Posterior representations of hierarchical completely random measures in trait allocation models

Lorenzo Masoero[1], Federico Camerlenghi[2], Stefano Favaro[3], Tamara Broderick[1] ▲ [1]MIT, CSAIL, [2]University of Milano, [3]University of Torino

Bayesian nonparametric latent trait models

- Many data analysis problems can be phrased as **discovering a set of latent traits** within a population:
  - *topics* in text data, *ancestral populations* in genomic data, *interest groups* in social network data
- BNP models based on completely random measures [**CRMs**] provide an extremely flexible modeling framework, amenable to efficient inference via exponential family conjugacy
- **Problem**: no general formalism for **hierarchical** BNP models based on CRMs is available

> **Our contribution:**
> 1. General formalism for hierarchical BNP models with CRM priors
> 2. Characterization of the number of unique traits in such models
> 3. Derivation of the posterior distribution of the latent trait frequencies
>   → Provide building blocks for efficient inference schemes in hierarchical BNP latent trait models
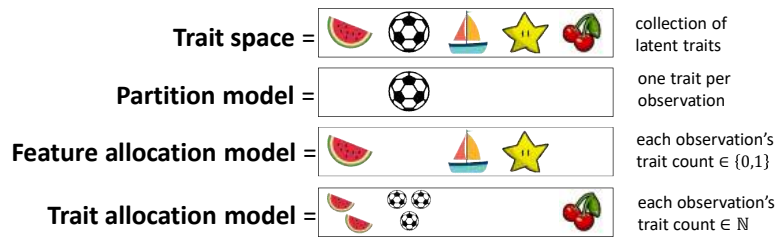
Background: CRMs

Most BNP priors are constructed using Poisson processes to obtain **CRMs**: these are random measures which couple (random) **rates** with (random) **traits**:
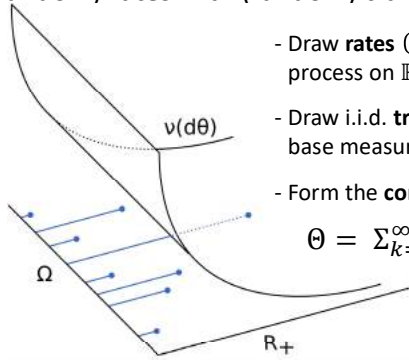
- Draw **rates** $(\theta_k)_{k\geq 1}$ from a Poisson process on $\mathbb{R}_+$ with mean measure $\nu(d\theta)$
- Draw i.i.d. **traits** $(\psi_k)_{k\geq 1}$ from a non-atomic base measure P_0 on the trait space Ω
- Form the **completely random measure:**

$$\Theta = \Sigma_{k=1}^{\infty}\theta_k\delta_{\psi_k} \sim CRM(\nu, P_0)$$

Hierarchical CRMs

Partially exchangeable setting: D populations share common traits $(\psi_k)_{k\geq 1}$ but rates $(\theta_{d,k})_{k\geq 1}$ differ across populations $d = 1, \dots, D$:

- text data from different corpora share same topics
- genomic data from different geographical regions share same structural variations
- social networks data from different platforms show same social circles

Draw shared base measure: $\Theta_0 \sim CRM(\nu, P_0)$
For every population d, draw conditionally i.i.d. rates

$$\theta_{d,k}|\theta_{0,k} \sim \rho(.\,|\theta_{0,k}, r_d)\, d\theta_{d,k} \longrightarrow \Theta_{1:D}|\Theta_0 \sim hCRM\left(\rho(\cdot\,|\theta_{0,k}, r_d)\right)_{d,k}$$

$$\Theta_0 \sim CRM(\nu, P_0)$$

$$\Theta_1|\Theta_0 \qquad \Theta_D|\Theta_0$$

Hierarchical trait allocation models

Trait space =		collection of latent traits
Partition model =		one trait per observation
Feature allocation model =		each observation's trait count ∈ {0,1}
Trait allocation model =		each observation's trait count ∈ ℕ

Describe **any of these allocation models** via integer random measure:

$$X \sim LP(h, \Theta) = \Sigma_{k=1}^{\infty} x_k\delta_{\psi_k}, \text{ i.e. } x_k|\theta_k \sim h(\theta_k)$$
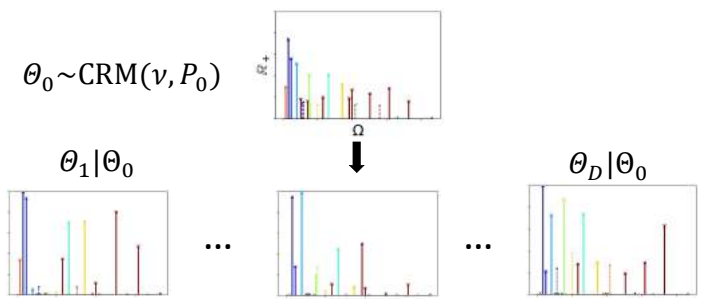
Hierarchical trait allocation model:
$$\begin{aligned}\Theta_0 &\sim CRM(\nu, P_0)\\ \Theta_{1:D}|\Theta_0 &\sim hCRM\left(\rho(\cdot\,|\theta_{0,k}, r_d)_{d,k}\right)\\ X_{n,d}|\Theta_d &\sim LP(h, \Theta)\end{aligned}$$

Theoretical results

Theorems 1 and 2 provide basic building blocks for MCMC posterior inference schemes in hierarchical BNP trait models. As above, X is the allocation of data points to traits. In single-level models

→ The marginal distribution $p(X)$ is often useful for samplers that integrate out the trait frequencies (cf. the Chinese restaurant process)

→ The conditional distribution $p(\Theta|X)$ is often useful for samplers that instantiate the trait frequencies

> **Theorem 1** (sketch): For any **hierarchical** CRM model, and any **hierarchical** allocation X^*, we characterize $p(X^*)$.

> **Theorem 2** (sketch): We obtain the posterior distribution of the hierarchical CRM $\Theta_{1:D}|\Theta_0, X^*$, as well as the distribution of $\Theta_0|X^*$.

Future work will focus on developing efficient algorithms for posterior inference in hierarchical trait allocation models