

## Rejoinder

Peter Müller \* and Riten Mitra †

We appreciate the many insightful comments and critiques in the discussions. Several discussions pointed out important BNP models and classes of problems that we missed in the paper. **Gelman** proposes consideration of models for classification trees (CART, BART). We actually considered including the Bayesian CART ([Chipman et al. 1998](#); [Denison et al. 1998](#)) in the review paper, and strongly agree that models like the BART model ([Chipman et al. 2010](#)) turn out to be amazingly versatile in many applications. **Kolossiatis** draws attention to recent literature on correlated NRMIs as an alternative to DDP priors for multiple related random probability measures. We appreciate the discussed models as alternatives to the DDP, and also for their elegance. See Sections 2.3. and 4.3. of the main paper for the definition of NRMIs and the DDP model. **Kottas, DeYoreo** and **Poynor** highlight curve fitting regression approaches as another important alternative implementation of fully nonparametric regression. We strongly agree and appreciate that their discussion added this important class of approaches to the review. **Perron** mentions models for copulas. **Tokdar** reviews quantile curves as another great example for problems where principled BNP inference can address limitations in currently used approaches.

Several discussions highlight some features and limitations of BNP inference, beyond what is already discussed in the paper. **Robert** and **Rousseau** point out that while asymptotic properties of the estimation of the random curve or probability measure are understood for many BNP priors, asymptotics for other important inference summaries are not. We agree that this is an important current limitation of BNP and thank the discussants for highlighting this issue. **Carlin** and **Murray** give a spirited discussion as die hard Bayesian parametricians. By re-analyzing some of the data used in the paper they argue for alternative parametric models. We comment on details for the specific examples below. But we agree with the overall assertion that well chosen parametric models can often achieve similarly flexible inference. **Parmigiani** and **Trippa** make a related comment, by pointing out the sometimes blurred nature of the boundary between parametric and BNP methods in Bayesian inference. **Hoff** argues that BNP priors in practice rarely represent actual prior beliefs. In many cases, and with respect to many details of the BNP prior, this is probably true. However, several steps can be taken, and are used by many authors, to mitigate this concern. Many BNP priors allow convenient prior centering. In the manuscript we discussed this for the DP and the PT prior. Also, investigators can use prior simulation to verify that typical prior realizations do in fact match actual prior beliefs. We did this, for example, when setting up the prior in Example 4 ([Berger et al. 2012](#)). However, in many cases BNP priors include features that are not directly related to actual prior information. For example, the hierarchical prior on the partition boundaries in mixture of PT models is used only to reduce the posterior sensitivity with respect to partition boundaries. Hoff's discussion

---

\*Department of Mathematics, University of Texas, [pmueller@math.utexas.edu](mailto:pmueller@math.utexas.edu)

†ICES, University of Texas, [riten82@gmail.com](mailto:riten82@gmail.com)

offers an interesting approach to avoid related problems by using a pseudo likelihood for some low-dimensional summary of interest  $\theta$  of the infinite dimensional quantity only. We agree with this recommendation when it is feasible. Such approaches are genuinely nonparametric, in the sense of avoiding a full specification and parametrization of the sampling model. For example, inference under the quantile pyramid model of [Hjort and Walker \(2009\)](#) can be implemented with a pseudo likelihood based on quantiles only. Also **O'Hagan** critically discusses prior specification in BNP models. He argues that it is difficult to model judgements about quantities that are meaningful in the original problem, such as, for example, the mean  $\mu(F)$  of a random probability measure  $F$ . We agree. For prior specification, simple prior simulation can be helpful. O'Hagan also comments that  $\mu(F)$  is random, even when the prior mean were fixed, say at  $E(\mu(F)) = 0$ . This is important when modeling random effects distributions or residual distributions. The non-zero mean can complicate the interpretation of posterior inference. We commented on this issue in the paper, in the paragraph following equation (9). **Clarke** and **Holt** argue that an interesting and in the literature mostly overlooked application of BNP occurs in testing. We appreciate the example that they construct to make the point, and strongly agree with the conclusion. We would add that some interesting uses of BNP arguments also occur in inference for multiple comparison problems ([Guindani et al. 2009](#); [Leon-Novelo et al. 2013](#)) **Kárný** argues that also inference under BNP priors is still making specific model assumptions. Inference should be interpreted as providing the best projection of reality to the assumed model class. This important point puts the argument from the opening paragraph of the paper in perspective. **Arbel** and **Nipoti** discuss an important and perhaps little known property of DPM models by arguing that inference is amazingly robust with respect to different scales of the data.

Finally, some discussions critically review the examples from the paper and add some more. **Hoff** argues that in Example 1 (T-cell receptors) the data provides little information about  $F(0)$ . We disagree. Consider the counts  $f_j$  in Figure 1a, for  $j = 4, 3, 2, 1$ . In an ad-hoc analysis one could consider an extrapolation to  $f_0$ . Adding some uncertainty one could argue for inference on  $f_0$ , and thus  $N$ . The BNP model formalizes this argument. Related, **Carlin** and **Murray** compare BNP inference in Example 1 with several alternative parametric models. However, the hurdle Poisson model parametrizes the model such that the implied prior on  $F(0)$  is independent of  $F(j)/(1 - F(0))$ . This implies that one can not possibly learn about  $F(0)$  from the data under the hurdle Poisson model. The model does not formalize the informal notion of extrapolation that we mentioned before. The zero-inflated Poisson model, on the other hand, does not impose independence, but still fails to formalize extrapolation of a trend in  $F(4), \dots, F(1)$ . The seemingly poor fit of  $F(1)$  in Figure 1a to the empirical frequency  $f_1$  is due to sampling conditional on  $y_i > 0$ . For a better comparison we should have perhaps plotted the estimates for  $F(1)/(1 - F(0))$ . **Aitkin** and **Polak** also discuss Example 1 and question the use of overly complex models for this application with limited data. While we appreciate much of the argument, in particular the useful references, we feel that  $n = 55$  observations might allow some flexibility in inference for a discrete distribution. The BNP model only formalizes what one could do by ad-hoc extrapolation of the frequencies  $f_j$  to  $f_0$ . **O'Hagan** draws attention to the fact that the use of the DP mixture model in Example 1 is not motivated by the need for a

continuous distribution. In fact,  $F(y)$  in Example 1 is a discrete distribution. Rather, the motivation for the DP mixture in Example 1 is the fact that it formalizes the notion of extrapolating the trend in  $F(4), \dots, F(1)$  to the unknown  $F(0)$ . We should have pointed that out more prominently in the discussion. Also, while we agree with O'Hagan's cautioning advice against extrapolation in general, in this particular example we feel that the extrapolation to  $y = 0$  is reasonable. It is based on the assumption that the change in  $F(y)$  from  $y = 1$  to 0 should be commensurate with the changes from  $y = 4$  through 1. **Carlin** and **Murray** also briefly discuss inference in Example 2 (prostate cancer study). Inference that is reported in Figure 2 is based on a PT prior for the distribution of the time to progression times, without use of the longitudinal covariate and without additional frailties. The discrepancy of the estimated survival functions  $S_j(y)$  and the Kaplan-Meier (KM) estimate is only due to the extensive censoring for large  $y$ . The KM estimate eventually levels off, while the model-based estimate under the PT prior includes an imputation of the censored event times. The estimated survival functions continue to decrease towards 0. We thank the discussants for pointing out the switched labels. An earlier version of the manuscript had  $G_1$  and  $G_2$  wrongly matched with the two treatments CH and AA. We corrected the error. **Carlin** and **Murray** also consider inference in Example 3 (sarcoma trial) and point out that an exchangeable parametric prior leads to very similar estimates as we show for the BNP prior in Figure 4. The example highlights a common feature of BNP inference. In hindsight one can often find a parametric model that could have matched BNP inference, as is the case here using the exchangeable model proposed by Carlin and Murray. However, the exchangeable parametric model would have probably not been chosen up front to set up a regression on prognosis. We feel that a partially exchangeable model might have been a more natural model choice, perhaps even including a monotonicity constraint across different prognoses. Surprisingly, posterior inference actually shows the opposite trend from what one might have expected. Posterior mean success probabilities for sarcomas with "good" prognosis are below the posterior means of several sarcomas with "intermediate" prognosis. A strength of the BNP model is that we can set up a prior model that is centered around some parametric assumptions, while allowing for posterior inference to deviate from these assumptions when the data so require. While similar flexibility can be imitated in parametric models with model selection one would always have to anticipate the nature of the model alternatives. **Paddock** and **Savitsky** introduce an interesting application of BNP methods to modeling random effects distributions, highlighting an important feature of typical BNP priors in this context. Using discrete random probability measures gives rise to a clustering of experimental units. This becomes important when the number of observations per experimental unit would be too small for the desired inference without some borrowing of strength across experimental units. The latter can be achieved by forming clusters of experimental units that share common random effects. We agree with this observation. As Paddock and Savitsky point out, the use of the BNP prior in Example 6 is similarly motivated. A single triple  $(y_{i1}, y_{i2}, y_{i3})$  would not allow meaningful inference on and decisions related to the tripeptide and organ-specific random effects  $(\mu_i, \beta_i, \delta_i)$ . Only by borrowing strength across similar experimental units by way of clustering are we able to report meaningful posterior probabilities of increasing mean counts.

In summary, we appreciate the many insightful and clarifying comments in the discussion. In particular, we agree with some of the cautioning comments, including Carlin and Murray’s advice to avoid “flexibility for flexibility’s sake.” And we agree with Hoff’s comment that BNP procedures are not without strong modeling assumptions. And we share O’Hagan’s concerns about the importance of prior specification in BNP models.

## References

- Berger, J., Jefferys, W. H., and Müller, P. (2012). “Bayesian Nonparametric Shrinkage Applied to Cepheid Star Oscillations.” *Statistical Science*, 27(1): 3–10. 357
- Chipman, H. A., George, E. I., and McCulloch, R. E. (1998). “Bayesian CART model search (C/R: p948-960).” *Journal of the American Statistical Association*, 93: 935–948. 357
- (2010). “BART: Bayesian Additive Regression Trees.” *The Annals of Applied Statistics*, 4(1): 266–298. 357
- Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). “A Bayesian CART algorithm.” *Biometrika*, 85: 363–377. 357
- Guindani, M., Müller, P., and Zhang, S. (2009). “A Bayesian Discovery Procedure.” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 71(5): 905–925. 358
- Hjort, N. L. and Walker, S. G. (2009). “Quantile Pyramids for Bayesian Nonparametrics.” *The Annals of Statistics*, 37(1): pp. 105–131. 358
- Leon-Novelo, L. G., Müller, P., Arap, W., Kolonin, M., Sun, J., Pasqualini, R., and Do, K.-A. (2013). “Semiparametric Bayesian Inference for Phage Display Data.” *Biometrics*, to appear.  
URL <http://dx.doi.org/10.1111/j.1541-0420.2012.01817.x> 358