Probability and statistical thinking

Biofísica Guatemala 2025

Javier Aguilar

Content

- Introduction
  -Aim
  -History
  -Philosophy

- Random variables
  - Statistics and probability
  - Distributions and histograms
  - Expectations and estimations
  - Conditioned probability
  - Moments
  - Central limit theorem*

- Inference
  - Method of moments
  - Maximum likelihood estimation
  - Bayes approach, distribution of parameters
  - Model selection
  - Inference in stochastic processes

- Stochastic processes
  - Markov chains (Branching process ?)
  - Diffusions (Brownian motion): First passage times
  - Pure jumping process (Birth-death process)
  - Van Kampen's system expansion*

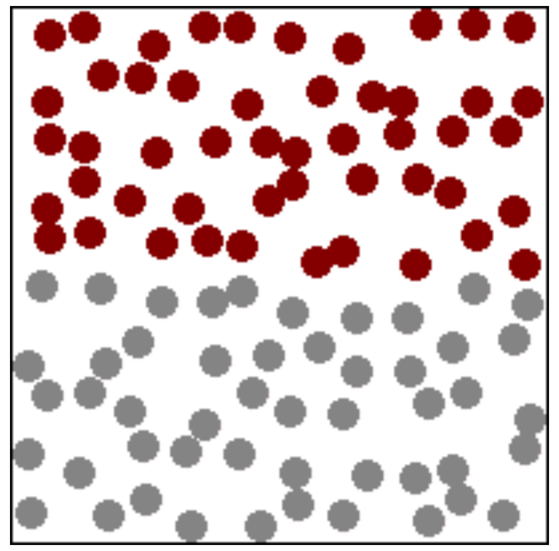
Interpretation of Bayes theorem: Have some prior knowledge/belief about event "A",update this based on observing another event "B".

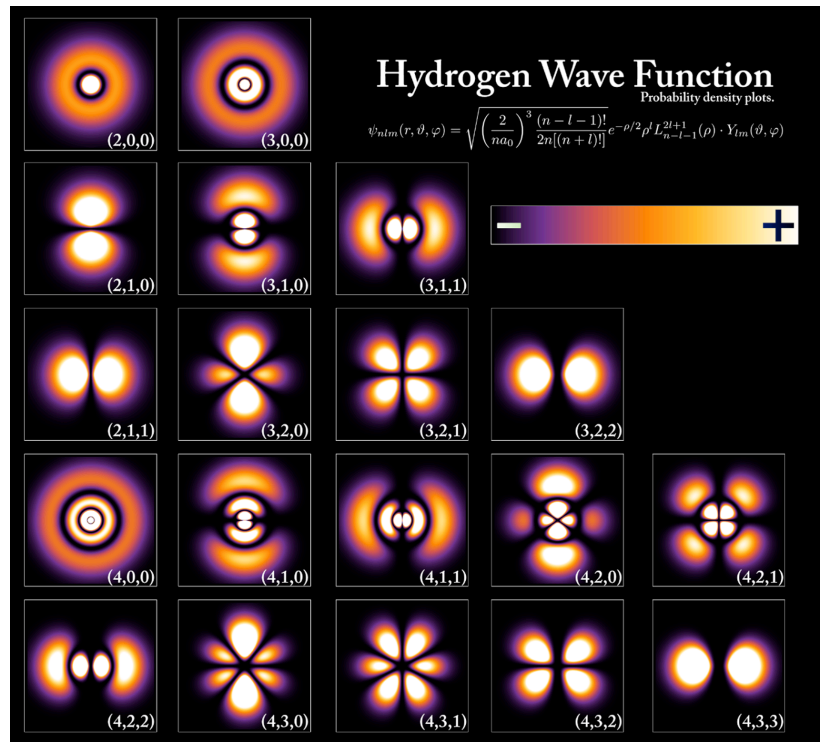
Aim and success

Probability is the logic of uncertainty

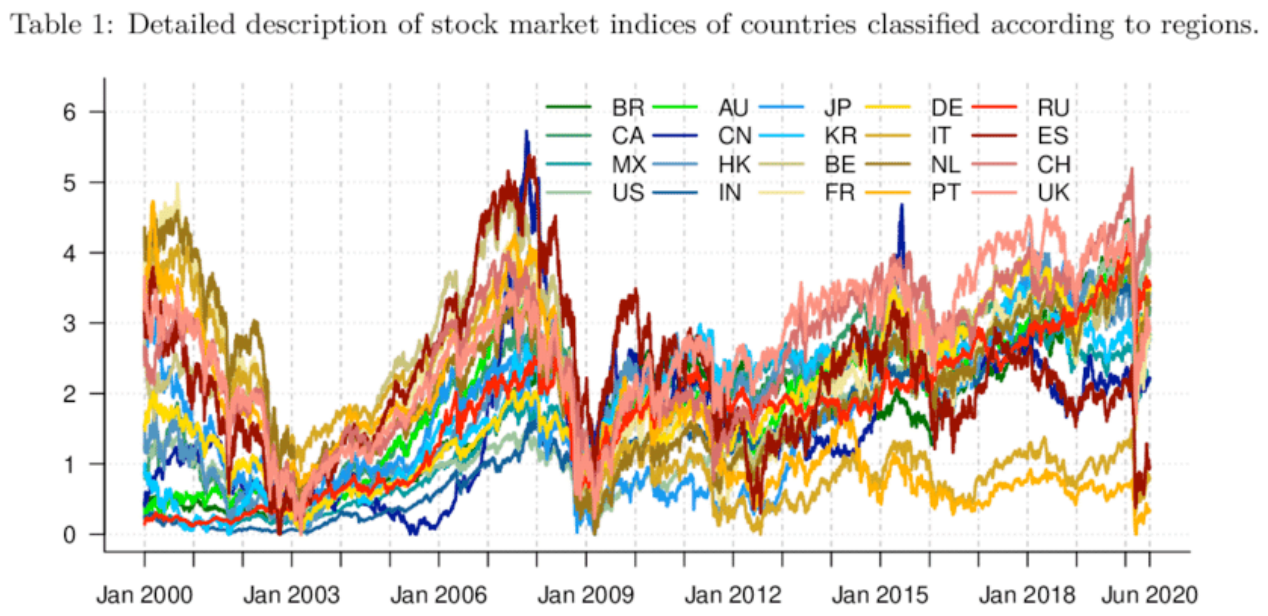
Luck. Coincidence. Randomness. Risk. Doubt. Fortune. Chance.

Physics

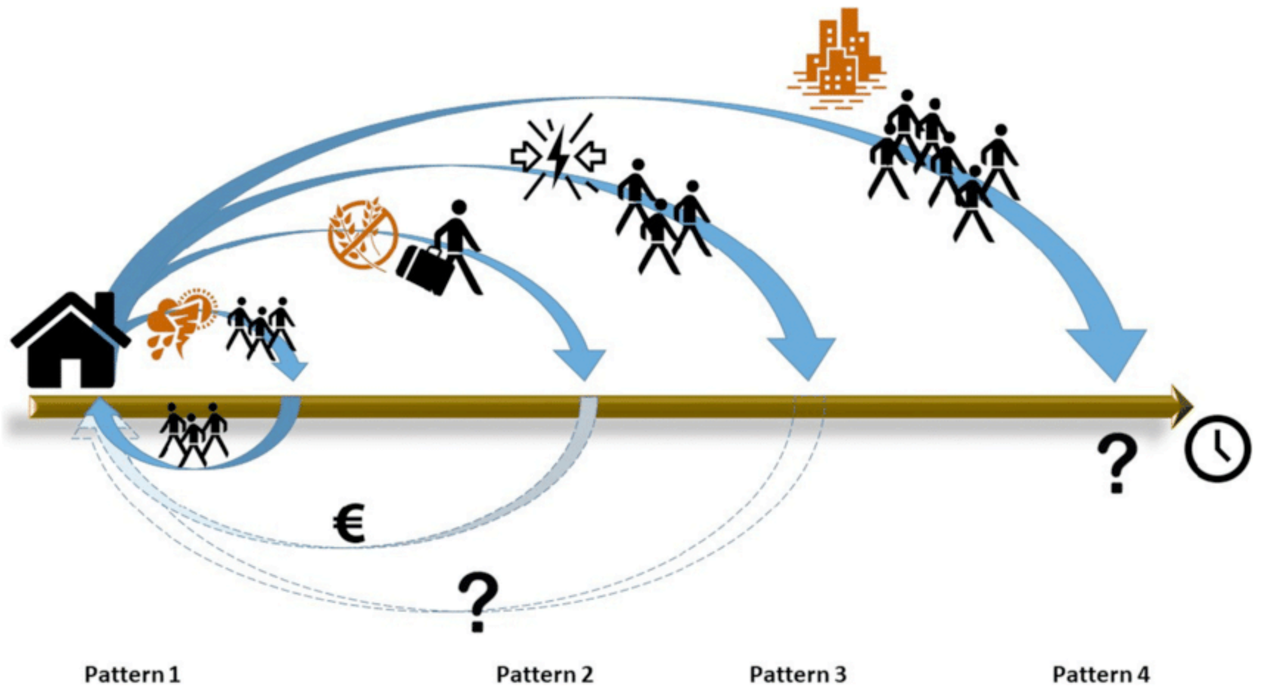




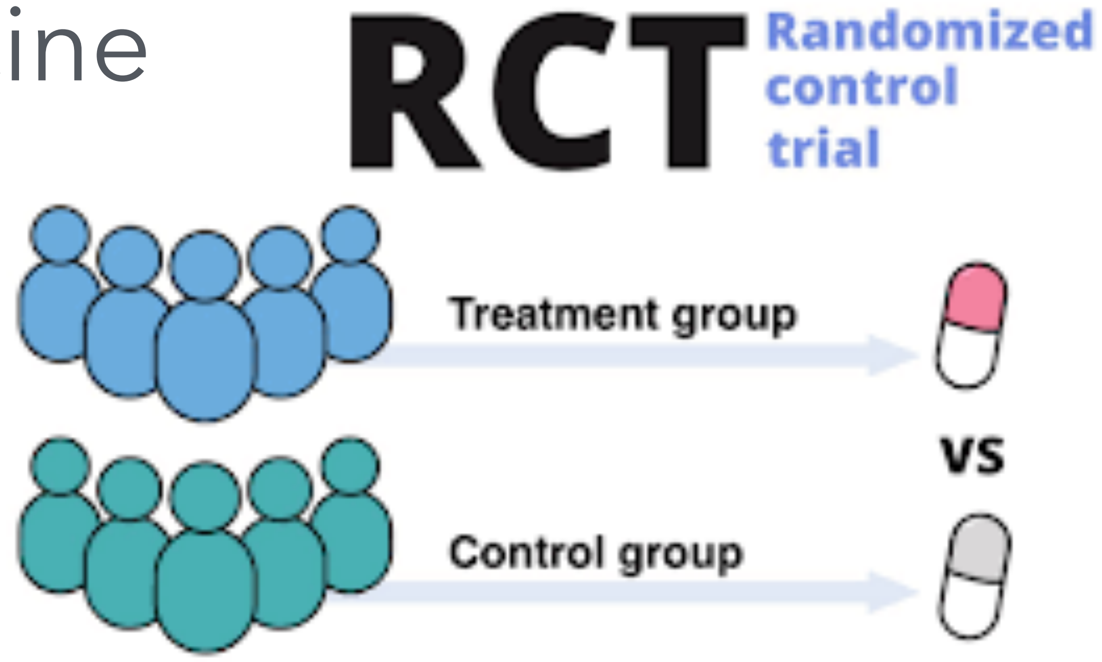
Finance



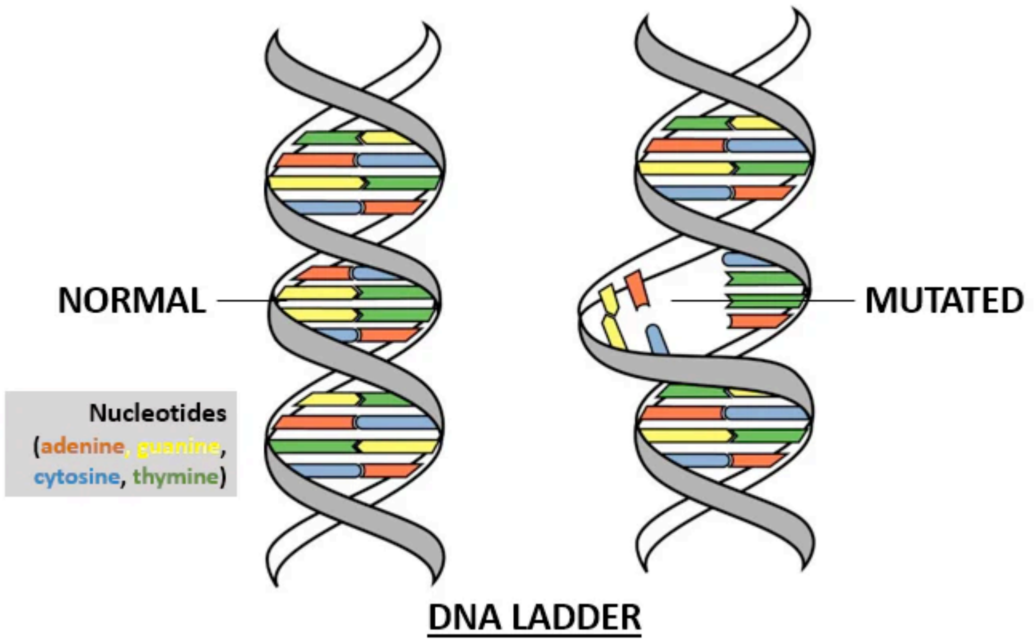
Social sciences



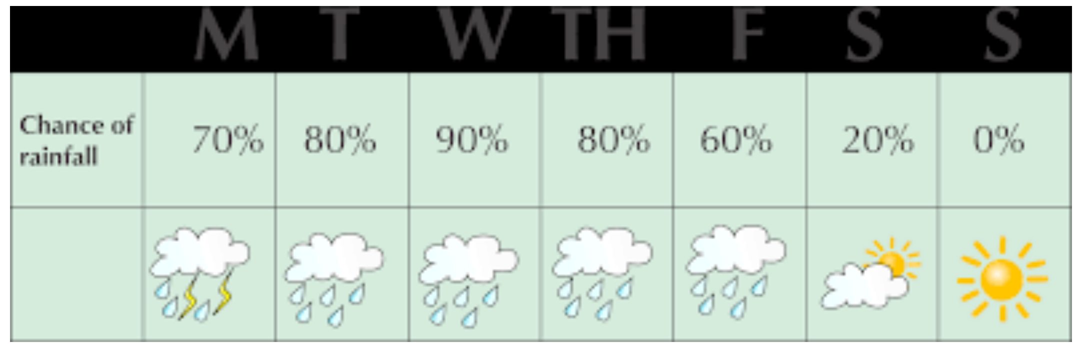
Medicine



Biology



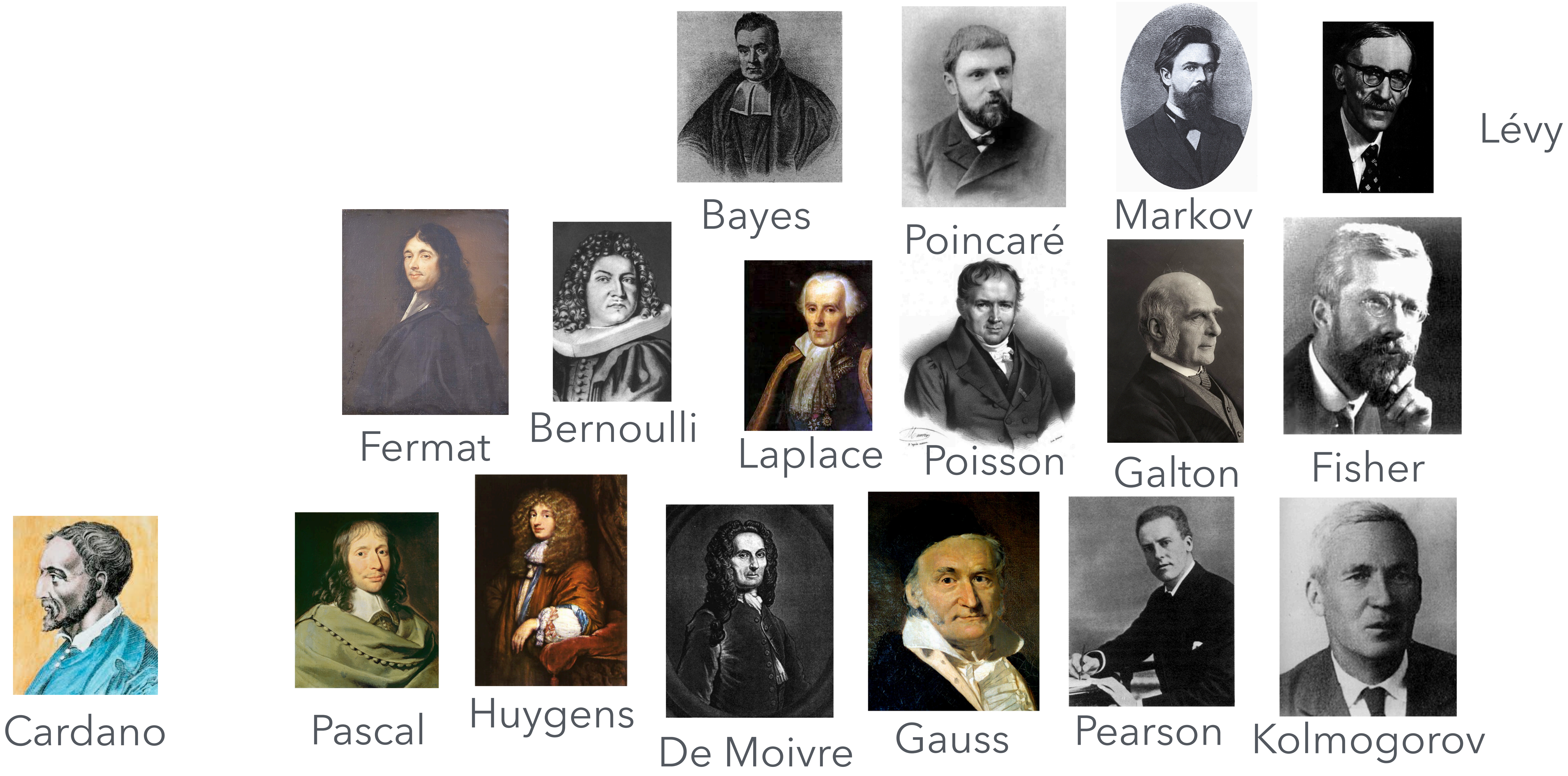
Weather forecast



History

"Since traveling was onerous (and expensive), and eating, hunting and wenching generally did not fill the 17" century gentleman's day, two possibilities remained to occupy the empty hours, praying and **gambling**; **many preferred the latter**."

Chris Van den Broeck



The rising of probability and statistical thinking

Gombaud

In 1654 a French nobleman poses a gambling question: Is it better to bet on a double six in 24 throws of two dices or one six in 4 throws?

Law of large numbers

Bayes theorem

Notion of standard deviation

Systematic use of statistics

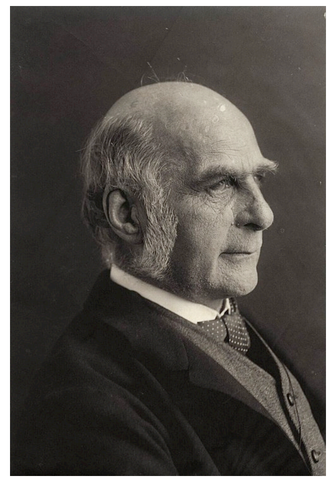
First book on probabilities: Generating functions

Probability postulates

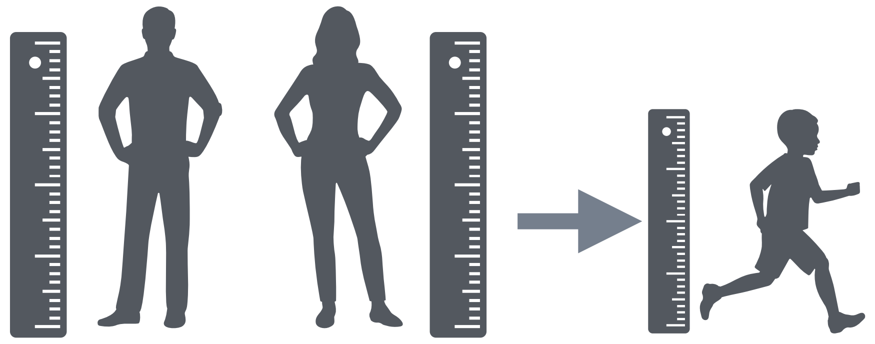
Statistics

Characterizing observed randomness

Sir Francis Galton
(1822-1911)



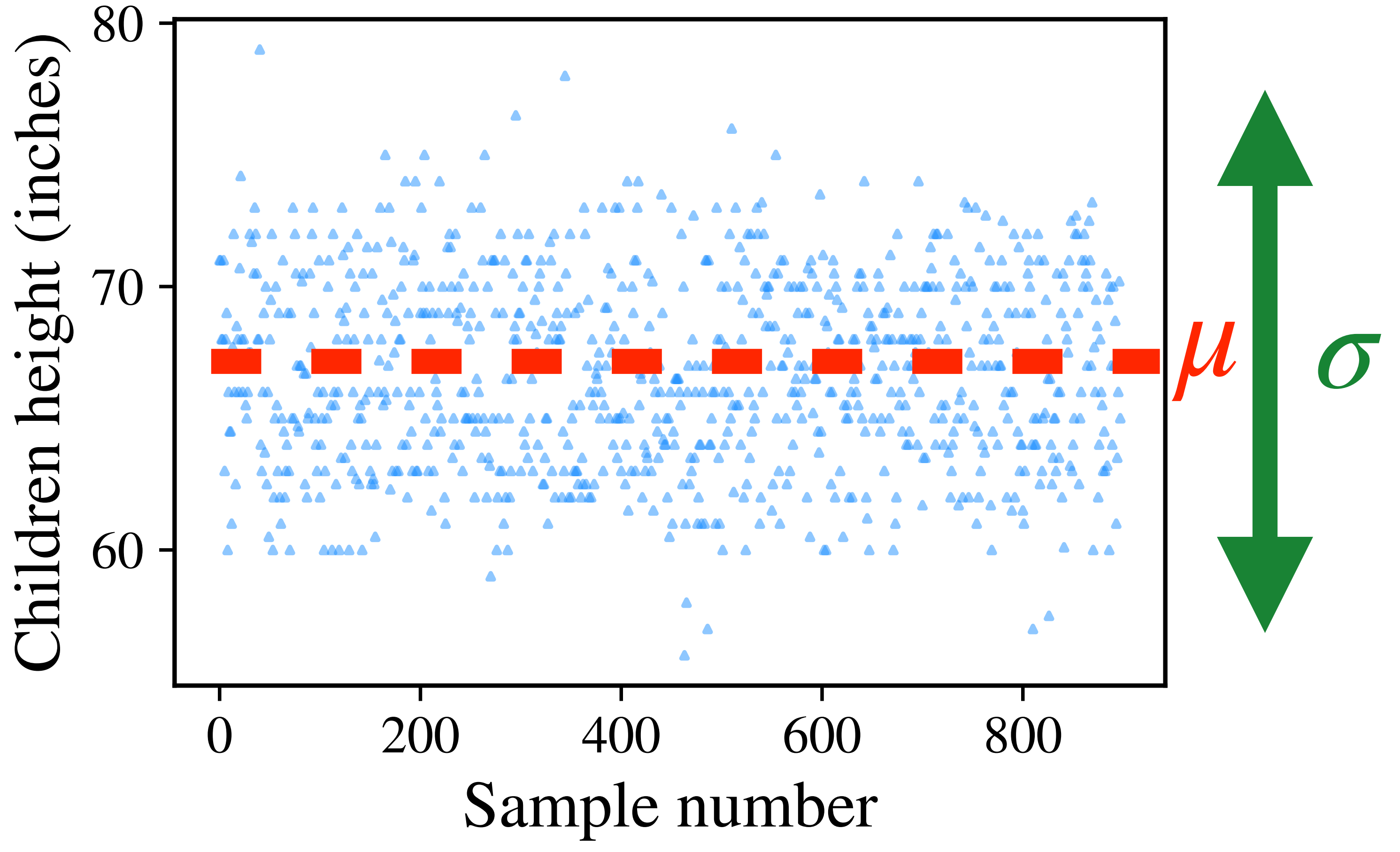
Case of study:
Human height heritage
Mean, variance, histogram, correlations



Data

	family	father	mother	gender	height	kids	male	female
0	1	78.5	67.0	M	73.2	4	1.0	0.0
1	1	78.5	67.0	F	69.2	4	0.0	1.0
2	1	78.5	67.0	F	69.0	4	0.0	1.0
3	1	78.5	67.0	F	69.0	4	0.0	1.0
4	2	75.5	66.5	M	73.5	4	1.0	0.0
...	...	...	...	...	...	...	...	...
893	136A	68.5	65.0	M	68.5	8	1.0	0.0
894	136A	68.5	65.0	M	67.7	8	1.0	0.0
895	136A	68.5	65.0	F	64.0	8	0.0	1.0
896	136A	68.5	65.0	F	63.5	8	0.0	1.0
897	136A	68.5	65.0	F	63.0	8	0.0	1.0

898 rows × 8 columns



Summary statistics

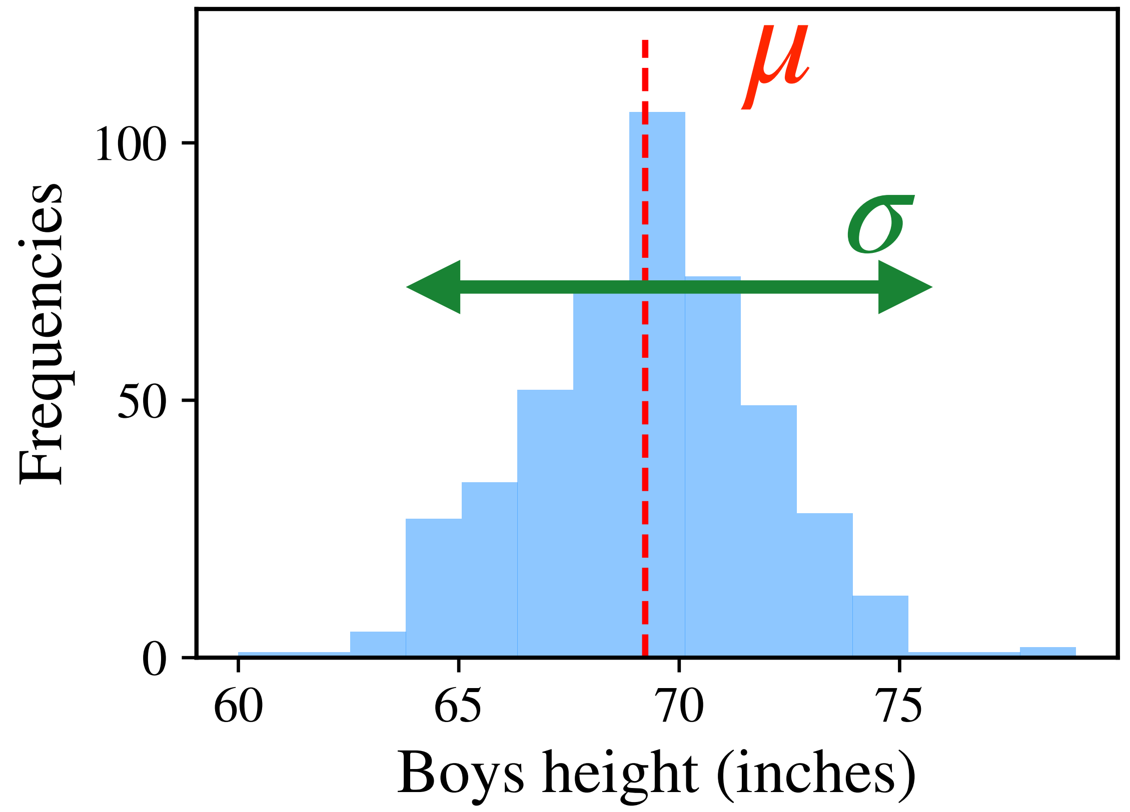
measures: N

measures: $x_1, x_2, \ldots, x_N$

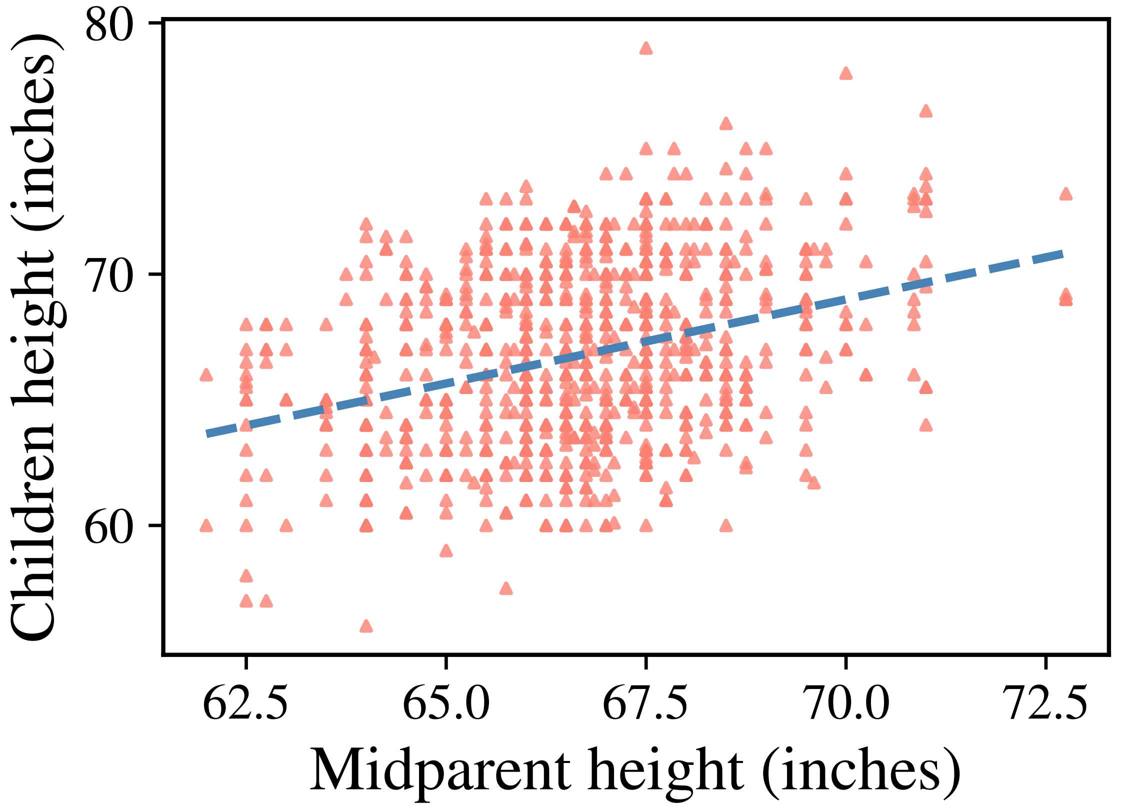
Sample Mean: $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$

Sample Variance: $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \hat{\mu}^2$

Frequency: (histogram) $f(x) = \sum_{i=1}^{N} I(x_i \in [x, x+dx])$



Correlations and linear regression



"Midparent" height : $h_m = \frac{1}{2}\left(h_{father} + h_{mother}\right)$

Linear regression: $h_c = \alpha h_m + \beta$

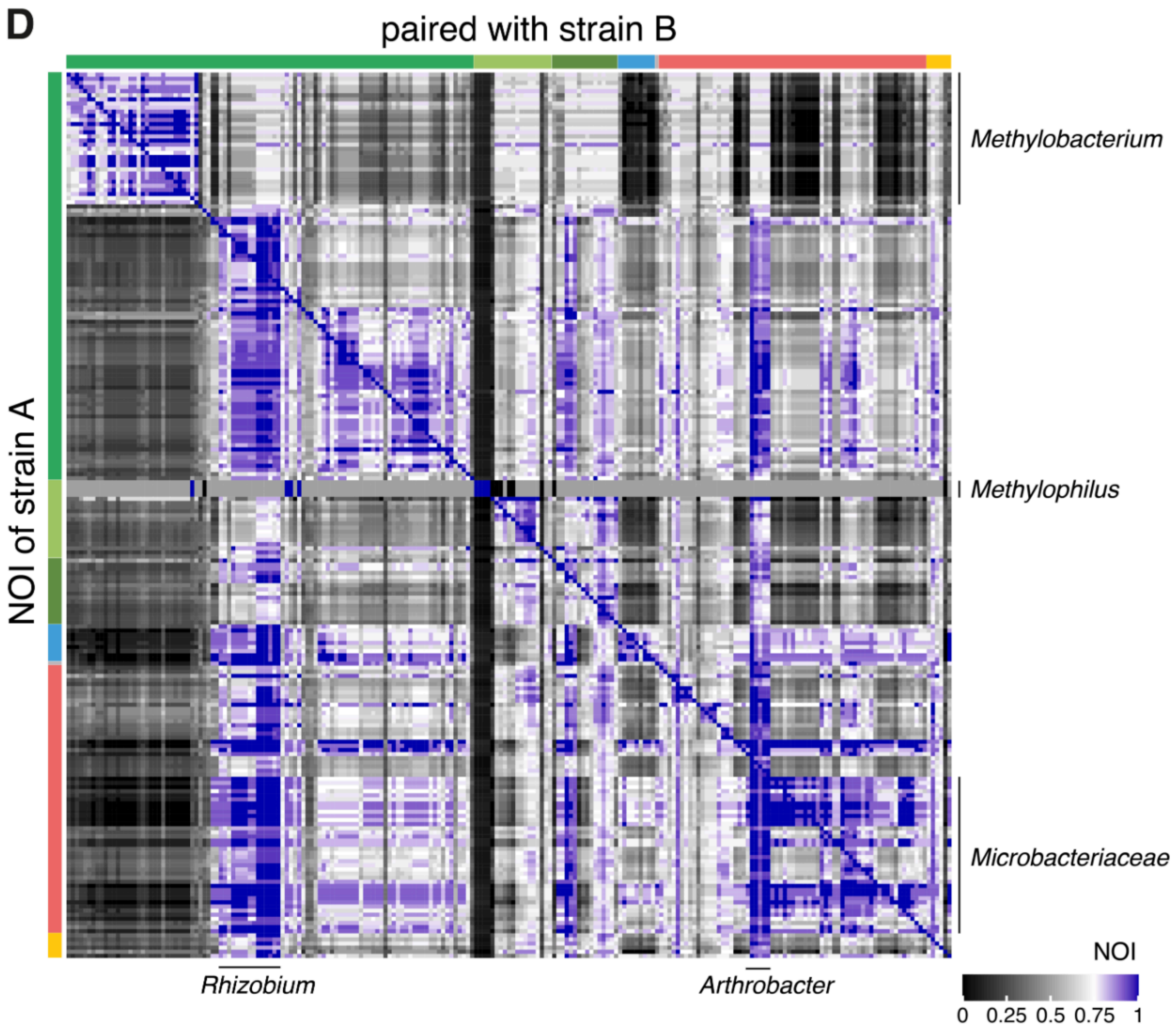
Sample Covariance: $\hat{C} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$

Pearson's coefficient: $\hat{\rho} = \frac{\hat{C}}{\hat{\sigma}_x \hat{\sigma}_y}$

Statistics

Characterizing observed randomness

How to visualize complex data?



How to interpret summary statistics?

What measures the central tendency?

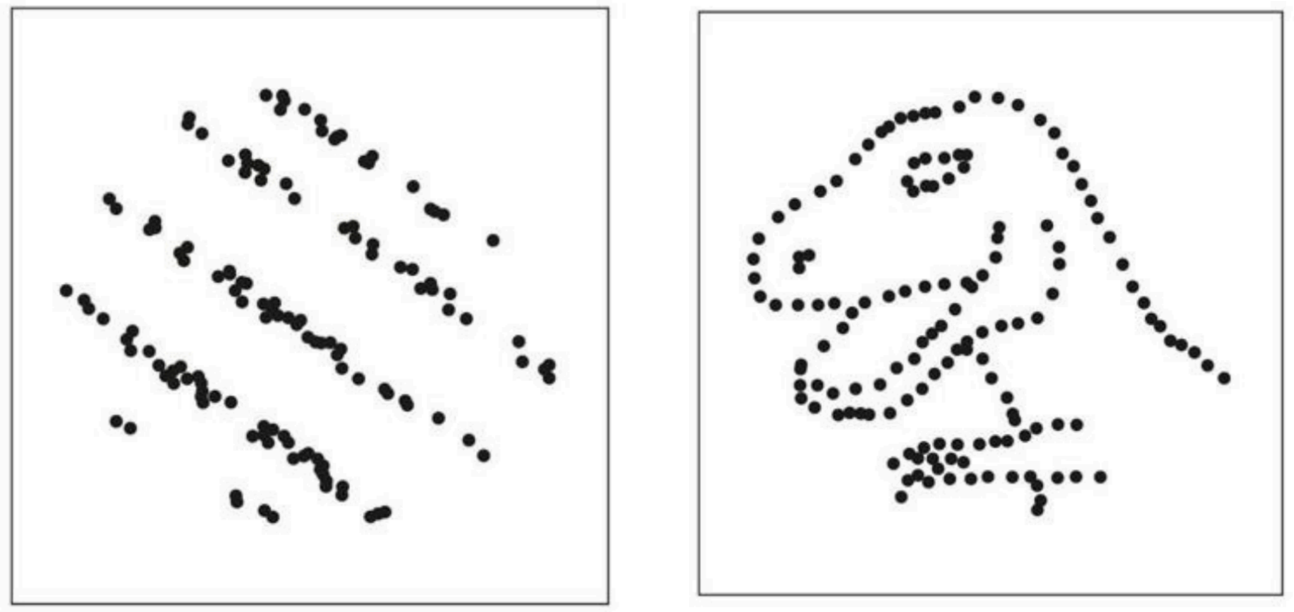
Mean, mode or median

How many sigmas?

Chebyshev’s theorem

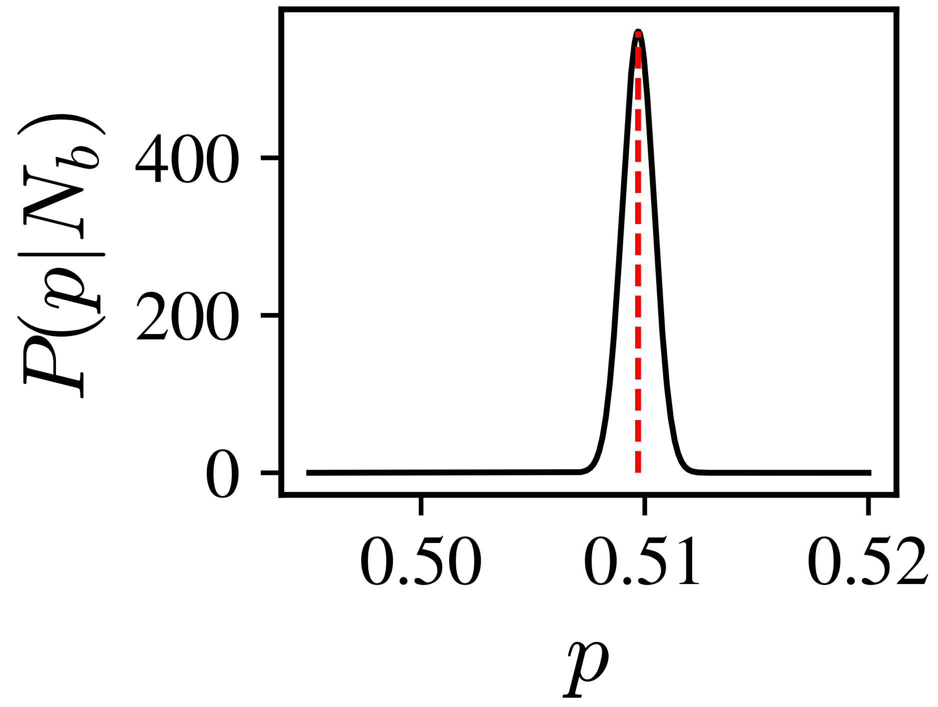
Is it correlation or causation?

Modeling

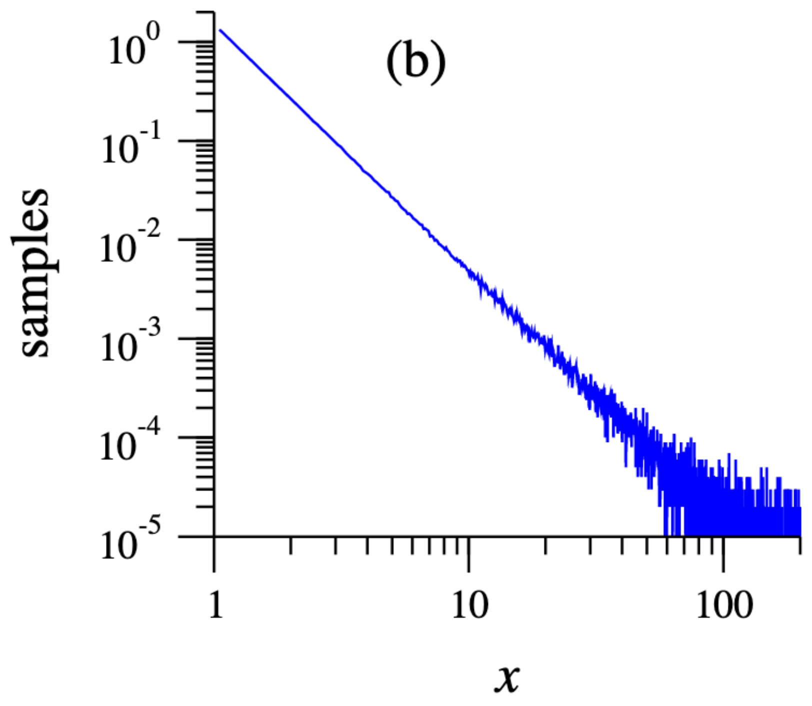


How to deal with fluctuations?

Confidence intervals



Logarithmic vs. linear binning



Statistics

Characterizing observed randomness

Probability estimators: normalized histograms

THE DEVIL IS IN THE DETAIL

Discrete data

$\hat{x}_i \in \{a, b, c, \dots\}$

$\hat{x}_i \in \{1,2,3,4,5,6\}$

Frequency estimate

$$\hat{f}(x) = \sum_{i=1}^{N} I(\hat{x}_i = x) \in [1,N]$$

Probability estimate

$$\hat{p}(x) = \frac{1}{N}\hat{f}(x) \in [0,1]$$

Example:

$\hat{x}_1 = 4 \quad \hat{x}_2 = 6$

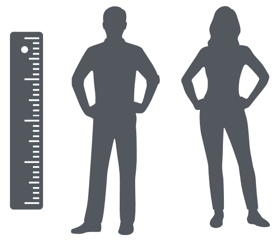
$\hat{x}_3 = 1 \quad \hat{x}_4 = 4$

$$\hat{f}(1) = 1 \quad \hat{f}(4) = 2 \quad \hat{f}(3) = 0$$

$$\hat{p}(1) = \frac{1}{4} \quad \hat{f}(4) = \frac{1}{2} \quad \hat{p}(3) = 0$$

Continuous data

$\hat{x}_i \in [a, b]$



$\hat{x}_i \in [0\text{ m}, 4\text{ m}]$

$$\boxed{I(\hat{x}_i = x) = 0}$$

Two real numbers are never equal: real numbers have zero measure

$$\hat{f}(x, \Delta x) = \sum_{i=1}^{N} I(\hat{x}_i \in [x, x + \Delta x])$$ Frequency estimate

$$\hat{p}(x, \Delta x) = \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x, \Delta x)$$ Probability estimate

Example:

$\hat{f}(x = 55\text{ in.}, \Delta x = 30\text{ in.}) = 896$

$\hat{p}(x = 55\text{ in.}, \Delta x = 30\text{ in.}) = 1$

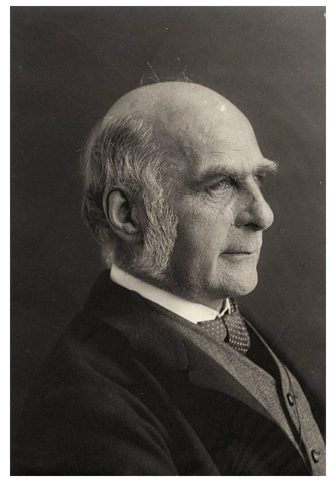
$\hat{f}(x = 55\text{ in.}, \Delta x = 5\text{ in.}) = 31$

$\hat{p}(x = 55\text{ in.}, \Delta x = 5\text{ in.}) \approx 0.03$

$\hat{f}(x = 55\text{ in.}, \Delta x = 0\text{ in.}) = 0$

$\hat{p}(x = 55\text{ in.}, \Delta x = 5\text{ in.}) \approx 0$

Galton's data



Getting rid of bins

Probability density function (estimate)

$$\hat{\rho}(x) = \frac{\hat{\rho}(x, \Delta x)}{\Delta x} \in [0,\infty]$$

Cumulative distribution function (estimate)

$$\hat{F}_<(x) = \frac{1}{N}\sum_{i=1}^{N} I(\hat{x}_i < x) \in [0,1]$$

$$\hat{F}_>(x) = \frac{1}{N}\sum_{i=1}^{N} I(\hat{x}_i > x) \in [0,1]$$

$$\hat{F}_<(x) + \hat{F}_>(x) = 1$$

Probability

Modeling randomness

Jacob Bernuilli
(1654 - 1705)

Case of study:

Flipping coins
Bernoulli distribution
Law of large numbers

Kolgomorov's receipt for probability theory

Andréi Kolmogórov
(1903 - 1987)

Ingredient 1: Sample space
(all possible outcomes)

H: heads

T: tails

Ingredient 2: Random variable
(Each outcome has a number)

X(H)=1 X(T)=0

Ingredient 3: Define probabilities
(each outcome has a probability)

$P(H) \in [0,1]$

$P(T) \in [0,1]$

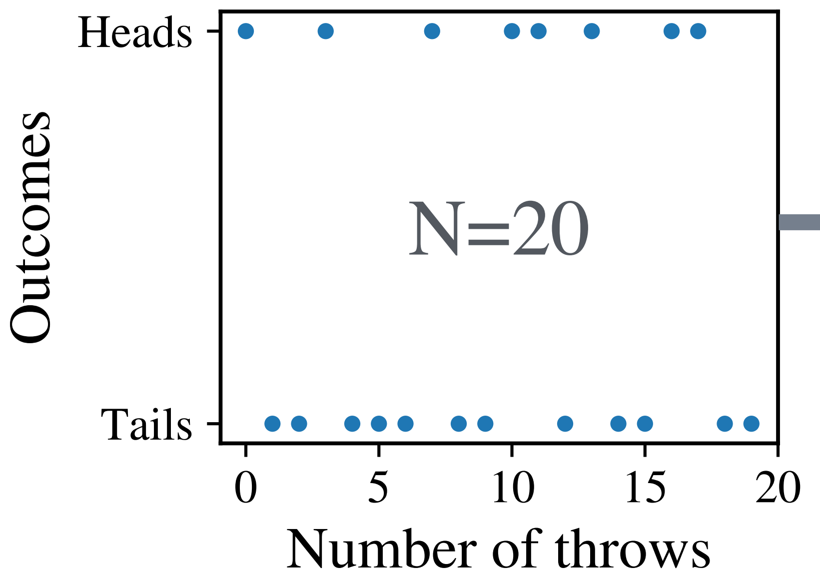
$P(H) + P(T) = 1$

How to assign probabilities?

Method I: Maximum ignorance
Symmetric guess

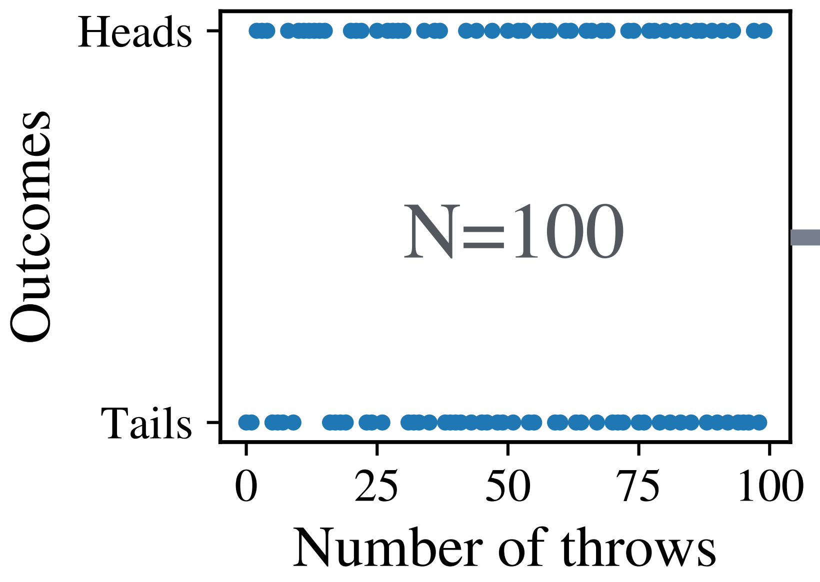
$$P(H) = P(T) = \frac{1}{2}$$

Moments



$\hat{\mu} = 0.40$

$\hat{\sigma} = 0.24$



$\hat{\mu} = 0.48$

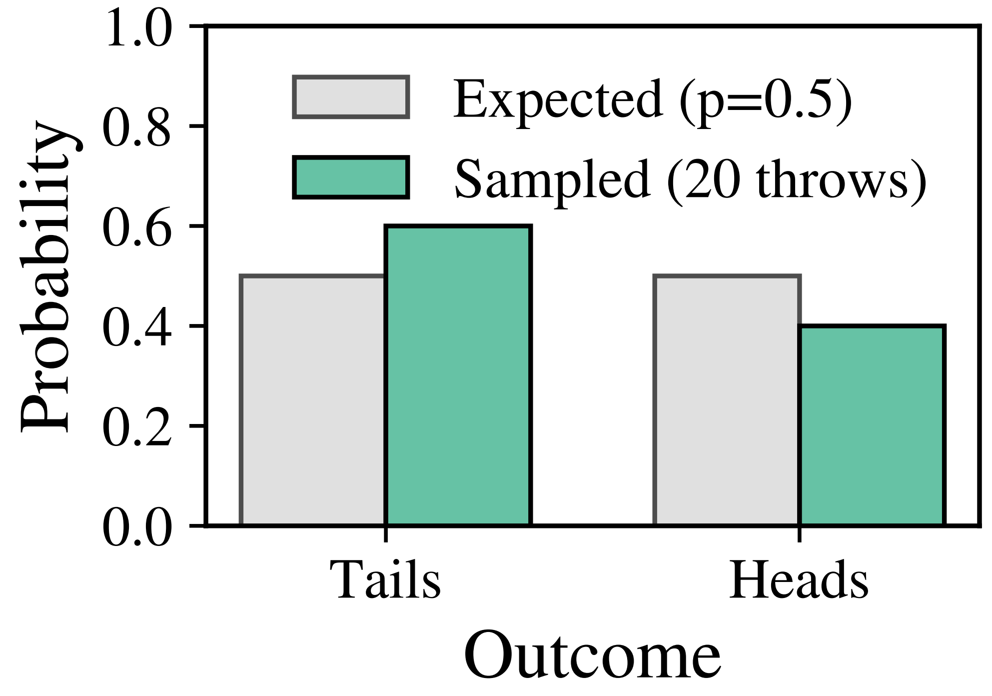
$\hat{\sigma} = 0.2496$

$$\langle X \rangle = E(X) = \mu = \sum_{x \in \Omega} xP(X = x) = 0.5 \longrightarrow \hat{\mu} \approx \mu$$

$$\langle (X-\mu)^2 \rangle = \mathrm{var}(X) = \sigma^2 = E(X^2) - \mu^2 = 0.25 \longrightarrow \hat{\sigma} \approx \sigma$$

Probabilities

$\hat{P} \approx P$



Law of large numbers

$$\lim_{N \to \infty} \hat{P}(X \in A) = P(X \in A)$$

$$Z = z(X)$$

$$\downarrow$$

$$\lim_{N \to \infty} \hat{Z} = E(Z)$$

Probability

Modeling randomness

Antoine Gombaud (1607-1684) | Blaise Pascal (1623-1662) | Pierre Fermat (1607-1665)

Case of study:

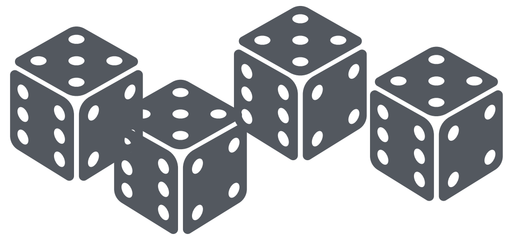
Throwing dices

Uniform distribution

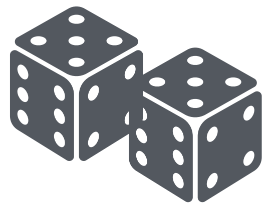
Binomial distribution

Which is more likely?

Game A: Rolling, at least, one six in 4 throws of one die

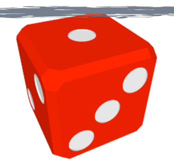


Game B: Rolling, at least, one double six in 24 throws of two dice



$\times\, 24$

How to assign probabilities of one die?



Method I: Maximum entropy/ignorance

Symmetric guess

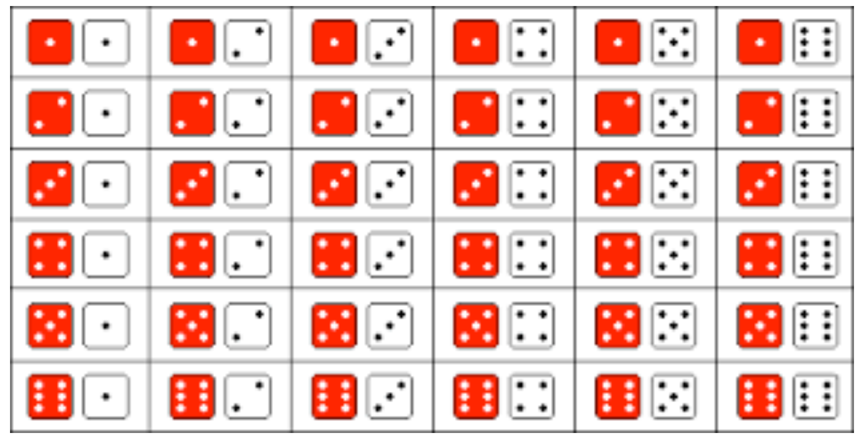
$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

How to assign probabilities of two dice?

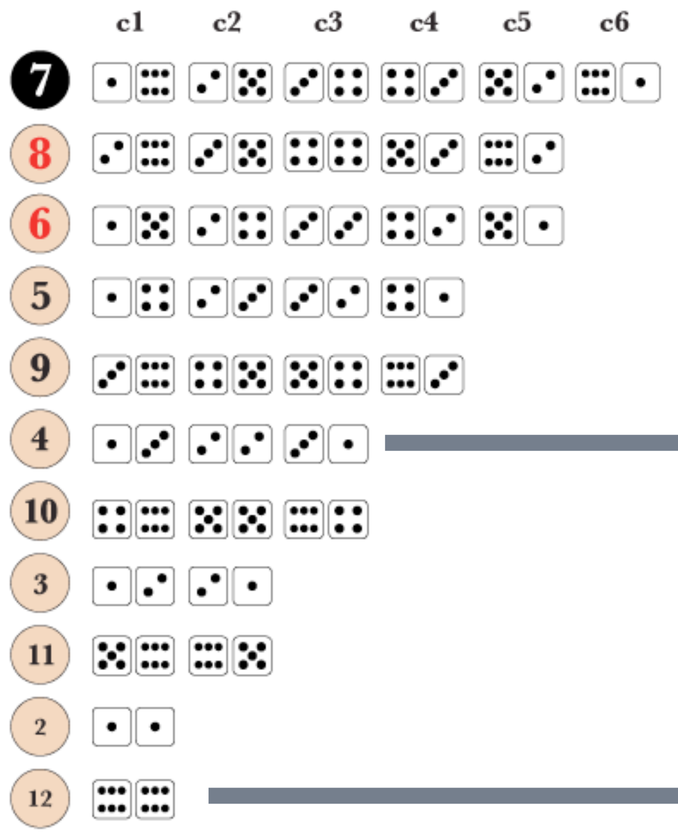
Method II: Intuitive definition of probability

$$P(A) = \frac{\#A}{\#\Omega}$$

$$\#\Omega = 6^2 = 36$$



$\#A$



$$P(7) = \frac{6}{36} = \frac{1}{6}$$

$$P(4) = \frac{3}{36} = \frac{1}{12}$$

$$P(12) = \frac{1}{36}$$

Computing probabilities

$P(V) + P(V^C) = 1 \longrightarrow P(V) = 1 - P(V^C)$ (Either it rains or it doesn't)

$V_1, \ldots, V_n$ Independent events $\longrightarrow P(V_1, \ldots, V_n) = P(V_1)P(V_2)\ldots P(V_n)$

$$\boxed{P(A \text{ wins}) = 1 - P(A \text{ loses}) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.52}$$

$$P(B \text{ wins}) = 1 - P(B \text{ loses}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.49$$

Game B loses on average, game A wins on average

Probability

Modeling randomness

Antoine Gombaud (1607-1684)

Blaise Pascal (1623-1662)

Pierre Fermat (1607-1665)

Case of study:

Throwing dices

Uniform distribution

Binomial distribution

Central limit theorem:

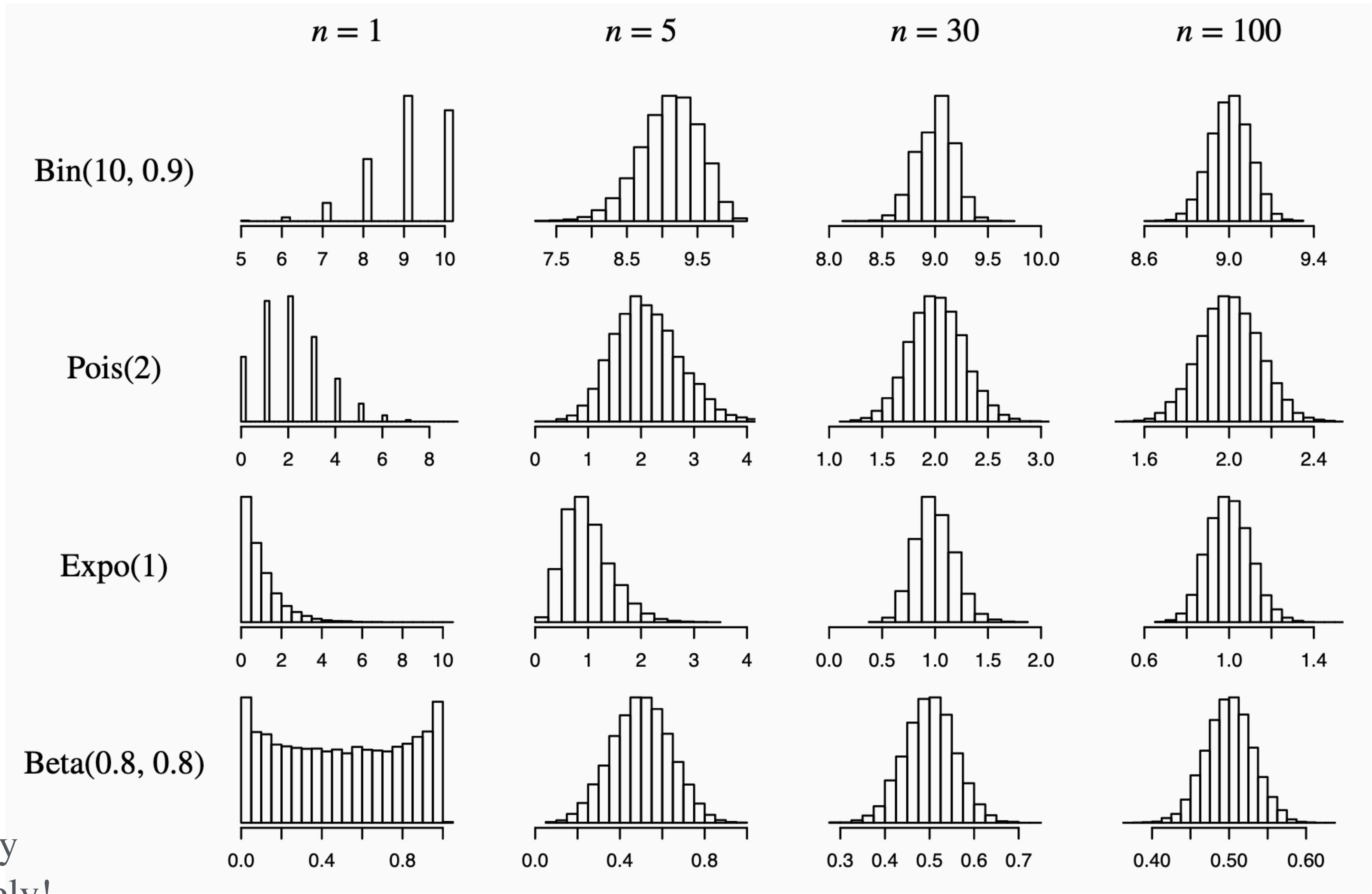
For any R.V. X with $\mu = E(X), \sigma^2 = Var(X) < \infty$

Let us define $Y := \frac{1}{N}\sum_{i=1}^{N} X^{(i)}$, Then

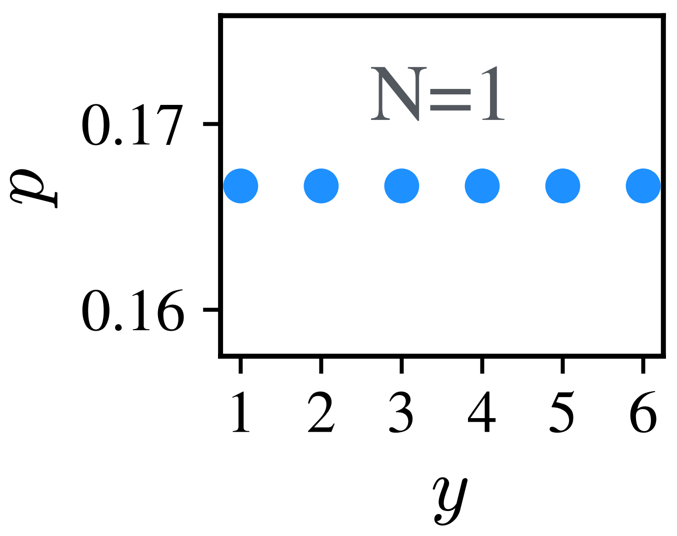
$$E(Y) \approx \mu \quad Var(Y) \approx \sigma^2/N \quad \rho(y) \approx G\left(x; \mu, \frac{\sigma}{\sqrt{N}}\right)$$

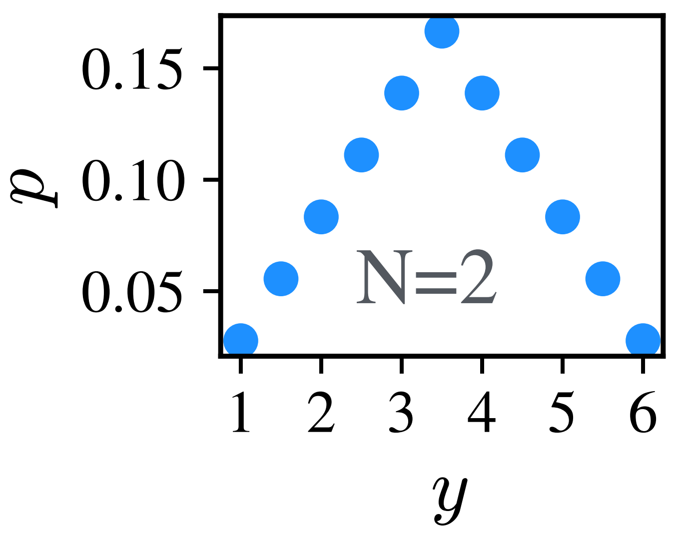
$$G(x; \mu, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Robustness of statistical patterns

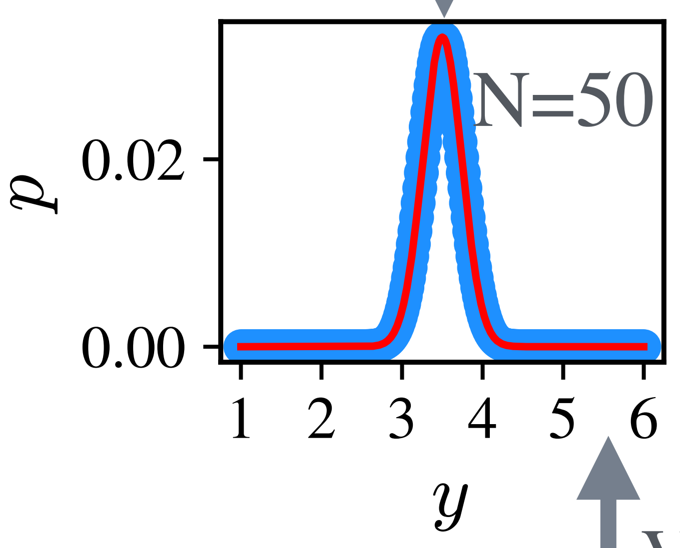


Betting on the tails of the distribution becomes riskier!





Very likely!



Very unlikely!

Probability

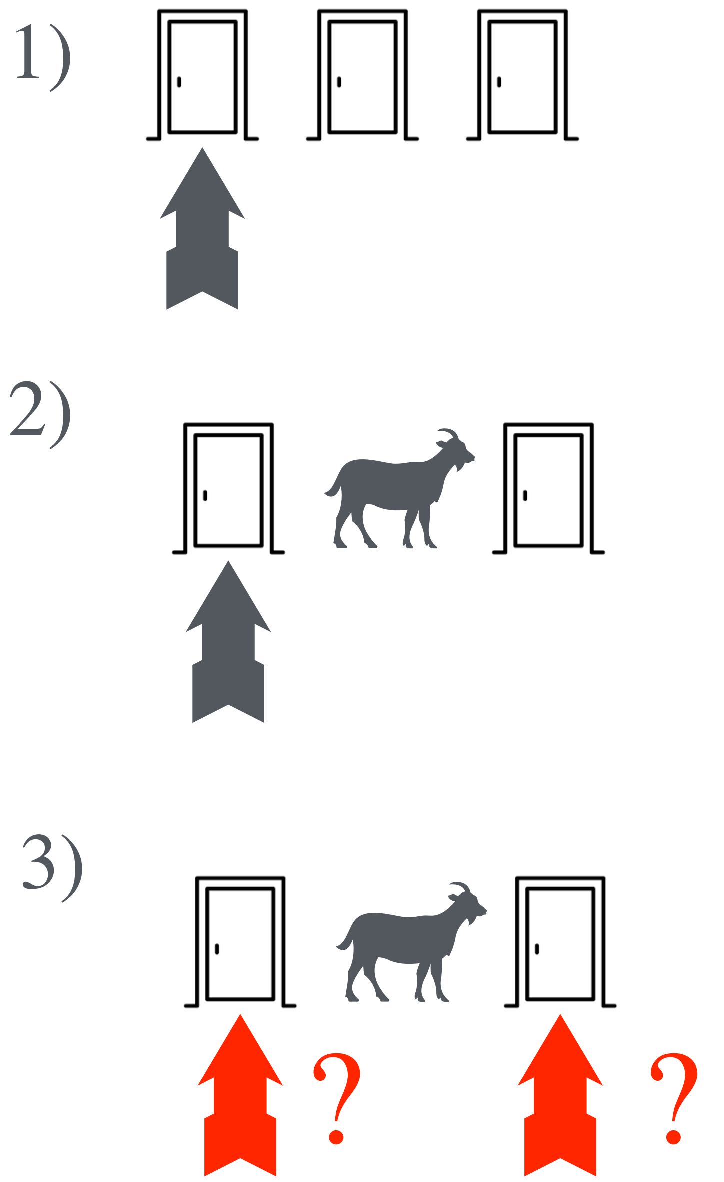
Modeling randomness

Marilyn vos Savant
(1946 -)

Case of study:

Monty Hall paradox

Conditioned probabilities



Rules:

1) Choose one door.
2) M.H. opens another door with no prize.
3) Stay with your initial choice or change.

Q: Is there a winner strategy?

- It doesn't matter if we stay or change.
- It is better to stay.
- It is better to change.

Reasoning 1:

M.H. provides information that favors one choice.
P(win) > 1/2.

Reasoning 2:

M.H. is just goofing around.
P(win) =1/2

C: Where is the car M: Where is M.H $C, M \in \{a, b, c\}$

$P(C = a \,|\, M = b) = P(\text{win staying}) = ?$

$P(C = b \,|\, M = b) = 0$

Say we choose the door "a" and M.H. shows a goat in door "b"

$P(C = c \,|\, M = b) = P(\text{win changing}) = ?$

What do we know? $P(M = b \,|\, C = a) = 1$ $P(M = b \,|\, C = B) = 0$ $P(M = b \,|\, C = B) = 1/2$

Bayes: $P(C = a \,|\, M = b) = \dfrac{P(M = b \,|\, C = a) P(C = a)}{P(M = B)} = \dfrac{2}{3} \longrightarrow P(C = a \,|\, M = b) = \dfrac{1}{3}$

A: It is more likely to win if we change!!

Probability

Modeling randomness

William Feller
(1906 - 1970)

Case of study:

Waiting time paradox

Exponential distribution
Poisson distribution

Exponential distribution

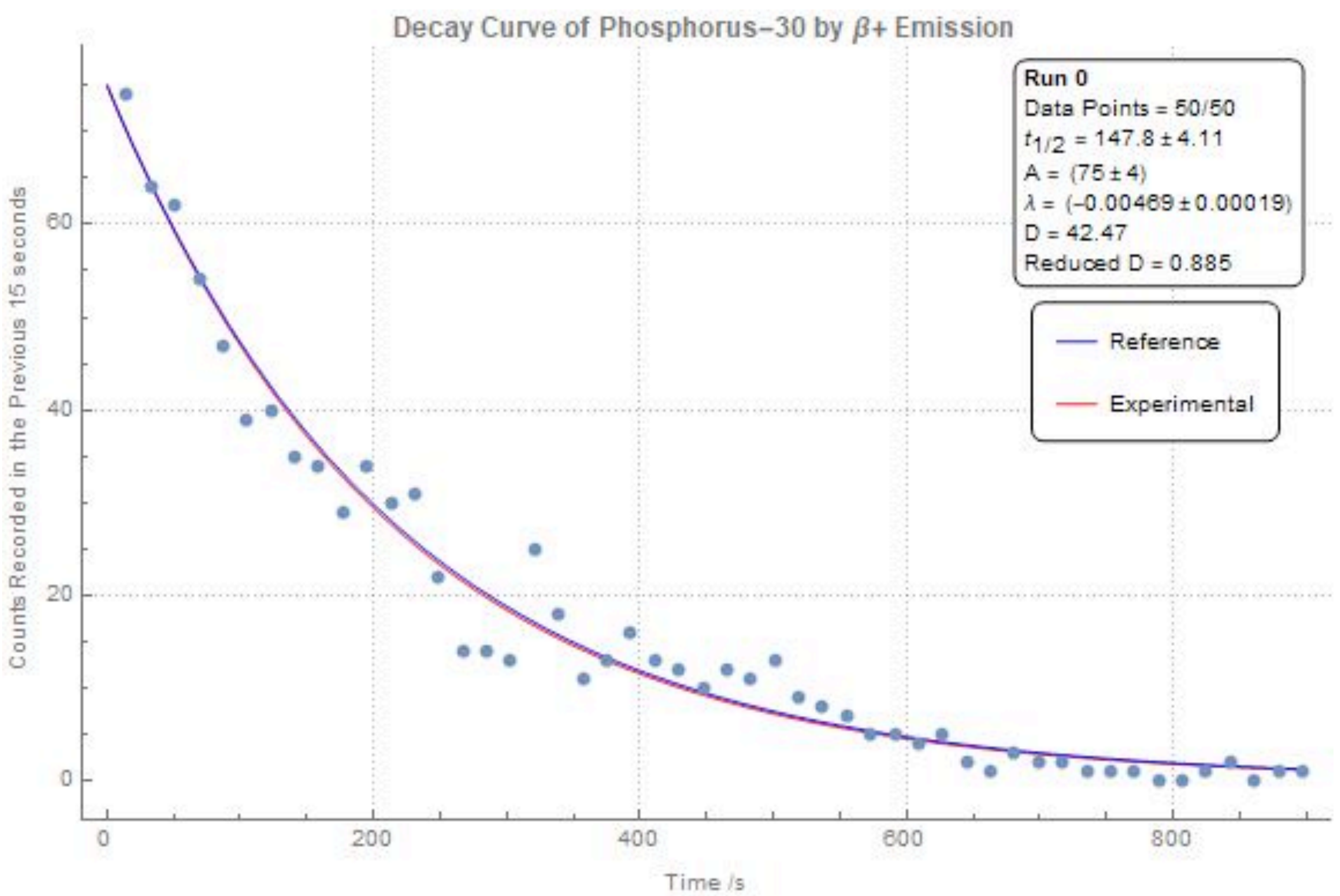
$$P(\Delta t) = w e^{-w\Delta t}$$

$$\boxed{E(\Delta t) = \frac{1}{w}}$$

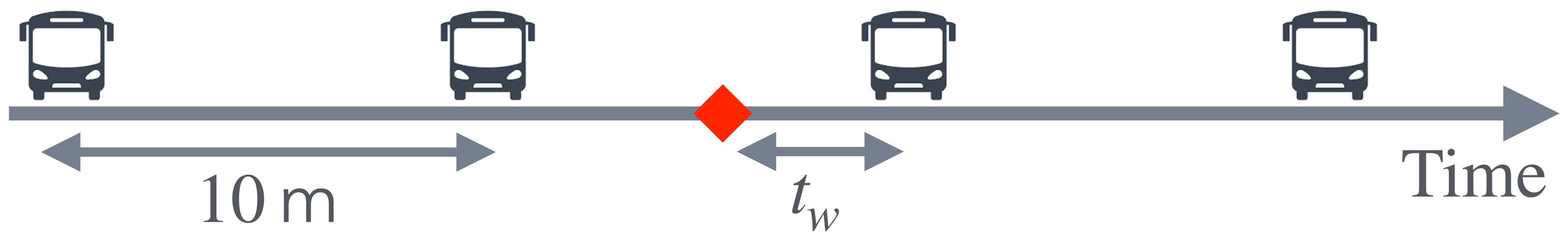
Typical time scale

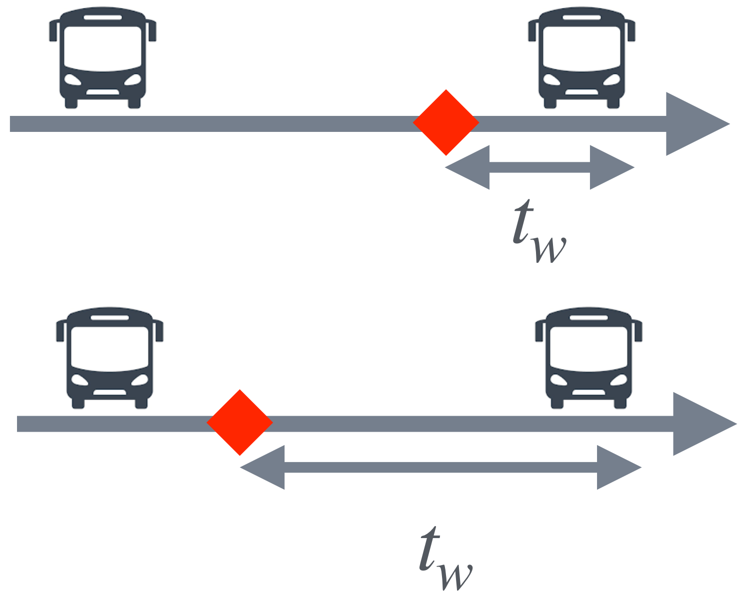
Memoryless property: $\frac{P(\Delta t)}{P(0)} = \frac{P(\Delta t + t)}{P(t)}$

Example: Radioactive decay



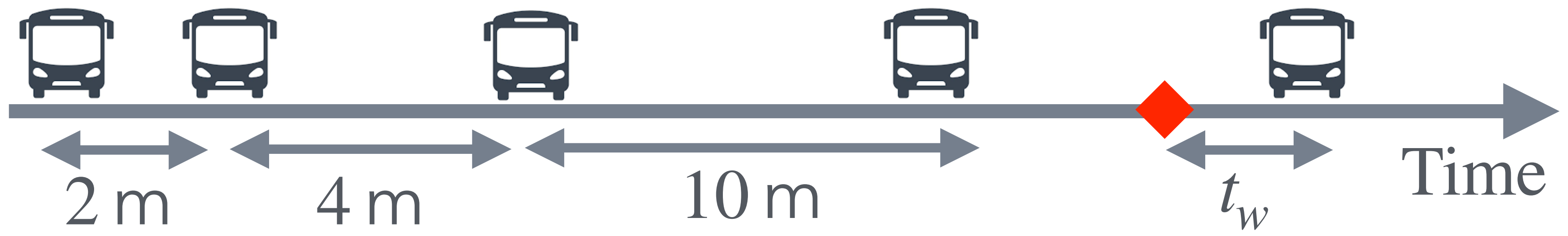
Waiting times with deterministic arrivals





Bad luck -good luck compensation: $E[t_w] = \frac{10\text{ m}}{2} = 5\text{ m}$

Waiting times with random arrivals



$$P(\Delta t) = 10\, e^{-10\Delta t}\,\text{m}^{-1}$$

Bad luck -good luck compensation: $E[t_w] = \frac{E[\Delta t]}{2} = 5\text{ m}$

Memoryless property: Waiting time doesn't depend on arrival time $\boxed{E[t_w] = 10\text{ m}}$

Probability

Modeling randomness

William Feller
(1906 – 1970)

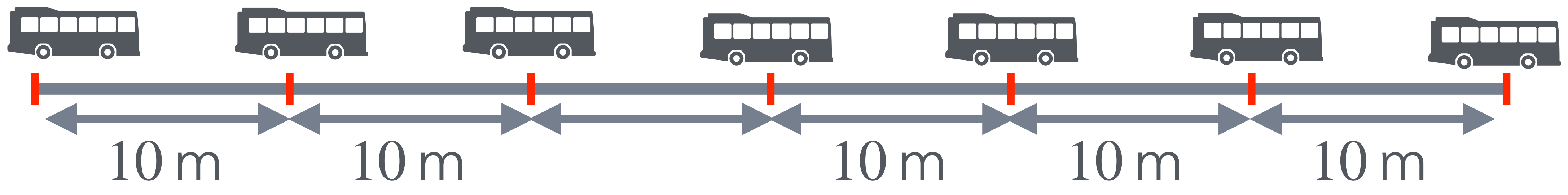
Case of study:

Waiting time paradox

Exponential distribution

Poisson distribution

Intuitive understanding 6 arrivals in one hour:



5 arrivals with 2 m

Average time between arrivals = $\frac{1}{6}\left(5 \cdot 2 + 50\right) = \boxed{10\text{ m}}$ → Event-based observer

Prob. Of hitting 2- and 50- time intervals

Average observed T.B.A.= $5\boxed{\frac{2}{60}}2 + \boxed{\frac{50}{60}}50 = \boxed{42\text{ m}}$ → Random-time observer

Deterministic case bounds random one

$$E[t_w] \geq = 10\text{ m}$$

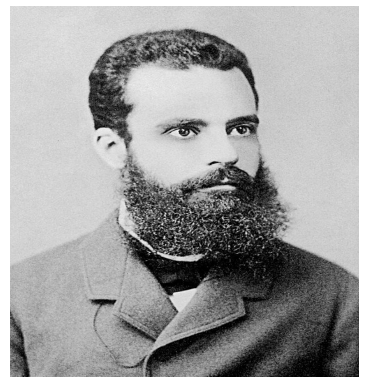
How we gather data alter statistics (Length bias)

$\neq$

It is more likely to arrive in long intervals between arrivals !!!

Probability

Modeling randomness



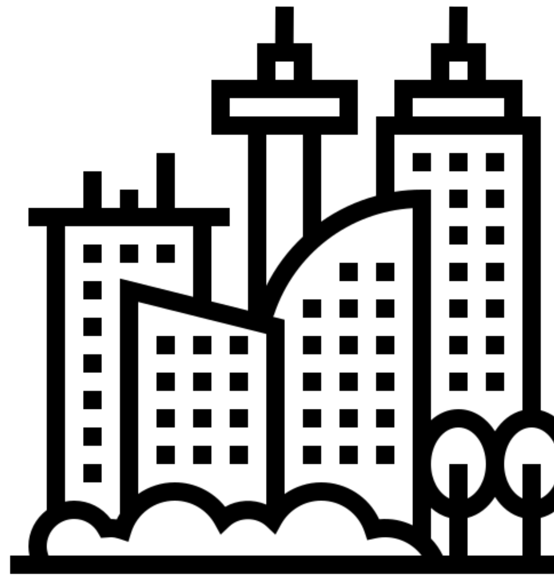
Vilfredo Pareto (1848-1923)

George Zipf (1902-1950)

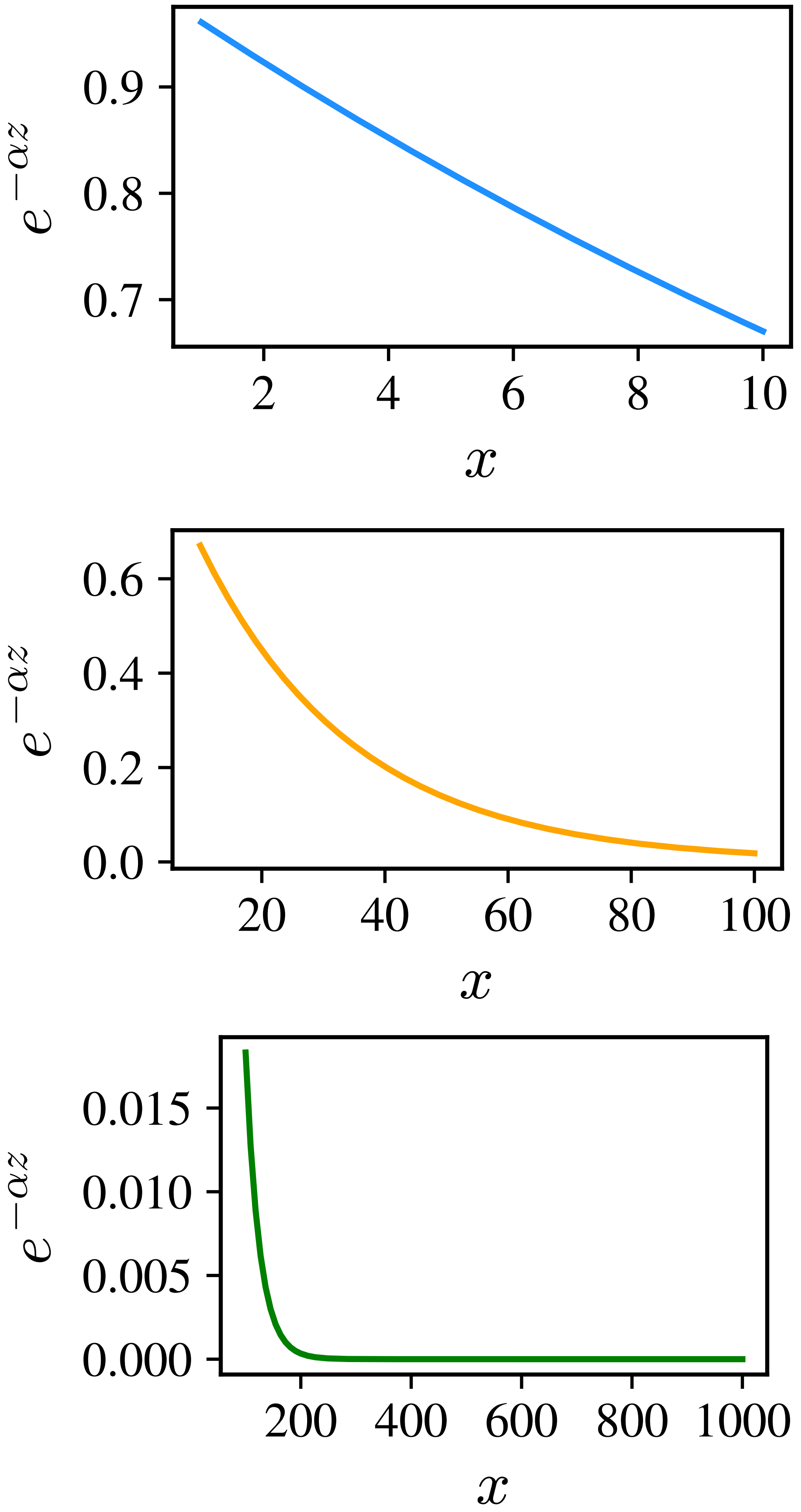
Case of study:

Frequency of words, size of cities, critical phenomena…

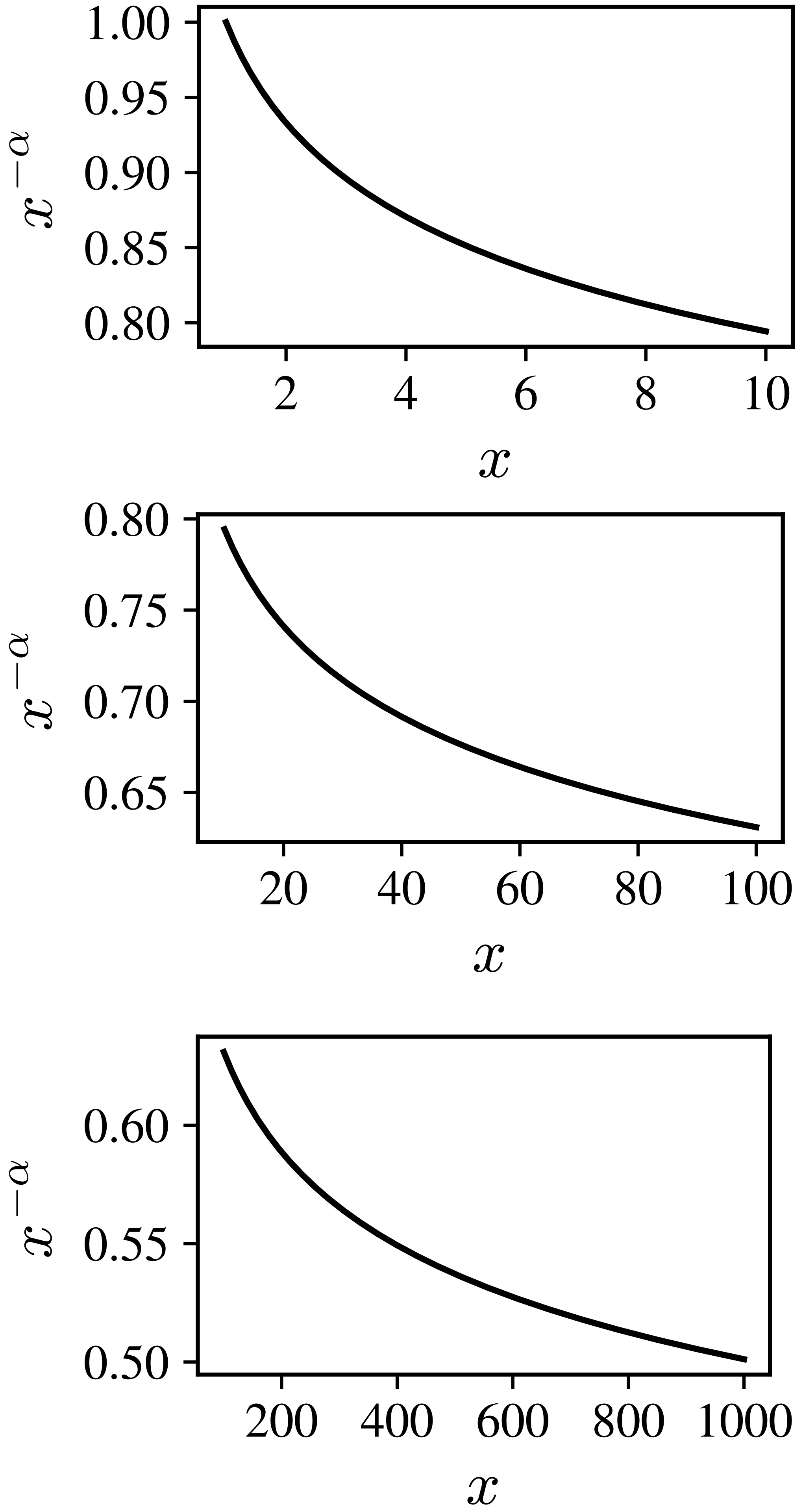
Power-law distribution



Typical scales

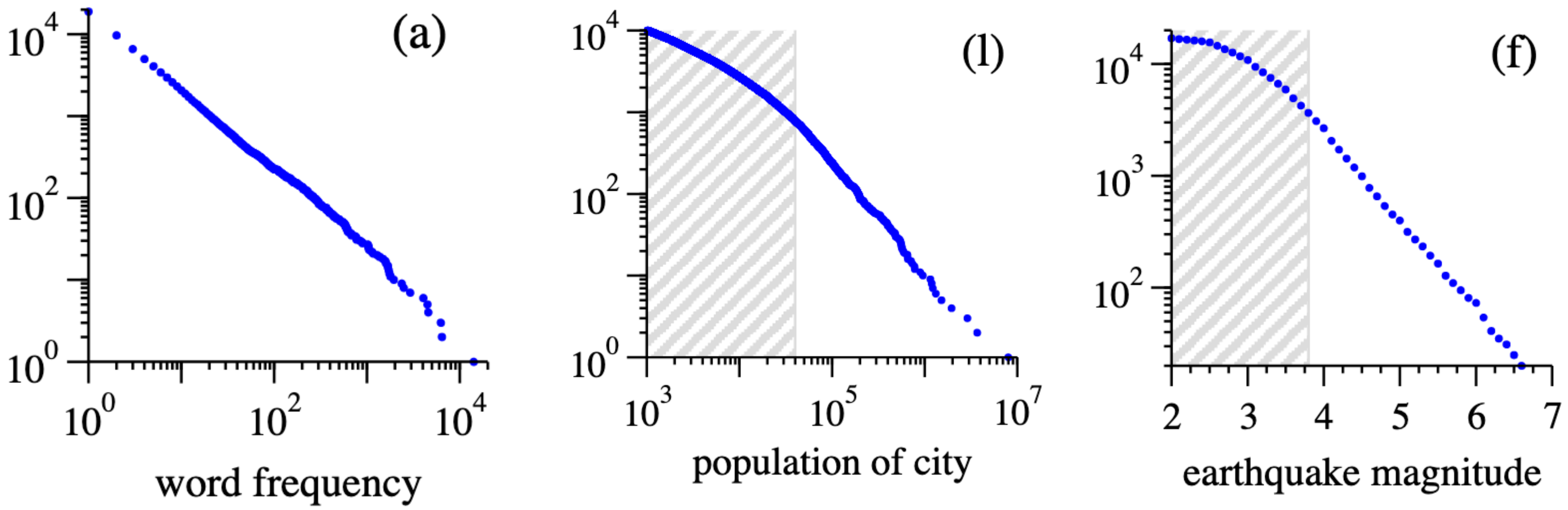


Scale-free

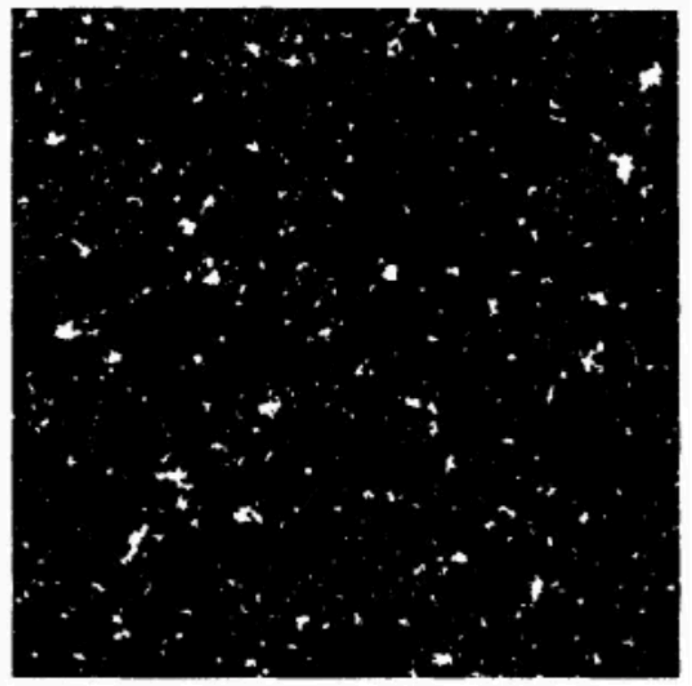


$P(x) \propto x^{-\alpha}$

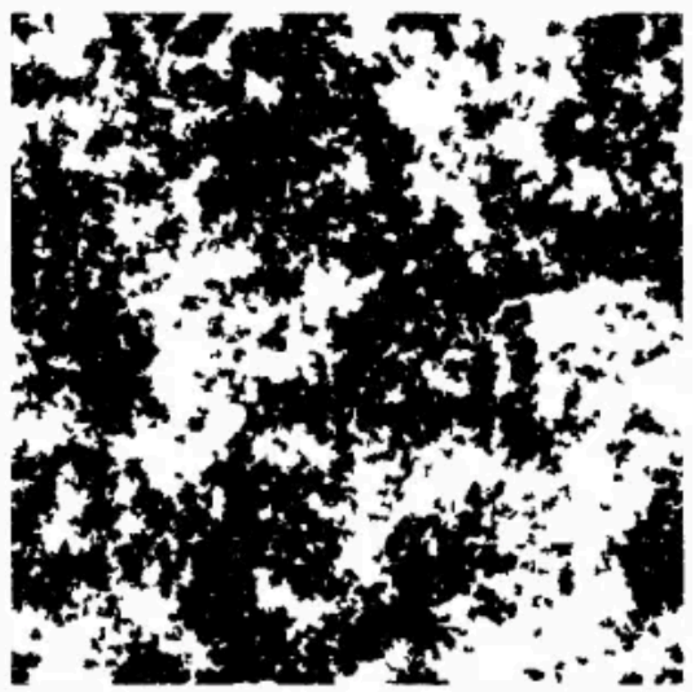
Scale-free property: $P(hx) = g(h)P(x)$



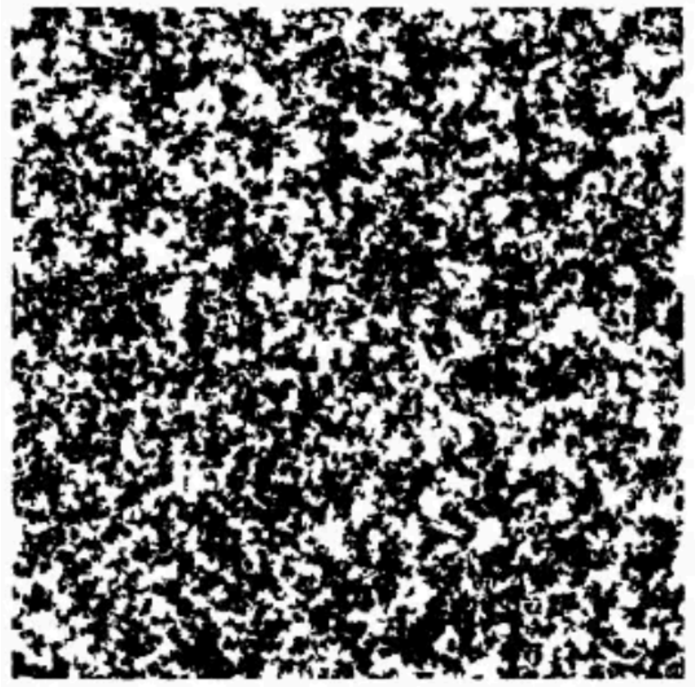
Order



Criticallity



Disorder



Inference

Connecting experiments and models

Karl Pearson
(1857-1936)

Case of study:

German tank problem

Method of moment

Maximum likelihood estimator (MLE)

Data:

In WWII Allies captured n serial numbers of German tanks

$$0 > s_1 > s_2 > \ldots > s_n$$

Q: What was the total number of tanks?

A (Naive): using the uniform distribution

H: It is equally likely to “sample” any serial number.

H: Assuming continuous range.

H: Counts with replacement.

$$s_i \sim U_{[0,N]}$$

$$P(s_i \in [s, s+ds]) = \frac{ds}{N}, \quad s \in [0,N]\,.$$

Method of moments:

$$E(s_i) = \frac{N}{2}$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} s_i \longrightarrow \hat{N}_m = 2\hat{\mu}$$

Maximum likelihood estimator:

$$L\left(s_1, s_2, \ldots, s_n \,|\, N\right) = N^{-n}, \quad N \geq s_n, \quad 0 > s_1$$

$$\ell = \log(L) = -n\log(N)$$

$$\partial_N \ell = -\frac{n}{N} < 0 \longrightarrow \hat{N}_{MLE} = s_n$$

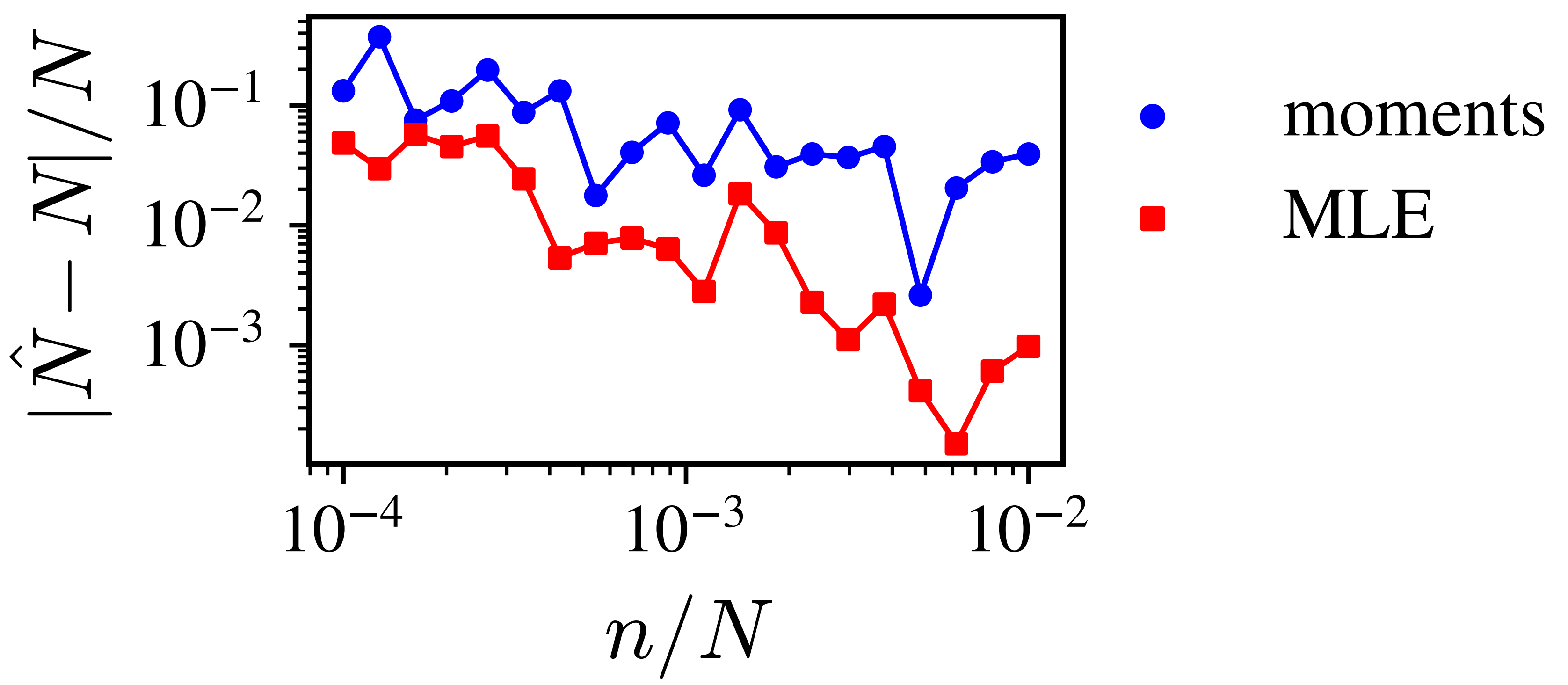
Comparing the methods:

Method of moments is unbiased

$$E(\hat{N}_m) = N$$

MLE is consistent

$$E(\hat{N}_{MLE}) = \frac{n}{n+1}N$$



Inference
Connecting experiments and models

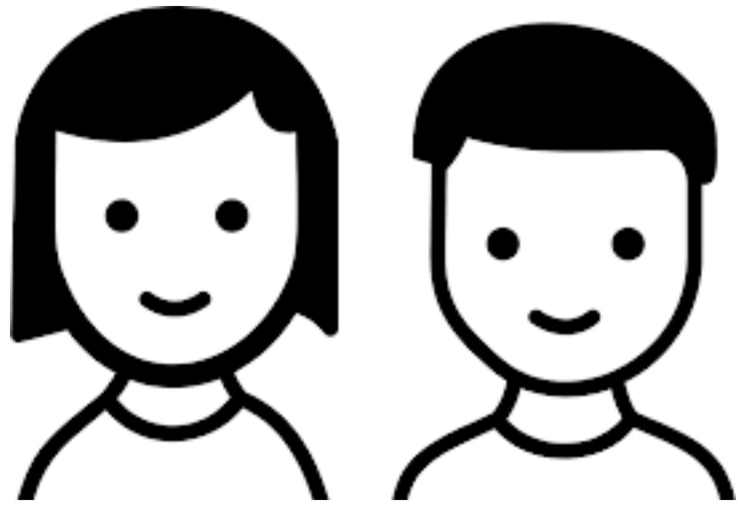
Pierre-Simon Laplace (1749-1827)

Case of study:

Asymmetry in birth rates

Bayesian inference

Beta distribution



Data:

There are more male than female newborns:
Laplace data: 251527 boys and 241945 girls

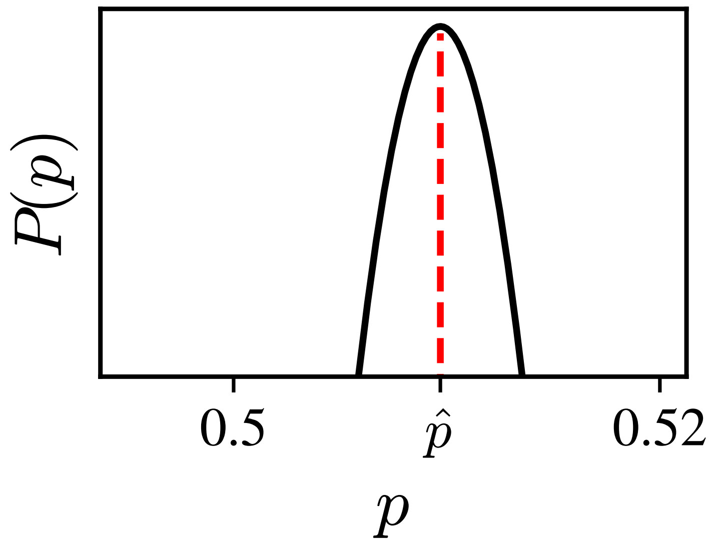
Q: How **robust** is this observation?

$N = N_b + N_g = 493472$

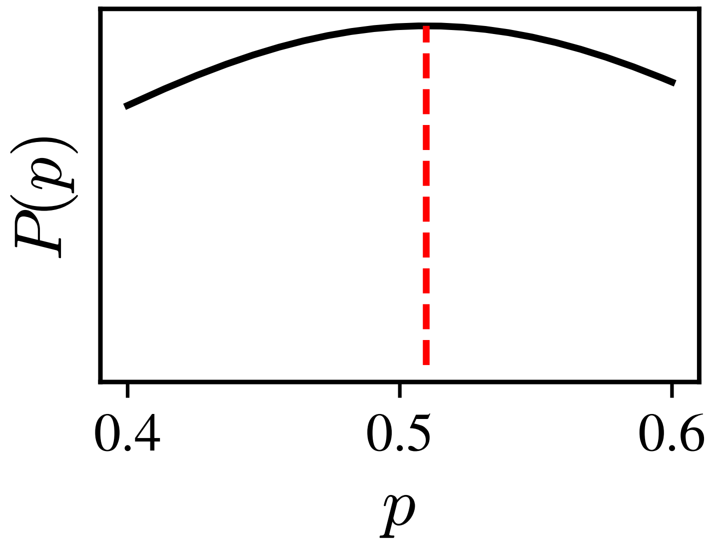
$N_b \sim B(N, p)$

$E(N_b) = Np \rightarrow \hat{p} = N_b/N = 0.5097$

Real feature



Due to fluctuations



Bayes method

Posterior, Likelihood, Prior, Normalization

$$\boxed{P(p\,|\,N_b)} = \boxed{P(N_b\,|\,p)}\frac{\boxed{P(p)}}{\boxed{P(N_b)}}$$

Prior $P(p) = 1, \quad p \in [0,1]$ **Assumed:** based on experimenter's belief.

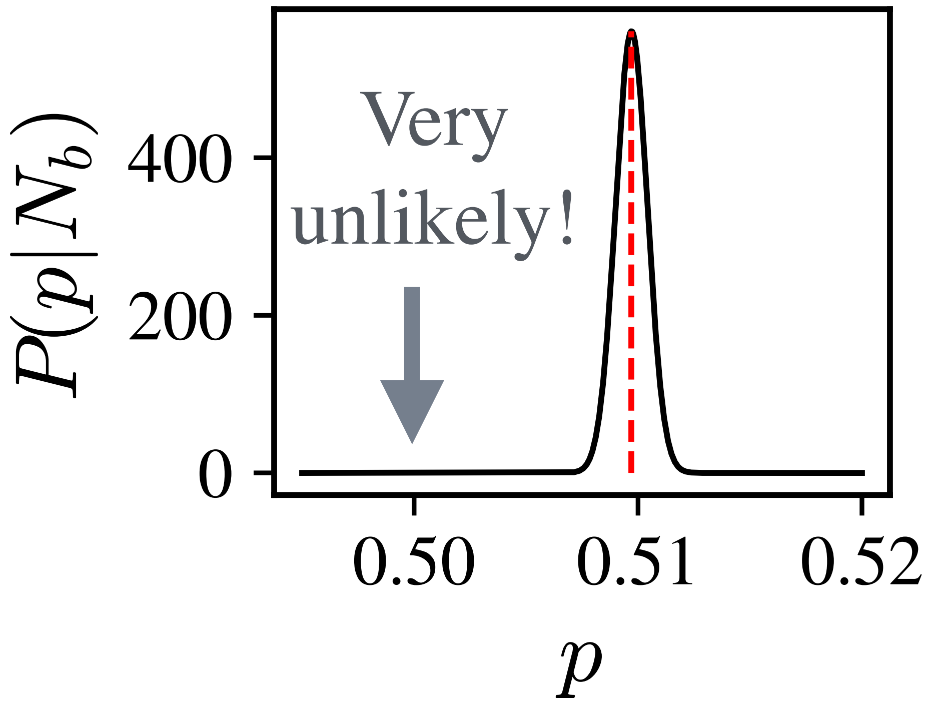
Likelihood $P(N_b\,|\,p) = p^{N_b}(1-p)^{N-N_b}\binom{N}{N_b}$ **Known:** model evaluated on data.

Normalization $P(N_b) = \int dp\, P(p)\, P(N_b\,|\,p)$ **Computed:** independent of p.

Uniform prior (no prior information) $\rightarrow$ $P(p\,|\,N_b) = Z^{-1}\, p^{N_b}(1-p)^{N-N_b}$ (Beta distribution)

We can trust the result!!

$P(p <= 0.5) \sim 10^{-42}$

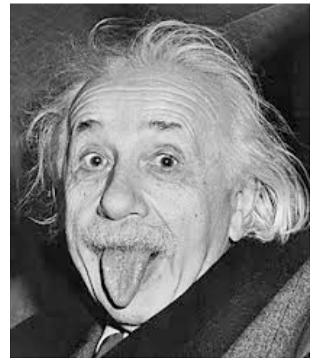


Stochastic processes

Branching process

Stochastic processes

Marian Smoluchowski (1872-1917)



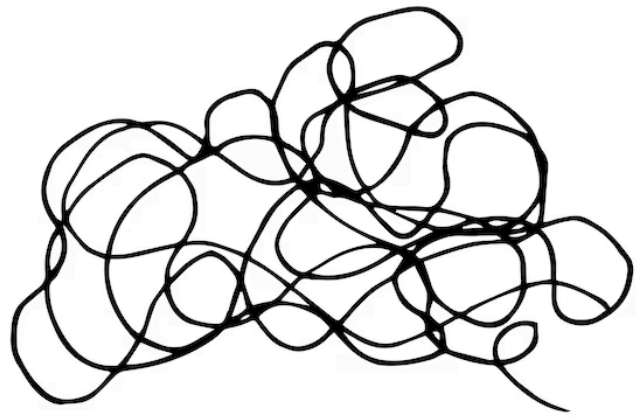
Albert Einstein (1879-1955)

Robert Brown (1857-1936)

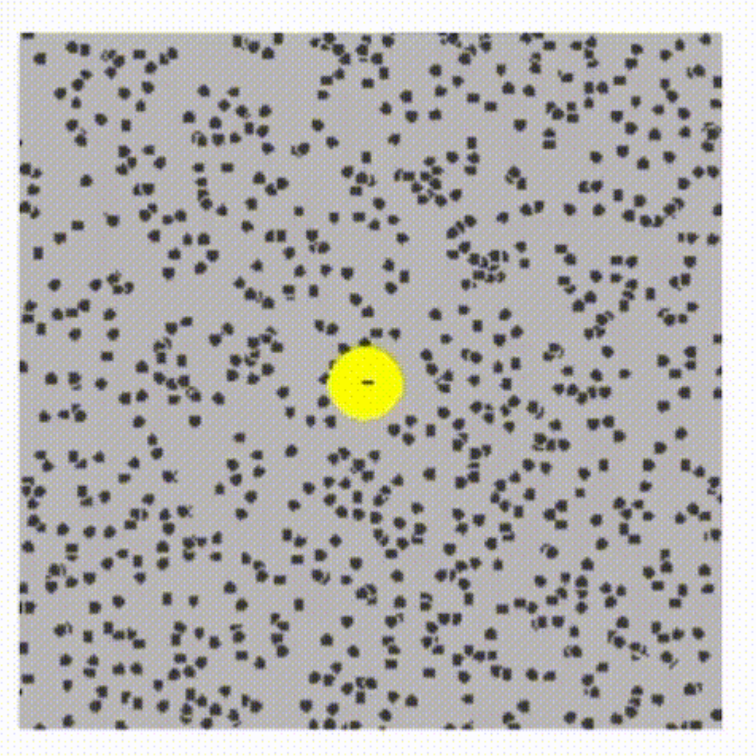
Case of study:

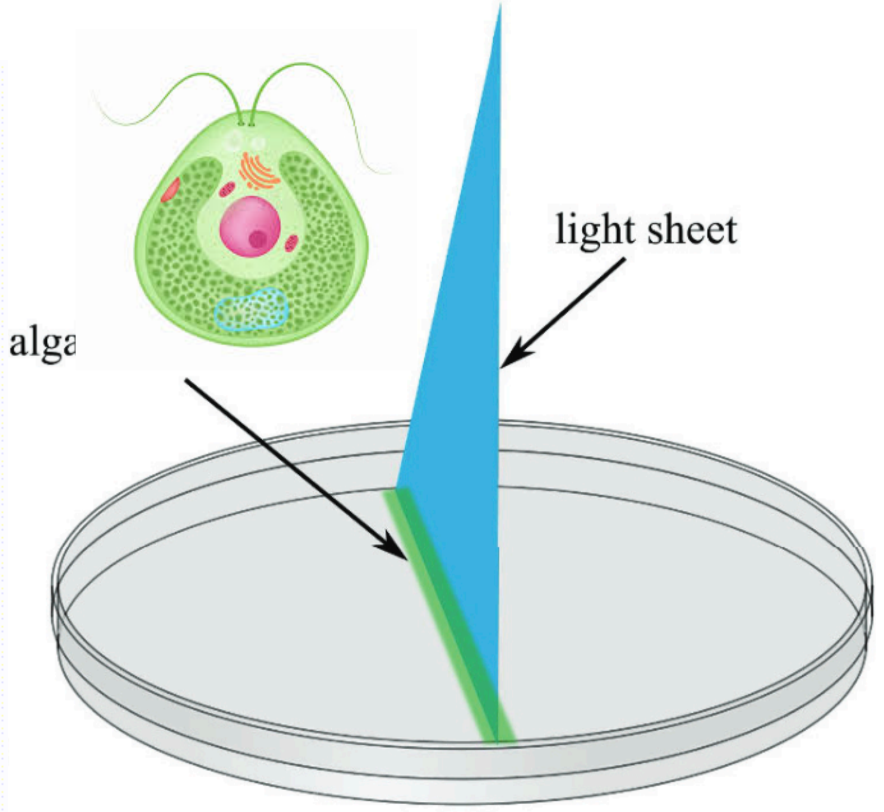
Diffusion, random walks

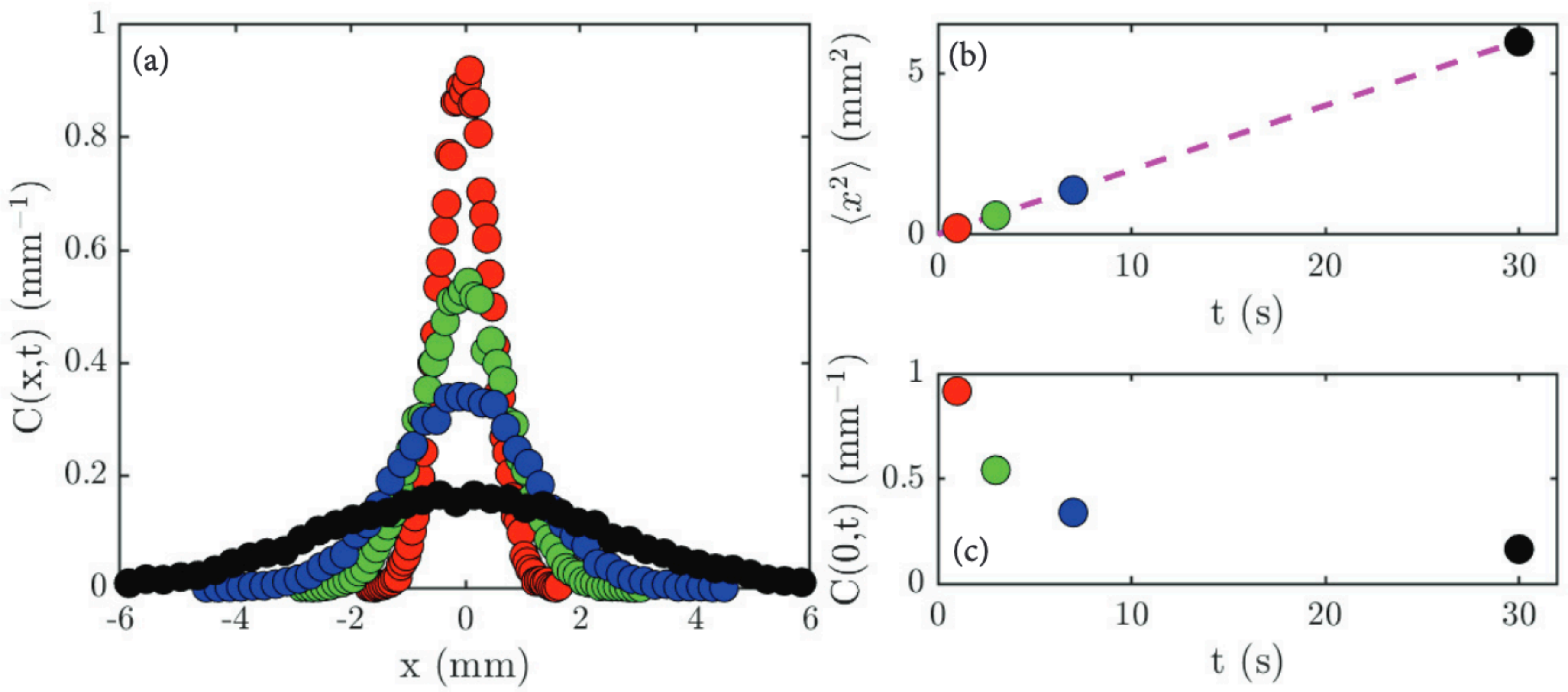
Markov chains, Fokker-Planck eq.



Q: Origin of statistical patterns in erratic movements?

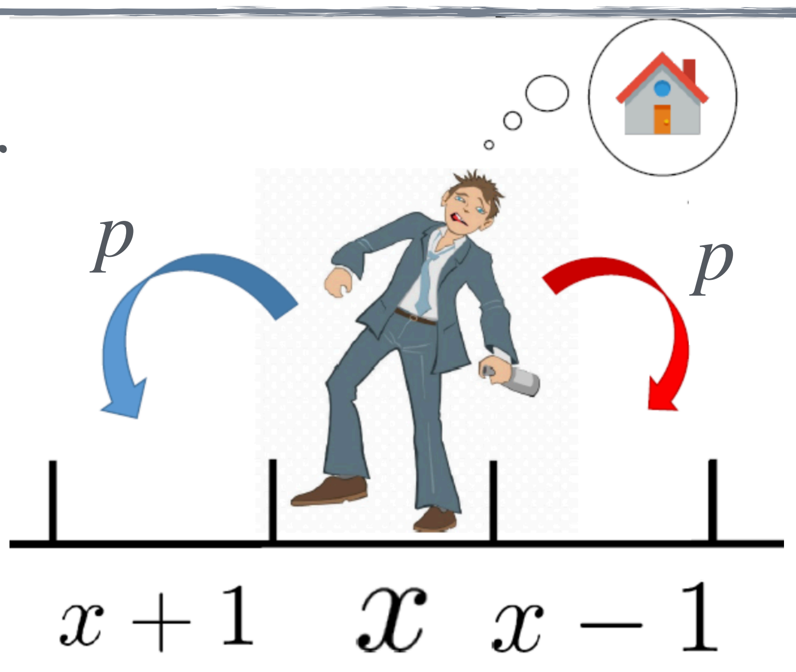






Random walk

$X_t \longrightarrow$ Position of the (drunk) walker



Transition probabilities only depend on current state: **Markov process**

$$p = P(X_{t+\Delta t} = x \pm \Delta x \,|\, X_t = x) \qquad 1 - p = P(X_{t+\Delta t} = x \,|\, X_t = x)$$

$$P_t(x) = P(X_t = x \,|\, X_0 = x_0) = ??$$

$$P_{t+\Delta t}(x) = \sum_y P_t(y) P(X_{t+\Delta t} = x \,|\, X_t = y) = (1-p)P_t(x) + pP_t(x+\Delta x) + qP_t(x-\Delta x)$$

Diffusion eq. Jumping prob. Per unit of time

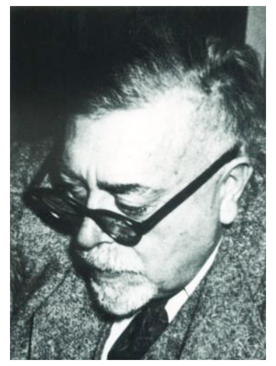
Continuum limit $\longrightarrow$ $\boxed{\partial_t P_t(x) = \frac{D^2}{2} \partial_x^2 P_t(x)}$ $\quad D^2 = \boxed{\frac{p}{\Delta t}}\boxed{\Delta x^2}$ Size of jumps

$$P_t(x) = \frac{1}{\sqrt{2\pi D^2 t}} e^{\frac{(x-x_0)^2}{2D^2 t}} \longrightarrow \begin{array}{l} E(X_t) = x_0 \\ \sigma^2(X_t) = D^2 t \\ \max_x \{P_t(x)\} = t^{-\frac{1}{2}} \end{array} \longrightarrow \boxed{\Delta l \sim D\sqrt{t}}$$

Signature of Diffusion

Stochastic processes

Norbert Wiener (1894-1964)

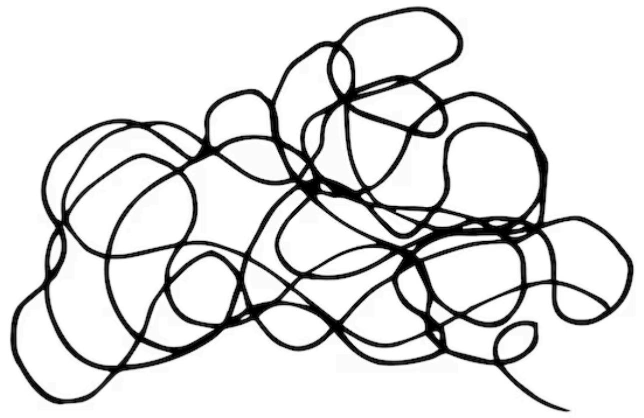


Paul Langevin (1872-1946)

Case of study:

Simulation of SDEs

Langevin Eq., Fokker-Planck eq.



Q: Relation between probabilities and processes?

Uncorrelated increments

$$\dot{X}_t = D\xi(t) \longrightarrow X_t = X_0 + D\int_0^t \xi(s)ds = X_0 + W_t \longrightarrow \boxed{\begin{array}{l} E(X_t) = 0 \\ E\left[(X_t - X_0)^2\right] = D^2 t \end{array}}$$

Signature of Diffusion

White noise/Wiener process

Random velocity =
multiple, fast degrees of freedom

$$v(X_t, t) = \xi(t)$$

$$W_t = \int_0^t \xi(s)ds$$

$$W_t \sim G(0,t) \longrightarrow \begin{array}{l} E(\Delta W_t) = 0 \\ \sigma(\Delta W_t) = t \end{array}$$

Stochastic processes

Branching process

Stochastic processes

Case of study:

Active matter

Brownian motion, Wiener process

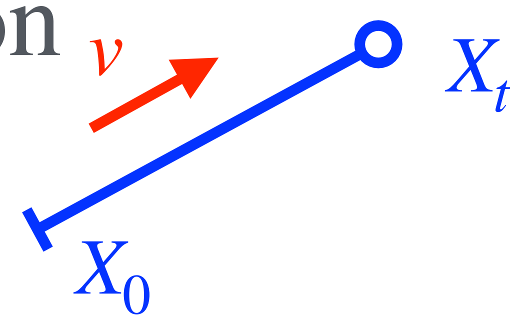
Overdamped dynamics

$$m\ddot{X}_t = -\gamma \dot{X}_t + f(X_t, t) \longrightarrow \dot{X}_t \approx v(X_t, t)$$

No inertia: low mass, big friction, ….

Deterministic motion

$$v(X_t, t) = v$$

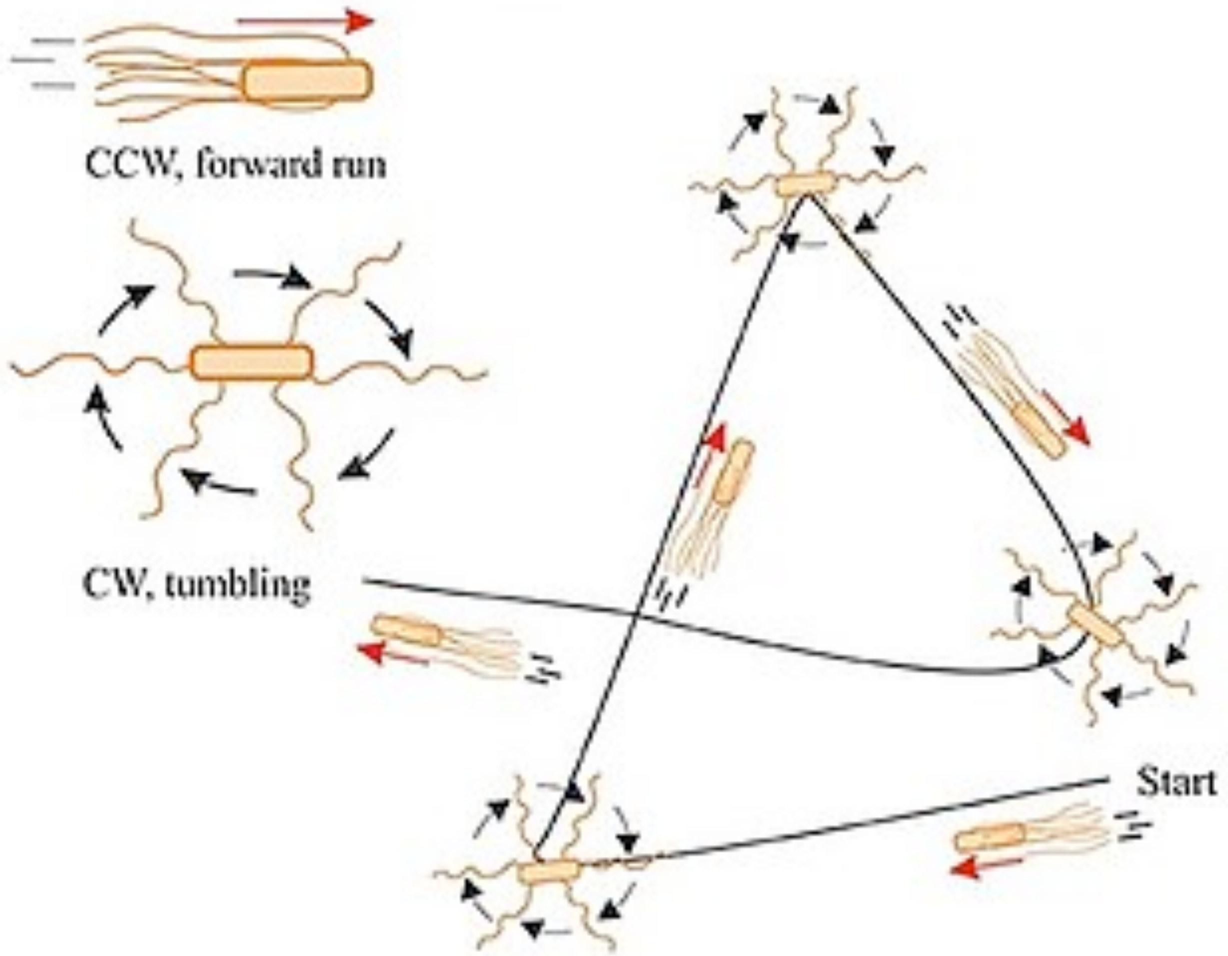


$$X_t = X_0 + vt$$

$$\boxed{E(X_t - X_0) = vt}$$

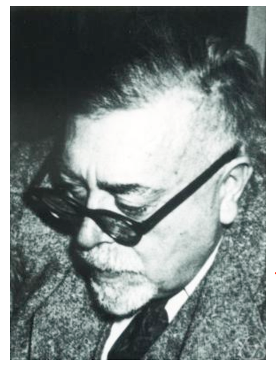
Ballistic regime

Active matter



Stochastic processes

Norbert Wiener (1894-1964)



Robert Brown 1857-1936)

Paul Langevin (1872-1946)

Case of study:

Diffusion, active matter

Brownian motion, Wiener process

Uncorrelated increments

$$\dot{X}_t = D\xi(t) \longrightarrow X_t = X_0 + D\int_0^t \xi(s)ds = X_0 + W_t \longrightarrow \boxed{\begin{array}{l} E(X_t) = 0 \\ E\left[(X_t - X_0)^2\right] = D^2 t \end{array}}$$

Signature of Diffusion

White noise/Wiener process

Random velocity = multiple, fast degrees of freedom

$$v(X_t, t) = \xi(t)$$

$$W_t = \int_0^t \xi(s)ds$$

$$W_t \sim G(0,t) \longrightarrow \begin{array}{l} E(\Delta W_t) = 0 \\ \sigma(\Delta W_t) = t \end{array}$$

Stochastic processes

Epidemic models

Stochastic processes

Inference

Philosophy

Ignorance

Complexity

Intrinsic randomness

Bayes vs. Frequentists

De Finetti

Probability does not exist!

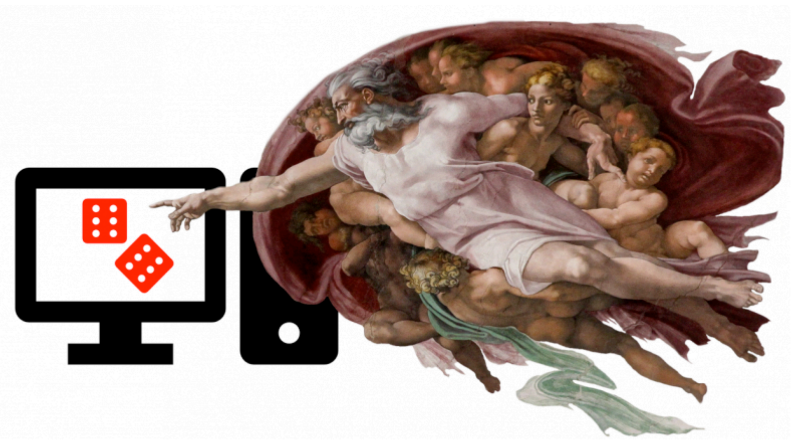
Bibliography

- Blitzstein, J. K.; Hwang, J. Introduction to Probability.
- Roussas, G. G.; Probability and statistics through the centuries.
- Florescu, I.; Probability. A (very) brief history.
- Spiegelhalter D.; The Art of Statistics
- Galton, F. Regression Towards Mediocrity in Hereditary Stature.
- Cole, T. J. Galton's Midparent Height Revisited.
- Newman, M. E. J. Power Laws, Pareto Distributions and Zipf's Law.
- Feynman, R. P. Statistical mechanics: A set of lectures.
- Van Kampen, N.G. Stochastic Processes in Physics and Chemistry.
- Risken, H. The Fokker-Planck Equation: Methods of Solution and Application.
- Wolfgang, P. and Baschnagel, J. Stochastic Processes. From physics to finance.
- Piantadosi, S. T. Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions.
- Feller, W. An introduction to probability theory and its applications. Vol. {II}.
- Casella, G. And Berger R. L. Statistical inference.
- Cocco, S. , Monasson, R. and Zamponi F. From statistical physics to data-driven modeling.
- Goldstein R.E. Are theoretical results 'Results'?

Datasets

- Kaggle Male & Female height and weight
- Galton height dataset

Probability and statistics

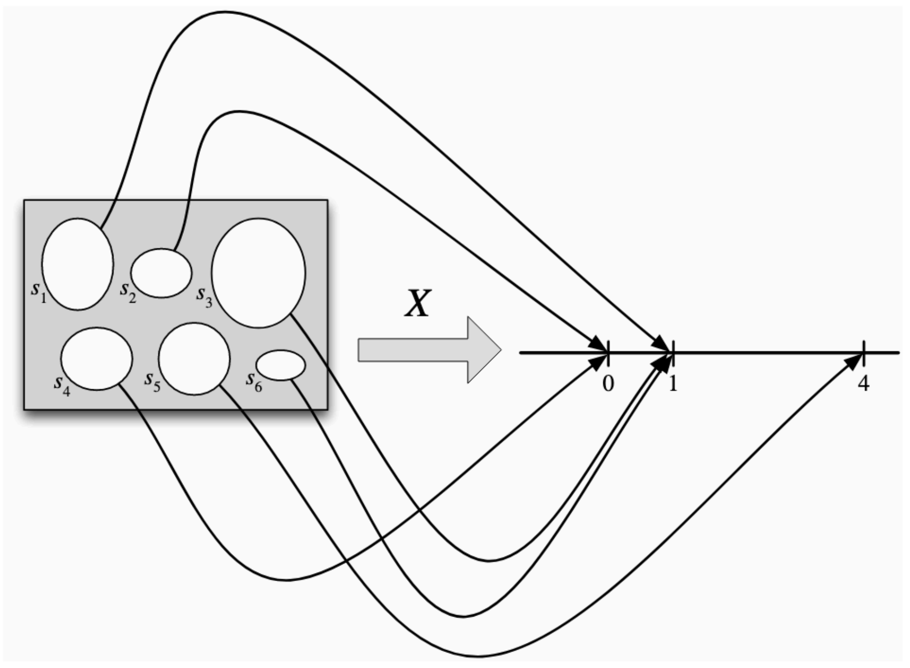


Summary

Statistics

Tools to analyze data

- Visualization of raw data
- Summary statistics (moments and correlations)
- Histograms



Probability

Tools to rationalize statistics

- Sample space
- Random variables
- Probability / probability distributions
- Moments

Perspective: We see data as realizations of a random variable

Relation between estimated probabilities (normalized histograms) and probability distributions

$$\hat{P}_A = \frac{1}{N}\sum_{i=1}^{N} I(X^{(i)} \in A) \qquad P(A) = \sum_x P(X = x)\, I(X \in A) \qquad \hat{P}_A \approx P(A)$$

Relation between moments of data and moments of probability distributions

$$\hat{\mu}_X = \frac{1}{N}\sum_{i=1}^{N} X^{(i)} \qquad E(X) = \sum_x P(X = x)\, x \qquad \hat{\mu}_x \approx E(X)$$

In general, if Z = f(X) $\quad \hat{\mu}_Z = \frac{1}{N}\sum_{i=1}^{N} f(X^{(i)}) \qquad E(Z) = \sum_x P(X = x) f(x) \qquad \hat{\mu}_Z \approx E(Z)$

Probability

Modeling randomness

Usefulness of Gaussian distributions

How moments and probabilities are related to estimators?

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X^{(i)} \longrightarrow P(\hat{\mu} \in [u, u + du]) = G\left(x; \mu_X, \frac{\sigma_X}{\sqrt{N}}\right) dx$$

$$\hat{P} = \frac{frec(A)}{N} = \frac{1}{N}\sum_{i=1}^{N} I(X^{(i)} \in A) \longrightarrow P(\hat{p} \in [p, p + dp]) = G\left(p; p_A, \frac{p_A(1 - p_A)}{\sqrt{N}}\right) dp$$

How to interpret the variance?

$$P(X \in [x, x + dx]) = G\left(x; \mu, \sigma\right) dx \longrightarrow$$

$$P(X - \mu \in [x - \sigma, x + \sigma]) = \ldots$$

$$P(X - \mu \in [x - 2\sigma, x + 2\sigma]) = \ldots$$

$$P(X - \mu \in [x - 3\sigma, x + 3\sigma]) = \ldots$$

Probability

Modeling randomness

Other important distributions: gamma and power law

Gamma distribution

$$\rho(x) = \frac{?}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}$$

Power law distribution

$$\rho(x) = \frac{?}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}$$