

## Comment by Miroslav Kárný<sup>7,8</sup>

The authors have done a great job in describing the state of the art of Bayesian Non-Parametrics and have illustrated the ideas by interesting examples. Their presentation has one (quite wide-spread) methodological flaw I want to point to. Essentially, their paper answers the question “how” and misleads in answering the question “why”. They do not take seriously Box’s statement they cite (*all models are wrong*). Taking it literally, it would mean that the prior distribution should have support out of the model class (irrespective of finite or massive parametrisation) and no inference would be possible. Luckily enough, a straightforward inspection of the Bayes rule leads to the Sanov-type view, [Sanov \(1957\)](#); [Berec and Kárný \(1997\)](#), that the posterior distribution is to be interpreted as the probability that a model, within the considered model class, which does not contain reality in the generic case, is the *best projection of reality to this class*. Consequently,

- the non-parametric (massive parametric) inference is susceptible to the same problems as the standard parametrisation (for instance, ignoring continuity of the estimated distribution can cause non-acceptable modelling errors);
- the information about concentration of the posterior distribution is the information regarding how close we are to the best projection and not how close we are to reality: it is increased due to the massive parametrisation but *not* due to better information about closeness to reality;
- the entropy rate, which often reduces to the Kullback-Leibler divergence, is the only adequate Bayes-rule induced measure of closeness.

Technically, the objection against mixture-type modelling is not completely correct as progress in this respect is enormous and counteracts the curse of dimensionality (R. Bellman, [Bellman \(1961\)](#)), which is an inherent barrier of non-parametric inference. Please, take our work [Kárný et al. \(2006\)](#) as an example of a strong research and development stream in this respect.

## References

- Bellman, R. (1961). *Adaptive Control Processes*. Princeton University Press, NJ.
- Berec, L. and Kárný, M. (1997). “Identification of reality in Bayesian context.” In Warwick, K. and Kárný, M. (eds.), *Computer-Intensive Methods in Control and Signal Processing*, 181–193. Birkhäuser.
- Kárný, M., Böhm, J., Guy, T. V., Jirsa, L., Nagy, I., Nedoma, P., and Tesař, L. (2006). *Optimized Bayesian Dynamic Advising: Theory and Algorithms*. London: Springer.

<sup>7</sup>Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic

<sup>8</sup>The work was supported by the grant GAČR 13-13502S

Sanov, I. (1957). “On probability of large deviations of random variables.” *Matematičeskij Sbornik*, 42: 11–44. (in Russian), also in Selected Translations Mathematical Statistics and Probability, I, 1961, 213–244.