Motivation: Computational Pipelines

Numerical analysis for the “drag and drop” era of computational pipelines:

![IBM High Performance Computation](image)

The sophistication and scale of modern computer models creates an urgent need to better understand the propagation and accumulation of numerical error within arbitrary - often large - pipelines of computation, so that “numerical risk” to end-users can be controlled.
Motivation: Solution of Poisson’s Equation

Consider numerical solution for $x \in \mathcal{X}$ of the Poisson equation

$$-\Delta x = f \quad \text{in } D$$

$$x = g \quad \text{on } \partial D$$

based on (noiseless) information of the form

$$A(x) = \begin{bmatrix}
-\Delta x(t_1) \\
\vdots \\
-\Delta x(t_m) \\
x(t_{m+1}) \\
\vdots \\
x(t_n)
\end{bmatrix} = \begin{bmatrix}
f(t_1) \\
\vdots \\
f(t_m) \\
x(t_{m+1}) \\
\vdots \\
x(t_n)
\end{bmatrix}, \quad \{t_i\}_{i=1}^m \in D, \quad \{t_i\}_{i=m+1}^d \in \partial D.$$

This is an ill-posed inverse problem and must be regularised.

The onus is on us to establish principled statistical foundations that are general.
Motivation: Solution of Poisson’s Equation

Consider numerical solution for $x \in \mathcal{X}$ of the Poisson equation

$$-\Delta x = f \quad \text{in } D$$

$$x = g \quad \text{on } \partial D$$

based on (noiseless) information of the form

$$A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_m) \\ x(t_{m+1}) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_m) \\ g(t_{m+1}) \\ \vdots \\ g(t_n) \end{bmatrix}, \quad \{t_i\}_{i=1}^m \in D, \quad \{t_i\}_{i=m+1}^d \in \partial D.$$

This is an ill-posed inverse problem and must be regularised.

The onus is on us to establish principled statistical foundations that are general.
Motivation: Solution of Poisson’s Equation

Consider numerical solution for $x \in \mathcal{X}$ of the Poisson equation

$$-\Delta x = f \quad \text{in } D$$

$$x = g \quad \text{on } \partial D$$

based on (noiseless) information of the form

$$A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_m) \\ x(t_{m+1}) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_m) \\ x(t_{m+1}) \\ \vdots \\ x(t_n) \end{bmatrix}, \quad \{t_i\}_{i=1}^m \in D, \quad \{t_i\}_{i=m+1}^d \in \partial D.$$

This is an ill-posed inverse problem and must be regularised.

The onus is on us to establish principled statistical foundations that are general.
The *Bayesian* approach, popularised in Stuart (2010), can be used:

- a *prior* measure $P_x$ is placed on $\mathcal{X}$
- a *posterior* measure $P_{x|a}$ is defined as the "restriction of $P_x$ to those functions $x \in \mathcal{X}$ for which

\[ A(x) = a \]

e.g. \[ A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a \]

is satisfied" (to be formalised).

⇒ **Principled and general** uncertainty quantification for numerical methods.
The Bayesian approach, popularised in Stuart (2010), can be used:

- a prior measure $P_x$ is placed on $\mathcal{X}$
- a posterior measure $P_{x \mid a}$ is defined as the “restriction of $P_x$ to those functions $x \in \mathcal{X}$ for which

$$A(x) = a$$

e.g. $A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a$

is satisfied” (to be formalised).

$\implies$ Principled and general uncertainty quantification for numerical methods.
The Research Agenda

Part I

1. First Job: Elicit the Abstract Structure
2. Second Job: Check Well-Defined, Existence and Uniqueness
3. Third Job: Characterise Optimal Information

Part II

4. Fourth Job: Algorithms to Access $P_{x|a}$
5. Fifth Job: Extend to Pipelines of Computation
First Job: Elicit the Abstract Structure
Abstractly, consider an unobserved state variable $x \in \mathcal{X}$ together with:

- A *quantity of interest*, denoted $Q(x) \in Q$.
- An *information operator*, denoted $x \mapsto A(x) \in A$.

**Examples:**

<table>
<thead>
<tr>
<th>Task</th>
<th>$Q(x)$</th>
<th>$A(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration</td>
<td>$\int x(t)\nu(dt)$</td>
<td>${x(t_i)}_{i=1}^n$</td>
</tr>
<tr>
<td>Optimisation</td>
<td>$\arg\max x(t)$</td>
<td>${x(t_i)}_{i=1}^n$</td>
</tr>
<tr>
<td>Solution of Poisson Eqn</td>
<td>$x(\cdot)$</td>
<td>${-\Delta x(t_i)}<em>{i=1}^m \cup {x(t_i)}</em>{i=m+1}$</td>
</tr>
</tbody>
</table>
Abstractly, consider an unobserved state variable $x \in \mathcal{X}$ together with:

- A *quantity of interest*, denoted $Q(x) \in Q$
- An *information operator*, denoted $x \mapsto A(x) \in A$.

Examples:

<table>
<thead>
<tr>
<th>Task</th>
<th>$Q(x)$</th>
<th>$A(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integration</td>
<td>$\int x(t) \nu(dt)$</td>
<td>${x(t_i)}_{i=1}^n$</td>
</tr>
<tr>
<td>Optimisation</td>
<td>$\arg\max x(t)$</td>
<td>${x(t_i)}_{i=1}^n$</td>
</tr>
<tr>
<td>Solution of Poisson Eqn</td>
<td>$x(\cdot)$</td>
<td>${-\Delta x(t_i)}<em>{i=1}^m \cup {x(t_i)}</em>{i=m+1}^n$</td>
</tr>
</tbody>
</table>
Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $M_\#\mu$ denote the “pushforward” measure, st $(M_\#\mu)(S) = \mu(M^{-1}(S))$.

<table>
<thead>
<tr>
<th></th>
<th>Classical Numerical Method</th>
<th>Probabilistic Numerical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>Assumed</td>
<td>$P_x \in \mathcal{P}_X$</td>
</tr>
<tr>
<td>Information</td>
<td>$a \in \mathcal{A}$</td>
<td>$a \in \mathcal{A}$</td>
</tr>
<tr>
<td>Output</td>
<td>$b(a) \in \mathcal{Q}$</td>
<td>$B(P_x, a) \in \mathcal{P}_Q$</td>
</tr>
</tbody>
</table>

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\#P_{x|a}$.
Let $\mathcal{P}$ denote the set of distributions on $\bullet$.

Let $M\#\mu$ denote the “pushforward” measure, st $(M\#\mu)(S) = \mu(M^{-1}(S))$.

<table>
<thead>
<tr>
<th></th>
<th>Classical Numerical Method</th>
<th>Probabilistic Numerical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>Assumed</td>
<td>$P_x \in \mathcal{P}_x$</td>
</tr>
<tr>
<td>Information</td>
<td>e.g. smoothness</td>
<td>$a \in \mathcal{A}$</td>
</tr>
<tr>
<td>Output</td>
<td>$b(a) \in \mathcal{Q}$</td>
<td>$B(P_x, a) \in \mathcal{P}_Q$</td>
</tr>
</tbody>
</table>

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q\#P_x|_a$. 
Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $M_\#\mu$ denote the “pushforward” measure, st $(M_\#\mu)(S) = \mu(M^{-1}(S))$.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Classical Numerical Method</th>
<th>Probabilistic Numerical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumed</td>
<td>e.g. smoothness</td>
<td>$P_x \in \mathcal{P}_\mathcal{X}$</td>
</tr>
<tr>
<td>Information</td>
<td>$a \in \mathcal{A}$</td>
<td>$a \in \mathcal{A}$</td>
</tr>
<tr>
<td>Output</td>
<td>$b(a) \in Q$</td>
<td>$B(P_x, a) \in \mathcal{P}_Q$</td>
</tr>
</tbody>
</table>

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_x|a$. 

Chris. J. Oates
Probabilistic Numerical Methods (I)
June 2017 @ ICERM
Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $M_\#\mu$ denote the “pushforward” measure, st $(M_\#\mu)(S) = \mu(M^{-1}(S))$.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Classical Numerical Method</th>
<th>Probabilistic Numerical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumed</td>
<td>e.g. smoothness</td>
<td>$P_x \in \mathcal{P}_x$</td>
</tr>
<tr>
<td>Information</td>
<td>$a \in \mathcal{A}$</td>
<td>$a \in \mathcal{A}$</td>
</tr>
<tr>
<td>Output</td>
<td>$b(a) \in \mathcal{Q}$</td>
<td>$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$</td>
</tr>
</tbody>
</table>

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_x|a$. 
Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $M_\# \mu$ denote the “pushforward” measure, st $(M_\# \mu)(S) = \mu(M^{-1}(S))$.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Classical Numerical Method</th>
<th>Probabilistic Numerical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumed</td>
<td>e.g. smoothness</td>
<td>$P_x \in \mathcal{P}_X$</td>
</tr>
<tr>
<td>Information</td>
<td>$a \in \mathcal{A}$</td>
<td>$a \in \mathcal{A}$</td>
</tr>
<tr>
<td>Output</td>
<td>$b(a) \in \mathcal{Q}$</td>
<td>$B(P_x, a) \in \mathcal{P}_Q$</td>
</tr>
</tbody>
</table>

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$. 
Let $\mathcal{P}$ denote the set of distributions on $\bullet$.

Let $M_#\mu$ denote the “pushforward” measure, st $(M_#\mu)(S) = \mu(M^{-1}(S))$.

<table>
<thead>
<tr>
<th></th>
<th>Classical Numerical Method</th>
<th>Probabilistic Numerical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs</strong></td>
<td>Assumed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>e.g. smoothness</td>
<td>$P_x \in \mathcal{P}_X$</td>
</tr>
<tr>
<td><strong>Information</strong></td>
<td>$a \in \mathcal{A}$</td>
<td>$a \in \mathcal{A}$</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>$b(a) \in \mathcal{Q}$</td>
<td>$B(P_x, a) \in \mathcal{P}_Q$</td>
</tr>
</tbody>
</table>

A Probabilistic Numerical Method is **Bayesian** iff $B(P_x, a) = Q_#P_x|_a$. 
<table>
<thead>
<tr>
<th>Method</th>
<th>QoI $Q(x)$</th>
<th>Information $A(x)$</th>
<th>Non-Bayesian PNMs</th>
<th>Bayesian PNMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrator</td>
<td>$\int x(t)\nu(dt)$</td>
<td>${x(t_i)}_{i=1}^n$</td>
<td>Approximate Bayesian Quadrature Methods [Osborne et al., 2012b,a, Gunter et al., 2014] Kong et al. [2003], Tan [2004], Kong et al. [2007]</td>
<td>Bayesian Quadrature [Diaconis, 1988, O’Hagan, 1991] Oates et al. [2016]</td>
</tr>
<tr>
<td></td>
<td>$\int f(t)x(dt)$</td>
<td>${t_i}_{i=1}^n$ s.t. $t_i \sim x$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\int x_1(t)x_2(dt)$</td>
<td>${(t_i, x_1(t_i))}_{i=1}^n$ s.t. $t_i \sim x_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimiser</td>
<td>$\arg\min x(t)$</td>
<td>${x(t_i)}_{i=1}^n$</td>
<td>Bayesian Optimisation [Mockus, 1989] Hennig and Kiefel [2013] Probabilistic Line Search [Mahsereci and Hennig, 2015] Probabilistic Bisection Algorithm [Horstein, 1963]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>${\nabla x(t_i)}_{i=1}^n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>${(x(t_i), \nabla x(t_i))}_{i=1}^n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>${[t_{min} &lt; t_i]}_{i=1}^n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>${[t_{min} &lt; t_i] + \text{error}}_{i=1}^n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ODE Solver</td>
<td>$x$</td>
<td>${\nabla x(t_i)}_{i=1}^n$</td>
<td>Skilling [1992]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$x(t_{\text{end}})$</td>
<td>$\nabla x + \text{rounding error}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>${\nabla x(t_i)}_{i=1}^n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDE Solver</td>
<td>$x$</td>
<td>${Dx(t_i)}_{i=1}^n$</td>
<td>Chkrebtii et al. [2016]</td>
<td>Probabilistic Meshless Methods [Owhadi, 2015a,b, Cockayne et al., 2016, Raissi et al., 2016]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Dx + \text{discretisation error}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Second Job: Check Well-Defined, Existence and Uniqueness
Limitations of existing Bayesian probabilistic numerical methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_x$
- 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:

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

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

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

But when “$p(a|x) = \delta(a - A(x))$”, the posterior $P_{x|a}$ will not be absolutely continuous wrt the prior $P_x$, so no Radon-Nikodym theorem!
Limitations of existing Bayesian probabilistic numerical methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_X$
- 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:

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

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

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

But when “$p(a|x) = \delta(a - A(x))$”, the posterior $P_{x|a}$ will not be absolutely continuous wrt the prior $P_x$, so no Radon-Nikodym theorem!
Well-Defined?

Limitations of existing Bayesian probabilistic numerical methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_x$
- 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:

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

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

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

But when “$p(a|x) = \delta(a - A(x))$”, the posterior $P_{x|a}$ will not be absolutely continuous wrt the prior $P_x$, so no Radon-Nikodym theorem!
Limitations of existing Bayesian probabilistic numerical methods:
- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_x$
- 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:

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

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

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

But when “$p(a|x) = \delta(a - A(x))$”, the posterior $P_{x|a}$ will not be absolutely continuous wrt the prior $P_x$, so no Radon-Nikodym theorem!
Well-Defined?

Limitations of existing Bayesian probabilistic numerical methods:

- Restriction to Gaussian prior distributions $P_x \in \mathcal{P}_X$
- 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:

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

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

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

But when \(p(a|x) = \delta(a - A(x))\), the posterior $P_{x|a}$ will not be absolutely continuous wrt the prior $P_x$, so no Radon-Nikodym theorem!
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
Borel-Kolmogorov paradox\textsuperscript{1}:

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

\footnote{Figures from Greg Gandenberger’s blog post}
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 $(Q, \Sigma_Q)$ 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 $\{P_{x|a}\}_{a \in \mathcal{A}} \subseteq \mathcal{P}_\mathcal{X}$ 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:) $P_x(f) = \int P_{x|a}(f) A \# P_x(da)$.
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 $(Q, \Sigma_Q)$ 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:) $P_x(f) = \int P_{x|a}(f) A \# P_x(da)$. 

Chris. J. Oates
Probabilistic Numerical Methods (I)
June 2017 @ ICERM 13 / 25
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 $(Q, \Sigma_Q)$ 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:) $P_x(f) = \int P_{x|a}(f) A\#P_x(da)$.
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 $\sigma$-algebra.
- Let $P_x \in \mathcal{P}_{\mathcal{X}}$ be Radon.
- Let $\Sigma_A$ be a countably generated $\sigma$-algebra that contains singletons $\{a\}$ for $a \in A$.

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

Thus 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 $a \in A$. 

Chris. J. Oates
Probabilistic Numerical Methods (I)
June 2017 @ ICERM
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 $\sigma$-algebra.
- Let $P_x \in \mathcal{P}_{\mathcal{X}}$ be Radon.
- Let $\Sigma_A$ be a countably generated $\sigma$-algebra that contains singletons $\{a\}$ for $a \in A$.

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

Thus 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 $a \in A$. 

Example: Solution of a Non-linear ODE

Consider Painlevé’s first transcendental:

\[
x''(t) = x(t)^2 - t, \quad t \in \mathbb{R}_+
\]

\[
x(0) = 0
\]

\[
t^{-1/2}x(t) \to 1 \text{ as } t \to \infty
\]

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 \( P_x \in \mathcal{P}_X \) 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:

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

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 \( P_x \in \mathcal{P}_X \) 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:

\[ x''(t) = x(t)^2 - t, \quad t \in \mathbb{R}_+ \]

\[ x(0) = 0 \]

\[ t^{-1/2}x(t) \to 1 \text{ as } t \to \infty \]

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 \( P_x \in \mathcal{P}_X \) 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

For this illustration the information, $n = 10$, is fixed.

[samples via *Numerical Disintegration* algorithm; see Part II]
Third Job: Characterise Optimal Information
Recall the contribution of Kadane and Wasilkowski [1985]:

Consider a classical numerical method \((A, b)\) with information operator \(A: \mathcal{X} \to \mathcal{A}\), such that \(A \in \Lambda\) for some set \(\Lambda\), and estimator \(b: \mathcal{A} \to \mathcal{Q}\). Let \(L: \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}\) be a loss function that is pre-specified. Then consider the minimal average case error

\[
\inf_{A \in \Lambda, b} \int L(b(A(x)), Q(x)) \, dP_x.
\]

The minimiser \(b(\cdot)\) is a non-randomised Bayes rule and the minimiser \(A\) is “optimal information” over \(\Lambda\), or optimal experimental design for this numerical task.

Generalisation of optimal information to probabilistic numerical methods?
Recall the contribution of Kadane and Wasilkowski [1985]:

Consider a classical numerical method \((A, b)\) with information operator \(A : \mathcal{X} \to \mathcal{A}\), such that \(A \in \Lambda\) for some set \(\Lambda\), and estimator \(b : \mathcal{A} \to \mathcal{Q}\). Let \(L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}\) be a loss function that is pre-specified. Then consider the minimal average case error

\[
\inf_{A \in \Lambda, b} \int L(b(A(x)), Q(x)) \, dP_x.
\]

The minimiser \(b(\cdot)\) is a non-randomised Bayes rule and the minimiser \(A\) is “optimal information” over \(\Lambda\), or optimal experimental design for this numerical task.

Generalisation of optimal information to probabilistic numerical methods?
For Bayesian probabilistic numerical methods $B(P_x, a) = Q#P_{x|a}$, optimal information is defined as

$$\arg \inf_{A \in \Lambda} \int \int L(Q#P_{x|A(x)}(\omega), Q(x))dP_x \, d\omega.$$ 

Important point: The Bayesian probabilistic numerical method output $Q#P_{x|a}$ will not in general be supported on the set of Bayes acts. This presents a non-trivial constraint on the risk set...
Optimal Information

Average Case Analysis $\leftrightarrow$ Bayesian Decision Theory $\leftrightarrow$ Bayesian Probabilistic Numerical Methods

- Risk set (classical)
- Risk set (BPNM)
- Optimal (BPNM)
- Contours of constant average risk

Chris. J. Oates
Probabilistic Numerical Methods (I)
June 2017 @ ICERM
Optimal Information

We have established the following (new) result:

Let \((Q, \langle \cdot, \cdot \rangle_Q)\) be an inner-product space with associated norm \(\| \cdot \|_Q\) and consider the canonical loss \(L(q, q') = \| q - q' \|^2_Q\). Then optimal information for Bayesian probabilistic numerical methods coincides with average-case optimal information.

The assumption is non-trivial:

Consider the following counter-example:

- \(\mathcal{X} = \{b, c, d, e\}\),
- \(Q(x) = 1[x = b]\),
- \(P_x\) uniform,
- \(A(x) = 1[x \in S]\), where we are allowed either \(S = \{b, c\}\) or \(\{b, c, d\}\),
- \(L(q, q') = 1[q \neq q']\).

Then average-case optimal information can be either \(S = \{b, c\}\) or \(\{b, c, d\}\). On the other hand, optimal information in the Bayesian probabilistic numerical context is just \(S = \{b, c\}\).
We have established the following (new) result:

Let \((Q, \langle \cdot, \cdot \rangle_Q)\) be an inner-product space with associated norm \(\| \cdot \|_Q\) and consider the canonical loss \(L(q, q') = \| q - q' \|_Q^2\). Then optimal information for Bayesian probabilistic numerical methods coincides with average-case optimal information.

The assumption is non-trivial:

Consider the following counter-example:

- \(\mathcal{X} = \{b, c, d, e\}\),
- \(Q(x) = 1[x = b]\),
- \(P_x\) uniform,
- \(A(x) = 1[x \in S]\), where we are allowed either \(S = \{b, c\}\) or \(\{b, c, d\}\),
- \(L(q, q') = 1[q \neq q']\).

Then average-case optimal information can be either \(S = \{b, c\}\) or \(\{b, c, d\}\). On the other hand, optimal information in the Bayesian probabilistic numerical context is just \(S = \{b, c\}\).
We have established the following (new) result:

Let $(Q, \langle \cdot, \cdot \rangle_Q)$ be an inner-product space with associated norm $\| \cdot \|_Q$ and consider the canonical loss $L(q, q') = \| q - q' \|_Q^2$. Then optimal information for Bayesian probabilistic numerical methods coincides with average-case optimal information.

The assumption is non-trivial:

Consider the following counter-example:

- $\mathcal{X} = \{b, c, d, e\}$,
- $Q(x) = 1[x = b]$,
- $P_x$ uniform,
- $A(x) = 1[x \in S]$, where we are allowed either $S = \{b, c\}$ or $\{b, c, d\}$,
- $L(q, q') = 1[q \neq q']$.

Then average-case optimal information can be either $S = \{b, c\}$ or $\{b, c, d\}$. On the other hand, optimal information in the Bayesian probabilistic numerical context is just $S = \{b, c\}$. 
Conclusion
In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, $P_\mathcal{X}$ Radon, $\Sigma_\mathcal{A}$ countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
Conclusion

In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, $P_x$ radon, $\Sigma_A$ countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions (\(\mathcal{X}\) metric space, \(P_\mathcal{X}\) radon, \(\Sigma_\mathcal{A}\) countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, $P_x$ radon, $\Sigma_A$ countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, $P_x$ radon, $\Sigma_A$ countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, $P_X$ radon, $\Sigma_A$ countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- Bayesian probabilistic numerical methods (BPNM) are well-defined under weak conditions ($\mathcal{X}$ metric space, $P_X$ radon, $\Sigma_A$ countably generated).
- Optimal information for BPNM is not always equivalent to optimal information in Average Case Analysis.

Full details (Parts I and II) can be found in the preprint:


Thank you for your attention!
References


