## Comment by Athanasios Kottas[10], Maria DeYoreo[10] and Valerie Poynor[10]

We commend the authors for an interesting review of applications of Bayesian nonparametric modeling and inference. Here, we offer some additional discussion, results and references on fully nonparametric regression, which we believe is a key success story of Bayesian nonparametrics.

As the authors discuss in Section 4, two dominant trends in the Bayesian regression literature have been to develop flexible regression function models and to accompany the regression relationship with more comprehensive uncertainty quantification. For problems involving a small to moderate number of random covariates, the *curve fitting regression* approach is an appealing alternative. Specifically, a DP mixture model, $f(y, \boldsymbol{x}; G) = \int k(y, \boldsymbol{x}; \boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta})$, $G \sim \mathrm{DP}(\alpha, G_0)$, is used for the joint distribution of the response, $y$, and covariates, $\boldsymbol{x}$, from which inference emerges for the conditional response distribution, $f(y \mid \boldsymbol{x}; G)$. Modeling the joint response-covariate distribution is natural for many applications, especially in the environmental and biomedical sciences.

Although the approach (based on normal mixtures) has been proposed in Müller et al. (1996), it has been overlooked as a general nonparametric regression framework until relatively recently. This may be attributed to the limitations of posterior predictive estimation for full inference about the conditional distribution $f(y \mid \boldsymbol{x}; G)$. However, with posterior simulation extended to the mixing distribution $G$, the DP mixture curve fitting approach enables rich inference for response densities that can change in non-trivial fashion across the covariate space, and for non-linear regression relationships built from the mean or from percentiles of the response distribution (Taddy and Kottas 2009, 2010). Moreover, the methodology can be extended to handle categorical responses (Shahbaba and Neal 2009; Dunson and Bhattacharya 2011), and looking beyond the standard regression setting, to develop emulation and calibration techniques for stochastic computer simulators (Farah 2011) as well as modeling for marked Poisson processes (Taddy and Kottas 2012).

A particularly promising direction involves problems with (possibly multivariate) ordinal responses, $y$, which can be represented as discretized versions of latent continuous responses, $z$, with a DP mixture model employed for the joint distribution of $z$ and $\boldsymbol{x}$. For continuous covariates, the mixture kernel can be built from a multivariate normal which, in the presence of binary responses, requires identifiability restrictions for its covariance matrix (DeYoreo and Kottas 2013). This modeling approach enables flexible nonparametric inference for the implied response classification probabilities, $\Pr(y = j \mid \boldsymbol{x}; G)$, the number of which increases significantly in multivariate ordinal regression problems rendering semiparametric modeling infeasible. Figure 3 illustrates the capacity of the model to uncover both relatively standard and non-monotonic shapes for the ordinal regression relationships as well as non-trivial interactions among covariates.

The curve fitting regression framework can be enhanced with hierarchically depen-

[10]Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA

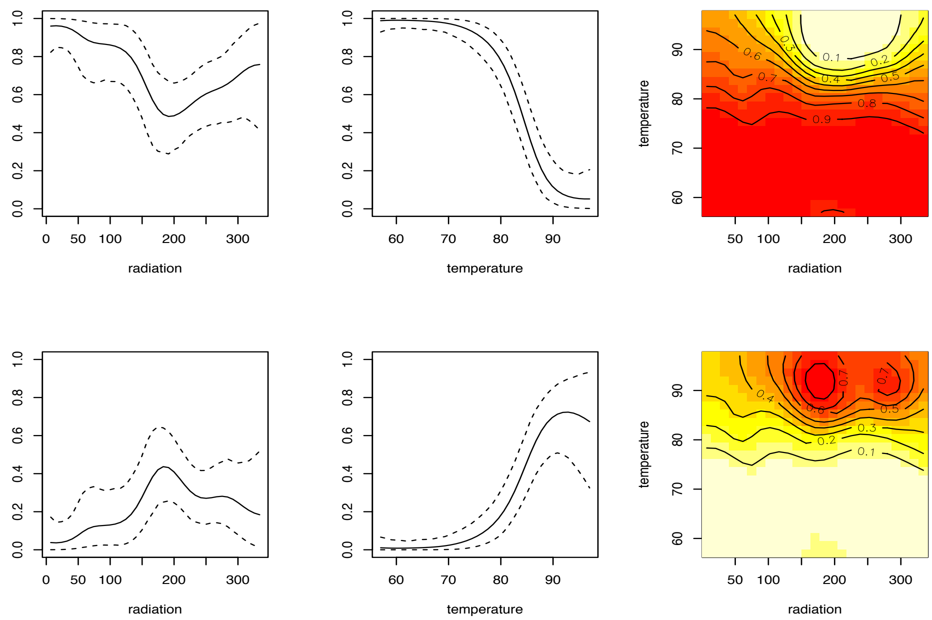


Figure 3: Ordinal regression example using data (available in `R`) on ozone concentration (variable $z$ measured in ppb), radiation (variable $x_1$ in langley units) and temperature (variable $x_2$ in degrees Fahrenheit) recorded over 111 days from May to September of 1973 in New York. Ozone concentration is discretized to construct an ordinal response, $y$, with classifications of "low", "medium", and "high" concentration corresponding respectively to: $y = 1$ for $z \leq 50$ ppb; $y = 2$ for 50 ppb $< z \leq 100$ ppb; and $y = 3$ for $z > 100$ ppb. Inference results are based on a trivariate normal DP mixture model for $(z, x_1, x_2)$. The top and bottom row panels include inference results for the "low" and "medium" categories; specifically, for $j = 1, 2$, the left and middle columns show posterior mean and 95% interval estimates for $\Pr(y = j \mid x_1; G)$ and $\Pr(y = j \mid x_2; G)$, and the right column plots the posterior mean estimate of $\Pr(y = j \mid x_1, x_2; G)$.

dent nonparametric priors (e.g., Rodriguez et al. 2009). More generally, it can be utilized complementary to DDP regression, with survival analysis providing a practically important area of application. Survival regression problems typically include a treatment categorical factor in addition to random covariates $\boldsymbol{x}$. For instance, for a generic treatment/control setting (indicated by $s \in \{T, C\}$), the joint response-covariate distribution can be modeled with $f(y, \boldsymbol{x}; G_s) = \int k(y, \boldsymbol{x}; \boldsymbol{\theta}) \mathrm{d}G_s(\boldsymbol{\theta})$. The choice of the kernel is important to ensure desirable properties for key functions of the survival response distribution, such as the hazard and mean residual life functionals (Poynor and Kottas 2013). Assigning a DDP prior to the pair of mixing distributions $(G_C, G_T)$ results in related regression relationships through the dependent T/C response distributions $f(y \mid \boldsymbol{x}; G_s)$. Here, a variable-weights DDP prior, $G_s = \sum_{h=1}^{\infty} \pi_{sh} \delta_{\tilde{\theta}_h}$, is an attractive

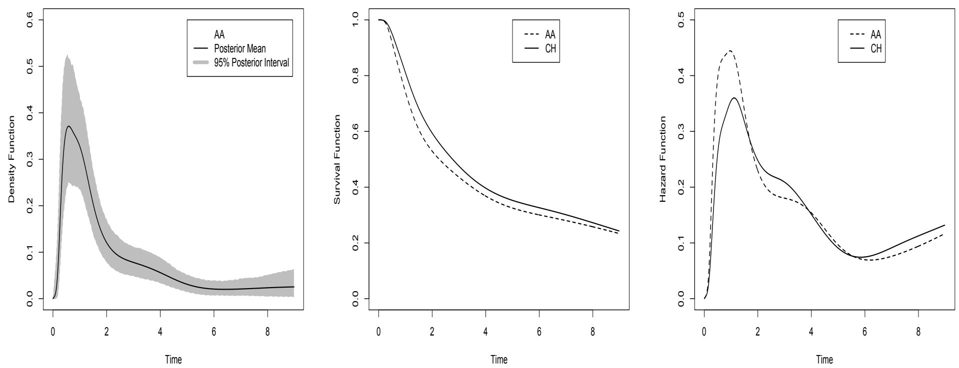


Figure 4: Prostate cancer study (Example 2 of the paper). Inference results are based on a gamma DDP mixture model, $f(y; G_s) = \int \text{gamma}(y; \theta, \phi) \mathrm{d}G_s(\theta, \phi)$, where $s \in \{\text{AA}, \text{CH}\}$, with a variable-weights DDP prior assigned to $(G_{\text{AA}}, G_{\text{CH}})$, using one of the bivariate beta distributions from Nadarajah and Kotz (2005) for the DDP stick-breaking weights. The left panel plots the posterior mean and 95% uncertainty bands for the treatment AA density function. The middle and right panels show the posterior mean estimates under the two treatments for the survival functions and for the hazard rate functions, respectively.

alternative to the basic DDP model (expression (8) of the paper); incorporating dependence through the DDP weights is invariant to the mixture kernel dimensionality, and for this application, it may be more natural to envision similar mixture locations with prevalence varying according to the T/C groups. To retain the DP marginally, we need an appropriate bivariate beta distribution for the latent variables $(\upsilon_{Ch}, \upsilon_{Th})$ that define the stick-breaking weights. For an illustration, Figure 4 shows results based on the portion of the data from the prostate cancer study made available on-line. Note that the point estimates for the hazard functions suggest a non-proportional hazards relationship for TTP under the two treatments providing further demonstration for the practical utility of flexible Bayesian nonparametric modeling relative to traditional parametric or semiparametric regression models.

## References

DeYoreo, M. and Kottas, A. (2013). "A fully nonparametric modeling approach to binary regression." Technical Report UCSC-SOE-13-03, Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

Dunson, D. B. and Bhattacharya, A. (2011). "Nonparametric Bayes regression and classification through mixtures of product kernels." In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M.

(eds.), *Bayesian Statistics 9, Proceedings of the Ninth Valencia International Meeting*, 145–164. Oxford University Press.

Farah, M. (2011). "Bayesian nonparametric methods for emulation, sensitivity analysis, and calibration of computer simulators." Ph.D. thesis, *Statistics and Stochastic Modeling*, University of California, Santa Cruz.

Müller, P., Erkanli, A., and West, M. (1996). "Bayesian curve fitting using multivariate normal mixtures." *Biometrika*, 83: 67–79.

Nadarajah, S. and Kotz, S. (2005). "Some bivariate beta distributions." *Statistics*, 39: 457–466.

Poynor, V. and Kottas, A. (2013). "Nonparametric Bayesian inference for mean residual life functions in survival analysis." Technical Report UCSC-SOE-13-04, Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2009). "Bayesian nonparametric functional data analysis through density estimation." *Biometrika*, 96: 149–162.

Shahbaba, B. and Neal, R. (2009). "Nonlinear models using Dirichlet process mixtures." *Journal of Machine Learning Research*, 10: 1820–1850.

Taddy, M. and Kottas, A. (2009). "Markov switching Dirichlet process mixture regression." *Bayesian Analysis*, 4: 793–816.

— (2010). "A Bayesian nonparametric approach to inference for quantile regression." *Journal of Business and Economic Statistics*, 28: 357–369.

— (2012). "Mixture modeling for marked Poisson processes." *Bayesian Analysis*, 7: 335–362.