Feature allocations, probability functions, and paintboxes

Tamara Broderick
UC Berkeley
(MIT starting Spring 2015)
Clustering/Partition
Clustering/Partition

“clusters”,
“classes”,
“blocks (of a partition)”
Clustering/Partition

"clusters", "classes", "blocks (of a partition)"
## Clustering/Partition

<table>
<thead>
<tr>
<th></th>
<th>Cat</th>
<th>Dog</th>
<th>Mouse</th>
<th>Lizard</th>
<th>Sheep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Latent feature allocation

<table>
<thead>
<tr>
<th></th>
<th>Cat</th>
<th>Dog</th>
<th>Mouse</th>
<th>Lizard</th>
<th>Sheep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture 1</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Picture 2</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 3</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Picture 4</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Picture 5</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“features”, “topics”
Latent feature allocation

- Exchangeable
- Finite # of features per data point

“features”, “topics”
Characterizations

• Exchangeable cluster distributions are characterized

• What about exchangeable feature distributions?
Exchangeable probability functions

\[ P(\mathbf{N}, 1, \ldots, N, K) \]
Exchangeable probability functions

\[ P(1, 2, \ldots, K) = p(S_{N,1}, \ldots, S_{N,K}) \]
Exchangeable probability functions

\[
\mathbb{P}(1 \ 2 \ \ldots \ K) = p(S_{N,1}, \ldots, S_{N,K})
\]
Exchangeable probability functions

Exchangeable partition probability function (EPPF)

\[ P(1, 2, \ldots, K) = p(S_{N,1}, \ldots, S_{N,K}) \]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?
Example: Indian buffet process
Example: Indian buffet process

\[
\begin{array}{cccc}
  & k = 1 & 2 & \ldots & K \\
  n = 1 & & & & \\
   2 & & & & \\
   \vdots & & & & \\
  N & & & & \\
\end{array}
\]

[Griffiths, Ghahramani 2006]
For $n = 1, 2, \ldots, N$

Example: Indian buffet process

For $n = 1, 2, \ldots, N$

$\text{Example: Indian buffet process}$

[Griffiths, Ghahramani 2006]
Example: Indian buffet process

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability

$$
\frac{S'_{n-1,k}}{\theta + n - 1}
$$

$S'$ indicates the number of new features for data point $n$. 

Example: Indian buffet process

[Griffiths, Ghahramani 2006]
For \( n = 1, 2, \ldots, N \)
1. Data point \( n \) has an existing feature \( k \) that has already occurred \( S_{n-1,k} \) times with probability \( \frac{S_{n-1,k}}{\theta + n - 1} \)
2. Number of new features for data point \( n \):
\[
K_n^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)
\]
Example: Indian buffet process

For $n = 1, 2, \ldots, N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability

$$
\frac{S_{n-1,k}}{\theta + n - 1}
$$

2. Number of new features for data point $n$:

$$
K_{n}^{+} = \text{Poisson} \left( \frac{\theta}{\theta + n - 1} \right)
$$

\[\text{[Griffiths, Ghahramani 2006]}\]
Example: Indian buffet process

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point $n$: $K_n^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)$

[Griffiths, Ghahramani 2006]
Example: Indian buffet process

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability

$$\frac{S_{n-1,k}}{\theta + n - 1}$$

2. Number of new features for data point $n$:

$$K^+_n = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)$$

[Griffiths, Ghahramani 2006]
Example: Indian buffet process

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point $n$: $K_n^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)$

[Griffiths, Ghahramani 2006]
Example: Indian buffet process

For $n = 1, 2, ..., N$
1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point $n$: $K^+_n = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)$
Example: Indian buffet process

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability

$$\frac{S_{n-1,k}}{\theta + n - 1}$$

2. Number of new features for data point $n$: $K_n^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]
Example: Indian buffet process

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has already occurred $S_{n-1,k}$ times with probability

$$\frac{S_{n-1,k}}{\theta + n - 1}$$

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)$$

[Griffiths, Ghahramani 2006]
Example: Indian buffet process

For \( n = 1, 2, \ldots, N \)

1. Data point \( n \) has an existing feature \( k \) that has already occurred \( S_{n-1,k} \) times with probability \( \frac{S_{n-1,k}}{\theta + n - 1} \)

2. Number of new features for data point \( n \):

\[
K_n^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + n - 1} \right)
\]

[Griffiths, Ghahramani 2006]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[
P(k = 1, 2, \ldots, K | n = 1, 2, \ldots, N) \]

\[
\begin{array}{cccc}
  n = 1 & 2 & \ldots & K \\
  \ \ 1 & \blacksquare & \blacksquare & \blacksquare \\
  \ \ 2 & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\
  \ \ \vdots & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\
  \ \ \ \ \ \ \ N & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\
\end{array}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[
P(\vdots) = \frac{1}{K_N!} (\theta \gamma)^{K_N} \exp \left( -\theta \gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k}) \Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[ P(\vdots) = \begin{array}{cccc}
    n = 1 & 2 & \cdots & K \\
    k = 1 & & & \\
    & & & \\
    \vdots & \cdots & \cdots & \\
    \text{N} & & & \\
\end{array} \]

\[
= \frac{1}{K_N!} (\theta \gamma)^{K_N} \exp \left( -\theta \gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[
\mathbb{P}(n_k = 1, 2, \ldots, K, n_1 = 1, 2, \ldots, N) = \frac{1}{K_N!} (\theta \gamma)^{K_N} \exp \left( -\theta \gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k}) \Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[ P(\vdots) \]

Number of data points

Number of features

Size of kth feature

\[
= \frac{1}{K_N!} (\theta \gamma)^{K_N} \exp \left( - \theta \gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k}) \Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[
P(\ldots) = \frac{1}{K_N!} (\theta \gamma)^{K_N} \exp \left( -\theta \gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k}) \Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}
\]

\[
= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Example: Indian buffet process (IBP)

\[
\mathbb{P}(\cdot) = \frac{1}{K_N!} (\theta \gamma)^{K_N} \exp \left( -\theta \gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k}) \Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}
\]

\[
= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})
\]

“EFPF”

[Broderick, Jordan, Pitman 2013]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

$n = 1$

\[
\begin{array}{ccc}
\text{2} & \text{0} & \text{0} \\
\vdots & \vdots & \vdots \\
\text{N} & \text{0} & \text{0} \\
\end{array}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

\[
\begin{align*}
\mathbb{P}(\text{row } = \text{ 1}) &= p_1 \\
\mathbb{P}(\text{row } = \text{ 2}) &= p_2 \\
\mathbb{P}(\text{row } = \text{ 3}) &= p_3 \\
\mathbb{P}(\text{row } = \text{ 4}) &= p_4
\end{align*}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

\[
\begin{align*}
\mathbb{P}(\text{row } = \begin{array}{c} \blacksquare \end{array}) &= p_1 \\
\mathbb{P}(\text{row } = \begin{array}{c} \bullet \blacksquare \end{array}) &= p_2 \\
\mathbb{P}(\text{row } = \begin{array}{c} \blacksquare \bullet \end{array}) &= p_3 \\
\mathbb{P}(\text{row } = \begin{array}{c} \bullet \bullet \end{array}) &= p_4
\end{align*}
\]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

\[ \mathbb{P}(\text{row } = \begin{array}{c} \blacksquare \\ \square \end{array}) = p_1 \]
\[ \mathbb{P}(\text{row } = \begin{array}{c} \square \\ \blacksquare \end{array}) = p_2 \]
\[ \mathbb{P}(\text{row } = \begin{array}{c} \blacksquare \\ \blacksquare \end{array}) = p_3 \]
\[ \mathbb{P}(\text{row } = \begin{array}{c} \square \\ \square \end{array}) = p_4 \]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

\[ P(\text{row } = \begin{array}{c}
  \text{black} \\
  \text{white}
\end{array}) = p_1 \]
\[ P(\text{row } = \begin{array}{c}
  \text{white} \\
  \text{black}
\end{array}) = p_2 \]
\[ P(\text{row } = \begin{array}{c}
  \text{black} \\
  \text{black}
\end{array}) = p_3 \]
\[ P(\text{row } = \begin{array}{c}
  \text{white} \\
  \text{white}
\end{array}) = p_4 \]

\[ P\left(\begin{array}{c}
  \text{black} \\
  \text{black} \\
  \text{white} \\
  \text{white}
\end{array}\right) = p_1 p_2 \]
\[ P\left(\begin{array}{c}
  \text{black} \\
  \text{black} \\
  \text{white} \\
  \text{white}
\end{array}\right) = p_3 p_4 \]

[Broderick, Jordan, Pitman 2013]
Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

\[ P(\text{row } = \begin{array}{c} 0 \\ 1 \\ \vdots \\ N \end{array} ) = p_1 \]
\[ P(\text{row } = \begin{array}{c} 1 \\ 0 \\ \vdots \\ N \end{array} ) = p_2 \]
\[ P(\text{row } = \begin{array}{c} 1 \\ 1 \\ \vdots \\ N \end{array} ) = p_3 \]
\[ P(\text{row } = \begin{array}{c} 0 \\ 0 \\ \vdots \\ N \end{array} ) = p_4 \]

\[ P( \begin{array}{c} 1 \\ 1 \\ \vdots \\ N \end{array} ) \neq P( \begin{array}{c} 0 \\ 0 \\ \vdots \\ N \end{array} ) \]

\[ p_1 p_2 \neq p_3 p_4 \]
Exchangeable probability functions

Exchangeable cluster distributions
equal Cluster distributions with EPPFs

Exchangeable feature distributions

Two-feature example

Feature distributions with EFPFs

IBP

[Broderick, Jordan, Pitman 2013]
Paintboxes

Exchangeable partition: Kingman paintbox
Paintboxes

Exchangeable partition: Kingman paintbox
Paintboxes

Exchangeable partition: Kingman paintbox
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]
Paintboxes

Exchangeable partition: Kingman paintbox

[Diagram showing a partition with four segments labeled 1 to 4, and a grid below with shaded boxes representing the partition.]
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]
Paintboxes

Exchangeable partition: Kingman paintbox

[Diagram showing a sequence of numbers and a grid with colored boxes]
Paintboxes

Exchangeable partition: Kingman paintbox

[Diagram of paintboxes with numbers and colors]

[Kingman 1978]
Paintboxes

Exchangeable partition: Kingman paintbox

[Diagram of paintboxes with labels and colors indicating Cat and Dog clusters]

[Kingman 1978]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Image of grid with feature paintboxes for Cat, Dog, Mouse, Lizard, Sheep, and Horse features]

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Diagram showing feature allocation with paintboxes for Cat, Dog, Mouse, Lizard, Sheep, and Horse features]

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

Cat feature
Dog feature
Mouse feature
Lizard feature
Sheep feature
Horse feature

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable feature allocation: feature paintbox

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs

Exchangeable feature distributions

Two-feature example

Feature distributions with EFPFs

[Broderick, Pitman, Jordan 2013]
Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Two-feature example

Feature distributions with EFPFs

[Broderick, Pitman, Jordan 2013]
Paintboxes

Two feature example

\[ \mathbb{P}(\text{row } = \begin{array}{c} \text{■} \text{□} \end{array}) = p_1 \]
\[ \mathbb{P}(\text{row } = \begin{array}{c} \text{□} \text{■} \end{array}) = p_2 \]
\[ \mathbb{P}(\text{row } = \begin{array}{c} \text{■} \text{■} \end{array}) = p_3 \]
\[ \mathbb{P}(\text{row } = \begin{array}{c} \text{□} \text{□} \end{array}) = p_4 \]
Paintboxes

Indian buffet process: beta feature frequencies
Paintboxes

Indian buffet process: beta feature frequencies

For $m = 1, 2, ...$
1. Draw $K_m^+ = \text{Poisson} \left( \frac{\gamma}{\theta + m - 1} \right)$

[Thibaux, Jordan 2007]
**Paintboxes**

Indian buffet process: beta feature frequencies

For $m = 1, 2, \ldots$

1. Draw $K_m^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size $q_k \sim \text{Beta}(1, \theta + m - 1)$
Indian buffet process: beta feature frequencies

For $m = 1, 2, \ldots$
1. Draw $K_m^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$

Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$
   Draw a frequency of size $q_k \sim \text{Beta}(1, \theta + m - 1)$
For $m = 1, 2, ...$

1. Draw $K^+_m = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$

Set $K_m = \sum_{j=1}^{m} K^+_j$

2. For $k = K_{m-1} + 1, \ldots, K_m$

Draw a frequency of size

$q_k \sim \text{Beta}(1, \theta + m - 1)$

[Thibaux, Jordan 2007]
Paintboxes

Indian buffet process: beta feature frequencies

For $m = 1, 2, \ldots$

1. Draw $K_m^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$

Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$
   Draw a frequency of size
   
   \[ q_k \sim \text{Beta}(1, \theta + m - 1) \]
Indian buffet process: beta feature frequencies

For $m = 1, 2, ...$
1. Draw $K_m^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$
   
   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$
   Draw a frequency of size
   
   $q_k \sim \text{Beta}(1, \theta + m - 1)$
Indian buffet process: beta feature frequencies

For $m = 1, 2, ...$

1. Draw $K_m^+ = \text{Poisson} \left( \frac{\theta}{\gamma \theta + m - 1} \right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size $q_k \sim \text{Beta}(1, \theta + m - 1)$

[Thibaux, Jordan 2007]
For $m = 1, 2, \ldots$

1. Draw $K_m^+ = \text{Poisson} \left( \frac{\theta}{\gamma \theta + m - 1} \right)$

Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$

Draw a frequency of size $q_k \sim \text{Beta}(1, \theta + m - 1)$
For $m = 1, 2, \ldots$

1. Draw $K_m^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$

Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$
   Draw a frequency of size
   
   $q_k \sim \text{Beta}(1, \theta + m - 1)$
For $m = 1, 2, \ldots$
1. Draw $K_m^+ = \text{Poisson} \left( \gamma \frac{\theta}{\theta + m - 1} \right)$
   Set $K_m = \sum_{j=1}^{m} K_j^+$
2. For $k = K_{m-1} + 1, \ldots, K_m$
   Draw a frequency of size $q_k \sim \text{Beta}(1, \theta + m - 1)$
For $m = 1, 2, ...$

1. Draw $K_m^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For $k = K_{m-1} + 1, \ldots, K_m$

Draw a frequency of size

$q_k \sim \text{Beta}(1, \theta + m - 1)$
Indian buffet process: beta feature frequencies
Paintboxes

Indian buffet process: beta feature frequencies
Paintboxes

Indian buffet process: beta feature frequencies
Paintboxes

Indian buffet process: beta feature frequencies
Paintboxes

Indian buffet process: beta feature frequencies
Paintboxes

Indian buffet process: beta feature frequencies
Paintboxes
Paintboxes

“Feature frequency models”
Paintboxes

Two feature example

\[ \mathbb{P}(row = \begin{array}{l} \text{black} \end{array} ) = p_1 \]
\[ \mathbb{P}(row = \begin{array}{l} \text{white} & \text{black} \end{array} ) = p_2 \]
\[ \mathbb{P}(row = \begin{array}{l} \text{black} \end{array} ) = p_2 \]
\[ \mathbb{P}(row = \begin{array}{l} \text{black} \end{array} ) = p_3 \]
\[ \mathbb{P}(row = \begin{array}{l} \text{white} & \text{white} \end{array} ) = p_4 \]
Paintboxes

Two feature example

Not a feature frequency model

\[
P(p_1) = p_1 \quad P(p_3) = p_2 \quad P(p_2) = p_3 \quad P(p_4) = p_4
\]
Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Two-feature example

Feature distributions with EFPFs

(IBP)

[Broderick, Pitman, Jordan 2013]
Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Two-feature example

Feature frequency models

[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?
Feature frequency models: EFPFs?
Feature frequency models: EFPFs?
Feature frequency models: EFPFs?
Feature frequency models: EFPFs?

\[ \mathbb{P}(k = 1, 2, \ldots, K, n = 1, 2, \ldots, N) \]
Feature frequency models: EFPFs?
Feature frequency models: EFPFs?

\[
P(q_1, \ldots, q_K; n, k) \propto q_1^{S_{n,k}} (1 - q_1)^{N - S_{n,k}}
\]
Feature frequency models: EFPFs?
Feature frequency models: EFPFs?

\[ P(N, k) = \sum_{i=1}^{K} q_{i_k}^{S_{N,k}} \left(1 - q_{i_k}\right)^{N - S_{N,k}} \]
Feature frequency models: EFPFs?

\[ \mathbb{P}(S_{N,k}) = \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \]

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

\[ n = 1, 2, \ldots, N \]

\[ 0 \leq q_1, q_2, q_3, q_4, q_5, q_6 \leq 1 \]
Feature frequency models: EFPFs?

\[ \mathbb{P}(i_k \neq j \mid i_k) = \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \]

\[ \cdot \prod_{j \notin \{i_k\}}^{K} (1 - q_j)^N \]
Feature frequency models: EFPFs?

\[
\mathbb{E}
\left[
\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N-S_{N,k}} \right]
\cdot \prod_{j \not\in \{i_k\}}^{K} (1 - q_j)^N
\]
Feature frequency models: EFPFs?

\[
P(\ldots) = \mathbb{E}\left[ \sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \right]
\]

[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?

\[
\mathbb{P}(
\begin{array}{cccc}
  k = 1 & 2 & \cdots & K \\
  n = 1 & \ast & \ast & \ast \\
  2 & \ast & \ast & \ast \\
  N & \ast & \ast & \ast
\end{array}
\)
\]

\[
\mathbb{E}
\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \\
\cdot \prod_{j \notin \{i_k\}^{K}_{k=1}} (1 - q_j)^N
\]

Size of \(k\)th feature

[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?

\[ \mathbb{P}( \cdot ) = \mathbb{E}[ \sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \cdot \prod_{j \notin \{i_k\}}^{K} (1 - q_j)^N ] \]

[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?

\[ P(\vdots) \]

\[ = \mathbb{E} \left[ \sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \right] \]

Number of features

Number of data points

Size of kth feature

[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?

\[
P(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K}) = \mathbb{E}\left[ \sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \right]
\]

\[
\cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N
\]

[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Two-feature example

Feature frequency models: EFPFs?
[Broderick, Pitman, Jordan 2013]
Feature frequency models: EFPFs?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Two-feature example

Feature frequency models

Feature distributions with EFPFs

[Broderick, Pitman, Jordan 2013]
Distributions with EFPFs: frequencies?
Distributions with EFPFs: frequencies?

- Any number (+unbounded case) of features

[Broderick, Pitman, Jordan 2013]
Distributions with EFPFs: frequencies?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Feature frequency models
Two-feature example
IBP

Feature distributions with EFPFs
[Broderick, Pitman, Jordan 2013]
Distributions with EFPFs: frequencies?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

Two-feature example

Feature distributions with EFPFs
= Feature frequency models

[Broderick, Pitman, Jordan 2013]
Theory conclusions

• Feature paintbox: characterization of exchangeable feature models
• Limits of clustering characterizations in feature case?
• Remaining connections to fill in
• Other combinatorial structures

Exchangeable features
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?

Exchangeable features; feature paintbox
Models with EFPFs
Theory conclusions

• Feature paintbox: characterization of exchangeable feature models
• Limits of clustering characterizations in feature case?
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure

Models with EFPFs; feature frequency models

Exchangeable features; feature paintbox

Exchangeable clusters; models with EPPFs; Kingman paintbox

Two-feature example

Models with EFPFs; feature frequency models
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections

Models with EFPFs; feature frequency models

Exchangeable features; feature paintbox

Exchangeable clusters; models with EPPFs; Kingman paintbox

Two-feature example

Models with EFPFs; feature frequency models

IBP
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs)

Models with EFPFs; feature frequency models

Exchangeable clusters; models with EPPFs; Kingman paintbox

Two-feature example

Models with EFPFs; feature frequency models

Completely random measures

Exchangeable features; feature paintbox

IBP
Theory conclusions

• Feature paintbox: characterization of exchangeable feature models
• Characterization of alternative correlation structure
• Remaining connections (CRMs)
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs)
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust)
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust, etc)
Theory conclusions

• Feature paintbox: characterization of exchangeable feature models
• Characterization of alternative correlation structure
• Remaining connections (CRMs, dust, etc)
• Other combinatorial structures
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust, etc)
- Other combinatorial structures
References


Further References


