Bayesian Analysis (2012) **7**, Number 1, pp. 47–50

Rejoinder

Alessio Sancetta*

I thank Professor Clarke for his sharp comments. He stresses the important fact that in most regular cases,

$$\mathbb{E}^{\theta} D_t \left(P_\theta \| P_w\right) = O\left(d/t\right),$$

where d is the dimension of θ (I am using the notation in my text). The intuition is based on his eq. (1), which is one of the results in Clarke and Barron (1994). The analytical case for the exponential family with conjugate priors shows that this is indeed the case. The crucial ingredient is eq. (12). Effectively, the argument should fail as soon as the posterior does not satisfy the Bernstein–von Mises Theorem. That is, the argument relies on asymptotic normality of the posterior distribution, or more precisely on the posterior concentrating around the MLE when the MLE is asymptotically normal. This is clearly stated by Professor Clarke in the last paragraph of his Section 2.

Of course, the behaviour of the t^{th} stage risk is what really matters for the practical problem of prediction. This is not directly addressed in the paper. The paper points out that in most regular cases, the cumulative expected risk is $O\left(\ln T\right)$, however, this begs the question of one example when this is not the case. As remarked by Professor Liang, one does not need to look at uncommon circumstances for the supremum of the resolvability index to be infinite. Hence, it is of interest to find examples where this is finite, but grows faster than $\ln T$. One such case is when the prior gives too little weight to some regions in the parameter space. The following is rather artificial: $\Theta = [0,1]$, $w\left(d\theta\right) = C \exp\left\{-\theta^{-c}\right\}$ for $C, c > 0$. When

$$\mathbb{E}^{\theta}\left[\ln p_{\theta'}\left(Z_t|\mathcal{F}_{t-1}\right) - \ln p_{\theta}\left(Z_t|\mathcal{F}_{t-1}\right)\right] = O\left(|\theta - \theta'|\right),$$

we need $|\theta - \theta'| \leq \delta/T$, but the prior gives weight

$$w\left(B_T\left(\theta\right)\right) = O\left(\exp\left\{-\left(\frac{T}{\delta}\right)^c\right\}\left(\frac{\delta}{T}\right)\right)$$

when $\theta = 0$. Hence, taking logs, the resolvability index at 0 is

$$R_T\left(0\right) = O\left(\inf_{\delta>0}\left\{\delta + \left(\frac{T}{\delta}\right)^c - \ln\delta + \ln T\right\}\right)$$

which grows faster than $\ln T$, but is still $o\left(T\right)$, so that universality holds.

Universality may of course fail, but the resolvability index still be finite uniformly in the parameter space. Consider an AR(1) with autoregressive coefficient $\theta \in [0, 1+\epsilon]$

*The author was lecturer at the University of Cambridge until 2008. He has then worked on several algorithmic trading positions in the City of London. He now works as a freelancer. http://sites.google.com/site/wwwsancetta/

 DOI:10.1214/12-BA701REJ