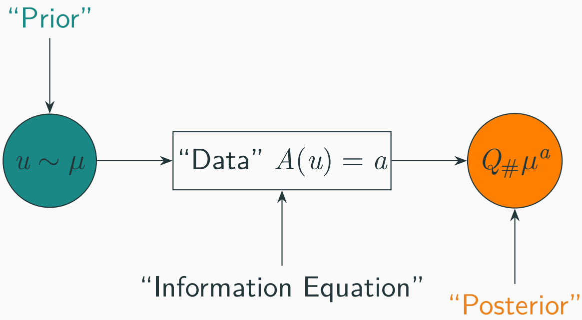
Bayesian Probabilistic Numerical Methods

Numerical Disintegration and Pipelines

Jon Cockayne

June 6, 2017

(Re)introduction

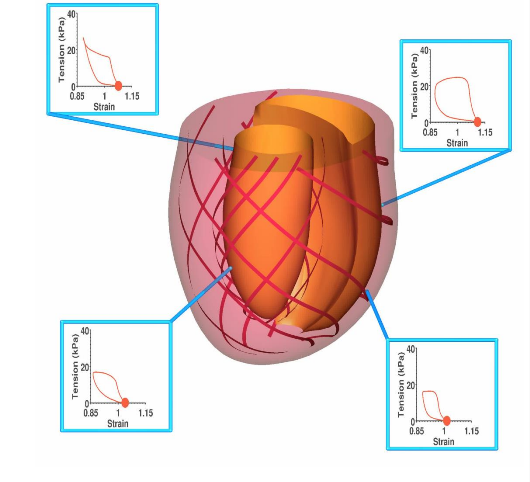

"Prior"
$u \sim \mu$
"Data" $A(u) = a$
$Q_{\#}\mu^a$
"Information Equation"
"Posterior"

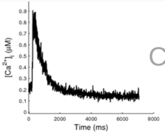
Q1: How can we access μ^a?

(Re)introduction

Unless probabilistic numerical methods “agree” about what their uncertainty means, they cannot be composed coherently.

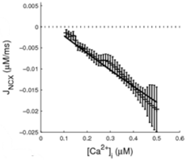
Modelling Electro-Mechanics in the Heart

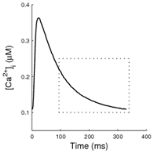




Ca Flux during Caffeine Ca transient

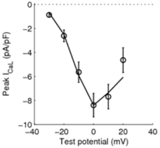
Fit NCX Model

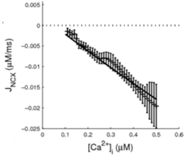




Ca Flux during tail of Field Stimulation Ca Transient. Less Ca Flux through NCX (calculated)

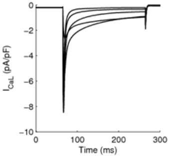
I_{CaL} Voltage clamp traces

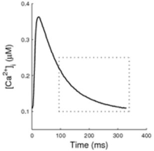




Fit SERCA Model

Fit I_{CaL} Model





Ca flux during start of Field Stimulation Ca Transient. Less Ca Flux through NCX, SERCA and I_{CaL} (calculated)

Fit RyR Model

Q2: when is it “legal” to compose Bayesian PNM in pipelines?

Numerical Disintegration

Numerical Disintegration

Recall, the issue:

$$\mathcal{X}^a = \{u \in \mathcal{X} : A(u) = a\}$$

$$\mu(X^a) = 0$$

Numerical Disintegration

Recall, the issue:

$$\mathcal{X}^a = \{u \in \mathcal{X} : A(u) = a\}$$

$$\mu(X^a) = 0$$

which means...

$$\nexists \frac{\mathrm{d}\mu^a}{\mathrm{d}\mu}$$

Our Approach

Design an algorithm for approximately sampling μ^a.

Our Approach

Design an algorithm for approximately sampling μ^a.

Two sources of error

- Intractability of μ^a ("Numerical Disintegration")

Our Approach

Design an algorithm for approximately sampling μ^a.

Two sources of error

- Intractability of μ^a (“Numerical Disintegration”)
- Intractability of non-Gaussian priors (“prior truncation”)

Three Considerations

Numerical Disintegration

Prior Truncation

Three Considerations

Numerical Disintegration

Prior Truncation

Sampler Convergence

Three Considerations

Numerical Disintegration

Prior Truncation

Sampler Convergence

Numerical Disintegration

Introduce the δ-relaxed measure μ_δ^a...

$$\frac{\mathrm{d}\mu_\delta^a}{\mathrm{d}\mu} \propto \phi\left(\frac{\|A(u) - a\|_{\mathcal{A}}}{\delta}\right)$$

Numerical Disintegration

Introduce the δ-relaxed measure μ_δ^a...

$$\frac{\mathrm{d}\mu_\delta^a}{\mathrm{d}\mu} \propto \phi\left(\frac{\|A(u) - a\|_{\mathcal{A}}}{\delta}\right)$$

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

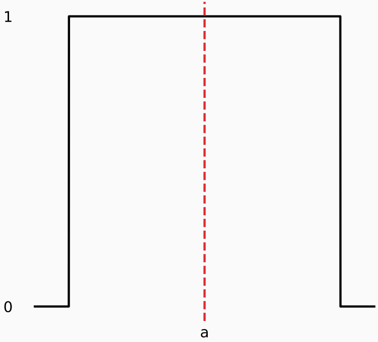
- $\phi(0) = 1$
- $\phi(r) \to 0$ as $r \to \infty$.

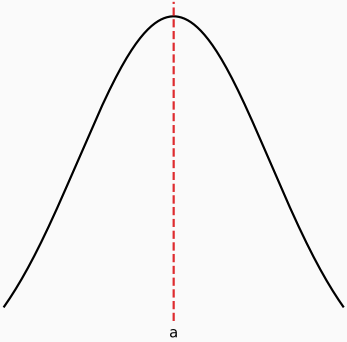
Numerical Disintegration: Intuition

"Ideal" Radon–Nikodym derivative

$$\text{``}\frac{\mathrm{d}\mu^a}{\mathrm{d}\mu} \propto \mathbb{I}(u \in \mathcal{X}^a)\text{''}$$

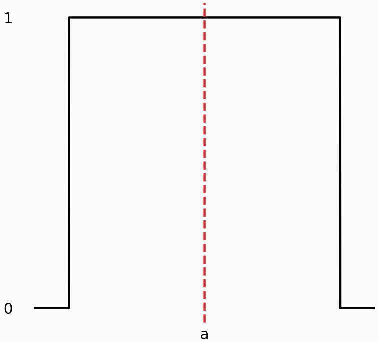
Example Relaxation Functions





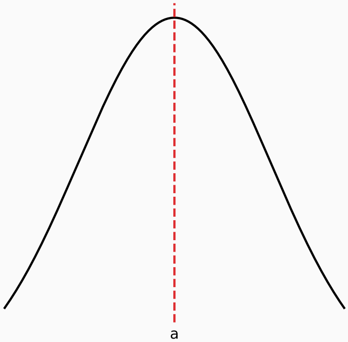
$$\phi(r) = \mathbb{I}(r < 1) \qquad \phi(r) = \exp(-r^2)$$

Example Relaxation Functions



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

Uniform noise over $B_\delta(a)$



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

Gaussian noise with s.d. $\propto \delta$

Tempering for Sampling μ_δ^a

To sample μ_δ^a we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Tempering for Sampling μ_δ^a

To sample μ_δ^a we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\delta_0 > \delta_1 > \ldots > \delta_N$ and consider

$$\mu_{\delta_0}^a,\ \mu_{\delta_1}^a,\ \ldots,\ \mu_{\delta_N}^a$$

Tempering for Sampling μ_δ^a

To sample μ_δ^a we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\delta_0 > \delta_1 > \ldots > \delta_N$ and consider

$$\mu_{\delta_0}^a,\ \mu_{\delta_1}^a,\ \ldots,\ \mu_{\delta_N}^a$$

- $\mu_{\delta_0}^a$ is the prior and easy to sample.
- $\mu_{\delta_N}^a$ has δ_N close to zero and is hard to sample.
- Intermediate distributions define a "ladder" which takes us from prior to posterior.

Example: Poisson's Equation

Consider

$$
\begin{aligned}
-\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) &= \sin(2\pi x) \qquad && x \in (0,1) \\
u(x) &= 0 && x = 0, x = 1
\end{aligned}
$$

Example: Poisson's Equation

Consider

$$\begin{aligned} -\frac{\mathrm{d}^2}{\mathrm{d}x^2} u(x) &= \sin(2\pi x) & \qquad x \in (0,1) \\ u(x) &= 0 & \qquad x = 0, x = 1 \end{aligned}$$

- Use a Gaussian prior on $u(x)$.

Example: Poisson's Equation

Consider

$$
\begin{aligned}
-\frac{\mathrm{d}^2}{\mathrm{d}x^2} u(x) &= \sin(2\pi x) && x \in (0,1) \\
u(x) &= 0 && x = 0, x = 1
\end{aligned}
$$

- Use a Gaussian prior on $u(x)$.
- Impose boundary conditions explicitly.
- Impose interior conditions at $x = 1/3$, $x = 2/3$.

Example: Poisson's Equation

Consider

$$
\begin{aligned}
-\frac{\mathrm{d}^2}{\mathrm{d}x^2} u(x) &= \sin(2\pi x) && x \in (0,1) \\
u(x) &= 0 && x = 0, x = 1
\end{aligned}
$$

- Use a Gaussian prior on $u(x)$.
- Impose boundary conditions explicitly.
- Impose interior conditions at $x = 1/3$, $x = 2/3$.
- Construct the posterior using ND with $\delta \in \{1.0, 10^{-2}, 10^{-4}\}$.
- Use $\phi(r) = \exp(-r^2)$.

Example: Poisson's Equation

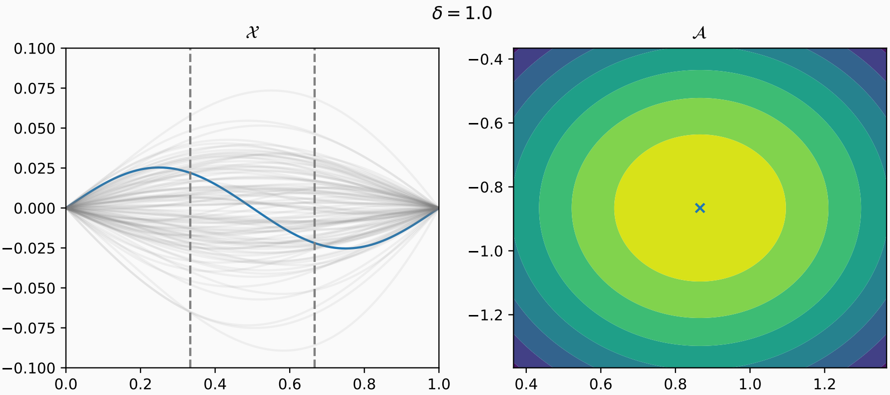
In what follows, on the **left** are samples from the posterior μ_δ^a in $\mathcal{X}$-space.

On the **right** are contours of

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

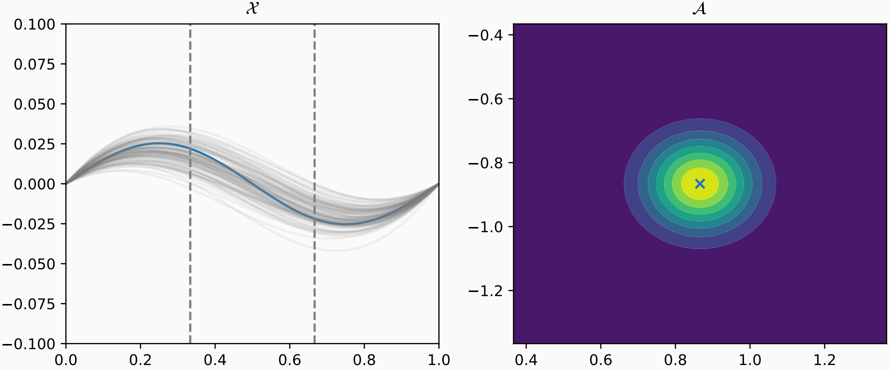
in $\mathcal{A}$-space.

Example: Poisson's Equation

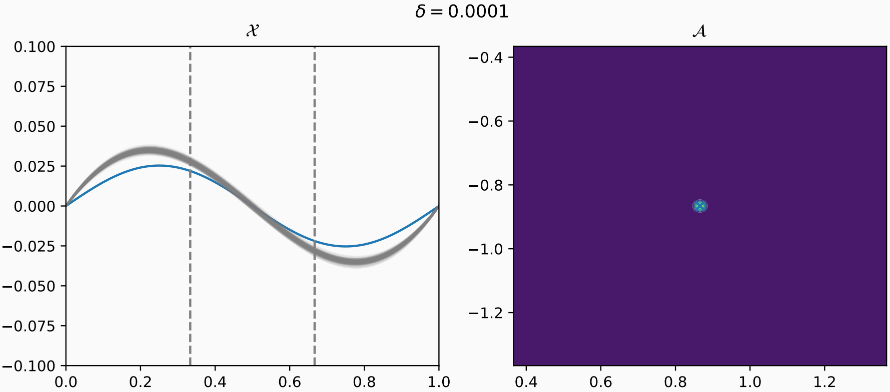


Example: Poisson's Equation

$\delta = 0.01$



Example: Poisson's Equation



Three Considerations

Numerical Disintegration

Prior Truncation

Sampler Convergence

Prior Construction

Assume $\mathcal{X}$ has a countable basis $\{\phi_i\}$, $i = 0, \dots, \infty$. Then for any $u \in \mathcal{X}$

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

Prior Construction

Assume $\mathcal{X}$ has a countable basis $\{\phi_i\}$, $i = 0, \ldots, \infty$. Then for any $u \in \mathcal{X}$

$$u(x) = \sum_{i=0}^{\infty} \gamma_i \xi_i \phi_i(x)$$

Different ξ_i require different γ for almost-sure convergence...

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

Prior Construction

Assume $\mathcal{X}$ has a countable basis $\{\phi_i\}$, $i = 0, \dots, \infty$. Then for any $u \in \mathcal{X}$

$$u^N(x) = \sum_{i=0}^{N} \gamma_i \xi_i \phi_i(x)$$

Different ξ_i require different γ for almost-sure convergence...

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

For practical computation we truncate to N terms.

Three Considerations

Numerical Disintegration

Prior Truncation

Sampler Convergence

Convergence, but in what metric?

All results show weak convergence framed in terms of an abstract integral probability metric[1]:

$$d_{\mathcal{F}}(\nu, \nu') = \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| \nu(f) - \nu'(f) \right|$$

[1] Müller [1997]

Convergence, but in what metric?

All results show weak convergence framed in terms of an abstract integral probability metric[1]:

$$d_{\mathcal{F}}(\nu, \nu') = \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| \nu(f) - \nu'(f) \right|$$

Results are generic to $A(u)$, μ.

[1] Müller [1997]

Convergence, but in what metric?

All results show weak convergence framed in terms of an abstract integral probability metric[1]:

$$d_{\mathcal{F}}(\nu, \nu') = \sup_{\|f\|_{\mathcal{F}} \leq 1} \left| \nu(f) - \nu'(f) \right|$$

Results are generic to $A(u)$, μ.

Examples: Total Variation, Wasserstein

[1] Müller [1997]

Convergence of μ_δ^a

Theorem

Assume that

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

for some C_μ, α constant and $A_\# \mu$-almost-all $a, a' \in \mathcal{A}$.

Convergence of μ_δ^a

Theorem

Assume that

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

for some C_μ, α constant and $A_\# \mu$-almost-all $a, a' \in \mathcal{A}$.

Then, for small δ

$$d_{\mathcal{F}}(\mu_\delta^a, \mu^a) \leq C_\mu \left(1 + C_\phi\right) \delta^\alpha$$

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

Total Error

Denote by $\mu^a_{\delta,N}$ the posterior distribution given by

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

Total Error

Denote by $\mu^a_{\delta,N}$ the posterior distribution given by

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

Total Error

Denote by $\mu^a_{\delta,N}$ the posterior distribution given by

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

Assume that

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

Total Error

Denote by $\mu^a_{\delta,N}$ the posterior distribution given by

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

Assume that

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

Then under certain assumptions it can be shown[2] that:

$$d_{\mathcal{F}}(\mu^a, \mu^a_{\delta,N}) \leq C_\mu(1 + C_\phi)\delta^\alpha + C_\delta\Phi(N)$$

[2] Cockayne et al. [2017]

Total Error

Denote by $\mu^a_{\delta,N}$ the posterior distribution given by

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

Assume that

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

Then under certain assumptions it can be shown[2] that:

$$d_{\mathcal{F}}(\mu^a, \mu^a_{\delta,N}) \leq C_\mu(1 + C_\phi)\delta^\alpha + C_\delta \Phi(N)$$

Thus, we have convergence with δ provided $C_\delta\Phi(N)$ is controlled.

[2] Cockayne et al. [2017]

Numerical Disintegration

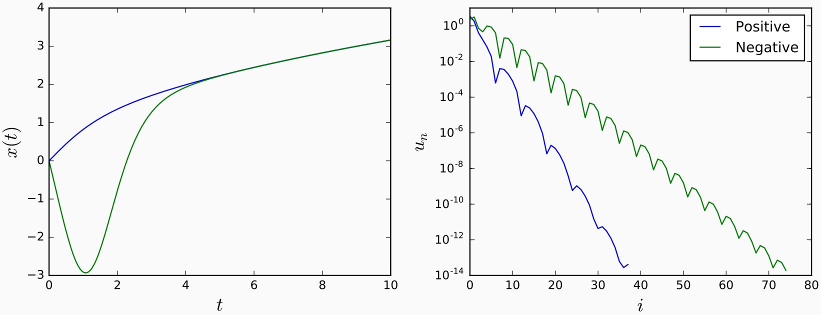
Numerical Example

Painlevé's First Transcendental

$$
\begin{aligned}
u''(x) - u(x)^2 &= -x \\
u(0) &= 0 \\
u(x) &\to \sqrt{x} \text{ as } x \to \infty
\end{aligned}
$$

Painlevé's First Transcendental

$$u''(x) - u(x)^2 = -x$$
$$u(0) = 0$$
$$u(10) = \sqrt{10}$$

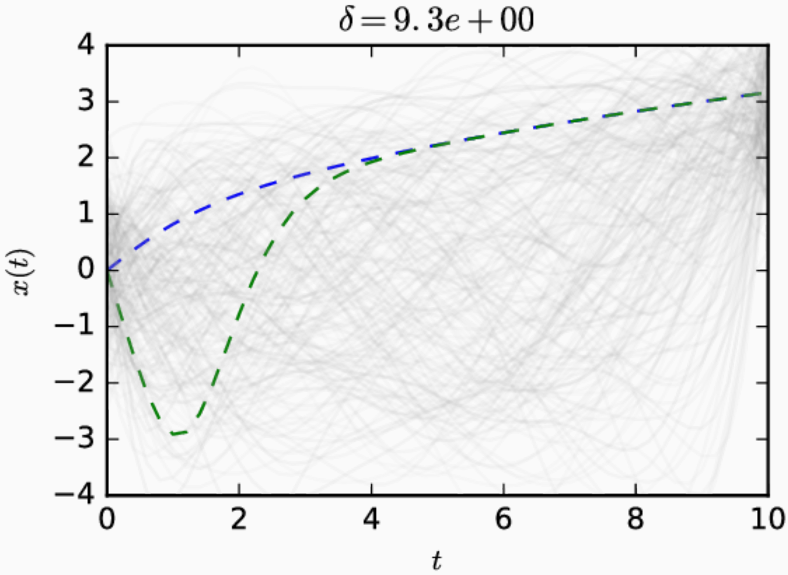


Painlevé's First Transcendental

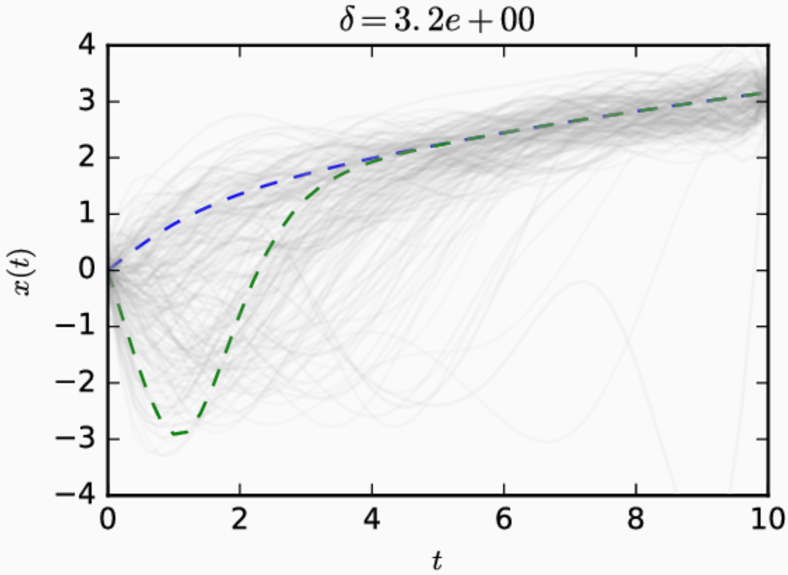
$$u''(x) - u(x)^2 = -x$$
$$u(0) = 0$$
$$u(10) = \sqrt{10}$$

We use $\phi(x) = \exp(-x^2)$, and define a schedule of 1600 δ from 10 to 10^{-4}. Following results are based on equi-spaced t_i, $i = 1, \dots, 15$, and generated with an SMC algorithm based upon a Cauchy prior.

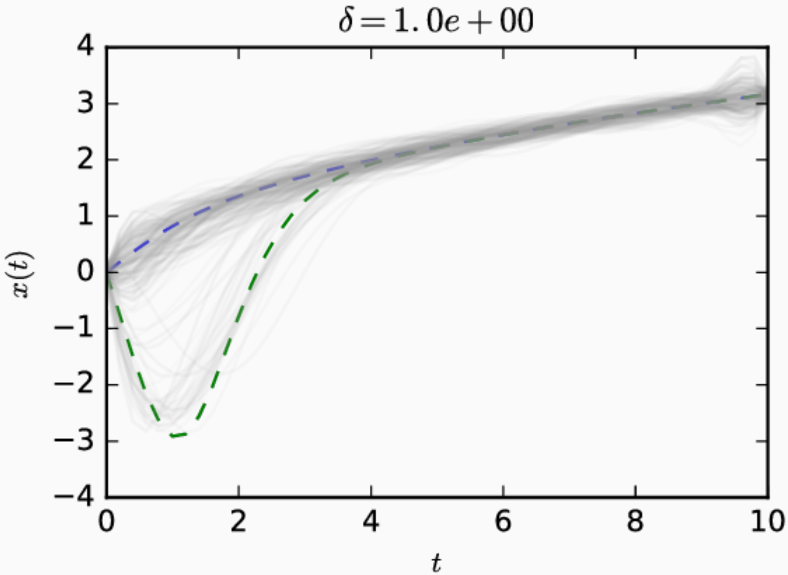
Painlevé: Posterior Measures



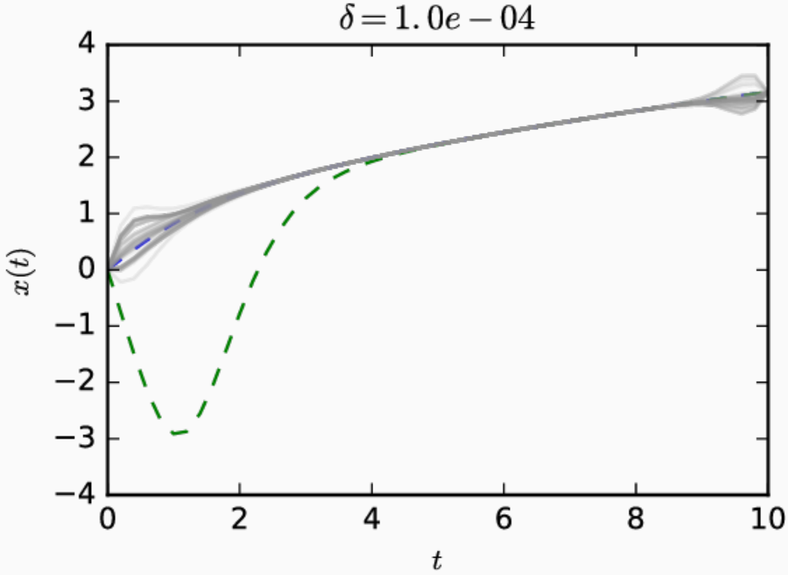
Painlevé: Posterior Measures



Painlevé: Posterior Measures



Painlevé: Posterior Measures



Pipelines

Example: Split Integration

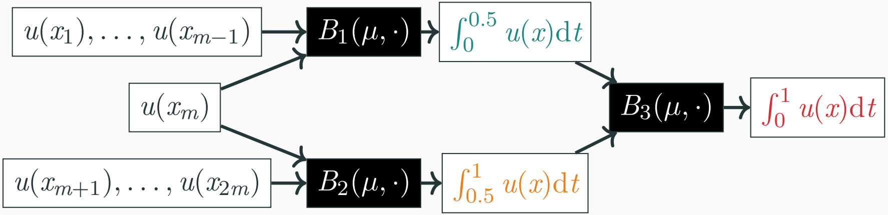
$$\int_0^1 u(x)\mathrm{d}x = \int_0^{0.5} u(x)\mathrm{d}x + \int_{0.5}^1 u(x)\mathrm{d}x$$

Observations $\{u(x_1), \ldots, u(x_{2m})\}$, where $u_1 = 0$, $u_m = 0.5$, $u_{2m} = 1$

Example: Split Integration

$$\int_0^1 u(x)\mathrm{d}x = \int_0^{0.5} u(x)\mathrm{d}x + \int_{0.5}^1 u(x)\mathrm{d}x$$

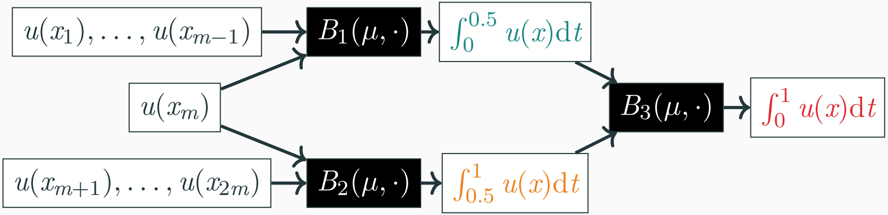
Observations $\{u(x_1), \dots, u(x_{2m})\}$, where $u_1 = 0$, $u_m = 0.5$, $u_{2m} = 1$



Example: Split Integration

$$\int_0^1 u(x)\mathrm{d}x = \int_0^{0.5} u(x)\mathrm{d}x + \int_{0.5}^1 u(x)\mathrm{d}x$$

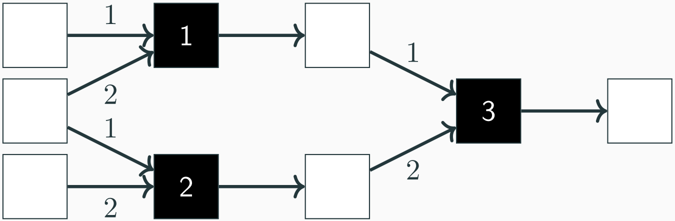
Observations $\{u(x_1), \ldots, u(x_{2m})\}$, where $u_1 = 0$, $u_m = 0.5$, $u_{2m} = 1$



When is the output of the pipeline Bayesian?

Dependence Graphs

The abstract structure of the graph allows us to establish a coherence condition

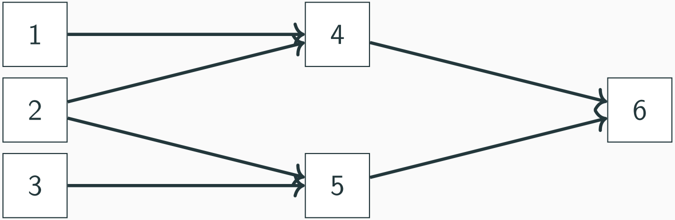


Pipeline

The dependency graph of a pipeline is obtained by deleting the method nodes and connecting their inputs directly to their outputs.

Dependence Graphs

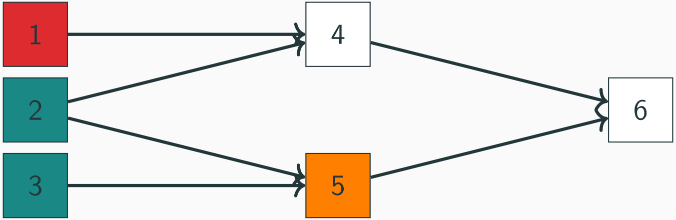
The abstract structure of the graph allows us to establish a coherence condition



Dependency Graph

The dependency graph of a pipeline is obtained by deleting the method nodes and connecting their inputs directly to their outputs.

Coherence



Definition

A prior is coherent for the dependency graph if Y_k is conditionally independent of Y_i given Y_j.

Here $i, j < k$, i are non-parent nodes and j are parent nodes.

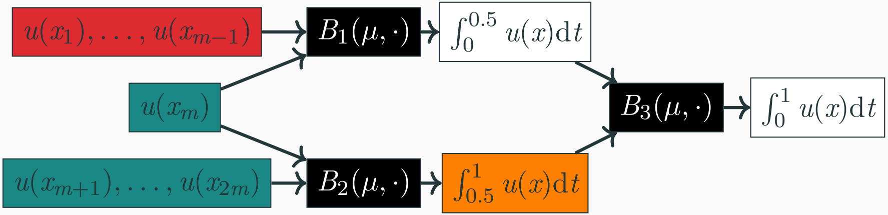
Bayesian Pipelines

Theorem

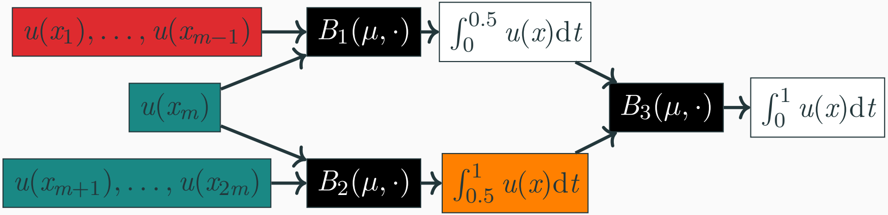
A pipeline is Bayesian for its output QoI if:

1. *The prior is coherent for the dependence graph.*
2. *The composite PNM are Bayesian.*

Split Integration: Coherence

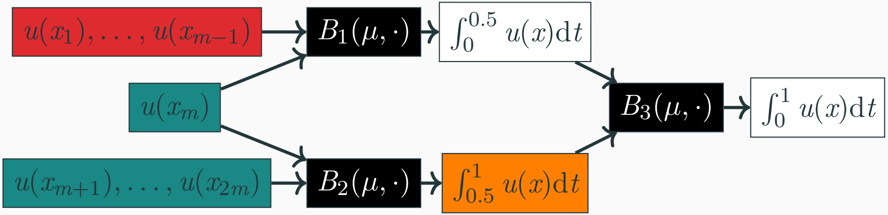


Split Integration: Coherence



Is $\int_{0.5}^1 u(x)\,\mathrm{d}t$ independent of $u(x_1), \ldots, u(x_{m-1})$?

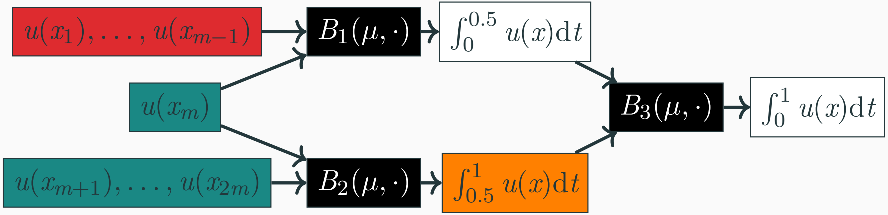
Split Integration: Coherence



Is $\int_{0.5}^{1} u(x)\, \mathrm{d}t$ independent of $u(x_1), \ldots, u(x_{m-1})$?

Sometimes - e.g. a Wiener process prior.

Split Integration: Coherence



Is $\int_{0.5}^1 u(x)\,\mathrm{d}t$ independent of $u(x_1), \dots, u(x_{m-1})$?

Sometimes - e.g. a Wiener process prior.

Sometimes not - e.g. if μ implies a Wiener process on $u^{(s)}(x)$.

Conclusions

Conclusions

We have seen...

- A method for approximately sampling from μ^a.
- Theoretical results proving asymptotic convergence of that sampler.
- Coherence conditions for composing Bayesian PNM into a Bayesian pipeline.

More to come

Numerical disintegration is highly inefficient compared to classical numerical methods (and even conjugate PNM).

More to come

Numerical disintegration is highly inefficient compared to classical numerical methods (and even conjugate PNM).

This is not intended to be a practical numerical method, but a framework for comparing other "approximate Bayesian" methods to the ideal.

More to come

Numerical disintegration is highly inefficient compared to classical numerical methods (and even conjugate PNM).

This is not intended to be a practical numerical method, but a framework for comparing other "approximate Bayesian" methods to the ideal.

Next steps:

- Make the algorithm more efficient?
- Explore more efficient approximations to the posterior

Thanks!

References I

J. Cockayne, C. Oates, T. Sullivan, and M. Girolami. Bayesian probabilistic numerical methods, 2017.

A. Müller. Integral probability metrics and their generating classes of functions. *Adv. in Appl. Probab.*, 29(2):429–443, 1997. ISSN 0001-8678. doi: 10.2307/1428011. URL `http://dx.doi.org/10.2307/1428011`.