Part II

Gaussian Measure in Hilbert Space and Applications in Numerical Analysis

F. M. LARKIN, Queen's University, Kingston, Ontario

Rocky Mountain Journal of Mathematics, 2(3), 379–422, 1972.

The numerical analyst is often called upon to estimate a function from a very limited knowledge of its properties (e.g. a finite number of ordinate values). This problem may be made well posed in a variety of ways, but an attractive approach is to regard the required function as a member of a linear space on which a probability measure is constructed, and then use established techniques of probability theory and statistics in order to infer properties of the function from the given information. This formulation agrees with established theory, for the problem of optimal linear approximation (using a Gaussian probability distribution), and also permits the estimation of nonlinear functionals, as well as extension to the case of "noisy" data.

Gaussian Measure in Hilbert Space and Applications in Numerical Analysis

F. M. LARKIN, Queen's University, Kingston, Ontario

Rocky Mountain Journal of Mathematics, 2(3), 379–422, 1972.

> *The numerical analyst is often called upon to estimate a function from a very limited knowledge of its properties (e.g. a finite number of ordinate values). This problem may be made well posed in a variety of ways, but an attractive approach is to regard the required function as a member of a linear space on which a probability measure is constructed, and then use established techniques of probability theory and statistics in order to infer properties of the function from the given information. This formulation agrees with established theory, for the problem of optimal linear approximation (using a Gaussian probability distribution), and also permits the estimation of nonlinear functionals, as well as extension to the case of "noisy" data.*

Fourth Job: Check Well-Defined, Existence and Uniqueness

Recap of Part I

Recall our set-up:

- Consider an unobserved state $x \in \mathcal{X}$ and a quantity of interest $Q(x)$.
- Given an information operator $A : \mathcal{X} \to \mathcal{A}$.
- Given a prior distribution $P_x \in \mathcal{P}_\mathcal{X}$.
- A Bayesian Probabilistic Numerical Method returns $B(a, P_x) = Q_{\#} P_{x|a}$.

But what is $P_{x|a}$?

Recall our set-up:

- Consider an unobserved state $x \in \mathcal{X}$ and a quantity of interest $Q(x)$.
- Given an information operator $A : \mathcal{X} \to \mathcal{A}$.
- Given a prior distribution $P_x \in \mathcal{P}_\mathcal{X}$.
- A Bayesian Probabilistic Numerical Method returns $B(a, P_x) = Q_\# P_{x|a}$.

But what is $P_{x|a}$?

Well-Defined?

The need to ensure that $P_{x|a}$ is well defined has, in part, motivated conjugate Gaussian process methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_\mathcal{X}$
- Often focused just on linear information operator $x \mapsto A(x)$

Outside of this context even existence of Bayesian probabilistic numerical methods is non-trivial when $\dim(\mathcal{X}) = \infty$:

$$p(x|a) = \frac{p(a|x)p(x)}{p(a)}$$

No Lebesgue measure $\implies$ work instead with Radon-Nikodym derivatives:

$$\frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x} = \frac{p(a|x)}{p(a)}$$

Let's define this object.

Well-Defined?

The need to ensure that $P_{x|a}$ is well defined has, in part, motivated conjugate Gaussian process methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_\mathcal{X}$
- Often focused just on linear information operator $x \mapsto A(x)$

Outside of this context even existence of Bayesian probabilistic numerical methods is non-trivial when $\dim(\mathcal{X}) = \infty$:

$$p(x|a) = \frac{p(a|x)p(x)}{p(a)}$$

No Lebesgue measure $\implies$ work instead with Radon-Nikodym derivatives:

$$\frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x} = \frac{p(a|x)}{p(a)}$$

Let's define this object.

Well-Defined?

The need to ensure that $P_{x|a}$ is well defined has, in part, motivated conjugate Gaussian process methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_{\mathcal{X}}$
- Often focused just on linear information operator $x \mapsto A(x)$

Outside of this context even existence of Bayesian probabilistic numerical methods is non-trivial when $\dim(\mathcal{X}) = \infty$:

$$p(x|a) = \frac{p(a|x)p(x)}{p(a)}$$

No Lebesgue measure $\implies$ work instead with Radon-Nikodym derivatives:

$$\frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x} = \frac{p(a|x)}{p(a)}$$

Let's define this object.

Well-Defined?

The need to ensure that $P_{x|a}$ is well defined has, in part, motivated conjugate Gaussian process methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_\mathcal{X}$
- Often focused just on linear information operator $x \mapsto A(x)$

Outside of this context even existence of Bayesian probabilistic numerical methods is non-trivial when $\dim(\mathcal{X}) = \infty$:

$$p(x|a) = \frac{p(a|x)p(x)}{p(a)}$$

No Lebesgue measure $\implies$ work instead with Radon-Nikodym derivatives:

$$\frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x} = \frac{p(a|x)}{p(a)}$$

Let's define this object.

Well-Defined?

The need to ensure that $P_{x|a}$ is well defined has, in part, motivated conjugate Gaussian process methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_{\mathcal{X}}$
- Often focused just on linear information operator $x \mapsto A(x)$

Outside of this context even existence of Bayesian probabilistic numerical methods is non-trivial when $\dim(\mathcal{X}) = \infty$:

$$p(x|a) = \frac{p(a|x)p(x)}{p(a)}$$

No Lebesgue measure $\implies$ work instead with Radon-Nikodym derivatives:

$$\frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x} = \frac{p(a|x)}{p(a)}$$

Let's define this object.

Well-Defined?

Standard tools of infinite dimensional statistics:

A probability measure ν on $(\mathcal{X}, \Sigma_{\mathcal{X}})$ is said to be *absolutely continuous* with respect to another probability measure ν' (written $\nu \ll \nu'$) on the same space if

$$\nu'(A) = 0 \implies \nu(A) = 0$$

Radon-Nikodym Theorem

If $\nu \ll \nu'$ then there exists a measurable function $\frac{\mathrm{d}\nu}{\mathrm{d}\nu'} : \mathcal{X} \to \mathbb{R}^+$ such that, for all $A \in \Sigma_{\mathcal{X}}$,

$$\nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\nu'}(x)\mathrm{d}\nu'(x)$$

For ν' the Lebesgue measure, we would usually call $\mathrm{d}\nu/\mathrm{d}\nu'$ the density of the random variable $X \sim \nu$.

Well-Defined?

Standard tools of infinite dimensional statistics:

A probability measure ν on $(\mathcal{X}, \Sigma_{\mathcal{X}})$ is said to be *absolutely continuous* with respect to another probability measure ν' (written $\nu \ll \nu'$) on the same space if

$$\nu'(A) = 0 \implies \nu(A) = 0$$

Radon-Nikodym Theorem

If $\nu \ll \nu'$ then there exists a measurable function $\frac{d\nu}{d\nu'} : \mathcal{X} \to \mathbb{R}^+$ such that, for all $A \in \Sigma_{\mathcal{X}}$,

$$\nu(A) = \int_A \frac{d\nu}{d\nu'}(x) d\nu'(x)$$

For ν' the Lebesgue measure, we would usually call $d\nu/d\nu'$ the density of the random variable $X \sim \nu$.

Well-Defined?

Standard tools of infinite dimensional statistics:

A probability measure ν on $(\mathcal{X}, \Sigma_{\mathcal{X}})$ is said to be *absolutely continuous* with respect to another probability measure ν' (written $\nu \ll \nu'$) on the same space if

$$\nu'(A) = 0 \implies \nu(A) = 0$$

Radon-Nikodym Theorem

If $\nu \ll \nu'$ then there exists a measurable function $\frac{\mathrm{d}\nu}{\mathrm{d}\nu'} : \mathcal{X} \to \mathbb{R}^+$ such that, for all $A \in \Sigma_{\mathcal{X}}$,

$$\nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\nu'}(x) \mathrm{d}\nu'(x)$$

For ν' the Lebesgue measure, we would usually call $\mathrm{d}\nu/\mathrm{d}\nu'$ the density of the random variable $X \sim \nu$.

Well-Defined?

Standard tools of infinite dimensional statistics:

A probability measure ν on $(\mathcal{X}, \Sigma_{\mathcal{X}})$ is said to be *absolutely continuous* with respect to another probability measure ν' (written $\nu \ll \nu'$) on the same space if

$$\nu'(A) = 0 \implies \nu(A) = 0$$

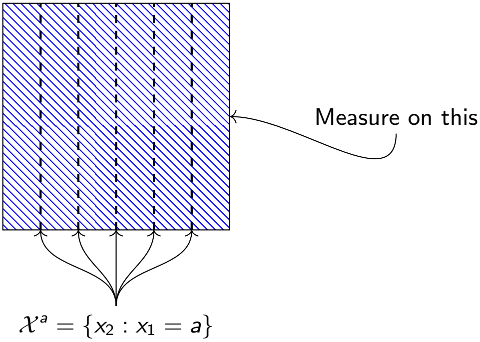
Radon-Nikodym Theorem

If $\nu \ll \nu'$ then there exists a measurable function $\frac{\mathrm{d}\nu}{\mathrm{d}\nu'} : \mathcal{X} \to \mathbb{R}^+$ such that, for all $A \in \Sigma_{\mathcal{X}}$,

$$\nu(A) = \int_A \frac{\mathrm{d}\nu}{\mathrm{d}\nu'}(x)\mathrm{d}\nu'(x)$$

For ν' the Lebesgue measure, we would usually call $\mathrm{d}\nu/\mathrm{d}\nu'$ the density of the random variable $X \sim \nu$.

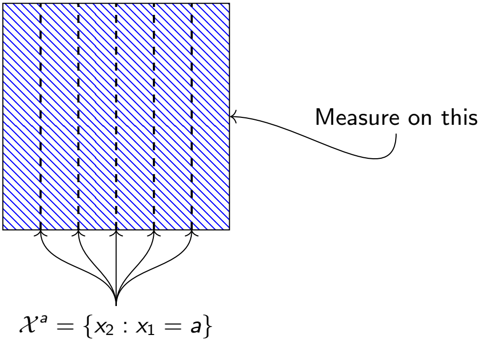
Consider, for now, $\dim(\mathcal{X}) = 2$ and condition a uniform measure P_x over $\mathcal{X} = [-1, 1]^2$ on the information that $x_1 = a$, for some fixed $a \in [-1, 1]$.



Informal answer: the conditional measure $P_{x|a}$ is "obviously" uniform over $[-1, 1]$

How to generalise this to infinite dimensional state spaces $\mathcal{X}$? It is not clear, because $\mathcal{X}^a$ is not easy to parametrise in general!

Consider, for now, $\dim(\mathcal{X}) = 2$ and condition a uniform measure P_x over $\mathcal{X} = [-1, 1]^2$ on the information that $x_1 = a$, for some fixed $a \in [-1, 1]$.

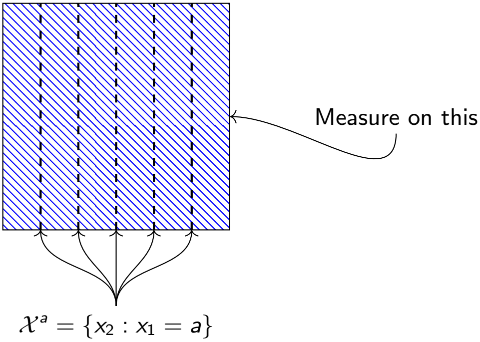


Informal answer: the conditional measure $P_{x|a}$ is "obviously" uniform over $[-1, 1]$

How to generalise this to infinite dimensional state spaces $\mathcal{X}$? It is not clear, because $\mathcal{X}^a$ is not easy to parametrise in general!

Conditioning on Null Sets

Consider, for now, $\dim(\mathcal{X}) = 2$ and condition a uniform measure P_x over $\mathcal{X} = [-1, 1]^2$ on the information that $x_1 = a$, for some fixed $a \in [-1, 1]$.



Informal answer: the conditional measure $P_{x|a}$ is "obviously" uniform over $[-1, 1]$

How to generalise this to infinite dimensional state spaces $\mathcal{X}$? It is not clear, because $\mathcal{X}^a$ is not easy to parametrise in general!

In our toy setting we want the support of the posterior to be

$$\mathcal{X}^a = \{x_2 : x_1 = a\}$$

However

$$P_{x|a}(\mathcal{X}^a) = 1$$

but...

$$P_x(\mathcal{X}^a) = 0$$

and this is the case for generic prior measures on $\mathcal{X}$ because $\mathcal{X}^a$ defines a submanifold of $\mathcal{X}$.

Thus $P_{x|a}$ will not be absolutely continuous wrt the prior P_x, and we cannot rely on the standard tools based on Radon-Nikodym derivatives in general.

In our toy setting we want the support of the posterior to be

$$\mathcal{X}^a = \{x_2 : x_1 = a\}$$

However

$$P_{x|a}(\mathcal{X}^a) = 1$$

but...

$$P_x(\mathcal{X}^a) = 0$$

and this is the case for generic prior measures on $\mathcal{X}$ because $\mathcal{X}^a$ defines a submanifold of $\mathcal{X}$.

Thus $P_{x|a}$ will not be absolutely continuous wrt the prior P_x, and we cannot rely on the standard tools based on Radon-Nikodym derivatives in general.

In our toy setting we want the support of the posterior to be

$$\mathcal{X}^a = \{x_2 : x_1 = a\}$$

However

$$P_{x|a}(\mathcal{X}^a) = 1$$

but...

$$P_x(\mathcal{X}^a) = 0$$

and this is the case for generic prior measures on $\mathcal{X}$ because $\mathcal{X}^a$ defines a submanifold of $\mathcal{X}$.

Thus $P_{x|a}$ will not be absolutely continuous wrt the prior P_x, and we cannot rely on the standard tools based on Radon-Nikodym derivatives in general.

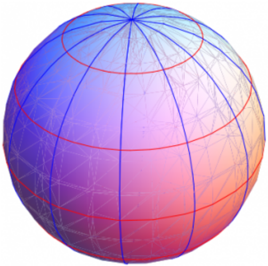
Conditioning on Null Sets

"a conditional probability relative to an isolated hypothesis whose probability equals zero is inadmissible"

—*Kolmogorov [1933]*

Conditioning on Null Sets

Borel-Kolmogorov paradox[1]:



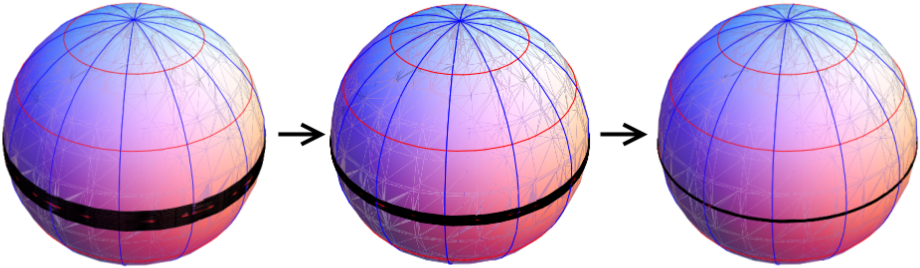
(latitude = red, longitude = blue)

To make progress it is required to introduce measure-theoretic detail.

[1] Figures from Greg Gandenberger's blog post

Conditioning on Null Sets

Borel-Kolmogorov paradox[1]:

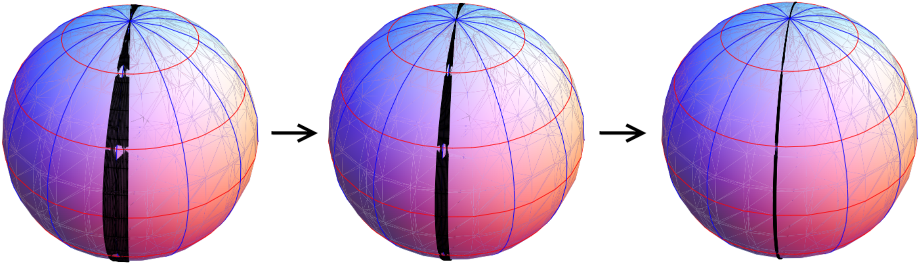


To make progress it is required to introduce measure-theoretic detail.

[1] Figures from Greg Gandenberger's blog post

Conditioning on Null Sets

Borel-Kolmogorov paradox[1]:



To make progress it is required to introduce measure-theoretic detail.

[1] Figures from Greg Gandenberger's blog post

Disintegration

High-level idea: Additional structure on $\mathcal{X}$, $\mathcal{A}$ and $A : \mathcal{X} \to \mathcal{A}$ is needed:

Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$ and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and A, Q be measurable.

Due to Dellacherie and Meyer [1978, p.78]:

For $P_x \in \mathcal{P}_{\mathcal{X}}$, a collection $\{P_{x|a}\}_{a \in \mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a disintegration of P_x with respect to the map $A : \mathcal{X} \to \mathcal{A}$ if:

1 (Concentration:) $P_{x|a}(\mathcal{X} \setminus \{x \in \mathcal{X} : A(x) = a\}) = 0$ for $A_{\#}P_x$-almost all $a \in \mathcal{A}$;

and for each measurable $f : \mathcal{X} \to [0, \infty)$ it holds that

2 (Measurability:) $a \mapsto P_{x|a}(f)$ is measurable;

3 (Conditioning:) $P_x(f) = \int P_{x|a}(f) A_{\#}P_x(\mathrm{d}a)$.

Disintegration

High-level idea: Additional structure on $\mathcal{X}$, $\mathcal{A}$ and $A : \mathcal{X} \to \mathcal{A}$ is needed:

Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$ and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and A, Q be measurable.

Due to Dellacherie and Meyer [1978, p.78]:

For $P_x \in \mathcal{P}_{\mathcal{X}}$, a collection $\{P_{x|a}\}_{a \in \mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a disintegration of P_x with respect to the map $A : \mathcal{X} \to \mathcal{A}$ if:

1. (Concentration:) $P_{x|a}(\mathcal{X} \setminus \{x \in \mathcal{X} : A(x) = a\}) = 0$ for $A_{\#}P_x$-almost all $a \in \mathcal{A}$;

and for each measurable $f : \mathcal{X} \to [0, \infty)$ it holds that

2. (Measurability:) $a \mapsto P_{x|a}(f)$ is measurable;
3. (Conditioning:) $P_x(f) = \int P_{x|a}(f) A_{\#}P_x(\mathrm{d}a)$.

Disintegration

High-level idea: Additional structure on $\mathcal{X}$, $\mathcal{A}$ and $A : \mathcal{X} \to \mathcal{A}$ is needed:

Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$ and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and A, Q be measurable.

Due to Dellacherie and Meyer [1978, p.78]:

For $P_x \in \mathcal{P}_{\mathcal{X}}$, a collection $\{P_{x|a}\}_{a \in \mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a disintegration of P_x with respect to the map $A : \mathcal{X} \to \mathcal{A}$ if:

1. (Concentration:) $P_{x|a}(\mathcal{X} \setminus \{x \in \mathcal{X} : A(x) = a\}) = 0$ for $A_{\#}P_x$-almost all $a \in \mathcal{A}$;

and for each measurable $f : \mathcal{X} \to [0, \infty)$ it holds that

2. (Measurability:) $a \mapsto P_{x|a}(f)$ is measurable;
3. (Conditioning:) $P_x(f) = \int P_{x|a}(f) A_{\#}P_x(\mathrm{d}a)$.

Disintegration

High-level idea: Additional structure on $\mathcal{X}$, $\mathcal{A}$ and $A : \mathcal{X} \to \mathcal{A}$ is needed:

Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$ and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and A, Q be measurable.

Due to Dellacherie and Meyer [1978, p.78]:

For $P_x \in \mathcal{P}_{\mathcal{X}}$, a collection $\{P_{x|a}\}_{a \in \mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a disintegration of P_x with respect to the map $A : \mathcal{X} \to \mathcal{A}$ if:

1. (Concentration:) $P_{x|a}(\mathcal{X} \setminus \{x \in \mathcal{X} : A(x) = a\}) = 0$ for $A_{\#}P_x$-almost all $a \in \mathcal{A}$;

and for each measurable $f : \mathcal{X} \to [0, \infty)$ it holds that

2. (Measurability:) $a \mapsto P_{x|a}(f)$ is measurable;
3. (Conditioning:) $P_x(f) = \int P_{x|a}(f) A_{\#}P_x(\mathrm{d}a)$.

High-level idea: Additional structure on $\mathcal{X}$, $\mathcal{A}$ and $A : \mathcal{X} \to \mathcal{A}$ is needed:

Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$ and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and A, Q be measurable.

Due to Dellacherie and Meyer [1978, p.78]:

For $P_x \in \mathcal{P}_{\mathcal{X}}$, a collection $\{P_{x|a}\}_{a \in \mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a disintegration of P_x with respect to the map $A : \mathcal{X} \to \mathcal{A}$ if:

1. (Concentration:) $P_{x|a}(\mathcal{X} \setminus \{x \in \mathcal{X} : A(x) = a\}) = 0$ for $A_\# P_x$-almost all $a \in \mathcal{A}$;

and for each measurable $f : \mathcal{X} \to [0, \infty)$ it holds that

2. (Measurability:) $a \mapsto P_{x|a}(f)$ is measurable;
3. (Conditioning:) $P_x(f) = \int P_{x|a}(f) A_\# P_x(\mathrm{d}a)$.

Disintegration

High-level idea: Additional structure on $\mathcal{X}$, $\mathcal{A}$ and $A : \mathcal{X} \to \mathcal{A}$ is needed:

Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$ and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and A, Q be measurable.

Due to Dellacherie and Meyer [1978, p.78]:

For $P_x \in \mathcal{P}_{\mathcal{X}}$, a collection $\{P_{x|a}\}_{a \in \mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a disintegration of P_x with respect to the map $A : \mathcal{X} \to \mathcal{A}$ if:

1 (Concentration:) $P_{x|a}(\mathcal{X} \setminus \{x \in \mathcal{X} : A(x) = a\}) = 0$ for $A_{\#}P_x$-almost all $a \in \mathcal{A}$;

and for each measurable $f : \mathcal{X} \to [0, \infty)$ it holds that

2 (Measurability:) $a \mapsto P_{x|a}(f)$ is measurable;

3 (Conditioning:) $P_x(f) = \int P_{x|a}(f) A_{\#}P_x(\mathrm{d}a)$.

Existence and Uniqueness

Disintegration Theorem; statement from Thm. 1 of Chang and Pollard [1997]:

- Let $\mathcal{X}$ be a metric space, $\Sigma_{\mathcal{X}}$ be the Borel σ-algebra.
- Let $P_x \in \mathcal{P}_{\mathcal{X}}$ be Radon.
- Let $\Sigma_{\mathcal{A}}$ be a countably generated σ-algebra that contains singletons $\{a\}$ for $a \in \mathcal{A}$.

Then there exists an (essentially) unique disintegration $\{P_{x|a}\}_{a \in \mathcal{A}}$ of P_x with respect to A.

Let $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be a measurable spaces and Q be measurable.

Then Bayesian probabilistic numerical methods $B(P_x, a) = Q_{\#}P_{x|a}$ are well-defined under quite general conditions.

In particular, $Q_{\#}P_{x|a}$ exists and is unique for $A_{\#}P_x$ almost all $a \in \mathcal{A}$.

Disintegration Theorem; statement from Thm. 1 of Chang and Pollard [1997]:

- Let $\mathcal{X}$ be a metric space, $\Sigma_{\mathcal{X}}$ be the Borel σ-algebra.
- Let $P_x \in \mathcal{P}_{\mathcal{X}}$ be Radon.
- Let $\Sigma_{\mathcal{A}}$ be a countably generated σ-algebra that contains singletons $\{a\}$ for $a \in \mathcal{A}$.

Then there exists an (essentially) unique disintegration $\{P_{x|a}\}_{a \in \mathcal{A}}$ of P_x with respect to A.

Let $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be a measurable spaces and Q be measurable.

Then Bayesian probabilistic numerical methods $B(P_x, a) = Q_{\#}P_{x|a}$ are well-defined under quite general conditions.

In particular, $Q_{\#}P_{x|a}$ exists and is unique for $A_{\#}P_x$ almost all $a \in \mathcal{A}$.

Existence and Uniqueness

Disintegration Theorem; statement from Thm. 1 of Chang and Pollard [1997]:

- Let $\mathcal{X}$ be a metric space, $\Sigma_{\mathcal{X}}$ be the Borel σ-algebra.
- Let $P_x \in \mathcal{P}_{\mathcal{X}}$ be Radon.
- Let $\Sigma_{\mathcal{A}}$ be a countably generated σ-algebra that contains singletons $\{a\}$ for $a \in \mathcal{A}$.

Then there exists an (essentially) unique disintegration $\{P_{x|a}\}_{a \in \mathcal{A}}$ of P_x with respect to A.

Let $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be a measurable spaces and Q be measurable.

Then Bayesian probabilistic numerical methods $B(P_x, a) = Q_\# P_{x|a}$ are well-defined under quite general conditions.

In particular, $Q_\# P_{x|a}$ exists and is unique for $A_\# P_x$ almost all $a \in \mathcal{A}$.

Fifth Job: Algorithms to Access $P_{x|a}$

Overview

The aim of this section is to develop an algorithm to approximate $P_{x|a}$ and hence $B(a, P_x) = Q_\# P_{x|a}$.

This will be achieved by designing a *sampler* for $P_{x|a}$.

Sampling $P_{x|a}$ Challenge

Non-Conjugate Challenge

The aim of this section is to develop an algorithm to approximate $P_{x|a}$ and hence $B(a, P_x) = Q_\# P_{x|a}$.

This will be achieved by designing a *sampler* for $P_{x|a}$.

Sampling $P_{x|a}$ Challenge

Non-Conjugate Challenge

Overview

The aim of this section is to develop an algorithm to approximate $P_{x|a}$ and hence $B(a, P_x) = Q_{\#} P_{x|a}$.

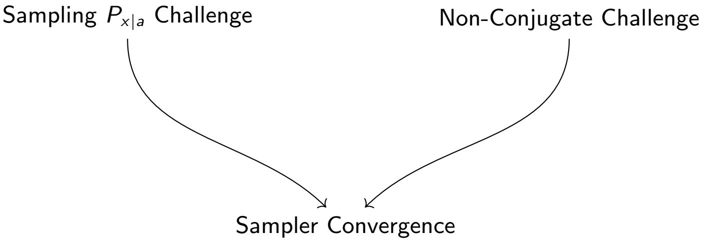
This will be achieved by designing a *sampler* for $P_{x|a}$.

Sampling $P_{x|a}$ Challenge

Non-Conjugate Challenge

The aim of this section is to develop an algorithm to approximate $P_{x|a}$ and hence $B(a, P_x) = Q_{\#} P_{x|a}$.

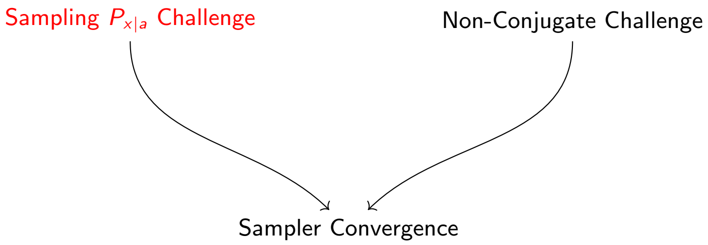
This will be achieved by designing a *sampler* for $P_{x|a}$.



Overview

The aim of this section is to develop an algorithm to approximate $P_{x|a}$ and hence $B(a, P_x) = Q_{\#} P_{x|a}$.

This will be achieved by designing a *sampler* for $P_{x|a}$.



Numerical Disintegration

Recall

$$\left.\begin{array}{c}\mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\} \\ P_x(\mathcal{X}^a) = 0\end{array}\right\} \implies \nexists \frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x}$$

Thus, standard techniques from infinite-dimensional statistics cannot be directly applied.

Our approach is to force the problem into the standard context, by approximating $P_{x|a}$ with a relaxed measure $P_{x|a}^{\delta}$ for which a Radon-Nikodym derivative is defined:

$$\frac{\mathrm{d}P_{x|a}^{\delta}}{\mathrm{d}P_x} \propto \phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

$\phi : \mathbb{R}^+ \to \mathbb{R}^+$ a *relaxation function* chosen so that:

- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

Idea is that this approaches $P_{x|a}$ as $\delta \downarrow 0$ (to be formalised).

Note that a norm structure has now been assumed on $\mathcal{A}$ (e.g. $\mathcal{A} = \mathbb{R}^n$).

Numerical Disintegration

Recall

$$\left.\begin{array}{c}\mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\} \\ P_x(\mathcal{X}^a) = 0\end{array}\right\} \implies \nexists \frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x}$$

Thus, standard techniques from infinite-dimensional statistics cannot be directly applied.

Our approach is to force the problem into the standard context, by approximating $P_{x|a}$ with a relaxed measure $P^{\delta}_{x|a}$ for which a Radon-Nikodym derivative is defined:

$$\frac{\mathrm{d}P^{\delta}_{x|a}}{\mathrm{d}P_x} \propto \phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

$\phi : \mathbb{R}^+ \to \mathbb{R}^+$ a *relaxation function* chosen so that:

- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

Idea is that this approaches $P_{x|a}$ as $\delta \downarrow 0$ (to be formalised).

Note that a norm structure has now been assumed on $\mathcal{A}$ (e.g. $\mathcal{A} = \mathbb{R}^n$).

Recall

$$\left.\begin{array}{c} \mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\} \\ P_x(\mathcal{X}^a) = 0 \end{array}\right\} \implies \nexists \, \frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x}$$

Thus, standard techniques from infinite-dimensional statistics cannot be directly applied.

Our approach is to force the problem into the standard context, by approximating $P_{x|a}$ with a relaxed measure $P_{x|a}^{\delta}$ for which a Radon-Nikodym derivative is defined:

$$\frac{\mathrm{d}P_{x|a}^{\delta}}{\mathrm{d}P_x} \propto \phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

$\phi : \mathbb{R}^+ \to \mathbb{R}^+$ a *relaxation function* chosen so that:

- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

Idea is that this approaches $P_{x|a}$ as $\delta \downarrow 0$ (to be formalised).

Note that a norm structure has now been assumed on $\mathcal{A}$ (e.g. $\mathcal{A} = \mathbb{R}^n$).

Recall

$$\left.\begin{array}{c}\mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\} \\ P_x(\mathcal{X}^a) = 0\end{array}\right\} \implies \nexists \frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x}$$

Thus, standard techniques from infinite-dimensional statistics cannot be directly applied.

Our approach is to force the problem into the standard context, by approximating $P_{x|a}$ with a relaxed measure $P_{x|a}^{\delta}$ for which a Radon-Nikodym derivative is defined:

$$\frac{\mathrm{d}P_{x|a}^{\delta}}{\mathrm{d}P_x} \propto \phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

$\phi : \mathbb{R}^+ \to \mathbb{R}^+$ a *relaxation function* chosen so that:

- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

Idea is that this approaches $P_{x|a}$ as $\delta \downarrow 0$ (to be formalised).

Note that a norm structure has now been assumed on $\mathcal{A}$ (e.g. $\mathcal{A} = \mathbb{R}^n$).

Recall

$$\left.\begin{array}{c}\mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\} \\ P_x(\mathcal{X}^a) = 0\end{array}\right\} \implies \nexists \, \frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x}$$

Thus, standard techniques from infinite-dimensional statistics cannot be directly applied.

Our approach is to force the problem into the standard context, by approximating $P_{x|a}$ with a relaxed measure $P^{\delta}_{x|a}$ for which a Radon-Nikodym derivative is defined:

$$\frac{\mathrm{d}P^{\delta}_{x|a}}{\mathrm{d}P_x} \propto \phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

$\phi : \mathbb{R}^+ \to \mathbb{R}^+$ a *relaxation function* chosen so that:

- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

Idea is that this approaches $P_{x|a}$ as $\delta \downarrow 0$ (to be formalised).

Note that a norm structure has now been assumed on $\mathcal{A}$ (e.g. $\mathcal{A} = \mathbb{R}^n$).

Recall

$$\left.\begin{array}{c} \mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\} \\ P_x(\mathcal{X}^a) = 0 \end{array}\right\} \implies \nexists \frac{\mathrm{d}P_{x|a}}{\mathrm{d}P_x}$$

Thus, standard techniques from infinite-dimensional statistics cannot be directly applied.

Our approach is to force the problem into the standard context, by approximating $P_{x|a}$ with a relaxed measure $P^{\delta}_{x|a}$ for which a Radon-Nikodym derivative is defined:

$$\frac{\mathrm{d}P^{\delta}_{x|a}}{\mathrm{d}P_x} \propto \phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

$\phi : \mathbb{R}^+ \to \mathbb{R}^+$ a *relaxation function* chosen so that:

- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

Idea is that this approaches $P_{x|a}$ as $\delta \downarrow 0$ (to be formalised).

Note that a norm structure has now been assumed on $\mathcal{A}$ (e.g. $\mathcal{A} = \mathbb{R}^n$).

Example Relaxation Functions

$\phi(r) = \mathbb{I}(r < 1)$

- $P^{\delta}_{x|a}$ is the conditional distribution $X|A(X) \in B_\delta(a)$
- Equivalent to assuming uniform noise over $B_\delta(a)$ on the information $A(X)$.
- Equivalent to rare event simulation: $A_\# P_x(B_\delta(a)) \to 0$ as $\delta \to 0$.
- Equivalent to approximate Bayesian computation (ABC) rejection algorithm.

$\phi(r) = \exp(-r^2)$

- Equivalent to assuming IID Gaussian noise $N(0, 2\delta)$ on the information $A(X)$.
- Gives access to nontrivial gradient information in the samplers.

Example Relaxation Functions

$\phi(r) = \mathbb{I}(r < 1)$

- $P^{\delta}_{x|a}$ is the conditional distribution $X|A(X) \in B_\delta(a)$
- Equivalent to assuming uniform noise over $B_\delta(a)$ on the information $A(X)$.
- Equivalent to rare event simulation: $A_\# P_x(B_\delta(a)) \to 0$ as $\delta \to 0$.
- Equivalent to approximate Bayesian computation (ABC) rejection algorithm.

$\phi(r) = \exp(-r^2)$

- Equivalent to assuming IID Gaussian noise $N(0, 2\delta)$ on the information $A(X)$.
- Gives access to nontrivial gradient information in the samplers.

Example Relaxation Functions

$\phi(r) = \mathbb{I}(r < 1)$

- $P^{\delta}_{x|a}$ is the conditional distribution $X | A(X) \in B_\delta(a)$
- Equivalent to assuming uniform noise over $B_\delta(a)$ on the information $A(X)$.
- Equivalent to rare event simulation: $A_{\#} P_x(B_\delta(a)) \to 0$ as $\delta \to 0$.
- Equivalent to approximate Bayesian computation (ABC) rejection algorithm.

$\phi(r) = \exp(-r^2)$

- Equivalent to assuming IID Gaussian noise $N(0, 2\delta)$ on the information $A(X)$.
- Gives access to nontrivial gradient information in the samplers.

Example Relaxation Functions

$\phi(r) = \mathbb{I}(r < 1)$

- $P^{\delta}_{x|a}$ is the conditional distribution $X|A(X) \in B_\delta(a)$
- Equivalent to assuming uniform noise over $B_\delta(a)$ on the information $A(X)$.
- Equivalent to rare event simulation: $A_\# P_x(B_\delta(a)) \to 0$ as $\delta \to 0$.
- Equivalent to approximate Bayesian computation (ABC) rejection algorithm.

$\phi(r) = \exp(-r^2)$

- Equivalent to assuming IID Gaussian noise $N(0, 2\delta)$ on the information $A(X)$.
- Gives access to nontrivial gradient information in the samplers.

Example Relaxation Functions

$\phi(r) = \mathbb{I}(r < 1)$

- $P^{\delta}_{x|a}$ is the conditional distribution $X | A(X) \in B_{\delta}(a)$
- Equivalent to assuming uniform noise over $B_{\delta}(a)$ on the information $A(X)$.
- Equivalent to rare event simulation: $A_{\#} P_x(B_{\delta}(a)) \to 0$ as $\delta \to 0$.
- Equivalent to approximate Bayesian computation (ABC) rejection algorithm.

$\phi(r) = \exp(-r^2)$

- Equivalent to assuming IID Gaussian noise $N(0, 2\delta)$ on the information $A(X)$.
- Gives access to nontrivial gradient information in the samplers.

Example Relaxation Functions

$\phi(r) = \mathbb{I}(r < 1)$

- $P^{\delta}_{x|a}$ is the conditional distribution $X|A(X) \in B_\delta(a)$
- Equivalent to assuming uniform noise over $B_\delta(a)$ on the information $A(X)$.
- Equivalent to rare event simulation: $A_\# P_x(B_\delta(a)) \to 0$ as $\delta \to 0$.
- Equivalent to approximate Bayesian computation (ABC) rejection algorithm.

$\phi(r) = \exp(-r^2)$

- Equivalent to assuming IID Gaussian noise $N(0, 2\delta)$ on the information $A(X)$.
- Gives access to nontrivial gradient information in the samplers.

Key Idea: Tempering

Consider a standard Bayesian inference problem for unknown θ with data y.

- Prior $p(\theta)$, which is easy to sample.
- Posterior $p(\theta|y) \propto p(y|\theta)p(\theta)$, which is hard to sample.

Define intermediate distributions by tempering

$$p_t(\theta|y) \propto p(y|\theta)^t p(\theta)$$

The idea is to interpolate between the easy and the hard problem.

Key Idea: Tempering

Consider a standard Bayesian inference problem for unknown θ with data y.

- Prior $p(\theta)$, which is easy to sample.
- Posterior $p(\theta|y) \propto p(y|\theta)p(\theta)$, which is hard to sample.

Define intermediate distributions by tempering

$$p_t(\theta|y) \propto p(y|\theta)^t p(\theta)$$

The idea is to interpolate between the easy and the hard problem.

Key Idea: Tempering

Consider a standard Bayesian inference problem for unknown θ with data y.

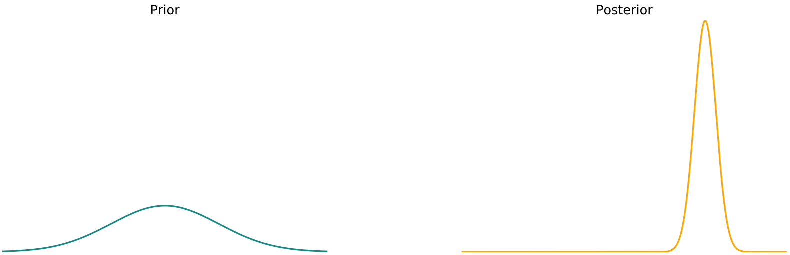
- Prior $p(\theta)$, which is easy to sample.
- Posterior $p(\theta|y) \propto p(y|\theta)p(\theta)$, which is hard to sample.

Define intermediate distributions by tempering

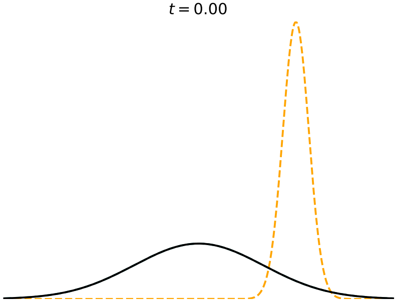
$$p_t(\theta|y) \propto p(y|\theta)^t p(\theta)$$

The idea is to interpolate between the easy and the hard problem.

Key Idea: Tempering



Key Idea: Tempering



Tempering for Sampling $P^{\delta}_{x|a}$

To sample $P^{\delta}_{x|a}$ we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\infty = \delta_0 > \delta_1 > \ldots > \delta_N = \delta$ and consider

$$P_x = P^{\delta_0}_{x|a},\ P^{\delta_1}_{x|a},\ \ldots,\ P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$$

- $P_x = P^{\delta_0}_{x|a}$ is the prior distribution (often easy to sample).
- $P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$ is the target distribution.
- Intermediate distributions define a "ladder" which smoothly interpolates from prior to target.

For P_x a Gaussian prior, efficient Monte Carlo methods are available based on pre-conditioned Crank Nicholson and its extensions [Cotter et al., 2013]. Not going to discuss further - too much detail - but remember this point for later!

Tempering for Sampling $P^{\delta}_{x|a}$

To sample $P^{\delta}_{x|a}$ we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\infty = \delta_0 > \delta_1 > \ldots > \delta_N = \delta$ and consider

$$P_x = P^{\delta_0}_{x|a},\ P^{\delta_1}_{x|a},\ \ldots,\ P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$$

- $P_x = P^{\delta_0}_{x|a}$ is the prior distribution (often easy to sample).
- $P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$ is the target distribution.
- Intermediate distributions define a "ladder" which smoothly interpolates from prior to target.

For P_x a Gaussian prior, efficient Monte Carlo methods are available based on pre-conditioned Crank Nicholson and its extensions [Cotter et al., 2013]. Not going to discuss further - too much detail - but remember this point for later!

Tempering for Sampling $P^{\delta}_{x|a}$

To sample $P^{\delta}_{x|a}$ we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\infty = \delta_0 > \delta_1 > \ldots > \delta_N = \delta$ and consider

$$P_x = P^{\delta_0}_{x|a},\ P^{\delta_1}_{x|a},\ \ldots,\ P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$$

- $P_x = P^{\delta_0}_{x|a}$ is the prior distribution (often easy to sample).
- $P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$ is the target distribution.
- Intermediate distributions define a "ladder" which smoothly interpolates from prior to target.

For P_x a Gaussian prior, efficient Monte Carlo methods are available based on pre-conditioned Crank Nicholson and its extensions [Cotter et al., 2013]. Not going to discuss further - too much detail - but remember this point for later!

Tempering for Sampling $P^{\delta}_{x|a}$

To sample $P^{\delta}_{x|a}$ we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\infty = \delta_0 > \delta_1 > \ldots > \delta_N = \delta$ and consider

$$P_x = P^{\delta_0}_{x|a}, \; P^{\delta_1}_{x|a}, \; \ldots, \; P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$$

- $P_x = P^{\delta_0}_{x|a}$ is the prior distribution (often easy to sample).
- $P^{\delta_N}_{x|a} = P^{\delta}_{x|a}$ is the target distribution.
- Intermediate distributions define a "ladder" which smoothly interpolates from prior to target.

For P_x a Gaussian prior, efficient Monte Carlo methods are available based on pre-conditioned Crank Nicholson and its extensions [Cotter et al., 2013]. Not going to discuss further - too much detail - but remember this point for later!

Example: Poisson's Equation

Consider

$$-\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t) = \sin(2\pi t) \qquad t \in (0,1)$$
$$x(t) = 0 \qquad t = 0, t = 1$$

- Use a Gaussian prior on x.
- Impose boundary conditions explicitly.
- Impose interior conditions at $t = 1/3$, $t = 2/3$.
- Construct the posterior using numerical disintegration with $\delta \in \{1.0, 10^{-2}, 10^{-4}\}$.
- Use relaxation function $\phi(r) = \exp(-r^2)$.
- Facilitated with pre-conditioned Crank-Nicholson.

Example: Poisson's Equation

Consider

$$-\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t) = \sin(2\pi t) \qquad t \in (0,1)$$

$$x(t) = 0 \qquad t = 0, t = 1$$

- Use a Gaussian prior on x.
- Impose boundary conditions explicitly.
- Impose interior conditions at $t = 1/3$, $t = 2/3$.
- Construct the posterior using numerical disintegration with $\delta \in \{1.0, 10^{-2}, 10^{-4}\}$.
- Use relaxation function $\phi(r) = \exp(-r^2)$.
- Facilitated with pre-conditioned Crank-Nicholson.

Consider

$$-\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t) = \sin(2\pi t) \qquad t \in (0,1)$$

$$x(t) = 0 \qquad t = 0, t = 1$$

- Use a Gaussian prior on x.
- Impose boundary conditions explicitly.
- Impose interior conditions at $t = 1/3$, $t = 2/3$.
- Construct the posterior using numerical disintegration with $\delta \in \{1.0, 10^{-2}, 10^{-4}\}$.
- Use relaxation function $\phi(r) = \exp(-r^2)$.
- Facilitated with pre-conditioned Crank-Nicholson.

Example: Poisson's Equation

Consider

$$-\frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t) = \sin(2\pi t) \qquad t \in (0,1)$$

$$x(t) = 0 \qquad t = 0, t = 1$$

- Use a Gaussian prior on x.
- Impose boundary conditions explicitly.
- Impose interior conditions at $t = 1/3$, $t = 2/3$.
- Construct the posterior using numerical disintegration with $\delta \in \{1.0, 10^{-2}, 10^{-4}\}$.
- Use relaxation function $\phi(r) = \exp(-r^2)$.
- Facilitated with pre-conditioned Crank-Nicholson.

Example: Poisson's Equation

In what follows, on the **left** are samples from the relaxed posterior $P^{\delta}_{x|a}$ in $\mathcal{X}$-space.

On the **right** are contours of

$$\phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

in $\mathcal{A}$-space.

All tempering is left "under the hood"; we will just consider the effect of $\delta \downarrow 0$.

(Monte Carlo error was negligible in this example when tempering and pre-conditioned Crank-Nicholson was used).

Example: Poisson's Equation

In what follows, on the **left** are samples from the relaxed posterior $P^{\delta}_{x|a}$ in $\mathcal{X}$-space.

On the **right** are contours of

$$\phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

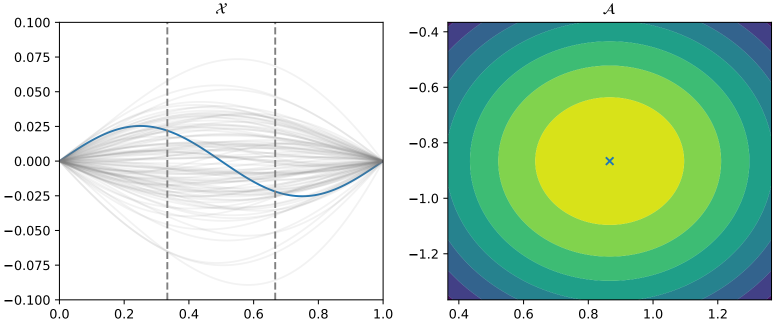
in $\mathcal{A}$-space.

All tempering is left "under the hood"; we will just consider the effect of $\delta \downarrow 0$.

(Monte Carlo error was negligible in this example when tempering and pre-conditioned Crank-Nicholson was used).

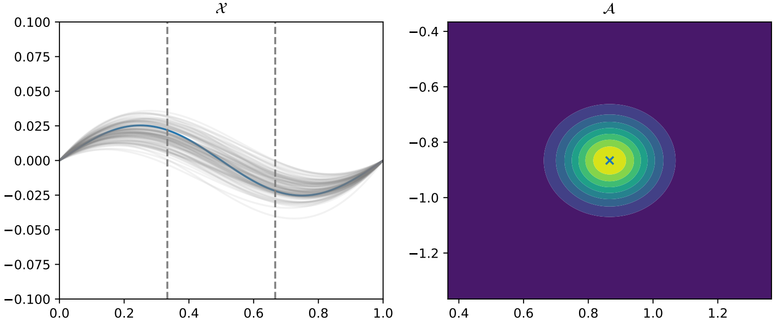
Example: Poisson's Equation

$\delta = 1.0$



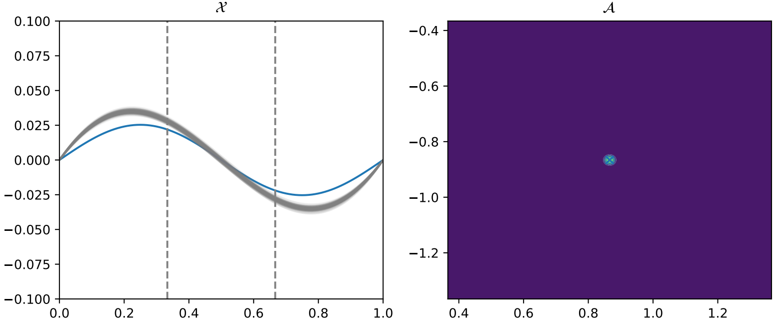
Example: Poisson's Equation

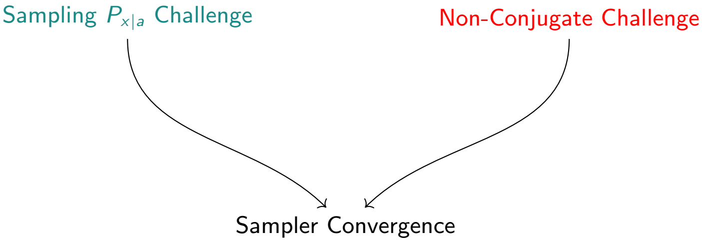
$\delta = 0.01$



Example: Poisson's Equation

$\delta = 0.0001$




Sampling $P_{x|a}$ Challenge
Non-Conjugate Challenge
Sampler Convergence

Non-Conjugate Challenge

Assume $\mathcal{X}$ admits a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$, so that for any $x \in \mathcal{X}$

$$x(t) = \sum_{i=0}^{\infty} \alpha_i \phi_i(t)$$

Recall that different u_i require different γ_i for the sum to exist:

- u_i IID Uniform, $\gamma \in \ell^1$
- u_i IID Gaussian, $\gamma \in \ell^2$
- u_i IID Cauchy, $\gamma \in \ell^2$

Key Idea: Update only the first N terms of the series based on the information $A(x) = a$.

Equivalent to consider the information operator $A_N = A \circ P_N$ where P_N is orthogonal projection onto $\{\phi_i\}_{i=0}^{N}$ (assumes a Hilbert structure on $\mathcal{X}$).

More sophisticated ("likelihood informed") alternatives to A_N; [Cui et al., 2014].

Non-Conjugate Challenge

Assume $\mathcal{X}$ admits a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$, so that for any $x \in \mathcal{X}$

$$x(t) = \sum_{i=0}^{\infty} \gamma_i u_i \phi_i(t)$$

Recall that different u_i require different γ_i for the sum to exist:

- u_i IID Uniform, $\gamma \in \ell^1$
- u_i IID Gaussian, $\gamma \in \ell^2$
- u_i IID Cauchy, $\gamma \in \ell^2$

Key Idea: Update only the first N terms of the series based on the information $A(x) = a$.

Equivalent to consider the information operator $A_N = A \circ P_N$ where P_N is orthogonal projection onto $\{\phi_i\}_{i=0}^{N}$ (assumes a Hilbert structure on $\mathcal{X}$).

More sophisticated ("likelihood informed") alternatives to A_N; [Cui et al., 2014].

Non-Conjugate Challenge

Assume $\mathcal{X}$ admits a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$, so that for any $x \in \mathcal{X}$

$$x(t) = \sum_{i=0}^{\infty} \gamma_i u_i \phi_i(t)$$

Recall that different u_i require different γ_i for the sum to exist:

- u_i IID Uniform, $\gamma \in \ell^1$ $\leftarrow$?
- u_i IID Gaussian, $\gamma \in \ell^2$ $\leftarrow$ pre-conditioned Crank-Nicholson
- u_i IID Cauchy, $\gamma \in \ell^2$ $\leftarrow$?

Key Idea: Update only the first N terms of the series based on the information $A(x) = a$.

Equivalent to consider the information operator $A_N = A \circ P_N$ where P_N is orthogonal projection onto $\{\phi_i\}_{i=0}^{N}$ (assumes a Hilbert structure on $\mathcal{X}$).

More sophisticated ("likelihood informed") alternatives to A_N; [Cui et al., 2014].

Non-Conjugate Challenge

Assume $\mathcal{X}$ admits a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$, so that for any $x \in \mathcal{X}$

$$x(t) = \sum_{i=0}^{\infty} \gamma_i u_i \phi_i(t)$$

Recall that different u_i require different γ_i for the sum to exist:

- u_i IID Uniform, $\gamma \in \ell^1$ $\leftarrow$?
- u_i IID Gaussian, $\gamma \in \ell^2$ $\leftarrow$ pre-conditioned Crank-Nicholson
- u_i IID Cauchy, $\gamma \in \ell^2$ $\leftarrow$?

Key Idea: Update only the first N terms of the series based on the information $A(x) = a$.

Equivalent to consider the information operator $A_N = A \circ P_N$ where P_N is orthogonal projection onto $\{\phi_i\}_{i=0}^{N}$ (assumes a Hilbert structure on $\mathcal{X}$).

More sophisticated ("likelihood informed") alternatives to A_N; [Cui et al., 2014].

Assume $\mathcal{X}$ admits a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$, so that for any $x \in \mathcal{X}$

$$x(t) = \sum_{i=0}^{\infty} \gamma_i u_i \phi_i(t)$$

Recall that different u_i require different γ_i for the sum to exist:

- u_i IID Uniform, $\gamma \in \ell^1$ $\leftarrow$?
- u_i IID Gaussian, $\gamma \in \ell^2$ $\leftarrow$ pre-conditioned Crank-Nicholson
- u_i IID Cauchy, $\gamma \in \ell^2$ $\leftarrow$?

Key Idea: Update only the first N terms of the series based on the information $A(x) = a$.

Equivalent to consider the information operator $A_N = A \circ P_N$ where P_N is orthogonal projection onto $\{\phi_i\}_{i=0}^{N}$ (assumes a Hilbert structure on $\mathcal{X}$).

More sophisticated ("likelihood informed") alternatives to A_N; [Cui et al., 2014].

Non-Conjugate Challenge

Assume $\mathcal{X}$ admits a Schauder basis $\{\phi_i\}_{i=1}^{\infty}$, so that for any $x \in \mathcal{X}$

$$x(t) = \sum_{i=0}^{\infty} \gamma_i u_i \phi_i(t)$$

Recall that different u_i require different γ_i for the sum to exist:

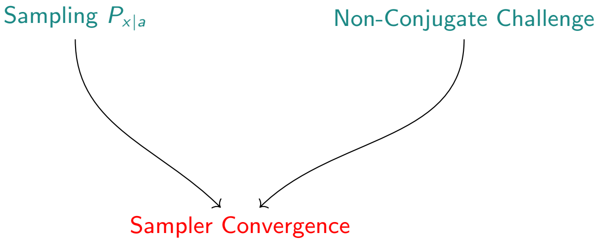
- u_i IID Uniform, $\gamma \in \ell^1$ $\leftarrow$?
- u_i IID Gaussian, $\gamma \in \ell^2$ $\leftarrow$ pre-conditioned Crank-Nicholson
- u_i IID Cauchy, $\gamma \in \ell^2$ $\leftarrow$?

Key Idea: Update only the first N terms of the series based on the information $A(x) = a$.

Equivalent to consider the information operator $A_N = A \circ P_N$ where P_N is orthogonal projection onto $\{\phi_i\}_{i=0}^{N}$ (assumes a Hilbert structure on $\mathcal{X}$).

More sophisticated ("likelihood informed") alternatives to A_N; [Cui et al., 2014].

Overview


Sampling $P_{x|a}$
Non-Conjugate Challenge
Sampler Convergence

Convergence, but in what metric?

The aim here is to show that the two approximations

- $P_{x|a} \approx P^{\delta}_{x|a}$
- $A \approx A_N$

combine to produce an approximation $P^{\delta,N}_{x|a}$ to the distribution $P_{x|a}$ of interest.

The results that we consider are formulated in terms of integration error:

$$d_{\mathcal{F}}(P^{\delta,N}_{x|a}, P_{x|a}) \quad = \quad \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| P^{\delta,N}_{x|a}(f) - P_{x|a}(f) \right|$$

where we use the notation $\nu(f) = \int f \mathrm{d}\nu$.

The test functions f come from a normed space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$. This can be chosen to induce Wasserstein, total variation, etc.

NB: This is only useful when $\mathcal{F}$ is not "too rich".

Convergence, but in what metric?

The aim here is to show that the two approximations

- $P_{x|a} \approx P^{\delta}_{x|a}$
- $A \approx A_N$

combine to produce an approximation $P^{\delta,N}_{x|a}$ to the distribution $P_{x|a}$ of interest.

The results that we consider are formulated in terms of integration error:

$$d_{\mathcal{F}}(P^{\delta,N}_{x|a}, P_{x|a}) \quad = \quad \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| P^{\delta,N}_{x|a}(f) - P_{x|a}(f) \right|$$

where we use the notation $\nu(f) = \int f \mathrm{d}\nu$.

The test functions f come from a normed space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$. This can be chosen to induce Wasserstein, total variation, etc.

NB: This is only useful when $\mathcal{F}$ is not "too rich".

Convergence, but in what metric?

The aim here is to show that the two approximations

- $P_{x|a} \approx P_{x|a}^{\delta}$
- $A \approx A_N$

combine to produce an approximation $P_{x|a}^{\delta,N}$ to the distribution $P_{x|a}$ of interest.

The results that we consider are formulated in terms of integration error:

$$d_{\mathcal{F}}(P_{x|a}^{\delta,N}, P_{x|a}) = \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| P_{x|a}^{\delta,N}(f) - P_{x|a}(f) \right|$$

where we use the notation $\nu(f) = \int f \mathrm{d}\nu$.

The test functions f come from a normed space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$. This can be chosen to induce Wasserstein, total variation, etc.

NB: This is only useful when $\mathcal{F}$ is not "too rich".

Convergence, but in what metric?

The aim here is to show that the two approximations

- $P_{x|a} \approx P_{x|a}^{\delta}$
- $A \approx A_N$

combine to produce an approximation $P_{x|a}^{\delta,N}$ to the distribution $P_{x|a}$ of interest.

The results that we consider are formulated in terms of integration error:

$$d_{\mathcal{F}}(P_{x|a}^{\delta,N}, P_{x|a}) = \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| P_{x|a}^{\delta,N}(f) - P_{x|a}(f) \right|$$

where we use the notation $\nu(f) = \int f \mathrm{d}\nu$.

The test functions f come from a normed space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$. This can be chosen to induce Wasserstein, total variation, etc.

NB: This is only useful when $\mathcal{F}$ is not "too rich".

Convergence of $P^{\delta}_{x|a}$ to $P_{x|a}$

Assume that:

- $\exists \alpha > 0$ s.t. $C_\phi^\alpha := \int r^{\alpha+n-1}\phi(r)\mathrm{d}r < \infty$
- $\exists C_\mu > 0$ s.t.

$$d_{\mathcal{F}}(P_{x|a}, P_{x|a'}) \quad \leq \quad C_\mu \left\| a - a' \right\|^\alpha$$

for $A_\# \mu$-almost-all $a, a' \in \mathcal{A}$.

Then, for $\delta \ll 1$,

$$d_{\mathcal{F}}(P^{\delta}_{x|a}, P_{x|a}) \leq C_\mu \left(1 + \frac{C_\phi^\alpha}{C_\phi^0}\right) \delta^\alpha$$

for $A_\# \mu$-almost-all $a \in \mathcal{A}$ Proof in Cockayne et al. [2017].

Convergence of $P^{\delta}_{x|a}$ to $P_{x|a}$

Assume that:

- $\exists \alpha > 0$ s.t. $C_\phi^\alpha := \int r^{\alpha+n-1}\phi(r)\mathrm{d}r < \infty$
- $\exists C_\mu > 0$ s.t.

$$d_{\mathcal{F}}(P_{x|a}, P_{x|a'}) \quad \leq \quad C_\mu \left\| a - a' \right\|^\alpha$$

for $A_{\#}\mu$-almost-all $a, a' \in \mathcal{A}$.

Then, for $\delta \ll 1$,

$$d_{\mathcal{F}}(P^{\delta}_{x|a}, P_{x|a}) \leq C_\mu \left(1 + \frac{C_\phi^\alpha}{C_\phi^0}\right) \delta^\alpha$$

for $A_{\#}\mu$-almost-all $a \in \mathcal{A}$ Proof in Cockayne et al. [2017].

Denote by $P_{x|a}^{\delta,N}$ the approximation

$$\frac{\mathrm{d}P_{x|a}^{\delta,N}}{\mathrm{d}P_x}(x) \propto \phi\left(\frac{\|A \circ P_N(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

Assume that:

- $\forall R > 0\ \exists C_R$ s.t. $|\log\phi(r) - \log\phi(r')| < C_R|r - r'|$ for all $r, r' < R$.
- $\exists$ measurable m s.t.

  $$\|A(u) - A \circ P_N(u)\| \leq \exp(m\,\|u\|_{\mathcal{X}})\Phi(N)$$

  where $\Phi(N) \downarrow 0$ and $\mathbb{E}_{X \sim P_x}[\exp(2m(\|X\|_{\mathcal{X}}))] < \infty$.
- $\sup_{x \in \mathcal{X}} \|A(x)\|_{\mathcal{A}} < \infty$
- $\exists C_{\mathcal{F}}$ s.t. $\|f\|_\infty \leq C_{\mathcal{F}}\|f\|_{\mathcal{F}}$ for all $f \in \mathcal{F}$.

Then $d_{\mathcal{F}}(P_{x|a}^{\delta,N}, P_{x|a}^{\delta}) \leq C_\delta \Phi(N)$. Proof in Cockayne et al. [2017], builds on Stuart [2010].

Convergence of $P^{\delta,N}_{x|a}$ to $P^{\delta}_{x|a}$

Denote by $P^{\delta,N}_{x|a}$ the approximation

$$\frac{\mathrm{d}P^{\delta,N}_{x|a}}{\mathrm{d}P_x}(x) \propto \phi\left(\frac{\|A \circ P_N(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

Assume that:

- $\forall R > 0 \; \exists C_R$ s.t. $|\log\phi(r) - \log\phi(r')| < C_R|r - r'|$ for all $r, r' < R$.
- $\exists$ measurable m s.t.

  $$\|A(u) - A \circ P_N(u)\| \leq \exp(m\,\|u\|_{\mathcal{X}})\Phi(N)$$

  where $\Phi(N) \downarrow 0$ and $\mathbb{E}_{X \sim P_x}[\exp(2m(\|X\|_{\mathcal{X}}))] < \infty$.
- $\sup_{x \in \mathcal{X}} \|A(x)\|_{\mathcal{A}} < \infty$
- $\exists C_{\mathcal{F}}$ s.t. $\|f\|_\infty \leq C_{\mathcal{F}}\|f\|_{\mathcal{F}}$ for all $f \in \mathcal{F}$.

Then $d_{\mathcal{F}}(P^{\delta,N}_{x|a}, P^{\delta}_{x|a}) \leq C_\delta \Phi(N)$. Proof in Cockayne et al. [2017], builds on Stuart [2010].

Convergence of $P^{\delta,N}_{x|a}$ to $P^{\delta}_{x|a}$

Denote by $P^{\delta,N}_{x|a}$ the approximation

$$\frac{\mathrm{d}P^{\delta,N}_{x|a}}{\mathrm{d}P_x}(x) \propto \phi\left(\frac{\|A \circ P_N(x) - a\|_{\mathcal{A}}}{\delta}\right)$$

Assume that:

- $\forall R > 0 \; \exists C_R$ s.t. $|\log \phi(r) - \log \phi(r')| < C_R |r - r'|$ for all $r, r' < R$.
- $\exists$ measurable m s.t.

$$\|A(u) - A \circ P_N(u)\| \leq \exp(m \, \|u\|_{\mathcal{X}}) \Phi(N)$$

  where $\Phi(N) \downarrow 0$ and $\mathbb{E}_{X \sim P_x}[\exp(2m(\|X\|_{\mathcal{X}}))] < \infty$.
- $\sup_{x \in \mathcal{X}} \|A(x)\|_{\mathcal{A}} < \infty$
- $\exists C_{\mathcal{F}}$ s.t. $\|f\|_\infty \leq C_{\mathcal{F}} \|f\|_{\mathcal{F}}$ for all $f \in \mathcal{F}$.

Then $d_{\mathcal{F}}(P^{\delta,N}_{x|a}, P^{\delta}_{x|a}) \leq C_\delta \Phi(N)$. Proof in Cockayne et al. [2017], builds on Stuart [2010].

Example: Solution of a Non-linear ODE

Consider Painlevé's first transcendental:

$$\begin{aligned} x''(t) &= x(t)^2 - t, \quad t \in \mathbb{R}_+ \\ x(0) &= 0 \\ t^{-1/2}x(t) &\to 1 \text{ as } t \to \infty \end{aligned}$$

The information operator is

$$A(x) = \begin{bmatrix} x''(t_1) - x(t_1)^2 \\ \vdots \\ x''(t_n) - x(t_n)^2 \\ x(0) \\ \lim_{t\to\infty} t^{-1/2}x(t) \end{bmatrix} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \\ 0 \\ 1 \end{bmatrix}.$$

Construct an infinite-dimensional prior $P_x \in \mathcal{P}_\mathcal{X}$ as

$$x(t) = \sum_{i=0}^{\infty} u_i \gamma_i \phi_i(t)$$

with u_i i.i.d. std. Cauchy coefficients, weights $\gamma_i = (i+1)^{-2}$ and $\phi_i(t)$ (normalized) Chebyshev polynomials of the first kind. [See Sullivan, 2016, for mathematical details.]

Consider Painlevé's first transcendental:

$$\begin{aligned} x''(t) &= x(t)^2 - t, \quad t \in \mathbb{R}_+ \\ x(0) &= 0 \\ t^{-1/2}x(t) &\to 1 \text{ as } t \to \infty \end{aligned}$$

The information operator is

$$A(x) = \begin{bmatrix} x''(t_1) - x(t_1)^2 \\ \vdots \\ x''(t_n) - x(t_n)^2 \\ x(0) \\ \lim_{t\to\infty} t^{-1/2}x(t) \end{bmatrix} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \\ 0 \\ 1 \end{bmatrix}.$$

Construct an infinite-dimensional prior $P_x \in \mathcal{P}_\mathcal{X}$ as

$$x(t) = \sum_{i=0}^{\infty} u_i \gamma_i \phi_i(t)$$

with u_i i.i.d. std. Cauchy coefficients, weights $\gamma_i = (i+1)^{-2}$ and $\phi_i(t)$ (normalized) Chebyshev polynomials of the first kind. [See Sullivan, 2016, for mathematical details.]

Example: Solution of a Non-linear ODE

Consider Painlevé's first transcendental:

$$\begin{aligned} x''(t) &= x(t)^2 - t, \quad t \in \mathbb{R}_+ \\ x(0) &= 0 \\ t^{-1/2}x(t) &\to 1 \text{ as } t \to \infty \end{aligned}$$

The information operator is

$$A(x) = \begin{bmatrix} x''(t_1) - x(t_1)^2 \\ \vdots \\ x''(t_n) - x(t_n)^2 \\ x(0) \\ \lim_{t\to\infty} t^{-1/2}x(t) \end{bmatrix} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \\ 0 \\ 1 \end{bmatrix}.$$

Construct an infinite-dimensional prior $P_x \in \mathcal{P}_\mathcal{X}$ as

$$x(t) = \sum_{i=0}^{\infty} u_i \gamma_i \phi_i(t)$$

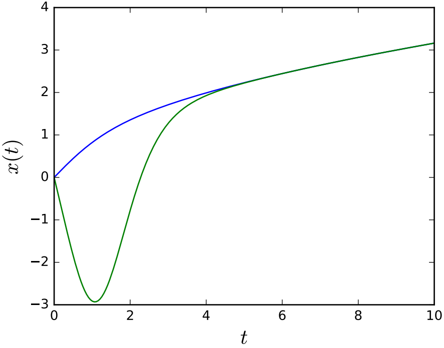
with u_i i.i.d. std. Cauchy coefficients, weights $\gamma_i = (i+1)^{-2}$ and $\phi_i(t)$ (normalized) Chebyshev polynomials of the first kind. [See Sullivan, 2016, for mathematical details.]

Example: Solution of a Non-linear ODE

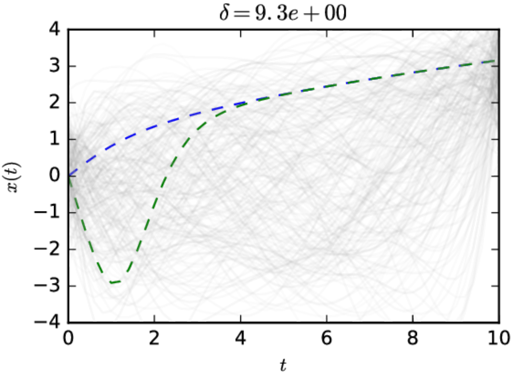
Consider Painlevé's first transcendental:

$$\begin{aligned} x''(t) &= x(t)^2 - t, \quad t \in \mathbb{R}_+ \\ x(0) &= 0 \\ t^{-1/2}x(t) &\to 1 \text{ as } t \to \infty \end{aligned}$$

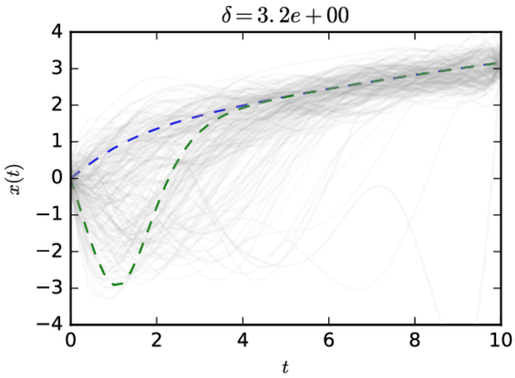
Exact "positive" and "negative" solutions:



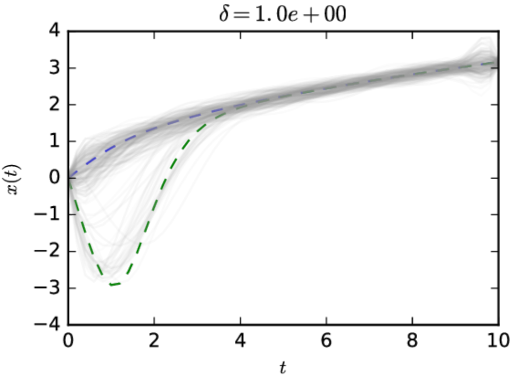
Painlevé: Posterior Measures



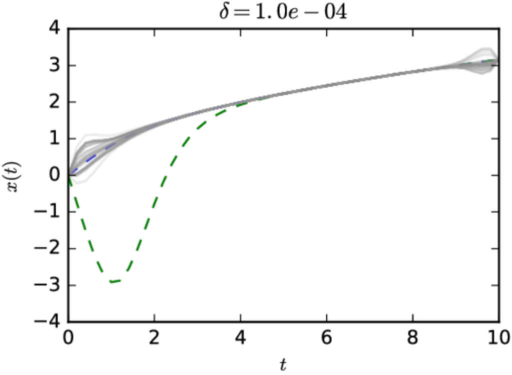
Painlevé: Posterior Measures



Painlevé: Posterior Measures

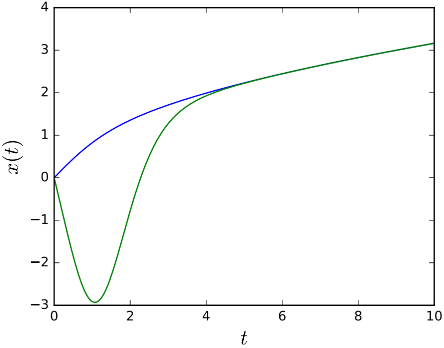


Painlevé: Posterior Measures



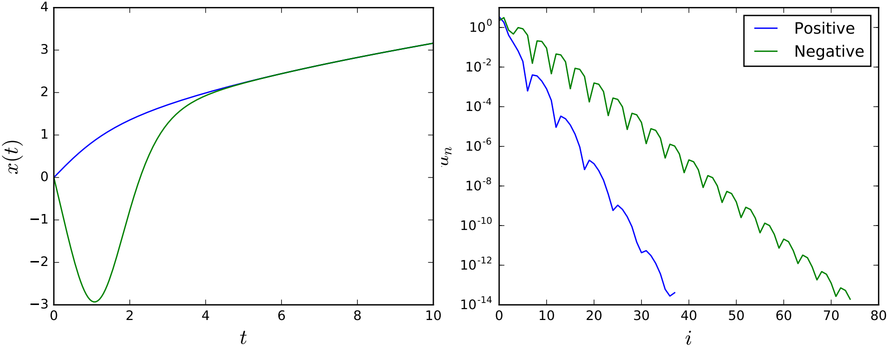
How might we explain the collapse of the posterior onto one solution?

Consider the spectra $\{u_i\}_{i=0}^{\infty}$ corresponding to the true solutions:



How might we explain the collapse of the posterior onto one solution?

Consider the spectra $\{u_i\}_{i=0}^{\infty}$ corresponding to the true solutions:



Conclusion

In Part II it has been argued that:

- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, P_x radon, $\Sigma_{\mathcal{A}}$ countably generated).
- The mathematical properties of the posterior $P_{x|a}$ are hard to understand in general.
- Wide open area for research!

END OF PART II

Conclusion

In Part II it has been argued that:

- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, P_x radon, $\Sigma_{\mathcal{A}}$ countably generated).
- The mathematical properties of the posterior $P_{x|a}$ are hard to understand in general.
- Wide open area for research!

END OF PART II

Conclusion

In Part II it has been argued that:

- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, P_x radon, $\Sigma_{\mathcal{A}}$ countably generated).
- The mathematical properties of the posterior $P_{x|a}$ are hard to understand in general.
- Wide open area for research!

END OF PART II

Conclusion

In Part II it has been argued that:

- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, P_x radon, $\Sigma_{\mathcal{A}}$ countably generated).
- The mathematical properties of the posterior $P_{x|a}$ are hard to understand in general.
- Wide open area for research!

END OF PART II

Conclusion

In Part II it has been argued that:

- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, P_x radon, $\Sigma_{\mathcal{A}}$ countably generated).
- The mathematical properties of the posterior $P_{x|a}$ are hard to understand in general.
- Wide open area for research!

END OF PART II