MANCHESTER 1824

The University of Manchester

M\cr NA
Manchester Numerical Analysis

MCMC for Bayesian Inverse Problems

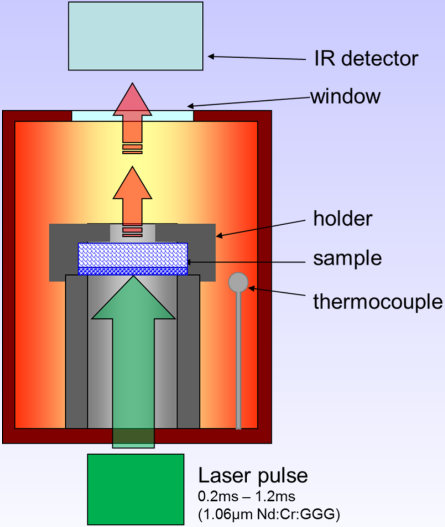
James Rynn
School of Mathematics
The University of Manchester

james.rynn@manchester.ac.uk

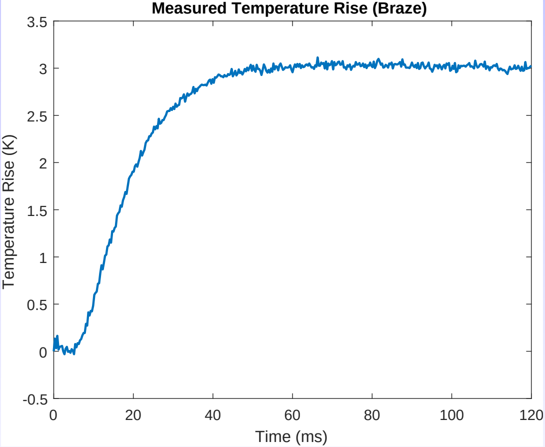
Informal Applied, 24/03/17

Outline

Motivation


IR detector
window
holder
sample
thermocouple
Laser pulse
0.2ms – 1.2ms
(1.06µm Nd:Cr:GGG)

Motivation


Measured Temperature Rise (Braze)
Temperature Rise (K)
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
0
20
40
60
80
100
120
Time (ms)

Motivation

Unknowns:

- λ — thermal conductivity,
- I — laser intensity,
- k — boundary condition parameter,
- σ — measurement noise,

Distributions

What is a distribution?

It describes the probability of an event occurring/random variable taking a certain value.

We write $X \sim \mathcal{D}$ to denote that the random variable X follows the distribution $\mathcal{D}$ and

$$\mathbb{P}(X \in A) = \int_{\mathcal{X}} \mathbf{1}_{\{x \in A\}} f_X(x) \mathrm{d}x = \int_A f_X(x) \mathrm{d}x,$$

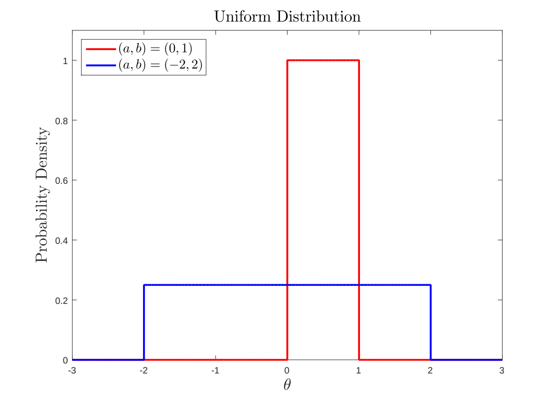
to denote the probability of the event A occurring.

Here $f_X(x)$ is the probability density function of the random variable X, which uniquely determines the distribution of X.

Some common distributions are:

- Uniform Distribution

$$X \sim \mathcal{U}(a, b), \qquad f_X(x) = \frac{1}{b-a}.$$

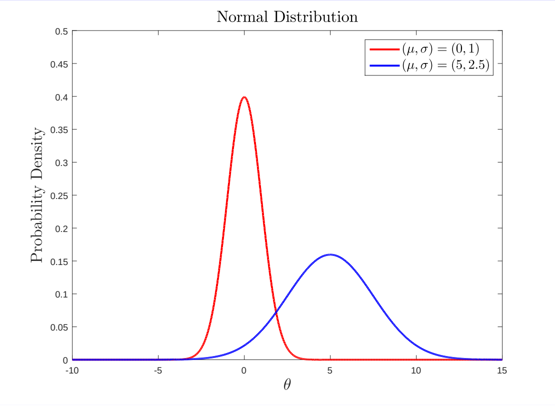


Distributions

Some common distributions are:

- Normal Distribution

$$X \sim \mathcal{N}(\mu, \sigma^2), \qquad f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

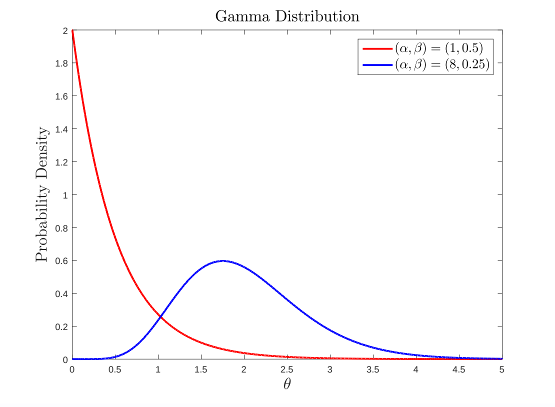


Distributions

Some common distributions are:

- Gamma Distribution

$$X \sim \text{Gamma}(\alpha, \beta), \qquad f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$



Monte Carlo Sampling

We often want to compute quantities of interest about the random variable X, for example

$$\mathbb{E}[X] = \int_{\mathcal{X}} x f_X(x) \mathrm{d}x,$$
$$\mathrm{Var}(X) = \int_{\mathcal{X}} (x - \mu)^2 f_X(x) \mathrm{d}x,$$
$$\mathbb{P}(X > a) = \int_{\mathcal{X}} \mathbf{1}_{\{x>a\}} f_X(x) \mathrm{d}x.$$

What if this integral is intractable? We need to somehow approximate it.
Q: Can we use standard quadrature approaches?
A: No. . . , such routines are far too expensive in high dimensions!

Monte Carlo Sampling

The answer is to use the method of Monte Carlo sampling. Simply put, to do this we sample from the given distribution of X and approximate the integral as a sum.

For example, to estimate the expectation $\mathbb{E}[X]$, using samples $x_1, x_2, \ldots, x_N$ drawn from the distribution of X we use the approximation

$$\mathbb{E}[X] = \int_{\mathcal{X}} x f_X(x) \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} x_i,$$

and for a general function $\phi \colon \mathcal{X} \to \mathbb{R}$,

$$\int_{\mathcal{X}} \phi(x) f_X(x) \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} \phi(x_i).$$

Monte Carlo Sampling

Q: Why is this better?

A1: Does not suffer from the ‘curse of dimensionality’!

A2: We can always perform this form of estimate whenever we can sample from the distribution of X and evaluate the function ϕ.

The error in this technique is $\mathcal{O}(N^{-\frac{1}{2}})$.

Monte Carlo Sampling

Q: What if the distribution of X is difficult/impossible to generate from directly? What if the distribution of X is only known up to a constant of proportionality?

A: We then require the use of Markov chain Monte Carlo (MCMC) methods.

First, we revisit, the key properties of a Markov chain...

Markov Chains

A stochastic process $X = (X_k)_{k\in\mathbb{N}}$ is a discrete–time Markov chain if it satisfies the 'memoryless property':

$$\mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k, \ldots, X_1 = x_1) = \mathbb{P}(X_{k+1} = x_{k+1} \mid X_k = x_k).$$

A Markov chains movement through state space $\mathcal{X}$ is described by a transition density ν, which satisfies

$$\nu(x,y) \geq 0 \quad \forall\, x, y \in \mathcal{X},$$
$$\int_{\mathcal{X}} \nu(x,y)\,\mathrm{d}y = 1 \quad \forall\, x \in \mathcal{X}.$$

Under certain (technical) conditions, a Markov chain has a stationary distribution, π, satisfying

$$\int_{\mathcal{X}} \pi(x)\nu(x,y)\,\mathrm{d}x = \pi(y) \quad \forall\, y \in \mathcal{X}.$$

Markov Chain Monte Carlo

Suppose we wish to draw samples $(X_k)_{k\in\mathbb{N}}$ from a given distribution X with density π.
It can often be difficult to produce samples from a prescribed distribution, or we may only know the density π up to a constant of proportionality.

The next best thing to independent identically distributed (i.i.d.) samples are the states of a Markov chain, which have low correlation due to the memoryless property.

Key Idea: Producing a Markov chain with stationary density equal to π is equivalent to generating samples from a distribution with density π.

Markov Chain Monte Carlo

Markov chain Monte Carlo algorithms can be summed up as:

- generate proposals using the current state in a clever (efficient) way,
- accept the proposal as the next state or reject the proposal and remain at the current state in a clever way,
- repeat to produce the desired number of samples by taking these as the states of the chain once it has reached stationarity.

Markov Chain Monte Carlo

Now for an example in MATLAB. . .

Bayesian Inverse Problems

Find the unknown θ given n_{obs} observations $\boldsymbol{D}$, satisfying

$$\boldsymbol{D} = \mathcal{G}(\theta) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

where

- $\boldsymbol{D} \in \mathbb{R}^{n_{\text{obs}}}$ is a given vector of **observations**,
- $\mathcal{G} : \mathbb{R} \to \mathbb{R}^{n_{\text{obs}}}$ is the **observation operator**,
- θ is the **unknown**,
- $\boldsymbol{\eta} \in \mathbb{R}^{n_{\text{obs}}}$ is a vector of **observational noise**.

We treat this as a probabilistic problem and search for a posterior distribution for θ.

Bayesian Inverse Problems

Crucially, we have

$$\pi(\theta|\boldsymbol{D}) \propto \mathcal{L}(\boldsymbol{D}|\theta)\pi_0(\theta),$$

so we can sample from the posterior $\pi(\theta|\boldsymbol{D})$ using our favourite MCMC method!

PDE Example

Consider the transient heat equation with source term,

$$\frac{\partial u}{\partial t}(x,t) = \lambda \frac{\partial^2 u}{\partial x^2}(x,t) + Q(x,t), \quad x \in [0,H],\ t \in [0,T],$$

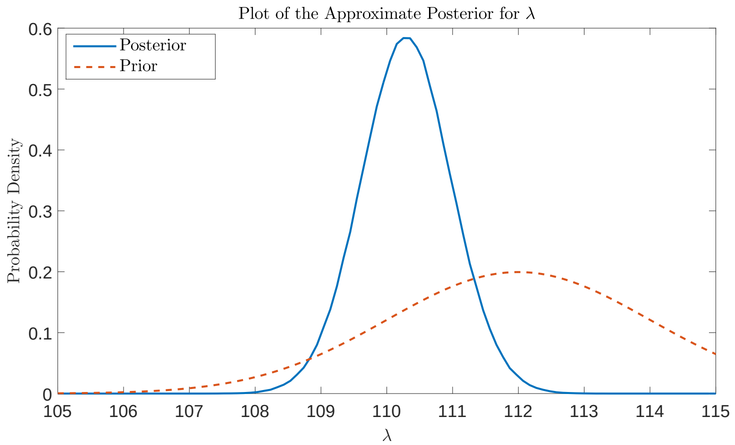
with boundary conditions

$$\frac{\partial u}{\partial x}(0,t) = g_1(t), \quad \frac{\partial u}{\partial x}(H,t) = g_2(t), \quad \forall\, t \in [0,T].$$

where λ is the unknown thermal conductivity.

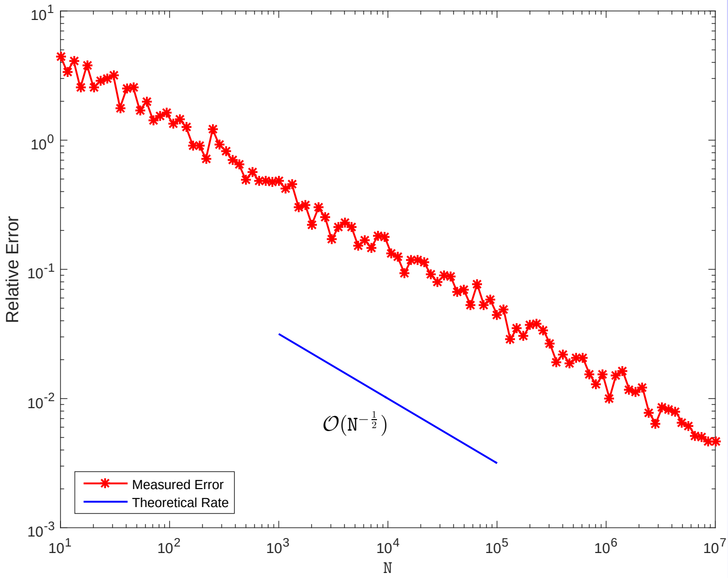
We now want to find a posterior distribution for λ, given observations of $u(H,t)$ at discrete time points.

PDE Example



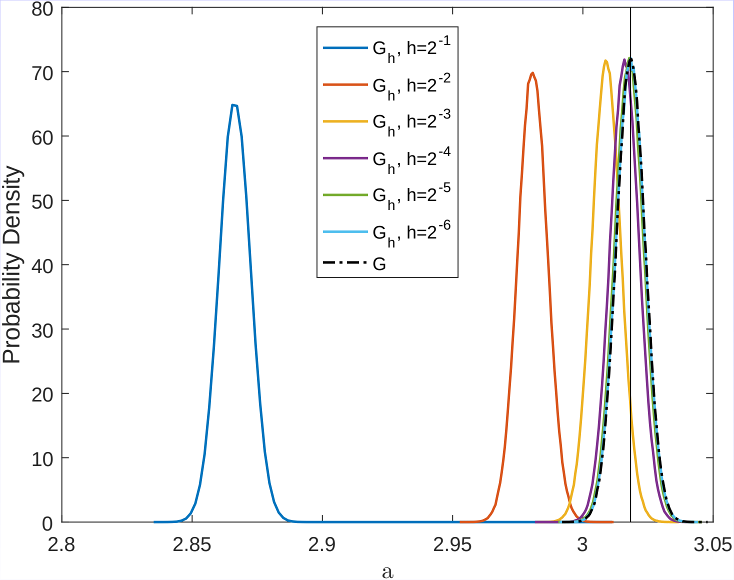
Plot of the Approximate Posterior for λ
Posterior
Prior
Probability Density
0
0.1
0.2
0.3
0.4
0.5
0.6
105
106
107
108
109
110
111
112
113
114
115
λ

PDE Example



Relative Error
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
10^{6}
10^{7}
N
$\mathcal{O}(N^{-\frac{1}{2}})$
Measured Error
Theoretical Rate

PDE Example



G_h, $h=2^{-1}$
G_h, $h=2^{-2}$
G_h, $h=2^{-3}$
G_h, $h=2^{-4}$
G_h, $h=2^{-5}$
G_h, $h=2^{-6}$
G
Probability Density
0
10
20
30
40
50
60
70
80
2.8
2.85
2.9
2.95
3
3.05
a