Inference Control and Driving of Natural Systems

MSci/MSc/MRes

Nick Jones

nick.jones@imperial.ac.uk

The fields at play

We will be drawing on ideas from Bayesian Cognitive Science (psychology and neuroscience), Biological Physics and non-equilbrium phenomena, Quantitative Cell Biology and Control theory and Robotics.

Projected topics to cover

Topics in Bayesian inference and connections to implementations of samplers, chemical reactions and neural coding. Topics in optimal control including Markov decision processes, dynamic programming, HJB equations, linearly solvable stochastic optimal control and its path integral formulation. Links between control, inference and reinforcement learning. Theory for driving of (living) systems including (metabolic) control theory, polytopes and flux balance analysis, and ageing and error correction.

How the course will be structured:

For this course the structure will be distinct from standard undergraduate courses. Each day's lecture will be followed by a second 'applied' lecture in which, alongside more lecture material (structured around examples) we will work through examples together.

The examples will be in MATLAB and you will be expected to bring a laptop with MATLAB.

`/www3.imperial.ac.uk/ict/services/softwarehardware/softwarepurchase/softwareshop/priceindex/studentpricelist`

If you don't have a laptop please let me know.

At the moment the course occurs on four Fridays at 3pm. This clashes with Dr Buck's course on bio-structure. Affects anyone?

Philosophy: This is a graduate course, and a research course, during it you will be implementing contemporary research. Use each other and the range of attendees from MRes, PhD students and Post-Docs.

How the course will be structured:

We will be adopting an intensive format that reflects the way you learn. Default: 1pm more theoretical. 3pm applying the theory.
Inference: First week has two days (4 Lectures), Second week has three days (6 lectures).
Two week pause for consolidation.
Control: Five day week. 10 Lectures
Two week pause for consolidation.
Driving: Five day week. 10 Lectures
Logistics: Regular exam at the end - but 10 percent will come from turning in course work: providing code that shows participation in the practicals.
Circulate a paper recording your names if you might take this as an exam (MRes, MSci, MSc).

The motivation for the course: Inference and Driving

Living systems are out of equilibrium and, as such, must be driven. These systems must thus have a supply of energy. One can think of this as a setting with two statistical models: a model generating the world and the organism's approximation to relevant parts of its world. The organism can make a more accurate model of the world but this will have a corresponding energetic cost. If the organism can construct a sufficiently accurate representation of the world then this can be exploited to help it find energy to drive itself. There is thus a very natural relationship between inference about the world and driving of natural systems.

The motivation for the course: Inference and Control

It is often appropriate to think of the model describing the world having dynamics in its parameters: this means that the organism cannot make a static statistical model for the world but ideally controls that representation: e.g. by appropriately weighting its past information. As such we need to think of control and inference as coupled.

The motivation for the course: Inference-Control through Driving

We finally note that an effective means of controlling a complicated system with an evolving structure is to modulate its power supply. If we want to have a robust understanding of biological systems we thus need to have a unified understanding of inference, control and driving in natural systems. The scientific literature is fragmented on this topic with only moderate coupling between statistics, engineering and physics in these respective areas: unfortunately good theory for biological systems does not respect disciplinary boundaries.

The kinds of questions we'll answer:

- How can one perform inference with a noisy molecular architecture?
- How can one perform inference with ensembles of noisy neurons?
- How are E. Coli Kalman Filters?
- What are optimal ways of making a robot move?
- How do living systems get their power?

How do living systems perform inference?

- We will be interested in physical architectures which allow inference
- How being intrinsically noisy is not a problem

The topics covered in the first unit on inference.

- Introduction to Bayesian Inference
- Bayesian inference with chemistry
- Bayesian inference with neurons
- Sampling approaches to Bayesian inference.

What we'll cover in this lecture

A basic introduction that makes sure we're on the same page.

- Bayes Theorem
- Evaluating Integrals
- Bayes Factors
- Examples of the above (applied lecture)

Components for Bayes Theorem

Data, D. Hypothesized model structure, $\mathcal{H}$. Model parameters, θ.

- Model

$$p(D|\theta, \mathcal{H})$$

Components for Bayes Theorem

Data, D. Hypothesized model structure, $\mathcal{H}$. Model parameters, θ.

- Model

$$p(D|\theta, \mathcal{H})$$

- Prior

$$\pi(\theta|\mathcal{H})$$

Components for Bayes Theorem

Data, D. Hypothesized model structure, $\mathcal{H}$. Model parameters, θ.

- Model $$p(D|\theta, \mathcal{H})$$
- Prior $$\pi(\theta|\mathcal{H})$$
- Posterior $$\pi(\theta|D, \mathcal{H})$$

Components for Bayes Theorem

Data, D. Hypothesized model structure, $\mathcal{H}$. Model parameters, θ.

- Model
$$p(D|\theta, \mathcal{H})$$
- Prior
$$\pi(\theta|\mathcal{H})$$
- Posterior
$$\pi(\theta|D, \mathcal{H})$$
- Evidence or The Marginal Likelihood
$$p(D|\mathcal{H})$$

Theorem

Bayes Theorem

$$\pi(\theta|D,\mathcal{H}) = \frac{p(D|\theta,\mathcal{H})\pi(\theta|\mathcal{H})}{p(D|\mathcal{H})}$$

Theorem

Posterior over parameters = Data likelihood given parameters × Prior over parameters/Chance of data given model

Example elements of Bayes Theorem

Data, D 3 Heads, 2 tails. Hypothesized model structure, $\mathcal{H}$, it's a coin (Bernoulli process). Model parameters, θ the bias of the coin (probability of a head).

- Model

$$p(D|\theta, \mathcal{H}) = \theta^3(1-\theta)^2$$

Example elements of Bayes Theorem

Data, D 3 Heads, 2 tails. Hypothesized model structure, $\mathcal{H}$, it's a coin (Bernoulli process). Model parameters, θ the bias of the coin (probability of a head).

- Model
$$p(D|\theta, \mathcal{H}) = \theta^3(1-\theta)^2$$
- Prior
$$\pi(\theta|\mathcal{H}) = Beta(1,1) \propto \theta^{1-1}(1-\theta)^{1-1}$$

  (uniform on the range $[0, 1]$)

Example elements of Bayes Theorem

Data, D 3 Heads, 2 tails. Hypothesized model structure, $\mathcal{H}$, it's a coin (Bernoulli process). Model parameters, θ the bias of the coin (probability of a head).

- Model
$$p(D|\theta, \mathcal{H}) = \theta^3(1-\theta)^2$$
- Prior
$$\pi(\theta|\mathcal{H}) = Beta(1,1) \propto \theta^{1-1}(1-\theta)^{1-1}$$
(uniform on the range $[0,1]$)
- Posterior
$$\pi(\theta|D, \mathcal{H}) = Beta(1+3, 1+2) \propto \theta^{4-1}(1-\theta)^{3-1}$$

The Marginal Likelihood

While the first three expressions listed above are always presented and discussed, less is said about the interpretation of the Marginal likelihood, or Integrated likelihood or Evidence (note, however, [3]). See Ref. [4] We can interpret $p(D|\mathcal{H})$ as being the probability of our data if we accept our hypothesized model structure $\mathcal{H}$ (namely $\pi(\theta, D|\mathcal{H})$) and integrate out our parameters $p(D|\mathcal{H}) = \int p(D|\theta, \mathcal{H})\pi(\theta|\mathcal{H})d\theta$. I'll draw an example of $p(D|\mathcal{H}_1)$ and $p(D|\mathcal{H}_2)$ and discuss Occam's razor [2].

Model Comparison and Bayes Factors

Given two models which are equivalent to two different hypotheses about the data, $\mathcal{H}_1$ and $\mathcal{H}_2$ with respective prior probabilities $p(\mathcal{H}_1)$ and $p(\mathcal{H}_2)$, we can ask about the relative probabilities of the data under these hypotheses. This is given by the Bayes Factor.

Bayes Factor

$$B = \frac{p(D|\mathcal{H}_1)}{p(D|\mathcal{H}_2)}$$

This is a ratio of the marginal likelihoods or model evidence. A rule of thumb is that if $B > 3$ then there is substantial evidence against hypothesis 2 relative to 1.

Bayesian Protocol for Humans

1. Understand your problem and its context.
2. Formulate an appropriate probabilistic model which allows you to write down $p(D|\theta, \mathcal{H})$ a likelihood for your data D.
3. Pick an appropriate choice of priors $p(\theta|\mathcal{H})$ given your model.
4. Reason about some actual data and so find the posterior over θ.

Remark on Priors and Hyperpriors

In the literature there is a great deal of discussion about appropriate choice of (un/informatative) priors and debate about the role of confidence intervals etc. [1]. This hasn't stopped Bayesian inference being very influential and widely industrially applied in Machine Learning.

We are going to hope that correctly characterizing the prior information that the biological system has been exposed to will hand us the priors in a natural manner: quantifying the prior experience of a system is part of describing its structure and understanding its function. Prior information and model parameters can have *physical* interpretations in terms of changes in RAM or disk magnetization or chemical modifications.

In practice Bayesians often write their priors $p(\theta)$ as parameterized distributions. These parameters are called hyper-parameters e.g. the μ and σ in the prior $p(\theta) \sim \mathcal{N}(\sigma, \mu)$.

Sampling and Posteriors

Here is a crude method of performing inference even if relevant integrals are intractable.

Algorithm

1. Use $\pi(\theta|\mathcal{H})$ to make many, N, draws of possible model parameters: $\{\theta_1, ...\theta_N\}$.
2. For each $\theta_i \in \{\theta_1, ...\theta_N\}$ plug it into the likelihood function (given data) $f_i = p(D|\theta_i, \mathcal{H})$.
3. Given the set of $\{f_1, ...f_N\}$ and using an appropriate density estimation (e.g. a histogram, [1]) we will approximate $\pi(\theta|D, \mathcal{H})$ (if we normalize) increasingly well with N.

We will be investigating more common methods in subsequent lectures.

Not covered but relevant to understanding inference

1. Improper priors: flat priors are not transformation invariant (not always a problem).
2. Jeffrey's priors: are transformation invariant. [1]
3. The connection between The Marginal Likelihood and the Partition function in Statistical Physics. [3]
4. Bayesian Non-parametrics: Gaussian, Chinese Restaurant and Dirichlet Processes.

Topics covered

- Introduced Bayes Theorem with an emphasis on the Marginal Likelihood and an operational approach
- Elements of model comparison
- Simple evaluation of posteriors through simulation

Bibliography

[1] L. Wasserman (2004). All of statistics: a concise course in statistical inference. Springer.

[2] I. Murray and Z. Ghahramani Technical note: A note on the evidence and Bayesian Occams razor

[3] D. J. C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.

[4] N. Friel and J. Wyse. Estimating the evidencea review. Statistica Neerlandica 66.3 (2012): 288-308.