Sampling and the Brain: Inference Control and Driving

Nick Jones

nick.jones@imperial.ac.uk

So far

What we have covered so far.

Inference with chemicals and point process models for chemistry (and neurons).

The problem of normalization.

Sampling (Rejection, Importance, Metropolis, Gibbs).

Encoding uncertain information using probabilistic population coding.

Today: Binocular rivalry and Gibbs sampling

We will now illustrate the preceding ideas with a study and implementation of the following paper:
Multistability and Perceptual Inference - Samuel J. Gershman Edward Vul Joshua B. Tenenbaum - Neural Computation 2012
You can read about the basics of binocular rivalry here:
http://www.scholarpedia.org/article/Binocular_rivalry

1. MCMC in the form of Gibbs sampling
2. (in passing an example of a Markov Random Field)
3. An alternative view to Probabilistic Population Coding

In this example we are not attempting to encode entire distributions (this is what PPC does).
Instead we are merely sampling from our posterior as in MCMC.

Plan + Rivalry

We'll cover:

1. Binocular rivalry
2. The model
3. Implementation tips

If we are presented with different images to our left eye $\mathbf{x^L}$ and to our right eye $\mathbf{x^R}$ then what image $\mathbf{s}$ do we perceive?

Images and Perceptions

If we are presented with different images to our left eye $\mathbf{x^L}$ and to our right eye $\mathbf{x^R}$ then what image $\mathbf{s}$ do we perceive?
In practice you appear to see the image from one eye dominating for a while and then flipping to the other eye.
Each image has N binary pixels and $x^i \in \{0, 1\}$ where $i \in \{L, R\}$.
We suppose that our perception of the n^{th} pixel of the image is a linear combination $s_n = w_n x_n^L - (1 - w_n) x_n^R$.
For simplicity we'll treat $\mathbf{w}$ as a binary vector. Perceiving the image as only that from the left eye would be $\mathbf{w} = \mathbf{1}$.

Occlusion

Left eye $\mathbf{x^L}$; Right eye $\mathbf{x^R}$; Perception $\mathbf{s}$; $s_n = w_n x_n^L - (1 - w_n) x_n^R$; $\mathbf{w}$ is a binary vector.

For fixed images we might expect a statistical model to yield a posterior distributed about a consensus version of $\mathbf{w}$ e.g. $\mathbf{w} = \frac{\mathbf{1}}{\mathbf{2}}$. *A fusion not a rivalry*.

Suppose that each eye's input is partly blocked by some dead pixels (e.g. intervening objects or vessels or eye's blindspot).

Encode these as images:

Left eye occlusion $\pi^{\mathbf{L}}$

Right eye occlusion $\pi^{\mathbf{R}}$

Suppose, further, that these occlusions undergo Glauber dynamics.If we couple this noise in appropriately *it can induce a bimodal distribution on the marginal posteriors over s_n with one mode associated with x_n^L and one associated with x_n^R*.

Graphic of the Graphical Model

I'll present a drawing on the board of the notional model.
I'll suggest why this is sensible.

Markov Random Field I

Define two Energies H_1 and H_2 which are minimized at zero temperature. We seek to minimize each of the following terms (which are coupled):

$H_1 =$ discord of perception with left eye data +
discord of perception with right eye data +
non-smoothness of perception in space +
degree of departure of perception from prior expectation

$H_2 =$ non-smoothness of occlusion in space (left + right eye) and amount of occlusion

I.e. a low energy choice of perception and occlusion has a smooth perception which is a good fit to both eyes' data which meets prior expectations and is constrained by a smooth and small occlusion in space.

Markov Random Field I continued

Define energies H_1, H_2 which are minimized at zero temperature. We seek to minimize each of the following terms (which are coupled):

discord of perception with i^{th} eye data:

$\sum_{n=1}^{N} \pi_n^i \frac{(x_n^i - s_n)^2}{2\sigma_i^2}$

non-smoothness of perception in space (where $\mathcal{C}_n$ encodes the neigbourhood of pixel n):

$\sum_{n=1}^{N} \sum_{j \in \mathcal{C}_n} (w_n - w_j)^2$

degree of departure of perception from prior expectation (b_n is the prior expected pixel value):

$\sum_{n=1}^{N} (b_n - s_n)^2$

non-smoothness of occlusion in space (left + right eye) and amount of occlusion

$\sum_{n=1}^{N} (\alpha(1 - \pi_n^i) + \sum_{j \in \mathcal{C}_n} (\pi_n^i - \pi_j^i)^2)$

Markov Random Field II

$H_1 = \sum_{n=1}^{N}((b_n - s_n)^2 + \sum_{j \in \mathcal{C}_n} \beta(w_n - w_j)^2) + \sum_i (\sum_{n=1}^{N} \frac{\pi_n^i (x_n^i - s_n)^2}{2\sigma_i^2})$
$H_2 = (\alpha(1 - \pi_n^i) + \gamma \sum_{j \in \mathcal{C}_n} (\pi_n^i - \pi_j^i)^2)$

We let $P(\mathbf{s}|\mathbf{x}, \pi) \propto exp(-\tau H_1)$ and recall $s_n = w_n x_n^L - (1 - w_n) x_n^R$.

System Dynamics is Sampling Dynamics

We suppose that our perceptions come as a result of a Gibbs sampling strategy. This can be interpreted as dynamics in both the perceptual field and the two occlusion fields. The interplay of occlusion and perception leads to a temporary dominance of one eye over the other.

Gibbs Sampling MCMC Strategy - as discussed in the lectures - sample from the marginal distributions on w_n and π_n^i. The strategy will prove simple as the marginals are easy to sample from (normalize) directly.

System Dynamics is Sampling Dynamics

Gibbs Sampling MCMC Strategy - as discussed in the lectures - sample from the marginal distributions on w_n and π_n^i.

$P(w_n|\mathbf{x}, \pi, \mathbf{w}_{/n}) \propto$

$exp(-\tau(b_n - s_n)^2 + \sum_{j \in \mathcal{C}_n} \beta(w_n - w_j)^2 - \tau \sum_i (\frac{\pi_n^i (x_n^i - s_n)^2}{2\sigma_i^2}))$

Where $\mathbf{w}_{/n}$ means holding all $\mathbf{w}$ constant save the n^{th} element the other terms are independent of w_n and so need not be considered. This is defined for w continuous though we'll treat it as binary.

And similarly:

$P(\pi_n^i|\pi_{/n}^i) \propto exp(-\tau(\alpha\pi_n^i + \gamma \sum_{j \in \mathcal{C}_n} (\pi_n^i - \pi_j^i)^2))$

Noting that the second distribution is defined only for $\pi_n^i \in \{0, 1\}$ (and that it is independent of $\mathbf{x}$ and $\mathbf{w}$).

Sampling and Sampling Dynamics

In this view we draw from the conditionals and (after burn-in) the set of sampled states approximates a sample from $P(\mathbf{s}, \pi|\mathbf{x})$. However when we interpret this as a process in the world then the claim is not just that the brain sensibly samples from $P(\mathbf{s}, \pi|\mathbf{x})$ it is that the dynamics of this sampling have psychological/physiological observables.

Code

- Look over the code and explore it.
- How long should burn-in be?
- Investigate the effects of increasing the trustworthiness of one of the eyes.
- Currently the two images are composed of random bits. Experiment with different types of images.
- Write down the Hamiltonian that is implemented in the code.
- Advanced: in the original W is a continuous variable drawn from a Gaussian - in the implementation it is binary - re-code this.
- Advanced: study the original paper - they were able to generate traveling waves in the perceptual field. Can this be reproduced?
- Advanced: The authors consider the dynamics of π to depend on W consider this coupled model.

Concluding

What have we covered.
What can we say about modern implementations of inference using neural and chemical architectures.
Big questions remain regarding sensible chemical inference, the right choice of neural architecture, how to implement complex models and energy costs.
Bayesian inference can be burdened by prior information in an un-modeled non-stationary world.
NEXT: Control. Laplace transforms to study ODE systems.

Bibliography

The paper below is the basis for this lecture. There are a few bugs but it's both interesting and well written. There are a couple of implementation disconnects. In their paper: a) the term $\alpha\pi^i_j$ in the Hamiltonian should be $-\alpha\pi^i_n$ b) they actually used a smaller lattice than they state (a 5 by 5 image - that helps explain Fig. 2a) c) they suppose the input image has values $+1$ or -1 and they used w in a discretized fashion.

There are also some differences between the form we discuss and implement mostly for reasons of simplicity.

Multistability and Perceptual Inference - Samuel J. Gershman Edward Vul Joshua B. Tenenbaum - Neural Computation 2012 Society of America A, 20(7):14341448, 2003.