Bayesian methods in neuroscience: a transition tutorial

The Fab Five
<***@ntnu.no>

Draft of 15 May 2019 (first drafted 14 May 2019)

A tutorial for neuroscientists about the Bayesian way of thinking
Note: Dear Reader & Peer, this manuscript is being peer-reviewed by you. Thank you.

Tentative scheme

Luca:

Purpose

The purpose of this paper in not so much to present the maths behind Bayesian methods, as to explain the *different way of thinking* they require, compared to old probabilistic thinking. Here are some differences:

- Old thinking often considered dichotomous hypotheses that, on closer scrutiny, weren't so distinguished after all. For example: 'is such-and-such difference zero, or different from zero?'. Well, what about 10^{-100}? that's different from zero; but wouldn't it be zero for all practical purposes?

  New thinking instead asks about a continuum of hypotheses: 'what's the value of such-and-such difference?'. Once we have the most probable answer we can decide whether for us it can be considered as zero or not – and this depend on the purpose of our question.

- Old thinking often asked to estimate quantities that can't be measured, such as the 'probability' of something or a parameter of some probability model.

  New thinking asks about quantities that can in principle be measured. For example, the frequency of some feature in a population of a given size, or in a sequence of a given number of experiments, or the average in such a sequence.

- Old thinking calculated probabilities to *predict* something.

  New thinking doesn't want to 'predict' – it leaves that magical task to coffee grounds. It just wants to *quantify our uncertainty* about something. If an action or choice is required – but that's not always the case – then from our quantified uncertainty and from the costs of performing available actions we can decide which action is best to choose. Compare the remark by Rasch (Lauritzen 1974a p. 268):

  > The "prediction" is suggestive of the statistician as a magician who can tell the future. [. . .] To speak seriously: you do not really *predict* anything. What you do, is to calculate the distribution of the variate in question, possibly offering its mean value or the like as a likely event – but only on the assumption that the model – or a characteristic feature of it – on which you based this forecasting, still holds, i.e. confronted with what eventually does happen *you are faced with a test of this hypothesis* and nothing else – *you were not telling what the future would be*!

- 'Significance'...

Layout

My idea would be to take a simple but realistic research question and show how to address it from beginning to end, pointing out the differences from the old way of thinking.

An example: we wonder whether some feature – say, neuronal activity or behaviour – changes in two different settings, or with two different stimuli, or at two different ages, or with two different conditionings, or similar.

Traditionally this is stated as the question 'is the difference between the two cases zero or significant?'. We show that it can instead be phrased as 'how much is the difference?', and the answer is a probability distribution for all possible values of the difference.

In addition to this we can show the following:

- There may be choices to make based on the difference; for example, further experiments, or buying new equipment, or something else. The choices have different costs. Based on the costs we may end up making very different choices, even if the probability distribution for the difference is always the same. This example would show that speaking of 'significance' at the start was a bit preemptive and too sweeping, because it wasn't clear for what it was significant.
- The calculation of the final probability distribution requires a pre-data probability distribution about the hypotheses. We can consider

very different (but still sensible) pre-data distributions – for example representing 'sceptics' and 'believers' in the hypotheses. Then we show how to roughly calculate how many data points would be necessary for the final distribution to be more or less the same, irrespective of the pre-data ones. This amount of data would the be sufficient to convince a 'sceptic'.

On the other hand, it sometimes happens that the scientific community has already a strong belief in some hypothesis, because of some general biological reasons or indirect experiments. In this case, only a few datapoints are needed to make the community almost certain about the hypothesis. Again, if we represent the initial belief with the pre-data distribution, we can estimate the amount of datapoints necessary to feel satisfied about the hypothesis, leading to some resource savings.

1 Intro

Bibliography

('de *X*' is listed under D, 'van *X*' under V, and so on, regardless of national conventions.)

Barndorff-Nielsen, O. E., Blæsild, P., Schou, G., eds. (1974): *Proceedings of Conference on Foundational Questions in Statistical Inference: Aarhus, May 7–12, 1973*. (University of Aarhus, Aarhus).

Lauritzen, S. L. (1974a): *Sufficiency, prediction and extreme models*. In: Barndorff-Nielsen, Blæsild, Schou (1974), 249–269. With discussion. Repr. without discussion in Lauritzen (1974b).

— (1974b): *Sufficiency, prediction and extreme models*. Scand. J. Statist. **1**[3], 128–134.