COMPUTING BAYES IN BIG SPACES

John Skilling
Maximum Entropy Data Consultants Ltd, Kenmare, Ireland
(skilling@eircom.net)

Abstract

The traditional aim of Bayesian computation has been to compute probability distributions (as a set of samples) with accuracy

$$\delta\big(\log \texttt{Evidence}\big) \le O(1)$$

sufficient to enable models to be ranked whenever the Bayes factors are $O(1)$ or more. In large-scale problems, this ideal is increasingly unattainable except at prohibitive cost. Practicality dictates that computing cost should grow with size only linearly, or not much more. In turn, this dictates compromise with precision.

Statistical mechanics offers an analogy. Here, the number of degrees of freedom is huge, of the order of Avogadro's number ($N \sim 10^{24}$). Entropy $S = \log \Omega$ is a linear ("extensive") variable, so that a thermodynamic system with N degrees of freedom has entropy $S = O(N)$. Fluctuations tend to be $O(N^{1/2})$, so that measuring to precision $\Delta S = O(1)$ would be meaningless. It would also be impossible in practice to attain a precision of 1 part in 10^{24}. Despite this, statistical mechanics is a productive discipline.

In Bayesian calculus, the controlling variable which corresponds to degeneracy Ω is the evidence Z, and the extensive form corresponding to entropy is $\log Z$. For most purposes, when comparing models, the only differences of $\log Z$ that really matter are large,

$$\Delta\big(\log \texttt{Evidence}\big) = O(N)$$

Any difference less than $N^{1/2}$ tends to reflect the particular realisation of noise in the data rather than an important difference between the models, so that computation to the traditional $O(1)$ accuracy would be largely meaningless. It may also be impractically expensive in practice. Despite this, we require Bayesian computation to be a productive methodology, even in large spaces.

On encountering statistical mechanics, most students have a sense of intellectual vertigo at the enormous numbers involved with Ω until they learn to think logarithmically in terms of entropy S. A similar adjustment of mindset is needed for Bayesian computation whenever $N \gg 1$.

I will argue that nested sampling is the natural algorithm for the exploration and quantification of large spaces. Gradient information, where available, can be used. There is also a rather unexpected rôle for curvature, which promises a distinctive new avenue of research.