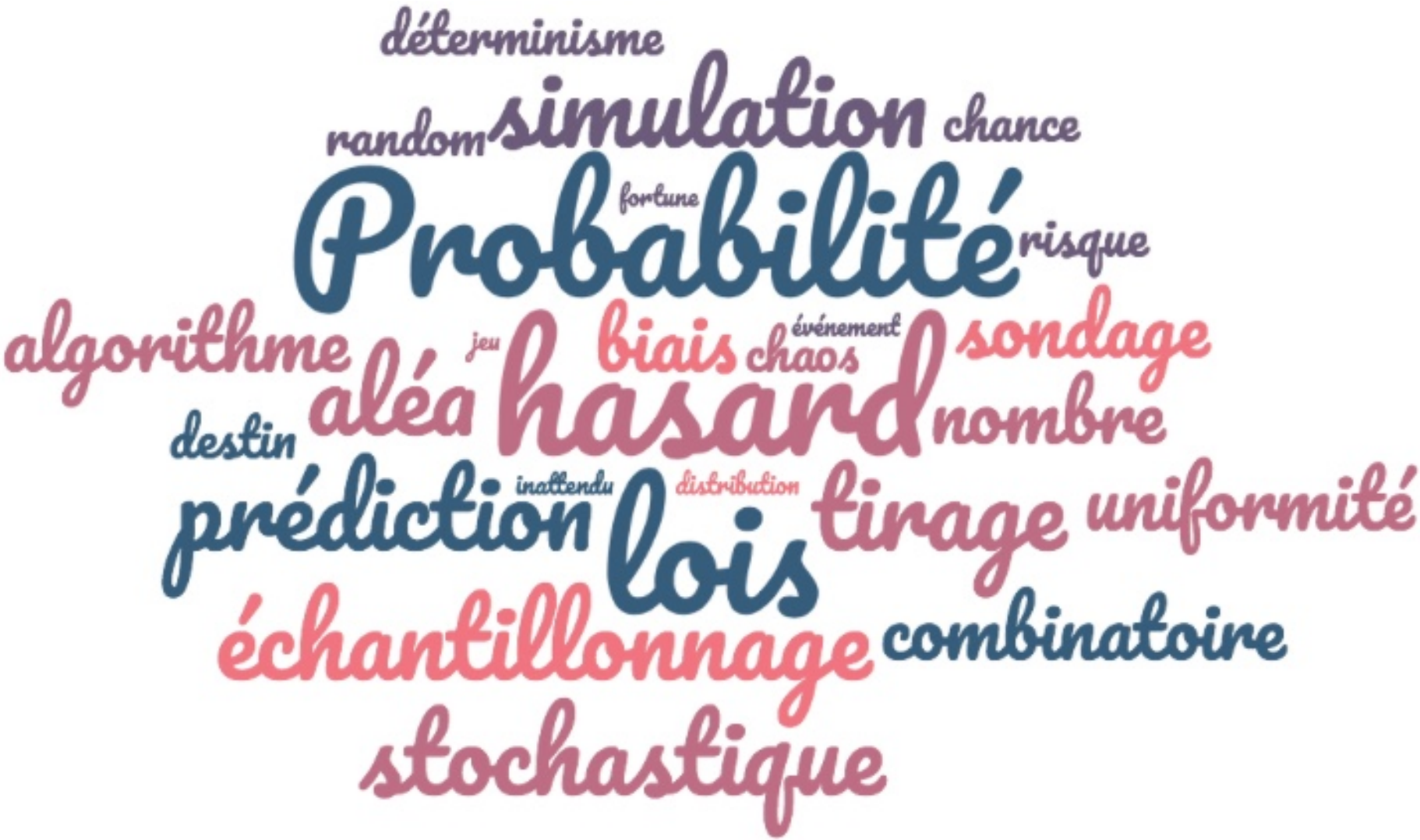
Probabilités et Simulation

Florence Perronnin (UGA) & Arnaud Legrand (CNRS)

Université Grenoble Alpes

September 7, 2018


déterminisme
random
simulation
chance
Probabilité
fortune
risque
algorithme
jeu
biais
chaos
événement
sondage
aléa
hasard
destin
nombre
prédiction
inattendu
distribution
lois
tirage
uniformité
échantillonnage
combinatoire
stochastique

Outline

https://www.youtube.com/watch?v=vEvHwO6fd3c

Challenger as a probability problem #1

It appears that there are enormous differences of opinion as to the probability of a failure with loss of vehicle and of human life. The estimates range from roughly 1 in 100 to 1 in 100,000. The higher figures come from the working engineers, and the very low figures from management. What are the causes and consequences of this lack of agreement? Since 1 part in 100,000 would imply that one could put a Shuttle up each day for 300 years expecting to lose only one, we could properly ask "What is the cause of management's fantastic faith in the machinery?" R.P. Feynman, Rogers commission [1]

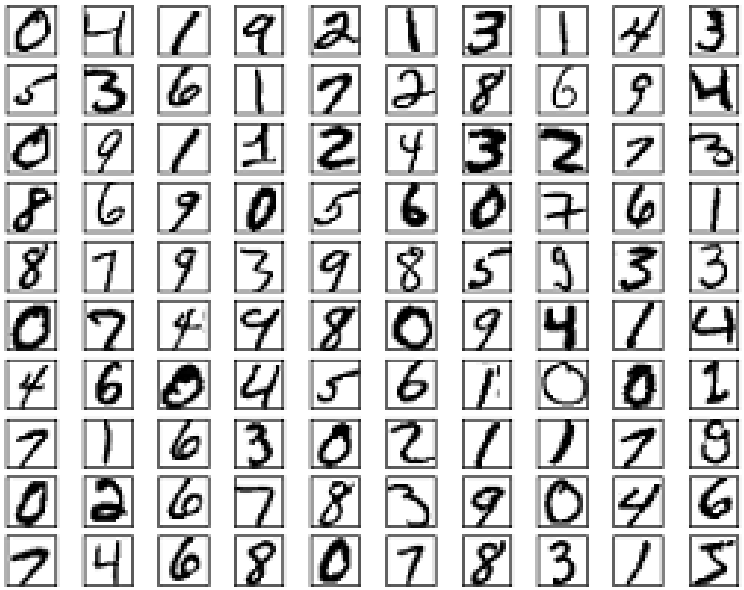
[1] "Report of the Presidential Commission on the Space Shuttle Challenger Accident". NASA. 1986-06-06.

Challenger as a probability problem #2

- O-ring concerns since 1977 (launch in 1986)
- No test data below 40F degrees to support risk assessment

https://inldigitallibrary.inl.gov/sites/sti/sti/3901032.pdf
https://en.wikipedia.org/wiki/Space_Shuttle_Challenger_disaster

Handwritten character recognition



MNIST database

Training data:

- 1 digit is a 28×28 pixels image (State space $\mathcal{S} = \{0, 1, \dots 255\}^{784}$)
- unknown underlying distribution $\mathbb{P}[]_{data}$

Probability model:

- Probability distribution $\mathbb{P}[]_\theta : \mathcal{S} \to [0, 1]$

- Hypothesis on $\mathbb{P}[]_\theta$ (e.g.: gaussian distribution)
- Maximize likelihood to estimate θ (parameters)

THE PROBLEM WITH AVERAGING STAR RATINGS [2]

[2] https://xkcd.com/937/

Quantifying certainty

The weather forecast: "Avalanche risk 3 (on a scale of 5)". An avalanche did occur. Was the forecast right?

In a maternity hospital, 44% of babies are girls. What happened?

Biased coin?

Ten trials to test a coin. 1 1 1 1 1 0 1 0 1 0
Is the coin fair?

Statistical multiplexing

Office space is costly in Manhattan. Your company decides to share desks.

For a 200-person company where people show up with probability 85%[a], how many seats do you need?

[a] 10% vacation and 5% sick leave

What is the probability that a worker arrives and finds all seats occupied?

Other engineering problems

- A/B testing
- Lock-in effet on market shares
- Packet routing algorithms in sparse networks
- Junk e-mail classification
- Security issues (unpredictability)
- Miller-Rabin primality test
- Complex software validation

As well as real-life problems

- gambler's fallacy
- poll results
- choosing a medical treatment

Outline

Probability literacy

- fluency in probability modeling language
- random sampling of classical distributions
- safely use RNGs
- critical thinking when reading statistics ("margin of error")
- design and analyze randomized algorithms
- extract relevant information from experimental data
- provide prediction accuracy

Target Skills

- Knowledge of basic tools:
  - Conditional probability and Bayes formula
  - Classical distributions
  - Mean and variance
- Randomness testing
- Sampling algorithms
- Use of `R` language
- Confidence intervals
- χ^2 test
- Maximum likelihood estimation

Outline

Statistics...

- En essayant continuellement, on finit par réussir. Donc plus ça rate, plus on a de chances que ça marche. Les Shadocks
- `http://images.math.cnrs.fr/IMG/jpg/3.jpg`

Probability or statistics?

Goal

Reconcile probability, statistics, simulation... and information sciences.

A simple experiment

Let us throw a die.

What is the probability the result was a 6?

- 1/6
- 0
- 1
- Depends on previous outcomes?

Probability model

Here, probability is used to model an unknown result.

$\Rightarrow$ Probability values do not only depend on the events but on the available information on the same event.

Another paradox

Mr and Mrs Smith

They have 2 kids. One is a girl. What is the probability that the other one is a boy?

- 1/2 ?
- 3/4 (no information)
- 1/2 revisited
- 0 or 1

There is no "true" probability. Only valid computation depending on available information.

Parametric estimation

Problem : we have some observed data $(x_1, \dots, x_n)$ (for instance: student height in cm) and we want to extract some information from it.

Probabilistic model

Statistics rely on the assumption that observed data is a realization of random variates with unknown (theoretic) distribution p_θ depending on unknown parameter(s) θ.

"(...)As if the observed sample was artificially sampled from p_θ" [3]

Statistics method

Estimate θ : find a number (or vector) $\hat{\theta}$ such that $p_{\hat{\theta}}$ explains best the observed data.
Need to make assumptions on the shape of p_θ.

[3] B. Ycart, SMEL

Simulation

Probabilistic simulation

Generate (pseudo)-random samples drawn from a probabilistic model of a (possibly non-random) phenomenon.

But then statistics are needed to make use of the sampled data...

Outline

Course overview

- Probability basics
  - Probability modeling
  - Independence
  - Conditional probability, Bayes
  - Distribution function
  - Must-know distibutions (geometric, binomial, exponential, gaussian...)
  - Mean and variance
- Random number generation
- Estimation and data analytics
  - KL divergence
  - Maximum Likelihood
  - Central limit theorem
  - χ^2 test

Possibly more, depending on what you want to learn...

Outline

Requirements and grading

"Do it yourself"

- Trying the math (on your own)
- Programming (on your own)
- Reading and researching
- Ask questions (and ask more)

Grading:

- mid-term 25%
- project 25%
- final exam 50%

Let's start

- Start with installing `R` and `RStudio`: `http://mescal.imag.fr/membres/arnaud.legrand/teaching/2017/RICM4_PS.php`
- Get used to R with this tutorial `http://swcarpentry.github.io/r-novice-gapminder/01-rstudio-intro/`

Reading

- Jean Bérard, Introduction aux probabilités et à la statistique. `http://math.univ-lyon1.fr/~jberard/cours-www.pdf`
- Page de J.P. Delahaye, `http://www.lifl.fr/~jdelahay/` et sa chronique "Logique et Calcul" dans *Pour la Science*.
- Ekeland, I. (1991), Au hasard, Le Seuil
- Michael Mitzenmacher and Eli Upfal, Probability and Computing. Cambridge University Press New York, 2017.
- R. Motwani et P. Raghavan, Randomized Algorithms, Cambridge University Press, 1995.
- Stochastic simulation S.M. Ross, Wiley
- Shannon, C. (1948), A mathematical theory of communications, Bells Systems Technical Journal.
- Page de Bernard Ycart, `https://ljk.imag.fr/membres/Bernard.Ycart/`