CPSC 535
Dirichlet Processes

FC

April 2007

Introduction

- Density estimation
- Set of data $\{\mathbf{z}_k\}_{k=1,\ldots,n}$ distributed from an unknown distribution F
- Example: Recession velocities of 82 galaxies

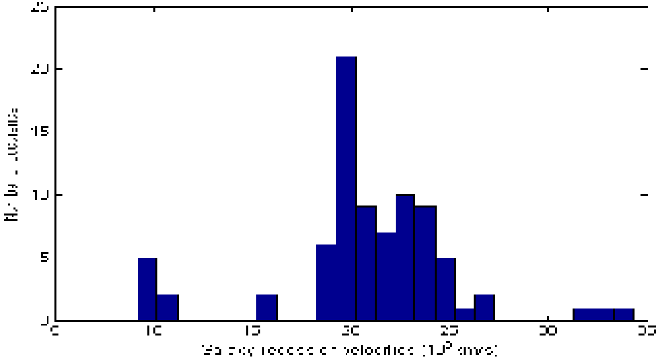
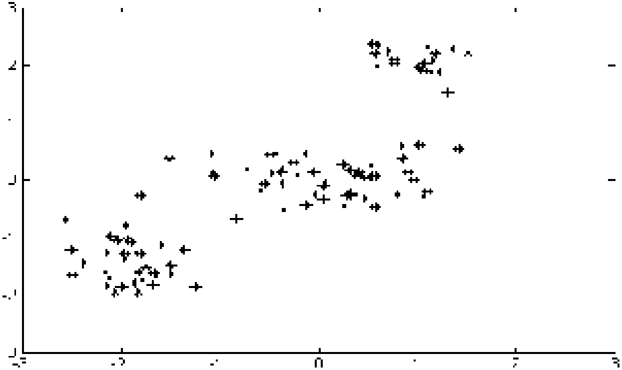


Figure: Histogram of velocities data

- Data clustering
- Set of data $\{\mathbf{z}_k\}_{k=1,\ldots,n}$ that we want to cluster into different groups and find each group centroid



Parametric prior

- Finite mixture

$$\mathbf{z}_k \sim \sum_{j=1}^{K} \pi_j f\left(\cdot|\mathbf{U}_j\right)$$

with $\sum_{j=1}^{K} \pi_j = 1$.

Parametric prior

- Finite mixture

$$\mathbf{z}_k \sim \sum_{j=1}^{K} \pi_j f\left(\cdot|\mathbf{U}_j\right)$$

with $\sum_{j=1}^{K} \pi_j = 1$.

- Equivalent formulation

$$\mathbf{z}_k \sim \int f(\cdot|\mathbf{U}) d\mathbb{G}(\mathbf{U})$$

where $\mathbb{G} = \sum_{j=1}^{K} \pi_j \delta_{U_j}$ is the mixing distribution

Parametric prior

- Finite mixture

$$\mathbf{z}_k \sim \sum_{j=1}^{K} \pi_j f\left(\cdot|\mathbf{U}_j\right)$$

  with $\sum_{j=1}^{K} \pi_j = 1$.

- Equivalent formulation

$$\mathbf{z}_k \sim \int f(\cdot|\mathbf{U}) d\mathbb{G}(\mathbf{U})$$

  where $\mathbb{G} = \sum_{j=1}^{K} \pi_j \delta_{U_j}$ is the mixing distribution

- Finite mixture of Gaussians

$$\mathbf{z}_k \sim \sum_{j=1}^{K} \pi_j \mathcal{N}(\mu_j, \sigma_j^2).$$

Parametric prior

- Prior distribution on the unknowns

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})$$

and for $j = 1, \ldots, K$

$$\mathbf{U}_j \overset{\text{i.i.d.}}{\sim} \mathbb{G}_0$$

Hierarchical model

- Prior distribution on the unknowns

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}).$$

- For $j = 1, \dots, K$

$$\mathbf{U}_j \overset{\text{i.i.d}}{\sim} \mathbb{G}_0$$

- For $k = 1, \dots, n$

$$\begin{aligned} c_k | \pi_{1:K} &\sim \mathcal{M}(\pi_{1:K}) \\ \mathbf{z}_k | \mathbf{U}_{c_k} &\sim f(\cdot | \mathbf{U}_{c_k}) \end{aligned}$$

Inference

- The mixing weights $\pi_{1:K}$ can be integrated out

$$\Pr(c_k = c | c_{1:k-1}) = \frac{n_c + \frac{\alpha}{K}}{k - 1 + \alpha}$$

Inference

- The mixing weights $\pi_{1:K}$ can be integrated out

$$\Pr(c_k = c | c_{1:k-1}) = \frac{n_c + \frac{\alpha}{K}}{k - 1 + \alpha}$$

- Polya urn interpretation

Proof

$$
\begin{aligned}
\Pr(c_k &= c|c_{1:k-1}) = \frac{\Pr(c_{1:k-1}, c_k = c)}{\Pr(c_{1:k-1})} \\
&= \frac{\int \pi_{c_1} \ldots \pi_{c_{k-1}} \pi_c \frac{\Gamma(\alpha)}{\Gamma(\frac{\alpha}{K})^K} \prod_{i=1}^K \pi_i^{\frac{\alpha}{K}-1} d\pi_{1:K}}{\int \pi_{c_1} \ldots \pi_{c_{k-1}} \frac{\Gamma(\alpha)}{\Gamma(\frac{\alpha}{K})^K} \prod_{i=1}^K \pi_i^{\frac{\alpha}{K}-1} d\pi_{1:K}} \\
&= \frac{\int \prod_{i=1}^K \pi_i^{n_i+\delta_i(c)+\frac{\alpha}{K}-1} d\pi_{1:K}}{\int \prod_{i=1}^K \pi_i^{n_i+\frac{\alpha}{K}-1} d\pi_{1:K}} \\
&= \frac{\frac{\prod_i \Gamma(\frac{\alpha}{K}+n_i+\delta_i(c))}{\Gamma(\alpha+k)}}{\frac{\prod_i \Gamma(\frac{\alpha}{K}+n_i)}{\Gamma(\alpha+k-1)}} \\
&= \frac{n_c + \frac{\alpha}{K}}{k-1+\alpha}
\end{aligned}
$$

MCMC

- Gibbs sampler to sample from $\pi\left(c_{1:n}, \mathbf{U}_{1:K} \,\middle|\, \mathbf{z}_{1:n}\right)$.

MCMC

- Gibbs sampler to sample from $\pi\left(c_{1:n}, \mathbf{U}_{1:K} \,\middle|\, \mathbf{z}_{1:n}\right)$.
  1. For $k = 1, \ldots, n$, Sample c_k with

$$\Pr(c_k = c | c_{-k}, \mathbf{z}_k, \mathbf{U}_{1:K}) \propto \frac{n_{-k,c} + \frac{\alpha}{K}}{n - 1 + \alpha} f(\mathbf{z}_k | \mathbf{U}_c)$$

MCMC

- Gibbs sampler to sample from $\pi\left(c_{1:n}, \mathbf{U}_{1:K} \,\middle|\, \mathbf{z}_{1:n}\right)$.
  1. For $k = 1, \ldots, n$, Sample c_k with

$$\Pr(c_k = c | c_{-k}, \mathbf{z}_k, \mathbf{U}_{1:K}) \propto \frac{n_{-k,c} + \frac{\alpha}{K}}{n - 1 + \alpha} f(\mathbf{z}_k | \mathbf{U}_c)$$

  2. For $j = 1, \ldots, K$, sample $\mathbf{U}_j | c_{1:n}, \mathbf{z}_{1:n}$

MCMC

- Gibbs sampler to sample from $\pi\left(c_{1:n}, \mathbf{U}_{1:K} \,\middle|\, \mathbf{z}_{1:n}\right)$.
  1. For $k = 1, \ldots, n$, Sample c_k with
  $$\Pr(c_k = c | c_{-k}, \mathbf{z}_k, \mathbf{U}_{1:K}) \propto \frac{n_{-k,c} + \frac{\alpha}{K}}{n - 1 + \alpha} f(\mathbf{z}_k | \mathbf{U}_c)$$
  2. For $j = 1, \ldots, K$, sample $\mathbf{U}_j | c_{1:n}, \mathbf{z}_{1:n}$
- In the conjugate case, we can integrate out the cluster locations $\mathbf{U}_{1:K}$

MCMC

- Gibbs sampler to sample from $\pi\left(c_{1:n}, \mathbf{U}_{1:K} \middle| \mathbf{z}_{1:n}\right)$.
  1. For $k = 1, \ldots, n$, Sample c_k with

$$\Pr(c_k = c | c_{-k}, \mathbf{z}_k, \mathbf{U}_{1:K}) \propto \frac{n_{-k,c} + \frac{\alpha}{K}}{n - 1 + \alpha} f(\mathbf{z}_k | \mathbf{U}_c)$$

  2. For $j = 1, \ldots, K$, sample $\mathbf{U}_j | c_{1:n}, \mathbf{z}_{1:n}$

- In the conjugate case, we can integrate out the cluster locations $\mathbf{U}_{1:K}$
  1. For $k = 1, \ldots, n$, Sample c_k with probability

$$\Pr(c_k = c | c_{-k}, \mathbf{z}_{1:n}) \propto \frac{n_{-k,c} + \frac{\alpha}{K}}{n - 1 + \alpha} \int f(\mathbf{z}_k | \mathbf{U}) p(\mathbf{U} | c_{-k}, c, \mathbf{z}_{-k}) d\mathbf{U}$$

Example: Latent Dirichlet Allocation for Topic Models

- Sample the topic weights

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})$$

- For $k = 1, \ldots, K$, sample the topics

$$\mathbf{U}_k \sim \mathcal{D}(\frac{\beta}{V}, \ldots, \frac{\beta}{V})$$

- For each of the n words w_k, $k = 1, \ldots, n$
  1. Choose a topic $c_k \sim \mathcal{M}(\pi_{1:K})$
  2. Choose a word $w_k \sim \mathcal{M}(\mathbf{U}_{c_k})$
- See Blei et al. (JMLR 2003) and Griffiths & Steyvers (PNAS 2004)

Limits of the parametric approach

- How to fix the number of clusters K ?

Limits of the parametric approach

- How to fix the number of clusters K ?
- Consider K as an hyperparameter with a given prior (reversible jumps)

Limits of the parametric approach

- How to fix the number of clusters K ?
- Consider K as an hyperparameter with a given prior (reversible jumps)
- Alternative solution: going toward Bayesian nonparametrics with Dirichlet Process Mixtures

Dirichlet Process

- Distribution over distributions

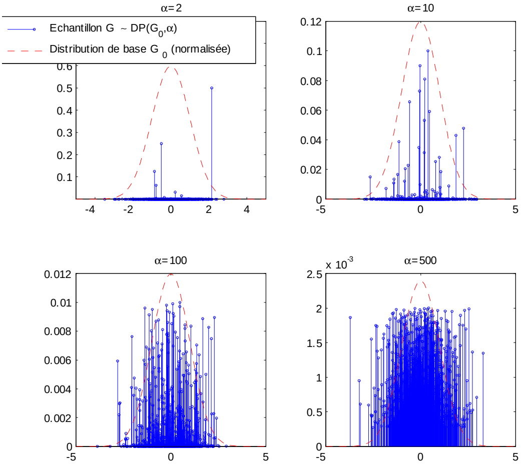
$$\mathbb{G} \sim DP(\alpha, \mathbb{G}_0)$$

- Realization of a DP is a.s. discrete and admits the following *stick-breaking* representation

$$\mathbb{G} = \sum_{j=1}^{\infty} \pi_j \delta_{\mathbf{U}_j}$$

with $\pi_j = \beta_j \prod_{k<j}(1-\beta_k)$, $\beta_j \sim \mathcal{B}(1, \alpha)$ and $\mathbf{U}_j \sim \mathbb{G}_0$.

- Some examples



Dirichlet Process Mixture

- The data $\mathbf{z}_k$ are supposed to be distributed from the following mixture model

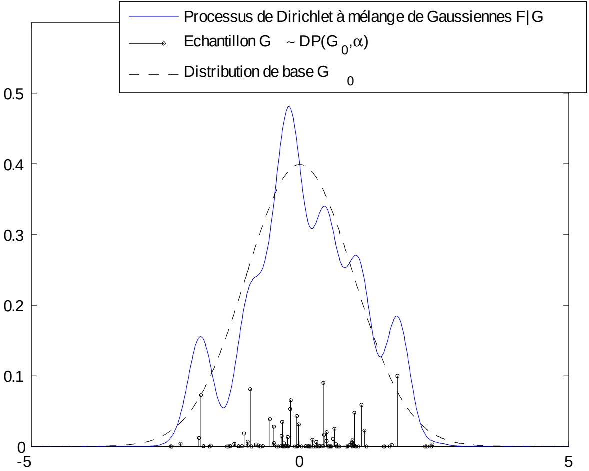
$$\mathbf{z}_k \sim \int f(\cdot|\mathbf{U}_j) d\mathbb{G}(\mathbf{U})$$

where the mixing distribution $\mathbb{G}$ is unknown

$$\mathbb{G} \sim DP(\alpha, \mathbb{G}_0)$$

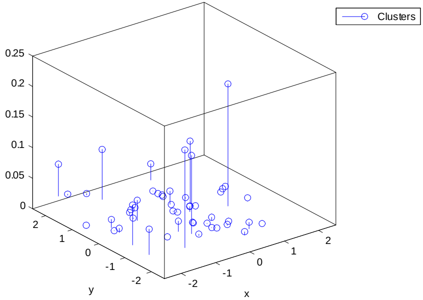
Dirichlet Process Mixture

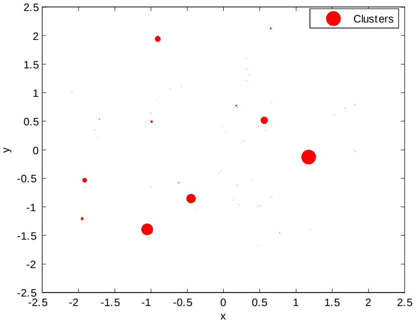
- Some examples



Dirichlet Process Mixture

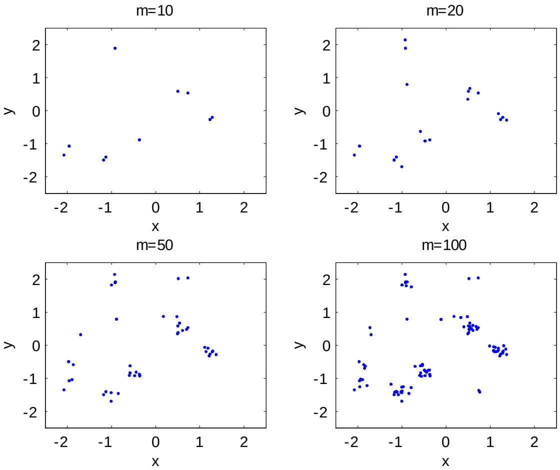
- Some examples





Dirichlet Process Mixture

- Some examples



Dirichlet Process Mixture

- Hierarchical model

$$\mathbb{G} \sim DP(\alpha, \mathbb{G}_0)$$

for $k = 1, \ldots, n$

$$\begin{aligned} \theta_k | \mathbb{G} &\sim \mathbb{G} \\ \mathbf{z}_k | \theta_k &\sim f(\cdot | \theta_k) \end{aligned}$$

Chinese Restaurant Process

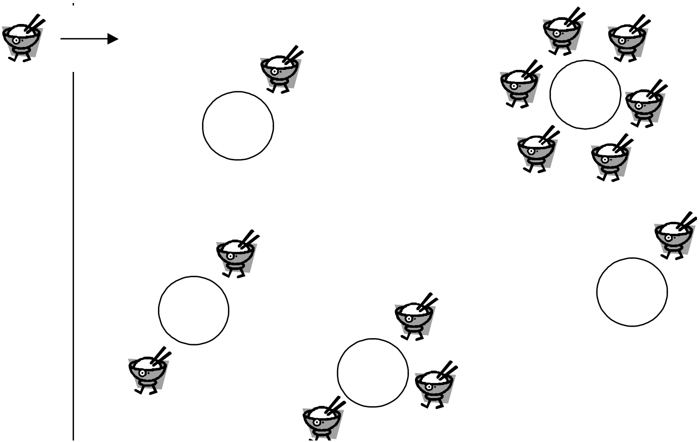
- Integration over $\mathbb{G}$

$$\theta_k|\theta_1,\ldots,\theta_{k-1} \sim \frac{1}{k-1+\alpha}\sum_{j=1}^{k-1}\delta_{\theta_j} + \frac{\alpha}{k-1+\alpha}\mathbb{G}_0$$

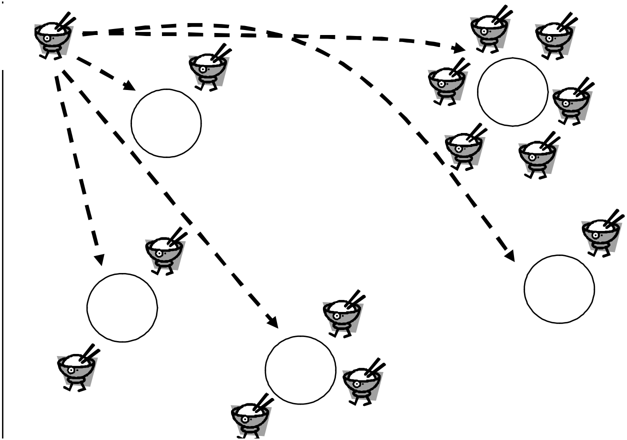
or

$$\theta_k|\theta_1,\ldots,\theta_{k-1} \sim \frac{1}{k-1+\alpha}\sum_{j=1}^{K} m_j^k\delta_{\theta'_j} + \frac{\alpha}{k-1+\alpha}\mathbb{G}_0$$

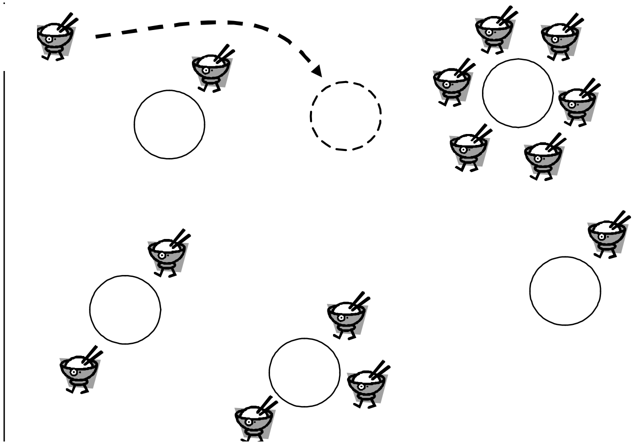
Chinese Restaurant Process



Chinese Restaurant Process



Chinese Restaurant Process



Dirichlet Process Mixtures

- Full posterior $p(\mathbb{G}, \theta_{1:n} | \mathbf{z}_{1:n})$

Dirichlet Process Mixtures

- Full posterior $p(\mathbb{G}, \theta_{1:n} | \mathbf{z}_{1:n})$
- CRP: $p(\theta_{1:n} | \mathbf{z}_{1:n})$

$$\theta_k | \theta_{-k} \sim \frac{1}{n-1+\alpha} \sum_{j \neq k} \delta_{\theta_j} + \frac{\alpha}{n-1+\alpha} \mathbb{G}_0$$

$$\begin{aligned}
p(\theta_k | \theta_{-k}, \mathbf{z}_k) &\propto p(\mathbf{z}_k | \theta_k, \theta_{-k}) p(\theta_k | \theta_{-k}) \\
&\propto \frac{1}{n-1+\alpha} \sum_{j \neq k} p(\mathbf{z}_k | \theta_j, \theta_{-k}) \delta_{\theta_j}(\theta_k) \\
&\quad + \frac{\alpha}{n-1+\alpha} \mathbb{G}_0(\theta_k) p(\mathbf{z}_k | \theta_k, \theta_{-k}) \\
&\propto \frac{1}{n-1+\alpha} \sum_{j \neq k} f(\mathbf{z}_k | \theta_j) \delta_{\theta_j}(\theta_k) \\
&\quad + \frac{\alpha}{n-1+\alpha} \mathbb{G}_0(\theta_k) f(\mathbf{z}_k | \theta_k)
\end{aligned}$$

MCMC algorithm

- So

$$p\left(\theta_k|\theta_{-k},\mathbf{z}_k\right) \propto \sum_{j\neq k} f\left(\mathbf{z}_k|\theta_j\right)\delta_{\theta_j} + \alpha \int f\left(\mathbf{z}_k|\theta\right) d\mathbb{G}_0\left(\theta\right) \times H_k$$

MCMC algorithm

- So

$$p\left(\theta_k|\theta_{-k},\mathbf{z}_k\right) \propto \sum_{j\neq k} f(\mathbf{z}_k|\theta_j)\delta_{\theta_j} + \alpha \int f(\mathbf{z}_k|\theta) d\mathbb{G}_0(\theta) \times H_k$$

- Conjugate case

MCMC algorithm

- So

$$p\left(\theta_k|\theta_{-k}, \mathbf{z}_k\right) \propto \sum_{j\neq k} f\left(\mathbf{z}_k|\theta_j\right)\delta_{\theta_j} + \alpha \int f\left(\mathbf{z}_k|\theta\right) d\mathbb{G}_0(\theta) \times H_k$$

- Conjugate case

  1. For $k = 1\ldots, n$, sample $\theta_k|\theta_{-k}, \mathbf{z}_k$

MCMC algorithm

- So

$$p\left(\theta_k|\theta_{-k},\mathbf{z}_k\right) \propto \sum_{j\neq k} f(\mathbf{z}_k|\theta_j)\delta_{\theta_j} + \alpha \int f(\mathbf{z}_k|\theta)d\mathbb{G}_0(\theta) \times H_k$$

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $\theta_k|\theta_{-k},\mathbf{z}_k$
- Non conjugate case

MCMC algorithm

- So

$$p\left(\theta_k|\theta_{-k},\mathbf{z}_k\right) \propto \sum_{j\neq k} f\left(\mathbf{z}_k|\theta_j\right)\delta_{\theta_j} + \alpha \int f\left(\mathbf{z}_k|\theta\right) d\mathbb{G}_0(\theta) \times H_k$$

- Conjugate case
  1. For $k = 1\ldots,n$, sample $\theta_k|\theta_{-k},\mathbf{z}_k$
- Non conjugate case
  1. For $k = 1\ldots,n$,

MCMC algorithm

- So

$$p(\theta_k | \theta_{-k}, \mathbf{z}_k) \propto \sum_{j \neq k} f(\mathbf{z}_k | \theta_j) \delta_{\theta_j} + \alpha \int f(\mathbf{z}_k | \theta) d\mathbb{G}_0(\theta) \times H_k$$

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $\theta_k | \theta_{-k}, \mathbf{z}_k$
- Non conjugate case
  1. For $k = 1 \ldots, n$,
     1. Sample $\theta_k^* | \theta_{-k} \sim \frac{1}{n-1+\alpha} \sum_{j \neq k} \delta_{\theta_j} + \frac{\alpha}{n-1+\alpha} \mathbb{G}_0$

MCMC algorithm

- So

$$p(\theta_k|\theta_{-k}, \mathbf{z}_k) \propto \sum_{j \neq k} f(\mathbf{z}_k|\theta_j)\delta_{\theta_j} + \alpha \int f(\mathbf{z}_k|\theta) d\mathbb{G}_0(\theta) \times H_k$$

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $\theta_k|\theta_{-k}, \mathbf{z}_k$
- Non conjugate case
  1. For $k = 1 \ldots, n$,
     1. Sample $\theta_k^*|\theta_{-k} \sim \frac{1}{n-1+\alpha} \sum_{j \neq k} \delta_{\theta_j} + \frac{\alpha}{n-1+\alpha} \mathbb{G}_0$
     2. With probability $a(\theta_k^*, \theta_k) = \min(1, \frac{f(\mathbf{z}_k|\theta_k^*)}{f(\mathbf{z}_k|\theta_k)})$, set $\theta_k \leftarrow \theta_k^*$.

Another parameterization of DPM

- Limit $K \to \infty$ of a finite mixture model

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})$$

and for $j = 1, \ldots, K$

$$\mathbf{U}_j | \mathbb{G}_0 \sim \mathbb{G}_0$$

and for $k = 1, \ldots, n$

$$\begin{aligned} c_k | \pi_{1:K} &\sim \mathcal{M}(\pi_{1:K}) \\ \mathbf{z}_k | \mathbf{U}_{c_k} &\sim f(\cdot | \mathbf{U}_{c_k}) \end{aligned}$$

Another parameterization of DPM

- Limit $K \to \infty$ of a finite mixture model

$$\pi_{1:K} \sim \mathcal{D}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})$$

and for $j = 1, \ldots, K$

$$\mathbf{U}_j | \mathbb{G}_0 \sim \mathbb{G}_0$$

and for $k = 1, \ldots, n$

$$\begin{aligned} c_k | \pi_{1:K} &\sim \mathcal{M}(\pi_{1:K}) \\ \mathbf{z}_k | \mathbf{U}_{c_k} &\sim f(\cdot | \mathbf{U}_{c_k}) \end{aligned}$$

- Equivalent to the former model with $\theta_k = \mathbf{U}_{c_k}$

Another parameterization of DPM

- Allocation variables

$$\begin{aligned}\Pr(c_k = c \text{ for } c \in c_{-k} | c_{-k}, \mathbf{z}_k, \mathbf{U}) &\propto \tfrac{n_{-k,c}}{n-1+\alpha} f(\mathbf{z}_k | U_c) \\ \Pr(c_k \neq c_j \text{ for all } j \neq k | c_{-k}, \mathbf{z}_k, \mathbf{U}) &\propto \tfrac{\alpha}{n-1+\alpha} \int f(\mathbf{z}_k | \theta) d\mathbb{G}_0(\theta)\end{aligned}$$

Another parameterization of DPM

- Allocation variables

$$\begin{aligned}\Pr(c_k = c \text{ for } c \in c_{-k} | c_{-k}, \mathbf{z}_k, \mathbf{U}) &\propto \tfrac{n_{-k,c}}{n-1+\alpha} f(\mathbf{z}_k | U_c) \\ \Pr(c_k \neq c_j \text{ for all } j \neq k | c_{-k}, \mathbf{z}_k, \mathbf{U}) &\propto \tfrac{\alpha}{n-1+\alpha} \int f(\mathbf{z}_k|\theta) d\mathbb{G}_0(\theta)\end{aligned}$$

- Cluster location
  For $c \in \{c_1, \ldots, c_n\}$

$$p(\mathbf{U}_c | c_{1:n}, \mathbf{z}_{1:n}) \propto \mathbb{G}_0(\mathbf{U}_c) \prod_{j|c_j=c} p(\mathbf{z}_j | \mathbf{U}_c)$$

MCMC algorithm

- Conjugate case

MCMC algorithm

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $c_k \big| c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k

MCMC algorithm

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $c_k | c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k
  2. For $c \in \{c_1, \ldots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$

MCMC algorithm

- Conjugate case
  1. For $k = 1\dots,n$, sample $c_k | c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k
  2. For $c \in \{c_1, \dots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$
- Non conjugate case

MCMC algorithm

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $c_k | c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k
  2. For $c \in \{c_1, \ldots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$
- Non conjugate case
  1. For $k = 1 \ldots, n$,

MCMC algorithm

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $c_k | c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k
  2. For $c \in \{c_1, \ldots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$
- Non conjugate case
  1. For $k = 1 \ldots, n$,
     1. Sample $c_k^* | c_{-k}$

MCMC algorithm

- Conjugate case
  1. For $k = 1 \ldots, n$, sample $c_k | c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k
  2. For $c \in \{c_1, \ldots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$
- Non conjugate case
  1. For $k = 1 \ldots, n$,
     1. Sample $c_k^* | c_{-k}$
     2. With probability $a(c_k^*, c_k)$, set $c_k \leftarrow c_k^*$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from $\mathbb{G}_0$

MCMC algorithm

- Conjugate case
  1. For $k = 1 \dots, n$, sample $c_k | c_{-k}, \mathbf{z}_k, \mathbf{U}$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from H_k
  2. For $c \in \{c_1, \dots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$
- Non conjugate case
  1. For $k = 1 \dots, n$,
     1. Sample $c_k^* | c_{-k}$
     2. With probability $a(c_k^*, c_k)$, set $c_k \leftarrow c_k^*$. If c_k takes a new value, sample $\mathbf{U}_{c_k}$ from $\mathbb{G}_0$
  2. For $c \in \{c_1, \dots, c_n\}$, sample $\mathbf{U}_c | \{\mathbf{z}_j\}_{c_j = c}$ or perform some update that leaves this distribution invariant

MCMC algorithm

- In the conjugate case, we may be able to integrate out $\mathbf{U}$

MCMC algorithm

- In the conjugate case, we may be able to integrate out $\mathbf{U}$

1. For $k = 1 \ldots, n$, sample $c_k | c_{-k}, \mathbf{z}_{1:n}$ with

$$\begin{aligned} \Pr(c_k = c \text{ for } c \in c_{-k} | c_{-k}, \mathbf{z}_{1:n}) &\propto \tfrac{n_{-k,c}}{n-1+\alpha} \int f(\mathbf{z}_k|\theta) dH_{-k,c}(\theta) \\ \Pr(c_k \neq c_j \text{ for all } j \neq k | c_{-k}, \mathbf{z}_{1:n}) &\propto \tfrac{\alpha}{n-1+\alpha} \int f(\mathbf{z}_k|\theta) d\mathbb{G}_0(\theta) \end{aligned}$$

where

$$H_{-k,c}(\theta) \propto \mathbb{G}_0(\theta) \prod_{j \neq k | c_j = c} f(\mathbf{z}_j|\theta)$$

More advanced algorithms in the non conjugate case

- Prior conditional

$$\Pr(c_k \neq c_j \text{ for all } j \neq k | c_{1:k-1}) = \frac{\alpha}{n-1+\alpha}$$

More advanced algorithms in the non conjugate case

- Prior conditional

$$\Pr(c_k \neq c_j \text{ for all } j \neq k | c_{1:k-1}) \quad = \frac{\alpha}{n-1+\alpha}$$

- If $c_k = c_j$ for some $j \neq k$, set c_k^* to a new value and sample $U_{c_k^*} \sim \mathbb{G}_0$

More advanced algorithms in the non conjugate case

- Prior conditional

$$\Pr(c_k \neq c_j \text{ for all } j \neq k | c_{1:k-1}) \quad = \tfrac{\alpha}{n-1+\alpha}$$

- If $c_k = c_j$ for some $j \neq k$, set c_k^* to a new value and sample $U_{c_k^*} \sim \mathbb{G}_0$
- If c_k is a singleton, sample c_k^* with

$$q(c_k^* = c \text{ for } c \in c_{-k}) = \frac{n_{-k,c}}{n-1}$$

More advanced algorithms in the non conjugate case

- Prior conditional

$$\Pr(c_k \neq c_j \text{ for all } j \neq k | c_{1:k-1}) \quad = \tfrac{\alpha}{n-1+\alpha}$$

- If $c_k = c_j$ for some $j \neq k$, set c_k^* to a new value and sample $U_{c_k^*} \sim \mathbb{G}_0$
- If c_k is a singleton, sample c_k^* with

$$q(c_k^* = c \text{ for } c \in c_{-k}) = \frac{n_{-k,c}}{n-1}$$

- If $c_k = c_j$ for some $j \neq k$

$$a(c_k, c_k^*) = \min\left(1, \frac{\alpha}{n-1} \frac{f(\mathbf{z}_k | \mathbf{U}_{c_k^*})}{f(\mathbf{z}_k | \mathbf{U}_{c_k^*})}\right)$$

More advanced algorithms in the non conjugate case

- Prior conditional

$$\Pr(c_k \neq c_j \text{ for all } j \neq k | c_{1:k-1}) \quad = \tfrac{\alpha}{n-1+\alpha}$$

- If $c_k = c_j$ for some $j \neq k$, set c_k^* to a new value and sample $U_{c_k^*} \sim \mathbb{G}_0$
- If c_k is a singleton, sample c_k^* with

$$q(c_k^* = c \text{ for } c \in c_{-k}) = \frac{n_{-k,c}}{n-1}$$

- If $c_k = c_j$ for some $j \neq k$

$$a(c_k, c_k^*) = \min\left(1, \frac{\alpha}{n-1} \frac{f(\mathbf{z}_k | \mathbf{U}_{c_k^*})}{f(\mathbf{z}_k | \mathbf{U}_{c_k^*})}\right)$$

- If c_k is a singleton

$$a(c_k, c_k^*) = \min\left(1, \frac{n-1}{\alpha} \frac{f(\mathbf{z}_k | \mathbf{U}_{c_k^*})}{f(\mathbf{z}_k | \mathbf{U}_{c_k^*})}\right)$$

More advanced algorithms in the non conjugate case

- Combined with partial Gibbs

More advanced algorithms in the non conjugate case

- Combined with partial Gibbs
- For $k = 1, \ldots, n$, if c_k is not a singleton, sample c_k from $c_{1:n}$ with probabilities

$$\Pr(c_k = c | c_{1:k-1}, \mathbf{z}_k, \mathbf{U}) \quad = b \frac{n_{-k,c}}{n-1} f(\mathbf{z}_k | \mathbf{U}_{c_k})$$

More advanced algorithms in the non conjugate case

- Combined with partial Gibbs
- For $k = 1, \ldots, n$, if c_k is not a singleton, sample c_k from $c_{1:n}$ with probabilities

$$\Pr(c_k = c | c_{1:k-1}, \mathbf{z}_k, \mathbf{U}) \quad = b \frac{n_{-k,c}}{n-1} f(\mathbf{z}_k | \mathbf{U}_{c_k})$$

- Algorithm 7 of Neal