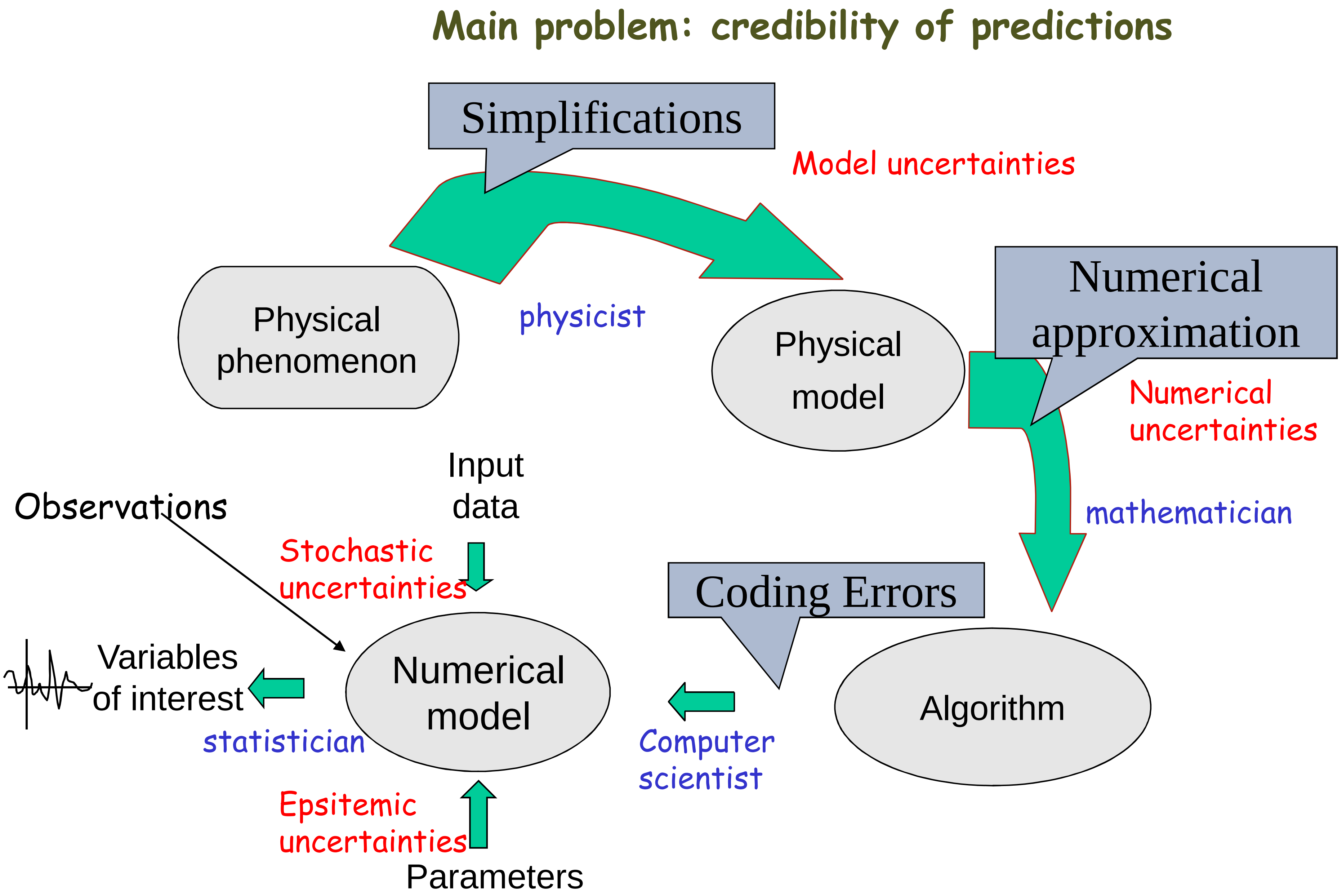
Treatment of uncertainties in numerical simulation

Fabrice Gamboa
Bertrand Iooss

07/03/2013

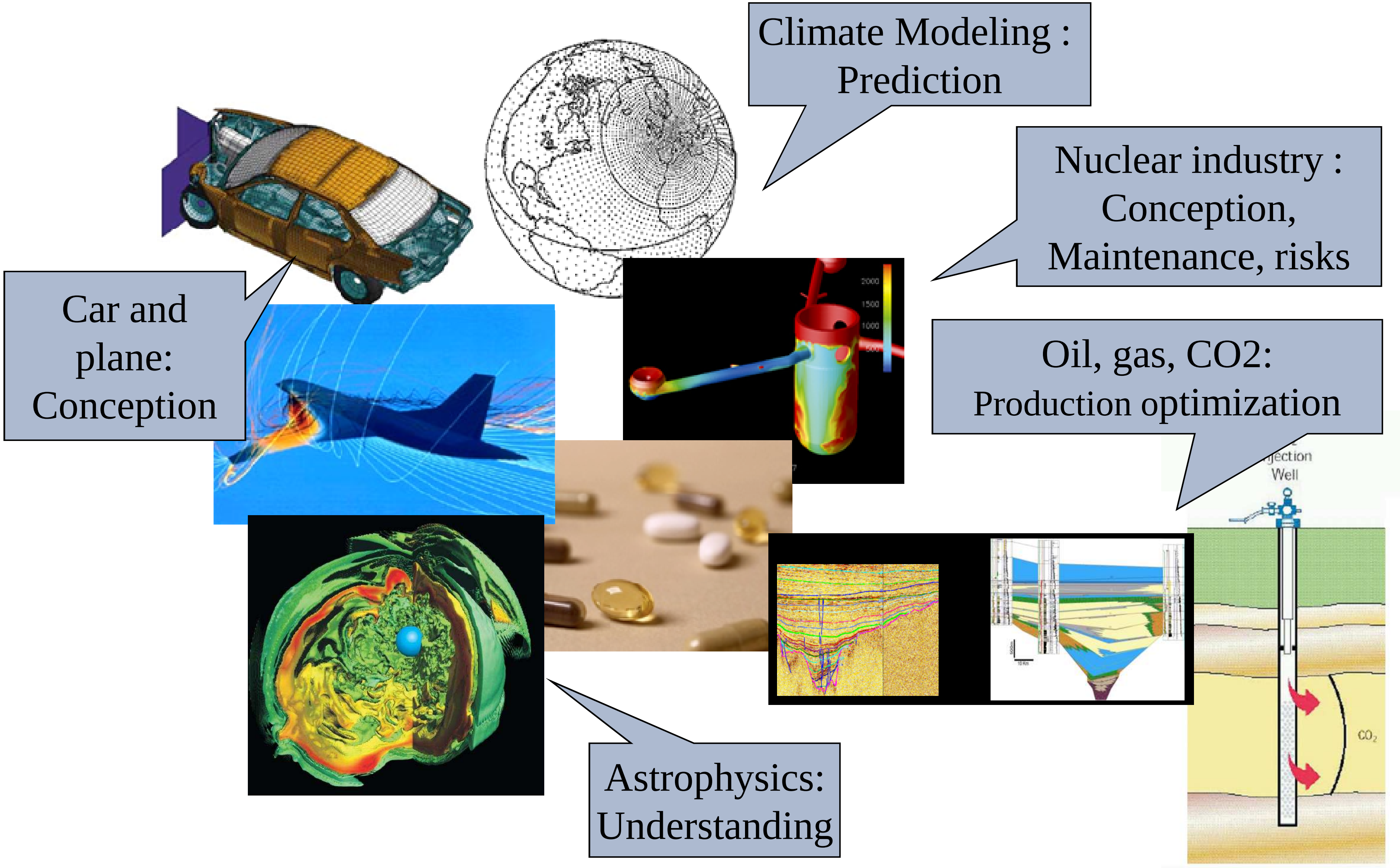
Introduction

Starting point: uncertainties everywhere in a modeling chain !



Similar safety and uncertainty issues in CS&E and Nature sciences

CS&E : Computational Science & Engineering



Exemple 1: particle dispersion in atmosphere (1/3)

Accidental scenario of pollutant release

Domain of study: 10 km around an industrial site

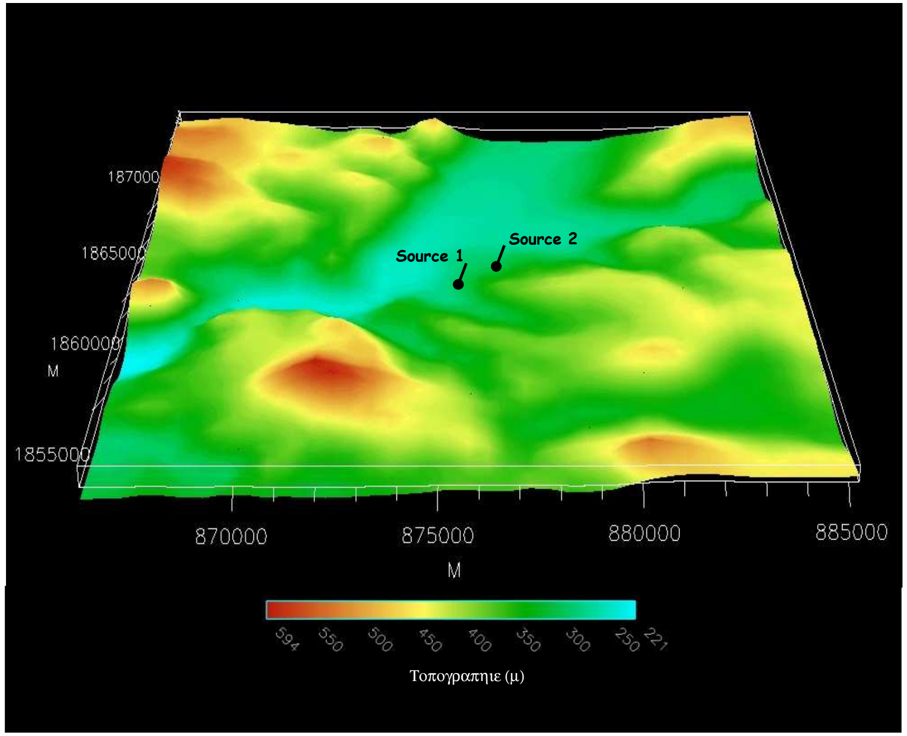
2 arbitrary sources (at ground level) :
* source 1 : tracer (gas)
* source 2 : iodine (particles)

Projection for 4 days

Meteorological data: wind, temperature, humidity, rain

Rugosity of the ground (vegetation)

Topography and location of sources

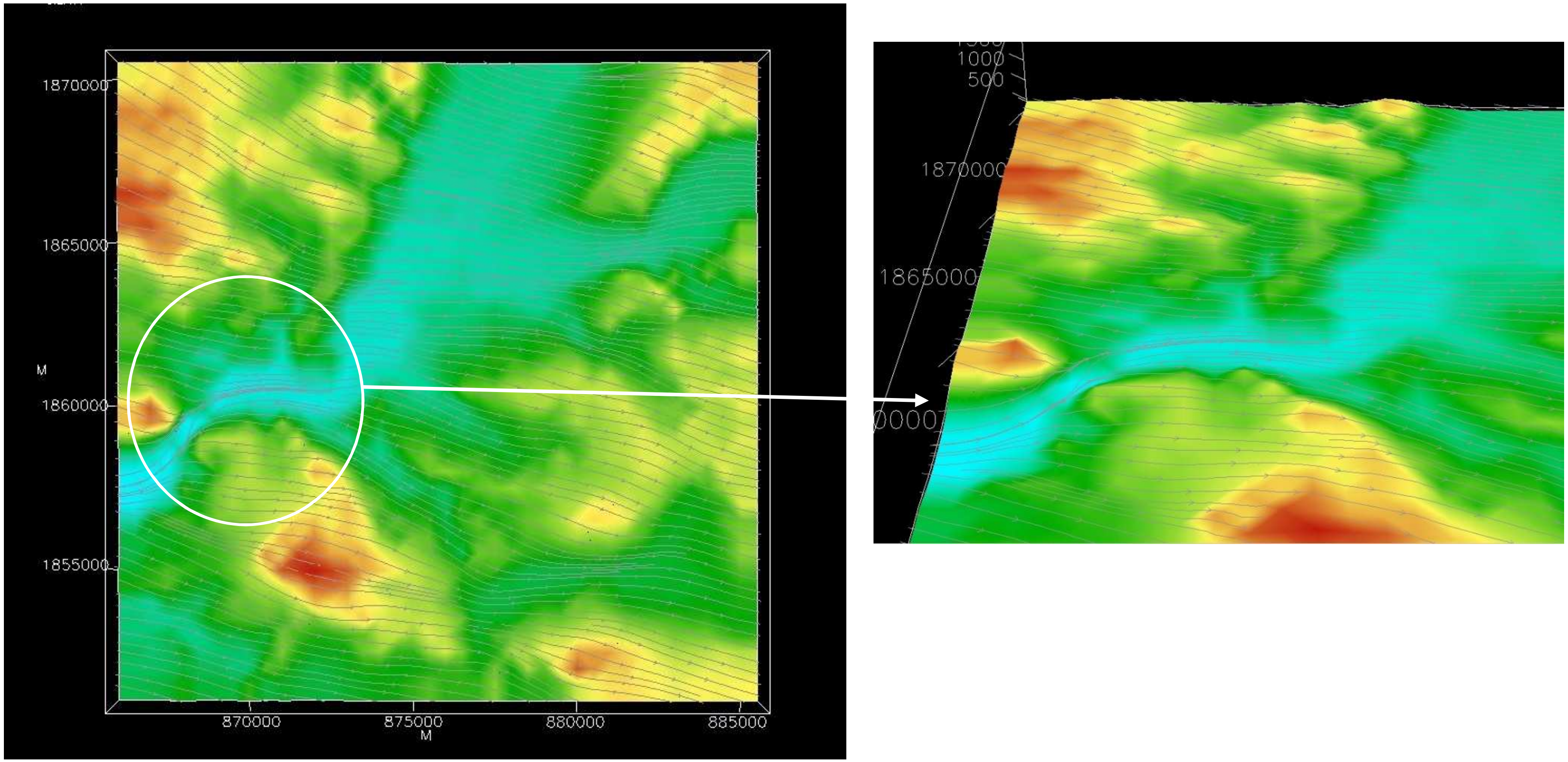


[Source : CEA]

Exemple 1: particle dispersion in atmosphere (2/3)

Computation of wind field (direction and amplitude)

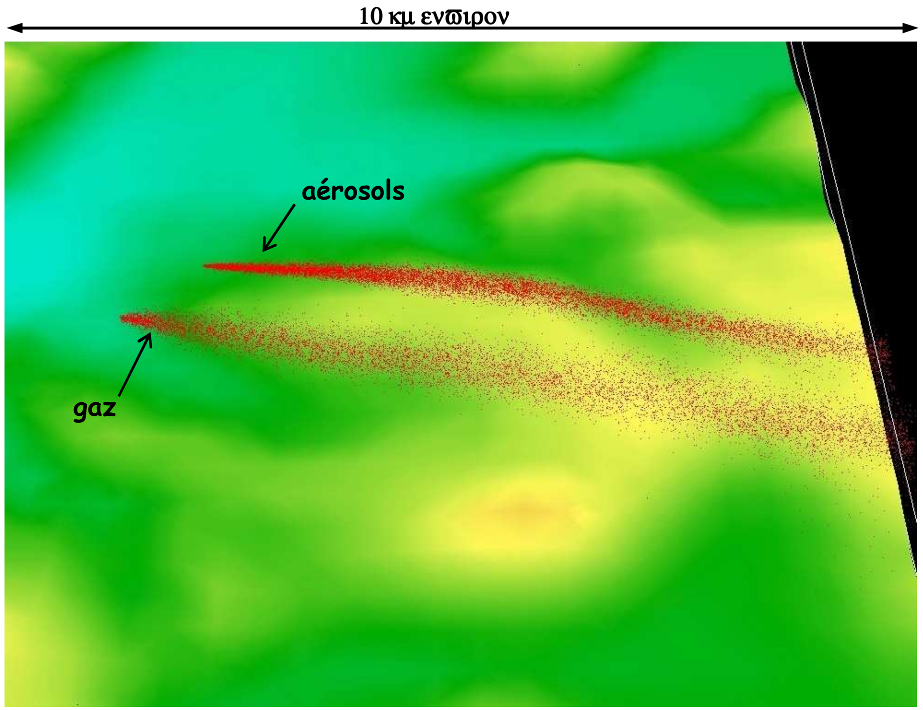
Visualization of the wind with flux lines



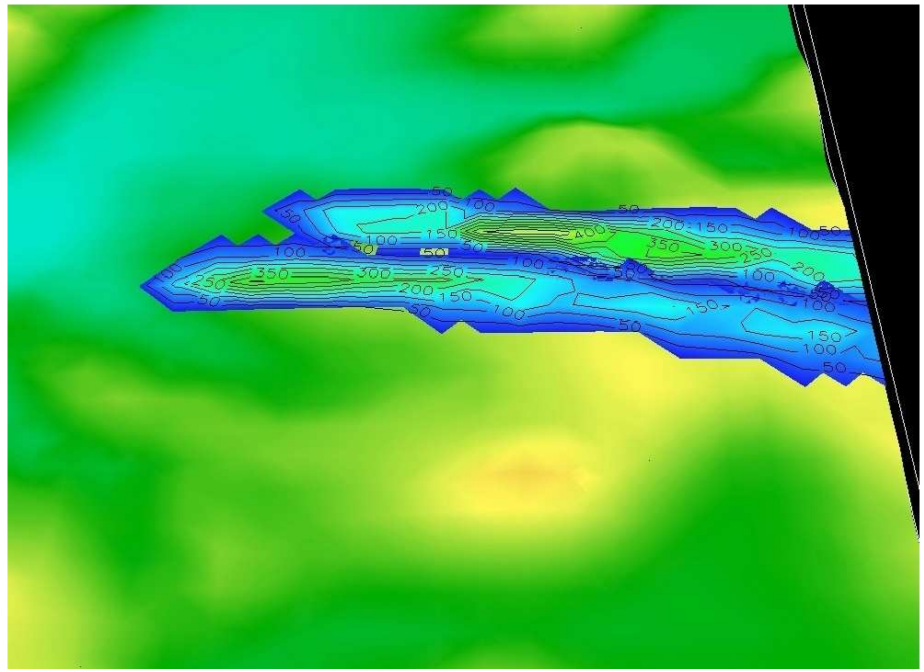
Exemple 1: particle dispersion in atmosphere (3/3)

Use of a **computer code** of lagrangian particle dispersion
(solving the Euler equations of fluid mechanics)

Visualisztion of gas concentrations en gaz after a 5 hours' release



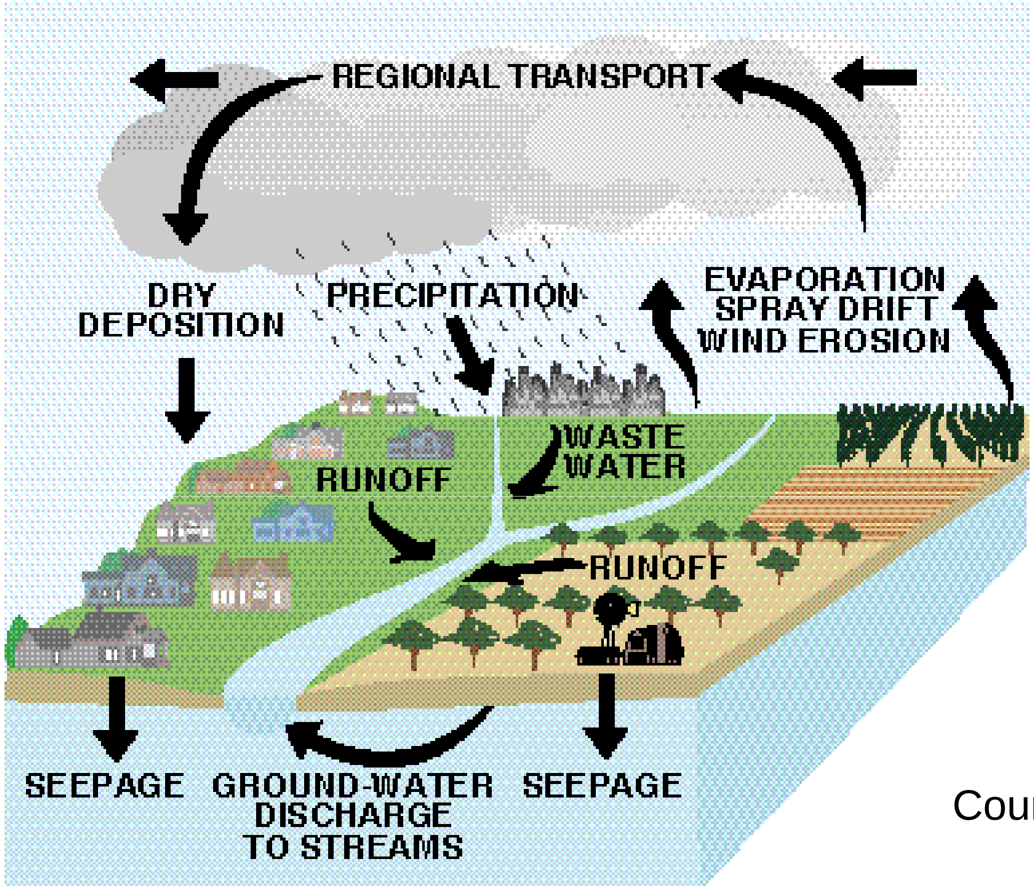
Plume under the particular form



Concentration plume (at 10 m level)

Results are strongly sensitive to meteorological data

Exemple 2: Models in hydrology

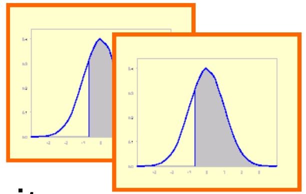


Courtesy of USGS

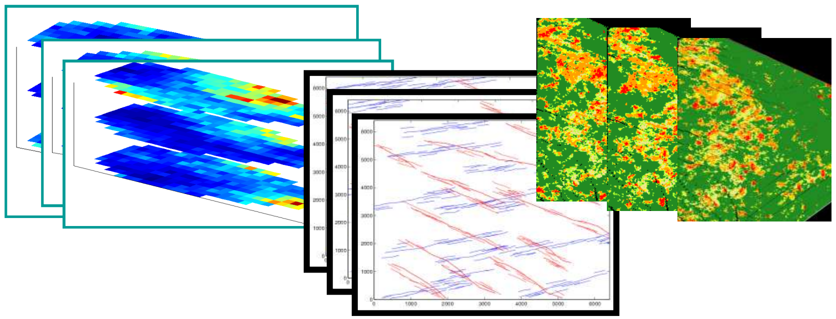
Uncertainties in model parameters that govern surface and ground water transport, …

Exemple 3: Uncertainties in oil reservoir characterization

- Scalar uncertain parameters :
  - Reservoir Geometry : limits, thickness, faults, etc...
  - Petrophysical properties : porosity, permeability,...
  - Fluid properties water/Oil/Gaz : contacts between fluids, viscosity,..
  - Rock/Fluid interactions, Well Data, etc...

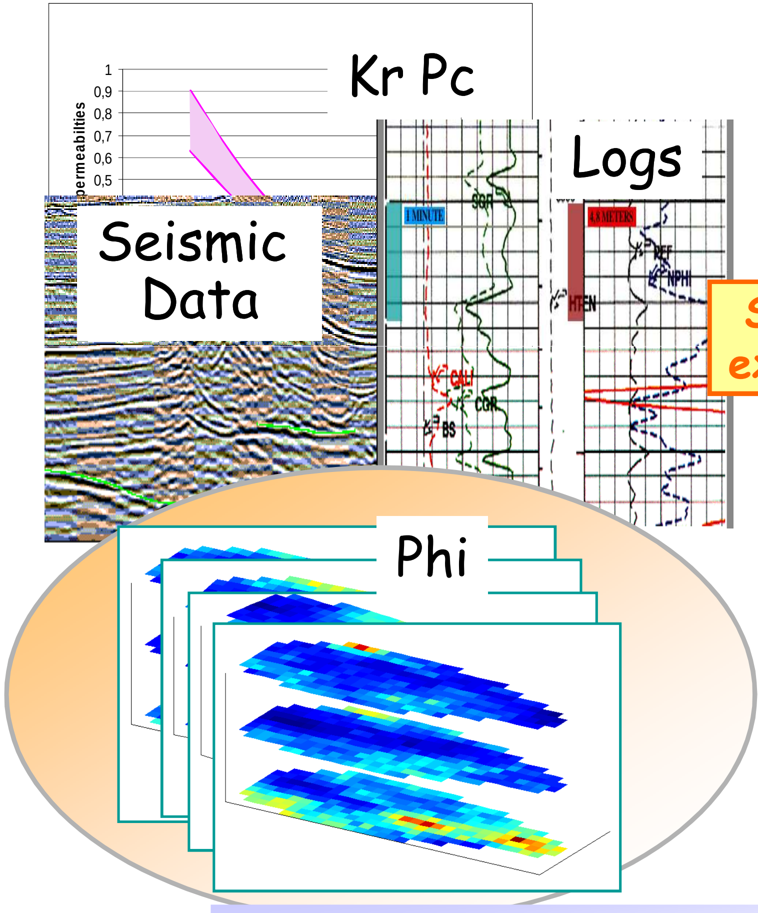


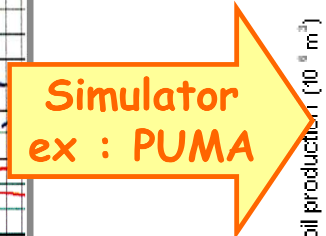
- Spatial uncertain parameters :
  - several realizations of a unique geological structure
  - geostatistical parameter $\Rightarrow$ represented as a "seed variable"
  - Exemples:
    - geostatistical seed
    - Structure maps
    - Stochastic fracture networks



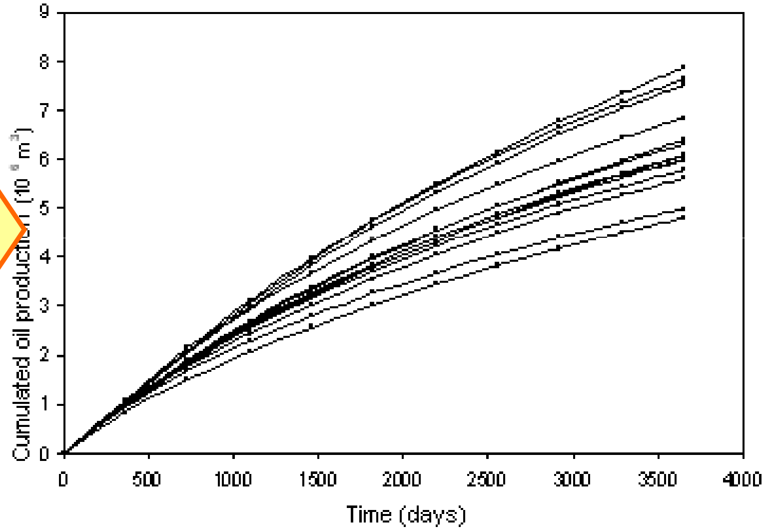
[Source : IFP EN]

Effect of geostatistical uncertainties ?





Uncertainty on Production Forecasts



Uncertainty on realizations

➔ How to characterize this effect of geostatistics ?

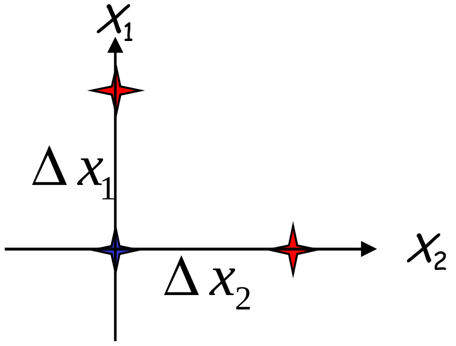
Main stakes of uncertainty management

- **Modeling phase:**
  - Improve the model
  - Explore the best as possible different input combinations
  - Identify the predominant inputs and phenomena in order to priorize R&D
- **Validation phase:**
  - Reduce prediction uncertainties
  - Calibrate the model parameters
- **Practical use of a model:**
  - **Safety studies:** assess a risk of failure (rare events)
  - **Conception studies:** optimize system performances and robustness

Uncertainties in simulation experiments $Y = a_1 x_1 + a_2 x_2$

Ancient way

$$\Delta Y = a_1 \Delta x_1 + a_2 \Delta x_2$$



Still learned in Schools

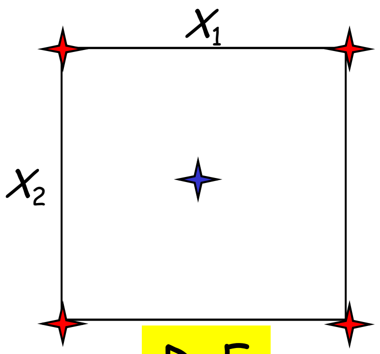
Pre-modern way

x's identified to R.V.
… but same algebra

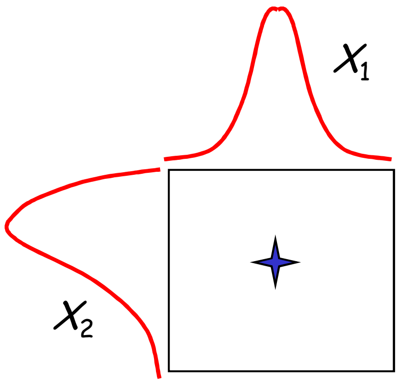
$$\sigma_Y = \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2}$$

Still used in metrology (GUM)

Really Modern way



DoE



Sampling

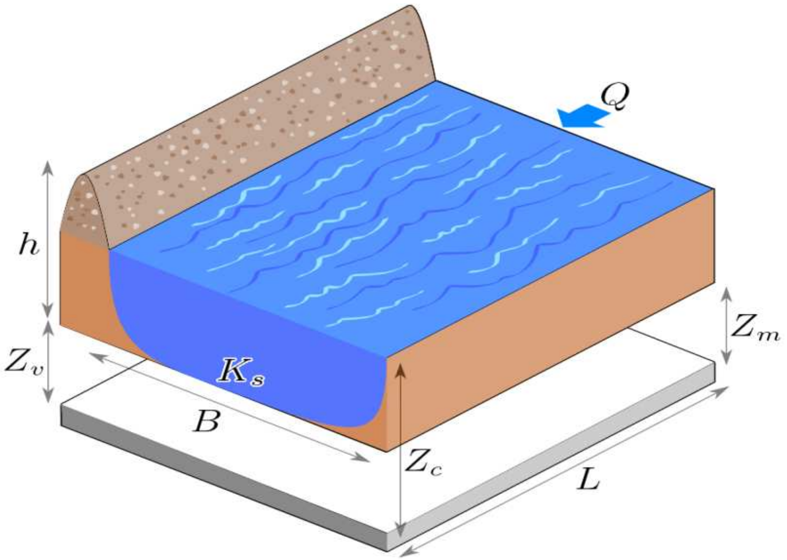
x's fully treated as R.V.

Can give moments, quantiles, and even pdf of Y …
…if fair waiting time

Which (parametric) uncertainty sources?

- Epistemic uncertainty
  - It is related to the lack of knowledge or precision about a parameter which is deterministic in itself (or can be considered deterministic under some accepted hypotheses). E.g. a characteristic of a material.
- Stochastic (or aleatory) uncertainty
  - It is related to the real variability of a parameter, which cannot be reduced (e.g. the discharge of a river in flood risk assessment of a riverside area). The parameter is stochastic in itself.
- Reducible vs non-reducible uncertainties
  - Epistemic uncertainties are (at least theoretically) reducible
  - Instead, stochastic uncertainties are (in general) irreducible (the discharge of a river will never be predicted with certainty)

A (very) simplified example: flood water level calculation



Uncertainty

$$Z_c = Z_v + \left[\frac{Q}{K_s \cdot \sqrt{(Z_m - Z_v)/L} \cdot B}\right]^{3/5}$$

Strickler's Formula

- Z_c : Flood level (variable of interest)
- Z_m et Z_v : level of the riverbed, upstream and downstream (random)
- Q : river discharge (random)
- K_s : Strickler's roughness coefficient (random)
- B, L : Width and length of the river cross section (deterministic)

General framework

Input Variables
Uncertain : *X*
Fixed : *d*

Model
G(X,d)

Output variables of interest
Z = G(X, d)

Which output variable of interest?

- Formally, we can link the output variable of interest *Z* to a number of continuous or discrete uncertain inputs *X* through the function G:

$$Z = G(X, d)$$

  - *d* denotes the "fixed" variables of the study, representing, for instance a given scenario. In the following we will simply note:

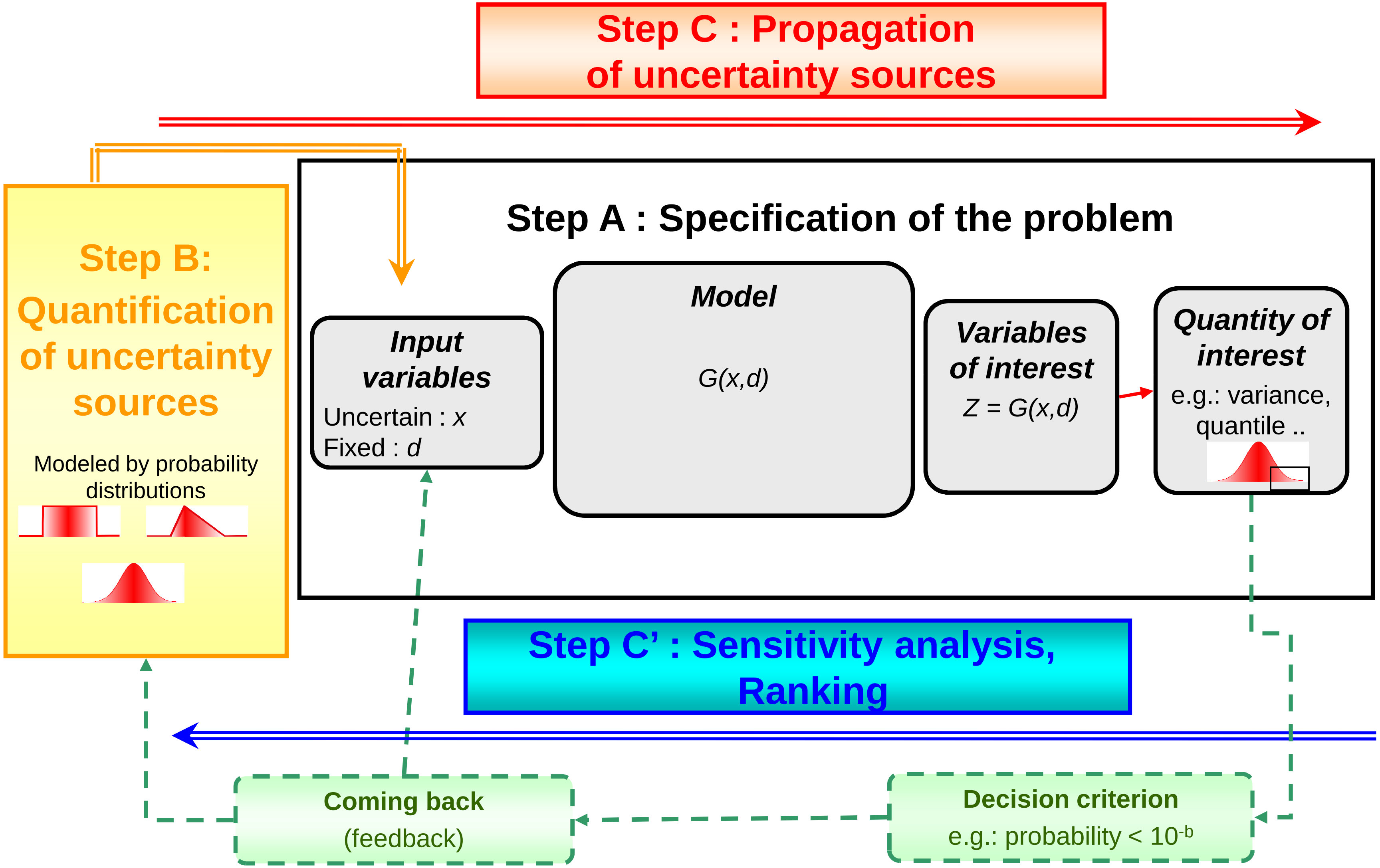
$$Z = G(X)$$

- The output variable of interest can be of dimension 1 or >1
- The function G can present itself as:
  - an analytical formula or a complex finite element code,
  - with high / low computational costs (measured by its CPU time),

- The uncertain inputs are modeled thanks to a random vector *X*, composed of *p* univariate random variables (X_1, X_2, …, X_p) linked by a dependence structure.

Methodology

The “global methodology” of uncertainty management



Let us focus on “step B”

- We stay in the case where uncertainty sources are modeled by random variable
  - Probabilistic setting
  - X is a multi-dimensional random variable
  - Its uncertainty is described by a joint distribution
  - A key question: the dependence between the components of X

- Situations encountered in common industrial practice:
  - No data → Expertise for assessing the distribution of X
  - Data available → Fitting parametric or non-parametric distributions
  - Indirectly observed data → inverse modeling

  - Bayesian approach → Combining expertise and data

No data

- In industrial practice, it may happen the only available information is an expert’s advice

- Elicitation methods
  – Formal translation of the expertise into a probability distribution
  – Particularly interesting problem in Bayesian statistics

- Open question, object of several research works

- A way to build probability distributions from minimalist information: the Maximum Entropy Method

Statistical Entropy (1/3)

- Definition given by Shannon (1948), then formalized by Jaynes (1957)

- Discrete case: X is a discrete r.v. the distribution of which is $P_X = \{p_1, p_2, \ldots, p_k\}$

$$H(X) = -\sum_{i=1}^{k} p_i \log(p_i)$$ ← Statistical Entropy

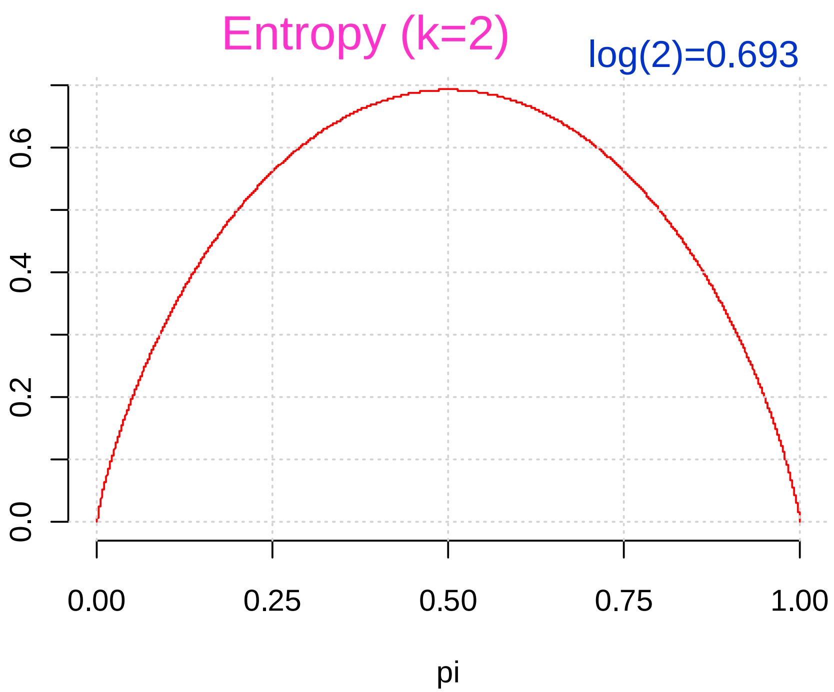
- Properties :
  - $H(X) \geq 0$
    $H(X) = 0 \Leftrightarrow \exists! p_i : p_i = 1, \quad \forall i \neq j \quad p_j = 0$
    ← Always positive, except for a particular case (minimum = 0)

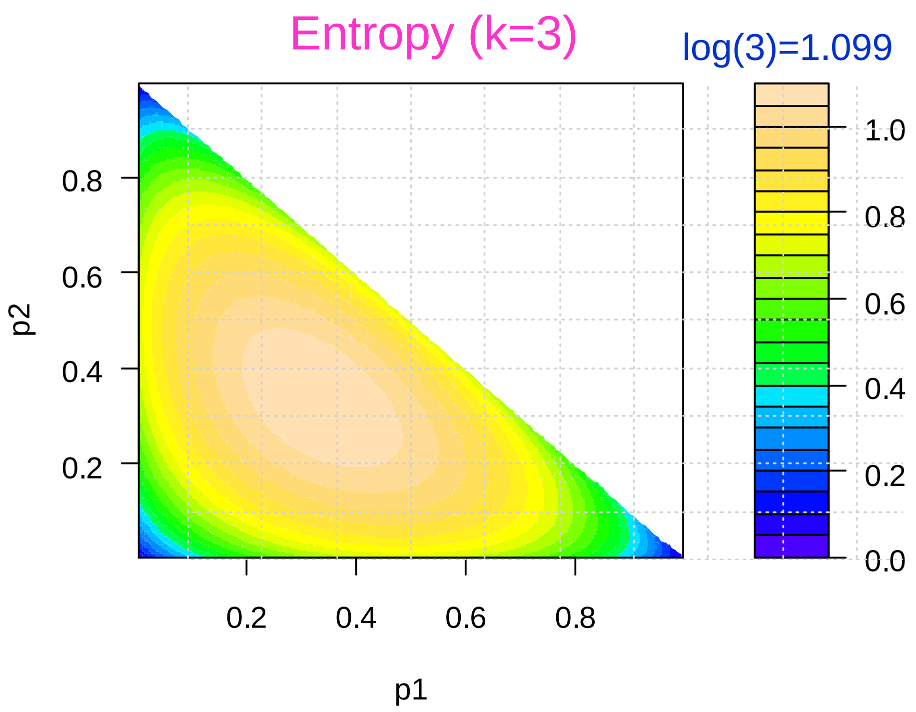
  - $H(X) \leq \log(k)$ ← Maximum de H
  - The maximum, equal to log(k), is reached in case of uniform distribution

$$p_i = \frac{1}{k} \forall i \Rightarrow H(X) = -\sum_{i=1}^{k} \frac{1}{k} \log\left(\frac{1}{k}\right) = -\frac{1}{k}\ k\ log(\frac{1}{k}) = \log(k)$$

Statistical Entropy (2/3)

- Intuitive interpretation of entropy
  - Minimum in case of “perfect” information → no doubt on the value of de X between the k possible values
  - Maximum in the case when the information given by the prob. is the most “vague” possible → each possible value of X is equiprobable
  - Entropy: (inverse) measure of the information on X brought by its prob. distribution





Statistical Entropy (3/3)

- Extension to a continuous r.v. : $H(X) = -\int_{\mathcal{X}} f(x) \log\left(f(x)\right) dx$

- Maximum Entropy Principle
  - Among all possible distributions, one chooses the one that brings the *minimum information* → i.e. the one maximizing the entropy
  - Justification : Research of "objectivity"
    - Do not add any information, except the one given by the expert

Maximum Entropy application

- Trivial Application: an expert tells that X is a discrete r.v that can take k values → choice of the discrete uniform distribution :

$$p_i = 1/k$$

- More generally, let us imagine an expert gives N pieces of information concerning X under the form:

$$\int_{\mathcal{X}} g_j(x) f(x) dx = c_j$$

  – The maximum entropy problem consists in finding a function f(x) maximizing H(X) et respecting the N +1 conditions:

$$\begin{cases} \int_{\mathcal{X}} f(x) dx = 1 \\ \int_{\mathcal{X}} g_j(x) f(x) dx = c_j \quad j = 1 ... N \end{cases}$$

Constrained Optimization

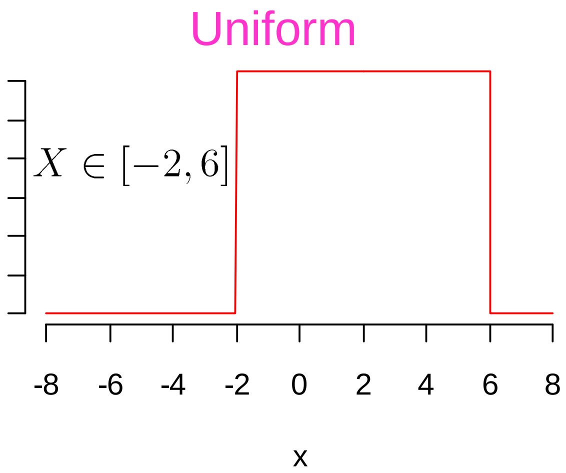
  – Justification : Among all possible distributions, one chooses the one that brings the minimum information compatible with available information

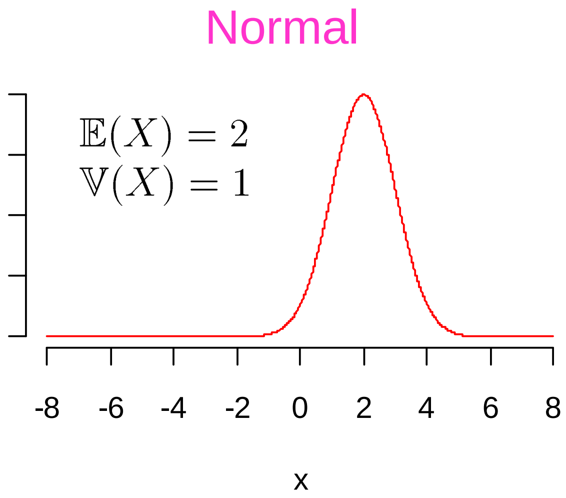
Maximum Entropy application - examples (1/2)

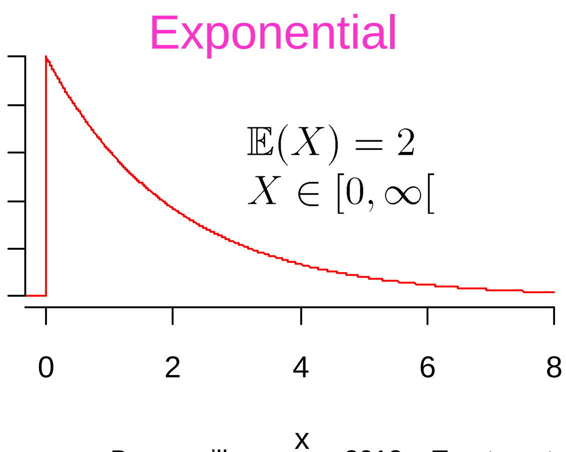
Given information	Distribution maximizing entropy	
$X \in [a,b]$	Uniform	$X \sim \mathcal{U}(a,b)$
$\mathbb{E}(X) = \mu$ $X \in [0, \infty[$	Exponential	$X \sim \mathcal{E}(1/\mu)$
$\mathbb{E}(X) = \mu$ $\mathbb{V}(X) = \sigma^2$	Normal	$X \sim \mathcal{N}(\mu, \sigma)$

- In spite of the justification, several objections can be made (e.g. on the choice of an uniform distribution…)
- Nevertheless, these choices are common in practice

Maximum Entropy application - examples (2/2)

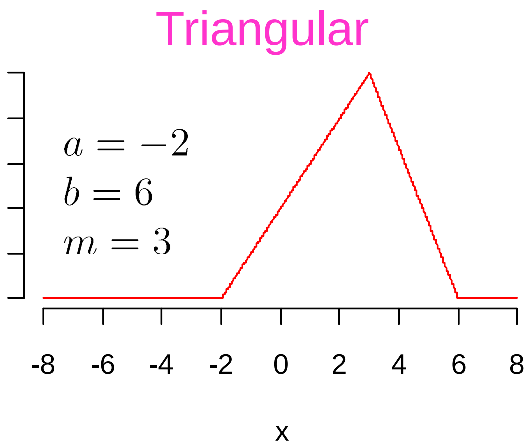






Other common distributions – no data (1/2)

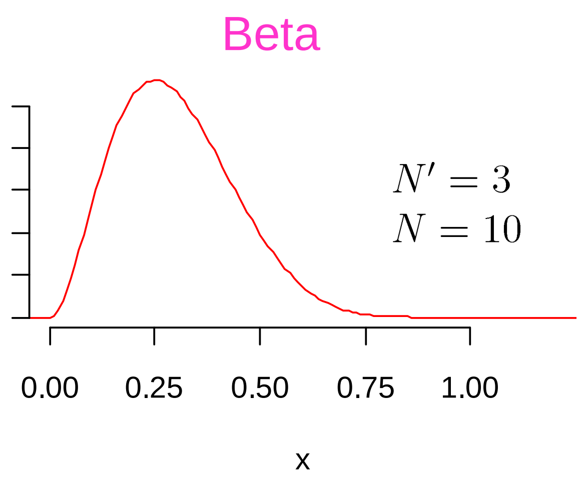
- Triangular distribution $\mathcal{T}(a,b,m)$
  - When the expert gives an interval and a mode m (most probable value)



Other common distributions – no data (2/2)

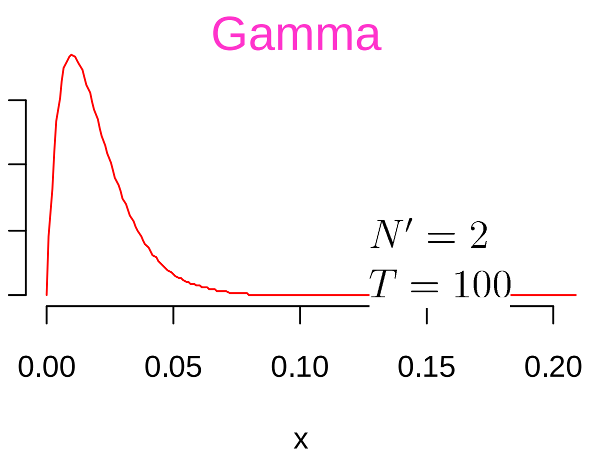
- Beta distribution $\mathcal{B}(\alpha, \beta)$
  - If the r.v. is the probability for a single event to occurr
  - The expert gives a number of "successes" N' over N virtual expériments

$$\alpha = N'$$
$$\beta = N - N'$$



- Gamma distribution $\mathcal{G}(\alpha, \beta)$
  - If the r.v. is a failure rate
  - The expert gives a number of "failures" N' observed over a "virtual" observation period T

$$\alpha = N' \qquad \beta = T$$



Data available

- Problem:
  - From an i.i.d. sample of the r.v. X:
  - Building the probability distribution of X, for:
    - Predicting its moments, quantiles etc.
    - Random sampling the r.v. X (e.g. Monte Carlo)
    - …

  - We will focus here on uni-dimensional variables

Independent and identically distributed

$$x^{(1)}, x^{(2)}, ..., x^{(n)}$$

- Non-parametric fitting
- Parametric fitting
- Verifying the quality of the fitting

Non-parametric fitting

- Empirical cumulated distribution function
- Empirical Histogram
  - “Basic” and well-known tools for the engineer

- Kernel smoothing techniques

Empirical cumulated distribution function

- i.i.d. sample of X, of size n:

$$x^{(1)}, x^{(2)}, ..., x^{(n)}$$

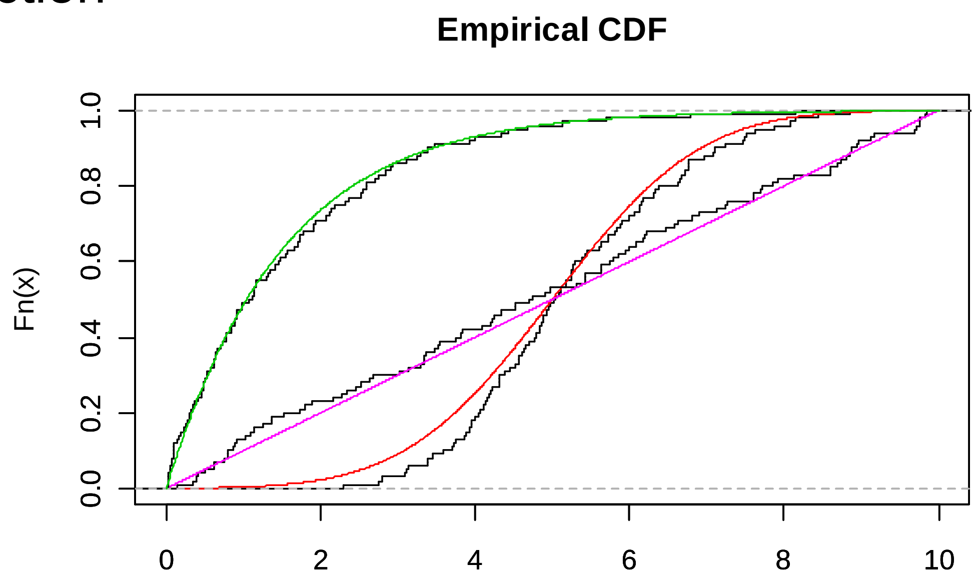
- Empirical cumulated distribution function:
  - Proportion of observations ≤ a fixed value x of la v.a. X

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x^{(i)} \leq x\}}$$

$$\hat{F}_n(x) \to F(x) \quad \text{a.s.}$$

- "Inversion" of the empirical cumulated distribution function
  - Empirical quantile:

$$\hat{x}_p = \inf\left(z : \hat{F}_n(x) \geq p\right)$$



Histogram approximation of the density

- Divide the domain of X in m intervals, of equal length h

- Approximate the density of X by the step function:

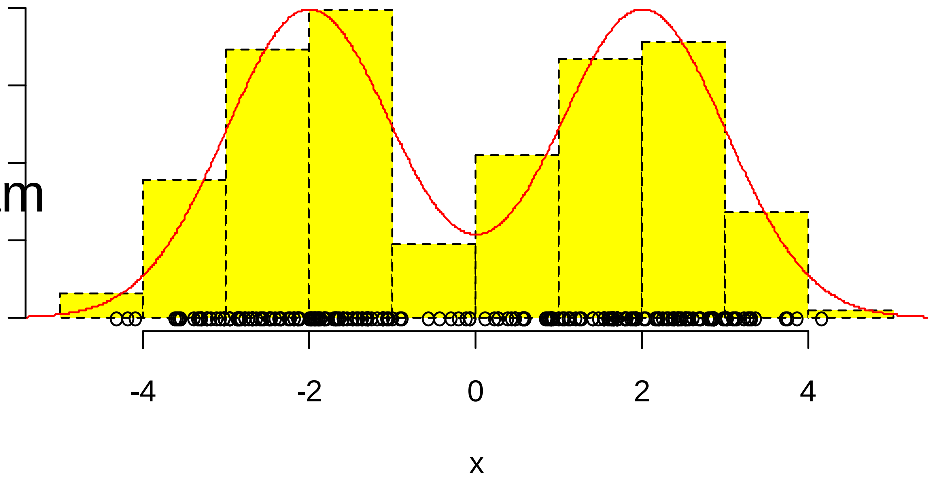
$]x^* + jh, x^* + (j+1)h] \quad j \in \mathbb{N}$

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{1}_{\{x^{(i)} \in \mathcal{I}(x)\}}$$

$\mathcal{I}(x)$ interval containing x

Number of elements of the sample which are in the same interval as x

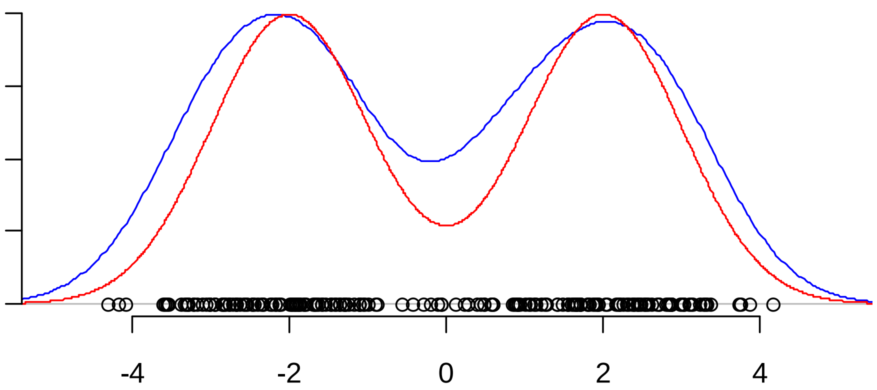
- Kernel approximation of the density is inspired by histogram approximation



Kernel approximation (1/4)

- Estimation of the density of X: $$\hat{f}_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{N} K\left(\frac{x - x^{(i)}}{h}\right)$$
  - h is called "bandwidth"
    - Smoothing parameter, the higher h, the "smoother" the density
  - K is a function, called "kernel", positive and such as: $\int_{\mathcal{X}} K(x)dx = 1$
  - K is, in general, a symmetric density, e.g. a normal distribution $\mathcal{N}(0,1)$
    In this case:

$$K\left(\frac{x - x^{(i)}}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}$$



- Other kernels: triangular, uniform, Epanechnikov …

Kernel approximation (2/4)

- Idea underlying this method
    - Histogram estimation: for x fixed, each point $x^{(i)}$ of the sample contributes to the value of f(x) in a “binary” way (yes/no)
    - Kernel estimation: the contribution is continuous and depends on the distance between x et $x^{(i)}$.

- The obtained function is continuous
  - The algorithm setting is made by:
    - The type of kernel
    - The value of h (smoothing parameter)

Kernel approximation (3/4)

- Quality of the approximation measured by:
  - Mean squared error : $\mathrm{MSE}[\hat{f}_{n,h}(x)] = \mathbb{E}\left[\left(\hat{f}_{n,h}(x) - f(x)\right)^2\right]$
  - Mean squared error integrated over the values de X: $\mathrm{MISE}[\hat{f}_{n,h}(x)] = \int_{\mathcal{X}} \mathrm{MSE}[\hat{f}_{n,h}(x)]dx$
  - Asymptotical value of MISE:

$$\mathrm{AMISE}[\hat{f}_{n,h}(x)]$$

  - By the expression of the limited development of the AMISE, it is possible, in dimension 1, to obtain a value of h minimizing it:

$$h_{\mathrm{AMISE}} = \left[\frac{R(K)}{n\mu_2^2(K)R(f'')}\right]^{1/5}$$

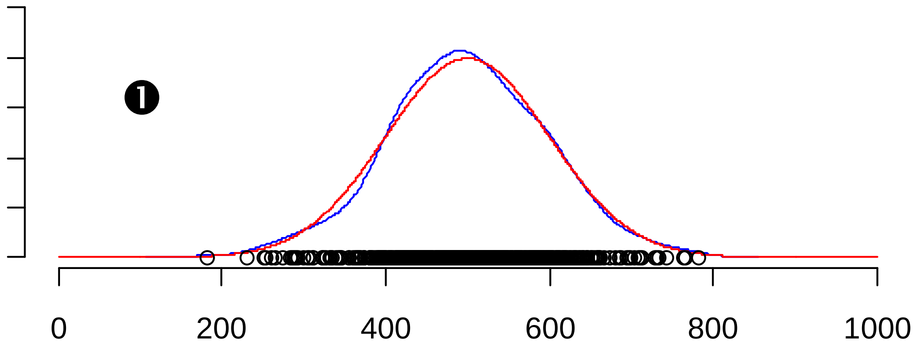
where

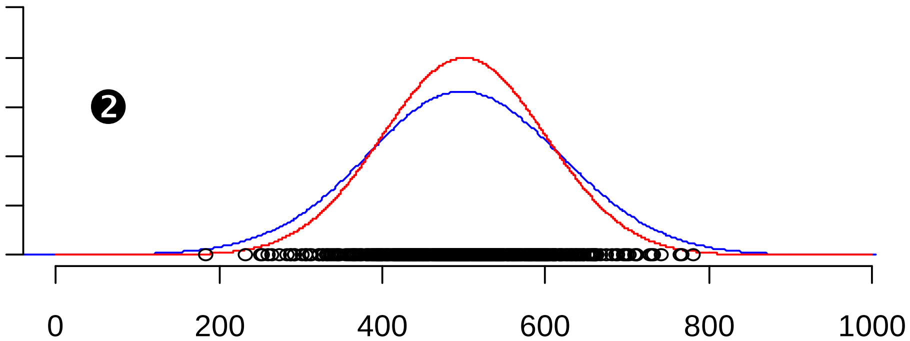
$$R(K) = \int_{\mathcal{X}} (K(x))^2 \, dx$$
$$\mu_2^2(K) = \int_{\mathcal{X}} x^2 K(x) dx$$

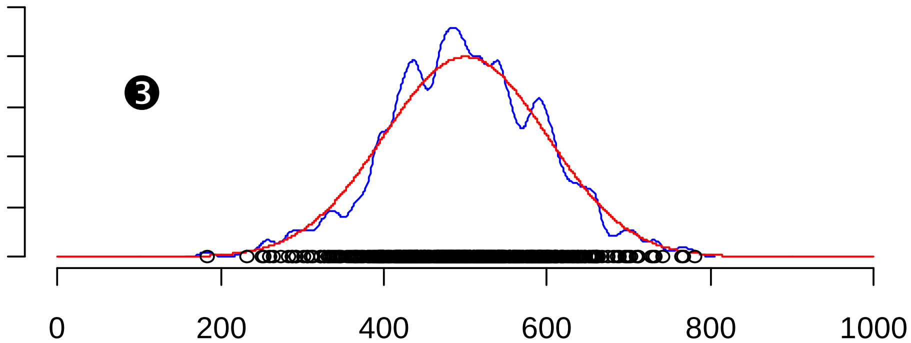
Kernel approximation (4/4)

- The most important parameter is the bandwidth h

- Examples obtained with the R function `density()`
- 500 i.i.d. values of $\mathcal{N}(\mu = 500, \sigma = 100)$
- ❶ Default algorithm in R (h*=24.9)
- ❷ H=h*3 → Oversmoothing (too smooth!)
- ❸H=h/3 → Undersmoothing







Parametric fitting

- Fundamental hypothesis: the distribution of X belongs to a given family of parametric distributions

$f(\cdot) \in \mathcal{D}_\theta$ We use the notation: $$X \sim f_\theta(x) \ \text{continuous case}$$ $$X \sim P_\theta(x) \ \text{discrete case}$$

- The distribution is completely determined by the value of its parameter θ (generally of dimension 1, 2 or 3 for the usual distributions)

- Parametric fitting consists in estimating, under the base of the available information on X, the value of the parameter θ of its under-lying distribution

Maximum Likelihood Estimation (1/4)

- Likelihood function:

$$\mathcal{L}(x^{(1)}, \ldots, x^{(n)}|\theta) = \prod_{i=1}^{n} f_\theta(x^{(i)}) \quad \text{continuous case}$$
$$\mathcal{L}(x^{(1)}, \ldots, x^{(n)}|\theta) = \prod_{i=1}^{n} P_\theta(x^{(i)}) \quad \text{discrete case}$$

- Maximum Likelihood Estimator (ML) :

$$\hat{\theta}_{\mathrm{ML}} = \underset{\theta}{\mathrm{ArgMax}} \left[\mathcal{L}(x^{(1)}, \ldots, x^{(n)}|\theta)\right] = \underset{\theta}{\mathrm{ArgMax}} \left[\log\left(\mathcal{L}(x^{(1)}, \ldots, x^{(n)}|\theta)\right)\right]$$

  - Value of θ that maximizes the likelihood (or the log-likelihood)
  - Intuitively, we look for the value which maximizes the "probability" to observe the given sample
  - It is an optimization problem

Maximum Likelihood Estimation (2/4)

- If the likelihood id differentiable (twice)
  - The two conditions below ensure $\hat{\theta}_{\mathrm{ML}}$ is a local maximizing point for the likelihood:

$$\left.\frac{\partial \mathcal{L}}{\partial \theta}\right|_{\theta=\hat{\theta}_{\mathrm{ML}}} = 0 \qquad \left.\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right|_{\theta=\hat{\theta}_{\mathrm{ML}}} \leq 0$$

- Some "classical" examples :

Normal $\mathcal{N}(\mu, \sigma)$	$\hat{\mu}_{\mathrm{ML}} = \overline{x}, \quad \hat{\sigma}_{\mathrm{ML}} = (1/n)\sum_i (x^{(i)} - \overline{x})^2$ $\text{avec } \overline{x} = (1/n)\sum_i x^{(i)}$
Exponential $\mathcal{E}(\lambda)$	$\hat{\lambda}_{\mathrm{ML}} = 1/\overline{x}$
Uniform $\mathcal{U}(a, b)$	$\hat{a}_{\mathrm{ML}} = \min(x^{(i)}), \quad \hat{b}_{\mathrm{ML}} = \max(x^{(i)})$ $i = 1 \dots n$ Case where de derivative is never null

Maximum Likelihood Estimation (3/4)

- Properties of the ML estimator

  – Convergent (consistent) : $n \to \infty, \quad \hat{\theta}_{\mathrm{ML}} \to \theta$ a.s.

  – Asymptotically normally distributed:

$$n \to \infty, \quad \sqrt{n}\left(\hat{\theta}_{\mathrm{ML}} - \theta\right) \sim \mathcal{N}(0,\sigma)$$

$$\sigma^2 = \left(\mathcal{I}(\theta)\right)^{-1}$$

Fisher information Matrix

$$\mathcal{I}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log\left(\mathcal{L}(X|\theta)\right)\right)^2\right]$$

  – Asymptotically unbiased $n \to \infty, \quad \mathbb{E}\left[\hat{\theta}_{\mathrm{ML}} - \theta\right] \to 0$

    - … but not necessarily unbiased for finite n
    - Example : estimator of the variance of a normal distribution

Maximum Likelihood Estimation (4/4)

- Properties of the ML estimator (more)

  - Asymptotically effective: among all unbiased estimators of θ, the ML estimator has a minimal variance

  - Invariant with respect to re-parameterization. Let us suppose to differently re-parameterize the distribution of X:
    - $$\xi = g(\theta) \qquad \hat{\xi}_{\mathrm{ML}} = g(\hat{\theta}_{\mathrm{ML}})$$

- Some good properties, but the ML estimator is not always calculable
  - Another usual estimation technique: the method of moments

Method of moments

- Under existence condition, let us consider the first r moments of the probability distribution of X:

$$m_j = \int_{\mathcal{X}} x^j f_\theta(x) dx, \qquad j = 1 \ldots r$$

  – Estimated by the empirical moments : $\hat{m}_j = (1/n) \sum_{i=1}^{n} \left[x^{(i)} \right]^j$

- Method of moments consists in solving in θ the system of equations:

$$\left\{ m_j = \hat{m}_j \qquad j = 1...r \right.$$

  – Properties : Convergence, asymptotic normality, but not efficiency → “less precise” estimator than the ML
  – Used when maximizing the likelihood is particularly tricky (e.g. Weibull)

Verifying the quality of the fitting

- Graphical verification
  - superimposition of theoretical and empirical cumulative distribution functions
  - QQ plot

- Goodness-of-fit tests
  - Kolmogorov - Smirnov
  - Cramer – Von Mises
  - Anderson – Darling
  - … many others

- Example: fitting a probability distribution on 149 data of maxima annual discharges of a river

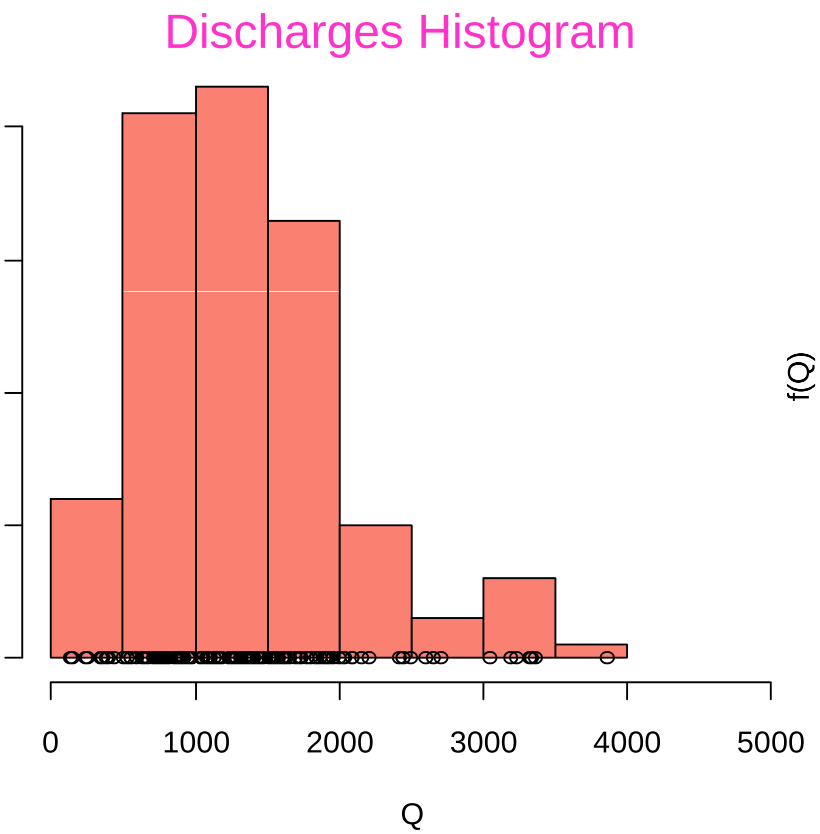
Parametric or non-parametric?

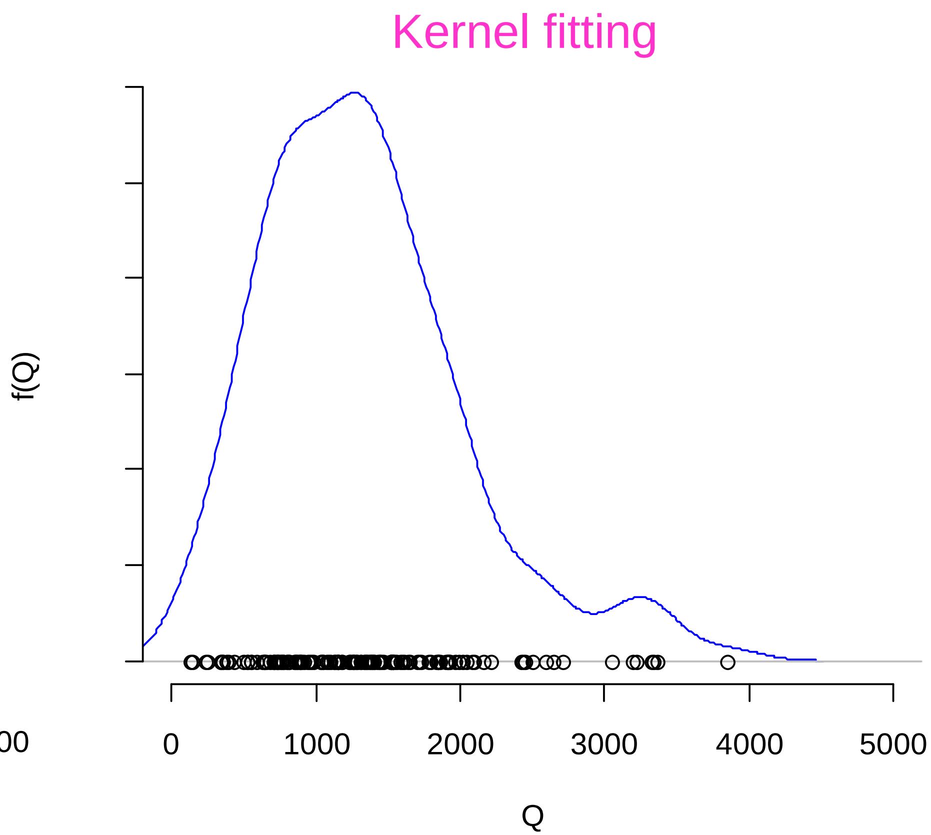
- No “unique” answer to the question!

- Generally, if possible, parametric fitting is chosen
  – Easier manipulation of the distribution

- Non-parametric fitting is interesting
  – When a great number of data is available
  – When the distribution is expected to be of “unusual” shape, e.g. multi-modal

Example of fitting (1/2)

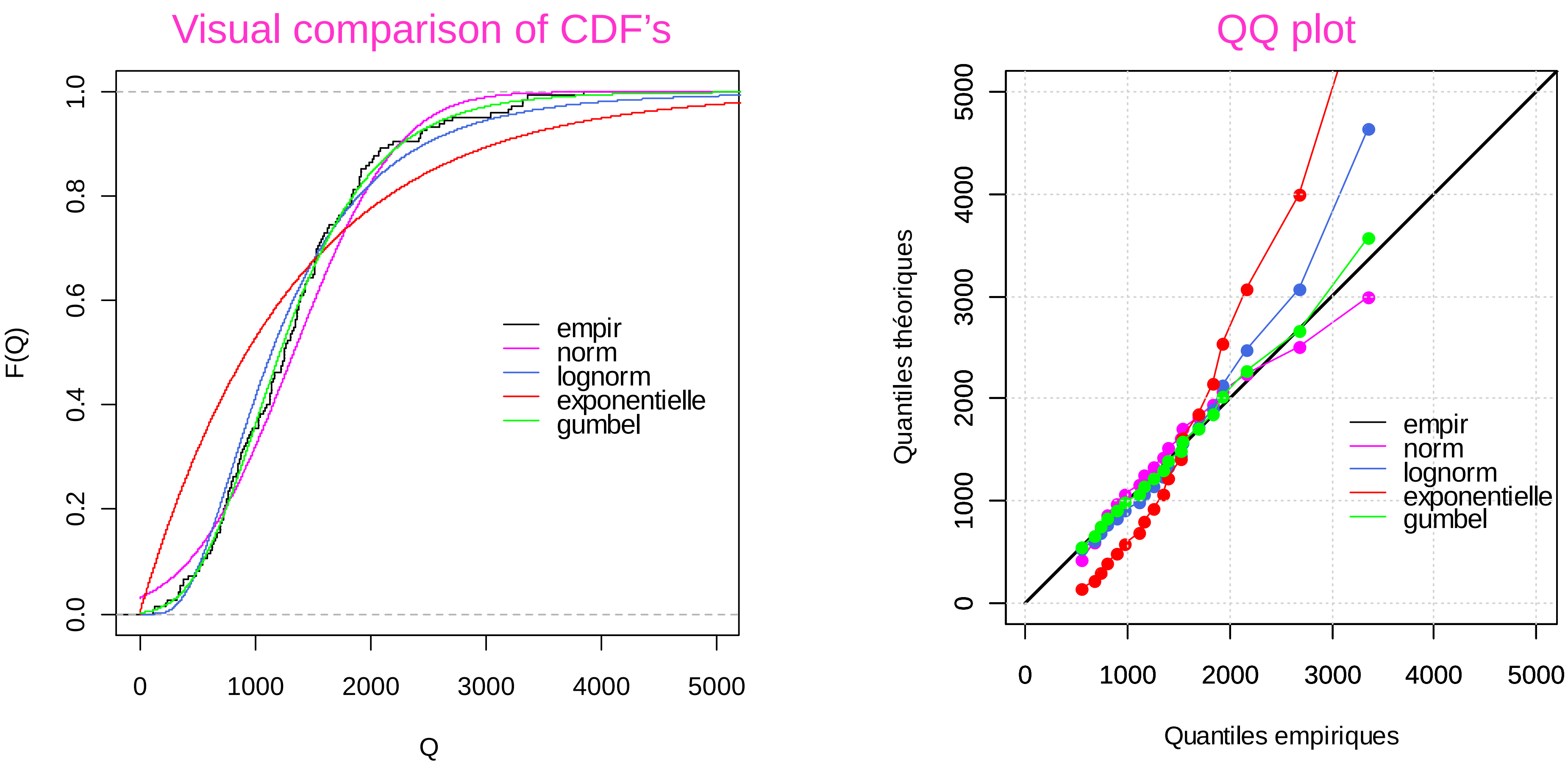




Example of fitting (2/2) – ML estimation

- Parametric fitting (maximum of likelihood)
  - Normal distribution
    - $(\hat{\mu}, \hat{\sigma}) = (1335, 711)$
  - Log-normal distribution
    - $(\hat{\mu}_{\log}, \hat{\sigma}_{\log}) = (7.0, 0.60)$
  - Exponential distribution
    - $\hat{\lambda} = 1/1335$
  - Gumbel distribution
    - $(\hat{\mu}, \hat{\beta}) = (1013, 557)$

Visual verification of the quality of fitting



Goodness-of-fit tests (1/4)

- **Goodness-of-fit test:**
  Hypothesis H_0 : the sample is an outcome of the given distribution
  Hypothesis H_1 : H_0 is not true

- These tests are based on the evaluation of a function of the data (named "test statistic") which, under the hypothesis H_0, is distributed according to a known distribution
- Significance level α : the probability to wrongly reject the null hypothesis H_0 (i.e. when H_0 is true)
- For a classical unilateral (at right) test, la decision rule is:

$$\text{Accept } H_0 \text{ if } \tau(x^{(1)}, \ldots, x^{(n)}) \leq \tau_{1-\alpha}$$

Value of the test statistic for the sample under investigation

$(1-\alpha)$ quantile of the test statistic, under the hypothesis H_0 → This quantity is known (tables, statistical software)

Goodness-of-fit tests (2/4)

- Quelques tests

  – Kolmogorov – Smirnov (KS)

    - $\tau_{KS} = \sup_x \sqrt{n} \left| F_n(x) - F(x) \right|$

  – Cramer – Von Mises (CM)

    - $\tau_{CM} = \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 \, dF(x) = \frac{1}{12n} \sum_{i=1}^{n} \left[\frac{2i-1}{2n} - F(x^{(i)}) \right]^2$ ← After an ordering of the sample

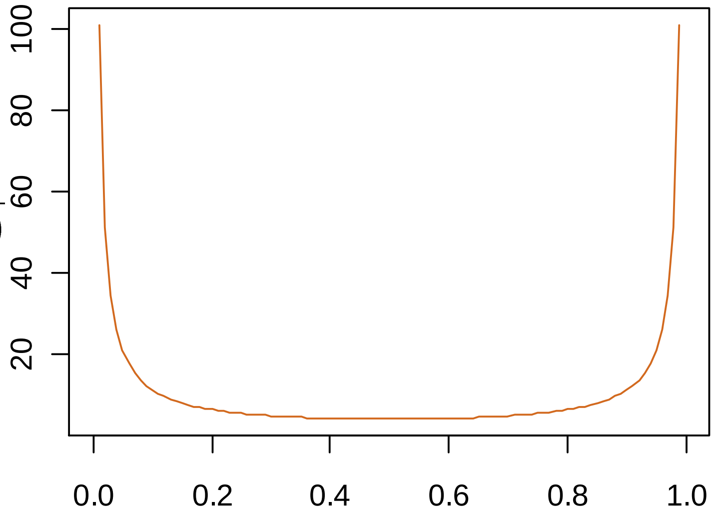
  – Anderson – Darling (AD)

    - $\tau_{AD^2} = n \int_{-\infty}^{+\infty} \frac{[F_n(x) - F(x)]^2}{F(x) \cdot (1 - F(x))} dF(x) = -n - \frac{1}{n} \sum_i \left[\log(F(x^{(i)}) + \log(1 - F(x^{(n-i+1)}) \right]$

Goodness-of-fit tests (3/4)

- The KS test takes into account the maximum deviation between empirical CDF and theoretical
- The CM test takes more into account the "global" fitting
- The AD test particularly consider the fitting on the tails of the distributions, by weighting the deviations with the factor:

$$\frac{1}{F(x)(1-F(x))}$$



Goodness-of-fit tests (4/4)

	Normal distrib.	Log-normal distrib.	Gumbel distrib.
Kolmogorov – Smirnov $\tau_{95\%}$ = 0.11	τ_{KS} = 0.091 p-val = 0.17	τ_{KS} = 0.087 p-val = 0.20	τ_{KS} = 0.043 p-val = 0.94
Cramer – Von Mises $\tau_{95\%}$ = 0.46	τ_{CM} = 0.29 p-val = 0.17	τ_{CM} = 0.23 p-val = 0.21	τ_{CM} = 0.038 p-val = 0.94
Anderson Darling	τ_{AD} = 2.08 p-val = 0	τ_{AD} = 1.44 p-val = 0.02	τ_{AD} = 0.25 p-val = 1

- Preference for the Gumbel distribution
- But other distribution could not be rejected. How to choose?
- Other selection tools (based on likelihoods ratio) :

$$\mathrm{AIC} = 2k - 2\log(\mathcal{L})$$
$$\mathrm{BIC} = k\log(n) - 2\log(\mathcal{L})$$

Just a word on Bayesian approach (1/2)

- That topic deserves an entire training course!

- Idea : updating, by data observation, a preliminary knowledge on the parameters of the statistical model, described by a *prior distribution*

$$\pi_0(\theta)$$ ← Prior distrib.

$$\pi(\theta|x^{(1)},\ldots,x^{(n)}) = \frac{\pi_0(\theta)\cdot\mathcal{L}(x^{(1)},\ldots,x^{(n)}|\theta)}{\int \pi_0(\theta)\cdot\mathcal{L}(x^{(1)},\ldots,x^{(n)}|\theta)d\theta}$$ ← Posterior distrib.

- Appealing approach in industrial practice, for incorporating expertise in statistical analysis
- Allows to explicitly account for uncertainty tainting the parameter θ, which is described here as a random variable…

Just a word on Bayesian approach (2/2)

- An unique probability distribution for describing the so-called “aleatory” (per se, irreducible) and “epistemic” (lack of knowledge, reducible) uncertainties

- A important point: the choice of the prior distribution
  - Non-informative distrib. → minimizing the information brought by the prior
  - Informative distrib. → properly modeling the expertise

- Bayesian computing
  - No analytical solution for the integral at denominator in Bayes formula
  - Use of simulation methods for getting a sample of the posterior distribution (the expression of which is always known up to a multiplicative constant)