Bayesian Brittleness

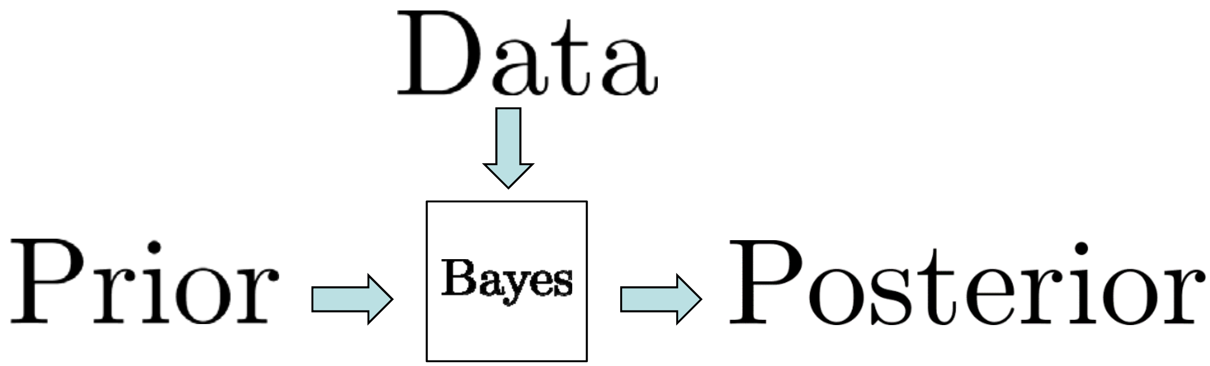
Houman Owhadi

- Bayesian Brittleness.
  H. Owhadi, C. Scovel, T. Sullivan. arXiv:1304.6772
- Brittleness of Bayesian inference and new Selberg formulas.
  H. Owhadi, C. Scovel. arXiv:1304.7046

Shanghai 2014

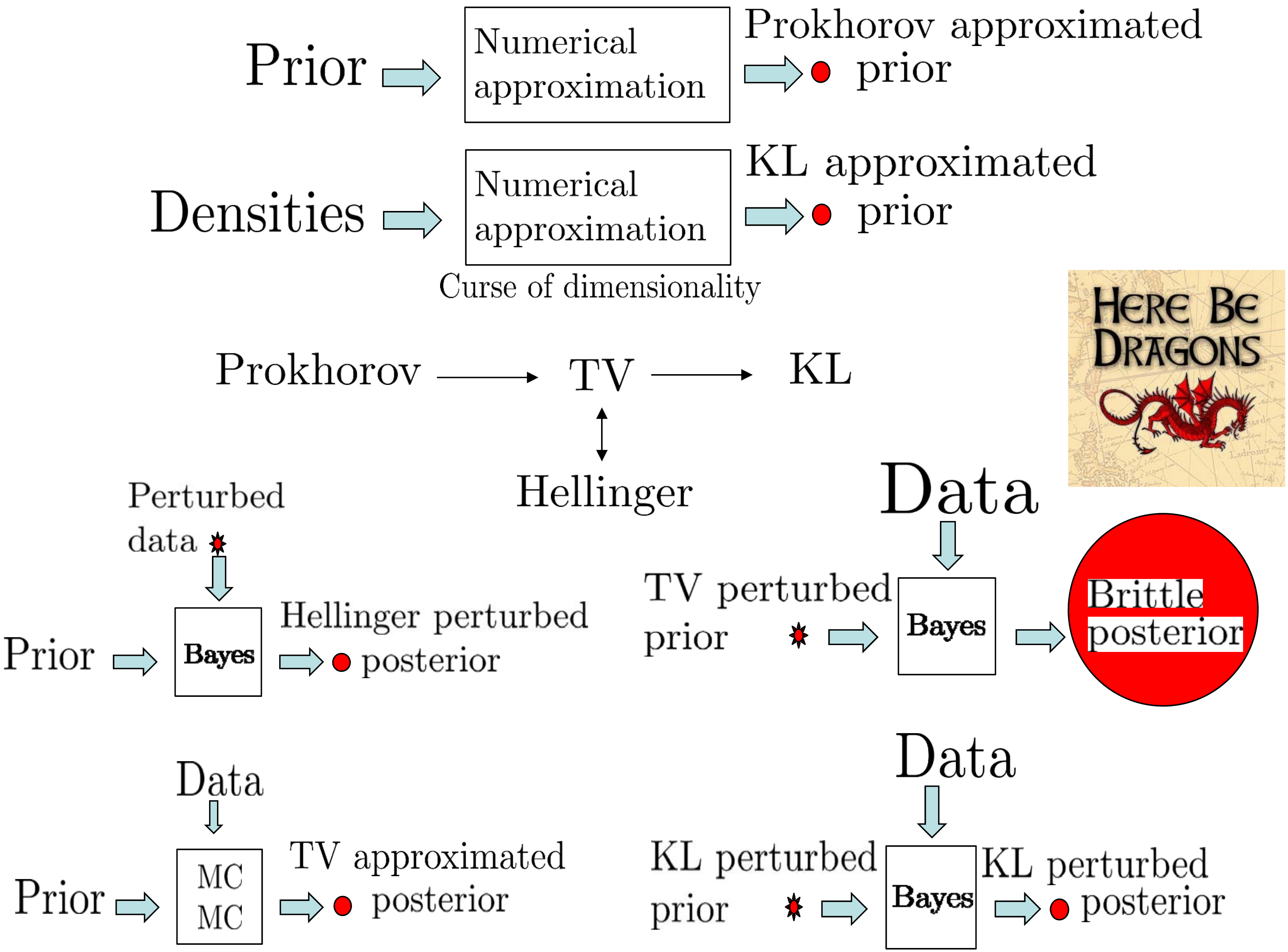
Conditioning in a continuous space

$$\theta_\pi(d) = \mathbb{E}_{\mu\sim\pi,\, d\sim\mu^n}\Big[\Phi(\mu)\Big|d\Big]$$



Worst case robustness questions

- What if the prior is a numerical approximation?
- What if the posterior is approximated and conditioning is used in a recursive manner?
- What if data is approximated and conditioning is used in a recursive manner?


Prior
Numerical approximation
Prokhorov approximated prior
Densities
Numerical approximation
KL approximated prior
Curse of dimensionality
Here Be Dragons
Prokhorov
TV
KL
Hellinger
Perturbed data
Prior
Bayes
Hellinger perturbed posterior
Data
TV perturbed prior
Bayes
Brittle posterior
Data
Prior
MC MC
TV approximated posterior
Data
KL perturbed prior
Bayes
KL perturbed posterior

Bayesian Inference in a Continuous world

Positive

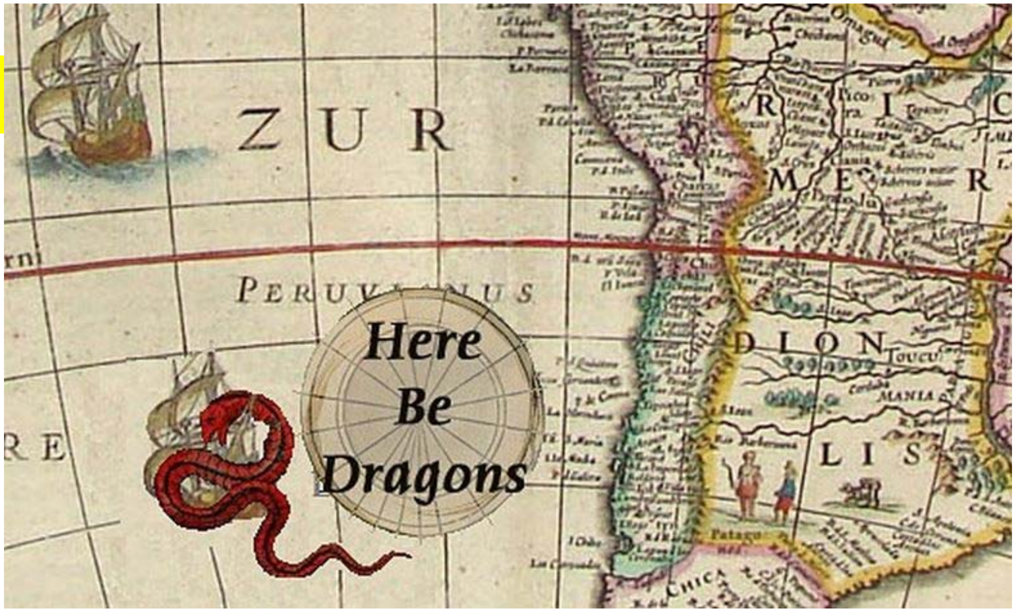
- Classical Bernstein Von Mises
- Wasserman, Lavine, Wolpert (1993)
- Castillo and Rousseau (2013)
- Castillo and Nickl (2013)
- Stuart & Al (2010+).
- ….

Negative

- Freedman (1963, 1965)
- P Gustafson & L Wasserman (1995)
- Diaconis & Freedman 1998
- Johnstone 2010
- Leahu 2011
- Belot 2013
- …
- Owhadi, Scovel, Sullivan (2013)

Other related negative results in Statistics

- Bahadur, Raghu Raj and Savage, Leonard J. (1956). The nonexistence of certain statistical procedures in nonparametric problems.
- Donoho, David L. (1988). One-sided inference about functionals of a density.



A warm-up problem

You have a bag containing 100 coins

99 coins are fair

1 always land on head

You pick one coin at random from the bag

You flip it 10 times and 10 times you get head

What is the probability that the coin that you have picked is the unfair one?

Answer

$$\mathbb{P}[A|B] = \mathbb{P}[B|A]\frac{\mathbb{P}[A]}{\mathbb{P}[B]} = \overbrace{\frac{1}{1+99\cdot(0.5)^{10}}}^{(1)} \approx 0.91$$

A: The coin is unfair

B: You observe 10 heads

Robustness If

bag contains 101 coins

and

fair coins are slightly unbalanced:
probability of a head is 0.51

Then

(1) still a good approximation of correct answer

What if random outcomes are not head or tail but decimal numbers, perhaps given to finite precision?

Problem 2

We want to estimate $\Phi(\mu^\dagger) = \mu^\dagger[X \geq a]$

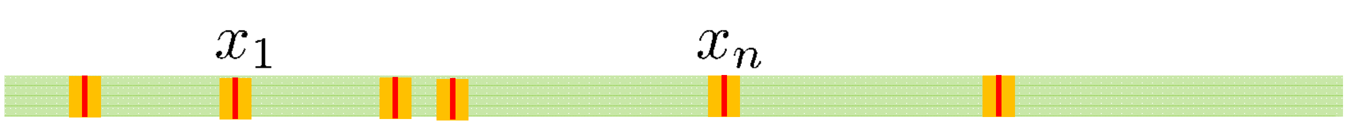
$\mu^\dagger$: Unknown or partially known measure of probability on $\mathbb{R}$

We observe $d = (d_1, \ldots, d_n) \in \mathbb{R}^n$

n i.i.d samples from $\mu^\dagger$

$$d \in B_\delta^n := \prod_{i=1}^n B_\delta(x_i)$$

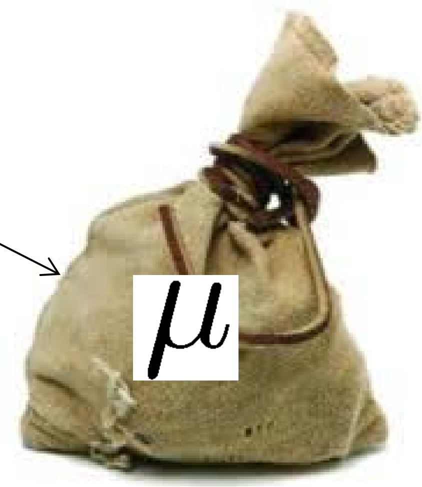
$B_\delta(x)$: open ball of radius δ centered on x



Bayesian Answer

Bayesian model class $\mathcal{A}$

$\mathcal{A} \subset \mathcal{M}(\mathbb{R})$



Assume that $\mu^\dagger$ is the realization of $\mu \sim \pi$

Prior $\pi \in \mathcal{M}(\mathcal{A})$

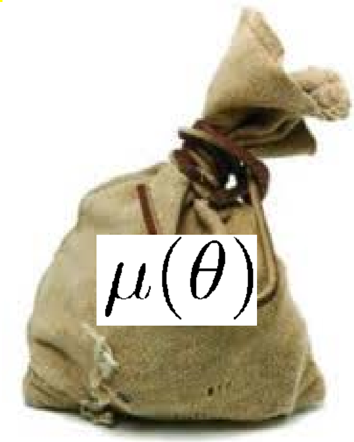
$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle| d \in B_\delta^n\right]$$

Probability space $\mathcal{A} \times \mathbb{R}^n$

$$\mathbb{E}_{\mu\sim\pi,\, d\sim\mu^n}\left[\phi(\mu)\middle| d \in B_\delta^n\right]$$

Parametric Bayesian Answer

Θ: Parameter space ($\subset \mathbb{R}^k$)



Define a map

$$\mathcal{P} : \Theta \to \mathcal{M}(\mathbb{R})$$
$$\theta \to \mu(\theta)$$

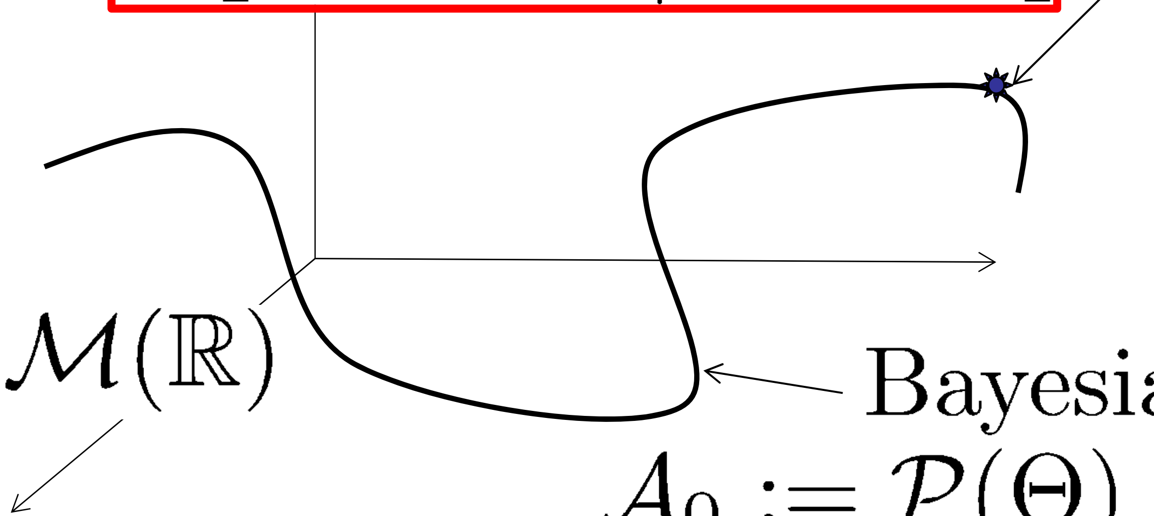
p_0: Prior distribution on Θ

$$\mathbb{E}\left[\Phi(\mu(\theta)) \,\middle|\, d \in B_\delta^n\right]$$

Bayesian model

$\mu(\theta)$: Random element of $\mathcal{A}_0$

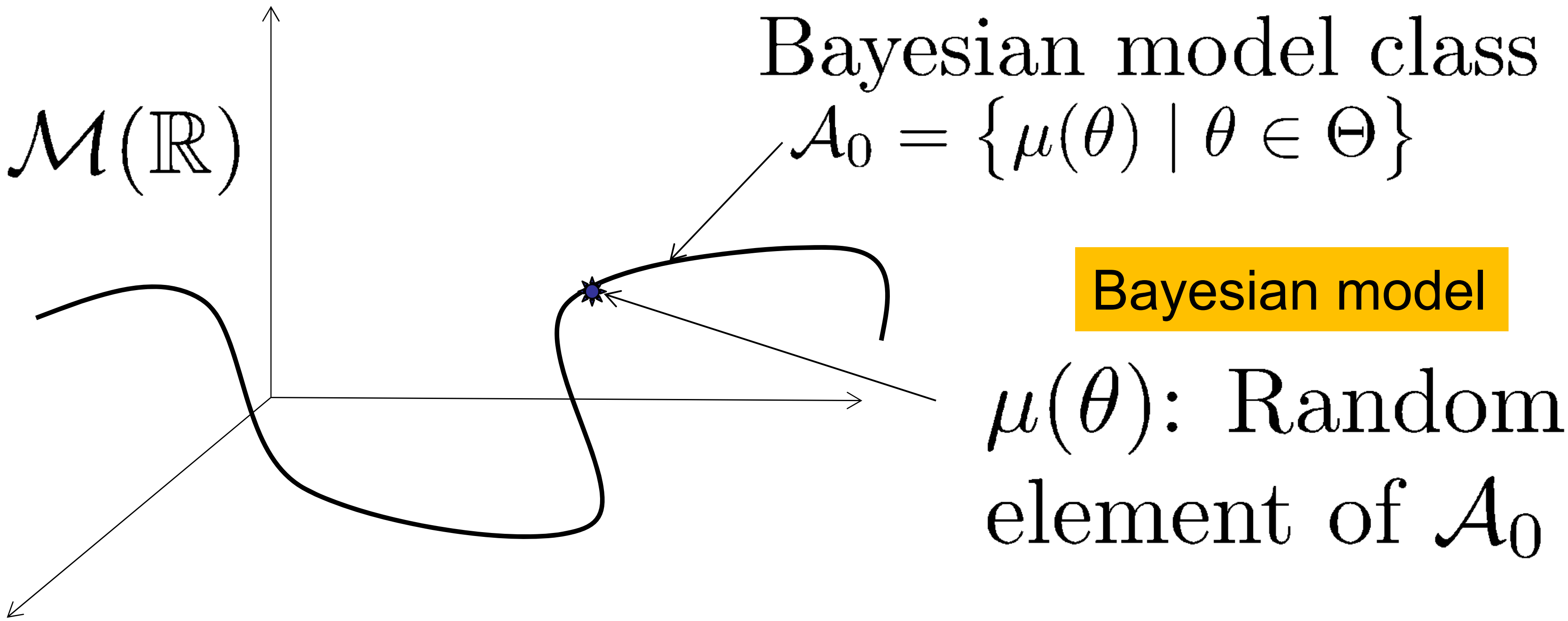
$$\pi_0 = \mathcal{P}(p_0)$$



Bayesian model class

$$\mathcal{A}_0 := \mathcal{P}(\Theta) = \{\mu(\theta) \mid \theta \in \Theta\}$$

Example



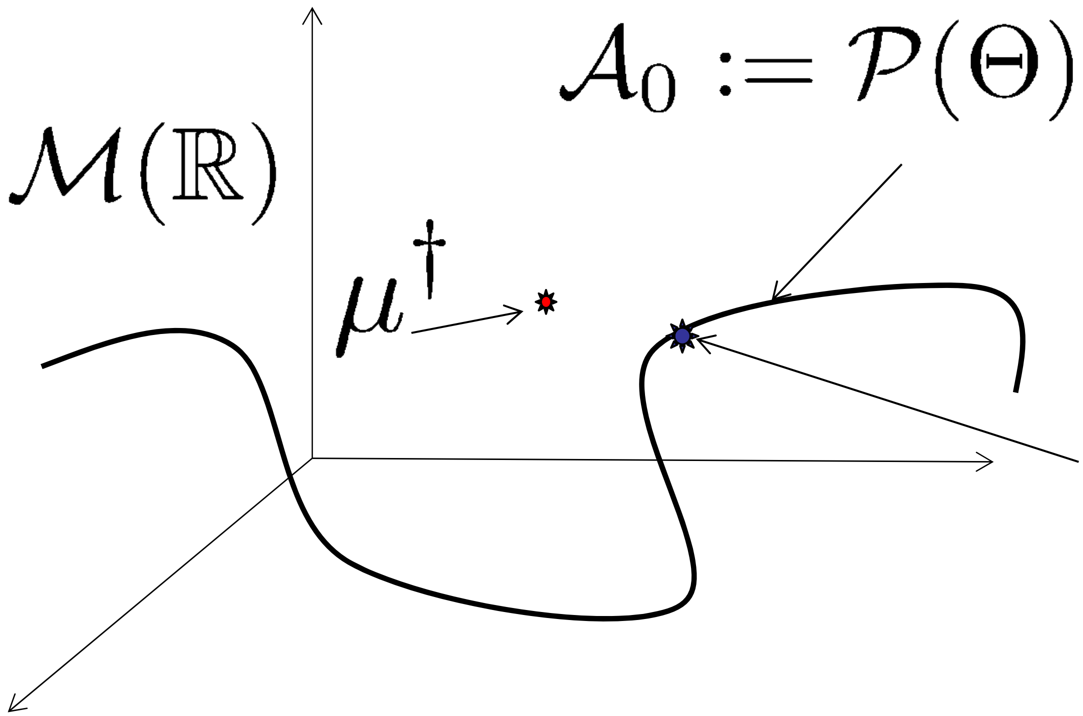
$\mu(\theta)$: Gaussian measure with mean c and SD σ

density: $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$

$$\theta = (c, \sigma) \in \Theta \subset \mathbb{R} \times \mathbb{R}_+$$

Bayesian model class

$$\mathcal{A}_0 := \mathcal{P}(\Theta) = \{\mu(\theta) \mid \theta \in \Theta\}$$



Bayesian model

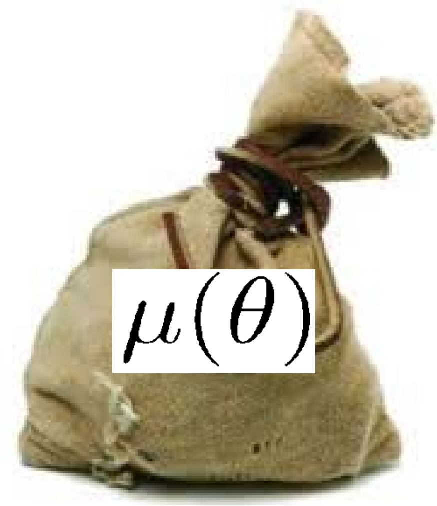
$\mu(\theta)$: Random element of $\mathcal{A}_0$

$$\mu^\dagger \in \mathcal{P}(\Theta)$$

Model is well specified

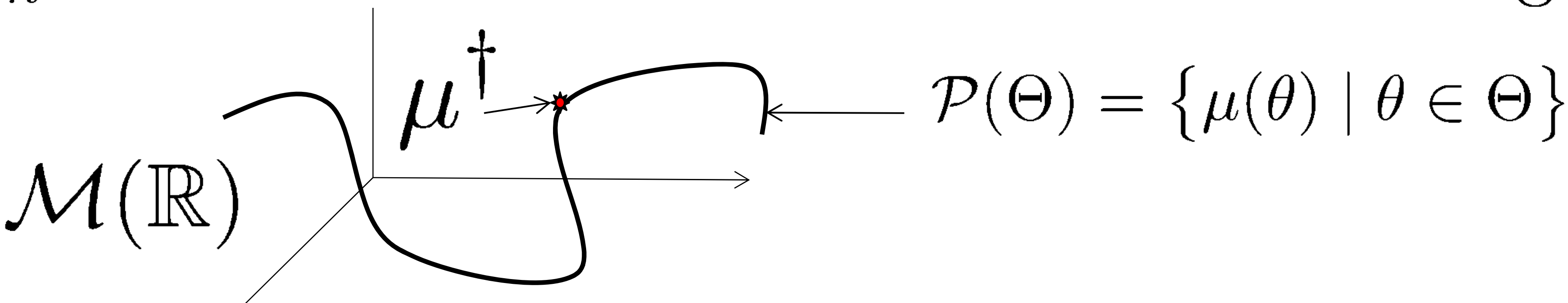
$$\mu^\dagger \notin \mathcal{P}(\Theta)$$

Model is misspecified



Asymptotic behavior of posterior estimates

p_n: Posterior distribution on Θ after observing d



If $\mu^\dagger \in \mathcal{P}(\Theta)$ (the model is well specified)

then (if $\Theta = \mathbb{R}^k$, under regularity conditions and under Crowmwell's rule)

$$\int_\Theta \Phi\big(\mu(\theta)\big) p_n(\mathrm{d}\theta) \xrightarrow{n\to\infty} \Phi(\mu^\dagger)$$

The Bayesian estimator is consistent

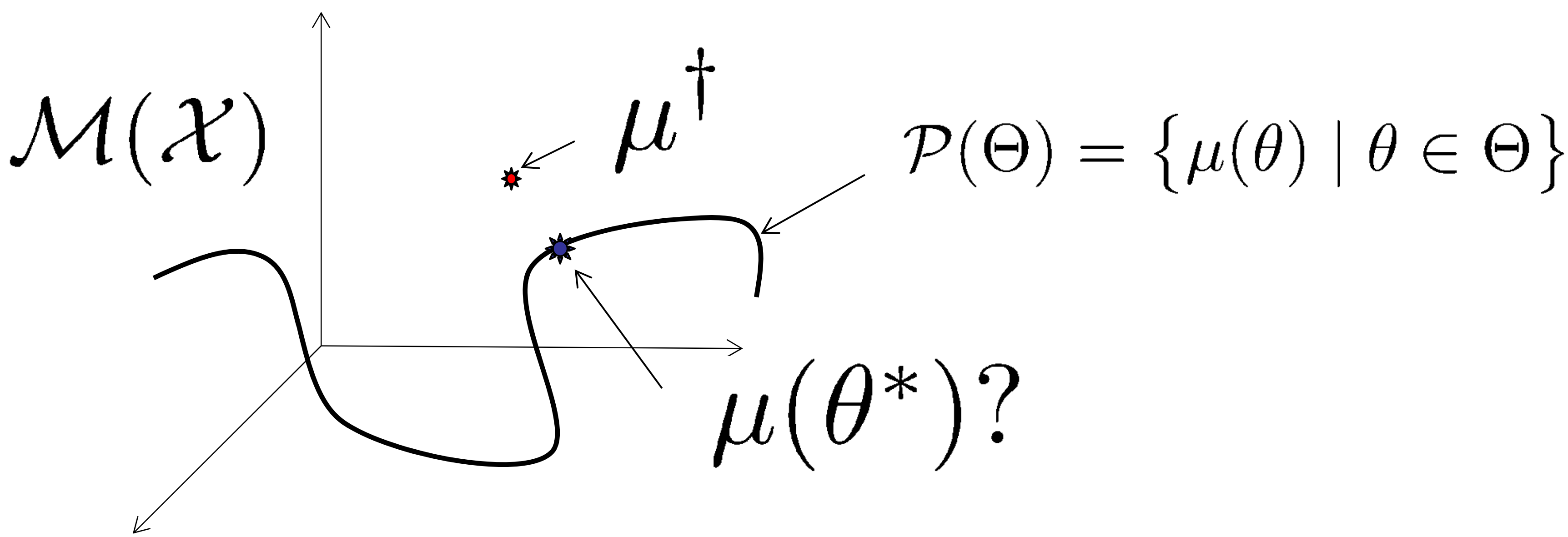
Bernstein-Von Mises CLTs (the rescaled limit is Normal)

If $\mu^\dagger \notin \mathcal{P}(\Theta)$ (the model is mis-specified)

then (if $\Theta = \mathbb{R}^k$, under regularity conditions and under Crowmwell's rule)

$$\int_\Theta \Phi\big(\mu(\theta)\big)p_n(\mathrm{d}\theta) \xrightarrow{n\to\infty} \Phi\big(\mu(\theta^*)\big)$$

θ^*: Minimizes $D_{\mathrm{KL}}\big(\mu^\dagger \| \mu(\theta)\big)$ over $\theta \in \Theta$



Example

$\mathcal{P}(\Theta)$: family of Gaussian models $\mu(\theta)$ with densities

$$\{\beta(\,\cdot\,,\theta) \mid \theta = (c,\sigma) \in \mathbb{R} \times \mathbb{R}_+\}$$

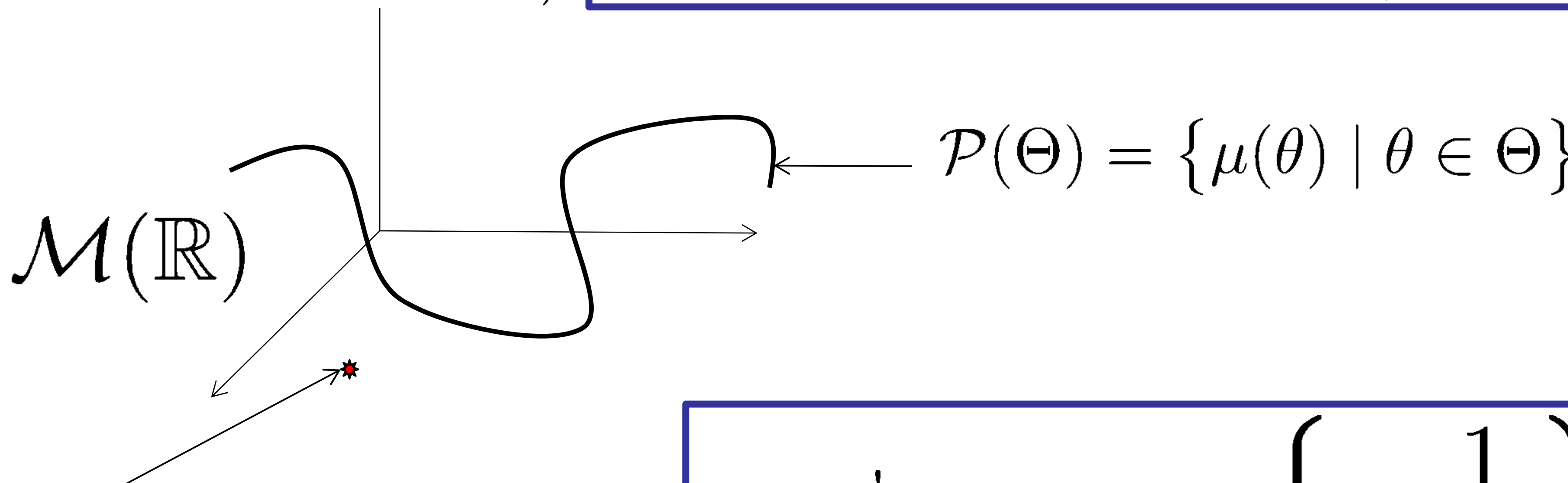
$$\beta(x,c,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$$

$\mu^\dagger$: potentially non-Gaussian with mean $c^\dagger$ and standard deviation $\sigma^\dagger$.

$$\int_\Theta \Phi\big(\mu(\theta)\big) p_n(\mathrm{d}\theta) \xrightarrow{n\to\infty} \Phi\big(\mu(c^\dagger,\sigma^\dagger)\big)$$

Example $\Phi(\mu) := \mu\left[|X - c_\mu| \geq t\sigma_\mu\right]$

Under the Gaussian model, $\Phi(\mu(c^\dagger, \sigma^\dagger)) = 1 + \operatorname{erf}\left(-\frac{t}{\sqrt{2}}\right)$,



$\mu^\dagger$: extreme case $\Phi(\mu^\dagger) = \min\left\{1, \frac{1}{t^2}\right\}$.

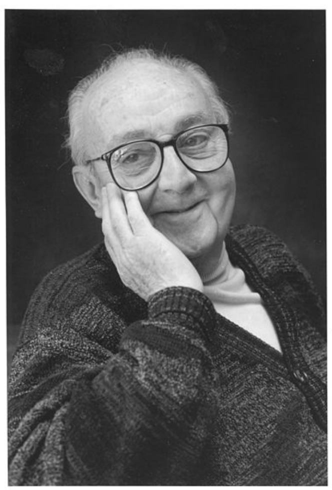
For $t = 6$ (the archetypically rare "6σ event"), the ratio between the two is approximately 1.4×10^7

Questions

What happens to posterior values if our Bayesian model is a little bit wrong?

How sensitive is Bayesian Inference to local misspecification?

G. E. P. Box



"Essentially, all models are wrong but some are useful"

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful?"

$$\mathcal{P} : \Theta \to \mathcal{M}(\mathbb{R})$$
$$\theta \to \mu(\theta)$$

ρ: TV distance on $\mathcal{M}(\mathbb{R})$ if Θ compact and $\mathcal{P}$ continuous

otherwise Prokhorov distance

Total variation distance

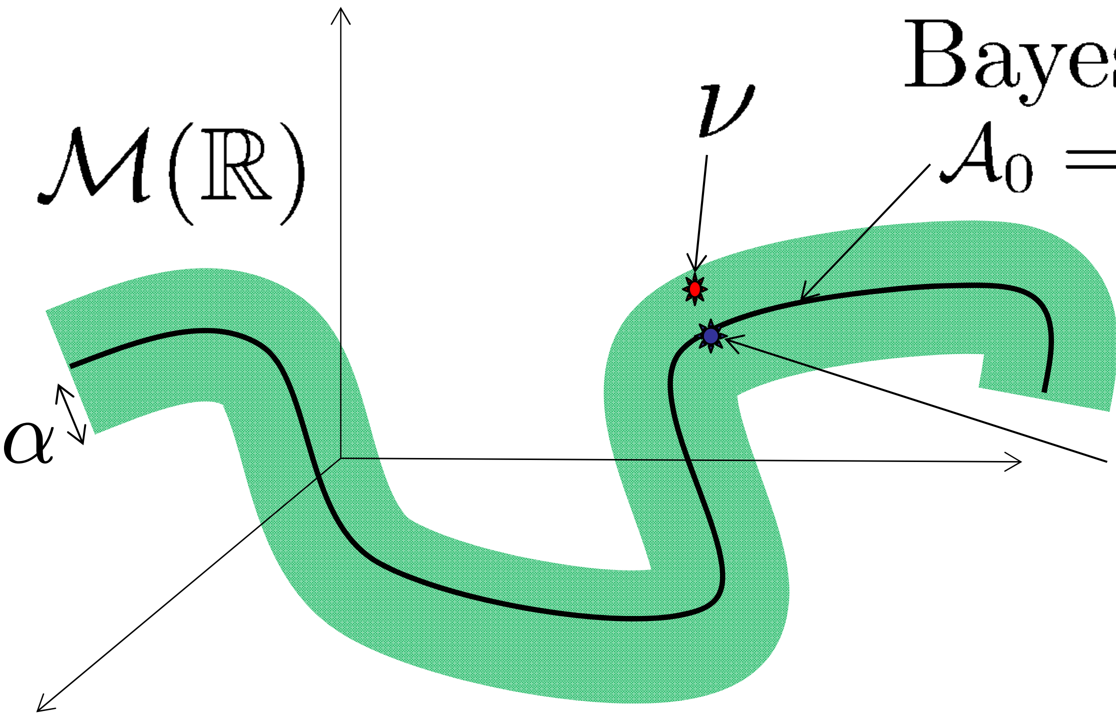
$$\rho(\mu,\nu) := \sup_{A\in\sigma(\mathbb{R})} \left|\mu(A) - \nu(A)\right|$$

Prokhorov distance

$$\rho(\mu,\nu) := \inf\left\{\epsilon > 0 \middle| \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for all } A \in \sigma(\mathbb{R})\right\}$$

$$A^\epsilon = \{x \in \mathbb{R} | d(x, A) \leq \epsilon\}$$

Perturbed Bayesian model



Bayesian model class
$\mathcal{A}_0 = \{\mu(\theta) \mid \theta \in \Theta\}$

Bayesian model

$\mu(\theta)$: Random element of $\mathcal{A}_0$

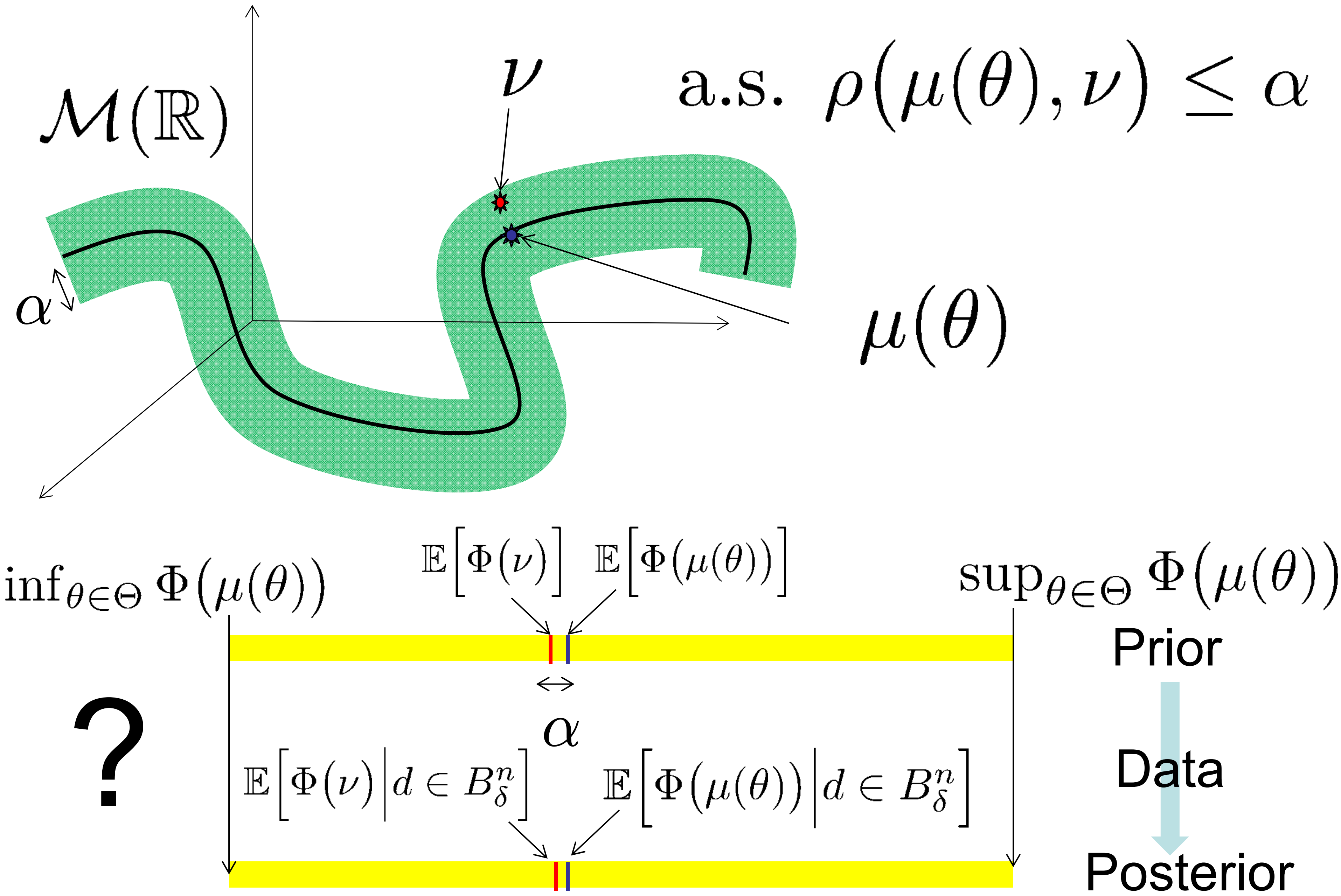
Perturbed Bayesian model

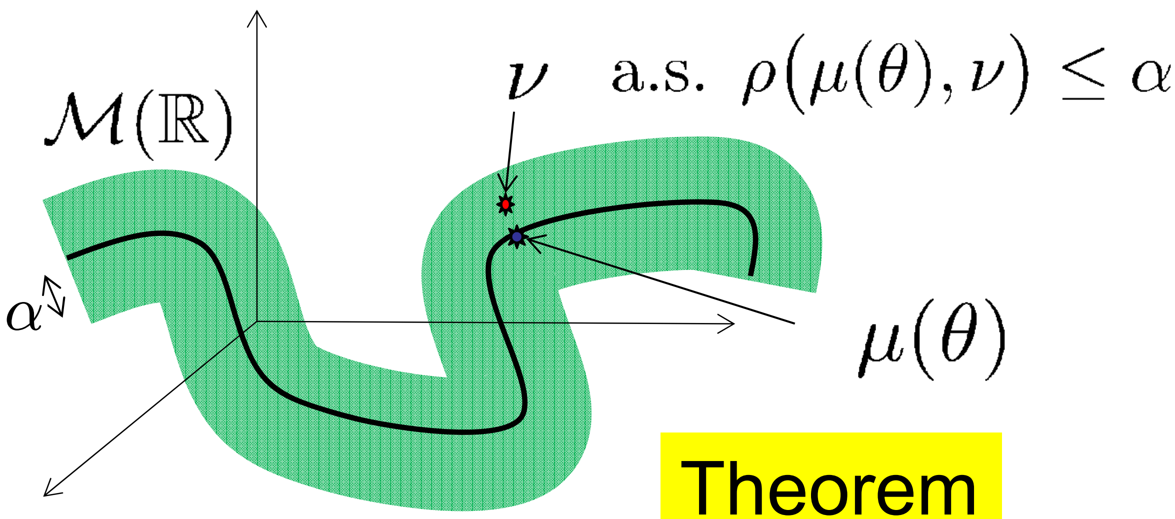
ν: Random element of $\mathcal{M}(\mathbb{R})$ such that a.s. $\rho\big(\mu(\theta), \nu\big) \le \alpha$

Total variation distance

$$\rho(\mu, \nu) := \sup_{E \in \sigma(\mathbb{R})} \big|\mu(E) - \nu(E)\big|$$

How Robust is the Bayesian Answer?





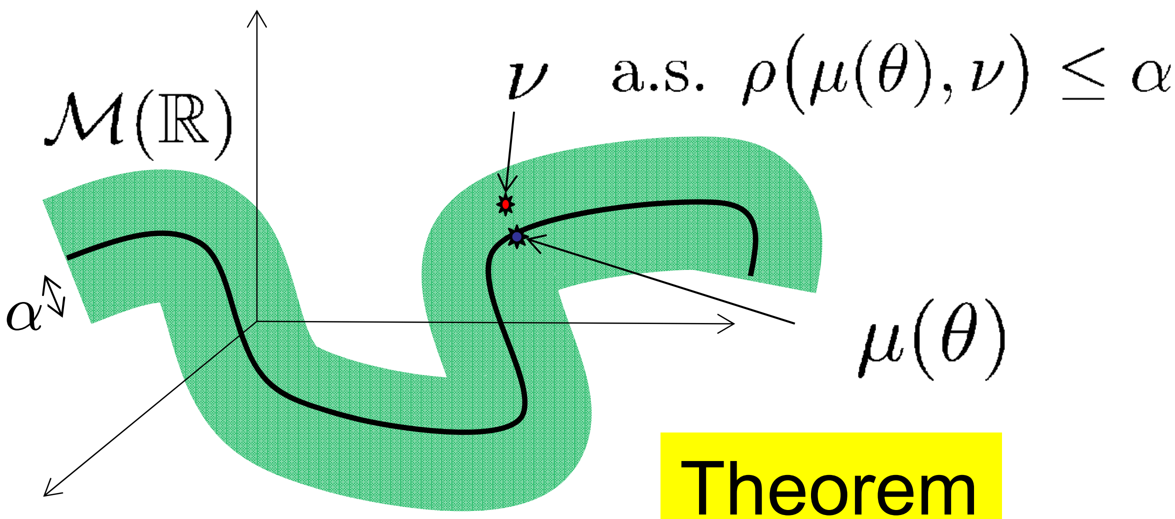
$$\mathcal{M}_\alpha := \{\nu\}$$

Theorem

If $\lim_{\delta\downarrow 0} \sup_{x\in\mathcal{X}} \sup_{\theta\in\Theta} \mathcal{P}(\theta)[B_\delta(x)] = 0$ then, for all $\alpha > 0$ there exists $\delta_c(\alpha) > 0$ such that for all $0 < \delta < \delta_c(\alpha)$ and all integers $n \ge 1$,

$$\operatorname*{esssup}_{p_0} \Phi\big(\mu(\theta)\big) \le \sup_{\nu\in\mathcal{M}_\alpha} \mathbb{E}\Big[\Phi\big(\nu\big)\Big| d \in B_\delta^n\Big]$$

$$\operatorname*{esssup}_{p_0} \Phi\big(\mu(\theta)\big) := \inf\{r \mid p_0\Big[\Phi\big(\mu(\theta)\big) > r\Big] = 0\}$$



$$\mathcal{M}_\alpha := \{\nu\}$$

Theorem

If $\lim_{\delta\downarrow 0}\sup_{x\in\mathcal{X}}\sup_{\theta\in\Theta}\mathcal{P}(\theta)[B_\delta(x)]=0$

then, for all $\alpha>0$ there exists $\delta_c(\alpha)>0$ such that for all $0<\delta<\delta_c(\alpha)$ and all integers $n\ge 1$,

$$\inf_{\nu\in\mathcal{M}_\alpha}\mathbb{E}\left[\Phi(\nu)\,\middle|\,d\in B^n_\delta\right]\le \operatorname*{essinf}_{p_0}\Phi\big(\mu(\theta)\big)$$

$$\operatorname*{essinf}_{p_0}\Phi\big(\mu(\theta)\big):=\sup\{r\mid p_0\big[\Phi\big(\mu(\theta)\big)<r\big]=0\}$$

Cromwell's rule

Every neighborhood of Θ has strictly positive mass under p_0

Implies consistency if the model is well specified

Implies maximal brittleness under local perturbations

$$\sup_{\theta\in\Theta}\Phi\big(\mu(\theta)\big) \leq \sup_{\nu\in\mathcal{M}_\alpha} \mathbb{E}\Big[\Phi(\nu)\Big| d\in B_\delta^n\Big]$$

$$\inf_{\nu\in\mathcal{M}_\alpha} \mathbb{E}\Big[\Phi(\nu)\Big| d\in B_\delta^n\Big] \leq \inf_{\theta\in\Theta}\Phi\big(\mu(\theta)\big)$$

Example

$$\mathcal{X} = \mathbb{R}$$

$$\Phi(\mu) = \mu[X \geq a]$$

$$\Theta \subset \mathbb{R}^k$$

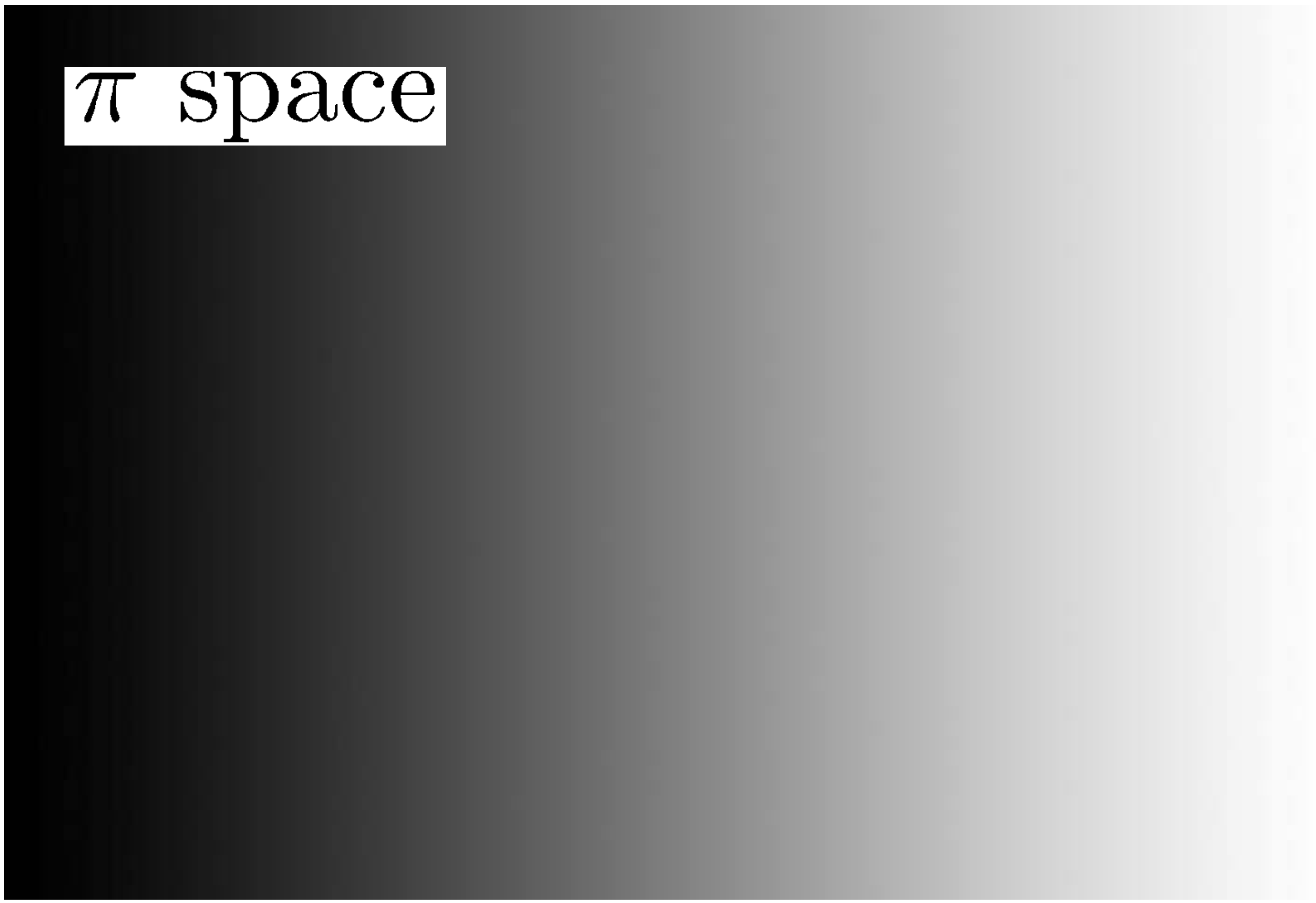
Generalization

$\mathcal{X}$: Polish space

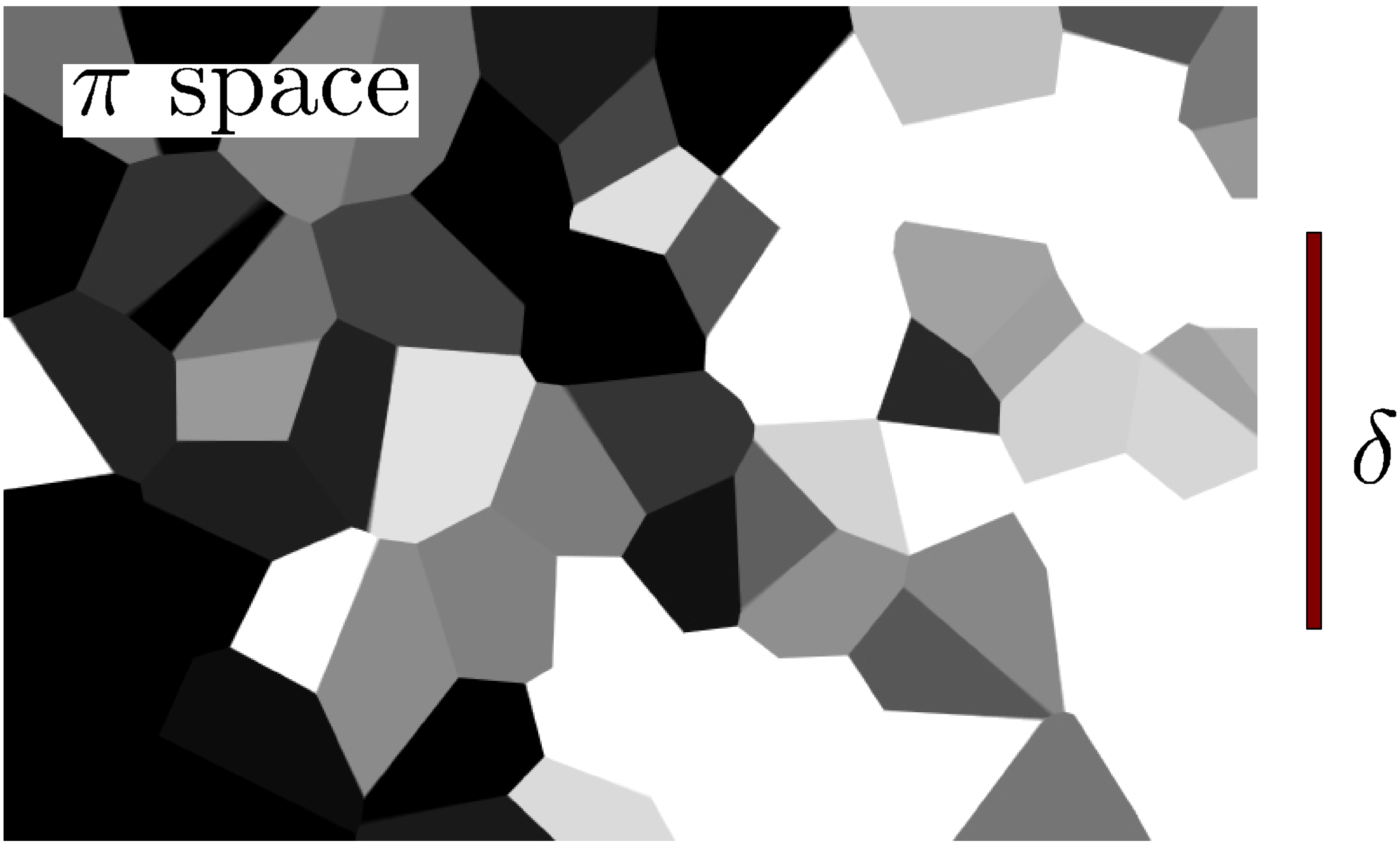
Φ: $\mathcal{M}(\mathcal{X}) \to \mathbb{R}$

Θ : Polish space

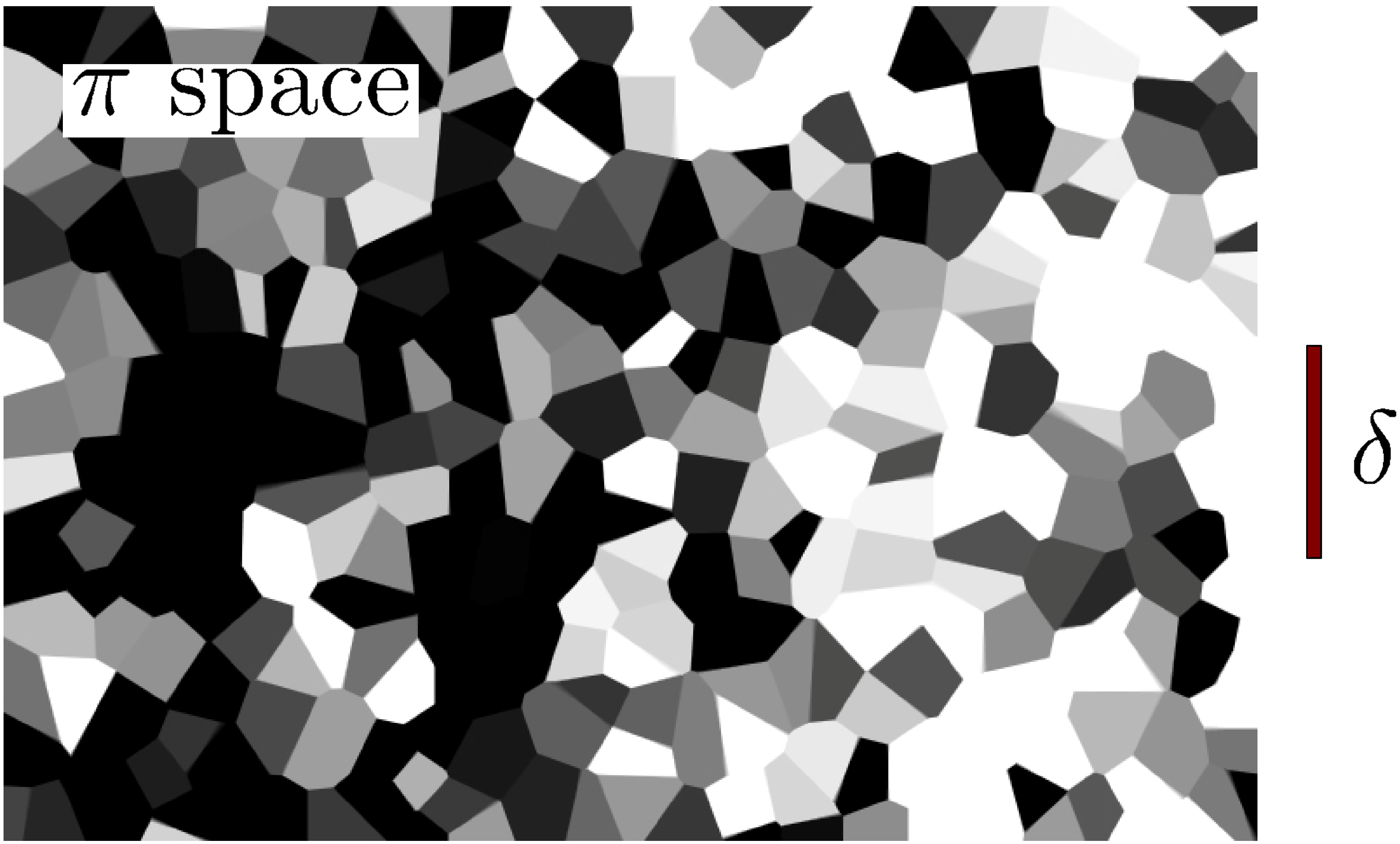
$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\right]$$



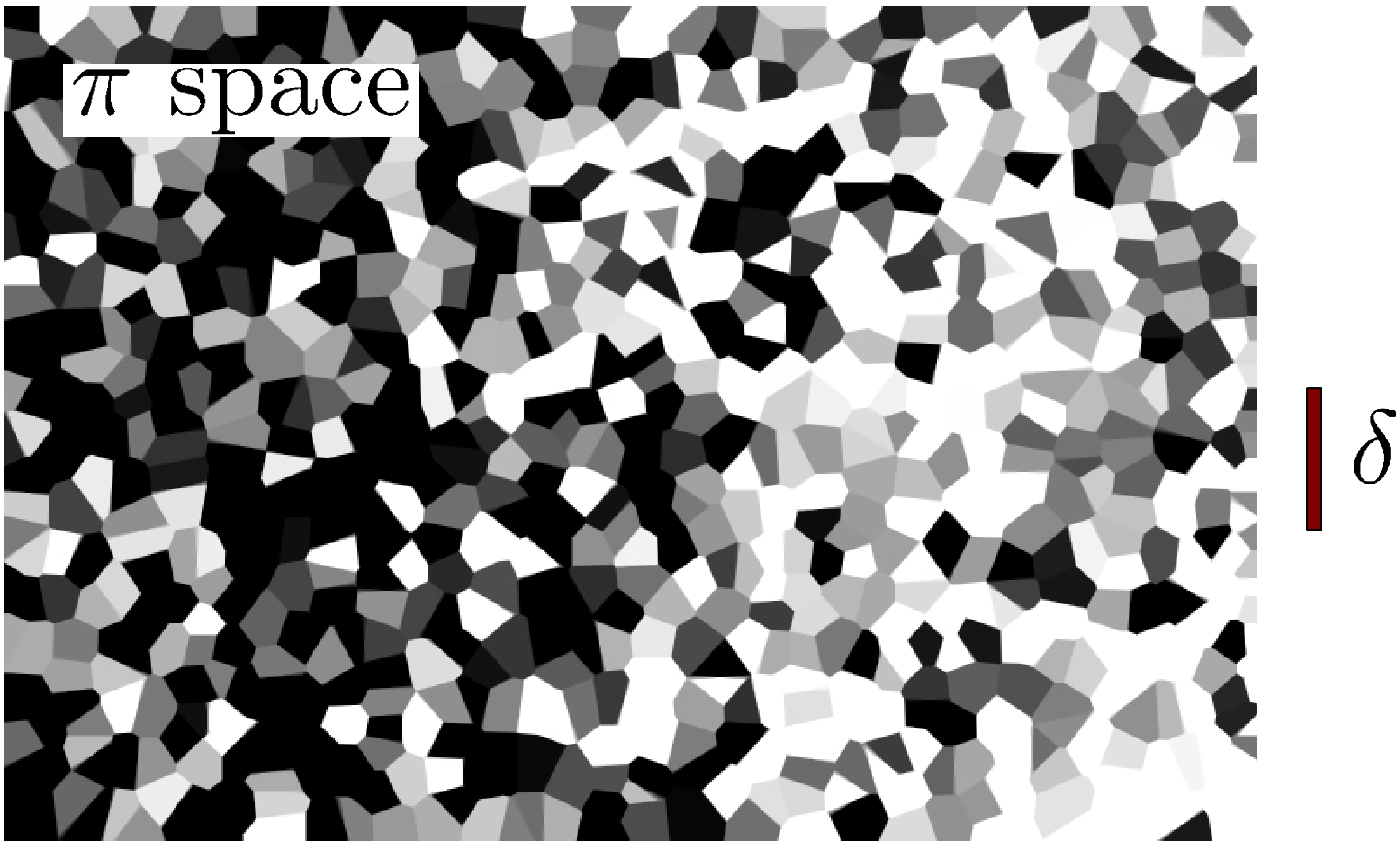
$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle| d \in B_\delta^n\right]$$



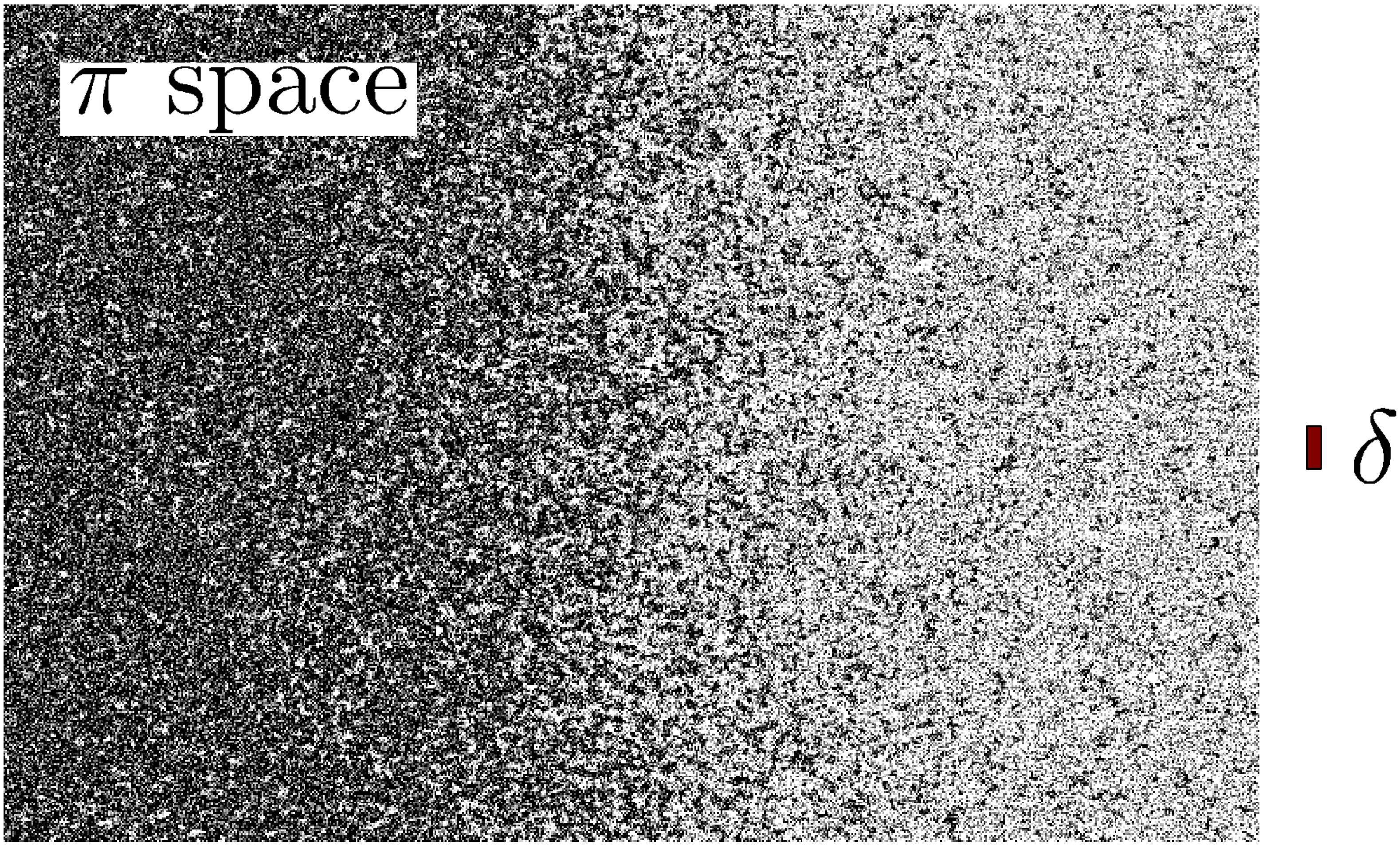
$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle| d\in B^n_\delta\right]$$



$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle| d\in B^n_\delta\right]$$



$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle| d \in B_\delta^n\right]$$



Hadamard well posed problem

1. A solution exists
2. The solution is unique
3. The solution's behavior hardly changes when there's a slight change in the initial condition

J. S. Hadamard
1865 –1963

Bayesian inference appears to be ill posed in the Hadamard sense (3)

Are these results compatible with classical Robust Bayesian Inference?

Framework is the same: Bayesian Sensitivity Analysis

Given a class Π of priors, compute

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\left[\Phi(\mu) \middle| d \in B_\delta^n\right]$$

$$\inf_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\left[\Phi(\mu) \middle| d \in B_\delta^n\right]$$

Classical Robust Bayesian Inference:

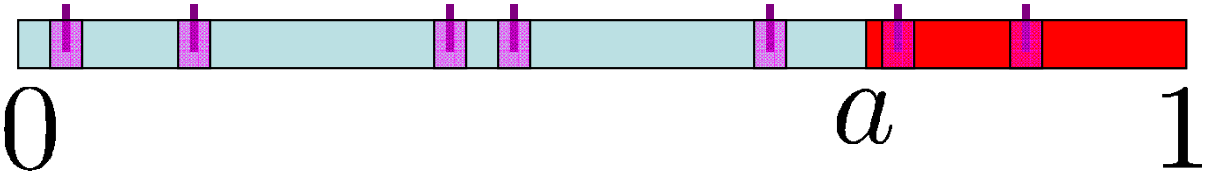
Box (1953)
Wasserman (1991)
Huber (1964)

Finite dimensional class of priors Π → Robustness

Our brittleness results:

Finite co-dimensional class of priors Π → Brittleness

Example



We want to estimate $\Phi(\mu^\dagger) = \mu^\dagger[X \geq a]$

$\mu^\dagger$: Unknown or partially known measure of probability on $[0, 1]$

We observe $d \in B_\delta^n := \prod_{i=1}^n B_\delta(x_i)$

Bayesian Answer

Assume $\mu^\dagger$ is the realization of a random measure on $[0, 1]$

π is the distribution of $\mu^\dagger$

$$\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle| d \in B_\delta^n\right]$$

Construction of π

Specify the distribution of $\left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots\right)$

Robustness?

$$\mathcal{M}([0,1]) \xrightarrow{\;\Psi\;} \mathbb{R}^k$$

$$\mu \longrightarrow \left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\right)$$

$$\Pi \xleftarrow{\;\Psi^{-1}\;} \mathbb{Q}$$

$\mathbb{Q}$: Uniform distribution on $\Psi\big(\mathcal{M}([0,1])\big)$

Π: Class of priors on $\mathcal{M}([0,1])$ such that if $\pi \in \Pi$ and $\mu \sim \pi$ then

$$\left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\right) \sim \mathbb{Q}$$

Π: Classes of priors on $\mathcal{M}([0,1])$ such that if $\pi \in \Pi$ and $\mu \sim \pi$ then

$$\Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \dots, \mathbb{E}_{X\sim\mu}[X^k]\Big) \sim \mathbb{Q}$$

Theorem As $\delta \downarrow 0$

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\Big[\Phi(\mu)\Big| d \in B_\delta^n\Big] \to 1$$

$$\inf_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\Big[\Phi(\mu)\Big| d \in B_\delta^n\Big] \to 0$$

sup and inf over $\pi \in \Pi$ s.t. $\mathbb{E}_{\mu\sim\pi}\Big[\mu^n\big[d \in B_\delta^n\big]\Big] > 0$

Generalization

$\mathcal{X}$: Polish space $\quad \mathcal{A} \subset \mathcal{M}(\mathcal{X})$

Ψ: Measurable map

$$\mathcal{M}(\mathcal{X}) \supset \mathcal{A} \xrightarrow{\quad \Psi \quad} \mathcal{Q}$$

Polish space

$$\mathcal{M}(\mathcal{A}) \supset \Pi \xleftarrow{\quad \Psi^{-1} \quad} \mathbb{Q} \quad \in \mathcal{M}(\mathcal{Q})$$

Π: Class of priors on $\mathcal{A}$

such that if $\pi \in \Pi$ and $\mu \sim \pi$

then $\Psi(\mu) \sim \mathbb{Q}$

Theorem

If

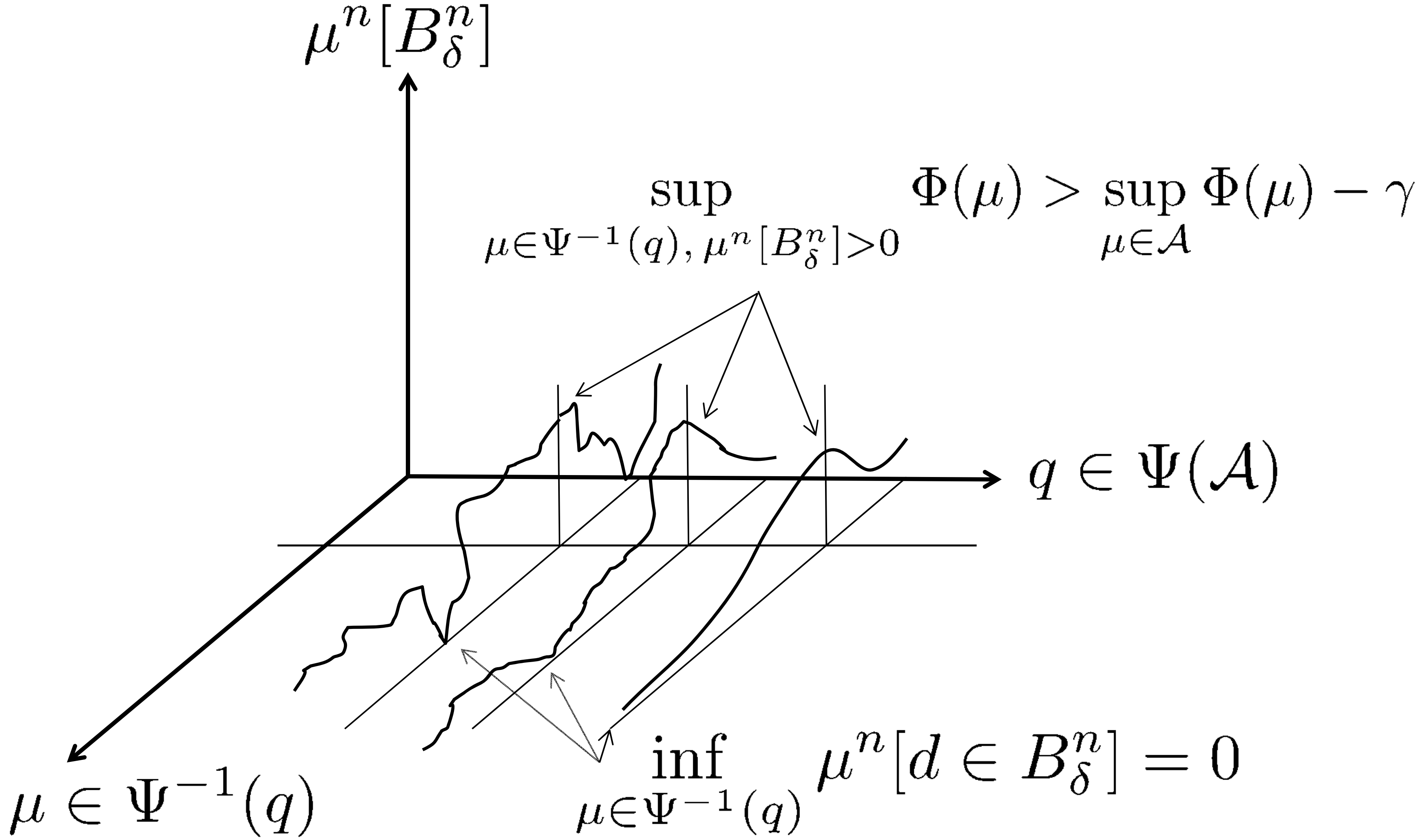
$$\mathbb{E}_{q\sim\mathbb{Q}}\left[\inf_{\mu\in\Psi^{-1}(q)}\mu^{n}[B_{\delta}^{n}]\right]=0$$

and for all $\gamma>0$

$$\mathbb{P}_{q\sim\mathbb{Q}}\left[\sup_{\mu\in\Psi^{-1}(q),\,\mu^{n}[B_{\delta}^{n}]>0}\Phi(\mu)>\sup_{\mu\in\mathcal{A}}\Phi(\mu)-\gamma\right]>0$$

Then

$$\sup_{\pi\in\Pi}\mathbb{E}_{\mu\sim\pi}\left[\Phi(\mu)\middle|d\in B_{\delta}^{n}\right]=\sup_{\mu\in\mathcal{A}}\Phi(\mu)$$



Example: $\Psi(\mu) = \Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\Big)$

Bayesian Sensitivity Analysis as it currently stands leads to Brittleness under finite information or local misspecification

Why?

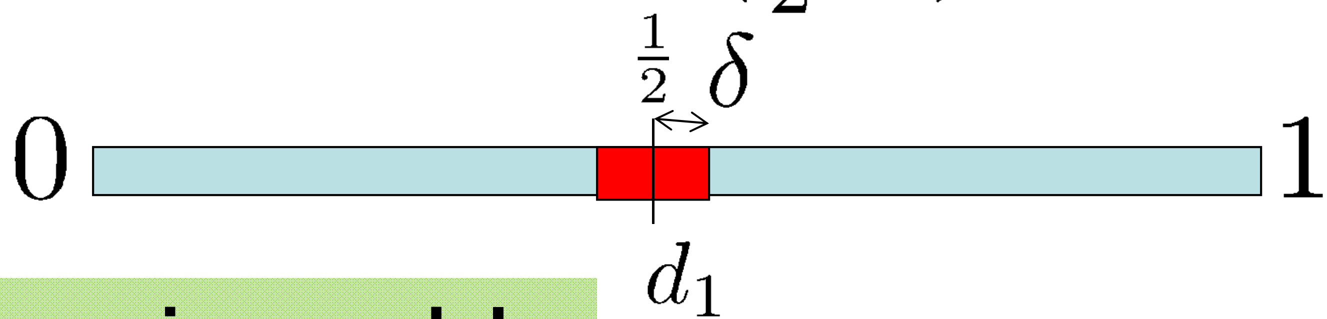
Let's look at one mechanism causing brittleness in a simple example

A simple example

We want to estimate $\Phi(\mu^\dagger) = \mathbb{E}_{X\sim\mu^\dagger}[X]$

$\mu^\dagger$: Unknown distribution on $[0,1]$

We observe $d_1 \in B(\frac{1}{2}, \delta)$

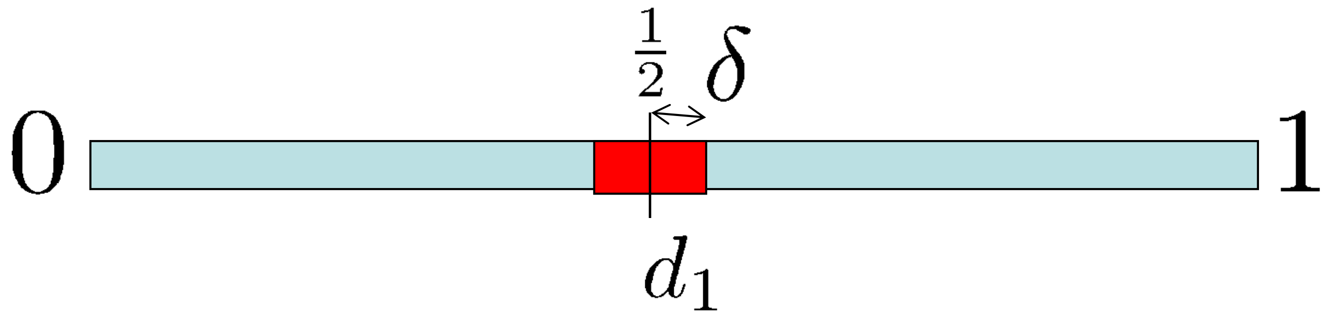


Two Bayesian models

$\mu^a(\theta)$: random measure on $[0,1]$

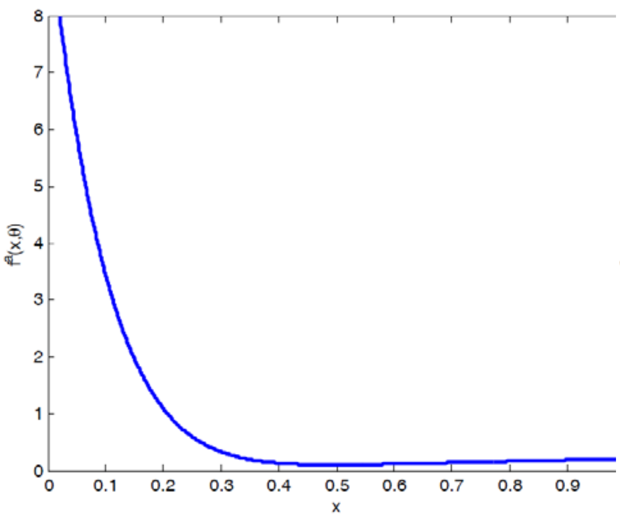
$\mu^b(\theta)$: random measure on $[0,1]$

θ: Uniformly distributed on $[0,1]$

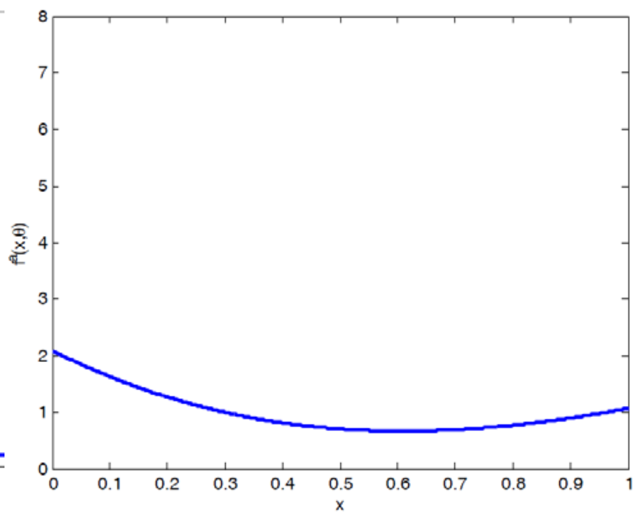


$\mu^a(\theta)$: Has density

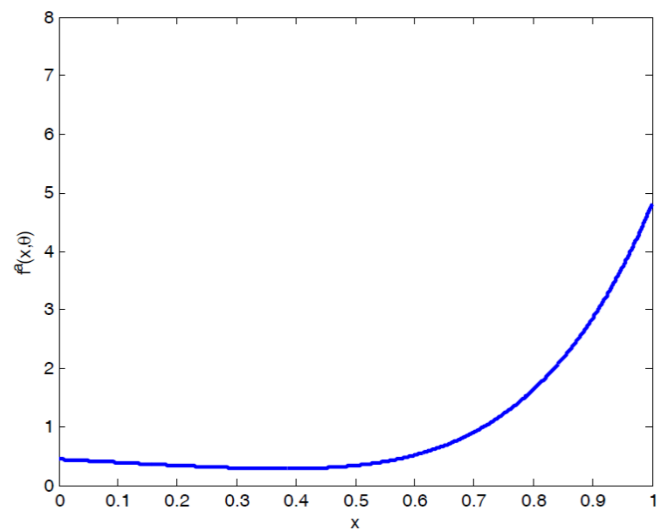
$$f^a(x,\theta) = (1-\theta)\frac{(1-x)^{\frac{1}{\theta}}}{1+1/\theta} + \theta\frac{x^{\frac{1}{1-\theta}}}{1+1/(1-\theta)}$$



$\theta = 0.1$



$\theta = 0.4$

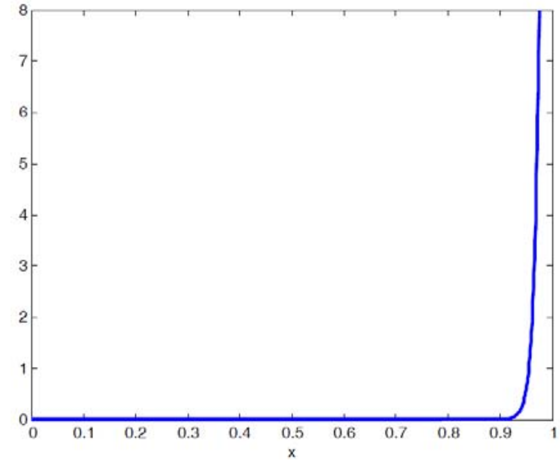


$\theta = 0.8$

$\mu^b(\theta)$ has density $f^b(x,\theta)$

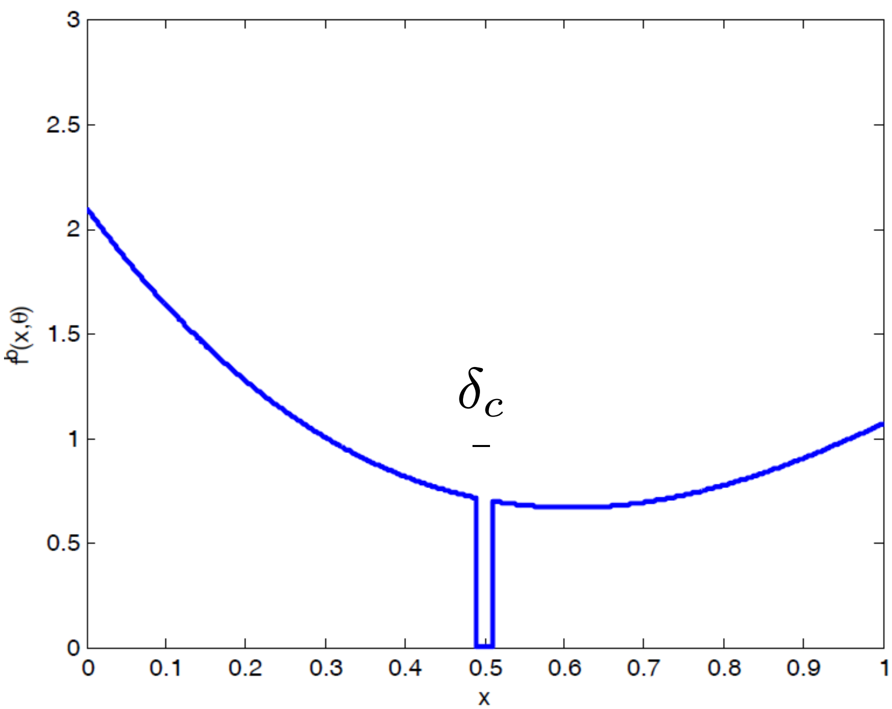
If $\theta \geq 0.999$

$$f^b(x,\theta) = f^a(x,\theta)$$



If $\theta < 0.999$

$$f^b(x,\theta) = f^a(x,\theta)\frac{1}{Z}\left(1_{\{x\notin(\frac{1}{2}-\delta_c,\frac{1}{2}+\delta_c)\}} + 10^{-9}1_{\{x\in(x_1-\delta_c,x_1+\delta_c)\}}\right)$$



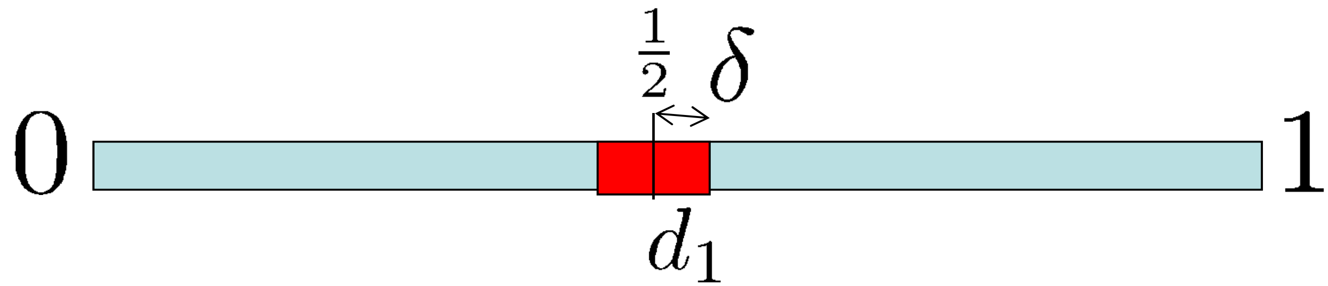
For all $\theta \in (0,1)$

$$TV\left(\mu^a(\theta),\mu^b(\theta)\right) \leq \delta_c$$

Prior Values

$$\mathbb{E}_\theta\left[\mathbb{E}_{X\sim\mu^a(\theta)}[X]\right] = \frac{1}{2}$$

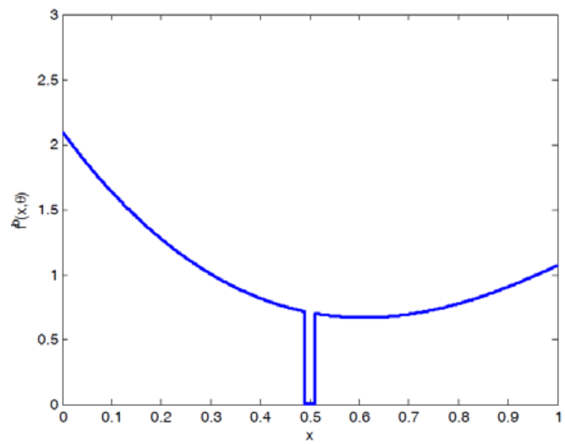
$$\mathbb{E}_\theta\left[\mathbb{E}_{X\sim\mu^b(\theta)}[X]\right] \approx \frac{1}{2}$$



Posterior Values

$$\mathbb{E}_\theta\left[\mathbb{E}_{X\sim\mu^a(\theta)}[X]\,\middle|\,d_1 \in \left(B\left(\tfrac{1}{2},\delta\right)\right)\right] = \frac{1}{2}$$

For $\delta < \delta_c$



$$TV\left(\mu^a(\theta),\mu^b(\theta)\right) \le \delta_c$$

$$\mathbb{E}_\theta\left[\mathbb{E}_{X\sim\mu^b(\theta)}[X]\,\middle|\,d_1 \in \left(B\left(\tfrac{1}{2},\delta\right)\right)\right] \approx 1$$

Bayesian Sensitivity Analysis as it currently stands leads to Brittleness under finite information or local misspecification

Why?

Bayesian Sensitivity Analysis as it currently stands is based on estimates posterior to the observation of data

Worst priors can achieve extreme values by making the probability of observing the data very small

Can we dismiss these priors because they depend on the data?

Problem

In the context of Bayesian Sensitivity analysis worst priors always depend on the data.

Dismissal of Bayesian Sensitivity Analysis

Tautology (circular reasoning) in its application

Can we dismiss these priors because they can ``look nasty’’ and make the probability of observing the data very small?

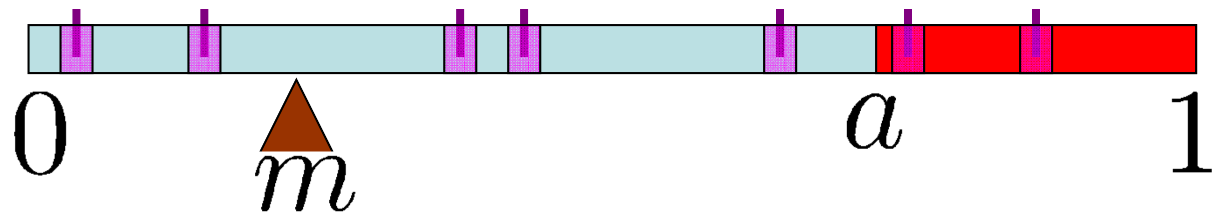
Problem

These priors are not isolated monsters but only directions of instability and these instabilities grow with the number of data points

How do we know that?

Let’s add a constraint on the probability of observing the data.

Example



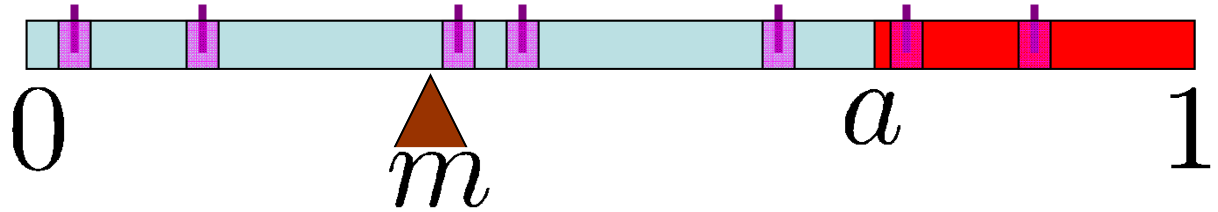
We want to estimate $\Phi(\mu^\dagger) = \mu^\dagger[X \geq a]$

$\mu^\dagger$: Unknown or partially known measure of probability on $[0,1]$

We observe $d \in B_\delta^n := \prod_{i=1}^n B_\delta(x_i)$

We believe

$\mu^\dagger$ is the realization of a random measure on $[0,1]$ whose mean value is on average m



Bayesian Model class $\mathcal{A} = \mathcal{M}\big([0,1]\big)$

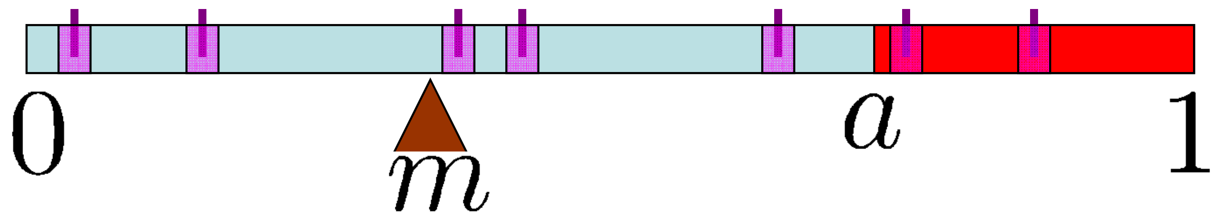
$$\Pi := \left\{\pi \in \mathcal{M}(\mathcal{A}) \mid \mathbb{E}_{\mu\sim\pi}\big[\mathbb{E}_\mu[X]\big] = m\right\}$$

Thm

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\big[\mu[X \geq a]\big] = \frac{m}{a}$$

For $\delta < \delta_c$

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\big[\mu[X \geq a] \big| d \in B^n_\delta\big] = 1$$



New Bayesian Model class

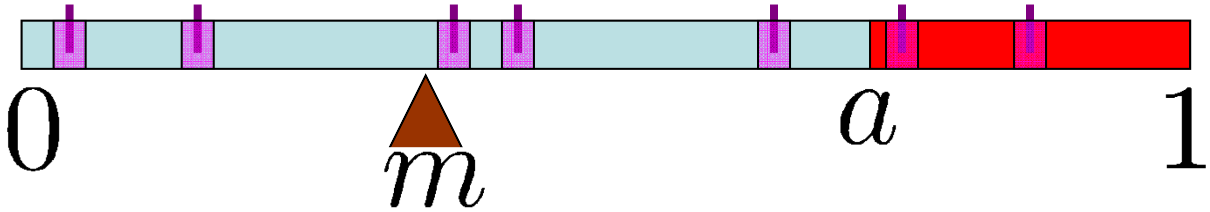
$$\mathcal{A} := \{\mu \in \mathcal{M}[0,1] \mid \frac{1}{\alpha}\mu_0^n[B_\delta^n] \leq \mu^n[B_\delta^n] \leq \alpha \mu_0^n[B_\delta^n]\}$$

μ_0: arbitrary distribution on $[0,1]$ with strictly positive density
e.g. uniform distribution on $[0,1]$

If $\alpha = 1$ then the data is equiprobable under all μ in the Bayesian model class and posterior values are equal to prior values.

$$\mathbb{E}_{\mu\sim\pi}[\Phi(\mu)] = \mathbb{E}_{\mu\sim\pi, d\sim\mu^n}[\Phi(\mu)|d \in B_\delta^n]$$

If $\alpha = 1$ then learning is not possible



New Bayesian Model class

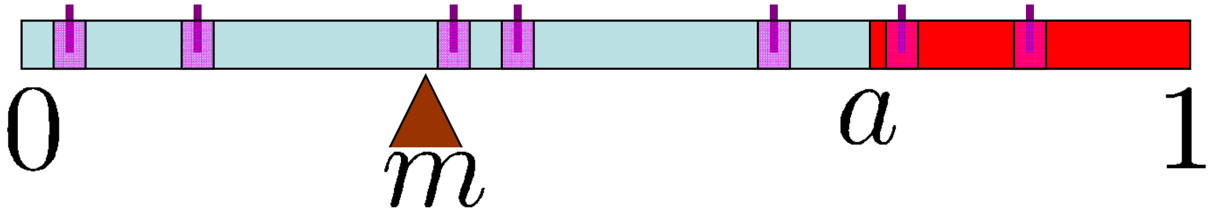
$$\mathcal{A} := \left\{ \mu \in \mathcal{M}[0,1] \mid \frac{1}{\alpha}\mu_0^n[B_\delta^n] \le \mu^n[B_\delta^n] \le \alpha \mu_0^n[B_\delta^n] \right\}$$

$$\Pi := \left\{ \pi \in \mathcal{M}(\mathcal{A}) \mid \mathbb{E}_{\mu\sim\pi}\big[\mathbb{E}_\mu[X]\big] = m \right\}$$

Thm $\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\big[\mu[X \ge a]\big] \le \frac{m}{a}$

$$\lim_{\delta\downarrow 0}$$

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\Big[\mu[X \ge a] \Big| d \in B_\delta^n\Big] = \frac{1}{1+\frac{1}{\alpha^2}\frac{a-m}{m}}$$



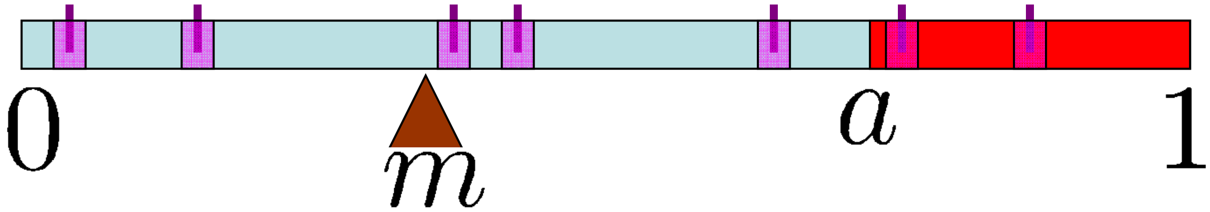
$$m = \frac{3}{8} \qquad a = \frac{3}{4}$$

$$\mathcal{A} := \left\{ \mu \in \mathcal{M}[0,1] \;\middle|\; \frac{1}{\alpha}\mu_0^n[B_\delta^n] \le \mu^n[B_\delta^n] \le \alpha\mu_0^n[B_\delta^n] \right\}$$

Thm $\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \ge a]\right] \le \frac{1}{2}$

$$\lim_{\delta\downarrow 0}$$

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \ge a] \middle| d \in B_\delta^n\right] = \frac{1}{1+\frac{1}{\alpha^2}}$$



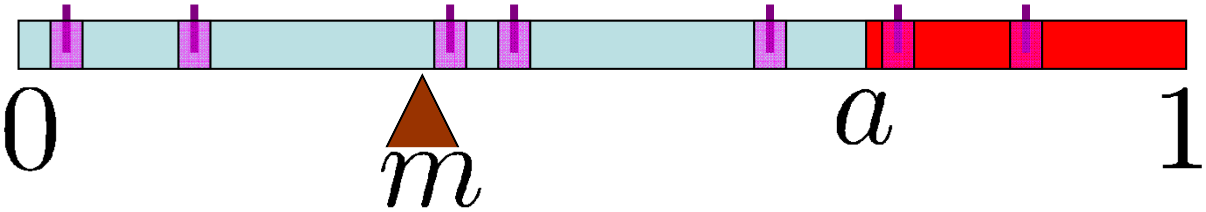
$\alpha = 2$ $\qquad m = \frac{3}{8} \qquad a = \frac{3}{4}$

$$\mathcal{A} := \left\{\mu \in \mathcal{M}[0,1] \mid \tfrac{1}{\alpha}\mu_0^n[B_\delta^n] \leq \mu^n[B_\delta^n] \leq \alpha\mu_0^n[B_\delta^n]\right\}$$

Thm $\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \geq a]\right] \leq \frac{1}{2}$

$\lim_{\delta\downarrow 0}$

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \geq a]\middle| d \in B_\delta^n\right] = 0.8$$



$\alpha = 10$

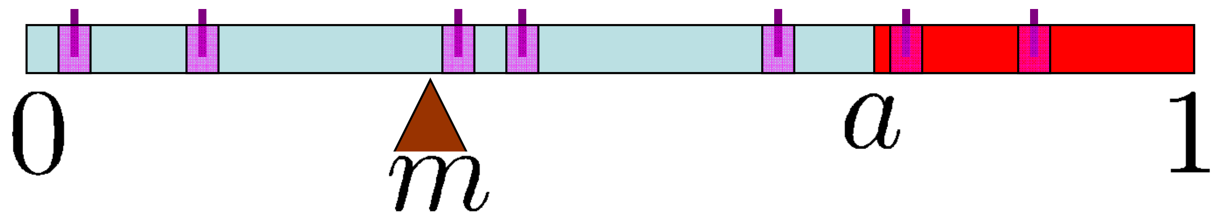
$m = \frac{3}{8}$ $a = \frac{3}{4}$

$\mathcal{A} := \{\mu \in \mathcal{M}[0,1] \mid \frac{1}{\alpha}\mu_0^n[B_\delta^n] \leq \mu^n[B_\delta^n] \leq \alpha\mu_0^n[B_\delta^n]\}$

Thm $\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\left[\mu[X \geq a]\right] \leq \frac{1}{2}$

$\lim_{\delta \downarrow 0}$

$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\left[\mu[X \geq a] \middle| d \in B_\delta^n\right] = 0.99$



$$m = \frac{3}{8} \qquad a = \frac{3}{4}$$

New Bayesian Model class

$$\mathcal{A} := \left\{ \mu \in \mathcal{M}[0,1] \mid \frac{1}{\gamma}\mu_0[B_\delta(x_i)] \le \mu[B_\delta(x_i)] \le \gamma \mu_0[B_\delta(x_i)] \right\}$$

Thm $\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[\mu[X \ge a] \right] \le \frac{1}{2}$

$$\lim_{\delta \downarrow 0}$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[\mu[X \ge a] \middle| d \in B_\delta^n \right] = \frac{1}{1 + \frac{1}{\gamma^{2n}}}$$

Effects of a uniform constraint on the probability of the data under finite information in the Bayesian model class

$$\mathcal{A} := \left\{\mu \in \mathcal{A}_0 \mid \frac{1}{\alpha}\mu_0^n[B_\delta^n] \leq \mu^n[B_\delta^n] \leq \alpha\mu_0^n[B_\delta^n]\right\}$$

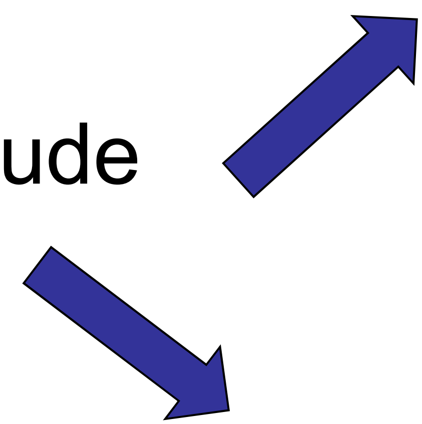
$\alpha = 1$ $\alpha \gg 1$

Learning not possible — Learning possible

Method is robust — Method is brittle

Learning Aptitude

Robustness



What is the stability condition for using Bayesian inference under finite information?

Numerically solving a PDE ⟷ CFL condition

Using Bayesian Inference under finite information ⟷ ?

What about using the KL-divergence (relative entropy)

Problem

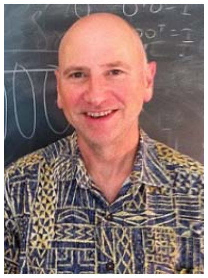
Closeness in KL divergence cannot be tested with discrete data.

Requires the non-singularity of the data generating distribution with respect to the Bayesian model.

Local Sensitivity Analysis (Frechet derivative) suggests blow-up with prob one as the number of data points goes to infinity

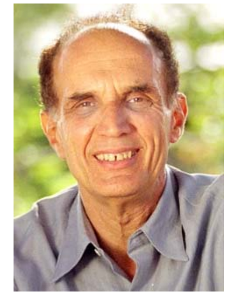
P Gustafson & L Wasserman 1995: Local Sensitivity Diagnostics for Bayesian Inference

Valid for the broader class of Φ-divergences (KL, Hellinger)



What about getting out of the strict Bayesian Inference framework for robustness/accuracy estimates?

Bradley Efron (2013): Bayes' theorem in the 21st century
Without genuine prior information "Bayesian calculations cannot be uncritically accepted and should be checked by other methods, which usually means frequentistically."



How do we do that with limited sample data?

We can compute sensitivity and accuracy estimates **before** the observation of the data.

Classical Bayesian Sensitivity Analysis

Compute robustness estimates after the observation of the data

Given the data d compute

$$\sup_{\pi,\pi'\in\Pi}\Big[\mathbb{E}_{\mu\sim\pi}\big[\Phi(\mu)\big|d\big]-\mathbb{E}_{\mu\sim\pi'}\big[\Phi(\mu)\big|d\big]\Big]$$

Alternative

Compute robustness estimates before the observation of the data

Take the average with respect to the distribution of the data

$$\sup_{\pi,\pi'\in\Pi}\mathbb{E}_{\mu\sim\pi,d\sim\mu^n}\Big|\mathbb{E}_{\mu\sim\pi}\big[\Phi(\mu)\big|d\big]-\mathbb{E}_{\mu\sim\pi'}\big[\Phi(\mu)\big|d\big]\Big|$$

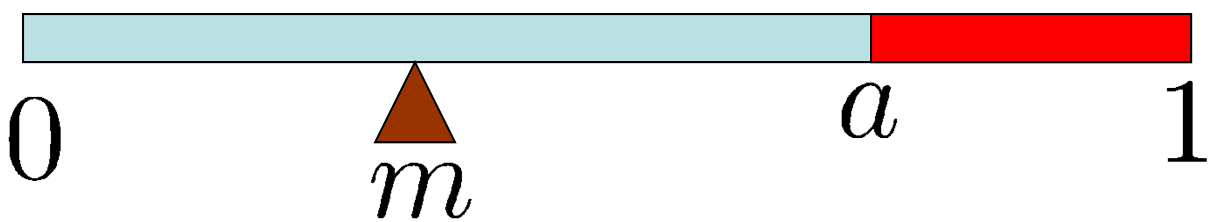
Problem Need a form of calculus allowing us to solve optimization problems over measures over spaces of measures

A simple example

What is the least upper bound on $\mathbb{P}[X \geq a]$

If all you know is $\mathbb{E}[X] \leq m$

and $\mathbb{P}[0 \leq X \leq 1] = 1$?

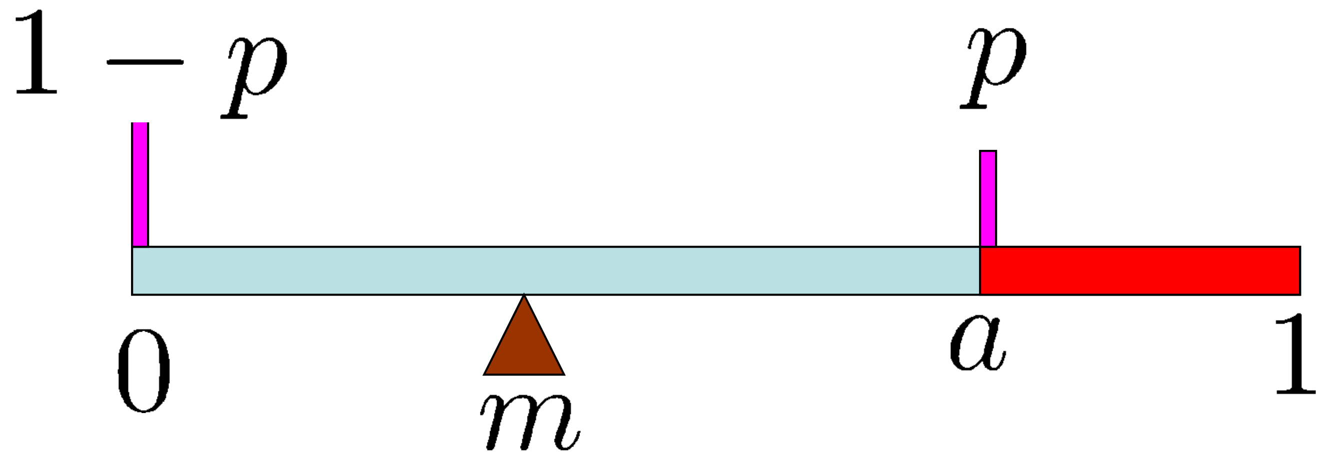


Answer

$$\sup_{\mu \in \mathcal{A}} \mu\left[X \geq a\right]$$

$$\mathcal{A} = \{\mu \in \mathcal{M}([0,1]) \,|\, \mathbb{E}_{\mu}[X] \leq m\}$$

You are given one pound of play-doh.
How much mass can you put above a while keeping the seesaw balanced around m?



Answer

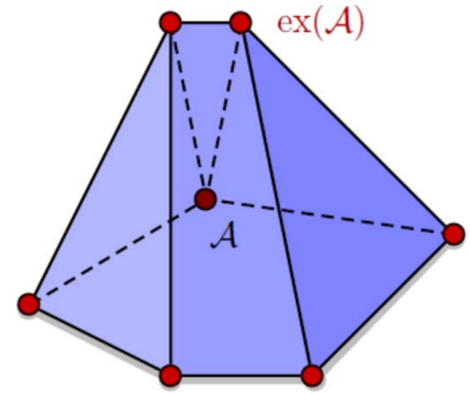
$$\begin{cases} \max p \\ \text{subject to } a\,p \leq m \end{cases}$$

Markov's inequality

$$\sup_{\mu \in \mathcal{A}} \mu\left[X \geq a\right] = \frac{m}{a}$$

$$\mathcal{A} = \{\mu \in \mathcal{M}([0,1]) \,|\, \mathbb{E}_\mu[X] \leq m\}$$

Generalization



$$\{f : \mathcal{X} \to \mathbb{R},\ \mu \in \mathcal{P}(\mathcal{X})\}$$

$$\left\{ f : \mathcal{X} \to \mathbb{R},\ \mu \in \mathcal{P}(\mathcal{X}) \,\middle|\, \mu = \sum_{i=1}^{k} \alpha_k \delta_{x_k} \right\}$$

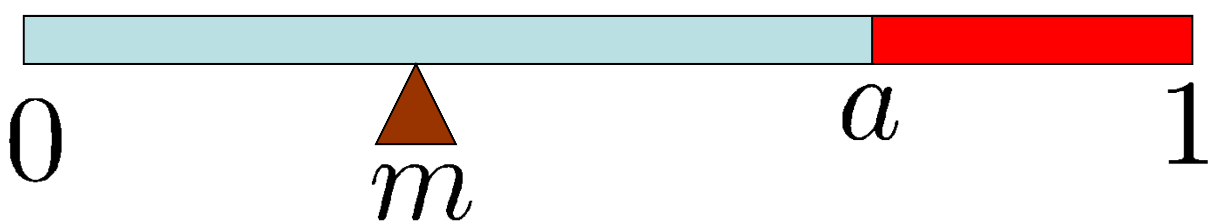
$$\{f : \{1,2,\ldots,n\} \to \mathbb{R},\ \mu \in \mathcal{P}(\{1,2,\ldots,n\})\}$$

$$\{\{1,2,\ldots,q\},\ \mu \in \mathcal{P}(\{1,2,\ldots,n\})\}$$

Optimal Uncertainty Quantification. Houman Owhadi, Clint Scovel, Tim Sullivan, Michael McKerns and Michael Ortiz. **SIAM Review** Vol. 55, No. 2 : pp. 271-345, 2013

New form of reduction calculus
A simple example

10.000 children are given one pound of play-doh. On average, how much mass can they put above a While, on average, keeping the seesaw balanced around m?



Paul is given one pound of play-doh. What can you say about how much mass he is putting above a if all you have is the belief that he is keeping the seesaw balanced around m?

What is the least upper bound on

$$\mathbb{E}_{\mu\sim\pi}\left[\mu[X \geq a]\right]$$

If all you know is $\mathbb{E}_{\mu\sim\pi}\left[\mathbb{E}_{\mu}[X]\right] = m$?

$$\mu \in \mathcal{A} := \mathcal{M}\big([0,1]\big)$$

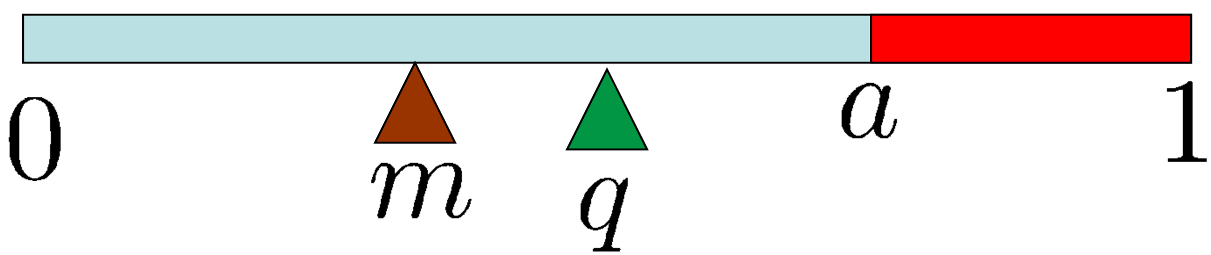
Answer

$$\sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\left[\mu[X \geq a]\right]$$

$$\Pi := \left\{\pi \in \mathcal{M}(\mathcal{A}) \,:\, \mathbb{E}_{\mu\sim\pi}\left[\mathbb{E}_{\mu}[X]\right] = m\right\}$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\left[\mu[X \geq a]\right]$$

$$\Pi := \left\{\pi \in \mathcal{M}\big(\mathcal{M}([0,1])\big) \ : \ \mathbb{E}_{\mu \sim \pi}\left[\mathbb{E}_{\mu}[X]\right] = m\right\}$$

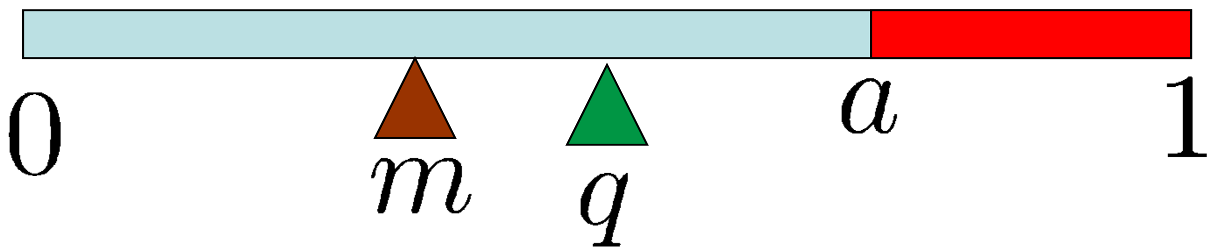


Theorem

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\left[\mu[X \geq a]\right] = \sup_{\mathbb{Q} \in \mathcal{M}([0,1]) \,:\, \mathbb{E}_{\mathbb{Q}}[q] = m} \mathbb{E}_{q \sim \mathbb{Q}}\left[\sup_{\mu \in \mathcal{M}([0,1]) \,:\, \mathbb{E}_{\mu}[X] = q} \mu[X \geq a]\right]$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\big[\mu[X \geq a]\big]$$

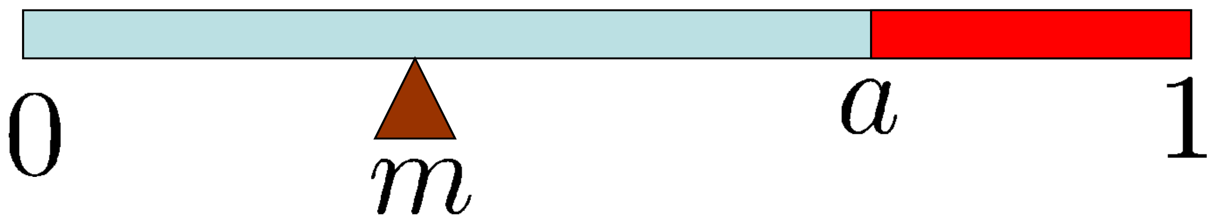
$$\Pi := \Big\{\pi \in \mathcal{M}\big(\mathcal{M}([0,1])\big) \;:\; \mathbb{E}_{\mu\sim\pi}\big[\mathbb{E}_{\mu}[X]\big] = m\Big\}$$



$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi}\big[\mu[X \geq a]\big] = \sup_{\mathbb{Q} \in \mathcal{M}([0,1]) \,:\, \mathbb{E}_{\mathbb{Q}}[q] = m} \mathbb{E}_{q \sim \mathbb{Q}}\Big[\min\big(\frac{q}{a}, 1\big)\Big]$$

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[\mu[X \geq a] \right]$$

$$\Pi := \left\{ \pi \in \mathcal{M}(\mathcal{M}([0,1])) \; : \; \mathbb{E}_{\mu \sim \pi} \left[\mathbb{E}_{\mu}[X] \right] = m \right\}$$



$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \left[\mu[X \geq a] \right] = \frac{m}{a}$$

New form of reduction calculus

$$\mathcal{M}(\mathcal{X}) \supset \mathcal{A} \xrightarrow{\Psi} \mathcal{Q}$$

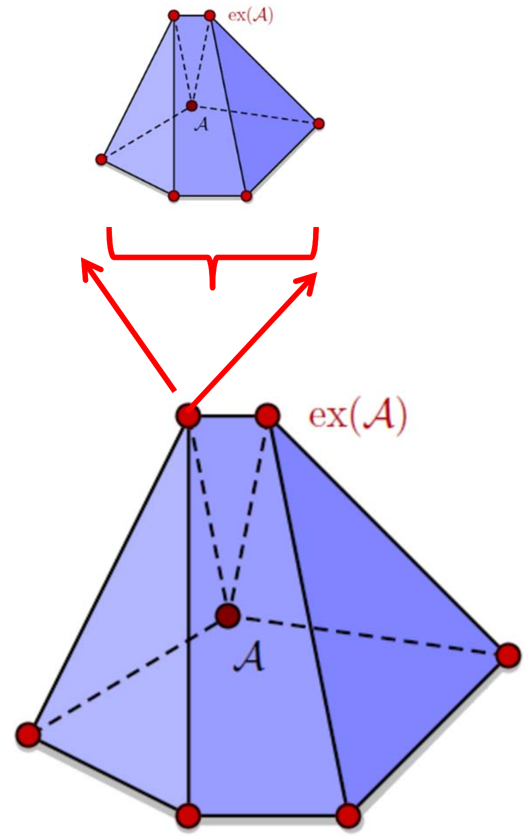
Polish space

$$\mathcal{M}(\mathcal{A}) \supset \Pi \xleftarrow{\Psi^{-1}} \mathfrak{Q} \subset \mathcal{M}(\mathcal{Q})$$

Theorem

$$\sup_{\pi \in \Psi^{-1}\mathfrak{Q}} \mathbb{E}_{\mu \sim \pi}\left[\Phi(\mu)\right] = \sup_{\mathbb{Q} \in \mathfrak{Q}} \left[\mathbb{E}_{q \sim \mathbb{Q}}\left[\sup_{\mu \in \Psi^{-1}(q)} \Phi(\mu)\right]\right]$$

Can we do some math with this form of calculus?



New Reproducing Kernel Hilbert Spaces and Selberg Integral formulas

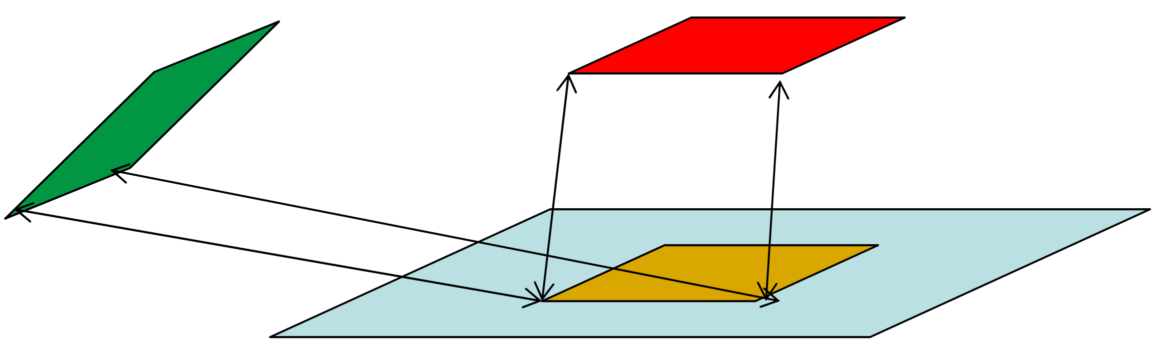
$$\int_{I^m} \Sigma t^{-1} \cdot \prod_{j=1}^m t_j^2 (1-t_j)^2 \Delta_m^4(t) dt = \frac{S_m(5,1,2) - S_m(3,3,2)}{2}$$

$$\int_{I^m} \Sigma t^{-1} \cdot \prod_{j=1}^m t_j^2 \cdot \Delta_m^4(t) dt = \frac{m}{2} S_{m-1}(5,3,2)$$

$$\Delta_m(t) := \prod_{j<k} (t_k - t_j) \quad I := [0,1]$$

$$\big(\Sigma\phi\big)(t) := \sum_{j=1}^m \phi(t_j), \quad t \in I^m$$

$$S_n(\alpha,\beta,\gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}$$



Infinite dim.

↕

Finite dim.

$$e_j(t) := \sum_{i_1<\cdots<i_j} t_{i_1}\cdots t_{i_j}$$

Π_0^n: n-th degree polynomials which vanish on the boundary of $[0,1]$

$M_n \subset \mathbb{R}^n$: set of $q = (q_1,\ldots,q_n) \in \mathbb{R}^n$ such that there exists a probability measure μ on $[0,1]$ with $\mathbb{E}_\mu[X^i] = q_i$ with $i \in \{1,\ldots,n\}$.

Theorem

Bi-orthogonal systems of Selberg Integral formulas

Consider the basis of Π_0^{2m-1} consisting of the associated Legendre polynomials $Q_j, j = 2,..,2m-1$ of order 2 translated to the unit interval I. For $k = 2,..,2m-1$ define

$$a_{jk} := \frac{(j+k+k^2)\Gamma(j+2)\Gamma(j)}{\Gamma(j+k+2)\Gamma(j-k+1)}, \quad k \le j \le 2m-1$$

$$\tilde{h}_k(t) := \sum_{j=k}^{2m-1} (-1)^{j+1} a_{jk} e_{2m-1-j}(t,t)\,.$$

Then for $j = k \bmod 2$, $j,k = 2,..,2m-1$, we have

$$\int_{I^{m-1}} \tilde{h}_k(t)\Sigma Q_j(t) \prod_{j'=1}^{m-1} t_{j'}^2 \cdot \Delta_{m-1}^4(t)dt = Vol(M_{2m-1})(2m-1)!(m-1)!\frac{(k+2)!}{(8k+4)(k-2)!}\delta_{jk}\,.$$

Forrester and Warnaar 2008

The importance of the Selberg integral

Used to prove outstanding conjectures in Random matrix theory and cases of the Macdonald conjectures

Central role in random matrix theory, Calogero-Sutherland quantum many-body systems, Knizhnik-Zamolodchikov equations, and multivariable orthogonal polynomial theory

The truncated moment problem

$$\mathcal{M}[0,1] \xrightarrow{\quad\Psi\quad} \mathbb{R}^k$$

$$\mu \longrightarrow \Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\Big)$$

Study of the geometry of $M_k := \Psi\big(\mathcal{M}([0,1])\big)$

P. L. Chebyshev
1821-1894

A. A. Markov
1856-1922

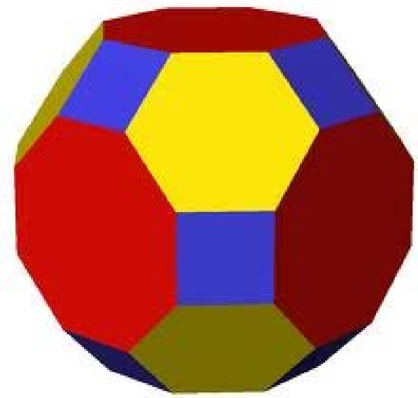
M. G. Krein
1907-1989

$$\mathcal{M}[0,1] \xrightarrow{\quad\Psi\quad} \mathbb{R}^k$$

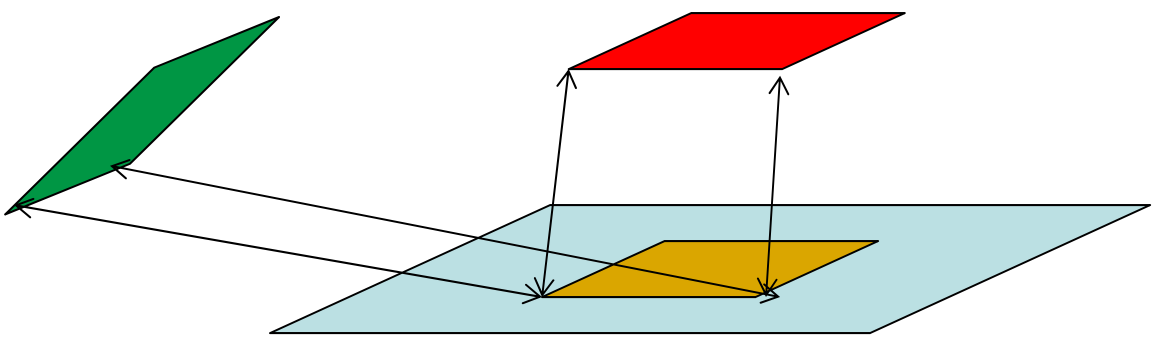
$$\mu \longrightarrow \left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \dots, \mathbb{E}_{X\sim\mu}[X^k]\right)$$

$$\boxed{M_k := \Psi\big(\mathcal{M}([0,1])\big)}$$

Origin of these new Selberg integral formulas and new RKHS



Compute $\mathrm{Vol}(M_k)$ using different (finite-dimensional) representations in $\mathcal{M}\big([0,1]\big)$



Infinite dim.

Finite dim.

$$\mathcal{M}[0,1] \xrightarrow{\quad\Psi\quad} \mathbb{R}^k$$

$$\mu \longrightarrow \left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\right)$$

$$\boxed{M_k := \Psi\big(\mathcal{M}([0,1])\big)}$$

Origin of these new Selberg integral formulas and new RKHS

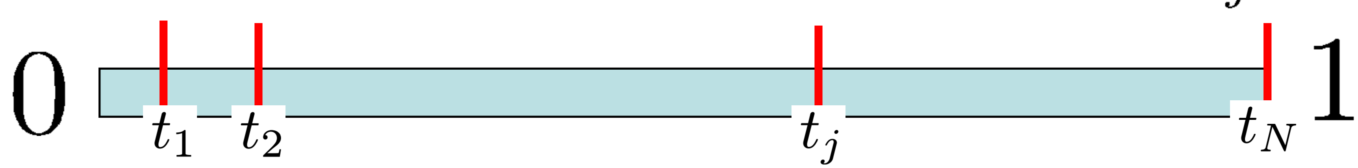
Compute $\mathrm{Vol}(M_k)$ using different (finite-dimensional) representations in $\mathcal{M}([0,1])$

$$0 \le t_1 < t_2 < \cdots < t_N \le 1$$

$$\lambda_1, \ldots, \lambda_N > 0, \; \sum_{j=1}^N \lambda_j = 1$$

$$\mu = \sum_{j=1}^N \lambda_j \delta_{t_j} \xrightarrow{\quad\Psi\quad} (q_1, \ldots, q_k)$$

$$q_i = \sum_{j=1}^N \lambda_j t_j^i$$



$$\mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j}$$

Index $i(\mu)$: Number of support points of μ

Counting interior points with weight 1 and boundary points with weight $\frac{1}{2}$

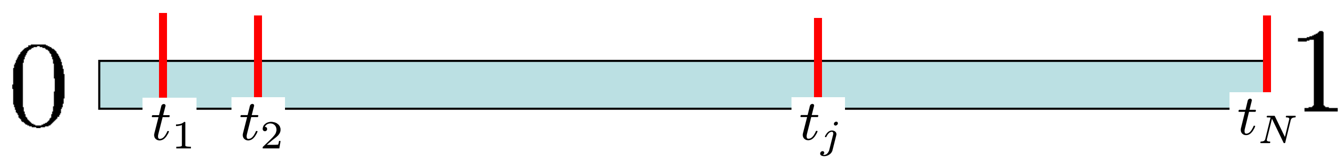
μ is called

- principal if $i(\mu) = \frac{k+1}{2}$
- canonical if $i(\mu) = \frac{k+2}{2}$
- upper if support points include 1
- lower if support points do not include 1

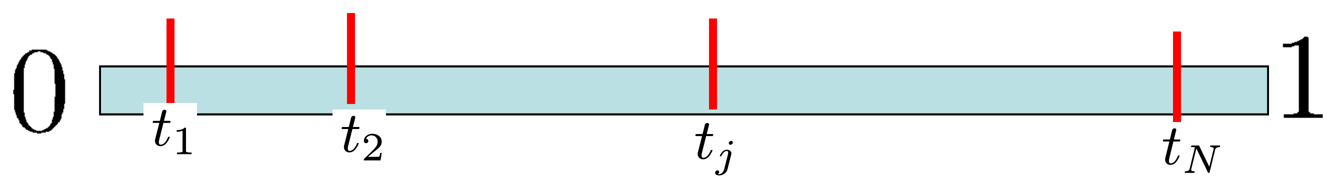
Theorem

Every point $q \in \mathrm{Int}(M_k)$ has a unique upper and lower principal representation.

Upper



Lower



$\mathrm{Vol}(M_{2m-1})$ using Upper Rep. $=$ $\mathrm{Vol}(M_{2m-1})$ using Lower Rep.

$$\frac{1}{(m-1)!} S_{m-1}(3,3,2) = \frac{1}{m!} S_m(1,1,2)$$

$\mathrm{Vol}(M_{2m})$ using Upper Rep. $=$ $\mathrm{Vol}(M_{2m})$ using Lower Rep.

$$S_m(1,3,2) = S_m(3,1,2)$$

Selberg Identities

$$S_n(\alpha,\beta,\gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}$$

$$S_n(\alpha,\beta,\gamma) := \int_{[0,1]^n} \prod_{j=1}^{n} t_j^{\alpha-1}(1-t_j)^{\beta-1}|\Delta(t)|^{2\gamma} dt\,.$$

$$\Delta(t) := \prod_{j<k} (t_k - t_j)$$

$$\mu = \sum_{j=1}^{N} \lambda_j \delta_{t_j}$$

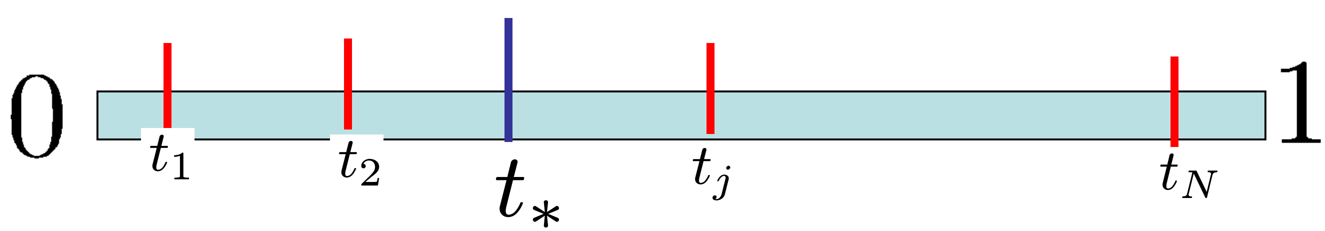
Index $i(\mu)$: Number of support points of μ

Counting interior points with weight 1 and boundary points with weight $\frac{1}{2}$

μ is called
- principal if $i(\mu) = \frac{k+1}{2}$
- canonical if $i(\mu) = \frac{k+2}{2}$
- upper if support points include 1
- lower if support points do not include 1

Theorem

For $t_* \in (0,1)$, every point $q \in \mathrm{Int}(M_k)$ has a unique canonical representation whose support contains t_*. When $t_* = 0$ or 1, there exists a unique principal representation whose support contains t_*.



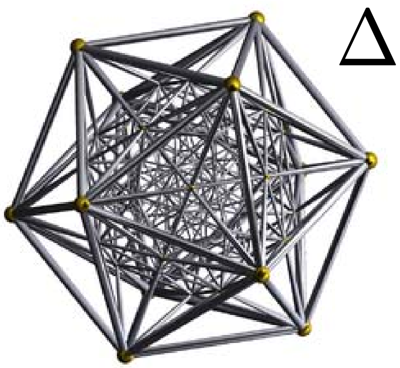
New Reproducing Kernel Hilbert Spaces and Selberg Integral formulas related to the Markov-Krein representations of moment spaces.

$$\mathcal{M}[0,1] \xrightarrow{\quad\Psi\quad} [0,1]^k$$

$$\mu \longrightarrow \left(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\right)$$

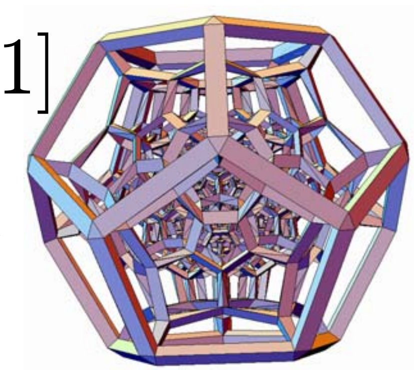
$$\int_{I^m} \Sigma t^{-1} \cdot \prod_{j=1}^m t_j^2(1-t_j)^2 \Delta_m^4(t)dt = \frac{S_m(5,1,2) - S_m(3,3,2)}{2}$$

$$\int_{I^m} \Sigma t^{-1} \cdot \prod_{j=1}^m t_j^2 \cdot \Delta_m^4(t)dt = \frac{m}{2} S_{m-1}(5,3,2)$$



$$\Delta_m(t) := \prod_{j<k} (t_k - t_j) \quad I := [0,1]$$

$$(\Sigma\phi)(t) := \sum_{j=1}^m \phi(t_j), \quad t \in I^m$$



$$S_n(\alpha,\beta,\gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}$$

$$e_j(t) := \sum_{i_1 < \cdots < i_j} t_{i_1} \cdots t_{i_j}$$

Π_0^n: n-th degree polynomials which vanish on the boundary of $[0,1]$

$M_n \subset \mathbb{R}^n$: set of $q = (q_1, \ldots, q_n) \in \mathbb{R}^n$ such that there exists a probability measure μ on $[0,1]$ with $\mathbb{E}_\mu[X^i] = q_i$ with $i \in \{1, \ldots, n\}$.

Theorem — Bi-orthogonal systems of Selberg Integral formulas

Consider the basis of Π_0^{2m-1} consisting of the associated Legendre polynomials $Q_j, j = 2, .., 2m-1$ of order 2 translated to the unit interval I. For $k = 2, .., 2m-1$ define

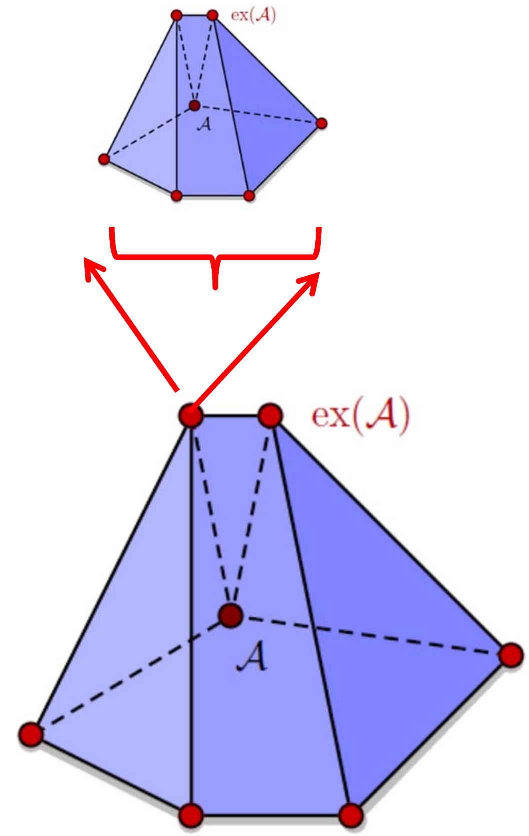
$$a_{jk} := \frac{(j+k+k^2)\Gamma(j+2)\Gamma(j)}{\Gamma(j+k+2)\Gamma(j-k+1)}, \quad k \le j \le 2m-1$$

$$\tilde{h}_k(t) := \sum_{j=k}^{2m-1} (-1)^{j+1} a_{jk} e_{2m-1-j}(t,t)\,.$$

Then for $j = k \bmod 2$, $j,k = 2, .., 2m-1$, we have

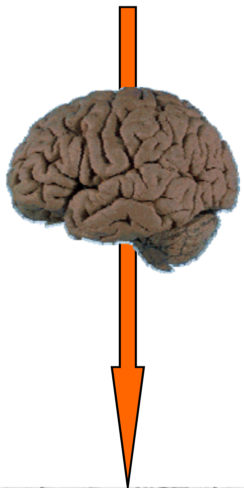
$$\int_{I^{m-1}} \tilde{h}_k(t) \Sigma Q_j(t) \prod_{j'=1}^{m-1} t_{j'}^2 \cdot \Delta_{m-1}^4(t) dt = Vol(M_{2m-1})(2m-1)!(m-1)! \frac{(k+2)!}{(8k+4)(k-2)!} \delta_{jk}\,.$$

Why develop this form of calculus? What else could we do?



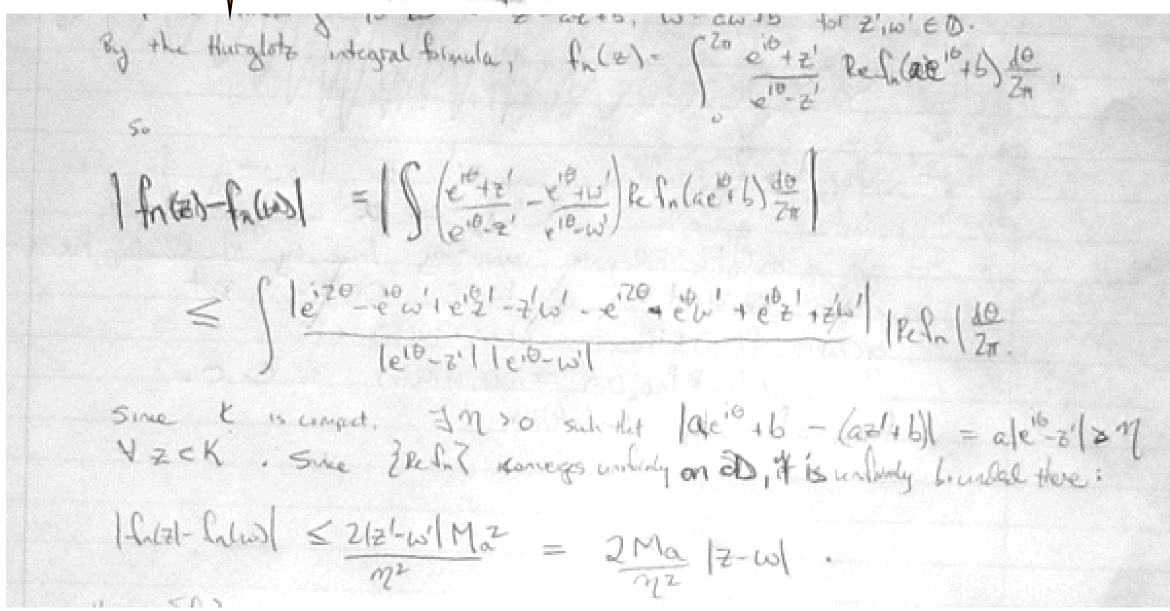
Solving PDEs: Two centuries ago

$$\Delta u = f$$



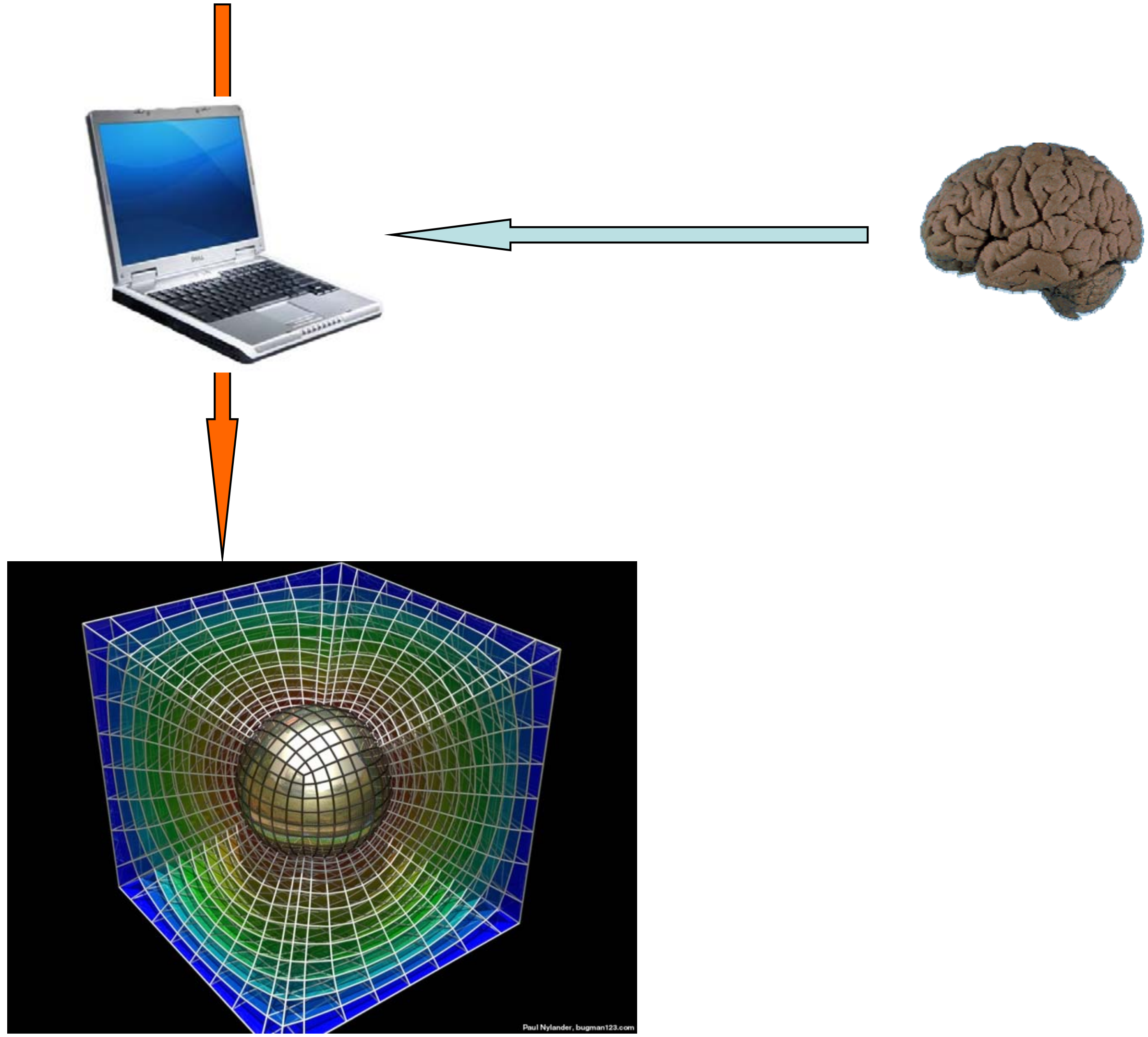
A. L. Cauchy
(1789-1857)

S. D. Poisson
(1781-1840)



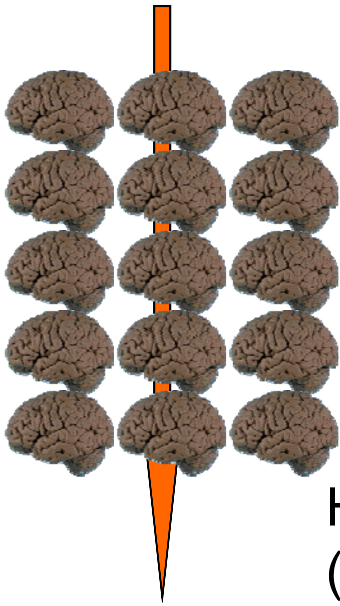
Solving PDEs: Now.

$$\Delta u = f$$

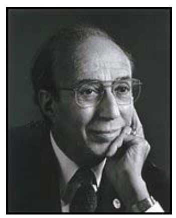


Paradigm shift

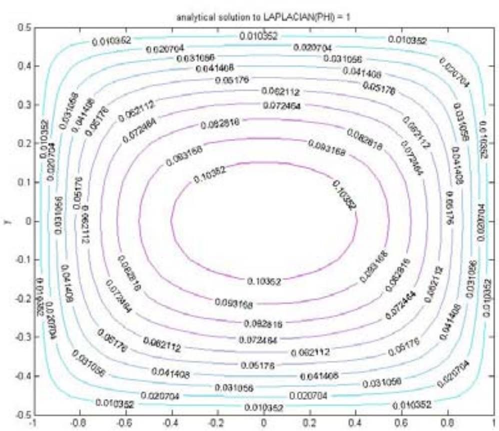
$$\Delta u = f$$



J. V. Neumann
(1903-1957)



H. Goldstine
(1913-2004)



Where are we at in finding statistical estimators?

$$X^2 = \sum \frac{(o-e)^2}{e}$$

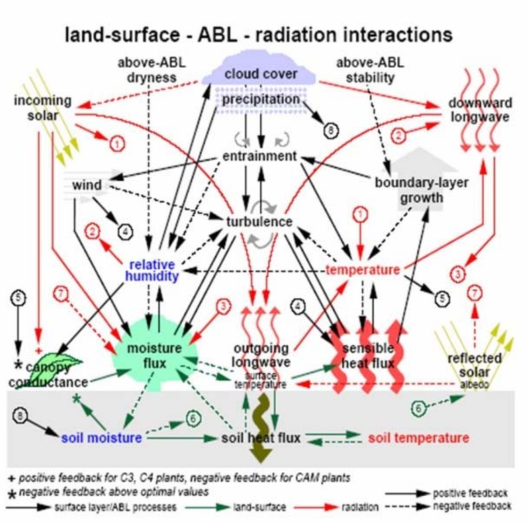
where

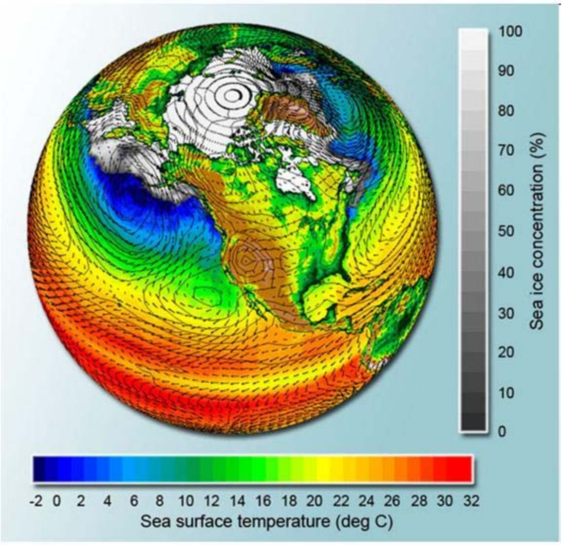
X^2 is Chi-squared,
$\sum$ stands for summation,
o is the observed values, a
e is the expected values.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2							
	0.99	**0.95**	**0.90**	**0.75**	**0.50**	**0.25**	**0.10**	**0.05**
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08

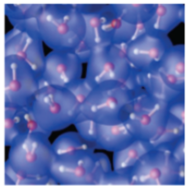
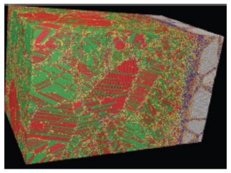
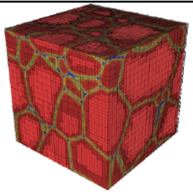
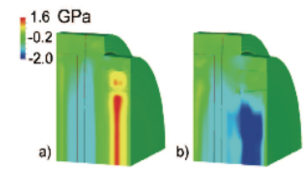
Find the best climate model given current information





Exascale Co-Design Center for Materials in Extreme Environments

ExMatEx

Ab-initio Methods	Molecular Dynamics	Phase-Field Modeling	Continuum Methods
Inter-atomic force model, equation of state,	**Defect and interface mobility, nucleation**	**Direct numerical simulation of multi-phase evolution**	**Multi-phase material response, experimental observables**
			
Code: Qbox/LATTE Motif: Particles and wavefunctions, plane wave DFT with nonlocal norm-conserving, ScaLAPACK, BLACS, and custom parallel 3D FFTs Prog. Model: MPI	Code: SPaSM/ddcMD Motif: Particles, domain decomposition, explicit time integration, neighbor and linked lists, dynamic load balancing, parity error recovery, and *in situ* visualization Prog. Model: MPI + Threads	Code: AMPE/GL Motif: Regular and adaptive grids, implicit time integration, real-space and spectral methods, complex order parameter (phase, crystal, species) Prog. Model: MPI	Code: VP-FFT/ALE3d Motif: Regular and irregular grids, implicit time integration, 3D FFTs, polycrystal and simgle crystal plasticity, Prog. Model: MPI

Where are we at in finding statistical estimators?

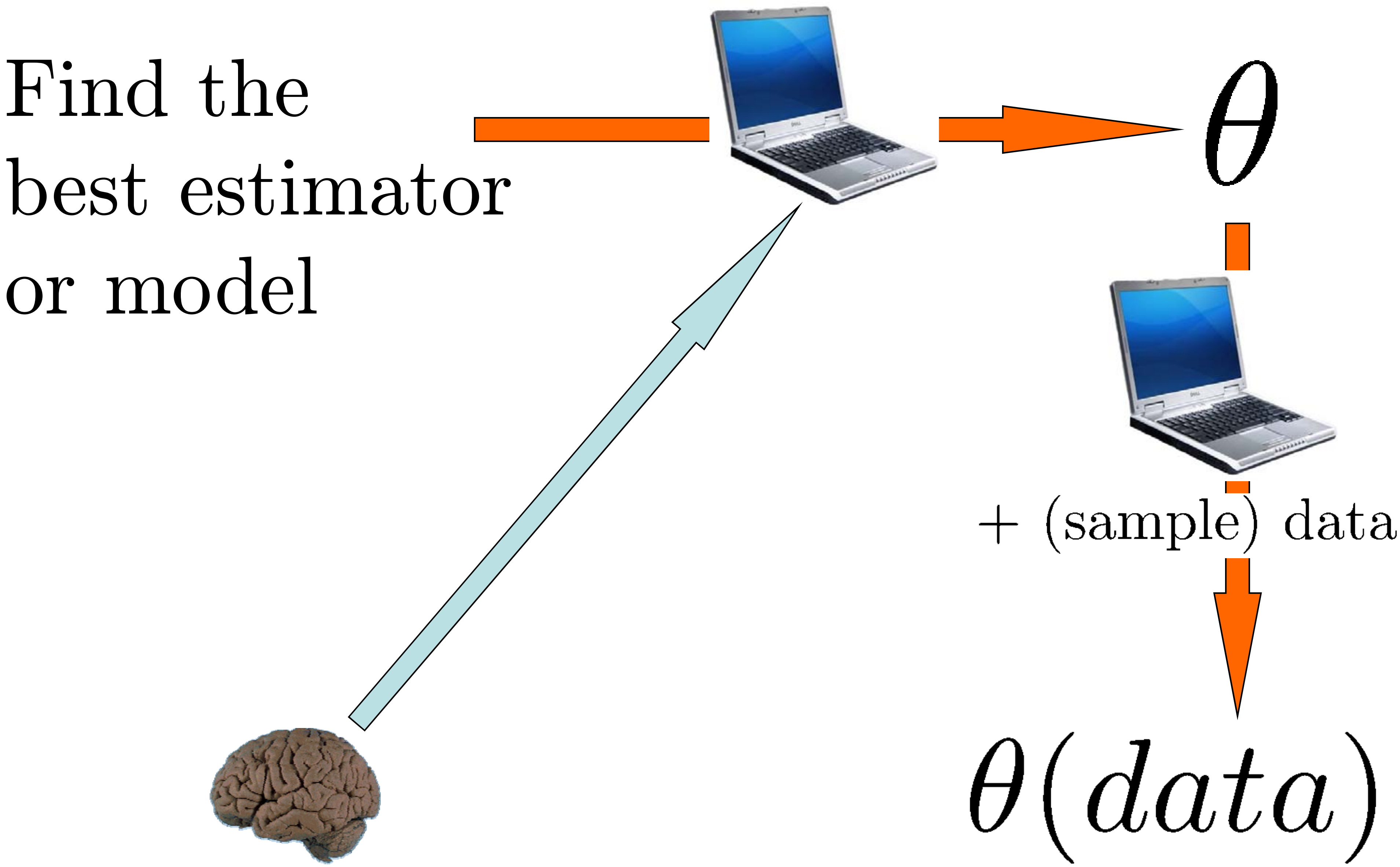
Find the
best estimator
or model

$$\theta$$

$+$ (sample) data

$$\theta(data)$$

Can we turn their design into a computation?



The UQ Problem with sample data

We want to estimate $\Phi(\mu^\dagger) = \mu^\dagger[X \geq a]$

$\mu^\dagger$: Unknown or partially known measure of probability on $\mathbb{R}$

You know $\mu^\dagger \in \mathcal{A}$

We observe $d = (d_1, \ldots, d_n) \in \mathbb{R}^n$

n i.i.d samples from $\mu^\dagger$

Your estimation: function of the data $\theta(d)$

Estimation error $\theta(d) - \Phi(\mu^\dagger)$

Statistical Error

$$\mathcal{E}(\theta, \mu^\dagger) = \mathbb{E}_{d \sim (\mu^\dagger)^n}\left[\left[\theta(d) - \Phi(\mu^\dagger)\right]^2\right]$$

Optimal bound on the statistical error

$$\max_{\mu \in \mathcal{A}} \mathcal{E}(\theta, \mu)$$

Optimal statistical estimators

$$\min_{\theta} \max_{\mu \in \mathcal{A}} \mathcal{E}(\theta, \mu)$$

Game theory and statistical decision theory

John Von Neumann

Abraham Wald

You

Estimator

θ

The universe

Measure of probability

μ

Loss/Statistical Error

$$\mathcal{E}(\theta, \mu)$$

Minimize

Maximize

Computer

Estimator

θ

The universe

Measure of probability

μ

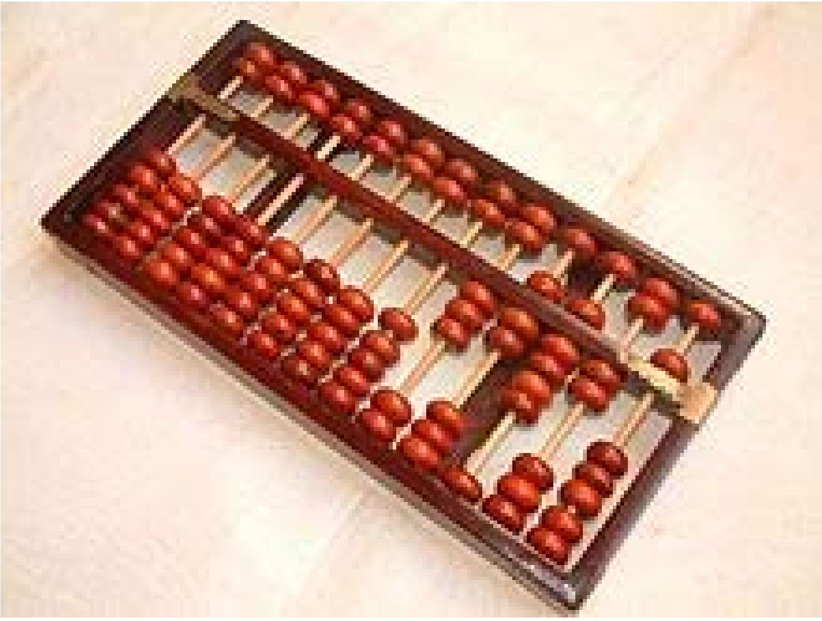
Loss/Statistical Error

$$\mathcal{E}(\theta, \mu)$$

Minimize

Maximize

The space of admissible scenarios along with the space of relevant information, assumptions, beliefs and models tend to be infinite dimensional, whereas calculus on a computer is necessarily discrete and finite



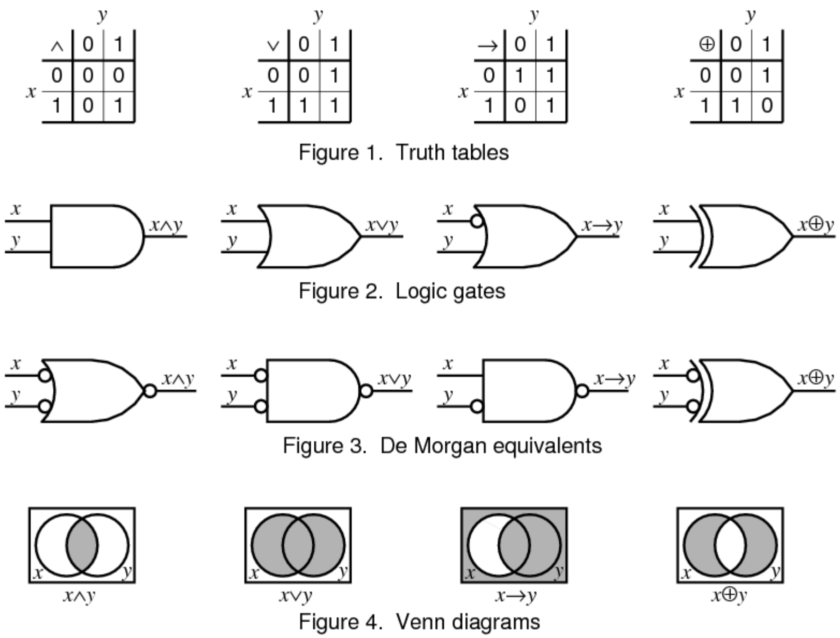


Figure 1. Truth tables

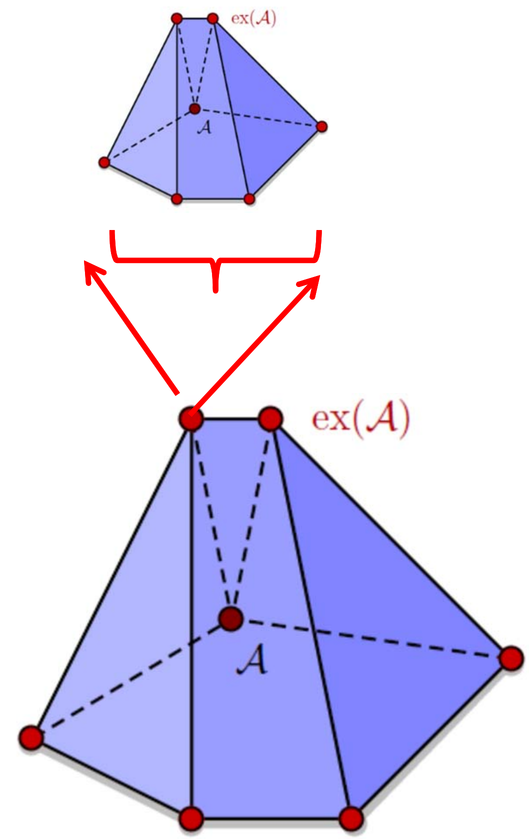
Figure 2. Logic gates

Figure 3. De Morgan equivalents

Figure 4. Venn diagrams

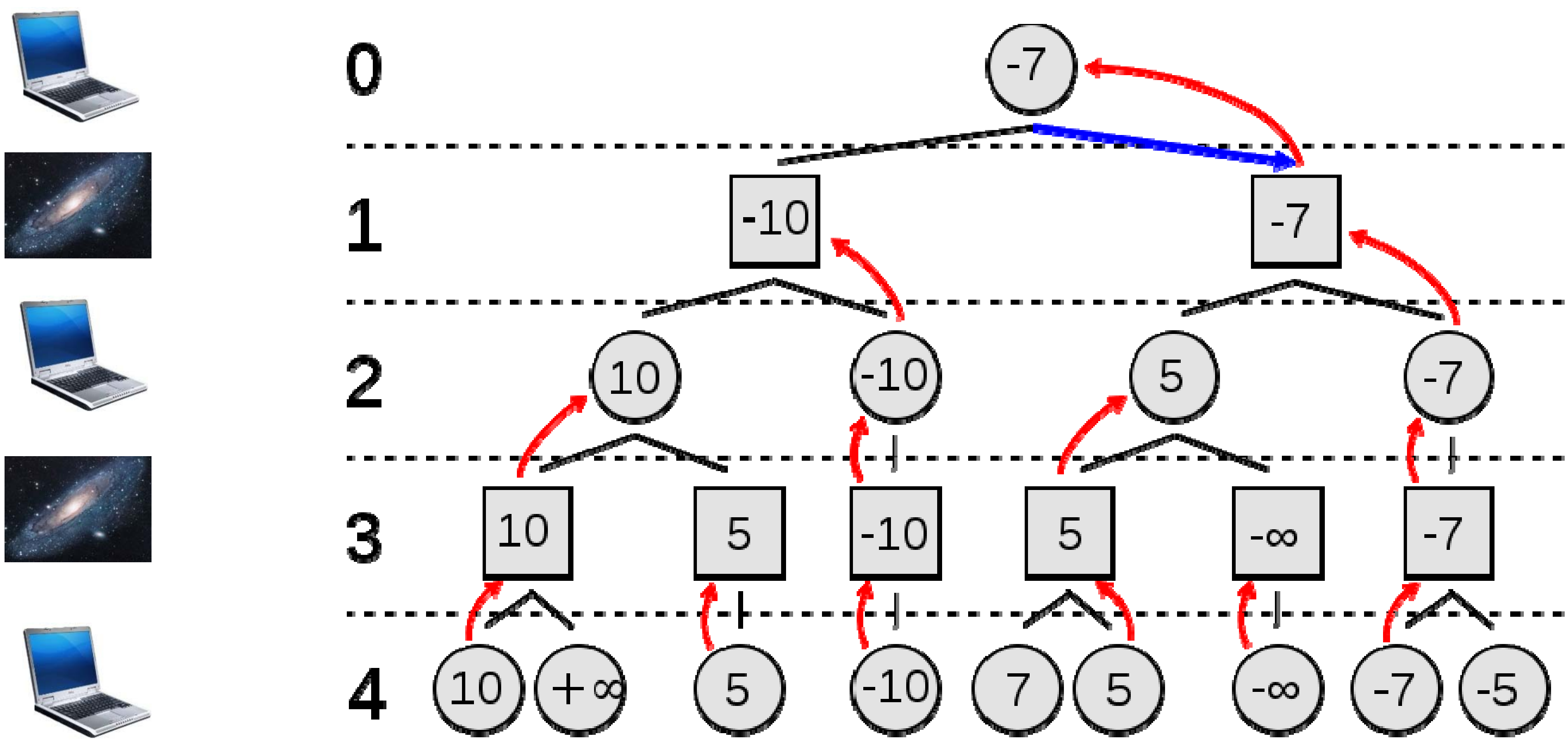
Arithmetic and Boolean logic

We need a form of calculus allowing us to manipulate infinite dimensional information structures



Min/Max Tree

Allows you to design optimal experimental campaigns
and turn the process of scientific discovery into a computation



Machine learning

Develop the best model of reality given available information

↓

Act based on That model

↓

Gather new information

↑ (back to: Develop the best model of reality given available information)

$p_0 \in \mathcal{M}(\Theta)$ Prior on Θ

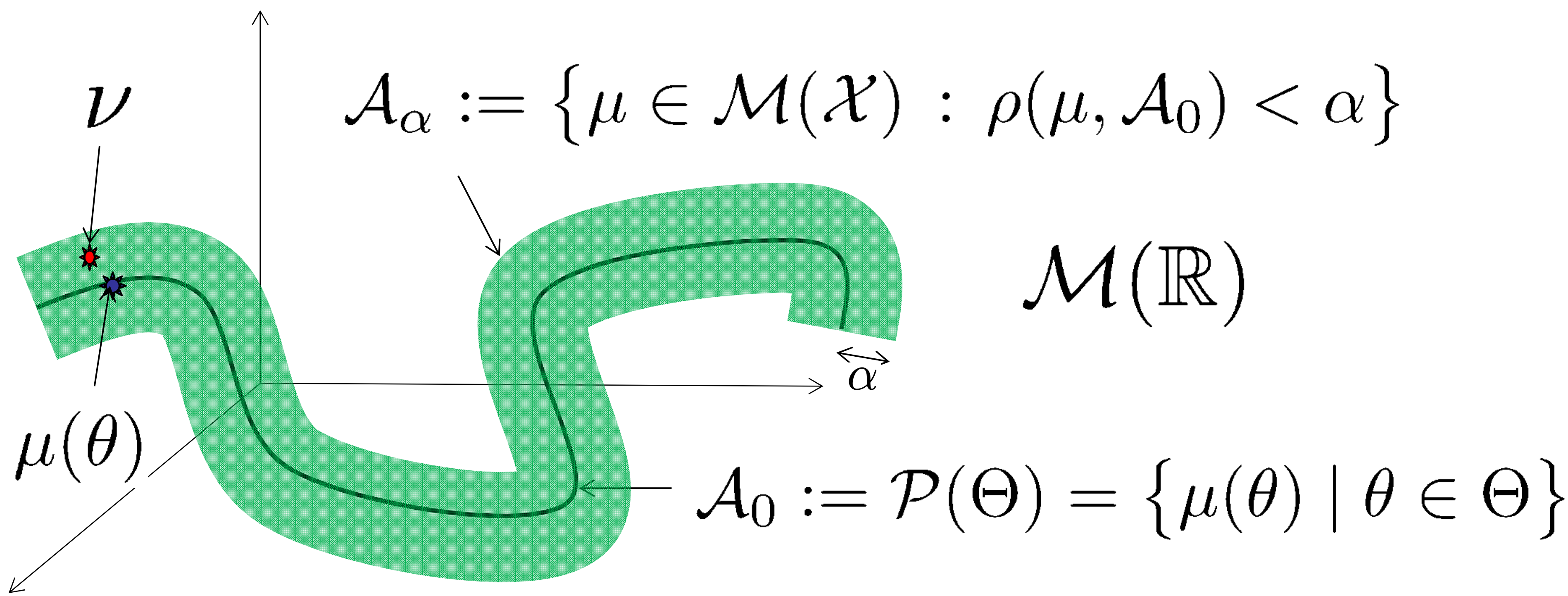
$\pi_0 := \mathcal{P}(p_0)$ Prior on $\mathcal{A}_0 \subset \mathcal{M}(\mathcal{X})$

Bayesian model

$\mu(\theta)$: Random element of $\mathcal{A}_0$ distributed according to π_0

Perturbed Bayesian model

ν: Random element of $\mathcal{A}_\alpha$ such that a.s. $\rho(\mu(\theta), \nu) \leq \alpha$



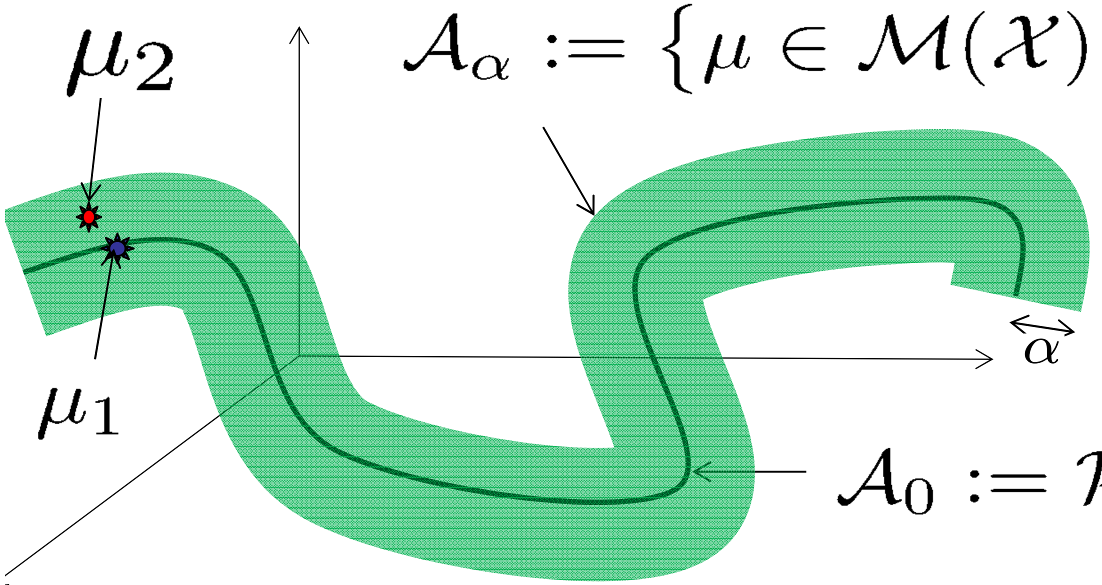
$$\mathcal{A}^* := \left\{(\mu_1,\mu_2) \in \mathcal{M}(\mathcal{X}) \times \mathcal{M}(\mathcal{X}) \middle| \mu_1 \in \mathcal{A}_0, \rho(\mu_2,\mu_1) < \alpha\right\},$$

$$P_0\mathcal{A}^* = \mathcal{A}_0 \qquad P_\alpha\mathcal{A}^* = \mathcal{A}_\alpha$$

$$\Pi_\alpha := \left\{\pi_\alpha \in \mathcal{M}(\mathcal{A}_\alpha) \middle| \text{for some } \pi \in \mathcal{M}(\mathcal{A}^*), P_0\pi = \pi_0 \text{ and } P_\alpha\pi = \pi_\alpha\right\}.$$

$$\sup_{\nu \in \mathcal{M}_\alpha} \mathbb{E}\left[\Phi(\nu)\middle| d \in B^n_\delta\right] := \sup_{\substack{\pi_\alpha \in \Pi_\alpha \\ \pi_\alpha[d \in B^n_\delta] > 0}} \mathbb{E}_{\mu \sim \pi_\alpha}\left[\Phi(\mu)\middle| d \in B^n_\delta\right]$$

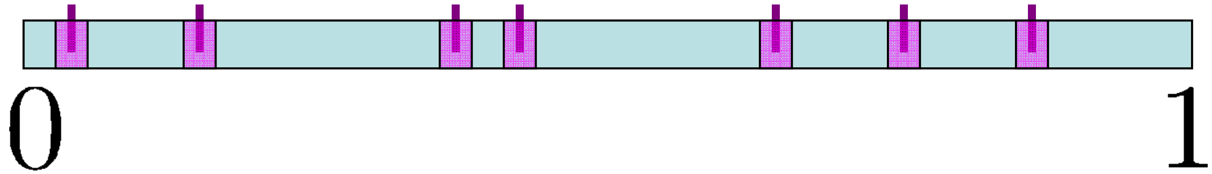
$$\mathcal{A}_\alpha := \{\mu \in \mathcal{M}(\mathcal{X}) \,:\, \rho(\mu, \mathcal{A}_0) < \alpha\}$$



$\mathcal{M}(\mathbb{R})$

$$\mathcal{A}_0 := \mathcal{P}(\Theta) = \{\mu(\theta) \mid \theta \in \Theta\}$$

Example



We want to estimate $\Phi(\mu^\dagger) = \mathbb{E}_{\mu^\dagger}[X]$

We observe $d \in B_\delta^n := \prod_{i=1}^n B_\delta(x_i)$

Π: Classes of priors on $\mathcal{M}([0,1])$ such that if $\pi \in \Pi$ and $\mu \sim \pi$ then

$$\Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\Big) \sim \mathbb{Q}$$

$\mathbb{Q}$ Uniform distribution on $\Psi\big(\mathcal{M}([0,1])\big)$

We observe $d \in B_\delta^1$

Π: Classes of priors on $\mathcal{M}([0,1])$ such that if $\pi \in \Pi$ and $\mu \sim \pi$ then

$$\Big(\mathbb{E}_{X\sim\mu}[X], \mathbb{E}_{X\sim\mu}[X^2], \ldots, \mathbb{E}_{X\sim\mu}[X^k]\Big) \sim \mathbb{Q}$$

Theorem $n = 1$, x_1 arbitrary, k arbitrary

$$1 - 4e\left(\frac{2k\delta}{e}\right)^{\frac{1}{2k+1}} \leq \sup_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\Big[\Phi(\mu)\Big| d \in B_\delta^n\Big] \leq 1$$

$$0 \leq \inf_{\pi\in\Pi} \mathbb{E}_{\mu\sim\pi}\Big[\Phi(\mu)\Big| d \in B_\delta^n\Big] \leq 4e\left(\frac{2k\delta}{e}\right)^{\frac{1}{2k+1}}$$

Stability of the method?

Numerically solving a PDE	⟷	CFL condition
Using Bayesian Inference under finite information	⟷	?

If we push Classical Bayesian Sensitivity Analysis the condition will depend on

- How much we already know
- Control on the probability of the data
- Resolution of the measurements

New form of reduction calculus

$$\mathcal{M}([0,1]) = \mathcal{A} \xrightarrow{\Psi(\mu) = \mathbb{E}_\mu[X]} \mathcal{Q} = [0,1]$$

$$\mathcal{M}(\mathcal{A}) \supset \Pi \xleftarrow{\Psi^{-1}} \mathfrak{Q}$$

$$= \{\mathbb{Q} \in \mathcal{M}(\mathcal{Q}) \mid \mathbb{E}_{\mathbb{Q}}[X] = m\}$$

Theorem

$$\sup_{\pi \in \Pi} \mathbb{E}_{\mu \sim \pi} \big[\mu[X \geq a]\big]$$

$$||$$

$$\sup_{\mathbb{Q} \in \mathcal{M}([0,1]) : \mathbb{E}_{\mathbb{Q}}[q] = m} \mathbb{E}_{q \sim \mathbb{Q}} \left[\sup_{\mu \in \mathcal{M}([0,1]) : \mathbb{E}_\mu[X] = q} \mu[X \geq a] \right]$$