High Performance Computing Uncertainty Quantification (Overview of Mathematical Tools)

P. Dossantos-Uzarralde

Option SGI

Uncertainty Sources

Epistemic Uncertainty

- Lack of knowledge about, say, the appropriate value to use for a quantity, or the proper model form to use.

”Reducible uncertainty” : can be reduced through increased understanding (research) or more, relevant data.

Aleatory Uncertainty

- ”Alea” = Latin for ”die” ; Latin aleator = ”dice player.”

Inherent randomness, intrinsic variability.
”Irreducible uncertainty” : cannot be reduced by additional data.
Usually modeled with probability distributions.

Uncertainty Quantification : Some facts

- UQ in computational science is the formal characterization, propagation, aggregation, comprehension, and communication of aleatory (variability) and epistemic (incomplete knowledge) uncertainties.
  E.g., demand fluctuations in a power grid are aleatory uncertainties.
  E.g., incomplete knowledge about the future (scenarios uncertainty), the validity of models, and inadequate ”statistics” are epistemic uncertainties.

- A huge range of technical issues arise in the Modeling & Simulation (M&S) components of problem definition and execution phases.

- Another huge range of technical issues arises in the delivery phase, especially in high-risk decision environments.

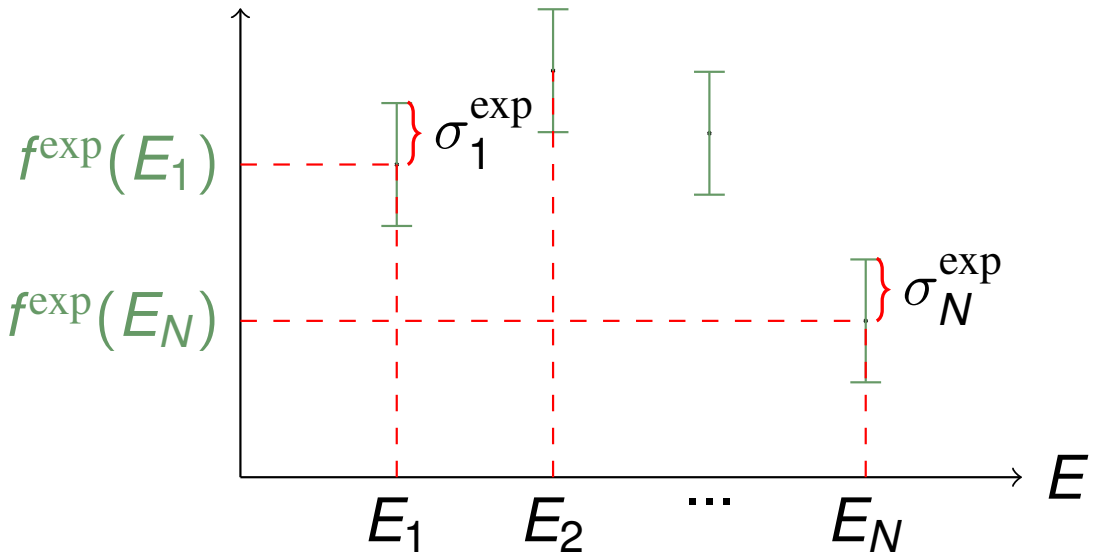
- ”Probability” is the main foundation for current ”quantification”.

UQ impact on HPC

- Any large-scale computational problem’s computing requirements increase (usually significantly) with UQ.

Problem

N measurements $f^{\exp} = (f^{\exp}(E_1), ..., f^{\exp}(E_N))$

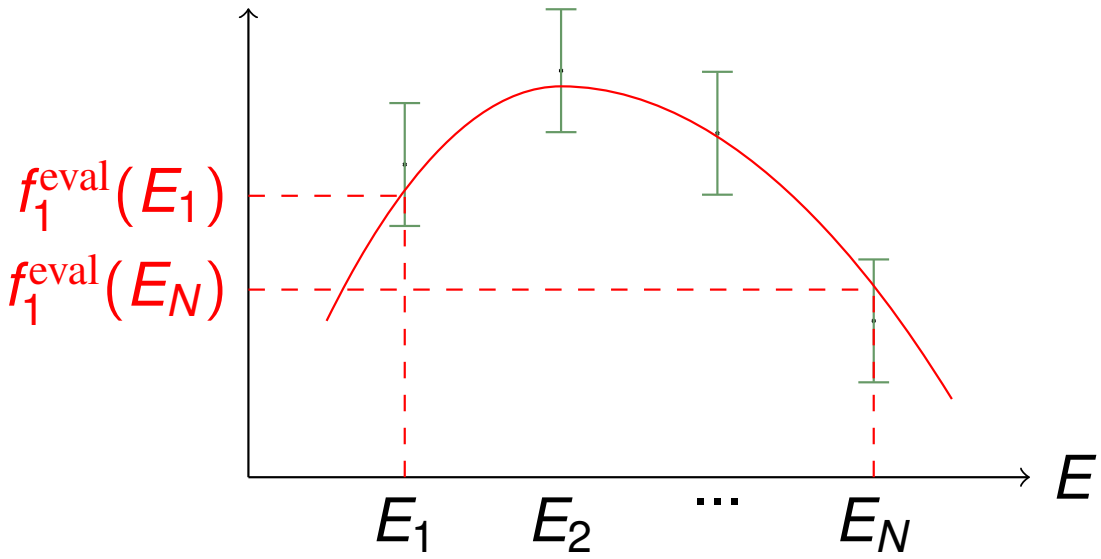


Problem

One possible model

N measurements $f^{\exp} = (f^{\exp}(E_1), ..., f^{\exp}(E_N))$

First set of input parameters: $p^1 = (p_1^1(E_1), ..., p_{n_p}^1(E_N))$

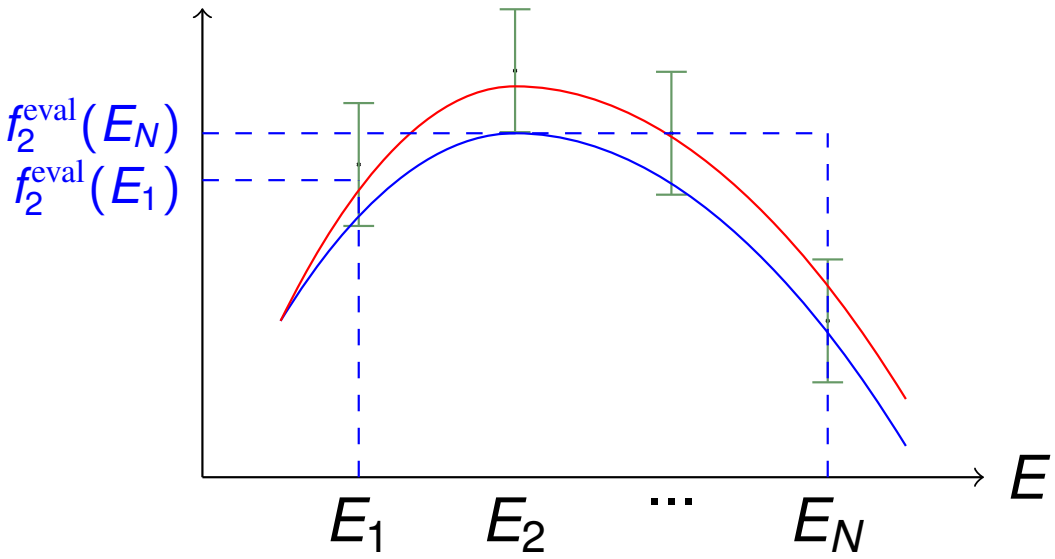


Problem

Two possible models

N measurements $f^{\exp} = (f^{\exp}(E_1), ..., f^{\exp}(E_N))$

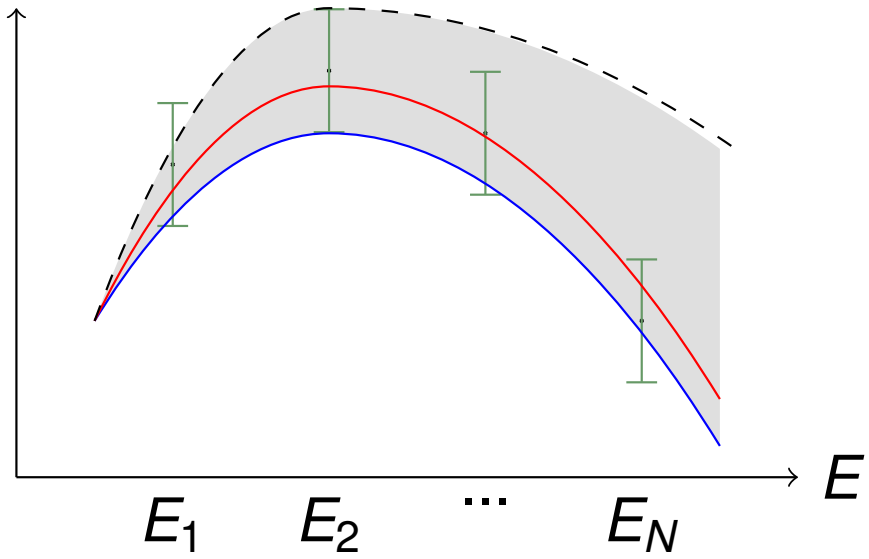
Second set of input parameters: $p^2 = (p_1^2(E_1), ..., p_{n_p}^2(E_N))$



Problem

N measurements $f^{\exp} = (f^{\exp}(E_1), ..., f^{\exp}(E_N))$

n sets of input parameters $\Rightarrow$ n possible models

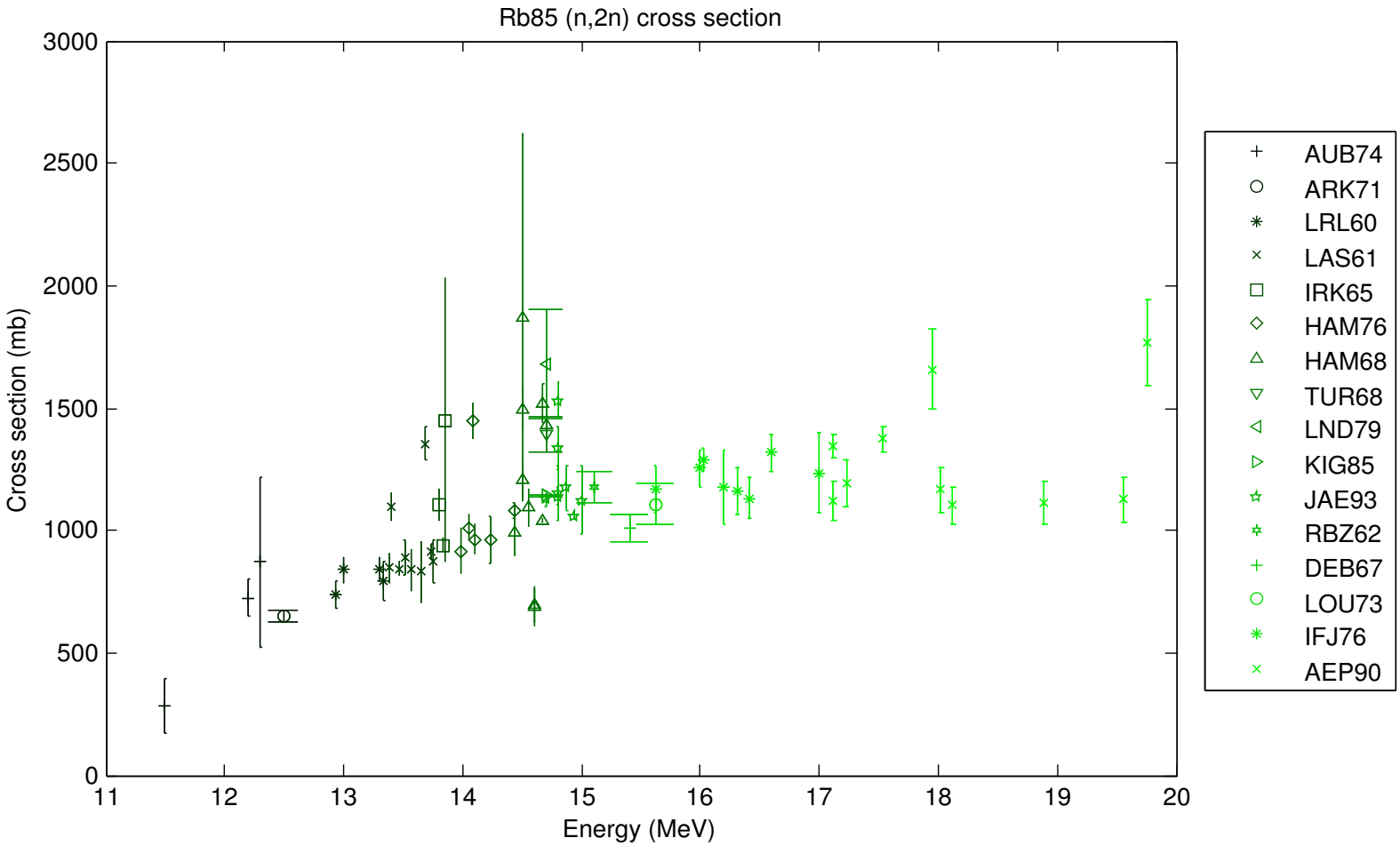


"Degrees of Goodness"

- When we use experimental results (such as property values) in an analytical solution, we should consider "how good" the data are and what influence that degree of goodness has on the interpretation and usefulness of the solution
- When we compare model predictions with experimental data, as in a validation process, we should consider the degree of goodness of the model results and the degree of goodness of the data.

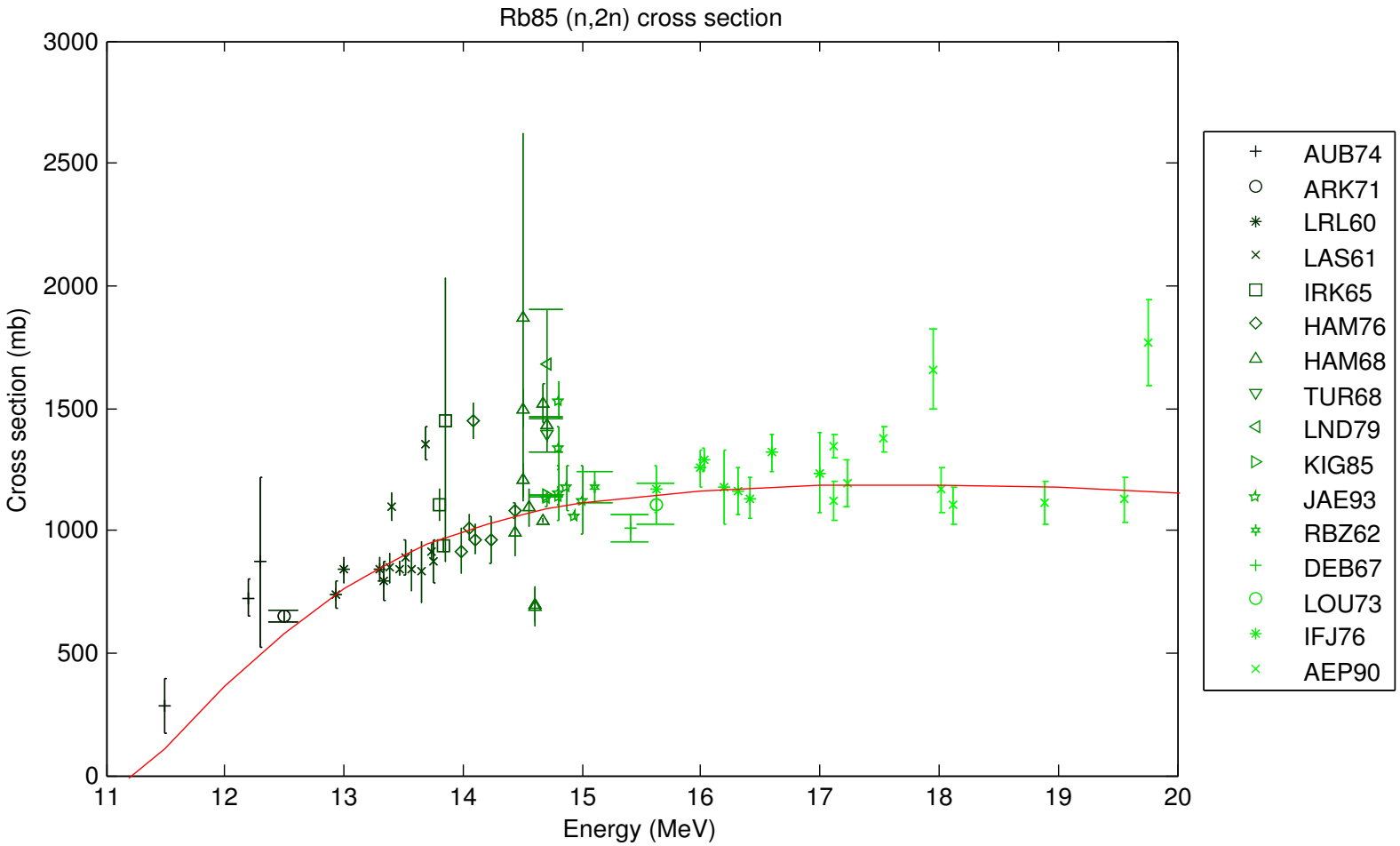
Transposition to the true life ?

Real Physical Problem



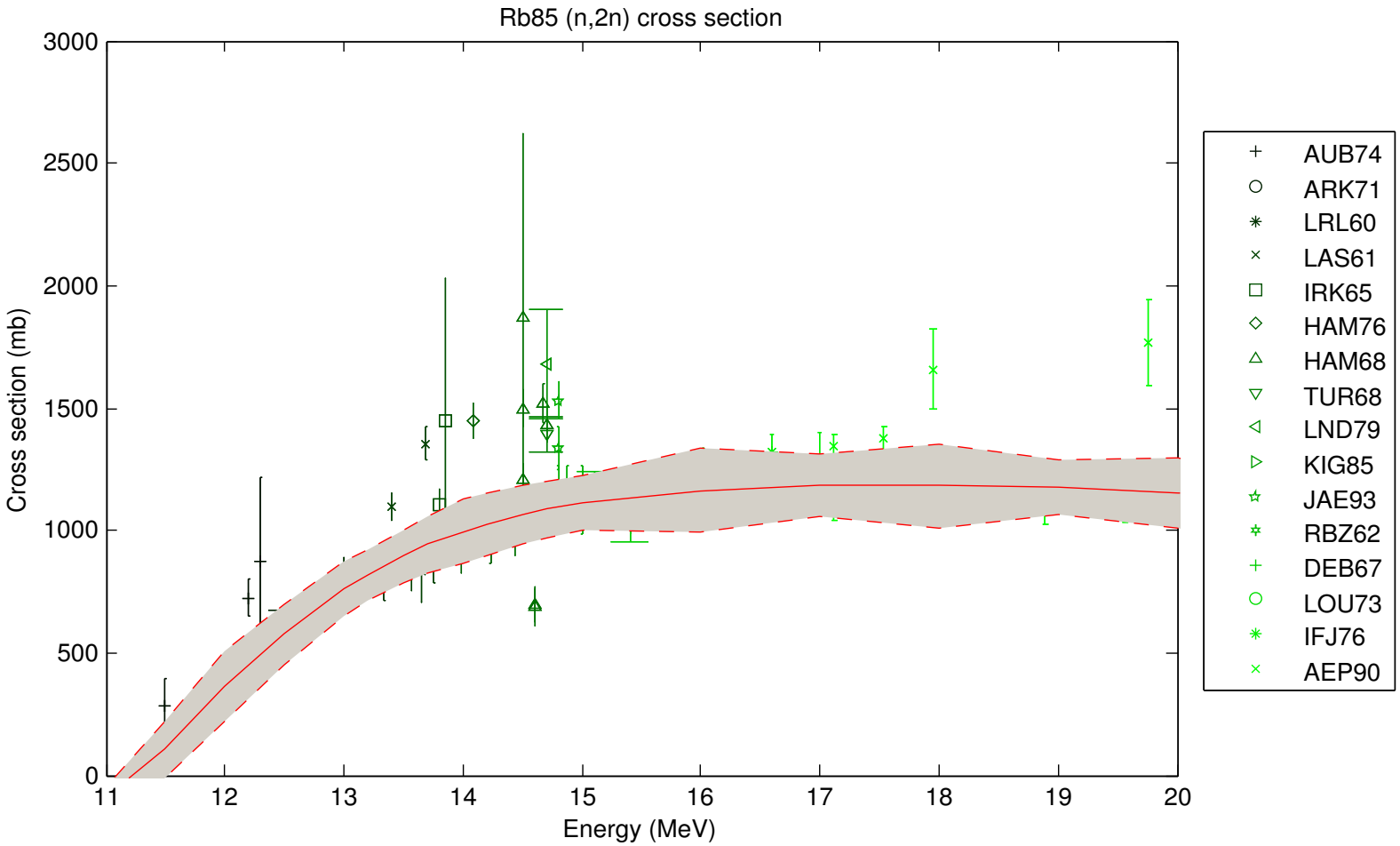
Evaluation of nuclear cross sections. *Experimental Data .*

Context: evaluation



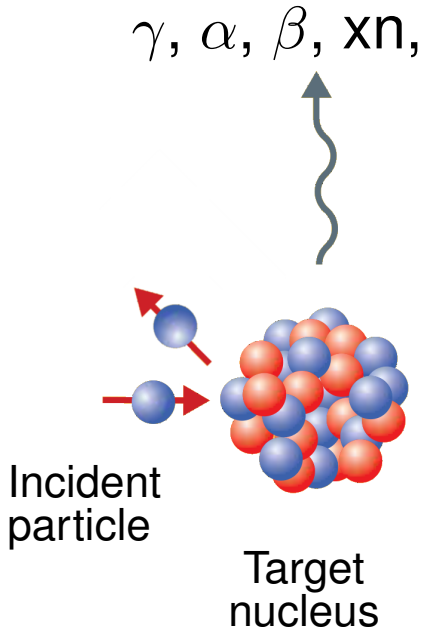
Find the good analytical model

Context: evaluation and covariance matrix



Find the model with the associated uncertainties estimations

What is the model ?



Shrödinger's Equation

Cross sections Evaluation

$$\nabla^2 \psi(\vec{r}) + \frac{2.m}{\hbar^2}.(E - \mathcal{U}(r, E)).\psi(\vec{r}) = 0$$

- ψ: wave function of the incident particle
- $\vec{r}$: position
- m: the reduced mass of the system $\{neutron + nucleus\}$
- E: energy
- $\mathcal{U}$: potential created by the nucleus
- r: radial coordinate of $\vec{r}$

Potentiel Optique

$$\mathcal{U}(r,E) = \underbrace{-\mathcal{V}_D(r,E) - i\mathcal{W}_D(r,E)}_{\text{volume}} \underbrace{-\mathcal{V}_S(r,E) - i\mathcal{W}_S(r,E)}_{\text{surface}} \underbrace{+\mathcal{V}_{SO}(r,E) + i\mathcal{W}_{SO}(r,E)}_{\text{spin-orbite}}$$

$\forall j \in \{D, S, SO\}$

$$\mathcal{V}_j(r,E) = V_j(E).g(r, R_{Vj}, a_{Vj}) \quad \text{and} \quad \mathcal{W}_j(r,E) = W_j(E).g(r, R_{Wj}, a_{Wj})$$

Paramètres du modèle optique

Potentiel Optique

$$\mathcal{U}(r,E) = \underbrace{-\mathcal{V}_D(r,E) - i\mathcal{W}_D(r,E)}_{\text{volume}} \underbrace{-\mathcal{V}_S(r,E) - i\mathcal{W}_S(r,E)}_{\text{surface}} \underbrace{+\mathcal{V}_{SO}(r,E) + i\mathcal{W}_{SO}(r,E)}_{\text{spin-orbite}}$$

$\forall j \in \{D, S, SO\}$

$$\mathcal{V}_j(r,E) = V_j(E).g(r, R_{Vj}, a_{Vj}) \quad \text{and} \quad \mathcal{W}_j(r,E) = W_j(E).g(r, R_{Wj}, a_{Wj})$$

Paramètres du modèle optique

18 paramètres: $p(E)$

Objectives

Introduce

- Challenges in UQ Science
- Various mathematical techniques

Convey

- To advance the application of mathematics and computational science to engineering, industry, science, and societ
- To promote research that will lead to effective new mathematical and computational methods and techniques for science, engineering, industry, and society

Outline

Outline

First order Taylor's development.

The Taylor series method is to approximate f_j by a linear function. The linearization greatly simplifies the error analysis, but at the expense of introducing an approximation error.

$$Y_j \approx f_j(\mu_1^X, \ldots, \mu_p^X) + \sum_{i=1}^{p} \left(\frac{\partial f_j}{\partial X_i}(\mu_1^X, \ldots, \mu_p^X) \right) \left(X_i - \mu_i^X \right). \quad (1)$$

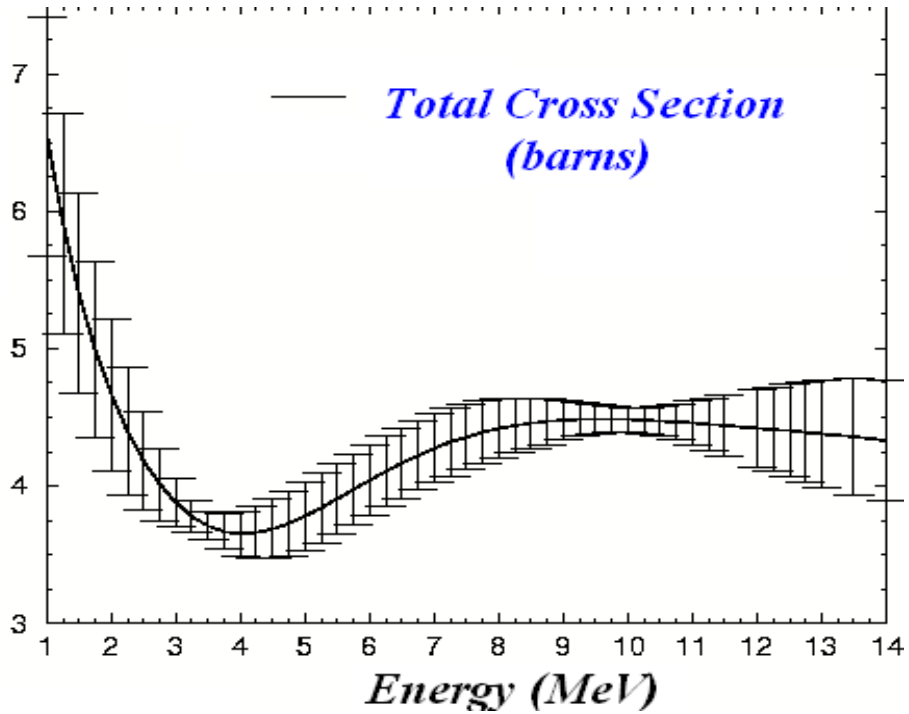
- Mean Value, Variance, Covariance :

$$\mu_j^Y \approx f_j(\mu_1^X, \ldots, \mu_p^X),\ \sigma_{jj}^Y \approx \sum_i \sum_{i'} \frac{\partial f_j}{\partial X_i} \frac{\partial f_j}{\partial X_{i'}} \sigma_{ii'}^X,\ \sigma_{jj'}^Y \approx \sum_i \sum_{i'} \frac{\partial f_j}{\partial X_i} \frac{\partial f_{j'}}{\partial X_{i'}} \sigma_{ii'}^X.$$

In a matrix form (Sandwich rule Cacuci[1981]) as : $\underline{\underline{V_Y}} = \underline{\underline{F_X}}\ \underline{\underline{V_X}}\ \underline{\underline{F_X}}^T$,
where $\underline{\underline{V_X}}$ here denotes the diagonal covariance matrix and $\underline{\underline{F_X}}$ the sensitivity matrix.

First order Taylor's development.

- Results



Second order Taylor's development

Goal : estimate how the second order affects the sensitivity of the results

$$\begin{aligned} Y_j &\approx f_j(\mu_1^X, \ldots, \mu_p^X) + \sum_{i=1}^{p} \left(\frac{\partial f_j}{\partial X_i}(\mu_1^X, \ldots, \mu_p^X) \right) (X_i - \mu_i^X) \\ &+ \frac{1}{2} \sum_{i=1}^{p} \sum_{i'=1}^{p} \left(\frac{\partial^2 f_j}{\partial X_i \partial X_{i'}}(\mu_1^X, \ldots, \mu_p^X) \right) (X_i - \mu_i^X)(X_{i'} - \mu_{i'}^X) \end{aligned} \tag{2}$$

Under the assumptions of independent parameters and continuous uniform distribution, the final covariance $Cov(Y_j, Y_{j'}) = \sigma_{jj'}^Y$ can be written:

$$\sigma_{jj'}^Y = \sum_{i=1}^{p} \frac{\partial f_j}{\partial X_i} \frac{\partial f_{j'}}{\partial X_i} \sigma_{ii}^X + \frac{1}{4} \sum_{i=1}^{p} \frac{\partial^2 f_j}{\partial X_i^2} \frac{\partial^2 f_{j'}}{\partial X_i^2} \left(E\left[(X_i - \mu_i^X)^4 \right] - \left(\sigma_{ii}^X \right)^2 \right).$$

Introducing the second derivatives and the 4^{th} central moment diagonal matrix:

$$\underline{\underline{F_X^2}} = \left(\frac{\partial^2 f_j}{\partial X_i^2} \right)_{\substack{1 \le j \le g \\ 1 \le i \le p}} \quad \underline{\underline{Q_X}} = \left(Q_{Xij} \right)_{\substack{1 \le i \le p \\ 1 \le j \le p}} = \begin{cases} 0 & \text{si } i \neq j \\ E\left[(X_i - \mu_i^X)^4 \right] & \text{si } i = j \end{cases}$$

The results above can be written in a matrix form as:

$$\underline{\underline{V_Y}} = \underline{\underline{F_X}}\, \underline{\underline{V_X}}\, \underline{\underline{F_X}}^T + \frac{1}{4} \underline{\underline{F_X^2}} \left(\underline{\underline{Q_X}} - \underline{\underline{V_X}}^2 \right) \underline{\underline{F_X^2}}^T \tag{3}$$

Outline

Monte Carlo Method

Monte Carlo simulation is categorized as a sampling method because the inputs are randomly generated from probability distributions.

- Uniform distribution

In the context of a comparison between both methods, we used the uniform distribution $\mathcal{U}(\alpha_i, \beta_i)$ as parameter probability function distribution (pdf).

$$f_{X_i}(x) = \frac{1}{\beta_i - \alpha_i} \mathbb{I}_{[\alpha_i, \beta_i]}(x), \quad \mathbb{I}_{[\alpha_i, \beta_i]}(x) = \begin{cases} 0 & \text{si } x \notin [\alpha_i, \beta_i] \\ 1 & \text{si } x \in [\alpha_i, \beta_i] \end{cases}$$

- Independent Parameters

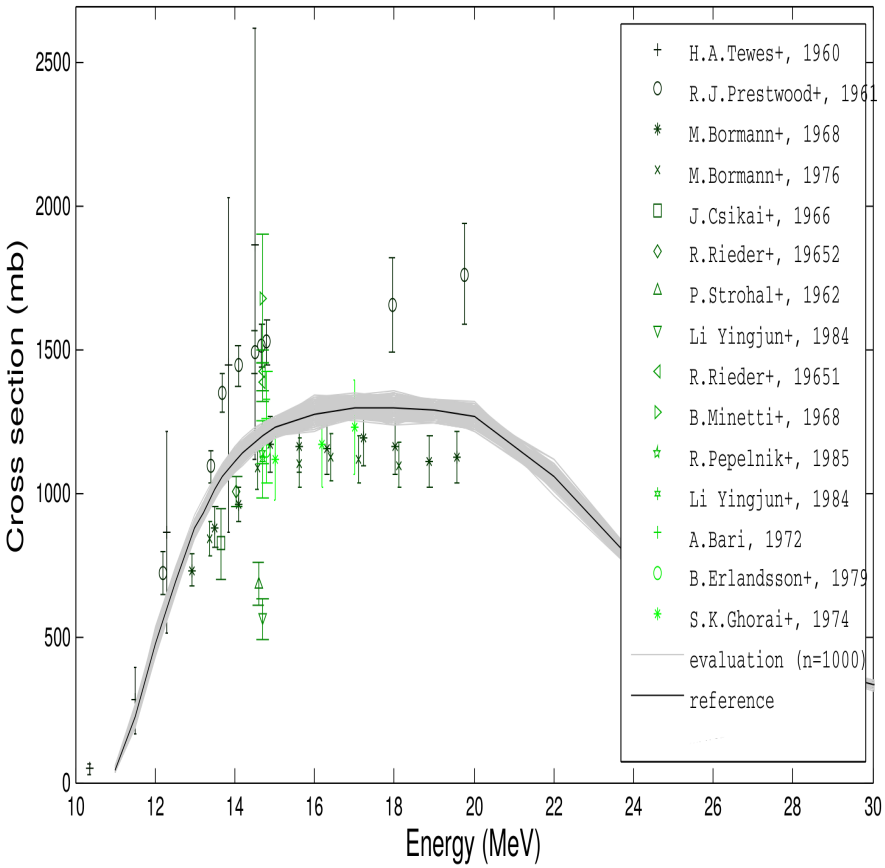
The pdf of the parameter vector $\underline{X}$ is the product of the marginal distributions f_{X_i} of each parameters X_i:

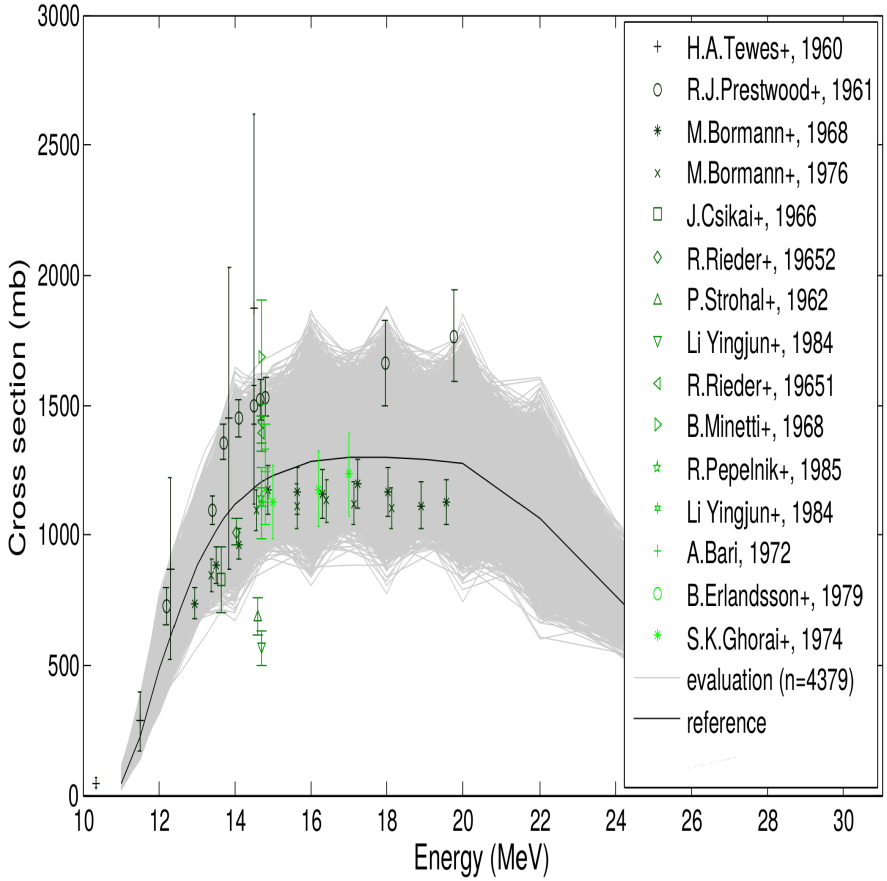
$$f_{\underline{X}} = \prod_{i=1}^{p} f_{X_i}.$$

Simulations Monte-Carlo

Parameters :

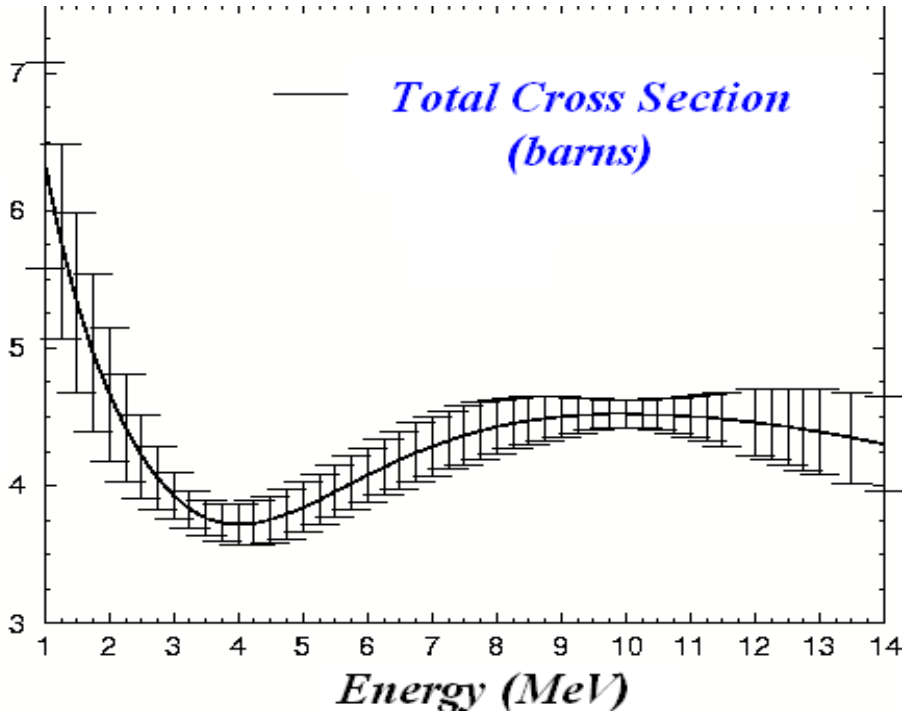
$$r_V, a_V, v_1, v_2, v_3, w_1, w_2, r_D, a_D, d_1, d_2, d_3, r_{SO}, a_{SO}, v_{SO_1}, v_{SO_2}, w_{SO_1} \text{ et } w_{SO_2}$$



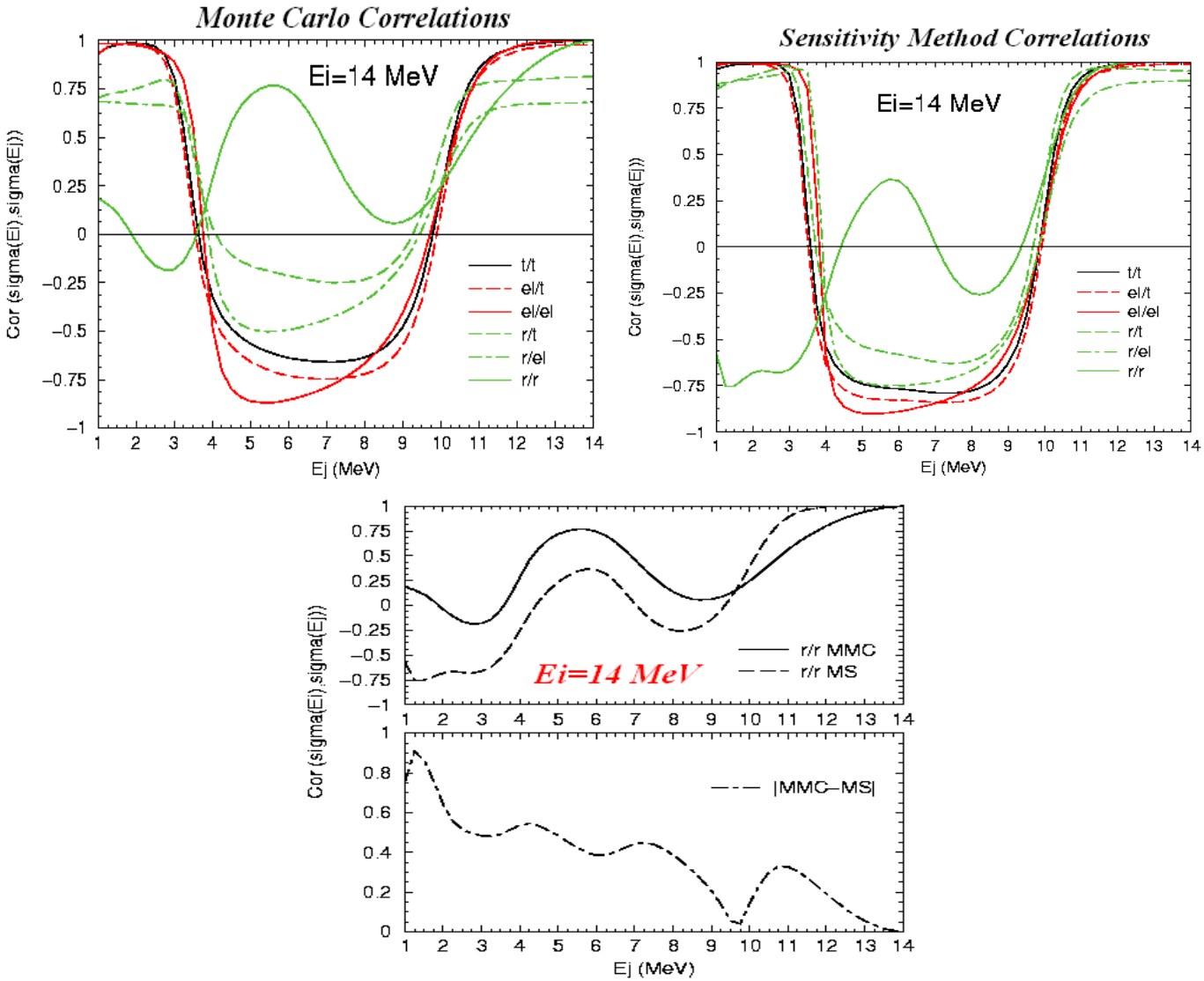


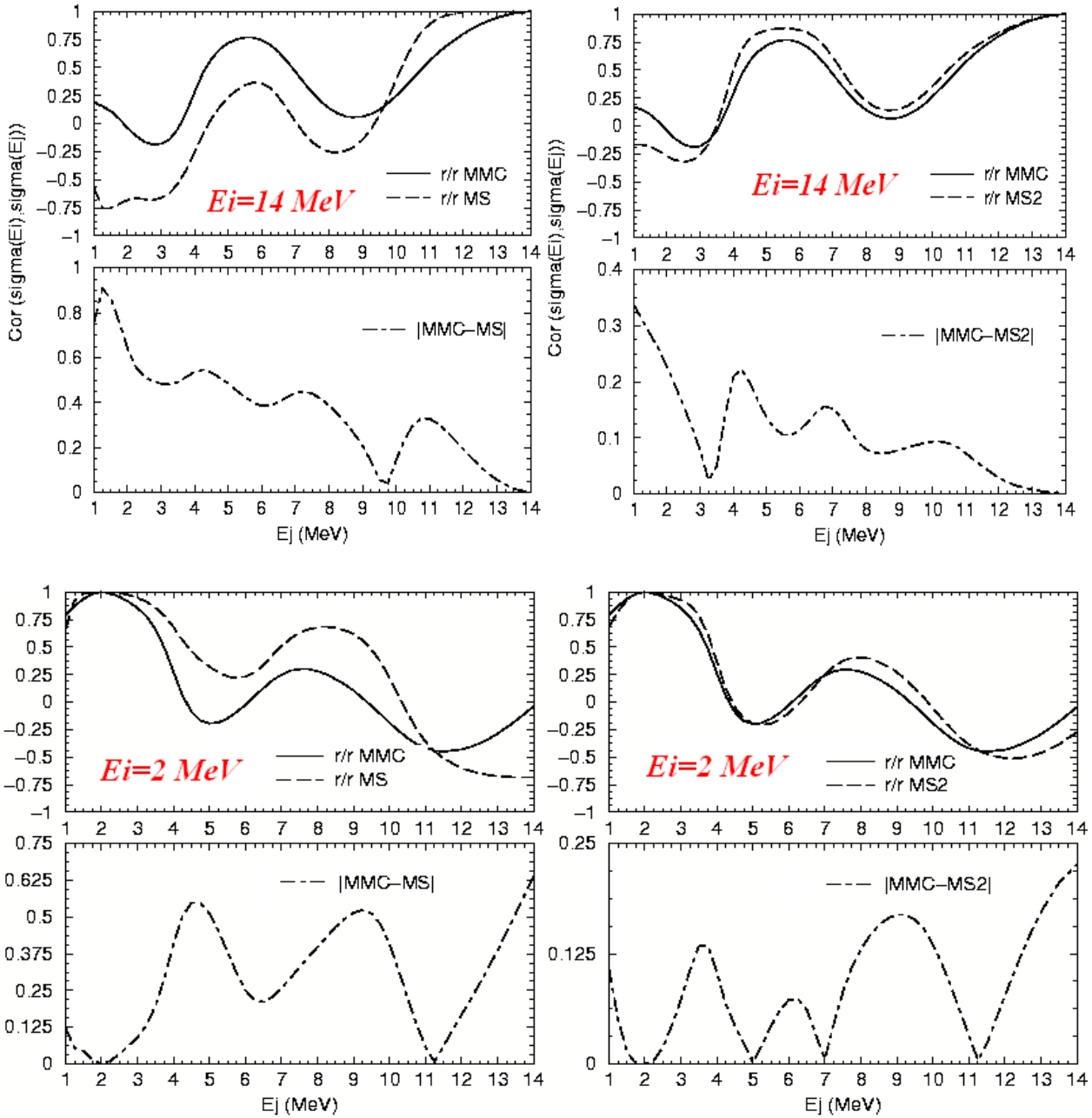
Monte Carlo Method

- Results



Comparisons




Cor (sigma(Ei),sigma(Ej))
Ei=14 MeV
r/r MMC
r/r MS
|MMC–MS|
Ej (MeV)
Ei=14 MeV
r/r MMC
r/r MS2
|MMC–MS2|
Ej (MeV)
Ei=2 MeV
r/r MMC
r/r MS
|MMC–MS|
Ej (MeV)
Ei=2 MeV
r/r MMC
r/r MS2
|MMC–MS2|
Ej (MeV)

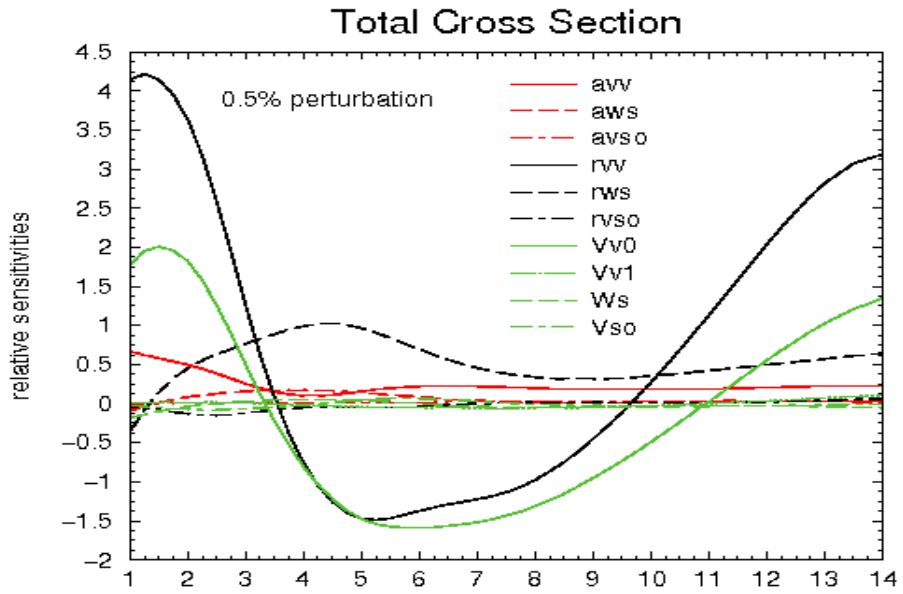
Local Sensitivity Analysis

The aim of sensitivity analysis : estimate the rate of change in the model output with respect to changes in model inputs - determine parameters for which it is important to have more accurate values, and understanding the behavior of the system being modeled.

Pb : the method involves an expansion of model outputs in terms of small random perturbations of model parameters.

- Different δx_i were tested in the interval $[0, \alpha_{X_i} = \pm 5\%]$.
- In this study, we use the normalized gradient $\frac{\partial f_j}{\partial X_i} \frac{\mu_i^X}{f_j}$.

Main limitation : require the perturbation terms be small.
These methods are in general difficult to apply along with the modeling of complex, nonlinear systems.



Outline

Sensitivity Analysis

Purpose of Sensitivity Analysis (SA)

Sensitivity analysis is the study of how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs. The process of recalculating outcomes under alternative assumptions to determine the impact of variable under sensitivity analysis can be useful for a range of purposes including :

- Testing the robustness of the results of a model or system in the presence of uncertainty.
- Increased understanding of the relationships between input and output variables in a system or model.
- Identifying model inputs that cause significant uncertainty in the output.
- Searching for errors in the model (by encountering unexpected relationships between inputs and outputs).
- Model simplification - fixing model inputs that have no effect on the output, or identifying and removing redundant parts of the model structure.
- Finding regions in the space of input factors for which the model output is either maximum or minimum or meets some optimum criterion (see optimization and Monte Carlo filtering).

Sobol Indices

- Let us assume the model function $y = f(x_1, \ldots, x_p)$, where $x_1, \ldots, x_p$ are independent input factors. Whenever $y = f(x_1, x_2, ..., x_p)$ is integrable over $[0,1]^p$, y can be decomposed (Sobol[1993]) as :

$$f(x_1, \ldots, x_p) = f_0 + \sum_{i=1}^{p} f_i(x_i) + \sum_{1 \le i \le j \le p} f_{ij}(x_i, x_j) + \ldots + f_{1,\ldots,p}(x_1, \ldots, x_p)$$

with the classical property of orthogonality.

- Consider that the input parameters are independent random variables, the joint pdf of the input factors is $P(X_1, X_2, \ldots, X_p) = \prod_{i=1}^{p} p(X_i)$.

$$E[f(X_1, \ldots, X_p)] = \int_{[0,1]^p} \left(f_0 + \sum_{i=1}^{p} f_i(X_i) + \sum_{1 \le i \le j \le p} f_{ij}(X_i, X_j) + \ldots + f_{12\ldots p}(X_1, X_2, \ldots, X_p) \right) \prod_{i=1}^{p} p(X_i)\, dX_i$$

$$V(Y) = V(f(X_1, \ldots, X_k)) = \int_{[0,1]^p} f^2(X_1, \ldots, X_k) \prod_{i=1}^{p} p(X_i) dX_i - f_0^2$$

- First order Sobol index (S_i) gives the influence of each parameter taken alone.
  Second order Sobol index (S_{ij}) a sensitivity measure due to the uncertainties in the set of input parameters$\{i, j\}$.
  Higher orders give mixed influence of parameters

$$S_i = \frac{V(E(Y|X_i))}{V(Y)}, \quad S_{ij} = \frac{V(E[Y/X_i, X_j]) - V(E[Y/X_i]) - V(E[Y/X_j])}{V(Y)}, \quad S_{ijk}, \ldots$$

Wiener-Askey Polynomial Chaos

The chaos expansion expresses the random process through a complete and orthogonal basis in terms of random variables. Polynomial Chaos {PC} is restricted to second-order stochastic processes, i.e. processes with finite second-order moments (Cameron, Martin[1947]).

- A second-order random process ($\mathcal{L}^2$) can be represented as :

$$\theta(\omega) = \sum_{k=0}^{\infty} \theta_k \, \Psi_k(\xi(\omega)), \quad \xi(\omega) = \{\xi_i(\omega)\}_{i=1}^{\infty} \qquad \textit{Wiener}[1938]$$

  - θ_k deterministic expansion coefficients,
  - $\Psi_k(\xi(\omega))$ random trial basis,
  - $\Psi_k's$ orthogonal polynomials from the Askey-scheme,
  - $\{\xi_i(\omega)\}_{i=1}^{\infty}$ independent random variables.

- Generalized {PC} forms a complete orthonormal basis of the Hilbert space with the inner product and the orthogonality relation.
- In pratical computations, the stochastic space must be truncated.
  The total number of expansion terms $P+1$ is determined by N ($\xi(\omega) = \{\xi_i(\omega)\}_{i=1}^{N}$) and p, which is the highest degree of the orthogonal polynomials :

$$\theta(\omega) = \sum_{k=0}^{P} \theta_k \Psi_k \left(\xi(\omega)\right) \qquad P+1 = \frac{(N+p)!}{N!p!}$$

- The combination of random vector and polynomials is carefully selected from the distribution of the random input.

Sobol Indices Calculations

It is possible to approximately represent the random response of a system as a {PC} expansion. Due to the orthogonality of the basis we can obtain the mean and the variance of the response

$$E[f(X)] = E[\sum_{k=0}^{\infty} f_k\, \Psi_k(X)] = f_0,\ V[f(X)] = E\left[\left(\sum_{k=0}^{\infty} f_k\, \Psi_k(X) - f_0\right)^2\right] = \sum_{k=1}^{\infty} f_k^2\, E\left[\Psi_k^2(X)\right]$$

- Expansion coefficients $\{f_k\}_{k=1}^{P-1}$ calculations:

$$< f(X)\Psi_k(X) > = < \sum_{l=0}^{\infty} f_l\, \Psi_l(X)\, \Psi_k(X) > = \sum_{l=0}^{\infty} f_l\, < \Psi_l(X)\, \Psi_k(X) > = f_k\, < \Psi_k^2(X) >$$

$$f_k = \frac{< f(X)\Psi_k(X) >}{< \Psi_k^2(X) >}$$

- Variance calculation

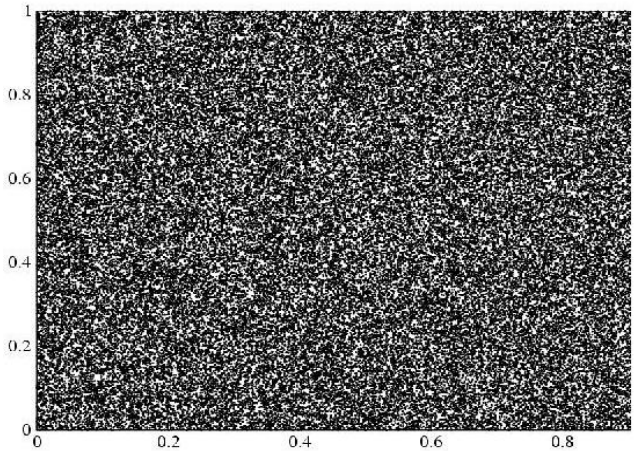
$$Var[f(X_{i_1}, ..., X_{i_s})] = E\left[(\sum_{k \in \Gamma_{i_1,\ldots,i_s}} f_k \Psi_k(X))^2\right] = \sum_{k \in \Gamma_{i_1,\ldots,i_s}} f_k^2\, E[\Psi_k^2(X)]$$

- Sobol Indices Calculations : from the {PC} chaos representation we can reach the full list of Sobol indices (Sudret[2006])

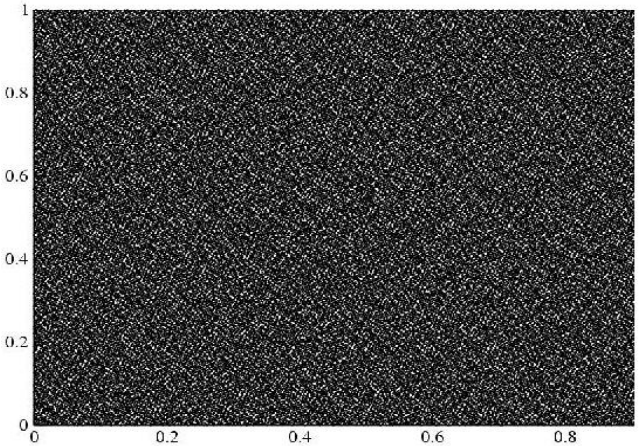
$$S_{i_1,\ldots,i_s} = \frac{\sum_{k \in \Gamma_{i_1,\ldots,i_s}} f_k^2\, E[\Psi_k^2(X)]}{\sum_{k=1}^{\infty} f_k^2\, E[\Psi_k^2(X)]}$$

Quasi-Monte Carlo

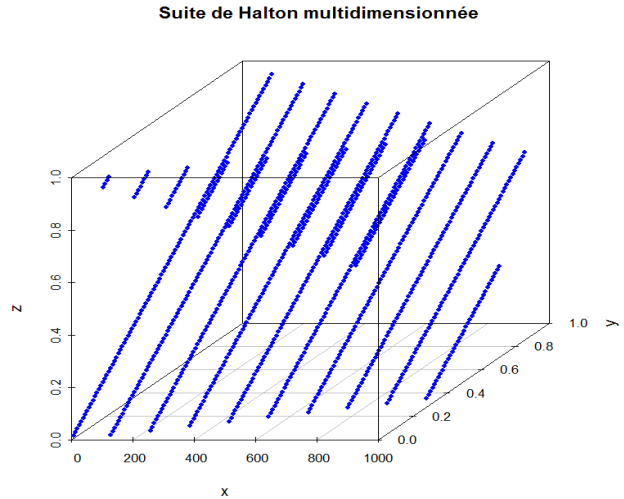
Congruential Generator



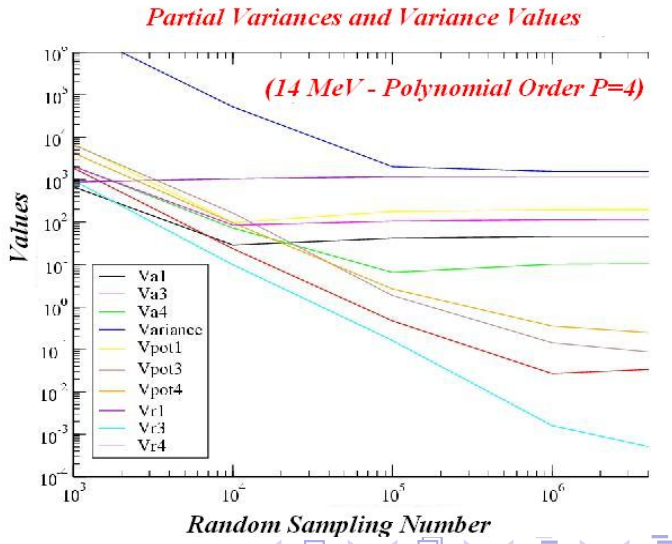
Halton Sequence



Curse of dimensionality



Rate of convergence



Outline

Least Square

- The most probable probabilty density function is the Gaussian multivariate distribution :

$$f_{\underline{X}}(\underline{X}) = \frac{1}{(2\pi)^{p/2} \mid V_X \mid^{1/2}} \exp\left(-\frac{1}{2}(\underline{X}-\underline{\mu})^T V_X^{-1}(\underline{X}-\underline{\mu})\right)$$

- V_X : in a rigourous way, we introduce the parameters correlations
  - V_X obtained with a minimum chi-square estimation.

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sigma_i^{exp} - \sigma_i^{calc}}{\Delta\sigma_i^{exp}}\right)^2.$$

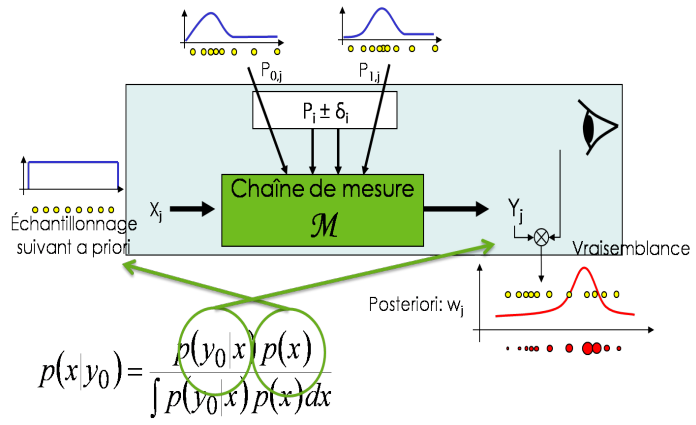
- V : experimental covariances matrice
  - σ_i^{exp} : no covariance estimation.
  - introduce χ^2 (generalized χ^2) with experimental covariances :

$$\chi^2 = \sum_{i,j=1}^{N}\left((\sigma_i^{exp} - \sigma_i^{calc}))(V^{-1})_{ij}(\sigma_j^{exp} - \sigma_j^{calc}))\right)$$

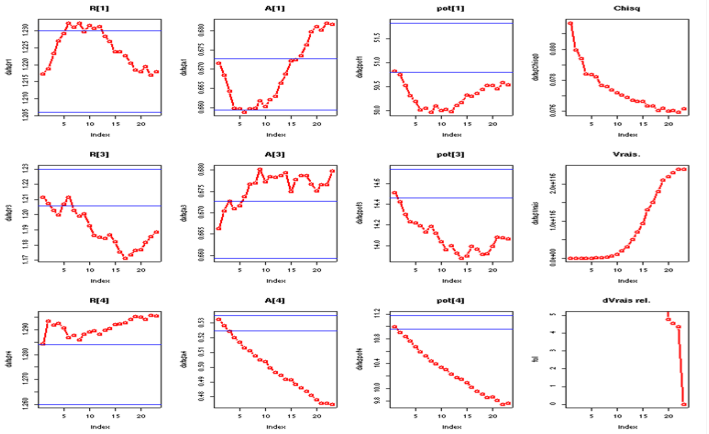
- Levenberg-Marquardt Minimization Algorithm (P. Dossantos-Uzarralde[2007])

Bayes inference

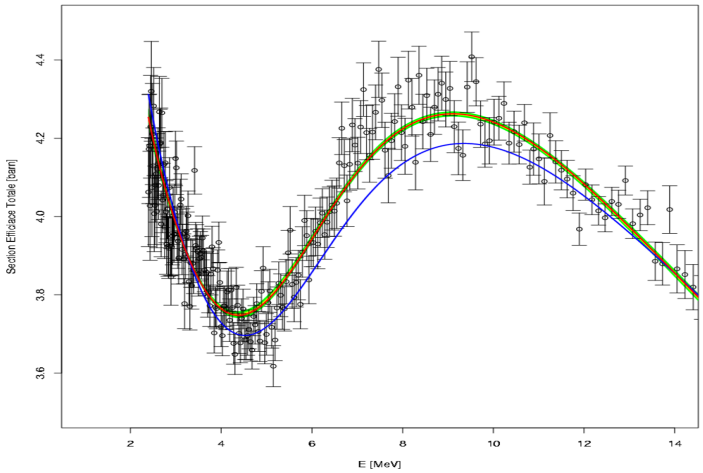
Principe



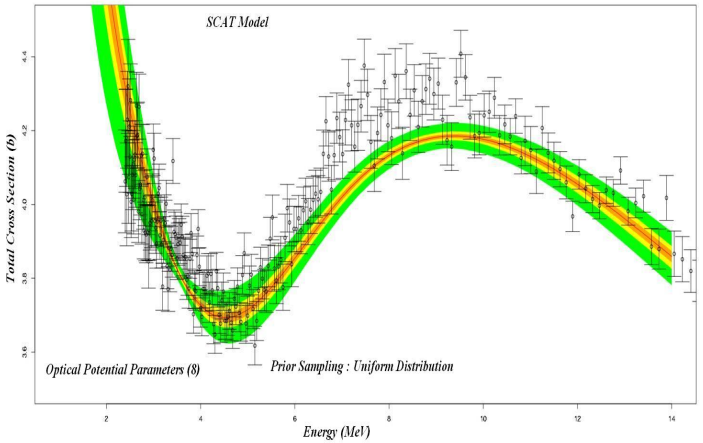
Calibration par méthode Bayésienne



Choix du Jeu de données exp.



Barres d'erreur- Dispersion des points exp.



- *On the right*, Bayesian Simulations.

Green : standard error obtained by least square optimisations - Brown : Bayesian optimisations - random parameter interval $\pm$ 1%, in yellow: random parameter interval $\pm$ 0.5%.)

Outline

Support Vector Machine

$\omega = (\omega_1, \cdots, \omega_d) \in \mathbb{R}^d$ un vecteur de paramètres du modèle f^{model}
ω_0 le paramètre d'ordonnée à l'origine.

Problème de régression

On cherche une application linéaire $f^{model}(E^{pred}) = {}^t\omega E^{pred} + \omega_0$ telle que pour tout point d'apprentissage de $\mathcal{A}_f$, σ_i^{exp} soit le plus proche possible de $f^{model}(E_i^{exp})$.

E_i^{exp} ($E_i^{exp} \in \mathbb{R}$) : vecteur caractérisant les valeurs des énergies. On pose cependant $E_i^{exp} \in \mathbb{R}^d$ uniquement pour des besoins mathématiques.
$< ., . >$ est le produit scalaire dans $\mathbb{R}^d$.

Support Vector Machine

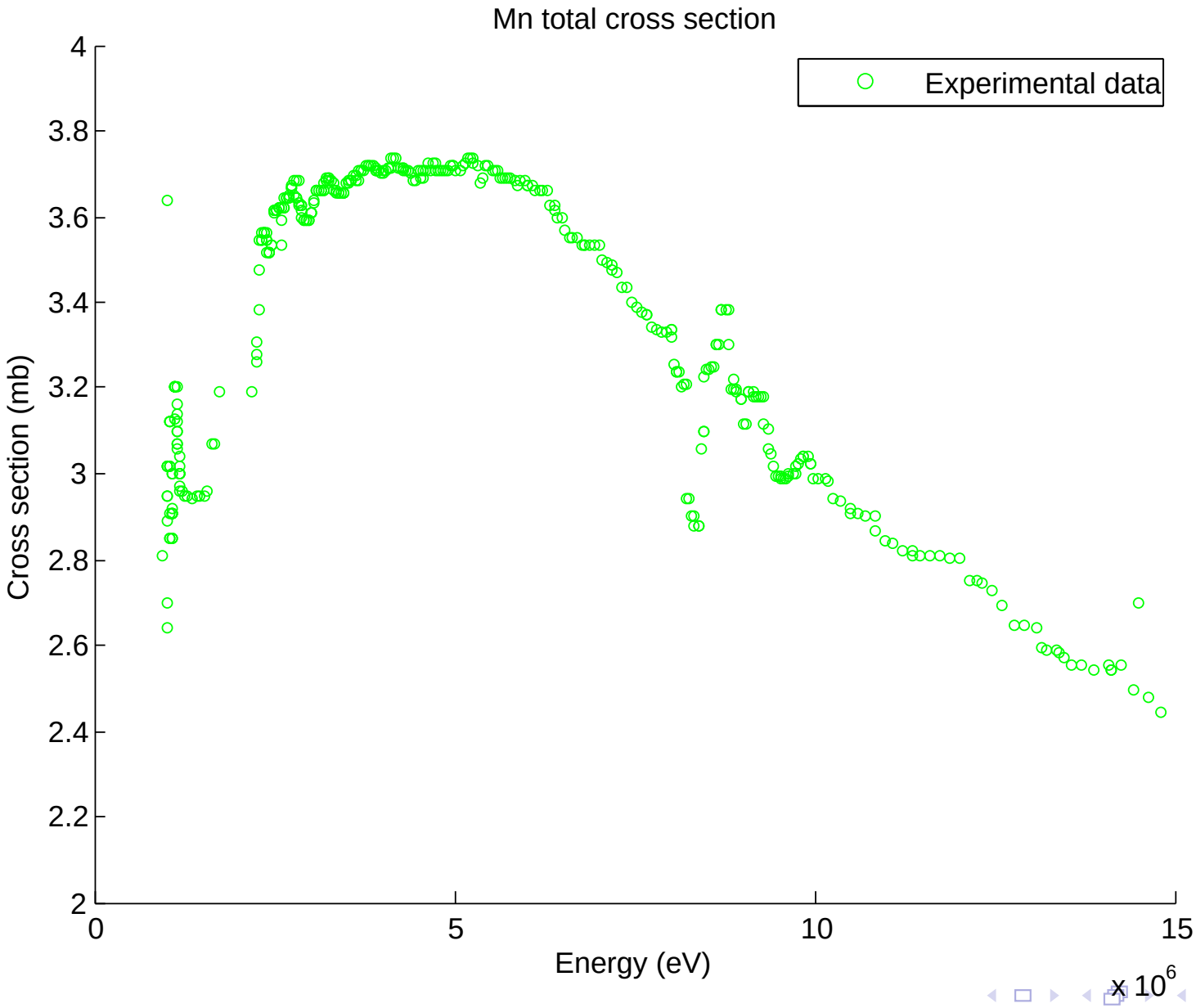
Fonction de perte

$$\begin{aligned} \forall \epsilon \in \mathbb{R}_+,\ l(\sigma_i^{exp}, f^{model}(E_i^{exp})) &= |\sigma_i^{exp} - f^{model}(E_i^{exp})|_\epsilon \\ &= \begin{cases} 0 & si\ |\sigma_i^{exp} - f^{model}(E_i^{exp})| \leq \epsilon \\ |\sigma_i^{exp} - f^{model}(E_i^{exp})| - \epsilon & sinon \end{cases} \end{aligned}$$

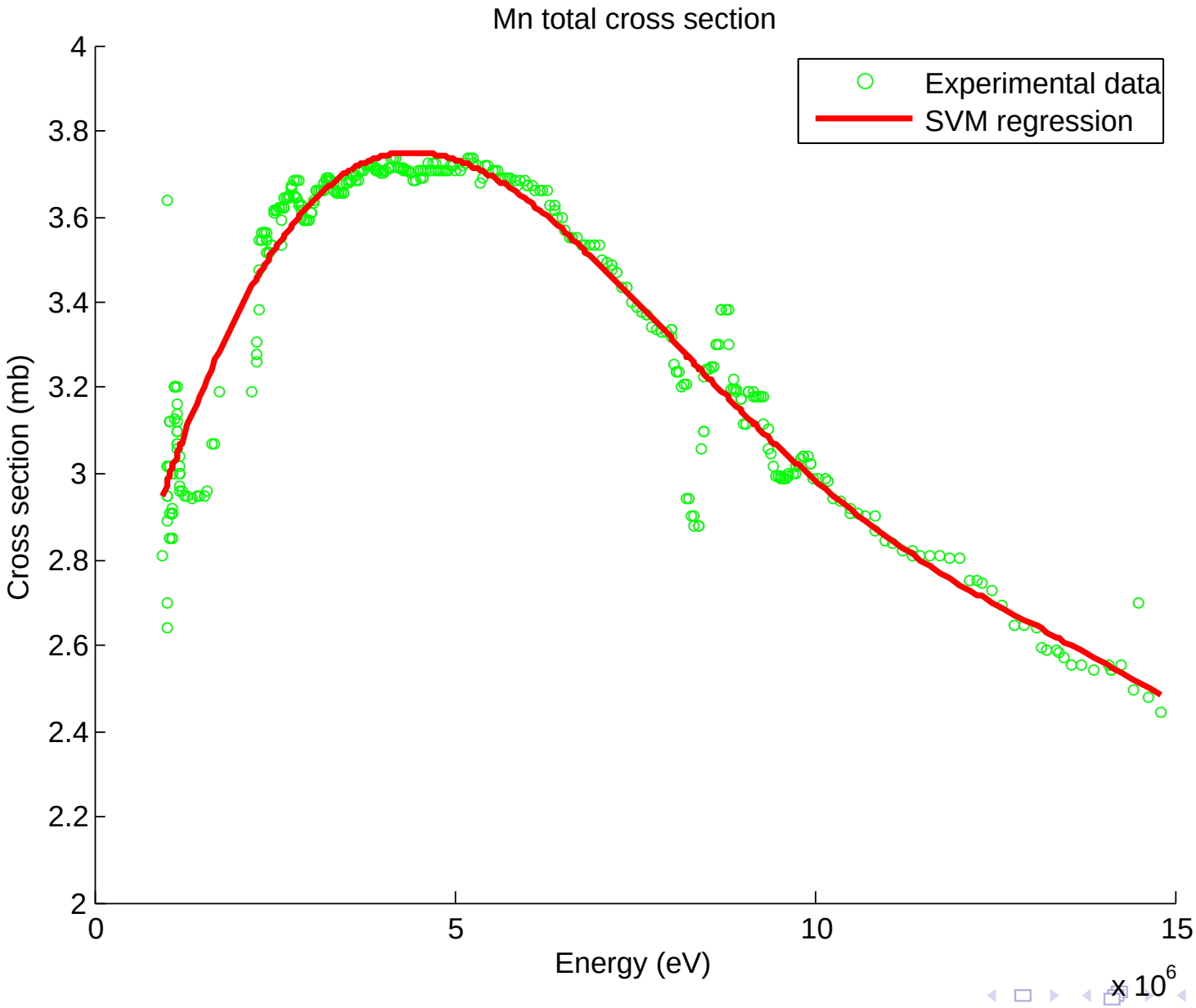
Cette fonction de perte est une norme L_1 epsilon insensible, c'est-à-dire qu'elle ne pénalise pas les mauvaises prédictions à epsilon près. Toutes les valeurs présentes dans le tube d'epsilon sont alors déterminantes pour la régression, les autres valeurs sont pénalisées. En ulilisant cette fonction de perte, on obtient le risque réel empirique suivant :

$$\hat{\mathcal{R}}_{Reel}(f^{model}, \mathcal{A}_f) = \frac{1}{m}\sum_{i=1}^{m} l(\sigma_i^{exp}, f_i^{model}) = \frac{1}{m}\sum_{i=1}^{m} |\sigma_i^{exp} - f^{model}(E^{pred})|_\epsilon. \quad (4)$$

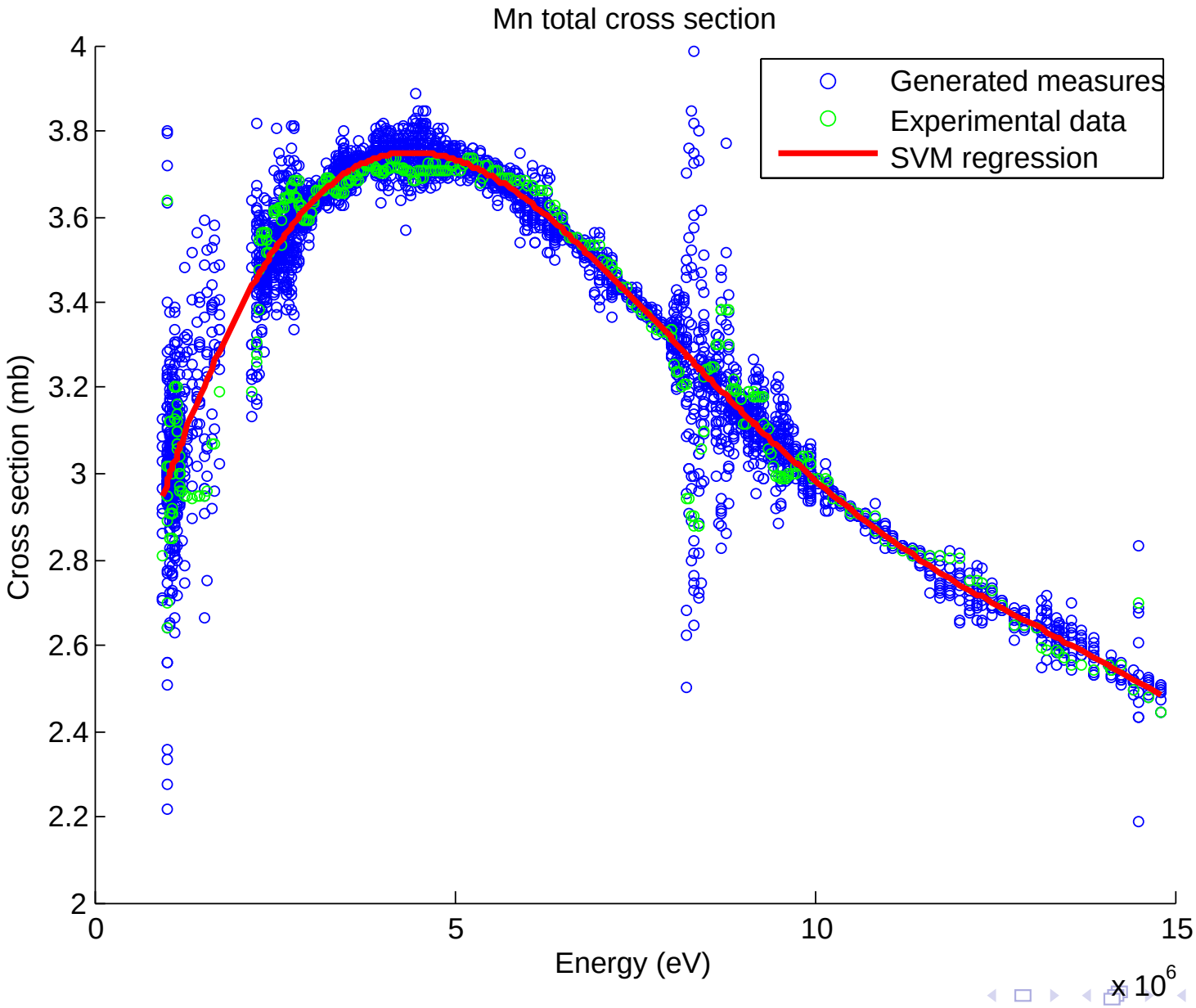
Support Vector Machine



Support Vector Machine

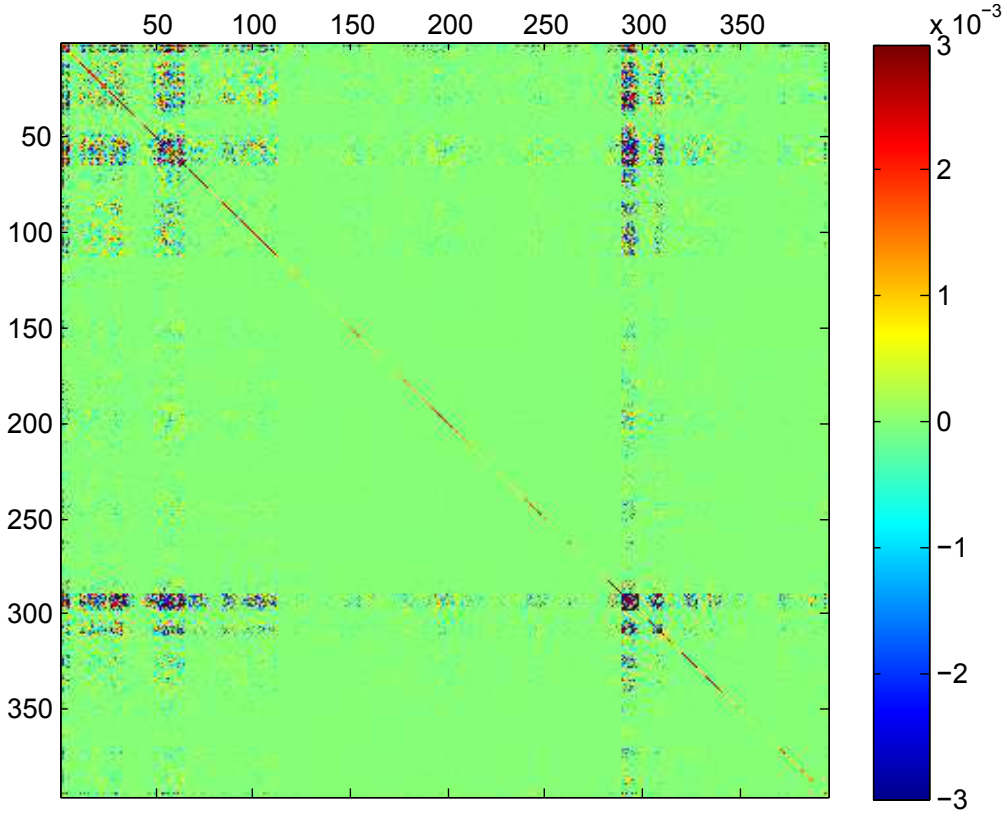


Support Vector Machine



Support Vector Machine

Matrice de covariance "expérimentale" Σ_F (400, 400)



Scoring Method

$f_1^{\text{eval}}, \ldots, f_n^{\text{eval}}$

n initial evaluations

Adequation to experimental data

$s(f_1^{\text{eval}}, f^{\text{exp}}), \ldots, s(f_n^{\text{eval}}, f^{\text{exp}})$

n score evaluations

Sort of the evaluations

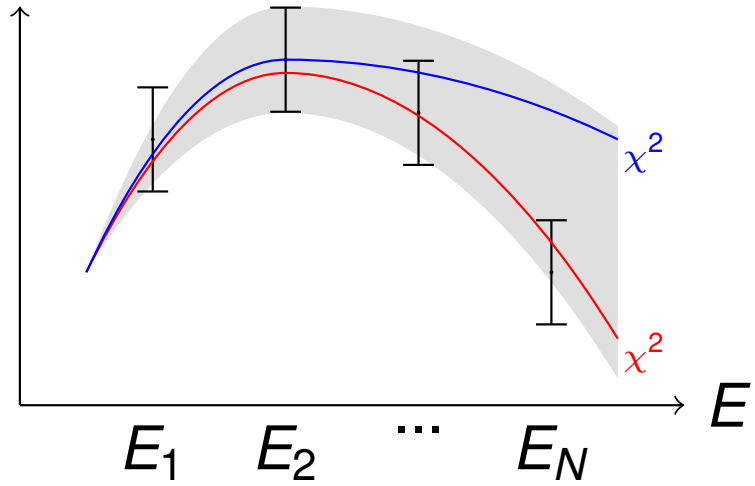
according to the score values

$f_{(1)}^{\text{eval}}, \ldots, f_{(n)}^{\text{eval}}$

n sorted evaluations

n_1 evaluations

Empirical estimators



$$\chi^2 = \sum_{j=1}^{N} \left(\frac{f^{\text{eval}}(E_j) - f^{\text{exp}}(E_j)}{\sigma_j^{\text{exp}}} \right)^2$$

Score functions used

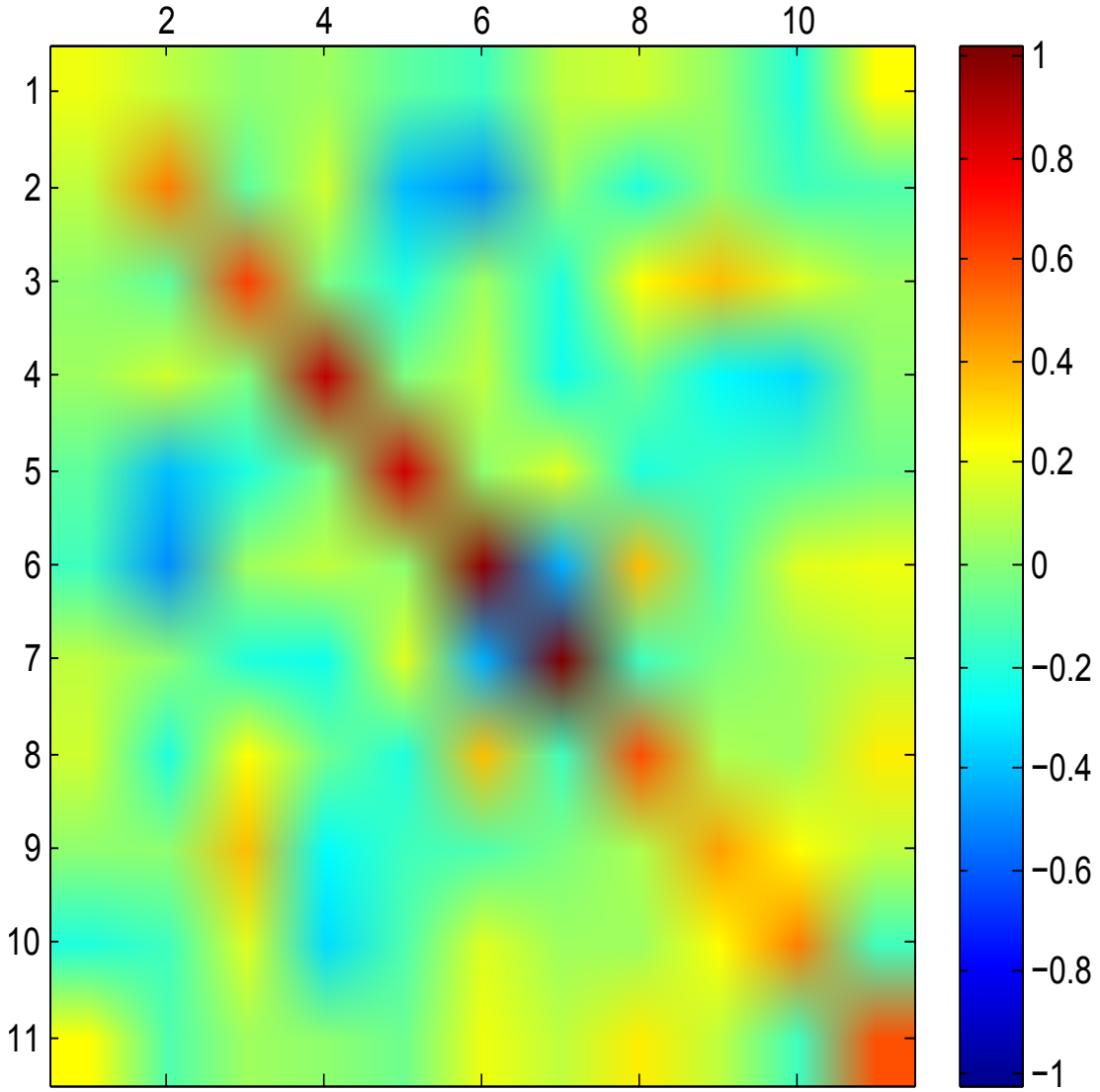
χ^2 $$S(f_i^{\text{eval}}, f^{\text{exp}}) = \sum_{j=1}^{N} \left(\frac{f_i^{\text{eval}}(E_j) - f^{\text{exp}}(E_j)}{\sigma_j^{\text{exp}}} \right)^2$$

BY $$S(f_i^{\text{eval}}, f^{\text{exp}}) = \sum_{j=1}^{N} \left(\frac{f_i^{\text{eval}}(E_j) - f^{\text{exp}}(E_j)}{\sigma_j^{\text{exp}}} \right)^2 + \sum_{k=1}^{n_p} \sum_{j=1}^{N} \left(\frac{p_k^i(E_j) - \overline{p_k(E_j)}}{\sigma_k^{\text{param}}(E_j)} \right)^2$$

HRBC $$S(f_i^{\text{eval}}, f^{\text{exp}}) = \sum_{j=1}^{N} \frac{\mathbf{I}_{[f^{\text{exp}}(E_j)-\sigma_j^{\text{exp}}; f^{\text{exp}}(E_j)+\sigma_j^{\text{exp}}]}(f_i^{\text{eval}}(E_j))}{\sigma_j^{\text{exp}}\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{f_i^{\text{eval}}(E_j) - f^{\text{exp}}(E_j)}{\sigma_j^{\text{exp}}} \right)^2 \right]$$

with $\mathbf{I}_A(x) = 1$ if $x \in A$
$\mathbf{I}_A(x) = 0$ if $x \notin A$

Example : Covariance Matrice Estimator



Outline

Distance between matrices

Classical distance:
Frobenius norm

Quality of an estimation $\widehat{\Sigma}$ of Σ

- $d_F(\widehat{\Sigma}, \Sigma) = ||\widehat{\Sigma} - \Sigma||_F = \sqrt{\text{trace}((\widehat{\Sigma} - \Sigma)^t.(\widehat{\Sigma} - \Sigma))}$

Quality of the inverse $\widehat{\Sigma}^{-1}$ for the estimation of Σ^{-1}

- $d_F(\widehat{\Sigma}^{-1}, \Sigma^{-1}) = ||\widehat{\Sigma}^{-1} - \Sigma^{-1}||_F$

Distance between matrices

Entropy and Kullback-Leibler

Quality of an estimation $\widehat{\Sigma}$ of Σ

- Entropy: $d_E(\widehat{\Sigma}, \Sigma) = \mathrm{tr}(\widehat{\Sigma}.\Sigma^{-1}) - \log(\det(\widehat{\Sigma}.\Sigma^{-1})) - N$

Quality of the inverse $\widehat{\Sigma}^{-1}$ for the estimation of Σ^{-1}

- Kullback-Leibler: $d_K(\widehat{\Sigma}, \Sigma) = \mathrm{tr}(\widehat{\Sigma}^{-1}.\Sigma) - \log(\det(\widehat{\Sigma}^{-1}.\Sigma)) - N$

James and Stein 1961, *Estimation with quadratic loss*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.

Levina & al. 2008, *Sparse estimation of large covariance matrices via a nested lasso penalty*, The Annals of Applied Statistics.

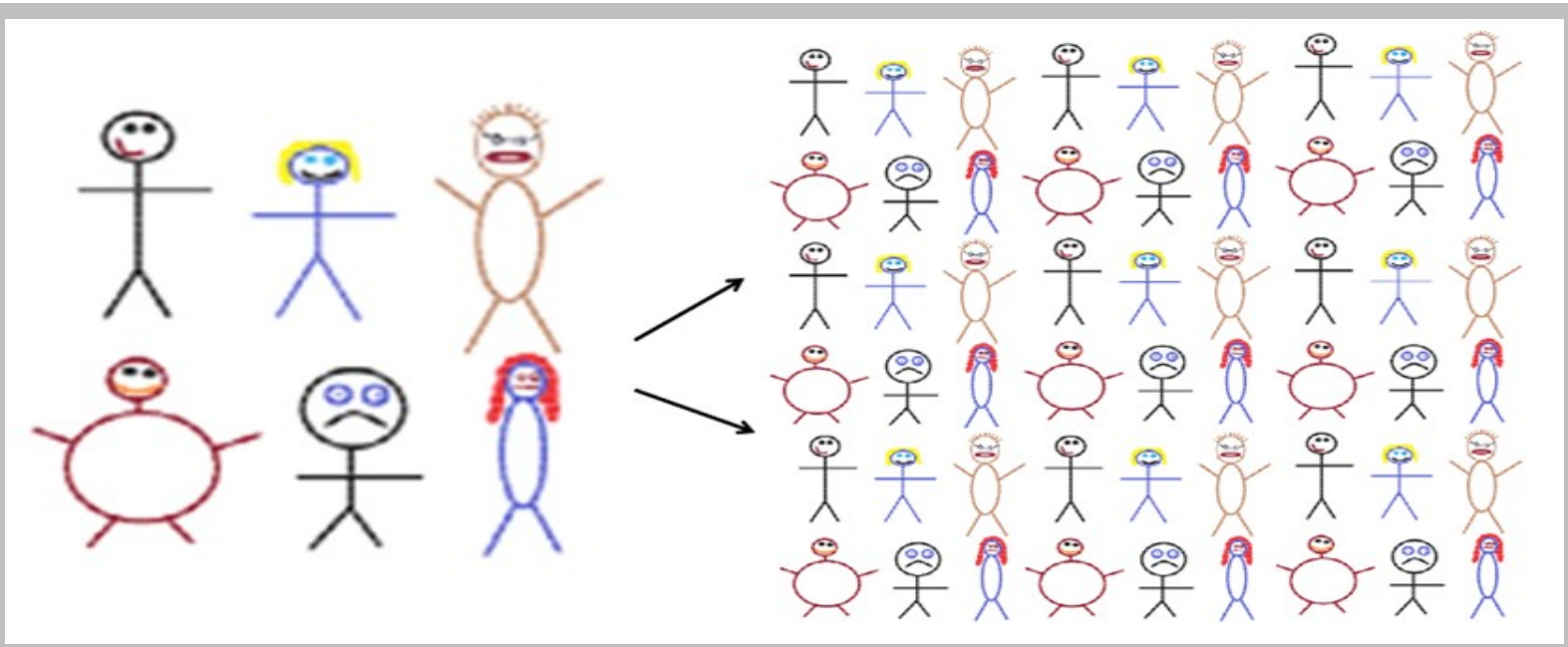
Tumminello & al. 2007, *Kullback-Leibler distance as a measure of the information filtered from multivariate data*, Phys. Review E.

Outline

Bootstrap definition

What is Bootstrap

In statistics, bootstrapping can refer to any test or metric that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy (defined in terms of bias, variance, confidence intervals, prediction error or some other such measure) to sample estimates.



Bootstrap: general principle

$$\left.\begin{array}{l} X \sim F \\ \Sigma = T(F) \end{array}\right\} \begin{array}{l} \text{what we want} \\ \text{unknown} \end{array}$$

Bootstrap: general principle

$$X \sim F$$
$$\Sigma = T(F)$$

} what we want — unknown

$$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$$
$$\text{plug-in: } \widehat{\Sigma} = T(\hat{F})$$

} what we have — observed

Example:

$(X_1, \ldots, X_n) = (1, 2, 3, 4, 5)$
(or $(X_1, \ldots, X_n) = n$ evaluations)

Bootstrap: general principle

$$X \sim F$$
$$\Sigma = T(F)$$

} what we want — unknown

$$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$$
$$\text{plug-in: } \widehat{\Sigma} = T(\hat{F})$$

} what we have — observed

$$(X_1^*, ..., X_n^*)_1 \sim \hat{F}^{\otimes n}$$
$$\hat{F}^{*1} \approx \hat{F}$$
$$\widehat{\Sigma}^{*1} = T(\hat{F}^{*1})$$

} what we do — resampling

Example:

$(X_1, \ldots, X_n) = (1, 2, 3, 4, 5)$
(or $(X_1, \ldots, X_n) = n$ evaluations)

$(X_1^*, \ldots, X_n^*)_1 = (1, 3, 4, 3, 5)$

Bootstrap: general principle

$$X \sim F$$
$$\Sigma = T(F)$$

} what we want — unknown

$$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$$
$$\text{plug-in: } \widehat{\Sigma} = T(\hat{F})$$

} what we have — observed

$$(X_1^*, ..., X_n^*)_1 \sim \hat{F}^{\otimes n} \quad , \cdots , \quad (X_1^*, ..., X_n^*)_B \sim \hat{F}^{\otimes n}$$
$$\hat{F}^{*1} \approx \hat{F} \qquad\qquad \hat{F}^{*B} \approx \hat{F}$$
$$\widehat{\Sigma}^{*1} = T(\hat{F}^{*1}) \qquad\qquad \widehat{\Sigma}^{*B} = T(\hat{F}^{*B})$$

} what we do — resampling

Example:

$$(X_1, \ldots, X_n) = (1, 2, 3, 4, 5)$$
(or $(X_1, \ldots, X_n) = n$ evaluations)

$$(X_1^*, \ldots, X_n^*)_1 = (1, 3, 4, 3, 5)$$
$$(X_1^*, \ldots, X_n^*)_2 = (4, 1, 2, 3, 4)$$
$$\vdots$$
$$(X_1^*, \ldots, X_n^*)_B = (5, 1, 3, 4, 2)$$

B resamplings allow statistics on the distribution of $\widehat{\Sigma}$:

$$\widehat{E}(\widehat{\Sigma}),\ \widehat{\mathrm{Var}}(\widehat{\Sigma}),\ \widehat{\mathrm{Bias}}(\widehat{\Sigma}),...$$

Bootstrap: general principle

$$X \sim F$$
$$\Sigma = T(F)$$

} what we want — unknown

$$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$$
$$\text{plug-in: } \widehat{\Sigma} = T(\hat{F})$$

} what we have — observed

$$(X_1^*, ..., X_n^*)_1 \sim \hat{F}^{\otimes n} \quad , \cdots , \quad (X_1^*, ..., X_n^*)_B \sim \hat{F}^{\otimes n}$$
$$\hat{F}^{*1} \approx \hat{F} \qquad\qquad \hat{F}^{*B} \approx \hat{F}$$
$$\widehat{\Sigma}^{*1} = T(\hat{F}^{*1}) \qquad\qquad \widehat{\Sigma}^{*B} = T(\hat{F}^{*B})$$

} what we do — resampling

Example:

$$(X_1, \ldots, X_n) = (1, 2, 3, 4, 5)$$
(or $(X_1, \ldots, X_n) = n$ evaluations)

$$(X_1^*, \ldots, X_n^*)_1 = (1, 3, 4, 3, 5)$$
$$(X_1^*, \ldots, X_n^*)_2 = (4, 1, 2, 3, 4)$$
$$\vdots$$
$$(X_1^*, \ldots, X_n^*)_B = (5, 1, 3, 4, 2)$$

B resamplings allow statistics on the distribution of $\widehat{\Sigma}$:

$$d_E(\widehat{\Sigma}, \Sigma) \approx \widehat{d_E}(\widehat{\Sigma}, \Sigma) = \frac{1}{B} \sum_{b=1}^{B} d_E(\widehat{\Sigma}^{*b}, \widehat{\Sigma}) \quad \text{and} \quad d_K(\widehat{\Sigma}, \Sigma) \approx \widehat{d_K}(\widehat{\Sigma}, \Sigma) = \frac{1}{B} \sum_{b=1}^{B} d_K(\widehat{\Sigma}^{*b}, \widehat{\Sigma})$$

Bootstrap: general principle

$$X \sim F$$
$$\Sigma = T(F)$$

} what we want — unknown

$$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$$
$$\text{plug-in: } \widehat{\Sigma} = T(\hat{F})$$

} what we have — observed

$$(X_1^*, ..., X_n^*)_1 \sim \hat{F}^{\otimes n} \quad , \cdots , \quad (X_1^*, ..., X_n^*)_B \sim \hat{F}^{\otimes n}$$
$$\hat{F}^{*1} \approx \hat{F} \qquad\qquad \hat{F}^{*B} \approx \hat{F}$$
$$\widehat{\Sigma}^{*1} = T(\hat{F}^{*1}) \qquad\qquad \widehat{\Sigma}^{*B} = T(\hat{F}^{*B})$$

} what we do — resampling

Example:

$(X_1, \ldots, X_n) = (1, 2, 3, 4, 5)$
(or $(X_1, \ldots, X_n) = n$ evaluations)

$(X_1^*, \ldots, X_n^*)_1 = (1, 3, 4, 3, 5)$
$(X_1^*, \ldots, X_n^*)_2 = (4, 1, 2, 3, 4)$
$\vdots$
$(X_1^*, \ldots, X_n^*)_B = (5, 1, 3, 4, 2)$

B resamplings allow statistics on the distribution of $\widehat{\Sigma}$:

$$d_E(\widehat{\Sigma}, \Sigma) \approx \widehat{d_E}(\widehat{\Sigma}, \Sigma) = \frac{1}{B} \sum_{b=1}^{B} d_E(\widehat{\Sigma}^{*b}, \widehat{\Sigma}) \quad \text{and} \quad d_K(\widehat{\Sigma}, \Sigma) \approx \widehat{d_K}(\widehat{\Sigma}, \Sigma) = \frac{1}{B} \sum_{b=1}^{B} d_K(\widehat{\Sigma}^{*b}, \widehat{\Sigma})$$

$\widehat{d_E}(\widehat{\Sigma}, \Sigma)$ and $\widehat{d_K}(\widehat{\Sigma}, \Sigma)$ are independant of the method used for the generation of $\widehat{\Sigma}$

Bootstrap: theoretical justification

Bickel and Freedman 1981

if $\sqrt{n}(T(\hat{F}) - T(F)) \xrightarrow{\mathcal{L}} P$ then $\sqrt{n}(T(\hat{F}^*) - T(\hat{F})) \xrightarrow{\mathcal{L}} P$

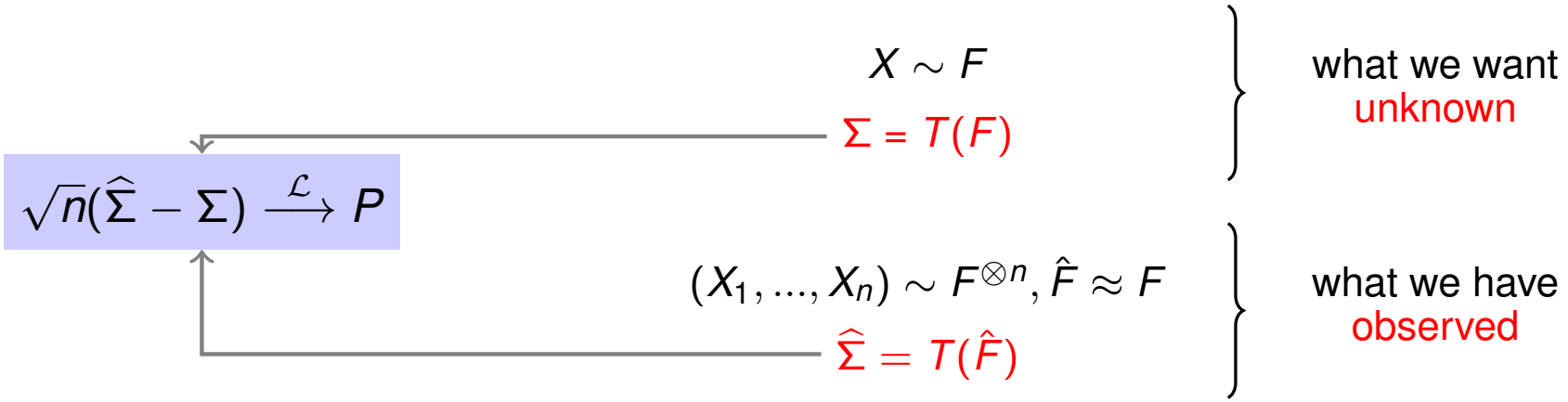
Bickel and Freedman, *Some asymptotic theory for the bootstrap*, the Annals of Statistics, 1981.

Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, 1981.

Bootstrap: theoretical justification

Bickel and Freedman 1981

if $\sqrt{n}(T(\hat{F}) - T(F)) \xrightarrow{\mathcal{L}} P$ then $\sqrt{n}(T(\hat{F}^*) - T(\hat{F})) \xrightarrow{\mathcal{L}} P$



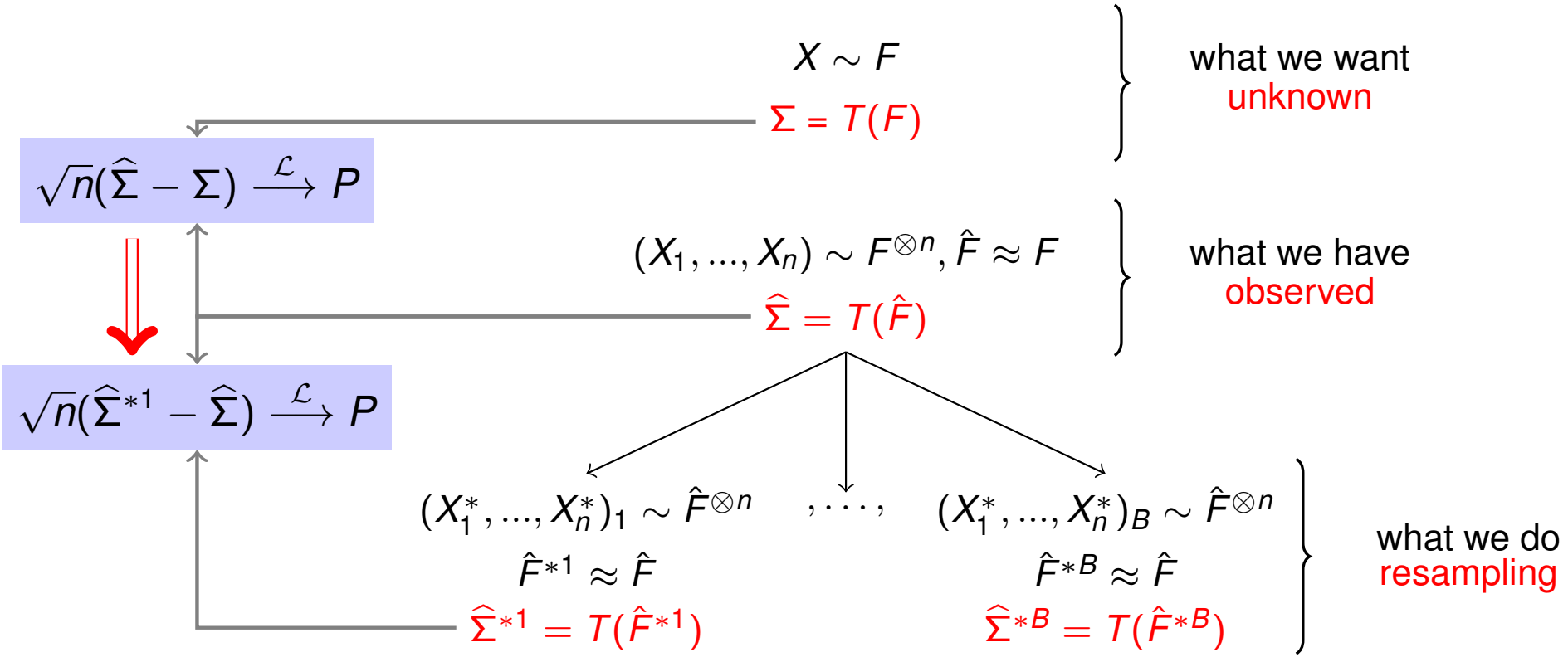
Bickel and Freedman, *Some asymptotic theory for the bootstrap*, the Annals of Statistics, 1981.

Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, 1981.

Bootstrap: theoretical justification

Bickel and Freedman 1981

if $\sqrt{n}(T(\hat{F}) - T(F)) \xrightarrow{\mathcal{L}} P$ then $\sqrt{n}(T(\hat{F}^*) - T(\hat{F})) \xrightarrow{\mathcal{L}} P$



Bickel and Freedman, *Some asymptotic theory for the bootstrap*, the Annals of Statistics, 1981.

Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, 1981.

Bootstrap: theoretical justification

Beran and Strivastava 1985

with $\hat{F}$ the empirical CDF, $\sqrt{n}(\widehat{\Sigma} - \Sigma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$ and $\sqrt{n}(\widehat{\Sigma}^* - \widehat{\Sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$

Beran and Srivastavia, *Bootstrap tests and confidence region for functions of covariance matrix*, The Annals of Statistics, 1985.

Bootstrap: theoretical justification

Beran and Strivastava 1985

with $\hat{F}$ the empirical CDF, $\sqrt{n}(\widehat{\Sigma} - \Sigma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$ and $\sqrt{n}(\widehat{\Sigma}^* - \widehat{\Sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$

$X \sim F$

$\Sigma = T(F)$

what we want
unknown

$\sqrt{n}(\widehat{\Sigma} - \Sigma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$

$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$

$\widehat{\Sigma} = T(\hat{F})$

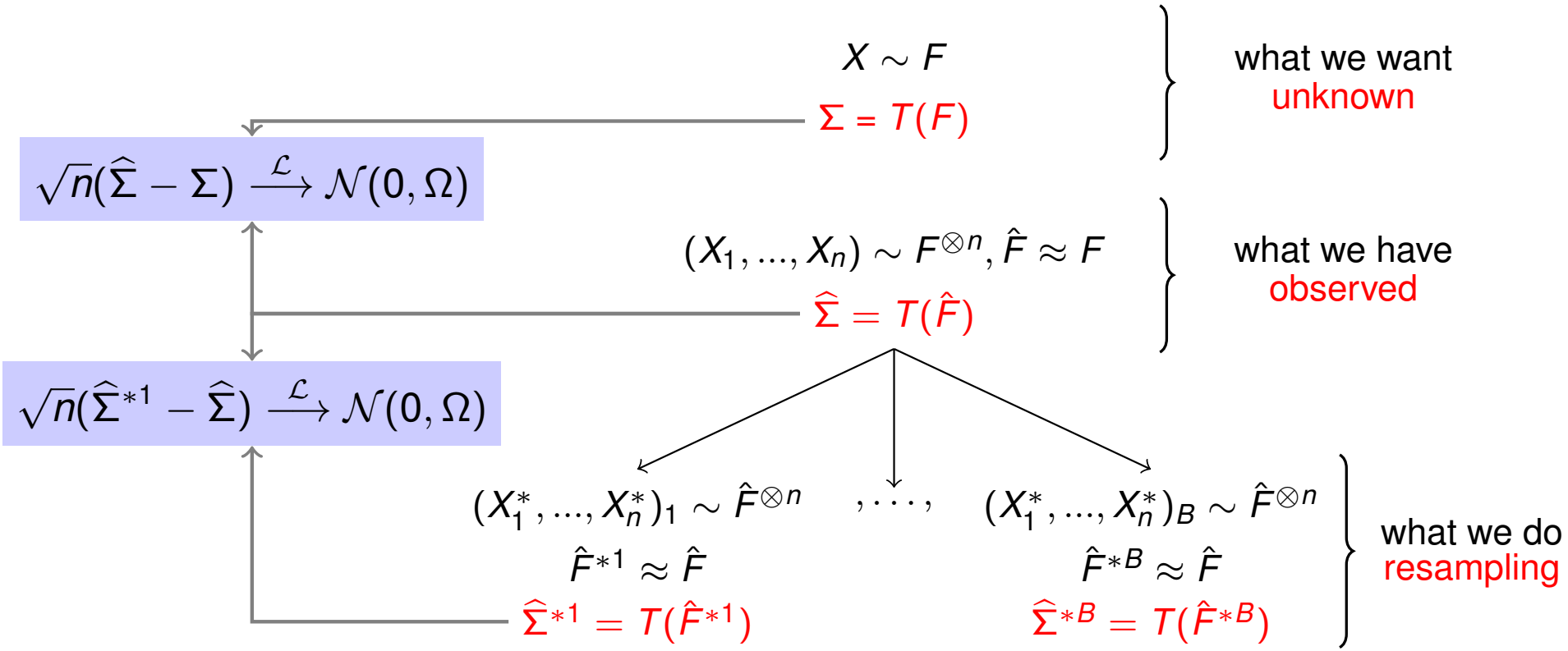
what we have
observed

Beran and Srivastavia, *Bootstrap tests and confidence region for functions of covariance matrix*, The Annals of Statistics, 1985.

Bootstrap: theoretical justification

Beran and Strivastava 1985

with $\hat{F}$ the empirical CDF, $\sqrt{n}(\widehat{\Sigma} - \Sigma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$ and $\sqrt{n}(\widehat{\Sigma}^* - \widehat{\Sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$



Beran and Srivastavia, *Bootstrap tests and confidence region for functions of covariance matrix*, The Annals of Statistics, 1985.

Bootstrap: theoretical justification

Beran and Strivastava 1985

with $\hat{F}$ the empirical CDF, $\sqrt{n}(\widehat{\Sigma} - \Sigma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$ and $\sqrt{n}(\widehat{\Sigma}^* - \widehat{\Sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega)$

$X \sim F$

$\Sigma = T(F)$

} what we want unknown

$(X_1, ..., X_n) \sim F^{\otimes n}, \hat{F} \approx F$

$\widehat{\Sigma} = T(\hat{F})$

} what we have observed

$(X_1^*, ..., X_n^*)_1 \sim \hat{F}^{\otimes n}$, ..., $(X_1^*, ..., X_n^*)_B \sim \hat{F}^{\otimes n}$

$\hat{F}^{*1} \approx \hat{F}$ $\quad$ $\hat{F}^{*B} \approx \hat{F}$

$\widehat{\Sigma}^{*1} = T(\hat{F}^{*1})$ $\quad$ $\widehat{\Sigma}^{*B} = T(\hat{F}^{*B})$

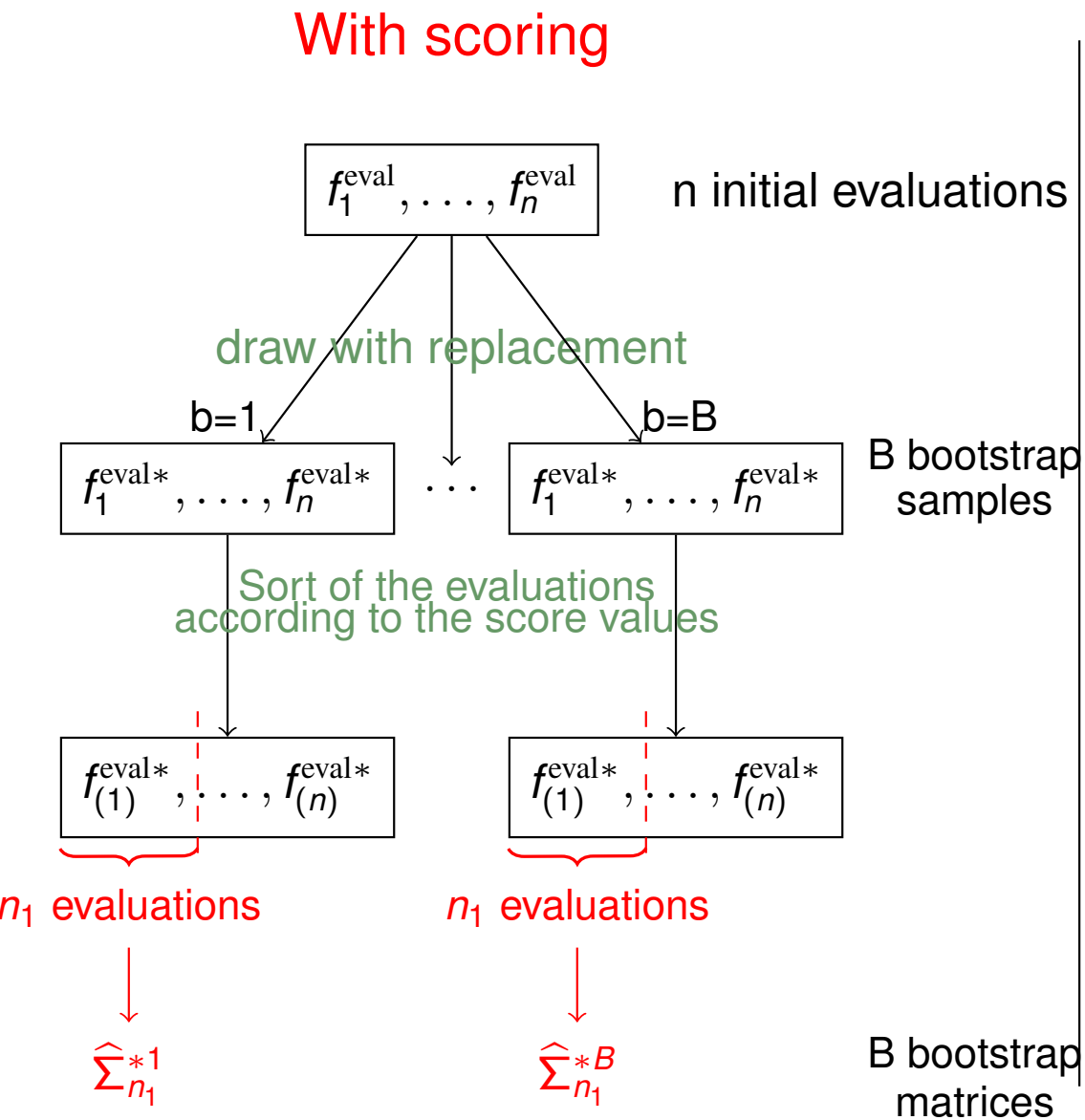
} what we do resampling

$d_E(\widehat{\Sigma}, \Sigma)$

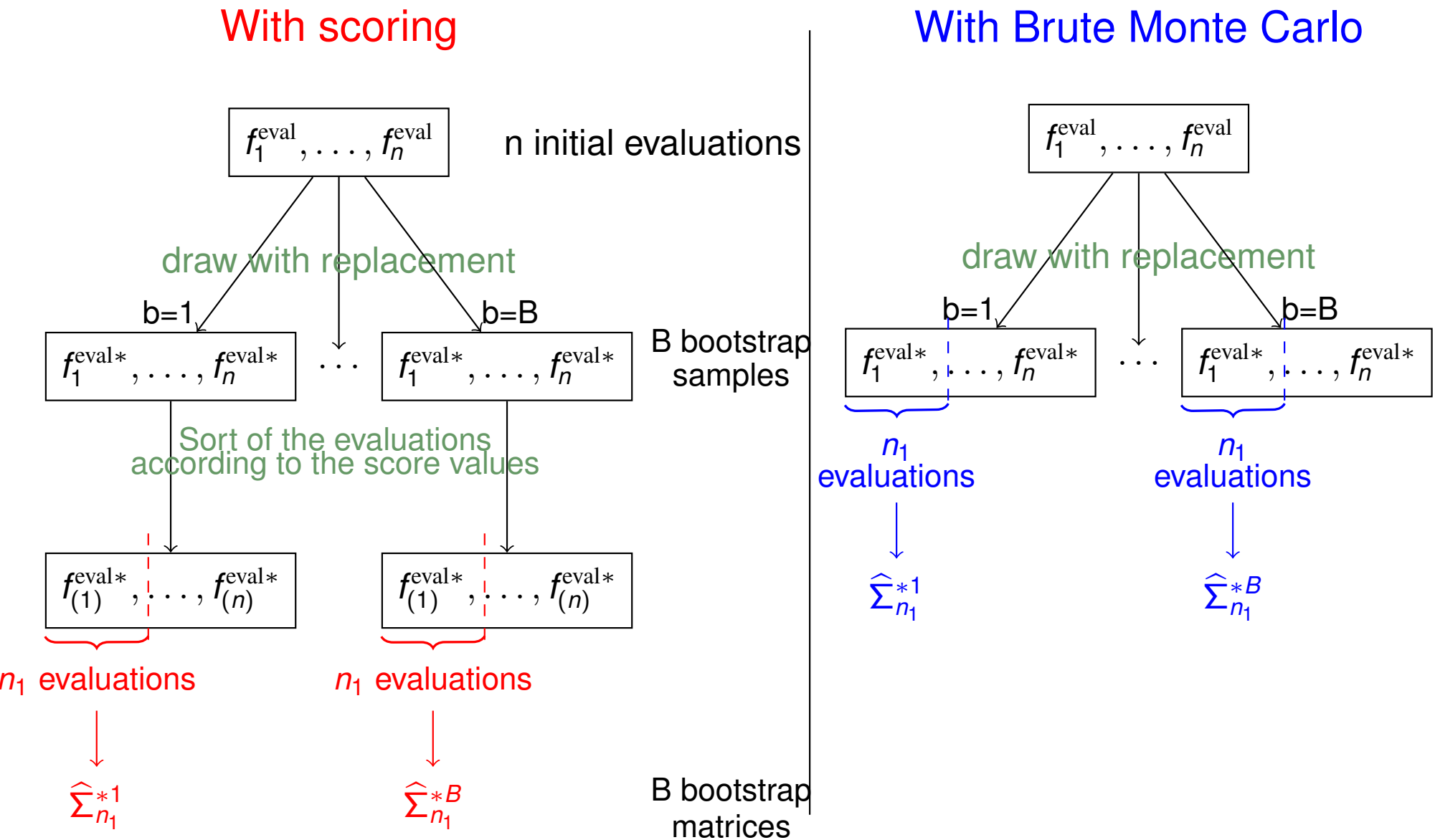
$d_E(\widehat{\Sigma}^*, \widehat{\Sigma})$

Beran and Srivastavia, *Bootstrap tests and confidence region for functions of covariance matrix*, The Annals of Statistics, 1985.

Bootstrap implementation

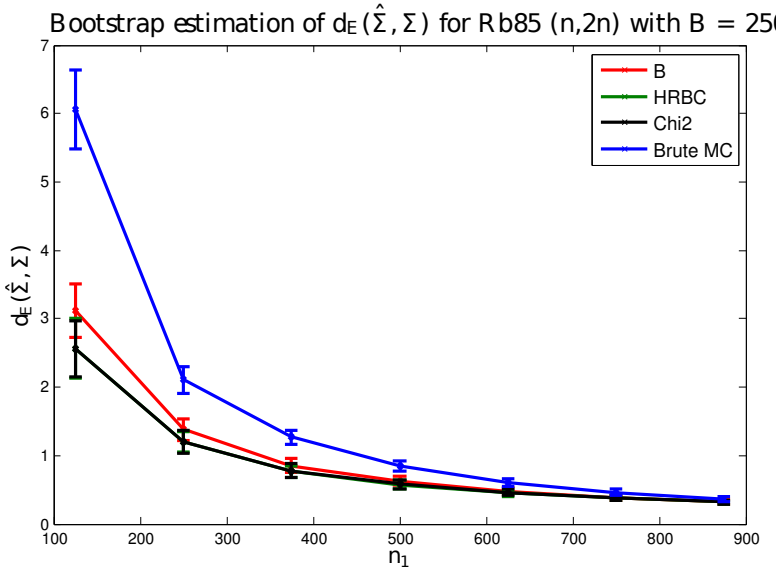


Bootstrap implementation

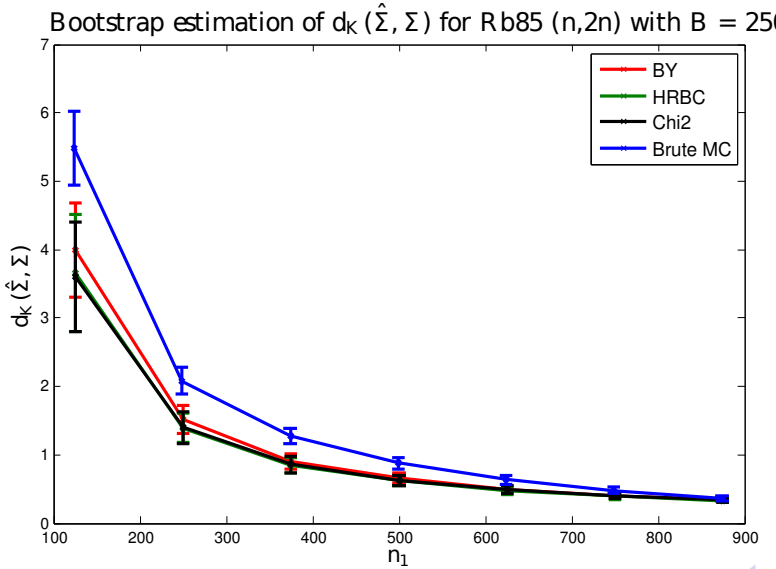


Entropy and Kullback-Leibler versus n_1

$\widehat{d_E}$



$\widehat{d_K}$



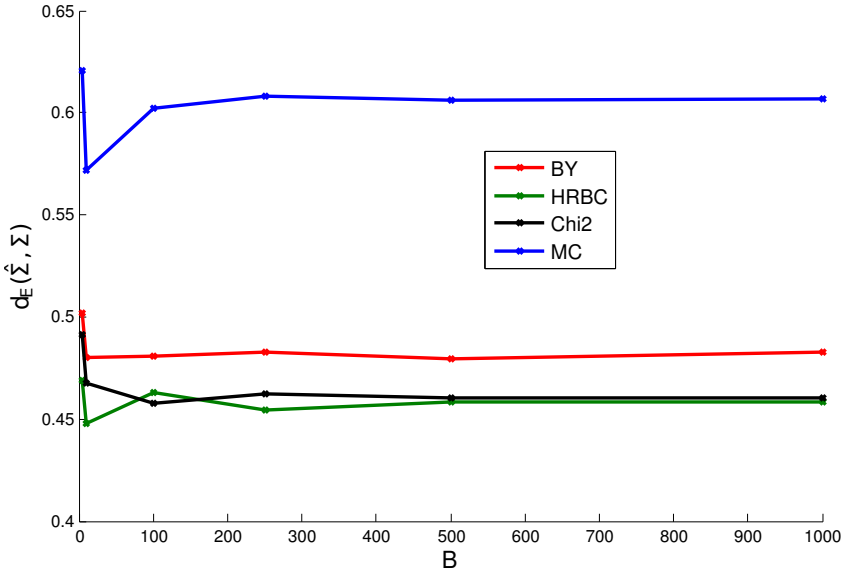
Entropy ratio and Kullback-Leibler ratio versus B

$n_1 = 625$

Local scoring

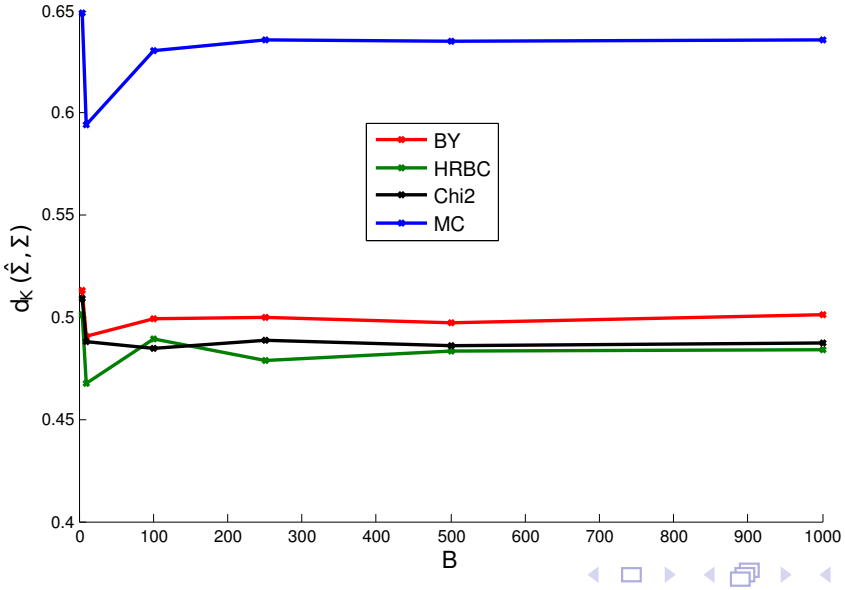
$d_E(\widehat{\Sigma}, \Sigma)$

for Rb85 (n,2n)



$d_K(\widehat{\Sigma}, \Sigma)$

for Rb85 (n,2n)



En. ratio and K-L ratio versus param. variation

$B = 500$,
$n_1 = 625$

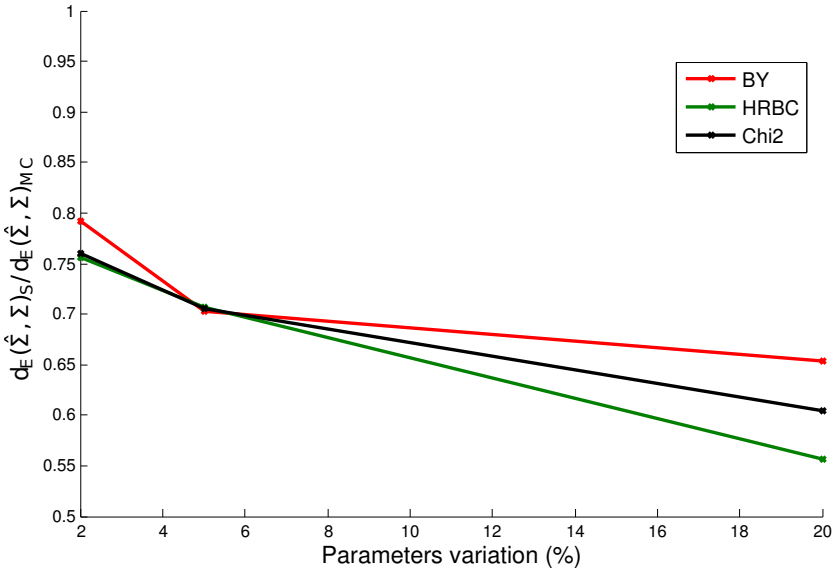
Local scoring

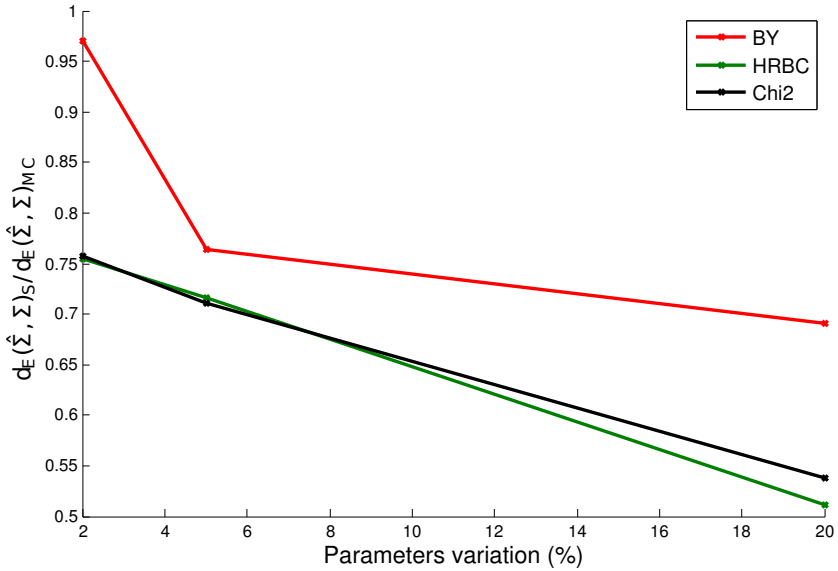
Global scoring ((n,2n),(n,γ),(n,α))

$\frac{d_E(\widehat{\Sigma}, \Sigma)_S}{d_E(\widehat{\Sigma}, \Sigma)_{MC}}$

for Rb85

(n,2n),

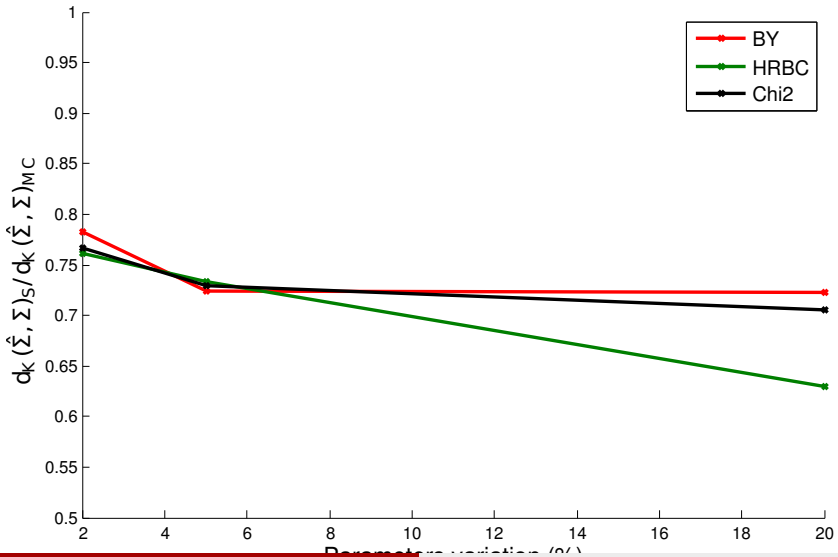


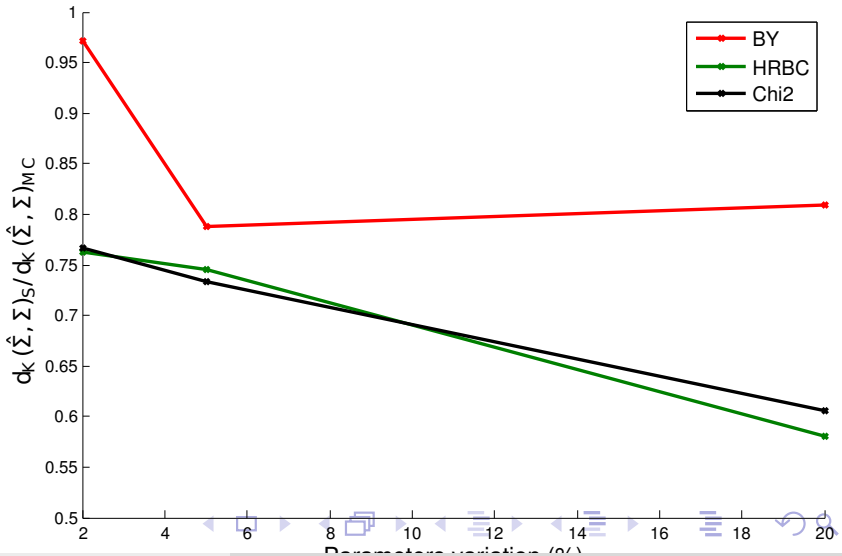


$\frac{d_K(\widehat{\Sigma}, \Sigma)_S}{d_K(\widehat{\Sigma}, \Sigma)_{MC}}$

for Rb85

(n,2n)





References


A. J. Koning, J. P. Delaroche, *Local and global nucleon optical models from 1 keV to 200 MeV*, Nucl. Phys. A713, 231 (2003).


O. Bersillon, *The computer Code SCAT2000*, Workshop on Applied Nuclear Theory and Nuclear Model Calculations for Nuclear Technology Applications, M.K. Mehta and J.J.Schmidt editors, 1988.


S. Wiener, *The Homogeneous Chaos*, American Journal of Mathematics, (1938).


R. Ghanem and P. Spanos. *Stochastic finite elements : a spectral approach.* Springer, Berlin, 1991.


O. Le Maître, O. Knio, H. Najm and R. Ghanem, *Multi-resolution analysis of Uncertainty propagation schemes using Wiener-type expansions*, Journal of Computational Physics, (2003).


R.H. Cameron, W.T. Martin, *The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals*, Ann.Math., 48 : 385-392 (1947).


B. Sudret, *Global sensitivity analysis using polynomial chaos expansion*, 2007.


M. B. Chadwick, T. Kawano, P. Talou, E. Bauge, S. Hilaire, P. Dossantos-Uzarralde, P. E. Garrett, *Yttrium ENDF/B-VII Data from Theory and LANSCE/GEANIE Measurements and Covariqnces Estimed using Bayesian and Monte-Carlo Methods* Nuclear Data Sheets 108 (2007) 2742 - 2755


P. Dossantos-Uzarralde, A. Guittet *A Polynomial Chaos Approach for Nuclear Data Uncertainties Evaluations* Nuclear Data Sheets 109 (2008) 2894 - 2899