Dirichlet Process Mixtures + Chinese Restaurant Processes

A Bayesian Nonparametric Approach to Clustering

Lia Ran

June 12, 2024

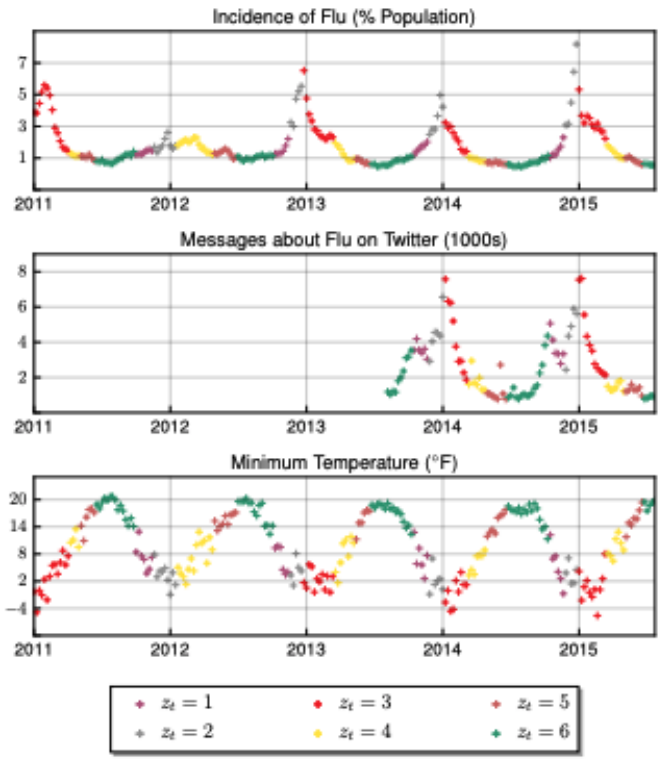
Table of Contents

Overview

- How many classes do we want to use in a mixture model?
- Bayesian nonparametric models fit a single model that can adapt its complexity to the data, **without specifying the number of clusters in advance**
- The hidden structure is determined as part of analyzing the data, or 'reversing' the generative model.
- These slides cover mathematical methodology behind Dirichlet Process models, as well as outline the main paper reviewed: **Temporally-Reweighted Chinese Restaurant Process**

Overview: TRCRP

- **Pointwise Clustering:** Find areas of similar behavior in a single time series.
- **Outer Clustering:** Find dependent time series by matching pointwise segmentation.



Bayesian Statistics

Bayesian Statistics

Broadly, Bayesian inference is used to update the probability for a hypothesis, or prior belief, as more data is observed.

- **Prior**: The belief about a parameter before any data is observed, denoted by $p(\theta)$.
- **Likelihood**: The probability of observing the data given a parameter value, θ, which is expressed as $p(x|\theta)$. It quantifies the "fit" of the data under different parameter values.
- **Posterior**: The updated belief about the parameter θ after observing data x. It combines the prior and the likelihood, calculated as $p(\theta|x) \propto p(x|\theta)p(\theta)$.

- **Predictive Distribution:** We predict the distribution of a new data point $\tilde{x}$ by integrating over the posterior, resulting in $p(\tilde{x}|x) = \int p(\tilde{x}|\theta)p(\theta|x)d\theta$.
- The integral in the predictive distribution *marginalizes* over θ, effectively summing over all possible parameter values to give a prediction that accounts for parameter uncertainty.
- By integrating θ out, we obtain a direct prediction for $\tilde{x}$, without having to commit to a specific value of θ.

Conjugate Prior

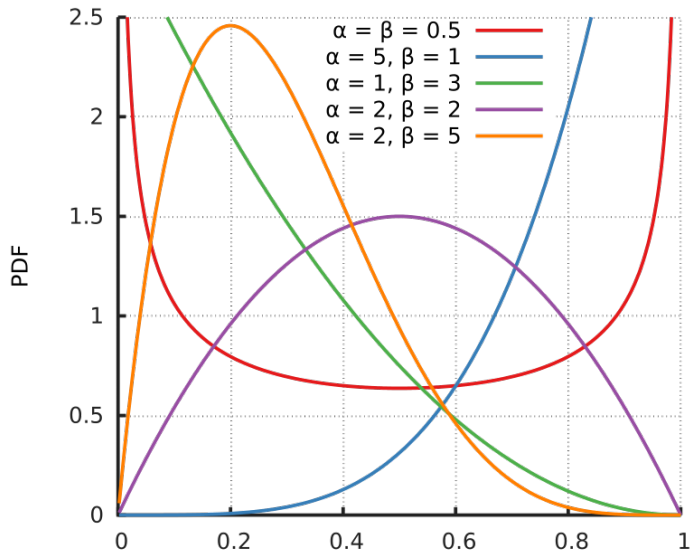
- **Conjugate Prior:** Prior distribution $p(\theta)$ and posterior $p(\theta|x)$ are in the same distribution family.
- **Posterior Formula:**

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta')p(\theta')d\theta'} \tag{1}$$

- With a conjugate prior, we can obtain a closed form expression for the posterior without the need for numerical integration.

Example: Beta-Binomial

- The **Beta distribution** Beta(α, β) represents distributions over [0, 1]
- Popular choice to model probabilities of random variables
- (α, β) can be thought of as shape parameters that affect the skew of the distribution



$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Example: Beta-Binomial

$$
\begin{aligned}
p &\sim \text{Beta}(\alpha, \beta) \\
X|p, n &\sim \text{Binomial}(n, p) \\
p|X &\sim \text{Beta}(\alpha + k, \beta + n - k)
\end{aligned}
\tag{2}
$$

- Say we are rolling an unfair die, and we want the probability that we will roll face k.
- We can model the probability of success p as a random variable.
- Conjugate prior gives us clean expression for the posterior.

Bayesian Inference

- **Objective**: Generate samples $p(\theta|x)$ using a Markov chain.
- **Strategy**:
  - Construct Markov chain with stationary distribution $p(\theta|x)$.
  - Apply an acceptance criterion to decide on the transitions, ensuring the chain converges to the target distribution.
- **Key Properties:**
  - Ergodicity, detailed balance.
- We can use empirical methods for summarizing the parameter's posterior distribution; ex. mean, confidence intervals, etc.

Gibbs Sampling

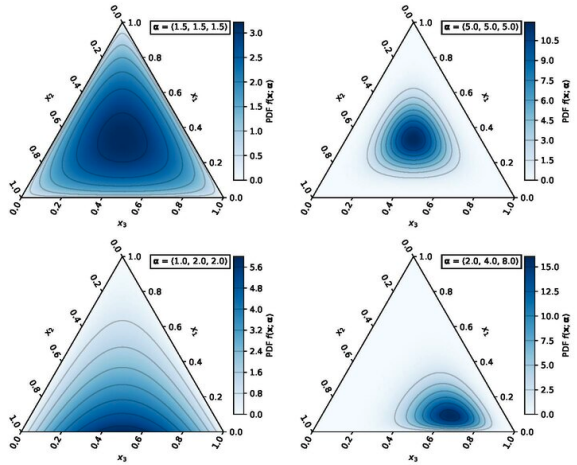
Gibbs Sampling constructs a Markov chain by sequentially sampling each variable from its conditional distribution, given the current values of all other variables.

- For a generative model $x \sim p(x)$:
  - Start with an initial set of values for x_i.
  - Iteratively, for each variable x_i in x:
    - Sample $x_i^{'}$ from $p(x_i|x_{-i})$, where x_{-i} represents all variables in x except x_i.
    - Update x_i to $x_i^{'}$ if 'acceptable'.
  - Continue the distribution of samples converges to the stationary distribution $p(x)$ (Checkable conditions).
- These samples can be used to approximate the full joint distribution $p(x)$, allowing for inference.

Dirichlet Distribution

Dirichlet Distribution

The Dirichlet Distribution $\pi \sim \text{Dir}(\alpha)$ represents distributions over a simplex (e.g., $\pi_i \in (0, 1), \sum \pi_i = 1$)



- $\pi = (\pi_1, \ldots, \pi_K)$ represent a probability vector on the $(K-1)$-simplex, where each π_i is the probability of the i-th outcome.
- **Domain:** The domain of the Dirichlet distribution is the set of K-dimensional discrete distributions, making it an effective model for the distribution of probabilities across K categories.
- **Note:** This is the multivariate extension of the Beta distribution ($K = 2$)!

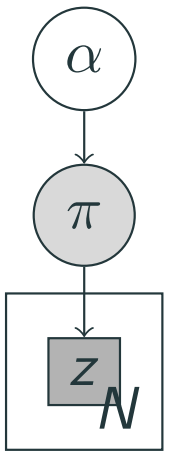
Dirichlet Distribution

- Think of Dir(α) as a bag containing an infinite $K = 6$-sided unfair die, with varying probabilities.
- $\pi \sim Dir(\alpha)$ represents pulling out a single die from the bag. π_i will represent the set of probabilities associated with each face of this specific die
- **Properties:**
  - *Marginalization:* Any subset of the components $(\pi_1, \ldots, \pi_K)$ follows a Dirichlet distribution.
    - i.e., regroup probabilities $(\pi_1, \pi_2 + \pi_3) \sim Dir(\alpha_1, \alpha_2 + \alpha_3)$
  - *Expectation:* The expected value of π_i under the Dirichlet distribution is $\mathbb{E}[\pi_i] = \frac{\alpha_i}{\sum_{j=1}^{K} \alpha_j}$.
    - ex. $\mathbb{E}[\pi_i] = \frac{1}{6}$ for a fair-sided die

Dirichlet-Multinomial Conjugacy

Multinomial is multivariate extension of beta-binomial; now we can model the probabilities of all outcomes.

$$\begin{aligned} \pi &\sim \text{Dir}(\alpha) \\ z &\sim \text{Mult}(\pi, n) \\ \pi|z &\sim \text{Dir}(z + \alpha) \end{aligned} \tag{3}$$



- $z \sim \text{Mult}(\pi, n)$ is a K dimensional vector that counts occurrences of all K outcomes over n trials
- π_k contains the mixture proportion for category z_k

Finite Mixture Model

- Pick one of K clusters with probability $\pi = (\pi_1, .., \pi_K)$
- Generate a data point from a cluster-specified probability distribution $x \sim p(x|\phi_k)$
- This yields $p(x|\phi, \theta) = \sum_{k=1}^{K} \pi_k p(x|\phi_k)$
- Can also capture the cluster assignment with a latent multinomial variable z:

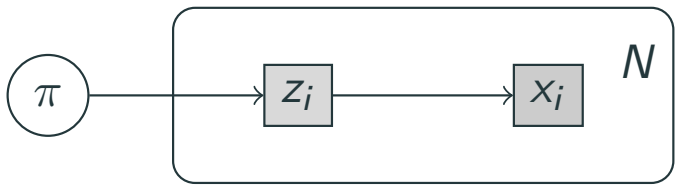
$$\begin{aligned} p(x|\phi, \pi) &= \sum_{k=1}^{K} p(z^k = 1|\pi) p(x|z^k = 1, \phi) \\ &= \sum_{k=1}^{K} \pi_k p(x|\phi_k) \end{aligned} \tag{4}$$

Finite Mixture Models

Finite Mixture Model

In a mixture model, each data point x_i are drawn from one of K source distributions with corresponding mixing proportions π_k

- The (unknown) values of z_i specify which distribution generates x_i, identifying cluster assignments
- What are the cluster assignments z? What are the mixing proportions π?
- We want to compute the posterior $P(z, \pi|x)$ to reveal the hidden structure of the data



Finite Mixture Model: Generative Model

- Let parameters ϕ_k define mixture component k. (ex; $\phi_k = (\theta_k, \sigma_k^2)$
- Let atoms δ_{ϕ_k} symbolize the concentration of probability mass at specific points ϕ_k
- The underlying measure $G = \sum_{k=1}^{K} \pi_k \delta_{\phi_k}$ creates a discrete probability distribution over the cluster parameters ϕ_k.
    - Sampling $\theta_i \sim G$ selects ϕ_k according to the mixture proportions π

$$\begin{aligned} \theta_i &\sim G \\ x_i &\sim p(\cdot|\theta_i) \end{aligned} \tag{5}$$

Bayesian Finite Mixture Model

Place priors on ϕ and π.

$$
\begin{aligned}
\phi_k &\sim G_0 \\
\pi_k &\sim \text{Dir}(\alpha_0/K, \ldots, \alpha_0/K) \\
G &= \sum_{k=1}^{K} \pi_k \, \delta_{\phi_k} \\
\theta_i &\sim G \\
x_i &\sim p(\cdot \mid \theta_i)
\end{aligned}
$$

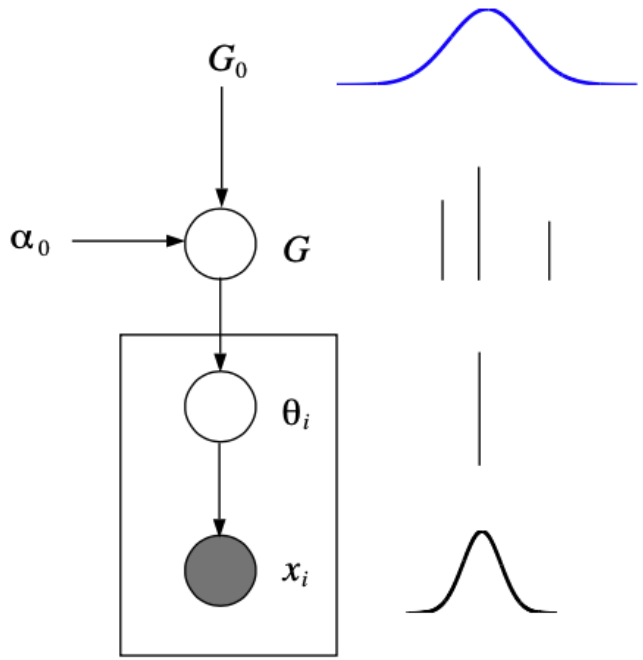


Figure 1: G is now random and captures the probabilistic structure of the mixture model

How do we find K?

Dirichlet Processes

Dirichlet Processes

Dirichlet Process

- The **Dirichlet Process** is the infinite-dimensional extension of the Dirichlet distribution.
- Consider $\pi \sim \lim_{K \to \infty} \text{Dir}(\frac{\alpha}{K}, .., \frac{\alpha}{K})$
- If $\theta_k \sim G_0$, $G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}$ is almost surely discrete but now infinite.
- Used as a prior on the mixing proportions $\pi \sim DP(\alpha, G_0)$, supporting an infinite number of mixture components.

Dirichlet Process

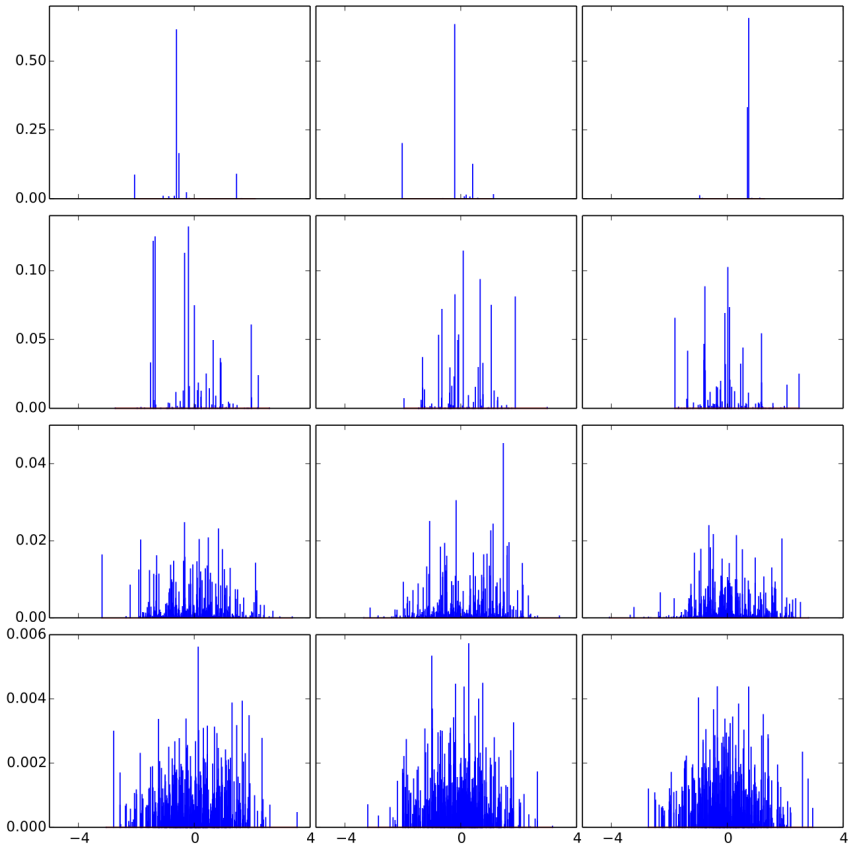


Figure 2: Draws from the $DP(N(0,1),\alpha)$. The four rows use different α (top to bottom: 1, 10, 100 and 1000) and each row contains three repetitions of the same experiment

Dirichlet Process: Properties

- **Expectation:** $G \sim DP(\alpha, G_0) \Rightarrow \mathbb{E}[G] = G_0$.
- **Exchangeability:** A sequence of random variables drawn from G is exchangeable, meaning the joint distribution is independent of the order of observations
- **Expected Number of Clusters:** After n observations, $\mathbb{E}[k_n] \approx \alpha \log(1 + \frac{n}{\alpha})$.
- **Posterior Predictive Distribution:** Integrate out the mixture weights π_k:

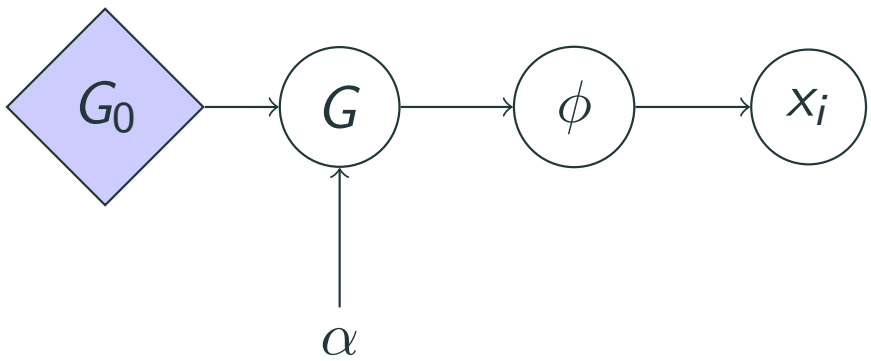
$$p(\tilde{\theta}|\theta) = \frac{\alpha}{\alpha + n} G_0(\tilde{\theta}) + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{\theta_i}(\tilde{\theta}) \tag{6}$$

Dirichlet Process: Pólya Urn Scheme

- The **Pólya Urn Scheme** illustrates the Dirichlet Process $G \sim \text{DP}(\alpha, G_0)$
- **Key Mechanisms:**
  - Starts with a special ball. Drawing this introduces a *new color* (new parameters $\phi_k \sim G_0$) with probability $\frac{\alpha}{n+\alpha}$.
  - Drawing an existing color k reinforces that color, adding another ball of the same color with probability $\frac{n_k}{n+\alpha}$
    - Indirectly increases the weight of δ_{ϕ_k}, affecting the mixture proportion π_k.
    - Enhances the probability of selecting the same cluster parameter ϕ_k in future draws.
- $G \sim \text{DP}(\alpha, G_0)$ is the probability measure over the each color in the urn.

Urn Scheme: Graphical Representation

$$
\begin{aligned}
G &:= \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \\
\phi_n &\sim G \\
x_n &\sim f(\phi_n)
\end{aligned}
\tag{7}
$$



Dirichlet Process: Chinese Restaurant Process

What cluster parameters θ_k will generate observation x_i?

- By exchangability property, we can consider customer i as the last customer without changing the joint probability
- **Chinese Restaurant Process (CRP):**
  - Suppose there are K distinct tables. For $i = 1, \ldots, n$, denote the table assignments of customer x_i by z_i.
  - The probability that a customer sits at an existing table $k < K$ or at a new table $k = K + 1$ is:

$$\mathbb{P}[z_{N+1} = k | z_{-N}] \propto \begin{cases} n_k & \text{if } k \leq K \\ \alpha & \text{if } k = K + 1 \end{cases} \tag{8}$$

Distribution of one variable updated while holding all others fixed... Gibbs sampling!

CRP: Visual Representation

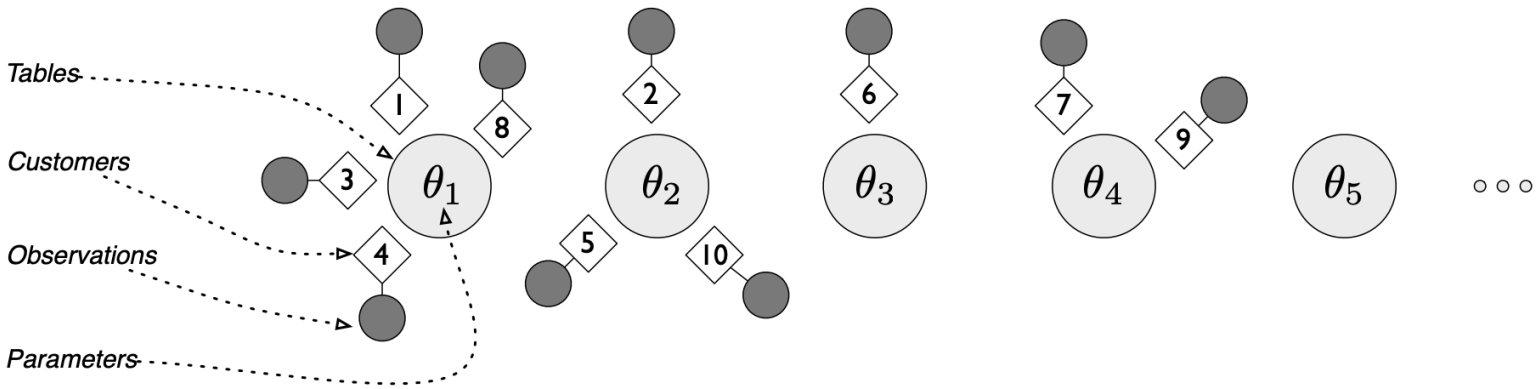


Figure 2: **The Chinese restaurant process**. The generative process of the CRP, where numbered diamonds represent customers, attached to their corresponding observations (shaded circles). The large circles represent tables (clusters) in the CRP and their associated parameters (θ). Note that technically the parameter values $\{\theta\}$ are not part of the CRP *per se*, but rather belong to the full mixture model.

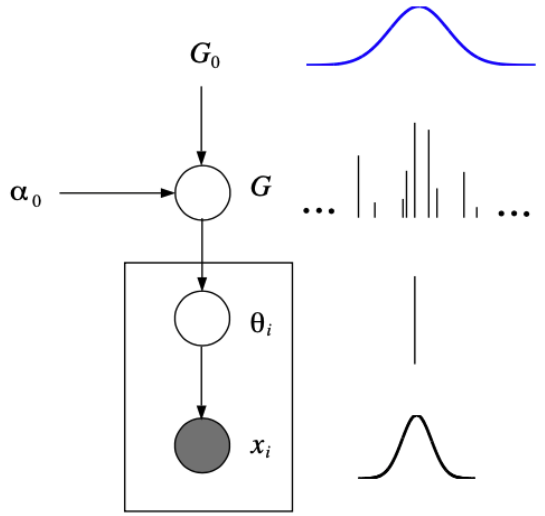
- CRP is a distribution over all partitions of $[N]$. The figure shows a CRP of $N = 10$: $\{\{1, 3, 4, 8\}, \{2, 5, 10\}, \{6\}, \{7, 9\}\}$

Infinite Mixture Models

Infinite Mixture Models

DP Mixture Models

- In DP Mixture models, the Dirichlet Process is used as a prior on the cluster parameters θ_i, and we complete the model by adding a likelihood on the data x_i.

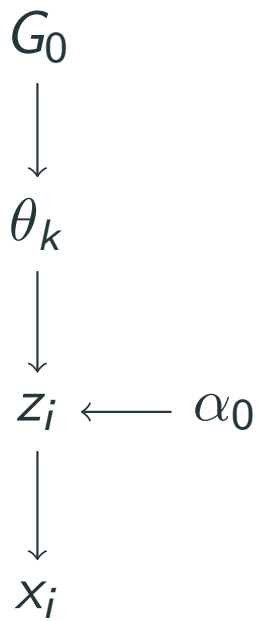


$$
\begin{aligned}
G &\sim \mathrm{DP}(\alpha_0 G_0) \\
\theta_i \,|\, G &\sim G \qquad\qquad i \in 1, \ldots, n \\
x_i \,|\, \theta_i &\sim F(x_i \,|\, \theta_i) \qquad i \in 1, \ldots, n
\end{aligned}
$$

Infinite Mixture Model: CRP Representation

- Model latent variables z_i of cluster assignments; z_i will denote the cluster id and θ_i will denote the (unique) cluster parameters
- With this scheme, a new θ is sampled only when we need to create a new cluster.

$$\begin{aligned} \theta_{1,\ldots,\infty} &\sim G_0 \\ z_{1,\ldots,n} &\sim CRP(\alpha_0) \\ x_{1,\ldots,n} &\sim F(\theta_{z_j}) \end{aligned} \tag{9}$$



MCMC

We want to uncover the cluster assignments $z_{1:N}$ with the posterior $P(z|x_{1:n})$

- We want the obtain posterior by building a Markov chain under certain conditions to ensure convergence (Neal, 1993)
- The key is to take advantage of *exchangeability*; we can always swap customer i with the final customer

Inference

MCMC: Collapsed Gibbs Sampling

- **For each data point:**
  - Temporarily remove x_i from its cluster
  - Re-evaluate the probabilities for x_i belonging to each cluster:

$$P(z_i = k|\mathbf{z}_{-i}) \propto \begin{cases} n_{-i,k} & \text{if } k \text{ is an existing cluster} \\ \alpha & \text{if } k \text{ is a new cluster} \end{cases}$$

  - Reassign x_i to a cluster k based on these updated probabilities from the mixture model likelihood:

$$P(x_i|\theta_{z_i}, \mathbf{x}_{-i}, \mathbf{z}_{-i}) \propto F(x_i|\theta_{z_i})$$

- **Parameter Resampling:**
  - For each cluster k, resample the cluster parameter θ_k conditioning on all data points $\mathbf{x}_k$ assigned to cluster k

Note: Requires conjugacy, only moves one data point at a time.

Inference on α_0: Empirical Bayes

- Recall that our prior on clustering is defined by CRP(α_0), where α_0 controls the expected number of clusters.
- *Empirial Bayes* fixes the hyperparameters that maximize the (marginal) likelihood.

(McAuliffe, Blei, Jordan, 2005)

Extensions

Hierarchical Bayesian Modeling

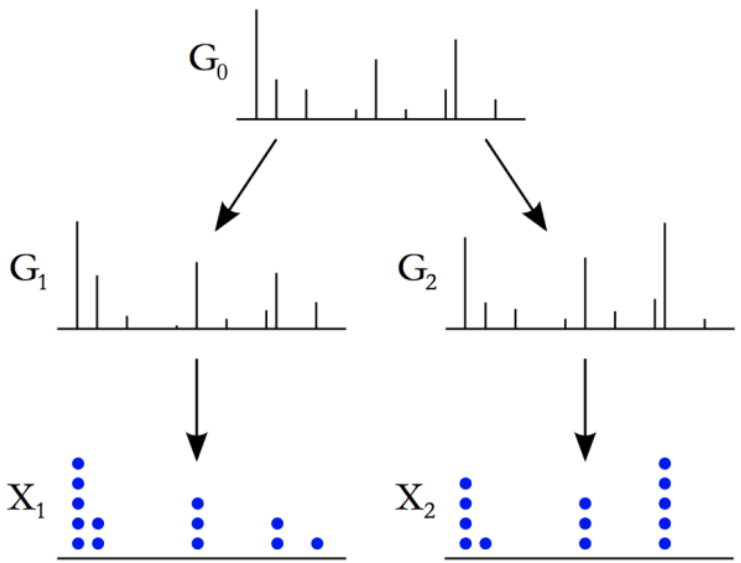
- View the parameters θ_i as random variables influenced by prior distributions.
- This approach allows for incorporating uncertainty about parameter values directly into the model.

Hierarchical Dirichlet Process

Now, we sample the base measure from a Dirichlet Process:

$$G_0 \sim \mathrm{DP}(\gamma, H)$$

$$G_M \sim DP(\alpha, G_0)$$



Temporally Reweighted CRP

Outline

1. Introduce recurrent version of Chinese restaurant process mixture to capture temporal dependencies.
2. Add a hierarchical prior to discover groups of time series whose underlying dynamics are modeled jointly.
3. Time series extension to nonparametric Bayesian regression models for cross-sectional data.
4. TRCRP builds empirical methods to detect patterns in the data, without customizing the structural theory of temporal dynamics.

Outline

- **Model Extension:** Adaptation of nonparametric Bayes for time series analysis without predefined temporal theories.
  - **Temporal CRP:** Introduction of a recurrent Chinese Restaurant Process to handle temporal dependencies.
  - **Hierarchical Priors:** Use of hierarchical priors for modeling groups of time series with shared dynamics.

TR-CRP: One Time Series

- **Temporal Adjustments:**
  - We incorporate temporal dependency by recording D_{tk}, which stores the past p lagged observations previously assigned to cluster k.
- **Cohesion Measure:**
  - The cohesion function G assesses how well new lagged observations fit with historical patterns stored in D_{tk}.
  - G uses historical cluster characteristics to evaluate current observations, ensuring continuity and temporal coherence.
- **Formulation:**

$$G(x_{t-p:t-1}; D_{tk}, \lambda_G) = \prod_{i=1}^{p} G_i(x_{t-i}; D_{tki}, \lambda_{Gi})$$

Cohesion Function

At step t, sample a cluster assignment z_t, whose probability of a cluster assignment z_t joining cluster k is a product of (i) the original CRP probability and (ii) the cohesion term G.

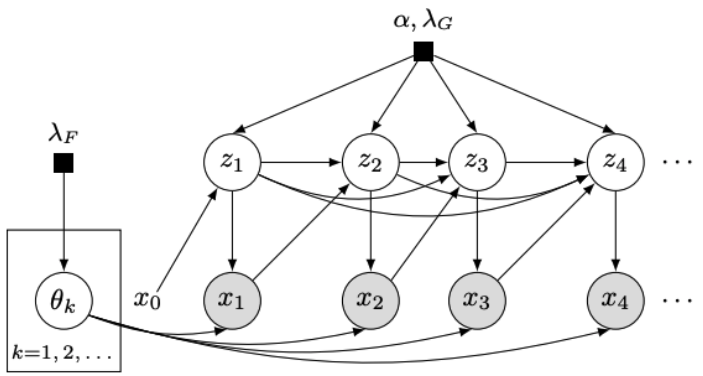


Figure 1: Graphical model for the TRCRP mixture in a single time series $\mathbf{x} = (x_1, x_2, \ldots)$ with lagged window size $p = 1$.

TR-CRP Model Specification

$$\begin{aligned} \theta_k &\sim G_0(\lambda_F), \\ z_j &\sim \mathrm{CRP}(\alpha) \cdot G(\cdot), \\ x_j | z_j, \theta_k &\sim F(\theta_{z_j}) \end{aligned} \tag{10}$$

- For a time series $(x_1, \ldots, x_n)$, points x_t are modeled as normally distributed within each cluster, with parameters $\theta_k = (\mu_k, \sigma_k^2)$.
- Priors for cluster parameters θ_k follow a Normal-InverseGamma distribution.
- The cohesion function G guide cluster assignments by considering the temporal compatibility of the historical data within each cluster.

Heuristics

- For a time series $(x_1, \ldots, x_n)$, points x_t are modeled as normally distributed within each cluster, with parameters $\theta_k = (\mu_k, \sigma_k^2)$.
- Priors for cluster parameters θ_k follow a Normal-InverseGamma distribution.
- G is defined as a product of p Student-T distributions whos parameters depend on lagged data D_{tk} in cluster k.

Multi-Variate TRCRP

- We generalize the mixture for one time series to handle a collection of N time series, assumed to be dependent
- All time series within a cluster share the same segmentation $z_{1:T}$ into various temporal regimes, although the parametric distributions $F(\cdot|\theta_k^n)$ vary
- G will now use lagged values of all N time series $\{x^n : n = 1, .., N\}$ in a given cluster:
-

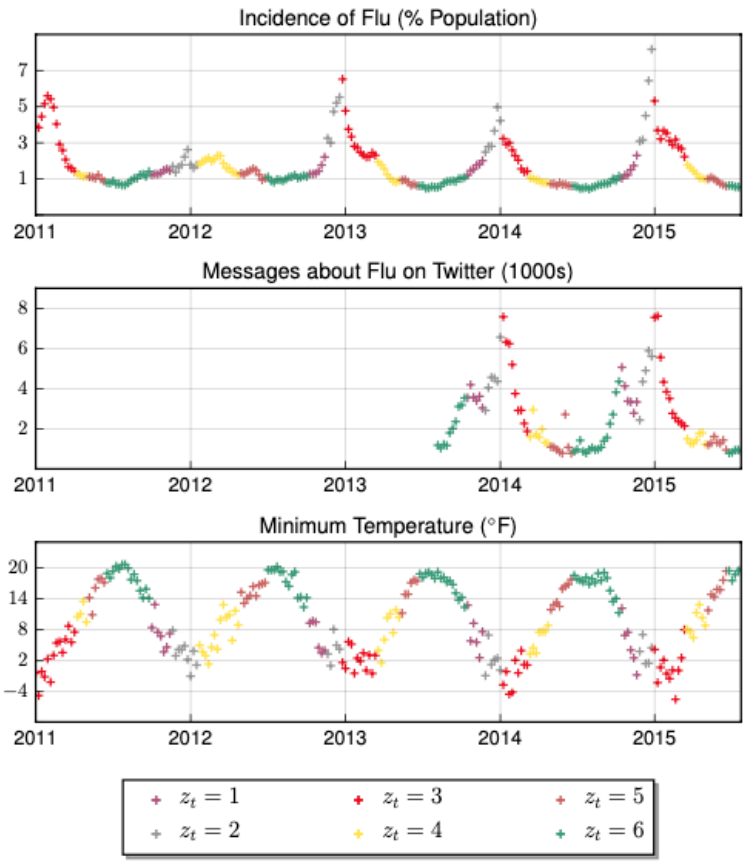
$$\begin{aligned} G &= \prod_{n=1}^{N} G(x^n_{t-p:t-1}; D^n_{tk}, \lambda^n_G) \\ &= \prod_{n=1}^{N} \prod_{i=1}^{p} G_i(x^n_{t-i}; D^n_{tki}, \lambda^n_{Gi}) \end{aligned} \tag{11}$$

- Q: How does combining across time series affect cluster probability weighting?

Generative Process for MV-TRCRP

1. Sample concentration parameter of CRP
   $\alpha \sim \text{Gamma}(1,1)$
2. Sample model hyperparameters $(n = 1, 2, \ldots, N)$
   $\lambda_G^n \sim H_G^n$
   $\lambda_F^n \sim H_F^n$
3. Sample distribution parameters of F $(n = 1, 2, \ldots, N)$
   $\theta_1^n, \theta_2^n, \ldots \overset{\text{iid}}{\sim} \pi_\Theta(\cdot|\lambda_F^n)$
4. Assume first p values are known $(n = 1, 2, \ldots, N)$
   $\mathbf{x}_{-p+1:0}^n := (x_{-p+1}^n, \ldots, x_0^n)$
5. Sample time series observations $(t = 1, 2, \ldots)$
   5.1 Sample temporal cluster assignment z_t
   $\Pr\left[z_t = k \mid \mathbf{z}_{1:t-1}, \mathbf{x}_{t-p:t-1}^{1:N}, \alpha, \lambda_G^{1:N}\right]$
   $\propto \text{CRP}(k|\alpha, \mathbf{z}_{1:t-1}) \prod_{n=1}^N G(\mathbf{x}_{t-p:t-1}^n; D_{tk}^n, \lambda_G^n)$
   where $D_{tk}^n := \left\{\mathbf{x}_{t'-p:t'-1}^n \mid z_{t'} = k, 1 \le t' < t\right\}$
   and $k = 1, \ldots, \max(\mathbf{z}_{1:t-1}) + 1$
   5.2 Sample data x_t^n $(n = 1, 2, \ldots, N)$
   $x_t^n \mid z_t, \{\theta_k^n\} \sim F(\cdot|\theta_{z_t}^n)$

(a) Generative process for the multivariate TRCRP mixture



(b) Discovering flu season dynamics with the method

Figure 3: In the final model, conjugacy of F and π_Θ mean we can marginalize $\{\theta_k^*\}$ out

Learning Dependence (Clusters) across Groups

- Introduce a hierarchical prior that allows the model to determine which subsets of the time series are probably well-describes by the joint TRCRPM model
- The prior nonparametrically partitions clusters using an 'outer' CRP

$$\begin{aligned} (c^1, .., c^N) &\sim CRP(\cdot|\alpha_0) \\ \{x^n : c^n = k\} &\sim TRCRP \end{aligned} \tag{12}$$

Posterior Inferences via MCMC

- Given window size p and initial observations $\{x^n_{-p+1:0} : n = 1, .., N\}$, the unnormalized posterior distribution of all latent variables given window p and beginning observations $\mathbf{x}^{1:N}_{-p+1:0}$ is:

$$
\begin{aligned}
&P\Big(\alpha_0, \mathbf{c}^{1:N},\ \alpha^{1:M}, \lambda_G^{1:N}, \lambda_F^{1:N}, \{\theta_j^n : 1 \le j \le K_{c^n}\}_{n=1}^N, \\
&\quad \mathbf{z}_{1:T}^{1:M}, \mathbf{x}_{1:T}^{1:N}\ ;\ \mathbf{x}_{-p+1:0}^{1:N}, p\Big) \\
&= \Gamma(\alpha_0; 1, 1)\mathrm{CRP}(\mathbf{c}^{1:N} \mid \alpha_0) \\
&\left(\prod_{n=1}^N H_G^n(\lambda_G^n)\right)\left(\prod_{n=1}^N H_F^n(\lambda_F^n)\right)\left(\prod_{n=1}^N \prod_{j=1}^{K_{c^n}} \pi_\Theta^n(\theta_j^n)\right) \\
&\prod_{m=1}^M \left(\Gamma(\alpha^m; 1, 1) \prod_{t=1}^T \left[b_t^m \mathrm{CRP}(z_t^m \mid \mathbf{z}_{1:t-1}^m, \alpha^m)\right.\right. \\
&\quad \left.\left.\prod_{n|c_n=m} G(\mathbf{x}_{t-p:t-1}^n; D_{tz_t^m}^n, \lambda_G^n) F(x_t^n \mid \theta_{z_t^m}^n)\right]\right), \qquad (6)
\end{aligned}
$$

Posterior Inference

- Creating a Markov chain that approached the desired distribution and repeatedly sampling from it, along with some acceptance/rejection rule for the sample.
- Time series cluster assignments $(c^n|\mathbf{c}^{1:N\backslash n}, ..)$ are transitioned by proposing to move x^n to either a new or existing cluster and computing the appropriate MH acceptance ratio for each case.
- Temporal regime assignments $(z_t^m|\mathbf{z}_{1:T\backslash t}...)$ are sampled using a form of Gibbs sampling

Predictive Inferences

Given approximate posterior samples $\{\hat{\xi^1}, .., \hat{\xi^S}\}$:

- **Forecasting:** to forecast over an h step horizon, draw a chain $\tilde{s}$ and use latent variables in $\xi^{\tilde{s}}$. For $t = T, .., T + h$:
  - Sample temporal cluster assignment z_t
  - Sample data x_t^n
- **Clustering** For two time series $\mathbf{x}^i, \mathbf{x}^k$, the posterior probability that they are dependent is the fraction of samples in which they are in the sample cluster

Inference

The Markov chain used for inference is a cycle over the following kernels (see 3.1 for more details):

1. Initialize by sampling from the prior Cycle kernel for inference for the outer CRP concentration α_0 and a cycle for inner concentration parameters $\{\alpha^m\}$ for the inner CRP
2. Infer over the Normal-Inverse Gamma hyperparameters λ_G, λ_F for each dimension: $\lambda \sim P(\lambda|x, z)$
3. Inner CRP: Update the pointwise categorization for a time series according to its conditioned distribution given other latent variables
4. Update the outer clustering

References and Further Reading i

- Introduction to Bayesian Nonparametrics
- Dirichlet Distribution / Finite Models, Dirichlet Process / Infinite Models
- Sampling Method for DPM
- Implementation: Implementation 1, Implementation 2
- TRCRPM paper