Bayesian Nonparametric Modelling with the Dirichlet Process Regression Smoother

J.E. Griffin and M. F. J. Steel

University of Kent and University of Warwick

Supplementary Material

S1 Proofs

Proof of Proposition 1

Let $Z \sim H$. Then

$$\mathrm{E}\left[\mu_x^{(k)}\right] = \mathrm{E}\left[\sum_{i=1}^{\infty} p_i(x)\theta_i^k\right] = \mathrm{E}[Z^k]\sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)\right] = \mathrm{E}[Z^k].$$

Similarly, $\mathrm{E}\left[\mu_y^{(k)}\right] = \mathrm{E}[Z^k]$.

$$\begin{aligned}
\mathrm{E}\left[\mu_x^{(k)}\mu_y^{(k)}\right] &= \mathrm{E}\left[\left(\sum_{i=1}^{\infty} p_i(x)\theta_i^k\right)\left(\sum_{i=1}^{\infty} p_i(y)\theta_i^k\right)\right]\\
&= \sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)p_i(y)\right]\mathrm{E}\left[\theta_i^{2k}\right] + \sum_{i=1}^{\infty}\sum_{j=1;j\neq i}^{\infty}\mathrm{E}\left[p_i(x)p_j(y)\right]\mathrm{E}\left[\theta_i^k\right]\mathrm{E}\left[\theta_j^k\right]\\
&= \mathrm{E}\left[Z^{2k}\right]\sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)p_i(j)\right] + \mathrm{E}\left[Z^k\right]^2\left(1 - \sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)p_i(y)\right]\right)\\
&= \mathrm{E}\left[Z^k\right]^2 + \mathrm{Var}\left[Z^k\right]\sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)p_i(y)\right],
\end{aligned}$$

so that

$$\mathrm{Cov}\left(\mu_x^{(k)},\mu_y^{(k)}\right) = \mathrm{Var}\left[Z^k\right]\sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)p_i(y)\right],$$

and so

$$\mathrm{Corr}\left(\mu_x^{(k)},\mu_y^{(k)}\right) = \frac{\sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)p_i(y)\right]}{\sum_{i=1}^{\infty}\mathrm{E}\left[p_i(x)^2\right]}$$

which follows from the stationarity of $p_1(x), p_2(x), p_3(x), \ldots$.

Proof of Theorem 1

It is easy to show (see Griffin and Steel, 2006) that for any measurable set B

$$\text{Corr}(F_x(B), F_y(B)) = (M+1)\sum_{i=1}^{\infty} \text{E}[p_i(x)p_i(y)].$$

In this case, if $x \notin S(\phi_i)$ or $y \notin S(\phi_i)$ then

$$\text{E}[p_i(x)p_i(y)] = 0.$$

Otherwise, let $R_i^{(1)} = \{j < i | x \in S(\phi_i) \text{ and } y \in S(\phi_i)\}$, $R_i^{(2)} = \{j < i | x \in S(\phi_i) \text{ and } y \notin S(\phi_i)\}$ and $R_i^{(3)} = \{j < i | x \notin S(\phi_i) \text{ and } y \in S(\phi_i)\}$

$$\begin{aligned}
\text{E}[p_i(x)p_i(y)] &= \text{E}\left[V_i^2 \prod_{j \in R_i^{(1)}} (1-V_i)^2 \prod_{j \in R_i^{(2)}} (1-V_i) \prod_{j \in R_i^{(3)}} (1-V_i)\right] \\
&= \frac{2}{(M+1)(M+2)}\text{E}\left[\left(\frac{M}{M+2}\right)^{\#R_i^{(1)}} \left(\frac{M}{M+1}\right)^{\#R_i^{(2)}} \left(\frac{M}{M+1}\right)^{\#R_i^{(3)}}\right],
\end{aligned}$$

and so

$$\text{Corr}(F_x(B), F_y(B)) = \frac{2}{M+2}\text{E}\left[\sum_{i=1}^{\infty} B_i \left(\frac{M}{M+1}\right)^{\sum_{j=1}^{i-1} A_j} \left(\frac{M+1}{M+2}\right)^{\sum_{j=1}^{i-1} B_j}\right].$$

Proof of Theorem 2

$$\text{E}\left[\sum_{i=1}^{\infty} B_i \left(\frac{M}{M+1}\right)^{\sum_{j=1}^{i-1} A_j} \left(\frac{M+1}{M+2}\right)^{\sum_{j=1}^{i-1} B_j}\right]$$

$$= \text{E}\left[\sum_{\{i | y \in S(\phi_i) \text{ or } x \in S(\phi_i)\}} B_i \left(\frac{M}{M+1}\right)^{\sum_{j=1}^{i-1} A_j} \left(\frac{M+1}{M+2}\right)^{\sum_{j=1}^{i-1} B_j}\right].$$

The set $\{i | y \in S(\phi_i) \text{ or } x \in S(\phi_i)\}$ must have infinite size since it is contained by the set $\{i | x \in S(\phi_i)\}$ which has infinite size. Let $\phi_1', \phi_2', \phi_3', \ldots$ be the subset of ϕ_1, ϕ_2, ϕ_3 for which

$\{i|y \in S(\phi_i) \text{ or } x \in S(\phi_i)\}$ and define $B'_i = \mathrm{I}\left(y \in S(\phi'_i) \text{ and } x \in S(\phi'_i)\right)$ then

$$\begin{aligned}
\mathrm{Corr}(F_s, F_v) &= \frac{2}{M+2}\mathrm{E}\left[\sum_{i=1}^{\infty} B'_i \left(\frac{M}{M+1}\right)^i \left(\frac{M+1}{M+2}\right)^{\sum_{j=1}^{i-1} B'_j}\right] \\
&= \frac{2}{M+2}\sum_{i=1}^{\infty}\left(\frac{M}{M+1}\right)^i \mathrm{E}\left[B'_i\right]\prod_{j=1}^{i-1}\mathrm{E}\left[\left(\frac{M+1}{M+2}\right)^{B'_j}\right] \\
&= \frac{2}{M+2}\sum_{i=1}^{\infty}\left(\frac{M}{M+1}\right)^i p_{s,v}\left[\left(\frac{M+1}{M+2}\right)p_{s,v} + (1-p_{s,v})\right]^{i-1} \\
&= \frac{2}{M+2}\left(\frac{M}{M+1}\right)\sum_{i=0}^{\infty}\left(\frac{M}{M+1}\right)^i p_{s,v}\left[\left(\frac{M+1}{M+2}\right)p_{s,v} + (1-p_{s,v})\right]^{i} \\
&= \frac{2}{M+2}\left(\frac{M}{M+1}\right)p_{s,v}\frac{1}{1-\left[\left(\frac{M}{M+2}\right)p_{s,v} + \frac{M}{M+1}(1-p_{s,v})\right]} \\
&= \frac{2\frac{M}{M+2}p_{s,v}}{1+\frac{M}{M+2}p_{s,v}}.
\end{aligned}$$

Proof of Theorem 3

Since (C, r, t) follows a Poisson process on $\mathbb{R}^p \times \mathbb{R}^2_+$ with intensity $f(r)$, $p(C_k|s \in S(\phi_k) \text{ or } v \in S(\phi_k), r_k)$ is uniformly distributed on $B_{r_k}(s) \cup B_{r_k}(v)$ and $p(r_k|s \in S(\phi_k) \text{ or } v \in S(\phi_k)) = \frac{\nu\left(B_{r_k}(s)\cup B_{r_k}(v)\right)f(r_k)}{\int \nu\left(B_{r_k}(s)\cup B_{r_k}(v)\right)f(r_k)\,dr_k}$ where $\nu(\cdot)$ is Lebesgue measure. Then

$$\begin{aligned}
p_{s,v} &= P(s, v \in S(\phi_k)|s \in S_k \text{ or } v \in S(\phi_k)) \\
&= \int\int_{B_{r_k}(s)\cap B_{r_k}(v)} p\left(C_k, r_k|s \in S(\phi_k) \text{ or } v \in S(\phi_k)\right)\,dC_k\,dr_k \\
&= \frac{\int \nu\left(B_{r_k}(s)\cap B_{r_k}(v)\right)f(r_k)\,dr_k}{\int \nu\left(B_{r_k}(s)\cup B_{r_k}(v)\right)f(r_k)\,dr_k}.
\end{aligned}$$

Proof of Theorem 4

The autocorrelation function can be expressed as $f(p_{s,s+u})$ where $f(x) = 2(\frac{M+1}{M+2})/(1+\frac{M}{M+2}x)$. Then by Faá di Bruno's formula

$$\frac{d^n}{du^n}f(p_{s,s+u}) = \sum \frac{n!}{m_1!m_2!m_3!\dots}\frac{d^{m_1+\dots+m_n}f}{dp_{s,s+u}^{m_1+\dots+m_n}}\prod_{\{j|m_j\neq 0\}}\left(\frac{d^j p_{s,s+u}}{du^j}\frac{1}{j!}\right)^{m_j},$$

where $m_1 + 2m_2 + 3m_3 + \dots + nm_n = n$ with $m_j \geq 0, j = 1, \dots, n$, and so

$$\lim_{u\to 0}\frac{d^n}{du^n}f(p_{s,s+u}) = \sum \frac{n!}{m_1!m_2!m_3!\dots}\lim_{u\to 0}\frac{d^{m_1+\dots+m_n}f}{dp_{s,s+u}^{m_1+\dots+m_n}}\prod_{\{j|m_j\neq 0\}}\left(\frac{d^j p_{s,s+u}}{du^j}\frac{1}{j!}\right)^{m_j}.$$

Since $\lim_{u\to 0}\frac{d^{m_1+\cdots+m_n}f}{dp_{s,s+u}^{m_1+\cdots+m_n}} = \lim_{p_{s,s+u}\to 1}\frac{d^{m_1+\cdots+m_n}f}{dp_{s,s+u}^{m_1+\cdots+m_n}}$ is finite and non-zero for all values of n, the degree of differentiability of the autocorrelation function is equal to the degree of differentiability of $p_{s,s+u}$. We can write $p_{s,s+u} = \left(\frac{4\mu}{a}-1\right)^{-1}$ with $a = 2\mu_2 - uI$. Now $\frac{d^k p_{s,s+u}}{a^k} = (k-1)!(4\mu-a)^{-k}$ and $\lim_{u\to 0}\frac{d^k p_{s,s+u}}{da^k} = (k-1)!(2\mu)^{-k}$ which is finite and non-zero. By application of Faá di Bruno's formula

$$\frac{d^n}{du^n}p_{s,s+u} = \sum \frac{n!}{m_1!m_2!m_3!\ldots}\frac{d^{m_1+\cdots+m_n}p_{s,s+u}}{da^{m_1+\cdots+m_n}}\prod_{\{j|m_j\neq 0\}}\left(\frac{d^j a}{du^j}\frac{1}{j!}\right)^{m_j}$$

and the degree of differentiability is determined by the degree of differentiability of a. If $p(r)\sim$ Ga(α,β) then $\frac{d\mu_2}{du} = -\frac{1}{2}\left(\frac{u}{2}\right)^{\alpha}\exp\{-u/2\}$ and $\frac{dI}{du} = -\frac{1}{2}\left(\frac{u}{2}\right)^{\alpha-1}\exp\{-u/2\}$ and it is easy to show that $\frac{d^n a}{du^n} = C_n u^{\alpha-n+1}\exp\{-u/2\}+\zeta$ where ζ contains terms with power of x greater than $\alpha-n+1$. If $\lim_{u\to 0}u^{\alpha}\exp\{-u/2\}$ is finite then so is $\lim_{u\to 0}u^{\alpha+k}\exp\{u/2\}$ for $k>0$ and so the limit will be finite iff $\alpha-n+1\geq 0$, *i.e.* $\alpha\geq n-1$.

S2 Computational Details

As we conduct inference on the basis of the Poisson process restricted to the set R, all quantities (C,r,t,V,θ) should have a superscript R. To keep notation manageable, these superscripts are not explicitly used in this Supplement.

Updating the centres

We update each centre $C_1,\ldots,C_K$ from its full conditional distribution Metropolis-Hastings random walk step. A new value C_i' for the i-th centre is proposed from N(C_i,σ_C^2) where σ_C^2 is chosen so that the acceptance rate is approximately 0.25. If there is no x_i such that $x_i\in(C_i'-r_i,C_i'+r_i)$ or if there is one value of j such that $s_j=i$ for which $x_i\notin(C_i'-r_i,C_i'+r_i)$ then $\alpha(C_i,C_i')=0$. Otherwise, the acceptance probability has the form

$$\alpha(C_i,C_i') = \frac{\prod_{j=1}^n\prod_{h<s_j \text{ and } C_h'-r_h<x_j<C_h'+r_h}(1-V_h)}{\prod_{j=1}^n\prod_{h<s_j \text{ and } C_h-r_h<x_j<C_h+r_h}(1-V_h)}.$$

Updating the distances

The distances can be updated using a Gibbs step since the full conditional distribution of r_k has a simple piecewise form. Recall that $d_{ik}=|x_i-C_k|$ and let $\mathcal{S}_k=\{j|s_j\geq k\}$. We define $\mathcal{S}_k^{ord}$ to be a version of $\mathcal{S}_k$ where the element have been ordered to be increasing in d_{ik}, *i.e.* if $i>j$ and $i,j\in\mathcal{S}_k^{ord}$ then $d_{ik}>d_{jk}$. Finally we define $d_k^\star=\max[\{x_{\min}-C_k,C_k-x_{\max}\}\cup\{d_{ik}|s_i=k\}]$ and $m^\star$ be such that $x_i\in S_k^{ord}$ and $x_{m^\star}>d_k^\star$ and $x_{m^\star-1}<d_k^\star$. Let l be the length of S_k^{ord}.

The full conditional distribution has density

$$f^\star(z) \propto \begin{cases} f(z) & \text{if } d_k^\star < z \le d_{S^{ord}_{m^\star}k} \\ f(z)(1-V_k)^{i-m^\star+1} & \text{if } d_{S^{ord}_i k} < z \le d_{S^{ord}_{i+1}k}, \quad i = m^\star, \ldots, l-1 \\ f(z)(1-V_k)^{l-m^\star+1} & \text{if } z > d_{S^{ord}_l k} \end{cases}.$$

Swapping the positions of atoms

The ordering of the atoms should also be updated in the sampler. One of the K included atoms, say $(V_i, \theta_i, C_i, r_i)$, is chosen at random to be swapped with the subsequent atom $(V_{i+1}, \theta_{i+1}, C_{i+1}, r_{i+1})$. If $i < K$, the acceptance probability of this move is $\min\{1, (1-V_{i+1})^{n_i}/(1-V_i)^{n_{i+1}}\}$. If $i = K$, then a new point $(V_{K+1}, \theta_{K+1}, C_{K+1}, r_{K+1})$ is proposed from their prior and the swap is accepted with probability $\min\{1, (1-V_{K+1})^{n_i}\}$.

Updating θ and V

The full conditional distribution of θ_i is proportional to $h(\theta_i)\prod_{\{j|s_i\}} k(y_j|\theta_i)$, where h is the density function of H. We update V_i from a Beta distribution with parameters $1+\sum_{j=1}^n \mathrm{I}(s_j = i)$ and $M + \sum_{j=1}^n \mathrm{I}(s_j > i, |x_j - C_i| < r_i)$.

Updating M

This parameter can be updated by a random walk on the log scale. Propose $M' = M\exp(\epsilon)$ where $\epsilon \sim \mathrm{N}(0, \sigma_M^2)$ with σ_M^2 a tuning parameter chosen to maintain an acceptance rate close to 0.25. The proposed value should be accepted with probability

$$\frac{M'^{K+1}\left[\prod_{i=1}^K(1-V_i)\right]^{M'} \beta(M')^{\alpha K}\exp\left\{-\beta(M')\sum_{i=1}^K r_i\right\} p(M')}{M^{K+1}\left[\prod_{i=1}^K(1-V_i)\right]^{M} \beta(M)^{\alpha K}\exp\left\{-\beta(M)\sum_{i=1}^K r_i\right\} p(M)},$$

where $\beta(M)$ is β expressed as a function of M, as in our suggested form

$$\beta = \frac{2}{x^\star}\log\left(\frac{1+M+\varepsilon}{\varepsilon(M+2)}\right).$$

Posterior inferences on $F_{\tilde{x}}$

We are often interested in inference at some point $\tilde{x} \in \mathcal{X}$ about the distribution $F_{\tilde{x}}$. We define $(\tilde{V}_1, \tilde{\theta}_1), (\tilde{V}_2, \tilde{\theta}_2), \ldots, (\tilde{V}_J, \tilde{\theta}_J)$ to be the subset of $(V_1, \theta_1), (V_1, \theta_2) \ldots, (V_K, \theta_K)$ for which $|\tilde{x}-$

$C_i| < r_i$. Then

$$F_{\tilde{x}} = \sum_{i=1}^{J} \delta_{\tilde{\theta}_i} \tilde{V}_i \prod_{j<i}(1-\tilde{V}_j) \prod_{j\leq i}\prod_{l=1}^{n_j}\left(1-V_l^{(j)}\right) + \prod_{i\leq J}\prod_{j=1}^{n_i}\left(1-V_j^{(i)}\right) \sum_{l=J+1}^{\infty} \delta_{\tilde{\theta}_l}\tilde{V}_l \prod_{m<l}(1-\tilde{V}_m)$$
$$+ \sum_{i=1}^{N}\sum_{j=1}^{n_i} \delta_{\theta_j^{(i)}} V_j^{(i)} \prod_{l<j}\left(1-V_l^{(i)}\right) \prod_{l<i}(1-V_l) \prod_{m=1}^{n_l}\left(1-V_m^{(l)}\right)$$

where n_j is a geometric random variable with success probability $1-\tilde{p}$, $\theta_j^{(i)} \sim H$, $V_j^{(i)} \sim \text{Be}(1, M)$, $\tilde{\theta}_m \sim H$ and $\tilde{V}_m \sim \text{Be}(1, M)$ for $m > N$. We calculate $\tilde{p}$ in the following way. If $x_{min} < \tilde{x} < x_{max}$, define i so that $x_{(i)} < \tilde{x} < x_{(i+1)}$, where $x_{(1)}, \ldots, x_{(n)}$ is an ordered version of $x_1, \ldots, x_n$, then $\tilde{p} = \frac{\beta}{2\alpha}\tilde{q}$ where

$$\tilde{q} =(x_{(i+1)} - x_{(i)})\mathcal{I}\left(\frac{x_{(i+1)} - x_{(i)}}{2}\right) + (x_{(i)} - \tilde{x})\mathcal{I}\left(\frac{\tilde{x} - x_{(i)}}{2}\right) - (x_{(i+1)} - \tilde{x})\mathcal{I}\left(\frac{x_{(i+1)} - \tilde{x}}{2}\right)$$
$$- 2\mu^\star\left(\frac{x_{(i+1)} - x_{(i)}}{2}\right) + 2\mu^\star\left(\frac{\tilde{x} - x_{(i)}}{2}\right) + 2\mu^\star\left(\frac{x_{(i+1)} - \tilde{x}}{2}\right)$$

with $\mathcal{I}(y) = \int_0^y f(r)\,dr$ and $\mu^\star(y) = \int_0^y rf(r)\,dr$. Otherwise if $\tilde{x} < x_{min}$

$$\tilde{q} = 2\mu^\star\left(\frac{x_{min} - \tilde{x}}{2}\right) + (x_{min} - \tilde{x})\left(1 - \mathcal{I}\left(\frac{x_{min} - \tilde{x}}{2}\right)\right)$$

and if $\tilde{x} > x_{max}$

$$\tilde{q} = 2\mu^\star\left(\frac{\tilde{x} - x_{max}}{2}\right) + (\tilde{x} - x_{max})\left(1 - \mathcal{I}\left(\frac{\tilde{x} - x_{max}}{2}\right)\right).$$

We use a truncated version of $F_{\tilde{x}}$ with h elements which are chosen so that $\sum_{i=1}^{h} p_i = 1 - \epsilon$ where ϵ is usually taken to be 0.001.

Model 2

This section is restricted to discussing the implementation when $m(x)$ follows a Gaussian process prior where we define $P_{ij} = \rho(x_i, x_j)$. We also reparametrise from u_i to $\phi_i = \sigma^2\psi_i$.

Updating $\psi_i|s$

The full conditional distribution has the density

$$p(\phi_i) \propto \phi_i^{0.5(1-\sum \mathrm{I}(s_j=i, 1\leq j\leq n))} \exp\{-0.5\phi_i/\sigma^2\}, \qquad \phi_i > \phi_{min}$$

where $\phi_{min} = \max\left\{(y_i - m(x_i))^2 \,|s_j = i, 1 \leq j \leq n\right\}$. A rejection sampler for this full conditional distribution can be constructed using the envelope

$$h^\star(\phi_i) \propto \begin{cases} \phi_i^{0.5(1-\sum \mathrm{I}(s_j=i,1\leq j\leq n))} & \phi_{min} < \phi_i < z \\ z^{0.5(1-\sum \mathrm{I}(s_j=i,1\leq j\leq n))} \exp\{-0.5(\phi_i - z)/\sigma^2\} & \phi_i > z \end{cases}$$

which can be sampled using inversion sampling. The acceptance probability is

$$\alpha(\phi_i) = \begin{cases} \exp\{-0.5(\phi_i - \phi_{min})/\sigma^2\} & \phi_{min} < \phi_i < z \\ \left(\frac{\phi_i}{z}\right)^{0.5-0.5k} \exp\{-0.5(z - \phi_{min})/\sigma^2\} & \phi_i > z \end{cases}$$

and the choice $z = \sigma^2 \sum \mathrm{I}(s_j = i, 1 \leq j \leq n)$ maximizes the acceptance rate.

Updating σ^{-2}

Using the prior Gamma(ν_1, ν_2), the full conditional distribution of σ^{-2} is again a Gamma distribution, where we define $P = (P_{ij})$

$$\sigma^{-2} \sim \text{Ga}\left(\nu_1 + \frac{3K}{2} + \frac{n}{2}, \nu_2 + \frac{1}{2}\sum_{i=1}^{K} \phi_i + \frac{1}{2\omega} m(x)^T P^{-1} m(x)\right).$$

Updating $m(x_1), \ldots, m(x_n)$

It is possible to update $m(x_i)$ using its full conditional distribution. However this tends to lead to slowly mixing algorithms. A more useful approach uses the transformation $m(x) = C^\star z$ where $C^\star$ is the Cholesky factor of $\sigma_0^{-2} P^{-1}$, where $z \sim \mathrm{N}(0, I)$. We then update z_j using their full conditional distribution which is a standard normal distribution truncated to the region $\cap_{i=1}^{n}(y_i - \sum_{k \neq j} C_{ik} z_k - \sqrt{\phi_i}, y_i - \sum_{k \neq j} C_{ik} z_k + \sqrt{\phi_i})$.

Updating ω

We define $\omega^2 = \sigma^2/\sigma_0^2$. If ω^2 follows a Gamma distribution with parameters a_0 and b_0 then the full conditional of σ_0^{-2} follows a Gamma distribution with parameters $a_0 + n/2$ and $b_0 + \sigma^{-2} m(x)^T P^{-1} m(x)/2$. A similar updating occurs for the Generalized inverse Gaussian prior used here.

Updating the Matèrn parameters

We update any parameters of the Matèrn correlation structure by a Metropolis-Hastings random walk. The full conditional distribution of the parameters (ζ, τ) would be proportional to

$$|P|^{-1/2} \exp\left\{-\sigma^{-2}\omega^{-2} m(x)^T P^{-1} m(x)\right\} p(\zeta, \tau).$$