Streaming Inference for Infinite Non-Stationary Clustering

Rylan Schaeffer [1,2] Gabrielle Kaili-May Liu [2] Yilun Du [3] Scott Linderman [4] Ila Rani Fiete [2]

[1]Stanford Computer Science [2]MIT Brain and Cognitive Sciences [3]MIT EECS [4]Stanford Statistics

Introduction

Biological intelligence operates in a radically different data regime than (most) artificial intelligence. In particular, biological intelligence must contend with data that are often:

(i) Unsupervised (ii) Streaming (iii) Non-Stationary

In this data regime, what is the right way to approach learning? Here, we attack the unsupervised problem of clustering on streaming, non-stationary data. Specifically, we:

1. Define the **Dynamical Chinese Restaurant Process** (Dynamical CRP), a novel, non-exchangeable distribution over partitions of a set.
2. Show how the Dynamical CRP provides a non-stationary prior over cluster assignments.
3. Construct an efficient streaming variational inference algorithm using the Dynamical CRP.
4. Demonstrate that our streaming clustering algorithm can be applied on diverse synthetic and real data with Gaussian and non-Gaussian likelihoods to achieve comparable or better performance than many common baselines.

Desiderata

An ideal algorithm for clustering non-stationary streaming data should:

- Be able to create an "infinite" number of clusters (albeit upper bounded by the total number of observations).
- Not assume that the observations are i.i.d., exchangeable or consistent.
- Leverage temporal information based on when observations are received.
- Admit an efficient streaming implementation.

Background: Clustering

- Observations: $o_{1:N}$ occurring at known times $t_{1:N}$
- Cluster assignments: $c_{1:N}$, with $c_n \in \{1, 2, ...\}$
- Cluster parameters: $\{\phi_c\}_{c=1}^{C}$, e.g. cluster means and covariances
- One common prior for cluster assignments is the **Chinese Restaurant Process** (CRP) [1]:

$$p^{CRP}(c_n = c | c_{<n}, \alpha) \propto \begin{cases} \sum_{n'<n} \mathbb{I}(c_{n'} = c) & \text{if } 1 \leq c \leq C_{n-1} \overset{\text{def}}{=} \max(c_1, ..., c_{n-1}) \\ \alpha & \text{if } c = C_{n-1} + 1 \end{cases} \quad (1)$$

- The CRP is ill-suited to streaming data because the CRP's conditional form requires knowing the entire history of cluster assignments, but the CRP can be adapted for streaming data [2] by rewriting the CRP in a recursive form:

$$p^{CRP}(c_n = c | \alpha) \propto \sum_{n'<n} p(c_{n'} = c | \alpha) + \alpha \, p(C_{n-1} = c - 1) \quad (2)$$

- The CRP has two undesirable properties for non-stationary data: it is (1) *exchangeable*, i.e. the order of the data does not matter, and (2) *consistent*, i.e. marginalizing out an observation is identical to the observation never having existed.
- Zhu, Ghahramani and Lafferty 2005 [3] defined the time-sensitive CRP:

$$p^{tsCRP}(c_n = c | c_{<n}, \alpha) \propto \begin{cases} \sum_{n'<n} e^{(t_n - t_{n'})/\tau} \, \mathbb{I}(c_{n'} = c) & \text{if } 1 \leq c \leq C_{n-1} \\ \alpha & \text{if } c = C_{n-1} + 1 \end{cases} \quad (3)$$

- The time-sensitive CRP narrowly restricts how time can play a role in affecting cluster assignments.

Dynamical Chinese Restaurant Process

- The sufficient statistics of the CRP are the "table occupancies" $N_c(t) \overset{\text{def}}{=} \sum_{n' \leq n} \mathbb{I}(c_n = c)$
- Idea: Embed the table occupancies in a linear dynamical system to evolve endogenously. The **Dynamical CRP** thereby gains rich time-dependent priors for cluster assignments.
- Let $\mathcal{H}$ be a Hilbert space and $\tilde{N}(t) \in \mathcal{H}$ contain both the "pseudo" table occupancies $\{N_c(t)\}$ and any desired higher-order temporal derivatives. Fix a linear dynamical system $\ell : \tilde{N} \rightarrow \tilde{N}$ and increment the c_n-th table N_{c_n} at time t_n by 1. The Dynamical CRP $D\text{-}CRP(\ell, \alpha)$, is defined as the conditional distribution

$$p^{D\text{-}CRP}(c_n = c | c_{<n}, t_{\leq n}, \ell, \alpha) \propto \begin{cases} N_c(t_n) & \text{if } 1 \leq c \leq C_{n-1} \\ \alpha & \text{if } c = C_{n-1} + 1 \end{cases} \quad (4)$$

- By design, the Dynamical CRP also has a recursive form for streaming data:

$$p^{D\text{-}CRP}(c_n = c | \ell, \alpha) \propto \sum_{n'<n} \ell(p(c_{n'} = c | \ell, \alpha), t_{n'}, t_n) + \alpha \, p(C_{n-1} = c - 1) \quad (5)$$

- For inference, we use an (approximate) filtering prior to construct an (approximate) filtering evidence lower bound that we maximize:

$$\mathbb{E}_{q(c_n, \{\phi\} | o_{\leq n}; \theta_n)}[\log p(o_n | c_n, \{\phi\}, o_{<n}) + \log q(c_n, \{\phi\} | o_{<n})] + H[q(c_n, \{\phi\} | o_{\leq n})] \quad (6)$$

Example Dynamics

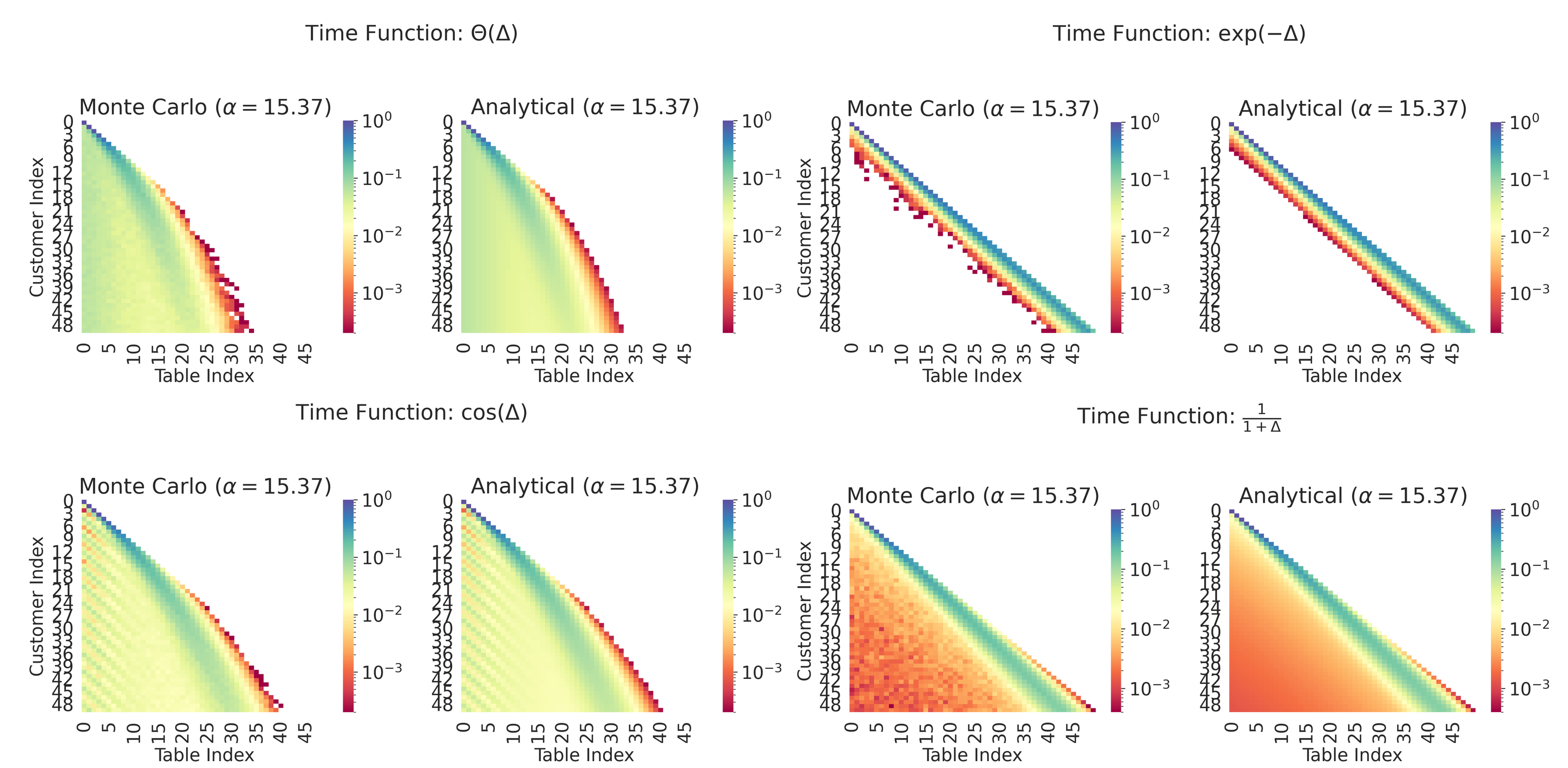

Figure 1. Cluster assignment priors under the Dynamical CRP with 4 dynamics: stationary, exponential, oscillatory and hyperbolic. $\Delta \overset{\text{def}}{=} t_n - t_{n-1}$ is elapsed time. The Dynamical CRP can produce the CRP, the time-sensitive CRP, or new priors over cluster assignments including oscillatory (Time Function: $\cos(\Delta)$) and hyperbolic (Time Function: $1/(1+\Delta)$). Columns 1 & 3 are Monte Carlo samples; Columns 2 & 4 are our analytical recursion (Eqn. 5).

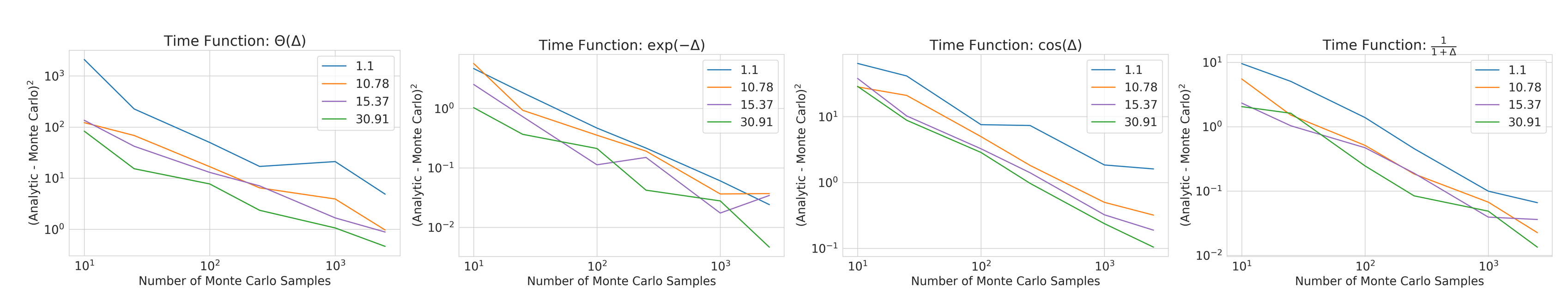

Figure 2. **Mean-Squared Error between analytical expression for $p(c_n | \ell, \alpha)$ and a Monte Carlo estimate.** Over a wide range of α values, the mean-squared error between our analytical expression and Monte Carlo estimates falls approximately as a power law, showing the exactness of Eqn. 5.

Experimental Results

We have begun testing D-CRP on synthetic and real datasets with Gaussian and non-Gaussian likelihoods. We are looking for feedback on benchmark tasks!

Synthetic Mixture of Gaussians

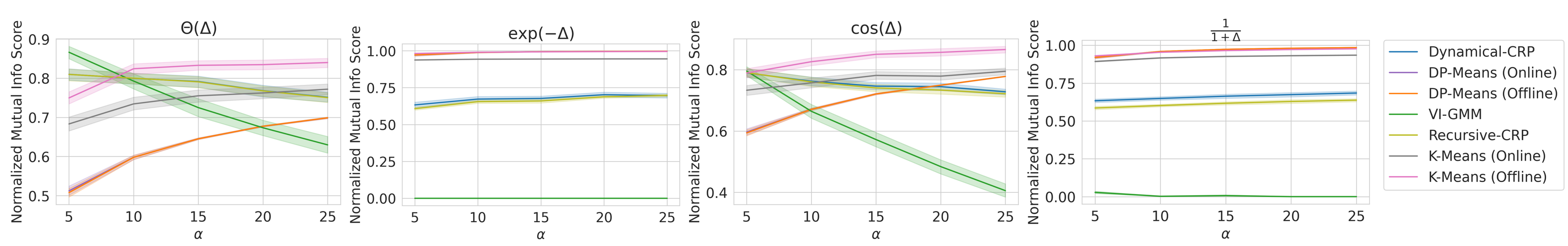

Figure 3. Normalized mutual information between true cluster assignments and inferred cluster assignments.

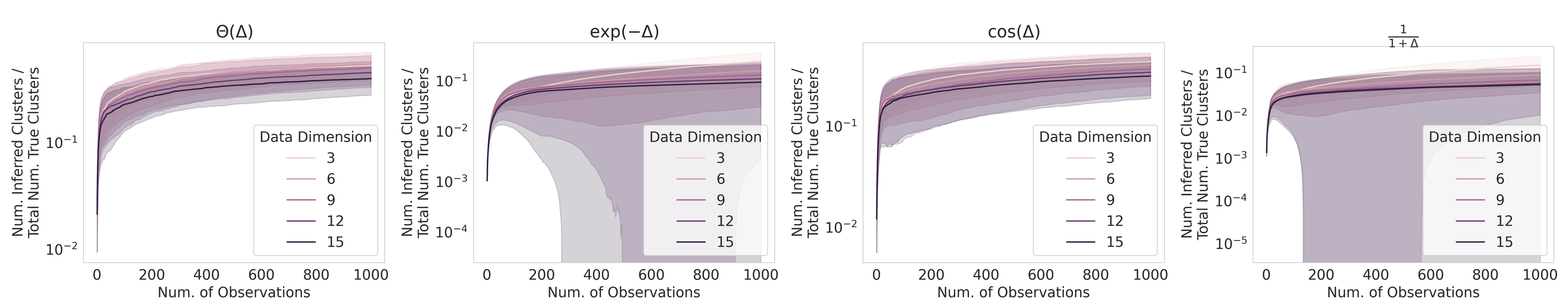

Figure 4. Dynamical CRP creates clusters over time, as necessitated by incoming data.

Synthetic Mixture of von Mises-Fishers

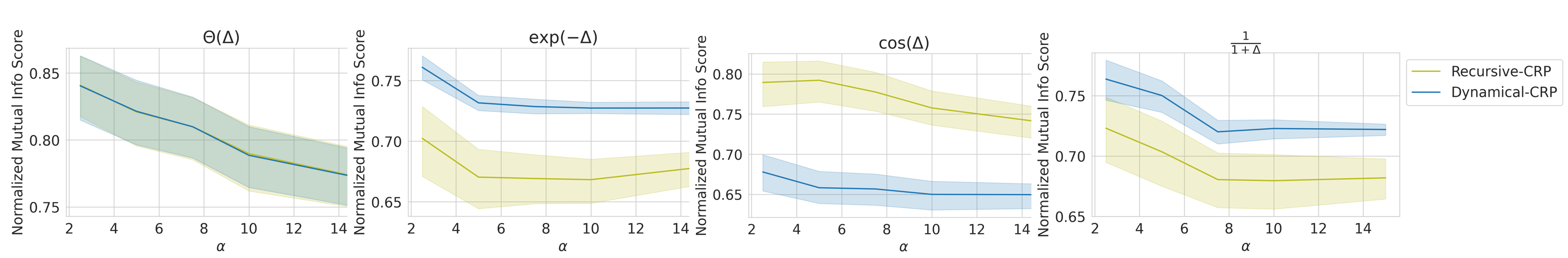

Figure 5. Normalized mutual information between true cluster assignments and inferred cluster assignments.

Room Clustering for Simultaneous Localization and Mapping

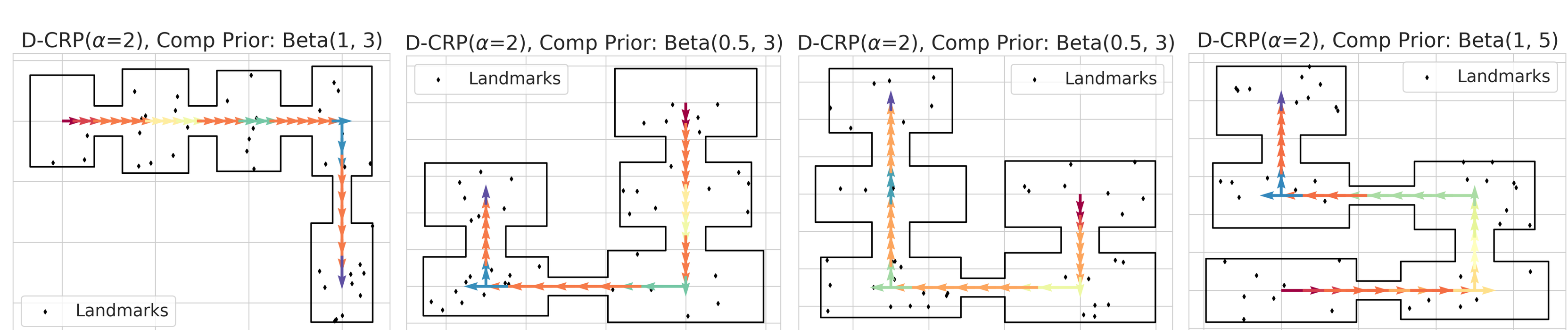

Figure 6. **Clusters inferred by Dynamical CRP in a 2D spatial navigation task.** Each color in each environment represents a unique cluster, inferred from visible landmarks (black diamonds) encountered along a single trajectory. The Dynamical CRP aggregates visually-distinguishable rooms (various colors) into distinct clusters and visually-identical hallways into the same cluster (orange).

References

[1] C. E. Antoniak. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *Annals of Statistics*, 2, 1974.

[2] R. Schaeffer, B. Bordelon, M. Khona, W. Pan, and I. R. Fiete. Efficient Online Inference for Nonparametric Mixture Models. *Uncertainty in Artificial Intelligence*, 2021.

[3] X. Zhu, Z. Ghahramani, and J. Lafferty. Time-Sensitive Dirichlet Process Mixture Models. 2005.