Hybrid Dirichlet mixture models for functional data

Sonia Petrone, Michele Guindani, and Alan E. Gelfand
Bocconi University, Italy, University of New Mexico, USA, Duke University, USA

Appendix

A.1. Proofs of the results in Section 3.3

To prove Proposition 1, we use the following lemma, that gives a general representation result by means of DP mixtures.

Lemma 1. *Let G be a r.p.m. on a measurable space Θ and q be a probability measure on $\{1, 2, \ldots\}$ with mass q_j on j.*

(i) Suppose that $G \sim DP(\alpha Q_0)$, where Q_0 is a discrete probability measure on Θ, $Q_0 = \sum_{j=1}^{\infty} q_j \delta_{\xi_j}$. Then a.s.

$$G = \sum_{j=1}^{\infty} p_j \delta_{\xi_j},$$

where the weights $(p_1, p_2, \ldots)$ define a r.p.m. p on $\{1, 2, \ldots\}$ having mass p_j on j, with $p \sim DP(\alpha q)$.

(ii) Suppose that $G \sim \int DP(\alpha Q) d\mu(Q)$, μ being the probability law of the r.p.m. $Q = \sum_{j=1}^{\infty} q_j \delta_{\theta_j^*}$, where $\theta_1^*, \theta_2^*, \ldots$ are random quantities with values in Θ. Then a.s.

$$G = \sum_{j=1}^{\infty} p_j \delta_{\theta_j^*},$$

where the weights $(p_1, p_2, \ldots)$ define a r.p.m. p on $\{1, 2, \ldots\}$ with $p \sim DP(\alpha q)$.

Proof. (i) If $G \sim DP(\alpha Q_0)$, then a.s. $G = \sum_{j=1}^{\infty} w_j \delta_{\theta_j^*}$, where the w_j's have a stick-breaking prior and the θ_j^* are i.i.d. according to Q_0. Since Q_0 is discrete,

$$G = \sum_{j=1}^{\infty} \Big(\sum_{i: \theta_i^* = \xi_j} w_j \Big) \delta_{\xi_j} = \sum_{j=1}^{\infty} p_j \delta_{\xi_j}.$$

For any measurable partition $(A_1, \ldots, A_c)$ of $\{1, 2, \ldots\}$, we can find a partition $(B_1, \ldots, B_c)$ of Θ such that B_i contains the uniques ξ_j with $j \in A_i$, $i = 1, \ldots, c$. Then, $p(A_i) = \sum_{j \in A_i} p_j = G(B_i)$, for any $i = 1, \ldots, c$. Thus, $(p(A_1), \ldots, p(A_c))$ has the same probability law as $(G(B_1), \ldots, G(B_c))$, i.e. a Dirichlet distribution $Dir(\alpha G_0(B_1), \ldots, \alpha G_0(B_c)) = Dir(\alpha q(A_1), \ldots, \alpha q(A_c))$, where $q(A_i) = \sum_{j \in A_i} q_j$. Thus $p \sim DP(\alpha q)$.

(ii) Conditionally on Q, the finite sum representation of G is obtained from (i) and the result follows. $\square$

As a special case, if q has support $\{1, \ldots, k\}$, $k < \infty$ and $\theta_1^*, \ldots, \theta_k^*$ are i.i.d. according to a probability measure G_0, then $G \sim \int DP(\alpha Q) d\mu(Q)$ implies $G = \sum_{j=1}^k p_j \delta_{\theta_j^*}$ a.s., with $(p_1, \ldots, p_k) \sim Dir(\alpha q_1, \ldots, \alpha q_k)$; that is, $G \sim DP_k((\alpha q_1, \ldots, \alpha q_k), G_0)$. In that sense, we say that the DP_k can be represented as a mixture of Dirichlet processes $\int DP(\alpha Q) d\mu(Q)$.

Proof of Proposition 1
(i) If $G \sim \int DP(\alpha Q) d\mu(Q)$, by Lemma **??** we have, a.s.,

$$G = \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} p(j_1, \ldots, j_m) \delta_{\theta^*_{1,j_1}, \ldots, \theta^*_{m,j_m}},$$

with $p \sim DP(\alpha q)$. Thus, G has the same probability law of a r.p.m. with a $hDP(\alpha q, G_0)$ prior.
Point (ii) follows from (i) by considering a probability measure q with support $\{1, \ldots, k\}^m$. $\diamond$

An analogous result can be proved for the functional hDP. Let $\mathbf{G}_0$ be a probability measure on $\mathbb{R}^D$, $\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots \overset{i.i.d.}{\sim} \mathbf{G}_0$ and $\mathbf{q}$ a probability measure on $\{1, 2, \ldots\}^D$. Let $\mathbf{Q}$ be the r.p.m. on $\mathbb{R}^D$ characterized by the family of finite-dimensional distributions

$$Q_{x_1, \ldots, x_m} = \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} \mathbf{q}_{x_1, \ldots, x_m}(j_1, \ldots, j_m) \delta_{\theta^*_{j_1}(x_1), \ldots, \theta^*_{j_m}(x_m)}, \tag{1}$$

for all $(x_1, \ldots, x_m)$, and let $\mathbf{Q}_k$ be defined analogously, replacing $\mathbf{q}$ with a probability measure $\mathbf{q}_k$ with support $\{1, \ldots, k\}^D$, $k < \infty$. Denote by μ and μ_k the probability laws of $\mathbf{Q}$ and $\mathbf{Q}_k$, respectively.

Proposition 3. *Let $\mathbf{G}$ be a random probability law on $\mathbb{R}^D$ and $\mathbf{Q}$ and $\mathbf{Q}_k$ be defined as above. Then,*

(i) a $hDP(\alpha \mathbf{q}, \mathbf{G}_0)$ prior for $\mathbf{G}$ can be represented as the mixture of functional Dirichlet processes $\int fDP(\alpha \mathbf{Q}) d\mu(\mathbf{Q})$;

(ii) a $hfPD_k(\alpha_k \mathbf{q}_k, \mathbf{G}_0)$ prior for $\mathbf{G}$ can be represented as the mixture of functional Dirichlet processes $\int fDP(\alpha_k \mathbf{Q}_k) d\mu_k(\mathbf{Q}_k)$.

Proof. (i) If $\mathbf{Q} \sim \int fDP(\alpha \mathbf{Q}) d\mu(\mathbf{Q})$, then the finite-dimensional distributions $Q_{x_1, \ldots, x_m}$ of $\mathbf{Q}$ have probability law $\int DP(\alpha Q_{x_1, \ldots, x_m}) d\mu(\mathbf{Q})$. Therefore, by Proposition 1, the prior process $\int fDP(\alpha \mathbf{Q}) d\mu(\mathbf{Q})$ a.s. selects probability measures on $\mathbb{R}^D$ characterized by the (consistent) family of finite-dimensional distributions

$$Q_{x_1, \ldots, x_m} = \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} p_{x_1, \ldots, x_m}(j_1, \ldots, j_m) \delta_{\theta^*_{j_1}(x_1), \ldots, \theta^*_{j_m}(x_m)},$$

where $\mathbf{p}_{x_1,\ldots,x_m} \sim DP(\alpha \mathbf{q}_{x_1,\ldots,x_m})$, and $p_{x_1,\ldots,x_m}$ are the finite-dimensional distributions of a r.p.m. $\mathbf{p} \sim DP(\alpha\mathbf{q})$. Therefore, we have that $\mathbf{Q}$ has the same probability law of a r.p.m. with distribution $hfDP(\alpha\mathbf{q}, \mathbf{G}_0)$.

Point (ii) follows from (i) by considering a probability measure $\mathbf{q}$ with support on $\{1, \ldots, k\}^D$.

□

To prove Proposition 2 we use the following lemma.

Lemma 2. *Let $G_k, k \geq 1$ and G be r.p.m.'s on Θ, with $G_k \sim \int DP(\alpha_k H) d\mu_k(H)$ and $G \sim \int DP(\alpha H) d\mu(H)$. Let $H_k \sim \mu_k$ and $E(H_k(\cdot)) = H_{0,k}$, such that the family of distributions $(H_{0,k}, k \geq 1)$ is tight. If $\alpha_k \to \alpha$, $0 < \alpha < \infty$ and μ_k converges weakly to μ as $k \to \infty$, then the sequence G_k converges to G in distribution.*

Proof. Let $\pi_k = \int DP(\alpha_k H) d\mu_k(H)$ and $\pi = \int DP(\alpha H) d\mu(H)$. For any partition $(B_1, \ldots, B_c)$ of Θ, we have

$$\pi_k(G(B_1) \leq t_1, \ldots, G(B_c) \leq t_c) = \int Dir(t_1, \ldots, t_c; \alpha_k H(B_1), \ldots, \alpha_k H(B_c)) d\mu_k(H),$$

where $Dir(\cdot, \ldots, \cdot; a_1, \ldots, a_c)$ denotes the Dirichlet d.f. with parameters $(a_1, \ldots, a_c)$. The Dirichlet d.f. is bounded, and is continuous in its parameters (being degenerate on the appropriate subspaces if some parameters go to zero). Thus, if $\alpha_k \to \alpha$ and μ_k converges weakly to μ for $k \to \infty$, then

$$\pi_k(G(B_1) \leq t_1, \ldots, G(B_c) \leq t_c) \to \int Dir(t_1, \ldots, t_c; \alpha H(B_1), \ldots, \alpha H(B_c)) d\mu(H),$$

that are the finite-dimensional laws of π. Since

$$E(G_k(x)) = E(G_k((x) \mid H) = E_{\mu_k}(H(x)) = H_{0,k}(x),$$

the family $\{\pi_k, k \geq 1\}$ is tight, by Theorem 2.5.1 in Ghosh and Ramamoorthi (2003). Therefore, $\pi_k \to \pi$ weakly, i.e. $G_k \to G$ in distribution.

□

Proof of Proposition 2

Let $\pi_k = hDP_k(\alpha_k q_k, G_0)$, for $k \geq 1$. Proposition 1 shows that π_k can be represented as $\int DP(\alpha_k Q_k) d\mu_k$. Therefore the results follow from Lemma 2 if we show that $\{\bar{G}_k(\cdot) \equiv E(G_k(\cdot)), k \geq 1\}$ is tight.

Let $G_{0,i_1,\ldots,i_c}, i_j = 1, \ldots, m$ be the marginal distribution of G_0 at coordinates $(i_1, \ldots, i_c)$. For any subset $A = (A_1 \times A_2 \times A_m)$ of $\mathbb{R}^m$, we have

$$\begin{aligned} E(G_k(A)) &= \sum_j q_k(j, \ldots, j) G_0(A) + \sum_{j_1} \sum_{j \neq j_1} q_k(j_1, j, \ldots, j) G_{0,1}(A_1) G_{0,2,\ldots,m}(A_2 \times \cdots \times A_m) \\ &+ \cdots + \sum_{j_1 \neq j_2 \neq \cdots \neq j_m} q_k(j_1, \ldots, j_m) G_{0,1}(A_1) \cdots G_{0,m}(A_m). \end{aligned}$$

For any $\epsilon > 0$, we can find a compact set $M = (M_1 \times M_2 \times M_m)$ such that M_i is compact and $G_{0,i}(M_i^c) < \epsilon/m$. This implies that, for any $k \geq 1$,

$$\bar{G}_k(M^c) = E(G_k(M^c)) = E(G_k(M_1^c \cup \cdots \cup M_m^c)) \leq E(\sum_{i=1}^m G_{k,i}(M_i^c)) = \sum_{i=1}^m G_{0,i}(M_i^c) \leq \epsilon.$$

$\diamond$

Proof of Theorem 1

Result (i) is known, see Ishwaran and Zarepour (2002), Theorem 3.

Let us prove case (ii). It is useful to write G_k as $G_k = \sum_{j=1}^{\infty} p_k(j)\delta_{\theta_j^*}$, where $(p_k(1), p_k(2), \ldots)$ define a random probability measure p_k on $\{1, 2, \ldots\}$ with support $\{1, \ldots, k\}$. By assumption, $(p_k(1), \ldots, p_k(k)) \sim \mathcal{D}(\alpha q_k(1), \ldots, \alpha q_k(k))$, and is easy to see that this is equivalent to $p_k \sim DP(\alpha q_k)$, where q_k is also regarded as a probability measure on $\{1, 2, \ldots\}$ with support $\{1, \ldots, k\}$.

Since, $q_k(j) \to q(j)$ for any j, by assumption, and q is a probability measure, from theorem 3.2.6 in Ghosh and Ramamoorthi (2003) it follows that $p_k \to p$, with $p \sim DP(\alpha q)$. Therefore, by Slutsky's theorem, $G_k = \sum_{j=1}^{\infty} p_k(j)\delta_{\theta_j^*} \to G = \sum_{j=1}^{\infty} p(j)\delta_{\theta_j^*}$, which completes the proof. $\diamond$

Proof of Theorem 2

We illustrate the proof of (i) for $m = 3$. Then

$$G_k = \sum_{j_1=1}^{k} \cdots \sum_{j_3=1}^{k} p(j_1, j_2, j_3)\ \delta_{\theta^*_{1,j_1}, \theta^*_{2,j_2}, \theta^*_{1,j_3}},$$

with weights $p \sim \mathcal{D}(\alpha q_k)$. Let

$$c_{k,1} = \sum_{j=1,2,\ldots} q_k(j,j,j), \qquad c_{k,2} = \sum_{i,j=1,2,\ldots;i\neq j} q_k(i,j,j), \quad c_{k,3} = \sum_{i,j=1,2,\ldots;i\neq j} q_k(j,i,j)$$
$$c_{k,4} = \sum_{i,j=1,2,\ldots;i\neq j} q_k(j,j,i), \qquad c_{k,5} = \sum_{i,j=1,2,\ldots;i\neq j} q_k(i,j,l).$$

Then we can write G_k as

$$G_k = \pi_{k,1}\, G^{(1,k)} + \pi_{k,2}\, G^{(2,k)} + \pi_{k,3}\, G^{(3,k)} + \pi_{k,4}\, G^{(4,k)} + \pi_{k,5}\, G^{(5,k)},$$

where

$$\begin{aligned}
\pi_{k,1} &= \sum_{j=1,2,\ldots} p(j,j,j) \sim \beta(\alpha c_{k,1}, \alpha(1-c_{k,1})), \\
\pi_{k,2} &= \sum_{i,j=1,2,\ldots;i\neq j} p(i,j,j) \sim \beta(\alpha c_{k,2}, \alpha(1-c_{k,2})), \\
&\ \vdots \\
\pi_{k,5} &= \sum_{i,j,l=1,2,\ldots;i\neq j\neq l} p(i,j,l) \sim \beta(\alpha c_{k,5}, \alpha(1-c_{k,5}))
\end{aligned}$$

where $\beta(a, b)$ denotes the beta density with parameters a, b, and

$$\begin{aligned}
G^{(1,k)} &= \sum_{j=1,2,\ldots} \frac{p(j,j,j)}{\pi_{k,1}}\, \delta_{\theta_j^*} \\
G^{(2,k)} &= \sum_{i,j=1,2,\ldots;i\neq j} \frac{p(i,j,j)}{\pi_{k,2}}\, \delta_{\theta_{i,1}^*,\theta_{j,2}^*,\theta_{j,3}^*} \\
G^{(3,k)} &= \sum_{i,j=1,2,\ldots;i\neq j} \frac{p(j,i,j)}{\pi_{k,3}}\, \delta_{\theta_{j,1}^*,\theta_{i,2}^*,\theta_{j,3}^*} \\
G^{(4,k)} &= \sum_{i,j=1,2,\ldots;i\neq j} \frac{p(j,j,i)}{\pi_{k,4}}\, \delta_{\theta_{j,1}^*,\theta_{j,2}^*,\theta_{i,3}^*} \\
G^{(5,k)} &= \sum_{i,j,l=1,2,\ldots;i\neq j\neq l} \frac{p(i,j,l)}{\pi_{k,5}}\, \delta_{\theta_{i,1}^*,\theta_{j,2}^*,\theta_{l,3}^*}.
\end{aligned}$$

Let $c_j = \lim_{k\to\infty} c_{k,j}$, $j = 1, \ldots, 5$. Note that each c_j is finite, and they cannot be all zero since $c_{k,1} + \cdots + c_{k,5} = 1$ for all k. By the continuity properties of the Beta distribution in its parameters, we have that, for each $j = 1, \ldots, 5$, $\pi_{k,j}$ converges in law to $\pi_j \sim \beta(\alpha c_j, \alpha(1 - c_j))$. Furthermore, by (i) of Theorem 1,
if $c_1 > 0$, $G^{(1,k)} \to G^{(1)} \sim DP(\alpha c_1 G_0)$;
if $c_2 > 0$, $G^{(2,k)} \to G^{(2)} \sim DP(\alpha c_2 G_{0,1} G_{0,23})$;
$\vdots$
if $c_5 > 0$, $G^{(5,k)} \to G^{(2)} \sim DP(\alpha c_5 G_{0,1} G_{0,2} G_{0,3})$, where the convergence is in distribution.

It follows that

$$G_k \to G = \pi_1\, G^{(1)} + \pi_2\, G^{(2)} + \cdots + \pi_5\, G^{(5)}$$

in distribution, and extending a known property of the Dirichlet distribution (see Ghosh and Ramamoorthi (2003), (4) on page 91) to Dirichlet processes, one can show that

$$G \sim DP(\alpha(c_1 G_0 + c_2 G_{0,1} G_{0,23} + c_3 G_{0,2} G_{0,13} + c_4 G_{0,3} G_{0,12} + c_5 G_{0,1} G_{0,2} G_{0,3})).$$

(ii) We can write

$$G_k = \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} p_k(j_1, \ldots, j_m) \delta_{\theta_{j_1,1}^*,\ldots,\theta_{j_m,m}^*},$$

regarding p_k as a probability measure on $\{1, 2, \ldots\}^m$ with support $\{1, \ldots, k\}^m$; analogously for q_k. By assumption, the weights $p_k(\cdot)$ have a joint Dirichlet distribution $\mathcal{D}(\alpha q_k)$, but we can equivalently say that $p_k \sim DP(\alpha q_k)$. Since $q_k \to q$ and $\alpha_k \to \alpha$, we have that $p_k \to p$ with $p \sim DP(\alpha q)$ and G_k converges in distribution to

$$G = \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} p(j_1, \ldots, j_m) \delta_{\theta_{j_1,1}^*,\ldots,\theta_{j_m,m}^*},$$

which has a $hDP(\alpha q, G_0)$ prior. $\diamond$

A.2. MCMC algorithm for the applications of Section 5

We detail the full conditionals for the approximate jittered posterior of the model in Section 5. The prior q is the discretization of a continuous distribution with uniform marginals. Using the notation in Sections 4.1 and 5, let $U_i(x_j) = F_x(L_i)$, with $L_i \sim N(0, \sigma_q^2 R(\phi_q))$. Here, the jittered approximation is a perturbation of the latent L_i, $\tilde{L}_i(x) = L_i(x) + \varepsilon_i(x)$, with $\varepsilon_i(x) \sim N(0, \eta^2)$. So, $\tilde{U}_i = F_x(\tilde{L}_i)$. The variance η^2 is fixed at a value that guarantees that $\tilde{U}_i$ and U_i share the same allocation structure with high probability. As a rule of thumb, we suggest to take $\eta = \sigma_q/10$, since for such a value $\sigma_q^2/(\eta^2 + \sigma_q^2) \approx 0.99$, and $P(\tilde{U}_i \in C_{j_1,\dots,j_m}) \approx P(U \in C_{j_1,\dots,j_m})$. The sensitivity analyses performed in our simulation studies show that the results don't show any significant improvement for smaller values of η. Following Section 4.3, given the one-to-one correspondence between U_i and L_i, we consider $L_i|P \sim P, P \sim DP(\alpha F_0^{(m)})$, with $F_0^{(m)} = N(0, \sigma_q^2 R(\phi_q))$. The jitter and the model for q specify the relevant distribution for the Gibbs sampler. The algorithm can be described as follows.

- The vectors $\tilde{L}_i$ are updated component by component. For notational convenience, let $\tilde{L}_{i,j} = \tilde{L}_i(x_j)$ and $L_{i,j} = L_i(x_j)$, $j = 1, \dots, m$, $i = 1, \dots, n$. Then, with probability w_r, $\tilde{L}_{i,j}$ is sampled from the univariate normal $N(L_{i,j}, \eta^2)$ truncated over the interval $\left(\sigma_q \Phi^{-1}\left(\frac{r-1}{k}\right), \sigma_q \Phi^{-1}\left(\frac{r}{k}\right)\right]$, $r = 1, \dots, k$, where Φ is the gaussian cdf. Here, for $r = 1, \dots, k$,

$$w_r \propto \left[\Phi\left(\left(\sigma_q \Phi^{-1}\left(\frac{r}{k}\right) - L_{i,j}\right)/\eta\right) - \Phi\left(\left(\sigma_q \Phi^{-1}\left(\frac{r-1}{k}\right) - L_{i,j}\right)/\eta\right)\right] \times$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}(Y_i(x_j) - \mu_i(x_j) - \theta_r^*(x_j))^2\right\}$$

- Since the $L_i = (L_i(x_1), \dots, L_i(x_m))^T$ are samples from a DP, they are updated through a Pólya Urn scheme. Accordingly, the vector L_i is either sampled from the multivariate normal $N(\frac{\Lambda}{\eta^2} L_i, \Lambda)$, $\Lambda = \left(\frac{1}{\eta^2} I_m + \frac{1}{\sigma_q^2} R^{-1}(\phi_q)\right)^{-1}$ with probability ω_0 or is one of the existing L_j^* in the DP urn (see (**??**)) with probability ω_j, where

$$\omega_0 = \frac{\alpha}{(2\pi)^{n/2} det(\Lambda)^{1/2}} \exp\{-\frac{1}{2} L_i^T \Lambda^{-1} L_i\},$$
$$\omega_j = \frac{m_j}{\sqrt{2\pi\eta^2}} \exp\{-\frac{1}{2} L_j^{*T} L_j^*\}, \quad j = 1, \dots, k,$$

  where m_j is the frequency in cluster j.

- Update $\theta_1^*, \dots, \theta_k^*$ as described in Section 4.3 and the other hyperparameters according to their (standard) full conditionals.

Additional References

Ghosh, J.K. and Ramamoorthi, R.V. (2003). *Bayesian nonparametrics*. Springer, New York.