Supplementary material for Bayesian nonparametric model for arbitrary cubic partitioning

Masahiro Nakano MASAHIRO.NAKANO.PR@HCO.NTT.CO.JP
Yasuhiro Fujiwara YASUHIRO.FUJIWARA.KH@HCO.NTT.CO.JP
Akisato Kimura AKISATO.KIMURA.XN@HCO.NTT.CO.JP
Takeshi Yamada TAKESHI.YAMADA.BC@HCO.NTT.CO.JP
Naonori Ueda NAONORI.UEDA.FR@HCO.NTT.CO.JP
NTT Communication Science Laboratories, NTT Corporation

Editors: Vineeth N Balasubramanian and Ivor Tsang

1. Appendix

Let $\mathcal{B}$ be a set of all cubic partitions of $[0,1]\times[0,1]\times[0,1]$. The distribution of a Markov process $(B_t; t \geq 0)$ is completely determined by (1) the distribution of its initial state B_0 and (2) its rate kernel $K : \mathcal{S}\times\mathcal{S} \to [0,1]$, which satisfies

$$K(x, A) := \lim_{h\downarrow 0} \frac{\mathbb{P}[B_{t+h} \in A \mid B_t = x]}{h}, \tag{1}$$

for all $t > 0$, all $x \in \mathcal{S}$, and all measureble subset $A \subset \mathcal{S}$. Specifically, we set B_0 as a single rectangle of $[0,1]\times[0,1]$, and employ the rate kernel corresponding to the stochastic enumeration algorithm described in Section 4.

Theorem 1 (Existence of CPP) *Let $(B_t; t \geq 0)$ be a Markov process whose initial state and rate kernel can be defined in Section 4. Then there exists such a Markov process.*

Proof For the proof, it is enough to show that the growth rate of the intensity of the Markov process is not too fast and does not diverge infinitely, as in the existence proof of the Mondrian process (Roy, 2011). In fact, we can see that, in a single jump, the intensity can only increase by a constant. Therefore, the intensity of the Markov process does not diverge, and we can show the existence of CPP.

Let x_i and t_i $(i = 0, 1, 2, \dots)$ be the Markov chain of cubic partitioning and jump times, respectively. For any $x_{i+1} \in \mathcal{S}$, by construction, we have

$$\eta(x_{i+1}) \leq \eta(x_i) + 6. \tag{2}$$

Then, we apply the law of iterated expectations, and obtain

$$\begin{aligned}
\mathbb{E}[\eta(x_{i+1})\mathcal{P}[\cdot \mid B_{t_{i+1}} = x_{i+1}]] &= \mathbb{E}\Big[\mathbb{E}\big[\eta(x_{i+1})\mathcal{P}[\cdot \mid B_{t_{i+1}} = y]\big] \mid B_{t_i} = x_i\Big] && (3)\\
&< \mathbb{E}[\eta(x_i)\mathcal{P}[\cdot \mid B_{t_i} = x_i]] + 6. && (4)
\end{aligned}$$

Therefore, we can immediately obtain

$$\mathbb{E}[\eta(x_n)\mathcal{P}[\cdot \mid B_{t_n} = x_n]] = \mathbb{E}[\eta(x_0)\mathcal{P}[\cdot \mid B_{t_0} = x_0]] + 6n. \tag{5}$$

Then, it follows from Fatou's lemma that

$$\begin{aligned} \mathbb{E}\Big[\liminf_{n\to\infty}\frac{\eta(x_n)\mathcal{P}[\cdot \mid B_{t_n} = x_n]}{n}\Big] &\leq \liminf_{n\to\infty}\mathbb{E}\Big[\frac{\eta(x_n)\mathcal{P}[\cdot \mid B_{t_n} = x_n]}{n}\Big] && (6)\\ &< \liminf_{n\to\infty}\mathbb{E}\Big[\frac{\eta(x_0)\mathcal{P}[\cdot \mid B_{t_0} = x_0] + n}{n}\Big] && (7)\\ &< \infty. && (8)\end{aligned}$$

Therefore, we obtain, almost surely,

$$\liminf_{n\to\infty}\frac{\eta(x_n)\mathcal{P}[\cdot \mid B_{t_n} = x_n]}{n} < \infty, \tag{9}$$

and complete the proof. ■

Theorem 2 (Support of CPP) *Let $\boldsymbol{B} = (B_t; t \geq 0)$ be CPP as defined in Section 4, and let μ_B be the probability distribution induced from $\boldsymbol{B}$. The support of μ_B is all cubic partitions $\mathcal{B}$ of $[0,1] \times [0,1] \times [0,1]$.*

Proof By mathematical induction. Now we assume that CPP is capable of generating all cubic partitions consisting of no more than n blocks. We need to show that any rectangular partition with $n+1$ blocks can be generated from CPP. For any cubic partition B of $[0,1] \times [0,1] \times [0,1]$ with $n+1$ blocks, we consider the following operation. Without loss of generality, we can suppose that the block of B that contains the origin is denoted by $[0,P]\times[0,Q]\times[0,R]$. Let F be the set of all blocks of B that have a face in $P\times[0,Q]\times[0,R]$. For every block c_j in F, we replace its vertex $(v_j^{(1)}, v_j^{(2)}, v_j^{(3)})$ with $(0, v_j^{(2)}, v_j^{(3)})$. Next, for all blocks c_j in B that are not in F, if the block c_j has a region that overlaps with F, we delete that region. After performing the above operations, a new cubic partition B' of $[0,1] \times [0,1] \times [0,1]$ is obtained. We recall here that the rectangular partitions obtained by slicing the possible cubic partitions in the yz-, zx-, and xy-planes can be enumerated by CPP (Merino and Mütze, 2021). Thus, by construction, we can see that the number of blocks in the cubic partition B' is less than or equal to n. Therefore, from the assumption of mathematical induction, we can say that B' can be generated by CPP. Furthermore, from the CPP construction procedure, we can see that B can be generated from CPP given B'. This completes the proof from mathematical induction. ■

Bayesian inference - Broadly, the Bayesian inference algorithm for the CPP-based relational model can be viewed as a problem of inferring (1) **Element coordinates:** Uniform random variables, $\boldsymbol{U}^{(1)} := (U_1^{(1)}, \ldots, U_L^{(1)})$, $\boldsymbol{U}^{(2)} := (U_1^{(2)}, \ldots, U_M^{(2)})$, and $\boldsymbol{U}^{(3)} := (U_1^{(3)}, \ldots, U_N^{(3)})$, and (2) **Cubic partitioning:** A random sample of CPP, $(B_t; t \geq 0)$, given the input data and hyper-parameters,that is, (a) **Input matrix:** $\boldsymbol{Z} := \{Z_{l,m,n} \mid$

$l = 1, \dots, L,\ m = 1, \dots, M, n = 1, \dots, N\}$, consisting of categorical elements, $Z_{l,m,n} \in \{1, 2, \dots, D\}$ ($D \in \mathbb{N}$), and (b) **Hyper parameters:** $\lambda > 0$ and $\alpha_0 > 0$. Then, the joint probability density is expressed as follows:

$$p\left(\boldsymbol{Z}, \boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)} \mid \lambda, \alpha_0\right) = p_{\mathrm{BP}}(B_t \mid \lambda) \cdot \left(\prod_{l=1}^{L} p_{\mathrm{uniform}}(U_l^{(1)})\right) \cdot \left(\prod_{m=1}^{M} p_{\mathrm{uniform}}(U_m^{(2)})\right) \cdot \left(\prod_{n=1}^{N} p_{\mathrm{uniform}}(U_n^{(3)})\right) \cdot p_{\mathrm{obs.}}(\boldsymbol{Z} \mid \boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)}, \alpha_0), \quad (10)$$

where $p_{\mathrm{BP}}(\cdot)$ is the probability law induced from CPP, $p_{\mathrm{uniform.}}(\cdot)$ is the probability density of the uniform distribution on $[0, 1]$, and the third term is

$$p_{\mathrm{obs.}}(\boldsymbol{Z} \mid \boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)}, \alpha_0) \propto \prod_{k=1}^{\infty} \left(\frac{\Gamma(D\alpha_0)}{\Gamma(D\alpha_0 + \sum_{d=1}^{D} \mathcal{N}_{k,d})} \prod_{d=1}^{D} \frac{\Gamma(\alpha_0 + \mathcal{N}_{k,d})}{\Gamma(\alpha_0)} \right), \quad (11)$$

where $\mathcal{N}_{k,d}$ denotes the number of elements in both the k-th block and the d-th category of the categorical distribution.

The most straightforward Bayesian inference algorithm can be derived as follows. We iteratively repeat the following two update rules:

- **Update of element coordinates -** For each element of $\boldsymbol{U}^{(1)}, \boldsymbol{U}^{(2)}, \boldsymbol{U}^{(3)}$, we generate a new candidate from the prior distribution, that is, the uniform distribution Uniform($[0, 1]$), and decide whether to accept or reject it by the Metropolis-Hastings (MH) algorithm using Equation (10).

- **Update of cubic partitioning -** For cubic partitioning of $(B_t; t \geq 0)$, we can consider two possible methods: (1) the reversible jump MCMC (RJMCMC) method (Wang et al., 2011) and (2) the particle MCMC (PMCMC) method (Fan et al., 2018, 2019; Ge et al., 2019). **(1) RJMCMC:** We assume that in the current iteration of MCMC, $\boldsymbol{B}$ has n jumps. RJMCMC has a proposal distribution that allows a transition to the case of $n - 1$ and $n + 1$ jumps in the next update, and a new candidate $\boldsymbol{B}$ is generated from the proposal distribution. Finally, the MH scheme decides whether to accept or reject it. **(2) PMCMC:** We have p particles, and each particle samples its own current state at the next jump of the Markov process by choosing its own current state according to the ratio of posterior probabilities using Equation (10) from each particle in the previous jump.

For hyper-parameters, we set the number of particles to 30. We compare the CPP-based relational model with the BNP stochastic block models based on cubic partitioning: (1) IRM: the intermediate random function of the AHK representation is drawn from the product of the SBPs, and the concentration parameter is drawn from the Gamma$(1, 1)$ prior. (2) MP: the intermediate random function of the AHK representation is drawn from the MP, the budget parameter of which is drawn from Gamma$(3, 1)$.

References

X. Fan, B. Li, and S. A. Sisson. The binary space partitioning-tree process. In *International Conference on Artificial Intelligence and Statistics*, pages 1859–1867, 2018.

Xuhui Fan, Bin Li, and Scott Anthony Sisson. Binary space partitioning forests. *arXiv:1903.09348*, 2019.

Shufei Ge, Shijia Wang, Yee Whye Teh, Liangliang Wang, and Lloyd Elliott. Random tessellation forests. In *Advances in Neural Information Processing Systems 32*, pages 9575–9585. 2019.

Arturo Merino and Torsten Mütze. Combinatorial generation via permutation languages. iii. rectangulations. *arXiv:2103.09333*, 2021.

D. M. Roy. *Computability, inference and modeling in probabilistic programming.* PhD thesis, Massachusetts Institute of Technology, 2011.

P. Wang, K. B. Laskey, C. Domeniconi, and M. I. Jordan. Nonparametric bayesian co-clustering ensembles. In *SIAM International conference on Data Mining*, pages 331–342, 2011.