Stochastic methods in Mathematical Modelling

Prof. Vladimir V. Palyulin, V.Palyulin@skoltech.ru

TA Maria Larchenko, Maria.Larchenko@skoltech.ru

Lectures and recitations

Monday 16:00pm – 19:00pm

Tuesday 16:00pm – 19:00pm

Friday 16:00pm –19:00pm

Course details

Assessment

Homework – 40 %

Written Exam – 20%

Project – 40 %

Telegram group of the course:
https://t.me/+DXUPXQ9ol-RlNjMy

Videos/presentations

All sessions will be recorded and the links uploaded to canvas. The presentations will be uploaded to canvas as well

Course details

- **Homework policy**

4 HW assignments in total, each 10 points, i.e. 1 point for the homework makes 1 percent of the total grade.

The deadline for the first homework to be submitted through the canvas is 9th of November EOD

Topics

- Random variables, distributions
- Laws of large numbers, extreme values
- Information-theoretic view on randomness
- Markov chains. Qeueuing theory
- Stochastic differential equations
- Anomalous stochastic processes
- Monte-Carlo algorithms (direct and MCMC)
- Markov decision processes
  and reinforcement learning

Literature will be provided in every section separately

Projects

Consist of a careful paper reading with reproduction of some of the results, the defence and writing a report

Single person assignment!! 40 points.

Examples of topics (alternative course-related papers could also be selected)

- **Sequential Monte Carlo for Importance Sampling & Inference**
- **Aging and ergodicity breaking in subdiffusion (TAMSD vs EAMSD)**
- **Resonant activation over a fluctuating barrier**
- **Jackson Networks of Queues**
- **Simulated Annealing Sampling**
- **The Noisy Channel Coding (Shannon) Theorem**

The full list of suggestions is expected to appear in the middle of November

Tentative project defence dates is 19^{th}-20^{th} of December with a report submission deadline on 23^{rd}

Exam.
9^{th} of December. 16:00-19:00

On the 2^{nd} of December (Friday) the mock exam paper will be released

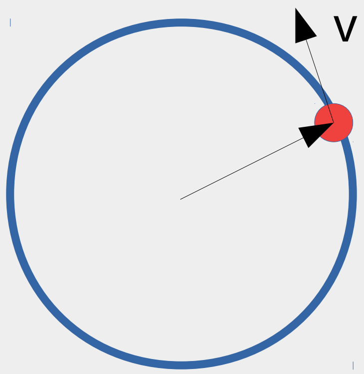
At 16:00 on the 9^{th} of Dec the exam problems will appear on Canvas and a couple of minutes later in the Telegram group. After that you will have 3 hours to submit the pictures of the solutions to Canvas. The exam brings another 20 points.

Randomness

Lack of predictability of events

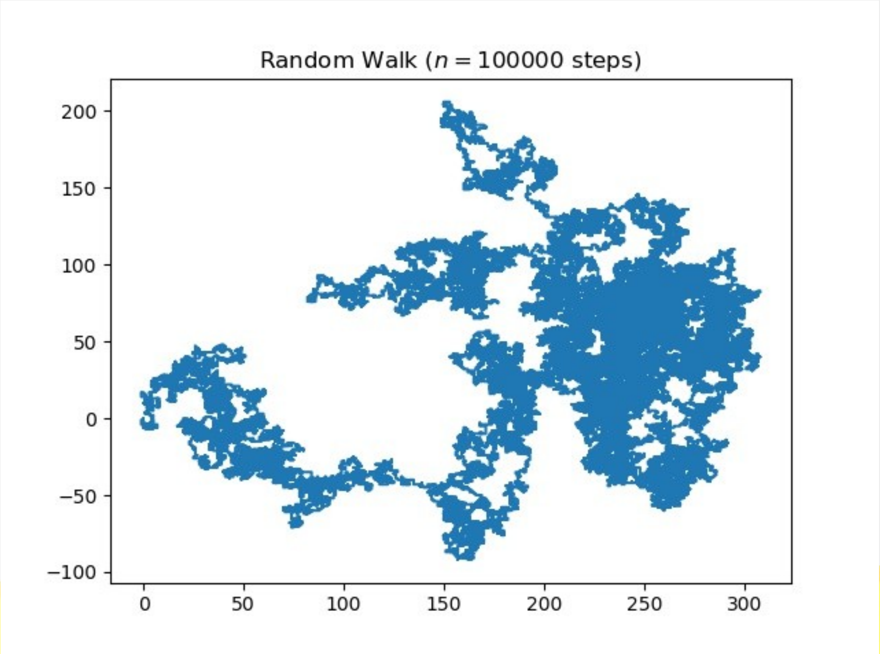
Deterministic

$$\frac{d\vec{v}}{dt} = \frac{\vec{F}}{m}$$



Random

$$\frac{d\vec{v}}{dt} = \frac{\overrightarrow{\xi(t)}}{m}$$



Randomness. History intro

Antiquity. The prevalent view: Outcomes of games of chance is a fate
Aristotle: Concept of unknowable events that happen by pure chance
Early christianity: Unknowable to man was considered to be predetermined by God

8th-13th century – Arab mathematicians (Al-Khalil, Al-Kindi, Ibn Adlan)
Forms of statistics, permutations, combinations, first code-breaking algorithms, sample size for frequency analysis

Mid-XVIth century – Gerolamo Cardano. Probabilities of dice throws

XVIIth century – Blaise Pascal and Pierre Fermat. Theory of probability

XVIIIth century – Jacob Bernoulli. Law of large numbers. T. Bayes: Bayes' theorem

XIXth century – 1801 Gauss predicts the orbit of Ceres using a line of best fit,
Early 1800s – normal distribution derived by Gauss to describe the pattern of random measurement error in observational data
1805 – Adrien-Marie Legendre introduces the method of least squares
1865 – concept of entropy introduced by Rudolf Clausius
1866 – John Venn publishes Logic of Chance
1881 he publishes Symbolic Logic with Venn diagrams
ca. 1880 - Peirce outlines frequentist statistics, the use of objective randomization in experiments and in sampling

Randomness. History intro

XIXth century – 1888 – Galton introduces the concept of correlation
1900 - Bachelier analyses stock price movements as a stochastic process

XXth century

1905 – Karl Pearson, Lord Rayleigh introduce the concept of random walk
1906 – Andrey Markov begins his studies of Markov chains
1925 – stable distributions by Paul Lévy
1928 – Tippett and Fisher develop extreme value theory
1933 – A.N. Kolmogorov publishes his book “Basic notions of the calculus of probability” (Grundbegriffe der Wahrscheinlichkeitsrechnung)
1948 – Shannon's Mathematical Theory of Communication defines capacity of communication channels in terms of probabilities,
1953 – Nicholas Metropolis introduces the idea of thermodynamic simulated annealing methods, MCMC
1970s – Non-Gaussian stochastic processes
1987 – hybrid (Hamiltonian) Monte Carlo
Starting from 1980s – randomised algorithms

XXIth century

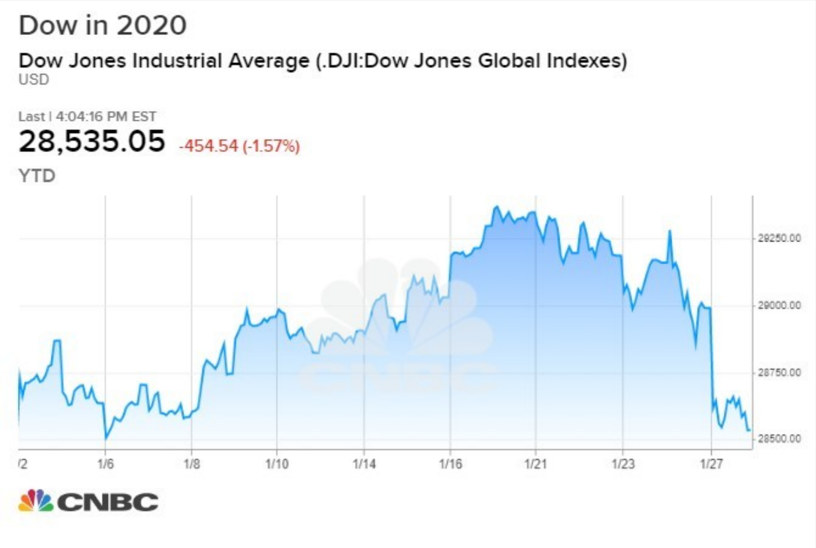
Multiple MC-based sampling methods, molecular simulations, further advances in MDPs, ML methods
Hyperuniformity
Etc. etc.

Stochasticity. Examples
Time series

Games of chance

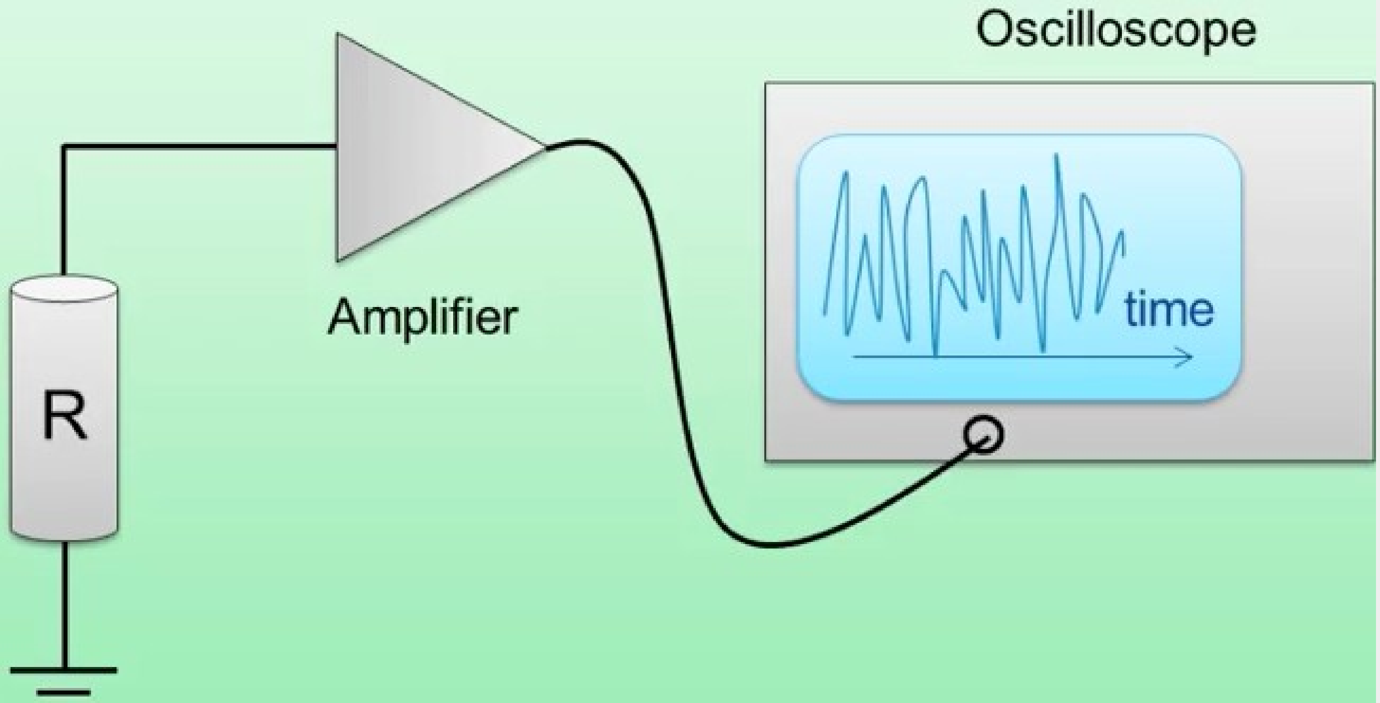
Stochasticity. Examples

Economic time series (Stock markets. Finance)



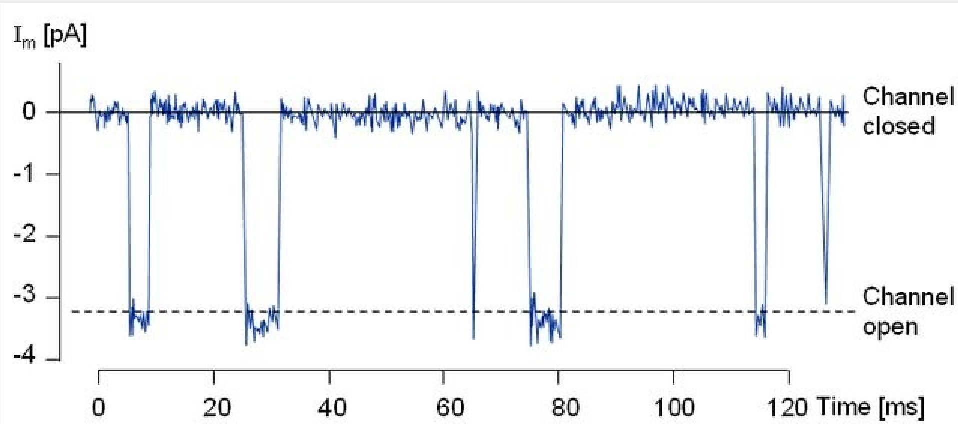
Stochasticity. Examples

Thermal noise in electric circuits (Johnson–Nyquist noise)

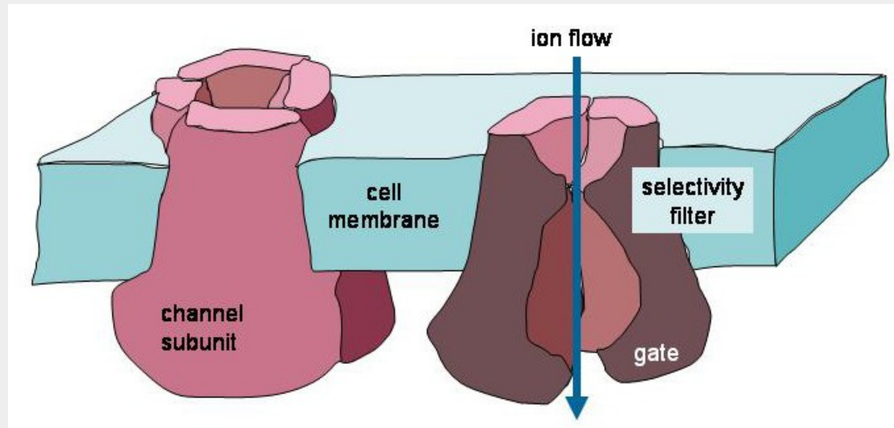


Stochasticity. Examples

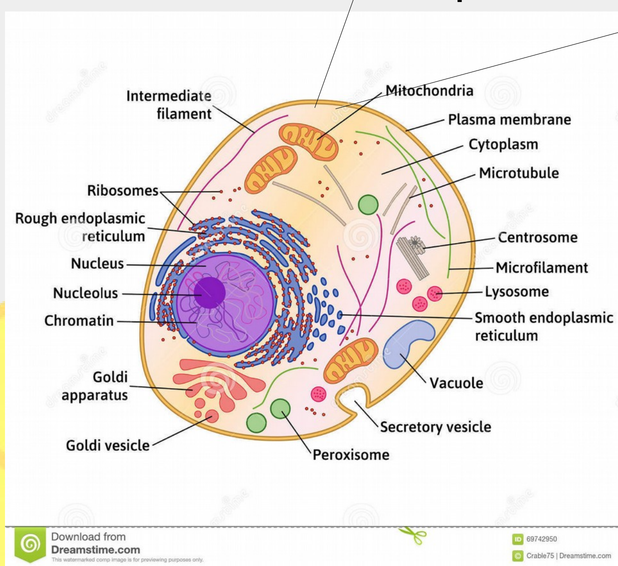
Ionic channel current fluctuations



Registration of the flow of current through a single ion channel at the neuromuscular endplate of frog muscle fiber with patch clamp method.
(From Sakmann and Neher, 1984.)

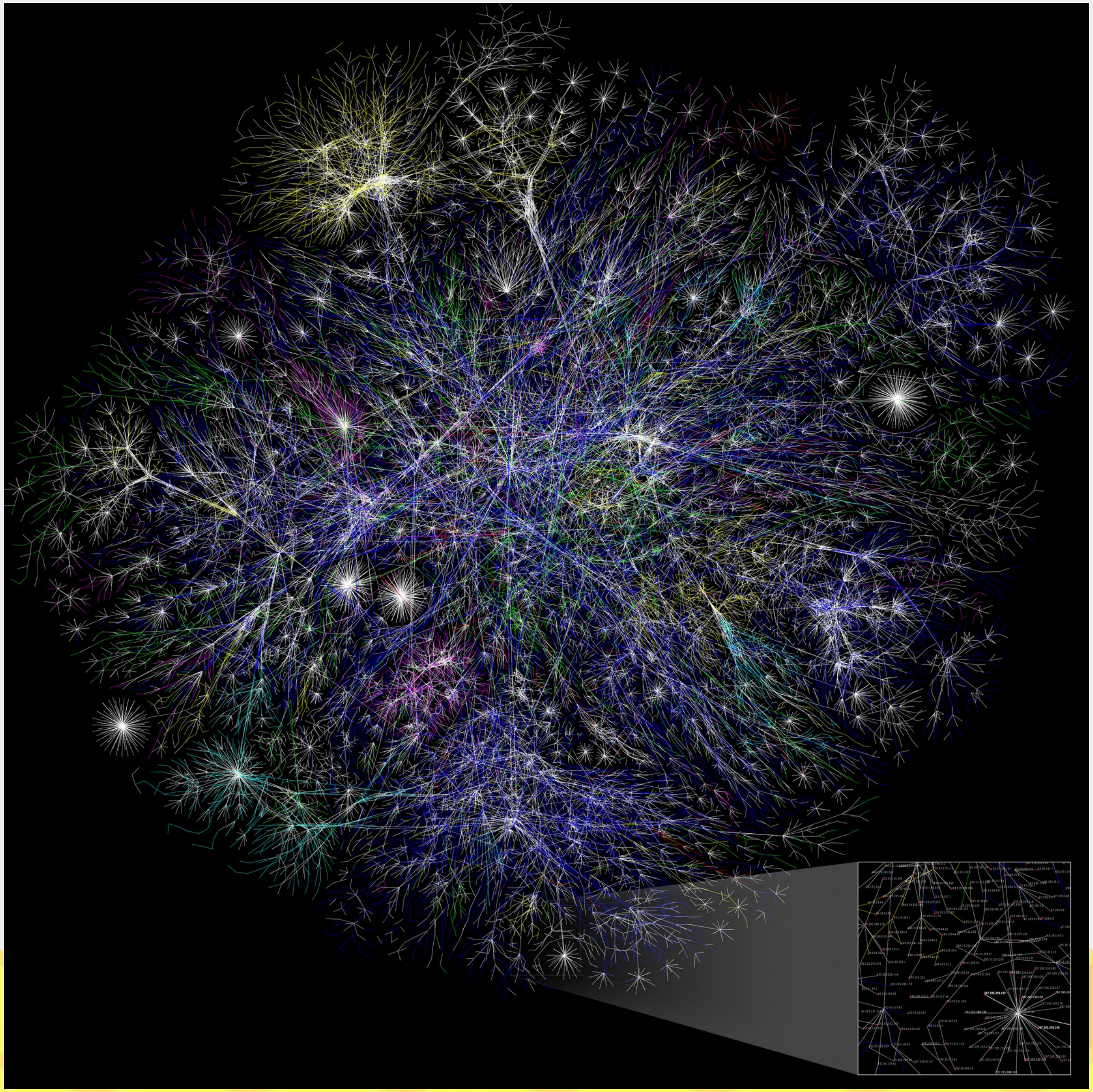


Taken from
http://www.scholarpedia.org/article/Ion_channels



Stochasticity

Random networks. Internet



Picture from https://en.wikipedia.org/wiki/Hub_(network_science)

Skoltech
Skolkovo Institute of Science and Technology

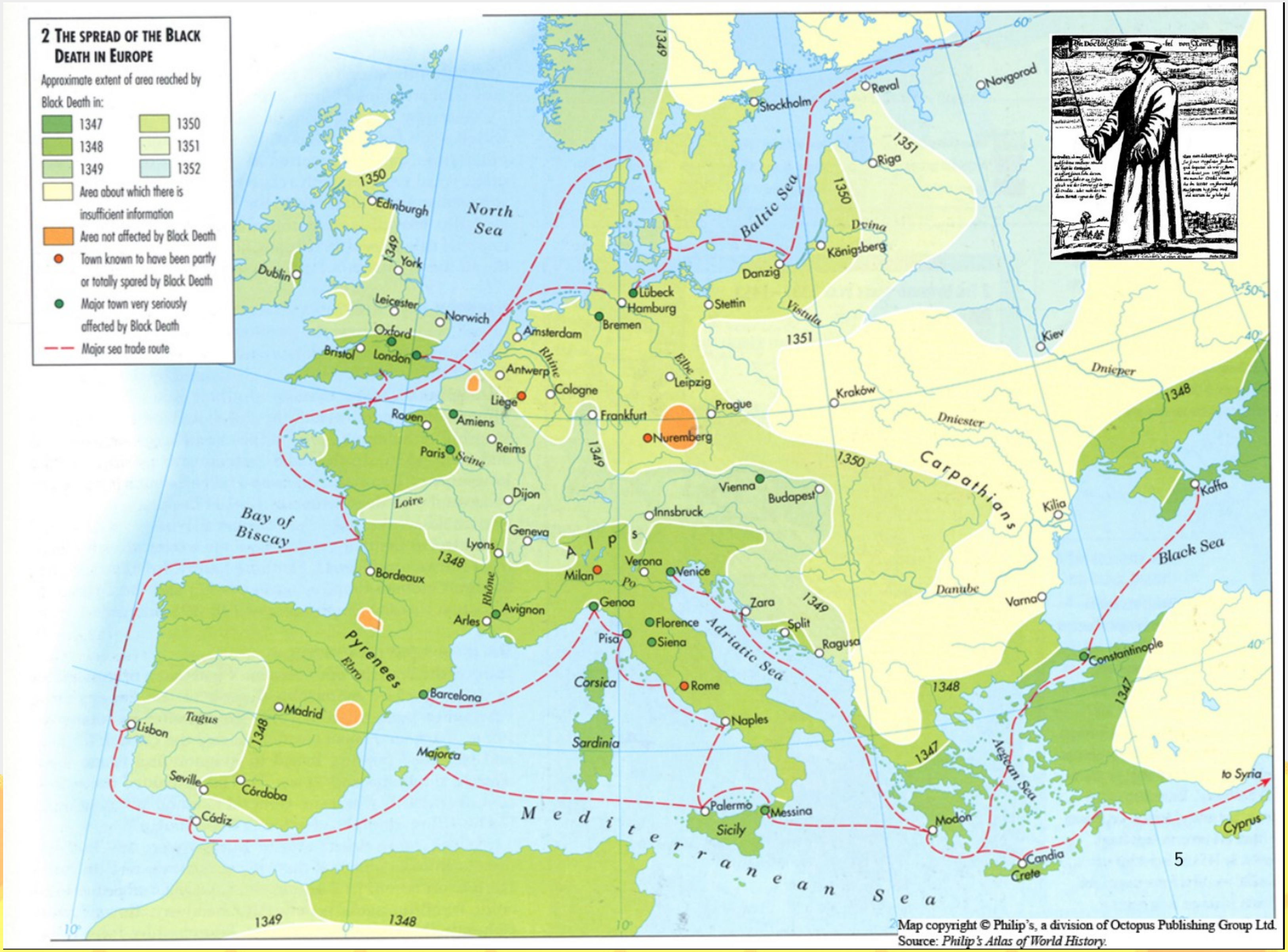
Stochasticity

Random networks. Flight routes across the world

Skoltech
Skolkovo Institute of Science and Technology

Stochasticity. Trajectories (time+space)

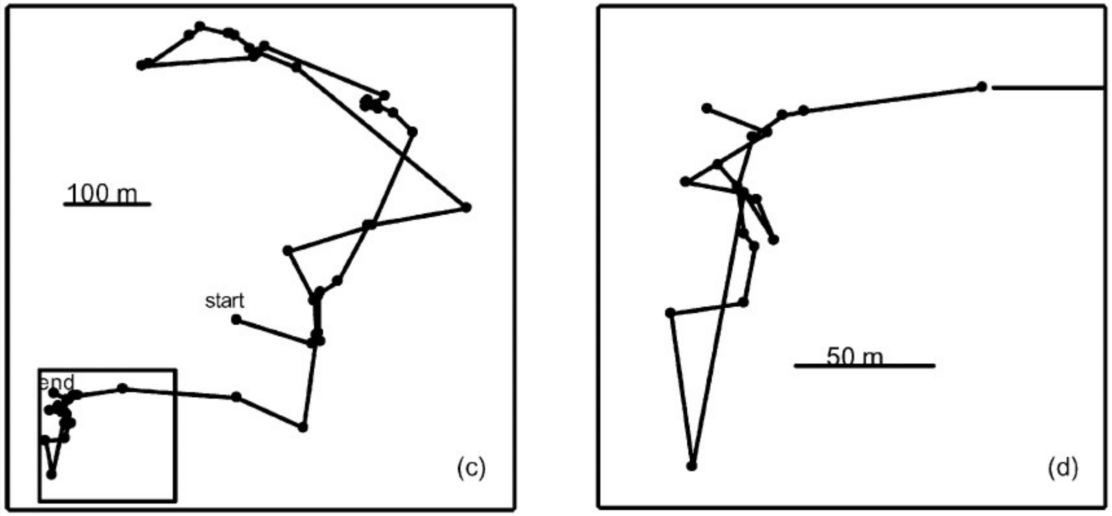
Black death epidemics of 14th century

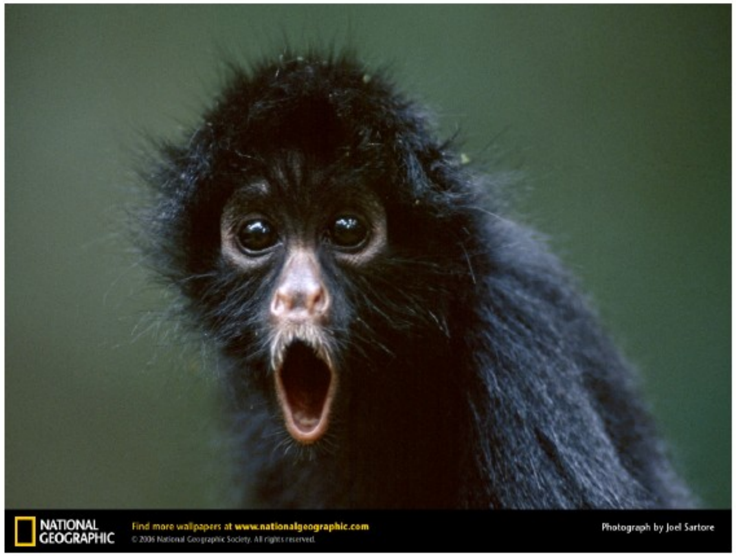


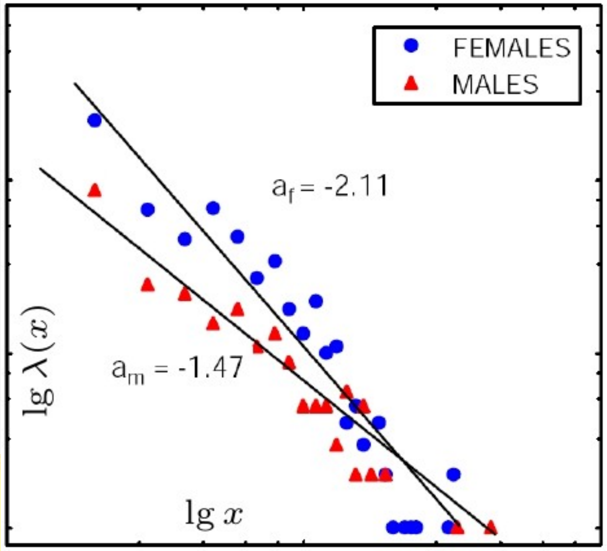
Skolkovo Institute of Science and Technology

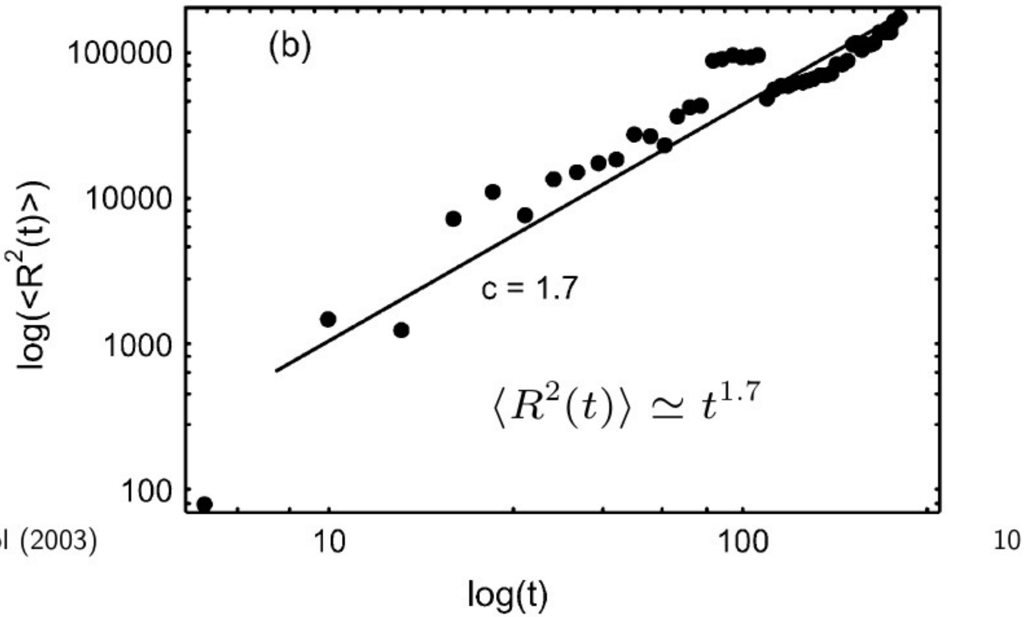
Stochasticity. Trajectories (time+space)

Jumps of spider monkeys





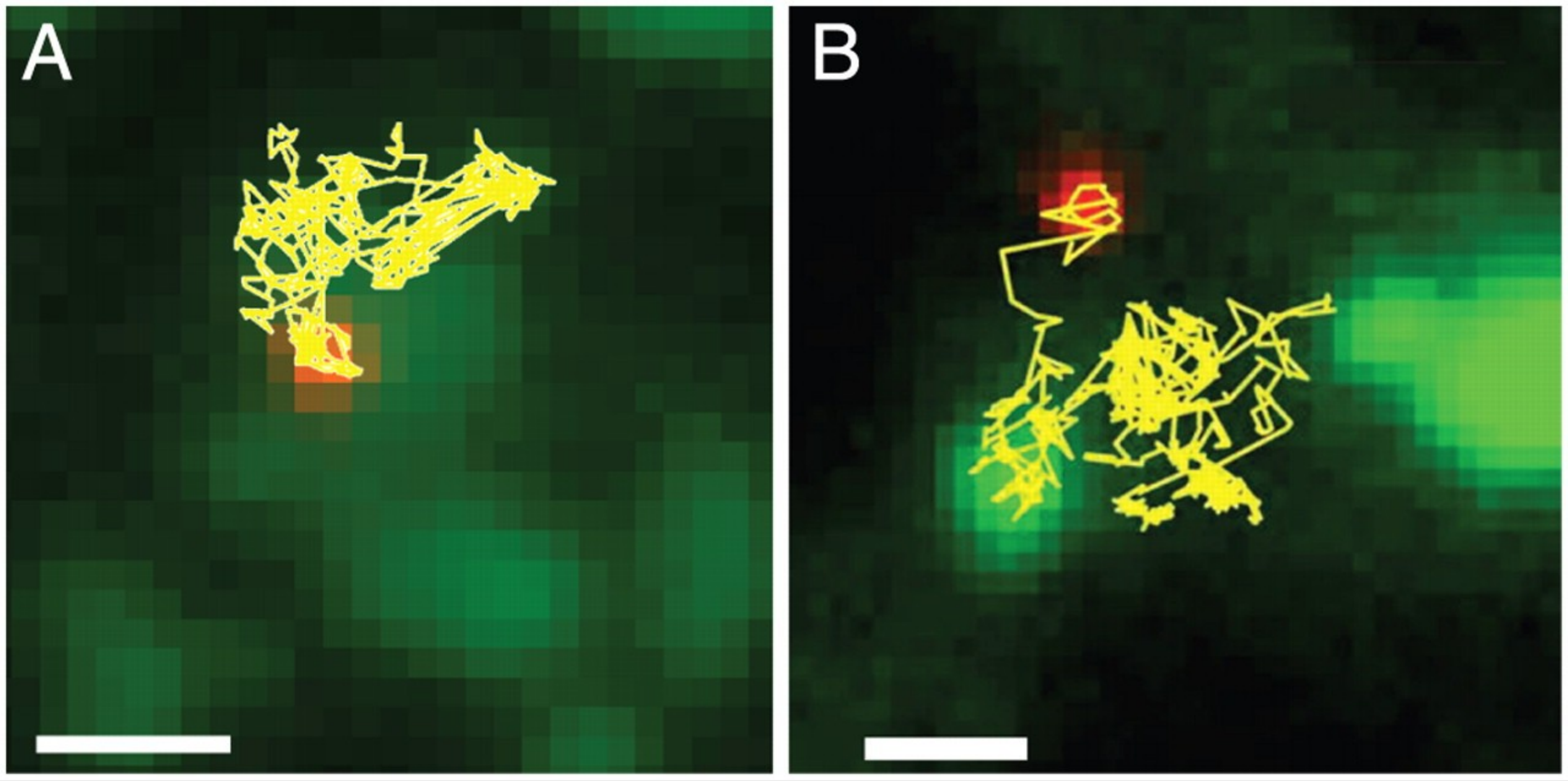




Ramos-Fernandez et al, Behav Ecol Sociobiol (2003)

Stochasticity. Trajectories (time+space)

Single particle tracking in human cells



Two-color TIRF images of human embryonic kidney (HEK) cells expressing GFP-Kv2.1 channels, labeled with QDs. A.V. Weigel, B. Simon, M.M. Tamkun, D. Krapf, Proceedings of the National Academy of Sciences, 2011, 108 (16) 6438-6443;

Why does lack of understanding of randomness lead to problems?

Gambler's fallacy& hot hand phenomenon

GF: Deceptive biases which make a person to anticipate the patterns which do not really exist

HH: Belief that a person who experiences a success has a greater chance of further successes

Gambler's fallacy&
hot hand phenomenon. Examples

1) ***53 fever*** (Venice lottery wheel) 2003-2005.
53 did not come up in almost 2 years.
Est. 4 bln Euro bets lost

2) ***Monte-Carlo fallacy. 1913***. One of the roulettes had 26 blacks in a row
According to some data still affected the gambling up till recently

3) Asylum ***Judges***, ***Loan Officers***, and ***Baseball Umpires*** make consistently negatively correlated decisions (Chen, Moskowitz, Shue, Q. J. Econ., 2016)

https://www.bbc.com/worklife/article/20200217-the-simple-maths-error-that-can-lead-to-bankruptcy

More subtle relevant examples: wrong MC results

Monte Carlo simulations are based on random sampling

If the sampling is incorrect the results may be wrong

Reasons: hidden correlations in the random number generation

These correlations could manifest themselves in some algorithms/modelling but work perfectly well in the other cases

If there are doubts one could test the correlations
(I. Vattulainen, T. Ala-Nissila, and K. Kankaala Phys. Rev. Lett. 73, 2513)

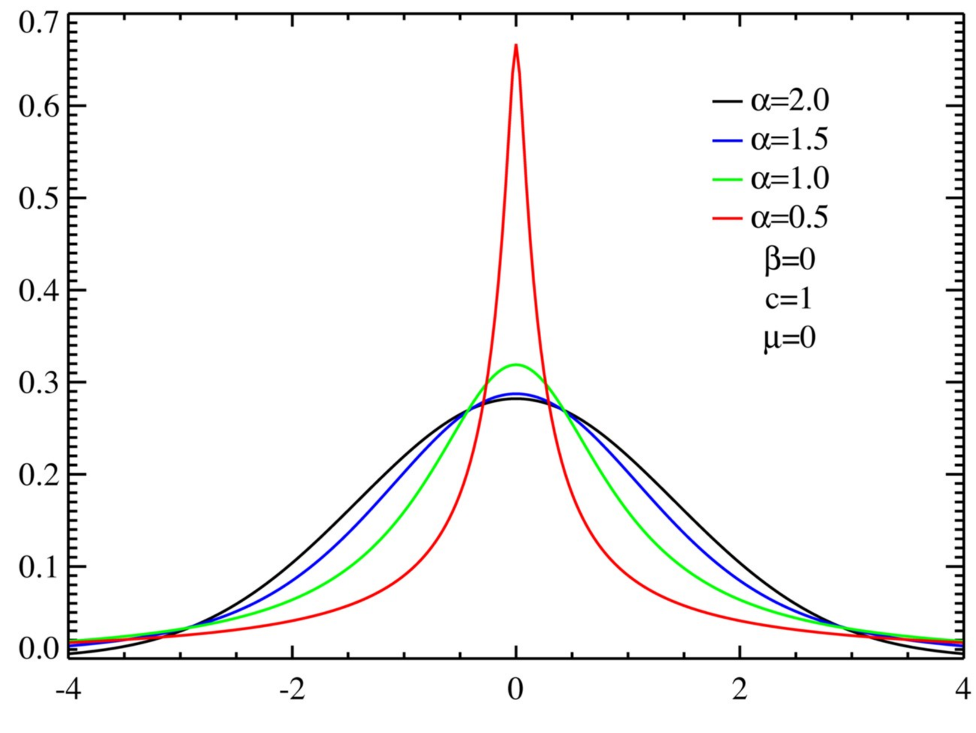
Conclusion
The proper sampling is very important for drawing statistical inferences

More subtle relevant examples: Gaussian vs heavy-tails

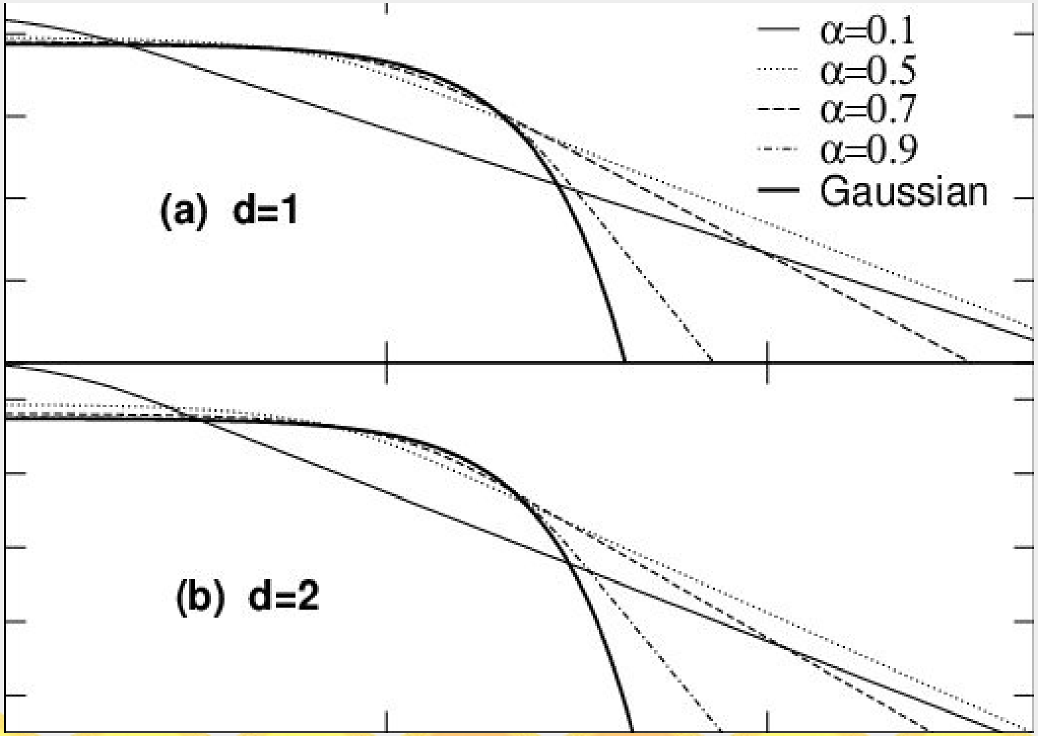
Extreme value analysis

A field of statistics analysing extreme deviation from the median of probability distributions

Alpha-stable distributions



Heavy tails vs Gaussian



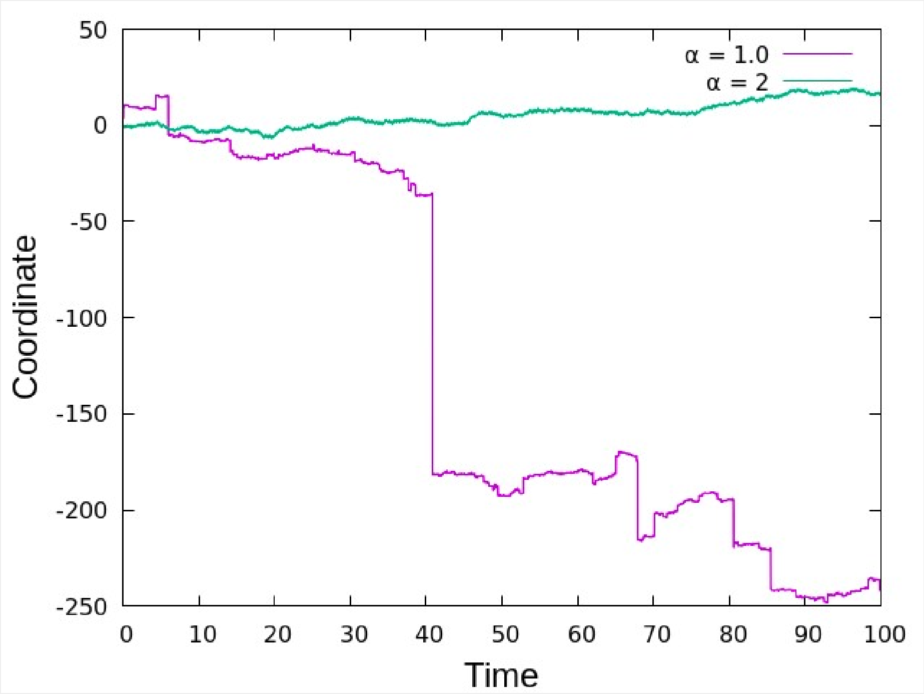
Trizac, Barrat, Ernst, Physical Review E 76, 031305

DOI: 10.1103/PhysRevE.76.031305

Skolkovo Institute of Science and Technology

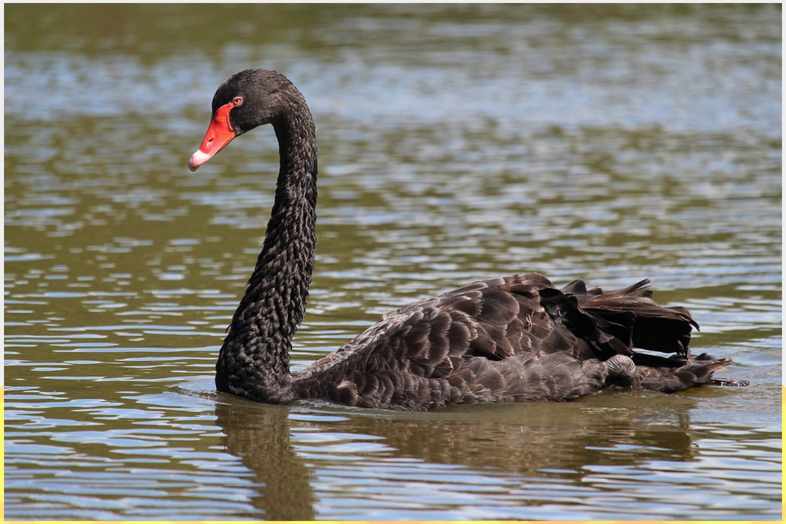
More subtle relevant examples: Gaussian vs heavy-tails

Example: Lévy flights



Extreme value statistics
Applications

- Equity risks. Day to day market risk
- Extreme weather events
- Performances in a sport (what would be the next 100 m sprint record?)
- Mutational events during evolution
- Side effects of drugs
  etc. etc.



Course goals

- Understanding stochasticity in life, science and technology
- Looking at randomness from Gaussian and a non-Gaussian perspective
- Learning how to deal with stochasticity in applications

Generation of random numbers

Getting “good enough” random numbers is essential for simulations (MC and other stochastics) and cryptography. However, the task is not as easy as it seems.

Types of random numbers

- ***Truly random***: impossible to predict what the next random number is
- ***Pseudorandom***: a sequence of algorithmically crafted numbers. Not truly random. The sequence can be fully repeated if the initial conditions and the algorithm are known
- ***Quasirandom***: act as random numbers in some kinds of simulations, but are well-ordered in other cases

Generation methods

Non-automated means

Almost never used these days. Still it is good to be aware of keep in mind that there do exist non-automated means to generate random numbers

Historically used example: To select numbers from some number sequence, e.g. the phone book or the decimals of π.

• The first example is not advisable, as there can be strong non-random features in phone numbers
• The second is not so bad, since the decimals of π are not supposed to have correlations

One could for instance get rough random numbers between 0 and 1 by selecting always 4 numbers at a time from π and divide these by 10000:

3.141592653589793238462643383279502884197169399375105

These non-automated random numbers are essentially pseudorandom

Generation methods

Mechanically generated random numbers

Random numbers produced by mechanical devices (obsolete)

- The machines used in lottery or roulette
- Specifically designed machines capable of generation of thousands of numbers. In 1955 RAND corporation published a book with 1 million random numbers

Mechanical random numbers can be both true random (lotto machine or a roulette) numbers, or pseudorandom.

Generation methods

Algorithmicaly-generated random numbers

Without doubt the most frequently used and practical approach for the random number generation

An algorithm produces a sequence of seemingly uncorrelated sequence of numbers. However, the sequence will be always the same as long as the recipe and the so-called seed (normally a single integer) are the same. That is why there are called *pseudorandom*.

Any pseudorandom number generator makes a sequence which eventually *repeats* itself. The period of the repeats is a characteristic of generator's quality.

Obviously, the period should be larger than the quantity of numbers used in a simulation

Useful feature: Since the sequence can be reproduced exactly. So can the simulation. This repeatability is important for verification and reproduction of results!

Skoltech
Skolkovo Institute of Science and Technology

Generation methods

Algorithmicaly-generated random numbers

Possible artefacts

- A period being too short
- Lack of uniformity of distribution
- Correlation of successive values
- Different periods for different seed states

i.e. for every random number generator the tests should be conducted to check for the above mentioned issues

Rukhin A. et al. A statistical test suite for random and pseudorandom number generators for cryptographic applications. NIST, Special Publication 800–822, (2010).

Skoltech
Skolkovo Institute of Science and Technology

Generation methods

Algorithmicaly-generated random numbers Mersenne twister

- Important features of a generator are the speed and the quality. These days the most frequently used one is ***Mersenne twister*** (1997, M. Matsumoto and T. Nishimura). It is used as a default random number generator in python (Note*, in C++ the default generator rand() produces rather poor results, but there are available codes for Mercenne twister*)

- ***Mersenne twister:*** Named after Mersenne primes, $\boldsymbol{2^n-1}$

- Period $2^{19937}-1$ compared with $\boldsymbol{2^{32}}$ in some old generators

More information about the generation of random numbers can be found here:
http://beam.helsinki.fi/~knordlun/mc/mc4nc.pdf
List of other generators with comments can be found here:
https://en.wikipedia.org/wiki/List_of_random_number_generators

Generation methods

Hardware-generated random numbers

True random number generators (TRNGs)

Pseudorandom number generators are actually intrinsically deterministic!

Therefore, they are normally not suitable for security applications, where it is crucial that a sequence of random numbers cannot be reproduced.

Generation methods

Hardware-generated random numbers

An electronic device could be used to produce random numbers (thermal noise in a circuit). It is very rarely done since small disturbances from an environment could influence the results

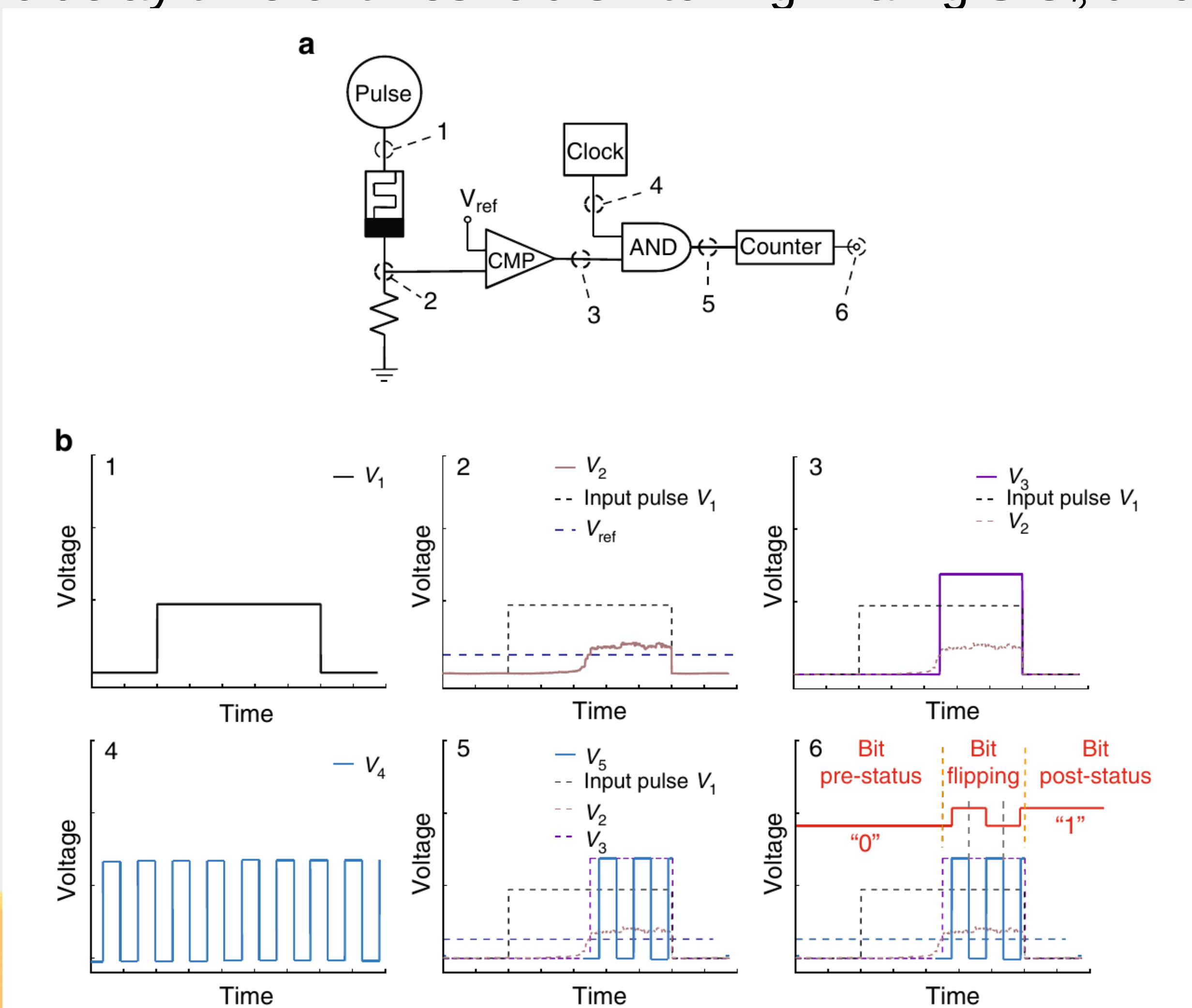
Radioactive decay is a process which does not depend on electrical features of an everyday environment. Hence it should provide a truly unpredictable sequences of random numbers

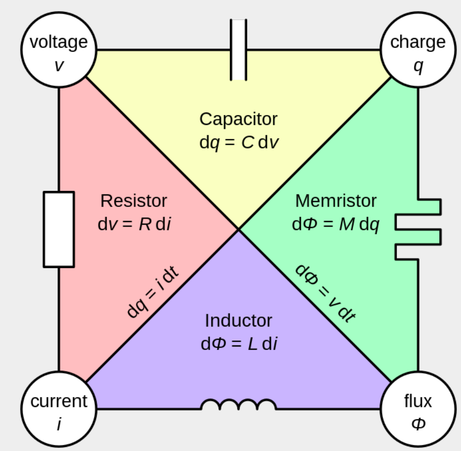
- Example: project http://www.fourmilab.ch/hotbits/ provides the random numbers produced by a radioactive source (shuts down at the end of the year).
  Potential drawback: downloading of data from the internet can be rather slow for most simulations

Generation methods

Hardware-generated random numbers

Stochastic delay time of threshold switching in a Ag:SiO_2 diffusive memristor (2017)



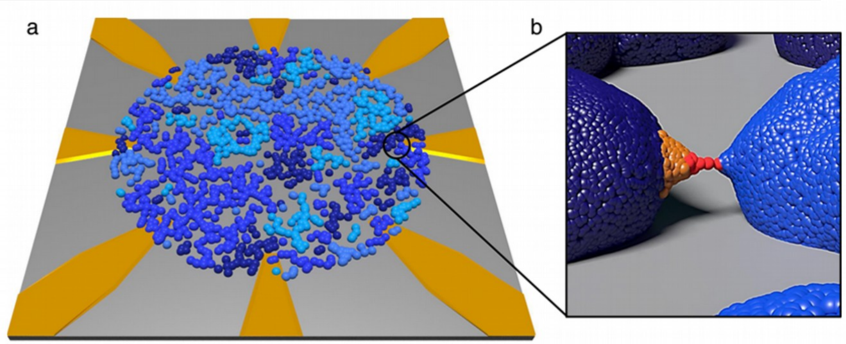


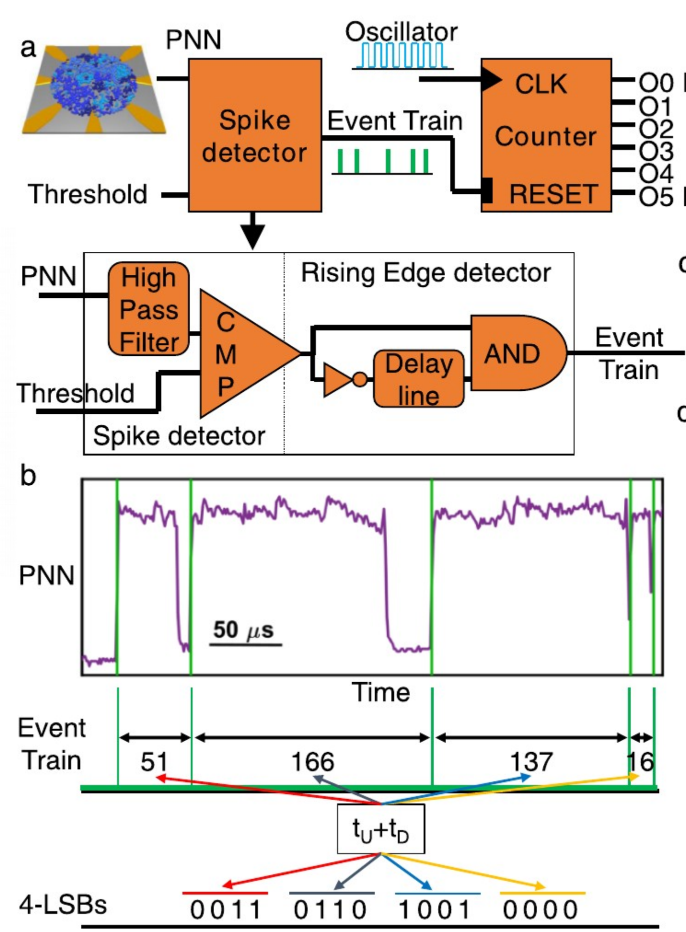
H. Jiang, Nature Communications volume 8, Article number: 882 (2017)

Generation methods

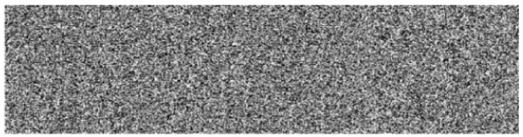
Hardware-generated random numbers

Stochastic spiking behaviour in percolating networks of tin nanoparticles





c



d

Device 2	NIST STATISTICAL TEST SUITE SP 800-22	
Test	P-value	Pass-Rate
Frequency	0.0669	97/100
Block Frequency	0.0006	99/100
Cumulative Sums	0.1453 0.3838	97/100 97/100
Runs	0.6993	99/100
Longest Run	0.7981	99/100
Rank	0.0352	45/46
FFT	0.8343	100/100
Non-Overlapping Template	> 0.0032	144/148
Overlapping Template	0.1088	96/100
Universal	0.4826	4/4
Approximate Entropy	0.2493	100/100
Serial	0.1917 0.4559	100/100 100/100

Susant K. Acharya et al., ACS Appl. Mater. Interfaces **2021**, 13, 52861−52870