A nonparametric dependent process for Bayesian regression

Ruth Fuentes–García*, Ramsés H. Mena** and Stephen G. Walker***

*[1] Facultad de Ciencias, Universidad Nacional Autónoma de México. México, D.F. 04510, México.
** IIMAS, Universidad Nacional Autónoma de México. México, D.F. 04510, México.
*** University of Kent, Canterbury, Kent, CT2 7NZ, UK.

Abstract

This article investigates the problem of Bayesian nonparametric regression. The proposed model is based on a recently introduced random distribution function which is based on a decreasing set of weights. The approach is surprisingly of a much simpler form than alternative models described in the literature. A Gibbs sampler algorithm is provided for posterior analysis.

Keywords: Bayesian nonparametric regression; Conditional density estimation; Dependent nonparametric process.

1. Introduction. Bayesian non-parametric inference is by now becoming quite routine with one of the fundamental models being the mixture of Dirichlet process (MDP), which is based on the Dirichlet process, introduced by Ferguson (1973). We use $\mathsf{G} \sim \mathcal{D}(a\,G_0)$, to indicate that G is a Dirichlet process random distribution function with scale parameter $a > 0$ and expected distribution G_0. A constructive representation of the Dirichlet process is given by Sethuraman (1994) through

$$\mathsf{G}(\cdot) = \sum_{k=1}^{\infty} \omega_k \, \delta_{x_k}(\cdot), \tag{1}$$

where $x_k \overset{\text{iid}}{\sim} G_0$, $\{\omega_k\}_{k=1,2,\ldots}$ is a sequence of weights, independent of the x_k's, defined as follows: $\omega_1 := v_1$ and $\omega_k := v_k \prod_{j<k}(1 - v_j)$ with $v_i \overset{\text{iid}}{\sim} \mathrm{Be}(1, a)$, $a > 0$, for each $k = 1, 2, \ldots$. This latter specification of the weights, termed the stick–breaking construction, ensures that $\sum_{k=1}^{\infty} \omega_k = 1$ a.s.

The discreteness of the Dirichlet process, see for example Blackwell (1973), led Lo (1984) to consider the mixture of Dirichlet process model, which can be written in hierarchical form as

$$
\begin{aligned}
y_i \,|\, x_i &\overset{\text{ind}}{\sim} & K(\,\cdot\,|x_i) \\
x_i | \mathsf{G} &\overset{\text{iid}}{\sim} & \mathsf{G} \\
\mathsf{G} &\sim& \mathcal{D}(a\,G_0)
\end{aligned}
\tag{2}
$$

where $K(\,\cdot\,|x)$ is a density function for all x. This approach allows us to think of a random density function defined as

$$f_{\mathsf{G}}(y) = \int K(y \mid x)\mathsf{G}(\mathrm{d}x). \tag{3}$$

Note that, based on the Sethuraman construction, we can equivalently write

$$f(y) = \sum_{k=1}^{\infty} \omega_k \, K(y \mid x_k).$$

---

[1]E-mail for correspondence: rfuentes@matematicas.unam.mx

Current trends and interests have focused on generalizing the Dirichlet process to develop more flexible mixture models. See, for example, Regazzini *et al.* (2003) and Ishwaran and James (2001), among others. These are based on general G processes such as stick–breaking processes or normalized random measures.

Recent attention has also been paid to extending these mixture models to cover regression scenarios and the most popular approach has been to use random densities of the type

$$f(y \mid z) = \int K(y \mid x) \, \mathsf{G}_z(\mathrm{d}x),$$

where z denotes the independent variable, and it is the random distribution G_z which is allowed to change with z. There are by now many proposals for G_z based on the Dirichlet process, commonly known as dependent Dirichlet processes (DDPs), an idea first presented in MacEachern (1999). The general idea consists of defining dependent stochastic processes $\{\mathsf{G}_z\}_{z \in \mathcal{Z}}$, where $\mathcal{Z}$ denotes an index set, over which such a distribution–valued process is defined.

Most of the Bayesian literature in this direction centers on the following generalization of the Sethuraman (1994) construction

$$\mathsf{G}_z(\cdot) = \sum_{k=1}^{\infty} \omega_k(z) \, \delta_{x_k(z)}(\cdot), \tag{4}$$

where the $\{\omega_k(z)\}_{k=1,2,\ldots}$ and the $\{x_k(z)\}_{k=1,2,\ldots}$ are infinite collections of processes indexed by the z–space. Examples are provided in; De Iorio *et al.* (2004), for the analysis of variance in regression problems; Caron *et al.* (2006), Griffin and Steel (2006) and Rodriguez and Ter Horst (2008) for dynamic and time series models. Some further developments, also based on the Dirichlet process, are introduced in Dunson *et al.* (2007) and Dunson and Park (2008).

In this paper, we will contribute to the problem of Bayesian nonparametric regression based on a recent idea due to Fuentes-García *et al.* (2008). In Section 2 we review the model of Fuentes-García *et al.* (2008) and the dependent and regression version is described in Section 3. Section 4 applies the proposed approach to simulated and real data sets.

2. Nonparametric priors with decreasing weights. In Fuentes-García *et al.* (2008), a nonparametric mixture model was introduced by considering

$$f(y|A) = |A|^{-1} \sum_{l \in A} k(y; x_l), \tag{5}$$

where A is a random subset of the positive integers, $|A|$ denoting the cardinality of A, which is always by design finite. For example, the Dirichlet process arises when $A = \{k : \omega_k > u\}$ and the $\{\omega_k\}$ are precisely the weights described for the Dirichlet process in Section 1. The distribution of u, which generates the random A, is given by $p(u) = \sum_k \mathbb{I}(u < \omega_k)$; see Walker (2007). However, two key points were noted in Fuentes-García *et al.* (2008); the first is that it would be useful to have A to be always of the type $A = \{1, \ldots, N\}$, and second is that it is not necessary to have both the infinite collection of $\{\omega_k\}_{k=1,2,\ldots}$ and $\{x_k\}_{k=1,2,\ldots}$ to guarantee flexible models. The $\{x_k\}_{k=1,2,\ldots}$ are enough and the weights can take a simple form.

These considerations led to the following model

$$f(y) = \sum_{N=1}^{\infty} \frac{1}{N} \sum_{l=1}^{N} k(y; x_l) \, q(N), \tag{6}$$

where $q(\cdot)$ is an un–yet specified distribution on the positive integers. Equation (6) can be seen as a mixture of a nonparametric process, as in (3), with random probability measure defined as in (1), but with the simpler structured weights given by

$$\omega_l = \sum_{N=l}^{\infty} q(N)/N. \tag{7}$$

A convenient choice for $q(N)$ results when we take $q(N \mid \lambda) = N\lambda^2(1 - \lambda)^{N-1}$, since the resulting weights are geometric, i.e. $\omega_l = \lambda(1 - \lambda)^{l-1}$. In Fuentes-García *et al.* (2008) this model was shown to be perfectly adequate and comparable with models like the mixture of Dirichlet process for density estimation purposes.

That model can be written instructively in hierarchical form as

$$
\begin{aligned}
y_i \,|\, N_i &\overset{\text{ind}}{\sim} \frac{1}{N_i}\sum_{l=1}^{N_i} K(\,\cdot \mid x_l) \\
N_i | \lambda &\overset{\text{ind}}{\sim} q(\,\cdot \mid \lambda) \\
x_l &\overset{\text{iid}}{\sim} G_0 \\
\lambda &\sim \pi,
\end{aligned}
\tag{8}
$$

where π can be taken as beta distribution for conjugacy purposes. In this paper we demonstrate that this model can be extended to a regression situation and the key is to allow λ to depend on the covariate z. Hence the difference is that now we will have to define a process for λ on the covariate space. This is detailed in Section 3.

3. Dependent random probability measure. We define the dependent process based on the Fuentes-García *et al.* (2008) construction, as

$$\mathsf{G}_z(\cdot) = \sum_{k=1}^{\infty} \omega_k(z)\delta_{x_k}(\cdot), \tag{9}$$

where $x_k \overset{\text{iid}}{\sim} G_0$ and

$$\omega_k(z) = \sum_{N=k}^{\infty} q(N \mid z)/N, \tag{10}$$

with

$$q(N \mid \lambda(z)) = N\left[1 - \lambda(z)\right]^{N-1}\lambda(z)^2. \tag{11}$$

Hence, for each z we have that $f(y \mid z)$ is a model described in Section 2. In order to complete the dependent process we need to describe a $\lambda(z)$ process. A sensible choice is then to take

$$\lambda(z) = \frac{e^{\xi(z)}}{1 + e^{\xi(z)}} \tag{12}$$

where $\xi := \{\xi(z)\}_{z \in \mathcal{Z}}$ is a Gaussian process with continuous mean function μ and continuous covariance function σ. The logistic transformation (12) ensures that $0 < \lambda(z) < 1$, as required. Hence, we can write

$$f_\lambda(y \mid z) = \sum_k \lambda(z)(1 - \lambda(z))^{k-1} K(y \mid x_k).$$

The marginal properties of the model for each z are known and so now we look at dependencies between two models one at z and one at z'.

Proposition 1. For any two points $z, z' \in \mathcal{Z}$ and set B, it is that

$$\mathrm{Corr}(\mathsf{G}_z(B), \mathsf{G}_{z'}(B)) = \frac{\tau(z, z')}{\sqrt{\tau(z)}\sqrt{\tau(z')}} \tag{13}$$

where

$$\tau(z, z') = \mathsf{E}\left\{\frac{\lambda(z)\lambda(z')}{1 - (1 - \lambda(z))(1 - \lambda(z'))}\right\} \tag{14}$$

and

$$\tau(z) = \mathsf{E}\left\{\frac{\lambda(z)^2}{1 - (1 - \lambda(z))^2}\right\} \tag{15}$$

Proof. First, let us notice that for any set B it is that

$$\mathsf{G}_z(B) = \lambda(z)\sum_{j=1}^{\infty}(1 - \lambda(z))^{j-1}\delta_{X_j}(B).$$

We can write

$$
\begin{aligned}
\mathsf{G}_z(B)\mathsf{G}_{z'}(B) &= \lambda(z)\lambda(z')\sum_{j=1}^{\infty}\left[(1 - \lambda(z))(1 - \lambda(z'))\right]^{j-1}\delta_{X_j}(B) \\
&+ \lambda(z)\lambda(z')\sum_{j \neq k}\left[1 - \lambda(z)\right]^{j-1}\left[1 - \lambda(z')\right]^{k-1}\delta_{X_j}(B)\delta_{X_k}(B)
\end{aligned} \tag{16}
$$

and so the conditional expectation of $\mathsf{G}_z(B)\mathsf{G}_{z'}(B)$ given $\lambda(\cdot)$ is given by

$$\frac{\lambda(z)\lambda(z')}{1 - (1 - \lambda(z))(1 - \lambda(z'))}G_0(B) + \left\{1 - \frac{\lambda(z)\lambda(z')}{1 - (1 - \lambda(z))(1 - \lambda(z'))}\right\}G_0(B)^2.$$

Therefore,

$$\mathrm{Cov}(\mathsf{G}_z(B), \mathsf{G}_{z'}(B)) = G_0(B)[1 - G_0(B)]\,\mathsf{E}\left\{\frac{\lambda(z)\lambda(z')}{\lambda(z) + \lambda(z') - \lambda(z)\lambda(z')}\right\}. \tag{17}$$

Now

$$\mathrm{Var}(\mathsf{G}_z(B)) = G_0(B)[1 - G_0(B)]\,\mathsf{E}\left\{\frac{\lambda(z)}{2 - \lambda(z)}\right\} \tag{18}$$

and therefore the result follows. $\square$

In the particular case of the logistic transformation (12) the τ functions simplify to

$$\tau(z, z') = \mathrm{E}\left\{\frac{e^{\xi(z)+\xi(z')}}{e^{\xi(z)} + e^{\xi(z')} + e^{\xi(z)+\xi(z')}}\right\}$$

and

$$\tau(z) = \mathrm{E}\left\{\frac{e^{2\xi(z)}}{2e^{\xi(z)} + e^{2\xi(z)}}\right\}.$$

Given our choices for $q(N \mid \lambda)$ and the logistic Gaussian process for $\{\lambda(z)\}_{z \in \mathcal{Z}}$, we can also determine that as $z' \to z$ the convergence of $f(\cdot \mid z)$ to $f(\cdot \mid z')$ is satisfied. In fact, if d denotes the L_1 distance, and $f_\lambda(y \mid z)$ is the model with λ given, it is easy to see that

$$d(f_\lambda(\cdot \mid z), f_\lambda(\cdot \mid z')) = \sum_k \left|\lambda(z)(1 - \lambda(z))^{k-1} - \lambda(z')(1 - \lambda(z'))^{k-1}\right|,$$

which is the L_1 distance between the two geometric distributions. Hence provided that the mean and covariance functions of the Gaussian process are continuous, it follows that $\xi(\cdot)$ and $\lambda(\cdot)$ are a.s. continuous and hence $d(f_\lambda(\cdot \mid z), f_\lambda(\cdot \mid z')) \to 0$ a.s.

3.1. Bayesian nonparametric regression. As a particular application of the model defined in (9), we consider the problem of estimating the conditional density of a response variable, y, given a predictor value $z = (z_1, \ldots, z_p)'$. We assume we observe n responses, y_i, $i = 1, \ldots, n$, with corresponding predictor, or covariate, values $z_i = (z_{i1}, \ldots, z_{ip})'$.

For this purpose, it is convenient to consider the mixture modeling approach given by

$$f(y \mid z) = \sum_{N=1}^{\infty} \frac{1}{N} \sum_{l=1}^{N} K(y \mid x_l) q(N \mid z). \tag{19}$$

If $\{\mathsf{G}_z\}_{z \in \mathcal{Z}}$ is defined as in (9), this model can also be written as

$$f(y \mid z) = \int K(y \mid x) \mathsf{G}_z(\mathrm{d}x).$$

Given N_i and a predictor value z_i, let us denote by d_i a random variable indicating from which component y_i came from, hence the following hierarchical model is deduced and is convenient for understanding and exposing the Gibbs sampler for estimating the model;

$$\begin{aligned}
y_i \mid d_i, (x_1, x_2, \ldots) &\sim K(\cdot \mid x_{d_i}) \\
d_i \mid N_i &\sim \mathrm{Unif}(\{1, \ldots, N_i\}) \\
N_i \mid \lambda, z_i &\sim q(\cdot \mid \lambda(z_i)) \\
\lambda(\cdot) &\sim \mathcal{LGP}(\mu, \sigma) \\
x_l &\sim G_0
\end{aligned}$$

where $\mathcal{LGP}(\mu, \sigma)$ denotes the logistic Gaussian process defined by (12).

3.2. Gibbs sampler algorithm. The above hierarchical interpretation allows us to construct a Gibbs sampler algorithm where an iteration consists of sampling from the full conditional distributions, which are given by:

$$\pi(x_j \mid \cdots) \propto \pi(x_j) \prod_{d_i = j} K(y_i \mid x_j)$$

for $j = 1, \ldots, N_{max}$, where $N_{max} := \max_i\{N_i\}$.

$$\mathsf{P}(d_i = l \mid N_i) = \frac{K(y_i \mid x_l)}{\sum_{k=1}^{N_i} K(y_i \mid x_k)},$$

5

for $l = 1, \ldots, N_i$.

$$\mathsf{P}(N_i = N \mid d_i) \propto [1 - \lambda(z_i)]^{N-1} \mathbb{I}(N \geq d_i)$$

with $\lambda(z)$ defined as in (12). Hence, the full conditional for $\xi = (\xi(z_1), \ldots, \xi(z_n))$ is given by

$$\pi(\xi \mid \cdots) \propto \pi(\xi) \prod_{i=1}^{n} \frac{e^{2\xi(z_i)}}{(1 + e^{\xi(z_i)})^{N_i+1}}.$$

If we update component-wise, then we have that for $i = 1, \ldots, n$,

$$\pi(\xi_i \mid \xi_{-i}, \ldots) \propto \frac{1}{(1 + e^{\xi_i})^{N_i+1}} \mathrm{N}\left(\xi_i; \mu_i - \frac{1}{c_{ii}} \sum_{j \neq i}(z_j - \mu_j)c_{ij} + \frac{1}{c_{ii}}, \frac{1}{c_{ii}}\right), \qquad (20)$$

where c_{ij} is the ij-th term of the precision matrix Σ^{-1}, $\Sigma = \{\sigma(z_i, z_j); i, j = 1, \ldots, n\}$. It is worth mentioning that (20) is log-concave, so it can be easily sampled via the adaptive rejection sampling algorithm of Gilks and Wild (1992).

4. Examples. *Simulated data.* For the sake of illustration we first consider a simulated data set with 61 observations coming from

$$Y_i = 0.2\, z_i^3 + \varepsilon_i$$

where $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, 0.25)$ and $z = (-3, -2.9, \ldots, 2.9, 3)$. Using our model we fixed $K(y \mid x) = \mathrm{N}(m, 1/v)$, so $x := (m, v)$, and a conjugate prior distribution given by $g_0(x) = \mathrm{N}(m; \mu, \gamma v^{-1})\, \mathrm{Ga}(v; \alpha, \beta)$ where $\gamma, \alpha, \beta > 0$. Hence, the full conditional for the x_j is given by

$$\pi(m_j, v_j \mid \ldots) = \mathrm{N}\left(m_j \mid \frac{\gamma\, n_j\, \bar{y}_j + \mu}{\gamma\, n_j + 1}; \frac{\gamma}{v_j\,(\gamma\, n_j + 1)}\right) \mathrm{Ga}\left(v_j \mid \frac{n_j}{2} + \alpha; \frac{n_j(\bar{y}_j - \mu)^2}{2(\gamma\, n_j + 1)} + \frac{D_j}{2} + \beta\right)$$

where $n_j := \sum \mathbb{I}(d_i = j)$, $s_j := \sum_{d_i=j} y_i$, $\bar{y}_j = s_j/n_j$ and $D_j = \sum_{d_i=j}(y_i - \bar{y}_j)^2$. For the choice of ξ we selected a Gaussian process with mean function $\mu(z) = -|z|$ and stationary covariance function given by $\sigma(z_i, z_j) = e^{-\phi\|z_i - z_j\|}$.

Commonly regression analysis focuses on the predictive mean $\mathsf{E}[Y \mid z_i]$, which in this context is clearly a random quantity. In general, we can infer about the whole density $f(y \mid z)$ or the distribution of any derived functional $\eta_z(h) := \int h(y)f(y \mid z)\mathrm{d}y$. In particular, we can approximate the latter distribution through the Rao-Blackwellized Monte Carlo (MC) estimator resulting from the Gibbs sampler described above. That is

$$\eta_{z_i}(h) \approx \frac{1}{M} \sum_{l=1}^{M} E_l[h(y) \mid z_i],$$

where the expectation is taken exactly within each iteration $l = 1, \ldots, M$ of the Gibbs sampler.

For example, if we are interested in $\mu_{z_i}(y) := \mathsf{E}[y \mid z_i]$, its mean value could be approximated by

$$\bar{\eta}_{z_i}(y) \approx \frac{1}{M} \sum_{l=1}^{M} m_{d_i}^l,$$

where $m_{d_i}^l$ is the value m evaluated at d_i at iteration l.

In Figure 1 we observe the MC estimation for the distribution of $\mu_z(y)$ together with the observed data. Figure 2 shows the mean and 95% high posterior density (HPD) intervals corresponding this distribution. It is worth emphasizing that the results shown in these figures are not to be interpreted in the usual sense for credible intervals, but rather as the distribution of $\mathsf{E}[y \mid z]$. A similar MC estimation could be done, for example, for the $\mathsf{Var}[y \mid z]$. In Figure 3, a combination of these, given by $\mathsf{E}\{\mathsf{E}[y \mid z]\} \pm \mathsf{E}\{\mathsf{Var}[y \mid z]\}$ is displayed. In this example, we have chosen relatively simple forms for the mean and covariance functions corresponding to the logistic Gaussian process, however a more elaborated choice, possibly with some hyperparameters involved could also be implemented. An important issue to consider while doing this is that the logistic transformation tends to tie down large values of the Gaussian process.

Great Barrier Reef data. We now consider a data set analysed in Bowman and Azzalini (1997). It considers the weight of fauna, given by a score in the log scale, mainly collected in the Far Northern Section of the Great Barrier Reef Marine Park, at different locations. The measurements were taken in an area that had location approximately: 11-12 Degrees South Latitude, 143-144 Degrees East Longitude; see Table 1. The interest is to describe the relationship of the 42 scores with the longitude of the sampling location.

Without any further tuning to the model specifications used for the simulated data, i.e. the same normal-gamma choice for G_0 and same mean and covariance functions, we are able to produce the MC density estimators at several covariate values, shown in Figure 4. The points were chosen so that one is able to witness the evolution of the densities as the covariates change, clearly picking up the change point that occurs around $z = 143.35$.

5. Discussion. In the non-regression case we can see why the choice of simple weights work. Essentially, the infinite number of locations available do all the work by controlling how much weight is at a particular location, e.g. by how many x's are placed at a particular location.

In the regression case, this principle still applies since we are only considering data from a finite number of random densities. The x's will need to work harder but will always have the infinite number at their disposal to allocate the "correct" weights at the right places.

Having the first, and largest weight for each z to be arbitrary in $(0,1)$ combined with the infinite choice of the x's is sufficient for regression purposes.

References.

Blackwell, D. (1973). Discreteness of Ferguson selections. *Annals of Statistics*, **1**, 356–358.

Bowman, A.W. and Azzalini A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford: Clarendon Press.

Caron, F., Davy, M., Doucet, A., Duflos, E. and Vanheeghe, P. (2006). Bayesian inference for dynamic models with Dirichlet process mixtures. In *International Conference on Information Fusion.*, pp. 1–8. Florence, Italy: INRIA-CCSd-CNRS.

De Iorio, M., Müller, P., Rosner, G.L. and MacEachern, S.N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, **99**, 205–215.

Dunson, D. B., Pillai, N. and Park, J-H. (2007). Bayesian density regression. *Journal of the Royal Statistical Society*, **69**, 163–183.

Dunson, D. B. and Park, J-H. (2008). Kernel stick-breaking processes. *Biometrika*, **95**, 307–323.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.

Fuentes-García, R., Mena, R. H. and Walker, S. G. (2008) A new Bayesian nonparametric mixture model. *Submitted manuscript.*

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, **41**, 337–348.

Griffin, J. E. and Steel, M. F. J. (2006). Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, **101**, 179–194.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of The American Statistical Association*, **96**, 161-173.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I Density estimates. *Annals of Statistics*, **12**, 351–357.

MacEachern, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the section on Bayesian Statistical Science.*, pp. 50-5. Alexandria, VA: American Statistical Association.

Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Annals of Statistics*, **31**, 560-585.

Rodriguez, A. and Ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis*, **3**, 339-366.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639–650.

Walker, S.G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics*, **36**, 45–54.

y	z	y	z
0.18	142.88	1.83	143.18
2.06	142.89	1.54	143.21
0.90	142.91	1.85	143.21
1.46	142.93	1.49	143.25
1.93	142.97	1.58	143.27
1.75	142.98	-0.06	143.35
1.68	142.98	0.45	143.41
1.95	143.00	-0.14	143.42
0.86	143.02	0.27	143.52
1.45	143.02	0.12	143.56
1.44	143.02	-0.30	143.59
1.44	143.05	-0.30	143.64
1.74	143.06	-0.30	143.64
0.89	143.06	0.26	143.66
1.04	143.10	-0.30	143.71
1.68	143.10	-0.30	143.78
1.73	143.11	-0.28	143.79
1.38	143.13	-0.30	143.82
1.53	143.14	0.02	143.87
0.90	143.15	-0.30	143.89
1.36	143.18	-0.30	143.93

Table 1: Far Northern Section of the Great Barrier Reef Marine Park data.
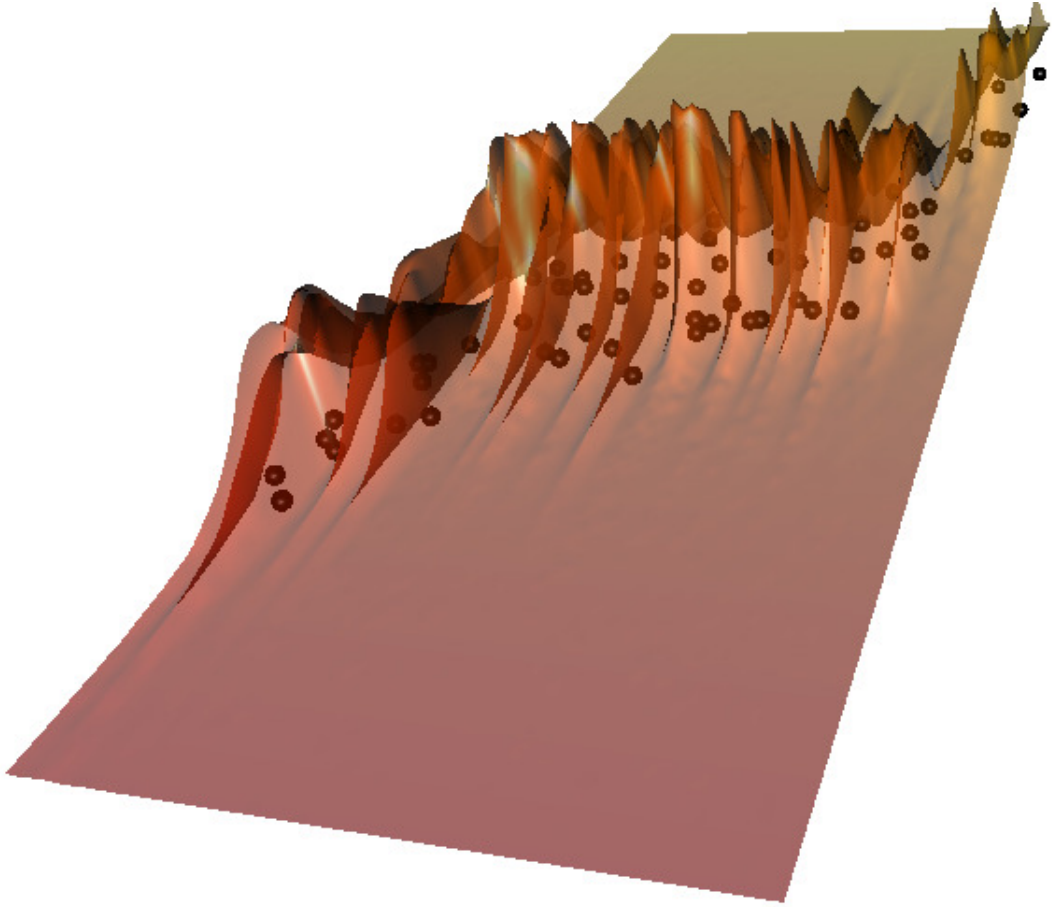
Figure 1: MC estimator for the density of $\eta_z(y)$ for simulated data set. The spheres represent the observed data and the surface the Rao-Blackwellized MC estimator for $\eta_z(y)$. The results are based on 10000 iterations of the Gibbs sampler algorithm.
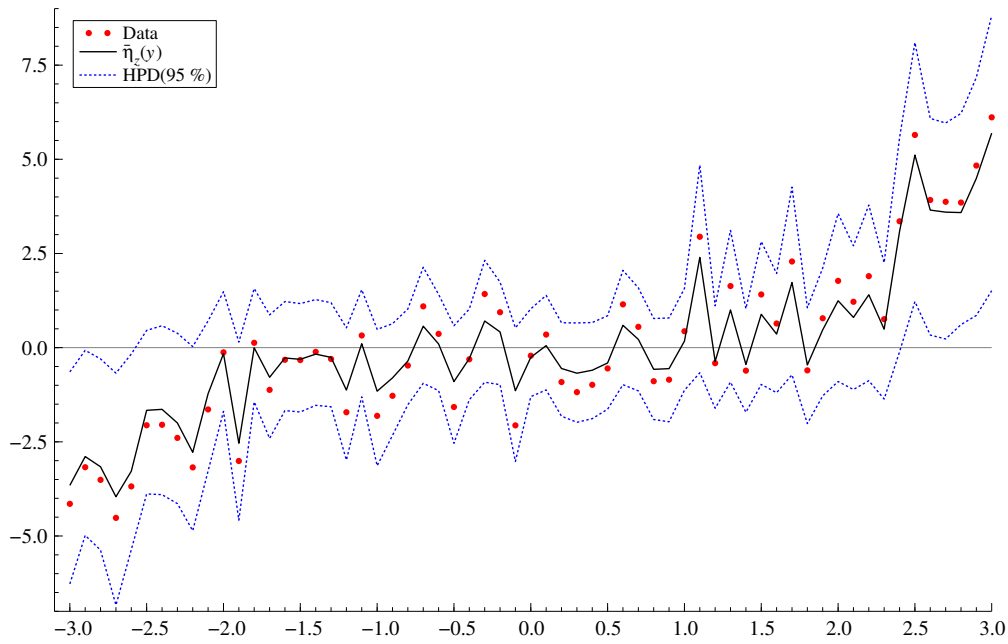
Figure 2: MC estimator for $\mathsf{E}[\eta_z(y)]$ for simulated data set. The solid line shows the mean of the random functional $\mathsf{E}[\eta_z(y)]$ and the dotted lines the corresponding HPD(95%)
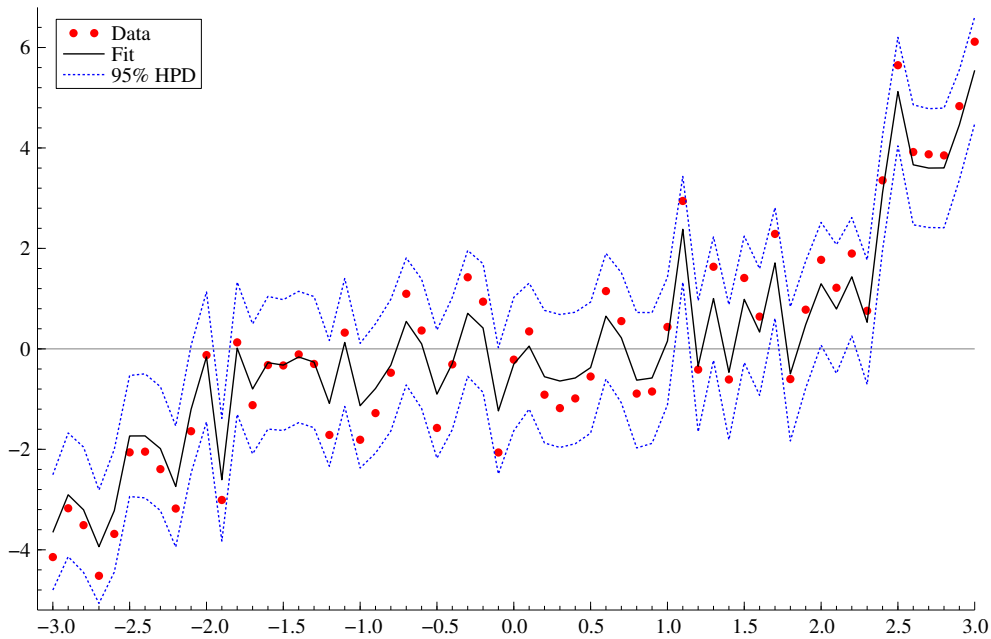

Figure 3: MC estimation for $\mathsf{E}\{\mathsf{E}[y \mid z]\} \pm \mathsf{E}\{\mathsf{Var}[y \mid z]\}$.
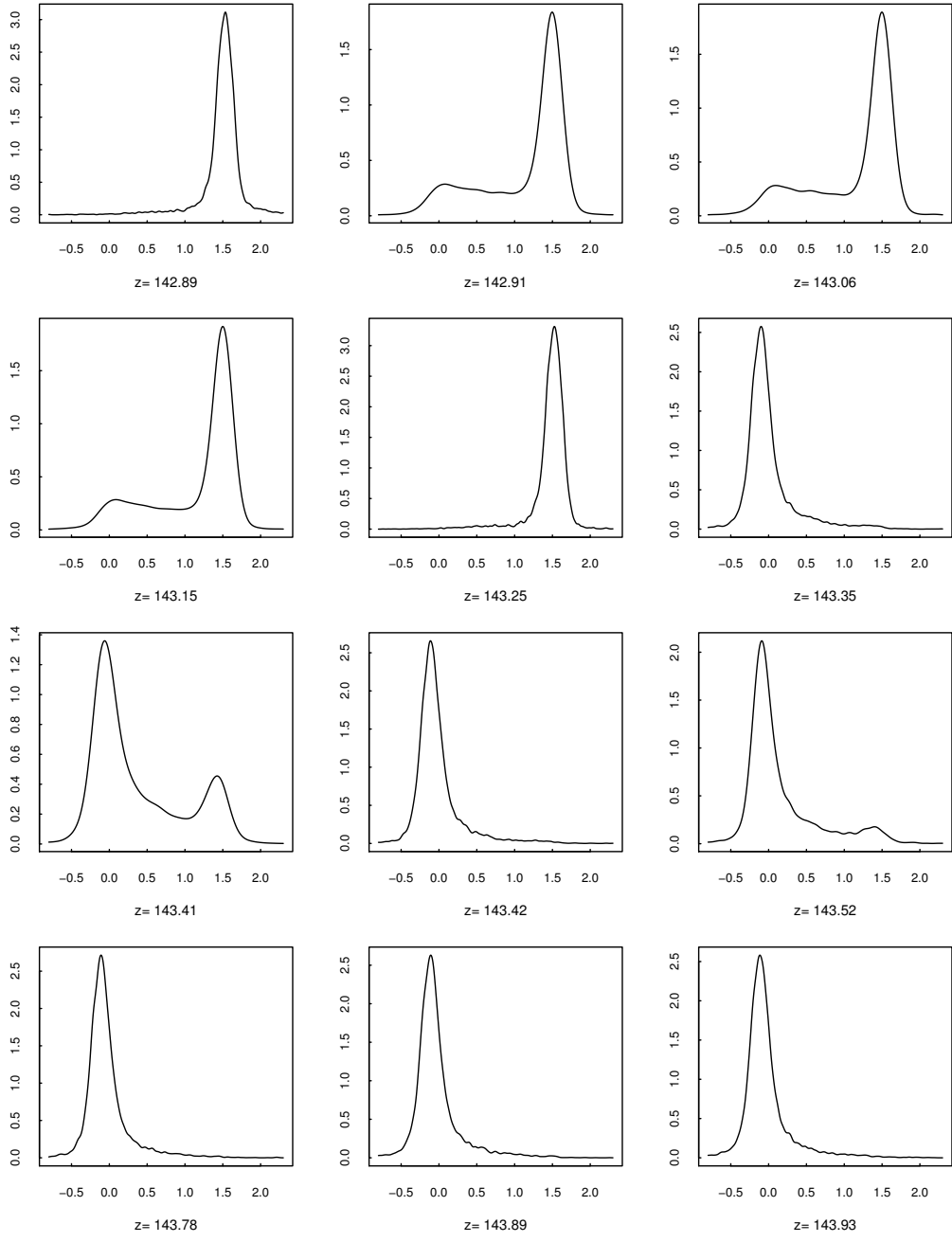
Figure 4: MC estimator for the density of $\eta_z(y)$ for Great Barrier Reef data set. The Rao-Blackwellized MC estimator for $\eta_z(y)$ is displayed for 12 different covariate values. The results are based on 10000 iterations of the Gibbs sampler algorithm.