Advanced Image Processing

Part III:
Random Signals

S. Voloshynovskiy

Course Outline

- Recall of Linear Algebra.
- Introduction. Human Visual System.
- Image Representation: pyramids and wavelets.
- Random Signals.
- Image Modeling.
- Image Sensor Models. Noise Models.
- Image Denoising.
- Image Restoration.
- Image Compression.
- Video Modeling and Compression.
- Digital Data Hiding.

Recommended books

- H. Stark and J. W. Woods, Probability, Random Processes, and Estimation Theory for Engineers, Prentice-Hall, 1994.

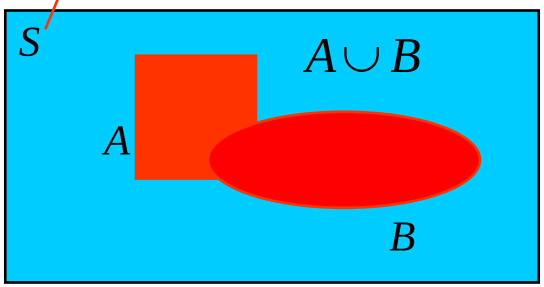
- A. Papoulis. Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, third edition, 1991.

Roadmap:

1. Probability
2. Random Variables
3. Random Processes

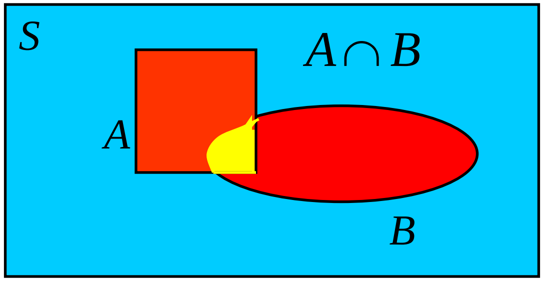
1. Probability: Notations and Venn Diagrams

Universal set



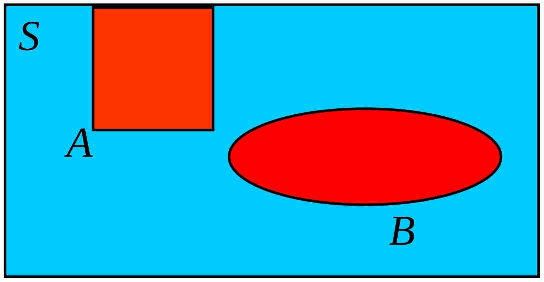
The union of sets A and B is the set of all elements that are either in A or in B, or in both.

Or



The intersection of two sets A and B is the set of all elements which are contained both in A and B.

And



A collection of sets A and B is mutually exclusive iff

$$A \cap B = \emptyset$$

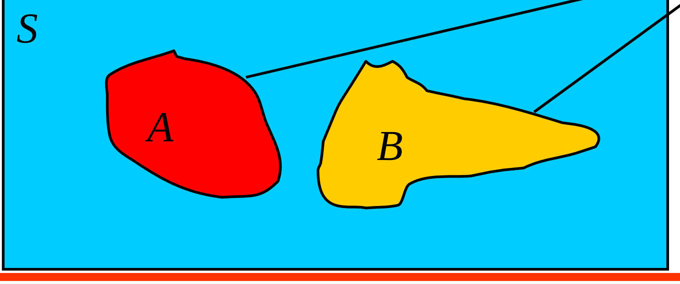
1. Probability

The probability *P(A)* is a number, which measures the likelihood of random event A.

Axioms of probability:

- $P(A) \geq 0$
- $P(A) \leq 1$ and $P(A) = 1$ only if $A = S$ (the certain event)
- If A and B are two events such that $A \cap B = \varnothing$,
  then $P(A \cup B) = P(A) + P(B)$

A and B are ***mutually exclusive*** $i.e.$ A and B don't overlap



1. Joint and Conditional Probability

- ***Joint probability*** is the probability that both A and B occur:

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

- ***Conditional probability*** is the probability that A will occur given that B has occurred:

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

- ***Bayes' theorem***:

$$P(A,B) = P(B)P(A|B) = P(A)P(B|A)$$

$$P(A) = \frac{P(A|B)P(B)}{P(B|A)} \quad \text{and} \quad P(B) = \frac{P(B|A)P(A)}{P(A|B)}$$

1. Statistical Independence

- Events A and B are ***statistically independent***

  if $\boxed{P(A,B) = P(A)P(B).}$

- If A and B are ***independent***, then:

  $P(A|B) = P(A)$ and $P(B|A) = P(B)$

- Example:
  - Flip a coin, call result A={heads}.
  - Flip it again, call result B={tails}.
  - Are A and B mutually exclusive?

2. Random Variable

- A ***random variable*** $X(s)$ is a real-valued function of the underlying event space $s \in S$.

  (typically, we just denote it is as X,

  i.e. we suppose the dependence on s.)

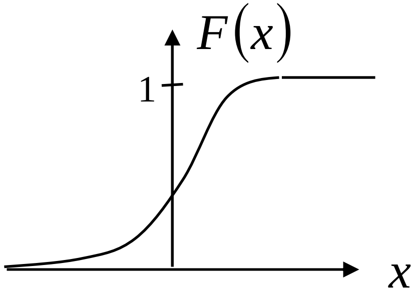
- Random variables (r.v.'s) can be either discrete or continuous:
  - A **discrete r.v.** can only take on a countable number of values.

    Ex.: The number of students in the class; discrete number of gray scale levels.
  - A **continuous r.v.** can take on a continuous range of values.

    Ex.: The current or voltage in the circuit.

2. Cumulative Distribution Function

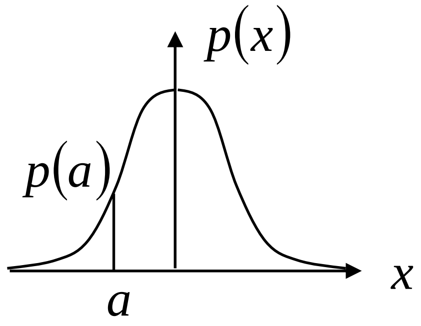
- Abbreviated ***CDF***.
- Also called Probability Distribution Function.
- Definition: $\boxed{F_x(a) = P(X \le x)}$
- Properties:
    - $F(x)$ is monotonically nondecreasing.
    - $F(-\infty) = 0$
    - $F(+\infty) = 1$
    - $P[a < X \le b] = F(b) - F(a)$



- The CDF completely defines the random variable, but is cumbersome to work with it.
- Instead, we will use the pdf ......

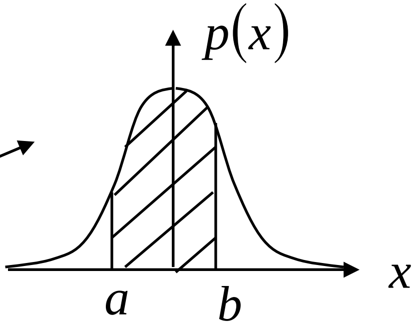
2. Probability Density Function

- Abbreviated ***pdf***.
- Definition: $$p_x(x) = \frac{d}{dx} F_x(x)$$



Properties:

- $p(x) \geq 0$
- $\int_{-\infty}^{+\infty} p(x)dx = 1$
- $\int_{a}^{b} p(x)dx = P[a < X \leq b] = F(b) - F(a)$

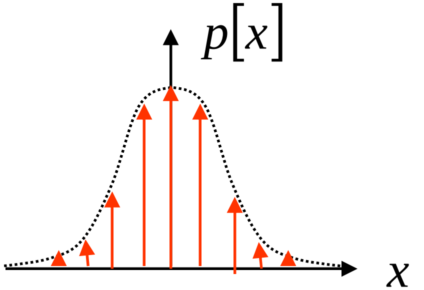


- Interpretation:
  - Measures how fast the CDF is increasing.
  - Measures how likely a r.v. is to lie at a particular value or within a range of values.

2. Probability Mass Function

- The pdf of discrete r.v.'s consists of a set of weighted dirac delta functions.
  - Delta functions can be cumbersome to work with.
- Instead, we can define the ***probability mass function*** (pmf) for discrete random variables:

$$p[x] = P[X = x]$$



- Properties of the pmf:
  - $p[x] \geq 0$
  - $\sum_{x=-\infty}^{+\infty} p[x] = 1$
  - $\sum_{x=a}^{b} p[x] = P[a \leq X \leq b]$

2. Expected Values

- Sometimes the pdf is described by the moments.
- *Expected values* are a shorthand way of describing a random variable.
- The most used are:

Discrete analog

- Mean: $\boxed{E[X] = m_x = \bar{x} = \int_{-\infty}^{+\infty} x p(x) dx}$ $\quad m_x = \bar{x} = \sum_{i=1}^{M} x_i P(X = x_i)$
- Variance: $\boxed{\sigma_x^2 = E\left[(X - m_x)^2\right] = \int_{-\infty}^{+\infty} (x - m_x)^2 p(x) dx = E\left[X^2\right] - m_x{}^2 = Var[X]}$

- The expectation operator works with any function $Y = g(X)$

$$\boxed{E[Y] = m_y = \bar{y} = \int_{-\infty}^{+\infty} g(x) p(x) dx}$$

2. Expected Values

- Derive that: $\boxed{Var[X] = E\left[(X - m_x)^2\right] = E\left[X^2\right] - m_x^2}$

$$Var[X] = E\left[(X - m_x)^2\right] = \int_{-\infty}^{+\infty} (x - m_x)^2 p(x) dx =$$

$$= \int_{-\infty}^{+\infty} x^2 p(x) dx - \int_{-\infty}^{+\infty} 2m_x x p(x) dx + \int_{-\infty}^{+\infty} m_x^2 p(x) dx =$$

$$= E\left[X^2\right] - 2m_x \int_{-\infty}^{+\infty} x p(x) dx + m_x^2 \int_{-\infty}^{+\infty} p(x) dx =$$

$$= E\left[X^2\right] - 2m_x^2 + m_x^2$$

$$= E\left[X^2\right] - m_x^2$$

2. Expected Values

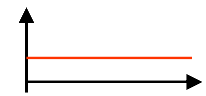
- For a random variable X, with expected value m_x

and variance $Var[X] = E\left[(X - m_x)^2\right]$.

- $$Var[X] = E[X^2] - (E[X])^2 = E[X^2] - m_x^2$$

- *If X always takes the value b (constant), then* $Var[X] = 0$

  $E[X] = b$



- *If* $Y = X + b$, *then* $Var[Y] = Var[X]$

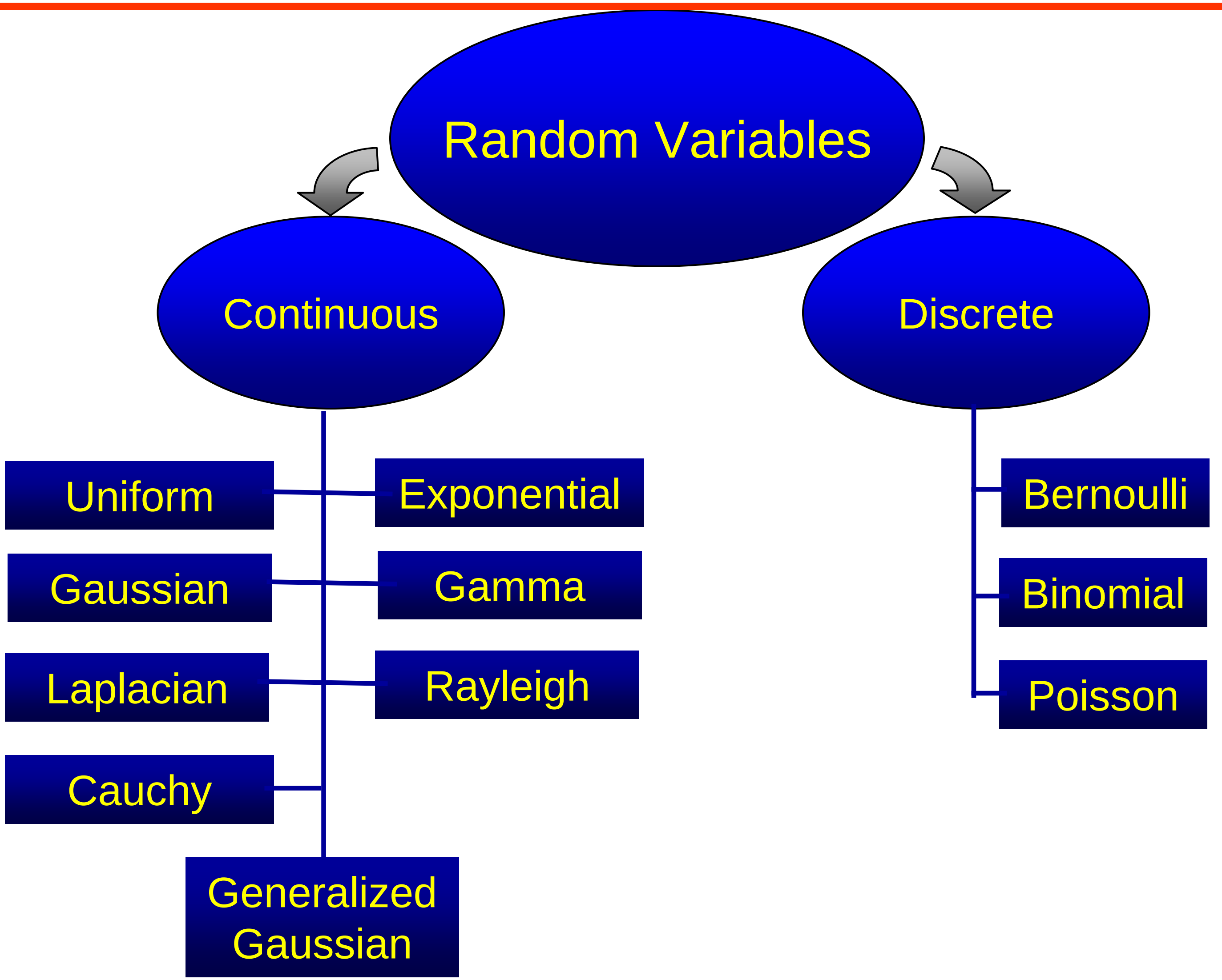
  $$Var[Y] = E\left[(Y - m_y)^2\right] = E\left[((X + b) - (E[X] + b))^2\right] = E\left[(X - E[X])^2\right] = Var[X]$$

- *If* $Y = aX$, $Var[Y] = a^2 Var[X]$

  $$Var[Y] = E\left[(aX - aE[X])^2\right] = E\left[a^2 (X - E[X])^2\right] = a^2 Var[X]$$

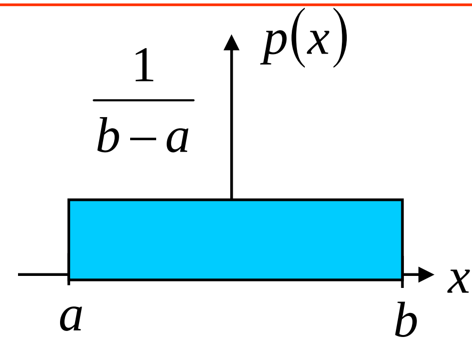
  $$E[Y] = aE[X]$$

2. Examples of distributions



2. Uniform Distribution

$$p(x)=\begin{cases}1/(b-a), & a \le x < b,\\ 0, & otherwise.\end{cases}$$



Often used for phase modeling:

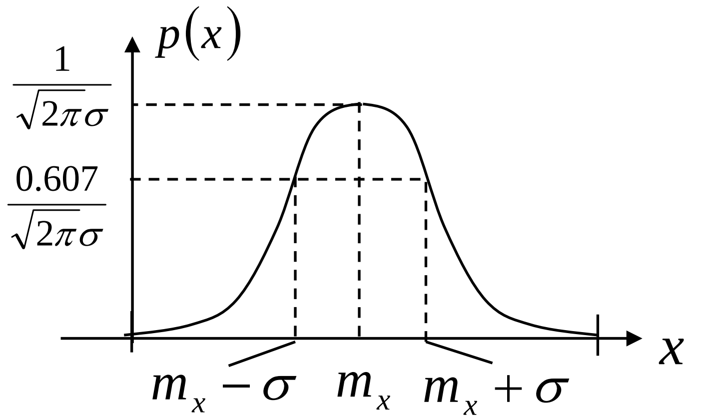
$$a = 0, b = 2\pi$$

$$E[X] = (b+a)/2, \quad Var[X] = (b-a)^2/12$$

$$E[X] = \int_a^b \frac{1}{b-a} x dx = (b+a)/2$$

2. Gaussian Distribution

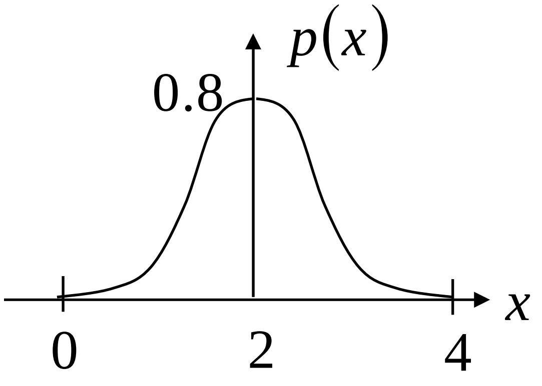
$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m_x)^2/2\sigma^2} \Rightarrow N(m_x, \sigma^2)$$



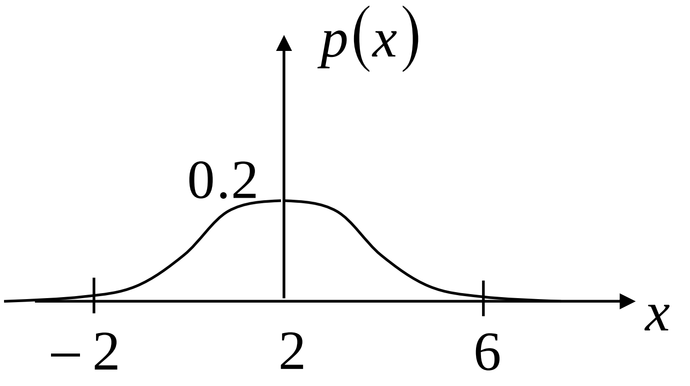
$$E[X] = m_x, \quad Var[X] = \sigma^2$$

$$Y = aX + b \Rightarrow Y \sim N\left(am_x + b, (a\sigma)^2\right)$$

2. Gaussian Distribution



$m_x = 2, \sigma = 0.5$



$m_x = 2, \sigma = 2$

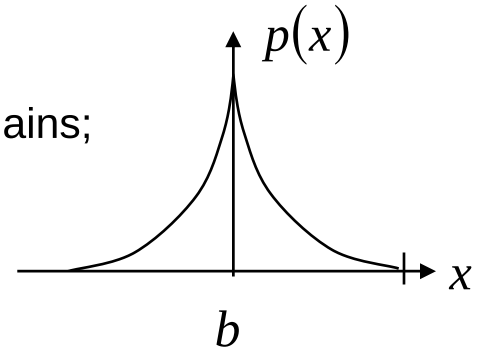
2. Laplacian Distribution: double exponential

$$p(x) = \frac{a}{2} e^{-a|x-b|}$$

$a > 0, \quad -\infty < b < \infty$

Laplacian pdf is used for modeling of:

- images in the transform domains;
- sparse data;
- outliers and impulse noise.



$$E[X] = b, \quad Var[X] = 2/a^2$$

2. Generalized Gaussian Distribution

$$p(x) = \left(\frac{\gamma \eta(\gamma)}{2\Gamma\left(\frac{1}{\gamma}\right)} \right) \cdot \frac{1}{\sigma_n} \cdot \exp\left\{ -\eta(\gamma) \left| \frac{x}{\sigma_n} \right|^{\gamma} \right\}$$

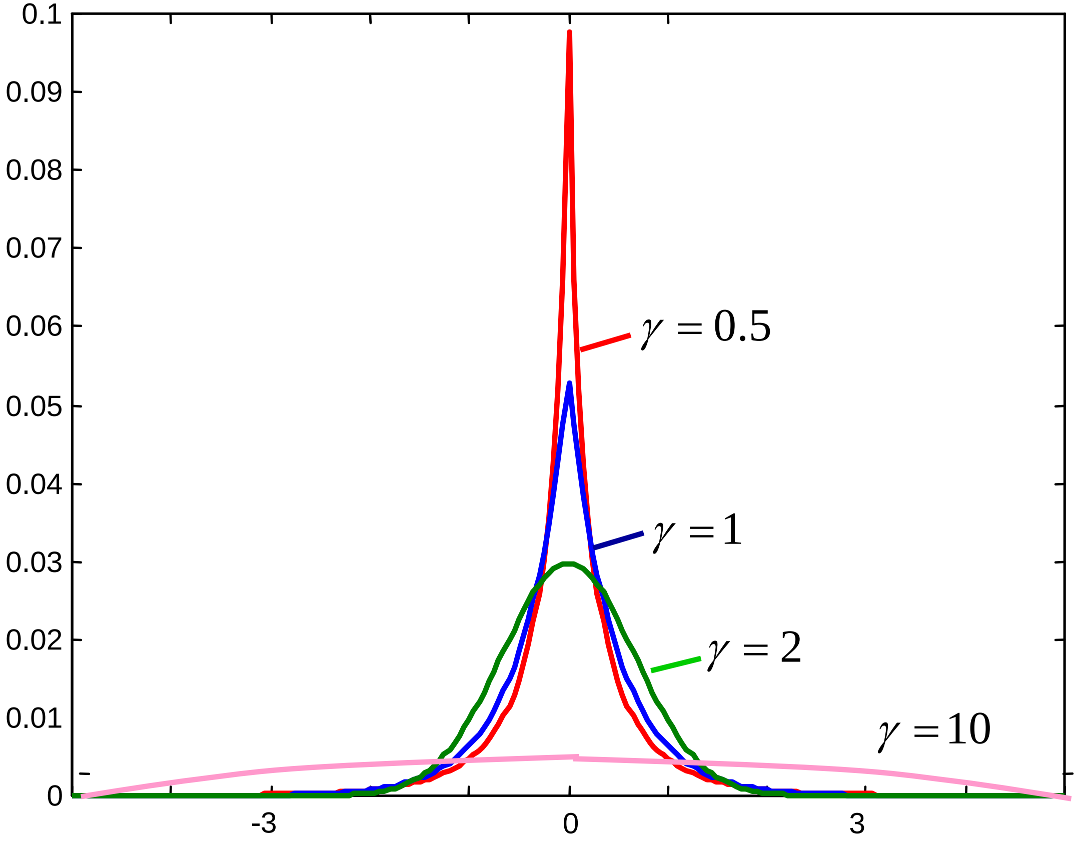
$$\eta(\gamma) = \sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}$$

$$\Gamma(t) = \int_0^{\infty} e^{-u} u^{t-1} du$$

Features:

- Generalized model for many distributions from the exponential family:
  - $\gamma = 2$ Gaussian
  - $\gamma = 1$ Laplacian
  - $\gamma \to \infty$ Uniform

2. Generalized Model: Generalized Gaussian Noise

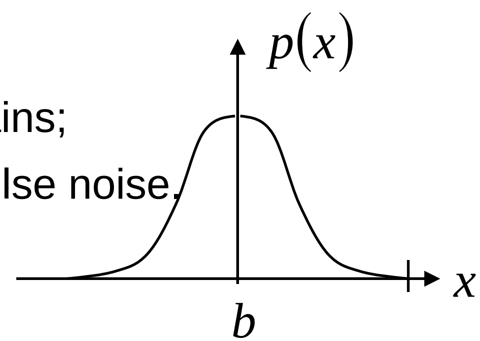


2. Cauchy Distribution

$$\boxed{p(x) = \frac{1}{\pi} \frac{a}{a^2 + (x-b)^2}} \qquad a > 0, \quad -\infty < b < \infty$$

Cauchy pdf is used for modeling of:

- images in the transform domains;
- heavy-tailed outliers and impulse noise.



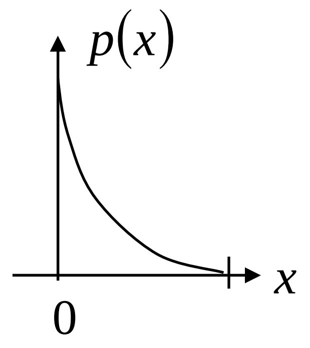
$$\boxed{E[X] \equiv b, \quad Var[X] = \infty}$$

due to symmetry

Difficult to define

2. Exponential Distribution

$$p(x) = \begin{cases} ae^{-ax}, \; x \geq 0, \\ 0, \quad otherwise. \end{cases} \qquad a > 0$$



$$E[X] = 1/a, \quad Var[X] = 1/a^2$$

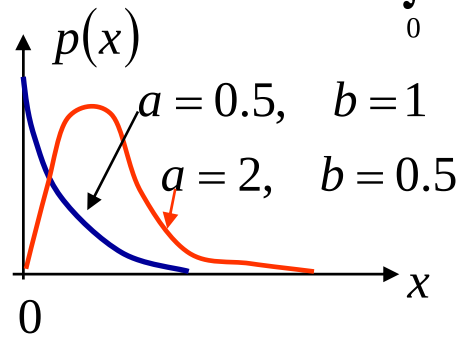
2. Gamma Distribution

$$p(x)=\begin{cases}\frac{b^{-a}}{\Gamma(a)}x^{a-1}e^{-x/b}, x>0,\\ 0, \quad otherwise.\end{cases} \qquad a>0, \quad b>0$$

$$\Gamma(x)=\int_0^{\infty}t^{x-1}e^{-t}dt$$

Gamma pdf is used for
modeling of:

- non-negative r.v.s;
- doubly-stochastic processes ;
- variance.



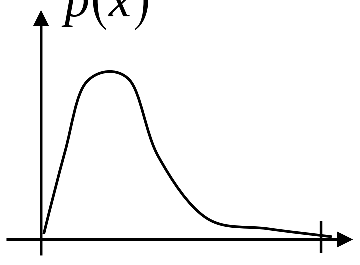
$$\boxed{E[X]=ab, \quad Var[X]=ab^2}$$

2. Rayleigh Distribution

$$p(x)=\begin{cases} a^2 x e^{-a^2x^2/2}, \; x \geq 0, & a>0 \\ 0, \quad otherwise. \end{cases}$$

Rayleigh pdf is used for
 modeling of:

- magnitude of complex valued r.v.s;
- fading in the communications channels;
- mapping from cartesian to polar coordinate.



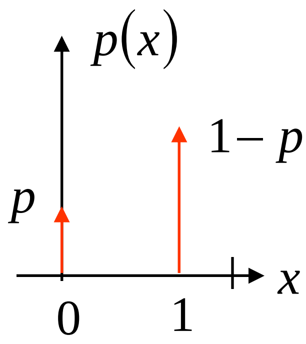
$$\boxed{E[X]=\sqrt{\frac{\pi}{2a^2}}, \quad Var[X]=\frac{2-\pi/2}{a^2}}$$

$$\begin{array}{c} x=\sqrt{r^2+i^2} \\ \phi=\arctan\left(\dfrac{i}{r}\right) \end{array} \Rightarrow \begin{array}{c} r=x\cos\phi \\ i=x\sin\phi \end{array} \Rightarrow p_{x\phi}(x,\phi)=p_x(x)p_\phi(\phi) \Rightarrow \begin{array}{c} p_\phi(\phi)=1/2\pi \\ p_x(x)=p(x) \end{array}$$

2. Binary or Bernoulli Distribution

$$p(x) = \begin{cases} p, & x = 0, \\ 1-p, & x = 1, \\ 0, & otherwise. \end{cases} \qquad 0 \le p \le 1$$

$$p(x) = p\delta(x-0) + (1-p)\delta(x-1)$$



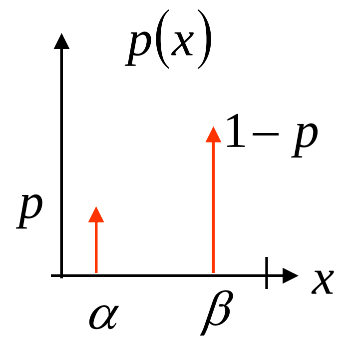
$$\boxed{E[X] = 0 \cdot \mathbf{Prob}(0) + 1 \cdot \mathbf{Prob}(1) = 0(p) + 1(1-p) = 1-p, \quad Var[X] = p(1-p)}$$

2. Binary or Bernoulli Distribution: General Case

Salt and pepper noise:

$$\alpha = x_{\min} = 0$$
$$\beta = x_{\max} = 255$$

$$p(x) = \begin{cases} p, & x = \alpha, \\ 1-p, & x = \beta, \\ 0, & otherwise. \end{cases} \qquad 0 \le p \le 1$$



$$p(x) = p\delta(x-\alpha) + (1-p)\delta(x-\beta)$$

$$\boxed{E[X] = \alpha p + \beta(1-p), \quad Var[X] = (\alpha-\beta)^2 p(1-p)}$$

$$m_x = E[X] = \int xp(x)dx = \int x\left(p\delta(x-\alpha) + (1-p)\delta(x-\beta)\right)dx = \alpha p + \beta(1-p)$$

2. Binomial Distribution

$$p(x)=\begin{cases}\binom{n}{x}p^x(1-p)^{n-x}, & x=0,1,2,\ldots,n\\ 0, & otherwise.\end{cases}\qquad 0\le p\le 1,\ n-\text{int}$$

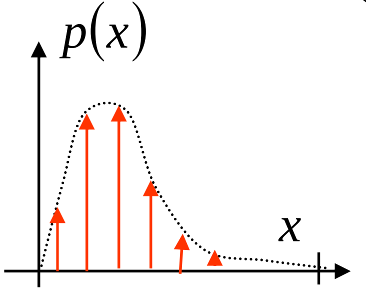
$$C_x^n=\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

<u>Given:</u>

a sequence of n idependent trials

each with success probability p.

The number of success is a binomial r.v.



$$\boxed{E[X]=np,\quad Var[X]=np(1-p)}$$

If $n=1$ Bernoulli r.v. is Binomial r.v.

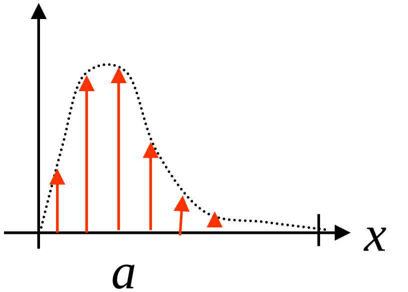
2. Poisson Distribution

$$p(x) = \begin{cases} \dfrac{a^x e^{-a}}{x!}, & x = 0,1,2,\ldots \\ 0, & otherwise. \end{cases} \qquad a > 0$$

a - average rate

(appearance per second or per space unit)

T interval



Poisson pdf allows to find:

- probability of certain number of appearance per given time interval.

It is used for:

- demand or request modeling;
- medical imaging (photon counting devices).

$$E[X] = a, \quad Var[X] = a$$

If $p \to 0$ and, $n \to \infty$

and constraint that $np \to m$

then Binomial → Poisson.

2. Multiple Random Variables

- The experimant produces not a single r.v. as before, but several r.v.s.
- Joint pdf of two r.v.s: $p_{XY}(x, y)$
- Conditional pdf:

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

Bayes' rule

$$\boxed{p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y)p_Y(y)}{p_X(x)}}$$

- Marginal pdf:

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

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

2. Independent, Orthogonal and Uncorrelated R.V.s

- Two r.v.s are called **independent** iff their joint pdf is a product of their individual pdf's, i.e.

$$\boxed{p_{XY}(x,y) = p_X(x)p_Y(y)}$$

Independent and identically distributed

$$p_{XY}(x,y) = p_X(x)p_X(y)$$

- Two r.v.s are **orthogonal**, if:

$$\boxed{E[XY] = 0}$$

- and **uncorrelated**, if:

$$\boxed{E[XY] = E[X]E[Y]}$$

- Gaussian r.v.s which are uncorrelated are also independent.

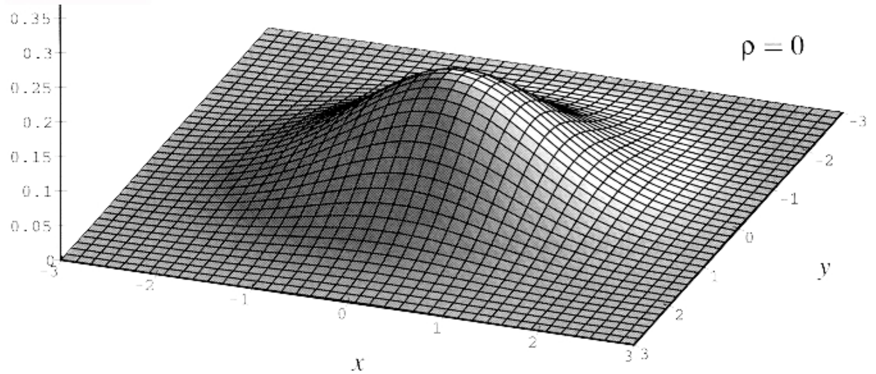
2. Bivariate Gaussian pdf

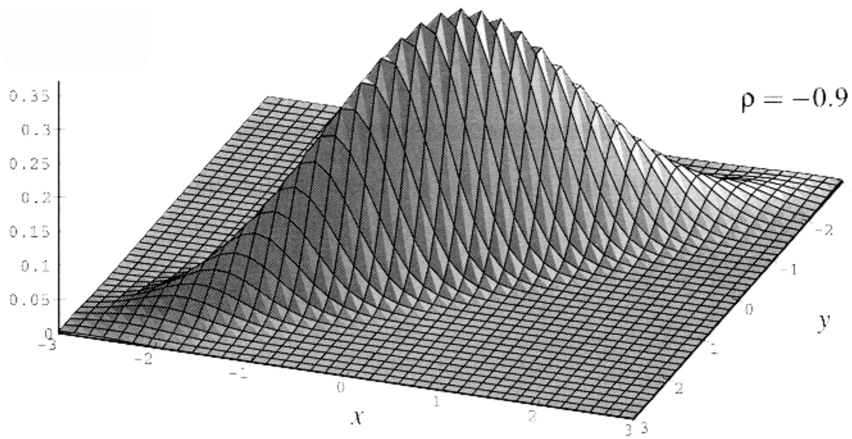
Joint pdf:

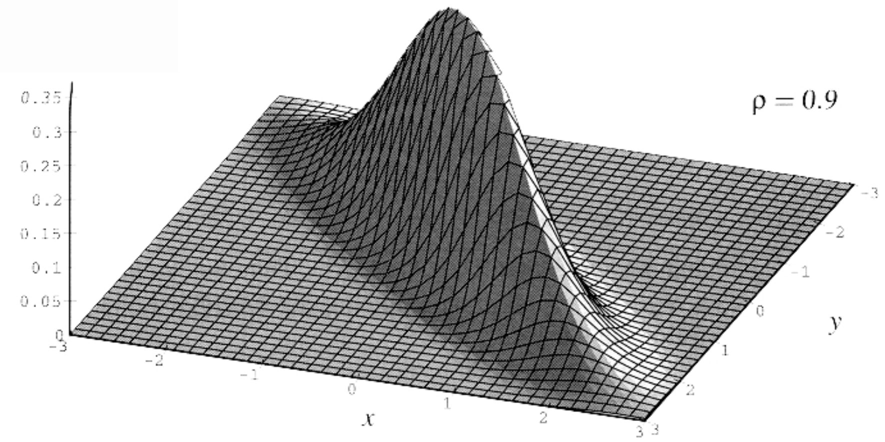
$$p_{XY}(x,y) = \frac{\exp\left[-\dfrac{\left(\dfrac{x-m_x}{\sigma_x}\right)^2 - \dfrac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \left(\dfrac{y-m_y}{\sigma_y}\right)^2}{2(1-\rho^2)} \right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

$$-1 \le \rho \le 1$$

2. Bivariate Gaussian pdf



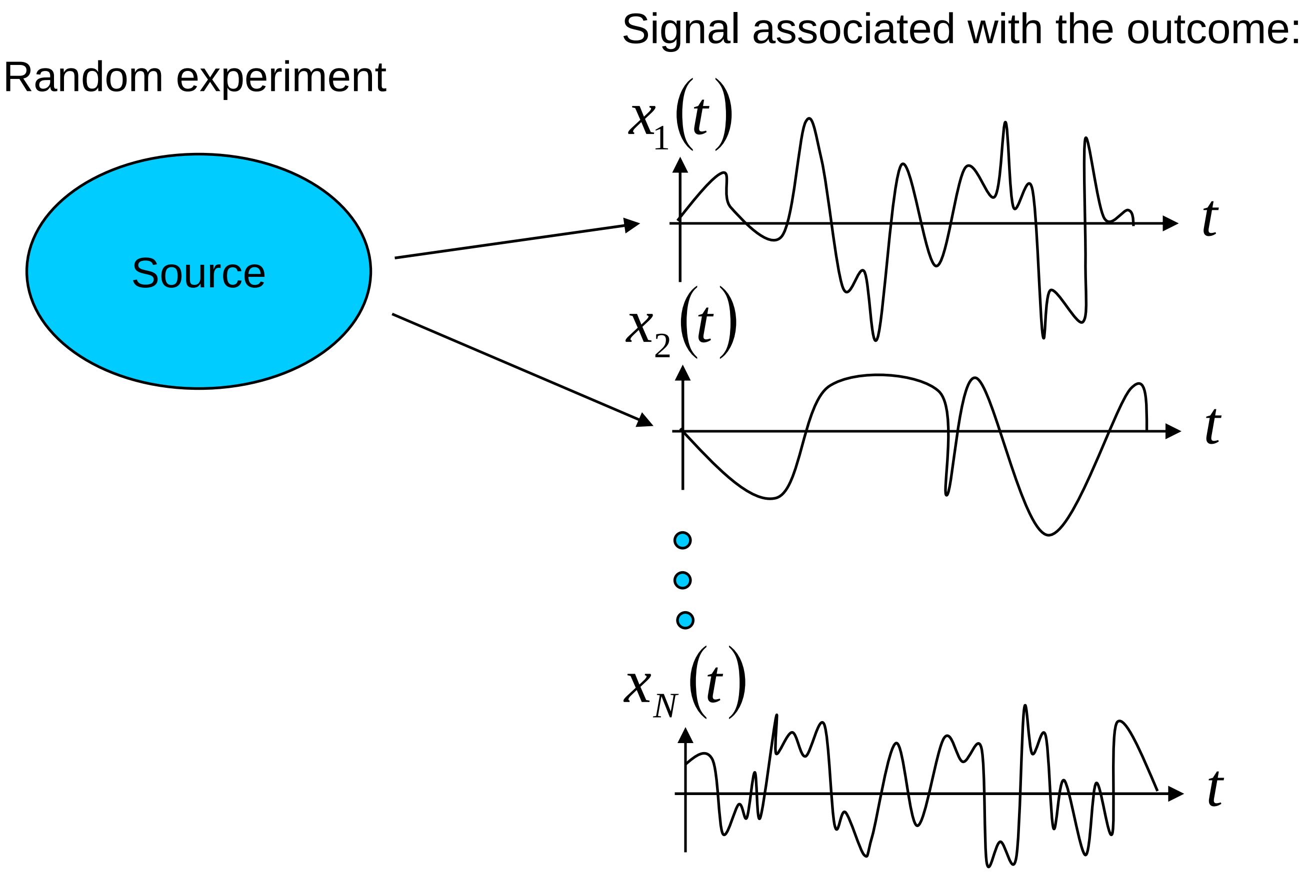




3. Random Processes

- Random variables model unknown values.
  - Random variables are numbers.
- Random processes model unknown signals.
  - Random processes are functions of time (space).
- A random process is just a collection of random variables.
  - A random process evaluated at a specific time t is a random variable.
  - If $x(t)$ is a random process then $x(1), x(1.5), \textit{and } x(20.5)$ are all random variables.

3. Random Processes



3. Random Processes [Stark&Woods Ch.7]

- Definition: random process is a collection of random variables

$\{x(t), \quad t \in T\}$ where $T =$ index set (parameter set)

e.g. $Z, Z^M, \Re, \Re^M, \ldots$

$x(t) =$ random variable (real-valued, binary-valued, vector-valued, complex-valued,..)

Difficulty: when $|T| = \infty$:

- notation of pdf breaks down;
- definition of ∞-dim distribution functions requires care.

3. Random Processes

- Examples: $T = Z^2$

  - $x(n_1, n_2) = i.i.d.$ random variables [spatially white noise]
  - $x(n_1, n_2) = A \quad \forall n_1, n_2$ where A= random variable
  - $x(n_1, n_2) \in \{0,1\}$ binary random field

3. Random Processes Terminology

- The **expected value** of random process plays a central role in modeling and processing of signals.
- Furthermore, the probability models of a random process are usually expressed as functions of the expected values.
  - Examples:
    - – Gaussian pdf is defined as an exponential function of the mean and the variance of the process;
    - – Poisson pdf is defined in terms of the mean of the process.
- The expected value of a function $g(x(t_1), x(t_2), \ldots, x(t_M))$ of a random process $x(t)$ is defined as:

$$\boxed{\begin{aligned} E[g(x(t))] &= \int_{-\infty}^{+\infty} g(x(t))p(x,t)dx \equiv \\ &\int \ldots \int g(x_1, x_2, \ldots, x_M) p_{x(t_1), x(t_2), \ldots, x(t_M)}(x_1, x_2, \ldots, x_M) dx_1 dx_2 \ldots dx_M \end{aligned}}$$

3. Moments of a random process

- The most important expected values or moments are:
  - **Mean** $\{m_x(t), t \in T\}$ where $m_x(t) = E_{p_t}[x(t)] = \int_{-\infty}^{+\infty} x(t)p(x,t)dx$
  - **AutoCorrelation** function $\{R_x(t_1,t_2), t_1,t_2 \in T\}$ where $R_x(t_1,t_2) = E_{p_{t_1},p_{t_2}}[x(t_1)x(t_2)]$
  - **AutoCovariance** function $\{K_x(t_1,t_2), t_1,t_2 \in T\} = \textit{correlation function for}$ $x(t) - m_x(t)$ where $K_x(t_1,t_2) = R_x(t_1,t_2) - m_x(t_1)m_x(t_2)$
  - **n-th order** moment = $E_{p_{t_1},\ldots,p_{t_n}}[x(t_1),\ldots,x(t_n)]$
    - **variance** $\sigma_x^2(t) = R_x(t,t) - (m_x(t))^2$

3. Mean and Autocorrelation

- Example: $x(t) = \sin(2\pi t + \theta)$
  - This is just a function $g(\theta)$ of θ :

$$g(\theta) = \sin(2\pi t + \theta)$$

  - The expected value of a function of a random variable:

$$E[x(t)] = E[g(\theta)] = E[\sin(2\pi t + \theta)]$$

    – To find this we need to know the pdf of θ.

3. Mean and Autocorrelation

- Example: if θ is uniform between 0 and π, then:

$$m_x(t) = E[\sin(2\pi t + \theta)] = \int_{-\infty}^{+\infty} \sin(2\pi t + \theta) p_\theta(\theta) d\theta$$

$$= \int_0^{\pi} \sin(2\pi t + \theta)\left(\frac{1}{\pi}\right) d\theta = \frac{2}{\pi}\cos(2\pi t)$$

$$R_x(t_1, t_2) = E[\sin(2\pi t_1 + \theta)\sin(2\pi t_2 + \theta)]$$

$$= \int_{-\infty}^{+\infty} \sin(2\pi t_1 + \theta)\sin(2\pi t_2 + \theta) p_\theta(\theta) d\theta$$

$$= \int_0^{\pi} \sin(2\pi t_1 + \theta)\sin(2\pi t_2 + \theta)\left(\frac{1}{\pi}\right) d\theta = \frac{1}{2}\cos(2\pi(t_2 - t_1))$$

3. Stationarity

- A process is **strict-sense stationary (SSS)** if all its joint densities are invariant to a time shift:

$$\boxed{\begin{aligned} &p_x(x(t)) = p_x(x(t+t_0)) \\ &p_x(x(t_1), x(t_2)) = p_x(x(t_1+t_0), x(t_2+t_0)) \\ &p_x(x(t_1), x(t_2), \ldots, x(t_N)) = p_x(x(t_1+t_0), x(t_2+t_0), \ldots, x(t_N+t_0)) \end{aligned}}$$

  - in general, it is difficult to prove that a random process is strict sense stationary.

- A process is **wide-sense stationary (WSS)** if:
  - The mean is a constant:

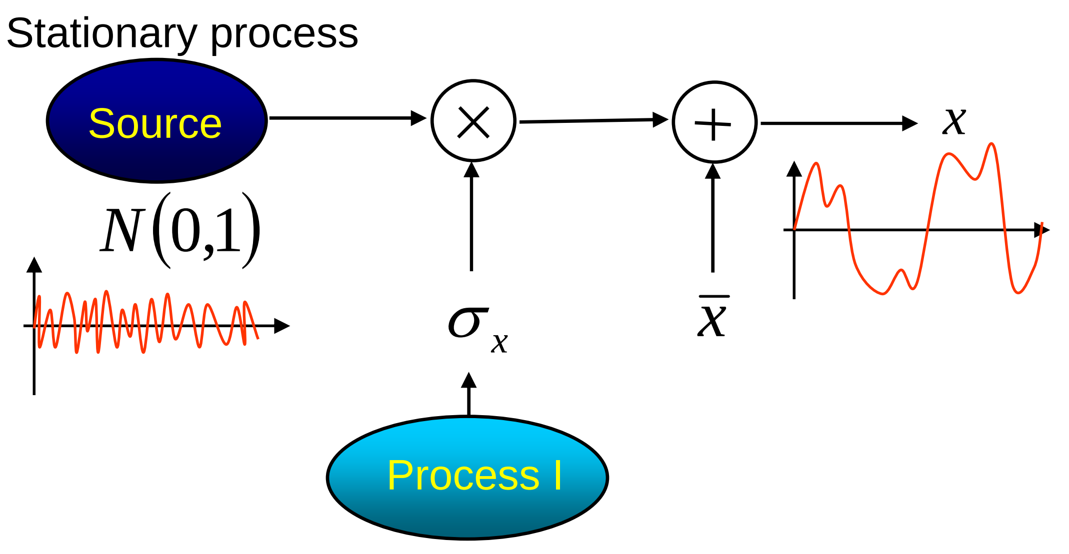
$$m_x(t) = m_x$$

  - The ACF is a function of time difference only:

$$R_x(t_1, t_2) = R_x(t_1 - t_2) = R_x(\tau)$$

3. Non-stationary Processes

A R.P. is non-stationary if its statistics vary in time or in space.



$$X \propto N(\bar{x}, R_x)$$

Doubly stochastic process: gamma, exponential, Jefrey

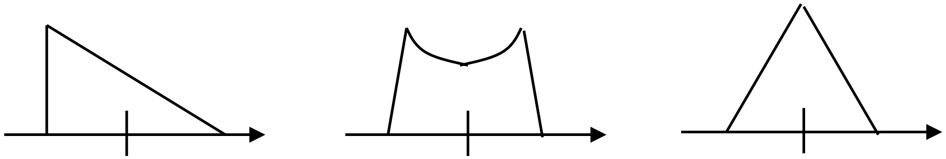
3. Properties of the Autocorrelation Function

- If $x(t)$ is Wide Sense Stationary, then its autocorrelation function has the following properties:

$R_x(0) = E\left[\lvert x(t)\rvert^2\right]$	this the second moment
$R_x(\tau) = R_x(-\tau)$	even symmetry
$R_x(0) \geq \lvert R_x(\tau)\rvert$	

- Examples:

Which of the following are valid ACF's?



3. Wiener-Khintchin Theorem

- We can find the **power spectral density** (PSD) for WSS random process.
- **Wiener-Khintchin theorem**:

if $x(t)$ is a wide sense stationary random process, then:

$$\boxed{P_x(f) = \Im\{R_x(\tau)\} = \int_{-\infty}^{+\infty} R_x(\tau) e^{-j2\pi f\tau} d\tau} \qquad \boxed{R_x(\tau) = \Im^{-1}\{P_x(f)\}}$$

i.e. the PSD is the Fourier Transform of the ACF.

3. Wiener-Khintchin Theorem

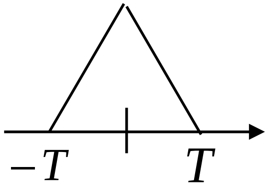
$$P_x(f) = \Im\{R_x(\tau)\} = \int_{-\infty}^{+\infty} R_x(\tau) e^{-j2\pi f\tau} d\tau$$

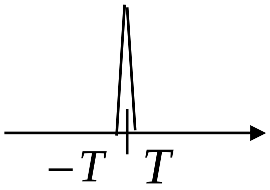
$$R_x(\tau) = \Im^{-1}\{P_x(f)\}$$

- Example for home exercise:

  Find the PSD of a WSS R.P. with the ACF:

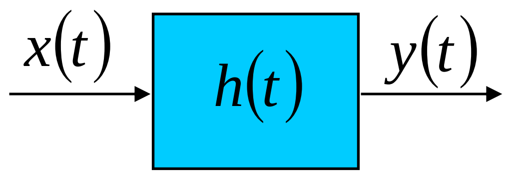
$$R_x(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{T}, & if\ |\tau| \le T, \\ 0, & if\ |\tau| > T. \end{cases}$$





3. Linear Systems

- The output of a linear time invariant (LTI) system is found by convolution.



$$y(t) = x(t) * h(t) \Longleftrightarrow Y(f) = X(f)H(f)$$

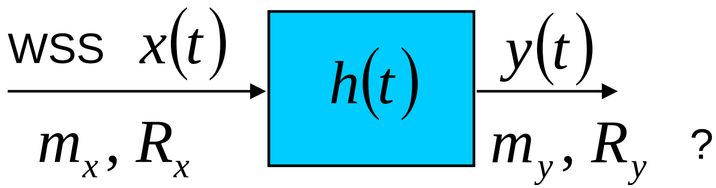
- However, if the input to the system is a random process, we can't find $X(f)$.

- Solution: use power spectral densities:

$$P_y(f) = P_x(f)\left|H(f)\right|^2$$

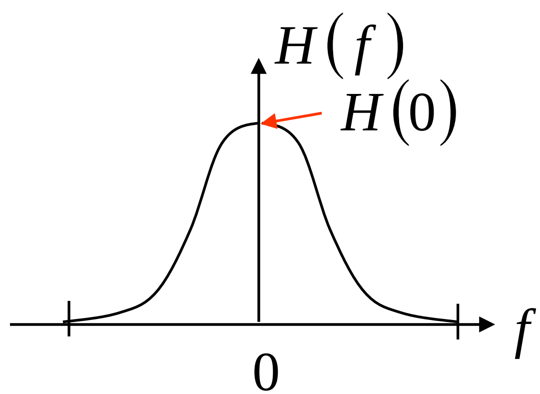
This implies that the output of a LTI system is WSS, if the input is WSS.

3. Linear Filtering of Random Processes



Mean: $m_y = E[Y] = \int_{-\infty}^{+\infty} E[x(\tau)]h(t-\tau)d\tau = m_x \int_{-\infty}^{+\infty} h(t-\tau)d\tau$

$= m_x \int_{-\infty}^{+\infty} h(\tau)d\tau = m_x H(0).$



3. Linear Filtering of Random Processes

ACF: $R_y(\tau) = E[y(\tau)y(t+\tau)] = \int \int_{-\infty}^{+\infty} E[x(\tau)x(\alpha)]h(t-\tau)h(t+\tau-\alpha)d\tau d\alpha$

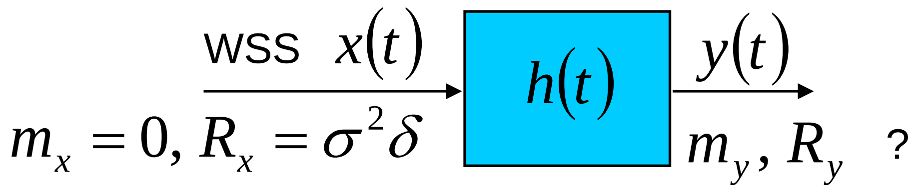
$$= \int \int_{-\infty}^{+\infty} R_x(t-\alpha)\underbrace{h(t-\tau)h(t+\tau-\alpha)}d\tau d\alpha \rightarrow \text{convolution}$$

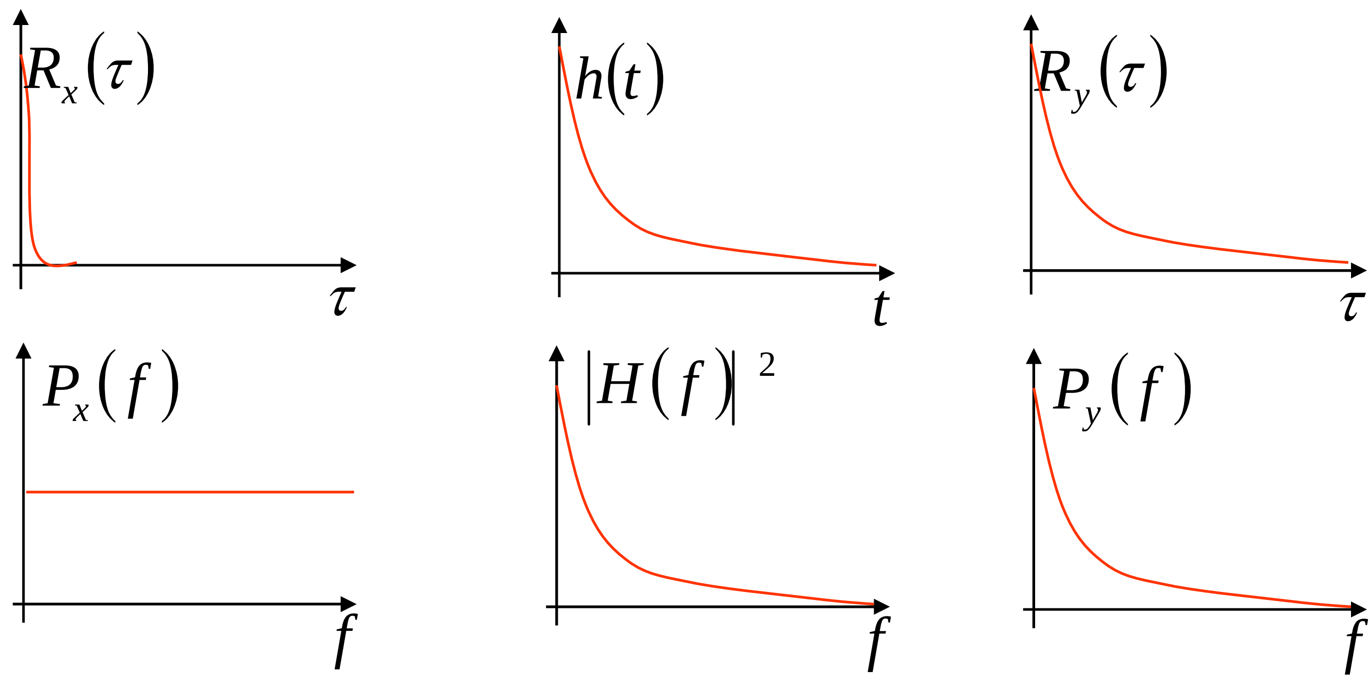
$$\boxed{R_y(t) = R_x(t) * R_h(t)}$$

In frequency domain:

$$P_y(f) = P_x(f)|H(f)|^2$$

3. Linear Filtering of Random Processes

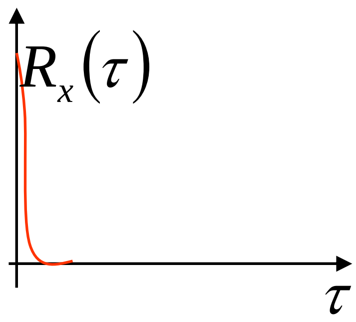


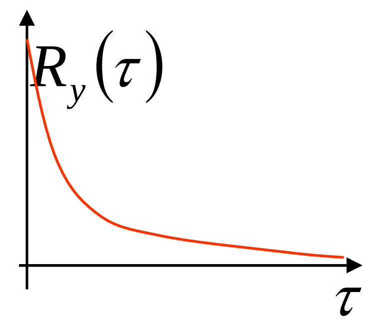


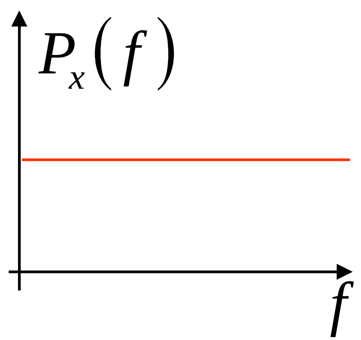
3. Inverse problem: System Identification

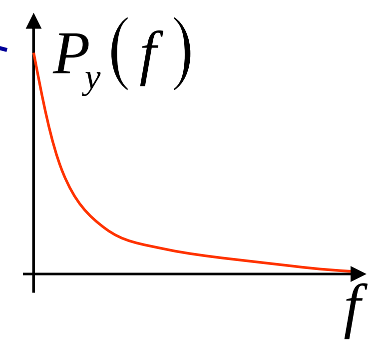
WSS $x(t)$ → $h(t)$ → $y(t)$

$m_x = 0,\ R_x = \sigma^2 \delta$ $\quad$? $\quad$ $m_y,\ R_y$ -is measured







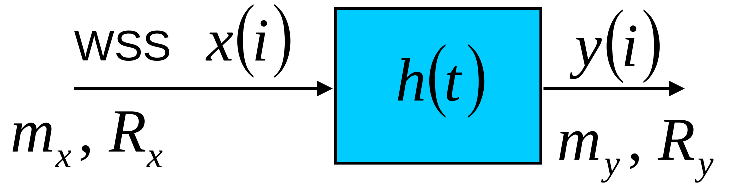


$$P_y(f) = P_x(f)\left|H(f)\right|^2$$

$$\left|H(f)\right|^2 = \frac{P_y(f)}{P_x(f)}$$

const

3. Digital Formulation

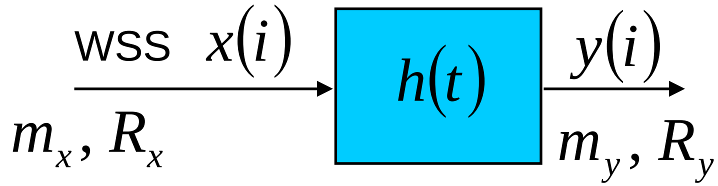


Convolution: $y(i) = x(i) * h(i) = \sum_{n=-\infty}^{+\infty} x_n h_{i-n}$

Mean: $m_y = m_x \sum_{n=-\infty}^{+\infty} h_n$

ACF: $R_y(n) = \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} h_i h_j R_x(n+i-j)$

3. Digital Formulation: example



Given: $h(i) = \begin{cases} 1, \; n = 0,1 \\ 0, \; otherwise \end{cases}$

$x(i)\; is\; WSS\; Gaussian: \quad m_x = 0.5 \quad R_x(i) = \begin{cases} 1, \; i = 0, \\ 0.5, \; i = 1, -1 \\ 0, \; |i| \geq 2. \end{cases}$

Find:

- the mean m_y
- the ACF R_y
- the variance $Var[Y(i)]$

3. Digital Formulation: example

Mean: $m_y = m_x \sum_{n=-\infty}^{+\infty} h_n = m_x (h_0 + h_1) = 2m_x = 1$

ACF: $R_y(n) = \sum_{i=0}^{1} \sum_{j=0}^{1} h_i h_j R_x(n+i-j) = 2R_x(n) + R_x(n-1) + R_x(n+1)$

Substituting R_x

$$R_y(n) = \begin{cases} 3, \ n = 0 \\ 2, \ |n| = 1 \\ 0.5, |n| = 2 \\ 0, \ otherwise. \end{cases}$$

Variance: $Var[Y] = E[Y^2] - m_y^{\ 2}$

$$E[Y^2] = R_y(0) = 3.$$

3. Multi-variate Gaussian process

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\left[\det(R_x)\right]^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_x)^T R_x^{-1}(\boldsymbol{x}-\boldsymbol{m}_x)\right]$$

$$R_x = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1N} \\ c_{21} & c_{22} & \cdots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \cdots & c_{NN} \end{bmatrix}$$

$$R_x = E\left[(X - M_X)(X - M_X)^T\right] = E\left[XX^T\right] - M_X M_X^T = \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \sigma_{ij}^2 & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}$$

3. Multi-variate Gaussian process

- Uncorrelated Gaussian process

$$R_x = \begin{bmatrix} \sigma_{x_1}^2 & & & 0 \\ & \sigma_{x_2}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{x_N}^2 \end{bmatrix}$$

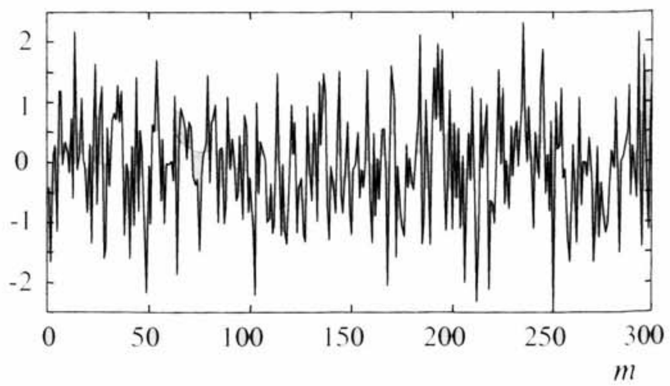
$$c_{ij} = E\left[\left(x_i - m_{x_i}\right)\left(x_j - m_{x_j}\right)^T\right] = 0 \qquad i \neq j$$

$$c_{ii} = E\left[\left(x_i - m_{x_i}\right)\left(x_i - m_{x_i}\right)^T\right] = \sigma_{x_i}^2$$

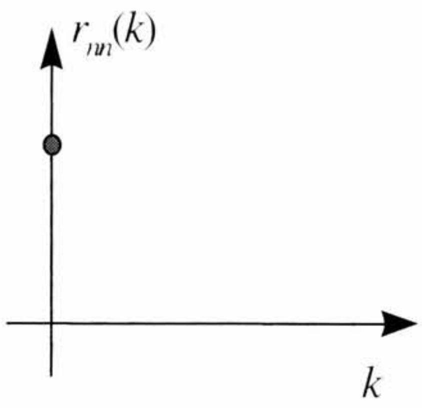
The uncorrelatedness implies independence for Gaussian processes!!!

3. White noise: stationary noise

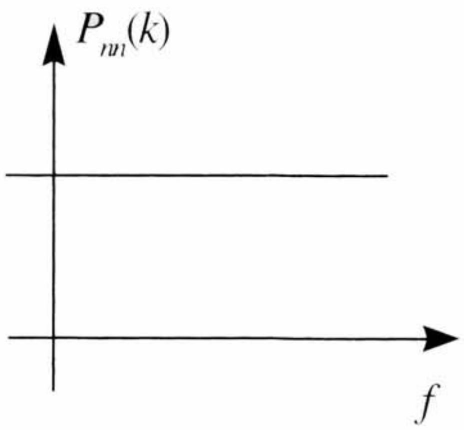
- White noise definition: uncorrelated process with equal power at all frequencies (theoretical concept).



White noise



Autocorrelation



Power spectrum

3. White noise

- Autocorrelation function of white noise $N(t)$:

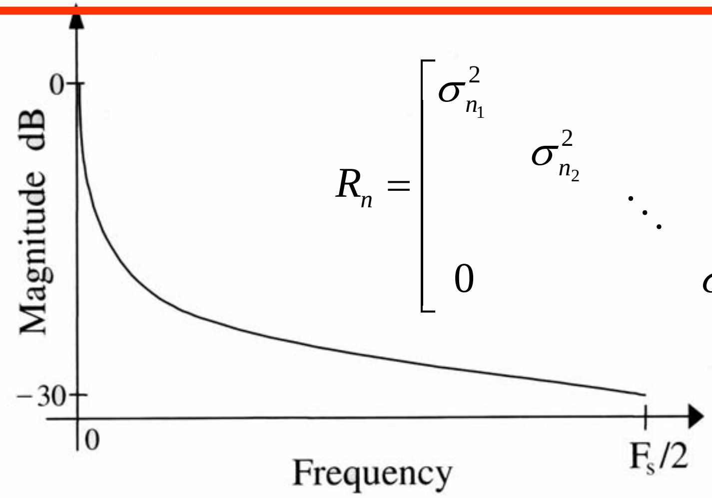
$$R_N(\tau) = E[N(t)N(t+\tau)] = \sigma_n^2 \delta(\tau)$$

$$R_N = \begin{bmatrix} \sigma_n^2 & & & 0 \\ & \sigma_n^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix}$$

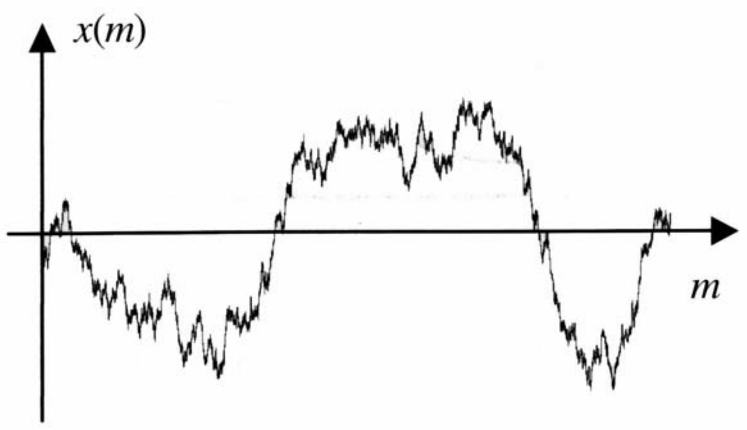
- Power spectrum of white noise:

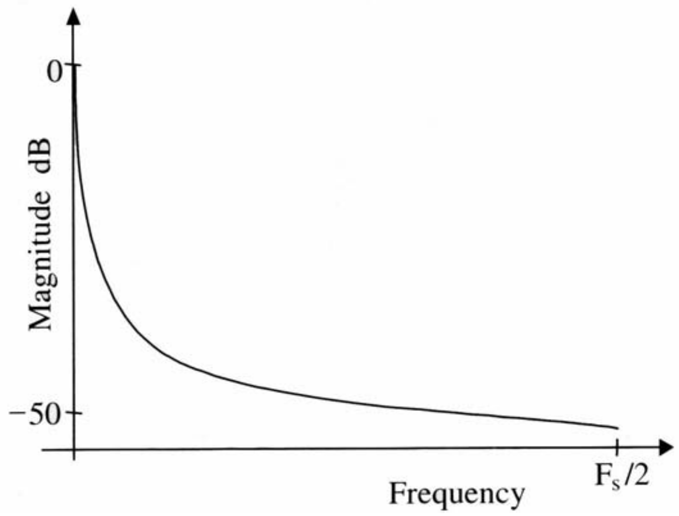
$$P_N(f) = \int_{-\infty}^{+\infty} R_N(t) e^{-j2\pi ft} dt = \sigma_n^2$$

3. Coloured noise: non-stationary noise



$$R_n = \begin{bmatrix} \sigma_{n_1}^2 & & & 0 \\ & \sigma_{n_2}^2 & & \\ & & \ddots & \\ 0 & & & \sigma_{n_N}^2 \end{bmatrix}$$



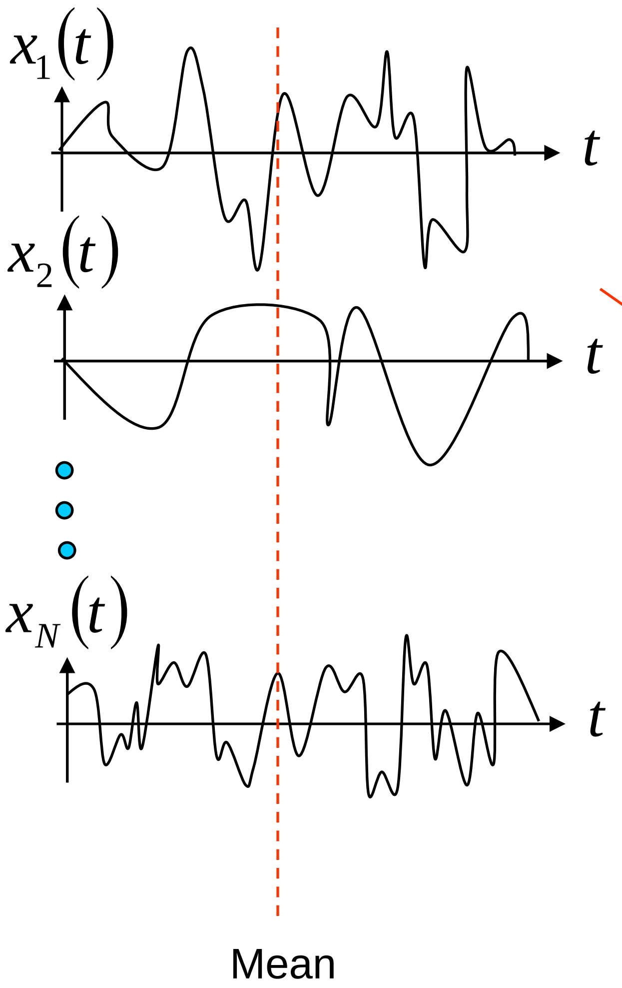


3. Ergodicity

- In many applications of DSP, there is only one single realization of a random process from which its statistical parameters, such as mean, the correlation, and the power spectrum can be estimated.

- In such cases time-averaged statistics, obtained from averages along the time dimension of a single realization of a process are used instead of the “true” ensemble averages obtained across the space of different realizations of the process.

3. Ergodicity



ERGODICITY

Estimation of mean from one realization

3. Ergodicity

In practice a statistical description of images (or any random process) is not available, therefore these quantities have to be estimated from the data.

Ex.: Consider the problem of estimating the mean m_x of a stationary R.P. given data $\{x(t), t \in \Lambda\}$

m_x can be estimated using the stochastic integral:

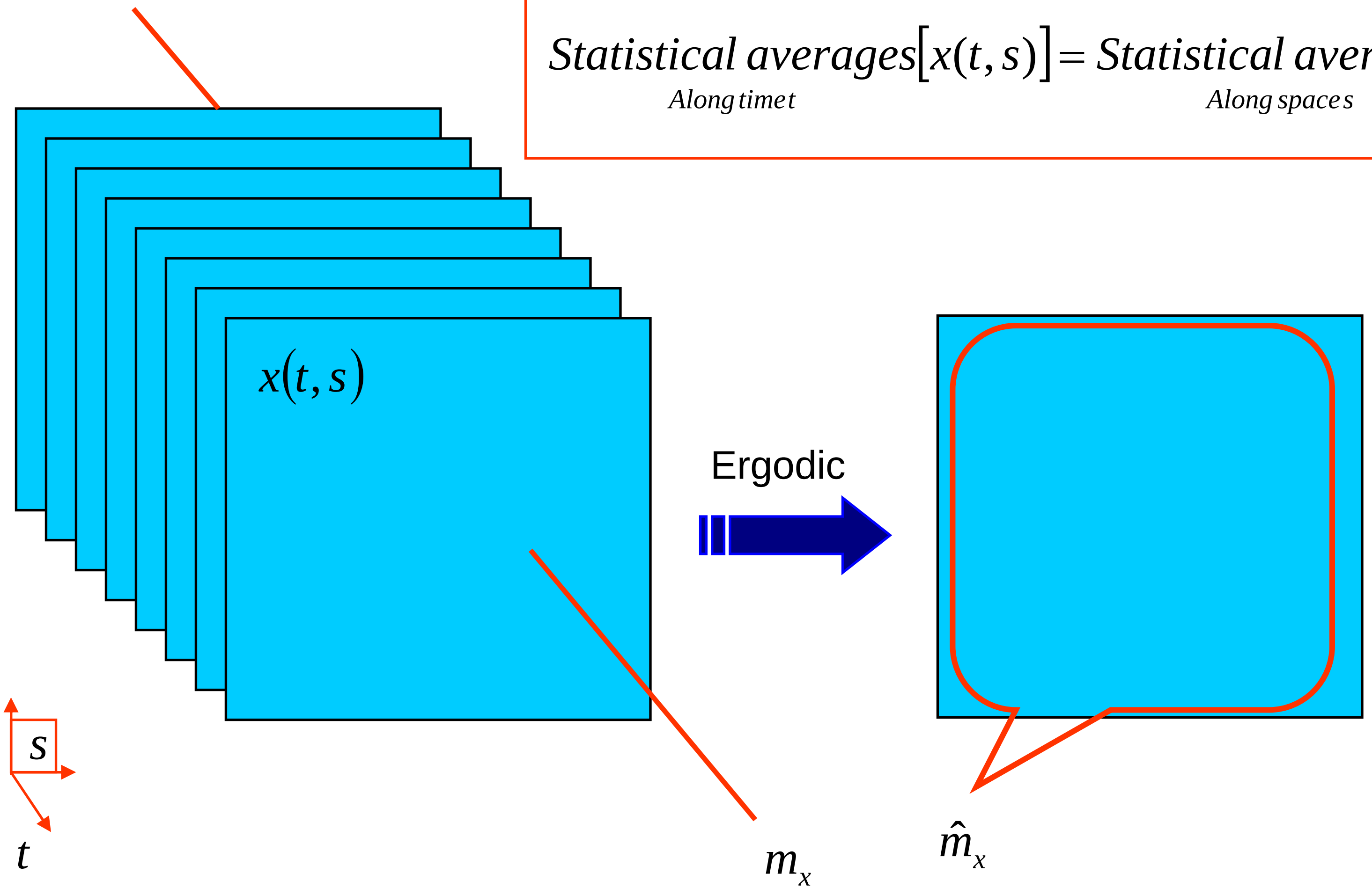
$$\boxed{\hat{m}_x = \frac{1}{|\Lambda|}\int_{\Lambda} x(t)dt} \qquad \hat{m}_x \to m_x \quad \text{as} \quad |\Lambda| \to \infty$$

Example: spatially white noise.

In this case, the process $x(t)$ is said to be ergodic in the mean.

3. Ergodicity: images

$$\underset{Along\,time\,t}{Statistical\ averages}[x(t,s)] = \underset{Along\,space\,s}{Statistical\ averages}[x(t,s)]$$



3. Ergodicity

Intuitively, $\hat{m}_x$ should be an average of many independent observations (so that a law of large numbers type result applies). This means that $x(t)$ should decorrelate rapidly enough with space shift, so the "correlation area" $\frac{1}{K_x(0)}\int_T K_x(t)dt$ should be small enough.

3. Ergodicity

- A random process is said to be "ergodic" if it is ergodic in the mean and ergodic in correlation:
  - **Ergodic in the mean**:

$$m_x = E[x(t)] = \langle x(t) \rangle$$

Time average operator:

$$\langle g(t) \rangle = \lim_{t \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} g(t)\,dt$$

Definition: A WSS process $\{x(t), t \in T\}$ is (m.s.) ergodic in the mean

iff $\hat{m}_x \to m_x$ as $|T| \to \infty$

3. Ergodicity

- **Ergodic in the correlation**:

$$R_x(\tau) = E[x(t)x(t+\tau)] = \langle x(t)x(t+\tau) \rangle$$

Definition: A WSS process $\{x(t), t \in T\}$ is (m.s.) ergodic in correlation at the shifts s iff

$$\hat{R}_x(s) \to R_x(s) \quad \text{as} \quad |T| \to \infty$$

where $\hat{R}_x = \frac{1}{|T|} \int_T x(t+s)x(t)dt$

- In order for a random process to be ergodic, it must first be **Wide Sense Stationary**.

3. Ergodicity

If a R.P. is ergodic, then we can compute power in three different ways:

- From any sample function: $$P_x = \lim_{t \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 dt = \left\langle |x(t)|^2 \right\rangle$$
- From the ACF: $$P_x = R_x(0)$$
- From the Power Spectral Density: $$P_x = \int_{-\infty}^{\infty} P_x(f) df$$

3. Cross-correlation

- If we have two random processes $x(t)$ and $y(t)$ we can define a **cross-correlation** function:

$$\{R_{xy}(t_1,t_2), t_1,t_2 \in T\} \quad \text{where} \quad R_{xy}(t_1,t_2) = E[x(t_1)y(t_2)]$$

- If $x(t)$ and $y(t)$ are **jointly stationary**, then the cross-correlation becomes:

$$R_{xy}(\tau) = E[x(t)y(t+\tau)]$$

- If $x(t)$ and $y(t)$ are **uncorrelated**, then:

$$R_{xy}(\tau) = m_x m_y$$

- If $x(t)$ and $y(t)$ are **independent**, then they are also uncorrelated, and thus:

$$E[x(t)y(t)] = E[x(t)]E[y(t)]$$