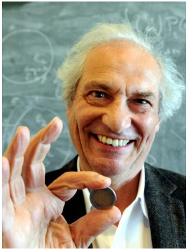
Part V

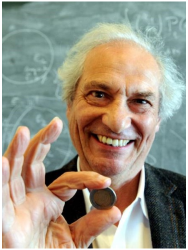


Bayesian Numerical Analysis

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).*



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).*

Tenth Job: Extension to More Challenging Integrals

Extension to More Challenging Integrals

Three extensions that we will discuss:

1. Integrals over manifolds:

$$\int_{\mathcal{M}} x(t) \mathrm{d}\pi(t)$$

2. Integrals with densities known up to normalisation:

$$\int x(t) \mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$

3. Integrals with unknown densities:

$$\int x(t) \mathrm{d}\pi(t), \quad \{t_i\}_{i=1}^n \overset{IID}{\sim} \pi$$

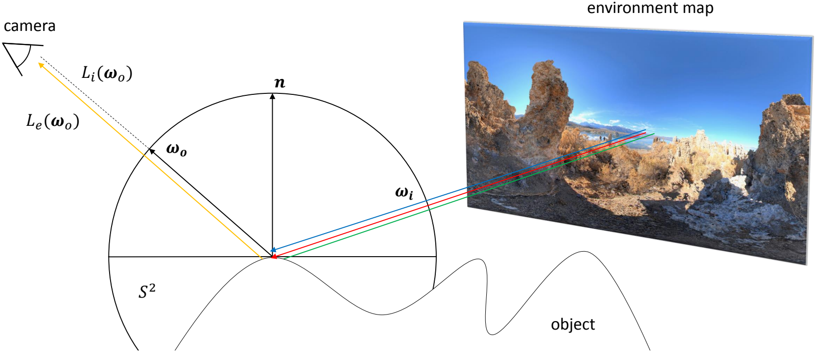
In each case the aim is to perform principled Bayesian uncertainty quantification for the value of the integral $Q = \int x(t) \mathrm{d}\pi(t)$.

Extension to More Challenging Integrals

Three extensions that we will discuss:

1. Integrals over manifolds:
$$\int_{\mathcal{M}} x(t)\mathrm{d}\pi(t)$$
2. Integrals with densities known up to normalisation:
$$\int x(t)\mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$
3. Integrals with unknown densities:
$$\int x(t)\mathrm{d}\pi(t), \quad \{t_i\}_{i=1}^n \overset{IID}{\sim} \pi$$

In each case the aim is to perform principled Bayesian uncertainty quantification for the value of the integral $Q = \int x(t)\mathrm{d}\pi(t)$.


environment map
camera
$L_i(\boldsymbol{\omega}_o)$
$L_e(\boldsymbol{\omega}_o)$
$\boldsymbol{n}$
$\boldsymbol{\omega_o}$
$\boldsymbol{\omega_i}$
S^2
object

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

- $L_o(\boldsymbol{\omega}_o)$ = 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 $\boldsymbol{\omega}_i$
- ρ = bidirectional reflectance distribution function
- π = uniform distribution on $\mathbb{S}^2$

To be computed

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

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

- $L_o(\boldsymbol{\omega}_o) =$ 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.

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 $\mathbb{S}^2$:

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

Extension 1: 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 $\mathbb{S}^2$:

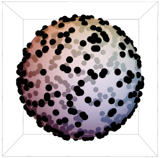
$$k(t, t') = \frac{8}{3} - \|t - t'\|_2 \text{ for all } t, t' \in \mathbb{S}^2.$$

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

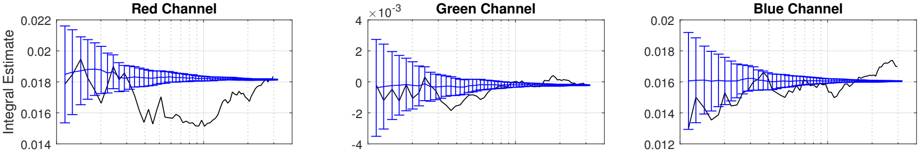
One such kernel, that leads to a Sobolev space of smoothness $\frac{3}{2}$ on $\mathbb{S}^2$:

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

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:



Full uncertainty quantification for integrals on manifolds:



Integrals with densities known up to normalisation

$$\int x(t)\mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$

occur in applications of Bayesian statistical methods:

$$p(\text{params}|\text{data}) = \frac{p(\text{data}|\text{params})\; p(\text{params})}{\int p(\text{data}|\text{params})\; \mathrm{d}p(\text{params})} \quad \begin{array}{l} \leftarrow \tilde{\pi} \\ \leftarrow \text{unknown } (*) \end{array}$$

Cannot compute with Bayesian quadrature, since relies on the following integrals having a closed form:

$$\int k(\cdot, t)\mathrm{d}\pi(t), \qquad \iint k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') \qquad (**)$$

MCMC? Compute the denominator $(*)$ with Bayesian Quadrature first?

To address these problems we will instead go to some effort to force $(**)$ to have a closed form... via Stein's Method.

Extension 2: Unknown Normalisation Constant

Integrals with densities known up to normalisation

$$\int x(t)\mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$

occur in applications of Bayesian statistical methods:

$$p(\text{params}|\text{data}) \quad = \quad \frac{p(\text{data}|\text{params})\ p(\text{params})}{\int p(\text{data}|\text{params})\ \mathrm{d}p(\text{params})} \quad \begin{array}{c} \leftarrow \tilde{\pi} \\ \leftarrow \text{unknown } (*) \end{array}$$

Cannot compute with Bayesian quadrature, since relies on the following integrals having a closed form:

$$\int k(\cdot, t)\mathrm{d}\pi(t), \qquad \iint k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') \qquad (**)$$

MCMC? Compute the denominator $(*)$ with Bayesian Quadrature first?

To address these problems we will instead go to some effort to force $(**)$ to have a closed form... via Stein's Method.

Integrals with densities known up to normalisation

$$\int x(t)\mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$

occur in applications of Bayesian statistical methods:

$$p(\text{params}|\text{data}) = \frac{p(\text{data}|\text{params})\ p(\text{params})}{\int p(\text{data}|\text{params})\ \mathrm{d}p(\text{params})} \quad \begin{array}{r} \leftarrow \tilde{\pi} \\ \leftarrow \text{unknown } (*) \end{array}$$

Cannot compute with Bayesian quadrature, since relies on the following integrals having a closed form:

$$\int k(\cdot, t)\mathrm{d}\pi(t), \qquad \iint k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') \qquad (**)$$

MCMC? Compute the denominator $(*)$ with Bayesian Quadrature first?

To address these problems we will instead go to some effort to force $(**)$ to have a closed form... via Stein's Method.

Extension 2: Unknown Normalisation Constant

Integrals with densities known up to normalisation

$$\int x(t)\mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$

occur in applications of Bayesian statistical methods:

$$p(\text{params}|\text{data}) = \frac{p(\text{data}|\text{params})\ p(\text{params})}{\int p(\text{data}|\text{params})\ \mathrm{d}p(\text{params})} \quad \begin{array}{c} \leftarrow \tilde{\pi} \\ \leftarrow \text{unknown } (*) \end{array}$$

Cannot compute with Bayesian quadrature, since relies on the following integrals having a closed form:

$$\int k(\cdot, t)\mathrm{d}\pi(t), \qquad \iint k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') \qquad (**)$$

MCMC? Compute the denominator $(*)$ with Bayesian Quadrature first?

To address these problems we will instead go to some effort to force $(**)$ to have a closed form... via Stein's Method.

Extension 2: Unknown Normalisation Constant

Integrals with densities known up to normalisation

$$\int x(t)\mathrm{d}\pi(t), \quad \tilde{\pi} \propto \pi$$

occur in applications of Bayesian statistical methods:

$$p(\text{params}|\text{data}) \quad = \quad \frac{p(\text{data}|\text{params})\ p(\text{params})}{\int p(\text{data}|\text{params})\ \mathrm{d}p(\text{params})} \quad \begin{array}{l} \leftarrow \tilde{\pi} \\ \leftarrow \text{unknown } (*) \end{array}$$

Cannot compute with Bayesian quadrature, since relies on the following integrals having a closed form:

$$\int k(\cdot, t)\mathrm{d}\pi(t), \qquad \iint k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') \qquad (**)$$

MCMC? Compute the denominator $(*)$ with Bayesian Quadrature first?

To address these problems we will instead go to some effort to force $(**)$ to have a closed form... via Stein's Method.

A BOUND FOR THE ERROR IN THE NORMAL APPROXIMATION TO THE DISTRIBUTION OF A SUM OF DEPENDENT RANDOM VARIABLES

CHARLES STEIN
STANFORD UNIVERSITY

A Brief History of Stein

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

A Brief History of Stein

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

A Brief History of Stein

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

A Brief History of Stein

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

A Brief History of Stein

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \dots$ is a stationary sequence of random variables.

- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

Original aim was a central limit theorem for correlated variables:

Stein, 1972

Suppose $X_1, X_2, \ldots$ is a stationary sequence of random variables.

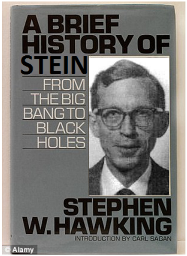
- Choose $A, B \subset \mathbb{N}$ such that $\inf_{i \in A, j \in B} |i - j| \geq k$.
- Choose arbitrary functions $Y \equiv Y(X_A)$, $Z \equiv Z(X_B)$.
- Assume that there exists α_k such that, for all such choices, $|\mathrm{Corr}(Y, Z)| \leq \alpha_k$.
- Assume that, for k sufficiently large, $\alpha_k \leq e^{-\lambda k}$.

Then

$$\left| \mathbb{P}\left[\frac{\sum_{i=1}^n X_i}{(\mathbb{V}(\sum_{i=1}^n X_i))^{1/2}} \leq a \right] - \Phi(a) \right| = O(n^{-1/2}).$$

A specific approach that led to some general methods for bounding the distance $d(\pi', \pi)$ between two distributions π, π'.

"I regret that, in order to complete this paper in time for publication, I have been forced to submit it with many defects remaining. In particular the proof of the concrete results of Section 3 is somewhat incomplete."



The essence of Stein's method is most clearly distilled in Ley et al. [2017]:

A p.d.f. π is characterised by the pair $(\mathcal{S}, \mathcal{F})$, consisting of a Stein Operator $\mathcal{S}$ and a Stein Class $\mathcal{F}$, if it holds that

$$X \sim \pi \quad \text{iff} \quad \mathbb{E}[\mathcal{S}f(X)] = 0 \quad \forall f \in \mathcal{F}.$$

Example 1 (Stein, 1972)

- π is the p.d.f. for $N(\mu, \sigma^2)$
- $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$
- $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} \text{ s.t. } f\pi \in W^{1,1} \text{ and } \lim_{t \searrow -\infty} f(t)\pi(t) = \lim_{t \nearrow +\infty} f(t)\pi(t)\}$.

The essence of Stein's method is most clearly distilled in Ley et al. [2017]:

A p.d.f. π is characterised by the pair $(\mathcal{S}, \mathcal{F})$, consisting of a Stein Operator $\mathcal{S}$ and a Stein Class $\mathcal{F}$, if it holds that

$$X \sim \pi \quad \text{iff} \quad \mathbb{E}[\mathcal{S}f(X)] = 0 \quad \forall f \in \mathcal{F}.$$

Example 1 (Stein, 1972)

- π is the p.d.f. for $N(\mu, \sigma^2)$
- $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$
- $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} \text{ s.t. } f\pi \in W^{1,1} \text{ and } \lim_{t \searrow -\infty} f(t)\pi(t) = \lim_{t \nearrow +\infty} f(t)\pi(t)\}$.

The essence of Stein's method is most clearly distilled in Ley et al. [2017]:

A p.d.f. π is characterised by the pair $(\mathcal{S}, \mathcal{F})$, consisting of a Stein Operator $\mathcal{S}$ and a Stein Class $\mathcal{F}$, if it holds that

$$X \sim \pi \quad \text{iff} \quad \mathbb{E}[\mathcal{S}f(X)] = 0 \quad \forall f \in \mathcal{F}.$$

Example 1 (Stein, 1972)

- π is the p.d.f. for $N(\mu, \sigma^2)$
- $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$
- $\mathcal{F} = \{f : \mathbb{R} \to \mathbb{R} \text{ s.t. } f\pi \in W^{1,1} \text{ and } \lim_{t \searrow -\infty} f(t)\pi(t) = \lim_{t \nearrow +\infty} f(t)\pi(t)\}$.

Our aim is to build a kernel k for which

$$\int_D k(\cdot, t)\mathrm{d}\pi(t) \quad = 0 \qquad \iint_{D \times D} k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') = 0$$

each have a (trivial) closed form, via Stein's method.

The kernel k will be associated with a RKHS of functions - this will be the set $\mathcal{SF}$ - that can be used within the Bayesian Quadrature method.

Full details in Oates et al. [2016a].

Extension 2: Unknown Normalisation Constant

Our aim is to build a kernel k for which

$$\int_D k(\cdot, t)\mathrm{d}\pi(t) \quad = 0 \qquad \iint_{D\times D} k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') = 0$$

each have a (trivial) closed form, via Stein's method.

The kernel k will be associated with a RKHS of functions - this will be the set $\mathcal{SF}$ - that can be used within the Bayesian Quadrature method.

Full details in Oates et al. [2016a].

Extension 2: Unknown Normalisation Constant

Our aim is to build a kernel k for which

$$\int_D k(\cdot, t)\mathrm{d}\pi(t) \quad = 0 \qquad \iint_{D \times D} k(t, t')\mathrm{d}(\pi \times \pi)(t \times t') = 0$$

each have a (trivial) closed form, via Stein's method.

The kernel k will be associated with a RKHS of functions - this will be the set $\mathcal{SF}$ - that can be used within the Bayesian Quadrature method.

Full details in Oates et al. [2016a].

Extension 2: Unknown Normalisation Constant

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$\begin{aligned} k(t,t') \;=\; \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') \;&=\; \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\ &\quad + \frac{\nabla_{t'}\pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'}\pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t'). \end{aligned}$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned} \int_D k(\cdot,t)\mathrm{d}\pi(t) \;&=\; \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &=\; \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &=\; \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \;\overset{k_{\mathcal{F}}\text{ "suff. reg."}}{=}\; \mathcal{S}.0 \;=\; 0 \end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$
\begin{aligned}
k(t,t') = \mathcal{S}_t \mathcal{S}_{t'} k_{\mathcal{F}}(t,t') &= \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\
&\quad + \frac{\nabla_{t'} \pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'} \pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t').
\end{aligned}
$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$
\begin{aligned}
\int_D k(\cdot,t)\mathrm{d}\pi(t) &= \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\
&= \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\
&= \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \overset{k_{\mathcal{F}} \text{ ``suff. reg.''}}{=} \mathcal{S}.0 = 0
\end{aligned}
$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{SF}$ can be endowed with RKHS structure, with kernel:

$$k(t,t') = \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') = \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_{t'} \pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'} \pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t').$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned}\int_D k(\cdot,t)\mathrm{d}\pi(t) &= \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \overset{k_{\mathcal{F}} \text{ “suff. reg.”}}{=} \mathcal{S}.0 = 0\end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Extension 2: Unknown Normalisation Constant

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$\begin{aligned} k(t,t') = \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') &= \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\ &+ \frac{\nabla_{t'}\pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'}\pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t'). \end{aligned}$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned} \int_D k(\cdot,t)\mathrm{d}\pi(t) &= \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \;\overset{k_{\mathcal{F}}\text{ ``suff. reg.''}}{=}\; \mathcal{S}.0 = 0 \end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Extension 2: Unknown Normalisation Constant

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$
\begin{aligned}
k(t,t') = \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') \quad &= \quad \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\
&\quad + \frac{\nabla_{t'}\pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'}\pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t').
\end{aligned}
$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$
\begin{aligned}
\int_D k(\cdot,t)\mathrm{d}\pi(t) \quad &= \quad \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\
&= \quad \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\
&= \quad \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \;\overset{k_{\mathcal{F}} \text{ “suff. reg.”}}{=}\; \mathcal{S}.0 \;=\; 0
\end{aligned}
$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Extension 2: Unknown Normalisation Constant

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$\begin{aligned} k(t,t') = \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') &= \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\ &+ \frac{\nabla_{t'}\pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'}\pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t'). \end{aligned}$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned} \int_D k(\cdot,t)\mathrm{d}\pi(t) &= \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}. \int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}. \oint_{\partial D} k_{\mathcal{F}}(\cdot,t) \cdot n(t)\mathrm{d}\pi(t) \overset{k_{\mathcal{F}} \text{ "suff. reg."}}{=} \mathcal{S}.0 = 0 \end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$\begin{aligned} k(t,t') = \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') &= \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\ &+ \frac{\nabla_{t'}\pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'}\pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t'). \end{aligned}$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned} \int_D k(\cdot,t)\mathrm{d}\pi(t) &= \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}. \int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}. \oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \overset{k_{\mathcal{F}} \text{ ``suff. reg.''}}{=} \mathcal{S}.0 = 0 \end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Extension 2: Unknown Normalisation Constant

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$\begin{aligned} k(t,t') = \mathcal{S}_t \mathcal{S}_{t'} k_{\mathcal{F}}(t,t') &= \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\ &+ \frac{\nabla_{t'} \pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'} \pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t'). \end{aligned}$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned} \int_D k(\cdot,t)\mathrm{d}\pi(t) &= \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \overset{k_{\mathcal{F}} \text{ ``suff. reg.''}}{=} \mathcal{S}.0 = 0 \end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Extension 2: Unknown Normalisation Constant

Let $\mathcal{S} : f \mapsto \nabla(f\pi)/\pi$ and let $\mathcal{F}$ be an RKHS with kernel $k_{\mathcal{F}}$.

Then (if $k_{\mathcal{F}}$ is sufficiently regular) the set $\mathcal{S}\mathcal{F}$ can be endowed with RKHS structure, with kernel:

$$\begin{aligned} k(t,t') \; = \; \mathcal{S}_t\mathcal{S}_{t'}k_{\mathcal{F}}(t,t') \quad &= \quad \nabla_t \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \nabla_{t'} k_{\mathcal{F}}(t,t') \\ &\quad + \frac{\nabla_{t'}\pi(t')}{\pi(t')} \cdot \nabla_t k_{\mathcal{F}}(t,t') + \frac{\nabla_t \pi(t)}{\pi(t)} \cdot \frac{\nabla_{t'}\pi(t')}{\pi(t')} k_{\mathcal{F}}(t,t'). \end{aligned}$$

Note that k can be computed from $\tilde{\pi}$! Moreover,

$$\begin{aligned} \int_D k(\cdot,t)\mathrm{d}\pi(t) \quad &= \quad \int_D \mathcal{S}.\mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \quad \mathcal{S}.\int_D \mathcal{S}_t k_{\mathcal{F}}(\cdot,t)\mathrm{d}\pi(t) \\ &= \quad \mathcal{S}.\oint_{\partial D} k_{\mathcal{F}}(\cdot,t)\cdot n(t)\mathrm{d}\pi(t) \;\; \overset{k_{\mathcal{F}} \text{ ``suff. reg.''}}{=} \;\; \mathcal{S}.0 \; = \; 0 \end{aligned}$$

Detail: The kernel $1 + k(t,t')$ is actually used for Bayesian Quadrature (to catch mean-shift).

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0} \left(\int x(t) \mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}} \right)^2 = O\left(n^{-1-\frac{2(a \wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0} \left(\int x(t) \mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}} \right)^2 = O\left(n^{-1-\frac{2(a \wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Extension 2: Unknown Normalisation Constant

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0}\left(\int x(t)\mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}}\right)^2 = O\left(n^{-1-\frac{2(a\wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Extension 2: Unknown Normalisation Constant

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0}\left(\int x(t)\mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}}\right)^2 = O\left(n^{-1-\frac{2(a \wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Extension 2: Unknown Normalisation Constant

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0}\left(\int x(t)\mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}}\right)^2 = O\left(n^{-1-\frac{2(a \wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0}\left(\int x(t)\mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}}\right)^2 = O\left(n^{-1-\frac{2(a \wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Suppose $\{t_i\}_{i=1}^n$ arise from a Markov chain that targets π.

- Assume D is bounded.
- Assume π is bounded away from 0 on D.
- Assume $\pi \in C^{2a+1}(D)$ and $k_{\mathcal{F}} \in C^{2b+2}(D \times D)$.
- Assume $k_{\mathcal{F}}$ is "sufficiently regular".
- Assume the Markov chain is uniformly ergodic.

Then, for $x \in \mathcal{SF}$, there exists $h_0 > 0$ such that

$$1_{h<h_0} \left(\int x(t) \mathrm{d}\pi(t) - \underbrace{b(a)}_{\text{BQ estimator}} \right)^2 = O\left(n^{-1-\frac{2(a \wedge b)}{d}+\epsilon}\right),$$

for arbitrary $\epsilon > 0$. Here h is the fill distance of $\{t_i\}_{i=1}^n$.

Full details in Oates et al. [2016b].

Extension 2: Unknown Normalisation Constant

Consider again Darcy's PDE

$$\begin{aligned} \nabla_t \cdot [c(t;\boldsymbol{\theta})\nabla_t x(t)] &= 0 && \text{if } t_1, t_2 \in (0,1) \\ x(t) &= \begin{cases} t_1 & \text{if } t_2 = 0 \\ 1 - t_1 & \text{if } t_2 = 1 \end{cases} \\ \nabla_{t_1} x(t) &= 0 && \text{if } t_1 \in \{0,1\}, \end{aligned}$$

Data are a grid of observations $y_{i,j} = x(t_{i,j}) + \epsilon_{i,j}$ and IID $\epsilon_{i,j} \sim N(0,\sigma^2)$. The field c is endowed with a prior

$$\log c(t;\boldsymbol{\theta}) = \sum_{i=1}^{d} \theta_i c_i(t),$$

where $\boldsymbol{\theta} \sim \mathrm{Unif}(D)$, $D = (-10,10)^d$ and c_i are orthonormal.

Aim: Estimate the posterior mean of the parameter $\boldsymbol{\theta}$.

Approach: Bayesian Probabilistic Numerical Method for the likelihood $\mathcal{L}_n(\boldsymbol{\theta};\boldsymbol{y})$ (to avoid exact solution of the PDE), followed by Stein's method for integration with respect to $\pi(\boldsymbol{\theta}) \propto \mathcal{L}_n(\boldsymbol{\theta};\boldsymbol{y})$.

Consider again Darcy's PDE

$$\begin{aligned} \nabla_t \cdot [c(t;\boldsymbol{\theta})\nabla_t x(t)] &= 0 && \text{if } t_1, t_2 \in (0,1) \\ x(t) &= \begin{cases} t_1 & \text{if } t_2 = 0 \\ 1 - t_1 & \text{if } t_2 = 1 \end{cases} \\ \nabla_{t_1} x(t) &= 0 && \text{if } t_1 \in \{0,1\}, \end{aligned}$$

Data are a grid of observations $y_{i,j} = x(t_{i,j}) + \epsilon_{i,j}$ and IID $\epsilon_{i,j} \sim N(0,\sigma^2)$. The field c is endowed with a prior

$$\log c(t;\boldsymbol{\theta}) = \sum_{i=1}^{d} \theta_i c_i(t),$$

where $\boldsymbol{\theta} \sim \text{Unif}(D)$, $D = (-10,10)^d$ and c_i are orthonormal.

Aim: Estimate the posterior mean of the parameter $\boldsymbol{\theta}$.

Approach: Bayesian Probabilistic Numerical Method for the likelihood $\mathcal{L}_n(\boldsymbol{\theta};\boldsymbol{y})$ (to avoid exact solution of the PDE), followed by Stein's method for integration with respect to $\pi(\boldsymbol{\theta}) \propto \mathcal{L}_n(\boldsymbol{\theta};\boldsymbol{y})$.

Consider again Darcy's PDE

$$\begin{aligned} \nabla_t \cdot [c(t;\boldsymbol{\theta})\nabla_t x(t)] &= 0 && \text{if } t_1, t_2 \in (0,1) \\ x(t) &= \begin{cases} t_1 & \text{if } t_2 = 0 \\ 1 - t_1 & \text{if } t_2 = 1 \end{cases} \\ \nabla_{t_1} x(t) &= 0 && \text{if } t_1 \in \{0,1\}, \end{aligned}$$

Data are a grid of observations $y_{i,j} = x(t_{i,j}) + \epsilon_{i,j}$ and IID $\epsilon_{i,j} \sim N(0, \sigma^2)$. The field c is endowed with a prior

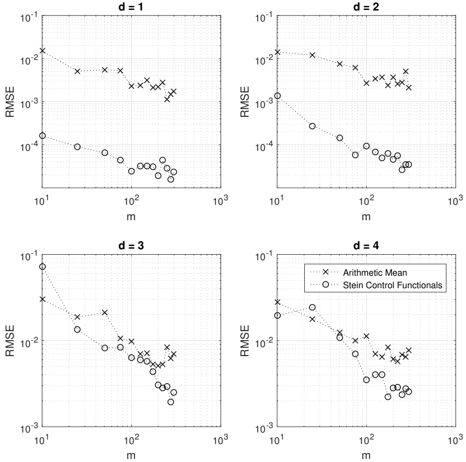
$$\log c(t;\boldsymbol{\theta}) = \sum_{i=1}^{d} \theta_i c_i(t),$$

where $\boldsymbol{\theta} \sim \text{Unif}(D)$, $D = (-10, 10)^d$ and c_i are orthonormal.

Aim: Estimate the posterior mean of the parameter $\boldsymbol{\theta}$.

Approach: Bayesian Probabilistic Numerical Method for the likelihood $\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{y})$ (to avoid exact solution of the PDE), followed by Stein's method for integration with respect to $\pi(\boldsymbol{\theta}) \propto \mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{y})$.

Performance of Bayesian Quadrature (via Stein's method) for estimation of $\int \theta_1 \mathrm{d}\pi(\boldsymbol{\theta})$:



Here m is the number of PDE forward-solves used.

Extension 3: Unknown p.d.f. π

Of course, knowing $\tilde{\pi}$ is mathematically equivalent to knowing π.

Consider now the situation where $t_i \sim \pi$ are IID and that is all that is known.

Idea:

Model both the integrand x and the p.d.f. π as unknown objects:

- $x \sim \mathcal{GP}$ (Gaussian process model - standard BQ)
- $\pi(t) = \int \psi(t; \varphi) P(\mathrm{d}\varphi)$ (hierarchical mixture model)
- $P \sim \mathcal{DP}(\alpha, P_0)$ (Dirichlet process model)

Extension 3: Unknown p.d.f. π

Of course, knowing $\tilde{\pi}$ is mathematically equivalent to knowing π.

Consider now the situation where $t_i \sim \pi$ are IID and that is all that is known.

Idea:

Model both the integrand x and the p.d.f. π as unknown objects:

- $x \sim \mathcal{GP}$ (Gaussian process model - standard BQ)
- $\pi(t) = \int \psi(t; \varphi) P(\mathrm{d}\varphi)$ (hierarchical mixture model)
- $P \sim \mathcal{DP}(\alpha, P_0)$ (Dirichlet process model)

Extension 3: Unknown p.d.f. π

Of course, knowing $\tilde{\pi}$ is mathematically equivalent to knowing π.

Consider now the situation where $t_i \sim \pi$ are IID and that is all that is known.

Idea:

Model both the integrand x and the p.d.f. π as unknown objects:

- $x \sim \mathcal{GP}$ (Gaussian process model - standard BQ)
- $\pi(t) = \int \psi(t; \varphi) P(\mathrm{d}\varphi)$ (hierarchical mixture model)
- $P \sim \mathcal{DP}(\alpha, P_0)$ (Dirichlet process model)

Recall: $P \sim \mathcal{DP}(\alpha, P_0)$ iff $(P(B_1), \dots, P(B_m)) \sim \mathrm{Dir}(\alpha P_0(B_1), \dots, \alpha P_0(B_m))$

Extension 3: Unknown p.d.f. π

Of course, knowing $\tilde{\pi}$ is mathematically equivalent to knowing π.

Consider now the situation where $t_i \sim \pi$ are IID and that is all that is known.

Idea:

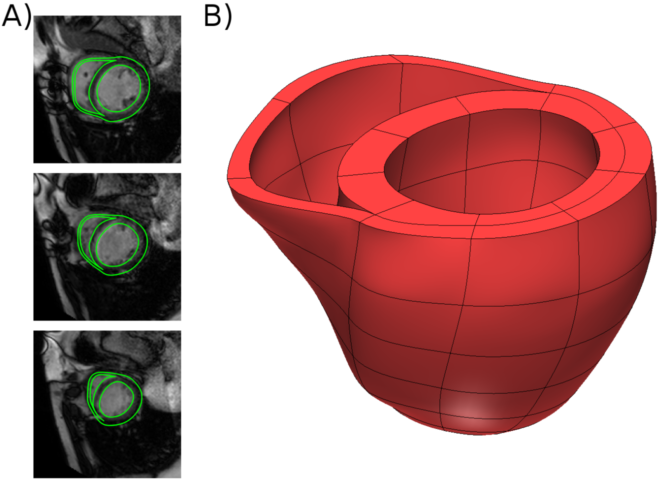
Model both the integrand x and the p.d.f. π as unknown objects:

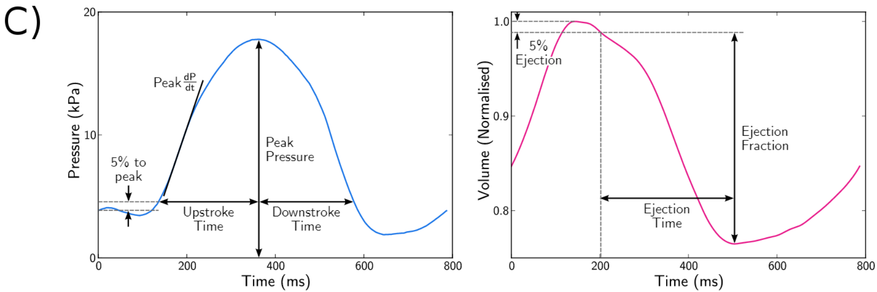
- $x \sim \mathcal{GP}$ (Gaussian process model - standard BQ)
- $\pi(t) = \int \psi(t; \varphi) P(\mathrm{d}\varphi)$ (hierarchical mixture model)
- $P \sim \mathcal{DP}(\alpha, P_0)$ (Dirichlet process model)

Then condition x on data $\{(t_i, x(t_i))\}_{i=1}^n$ and condition π on data $\{t_i\}_{i=1}^n$.

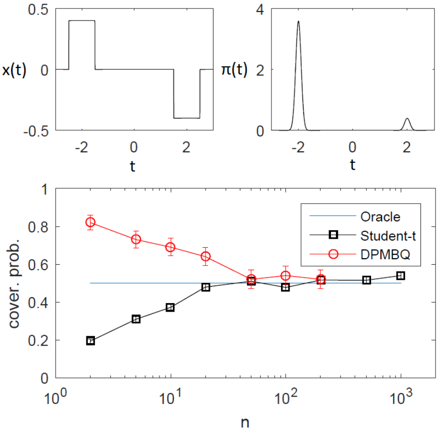
This implies to a posterior distribution over the integral $\int x(t) \mathrm{d}\pi(t)$ that accounts for uncertainty regarding both x and π.

Motivation: Assessment of Cardiac Models




C)
20
10
0
Pressure (kPa)
Peak $\frac{dP}{dt}$
5% to peak
Upstroke Time
Peak Pressure
Downstroke Time
0
200
400
600
800
Time (ms)
1.0
0.9
0.8
Volume (Normalised)
5% Ejection
Ejection Fraction
Ejection Time
0
200
400
600
800
Time (ms)

Motivation: Assessment of Cardiac Models



Suppose that:

- x belongs to the RKHS associated to a kernel k, bounded on $D \times D$, $D \subset \mathbb{R}$.
- $\pi(\cdot)$ is a location-scale mixture of Gaussians; $\psi(\cdot;\varphi) = \mathrm{N}(\cdot;\varphi_1,\varphi_2)$.
- Technical conditions on the Dirichlet process:
  - $\varphi_1 \in \mathbb{R}$ and $\varphi_2 \in [\underline{\sigma},\overline{\sigma}]$ for fixed $\underline{\sigma},\overline{\sigma} \in (0,\infty)$.
  - P, the true mixing distribution, has compact $\mathrm{supp}(P) \subset \mathbb{R} \times (\underline{\sigma},\overline{\sigma})$.
  - P_0 has positive and continuous density on a rectangle R such that $\mathrm{supp}(P_0) \subseteq R \subseteq \mathbb{R} \times [\underline{\sigma},\overline{\sigma}]$.
  - P_0 satisfies the tail condition $P_0(\{(\varphi_1,\varphi_2) : |\varphi_1| > t\}) \leq c\exp(-b|t|^\delta)$ for all $t > 0$.

Then the posterior distribution over the unknown value of the integral converges to the truth in Wasserstein metric at the rate

$$O_P(n^{-1/4+\epsilon}).$$

(Recall: $d_{\mathrm{Wass}} = \int |\theta - \theta_0| p_n(\theta) \mathrm{d}\theta$ where θ_0 is the true value of θ.)

Extension 3: Unknown p.d.f. π

Suppose that:

- x belongs to the RKHS associated to a kernel k, bounded on $D \times D$, $D \subset \mathbb{R}$.
- $\pi(\cdot)$ is a location-scale mixture of Gaussians; $\psi(\cdot;\varphi) = \mathrm{N}(\cdot;\varphi_1,\varphi_2)$.
- Technical conditions on the Dirichlet process:
  - $\varphi_1 \in \mathbb{R}$ and $\varphi_2 \in [\underline{\sigma}, \overline{\sigma}]$ for fixed $\underline{\sigma}, \overline{\sigma} \in (0,\infty)$.
  - P, the true mixing distribution, has compact $\mathrm{supp}(P) \subset \mathbb{R} \times (\underline{\sigma}, \overline{\sigma})$.
  - P_0 has positive and continuous density on a rectangle R such that $\mathrm{supp}(P_0) \subseteq R \subseteq \mathbb{R} \times [\underline{\sigma}, \overline{\sigma}]$.
  - P_0 satisfies the tail condition $P_0(\{(\varphi_1,\varphi_2) : |\varphi_1| > t\}) \leq c \exp(-b|t|^\delta)$ for all $t > 0$.

Then the posterior distribution over the unknown value of the integral converges to the truth in Wasserstein metric at the rate

$$O_P(n^{-1/4+\epsilon}).$$

(Recall: $d_{\mathrm{Wass}} = \int |\theta - \theta_0| p_n(\theta) \mathrm{d}\theta$ where θ_0 is the true value of θ.)

Extension 3: Unknown p.d.f. π

Suppose that:

- x belongs to the RKHS associated to a kernel k, bounded on $D \times D$, $D \subset \mathbb{R}$.
- $\pi(\cdot)$ is a location-scale mixture of Gaussians; $\psi(\cdot; \varphi) = \mathrm{N}(\cdot; \varphi_1, \varphi_2)$.
- Technical conditions on the Dirichlet process:
  - $\varphi_1 \in \mathbb{R}$ and $\varphi_2 \in [\underline{\sigma}, \overline{\sigma}]$ for fixed $\underline{\sigma}, \overline{\sigma} \in (0, \infty)$.
  - P, the true mixing distribution, has compact $\mathrm{supp}(P) \subset \mathbb{R} \times (\underline{\sigma}, \overline{\sigma})$.
  - P_0 has positive and continuous density on a rectangle R such that $\mathrm{supp}(P_0) \subseteq R \subseteq \mathbb{R} \times [\underline{\sigma}, \overline{\sigma}]$.
  - P_0 satisfies the tail condition $P_0(\{(\varphi_1, \varphi_2) : |\varphi_1| > t\}) \leq c \exp(-b|t|^\delta)$ for all $t > 0$.

Then the posterior distribution over the unknown value of the integral converges to the truth in Wasserstein metric at the rate

$$O_P(n^{-1/4+\epsilon}).$$

(Recall: $d_{\mathrm{Wass}} = \int |\theta - \theta_0| p_n(\theta) \mathrm{d}\theta$ where θ_0 is the true value of θ.)

Extension 3: Unknown p.d.f. π

Suppose that:

- x belongs to the RKHS associated to a kernel k, bounded on $D \times D$, $D \subset \mathbb{R}$.
- $\pi(\cdot)$ is a location-scale mixture of Gaussians; $\psi(\cdot;\varphi) = \mathrm{N}(\cdot;\varphi_1,\varphi_2)$.
- Technical conditions on the Dirichlet process:
  - $\varphi_1 \in \mathbb{R}$ and $\varphi_2 \in [\underline{\sigma}, \overline{\sigma}]$ for fixed $\underline{\sigma}, \overline{\sigma} \in (0,\infty)$.
  - P, the true mixing distribution, has compact $\mathrm{supp}(P) \subset \mathbb{R} \times (\underline{\sigma}, \overline{\sigma})$.
  - P_0 has positive and continuous density on a rectangle R such that $\mathrm{supp}(P_0) \subseteq R \subseteq \mathbb{R} \times [\underline{\sigma}, \overline{\sigma}]$.
  - P_0 satisfies the tail condition $P_0(\{(\varphi_1,\varphi_2) : |\varphi_1| > t\}) \leq c \exp(-b|t|^\delta)$ for all $t > 0$.

Then the posterior distribution over the unknown value of the integral converges to the truth in Wasserstein metric at the rate

$$O_P(n^{-1/4+\epsilon}).$$

(Recall: $d_{\mathrm{Wass}} = \int |\theta - \theta_0| p_n(\theta) \mathrm{d}\theta$ where θ_0 is the true value of θ.)

Extension 3: Unknown p.d.f. π

Suppose that:

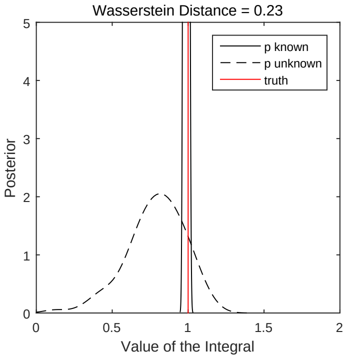
- x belongs to the RKHS associated to a kernel k, bounded on $D \times D$, $D \subset \mathbb{R}$.
- $\pi(\cdot)$ is a location-scale mixture of Gaussians; $\psi(\cdot;\varphi) = \mathrm{N}(\cdot;\varphi_1,\varphi_2)$.
- Technical conditions on the Dirichlet process:
  - $\varphi_1 \in \mathbb{R}$ and $\varphi_2 \in [\underline{\sigma}, \overline{\sigma}]$ for fixed $\underline{\sigma}, \overline{\sigma} \in (0,\infty)$.
  - P, the true mixing distribution, has compact $\mathrm{supp}(P) \subset \mathbb{R} \times (\underline{\sigma}, \overline{\sigma})$.
  - P_0 has positive and continuous density on a rectangle R such that $\mathrm{supp}(P_0) \subseteq R \subseteq \mathbb{R} \times [\underline{\sigma}, \overline{\sigma}]$.
  - P_0 satisfies the tail condition $P_0(\{(\varphi_1,\varphi_2) \ : \ |\varphi_1| > t\}) \leq c\exp(-b|t|^\delta)$ for all $t > 0$.

Then the posterior distribution over the unknown value of the integral converges to the truth in Wasserstein metric at the rate

$$O_P(n^{-1/4+\epsilon}).$$

(Recall: $d_{\mathrm{Wass}} = \int |\theta - \theta_0| p_n(\theta) \mathrm{d}\theta$ where θ_0 is the true value of θ.)

Extension 3: Unknown p.d.f. π



(a) $n = 10$ (b) $n = 100$

(Implementation is straight-forward with a stick-breaking construction. Exploits well-known conjugacy results for DP mixture models; Oates et al. [2017].)

Eleventh Job: Non-Bayesian Methods?

Hull and Swenson [1966] and others supposed that rounding, i.e. representation of a real number

$$x = 0.a_1 a_2 a_3 a_4 \dots \quad \in [0,1]$$

in a truncated form

$$\hat{x} = 0.a_1 a_2 a_3 a_4 \dots a_n,$$

is such that the error $e = x - \hat{x}$ can be reasonably modelled by a uniform random variable

$$e \sim \mathrm{Unif}(-5 \times 10^{-(n+1)}, 5 \times 10^{-(n+1)}).$$

This implies a distribution over the unknown value of x.

The proposal of Hull and Swenson [1966] and others was to replace the last digit a_n, in each stored number that arises in the numerical solution of an ODE, with a uniformly chosen element of $\{0, \dots, 9\}$.

NB: This work focused on rounding error, rather than e.g. the (time) discretisation error that is intrinsic to numerical ODE solvers; this could reflect the limited precision arithmetic that was available from the computer hardware of the period.

Hull and Swenson [1966] and others supposed that rounding, i.e. representation of a real number

$$x = 0.a_1 a_2 a_3 a_4 \dots \quad \in [0, 1]$$

in a truncated form

$$\hat{x} = 0.a_1 a_2 a_3 a_4 \dots a_n,$$

is such that the error $e = x - \hat{x}$ can be reasonably modelled by a uniform random variable

$$e \sim \mathrm{Unif}(-5 \times 10^{-(n+1)}, 5 \times 10^{-(n+1)}).$$

This implies a distribution over the unknown value of x.

The proposal of Hull and Swenson [1966] and others was to replace the last digit a_n, in each stored number that arises in the numerical solution of an ODE, with a uniformly chosen element of $\{0, \dots, 9\}$.

NB: This work focused on rounding error, rather than e.g. the (time) discretisation error that is intrinsic to numerical ODE solvers; this could reflect the limited precision arithmetic that was available from the computer hardware of the period.

Hull and Swenson [1966] and others supposed that rounding, i.e. representation of a real number

$$x = 0.a_1 a_2 a_3 a_4 \ldots \quad \in [0,1]$$

in a truncated form

$$\hat{x} = 0.a_1 a_2 a_3 a_4 \ldots a_n,$$

is such that the error $e = x - \hat{x}$ can be reasonably modelled by a uniform random variable

$$e \sim \text{Unif}(-5 \times 10^{-(n+1)}, 5 \times 10^{-(n+1)}).$$

This implies a distribution over the unknown value of x.

The proposal of Hull and Swenson [1966] and others was to replace the last digit a_n, in each stored number that arises in the numerical solution of an ODE, with a uniformly chosen element of $\{0, \ldots, 9\}$.

NB: This work focused on rounding error, rather than e.g. the (time) discretisation error that is intrinsic to numerical ODE solvers; this could reflect the limited precision arithmetic that was available from the computer hardware of the period.

Conrad et al. [2016] and others supposed that discretisation, i.e. representation of a infinite-dimensional object

$$x(\cdot)$$

in a discrete form

$$\hat{x}(\cdot) = a_1\phi_1(\cdot) + \cdots + a_m\phi_m(\cdot)$$

is such that the error $e = x - \hat{x}$ can be reasonably modelled by a random process, such as a Gaussian process:

$$e \sim \mathcal{GP}(0, k_e).$$

This implies a distribution over the unknown value of x.

In particular, when ϕ_i are finite elements, we can model

$$e(\cdot) = a_1 e_1(\cdot) + \cdots + a_m e_m(\cdot)$$

where e_i is a Gaussian process constrained to share the same support as ϕ_i and vanish at nodal points. This enables to "trivial" modification of finite element methods. (i.e. "Randomise the finite elements"; $\phi_i \mapsto \phi_i + e_i$.)

Conrad et al. [2016] and others supposed that discretisation, i.e. representation of a infinite-dimensional object

$$x(\cdot)$$

in a discrete form

$$\hat{x}(\cdot) = a_1\phi_1(\cdot) + \cdots + a_m\phi_m(\cdot)$$

is such that the error $e = x - \hat{x}$ can be reasonably modelled by a random process, such as a Gaussian process:

$$e \sim \mathcal{GP}(0, k_e).$$

This implies a distribution over the unknown value of x.

In particular, when ϕ_i are finite elements, we can model

$$e(\cdot) = a_1 e_1(\cdot) + \cdots + a_m e_m(\cdot)$$

where e_i is a Gaussian process constrained to share the same support as ϕ_i and vanish at nodal points. This enables to "trivial" modification of finite element methods. (i.e. "Randomise the finite elements"; $\phi_i \mapsto \phi_i + e_i$.)

Properties of (some) non-Bayesian methods:

- Often trivial modification of classical code, to "inject noise"
- Computationally competitive with classical methods
- However, simple models for error e can be inappropriate - and controversial

Properties of (some) Bayesian methods:

- Statistically well-founded
- Coherent framework in which to combine methods (see Part VI)
- Computationally very expensive (at the moment)

Properties of (some) non-Bayesian methods:

- Often trivial modification of classical code, to "inject noise"
- Computationally competitive with classical methods
- However, simple models for error e can be inappropriate - and controversial

Properties of (some) Bayesian methods:

- Statistically well-founded
- Coherent framework in which to combine methods (see Part VI)
- Computationally very expensive (at the moment)

Conclusion

In Part V it has been argued that:

- Several extensions of Bayesian Quadrature can be developed.
- Dirichlet process mixture models are a convenient means to construct a non-parametric distribution on the space of p.d.f.s π.
- Non-Bayesian probabilistic numerical methods have been developed - but are rather different to Bayesian probabilistic numerical methods (more like a perturbation analysis?)

Open question: In what sense are filtering methods for ODEs an (approximate) Bayesian method?

END OF PART V

Conclusion

In Part V it has been argued that:

- Several extensions of Bayesian Quadrature can be developed.
- Dirichlet process mixture models are a convenient means to construct a non-parametric distribution on the space of p.d.f.s π.
- Non-Bayesian probabilistic numerical methods have been developed - but are rather different to Bayesian probabilistic numerical methods (more like a perturbation analysis?)

Open question: In what sense are filtering methods for ODEs an (approximate) Bayesian method?

END OF PART V

Conclusion

In Part V it has been argued that:

- Several extensions of Bayesian Quadrature can be developed.
- Dirichlet process mixture models are a convenient means to construct a non-parametric distribution on the space of p.d.f.s π.
- Non-Bayesian probabilistic numerical methods have been developed - but are rather different to Bayesian probabilistic numerical methods (more like a perturbation analysis?)

Open question: In what sense are filtering methods for ODEs an (approximate) Bayesian method?

END OF PART V

Conclusion

In Part V it has been argued that:

- Several extensions of Bayesian Quadrature can be developed.
- Dirichlet process mixture models are a convenient means to construct a non-parametric distribution on the space of p.d.f.s π.
- Non-Bayesian probabilistic numerical methods have been developed - but are rather different to Bayesian probabilistic numerical methods (more like a perturbation analysis?)

Open question: In what sense are filtering methods for ODEs an (approximate) Bayesian method?

END OF PART V

Conclusion

In Part V it has been argued that:

- Several extensions of Bayesian Quadrature can be developed.
- Dirichlet process mixture models are a convenient means to construct a non-parametric distribution on the space of p.d.f.s π.
- Non-Bayesian probabilistic numerical methods have been developed - but are rather different to Bayesian probabilistic numerical methods (more like a perturbation analysis?)

Open question: In what sense are filtering methods for ODEs an (approximate) Bayesian method?

END OF PART V

In Part V it has been argued that:

- Several extensions of Bayesian Quadrature can be developed.
- Dirichlet process mixture models are a convenient means to construct a non-parametric distribution on the space of p.d.f.s π.
- Non-Bayesian probabilistic numerical methods have been developed - but are rather different to Bayesian probabilistic numerical methods (more like a perturbation analysis?)

Open question: In what sense are filtering methods for ODEs an (approximate) Bayesian method?

END OF PART V