathbf{Y}_{m\times 1}$$

$$\mathbf{x} = \mathbf{X}(\xi), \quad \mathbf{y} = \mathbf{Y}(\xi)$$

Suppose that

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}, \quad \mathbf{A}_{m\times n}, \mathbf{b}_{m\times 1}$$

then

$$\begin{aligned} \mathrm{E}[\mathbf{Y}] &= \mathbf{A}\mathrm{E}[\mathbf{X}] + \mathbf{b} \\ &\Longrightarrow \mathbf{m}_Y = \mathbf{A}\mathbf{m}_X + \mathbf{b} \end{aligned}$$

and

$$\begin{aligned} \mathrm{cov}[\mathbf{Y}] &= \mathrm{E}[\mathbf{A}(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}] \\ &\Longrightarrow \mathbf{\Sigma}_Y = \mathbf{A}\mathbf{\Sigma}_X\mathbf{A}^{\mathrm{T}} \end{aligned}$$

note: if $\mathbf{X}$ is Gaussian then $\mathbf{Y}$ is also Gaussian

Project 1

Record and mix signals obtained from two independent speech sources (30 second long recordings, 2 linear mixtures with different mixing coefficients).

Perform blind source separation using the FastICA algorithm:

1. Center the data to make its mean zero.
2. Whiten the data to obtain uncorrelated mixture signals.
3. Initialize $\mathbf{w}_0 = [w_{1,0}, w_{2,0}]^{\mathrm{T}}$, $||\mathbf{w}_0|| = 1$ (e.g. randomly)
4. Perform an iteration of a one-source extraction algorithm

$$\widetilde{\mathbf{w}}_i = \mathrm{avg}\left\{\mathbf{z}(t)\left[\mathbf{w}_{i-1}^{\mathrm{T}}\mathbf{z}(t)\right]^3\right\} - 3\mathbf{w}_{i-1}$$

where $\mathbf{z}(t) = [z_1(t), z_2(t)]^{\mathrm{T}}$ denotes the vector of whitened mixture signals and $\mathrm{avg}(\cdot)$ denotes time averaging

$$\mathrm{avg}\{\mathrm{x(t)}\} = \frac{1}{N}\sum_{t=1}^{N} x(t)$$

5. Normalize $\widetilde{\mathbf{w}}_i$ by dividing it by its norm

$$\mathbf{w}_i = \frac{\widetilde{\mathbf{w}}_i}{||\widetilde{\mathbf{w}}_i||}$$

6. Determine a unit-norm vector $\mathbf{v}_i$ orthogonal to $\mathbf{w}_i$.

$$\mathbf{w}_i^{\mathrm{T}}\mathbf{v}_i = 0, \quad ||\mathbf{v}_i|| = 1$$

7. Display and listen to the current results of source separation

$$\left[\begin{array}{c} \widehat{x}_1(t) \\ \widehat{x}_2(t) \end{array} \right] = \left[\begin{array}{c} \mathbf{w}_i^{\mathrm{T}} \\ \mathbf{v}_i^{\mathrm{T}} \end{array} \right] \left[\begin{array}{c} z_1(t) \\ z_2(t) \end{array} \right], \quad t = 1, \ldots, N$$

8. If $\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_{i-1}$ is not close enough to 1, go back to step 4. Otherwise stop.

Literature: A. Hyvärinen, E. Oja. "A fast fixed-point algorithm for independent component analysis", *Neural Computation*, vol. 9, pp. 1483-1492, 1997.

To obtain the highest grade (5), a procedure that extracts $n > 2$ source signals from n mixtures must be implemented.

Facts about matrices

Fact 1

Given a square matrix $\mathbf{A}_{n\times n}$, an eigenvalue λ and its associated eigenvector $\mathbf{v}$ are a pair obeying the relation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

The eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are the roots of the characteristic polynomial of $\mathbf{A}$

$$\det(\mathbf{A} - \lambda\mathbf{I})$$

For a positive definite matrix $\mathbf{A}$ it holds that $\lambda_1, \ldots, \lambda_n > 0$, i.e., all eigenvalues are positive real.

Fact 2

Let $\mathbf{A}$ be a matrix with linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Then $\mathbf{A}$ can be factorized as follows

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$$

where $\mathbf{\Lambda}_{n\times n} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ and $\mathbf{V}_{n\times n} = [\mathbf{v}_1|\ldots|\mathbf{v}_n]$.

Facts about matrices

Fact 3

Let $\mathbf{A} = \mathbf{A}^{\mathrm{T}}$ be a real symmetric matrix. Such matrix has n linearly independent real eigenvectors. Moreover, these eigenvectors can be chosen such that they are orthogonal to each other and have norm one

$$\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_j = 0, \ \forall i \neq j$$
$$\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_i = ||\mathbf{w}_i||^2 = 1, \ \forall i$$

A real symmetrix matrix $\mathbf{A}$ can be decomposed as

$$\mathbf{A} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{\mathrm{T}}$$

where $\mathbf{W}$ is an orthogonal matrix: $\mathbf{W}_{n\times n} = [\mathbf{w}_1|\ldots|\mathbf{w}_n]$, $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}$.

The eigenvectors $\mathbf{w}_i$ can be obtained by normalizing the eigenvectors $\mathbf{v}_i$

$$\mathbf{w}_i = \frac{\mathbf{v}_i}{||\mathbf{v}_i||^2}, \quad i = 1, \ldots, n$$

Facts about matrices

Fact 4

Consider a zero-mean vector random variable $\mathbf{X}$ with covariance matrix

$$\mathbf{\Sigma}_X = \mathrm{E}[\mathbf{X}\mathbf{X}^{\mathrm{T}}] = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^{\mathrm{T}}$$

where $\mathbf{\Lambda}$ is a diagonal matrix made up of eigenvalues of $\mathbf{\Sigma}_X$ and $\mathbf{W}$ is an orthogonal matrix.

Let

$$\mathbf{Y} = \mathbf{W}\mathbf{\Lambda}^{-1/2}\mathbf{W}^{\mathrm{T}}\mathbf{X}$$

Then it holds that

$$\mathbf{\Sigma}_Y = \mathbf{I}$$

i.e., $\mathbf{Y}$ is the vector random variable made up of uncorrelated components.

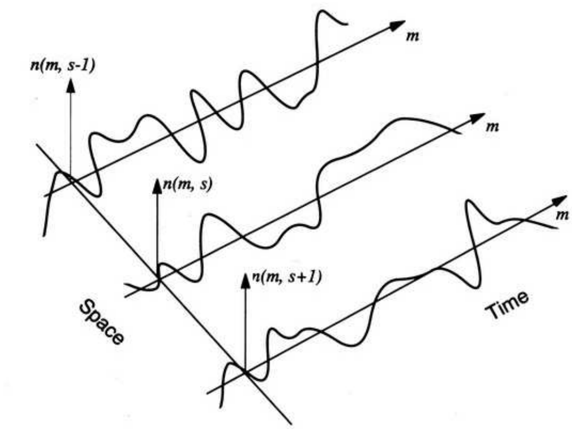
Discrete-time random (stochastic) processes
Procesy losowe (stochastyczne) z czasem dyskretnym

$$\{X(t,\xi),\ \xi \in \Xi,\ t \in \mathcal{T}\},\ \ \mathcal{T} = \{\ldots,-1,0,1,\ldots\}$$
$$X(t,\xi) = X_c(t_c = tT_s, \xi)$$

where t denotes the normalized (dimensionless) discrete time – the number of sampling intervals T_s – and $X_c(t_c, \xi)$ denotes a continuous-time random process

$$x(t) = X(t,\xi)$$

realization of a random process / realizacja procesu losowego



Two interpretations of a random process

- a sequence of random variables
- an ensemble (collection, family) of realizations, i.e., functions of time

Examples of random processes
Przykłady procesów losowych

Example 1: first-order autoregressive process

$$x(t) = ax(t-1) + n(t)$$

where $\{n(t)\}$ denotes a sequence of zero-mean uncorrelated random variables with variance σ_n^2.

Example 2: random sinusoidal signal

$$x(t) = A\sin(\omega t + \varphi)$$

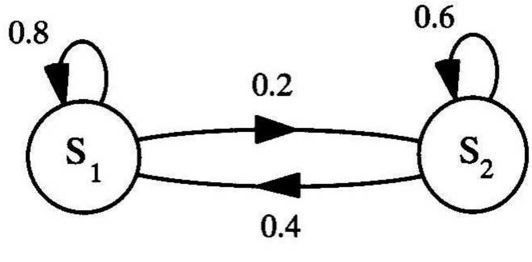
where φ, denoting initial phase, is a random variable with uniform distribution over $[0, 2\pi)$: $\varphi \sim \mathcal{U}(0, 2\pi)$.

note: this is an example of a strictly deterministic (singular) random process, perfectly predictable from its past

Example 3: pseudo-random binary signal (PRBS)

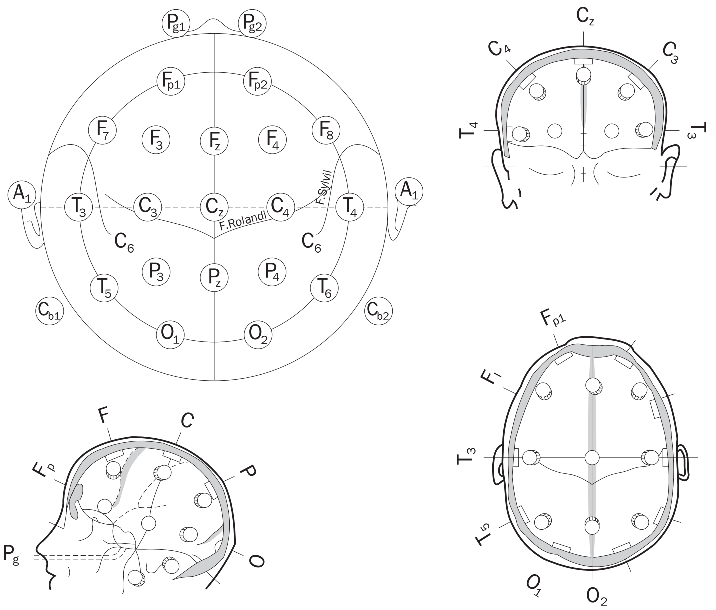
$$x(t) \in \{-A, A\}, \quad S_1 : -A,\ S_2 : A$$

transition probabilities:
$P_{11} = 0.8,\ P_{12} = 0.2,\ P_{22} = 0.6,\ P_{21} = 0.4$



Electroencephalographic (EEG) signals
Sygnały elektroencefalograficzne (EEG)

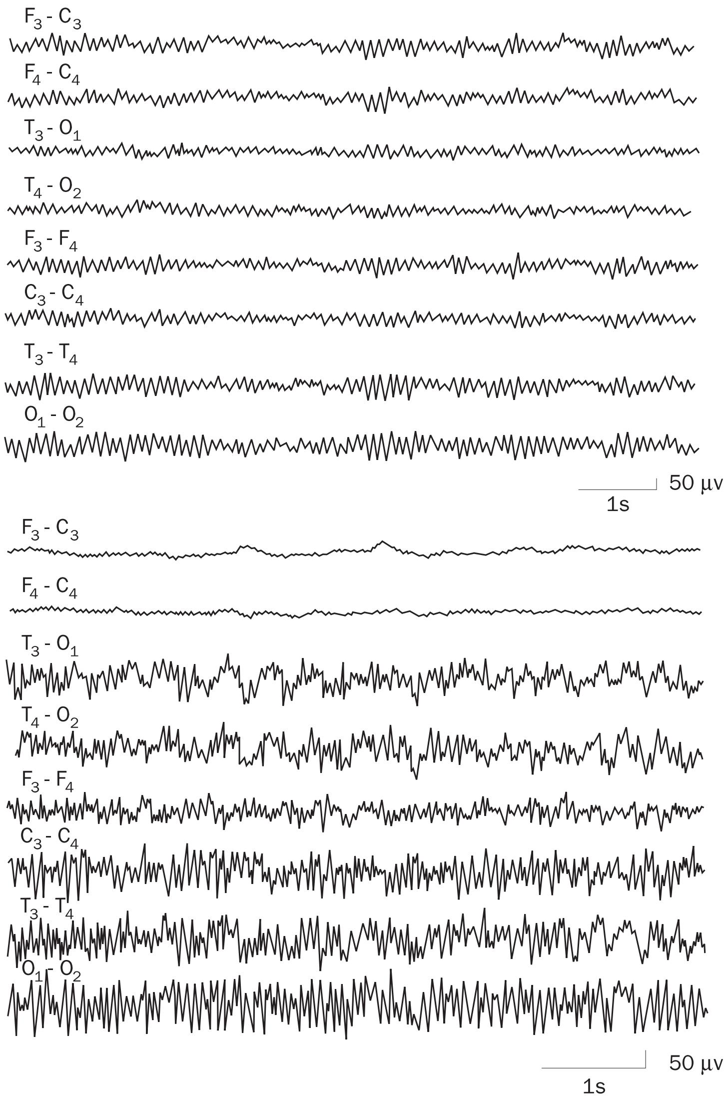
H. Berger, "Uber das Elektroenzephalogramm des Menschen", *Arch. Psychiatr. Nervenkrankheiten*, vol. 87, pp. 527–570, 1929.



Localization of electrodes on the patient's scull

EEG signals carry information about the state of the brain and its reaction to internal and external stimuli. Despite advances made recently in the area of computer-aided tomography (CAT), EEG remains a practically useful tool for studying the functional states of the brain (e.g. sleep-stage analysis) and for diagnosing functional brain disturbances (e.g. epilepsy).

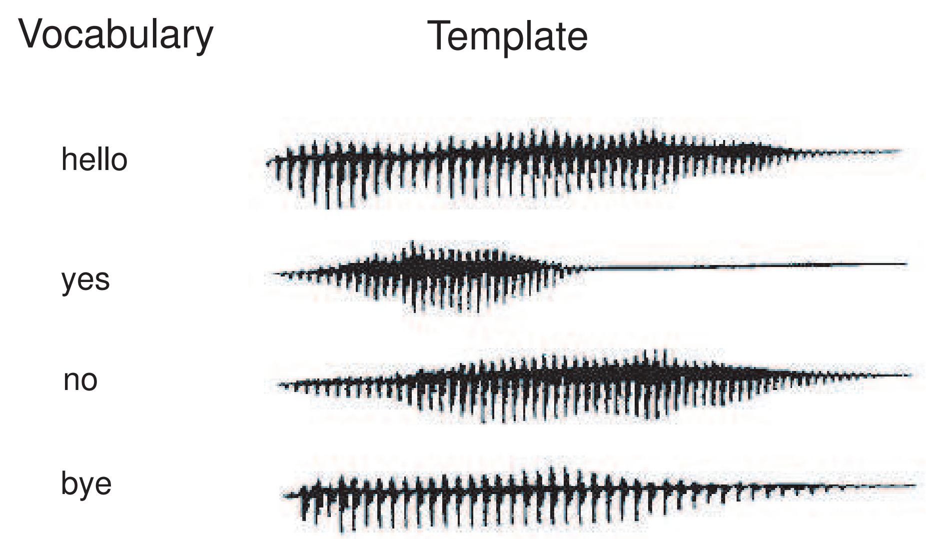
Electroencephalographic (EEG) signals



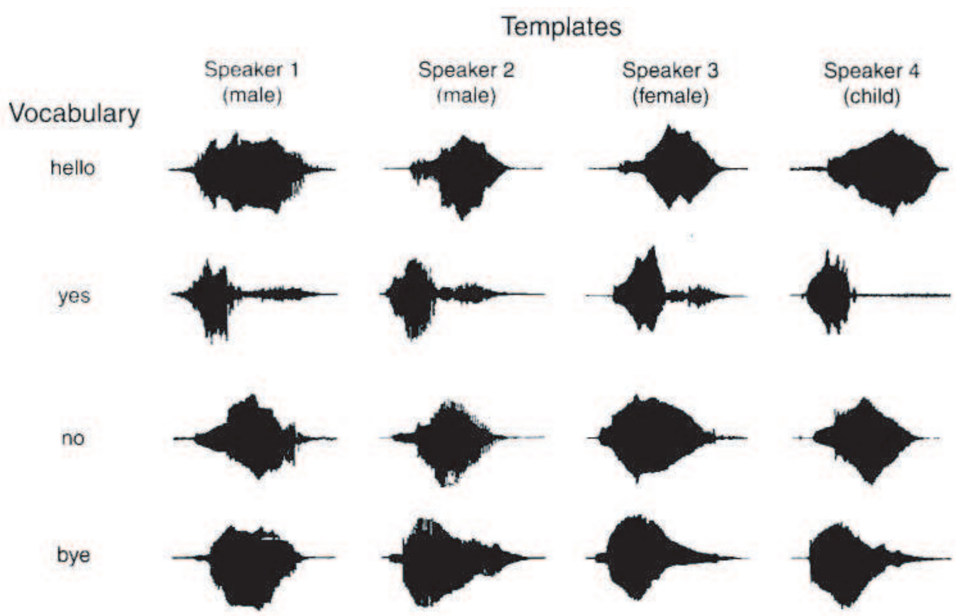
Example of normal (upper plots) and pathological (lower plots) EEG recordings

According to the traditional clinical nomenclature introduced in a pioneering work of Berger, EEG signals are described and classified in terms of several rhythmic activities called alpha (8–13 Hz), beta (14–30 Hz), delta (1–3 Hz) and theta (4–7 Hz). The normal EEG of an adult in a state called relaxed wakefulness is usually dominated by rhythmic alpha activity and a less pronounced beta activity. The contribution of delta and theta activities is arrhythmic and much smaller.

Speech signals
Sygnały mowy



A simple dictionary of speech templates for speech recognition



Variation in speech templates for different speakers

Description of a random processes
Opis procesów losowych

To fully describe a random process, one should specify all finite-dimensional ($n \in N$) joint distributions of random variables $X(t_1), \ldots, X(t_n)$ for all possible values of $t_1, \ldots, t_n \in \mathcal{T}$.

n-dimensional cumulative distribution function

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) \\ = P\left(X(t_1) < x_1, \ldots, X(t_n) < x_n\right)$$

n-dimensional probability density function

$$p(x_1, \ldots, x_n; t_1, \ldots, t_n) \\ = \frac{\partial^n F(x_1, \ldots, x_n; t_1, \ldots, t_n)}{\partial x_1 \cdots \partial x_n}$$

(assuming that all derivatives exist)

First-order and second-order characteristics
Charakterystyki pierwszego i drugiego rzędu

first-order characteristics: $X(t)$

first-order cumulative distribution and probability density functions (characterize the "amplitude" distribution)

$$F(x;t),\ \ p(x;t)$$

mean value / wartość oczekiwana

$$m_X(t) = \mathrm{E}[X(t)] = \int_{-\infty}^{\infty} x\, p(x;t) dx$$

second-order characteristics: $X(t_1), X(t_2)$

second-order cumulative distribution and probability density functions (characterize the joint "amplitude" distribution)

$$F(x_1, x_2; t_1, t_2),\ \ p(x_1, x_2; t_1, t_2)$$

autocorrelation function / funkcja autokorelacji

$$\begin{aligned} R_X(t_1, t_2) &= \mathrm{E}[X(t_1)X(t_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2; t_1, t_2) dx_1 dx_2 \end{aligned}$$

First-order and second-order characteristics

autocovariance function / funkcja autokowariancji

$$\begin{aligned} C_X(t_1, t_2) &= \mathrm{E}\{[X(t_1) - m_X(t_1)][X(t_2) - m_X(t_2)]\} \\ &= R_X(t_1, t_2) - m_X(t_1) m_X(t_2) \end{aligned}$$

variance / wariancja

$$\begin{aligned} \sigma_X^2(t) &= \mathrm{var}[X(t)] = \mathrm{E}\{[X(t) - m_X(t)]^2\} \\ &= \int_{-\infty}^{\infty} [x(t) - m_X(t)]^2 p(x; t) dx = C_X(t, t) \end{aligned}$$

normalized autocovariance function / unormowana funkcja autokowariancji

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sigma_X(t_1) \sigma_X(t_2)}$$

$$-1 \leq \rho_X(t_1, t_2) \leq 1$$

Stationary random processes
Stacjonarne procesy losowe

Qualitative (but not precise) description: a random process is said to be stationary (time stationary) if its density functions are invariant under a translation of time

strict sense stationarity / stacjonarność w ścisłym sensie

A discrete-time random process $X(t)$ is strict sense stationary if the joint distribution of random variables $X(t_1), \ldots, X(t_n)$ is identical with the joint distribution of random variables $X(t_1 + \Delta t), \ldots, X(t_n + \Delta t)$ for *any* values of $n \in N$, $\Delta t \in \mathcal{T}$ and $t_1, \ldots, t_n \in \mathcal{T}$:

$$\begin{aligned} & p(x_1, \ldots, x_n; t_1, \ldots, t_n) \\ & \qquad = p(x_1, \ldots, x_n; t_1 + \Delta t, \ldots, t_n + \Delta t) \end{aligned}$$

consequences of invariance to a time shift

$$\begin{aligned} & p(x; t) = p(x; t + \Delta t) \\ & \qquad \Longrightarrow p(x; t) = p(x) \end{aligned}$$

$$\begin{aligned} & p(x_1, x_2; t_1, t_2) = p(x_1, x_2; t_1 + \Delta t, t_2 + \Delta t) \\ & \qquad \Longrightarrow p(x_1, x_2; t_1, t_2) = p(x_1, x_2; \tau) \end{aligned}$$

where $\tau = t_2 - t_1$.

Stationary random processes

using the first property one obtains

$$m_X(t) = m_X$$

using the second property one obtains

$$\begin{aligned} R_X(t_1, t_2) &= R_X(\tau) \\ C_X(t_1, t_2) &= C_X(\tau) \\ \sigma_X^2(t) &= C_X(0) = \sigma_X^2 \end{aligned}$$

wide sense stationarity / stacjonarność w szerokim sensie

A discrete-time random process $X(t)$ is wide sense stationary if its mean function is constant and its autocorrelation function is a function only of τ

$$m_X(t) = m_X, \quad R_X(t_1, t_2) = R_X(\tau)$$

note: all strict sense stationary random processes are also wide sense stationary, provided that their mean and autocorrelation functions exist. The converse is not true.

Ergodic processes
Procesy ergodyczne

Qualitative (but not precise) description: a random process is said to be ergodic if its statistical properties (such as its mean and autocorrelation function) can be deduced from a single, sufficiently long sample (realization) of the process. One can discuss the ergodicity of various properties of a random process.

mean-square ergodicity in the first-order moment / ergodyczność, w sensie średniokwadratowym, względem wartości oczkiwanej

$$m_X = \mathrm{E}[X(t)] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X(t,\xi_i)$$

ensemble average

$$\widehat{m}_X(T,\xi) = \frac{1}{T} \sum_{t=1}^{T} X(t,\xi)$$

time average

What are the conditions under which

$$\lim_{T\to\infty} \mathrm{E}\left\{[\widehat{m}_X(T,\xi) - m_X]^2\right\} = 0 \quad ?$$

Ergodic processes

THEOREM

The necessary and sufficient condition of a mean-square ergodicity in the first moment of a wide sense stationary process $X(t)$ has the form

$$\lim_{T\to\infty} \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} C_X(t_2 - t_1) = 0$$

PROOF

$$\mathrm{E}\left\{[\widehat{m}_X(T) - m_X]^2\right\} = \mathrm{E}\left\{\left[\frac{1}{T}\sum_{t=1}^{T} X(t) - m_X\right]^2\right\}$$

$$= \mathrm{E}\left\{\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}[X(t_1) - m_X][X(t_2) - m_X]\right\}$$

$$= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{E}\left\{[X(t_1) - m_X][X(t_2) - m_X]\right\}$$

$$= \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} C_X(t_1, t_2) = \frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} C_X(t_2 - t_1)$$

sufficient condition

$$\lim_{\tau\to\infty} |C_X(\tau)| = 0$$

Ergodic processes

mean-square ergodicity in the second-order moments / ergodyczność, w sensie średniokwadratowym, względem momentów drugiego rzędu

$$R_X(\tau) = \mathrm{E}[X(t)X(t+\tau)]$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X(t;\xi_i)X(t+\tau;\xi_i)$$

$$\widehat{R}_X(\tau, T, \xi) = \frac{1}{T} \sum_{t=1}^{T} X(t,\xi)X(t+\tau,\xi)$$

What are the conditions under which

$$\lim_{T\to\infty} \mathrm{E}\left\{[\widehat{R}_X(\tau,T,\xi) - R_X(\tau)]^2\right\} = 0 \quad ?$$

for Gaussian processes the sufficient condition is

$$\lim_{\tau\to\infty} |C_X(\tau)| = 0$$

note: a process which is ergodic in the first and second moments is sometimes called ergodic in the wide sense

example of a process that is *not* ergodic:
$X(t,\xi) = A(\xi), \quad A \sim \mathcal{U}[0,1]$

Properties of the autocorrelation function
Własności funkcji autokorelacji

The autocorrelation function of a wide sense stationary random process $X(t)$

- is an even function, that is, $R_X(\tau) = R_X(-\tau)$, $\forall\tau$
- is maximum at the origin, that is, $|R_X(\tau)| \leq R_X(0), \forall\tau$
- is a nonnegative definite function, that is, for any $n \in N$, any selection of $t_1, \ldots, t_n \in \mathcal{T}$, and any selection of $z_1, \ldots, z_n \in R$, it holds that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_j R_X(t_i - t_j) \geq 0$$

which is equivalent to

$$\mathcal{F}[R_X(\tau)] \geq 0$$

where $\mathcal{F}[\cdot]$ denotes discrete Fourier transform (DFT)

note: the nonnegative definiteness condition
is equivalent to $\mathbf{R} = [R_X(t_i - t_j)]_{n\times n} \geq 0$

Introduction to spectral analysis
Wprowadzenie do analizy widmowej

Consider a continuous-time band-limited periodic signal

$$x_c(t_c) = \sin(2\pi f_c t_c) = \sin \omega_c t_c, \quad f_c \leq F_0$$

where $t_c \in R$ denotes continuous time measured in seconds [sec] and

f_c	-	frequency of a continuous-time signal expressed in cycles per second [Hz]
ω_c	-	angular frequency of a continuous-time signal expressed in radians per second [rad/sec]

Suppose that the signal $x_c(t_c)$ is sampled with the frequency $F_s = 1/T_s \geq 2F_0$

$$x(t) = x_c(t_c = tT_s) = \sin(2\pi f t) = \sin(\omega t)$$

where $t \in C$ denotes normalized discrete time measured in samples [sa] and

f	-	normalized frequency / unormowana częstotliwość expressed in cycles per sample [c/sa]
ω	-	normalized angular frequency / unormowana pulsacja expressed in radians per sample [rad/sa]

$$f = \frac{f_c}{F_s} \in \left[-\frac{1}{2}, \frac{1}{2}\right], \quad \omega = \frac{\omega_c}{F_s} \in [-\pi, \pi]$$

Introduction to spectral analysis

energy spectrum of a deterministic signal / widmo energii sygnału deterministycznego

Let $\{x(t)\}$ be a finite energy deterministic signal

$$E_x = \sum_{t=-\infty}^{\infty} x^2(t) < \infty$$

Then $\{x(t)\}$ posesses a discrete-time Fourier transform (DFT)

$$X(\omega) = \sum_{t=-\infty}^{\infty} x(t)e^{-j\omega t}$$

$$\phi_X(\omega) = |X(\omega)|^2$$

energy spectral density / gęstość widmowa energii
describes distribution of the signal energy over different frequency bands

Such interpretation stems from the fact that, according to the Parseval's theorem

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_X(\omega)d\omega = \sum_{t=-\infty}^{\infty} x^2(t) = E_x$$

Introduction to spectral analysis

naïve extension to random processes /
naiwne rozszerzenie na procesy losowe

Consider a zero-mean wide sense stationary and ergodic random process $X(t, \xi)$. It holds that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} X^2(t, \xi) = \sigma_X^2$$

power (mean energy) / moc (średnia energia)

Can we define power spectral density in the form analogous to the energy spectral density

$$S_X(\omega) = \lim_{N \to \infty} P_X(\omega, \xi; N) \quad ?$$

where

$$P_X(\omega, \xi; N) = \frac{1}{N} \left| \sum_{t=1}^{N} X(t, \xi) e^{-j\omega t} \right|^2$$

periodogram / periodogram

No, in the general case the limit above does not exist and/or it is realization-dependent, i.e., it is a random variable !

Power spectral density
Widmowa gęstość mocy

power spectral density shows distribution of the signal power over different frequency bands

Definition 1

$$S_X(\omega) = \lim_{N\to\infty} \mathrm{E}[P_X(\omega, \xi; N)]$$

mean periodogram

Definition 2 (Wiener - Khintchine)

$$S_X(\omega) = \mathcal{F}[R_X(\tau)] = \sum_{\tau=-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau}$$

Fourier transform of the autocorreletion function

THEOREM

Both definitions given above are equivalent provided that the autocorrelation function of $\{X(t)\}$ decays sufficiently rapidly with growing lag, namely

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-N}^{N} |\tau| R_X(\tau) = 0$$

Power spectral density

PROOF

$$
\begin{aligned}
&\lim_{N\to\infty} \mathrm{E}[P_X(\omega,\xi;N)] \\
&\quad = \lim_{N\to\infty} \mathrm{E}\left[\frac{1}{N}\left|\sum_{t=1}^{N} X(t)e^{-j\omega t}\right|^2\right] \\
&\quad = \lim_{N\to\infty} \frac{1}{N}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N} \mathrm{E}[X(t_1)X(t_2)]e^{-j\omega(t_1-t_2)} \\
&\quad = \lim_{N\to\infty} \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1} (N-|\tau|)R_X(\tau)e^{-j\omega\tau} \\
&\quad = \sum_{\tau=-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau} - \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=-(N-1)}^{N-1} |\tau|R_X(\tau) \\
&\quad = \sum_{\tau=-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau} = S_X(\omega)
\end{aligned}
$$

note: power spectral density function carries the same information about a random process as its autocorrelation function because

$$R_X(\tau) = \mathcal{F}^{-1}[S_X(\omega)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_X(\omega)e^{j\omega\tau}d\omega$$

Properties of power spectral density function
Własności funkcji widmowej gęstości mocy

$$\sigma_X^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_X(\omega)d\omega$$

The power spectral density function of a wide sense stationary random process $X(t)$

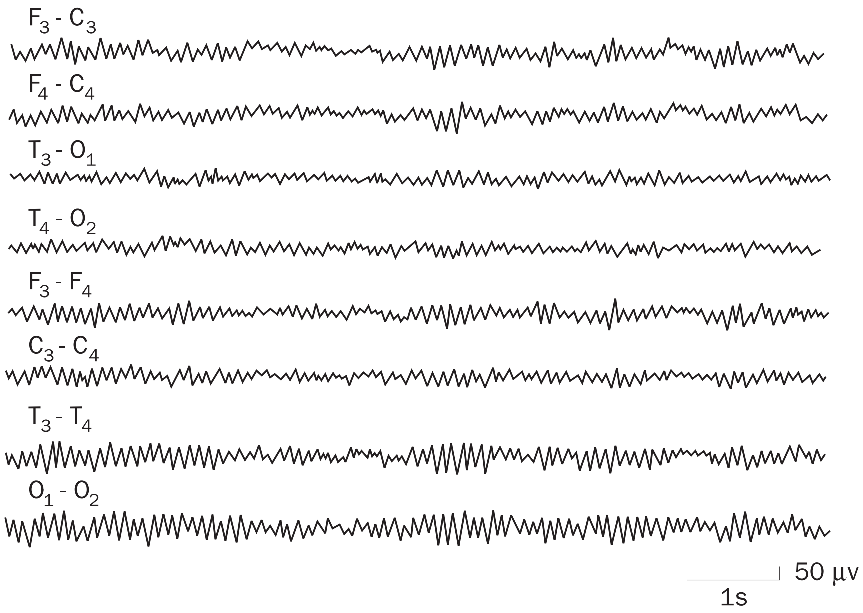
- is an even function, that is, $S_X(\omega) = S_X(-\omega)$, $\forall\omega$
- is a nonnegative function, that is, $S_X(\omega) \geq 0$, $\forall\omega$
- is a periodic function, namely $S_X(\omega + 2k\pi) = S_X(\omega)$, $\forall k \in C$

interpretation of different frequency bands:

$\omega \in [0, \pi]$	-	region of physical significance
$\omega \in [-\pi, 0)$	-	symmetric fold
$\lvert\omega\rvert > \pi$	-	periodic extension

DFT-free assessment of spectral density function

woman, age 19, healthy



Doctor's evaluation: persistent, regular alpha activity with a dominant 9 Hz component

DFT-free approaches:

- Count cycles per second; take into account the relative amplitudes of different rhythmic components [qualitative assessment].

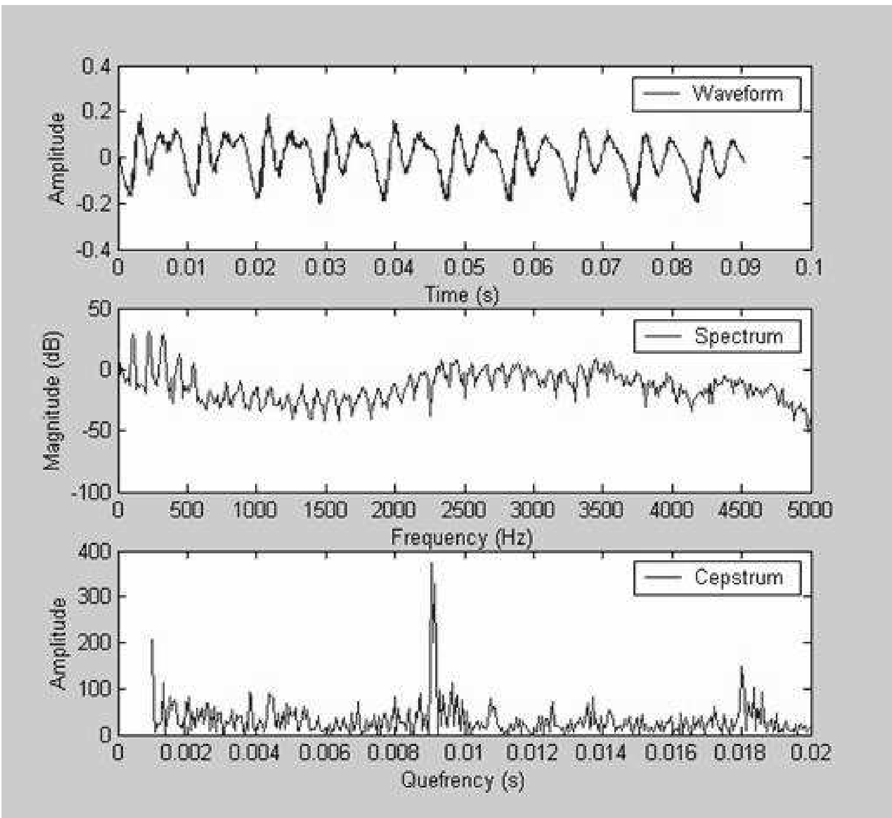
- Pass the analyzed signal through a bank of narrowband bandpass filters with different center frequencies $\omega_i \in [0, \pi], i = 1, \ldots, k$. Evaluate $S_X(\omega_i)$ as the mean power of the signal observed at the output of the i-th filter [quantitative assessment].

Power cepstrum of a signal
Cepstrum mocy sygnału

A cepstrum is the result of taking the Fourier transform of the log signal spectrum. Effectively we are treating the signal spectrum as another signal and looking for periodicity in the spectrum itself. The term *cepstrum* was derived by reversing the first four letters of "spectrum". The cepstrum is so-called because it turns the spectrum inside-out. The x-axis of the cepstrum has units of *quefrency*, and peaks in the cepstrum (which relate to periodicities in the spectrum) are called *rahmonics*.

$$K_X(q) = \mathcal{F}^{-1}\left[\log|X(\omega)|\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(\omega)| e^{j\omega q} d\omega$$

real cepstrum / cepstrum rzeczywiste



Relationships between random processes
Związki pomiędzy procesami losowymi

$$X(t,\xi) \longleftrightarrow Y(t,\xi)$$

crosscorelation function / funkcja korelacji wzajemnej

$$R_{XY}(t_1,t_2) = \mathrm{E}\{X(t_1)Y(t_2)\}$$

cross-covariance function / funkcja kowariancji wzajemnej

$$\begin{aligned} C_{XY}(t_1,t_2) &= \mathrm{E}\{[X(t_1) - m_X(t_1)][Y(t_2) - m_Y(t_2)]\} \\ &= R_{XY}(t_1,t_2) - m_X(t_1)m_Y(t_2) \end{aligned}$$

normalized cross-covariance function / unormowana funkcja kowariancji wzajemnej

$$\rho(t_1,t_2) = \frac{C_{XY}(t_1,t_2)}{\sigma_X(t_1)\sigma_Y(t_2)} \in [-1,1]$$

uncorrelated processes / procesy nieskorelowane

$$C_{XY}(t_1,t_2) = 0, \quad \forall t_1, t_2$$

orthogonal processes / procesy ortogonalne

$$R_{XY}(t_1,t_2) = 0, \quad \forall t_1, t_2$$

Jointly wide sense stationary random processes
Procesy losowe stacjonarne łącznie w szerokim sensie

Processes $X(t)$ and $Y(t)$ are called jointly wide sense stationary if each of them is wide sense stationary and

$$R_{XY}(t_1, t_2) = R_{XY}(\tau)$$

where $\tau = t_2 - t_1$.

For jointly wide sense stationary processes $X(t)$ and $Y(t)$ one can define

cross spectral density function / wzajemna widmowa gęstość mocy

$$S_{XY}(\omega) = \mathcal{F}[R_{XY}(\tau)] = \sum_{\tau=-\infty}^{\infty} e^{-j\omega\tau} R_{XY}(\tau)$$

note: since in general $R_{XY}(\tau) \neq R_{YX}(\tau)$, the cross spectral density function $S_{XY}(\omega)$ is *complex-valued*

Independent random processes
Procesy losowe niezależne

Two discrete-time random processes $X(t)$ and $Y(t)$ are statistically independent if the group of random variables $X(t_1), \ldots, X(t_n)$ is statistically independent of random variables $Y(s_1), \ldots, Y(s_m)$ for *any* values of $n, m \in N$, $t_1, \ldots, t_n \in \mathcal{T}$ and $s_1, \ldots, s_m \in \mathcal{T}$:

$$
\begin{aligned}
p(x_1, \ldots, x_n, y_1, \ldots, y_m; t_1, \ldots, t_n, s_1, \ldots, s_m)& \\
= p(x_1, \ldots, x_n; t_1, \ldots, t_n)& \\
\times\, p(y_1, \ldots, y_m; s_1, \ldots, s_m)&
\end{aligned}
$$

note: two jointly Gaussian and mutually uncorrelated random processes are independent

Relationships for the sum of two random processes
Zależności dla sumy dwóch procesów losowych

Suppose that $X(t)$ and $Y(t)$ are jointly wide sense stationary random processes and

$$Z(t) = X(t) + Y(t)$$

Then

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau) + R_{XY}(\tau) + R_{YX}(\tau)$$
$$S_Z(\tau) = S_X(\tau) + S_Y(\tau) + S_{XY}(\tau) + S_{YX}(\tau)$$

When the processes $X(t)$ and $Y(t)$ are also orthogonal, it holds that

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau)$$
$$S_Z(\tau) = S_X(\tau) + S_Y(\tau)$$

Linear transformations of random processes

Liniowe przekształcenia procesów losowych

$$Y(t,\xi) = L[X(t,\xi)]$$

$$m_X(t), R_X(t_1,t_2) \longrightarrow m_Y(t), R_Y(t_1,t_2), R_{XY}(t_1,t_2)$$

condition of linearity:

$$L[k_1 x_1(t) + k_2 x_2(t)] = k_1 L[x_1(t)] + k_2 L[x_2(t)]$$
$$\forall\ t, k_1, k_2,\ \{x_1(\cdot)\}, \{x_2(\cdot)\}$$

$$y(t) = \sum_{i=0}^{t} k(i) x(t-i)$$

$\{k(i), i = 1, 2, \ldots\}$ – impulse response
of a dynamic system

$$m_Y(t) = \mathrm{E}[Y(t,\xi)] = \mathrm{E}\left[\sum_{i=0}^{t} k(i) X(t-i,\xi)\right]$$
$$= \sum_{i=0}^{t} k(i) \mathrm{E}[X(t-i,\xi)] = \sum_{i=0}^{t} k(i) m_X(t-i)$$

Linear transformations of random processes

In an analogous way one can show that:

$$R_{XY}(t_1,t_2) = \mathrm{E}[X(t_1,\xi)Y(t_2,\xi)]$$
$$= \sum_{i=0}^{t_2} k(i) R_X(t_1, t_2 - i)$$
$$R_Y(t_1,t_2) = \mathrm{E}[Y(t_1,\xi)Y(t_2,\xi)]$$
$$= \sum_{i=0}^{t_1} \sum_{j=0}^{t_2} k(i)k(j) R_X(t_1 - i, t_2 - j)$$

steady state relationships for a wide sense stationary process $X(t)$: $m_X(t) = m_X, R_X(t_1,t_2) = R_X(\tau)$

$$y(t) = \sum_{i=0}^{\infty} k(i) x(t-i)$$

$$m_Y = m_X \sum_{i=0}^{\infty} k(i)$$

$$R_{XY}(\tau) = \sum_{i=0}^{\infty} k(i) R_X(\tau - i)$$

$$R_Y(\tau) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} k(i)k(j) R_X(\tau + i - j)$$

note: the output process $Y(t)$ is also wide sense stationary

Linear time-invariant systems
Układy liniowe niezmiennicze w czasie

conditions of preservation of stationarity
at the output of a linear system

THEOREM

Suppose that a linear system $L[\cdot]$ is excited by a strictly (or wide sense) stationary random process $X(t)$. When the system is time-invariant, i.e., the relationship $y(t) = L[x(t)]$ entails

$$y(t + \Delta t) = L[x(t + \Delta t)], \quad \forall t, \Delta t, \{x(\cdot)\}$$

then the output random signal $Y(t)$ is also strictly (or wide sense) stationary.

The system governed by $y(t) = L[x(t)] = \sum_{i=0}^{t} k(i)x(t-i)$ is *not* time-invariant because

$$y(t + \Delta t) = \sum_{i=0}^{t+\Delta t} k(i)x(t + \Delta t - i)$$

$$\neq L[x(t + \Delta t)] = \sum_{i=0}^{t} k(i)x(t + \Delta t - i)$$

Show that the system governed by $y(t) = L[x(t)] = \sum_{i=0}^{\infty} k(i)x(t-i)$ is time-invariant.

Relationships between spectral density functions

Związki pomiędzy funkcjami widmowej gęstości mocy

Suppose that $X(t)$ is a wide sense stationary random process with a spectral density function $S_X(\omega)$ and $Y(t)$ is the process observed at the output of a linear time-invariant system

$$y(t) = \sum_{i=0}^{\infty} k(i)x(t-i)$$

Derive expressions for $S_{XY}(\omega)$ and $S_Y(\omega)$.

$$S_X(\omega) = \sum_{\tau=-\infty}^{\infty} R_X(\tau)e^{-j\omega\tau}$$

$$\begin{aligned}
S_{XY}(\omega) &= \sum_{\tau=-\infty}^{\infty} R_{XY}(\tau)e^{-j\omega\tau} \\
&= \sum_{\tau=-\infty}^{\infty} e^{-j\omega\tau} \sum_{s=0}^{\infty} k(s)R_X(\tau-s) \\
&= \sum_{s=0}^{\infty} e^{-j\omega s}k(s) \left[\sum_{\tau=-\infty}^{\infty} e^{-j\omega(\tau-s)}R_X(\tau-s) \right] \\
&= \sum_{s=0}^{\infty} e^{-j\omega s}k(s)S_X(\omega) = K(e^{-j\omega})S_X(\omega)
\end{aligned}$$

Relationships between spectral density functions

where

$$K(e^{-j\omega}) = \sum_{s=0}^{\infty} e^{-j\omega s} k(s)$$

denotes transfer function of an excited system.

$$\begin{aligned}
S_Y(\omega) &= \sum_{\tau=-\infty}^{\infty} R_Y(\tau) e^{-j\omega\tau} \\
&= \sum_{\tau=-\infty}^{\infty} e^{-j\omega\tau} \sum_{s_1=0}^{\infty} \sum_{s_2=0}^{\infty} k(s_1) k(s_2) R_X(\tau - s_2 + s_1) \\
&= \sum_{s_1=0}^{\infty} e^{j\omega s_1} k(s_1) \sum_{s_2=0}^{\infty} e^{-j\omega s_2} k(s_2) \\
&\qquad \times \left[\sum_{\tau=-\infty}^{\infty} e^{-j\omega(\tau - s_2 + s_1)} R_X(\tau - s_2 + s_1) \right] \\
&= K(e^{j\omega}) K(e^{-j\omega}) S_X(\omega) = \left| K(e^{-j\omega}) \right|^2 S_X(\omega)
\end{aligned}$$

summary:

$$S_{XY}(\omega) = K(e^{-j\omega}) S_X(\omega)$$

$$S_Y(\omega) = \left| K(e^{-j\omega}) \right|^2 S_X(\omega)$$

Spectral subtraction
Metoda odejmowania widmowego

Goal:

Recover signal $x(t)$ based on its noisy measurements $y(t)$

$$y(t) = x(t) + z(t)$$

Assumptions:

1. $X(t)$ (recovered signal) and $Z(t)$ (noise) are *locally* wide-sense stationary random processes.

2. Processes $X(t)$ and $Z(t)$ are orthogonal.

3. Spectral density function of noise $S_Z(\omega)$ is known or can be estimated based on those fragments of $Y(t)$ that contain noise only.

S.F. Böll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 27, pp. 113-120, 1979.

Spectral subtraction

note: our perception of sounds depends primarily on their spectral content – two sounds with the same spectral density functions are perceived as (almost) identical

Under the assumptions made it holds that (locally)

$$S_Y(\omega) = S_X(\omega) + S_Z(\omega)$$

Can one design a linear filter $K(z^{-1})$ that would transform a process $Y(t)$ with spectral density function $S_Y(\omega)$ into the process $X(t)$ with spectral density function $S_X(\omega)$?

$$S_X(\omega) = \left|K(e^{-j\omega})\right|^2 S_Y(\omega)$$

Therefore the amplitude characteristic of the transforming filter can be obtained from

$$A(\omega) = \left|K(e^{-j\omega})\right| = \sqrt{\frac{S_X(\omega)}{S_Y(\omega)}}$$

$$= \sqrt{\frac{S_Y(\omega) - S_Z(\omega)}{S_Y(\omega)}}$$

power spectral subtraction

Discrete Fourier transform (DFT)
Dyskretna transformata Fouriera

$$X(\omega_i) = \sum_{t=0}^{N-1} x(t) e^{-j\omega_i t}$$

$$\omega_i = \frac{2\pi i}{N}\,, \quad i = 0, \ldots, N-1$$

$$\{X(\omega_0), \ldots, X(\omega_{N-1})\}$$
$$= \mathrm{DFT}\{x(0), \ldots, x(N-1)\}$$

Inverse Discrete Fourier Transform (IDFT)

$$x(t) = \frac{1}{N} \sum_{i=0}^{N-1} X(\omega_i) e^{j\omega_i t}$$
$$t = 0, \ldots, N-1$$

$$\{x(0), \ldots, x(N-1)\}$$
$$= \mathrm{IDFT}\{\mathrm{X}(\omega_0), \ldots, \mathrm{X}(\omega_{\mathrm{N}-1})\}$$

(Inverse) Fast Fourier Transform

$$N = 2^k \longrightarrow \mathrm{FFT},\ \mathrm{IFFT}$$

Power spectral subtraction

The signal is divided into frames (segments) of Identical widths N (N should be short enough to guarantee local signal stationarity, e.g., in case of speech signals it should cover no more than 10-20 ms of speech). Each frame is "denoised" separately and the results are combined for the entire recording. Consider any frame (to simplify notation the number of the frame was suppressed)

$$\{y(0), \dots, y(N-1)\}$$

<u>Step 1</u>

Use a fragment of the signal that contains noise only $[y(t) = z(t)]$ to estimate the power spectral density of the noise

$$\widehat{S}_Z(\omega_i) = \frac{1}{N} \, |Z(\omega_i)|^2, \quad i = 0, \dots, \left[\frac{N}{2}\right]$$

If many noise frames are available, apply averaging.

<u>Step 2</u>

Estimate the power spectral density function of the noisy signal

$$\widehat{S}_Y(\omega_i) = \frac{1}{N} \, |Y(\omega_i)|^2, \quad i = 0, \dots, \left[\frac{N}{2}\right]$$

Step 3

Estimate the power spectral density function of the noiseless signal

$$\widehat{S}_X(\omega_i) = \begin{cases} \widehat{S}_Y(\omega_i) - \widehat{S}_Z(\omega_i) & \text{when nonnegative} \\ 0 & \text{otherwise} \end{cases}$$

$$i = 0, \ldots, \left[\frac{N}{2}\right]$$

Step 4

Design denoising filter

$$\widehat{A}(\omega_i) = \sqrt{\frac{\widehat{S}_X(\omega_i)}{\widehat{S}_Y(\omega_i)}}, \quad i = 0, \ldots, \left[\frac{N}{2}\right]$$

$$\widehat{A}(\omega_i) = \widehat{A}(\omega_{N-1-i}), \quad i = \left[\frac{N}{2}\right] + 1, \ldots, N-1$$

Step 5

Evaluate denoised signal

$$\widehat{X}(\omega_i) = \widehat{A}(\omega_i) Y(\omega_i), \quad i = 0, \ldots, N-1$$

$$\begin{aligned} &\{\widehat{x}(0), \ldots, \widehat{x}(N-1)\} \\ &\qquad = \mathrm{IDFT}\{\widehat{\mathrm{X}}(\omega_0), \ldots, \widehat{\mathrm{X}}(\omega_{\mathrm{N}-1})\} \end{aligned}$$

Read next frame and return to Step 2.

Remarks and extensions

1. When using spectral subtraction one modifies the amplitude distribution of noisy speech in different frequency bands, but preserves its phase distribution.

2. When intensity of wideband noise corrupting speech is large, the results of denoising may suffer from the so-called *musical noise* (tin-like sound consisting of many short-lived narrowband components appearing at random frequencies). Musical noise can be reduced by eliminating tones (i.e., periodogram spectral lines) that appear only in isolated segments.

3. To eliminate signal discontinuities that may occur at the beginning/end of each denoised speech segment, one may use the *overlap-add technique*, e.g. for segments of length $N = 2n + 1$ and 50% overlap one can use the following data windows

$$w(t) = 1 - \frac{|t-n|}{n}\ , \quad t = 0, \ldots, 2n$$

or

$$w(t) = \frac{1}{2}\left[1 - \cos\frac{\pi t}{n}\right],\ t = 0, \ldots, 2n$$

$$w(t) + w(t-n) = 1, \quad t = n, \ldots, 2n$$

Overlap-add synthesis

$$\widehat{x}_w(t) = \begin{cases} \widehat{x}(t)w(t) & t = 0, \ldots, 2n \\ 0 & \text{elsewhere} \end{cases}$$

$$\widehat{x}_w^k(t) = \widehat{x}_w(t), \quad t = kn, \ldots, (k+2)n$$

$$\widehat{x}(t) = \ldots + \widehat{x}_w^{k-1}(t) + \widehat{x}_w^k(t) + \widehat{x}_w^{k+1}(t) + \ldots$$

4. Generalized spectral subtraction

$$\widehat{X}(\omega_i) = \left\{ \frac{\left[|Y(\omega_i)|^\beta - \alpha \overline{|Z(\omega_i)|^\beta} \right]_+}{|Y(\omega_i)|^\beta} \right\}^{1/\beta} Y(\omega_i)$$

where $\beta \geq 1$, $\alpha > 0$ is the user-dependent coefficient, $\overline{(\cdot)}$ denotes time averaging, and $[\cdot]_+$ denotes "round to nonnegative" operation.

$$\beta = 1 \longrightarrow \text{magnitude spectral substraction}$$

$$\beta = 2 \longrightarrow \text{power spectral substraction}$$

<u>NOTE</u>: in practice one often takes $\alpha > 1$.

Project 2

Create a noisy audio recording (add artificially generated noise to a clean music or speech signal). The first second of the recording should contain noise only. Then denoise recording using the method of spectral subtraction.

THE END