Probabilistic Numerical Computation: A Role for (Bayesian) Statisticians in Numerical Analysis?

Mark Girolami

Department of Mathematics
Imperial College London

The Alan Turing Institute

Lloyds Register Foundation Programme on Data-Centric Engineering

https://iconicmath.org/

Bayes on the Beach 2017
Probabilistic Numerical Computation
Consider the following school boy and girl differential equation

\[ \frac{du}{dt} = \theta u, \quad u(t = 0) = 1. \]

This is the simplest example model used to describe Malthusian population growth e.g. bacterial growth and radioactive decay. Simplest representation of compound interest in finance.

Every school boy and girl knows the solution:

\[ u(t; \theta) = \exp(\theta t) \]

Despite the function \( u(t; \theta) \) being implicitly defined it is a fully deterministic object.

Given the initial value then \( u(t; \theta) \) at any point in the future is fully determined.
Consider the following school boy and girl differential equation

\[ \frac{du}{dt} = \theta u, \quad u(t = 0) = 1. \]

This is the simplest example model used to describe Malthusian population growth e.g. bacterial growth and radioactive decay. Simplest representation of compound interest in finance.

Every school boy and girl knows the solution:

\[ u(t; \theta) = \exp(\theta t) \]

Despite the function \( u(t; \theta) \) being implicitly defined it is a fully deterministic object.

Given the initial value then \( u(t; \theta) \) at any point in the future is fully determined.
Consider the following school boy and girl differential equation

$$\frac{du}{dt} = \theta u, \quad u(t = 0) = 1.$$ 

This is the simplest example model used to describe Malthusian population growth e.g. bacterial growth and radioactive decay. Simplest representation of compound interest in finance.

Every school boy and girl knows the solution:

$$u(t; \theta) = \exp(\theta t)$$

Despite the function $u(t; \theta)$ being implicitly defined it is a fully deterministic object.

Given the initial value then $u(t; \theta)$ at any point in the future is fully determined.
Consider the following school boy and girl differential equation

\[ \frac{du}{dt} = \theta u, \quad u(t = 0) = 1. \]

This is the simplest example model used to describe Malthusian population growth e.g. bacterial growth and radioactive decay. Simplest representation of compound interest in finance.

Every school boy and girl knows the solution:

\[ u(t; \theta) = \exp(\theta t) \]

Despite the function \( u(t; \theta) \) being implicitly defined it is a fully deterministic object.

Given the initial value then \( u(t; \theta) \) at any point in the future is fully determined.
Consider the following school boy and girl differential equation

\[ \frac{du}{dt} = \theta u, \quad u(t = 0) = 1. \]

This is the simplest example model used to describe Malthusian population growth e.g. bacterial growth and radioactive decay. Simplest representation of compound interest in finance.

Every school boy and girl knows the solution:

\[ u(t; \theta) = \exp(\theta t) \]

Despite the function \( u(t; \theta) \) being implicitly defined it is a fully deterministic object.

Given the initial value then \( u(t; \theta) \) at any point in the future is fully determined.
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability.

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$.

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma)$ $\Rightarrow$ $u(t; \theta) \sim \log \mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object.

Uncertainty can also enter by being unable to solve the differential equation analytically.

What if the differential equation cannot be solved analytically?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability.

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$.

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object.

Uncertainty can also enter by being unable to solve the differential equation analytically.

What if the differential equation cannot be solved analytically?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability.

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$.

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object.

Uncertainty can also enter by being unable to solve the differential equation analytically.

What if the differential equation cannot be solved analytically?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter. This immediately introduces uncertainty into our deterministic world. Our uncertainty in $\theta$ can be described using the calculus of probability

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \Rightarrow u(t; \theta) \sim \log \mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object

Uncertainty can also enter by being unable to solve the differential equation analytically

What if the differential equation cannot be solved analytically?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$

$$\text{Uncertainty } \theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$$

With uncertainty our deterministic object becomes a probabilistic object

Uncertainty can also enter by being unable to solve the differential equation analytically

What if the differential equation cannot be solved analytically ?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability.

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object.

Uncertainty can also enter by being unable to solve the differential equation analytically.

What if the differential equation cannot be solved analytically?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object

Uncertainty can also enter by being unable to solve the differential equation analytically

What if the differential equation cannot be solved analytically?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object

Uncertainty can also enter by being unable to solve the differential equation analytically

What if the differential equation cannot be solved analytically ?
But wait.....

The rate parameter $\theta$ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in $\theta$ can be described using the calculus of probability

This uncertainty in $\theta$ propagates and induces uncertainty in $u(t; \theta)$

**Uncertainty** $\theta \sim \mathcal{N}(\mu, \sigma) \quad \Rightarrow \quad u(t; \theta) \sim \log\mathcal{N}(\mu t, \sigma t)$

With uncertainty our deterministic object becomes a probabilistic object

Uncertainty can also enter by being unable to solve the differential equation analytically

What if the differential equation cannot be solved analytically?
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a Known Unknown

For a general differential equation $\dot{u} = f(u; \theta)$ then the Euler method gives

$$U_{n+1} = U_n + h f(U_n; \theta)$$

For our school boy example with $U_0 = 1$ then

$$U_{n+1} = U_n + h \theta U_n = (1 + h \theta)^n$$

$$\theta \sim \log\mathcal{N}(\mu, \sigma)$$

$$\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}$$

$$\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}$$

The deterministic numerical procedure contributes further to uncertainty
The numerical procedure is now an inference procedure
Defines a measure from which approximate solutions can be drawn
Must resort to numerical methods to access approximations to the solution.

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a **Known Unknown**.

For a general differential equation $\dot{u} = f(u; \theta)$ then the Euler method gives

$$U_{n+1} = U_n + hf(U_n; \theta)$$

For our school boy example with $U_0 = 1$ then

$$U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n$$

$$\theta \sim \log\mathcal{N}(\mu, \sigma)$$

$$\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}$$

$$\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}$$

The deterministic numerical procedure contributes further to uncertainty.

The numerical procedure is now an inference procedure.

 Defines a measure from which approximate solutions can be drawn.
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a Known Unknown

For a general differential equation $\dot{u} = f(u; \theta)$ then the Euler method gives

$$U_{n+1} = U_n + hf(U_n; \theta)$$

For our school boy example with $U_0 = 1$ then

$$U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n$$

$$\theta \sim \log\mathcal{N}(\mu, \sigma)$$

$$\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}$$

$$\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}$$

The deterministic numerical procedure contributes further to uncertainty
The numerical procedure is now an inference procedure
Defines a measure from which approximate solutions can be drawn
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a Known Unknown

For a general differential equation $\dot{u} = f(u; \theta)$ then the Euler method gives

$$U_{n+1} = U_n + hf(U_n; \theta)$$

For our school boy example with $U_0 = 1$ then

$$U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n$$

$$\theta \sim \log\mathcal{N}(\mu, \sigma)$$

$$\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}$$

$$\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}$$

The deterministic numerical procedure contributes further to uncertainty

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a **Known Unknown**

For a general differential equation $\dot{u} = f(u; \theta)$ then the Euler method gives

$$U_{n+1} = U_n + hf(U_n; \theta)$$

For our school boy example with $U_0 = 1$ then

$$U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n$$

$$\theta \sim \log\mathcal{N}(\mu, \sigma)$$

$$\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}$$

$$\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}$$

The deterministic numerical procedure contributes further to uncertainty

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a Known Unknown

For a general differential equation $\dot{u} = f(u; \theta)$ then the Euler method gives

$$U_{n+1} = U_n + hf(U_n; \theta)$$

For our school boy example with $U_0 = 1$ then

$$U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n$$

$$\theta \sim \log\mathcal{N}(\mu, \sigma)$$

$$\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}$$

$$\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}$$

The deterministic numerical procedure contributes further to uncertainty

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a **Known Unknown**

For a general differential equation \( \dot{u} = f(u; \theta) \) then the Euler method gives

\[
U_{n+1} = U_n + hf(U_n; \theta)
\]

For our school boy example with \( U_0 = 1 \) then

\[
U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n
\]

\[
\theta \sim \log\mathcal{N}(\mu, \sigma)
\]

\[
\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\} \quad \mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}
\]

The deterministic numerical procedure contributes further to uncertainty

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Must resort to numerical methods to access approximations to the solution

We now have an additional layer of epistemic uncertainty in that the implicit function is unknown - we have a Known Unknown

For a general differential equation \( \dot{u} = f(u; \theta) \) then the Euler method gives

\[
U_{n+1} = U_n + hf(U_n; \theta)
\]

For our school boy example with \( U_0 = 1 \) then

\[
U_{n+1} = U_n + h\theta U_n = (1 + h\theta)^n
\]

\[
\theta \sim \log \mathcal{N}(\mu, \sigma)
\]

\[
\mathbb{E}\{U_n\} = \sum_{k=0}^{n-1} \binom{n-1}{k} h^k \mathbb{E}\{\theta^k\}
\]

\[
\mathbb{E}\{U_n^2\} = \sum_{k=0}^{2(n-1)} \binom{2(n-1)}{k} h^k \mathbb{E}\{\theta^k\}
\]

The deterministic numerical procedure contributes further to uncertainty
The numerical procedure is now an inference procedure
Defines a measure from which approximate solutions can be drawn
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that $\theta$ takes we only have the *Known Unknown* to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

$$e_{n+1} = e_n + h[u(t_n) - U_n] + R$$

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown

Subjectivist Probability - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that $\theta$ takes we only have the **Known Unknown** to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

$$e_{n+1} = e_n + h[u(t_n) - U_n] + R$$

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown.

Subjectivist Probability - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that $\theta$ takes we only have the Known Unknown to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

\[ e_{n+1} = e_n + h[u(t_n) - U_n] + R \]

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown

Subjectivist Probability - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that $\theta$ takes we only have the Known Unknown to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

\[ e_{n+1} = e_n + h[u(t_n) - U_n] + R \]

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown

Subjectivist Probability - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that \( \theta \) takes we only have the **Known Unknown** to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

\[
e_{n+1} = e_n + h[u(t_n) - U_n] + R
\]

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown.

**Subjectivist Probability** - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure.

Defines a measure from which approximate solutions can be drawn.
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that $\theta$ takes we only have the Known Unknown to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

$$e_{n+1} = e_n + h[u(t_n) - U_n] + R$$

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown

Subjectivist Probability - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Now then in the *unlikely* situation where we have complete knowledge of the initial value and value that $\theta$ takes we only have the Known Unknown to deal with.

Now everything is fully deterministic in the computation of our approximation. The evolution of the error is fully determined.

$$e_{n+1} = e_n + h[u(t_n) - U_n] + R$$

Nothing stochastic or random about this.

However we cannot compute the deterministic error or its equation of evolution - it is unknown

Subjectivist Probability - De Finetti, Ramsey, Jeffreys, Berger, Bernardo

The numerical procedure is now an inference procedure

Defines a measure from which approximate solutions can be drawn
Probabilistic Numerical Computation
In any prolonged computation it is generally assumed that the accumulated effect of roundoff errors is in some sense statistical. The purpose of this paper is to give precise descriptions of certain probabilistic models for roundoff error, and then to describe a series of experiments for testing the validity of these models. It is concluded that the models are in general very good. Discrepancies are both rare and mild. The test techniques can also be used to experiment with various types of special arithmetic.
Tests of Probabilistic Models for Propagation of Roundoff Errors

T. E. HULL, University of Toronto; J. R. SWENSEN, New York University (Ed: J. Traub)

In any prolonged computation it is generally assumed that the accumulated effect of roundoff errors is in some sense statistical. The purpose of this paper is to give precise descriptions of certain probabilistic models for roundoff error, and then to describe a series of experiments for testing the validity of these models. It is concluded that the models are in general very good. Discrepancies are both rare and mild. The test techniques can also be used to experiment with various types of special arithmetic.
Question: “Is numerical computation a statistical inference problem?”
GAUSSIAN MEASURE IN HILBERT SPACE
AND APPLICATIONS IN NUMERICAL ANALYSIS

F. M. LARKIN

Abstract. 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.
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.
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.
What is Probabilistic Numerics?¹

Definition (Probabilistic Numerics)

Probabilistic Numerics models the function uncertainty and propagates a probabilistic description of this error through subsequent computations.

¹[Hennig, Osborne, Girolami., 2015]
What is Probabilistic Numerics?¹

**Definition (Probabilistic Numerics)**

Probabilistic Numerics models the function uncertainty and propagates a probabilistic description of this error through subsequent computations.

- Produces probability measures over all unknowns.
- Structure in residuals can be propagated through later computations.
- Analysis of variance to determine the computational sticking points.
- New perspective leads to design of new algorithms.
- Safeguards against unwarranted optimism for decision making.

¹[Hennig, Osborne, Girolami., 2015]
Probabilistic numerics and uncertainty in computations

Philipp Hennig\textsuperscript{1}, Michael A. Osborne\textsuperscript{2} and Mark Girolami\textsuperscript{3}

\textsuperscript{1}Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Tübingen, Germany
\textsuperscript{2}Department of Engineering Science, University of Oxford, Oxford, UK
\textsuperscript{3}Department of Statistics, University of Warwick, Warwick, UK

We deliver a call to arms for probabilistic numerical methods: algorithms for numerical tasks, including linear algebra, integration, optimization and solving differential equations, that return uncertainties in their calculations. Such uncertainties, arising from the loss of precision induced by numerical calculation with limited time or hardware, are important for much contemporary science and industry. Within applications such as climate science and astrophysics, the need to make decisions on the basis of computations with large and complex data have led to a renewed focus on the management of numerical uncertainty. We describe how several
Differential Equations
Motivation - Data Driven Engineering
Motivation - Data Informed Medical and Life Sciences
Motivation - Computational Social Science

Burglaries

Drugs

Traffic

Violence
A “widely used” linear PDE. Given $g$, $\kappa$, $b$ find $u$

$$-\nabla \cdot (\kappa(x) \nabla u(x)) = g(x) \quad \text{in } D$$

$$u(x) = b(x) \quad \text{on } \partial D$$

For general $D$, $u(x)$ this cannot be solved analytically. The majority of PDE solvers produce an approximation like:

$$\hat{u}(x) = \sum_{i=1}^{N} w_i \phi_i(x)$$

We want to quantify the error from finite $N$ probabilistically.
PN for PDEs

A “widely used” linear PDE. Given $g$, $\kappa$, $b$ find $u$

$$-\nabla \cdot (\kappa(x) \nabla u(x)) = g(x) \quad \text{in } D$$

$$u(x) = b(x) \quad \text{on } \partial D$$

For general $D$, $u(x)$ this cannot be solved analytically.

The majority of PDE solvers produce an approximation like:

$$\hat{u}(x) = \sum_{i=1}^{N} w_i \phi_i(x)$$

We want to quantify the error from finite $N$ probabilistically.
History of Probabilistic Numerical Methods

Bayesian Numerical Analysis

P. DIACONIS, Stanford University.

Statistical Decision Theory and Related Topics IV, 1, 163-175, 1988.

Seeing standard procedures emerge from the Bayesian approach may convince readers the argument isn’t so crazy after all. The examples suggest the following program: Take standard numerical analysis procedures and see if they are Bayes (or admissible, or minimax). [...] The Bayesian approach yields more than the Bayes rule; it yields a posterior distribution. This can be used to give confidence sets as in Wahba (1983).
History of Probabilistic Numerical Methods

Bayesian Numerical Analysis

P. DIACONIS, Stanford University.


Seeing standard procedures emerge from the Bayesian approach may convince readers the argument isn’t so crazy after all. The examples suggest the following program: Take standard numerical analysis procedures and see if they are Bayes (or admissible, or minimax). [...] The Bayesian approach yields more than the Bayes rule; it yields a posterior distribution. This can be used to give confidence sets as in Wahba (1983).
Inverse Problem: Given partial information of $g, b, u$ find $\kappa$

\[-\nabla \cdot (\kappa(x)\nabla u(x)) = g(x) \quad \text{in } D\]

\[u(x) = b(x) \quad \text{on } \partial D\]
Inverse Problem: Given partial information of $g, b, u$ find $\kappa$

\[-\nabla \cdot (\kappa(x)\nabla u(x)) = g(x) \quad \text{in } D\]
\[u(x) = b(x) \quad \text{on } \partial D\]
Inverse Problem: Given partial information of $g, b, u$ find $\kappa$

$$-\nabla \cdot (\kappa(x) \nabla u(x)) = g(x) \quad \text{in } D$$

$$u(x) = b(x) \quad \text{on } \partial D$$
PN for PDEs

Inverse Problem: Given partial information of $g, b, u$ find $\kappa$

$$-
abla \cdot (\kappa(x) \nabla u(x)) = g(x) \quad \text{in } D$$
$$u(x) = b(x) \quad \text{on } \partial D$$

*Bayesian* Inverse Problem:

We want to account for an inaccurate forward solver in the inverse problem.
Forward Problem
Abstract Formulation

\[ Au(x) = g(x) \quad \text{in } D \]

**Forward** inference procedure:

\[ u \sim \Pi_u \quad \text{“Data” } Au(x_i) = g(x_i) \quad u|g \sim \Pi_u^g \]
Use a Gaussian Process prior $u \sim \Pi_u = \mathcal{GP}(0, k)$. Assuming linearity, the posterior $\Pi^g_u$ is available in closed-form\(^2\).

$$\Pi^g_u \sim \mathcal{GP}(m_1, \Sigma_1)$$

$$m_1(x) = \bar{A}K(x, X) \left[ \bar{A}A\bar{K}(X, X) \right]^{-1} g$$

$$\Sigma_1(x, x') = k(x, x') - \bar{A}K(x, X) \left[ \bar{A}A\bar{K}(X, X) \right]^{-1} A\bar{K}(X, x')$$

$\bar{A}$ the adjoint of $A$

**Observation:** The mean function is the same as in symmetric collocation!

\(^2\)Larkin 1972, [Cockayne et al., 2016, Owhadi, 2014]
Use a Gaussian Process prior \( u \sim \Pi_u = \mathcal{GP}(0, k) \). **Assuming linearity**, the posterior \( \Pi^g_u \) is available in closed-form\(^2\).

\[
\Pi^g_u \sim \mathcal{GP}(m_1, \Sigma_1)
\]

\[
m_1(x) = \bar{A}K(x, X) [A\bar{A}K(X, X)]^{-1} g
\]

\[
\Sigma_1(x, x') = k(x, x') - \bar{A}K(x, X) [A\bar{A}K(X, X)]^{-1} A K(X, x')
\]

\( \bar{A} \) the adjoint of \( A \)

**Observation:** The mean function is the same as in symmetric collocation!

---

\(^2\)Larkin 1972, [Cockayne et al., 2016, Owhadi, 2014]
Theoretical Results

Theorem (Forward Contraction)

For a ball $B_\varepsilon(u_0)$ of radius $\varepsilon$ centered on the true solution $u_0$ of the PDE, we have

$$1 - \prod_{u} [B_\varepsilon(u_0)] = O \left( \frac{h^{2\beta - 2\rho - d}}{\varepsilon} \right)$$

- $h$ the fill distance
- $\beta$ the smoothness of the prior
- $\rho < \beta - d/2$ the order of the PDE
- $d$ the input dimension
Toy Example

\[-\nabla^2 u(x) = g(x) \quad x \in (0, 1)\]
\[u(x) = 0 \quad x = 0, 1\]

To associate with the notation from before...

\[\Pi_u \sim \mathcal{GP}(0, k(x, y))\]
\[\mathcal{A} = -\frac{d^2}{dx^2} \quad \bar{\mathcal{A}} = -\frac{d^2}{dy^2}\]
Forward problem: posterior samples

\[ g(x) = \sin(2\pi x) \]
Forward problem: convergence

(a) Mean error from truth

(b) Trace of posterior covariance

Figure: Convergence
Inverse Problem
Recap

$$-\nabla \cdot (\kappa(x) \nabla u(x)) = g(x) \quad \text{in } D$$

$$u(x) = b(x) \quad \text{on } \partial D$$

Now we need to incorporate the forward posterior measure $\Pi^g_u$ into the posterior measure for the inverse problem, $\kappa$
Incorporation of Forward Measure

Assuming the data in the inverse problem is:

\[ y_i = u(x_i) + \xi_i \quad i = 1, \ldots, n \]
\[ \xi \sim N(0, \Gamma) \]

implies the standard likelihood:

\[ p(y|\kappa, u) \sim N(y; u, \Gamma) \]

But we don’t know \( u \)

Marginalise the forward posterior \( \prod^g_u \) to obtain a “PN” likelihood:

\[ p_{PN}(y|\kappa) \propto \int p(y|\kappa, u) d\prod^g_u \]
\[ \sim N(y; m_1, \Gamma + \Sigma_1) \]
Incorporation of Forward Measure

Assuming the data in the inverse problem is:

\[ y_i = u(x_i) + \xi_i \quad i = 1, \ldots, n \]

\[ \xi \sim N(0, \Gamma) \]

implies the standard likelihood:

\[ p(y|\kappa, u) \sim N(y; u, \Gamma) \]

But we don’t know \( u \)
Marginalise the forward posterior \( \Pi_u^g \) to obtain a “PN” likelihood:

\[ p_{\text{PN}}(y|\kappa) \propto \int p(y|\kappa, u) d\Pi_u^g \]

\[ \sim N(y; m_1, \Gamma + \Sigma_1) \]
Inverse Problem

Denote by \( \Pi^y_\kappa \) the posterior for \( \kappa \) from likelihood \( p \), and by \( \Pi^y_{\kappa,PN} \) the posterior for \( \kappa \) from likelihood \( p_{PN} \).

**Theorem (Inverse Contraction)**

Assume \( \Pi^y_\kappa \to \delta(\kappa_0) \) as \( n \to \infty \).

Then \( \Pi^y_{\kappa,PN} \to \delta(\kappa_0) \) provided

\[
h = o\left(n^{-1/(\beta-\rho-d/2)}\right)
\]
Inverse Problem

Back to the Toy Example

\[-\nabla \cdot (\kappa \nabla u(x)) = \sin(2\pi x) \quad x \in (0, 1)\]
\[u(x) = 0 \quad x = 0, 1\]

Infer $\kappa \in \mathbb{R}^+$; data generated for $\kappa = 1$ at $x = 0.25, 0.75$. Corrupted with independent Gaussian noise $\xi \sim N(0, 0.01^2)$.
Posterior Distributions for different numbers of design points.

<table>
<thead>
<tr>
<th>Design Points</th>
<th>Probabilistic</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Posterior Distributions for different numbers of design points.
Electrical Impedance Tomography

A medical imaging technique. Goal: reconstruct \textit{interior conductivity field} of a patient, to detect tumors.
Inverse Problem

Electrical Impedance Tomography

A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.

Many patterns of current $c_{ij}, j = 1, \ldots, N_c$ injected through boundary electrodes $t_i^{obs}, i = 1, \ldots, N_s$.
Electrical Impedance Tomography

A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.

Resulting voltage measured: $y_i = x(t_i^{obs}) - x(t_{ref}) + \epsilon_i$
Electrical Impedance Tomography

A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.

Governing equations are essentially Darcy’s law:

\[- \nabla \cdot (\theta(t) \nabla x(t)) = 0 \quad t \in D\]

\[\theta(t_i^{\text{obs}}) \frac{\partial x}{\partial n}(t_i^{\text{obs}}) = c_{ij} \quad i = 1, \ldots, N_S\]
Experimental Set-Up

Experiments due to Isaacson 2004.

- Tank filled with saline.
- Three targets:
  - “Heart shaped”: higher conductivity.
  - “Lung shaped”: lower conductivity.
- 32 equally spaced electrodes.
- Simultaneously stimulated for 31 different stimulation patterns.
Experimental Set-Up

Experiments due to Isaacson 2004.

- Tank filled with saline.
- Three targets:
  - “Heart shaped”: higher conductivity.
  - “Lung shaped”: lower conductivity.
- 32 equally spaced electrodes.
- Simultaneously stimulated for 31 different stimulation patterns.
A Hard Problem...

- High dimensional (992) observations.
- Observations are only of the boundary - weak information.
- Target $\theta(\cdot)$ is infinite-dimensional.
- The “ideal” likelihood $\mathcal{L}(\theta; y)$ requires exact solution of the PDE.

Posterior obtained using the PN likelihood

$$
\mathcal{L}_n(\theta; y) \propto \int p(y|\theta, x) dP_x|_a
$$

$$
y|\theta \sim N(m_1, \Gamma + \Sigma_1).
$$

Focus on varying the number $n$ of points in $T = \{t_i\}_{i=1}^n$ that are used.

Computation facilitated with Markov chain Monte Carlo, based on the preconditioned Crank-Nicholson proposal.
A Hard Problem...

- High dimensional (992) observations.
- Observations are only of the boundary - weak information.
- Target $\theta(\cdot)$ is infinite-dimensional.
- The “ideal” likelihood $\mathcal{L}(\theta; y)$ requires exact solution of the PDE.

Posteriors obtained using the PN likelihood

$$
\mathcal{L}_n(\theta; y) \propto \int p(y|\theta, x) dP_x|_a
$$

$$
\implies y|\theta \sim N(m_1, \Gamma + \Sigma_1).
$$

Focus on varying the number $n$ of points in $T = \{t_i\}_{i=1}^n$ that are used.

Computation facilitated with Markov chain Monte Carlo, based on the preconditioned Crank-Nicholson proposal.
Recovered Fields

Posterior means $m(t) = \mathbb{E}_y[\theta(t)]:$

(a) $n = 96$

(b) $n = 127$

(c) $n = 165$
Variance Analysis

Ratio of (pointwise) posterior variance $v(t) = \nabla_y[\theta(t)]$ computed from the PN posterior based on $L_n$ and the “standard” posterior based on $\hat{L}_N$ with $n = N = 96$:
Allen–Cahn

A prototypical nonlinear model.

\[-\theta \nabla^2 u(x) + \theta^{-1} (u(x)^3 - u(x)) = 0 \quad x \in (0, 1)^2\]

\[u(x) = 1 \quad x_1 \in \{0, 1\}; 0 < x_2 < 1\]

\[u(x) = -1 \quad x_2 \in \{0, 1\}; 0 < x_1 < 1\]

Goal: infer $\theta$
Allen–Cahn: Forward Solutions

Nonlinear PDE - non-GP posterior sampling schemes required, see [Cockayne et al., 2016].
Allen–Cahn: Inverse Problem

(a) Probabilistic.

(b) Standard
Illustrative Application - Integral over Manifold

![Diagram of a camera and environment map with vectors and labels: $L_i(\omega_o)$, $L_e(\omega_o)$, $\omega_o$, $\omega_i$, and $S^2$. The diagram shows the relationship between the camera, object, and environment map.]
Integrals Over Manifolds

\[ L_0(\omega_o) = L_e(\omega_o) + \int_{S^2} L_i(\omega_i) \rho(\omega_i, \omega_o) [\omega_i \cdot \mathbf{n}]_+ d\pi(\omega_i) \]

- \( L_0(\omega_o) \) = outgoing radiance
- \( L_e(\omega_o) \) = amount of light emitted by the object itself
- \( L_i(\omega_i) \) = amount of light reaching object from direction \( \omega_i \)
- \( \rho \) = bidirectional reflectance distribution function
- \( \pi \) = uniform distribution on \( S^2 \)

To be computed
- for each pixel, and
- for each RGB channel.
Integrals Over Manifolds

\[ L_o(\omega_o) = L_e(\omega_o) + \int_{S^2} L_i(\omega_i) \rho(\omega_i, \omega_o) [\omega_i \cdot n] + d\pi(\omega_i) \]

- \( L_o(\omega_o) \) = outgoing radiance
- \( L_e(\omega_o) \) = amount of light emitted by the object itself
- \( L_i(\omega_i) \) = amount of light reaching object from direction \( \omega_i \)
- \( \rho \) = bidirectional reflectance distribution function
- \( \pi \) = uniform distribution on \( S^2 \)

To be computed

- for each pixel, and
- for each RGB channel.
The Problem

Let $f$ be continuous and square-integrable, $\Pi$ be a probability measure and $\mathcal{X} \subseteq \mathbb{R}^d$. We want to compute (numerically):

$$\Pi[f] = \int_{\mathcal{X}} f \, d\Pi \approx \sum_{i=1}^{n} w_i f(x_i) = \hat{\Pi}[f] \quad (1)$$

High numerical uncertainty when $f$ is expensive or $n$ is small!
The Problem

Let $f$ be continuous and square-integrable, $\Pi$ be a probability measure and $\mathcal{X} \subseteq \mathbb{R}^d$. We want to compute (numerically):

$$\Pi[f] = \int_{\mathcal{X}} f \, d\Pi \approx \sum_{i=1}^{n} w_i f(x_i) = \hat{\Pi}[f]$$  \hspace{0.5cm} (1)

High numerical uncertainty when $f$ is expensive or $n$ is small!

**Probabilistic Numerics Solution:** Bayesian Quadrature\(^3\) (BQ) makes use of *prior information* about $f$ to guide our choice of $\{x_i, w_i\}_{i=1}^n$ (through a choice of function space/RKHS).

Measure on Integral push-forward of measure on function.

\(^3\)\cite{O'Hagan1991, RasmussenGhahramani2002, BriolEtAl2015a, BriolEtAl2015b}
Sketch of Bayesian Quadrature

\[ \mathbb{E}_n[\Pi[f]] = \hat{\Pi}_{\text{BQ}} = \Pi[k(\cdot, X)]K^{-1}f \]
\[ \nabla_n[\Pi[f]] = \Pi \Pi[k(\cdot, \cdot)] - \Pi[k(\cdot, X)]K^{-1}\Pi[k(X, \cdot)]. \]
**Theory for Bayesian Quadrature**

We consider Sobolev spaces, which are RKHS $\mathcal{H}^\alpha$ of varying levels of smoothness $\alpha$, which consist of functions in $L_2$ with associated inner product:

$$\langle f, g \rangle_{\mathcal{H}^\alpha} := \sum_{m=0}^{\alpha} \left\langle \frac{d^m f}{dx^m}, \frac{d^m g}{dx^m} \right\rangle_{L_2}$$

and finite norm $\|f\|_{\mathcal{H}^\alpha(\Pi)} := \langle f, f \rangle_{\mathcal{H}^\alpha}^{1/2}$.

We study the performance of the method in terms of worst-case error:

$$e(\hat{\Pi}; \Pi, \mathcal{H}) = \sup_{f: \|f\|_{\mathcal{H}} \leq 1} |\Pi[f] - \hat{\Pi}[f]|.$$
Theory for Bayesian Quadrature

Theorem (BQ in Sobolev spaces [Briol et al., 2015b])

Let $\mathcal{X} = [0, 1]^d$, $\Pi$ be $\text{Unif}(\mathcal{X})$ and $\Pi_{BQ}$ be a BQ rule whose states
\[ \{x_i\}_{i=1}^n \overset{i.i.d.}{\sim} \pi. \]
Then, whenever $\alpha > d/2$, we have:

\[
e(\hat{\Pi}_{BQ}; \Pi, \mathcal{H}) = \mathcal{O}_P(n^{-\alpha/d+\epsilon})
\]

where $\epsilon > 0$ can be arbitrarily small. Furthermore, let
\[ I_D = [\Pi[f] - D, \Pi[f] + D]. \]
Then:

\[ \mathbb{P}_n[I_D^c] = o_P(\exp(-Cn^{2\alpha/d-\epsilon})) \]
Integrals Over Manifolds

Idea: Construct a RKHS of functions $x : \mathbb{S}^2 \rightarrow \mathbb{R}$.

One such kernel, that leads to a Sobolev space of smoothness $\frac{3}{2}$ on $\mathbb{S}^2$:

$$k(t, t') = \frac{8}{3} - \|t - t'\|_2$$ for all $t, t' \in \mathbb{S}^2$. 

Integrals Over Manifolds

Idea: Construct a RKHS of functions $x : S^2 \rightarrow \mathbb{R}$.

One such kernel, that leads to a Sobolev space of smoothness $\frac{3}{2}$ on $S^2$:

$$k(t, t') = \frac{8}{3} - \|t - t'\|_2$$

for all $t, t' \in S^2$. 
Integrals Over Manifolds

Idea: Construct a RKHS of functions $x : \mathbb{S}^2 \to \mathbb{R}$.

One such kernel, that leads to a Sobolev space of smoothness $\frac{3}{2}$ on $\mathbb{S}^2$:

$$k(t, t') = \frac{8}{3} - \|t - t'\|_2^2 \text{ for all } t, t' \in \mathbb{S}^2.$$

For a certain spherical t-design $\{t_i\}_{i=1}^n$, a convergence rate of $e_{WCE}(M) = O(n^{-\frac{3}{4}})$ is achieved by the method $M = (A, b)$ where $b$ is the Bayesian Quadrature posterior mean - and this is worst-case optimal:
Integrals Over Manifolds

Full uncertainty quantification for integrals on manifolds:
We provide rates of $O_P(n^{-3/4})$ which is optimal for $\mathcal{H}^{3/2}(S^2)$!
Integration with Intractable Densities
Intractable Densities and Stein’s identity

What if \( \pi(x) \) is only known up to a constant?

\[
\pi(x) = \frac{\pi_c(x)}{c} \propto \pi_c(x)
\]

In those cases \( \Pi[k(\cdot, x)] \) is not available in closed form!

We can build an RKHS via kernel which takes into account information about \( \pi \), but does not require us to know \( c^4 \).

Let \( \phi(x) \) be twice differentiable, we can use the Stein transformation

\[
\mathcal{L}\phi(x) := \frac{\nabla[\phi(x)\pi(x)]}{\pi(x)}.
\]

Obtain an RKHS taking account of smoothness of both integrand and density of distribution - Control Functionals

---

\(^4\)Oates et al. [2017], Oates and Girolami [2016]
Control functionals for Monte Carlo integration

Chris J. Oates,

*University of Technology Sydney, Australia*

Mark Girolami

*University of Warwick, Coventry, and Alan Turing Institute, London, UK*

and Nicolas Chopin

*Centre de Recherche en Economie et Statistique and Ecole Nationale de la Statistique et de l'Administration Economique, Paris, France*

[Received October 2014. Final revision February 2016]

**Summary.** A non-parametric extension of control variates is presented. These leverage gradient information on the sampling density to achieve substantial variance reduction. It is not required that the sampling density be normalized. The novel contribution of this work is based on two important insights: a trade-off between random sampling and deterministic approximation and a new gradient-based function space derived from Stein's identity. Unlike classical control variates, our estimators improve rates of convergence, often requiring orders of magnitude fewer simulations to achieve a fixed level of precision. Theoretical and empirical results are presented, the latter focusing on integration problems arising in hierarchical models and models based on non-linear ordinary differential equations.

**Keywords:** Control variates; Non-parametrics; Reproducing kernel; Stein’s identity; Variance reduction
Theorem (Consistency of Control Functionals)

Suppose \( \{x_i\}_{i=1}^n \) arise from a Markov chain that targets a density \( \pi(x) \).

- Assume \( \mathcal{X} \) is bounded.
- Assume \( \pi(x) \) is bounded away from 0 on \( \mathcal{X} \).
- Assume \( \pi \in C^{2a+1}(\mathcal{X}) \) & \( k \in C^{2b+2}(\mathcal{X} \times \mathcal{X}) \).
- Assume \( k \) satisfies “certain boundary conditions”.
- Assume the Markov chain is uniformly ergodic.

Then, for \( f \in \mathcal{H}_k \), there exists \( h > 0 \) such that

\[
1_{h_n < h} \left( \prod[f] - \hat{\prod}[f] \right)^2 = \mathcal{O}_P \left( n^{-1-\frac{2(a+b)}{d}} + \epsilon \right),
\]

where \( \epsilon > 0 \) hides logarithmic factors.

---

5[Oates et al., 2016b]
Example: Computation of Marginal Likelihood

Consider computing the marginal likelihood for a non-linear ODE model

\[ \frac{d^2x}{dt^2} - \theta(1 - x^2) \frac{dx}{dt} + x = 0 \]

where \( \theta \in \mathbb{R} \) is an unknown parameter indicating the non-linearity and the strength of damping.

Observations \( y \) are made once every time unit, up to 10 units, and Gaussian measurement noise of standard deviation \( \sigma = 0.1 \) was added. A log-normal prior was placed on \( \theta \) such that \( \log(\theta) \sim N(0, 0.25) \).

Goal: Compute \( p(y) \).
Example: Computation of Marginal Likelihood

Thermodynamic integration is based on the identity

$$\log p(y) = \int_0^1 E_{\theta|y,t} [\log p(y|\theta)] dt.$$\n
where the “power posterior” for parameters $\theta$ given data $y$ is defined as $p(\theta|y,t) \propto p(y|\theta)^t p(\theta)$.

In TI, this integral is evaluated numerically over a discrete temperature ladder $0 = t_0 < t_1 < \cdots < t_m = 1$. e.g.

$$\log p(y) := \sum_{i=0}^{m-1} \frac{(t_{i+1} - t_i)}{2} \{E_{\theta|y,t_i} [\log p(y|\theta)] + E_{\theta|y,t_{i+1}} [\log p(y|\theta)]\}.$$\n
i.e. lots of integrals!
Example: Computation of Marginal Likelihood

Thermodynamic integration is based on the identity

$$\log p(y) = \int_0^1 E_{\theta|y,t} [\log p(y|\theta)] dt.$$ 

where the “power posterior” for parameters $\theta$ given data $y$ is defined as

$$p(\theta|y, t) \propto p(y|\theta)^t p(\theta).$$

In TI, this integral is evaluated numerically over a discrete temperature ladder $0 = t_0 < t_1 < \cdots < t_m = 1$. e.g.

$$\log p(y) := \sum_{i=0}^{m-1} \frac{(t_{i+1} - t_i)}{2} \{E_{\theta|y,t_i} [\log p(y|\theta)] + E_{\theta|y,t_{i+1}} [\log p(y|\theta)]\}.$$ 

i.e. lots of integrals!
Example: Computation of Marginal Likelihood

Thermodynamic integration is based on the identity

$$\log p(y) = \int_0^1 \mathbb{E}_{\theta|y,t} [\log p(y|\theta)] dt.$$  

where the “power posterior” for parameters $\theta$ given data $y$ is defined as $p(\theta|y,t) \propto p(y|\theta)^t p(\theta)$.

In TI, this integral is evaluated numerically over a discrete temperature ladder $0 = t_0 < t_1 < \cdots < t_m = 1$. e.g.

$$\log p(y) := \sum_{i=0}^{m-1} \frac{(t_{i+1} - t_i)}{2} \left\{ \mathbb{E}_{\theta|y,t_i} [\log p(y|\theta)] + \mathbb{E}_{\theta|y,t_{i+1}} [\log p(y|\theta)] \right\}.$$  

i.e. lots of integrals!
The Controlled Thermodynamic Integral for Bayesian Model Evidence Evaluation

Chris J. Oates, Theodore Papamarkou, and Mark Girolami

**Abstract**
Approximation of the model evidence is well known to be challenging. One promising approach is based on thermodynamic integration, but a key concern is that the thermodynamic integral can suffer from high variability in many applications. This article considers the reduction of variance that can be achieved by exploiting control variates in this setting. Our methodology applies whenever the gradient of both the log-likelihood and the log-prior with respect to the parameters can be efficiently evaluated. Results obtained on regression models and popular benchmark datasets demonstrate a significant and sometimes dramatic reduction in estimator variance and provide insight into the wider applicability of control variates to evidence estimation. Supplementary materials for this article are available online.

**Introduction**
In hypothesis-driven research, we are presented with data y that
Example: Computation of Marginal Likelihood

Figure: Computation of marginal likelihood for non-linear ordinary differential equations using thermodynamic integration (TI); van der Pol oscillator example. [Here we show the distribution of 100 independent realisations of each estimator for log \( p(y) \). “Standard TI” is based on arithmetic means. “Controlled TI” is based on ZV control variates.]
Intractable Densities and the Cone of Probability Measures

Ongoing work: BQ for densities $\pi(x)$ only available via samples Doubly Known Unknowns, optimal approximating projection in convex cone [Oates et al., 2016a].

$n = 10$:
Intractable Densities and the Cone of Probability Measures

**Ongoing work:** BQ for densities $\pi(x)$ only available via samples, **Doubly Known Unknowns**, optimal approximating projection in convex cone [Oates et al., 2016a].

$n = 100$: 

![Graphs showing True kernel mean $\mu$, draws from $[\mu | X]$, and distributions for $p$ known and $p$ unknown.](image)
Motivation: Assessment of Cardiac Models
Motivation: Assessment of Cardiac Models
Motivation: Assessment of Cardiac Models

Figure 2: Cardiac model results: (a) Computational cardiac model. A) Segmentation of the cardiac MRI. B) Computational model of the left and right ventricles. C) Schematic image showing the features of pressure (left) and volume transient (right). (b) Comparison of coverage frequencies, for each of 10 numerical integration tasks defined by functionals $g_j$ of the cardiac model output.
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
- Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
- Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
- An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
- Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
- Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
- An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
- Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
- Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
- An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
  - Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
  - Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
  - An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
- Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
- Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
- An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
- Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
- Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
- An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Conclusion

- A role for statistical science in numerical computation?
- A way to formally account for and quantify uncertainty in pipeline of computation
- Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
- Numerical computation increasingly resorted to in methods and applications
- Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
- Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
- An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
A role for statistical science in numerical computation?
A way to formally account for and quantify uncertainty in pipeline of computation
Contemporary Sciences and Engineering reliant on increasingly sophisticated mathematical objects
Numerical computation increasingly resorted to in methods and applications
Quantifying, accounting for uncertainty fundamental to support reasoning and subsequent decision making under uncertainty
Understanding the impacts of numerical uncertainty is essential for any application related to decision making and risk assessment
An exciting research area emerging at the intersection of mathematics, statistics and computing science - come and join us!
Integration with Intractable Densities

Application to Marginalisation

Research Group at Imperial - PAPER SUBMITTED

EPSRC
Engineering and Physical Sciences
Research Council

Mark Girolami (Imperial, ATI)
Integration with Intractable Densities
Application to Marginalisation

Research Group at Imperial - PAPER ACCEPTED

EPSRC
Engineering and Physical Sciences
Research Council

Mark Girolami (Imperial, ATI)  Probabilistic Numerical Computation  November 12, 2017  68 / 72


Funding Acknowledgments

- EPSRC Established Research Career Fellowship
- EPSRC Programme Grant - Enabling Quantification of Uncertainty in Inverse Problems
- EPSRC Programme Grant - Inference, Computation and Numerics for Insights into Cities
Hiring

- EPSRC Programme Grant - EP/P020720/1
- Inference, Computation, and Numerics for Insights into Cities - ICONIC
- https://iconicmath.org/
Hiring

- EPSRC Programme Grant - EP/P020720/1
- Inference, Computation, and Numerics for Insights into Cities - ICONIC
- https://iconicmath.org/