ITMO UNIVERSITY

Saint Petersburg, Russia

Methods and models for multivariate data analysis

Lecture 1. A bit about time (theory).

Dr. Alex Hvatov, Composite AI Lab Head

SPb
29.11.2022

Agenda

- What exactly stationarity is?
- Why is non-stationary processes are differ?
- Markov processes
- Impulse processes

Life without non-stationarity

Random processes (RP) types

- Reminder: two stationarity definitions
  - "Strong" – no time at all – stationary as a bus station

If $X_t = (x_1, \dots, x_n, t)$ –random process, F_X – distribution function, then

$\forall\tau \in \mathbb{R}\ \ F_X(x_1, \dots x_n, t){=}F_X(x_1, \dots x_n, t{+}\tau){=}F_X(x_1, \dots x_n)$

  - "Weak" – only with several parameters

$m_X(t) = m_X(t + \tau) = const$ –nonstationarity in ML

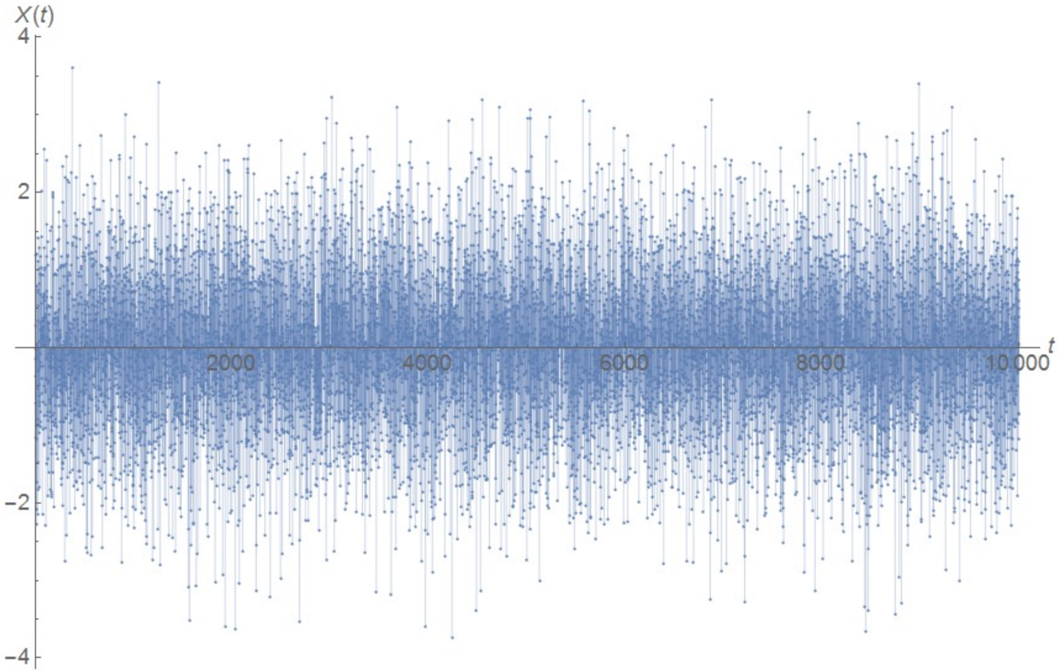
$C_X(t_1, t_2) = C_X(\tau)$

Weak>Strong (there are more weak stationary processes than strong ones)

Reminder: stationary RP example

- White noise –

Can one say, that stationarity roughly is the oscillation near zero (some constant)?



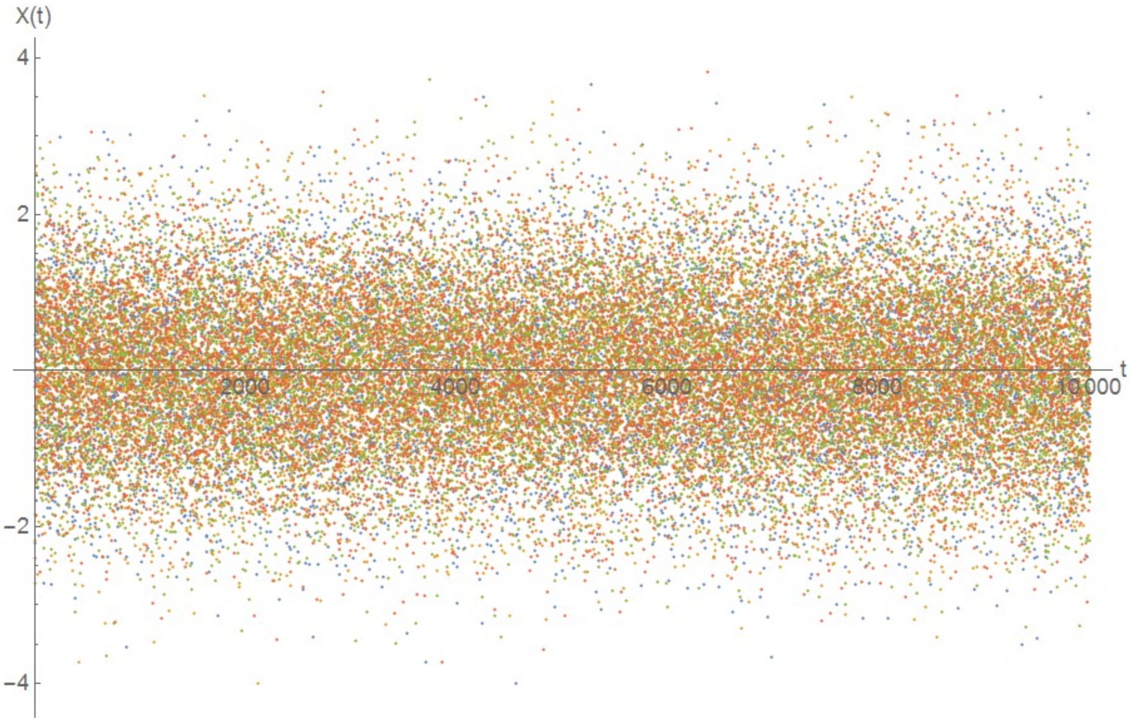
Reminder: What exactly stationarity is?

- White noises –

  (four realizations)

Let’s pretend that we measure something four times in a row and at every time point we get four measures.

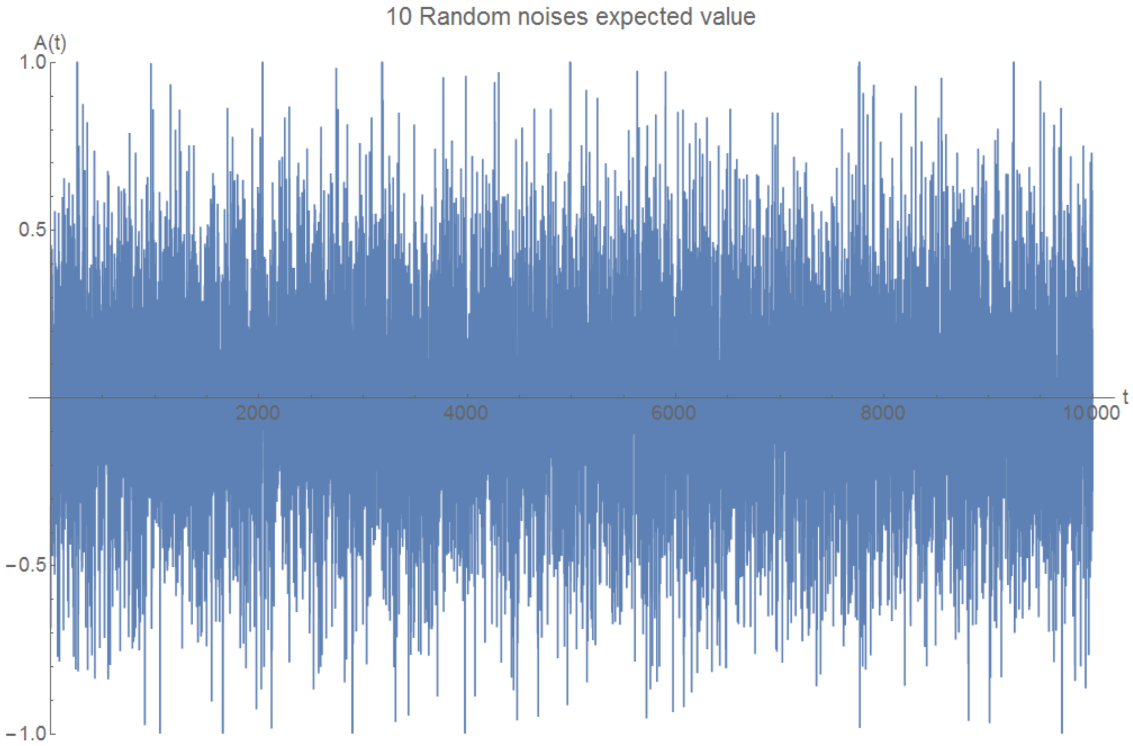
Is it still oscillating near zero? Can you say what ‘near zero’ exactly is?



Reminder: stationary RP example

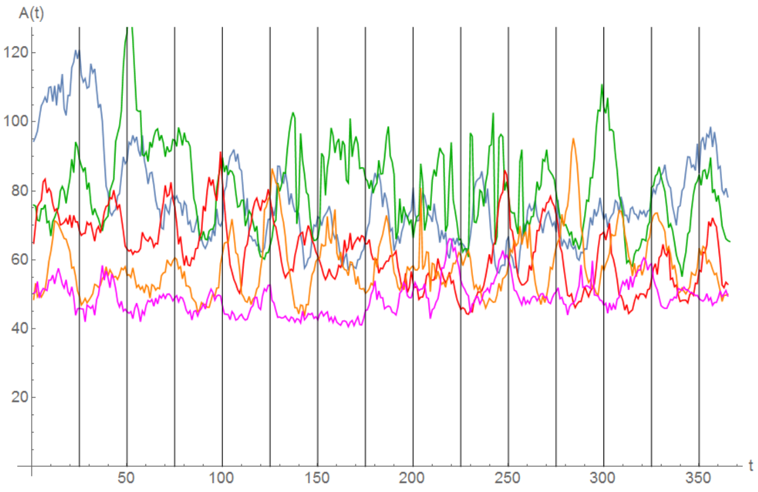
ITMO UNIVERSITY

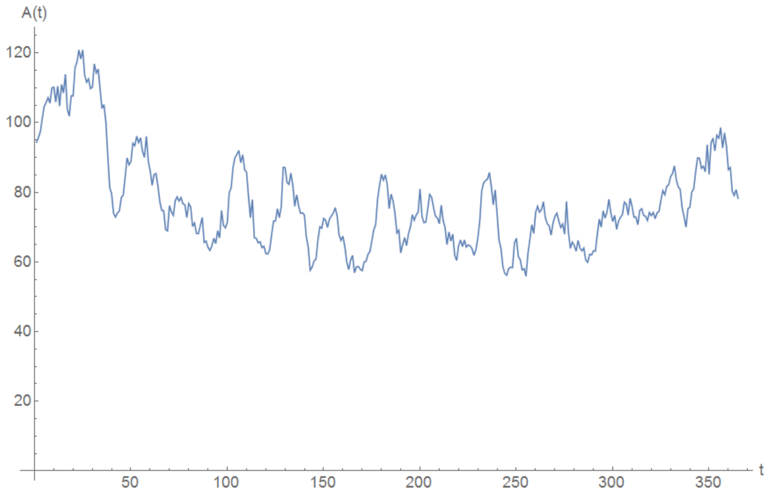
- Usually we use 'weak' non-stationarity and want to prove roughly $m_x(t) = const$
- What if we try to do that directly and make 10,20,…,3000 white noise experiments?
- Where is my $m_x(t) = 0$?
- (Can you measure air temperature for 3000 years?)



ITsMOre than a UNIVERSITY

Two time series approaches





"Stochastic" - to study the distribution of different realizations at the same point in time (suitable if you have a lot of time and diligence)

vs

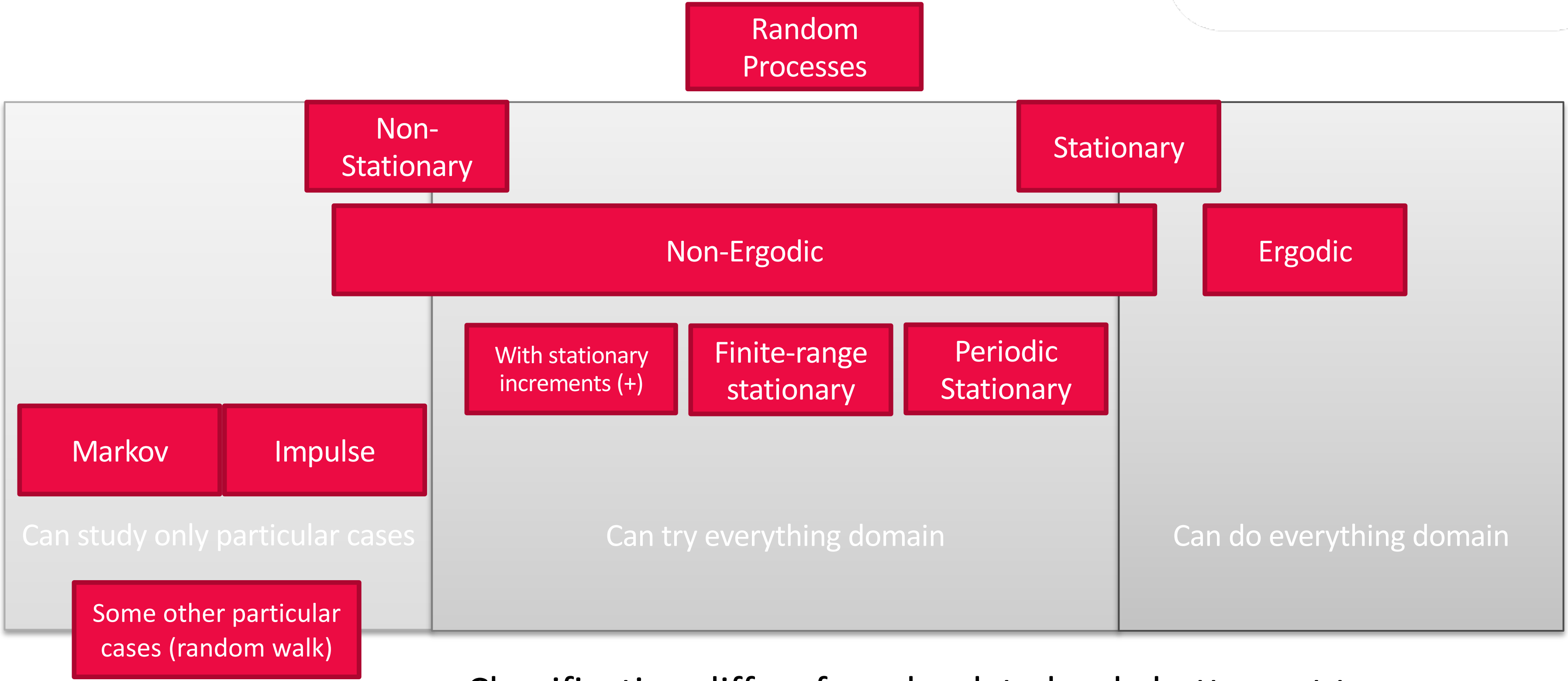
"Functional"- to study each implementation separately (we often have no other choice)

It is desirable to make both approaches give "the same" results

Life without time dependency

Divide et empera – divide and rule

(Time series classification)



- Classification differs from book to book, better not to even try to classify random processes

Why we don't like time-dependence?

- Every process can be written as $y_t = T(t) + S(t) + \varepsilon_t$
  - $T(t)$ is for trend
  - $S(t)$ is for seasonal component (low-frequency component)
  - ε_t is for random (stationary) noise (high-frequency component)
- Main trouble is $T(t)$, there is no solid answer what $T(t)$ actually is
- If $T(t)$ is somehow estimated, then $S(t)$ is "smoothed" $S(t) + \varepsilon_t$
- ε_t seems to be the white noise, there is no more troubles (?)

Stationarity check – I

(Know thy enemy - 101)

- Unit root test
  - Build autoregression model $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots a_p y_{t-p} + \varepsilon_p$
  - Build characteristic polynomial $m^p + a_1 m^{p-1} + \cdots a_{p-1} m + a_p$
  - If roos have property $\forall i \; abs(m_i) < 1$, then process is stationary
  - Otherwise in most cases $\exists i \; abs(m_i) = 1$ (there could be more that one such i)

In the second case, the process is nonstationary (generally speaking, weakly stationary). Moreover, the number of such roots $\#i$ is very important for further analysis.

If $\#i = 1$ then $\Delta y_t = y_{t+1} - y_t$ is stationary, and so on, if $\#i = 2$ then $\Delta\Delta y_t = \Delta^2 y_t = \Delta y_{t+1} - \Delta y_t$ is stationary.

Processes with this property are called integral with order of n where $n = \#i$

Stationarity check – II

(Advanced know thy enemy)

- $\#i$ helps one to determine T(t)
    - If $\#i = 1$ then trend could be chosen as linear
    - If $\#i > 1$ then linear trend is not fully descriptive, one should use polynomial of order at least of #i
- There is no problem to use higher order polynomials
- So, trend $T(t)$ is often considered as a polynomial (linear) part
- What if $abs(m) > 1$?
    - Explosive process
    - In most cases we say, that choice of experiment setup/model is inappropriate

Si fecisti, nega – commit crime and deny it

(Why you should remove trend beforehead)

- Non-constant trend is shifting the expected value
  - It has an effect on correlation function analysis
- Trend $T(t)$ and Seasonal component $S(t)$ have effect on spectrum (it follows from Wiener-Khichin theorem)
  - It has effect on a spectral analysis (shifting)
  - We have to filter spectrum – we lose information

Life without Markov processes

Markov property

- Simplest definition of Markov property can be defined in case of discrete state space $S = \{s_1, \dots, s_n\}$ and discrete time T=N
  - $P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1})$
  - Future depends only on a current moment, not on a past moments
  - No matter how agent has reached current state, probability of next state depends only on a current state
- There is also common definition for a continuous state space and time, but the main idea is the same as above

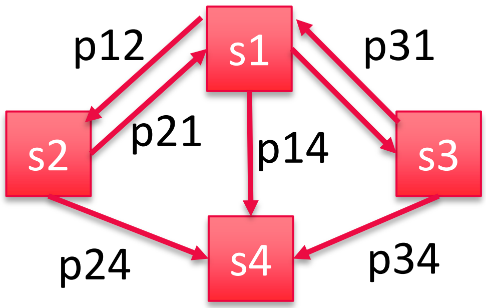
Markov chain - I

- System with discrete state space and discrete time with Markov property is called Markov's chain
- Markov's chain is fully defined by transition matrix $P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$

  and initial state distribution $p(0) = \big(p_1(0), \dots, p_n(0)\big)$

$$\sum_{i=1}^{n} p_{ji} = 1$$

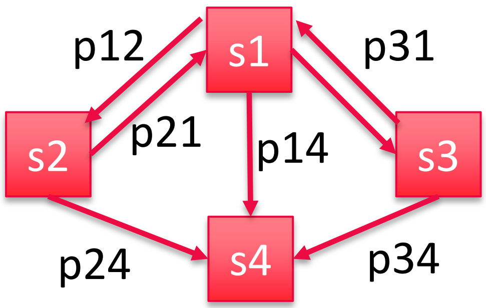
- We can also draw a scheme



Markov chain - II

- Let's introduce some numbers

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.2 & 0 \\ 0.1 & 0.6 & 0 & 0 \\ 0.1 & 0 & 0.5 & 0 \\ 0.1 & 0.2 & 0.3 & 1 \end{pmatrix}$$



- $p(0) = (1,0,0,0)$

$$p(1) = P\left(p(0)\right)^T = (0.7,0.1,0.1,0.1)$$

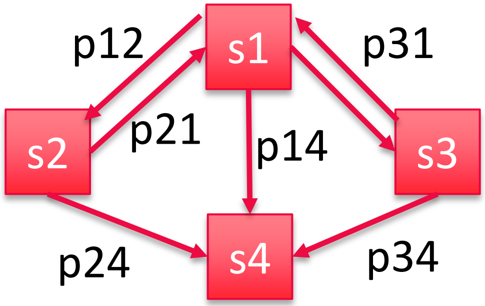
$$p(k) = P^k\left(p(0)\right)^T$$

- From scheme is seen that $p(k) \to (0,0,0,1)$ as $k \to \infty$
- Can we achieve it numerically?

Markov chain - III

ITMO UNIVERSITY

- Can we achieve it numerically?
  - $p_j(k) = \sum_{i=1}^{n} p_i(k-1) p_{ij}(k)$
  - Stationary mode is not dependent on k
  - $p_j = \sum_{i=1}^{n} p_i \, p_{ij}$, with property $\displaystyle\sum_{i=1}^{n} p_{ji} = 1$
  - $p_j \sum_{i=1}^{n} p_{ji} = \sum_{i=1}^{n} p_i \, p_{ij}$
  - $p_j p_{jj} + \sum_{i=1}^{n} p_{ji} p_j = p_j p_{jj} + \sum_{i=1}^{n} p_i \, p_{ij}, i \neq j$
  - $\sum_{i=1}^{n} p_{ji} p_j = \sum_{i=1}^{n} p_i \, p_{ij}, i \neq j$ – if this equation has the solution, then stationary mode exists
  - Actually, solution of equation above is not unique, one should add norm equation $\sum_{i=1}^{n} p_i = 1$



ITsMOre than a UNIVERSITY

Continuous time - I

- Transition matrix is now function of time

$$P(t) = \begin{pmatrix} \lambda_{11}(t) & \dots & \lambda_{1n}(t) \\ \vdots & \ddots & \vdots \\ \lambda_{n1}(t) & \dots & \lambda_{nn}(t) \end{pmatrix}$$

- Without proof, there are Kolmogoroff equations for stationary mode (now it is the system of ordinary differential equations,ODE)

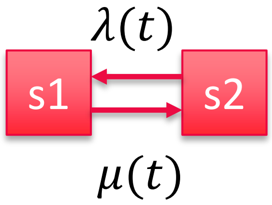
$$\frac{dp_i(t)}{dt} = \sum_{j=1}^{n} p_j(t)\lambda_{ji}(t) - p_i(t)\sum_{j=1}^{n} \lambda_{ij}(t) \,, i = 1, n$$

- Initial conditions have the form $p_i(0) = p_i \,, i = 1, n$
- In this case we need norm equation$\sum_{i=1}^{n} p_i(t) = 1 \,, t \geq 0$

Continuous time - II

ITMO UNIVERSITY

- Simple example

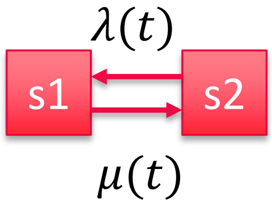


- $$p_1'(t) = p_2(t)\mu(t) - p_1(t)\lambda(t)$$
  $$p_2'(t) = p_1(t)\lambda(t) - p_2(t)\mu(t)$$
  $$p_1(t) + p_2(t) = 1$$
- With $p_2(t) = 1 - p_1(t)$ we obtain one differential equation
  $$p_1'(t) + \big(\lambda(t) + \mu(t)\big)p_1(t) = \mu(t)$$
- Let simplify it a bit more $\lambda(t) = \lambda, \mu(t) = \mu$
- $p_1'(t) + (\lambda + \mu)p_1(t) = \mu$

ITMOre than a UNIVERSITY

Continuous time - III

- $p_1'(t) + (\lambda + \mu)p_1(t) = \mu$
- $p_1'(t) = \mu - (\lambda + \mu)p_1(t)$
- $\frac{1}{\lambda+\mu}\frac{d(\lambda+\mu)p_1(t)-\mu}{(\lambda+\mu)p_1(t)-\mu} = dt$
- $-\frac{1}{\lambda+\mu}\log(\mu - (\lambda + \mu)p_1(t)) = t$
- $p(t) = \frac{\mu}{\lambda+\mu} + \exp\big(-t(\lambda + \mu)\big)C$
- Let $p(0) = 1$, then $p(t) = \frac{\mu}{\lambda+\mu} + \exp\big(-t(\lambda + \mu)\big)\frac{\lambda}{\lambda+\mu}$
- Equation has the stationary mode $p = \frac{\mu}{\lambda+\mu}, t \to \infty$

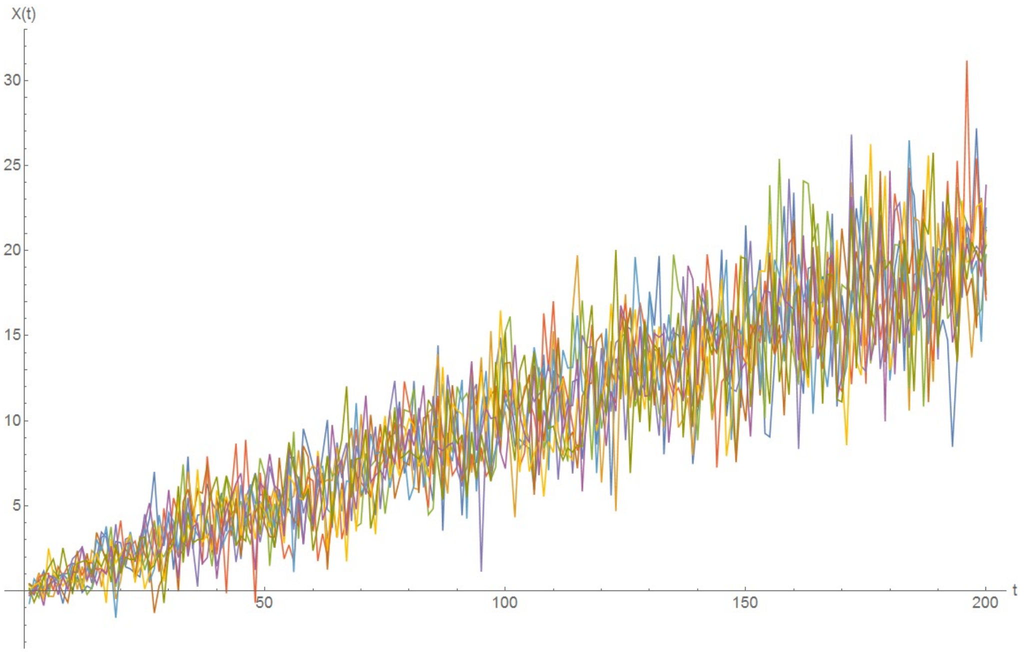


General Markov process

- General Markov process (with continuous time and state space) is defined by differential equation with stochastic right part $\xi(\boldsymbol{X}, t)$
$$\boldsymbol{X}'(t) = F(\boldsymbol{X}, t) + \xi(\boldsymbol{X}, t)$$
- In this case we have to integrate stochastic function $\xi(\boldsymbol{X}, t)$, which is of course not so easy process
- There are two definitions of stochastic integrals: Ito and Stratonovich integrals

General Markov process

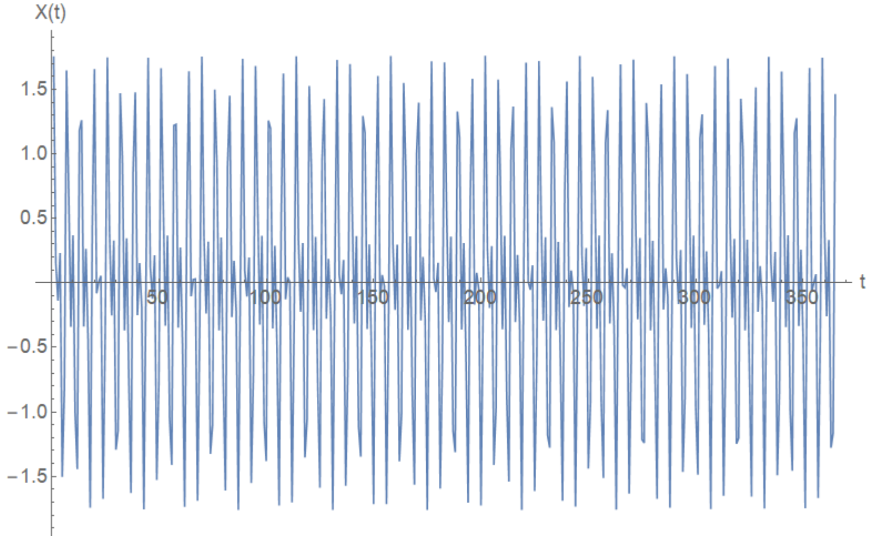
- Main idea is to consider motion as $\mathrm{X}(t) = x_0 + \mu t + \sigma\sqrt{t}\varepsilon$
- Then $dx \sim \mu dt + \sigma\varepsilon\sqrt{dt}$ - Wiener equation
- We are not able to integrate it directly

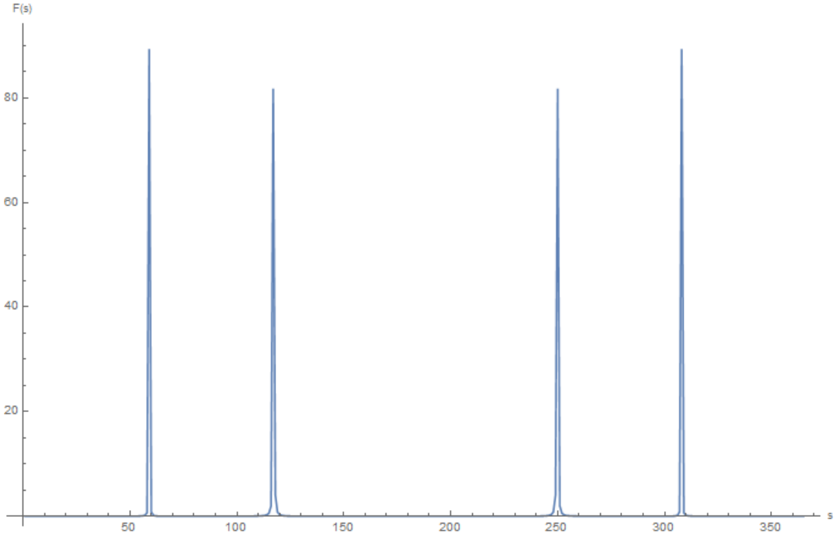


A lot of random walk

Life without wavelets

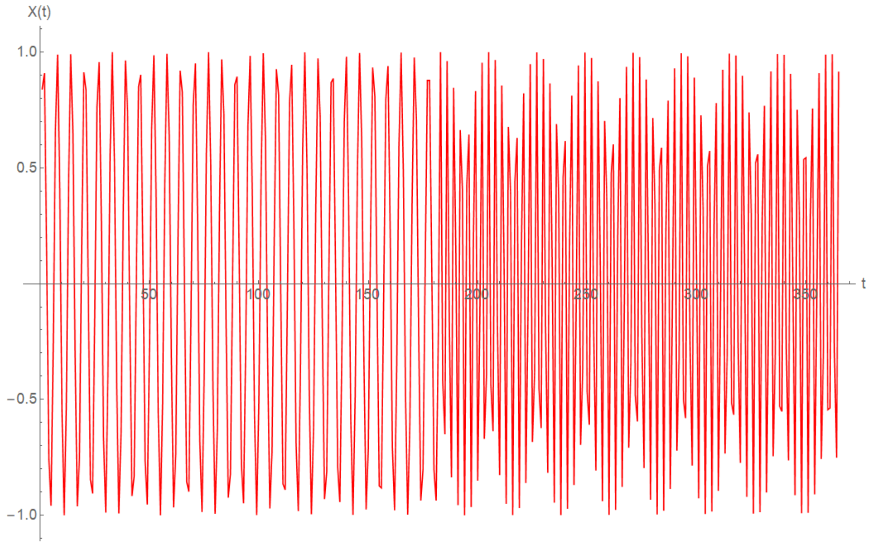
Impulse processes - I

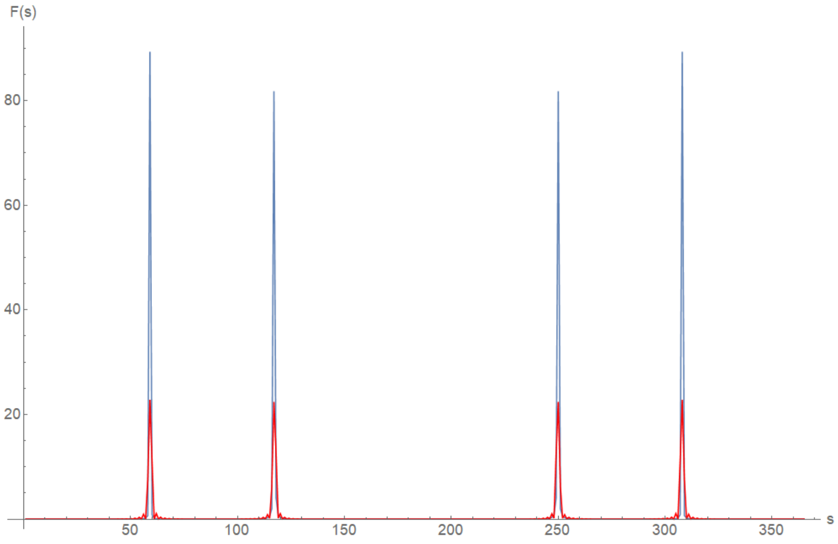




- Plot $\sin(t) + \sin(2t)$
- Fourier transform

Impulse processes - II





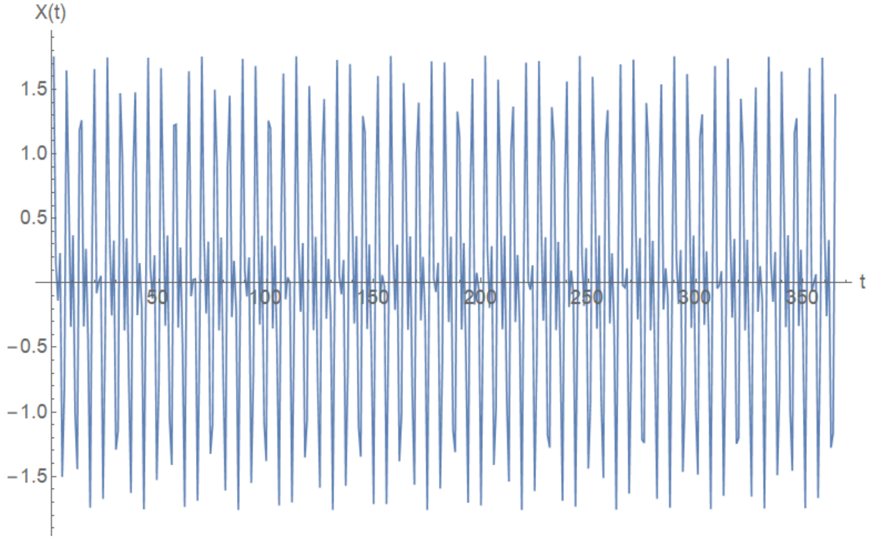
- Plot half $\sin(t)$,half $\sin(2t)$
- Fourier transform

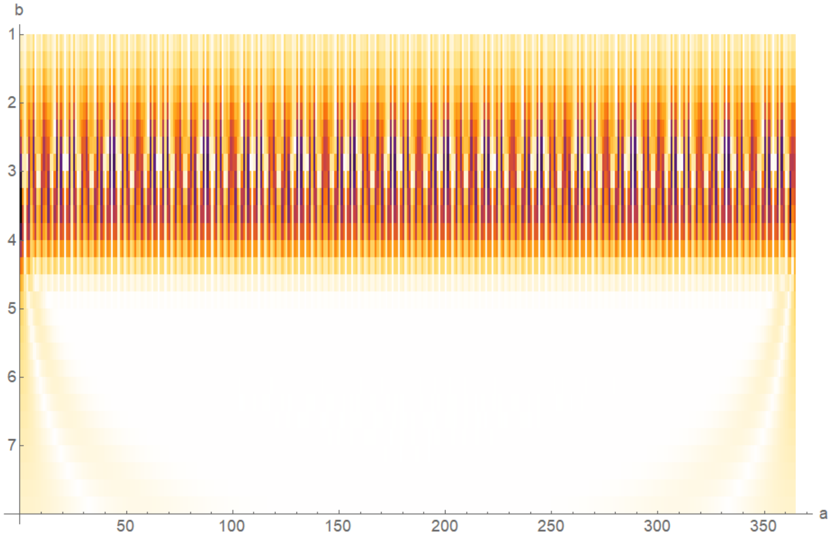
- They are not so different in Fourier specter

Wavelet transform

- $W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^{*}\left(\frac{t-b}{a}\right) dt$ - wavelet transform
  a- scale parameter (frequency analogue)
  b- shift parameter
- $\psi(t)$ - wavelet
  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ - main wavelet property
  $\int_{-\infty}^{\infty} \psi(t) t^m dt = 0$ - optional wavelet property
- Wavelet ignores linear (m degree polynomial, if optional property is fulfilled) trend
  - Still better to remove it

Scalogram-I

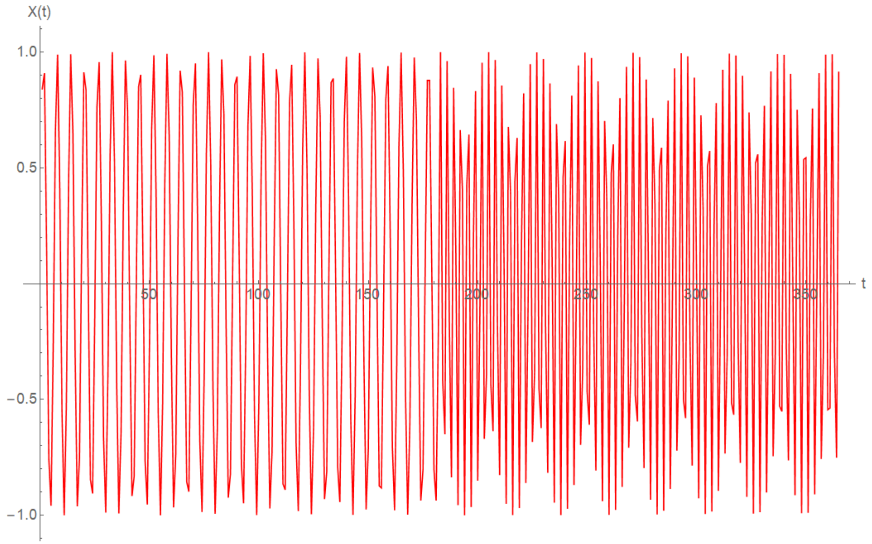


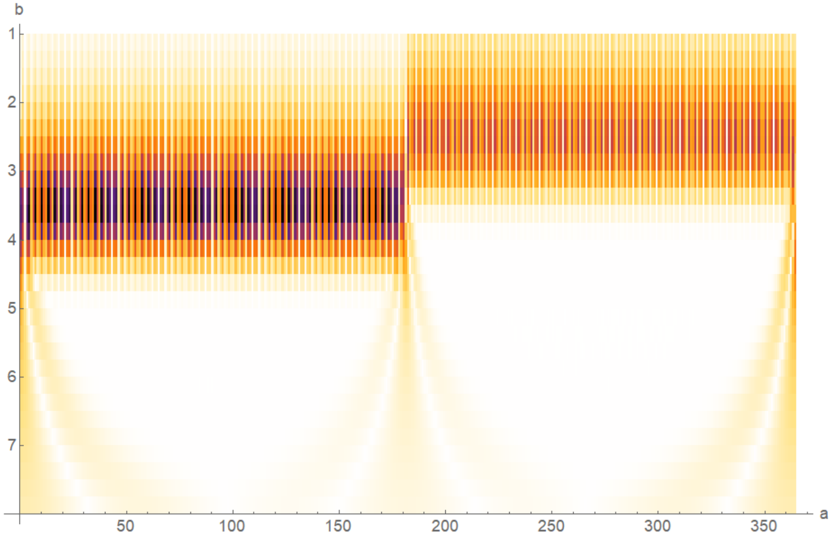


- Plot $\sin(t) + \sin(2t)$

- Wavelet transform

Scalogram-II





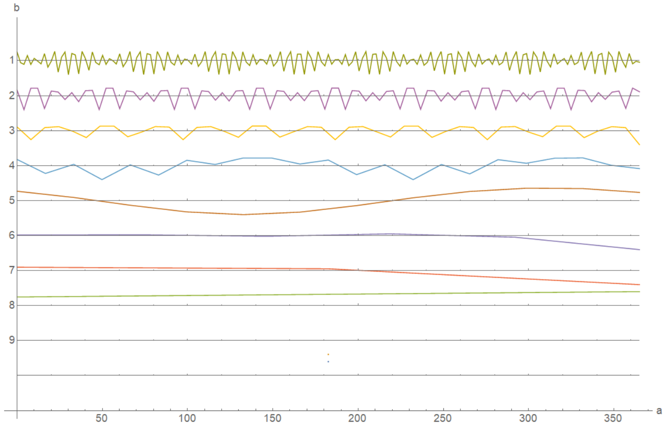
- Plot half $\sin(t)$,half $\sin(2t)$

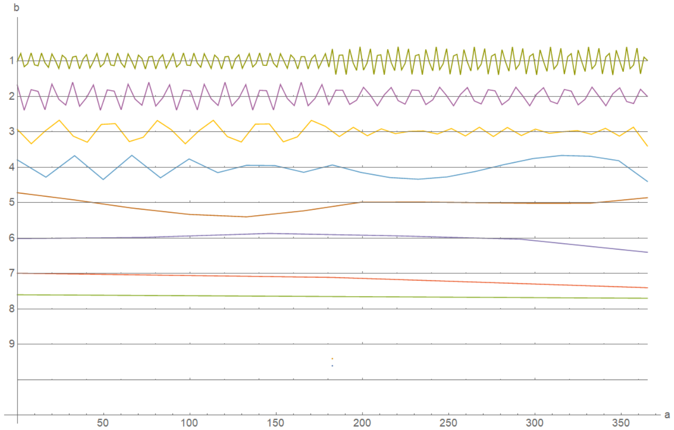
- Wevalet transform

- They are different in wavelet specter

Wavelet transform

- Wavelet analysis is used in signal processing
- There are a lot of different wavelets with many applications
- Scalogram is difficult to interpret and is kind of art





- Discrete wavelet transform

Thanks for coming!

www.ifmo.ru