Bayesian calibration

Pierre BARBILLON

AgroParisTech / INRA MIA UMR 518

11/06/13

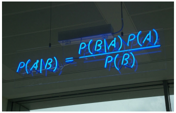
Journée Utilisateurs Open TURNS

Outline

Outline

History

- Thomas Bayes (1702-1761) 

  Posthumous publication of the formula in 1763 

- Pierre Simon Laplace (1749-1827) in 1771 independently from Bayes, discovered the formula and used it for inference.
  All causes assumed to be equally likely.

- Only possible for conjugate prior until computers !

- 1990's : Monte Carlo Markov Chain samplers for posterior distribution.

Bayes' formula

A and E two events such that $\mathbb{P}(E) > 0$:

$$\mathbb{P}(A|E) = \frac{\mathbb{P}(E|A) \cdot \mathbb{P}(A)}{\mathbb{P}(E|A) \cdot \mathbb{P}(A) + \mathbb{P}(E|A^c) \cdot \mathbb{P}(A^c)} = \frac{\mathbb{P}(E|A) \cdot \mathbb{P}(A)}{\mathbb{P}(E)}$$

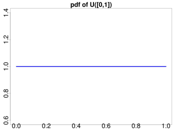
Symmetric formula but seen as an inversion of cause and consequence

- A cause unobserved, E consequence observed.
- $\mathbb{P}(E|A)$ and $\mathbb{P}(E|A^c)$ assumed by the stochastic model.
- $\mathbb{P}(A)$ and $\mathbb{P}(A^c)$ prior belief on the cause (equally probable ?)
- Inference on $\mathbb{P}(A|E)$ posterior: update on our knowledge on A given the observations/data.

An example from Laplace 1786

Proportion of male births in Paris ? Is it $p = 1/2$?

Data: $N_m = 251,527$ male births and $N_f = 241,945$ female births

Prior: $p \sim \pi(p) = \mathcal{U}([0,1])$ 

Stochastic Model: $N_m \sim \mathcal{B}(N_m + N_f = N_T, p) = \pi(N_m|p)$

Posterior:

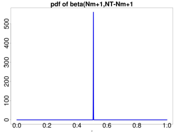
$$\begin{aligned} \pi(p|N_m) &= \frac{\pi(N_m|p) \cdot \pi(p)}{\pi(N_m)} \propto \pi(N_m|p) \cdot \pi(p) \\ &\propto C_{N_T}^{N_m} p^{N_m} (1-p)^{N_T - N_m} \cdot \mathbb{I}_{[0,1]}(p) \end{aligned}$$

An example from Laplace 1786

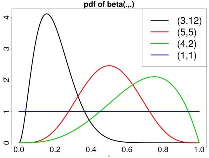
Posterior:

$$\pi(p|N_m) \propto p^{N_m}(1-p)^{N_T-N_m}\mathbb{I}_{[0,1]}(p)$$

Loi *beta*(N_m+1, N_T-N_m+1)



- More than an point-estimate a distribution.
  A possible choice: $\mathbb{E}(p|N_m) = \frac{N_m+1}{N_T+2} = 50.97\%$.
  But also, $\mathbb{P}(p < 1/2) = 1.1 \cdot 10^{-42}$.
- Weight of prior information: 2 compare with weight of data N_T !
- Explicit posterior distribution $\rightarrow$ Conjugate prior Beta-Binomial model !



$$\begin{aligned}\pi(p) &= beta(\alpha, \beta)\\ \pi(p|N_m) &= beta(N_m+\alpha, N_T-N_m+\beta)\end{aligned}$$

Likelihood function

Data:
$\mathbf{x} = (x_1, \ldots, x_n)$ observed

Stochastic model:
$\mathbf{x}$ is assumed to be a realisation of a random variable $\mathbf{X}$ with distribution $\mathbb{P}_\theta$ where $\theta \in \Theta \subset \mathbb{R}^d$ is unknown.

$f(\mathbf{x}|\theta)$ is "the probability of the data for a given θ":

- If $\mathbb{P}_\theta$ is discrete, $f(\mathbf{x}|\theta) = \mathbb{P}_\theta(\mathbf{X} = \mathbf{x})$,
- If $\mathbb{P}_\theta$ is continuous, $f(\mathbf{x}|\theta)$ is the density function with respect to the Lebesgue measure.

Likelihood:
Information on θ brought by $\mathbf{x}$:

$$l(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)\,.$$

Remark: If $\mathbf{x} = (x_1, \ldots, x_n)$ i.i.d. with model $f(\cdot|\theta)$ then:

$$l(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta)\,.$$

Formalism

- **Data:** $\mathbf{x} = (x_1, \ldots, x_n)$.
- **Stochastic Model:** $f(\mathbf{x}|\theta)$ where $\theta \in \Theta \subset \mathbb{R}^d$ (parametric statistic).
  likelihood function seen as a function of θ.

$$l(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

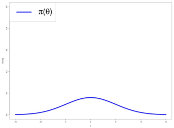
- **Prior:** $\theta \sim \pi(.)$.
- **Posterior:** by application of Bayes Formula:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta) \cdot \pi(\theta)}{\int_\Theta l(\theta|\mathbf{x})\pi(\theta)d\mathbb{P}(\theta)} = \frac{l(\theta|\mathbf{x}) \cdot \pi(\theta)}{\int_\Theta l(\theta|\mathbf{x})\pi(\theta)d\mathbb{P}(\theta)} .$$

  Normalizing constant: $\int_\Theta l(\theta|\mathbf{x})\pi(\theta)d\mathbb{P}(\theta)$
  - hard to compute in non-conjugate models,
  - depends only on $\mathbf{x}$.

An illustration of the Bayes machine

Prior
Expert judgment
previous knowledge



$\Rightarrow$

Modelisation
$f(\mathbf{x}|\theta) = l(\theta|\mathbf{x})$
$\Downarrow$

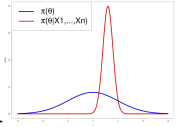
Bayes Formula

$$\pi(\theta|\mathbf{x}) = \frac{l(\theta|\mathbf{x}) \cdot \pi(\theta)}{\int_\Theta l(\theta|\mathbf{x})\pi(\theta)d\mathbb{P}(\theta)} \, .$$

$\Uparrow$
$\mathbf{x} = (x_1, \ldots, x_n)$
data

posterior

$\Rightarrow$



Remark: Can be used sequentially.

Main features of Bayesian statistic

- Allows to incorporate in prior distribution:
  - knowledge from previous experiments or experiments in different conditions,
  - experts' knowledge.

- Provides a distribution on the unknown parameters given the data and prior information $\Rightarrow$ Take naturally into account uncertainty on estimation in a non-asymptotic context.
  Credible interval can be obtained from the posterior distribution.

Topics in Bayesian statistic

- Choice of a prior
  - Elicitation Subjective Bayes
  - Non informative Prior Objective Bayes
  - Sensitivity to the prior

- Decision theory: choose of a cost / utility function ?
  Obtaining a point-estimate.
- Non parametric Bayesian statistic.

- Inference concerns
  Next section !

Outline

Plan

If no conjugate prior ?

Solution: Simulate according to the posterior distribution: $\pi(\theta|\mathbf{x})$ and using Monte Carlo techniques.
$\Rightarrow$ posterior distribution only known through a sample (as big as you want/need).

Main difficulties: Normalizing constant: $\int_\Theta l(\theta|\mathbf{x})\pi(\theta)d\mathbb{P}(\theta)$ is not tractable
Posterior distribution is known up to a constant:

$$\pi(\theta|\mathbf{x}) \propto l(\theta|\mathbf{x}) \cdot \pi(\theta)$$

Monte Carlo Markov Chain (MCMC)

Principle

Generate a Markov Chain $(\theta^t)_{t\in\mathbb{N}}$ with stationary distribution $\pi(\theta|\mathbf{x})$.

If enough iterations of the chain, the chain

- forgets its initial state,
- is distributed according to $\pi(\cdot|\mathbf{x})$
- Estimation: for any h measurable $\frac{1}{T}\sum_t h(\theta^t) \overset{p.s.}{\rightarrow} \mathbb{E}_{\pi(\cdot|\mathbf{x})}(h(\theta))$ (Ergodic Theorem)

2 well-known algorithms

- Metropolis-Hastings algorithm
- Gibbs sampler

Plan

Metropolis-Hastings algorithm

Initialisation: Generate θ^0 from a starting distribution (prior for example)
Algorithm : iterations $t = 1, \ldots$:

- propose $\tilde{\theta}^{t+1}$ according to $q(\theta^t, \cdot)$,
- Compute $\alpha(\theta^t, \tilde{\theta}^{t+1}) = \begin{cases} \min\left(\frac{\pi(\tilde{\theta}^{t+1}|\mathbf{x})}{\pi(\theta^t|\mathbf{x})} \frac{q(\tilde{\theta}^{t+1}, \theta^t)}{q(\theta^t, \tilde{\theta}^{t+1})}, 1\right), & \pi(\theta^t|\mathbf{x}) > 0 \\ 1 & \pi(\theta^t|\mathbf{x}) = 0 \end{cases}$,
- $\theta^{t+1} = \begin{cases} \tilde{\theta}^{t+1} & \text{with probability } \alpha(\theta^t, \tilde{\theta}^{t+1}) \\ \theta^t & \text{otherwise} \end{cases}$.

Remark : Note that

$$\frac{\pi(\tilde{\theta}^{t+1}|\mathbf{x})}{\pi(\theta^t|\mathbf{x})} = \frac{l(\tilde{\theta}^{t+1}|\mathbf{x}) \cdot \pi(\theta)}{l(\theta^t|\mathbf{x}) \cdot \pi(\theta)}$$

simplification of the normalising constant.

If $q(\theta, \tilde{\theta}) = q(\tilde{\theta}, \theta)$ (symmetric distribution),

$$\alpha(\theta, \tilde{\theta}) = \begin{cases} \min\left(\frac{\pi(\tilde{\theta}^{t+1}|\mathbf{x})}{\pi(\theta^t|\mathbf{x})}, 1\right), & \pi(\theta^t|\mathbf{x}) > 0 \\ 1 & \pi(\theta^t|\mathbf{x}) = 0 \end{cases}.$$

Metropolis-Hastings algorithm

Choice of the kernel $q(\cdot,\cdot)$: often a random walk

- $\tilde{\theta}^{t+1} \sim \theta^t + \mathcal{N}(0, \sigma^2 Id)$,
- $\tilde{\theta}^{t+1} \sim \theta^t + \mathcal{U}[-\gamma, \gamma]$.

Produce a sample $(\theta_M, \ldots, \theta_N)$:

- $M-1$ first iterations not used: burnin of the chain,
- Not independent (autocorrelation),
- distribution of $(\theta_M, \ldots, \theta_N)$ should be close to $\pi(\cdot|\mathbf{x})$.

Efficiency of the algorithm:

- Fast convergence toward the stationary distribution ?
- Small autocorrelation ?
- Good choice of $q(\cdot,\cdot)$?
  Link with the acceptation rate: proportion of accepted transition $\theta^{t+1} = \tilde{\theta}^{t+1}$.

An example

Metropolis-Hastings with random walk $\tilde{\theta}^{t+1} \sim \theta^t \mathcal{U}[-\gamma, \gamma]$ for $\pi = \mathcal{N}(0,1)$.

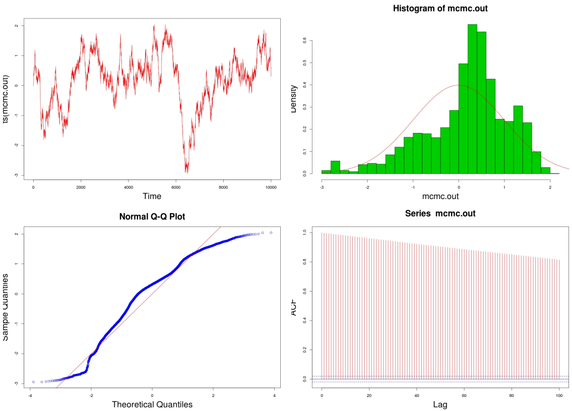


Figure: $\gamma = 0.1$, high acceptation rate (97%), bad mixing

An example

Metropolis-Hastings with random walk $\theta^{t+1} \sim \mathcal{U}[-\gamma, \gamma]$ for $\pi = \mathcal{N}(0, 1)$.

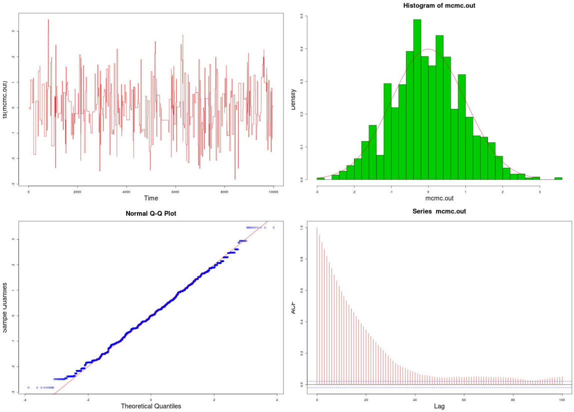


Figure: $\gamma = 40$, low acceptation rate (3.7%), bad mixing

An example

Metropolis-Hastings with random walk $\theta^{t+1} \sim \mathcal{U}[-\gamma, \gamma]$ for $\pi = \mathcal{N}(0, 1)$.

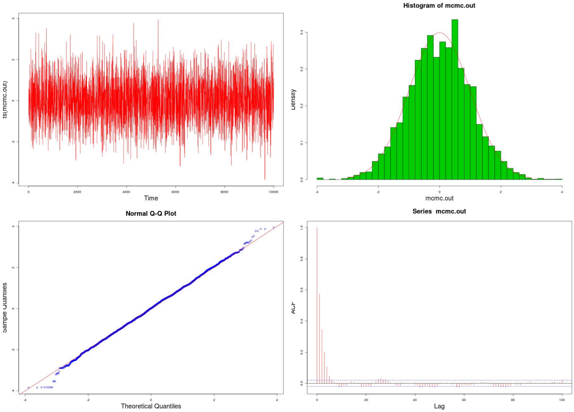


Figure: $\gamma = 5$, medium acceptation rate (38%), *good* mixing

Morality: Goldilocks' principle

Acceptation rate should be:

- not too big (small γ),
- not too small (large γ)....

Optimality: In particular situation theoretical works (Roberts et al., 1997) show optimal acceptation rate is **0.234**.
Often ok if in $[0.1; 0.6]$.

Plan

Gibbs sampler: conditional simulations

Generate $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ according to $\pi(\cdot|\mathbf{x})$ by using conditional distributions: $\pi(\cdot|\mathbf{x}, \boldsymbol{\theta}_{-j})$ where $\boldsymbol{\theta}_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_d)$.

Initialisation: Generate $\boldsymbol{\theta}^0$ from a starting distribution (prior for example).

Algorithm: Iterations $t = 1, \ldots$, for $j = 1, \ldots, d$
sample θ_j^{t+1} from:

$$\pi(\cdot|\mathbf{x}, \theta_1^{t+1}, \ldots, \theta_{j-1}^{t+1}, \theta_{j+1}^{t}, \ldots, \theta_d^{t})$$

Gibbs sampler: an example

Gaussian model:
$X_i \sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$, i.i.d.
($\boldsymbol{\theta} = (\mu, \sigma^2)$):

$$f(\mathbf{x}|\mu, \sigma^2) = 1/(2\pi\sigma^2)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i - \mu)^2\right)$$

Prior:
Mean $\mu \sim \mathcal{N}(m_0, \sigma_0^2)$
Variance $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$ (inverse Gamma)

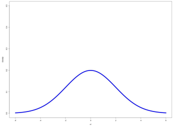
Posterior:
$\mu|\sigma^2, \mathbf{x} \sim \mathcal{N}(M, \Sigma^2)$ where
$M = \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\frac{1}{n}\left(\sum_i x_i\right) + \frac{\sigma^2}{n\sigma_0^2+\sigma^2} m_0$ and $\Sigma^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2+n\sigma_0^2}$

$\sigma^2|\mu, \mathbf{x} \sim \mathcal{IG}\left(\frac{n}{2} + \alpha, \frac{1}{2}\sum\left(x_i - \mu\right)^2 + \beta\right))$

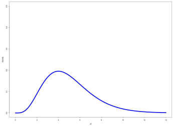
Gibbs sampler: an example

Prior:

$\mu \sim \mathcal{N}(0, 4)$



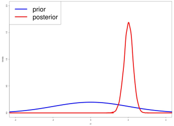
$\sigma^2 \sim \mathcal{IG}(5, 1)$



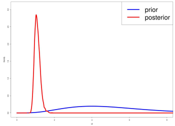
Data: 20 i.i.d. $x_i \sim \mathcal{N}(2, 1)$.

Gibbs sampling $\Rightarrow$ **Posterior**

μ



σ^2



Outline

Plan

Calibration of a computer code

Computer experiments:

Evaluations of the computer model (simulator) $f(\mathbf{x}, \boldsymbol{\theta}) \in \mathbb{R}^s$ where

- **physical parameters**: $\mathbf{x} \in \mathbb{R}^m$ observable and often controllable inputs
- **simulator parameters** $\boldsymbol{\theta} \in \mathbb{R}^d$ non-observable parameters, required to run the simulator.

  2 types:
  - "calibration parameters": physical meaning but unknown, necessary to make the code mimic the reality,
  - "tuning parameters": no physical interpretation.

Goal:
Calibrate the code: finding "best" or "true" $\boldsymbol{\theta}$ from real observations.

Observations / Data: For different inputs: $\mathbf{x}_1, \ldots, \mathbf{x}_n$
Noisy (measurement error) observations of reality: $y_1 = y(\mathbf{x}_1), \ldots, y_n = y(\mathbf{x}_n)$

Usual framework

Hypotheses:

- Observations are noisy realisations of a physical system:

$$y_i = \zeta(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \dots, n\,.$$

  - physical system does not depend on $\boldsymbol{\theta}$,
  - sometimes, distribution of the measurement errors ϵ_i is treated as known.

- Relationship between the simulator and the data

$$y(\mathbf{x}_i) = f(\mathbf{x}_i, \boldsymbol{\theta}^*) + \delta(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \dots, n\,.$$

  - $\boldsymbol{\theta}^*$ denotes the true parameter,
  - $\delta(\cdot)$ is the discrepancy between the simulator and reality: sometimes negligible.

A calibration example

Hypotheses:

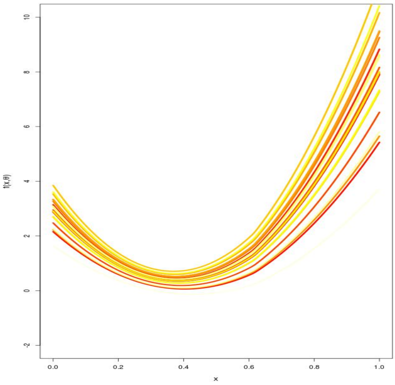
- The simulator represents sufficiently well the physical system:

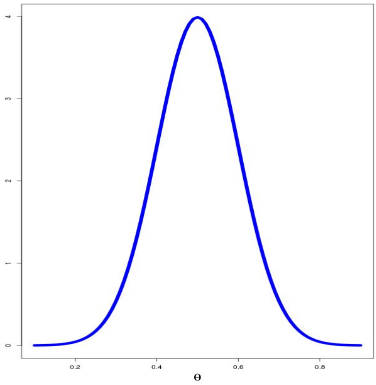
$$y(\mathbf{x}_i) = f(\mathbf{x}_i, \boldsymbol{\theta}^*) + \epsilon_i, \quad i = 1, \ldots, n.$$

- But unknown $\boldsymbol{\theta}^*$.
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ i.i.d. with known σ^2.

A calibration example

Prior:
prior distribution on unknown $\boldsymbol{\theta}$: $\pi(\cdot)$
from expert judgment, past experiments...
Possible choice $\pi(\boldsymbol{\theta}) = \mathcal{N}(\theta_0, \sigma_0^2)$.





A calibration example

Data:
Couples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ from physical experiments.

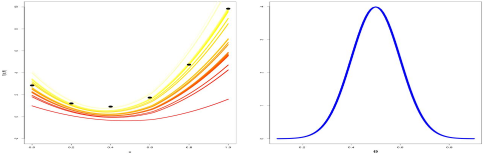
Posterior distribution:

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{y}) &\propto l(\boldsymbol{\theta}|\mathbf{y}) \cdot \pi(\boldsymbol{\theta}) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y(\mathbf{x}_i) - f(\mathbf{x}_i, \boldsymbol{\theta}))^2 - \frac{1}{2\sigma_0^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^2\right) \end{aligned}$$

- Analytical posterior if $\boldsymbol{\theta} \mapsto f(\mathbf{x}, \boldsymbol{\theta})$ is a linear map,
- Otherwise MCMC sampling to simulate according to the posterior distribution.

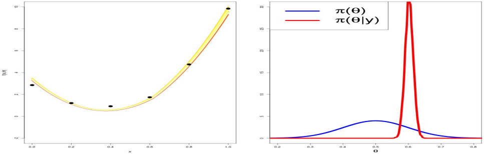
A calibration example

Prior with data:
$(\theta^* = 0.6)$



$\Downarrow$ Metropolis-Hastings algorithm $\Downarrow$

Posterior on θ:



More details on the MH algorithm

Initialisation:
θ^0 chosen.

Update:
iterations $t = 1, \ldots,$

1. Proposal: $\tilde{\theta}^{t+1} = \theta^t + \mathcal{N}(0, \tau^2)$.
2. Compute

$$\alpha(\theta^t, \tilde{\theta}^{t+1}) = \frac{\pi(\tilde{\theta}^{t+1}|\mathbf{y})}{\pi(\theta^t|\mathbf{y})}$$

3. Acceptation:

$$\theta^{t+1} = \begin{cases} \tilde{\theta}^{t+1} & \text{with probability } \alpha(\theta^t, \tilde{\theta}^{t+1}) \\ \theta^t & \text{otherwise.} \end{cases}$$

Note that the ratio $\alpha(\theta^t, \tilde{\theta}^{t+1})$ needs several computations of $f(\mathbf{x}, \boldsymbol{\theta})$ at each step since

$$\pi(\theta|\mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y(\mathbf{x}_i) - f(\mathbf{x}_i, \boldsymbol{\theta}))^2 - \frac{1}{2\sigma_0^2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^2\right).$$

Plan

Expensive black-box computer code

- Run the simulator for a given $(\mathbf{x}, \theta)$ is time-consuming / expensive.
- The simulator is a black-box, no intrusive methods are possible.

$\Rightarrow$ Only few runs of the simulator are possible then we cannot apply the MH algorithm.

Using an emulator / metamodel / coarse model / approximation of the simulator which is fast to compute, but:

- loss on precision of prediction,
- new uncertainty source: accuracy of the model approximation ?
- how to take it into account ?

Emulator using Gaussian Process:

- Very popular in computer experiments.
- integrated in a Bayesian framework: appears in the likelihood function and a prior on the parameters of the Gaussian process are chosen.
- model uncertainty coming from approximation of f.

Prior distribution on f:

- Gaussian process F with given mean and covariance function (a priori on the regularity of the function).
- Prior distribution on ϕ hyperparameters of the Gaussian process.

Posterior distribution on f:
Given some evaluations of f: $\mathbf{y}^c$.
$\Rightarrow$ Gaussian process F conditioned to interpolate these evaluations (and conditioned to ϕ), still a Gaussian process.

- Mean of the conditioned process approximates $f(\mathbf{x})$,
- Variance of the conditioned process provides a measure of uncertainty on this approximation.

Gaussian process emulator: illustration

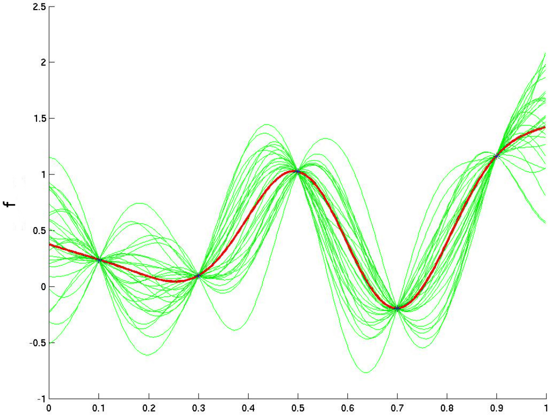


Figure: Posterior mean and realisations of the conditioned process

Calibration of $f(\cdot, \cdot)$

Uncertainty on f due to emulation integrated in likelihood $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^c, \phi)$:
($\mathbf{y}^c$ evaluations of the simulator, ϕ hyperparameters of the Gaussian process)

posterior distribution

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^c) \propto \int_{\Phi} l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^c, \phi) p(\boldsymbol{\theta}) p(\phi) d\phi\,,$$

or

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^c) \propto l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^c, \hat{\phi}) p(\boldsymbol{\theta})\,,$$

where $\hat{\phi}$ is an estimation regarded as fixed.

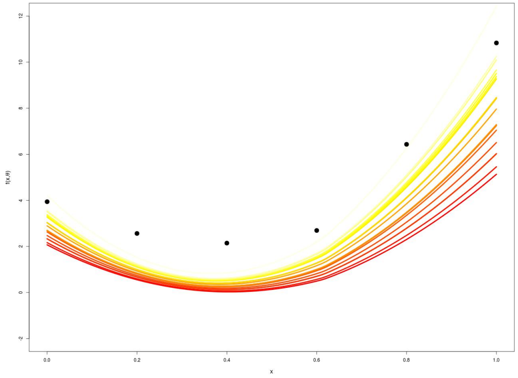
Plan

Model discrepancy

The model cannot capture the entire physical phenomenon:

$$y(\mathbf{x}_i) = f(\mathbf{x}_i, \boldsymbol{\theta}^*) + \delta(\mathbf{x}_i) + \epsilon_i, \quad i = 1, \dots, n,$$

where $\delta(\mathbf{x}_i) = \zeta(\mathbf{x}_i) - f(\mathbf{x}_i, \boldsymbol{\theta}^*)$ (recall $\zeta(\cdot)$ is the unobserved physical phenomenon) accounts for the model discrepancy.



Modelisation of δ:

Sensible to assume: $\delta(\mathbf{x}) \approx \delta(\mathbf{x} + d\mathbf{x})$

Gaussian Process hypothesis on δ with possible:

- zero mean,
- smooth a priori on covariance function,
- combining with Gaussian process hypothesis on f.

Meaning of $\boldsymbol{\theta}$:

- few information on $\boldsymbol{\theta}$ if there is a systematic discrepancy ?
- the model $f(\mathbf{x}, \boldsymbol{\theta})$ is informative through $\boldsymbol{\theta}$ on the shape of the physical phenomenon $\zeta(\cdot)$?

Plan

Prediction with a calibrated simulator:

Once the model is calibrated (in a Bayesian way):

Posterior distribution on $\boldsymbol{\theta}$: $\pi(\cdot|\mathbf{y}, ...)$

Prediction of the physical phenomenon $\zeta(\cdot)$, for $\mathbf{x}^{new}$?

- If no discrepancy, no emulator, $\zeta(\mathbf{x}^{new})$ can be estimated through

$$\hat{\zeta}(\mathbf{x}^{new}) = \int_{\Theta} f(\mathbf{x}^{new}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}\,.$$

- otherwise $\zeta(\mathbf{x}^{new})$ has a Gaussian process as posterior distribution with mean and covariance depending on $\boldsymbol{\theta}$.
  $\Rightarrow$ combining this distribution with $\pi(\cdot|\mathbf{y}, \mathbf{y}^c)$
  integration of the posterior mean of $\zeta(\mathbf{x}^{new})$:

$$\int_{\Theta} \mathbb{E}(\zeta(\mathbf{x}^{new})|\mathbf{y}, \mathbf{y}^c, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{y}^c)\,.$$

Difficulties and questions

Identifiability concerns

- If there is discrepancy, very little information on $\boldsymbol{\theta}$ and meaning of "best" or "true" $\boldsymbol{\theta}$?
- If measurement error distribution ($\epsilon_i \sim \mathcal{N}(0, \sigma^2)$) unknown $\Rightarrow$ lack of identifiability.
- Prediction can be accurate in a non-identifiable model...

Next step validation ?

- Validate a code is different from checking.
- Validation is a confrontation with the physical phenomenon.
- Validate with a model discrepancy ?

References

History:

- Sharon Bertsch McGrayne, 2011. The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy. Yale university press.

Bayesian methodology and application:

- Christian P. Robert, 2001. The Bayesian Choice. Springer.
- Jean-Michel Marin and Christian P. Robert, 2007. Bayesian Core. Springer.
- Jean-Michel Marin and Christian P. Robert (2009) Statistique bayésienne : les bases, Techniques de l'Ingénieur.
- Éric Parent and Jacques Bernier, 2007. Le raisonnement bayésien. Springer.

Inference by MCMC techniques:

- Christian P. Robert and Georges Casella, 2004. Monte Carlo Statistical Methods. Springer.

Calibration of computer models

- Dave Hidgon et al., 2005. Combining Field Data and Computer Simulations for Calibration and Prediction. SIAM 26(2).
- Marc Kennedy and Anthony O'Hagan, 2001. Bayesian Calibration of Computer Models. Journal of the Royal Statistical Society B 68.

Validation of computer models

- Susie Bayarri et al., 2002. A Framework for Validation of Computer Models. Technometrics 49(2).

Gaussian Process emulator:

- Thomas Santner et al., 2003. The Design and Analysis of Computer Experiments. Springer-Verlag.
- Kai-Tai Fang et al., 2006. Design and Modeling for Computer Experiments. Computer Science and Data Analysis. Chapman & Hall/CRC.