BUC5: MCMC for Inverse Problems in PDEs

Colin Fox
http://elec.otago.ac.nz/w/index.php/Colin_Fox

September 21, 2016

Preface

We scientists claim to be able to make predictions about the behaviour of the physical world – more than that, we claim to be able to make *quantitatively accurate* predictions, and make statements that are 'true', at least in the sense that we can quantify their accuracy. Can we actually do that? Can we do that in large-scale and complex settings where the answer really matters?

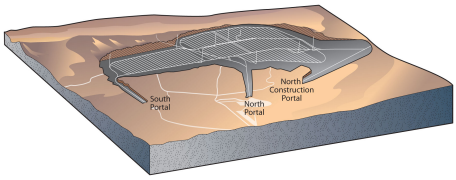
This short course provides some computational tools for one of the best current technologies we have for trying to fulfill that goal – that is MCMC for performing *sample-based inference*. The acronym 'MCMC' stands for *Markov chain Monte Carlo* which is a suite of methods and algorithms for calculating statistics over complex probability distributions, i.e., performing inference. One of the big goals these days is to avoid both the MCs, that is, not build a Markov chain, and do better than Monte Carlo integration. However we will still use the term MCMC (spoken em-cee-em-cee) because it seems to have stuck.

Many models for physics systems are written in terms of PDEs (partial differential equations). That's often because the governing equations are basically Newton's second law, which is a differential equation ($f = ma$, or $f = m\ddot{x}$) and coupled with a constitutive law (e.g. $f = -kx$) we find that the system dynamics satisfy a differential equation ($m\ddot{x} + kx = 0$ giving SHM). Another common class of PDEs is the *conservation equations* which take the form $\nabla \cdot J(x) = 0$ ($\nabla\cdot$ is spoken 'div', being short for 'divergence') when $J(x)$ is the spatially varying vector flow of some quantity that is conserved. More generally $\nabla \cdot J(x)$ is the effective source density of the quantity, and we have a *continuity equation* of the form $\nabla \cdot J(x) = \dot{\rho}$ where ρ is the density of stuff that flows according the J. An example of an equilibrium law is Kirchoff's law, when J is electrical current (electrons are conserved). Then coupled with a constitutive law, e.g. Ohm's law that $J = \sigma(x)\nabla u(x)$ (J equals sigma grad u) we end up with a typical equilibrium equation $\nabla \cdot \sigma(x)\nabla u(x) = 0$. Constitutive laws can be nonlinear, with coefficients that depend on the solution or the history, that lead to (way) more complicated PDEs.

Some big examples

Yucca mountain nuclear waste repository

Here's a large-scale and complex problem that really matters. (Source: Wikipedia)



The Yucca Mountain Nuclear Waste Repository was to be a deep geological repository storage facility for spent nuclear fuel and other high level radioactive waste.

The (USA) DOE began studying Yucca Mountain in 1978 to determine whether it would be suitable for long-term geologic repository for over 70,000 metric tons of spent nuclear fuel and high-level radioactive waste. Long term means 10,000 years to 1,000,000 years. In 2005, the United States Environmental Protection Agency proposed a limit of 350 millirem per year for that period. In 2007, the DOE issued a draft of the Supplemental Environmental Impact Statement in which it shows that for the first 10,000 years mean public dose would be 0.24 mrem/year and that thereafter to 1,000,000 years the median public dose would be 0.98 mrem/year, both of which are substantially below the proposed EPA limit.

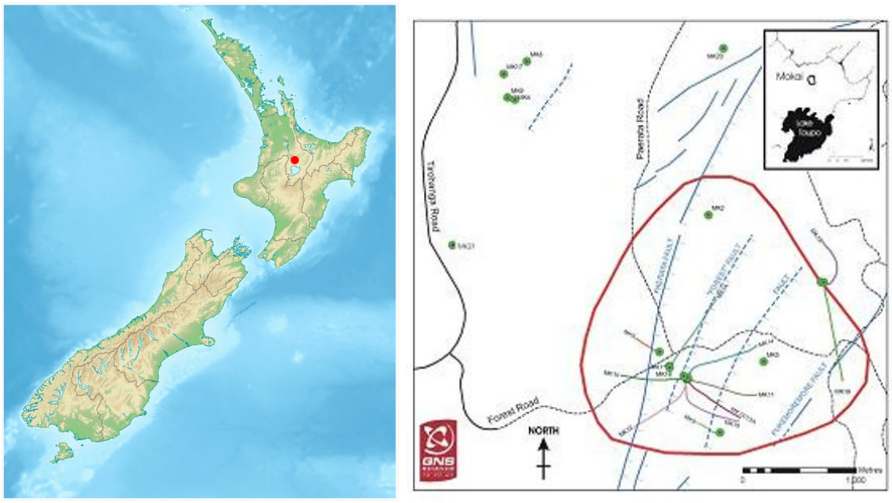
The formation that makes up Yucca Mountain was created by several large eruptions from a caldera volcano and is composed of alternating layers of ignimbrite (welded tuff), non-welded tuff, and semi-welded tuff. The volcanic tuff at Yucca Mountain is appreciably fractured and movement of water through an aquifer below the waste repository is primarily through fractures. Some site opponents assert that, after the predicted containment failure of the waste containers, these cracks may provide a route for movement of radioactive waste that dissolves in the water flowing downward from the desert surface. Officials state that the waste containers will be stored in such a way as to minimize or even nearly eliminate this possibility.

The area around Yucca Mountain received much more rain in the geologic past and the water table was consequently much higher than it is today, though well below the level of the repository.

Given all this information and uncertainties, how confident (and correct) can we be when making predictions about the bahaviour of such a complex system?

Mokai geothermal field modelling for electricity generation

On a much smaller scale, but nevertheless still large-scale by current standards in UQ, is the modelling and prediction of the behaviour of the geothermal field at Mokai (NZ), that I helped with. The resulting MCMC analysis of this problem is a contender for the largest correct and complete MCMC interpreting measured data for an inverse problem in a PDE. We will see the MCMC algorithm developed for this problem in Lecture 2.



The Mokai geothermal field is located 20 km north of Taupo. The Mokai power station was commissioned in 2000 with an initial installed capacity of 55 MW. A 39 MW expansion of similar design was commissioned in 2005. A further 17 MW binary plant extension was installed at the station in 2007 to take account of the changing steam/water ratios caused by exploitation. Annual generation is about 930 GWh. In 2011 there were 22 wells in total, 11 for production, six for injection and four others.

The continuity equation giving a multiphase non-isothermal flow model for this field is (in brief): Mass ($\alpha = m$) and energy ($\alpha = e$) balance equations

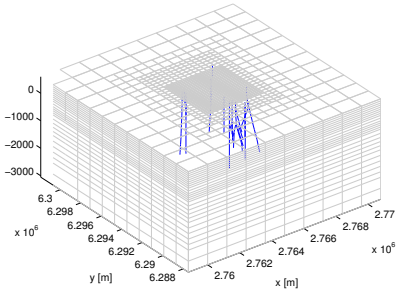
$$\frac{d}{dt}\int_\Omega M_\alpha dV = \int_{\partial\Omega} Q_\alpha \cdot \hat{n} d\Gamma + \int_\Omega q_\alpha dV$$

Flux Q_α calculated from

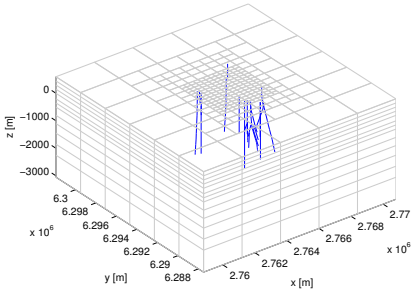
$$\begin{aligned}
M_m &= \phi\left(\rho_l S_l + \rho_v S_v\right) \\
M_e &= (1-\phi)\rho_r c_r T + \phi\left(\rho_l u_l S_l + pho_v u_v S_v\right) \\
Q_m &= \frac{kk_{rl}}{v_l}\left[\rho_l \underline{g} - \nabla p\right] + \frac{kk_{rv}}{v_v}\left[\rho_v \underline{g} - \nabla p\right] \\
Q_e &= -\frac{h_l kk_{rl}}{v_l}\left[\rho_l \underline{g} - \nabla p\right] - \frac{h_v kk_{rv}}{v_v}\left[\rho_v \underline{g} - \nabla p\right] - K\nabla T
\end{aligned}$$

for porosity ϕ, permeability k, and relative permeabilities k_{rl} k_{rv} given by van Genuchten-Mualem model.

Numerical solution by finite volume method (TOUGH2), using the fine and coarse discretizations (10,046 parameters in representation):

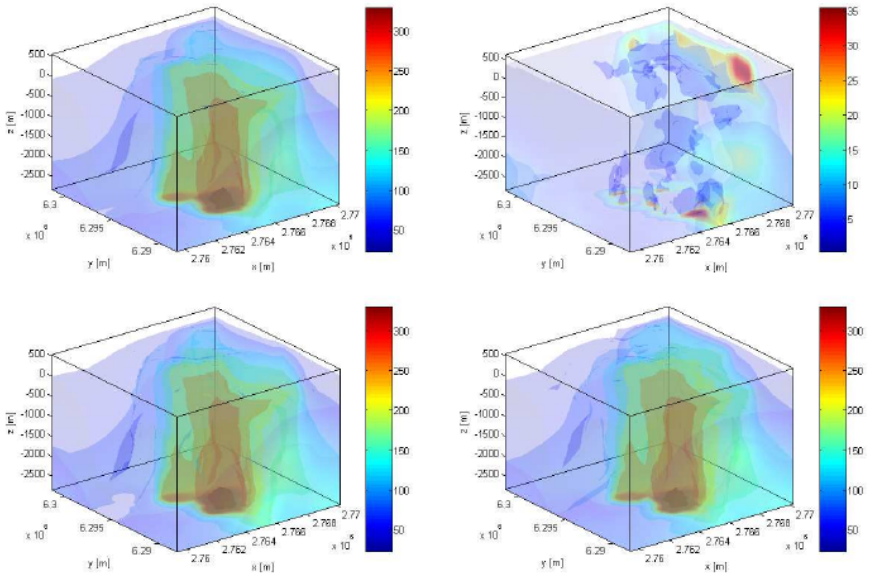


Fine mesh: 26,005 blocks
0.5 to 1 hours



Coarse mesh: 3,335 blocks,
1 to 1.5 minutes

Here are some sample-based results (left to right, top to bottom): the mean temperature, standard deviation of the sample mean, and two samples from the chain over (predicted) temperature.



For further details, see: Cui T Fox C and O'Sullivan M J 2011 Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. Water Resources Research, 47, W10521, doi:10.1029/2010WR010352.

The absurd range of skills required

The Berkeley stattician Le Cam has been quoted as saying of inverse problems

> the forward problem is probability and the inverse problem is statistics

which is true (and useful to observe), but it does not get anywhere near expressing the range of skills that are required in resolving real problems in uncertainty quantification. Indeed, by implying that inverse problems are nothing more than the fields that he is expert in, Le Cam displayed the myopia that is unfortunately quite common in inverse problems; often experts in one field produce 'solutions' that are based on highly technical and correct work in their own field, but that fail to account for critical issues that arise from other aspects of the problem. To genuinely make useful and correct statements in large-scale problems, one needs to be aware of all aspects of the problem, from problem statement through modelling and mathematical formulation, right through to output analysis of the computational statistics and statement of conclusions.

So what skills/fields are involved?

- Correct physical modeling to state the PDEs: In all real inverse problems that I know of, modelling error is greater than error (noise) in observations. So it is critical to understand the modelling process, and what assumptions and approximations are involved. This work is best done by an expert in the field, so one needs to be able to, at least, talk to domain-specific experts.

- Engineering of the measurement system: Often the measurement system contributes to the observation process. Good engineering tries to minimize the influence of the measurement system, but usually some biases or unintended physical processes remain. Modelling the observation process requires discovering all the processes that contribute to observations (up to some point), which often means having some knowledge of the measurement system.

- Probabilistic modelling of measurement uncertainties: The Bayesian formulation requires distributions over all uncertain quantities, from unknown parameters to uncertain observations. Getting those correct, or validated, is necessary if uncertainties are to be *quantified*.

- Adequate numerical solvers: Computational science for PDEs is a well developed and sophisticated field. Indeed, some people spend their whole careers studying nothing else. How best to merge the computational statistics and the numerical solvers remains a topic of research.

- Correct a priori representations: The problem statement puts no constraints on how we represent the unknowns. How do you represent something that you don't know, when you are free to use anything? Often the physical unknowns are spatially varying quantities, so the fields of spatial statistics, pattern theory, and stochastic geometry are very pertinent. We would love to be able to include

expert knowledge into representations (e.g., compute over the space of geological models that a geologist believes), but this is currently beyond our computational capability. Unfortunately, one often sees low-level representations (e.g., linear spaces) that are convenient for the mathematician but not physically plausible.

- Efficient MCMC. Sample-based inference in inverse problems with PDEs can be a compute intensive business. The development of better algorithms and computation is often needed to resolve specific applications.

One does not need to be expert in all these areas to make a contribution, but one certainly needs to be aware and conversant in all these aspects of the problem.

1 Lecture 1

1.1 Problems and goals

The aim of collecting data is to gain useful information about a physical system or phenomenon of interest. Inverse problems occur in those frequent situations where the quantities we can measure are not the ones that we are primarily interested in. Typically we can measure data that depends in some way on the quantities we want, so the data at least contains some information about those quantities. Starting with the data that we have measured, the problem of trying to reconstruct the quantities that we really want is called an *inverse problem.* We often say an inverse problem is where we measure an *effect* and want to determine the *cause*.

Most science and statistics is data-driven in this way, though not always called an 'inverse problem'. The quintessential setting is where the measurement process is a complex physical relationship, and inversion presents analytic difficulties.

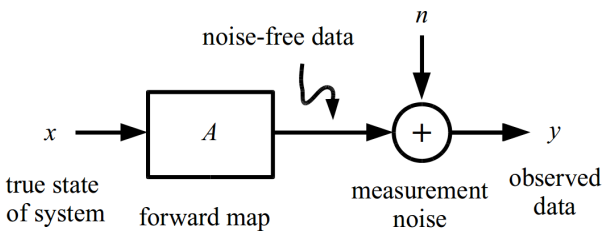
In physics, that cause-effect relationship is well defined, repeatable, and can be quantitatively modelled. In this course we focus on forward models that require the solution of a PDE (partial differential equation) – strictly, we should say a BVP (boundary value problem).

In a mathematical setting, we represent the measurement process by a family of models parameterized by x, where all necessary physical parameters are contained in x, including nuisance parameters. In the language of inverse problems, simulation of the model for given x defines the *forward map* $A : x \mapsto d$ giving data d in the absence of errors. Determining and simulating the map $A : x \mapsto d$ is the *forward problem*, whereas inferring x from d is the *inverse problem.*

We take a Bayesian route to the inverse problem, and recognize that all components of this process contain errors, or *uncertainties.* All available data contains measurement errors, so the estimated unknowns are to some degree uncertain. A natural question then arises: if measurement noise corrupting the data follows some statistics, what is the *distribution* over the possible solutions after the estimation procedure? Bayesian thinking explicitly allows for the unknown vector x to be interpreted as a *random variable* with a distribution of its own.

1.2 Forward maps and hierarchical models

In physics and signal processing, the observation process is depicted as



In statistics, a hierarchical model for the inverse problem is, in the general form,

$$y|x,\theta \sim \pi(y|x,\theta) \tag{1.1a}$$
$$x|\theta \sim \pi(x|\theta) \tag{1.1b}$$
$$\theta \sim \pi(\theta). \tag{1.1c}$$

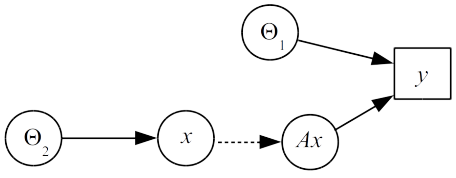
This hierarchical stochastic model occurs commonly in statistics, in which y is *observed data*, x is a *latent field*, and θ is a vector of *hyperparameters* that model uncertainties in the measurement noise and in modeling of the latent field process. In the language of Bayesian analysis, (1.1a) defines the *likelihood function* for unknown x and θ once data y is observed, (1.1b) is the *prior distribution* over latent field x with hyperparameters θ, and (1.1c) sets the *hyperprior distribution* over those hyperparameters.

A very common version of this model assumes additive noise, as above, that is zero mean Gaussian, and uses a GMRF (Gaussian Markov random field) as a model for x, giving the model

$$y|x,\theta \sim \mathrm{N}(Ax, \Sigma(\theta)) \tag{1.2a}$$
$$x|\theta \sim \mathrm{N}(\mu, Q^{-1}(\theta)) \tag{1.2b}$$
$$\theta \sim \pi(\theta). \tag{1.2c}$$

Here, x is modelled as a multivariate normal with mean μ, and θ is a vector of *hyperparameters* that model uncertainties in the measurement noise covariance Σ and in modeling of the precision (inverse of covariance) matrix Q of the latent field.

A nice representation of conditional dependencies is given by the DAG (directed acyclic graph)



1.3 Posterior exploration

The focus of inference is the *posterior distribution* over unknowns x and θ conditioned on measured y, given by Bayes' rule as

$$\pi\left(x,\theta|y\right)=\frac{\pi\left(y|x,\theta\right)\pi\left(x,\theta\right)}{\pi\left(y\right)}. \tag{1.3}$$

Note that we are performing the standard abuse of notation by using the symbol π to denote any probability density function, and associated distributions, with the particular function determined implicitly by the arguments. Solutions and uncertainties may be computed as the posterior expectation of some function g of x,

$$\mathrm{E}_{x,\theta|y}\left[g\left(x\right)\right]=\int g\left(x\right)\pi\left(x,\theta|y\right)\,\mathrm{d}x\,\mathrm{d}\theta$$

which implicitly averages over the nuisance parameter θ. Sample-based methods use a Monte Carlo estimate of the integral. When $(x,\theta)^{(1)},\ldots,(x,\theta)^{(N)}\sim\pi\left(x,\theta|y\right)$ are iterates of an ergodic Markov chain,

$$\mathrm{E}_{x,\theta|y}\left[h\left(x\right)\right]\approx\overline{g}_N\equiv\frac{1}{N}\sum_{i=1}^{N}g\left(x^{(i)}\right)$$

with convergence guaranteed by a central limit theorem (CLT).

When the $x^{(i)}$ are independent, an suitable CLT gives

$$\frac{\overline{g}_N-\mathrm{E}\left[g\right]}{\sqrt{\mathrm{Var}(\overline{g}_N)}}\sim\mathrm{N}(0,1)$$

where

$$\mathrm{Var}(\overline{g}_N)=\frac{\mathrm{Var}(g)}{N}.$$

When the $x^{(i)}$ come from a correlated Markov chain, instead we find (for large N)

$$\mathrm{Var}(\overline{g}_N)=\frac{\mathrm{Var}(g)}{N}\left(1+2\sum_{i=1}^{\infty}\rho_{gg}(i)\right)$$

where $\rho_{gg}(i)$ is the autocorrelation coefficient for the chain in g at lag i. Thus, the rate of variance reduction, compared to independent samples, is reduced by the factor

$$\tau=\left(1+2\sum_{i=1}^{\infty}\rho_{gg}(i)\right)$$

that is called the integrated autocorrelation time (for the statistic g). We can think of τ being the length of the correlated chain that produces the same variance reduction as one independent sample.

A CLT for correlated Markov chains is given in: C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. Comm. Math. Phys., 104(1):1–19, 1986.

1.3.1 Metropolis-Hastings MCMC

The Metropolis-Hastings (MH) algorithm is the basis of nearly all sampling algorithms that we currently use. This algorithm was originally developed for applications in statistical physics, and was later generalized to allow general proposal distributions (Hastings 1970), and then allowing transitions in state space with differing dimension (Green 1995). Even though we do not always use variable-dimension models, we prefer this Metropolis-Hastings-Green (MHG) 'reversible jump' formulation of MH as it greatly simplifies calculation of acceptance probabilities for the subspace moves that are frequently employed in inverse problems. One step of MHG dynamics can be written as:

Algorithm 1 (MHG) *Let the chain be in state $x_n = x$, then x_{n+1} is determined in the following way:*

1. *Propose a new candidate state x' from x depending on random numbers γ with density $q(\gamma)$.*
2. *With probability*

$$\alpha(x, x') = \min\left(1, \frac{\pi(x'|d)q(\gamma')}{\pi(x|d)q(\gamma)}\left|\frac{\partial(x', \gamma')}{\partial(x, \gamma)}\right|\right) \tag{1.4}$$

   accept the proposed state by setting $x_{n+1} = x'$. Otherwise reject by setting $x_{n+1} = x$.

The last factor in eqn (1.4) denotes the magnitude of the Jacobian determinant of the transformation from (x, γ) to (x', γ'), as implemented in computer code for the proposal. A few details remain to be specified such as the choice of starting state, and the details of the proposal step.

The only choice one has within the MHG algorithm, is *how* to propose a new state x' when at state x. The popular choice of Gibbs sampling is the special case where x' is drawn from a (block) conditional distribution, giving $\alpha(x, x') = 1$. The choice of the proposal density is largely arbitrary, with convergence guaranteed when the resulting chain is irreducible and aperiodic. However, the choice of proposal distribution critically affects efficiency of the resulting sampler. The most common MH variants employ *random walk* proposals that set $x' = x + \gamma$ where γ is a random variable with density $q(\cdot)$, usually centered about zero. In high-dimensional problems, global proposals that attempt to change all components of the state usually have vanishingly small acceptance probability, so are not used. Since ill-posedness results in extremely high correlations, single-component proposals result in slow mixing. Hence, a multi-component update is usually required, that is problem specific.

1.3.2 Some inverse problems in PDEs

Many potential imaging modalities such as optical diffusion 'tomography' involve strong scattering of the waves. In these cases the paths of propagation depend im-

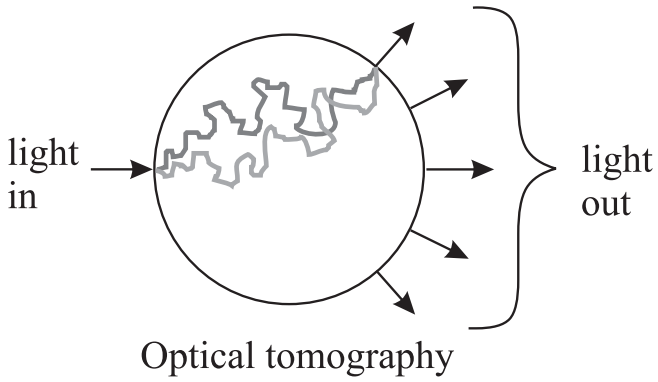


Figure 1.1: A schematic of the measurement process in optical diffusion tomography.

plicitly on the object being imaged. Then the wave field that propagates and scatters must be modelled by requiring that it satisfies a partial differential equation (PDE) in which the unknown object properties appear as spatially-varying coefficients. Three examples of governing PDEs, corresponding to three different types of energy being propagated, are given in the following table.

object property	governing PDE	PDE classification
electrical conductivity	$\nabla \cdot (\sigma \nabla \phi) = s$	elliptic
acoustic impedance	$\nabla \cdot (\sigma \nabla p) = \frac{\sigma}{c^2} \ddot{p}$	hyperbolic
thermal conductivity	$\nabla \cdot (\sigma \nabla u) = \dot{u}$	parabolic

In each case the object property being sought is denoted by σ and appears as the spatially-varying coefficient in the PDE governing the propagation of energy, or 'waves'. The measurements made are of the boundary values of electrical potential ϕ, sound pressure p, and temperature u, respectively in the three cases. If the scattered field can be well approximated as a small change about a known field then the (linear) Born approximation may be used to simulate the measurement process. Otherwise, the measurement process must be simulated by solving the PDE subject to boundary conditions that correspond to the wave irradiation. In that case the mapping from object property, σ, to measurements is nonlinear. This latter property presents a basic difficulty in these inverse problems and, as we will see later, is one reason why straightforward application of algorithms that have been successful for linear inverse problems, such regularization, do not provide quantitatively accurate solutions in this class of inverse problems.

An inverse problem in the heat equation

The first specific example we look at is in the heat equation, in 1 space variable. We will estimate the (scalar) diffusivity D from solution of a partial differential equation,

and consider related inverse problems.

Suppose we have a rod of length L, and that $u \equiv u(x,t)$ is the temperature of the rod at position $x \in [0,L]$ and time $t \geq 0$. (We could also be modelling the spread of a bug that diffuses, or any other conserved diffusing quantity.) The temperature is held at temperature 0 at the ends of the rod so that $u(0,t) = u(L,t) = 0$.

Local conservation of heat gives the diffusion equation

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} D \frac{\partial u}{\partial x}. \tag{1.5}$$

At time $t = 0$, the temperature is known to be

$$u(x,0) = \begin{cases} 1 & \text{for } 0.75L \leq x \leq 0.8L \\ 0 & \text{otherwise} \end{cases} \tag{1.6}$$

At time $t = T$, the temperature is measured at points $x_1, x_2, \ldots, x_K$. Since the measurements are inexact, the data vector is $\mathbf{y} = (y_i)$ where

$$y_i = u(x_i, T) + \varepsilon_i \tag{1.7}$$

and ε_i are normally distributed with standard deviation s. From these measurements, we wish to sample from the posterior distribution of D, assuming that the prior distribution of D is uniform on $D \geq 0$.

By Bayes' theorem,

$$\pi(D|\mathbf{y}) \propto \pi(\mathbf{y}|D)\,\pi(D) \tag{1.8}$$

The likelihood function is determined by the forward map and the noise process. Given the diffusivity is D and given the initial conditions, we may solve the partial differential equation and obtain the expected number density at the locations x_i after a time T, namely $u(x_i, T; D)$. The probability that we measure the data vector $\mathbf{y}$ is

$$\pi(\mathbf{y}|D) = \prod_{i=1}^{K} p(\varepsilon_i = y_i - u(x_i, T; D)) \tag{1.9}$$

$$\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{K}\left(\frac{y_i - u(x_i,T;D)}{s}\right)^2\right] \tag{1.10}$$

where we have absorbed into the proportionality quantities which do not depend on D.

For the prior, we may choose

$$\pi(D) \propto \begin{cases} 1 & \text{if } D \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.11}$$

As before, this is improper (i.e., not normalizable), but could be made proper without affecting the analysis, for all practical purposes by imposing a conservative upper bound $D_{\max}$ on D.

The posterior probability density is thus given by

$$\pi\left(D|\mathbf{y}\right) \propto \begin{cases} \exp\left[-\frac{1}{2}\sum_{i=1}^{K}\left(\frac{y_i - u(x_i,T;D)}{s}\right)^2\right] & \text{if } D \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1.12}$$

We can estimate D given y using sample-based inference.

We now specify a random walk Metropolis Hatings MCMC that targets $\pi\left(D|\mathbf{y}\right)$. Let $X_n = D$, X_{n+1} is given in the following way:

1. Let w be a positive constant. Draw r from a uniform distribution on $[0, 1]$ and set $D' = D + w\left(2r - 1\right)$
2. With probability

$$\alpha\left(D'|D\right) = \min\left\{1, \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{K}\left(\frac{y_i - u\left(x_i,T;D'\right)}{s}\right)^2\right]}{\exp\left[-\frac{1}{2}\sum_{i=1}^{K}\left(\frac{y_i - u(x_i,T;D)}{s}\right)^2\right]}\right\}$$

set $X_{n+1} = D'$, otherwise set $X_{n+1} = D$.

Notice that at each step of the MCMC algorithm, we must compute $u\left(x_i, T; D\right)$, (i.e., solve the boundary value problem for a trial value of D) to work out what the solution would have looked like at T if the true diffusivity were given.

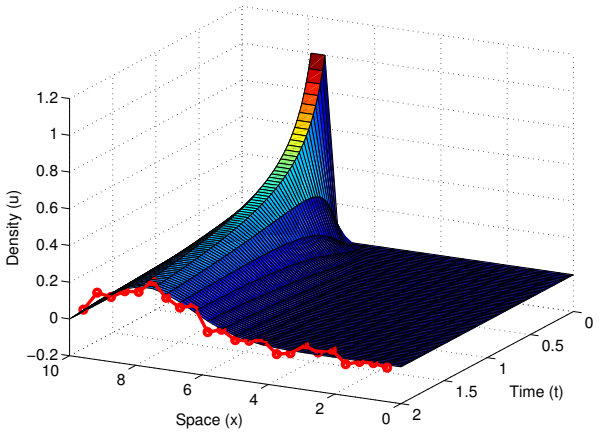


Figure 1.2: Surface is $u\left(x, t\right)$ with the true value of D and the points show the data measured at T.

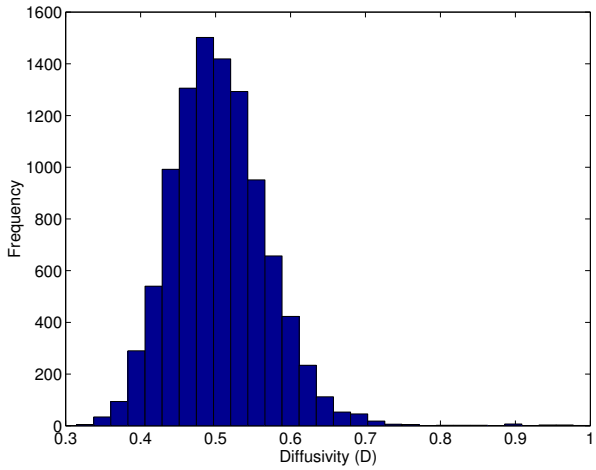


Figure 1.3: Output of Markov chain producing samples from posterior distribution of the diffusivity. True D value was 0.5.

1.4 Computing 1

All code listed here, is available at URL
`https://coursesupport.physics.otago.ac.nz/wiki/pmwiki.php/ELEC446/BUC5`, or search for "Colin Fox Otago" and follow the link from my home page.

1.4.1 RWM sampling from $\mathrm{N}(0,1)$

Matlab code for a random-walk Metropolis Hastings targeting $\mathrm{N}(\mu,\sigma^2)$:

```
 function X=mcgaus(mu,sig,x0,N,w)
%function X=mcgaus(mu,sig,x0,N,w)
%
% Return N samples of N(mu,sig^2) using a
% RWM with window w, starting at x0

X = zeros(1,N);
X(1)=x0;

for k=1:(N-1)
   %xp=X(k)+w*(2*rand-1); % uniform window
   xp=X(k)+w*randn; % normal window, variance w^2

   alpha=min(1,exp( (-(xp-mu)^2+(X(k)-mu)^2)/(2*sig^2) ));

   if rand<alpha
      X(k+1)=xp;     % accept
   else
      X(k+1)=X(k);   % reject
   end
end
```

Exercises:

1. Use the code in `mcgauss.m` to generate a chain of length 10^4 that targets $\mathrm{N}(0,1)$. Use windows $w = 0.3, 3, 30$.

2. Plot histograms of each chain.

3. Plot the chains, and see which window gives the most efficient mixing.

4. Adapt `mcgauss.m` to sample from the (unnormalized) pdf $x\exp(-x)$ for $x \geq 0$.

The following code evaluates the integrated autocorrelation time (need to multiply by 2 to transform from physics definition to statistics definition), using the `UWerr.m` function available at `http://www.physik.hu-berlin.de/com/ALPHAsoft/`.

```
% script file to plot IACT as a function of window size for
% MH MCMC (RWM) sampling from a standard normal

ws = linspace(1,8,100); %window sizes

mu = 0;
sig = 1;
x0 = 0;

n=1e5;

for cnt = 1:length(ws)
  w = ws(cnt);
  X = mcgaus(mu,sig,x0,n,w);
  [value,dvalue,ddvalue,tauint,dtauint,Qval] = UWerr(X',1.5,length(X),0);
  taus(cnt) = tauint*2; %to get stats definition of IACT
end

plot(ws,taus)
```

1.4.2 Sampling in inverse heat diffusivity problem