Bayesian Inference for Dynamical Systems

Hugh Coleman (Open University)

Introduction

I descibe here a computational appraoch to compare models of dynamical systems with observations. The methods described below are demonstrated by Infr (https://github.com/Homo-carbonis/infr), a Common Lisp library for performing Bayesian inference over over Markov chains.

We're going to take a model of a system in the form of an ordinary differential equation, approximate it as a Markov chain using the finite difference method and then use Bayes' Theorem to compare it to a sequence of data points. This will allow us to infer the probability distribution of the model's parameters and compare two competing models to see which is best supported by the data.

Bayes' Theorem

If we want to inferer the probability distribution of an unknown parameter β from an observed value, y. We need two things: A model telling us the likelihood of y given β $P(y|\beta)$ and a prior distribution $P(\beta)$, representing any prior knowledge we have of β. Bayes' theorem then tells us

$$P(\beta \mid y) = \frac{P(y\mid \beta)P(\beta)}{P(y)}$$

$P(y)$ is known as the marginal distribution of y given by

$$\int_{-\infty}^{+\infty} P(y\mid \beta)P(\beta)\ d\beta$$

and it normalises the joint posterior distribution $P(\beta, y)$ for our particular model. If we are are only interested in comparing relative probabilities of β, we can discard this and just use the relative posterior

$$P(\beta \mid y) \propto P(y\mid \beta)P(\beta)$$

A simple method to estimate the marginal probability so we can compare different models is presented under **Marginal**.

If we want find the probability of β given a sequence of observations of a process, we need a model which gives us the probability of the next observation given the ones which preceded it.

From the first two observations we get:

$$P(\beta \mid y_1, y_0) = \frac{P(y_1 \mid \beta, y_0)P(\beta)}{P(y_1 \mid y_0)}$$

This then becomes the prior proability for the next observation, so we have

$$P(\beta \mid y_2, y_1, y_0) = \frac{P(y_1 \mid \beta, y_0)P(\beta \mid y_1, y_0)}{P(y_1 \mid y_0)}$$

We can continue chaining applications of Bayes' Theorem like this to give a recursive definition

$$P(\beta \mid y_n, y_{n-1}, ...y_0) = \frac{P(y_n \mid \beta, y_{n-1}, ...y_0)P(\beta \mid y_{n-1}, y_{n-2}, ...y_0)}{P(y_n \mid y_{n-1}, y_{n-2}...y_0)}$$

This can be applied to any discrete process but the integration involved in computing the marginal probabilities make it impractical. To produce an equation we can solve, we constrain ourselves to a particular kind of process.

Markov Chains

A Markov Chain is a stochastic process with the property:

$$P(x_n \mid x_{n-1}, x_{n-2}, ...x_0) = P(x_n \mid x_0)$$

That is, it is a chain of random variables, each of which depends only on its predecessor.

For our purposes Markov Chains have the very useful property

$$P(x_n, x_{n-1}, ...x_0) = P(x_n|x_{n-1})P(x_{n-1} \mid x_{n-2})...P(x_1|x_0)P(x_0)$$

Proof

$$P(x_1, x_0) = P(x_1|x_0) \ P(x_0)$$

$$P(x_n, x_{n-1}, ...x_0) = P(x_n \mid x_{n-1}, x_{n-2}, ...x_0) \ P(x_{n-1}, x_{n-2}, ...x_0)$$

By induction this gives

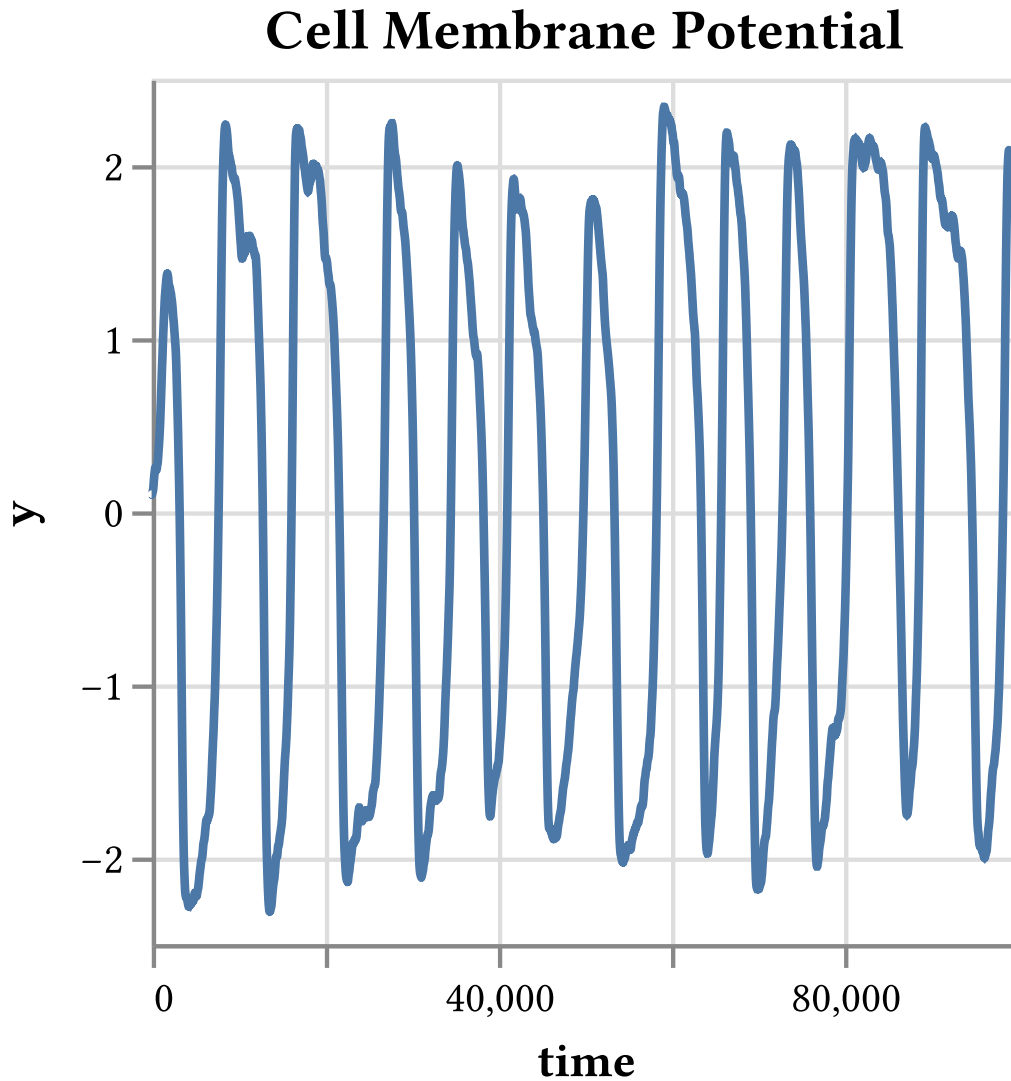$$P(x_n, x_{n-1}, ...x_0) = P(x_n|x_{n-1})P(x_{n-1} \mid x_{n-2})...P(x_1|x_0)P(x_0)$$

For a Markov chain our posterior distribution becomes

$$P(y_n, y_{n-1}, ...y_0) = \frac{P(y_n \mid \beta, y_{n-1})P(y_n - 1 \mid \beta, y_{n-2})...P(y_1 \mid \beta, y_0)P(\beta)}{P(y_n, y_{n-1}, ...y_0)}$$

$$\text{or} \quad P(\mathbf{y}) = \frac{P(\beta) \prod_{i=0}^{n} P(y_i|\beta, y_{i-1})}{P(\mathbf{y})}$$

The Model

We begin with some experimental data, perhaps we have measured the membrane potential of a cell a we have a plot like this

Cell Membrane Potential

We have a model which we think may be a good fit the observed process, in the form of a differential equation:

$$\frac{d^2y}{dt^2} - \beta(1 - x^2)\frac{dy}{dx} + x = 0$$

The Finite Difference Method allows us to approximate a continuous time system as a discrete time system. If the function of the original system is sufficiently smooth, we can choose time increments, h, small enough to apporximate it with arbitrary accuracy.

We substitute:

$$\frac{dx}{dt} \rightarrow \frac{x_i + 1 - x_i - 1}{2}h$$

$$\frac{d^2x}{dt^2} \rightarrow \frac{x_i - 1 - 2y_i + y_i + 1}{h^2}$$

This gives us

$$x_{i+1} = f(x_i, x_{i-1}) = \frac{-2x_i + x_{i-1}}{h^2} - \beta(1 - x_i^2)\left(\frac{-x_i - 1}{2h}\right) + x_i$$

Adding a certain quantity of Gaussian noise to this difference equation we have

$$x_{i+1} = f(x_i, x_{i-1}) + r, \quad \text{where} \quad r \sim \mathcal{N}(0, \sigma^2)$$

NB. This introduces a second parameter σ^2. This method can be applied to any number of paramters but in the interest of simplicity here we treat it as a constant.

We now have something approaching a Markov chain. It depends on x_i and x_{i-1} so it is not a simple Markov Chain, but for a sufficiently small h this should not affect the result.

Inference

Say from previous experiments we have an idea β is around 3. To reflect this For our prior we choose a normal distributon $\mathcal{N}(3, 1)$.

Now we have everything we need to find the posterior distribution of β.

$$P(\beta \mid \mathbf{y}) \propto \prod_{i=0}^{n} P(y_i \mid \beta, y_{i-1}) P(\beta)$$

P(beta)

beta
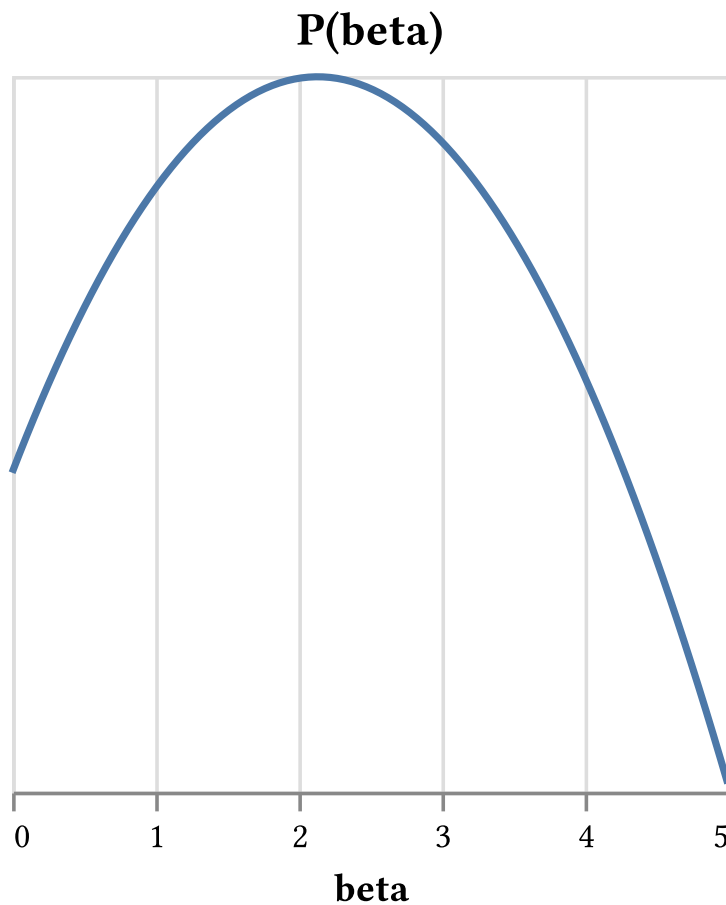
We can see β is somewhere around 2 but there is still a great deal of uncertainty. We will need more data to know the value with greater accuracy.

Marginal

The relative posterior probability allows to compare values of β for the same model but if we want to compare two different models, we also need to know the marginal probability for each:

$$P(\mathbf{y}|\mathcal{M}) = \int_{-\infty}^{+\infty} P(\mathbf{y}|\ \beta, \mathcal{M})P(\beta)\ d\beta$$

This can't be solved analytically except in the simplest cases. Numerical integration by quadrature is also impractical since the integral is form $-\infty$ to $+\infty$ and we may be considering several parameters leading to integration of many dimensions.

The method described below uses the prior distribution to optimize numerical integration It is a simpler varient of the particle filter method which uses only the prior distribution to generate samples.

To integrate a posterior function of beta, which is approximated by another function of β, the prior, we draw number of random samples from the prior distribution beta, $\beta_0, \beta_1, ...\beta_m$. We then divide the area under the function we are integrating into strips of with h_i, such that there is approximately one sample in each strip. If our prior is a reasonably good fit for our posterior the strips should be narrower near the mode, leading to much greater accuracy compared to quadrature, where the strips are all the same. We also get over the problem with not knowing the range to integrate over.
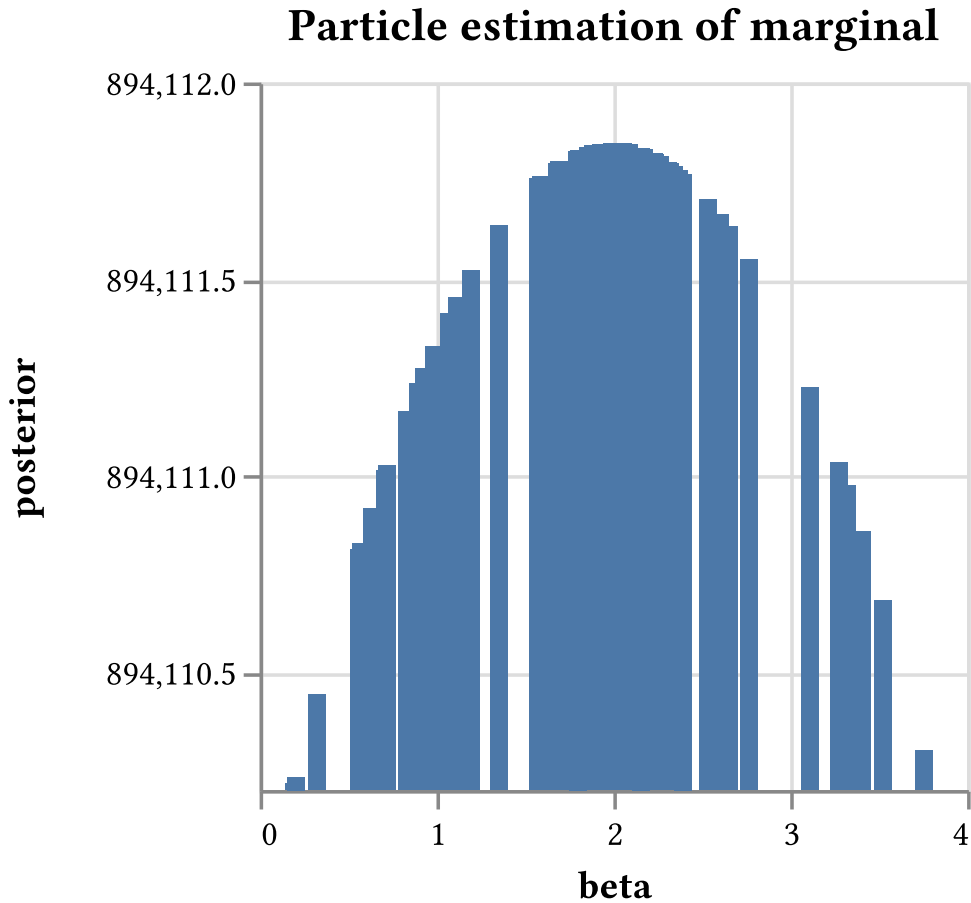

Figure 1: Plot showing samples drawn from $P(beta)$

h_i is proportional to $1/P(B_i)$ so we have

$$\int_{-\infty}^{+\infty} P(\mathbf{y}|\ \beta, \mathcal{M}) P(\beta)\ d\beta \approx \sum_i P(\mathbf{y}\mid\beta_i) P(\beta_i)\ \frac{k}{P(\beta_i)}$$

$$\approx K \sum_i P(\mathbf{y}\mid\beta_i)$$

We can find the constant K from the fact that

$$\sum_i h(\beta_i) \approx B_{\max} - B_{\min}$$

$$\text{so}\quad K \sum_i 1/P(\beta_i) \approx B_{\max} - B_{\min}$$

This gives us the formula

$$P(\mathbf{y}|\mathcal{M}) \approx \frac{B_{\max} - B_{\min}}{\sum_i \frac{1}{P(\beta_i)}} \sum_i P(\mathbf{y}\mid\beta_i)$$

Applying this to two different models we can find the Bayes' Factor

$$\frac{P(\mathbf{y}|\mathcal{M}_1)}{P(\mathbf{y}|\mathcal{M}_2)}$$

which allows us to make a direct comparison between the evidence for the two models.