Nonparametric Bayesian Methods: Models, Algorithms, and Applications (Day 3)

Tamara Broderick
ITT Career Development Assistant Professor
Electrical Engineering & Computer Science
MIT
Applications

[Wikipedia.org]

[Wikipedia.org]

[Us CDC PHIL; Futoma, Hariharan, Heller 2017]

[Chati, Balakrishnan 2017]

[Ed Bowlby, NOAA]

[Edens 1972; Hartl, Clark 2003]

[Saria et al 2010]

[US CDC PHIL; Futoma, Hariharan, Heller 2017]

[Deisenroth, Fox, Rasmussen 2015]

[Fox et al 2014]

[Kiefel, Schuler, Hennig 2014]

[Lloyd et al 2012; Miller et al 2010]

[Sudderth, Jordan 2009]

[Datta, Banerjee, Finley, Gelfand 2016]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
Generative model

\[ P(\text{parameters}|\text{data}) \propto P(\text{data}|\text{parameters})P(\text{parameters}) \]

- Finite Gaussian mixture model (\(K\) clusters)

\[
\begin{align*}
\rho_1:K & \sim \text{Dirichlet}(a_1:K) \\
\mu_k & \sim \mathcal{N}(\mu_0, \Sigma_0) \\
z_n & \sim \text{Categorical}(\rho_1:K) \\
x_n & \sim \mathcal{N}(\mu_{z_n}, \Sigma)
\end{align*}
\]
What if $K > N$?

- e.g. species sampling, topic modeling, groups on a social network, etc.

- Components: number of latent groups

- Clusters: number of components represented in the data

- [demo 1, demo 2]

- Number of clusters for $N$ data points is random

- Number of clusters grows with $N$
• Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
Choosing $K = \infty$

• Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

• How to generate $K = \infty$ strictly positive frequencies that sum to one?
Choosing \( K = \infty \)

- Here, difficult to choose finite \( K \) in advance (contrast with small \( K \)): don’t know \( K \), difficult to infer, streaming data
- How to generate \( K = \infty \) strictly positive frequencies that sum to one?
- Observation: \( \rho_{1:K} \sim \text{Dirichlet}(a_{1:K}) \)
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_1:K \sim \text{Dirichlet}(a_1:K)$
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

\[ \Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \]
Choosing $K = \infty$

• Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

• How to generate $K = \infty$ strictly positive frequencies that sum to one?

• Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$\iff \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1)$
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

\[ \Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K) \]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

- Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

\[
\rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1 - \rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)
\]
Choosing \( K = \infty \)

- Here, difficult to choose finite \( K \) in advance (contrast with small \( K \)): don’t know \( K \), difficult to infer, streaming data
- How to generate \( K = \infty \) strictly positive frequencies that sum to one?
  - Observation: \( \rho_{1:K} \sim \text{Dirichlet}(a_{1:K}) \)

\[ \Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp (\rho_2, \ldots, \rho_K) \overset{d}{=} \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K) \]
Choosing $K = \infty$  

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

\[
\Leftrightarrow \begin{array}{c}
\rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1 - \rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)
\end{array}
\]

- “Stick breaking”
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data.

- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

\[
\iff \quad \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)
\]

- “Stick breaking”

\[
v_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4)
\]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

  - Observation: $\mathbf{\rho}_{1:K} \sim \text{Dirichlet}(a_{1:K})$

  \[ \iff \quad \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K) \]

- “Stick breaking”

  \[ V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \quad \rho_1 = V_1 \]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)$$

- “Stick breaking”

$$V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \quad \rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, a_3 + a_4)$$
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$$
\Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)
$$

- “Stick breaking”
  $$
  V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \quad \rho_1 = V_1
  $$
  $$
  V_2 \sim \text{Beta}(a_2, a_3 + a_4) \quad \rho_2 = (1 - V_1)V_2
  $$
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
- Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

\[ \Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1 - \rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K) \]

- “Stick breaking”

\[ V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \]
\[ \rho_1 = V_1 \]
\[ V_2 \sim \text{Beta}(a_2, a_3 + a_4) \]
\[ \rho_2 = (1 - V_1)V_2 \]
\[ V_3 \sim \text{Beta}(a_3, a_4) \]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

  $$\Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp (\rho_2, \ldots, \rho_K) \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)$$

- “Stick breaking”
  
  \[
  \begin{align*}
  V_1 & \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \quad \rho_1 = V_1 \\
  V_2 & \sim \text{Beta}(a_2, a_3 + a_4) \quad \rho_2 = (1 - V_1)V_2 \\
  V_3 & \sim \text{Beta}(a_3, a_4) \quad \rho_3 = (1 - V_1)(1 - V_2)V_3
  \end{align*}
  \]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_1:K \sim \text{Dirichlet}(a_1:K)$

$$\Leftrightarrow \rho_1 \overset{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)$$

- “Stick breaking”

  $$V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \quad \rho_1 = V_1$$
  $$V_2 \sim \text{Beta}(a_2, a_3 + a_4) \quad \rho_2 = (1 - V_1)V_2$$
  $$V_3 \sim \text{Beta}(a_3, a_4) \quad \rho_3 = (1 - V_1)(1 - V_2)V_3$$
  $$\rho_4 = 1 - \sum_{k=1}^{3} \rho_k$$
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[ V_1 \sim \text{Beta}(a_1, b_1) \]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[ V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[
V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1
\]
\[
V_2 \sim \text{Beta}(a_2, b_2)
\]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[
V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \\
V_2 \sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2
\]
Choosing \( K = \infty \)

- Here, difficult to choose finite \( K \) in advance (contrast with small \( K \)): don’t know \( K \), difficult to infer, streaming data
- How to generate \( K = \infty \) strictly positive frequencies that sum to one?

\[
V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \\
V_2 \sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2
\]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[
V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \\
V_2 \sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2
\]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[
\begin{align*}
V_1 & \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \\
V_2 & \sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2 \\
\vdots & \quad V_k \sim \text{Beta}(a_k, b_k)
\end{align*}
\]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

\[
V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \\
V_2 \sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2 \\
\vdots \quad V_k \sim \text{Beta}(a_k, b_k) \quad \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j)\right] V_k
\]
Choosing \( K = \infty \)

- Here, difficult to choose finite \( K \) in advance (contrast with small \( K \)): don’t know \( K \), difficult to infer, streaming data
- How to generate \( K = \infty \) strictly positive frequencies that sum to one?

\[
\begin{align*}
V_1 & \sim \text{Beta}(a_1, b_1) & \rho_1 &= V_1 \\
V_2 & \sim \text{Beta}(a_2, b_2) & \rho_2 &= (1 - V_1)V_2 \\
\vdots & & \rho_k &= \left[ \prod_{j=1}^{k-1} (1 - V_j) \right] V_k
\end{align*}
\]

[Ishwaran, James 2001]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

- **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$

$$
\begin{align*}
V_1 &\sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1 \\
V_2 &\sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2 \\
\vdots & \\
V_k &\sim \text{Beta}(a_k, b_k) \quad \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j)\right] V_k \\
\end{align*}
$$

[Ishwaran, James 2001]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

- **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$
- Griffiths-Engen-McCloskey (GEM) distribution:

\[
\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
\]

\[
V_1 \sim \text{Beta}(a_1, b_1) \quad \rho_1 = V_1
\]

\[
V_2 \sim \text{Beta}(a_2, b_2) \quad \rho_2 = (1 - V_1)V_2
\]

\[
\vdots \quad V_k \sim \text{Beta}(a_k, b_k) \quad \rho_k = \left[ \prod_{j=1}^{k-1} (1 - V_j) \right] V_k
\]

[McCloskey 1965; Engen 1975; Patil and Taillie 1977; Ewens 1987; Sethuraman 1994; Ishwaran, James 2001]
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

- **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$

- Griffiths-Engen-McCloskey (GEM) distribution:

  $\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$

  $V_k \overset{iid}{\sim} \text{Beta}(1, \alpha)$  

  $\rho_k = \left[ \prod_{j=1}^{k-1} (1 - V_j) \right] V_k$

  ...
Choosing $K = \infty$

- Here, difficult to choose finite $K$ in advance (contrast with small $K$): don’t know $K$, difficult to infer, streaming data

- How to generate $K = \infty$ strictly positive frequencies that sum to one?

  - **Dirichlet process stick-breaking:** $a_k = 1, b_k = \alpha > 0$

  - Griffiths-Engen-McCloskey (GEM) distribution:
    \[
    \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
    \]
    \[
    V_k \overset{iid}{\sim} \text{Beta}(1, \alpha) \quad \rho_k = \left[ \prod_{j=1}^{k-1} (1 - V_j) \right] V_k
    \]

  [McCloskey 1965; Engen 1975; Patil and Taillie 1977; Ewens 1987; Sethuraman 1994; Ishwaran, James 2001]
Distributions
Distributions

- Beta $\rightarrow$ random distribution over 1, 2
Distributions

- Beta → random distribution over 1, 2
- Dirichlet → random distribution over 1, 2, ..., $K$
Distributions

- Beta → random distribution over 1, 2
- Dirichlet → random distribution over 1, 2, …, \( K \)
- GEM / Dirichlet process stick-breaking → random distribution over 1, 2, …
Distributions

• Beta $\rightarrow$ random distribution over $1, 2$

• Dirichlet $\rightarrow$ random distribution over $1, 2, \ldots, K$

• GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over $1, 2, \ldots$

• Infinity of parameters: components
• Growing number of parameters: clusters
Distributions

• Beta $\rightarrow$ random distribution over $1, 2$

• Dirichlet $\rightarrow$ random distribution over $1, 2, \ldots, K$

• GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over $1, 2, \ldots$
Dirichlet process mixture model
Dirichlet process mixture model

• Gaussian mixture model
Dirichlet process mixture model

- Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
Dirichlet process mixture model

- Gaussian mixture model
  \[
  \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
  \]
Dirichlet process mixture model

- Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \; k = 1, 2, \ldots \]
Dirichlet process mixture model

- Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \quad k = 1, 2, \ldots \]
Dirichlet process mixture model

- Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots \]
Dirichlet process mixture model

• Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \quad k = 1, 2, \ldots \]

• i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \)
Dirichlet process mixture model

- Gaussian mixture model
  \[
  \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
  \]
  \[
  \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \quad k = 1, 2, \ldots
  \]
- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)
Dirichlet process mixture model

- Gaussian mixture model
  
  \[
  \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
  \]

  \[
  \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \quad k = 1, 2, \ldots
  \]

- i.e.
  
  \[
  G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))
  \]
Dirichlet process mixture model

- Gaussian mixture model
  \[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
  \[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \; k = 1, 2, \ldots \]
  - i.e. \[ G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \]

- \[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]
Dirichlet process mixture model

- Gaussian mixture model
  \[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
  \[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \ k = 1, 2, \ldots \]
- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]
\[ \mu_n^* = \mu_{z_n} \]
Dirichlet process mixture model

- Gaussian mixture model
  \[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
  \[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \quad k = 1, 2, \ldots \]
  i.e. \[ G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \]

- \[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]
  \[ \mu_n^* = \mu_{z_n} \]
  i.e. \[ \mu_n^* \overset{iid}{\sim} G \]
Dirichlet process mixture model

- Gaussian mixture model
  \[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
  \[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \, k = 1, 2, \ldots \]
- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)

- \( z_n \overset{iid}{\sim} \text{Categorical}(\rho) \)
  \[ \mu^*_n = \mu_{z_n} \]
- i.e. \( \mu^*_n \overset{iid}{\sim} G \)

- \( x_n \overset{indep}{\sim} \mathcal{N}(\mu^*_n, \Sigma) \)
Dirichlet process mixture model

- Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \ k = 1, 2, \ldots \]
- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)

- \( z_n \overset{iid}{\sim} \text{Categorical}(\rho) \)
- \( \mu_n^* = \mu_{z_n} \)
- i.e. \( \mu_n^* \overset{iid}{\sim} G \)

- \( x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu_n^*, \Sigma) \)
Dirichlet process mixture model

• Gaussian mixture model

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \ k = 1, 2, \ldots \]

• i.e. \[ G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \]

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]
\[ \mu_n^* = \mu_{z_n} \]

• i.e. \[ \mu_n^* \overset{iid}{\sim} G \]

\[ x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma) \]

[demo]
Dirichlet process mixture model

- More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), \ k = 1, 2, \ldots \]
- i.e. \[ G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \]

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]
\[ \mu_n^* = \mu_{z_n} \]
- i.e. \[ \mu_n^* \overset{iid}{\sim} G \]

\[ x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma) \]
Dirichlet process mixture model

- More generally
  \[
  \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
  \]
  \[
  \phi_k \overset{iid}{\sim} G_0
  \]
  \[
  k = 1, 2, \ldots
  \]
  \[
  \text{i.e. } G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))
  \]

- \[
  z_n \overset{iid}{\sim} \text{Categorical}(\rho)
  \]
  \[
  \mu_n^* = \mu_{z_n}
  \]
  \[
  \text{i.e. } \mu_n^* \overset{iid}{\sim} G
  \]

- \[
  x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu_n^*, \Sigma)
  \]
Dirichlet process mixture model

• More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \phi_k \overset{iid}{\sim} G_0 \]

\[ k = 1, 2, \ldots \]

• i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]

\[ \mu^n_* = \mu_{z_n} \]

• i.e. \( \mu^n_* \overset{iid}{\sim} G \)

\[ x_n \overset{indep}{\sim} \mathcal{N}(\mu^n_*, \Sigma) \]
Dirichlet process mixture model

- More generally
  \[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

- i.e. \( \phi_k \overset{iid}{\sim} G_0 \)
  \[ k = 1, 2, \ldots \]

- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)

- \( z_n \overset{iid}{\sim} \text{Categorical}(\rho) \)

- \( \mu^*_n = \mu_{z_n} \)

- i.e. \( \mu^*_n \overset{iid}{\sim} G \)

- \( x_n \overset{indep}{\sim} \mathcal{N}(\mu^*_n, \Sigma) \)
Dirichlet process mixture model

More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \phi_k \overset{iid}{\sim} G_0 \]

\[ k = 1, 2, \ldots \]

- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0)) \)

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]

\[ \mu_n^* = \mu_{z_n} \]

- i.e. \( \mu_n^* \overset{iid}{\sim} G \)

\[ x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu_n^*, \Sigma) \]
Dirichlet process mixture model

- More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \phi_k \overset{iid}{\sim} G_0 \]

- i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0) \)

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]

\[ \mu_n^* = \mu_{z_n} \]

- i.e. \( \mu_n^* \overset{iid}{\sim} G \)

\[ x_n \overset{iid}{\sim} N(\mu_n^*, \Sigma) \]
Dirichlet process mixture model

- More generally
  
  \[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
  
  \[ \phi_k \overset{iid}{\sim} G_0 \quad k = 1, 2, \ldots \]
  
  - i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0) \)

- \( z_n \overset{iid}{\sim} \text{Categorical}(\rho) \)
  
  \( \theta_n = \phi_{z_n} \)
  
  - i.e. \( \mu_n^* \overset{iid}{\sim} G \)

- \( x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma) \)
More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \phi_k \overset{iid}{\sim} G_0 \quad k = 1, 2, \ldots \]

i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0) \)

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]

\[ \theta_n = \phi_{z_n} \]

i.e. \( \theta_n \overset{iid}{\sim} G \)

\[ x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu^*_n, \Sigma) \]
Dirichlet process mixture model

• More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \phi_k \overset{iid}{\sim} G_0 \quad k = 1, 2, \ldots \]

• i.e.

\[ G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0) \]

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]

\[ \theta_n = \phi_{z_n} \]

• i.e.

\[ \theta_n \overset{iid}{\sim} G \]

\[ x_n \overset{indep}{\sim} F(\theta_n) \]
Dirichlet process mixture model

More generally

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]

\[ \phi_k \overset{iid}{\sim} G_0, \quad k = 1, 2, \ldots \]

i.e. \( G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0) \)

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho) \]

\[ \theta_n = \phi_{z_n} \]

i.e. \( \theta_n \overset{iid}{\sim} G \)

\[ x_n \overset{indep}{\sim} F(\theta_n) \]

[Antoniak 1974; Ferguson 1983; West, Müller, Escobar 1994; Escobar, West 1995; MacEachern, Müller 1998]
Distributions

- Beta $\rightarrow$ random distribution over 1, 2
- Dirichlet $\rightarrow$ random distribution over 1, 2, $\ldots$, $K$
- GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over 1, 2, $\ldots$
Distributions

- Beta $\rightarrow$ random distribution over 1, 2
- Dirichlet $\rightarrow$ random distribution over 1, 2, \ldots, $K$
- GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over 1, 2, \ldots

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
Distributions

- Beta → random distribution over 1, 2
- Dirichlet → random distribution over 1, 2, ..., K
- GEM / Dirichlet process stick-breaking → random distribution over 1, 2, ...

\[ \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha) \]
\[ \phi_k \overset{iid}{\sim} G_0 \]
Distributions

- Beta $\rightarrow$ random distribution over 1, 2
- Dirichlet $\rightarrow$ random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over $1, 2, \ldots$

\[
\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
\]
\[
\phi_k \overset{iid}{\sim} G_0
\]
\[
G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}
\]
Distributions

- Beta $\rightarrow$ random distribution over $1, 2$

- Dirichlet $\rightarrow$ random distribution over $1, 2, \ldots, K$

- GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over $1, 2, \ldots$

$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$
$$\phi_k \overset{iid}{\sim} G_0$$
$$G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}$$
Distributions

- Beta $\rightarrow$ random distribution over 1, 2
- Dirichlet $\rightarrow$ random distribution over 1, 2, $\ldots$, $K$
- GEM / Dirichlet process stick-breaking $\rightarrow$ random distribution over 1, 2, $\ldots$

- **Dirichlet process** $\rightarrow$
  random distribution over $\Phi$:
  \[
  \rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)
  \]
  \[
  \phi_k \overset{iid}{\sim} G_0
  \]
  \[
  G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}
  \]

[Ferguson 1973]
DP or not DP, that is the question
DP or not DP, that is the question

- GEM:

- Compare to:
  - Finite (small $K$) mixture model
  - Finite (large $K$) mixture model

...
DP or not DP, that is the question

- GEM: 

- Compare to:

![Diagram with data points]
DP or not DP, that is the question

• GEM:  

• Compare to:
  • Finite (small $K$) mixture model
DP or not DP, that is the question

• GEM: ...

• Compare to:
  • Finite (small $K$) mixture model
  
  ![Diagram of finite small K mixture model]

  • Finite (large $K$) mixture model

  ![Diagram of finite large K mixture model]
DP or not DP, that is the question

- GEM:
- Compare to:
  - Finite (small $K$) mixture model
  - Finite (large $K$) mixture model
  - Time series
Calculating the posterior

\[ P(\text{parameters}|\text{data}) \propto P(\text{data}|\text{parameters})P(\text{parameters}) \]

- Finite Gaussian mixture model \((K\text{ clusters})\)

\[ \rho_{1:K} \sim \text{Dirichlet}(a_{1:K}) \]

\[ \mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0) \]

\[ z_n \overset{iid}{\sim} \text{Categorical}(\rho_{1:K}) \]

\[ x_n \overset{indep}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma) \]