Bayesian Analysis (2013) **8**, Number 2, pp. 319–322

Comment on Article by Müller and Mitra

Anthony O'Hagan*

Müller and Mitra have given us a superb paper, eloquently arguing for the many applications of Bayesian nonparametric (BNP) methods. There is no doubt that such techniques have enormous value in a wide range of contexts. I want to raise two linked areas of quite general concern, but these should not be read as detracting in any way from the quality and importance of this paper.

1 Prior information

Bayesian methods require a prior distribution that encodes genuine prior information about the model parameters. I find it quite depressing how rarely this fact is taken seriously in published work which professes to be Bayesian. To illustrate ideas I will look at the prior distribution in the authors' Example 1 and ask what genuine prior information it encodes.

All BNP methods involve specifying a prior distribution for a function. In Example 1, the unknown function in question is the probability mass function F, where $F(y)$ is the probability that a given type of T-cell will be observed y times in the probe. The problem requires that we specify our prior knowledge about F; that is, we need to specify a joint prior distribution for $\{F(0), F(1), F(2), \ldots\}$. This is an infinite-dimensional distribution (as will invariably be the case in BNP applications), so we are looking at a complex problem.

Complex problems benefit from being build up in stages, so first consider a simple parametric model. A natural choice in this problem is to suppose that F is a Poisson distribution $Po(\lambda)$, for some λ, and then to put a prior distribution on λ. The use of the word 'model' here is enlightening — all models involve some degree of simplification of reality. In this case, the parametric model is a simplification of our real prior beliefs. It states that the prior distribution for F gives zero prior probability for all possible distributions F that are not Poisson distributions. What the prior for λ says about F depends to some extent on what judgements were actually used to derive it. Typically, because λ is generally seen as the mean of the Poisson distribution (although of course it is also the variance), its prior distribution will be elicited by making judgements about the mean $\mu(F) = \sum_{y=0}^{\infty} yF(y)$ of F. A few specific judgements such as median and quartiles of $\mu(F)$ will have been made and a convenient distribution fitted to those judgements. The full distribution for F is then completed by a judgement that F is likely to be unimodal and similar to a Poisson distribution, and the parametric assumption of a Poisson distribution is then a convenient choice that is 'fitted' to this judgement.

The underlying approach applies to all pragmatic prior distribution specification: a

*Department of Probability and Statistics, University of Sheffield, Sheffield, UK, a.ohagan@sheffield.ac.uk

 DOI:10.1214/13-BA811C