UNIVERSITÉ TOULOUSE III PAUL SABATIER

Université de Toulouse

ROBUST BAYESIAN ANALYSIS AND OPTIMIZATION ON A MOMENT CLASS

Jérôme Stenger

AppliBUGS - 13/06/2019

Fabrice Gamboa (IMT)
Merlin Keller (EDF) - Bertrand Iooss (EDF)

EDF R&D - Université Paul Sabatier

INTRODUCTION

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

→ A posterior mean :

→ A posterior generalized moment :

→ A posterior quantile :

→ A posterior value associated to a loss function L :

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

→ A posterior mean :

$$\phi_D(\pi) = \frac{\int \theta \mathscr{L}(D|\theta)\, \pi(d\theta)}{\int \mathscr{L}(D|\theta)\, \pi(d\theta)}$$

→ A posterior generalized moment :

→ A posterior quantile :

→ A posterior value associated to a loss function L :

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

→ A posterior mean :

→ A posterior generalized moment :

$$\phi_D(\pi) = \frac{\int q(\theta)\, \mathscr{L}(D|\theta)\, \pi(d\theta)}{\int \mathscr{L}(D|\theta)\, \pi(d\theta)}$$

→ A posterior quantile :

→ A posterior value associated to a loss function L :

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

- → A posterior mean :
- → A posterior generalized moment :
- → A posterior quantile :

$$\phi_D(\pi) = \inf\left\{x : \frac{\int_{-\infty}^{x} \mathscr{L}(D|\theta)\,\pi(d\theta)}{\int \mathscr{L}(D|\theta)\,\pi(d\theta)} \geq \alpha\right\}$$

- → A posterior value associated to a loss function L :

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

- → A posterior mean :
- → A posterior generalized moment :
- → A posterior quantile :
- → A posterior value associated to a loss function L :

$$\phi_D(\pi) = \frac{\int L(\theta, a)\, \mathscr{L}(D|\theta)\, \pi(d\theta)}{\int \mathscr{L}(D|\theta)\, \pi(d\theta)}$$

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

- → A posterior mean :
- → A posterior generalized moment :
- → A posterior quantile :
- → A posterior value associated to a loss function L :

All are quasi-convex function of the prior distribution π,

i.e $\phi_D(\lambda\pi_1 + (1-\lambda)\pi_2) \leq \max\{\phi_D(\pi_1), \phi_D(\pi_2)\}$

HOW IS THE PRIOR CHOSEN

The prior is not arbitrarily chosen

- → Tradition : e.g. lognormal for engineers.
- → Suitable functional form : monotone, unimodal, heavy tails, etc.
- → Mathematical convenience : parametric distribution, weakly informative, etc.

From an expert opinion/data, we often possess informations on the prior distribution $\pi(\lambda)$:

- → Quantiles, i.e, $\alpha_i \leq P_\pi(X \in [a_i, b_i]) \leq \beta_i$.
- → Moments, $\int \theta^k \pi(d\theta) = c_k$.
- → Generalized moments, $\mathbb{E}_\pi[q] = 0$.
- → Support bounds.
- → $\ldots$

HOW IS THE PRIOR CHOSEN

The prior is not arbitrarily chosen

- → Tradition : e.g. lognormal for engineers.
- → Suitable functional form : monotone, unimodal, heavy tails, etc.
- → Mathematical convenience : parametric distribution, weakly informative, etc.

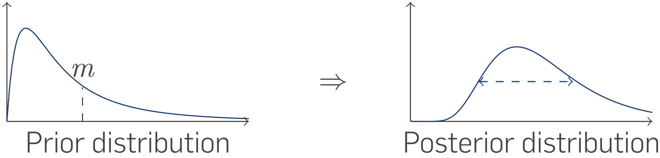
From an expert opinion/data, we often possess informations on the prior distribution $\pi(\lambda)$:

- → Quantiles, i.e, $\alpha_i \leq P_\pi(X \in [a_i, b_i]) \leq \beta_i$.
- → Moments, $\int \theta^k \, \pi(d\theta) = c_k$.
- → Generalized moments, $\mathbb{E}_\pi[q] = 0$.
- → Support bounds.
- → $\ldots$

IMPORTANCE OF PRIOR CHOICE

How is the statistical analysis affected by such uncertainty and, sometimes, arbitrariness in the prior choice.

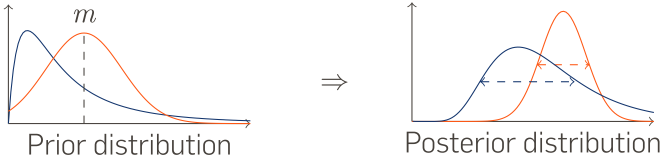
Suppose that we have information only on the prior mean, and that we are interested in the posterior variance :



IMPORTANCE OF PRIOR CHOICE

How is the statistical analysis affected by such uncertainty and, sometimes, arbitrariness in the prior choice.

Suppose that we have information only on the prior mean, and that we are interested in the posterior variance :



WHAT IS ROBUSTNESS ?

- → We model our uncertainty on the prior distribution through classes of priors $\pi \in \mathcal{A}$.
- → What is the worst impact the prior choice have on the quantity of interest $\phi(\pi)$?
  We compute $\overline{\phi} = \sup_{\pi \in \mathcal{A}} \phi(\pi)$ and $\underline{\phi} = \inf_{\pi \in \mathcal{A}} \phi(\pi)$
  - → If the range $\overline{\phi} - \underline{\phi}$ is "small", this means the prior choice has small impact on the quantity of interest $\rightsquigarrow$ robustness.
  - → If the range $\overline{\phi} - \underline{\phi}$ is "large", we must precise our prior distribution, i.e. find a smaller class $\mathcal{A}' \subset \mathcal{A}$.

WHAT IS ROBUSTNESS ?

→ We model our uncertainty on the prior distribution through classes of priors $\pi \in \mathcal{A}$.

→ What is the worst impact the prior choice have on the quantity of interest $\phi(\pi)$?
We compute $\overline{\phi} = \sup_{\pi\in\mathcal{A}} \phi(\pi)$ and $\underline{\phi} = \inf_{\pi\in\mathcal{A}} \phi(\pi)$

  → If the range $\overline{\phi} - \underline{\phi}$ is "small", this means the prior choice has small impact on the quantity of interest $\rightsquigarrow$ robustness.

  → If the range $\overline{\phi} - \underline{\phi}$ is "large", we must precise our prior distribution, i.e. find a smaller class $\mathcal{A}' \subset \mathcal{A}$.

CLASS OF PRIORS

A good class of priors should be :

→ easily interpretable,

→ compatible with the prior knowledge,

→ effectively representative of our uncertainty on the prior,

→ computationally friendly.

Some examples :

→ Density bounded class

→ Quantile class

→ (Symmetric) Unimodal class

→ Generalized Moment class

CLASS OF PRIORS

A good class of priors should be :

- → easily interpretable,
- → compatible with the prior knowledge,
- → effectively representative of our uncertainty on the prior,
- → computationally friendly.

Some examples :

- → Density bounded class

$$\mathcal{A} = \{\pi \in \mathcal{P}(\mathcal{X}) \mid \pi_l \leq \pi \leq \pi_u\} \ .$$

- → Quantile class
- → (Symmetric) Unimodal class
- → Generalized Moment class

CLASS OF PRIORS

A good class of priors should be :

→ easily interpretable,

→ compatible with the prior knowledge,

→ effectively representative of our uncertainty on the prior,

→ computationally friendly.

Some examples :

→ Density bounded class

→ Quantile class

$$\mathcal{A} = \{\pi \in \mathcal{P}(\mathcal{X}) \mid \alpha_i \leq \pi([a_i, b_i]) \leq \beta_i,\ i = 1, \ldots, n\}\ .$$

→ (Symmetric) Unimodal class

→ Generalized Moment class

CLASS OF PRIORS

A good class of priors should be :

- → easily interpretable,
- → compatible with the prior knowledge,
- → effectively representative of our uncertainty on the prior,
- → computationally friendly.

Some examples :

- → Density bounded class
- → Quantile class
- → (Symmetric) Unimodal class

$$\mathcal{A} = \{\pi \in \mathcal{P}(\mathcal{X}) \mid \pi \text{ is (symetric) unimodal}\} .$$

- → Generalized Moment class

CLASS OF PRIORS

A good class of priors should be :

- → easily interpretable,
- → compatible with the prior knowledge,
- → effectively representative of our uncertainty on the prior,
- → computationally friendly.

Some examples :

- → Density bounded class
- → Quantile class
- → (Symmetric) Unimodal class
- → Generalized Moment class

$$\mathcal{A} = \{\pi \in \mathcal{P}(\mathcal{X}) \mid \alpha_i \leq \mathbb{E}_\pi[\varphi_i] \leq \beta_i,\ i = 1, \dots, n\}\ .$$

THE MOMENT CLASS

We are interested in the following moment class :

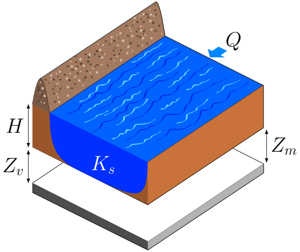
$$\mathcal{A} = \left\{ \pi \in \mathcal{P}(\mathcal{X}) \mid \alpha_i \leq \mathbb{E}_\pi[X^i] \leq \beta_i,\ i = 1, \ldots, n \right\} ,$$

where every priors satisfy moment constraints.

APPLICATION

PRESENTATION OF THE USE CASE

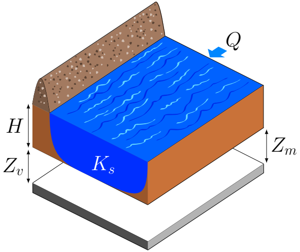
We study a simplified hydraulic code that calculates the water height H of a river.



$$H = \left(\frac{Q}{300 K_s \sqrt{\frac{Z_m - Z_v}{5000}}} \right)^{3/5}$$

PRESENTATION OF THE USE CASE

We study a simplified hydraulic code that calculates the water height H of a river.



$$H = \left(\frac{Q}{300 K_s \sqrt{\frac{Z_m - Z_v}{5000}}} \right)^{3/5}$$

Variable	Description	Distribution
Q	annual maximum flow rate	$Gumbel(\mu, \rho)$
K_s	Manning-Strickler coefficient	$\mathcal{N}(30, 7.5)$
Z_v	Depth measure of the river downstream	$\mathcal{U}(49, 51)$
Z_m	Depth measure of the river upstream	$\mathcal{U}(54, 55)$

BAYESIAN APPROACH

We'd like to compute a quantile or a probability not to excess a given height h

→ In a plug-in approach, the parameters μ and ρ of the Gumbel distribution are estimated by maximum likelihood based on a data set D of 47 maximal annual flow rate.

$$\mathbb{P}\left(H \geq h \mid \Theta\right) = \exp\left(-\exp\left\{\rho\left(\mu - 300 K_s \sqrt{\frac{Z_m - Z_v}{5000}}(h - Z_v)^{5/3}\right)\right\}\right),$$

with $\Theta = (\mu, \rho, K_s, Z_v, Z_m)$

→ In a Bayesian approach, we compute the following quantity :

$$\int \mathbb{P}\left(H \geq h \mid \Theta\right) \pi(\Theta|D)\, d\Theta,$$

where $\pi(\Theta|D) = \frac{\mathscr{L}(D|\Theta)\pi(\Theta)}{\int \mathscr{L}(D|\Theta)\pi(\Theta)\, d(\Theta)}$

BAYESIAN APPROACH

We'd like to compute a quantile or a probability not to excess a given height h

→ In a plug-in approach, the parameters μ and ρ of the Gumbel distribution are estimated by maximum likelihood based on a data set D of 47 maximal annual flow rate.

$$\mathbb{P}\left(H \geq h \mid \Theta\right) = \exp\left(- \exp\left\{ \rho \left(\mu - 300 K_s \sqrt{\frac{Z_m - Z_v}{5000}} (h - Z_v)^{5/3} \right) \right\} \right) ,$$

with $\Theta = (\mu, \rho, K_s, Z_v, Z_m)$

→ In a Bayesian approach, we compute the following quantity :

$$\int \mathbb{P}\left(H \geq h \mid \Theta\right) \pi(\Theta|D) \, d\Theta \, ,$$

where $\pi(\Theta|D) = \frac{\mathscr{L}(D|\Theta)\pi(\Theta)}{\int \mathscr{L}(D|\Theta)\pi(\Theta) \, d(\Theta)}$

ROBUST BAYESIAN APPROACH

What are the information on the prior distribution of μ, ρ?

→ We enforce their mean to be equal to their maximum likelihood estimation.

→ We fix bounds to *reasonable* values.

Variable	Bounds	Mean
μ	$[550, 700]$	626.14
ρ	$[150, 250]$	190

The optimization space is the moment class $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ with :

$$\mathcal{A}_1 = \{\pi_1 \in \mathcal{P}([550, 700]) \mid \mathbb{E}_{\pi_1}[X] = 626.14\} \;,$$
$$\mathcal{A}_2 = \{\pi_2 \in \mathcal{P}([150, 250]) \mid \mathbb{E}_{\pi_2}[X] = 190\} \;.$$

ROBUST BAYESIAN APPROACH

What are the information on the prior distribution of μ, ρ?

→ We enforce their mean to be equal to their maximum likelihood estimation.

→ We fix bounds to *reasonable* values.

Variable	Bounds	Mean
μ	$[550, 700]$	626.14
ρ	$[150, 250]$	190

The optimization space is the moment class $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ with :

$$\mathcal{A}_1 = \{\pi_1 \in \mathcal{P}([550, 700]) \mid \mathbb{E}_{\pi_1}[X] = 626.14\} \ ,$$
$$\mathcal{A}_2 = \{\pi_2 \in \mathcal{P}([150, 250]) \mid \mathbb{E}_{\pi_2}[X] = 190\} \ .$$

OPTIMAL QUANTITY OF INTEREST

We compute the optimal probability of failure

$$\sup_{\pi\in\mathcal{A}_1\otimes\mathcal{A}_2} F_\pi(h) = \sup_{\pi\in\mathcal{A}_1\otimes\mathcal{A}_2} \int \mathbb{P}(H \leq h|\Theta)\pi(\Theta|D)\, d\Theta$$

$$\inf_{\pi\in\mathcal{A}_1\otimes\mathcal{A}_2} F_\pi(h) = \inf_{\pi\in\mathcal{A}_1\otimes\mathcal{A}_2} \int \mathbb{P}(H \leq h|\Theta)\pi(\Theta|D)\, d\Theta$$

$\rightsquigarrow$ The moment space is a non parametric infinite dimensional space.

REDUCTION THEOREM

REDUCTION THEOREM

Reduction theorem

Let ϕ be a quasi-convex lower semicontinuous function on a locally convex topological vector space. Let $\mathcal{A}$ be a compact convex subset. Then

$$\sup_{\pi \in \mathcal{A}} \phi(\pi) = \sup_{\pi \in \Delta} \phi(\pi) \ ,$$

where Δ is the set of extreme points of $\mathcal{A}$.

Here, our posterior distribution is the ratio of two linear function of the prior distribution.

$$\pi(\theta \,|\, x) = \frac{l(x \,|\, \theta)\pi(\theta)}{\int l(x \,|\, \theta)\, \pi(d\theta)}$$

THE QUANTITY OF INTEREST

In a Bayesian analysis, it is fundamental to compute some quantity of interest on the posterior distribution :

- → A posterior mean
- → A posterior generalized moment
- → A posterior quantile
- → A posterior value associated to a loss function

All are quasi-convex lower semi-continuous function of the prior distribution π,

EXTEME POINTS OF MOMENT SETS

→ Let $\mathcal{X} = \{1, 2, 3\}$ be a finite sample space, so that $\mathcal{P}(\mathcal{X})$ is isomorphic to the simplex of $\mathbb{R}^3$,

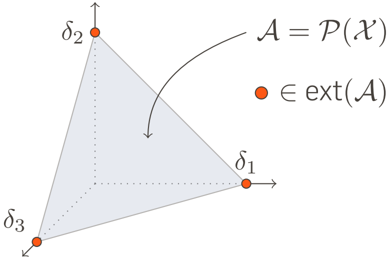
→ Admit that the objective function reaches its optimums on the extreme points.

⇝ Extreme points are Dirac mass.

EXTEME POINTS OF MOMENT SETS

→ Let $\mathcal{X} = \{1, 2, 3\}$ be a finite sample space, so that $\mathcal{P}(\mathcal{X})$ is isomorphic to the simplex of $\mathbb{R}^3$,

→ Admit that the objective function reaches its optimums on the extreme points.

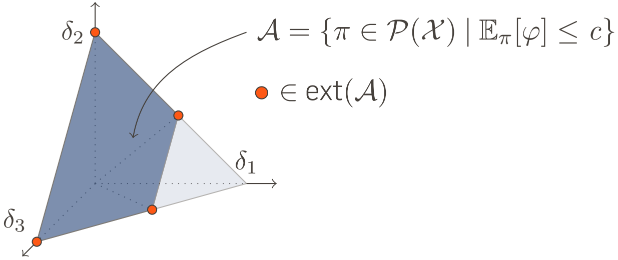


⇝ Extreme points are Dirac mass.

EXTEME POINTS OF MOMENT SETS

→ Let $\mathcal{X} = \{1, 2, 3\}$ be a finite sample space, so that $\mathcal{P}(\mathcal{X})$ is isomorphic to the simplex of $\mathbb{R}^3$,

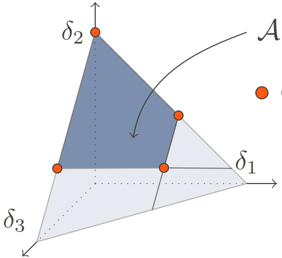
→ Admit that the objective function reaches its optimums on the extreme points.



⇝ After adding **one** constraint, the extreme points are convex combination of at most **two** Dirac masses.

EXTEME POINTS OF MOMENT SETS

- → Let $\mathcal{X} = \{1, 2, 3\}$ be a finite sample space, so that $\mathcal{P}(\mathcal{X})$ is isomorphic to the simplex of $\mathbb{R}^3$,
- → Admit that the objective function reaches its optimums on the extreme points.



$$\mathcal{A} = \{\pi \in \mathcal{P}(\mathcal{X}) \mid \mathbb{E}_\pi[\varphi_1] \leq c_1, \mathbb{E}_\pi[\varphi_2] \leq c_2\}$$

● $\in$ ext($\mathcal{A}$)

⇝ After adding **two** constraints, the extreme points are convex combination of at most **three** Dirac masses.

WINKLER'S CLASSIFICATION OF EXTREME POINTS

Heuristic

If you have N pieces of information relevant to the random variable X then it is enough to pretend that X takes at most $N+1$ values in $\mathcal{X}$.

Winkler theorem

The extreme measures of moment class

$$\{\pi \in \mathcal{P}(\mathcal{X}) \mid \mathbb{E}_\pi[\varphi_1] \leq 0, \ldots, \mathbb{E}_\pi[\varphi_n] \leq 0\}$$

are the discrete measures that are supported on at most $n+1$ points.

DISCRETE MEASURES

Let enforce N moment constraints on a measure $\mathbb{E}_\mu[X^j] = c_j$.
OUQ theorem guaranties the optimal measure to be supported on at most $N+1$ points :

$$\mu = \sum_{i=1}^{N+1} \omega_i \delta_{x_i}$$

We have the following system

$$\begin{cases} \omega_1 & + & \dots & + & \omega_{N+1} & = 1 \\ \omega_1 x_1 & + & \dots & + & \omega_{N+1} x_{N+1} & = c_1 \\ \vdots & & & & \vdots & \vdots \\ \omega_1 {x_1}^N & + & \dots & + & \omega_{N+1} {x_{N+1}}^N & = c_N \end{cases}$$

$\rightsquigarrow$ The weights are uniquely determined by the positions.

ADMISSIBLE MEASURE

We reparameterize the problem with the position of the support points. But generating a discrete measure having constraints on its moments is not easy...

Example : Let μ be supported on $[0,1]$ such that $\mathbb{E}_\mu[X] = 0.5$ and $\mathbb{E}_\mu[X^2] = 0.3$.

$$\Delta = \left\{ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \in \mathcal{P}([0,1]) \mid \mathbb{E}_\mu[X] = 0.5,\ \mathbb{E}_\mu[X^2] = 0.3 \right\} ,$$

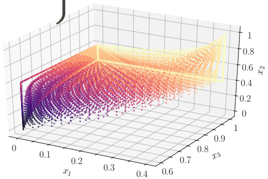
ADMISSIBLE MEASURE

We reparameterize the problem with the position of the support points. But generating a discrete measure having constraints on its moments is not easy...

Example : Let μ be supported on $[0,1]$ such that $\mathbb{E}_\mu[X] = 0.5$ and $\mathbb{E}_\mu[X^2] = 0.3$.

$$\Delta = \left\{ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \in \mathcal{P}([0,1]) \mid \mathbb{E}_\mu[X] = 0.5,\ \mathbb{E}_\mu[X^2] = 0.3 \right\},$$

$$\mathcal{V}_\Delta = \left\{ \mathbf{x} = (x_1, x_2, x_3) \in [0,1]^3 \mid \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \in \mathcal{A}_\Delta \right\}$$



ADMISSIBLE MEASURE

We reparameterize the problem with the position of the support points. But generating a discrete measure having constraints on its moments is not easy...

Example : Let μ be supported on $[0,1]$ such that $\mathbb{E}_\mu[X] = 0.5$ and $\mathbb{E}_\mu[X^2] = 0.3$.

$$\Delta = \left\{ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \in \mathcal{P}([0,1]) \mid \mathbb{E}_\mu[X] = 0.5,\ \mathbb{E}_\mu[X^2] = 0.3 \right\},$$

$$\rightsquigarrow \mu = \omega_1 \delta_{x_1} + \omega_2 \delta_{x_2} + \omega_3 \delta_{x_3}$$

$\mathbf{x} = (0.1, 0.4, 0.9)$ gives weights $\boldsymbol{\omega} = (0.05, 0.73, 0.22)$ ✓

$\mathbf{x} = (0.1, 0.3, 0.9)$ gives weights $\boldsymbol{\omega} = (-0.19, 0.92, 0.27)$ ✗

How to optimize over Δ ? How to explore the manifold $\mathcal{V}_\Delta$?

ADMISSIBLE MEASURE

We reparameterize the problem with the position of the support points. But generating a discrete measure having constraints on its moments is not easy...

Example : Let μ be supported on $[0,1]$ such that $\mathbb{E}_\mu[X] = 0.5$ and $\mathbb{E}_\mu[X^2] = 0.3$.

$$\Delta = \left\{ \mu = \sum_{i=1}^{3} \omega_i \delta_{x_i} \in \mathcal{P}([0,1]) \mid \mathbb{E}_\mu[X] = 0.5,\ \mathbb{E}_\mu[X^2] = 0.3 \right\} ,$$

$$\rightsquigarrow \mu = \omega_1 \delta_{x_1} + \omega_2 \delta_{x_2} + \omega_3 \delta_{x_3}$$

$\mathbf{x} = (0.1, 0.4, 0.9)$ gives weights $\boldsymbol{\omega} = (0.05, 0.73, 0.22)$ ✓
$\mathbf{x} = (0.1, 0.3, 0.9)$ gives weights $\boldsymbol{\omega} = (-0.19, 0.92, 0.27)$ ✗

How to optimize over Δ ? How to explore the manifold $\mathcal{V}_\Delta$?

POSSIBLE WAYS OF OPTIMIZING

→ Optimization under constraints : the position and the weight must satisfy the Vandermonde system.

→ Optimization by rewriting the objective function : changing the parameterization of the problem so that the constraint are naturally enforced in the objective function.

↳ Canonical moments allows to efficiently explore the set of optimization Δ.

POSSIBLE WAYS OF OPTIMIZING

→ Optimization under constraints : the position and the weight must satisfy the Vandermonde system.

→ Optimization by rewriting the objective function : changing the parameterization of the problem so that the constraint are naturally enforced in the objective function.

↳ Canonical moments allows to efficiently explore the set of optimization Δ.

CANONICAL MOMENTS PARAMETERIZATION

CLASSICAL MOMENTS PROBLEM

→ Moments of $\mathcal{U}[0,1]$:

$$\left(\frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \dots\right)$$

→ Moments of $\mathcal{U}[0,2]$:

$$\left(1,\ \frac{4}{3},\ 2,\ \dots\right)$$

There is no relation between the classical moments and the intrinsic structure of the distribution.

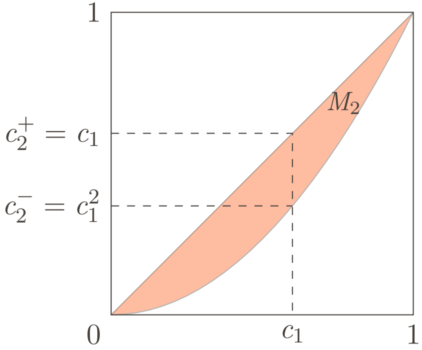
MOMENT SPACE

We define the moment space $M_n = \{\mathbf{c}_n(\mu) = (c_1, \ldots, c_n) \mid \mu \in \mathcal{P}([0,1])\}$

Given $\mathbf{c}_n \in \operatorname{int} M_n$ we define the extreme values

$$c_{n+1}^{+} = \max\{c : (c_1, \ldots, c_n, c) \in M_{n+1}\}$$
$$c_{n+1}^{-} = \min\{c : (c_1, \ldots, c_n, c) \in M_{n+1}\}$$

They represent the maximum and minimum values of the $(n+1)$th moment a measure can have, when its moments up to order n equals to c_n.



CANONICAL MOMENTS

The nth canonical moment is defined as

$$p_n = p_n(\mathbf{c}) = \frac{c_n - c_n^-}{c_n^+ - c_n^-}$$

Properties of canonical moments

- $p_n \in [0, 1]$,
- Canonical moments are defined up to degree $N = \min \{n \in \mathbb{N} \mid \mathbf{c}_n \in \partial M_n\}$ and $p_N \in \{0, 1\}$,
- The canonical moments are invariants by affine transformation. Which means we can always transform a measure supported on $[a, b]$ to $[0, 1]$

THE STIELTJES TRANSFORM

The Stieltjes transform is the analytic function on $\mathbb{C}\backslash\text{supp}(\mu)$

$$S(z) = S(z, \mu) = \int_a^b \frac{d\mu(x)}{z - x} \,,$$

If μ has a finite support : $S(z) = \sum_{i=1}^{n} \frac{\omega_i}{z - x_i} = \frac{Q_{n-1}(z)}{P_n^*(z)} \,,$

$P_n^* = \prod_{i=1}^n (z - x_i) \rightsquigarrow$ its roots are the support points of μ

Properties of the Stieltjes transform

P_n^* can be expressed recursively with the canonical moments :

$$P_{k+1}^*(x) = (x - a - (b-a)(\zeta_{2k} + \zeta_{2k+1}))P_k^*(x) - (b-a)^2 \zeta_{2k-1}\zeta_{2k} P_{k-1}^*(x)$$

where $\zeta_k = (1 - p_{k-1})p_k$

THE STIELTJES TRANSFORM

The Stieltjes transform is the analytic function on $\mathbb{C}\backslash\text{supp}(\mu)$

$$S(z) = S(z,\mu) = \int_a^b \frac{d\mu(x)}{z-x} \ ,$$

If μ has a finite support : $S(z) = \sum_{i=1}^{n} \frac{\omega_i}{z-x_i} = \frac{Q_{n-1}(z)}{P_n^*(z)}$,

$P_n^* = \prod_{i=1}^n (z-x_i) \rightsquigarrow$ its roots are the support points of μ

Properties of the Stieltjes transform

P_n^* can be expressed recursively with the canonical moments :

$$P_{k+1}^*(x) = (x-a-(b-a)(\zeta_{2k}+\zeta_{2k+1}))P_k^*(x)-(b-a)^2\zeta_{2k-1}\zeta_{2k}P_{k-1}^*(x)$$

where $\zeta_k = (1-p_{k-1})p_k$

POLYNOMIAL IDENTIFICATION

In summary, given a measure $\mu = \sum_{i=1}^{n+1} \omega_i \delta_{x_i}$, we have two representations of the polynomial P_{n+1}^*

→ Its roots are the measure support points :

$$P_n^*(z) = \prod_{i=1}^{n+1} (z - x_i) \ .$$

→ Its coefficients are function of a sequence of the measure canonical moments $\mathbf{c} = (c_1, \dots, c_{2n+1})$:

$$P_n^*(z) = \phi_0(\mathbf{c}) + \phi_1(\mathbf{c}) z + \cdots + \phi_{n+1}(\mathbf{c}) z^{n+1} \ .$$

POLYNOMIAL IDENTIFICATION

In summary, given a measure $\mu = \sum_{i=1}^{n+1} \omega_i \delta_{x_i}$, we have two representations of the polynomial P_{n+1}^*

→ Its roots are the measure support points :

$$P_n^*(z) = \prod_{i=1}^{n+1} (z - x_i) \ .$$

→ Its coefficients are function of a sequence of the measure canonical moments $\mathbf{c} = (c_1, \ldots, c_{2n+1})$:

$$P_n^*(z) = \phi_0(\mathbf{c}) + \phi_1(\mathbf{c}) z + \cdots + \phi_{n+1}(\mathbf{c}) z^{n+1} \ .$$

GENERATION OF ADMISSIBLE MEASURES

Theorem

Consider a sequence of moment $\mathbf{c}_n = (c_1, \dots, c_n) \in M_n$, *and the set of measure*

$$\Delta = \left\{ \mu = \sum_{i=1}^{n+1} \omega_i \delta_{x_i} \in \mathcal{P}([a,b]) \mid \mathbb{E}_\mu[X^j] = c_j,\ j = 1, \dots, n \right\} .$$

We define

$$\Gamma = \left\{ (p_{n+1}, \dots, p_{2n+1}) \in [0,1]^{n+1} \mid p_i \in \{0,1\} \Rightarrow p_k = 0,\ k > i \right\} .$$

Then there exists a bijection between Δ *and* Γ.

EFFECTIVE PARAMETERIZATION

Let $\mu \in \Delta = \left\{ \sum_{i=1}^{n+1} \omega_i \delta_{x_i} \in \mathcal{P}([a,b]) \mid \mathbb{E}_\mu[X^j] = c_j,\ 1 \leq j \leq n \right\}$

EFFECTIVE PARAMETERIZATION

$$\mu \in \Delta$$

$\downarrow$ The support of μ is the roots of a polynomial P^*_{n+1}

$$P^*_{n+1}$$

EFFECTIVE PARAMETERIZATION

$$\mu \in \Delta$$

$\downarrow$ The support of μ is the roots of a polynomial P^*_{n+1}

$$P^*_{n+1}$$

$\downarrow$ Coefficients are defined with the sequence of canonical moments

$$(p_1, \ldots, p_n, p_{n+1}, \ldots, p_{2n+1})$$

EFFECTIVE PARAMETERIZATION

$\mu \in \Delta$

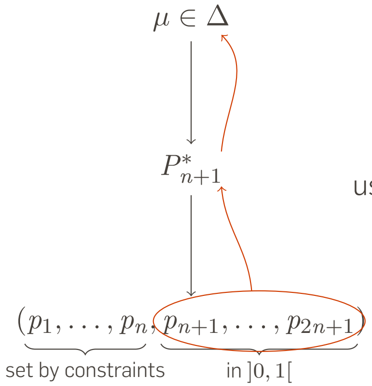
↓ The support of μ is the roots of a polynomial P^*_{n+1}

P^*_{n+1}

↓ Coefficients are defined with the sequence of canonical moments

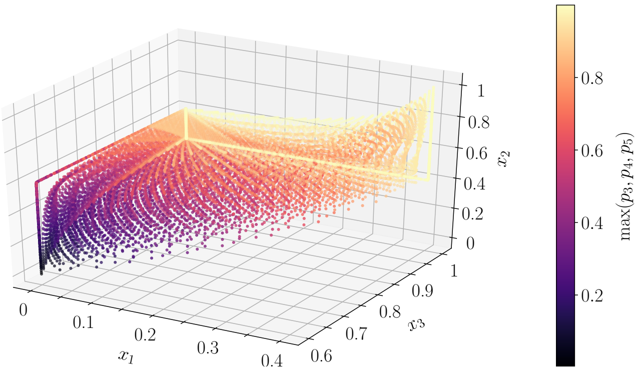
$$(\underbrace{p_1, \ldots, p_n}_{\text{set by constraints}}, \underbrace{p_{n+1}, \ldots, p_{2n+1}}_{\text{in }]0,1[})$$

EFFECTIVE PARAMETERIZATION



We can explore the whole set Δ using a parameterization in $]0, 1[^{n+1}$.

SET OF ADMISSIBLE MEASURES



Each point correspond to a measure μ on $[0, 1]$, we enforced $c_1 = 0.5$ and $c_2 = 0.3$ so that $p_1 = 0.5$ and $p_2 = 0.2$. We generated a regular grid where p_3, p_4 and p_5 goes from 0 to 1. The three Dirac masses corresponding to the roots of P_3^* are projected on each axis.

ALGORITHM

Algorithm 1 : P.O.F COMPUTATION

Inputs : - lower bounds, $\mathbf{l} = (l_1, \dots, l_d)$

- upper bounds, $\mathbf{u} = (u_1, \dots, u_d)$

- constraints sequences of moments, $\mathbf{c}_i = (c_i^{(1)}, \dots, c_i^{(N_i)})$ and its corresponding sequences of canonical moments, $\mathbf{p}_i = (p_i^{(1)}, \dots, p_i^{(N_i)})$ for $1 \leq i \leq d$.

function P.O.F$(p_1^{(N_1+1)}, \dots, p_1^{(2N_1+1)}, \dots, p_d^{(N_d+1)}, \dots, p_d^{(2N_d+1)})$

  for $i = 1, \dots, d$ **do**

    for $k = 1, \dots, N_i$ **do**

      $P_{i*}^{(k+1)} = (X - l_i - (u_i - l_i)(\zeta_i^{2k} + \zeta_i^{(2k+1)}))P_{i*}^{(k)} - (u_i - l_i)^2 \zeta_i^{(2k-1)} \zeta_i^{(2k)} P_{i*}^{(k-1)}$;

    $x_i^{(1)}, \dots, x_i^{(N_i+1)} =$ roots$(P_i^{*(N_i+1)})$;

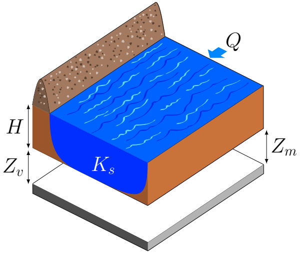
    $\omega_i^{(1)}, \dots, \omega_1^{(N_i+1)} =$ weight$(x_i^{(1)}, \dots, x_1^{(N_i+1)}, \mathbf{c}_i)$;

  return $\sum_{i_1=1}^{N_1+1} \dots \sum_{i_d=1}^{N_d+1} \omega_1^{(i_1)} \dots \omega_d^{(i_d)} \mathbb{1}_{\{G(x_1^{(i_1)}, \dots, x_d^{(i_d)}) \leq h\}}$;

RESULTS

PRESENTATION OF THE USE CASE

We recall the use case



$$H = \left(\frac{Q}{300 K_s \sqrt{\frac{Z_m - Z_v}{5000}}} \right)^{3/5}$$

Variable	Description	Distribution
Q	annual maximum flow rate	$Gumbel(\mu, \rho)$
K_s	Manning-Strickler coefficient	$\mathcal{N}(30, 7.5)$
Z_v	Depth measure of the river downstream	$\mathcal{U}(49, 51)$
Z_m	Depth measure of the river upstream	$\mathcal{U}(54, 55)$

PRESENTATION OF THE USE CASE

The prior distributions of μ, ρ are on the following moment classes

$$\mathcal{A}_1 = \{\pi_1 \in \mathcal{P}([550, 700]) \mid \mathbb{E}_{\pi_1}[X] = 626.14\} \ ,$$
$$\mathcal{A}_2 = \{\pi_2 \in \mathcal{P}([150, 250]) \mid \mathbb{E}_{\pi_2}[X] = 190\} \ .$$

We compute the optimal probability of failure

$$\sup_{\pi \in \mathcal{A}_1 \otimes \mathcal{A}_2} F_\pi(h) = \sup_{\pi \in \mathcal{A}_1 \otimes \mathcal{A}_2} \int \mathbb{P}(H \leq h|\Theta)\pi(\Theta|D)\, d\Theta$$

$$\inf_{\pi \in \mathcal{A}_1 \otimes \mathcal{A}_2} F_\pi(h) = \inf_{\pi \in \mathcal{A}_1 \otimes \mathcal{A}_2} \int \mathbb{P}(H \leq h|\Theta)\pi(\Theta|D)\, d\Theta$$

BAYES QUANTITY OF INTEREST

We compute $F_\pi(h) = \int \mathbb{P}(H \leq h|\Theta)\pi(\Theta|D)\, d\Theta$

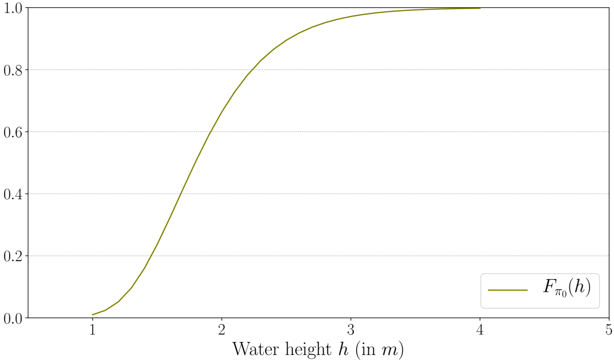


Figure : The initial prior distribution are $\mu \sim \mathscr{G}(1, 500), 1/\rho \sim \mathscr{G}(1, 200)$.

ROBUST BAYES QUANTITY OF INTEREST

We compute $F_\pi(h) = \int \mathbb{P}(H \leq h|\Theta)\pi(\Theta|D)\, d\Theta$

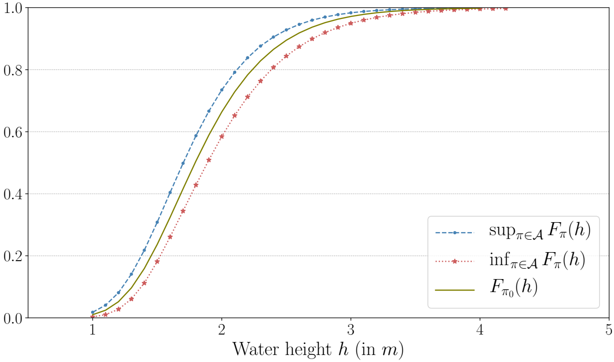


Figure : We only enforce mean and bounds on the prior distributions μ, ρ.

SAFETY MARGINS

In regards of the uncertainty on the prior distribution, here are the bounds on the quantile of the output distribution.

Quantile	Lower Bounds	Bayes estimation	Upper Bounds
0.95%	$2.62m$	$2.78m$	$3.00m$
0.99%	$3.16m$	$3.38m$	$3.67m$

REFERENCES

[1] J.O. Berger, *Robust Bayesian analysis : sensitivity to the prior*, JSPI, 25, p.303, (1990).

[2] F. Ruggeri, D. Ríos Insua, and J. Martin, *Robust Bayesian Analysis*, in Hanbook of Statistics, (2005).

[3] J. Stenger, F. Gamboa, M. Keller, B. Iooss, *Optimal Uncertainty Quantification of a risk measurement from a thermal-hydraulic code using canonical moments*, (2019).

[4] G. Winkler, *Extreme Points of Moment Sets*, Math. Oper. Res. 13, p.581 (1988).

[5] A. Pasanisi, M. Keller, and E. Parent *Estimation of a quantity of interest in uncertainty analysis : some help from Bayesian decision theory*, Reliability Engineering & System Safety, 100, p.93, (2012).

[6] H. Dette, W.J. Student, *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley-Blackwell, (1997).

THANK YOU FOR YOUR ATTENTION !