Supplementary material of Non-exchangeable feature allocation models with sublinear growth of the feature sizes

Giuseppe Di Benedetto
University of Oxford

François Caron
University of Oxford

Yee Whye Teh
University of Oxford
DeepMind

In this document we provide the proofs of the Propositions stated in the main paper, together with the posterior predictive distribution of the feature allocations.

A Posterior predictive distribution of feature allocations

The data augmentation described in Section 5 in the main paper allows us to compute the posterior predictive distribution of the $n+1$-th object given the first n ones.

$$Z_{n+1} \,|\, U_1,\dots,U_n \overset{d}{=} Z^*_{n+1} + \sum_{k=1}^{K_n} \widetilde{z}_{n+1,k}\delta_{\widetilde{\theta}_k}$$

where Z^*_{n+1} is independent of the $\widetilde{z}_{n+1,k}$ which are distributed as follows

$$\widetilde{z}_{n+1,k} \,|\, U_1,\dots,U_n \sim \mathrm{Ber}\left(1 - \frac{\kappa\left(\widetilde{m}_{nj}\,,\, \Delta_{n+1} + \sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k \le Y_i}\right)}{\kappa\left(\widetilde{m}_{nj}\,,\, \sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k \le Y_i}\right)}\right).$$

By the marking theorem for Poisson point processes and Equation (20), $Z^*_{n+1} = \sum_{k=K_n+1}^{K_n+K^*_{n+1}} \delta_{\widetilde{\theta}_k}$ is a Poisson random measure on $\mathbb{R}_+$ with mean measure $\mu^*_{n+1}(d\theta) = \int_0^\infty \left(1 - e^{-\omega\Delta_{n+1}1_{\theta\le Y_{n+1}}}\right) d\nu^*(d\omega, d\theta)$, where we recall that

$$\begin{aligned}\nu^*(d\omega, d\theta) &= e^{-\omega\sum_{i=1}^n \Delta_i 1_{\theta\le Y_i}}\rho(\omega)d\omega d\theta \\ &= \left(1_{\theta > Y_n} + \sum_{i=1}^n e^{-\omega(T_n - T_{i-1})} 1_{Y_{i-1}<\theta\le Y_i}\right) \\ &\quad \times \rho(\omega) d\omega d\theta\end{aligned}$$

where $T_0 = Y_0 = 0$. Therefore, the number $K^*_{n+1} = Z^*_{n+1}(\mathbb{R}_+)$ of new features of the $n+1$ object is Poisson distributed with mean

$$\mathbb{E}[Z^*_{n+1}(\mathbb{R}_+)] = \int_0^\infty \int_0^\infty \left(1 - e^{-\omega\Delta_{n+1}1_{\theta\le Y_{n+1}}}\right) d\nu^*(d\omega, d\theta)$$

If follows that

$$\mathbb{E}[Z^*_{n+1}(\mathbb{R}_+)] = \sum_{i=1}^{n+1} (Y_i - Y_{i-1}) \int_0^\infty \left(1 - e^{-\omega\Delta_{n+1}}\right) e^{-\omega(T_n - T_{i-1})}\rho(\omega)d\omega \tag{A.1}$$

The locations $\widetilde{\theta}_{K_n+1},\dots,\widetilde{\theta}_{K_n+K^*_{n+1}}$ are sampled iid from the piecewise constant distribution on $[0, Y_{n+1}]$ with pdf proportional to

$$\sum_{i=1}^{n+1} 1_{Y_{i-1}<\theta\le Y_i} \int_0^\infty \left(1 - e^{-\omega\Delta_{n+1}}\right) e^{-\omega(T_n - T_{i-1})}\rho(\omega)d\omega \tag{A.2}$$

If B is a GGP, the integral in Equations (A.1) and (A.2) is tractable and we have

$$\int_0^\infty \left(1 - e^{-\omega\Delta_{n+1}}\right) e^{-\omega(T_n - T_{i-1})}\rho(\omega)d\omega = \begin{cases} \frac{\eta}{\sigma}\left[(T_{n+1} - T_{i-1} + \zeta)^\sigma - (T_n - T_{i-1} + \zeta)^\sigma\right] & \sigma > 0 \\ \eta \log\left(1 + \frac{\Delta_{n+1}}{T_n - T_{i-1} + \zeta}\right) & \sigma = 0 \end{cases}$$

Proof. We have

$$\Pr(\widetilde{z}_{n+1,k} = 1 \,|\, \widetilde{\omega}_k, U_1,\dots,U_n) = 1 - e^{-\widetilde{\omega}_k\Delta_{n+1}}$$

and

$$p(\widetilde{\omega}_k \mid U_1,\dots,U_n) = \frac{\widetilde{\omega}_k^{\widetilde{m}_{n,k}} e^{-\widetilde{\omega}_k\sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k\le Y_i}} \rho(\widetilde{\omega}_k)}{\kappa\left(\widetilde{m}_{n,k}, \sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k\le Y_i}\right)}$$

Hence

$$\Pr(\widetilde{z}_{n+1,k} = 1 \mid U_1, \ldots, U_n) = 1 - \frac{\kappa\left(\widetilde{m}_{n,k}, \Delta_{n+1} + \sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k \leq Y_i}\right)}{\kappa\left(\widetilde{m}_{n,k}, \sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k \leq Y_i}\right)}$$

□

The conditional distribution of the latent point process U_{n+1} can be written as follows:

$$p(\widetilde{u}_{n+1,k} \mid \widetilde{z}_{n+1,k} = 1,\, \widetilde{u}_{1:n,k}) \propto \kappa\left(\widetilde{m}_{n,k} + 1, \Delta_{n+1}\widetilde{u}_{n+1,k} + \sum_{i=1}^n \Delta_i \widetilde{u}_{ik}\right) 1_{u_{n+1,j} < 1}$$

Proof. We have

$$p(\widetilde{u}_{n+1,k} \mid \widetilde{\omega}_k, \widetilde{z}_{n+1,k} = 1,\, \widetilde{u}_{1:n,k}) = \frac{\Delta_{n+1}\widetilde{\omega}_k e^{-\widetilde{u}_{n+1,k}\Delta_{n+1}\widetilde{\omega}_k} 1_{\widetilde{u}_{n+1,k} < 1}}{1 - e^{-\Delta_{n+1}\widetilde{\omega}_k}}$$

and

$$p(\widetilde{\omega}_k \mid \widetilde{z}_{n+1,k} = 1, U_1, \ldots, U_n) \propto (1 - e^{-\Delta_{n+1}\widetilde{\omega}_k})\widetilde{\omega}_k^{\widetilde{m}_{n,k}} e^{-\widetilde{\omega}_k \sum_{i=1}^n \Delta_i \widetilde{u}_{ij} 1_{\widetilde{\theta}_k \leq Y_i}} \rho(\widetilde{\omega}_k)$$

□

B Proofs

B.1 Proof of Proposition 1

By the marking theorem for Poisson point processes, the set of points $\{(\omega_j)_{j\geq 1} \mid z_{ij} = 1\}$ is drawn from a Poisson point process with mean measure $Y_i(1 - e^{-\Delta_i \omega})\rho(\omega)d\omega$. The total number of such points $Z_i(\mathbb{R}_+)$ is therefore Poisson distributed with mean $\mathbb{E}\left[Z_i(\mathbb{R}_+)\right] = Y_i \psi(\Delta_i)$. Using integration by part, we have

$$\psi(t) = t \int_0^\infty e^{-wt} \overline{\rho}(w) dw$$

where

$$\overline{\rho}(x) = \int_x^\infty \rho(w) dw. \tag{B.1}$$

Hence, by monotone convergence,

$$\lim_{t\to\infty} \frac{\psi(t)}{t} = \int_0^\infty \overline{\rho}(w) dw = \kappa(1, 0)$$

and it follows that

$$Y_n \psi(\Delta_n) \overset{n\to\infty}{\sim} Y_n \Delta_n \kappa(1, 0)$$

Finally note that $\Delta_n \overset{n\to\infty}{\sim} (1 + \xi)^{-1} n^{-\xi/(\xi+1)}$, hence $Y_n \Delta_n \to (1 + \xi)^{-1}$.

B.2 Proof of Proposition 2

The number of features observed in the first n objects can be written as

$$m_n := \sum_{i=1}^n \sum_{j\geq 1} z_{ij} = \sum_{i=1}^n Z_i(\mathbb{R}_+)$$

Since $\mathbb{E}[Z_n(\mathbb{R}_+)] = \mathbb{E}[m_n] - \mathbb{E}[m_{n-1}]$, it follows by Stolz-Cesàro theorem that

$$\mathbb{E}[m_n] \overset{n\to\infty}{\sim} (1 + \xi)^{-1}\kappa(1, 0)n.$$

In order to get the almost sure convergence of m_n to its expectation we can use the Kolmogorov strong law of large numbers which, under the assumption $\sum_{n\geq 1} \mathrm{Var}\left(\frac{Z_n(\mathbb{R}_+)}{n}\right) < \infty$, gives

$$\frac{m_n - \mathbb{E}[m_n]}{n} \to 0 \text{ almost surely.}$$

Recall that $\mathrm{Var}(Z_n(\mathbb{R}_+)) = \mathbb{E}[Z_n(\mathbb{R}_+)]$. Therefore the summability condition on the variance boils down to the convergence of the sum $\sum_{n\geq 1} \frac{1}{n^2} Y_n\, \psi(\Delta_n)$, which holds true since the elements of the sum are of order n^{-2}.

B.3 Proof of Proposition 3

Since $\mathbb{E}[m_{n,j}|B] = \sum_{i=1}^n \left(1 - e^{-\omega_j \Delta_i}\right) 1_{\theta_j < Y_i}$, we have $\frac{\mathbb{E}[m_{nj}|B] - \mathbb{E}[m_{n-1\,j}|B]}{T_n - T_{n-1}} = \frac{1 - e^{-\omega_j \Delta_n}}{\Delta_n} 1_{\theta_j < Y_n} \to \omega_j$, then by Stolz-Cesàro theorem we have that

$$\mathbb{E}[m_{nj}|B] \overset{n\to\infty}{\sim} \omega_j T_n.$$

We have

$$\mathrm{Var}(m_{n,j} \mid B) = \sum_{i=1}^n \left(1 - e^{-\omega_j \Delta_i}\right) e^{-\omega_j \Delta_i} 1_{\theta_j < Y_i} \leq \mathbb{E}[m_{n,j}|B].$$

Using the sandwiching argument in Proposition 2 of Gnedin et al. (2007) it follows that, conditionally on B, $\frac{m_{nj}}{\mathbb{E}[m_{nj} \mid B]} \to 1$ almost surely.

B.4 Proof of Proposition 4

Applying Campbell's theorem

$$\begin{aligned}
\mathbb{E}[K_n] &= \mathbb{E}[\mathbb{E}[K_n \mid B]] \\
&= \mathbb{E}\left[\sum_j \Pr(m_{n,j} > 0 \mid B)\right] \\
&= \int_0^\infty \int_0^\infty \left(1 - e^{-\omega f_n(\theta)}\right) \rho(\omega) d\omega d\theta \\
&= \int_0^{Y_n} \psi(T_n - g(\theta)) d\theta
\end{aligned}$$

where

$$f_n(\theta) := \sum_{i=1}^{n} \Delta_i 1_{\theta \le Y_i} = (T_n - g(\theta)) 1_{\theta \le Y_n} \qquad \text{(B.2)}$$

with $g(\theta) := \sum_{i=1}^{\infty} \Delta_i 1_{\theta > Y_i}$ is a monotone increasing step function satisfying, for all $\theta \ge 0$

$$\max(0, g_1(\theta)) \le g(\theta) \le g_2(\theta) \qquad \text{(B.3)}$$

where $g_1(\theta) = (\theta^{(\xi+1)/\xi} - 1)^{1/(\xi+1)}$ and $g_2(\theta) = \theta^{1/\xi}$. Note that $g_1(\theta) \overset{\theta\to\infty}{\sim} g_2(\theta) \overset{\theta\to\infty}{\sim} \theta^{1/\xi}$. As ψ is an increasing function, it follows

$$\int_0^{Y_n} \psi(T_n - g_2(\theta)) d\theta \le \mathbb{E}[K_n] \le \int_1^{Y_n} \psi(T_n - g_1(\theta)) d\theta + \psi(T_n).$$

Using a change of variable, we obtain

$$\int_0^{Y_n} \psi(T_n - g_2(\theta)) d\theta = \xi \int_0^{T_n} \psi(T_n - \theta)\, \theta^{\xi - 1} d\theta$$

Finally, noting that

$$\psi(t) \overset{t\to\infty}{\sim} t^{\sigma} \ell(t)$$

where

$$\ell(t) = \begin{cases} \eta \log(t) & \sigma = 0 \\ \frac{\eta}{\sigma} & \sigma \in (0,1) \end{cases}$$

and using (Di Benedetto et al., 2017, Lemma 14), we obtain

$$\int_0^{Y_n} \psi(T_n - g_2(\theta)) d\theta \overset{n\to\infty}{\sim} \frac{\Gamma(\xi+1)\Gamma(\sigma+1)}{\Gamma(\sigma+\xi+1)} n^{\frac{\xi+\sigma}{\xi+1}} \ell(n)$$

Similarly, we have

$$\int_1^{Y_n} \psi(T_n - g_1(\theta)) d\theta \overset{n\to\infty}{\sim} \frac{\Gamma(\xi+1)\Gamma(\sigma+1)}{\Gamma(\sigma+\xi+1)} n^{\frac{\xi+\sigma}{\xi+1}} \ell(n).$$

It follows by sandwiching that

$$\mathbb{E}[K_n] \overset{n\to\infty}{\sim} \frac{\Gamma(\xi+1)\Gamma(\sigma+1)}{\Gamma(\sigma+\xi+1)} n^{\frac{\xi+\sigma}{\xi+1}} \ell(n).$$

Using Campbell's theorem again,

$$\begin{aligned} \mathrm{Var}[K_n] &= \mathrm{Var}[\mathbb{E}[K_n \mid B]] + \mathbb{E}[\mathrm{Var}[K_n \mid B]] \\ &= \int \left(1 - e^{-\omega f_n(\theta)}\right) e^{-\omega f_n(\theta)} \rho(d\omega) d\omega d\theta \\ &\quad + \int \left(1 - e^{-\omega f_n(\theta)}\right)^2 \rho(d\omega) d\omega d\theta \\ &= \int \left(1 - e^{-\omega f_n(\theta)}\right) \rho(d\omega) d\omega d\theta \end{aligned}$$

therefore the almost sure asymptotic equivalence follows by Chebyshev inequality and the strong law of large numbers for K_n (see (Gnedin et al., 2007, Proposition 2)).

B.5 Proof of Proposition 5

We have the following inequality, for any $x \ge 0$

$$0 \le x - (1 - e^{-x}) \le \frac{x^2}{2}. \qquad \text{(B.4)}$$

Let us recall that

$$K_{n,r} = \sum_{j \ge 1} 1_{m_{n,j} = r}.$$

where $m_{n,j} = \sum_{i=1}^{n} z_{ij}$. Conditional on the CRM B we have

$$\mathbb{E}[K_{n,r} \mid B] = \sum_{j \ge 1} \Pr\left(m_{n,j} = r \,\middle|\, B\right).$$

Note that $\Pr\left(m_{n,j} = r \,\middle|\, B\right)$ only depends on (θ_j, ω_j). Write $S_{n,r}(\theta_j, \omega_j) = \Pr\left(m_{n,j} = r \,\middle|\, B\right)$. Let us denote $q_i(\theta_j, \omega_j) := \Pr(z_{ij} = 1 \mid B) = 1 - e^{-\Delta_i \omega_j 1_{\theta_j \le Y_i}}$; $\lambda_n(\theta_j, \omega_j) := \sum_{i=1}^{n} q_i(\theta_j, \omega_j)$. Conditional on B the random variable $m_{n,j}$ has a Poisson-Binomial distribution with parameters $(q_1(\theta_j, \omega_j), \ldots, q_n(\theta_j, \omega_j))$. For each fixed (ω_j, θ_j) Le Cam's inequality Le Cam (1960) and inequality (B.4) give

$$\begin{aligned} &\sum_{r \ge 0} \left| \Pr(m_{n,j} = r \mid B) - \mathrm{Poisson}\left(r; \lambda_n(\theta_j, \omega_j)\right) \right| \\ &\le 2 \sum_{i=1}^{n} q_i(\theta_j, \omega_j)^2 \le 2\omega_j^2 \sum_{i=1}^{n} \Delta_i^2 1_{\theta_j \le Y_i}. \qquad \text{(B.5)} \end{aligned}$$

where $\mathrm{Poisson}(r; \lambda)$ denote the probability mass function of a Poisson random variable with rate parameter λ evaluated at r. Note that for any $0 < \lambda_1 \le \lambda_2$, using coupling inequalities (see. e.g. (Roch, 2015, Example 4.10 p. 154))

$$\sum_{r \ge 0} \left| \mathrm{Poisson}(r; \lambda_1) - \mathrm{Poisson}(r; \lambda_2) \right| \le 2(\lambda_2 - \lambda_1).$$

Noting that $\lambda_n(\theta_j, \omega_j) \le \omega_j f_n(\theta_j)$, where f_n is defined in Equation (B.2), and using inequality (B.4), we obtain

$$\begin{aligned} &\sum_{r \ge 0} \left| \mathrm{Poisson}(r; \lambda_n(\theta_j, \omega_j)) - \mathrm{Poisson}(r; \omega_j f_n(\theta_j)) \right| \\ &\le 2 \sum_{i=1}^{n} \left(\omega_j \Delta_i - (1 - e^{-\omega_j \Delta_j}) \right) 1_{\theta_j \le Y_i} \\ &\le \omega_j^2 \sum_{i=1}^{n} \Delta_i^2 1_{\theta_j \le Y_i} \end{aligned}$$

Combining the above inequality with the inequality (B.5), we obtain the total variation bound

$$\sum_{r\geq 0}\left|\Pr(m_{n,j}=r\mid B)-\text{Poisson}\left(r;\omega_j f_n(\theta_j)\right)\right| \leq 3\omega_j^2\sum_{i=1}^n \Delta_i^2 1_{\theta_j\leq Y_i}. \tag{B.6}$$

Using Campbell's theorem,

$$\mathbb{E}\left[\sum_{j\geq 1}\omega_j^2\sum_{i=1}^n \Delta_i^2 1_{\theta_j\leq Y_i}\right]=\kappa(2,0)\sum_{i=1}^n Y_i\Delta_i^2 \overset{n\to\infty}{\asymp} n^{1/(1+\xi)}. \tag{B.7}$$

Using Campbell's theorem again,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{j\geq 1}\text{Poisson}\left(r;\omega_j f_n(\theta_j)\right)\right]\\
&=\frac{1}{r!}\int_0^\infty\int_0^\infty e^{-\omega f_n(\theta)}\omega^r f_n(\theta)^r\,\rho(\omega)\,d\omega d\theta\\
&=\frac{1}{r!}\int_0^\infty \kappa(r,f_n(\theta))f_n(\theta)^r d\theta\\
&=\frac{1}{r!}\int_0^{Y_n}\kappa(r,T_n-g(\theta))(T_n-g(\theta)))^r d\theta
\end{aligned}$$

We use again the inequality (B.3) to bound the above expression. The upper bound is given by

$$\frac{1}{r!}\int_0^{Y_n}\kappa(r,T_n-g_2(\theta))(T_n-g_1(\theta))^r d\theta$$

Using a change of variable, we obtain

$$\begin{aligned}
&\frac{\xi}{r!}\int_0^{T_n}\kappa(r,T_n-\theta)(T_n-g_1(\theta^\xi)))^r\theta^{\xi-1}d\theta\\
&=\frac{\xi}{r!}\int_0^{T_n}\kappa(r,\theta)(g_1(\theta^\xi))^r(T_n-\theta)^{\xi-1}d\theta.
\end{aligned} \tag{B.8}$$

Noting that $\kappa(r,\theta)\overset{\theta\to\infty}{\sim}\eta\theta^{\sigma-r}\frac{\Gamma(r-\sigma)}{\Gamma(1-\sigma)}$ and $(g_1(\theta^\xi))^r\overset{\theta\to\infty}{\sim}\theta^r$, and using (Di Benedetto et al., 2017, Lemma 14), we obtain that (B.8) is asymptotically equivalent to

$$\eta\frac{\Gamma(\xi+1)\Gamma(r-\sigma)}{r!\Gamma(\sigma+\xi+1)}\,n^{\frac{\xi+\sigma}{\xi+1}}.$$

A similar asymptotic equivalence is obtained for the lower bound, and we conclude by sandwiching that

$$\mathbb{E}\left[\sum_{j\geq 1}\text{Poisson}\left(r;\omega_j f_n(\theta_j)\right)\right] \overset{n\to\infty}{\sim}\eta\frac{\Gamma(\xi+1)\Gamma(r-\sigma)}{r!\Gamma(\sigma+\xi+1)}\,n^{\frac{\xi+\sigma}{\xi+1}}$$

Combining the above asymptotic result with Equations (B.6) and (B.7), and assuming $\xi+\sigma>1$, we conclude

$$\mathbb{E}[K_{n,r}]=\mathbb{E}\left[\sum_{j\geq 1}\Pr\left(m_{n,j}=r\,\middle|\,B\right)\right] \overset{n\to\infty}{\sim}\eta\frac{\Gamma(\xi+1)\Gamma(r-\sigma)}{r!\Gamma(\sigma+\xi+1)}\,n^{\frac{\xi+\sigma}{\xi+1}}.$$

The variance of $K_{n,r}$ can be written as

$$\begin{aligned}
\text{Var}[K_{n,r}]&=\text{Var}[\mathbb{E}[K_{n,r}\mid B]]+\mathbb{E}[\text{Var}[K_{n,r}\mid B]]\\
&=\int S_{n,r}(\theta,\omega)\rho(\omega)d\omega d\theta\\
&+\int S_{n,r}(\theta,\omega)(1-S_{n,r}(\theta,\omega))\,\rho(\omega)d\omega d\theta\\
&=\mathbb{E}[K_{n,r}].
\end{aligned}$$

Using the result below Proposition 2 in Gnedin et al. (2007), we obtain, almost surely, $\mathbb{E}\left[\sum_{r\geq j}K_{n,r}\right]\overset{n\to\infty}{\sim}\sum_{r\geq j}K_{n,r}$. Using a proof similar to that of Corollary 21 in Gnedin et al. (2007), we obtain

$$K_{n,r}\overset{n\to\infty}{\sim}\eta\frac{\Gamma(\xi+1)\Gamma(r-\sigma)}{r!\Gamma(\sigma+\xi+1)}\,n^{\frac{\xi+\sigma}{\xi+1}}. \tag{B.9}$$

Combining Equation (B.9) with Equation (17) gives the final result.

References

Di Benedetto, G., Caron, F., and Teh, Y. W. (2017). Non-exchangeable random partition models for microclustering. *arXiv preprint arXiv:1711.07287.*

Gnedin, A., Hansen, B., and Pitman, J. (2007). Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv*, 4(146-171):88.

Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10(4):1181–1197.

Roch, S. (2015). Modern discrete probability: An essential toolkit. chapter 4. https://www.math.wisc.edu/ roch/mdp/roch-mdp-chap4.pdf.