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

Collaborators in Probabilistic Numerical Computation

Chris J. Oates
Newcastle

Jon Cockayne
Warwick

F-X Briol
Warwick

Tim Sullivan
F.U. Berlin

Philipp Hennig
MPI Tuebingen

Mike Osborne
Oxford

Dino Sejdinovic
Oxford

Andrew Stuart
Caltech

Probabilistic Numerical Computation

Consider the following school boy and girl differential equation

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

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

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

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

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

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

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

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

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

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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

This uncertainty in θ 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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

This uncertainty in θ 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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

This uncertainty in θ 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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

This uncertainty in θ 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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

This uncertainty in θ 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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

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

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

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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

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

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

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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

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

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

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 θ may be an empirically derived parameter.

This immediately introduces uncertainty into our deterministic world.

Our uncertainty in θ can be described using the calculus of probability

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 θ 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 θ 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 θ 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 θ 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 θ 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 θ 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 θ 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.

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

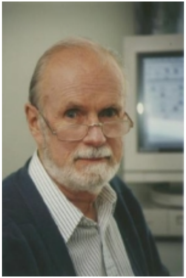
Nothing stochastic or random about this.

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

History of Probabilistic Numerical Methods

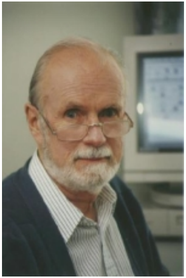


Tests of Probabilistic Models for Propagation of Roundoff Errors

T. E. HULL, University of Toronto; J. R. SWENSON, New York University (Ed: J. Traub) Communications of the ACM, 9(2):108113, 1966.

> *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.*

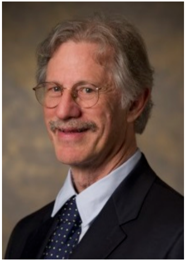
History of Probabilistic Numerical Methods



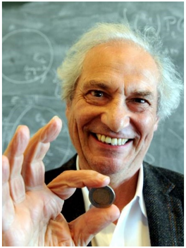
Tests of Probabilistic Models for Propagation of Roundoff Errors

T. E. HULL, University of Toronto; J. R. SWENSON, New York University (Ed: J. Traub) Communications of the ACM, 9(2):108113, 1966.

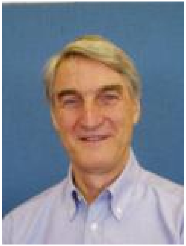
> *In any prolonged computation it is generally assumed that the accumulated effect of roundoff errors is in some sense statistical. The purpose of this paper is to give precise descriptions of certain probabilistic models for roundoff error, and then to describe a series of experiments for testing the 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.*



Joseph Kadane
Kadane [1985]



Persi Diaconis
Diaconis [1988]



Tony O'Hagan
O'Hagan [1992]

John Skilling
Skilling [1991]

Question: "Is numerical computation a statistical inference problem?"

ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 2, Number 3, Summer 1972

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.

History of Probabilistic Numerical Methods, F.M.Larkin

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.

History of Probabilistic Numerical Methods, F.M.Larkin

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]

PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Cite this article: Hennig P, Osborne MA, Girolami M. 2015 Probabilistic numerics and uncertainty in computations. *Proc. R. Soc. A* **471**: 20150142.
http://dx.doi.org/10.1098/rspa.2015.0142

Received: 2 March 2015
Accepted: 3 June 2015

Subject Areas:
statistics, computational mathematics, artificial intelligence

Keywords:
numerical methods, probability, inference, statistics

Probabilistic numerics and uncertainty in computations

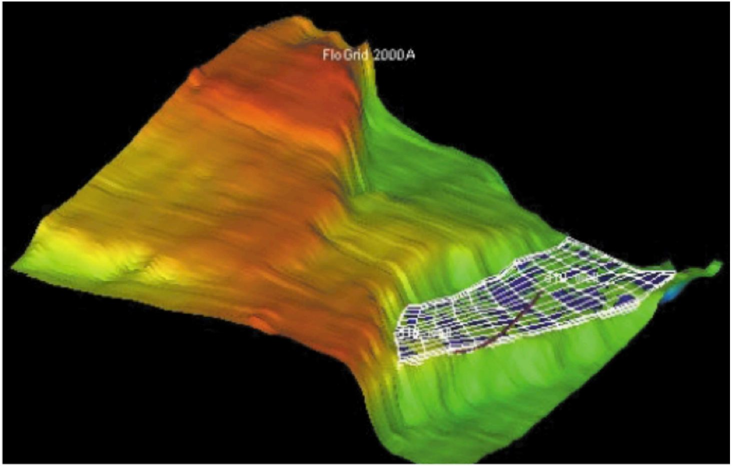
Philipp Hennig[1], Michael A. Osborne[2] and Mark Girolami[3]

[1]Department of Empirical Inference, Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]Department of Engineering Science, University of Oxford, Oxford, UK
[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

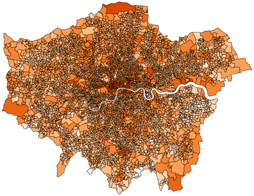
Differential Equations

Motivation - Data Driven Engineering

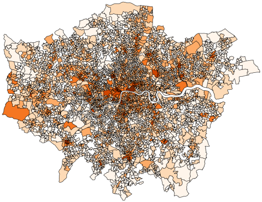


Motivation - Data Informed Medical and Life Sciences

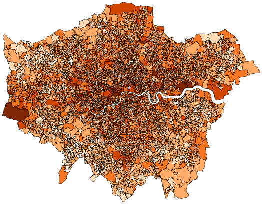
Motivation - Computational Social Science



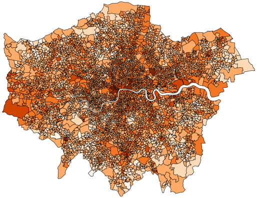
Burglaries



Drugs



Traffic



Violence

PN for PDEs

A "widely used" linear PDE. Given g, κ, b find u

$$\begin{aligned} -\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) &= g(\boldsymbol{x}) \quad \text{in } D \\ u(\boldsymbol{x}) &= b(\boldsymbol{x}) \quad \text{on } \partial D \end{aligned}$$

For general D, $u(\boldsymbol{x})$ this cannot be solved analytically.

The majority of PDE solvers produce an approximation like:

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

We want to quantify the error from finite N probabilistically.

PN for PDEs

A "widely used" linear PDE. Given g, κ, b find u

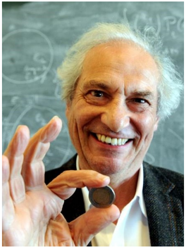
$$
\begin{aligned}
-\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) &= g(\boldsymbol{x}) \quad \text{in } D \\
u(\boldsymbol{x}) &= b(\boldsymbol{x}) \quad \text{on } \partial D
\end{aligned}
$$

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

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

We want to quantify the error from finite N probabilistically.

History of Probabilistic Numerical Methods



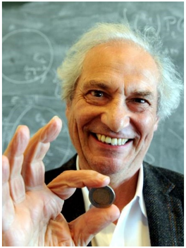
Bayesian Numerical Analysis

P. DIACONIS, Stanford University.

Statistical Decision Theory and Related Topics IV, 1, 163175, 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.

Statistical Decision Theory and Related Topics IV, 1, 163175, 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).*

PN for PDEs

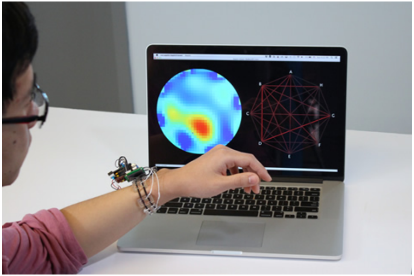
Inverse Problem: Given partial information of g, b, u find κ

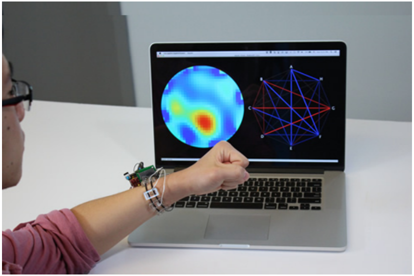
$$
\begin{aligned}
-\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) = g(\boldsymbol{x}) \quad & \text{in } D \\
u(\boldsymbol{x}) = b(\boldsymbol{x}) \quad & \text{on } \partial D
\end{aligned}
$$

PN for PDEs

Inverse Problem: Given partial information of g, b, u find κ

$$
\begin{aligned}
-\nabla \cdot (\kappa(\boldsymbol{x})\nabla u(\boldsymbol{x})) &= g(\boldsymbol{x}) \quad \text{in } D \\
u(\boldsymbol{x}) &= b(\boldsymbol{x}) \quad \text{on } \partial D
\end{aligned}
$$

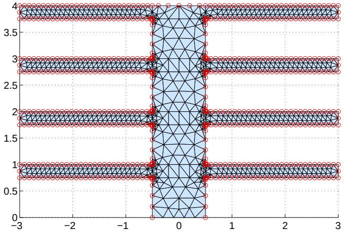


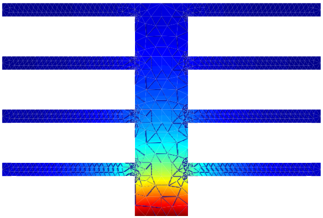


PN for PDEs

Inverse Problem: Given partial information of g, b, u find κ

$$
\begin{aligned}
-\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) &= g(\boldsymbol{x}) \quad \text{in } D \\
u(\boldsymbol{x}) &= b(\boldsymbol{x}) \quad \text{on } \partial D
\end{aligned}
$$





PN for PDEs

Inverse Problem: Given partial information of g, b, u find κ

$$
\begin{aligned}
-\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) &= g(\boldsymbol{x}) \quad \text{in } D \\
u(\boldsymbol{x}) &= b(\boldsymbol{x}) \quad \text{on } \partial D
\end{aligned}
$$

Bayesian Inverse Problem:

$$
\kappa \sim \Pi_\kappa \longrightarrow \boxed{\text{Data } u(\boldsymbol{x}_i) = y_i} \longrightarrow \kappa | \boldsymbol{y} \sim \Pi_\kappa^{\boldsymbol{y}}
$$

We want to account for an inaccurate forward solver in the inverse problem.

Forward Problem

Abstract Formulation

$$\mathcal{A}u(\boldsymbol{x}) = g(\boldsymbol{x}) \qquad \text{in } D$$

Forward inference procedure:

$$u \sim \Pi_u \longrightarrow \boxed{\text{“Data” } \mathcal{A}u(\boldsymbol{x}_i) = g(\boldsymbol{x}_i)} \longrightarrow u|\boldsymbol{g} \sim \Pi_u^{\boldsymbol{g}}$$

Posterior for the forward problem

Use a Gaussian Process prior $u \sim \Pi_u = \mathcal{GP}(0, k)$. Assuming linearity, the posterior $\Pi_u^{\boldsymbol{g}}$ is available in closed-form[2].

$$
\begin{aligned}
\Pi_u^{\boldsymbol{g}} &\sim \mathcal{GP}(m_1, \Sigma_1) \\
m_1(\boldsymbol{x}) &= \bar{\mathcal{A}} K(\boldsymbol{x}, X) \left[\mathcal{A}\bar{\mathcal{A}} K(X, X)\right]^{-1} \boldsymbol{g} \\
\Sigma_1(\boldsymbol{x}, \boldsymbol{x}') &= k(\boldsymbol{x}, \boldsymbol{x}') - \bar{\mathcal{A}} K(\boldsymbol{x}, X) \left[\mathcal{A}\bar{\mathcal{A}} K(X, X)\right]^{-1} \mathcal{A} K(X, \boldsymbol{x}')
\end{aligned}
$$

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

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

[2] Larkin 1972, [Cockayne et al., 2016, Owhadi, 2014]

Posterior for the forward problem

Use a Gaussian Process prior $u \sim \Pi_u = \mathcal{GP}(0, k)$. Assuming linearity, the posterior $\Pi_u^{\boldsymbol{g}}$ is available in closed-form[2].

$$\Pi_u^{\boldsymbol{g}} \sim \mathcal{GP}(m_1, \Sigma_1)$$

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

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

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

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 $B_\epsilon(u_0)$ of radius ϵ centered on the true solution u_0 of the PDE, we have

$$1 - \Pi_u^{\boldsymbol{g}}[B_\epsilon(u_0)] = \mathcal{O}\left(\frac{h^{2\beta-2\rho-d}}{\epsilon}\right)$$

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

Toy Example

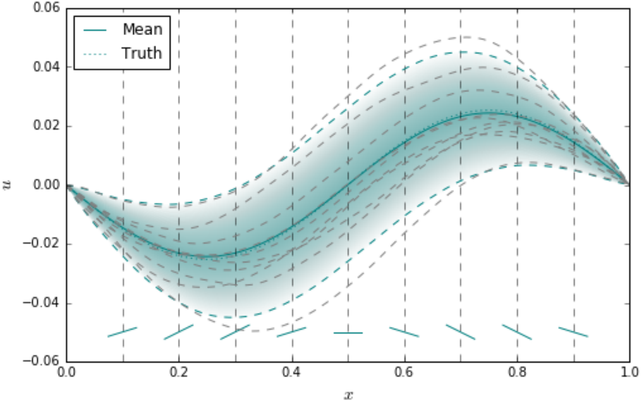
$$-\nabla^2 u(x) = g(x) \qquad x \in (0,1)$$
$$u(x) = 0 \qquad x = 0, 1$$

To associate with the notation from before...

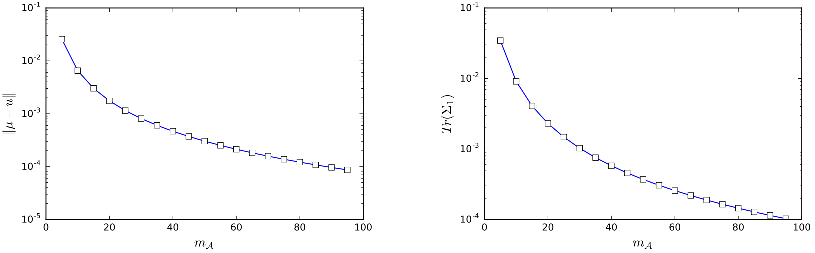
$$\Pi_u \sim \mathcal{GP}(0, k(x,y))$$
$$\mathcal{A} = -\frac{d^2}{dx^2} \quad \bar{\mathcal{A}} = -\frac{d^2}{dy^2}$$

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

$$
\begin{aligned}
-\nabla \cdot (\kappa(\boldsymbol{x}) \nabla u(\boldsymbol{x})) &= g(\boldsymbol{x}) \quad \text{in } D \\
u(\boldsymbol{x}) &= b(\boldsymbol{x}) \quad \text{on } \partial D
\end{aligned}
$$

Now we need to incorporate the forward posterior measure $\Pi_u^{\boldsymbol{g}}$ into the posterior measure for the inverse problem, κ

Incorporation of Forward Measure

Assuming the data in the inverse problem is:

$$y_i = u(\boldsymbol{x}_i) + \xi_i \quad i = 1, \ldots, n$$
$$\boldsymbol{\xi} \sim N(\boldsymbol{0}, \Gamma)$$

implies the standard likelihood:

$$p(\boldsymbol{y} | \kappa, u) \sim N(\boldsymbol{y}; \boldsymbol{u}, \Gamma)$$

But we don't know u

Marginalise the forward posterior $\Pi_u^{\boldsymbol{g}}$ to obtain a "PN" likelihood:

$$p_{\mathrm{PN}}(\boldsymbol{y}|\kappa) \propto \int p(\boldsymbol{y}|\kappa, u) d\Pi_u^{\boldsymbol{g}}$$
$$\sim N(\boldsymbol{y}; \boldsymbol{m}_1, \Gamma + \Sigma_1)$$

Incorporation of Forward Measure

Assuming the data in the inverse problem is:

$$y_i = u(\boldsymbol{x}_i) + \xi_i \quad i = 1, \ldots, n$$
$$\boldsymbol{\xi} \sim N(\boldsymbol{0}, \Gamma)$$

implies the standard likelihood:

$$p(\boldsymbol{y}|\kappa, u) \sim N(\boldsymbol{y}; \boldsymbol{u}, \Gamma)$$

But we don't know u

Marginalise the forward posterior $\Pi_u^{\boldsymbol{g}}$ to obtain a "PN" likelihood:

$$p_{\mathrm{PN}}(\boldsymbol{y}|\kappa) \propto \int p(\boldsymbol{y}|\kappa, u) d\Pi_u^{\boldsymbol{g}}$$
$$\sim N(\boldsymbol{y}; \boldsymbol{m}_1, \Gamma + \Sigma_1)$$

Inverse Contraction

Denote by $\Pi_\kappa^{\boldsymbol{y}}$ the posterior for κ from likelihood p, and by $\Pi_{\kappa,\mathrm{PN}}^{\boldsymbol{y}}$ the posterior for κ from likelihood p_{PN}.

Theorem (Inverse Contraction)

Assume $\Pi_\kappa^{\boldsymbol{y}} \to \delta(\kappa_0)$ *as* $n \to \infty$.
Then $\Pi_{\kappa,PN}^{\boldsymbol{y}} \to \delta(\kappa_0)$ *provided*

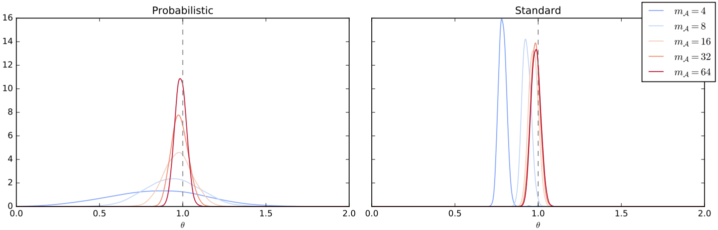
$$h = o\left(n^{-1/(\beta-\rho-d/2)}\right)$$

Back to the Toy Example

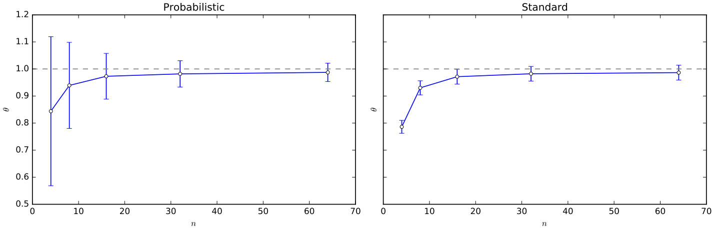
$$
\begin{aligned}
-\nabla \cdot (\kappa \nabla u(x)) &= \sin(2\pi x) \qquad && x \in (0,1) \\
u(x) &= 0 && x = 0, 1
\end{aligned}
$$

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

Posteriors for κ



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

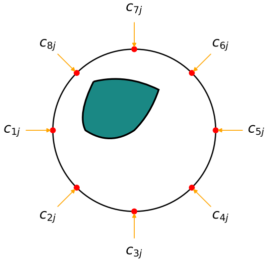


Electrical Impedance Tomography

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

Electrical Impedance Tomography

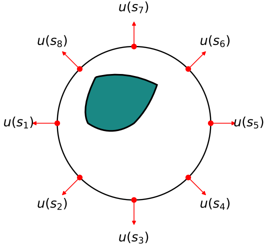
A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.



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: $y_i = x(t_i^{\text{obs}}) - x(t_{\text{ref}}) + \epsilon_i$

Electrical Impedance Tomography

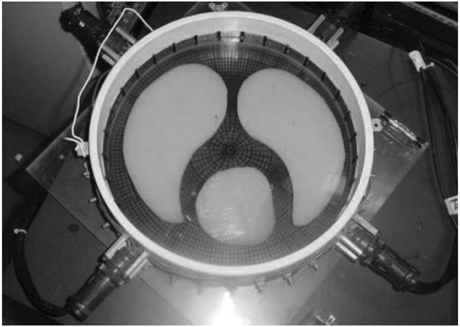
A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.

Governing equations are essentially Darcy's law:

$$
\begin{aligned}
-\nabla \cdot (\theta(t) \nabla x(t) &= 0 && t \in D \\
\theta(t_i^{\text{obs}}) \frac{\partial x}{\partial n}(t_i^{\text{obs}}) &= c_{ij} && i = 1, \dots, N_S
\end{aligned}
$$

Experimental Set-Up

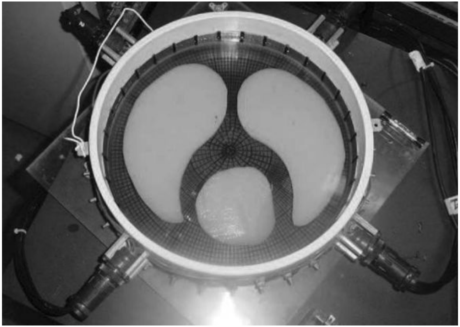
Experiments due to Isaacson 2004.



- 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 $\mathcal{L}(\theta; \boldsymbol{y})$ requires exact solution of the PDE.

Posteriors obtained using the PN likelihood

$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x)\mathrm{d}P_{x|a}$$
$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1).$$

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 $\mathcal{L}(\theta; \boldsymbol{y})$ requires exact solution of the PDE.

Posteriors obtained using the PN likelihood

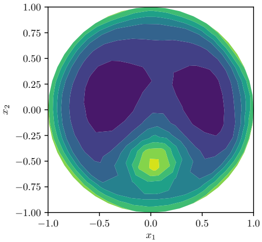
$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x) \mathrm{d}P_{x|a}$$
$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1).$$

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

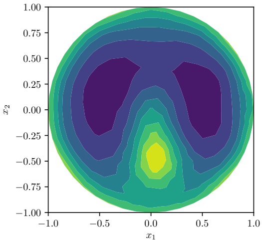
Computation facilitated with Markov chain Monte Carlo, based on the preconditioned Crank-Nicholson proposal.

Recovered Fields

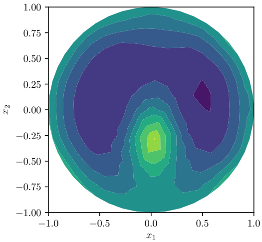
Posterior means $m(t) = \mathbb{E}_{\boldsymbol{y}}[\theta(t)]$:



(a) $n = 96$



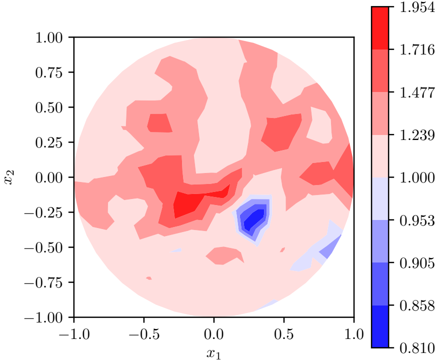
(b) $n = 127$



(c) $n = 165$

Variance Analysis

Ratio of (pointwise) posterior variance $v(t) = \mathbb{V}_{\boldsymbol{y}}[\theta(t)]$ computed from the PN posterior based on $\mathcal{L}_n$ and the "standard" posterior based on $\hat{\mathcal{L}}_N$ with $n = N = 96$:

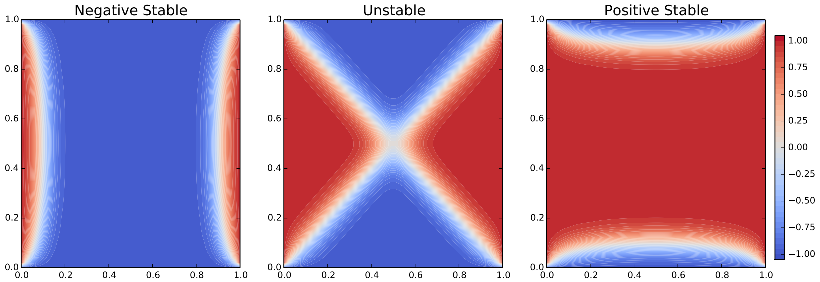


Allen–Cahn

A prototypical nonlinear model.

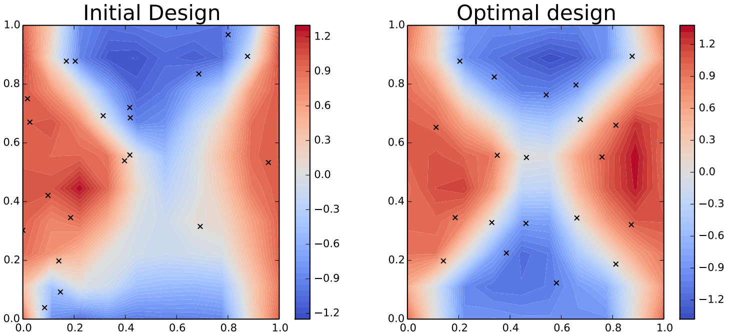
$$
\begin{aligned}
-\theta \nabla^2 u(\boldsymbol{x}) + \theta^{-1}(u(\boldsymbol{x})^3 - u(\boldsymbol{x})) &= 0 && \boldsymbol{x} \in (0,1)^2 \\
u(\boldsymbol{x}) &= 1 && x_1 \in \{0,1\}\,;\, 0 < x_2 < 1 \\
u(\boldsymbol{x}) &= -1 && x_2 \in \{0,1\}\,;\, 0 < x_1 < 1
\end{aligned}
$$

Goal: infer θ

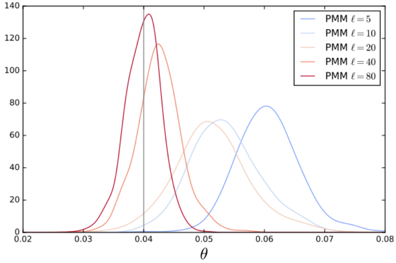


Allen–Cahn: Forward Solutions

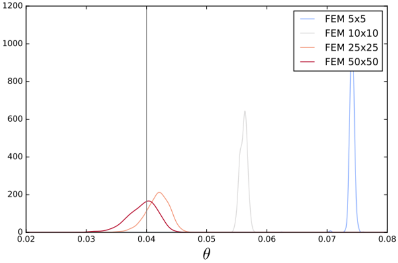
Nonlinear PDE - non-GP posterior sampling schemes required, see [Cockayne et al., 2016].



Allen–Cahn: Inverse Problem



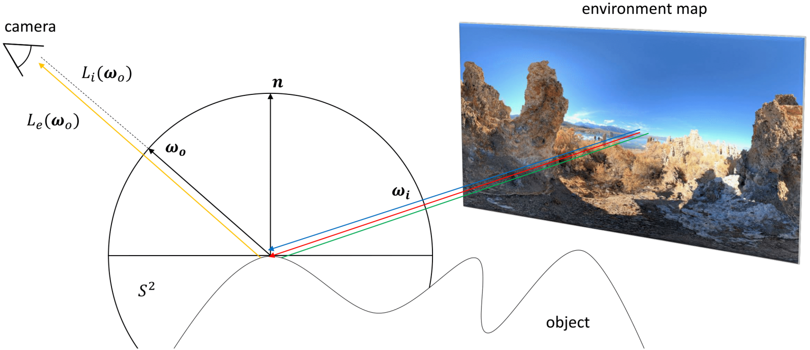
(a) Probabilistic.



(b) Standard

Integration

Illustrative Application - Integral over Manifold



Integrals Over Manifolds

$$L_o(\boldsymbol{\omega}_o) \quad = \quad L_e(\boldsymbol{\omega}_o) + \int_{\mathbb{S}^2} L_i(\boldsymbol{\omega}_i)\rho(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)[\boldsymbol{\omega}_i \cdot \boldsymbol{n}]_+ \mathrm{d}\pi(\boldsymbol{\omega}_i)$$

- $L_o(\boldsymbol{\omega}_o) =$ outgoing radiance
- $L_e(\boldsymbol{\omega}_o) =$ amount of light emitted by the object itself
- $L_i(\boldsymbol{\omega}_i) =$ amount of light reaching object from direction $\boldsymbol{\omega}_i$
- $\rho =$ bidirectional reflectance distribution function
- $\pi =$ uniform distribution on $\mathbb{S}^2$

To be computed

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

Integrals Over Manifolds

$$L_o(\boldsymbol{\omega}_o) \;\; = \;\; L_e(\boldsymbol{\omega}_o) + \int_{\mathbb{S}^2} L_i(\boldsymbol{\omega}_i)\rho(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)[\boldsymbol{\omega}_i \cdot \boldsymbol{n}]_+ \mathrm{d}\pi(\boldsymbol{\omega}_i)$$

- $L_o(\boldsymbol{\omega}_o) =$ outgoing radiance
- $L_e(\boldsymbol{\omega}_o) =$ amount of light emitted by the object itself
- $L_i(\boldsymbol{\omega}_i) =$ amount of light reaching object from direction $\boldsymbol{\omega}_i$
- $\rho =$ bidirectional reflectance distribution function
- $\pi =$ uniform distribution on $\mathbb{S}^2$

To be computed

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

The Problem

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

$$\Pi[f] = \int_{\mathcal{X}} f \, \mathrm{d}\Pi \approx \sum_{i=1}^{n} w_i f(\boldsymbol{x}_i) = \hat{\Pi}[f] \tag{1}$$

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

The Problem

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

$$\Pi[f] = \int_{\mathcal{X}} f \mathrm{d}\Pi \approx \sum_{i=1}^{n} w_i f(\boldsymbol{x}_i) = \hat{\Pi}[f] \tag{1}$$

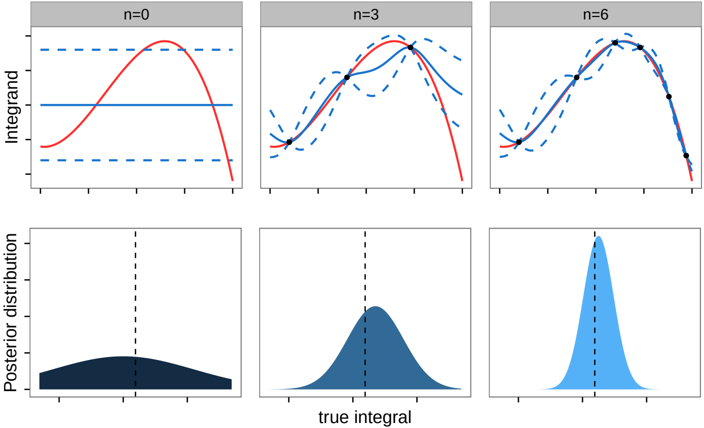
High numerical uncertainty when f is expensive or n is small!

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

Measure on Integral push-forward of measure on function.

[3] [O'Hagan, 1991, Rasmussen and Ghahramani, 2002, Briol et al., 2015a,b]

Sketch of Bayesian Quadrature



$$\mathbb{E}_n[\Pi[f]] = \hat{\Pi}_{\mathrm{BQ}} = \Pi[k(\cdot, \boldsymbol{X})]\boldsymbol{K}^{-1}\boldsymbol{f}$$
$$\mathbb{V}_n[\Pi[f]] = \Pi\Pi[k(\cdot,\cdot)] - \Pi[k(\cdot, \boldsymbol{X})]\boldsymbol{K}^{-1}\Pi[k(\boldsymbol{X}, \cdot)].$$

Theory for Bayesian Quadrature

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

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

and finite norm $\left\|f\right\|_{H^\alpha(\Pi)} := \langle f, f \rangle_{H^\alpha}^{1/2}$.
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 $\mathcal{X} = [0,1]^d$, Π be Unif($\mathcal{X}$) and Π_{BQ} be a BQ rule whose states $\{\mathbf{x}_i\}_{i=1}^n \overset{i.i.d.}{\sim} \pi$. Then, whenever $\alpha > d/2$, we have:

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

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

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

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 $\mathbb{S}^2$:

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

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 $\mathbb{S}^2$:

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

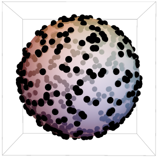
Integrals Over Manifolds

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 $\mathbb{S}^2$:

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

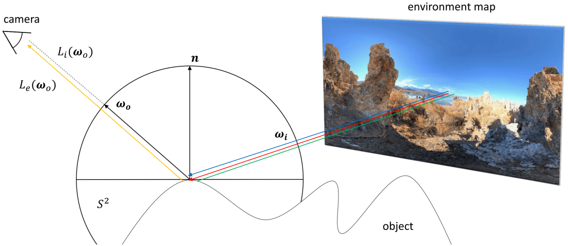
For a certain *spherical t-design* $\{t_i\}_{i=1}^n$, a convergence rate of $e_{\text{WCE}}(M) = O(n^{-\frac{3}{4}})$ is achieved by the method $M = (A, b)$ where b is the Bayesian Quadrature posterior mean - and this is worst-case optimal:

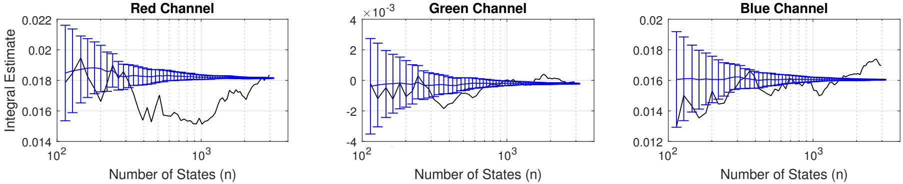


Integrals Over Manifolds

Full uncertainty quantification for integrals on manifolds:

Prob Integration in Comp Graphics [Briol et al., 2015b]





We provide rates of $\mathcal{O}_P(n^{-\frac{3}{4}})$ which is optimal for $\mathcal{H}^{\frac{3}{2}}(\mathbb{S}^2)$!

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 π, but does not require us to know c[4].

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

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

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]

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 3, *pp.* 695–718

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

Theory for Control Functionals[5]

Theorem (Consistency of Control Functionals)

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

- *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*

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

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 $\boldsymbol{y}$ 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 θ such that $\log(\theta) \sim N(0, 0.25)$.

Goal: Compute $p(\boldsymbol{y})$.

Example: Computation of Marginal Likelihood

Thermodynamic integration is based on the identity

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

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

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

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

i.e. lots of integrals!

Example: Computation of Marginal Likelihood

Thermodynamic integration is based on the identity

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

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

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

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

i.e. lots of integrals!

Example: Computation of Marginal Likelihood

Thermodynamic integration is based on the identity

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

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

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

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

i.e. lots of integrals!

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2016, VOL. 111, NO. 514, 634–645, Theory and Methods
http://dx.doi.org/10.1080/01621459.2015.1021006

Taylor & Francis
Taylor & Francis Group

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.

ARTICLE HISTORY
Received April 2014
Revised November 2014

KEYWORDS
Control variates; Model evidence; Temperature ladder

Introduction

In hypothesis-driven research, we are presented with data y that

Jeliazkov 2001), nested sampling (Skilling 2006), particle filters (Del Moral, Doucet, and Jasra 2006), multicanonical algorithms

Example: Computation of Marginal Likelihood

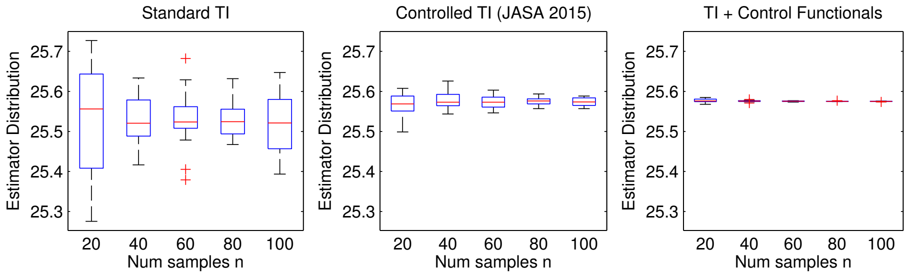
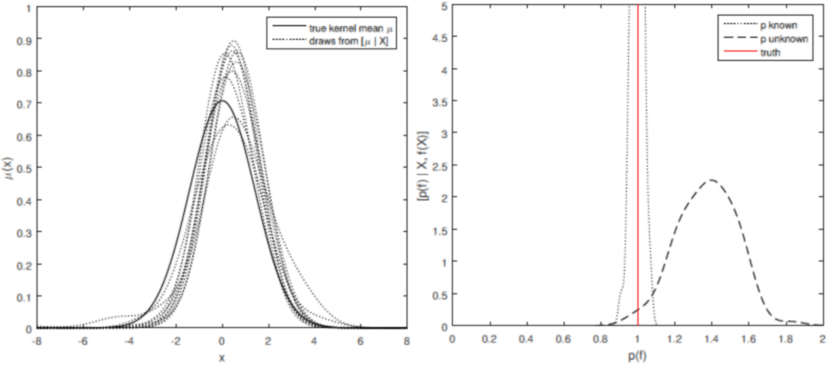


Figure: Computation of marginal likelihood for non-linear ordinary differential equations using thermodynamic integration (TI); van der Pol oscillator example. [Here we show the distribution of 100 independent realisations of each estimator for $\log p(\boldsymbol{y})$. "Standard TI" is based on arithmetic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

Intractable Densities and the Cone of Probability Measures

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

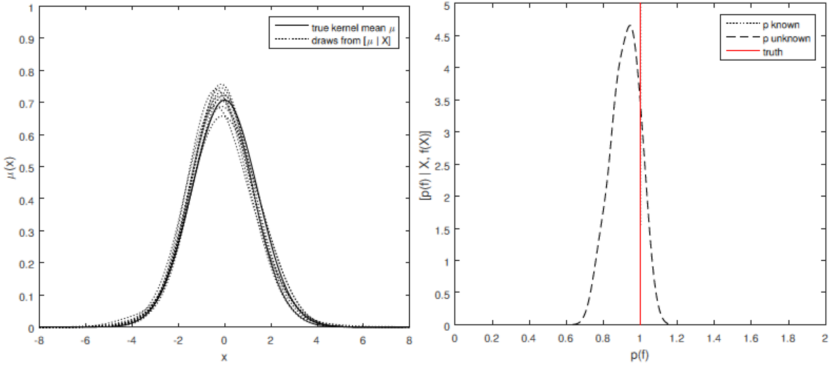
$n = 10$:



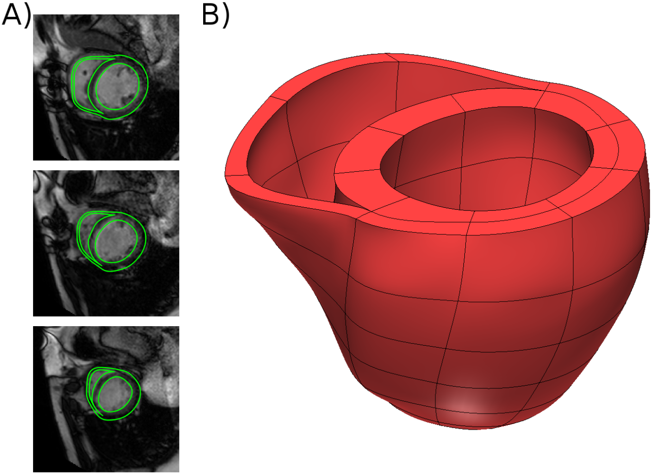
Intractable Densities and the Cone of Probability Measures

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

$n = 100$:

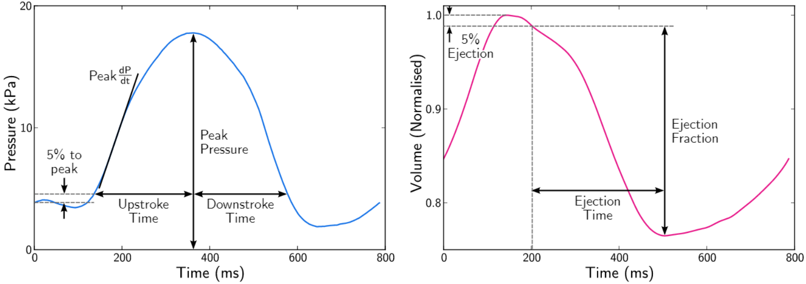


Motivation: Assessment of Cardiac Models

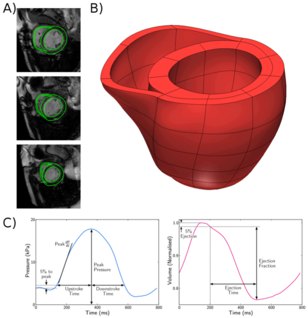


Motivation: Assessment of Cardiac Models

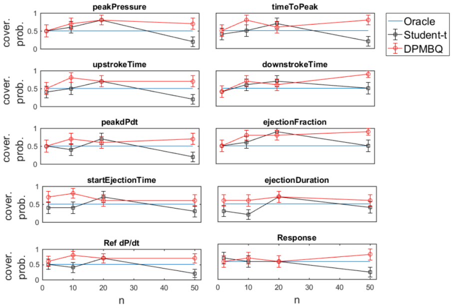
C)



Motivation: Assessment of Cardiac Models



(a)



(b)

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 !

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 !

Research Group at Imperial - PAPER SUBMITTED

EPSRC

Engineering and Physical Sciences
Research Council












Research Group at Imperial - PAPER ACCEPTED

EPSRC
Engineering and Physical Sciences Research Council












F.-X. Briol, C. J. Oates, M. Girolami, and M. A. Osborne. Frank-Wolfe Bayesian Quadrature: Probabilistic Integration with Theoretical Guarantees. In *Advances In Neural Information Processing Systems 28*, pages 1162–1170, 2015a.

F-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic. Probabilistic integration: A role for statisticians in numerical analysis? *arXiv preprint arXiv:1512.00933*, 2015b.

J. Cockayne, C. J. Oates, T. Sullivan, and M. Girolami. Probabilistic Meshless Methods for Partial Differential Equations and Bayesian Inverse Problems. *arXiv preprint arXiv:1605.07811*, 2016.

P. Diaconis. Bayesian numerical analysis. *Statistical decision theory and related topics IV*, 1:163–175, 1988.

J. B. Kadane. Parallel and Sequential Computation: A Statistician's view. *Journal of Complexity*, 1:256–263, 1985.

C. J. Oates and M. Girolami. Control Functionals for Quasi-Monte Carlo Integration. In *Proceedings of the 19th Conference on Artificial Intelligence and Statistics*, 2016.

C. J. Oates, F-X. Briol, and M. Girolami. Probabilistic Integration and Intractable Distributions. *arXiv preprint arXiv:1606.06841*, 2016a.

C. J. Oates, J. Cockayne, F-X. Briol, and M. Girolami. Convergence Rates for a Class of Estimators Based on Stein's Identity. *arXiv preprint arXiv:1603.03220*, 2016b.

C. J. Oates, M. Girolami, and N. Chopin. Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society B: Statistical Methodology*, 2017.

A. O'Hagan. BayesHermite quadrature. *Journal of Statistical Planning and Inference*, 29:245–260, 1991.

A. O'Hagan. Some Bayesian numerical analysis. *Bayesian Statistics*, 4:345–363, 1992.

H. Owhadi. Bayesian numerical homogenization. *arXiv preprint arXiv:1406.6668*, 2014.

C. Rasmussen and Z. Ghahramani. Bayesian Monte Carlo. In *Advances in Neural Information Processing Systems*, pages 489–496, 2002.

J. Skilling. Bayesian solution of ordinary differential equations. *Maximum Entropy and Bayesian Methods*, 50:23–37, 1991.

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/`