The Mondrian Process

Daniel Roy

Joint work with Yee Whye Teh

Massachusetts Institute of Technology

July 9, 2008

NPBayes Workshop @ ICML/UAI
Outline

DP Poisson Construction and Relational Modeling

Extending Exchangeability

The Mondrian Process
DP Poisson Construction

A less well-known representation for Dirichlet processes:

\[
\{\mu_1 > \mu_2 > \cdots\} \sim \text{Poisson}(\alpha \mu^{-1}, (0, 1))
\]

\[
\theta_k \sim \mathcal{H}
\]

\[
\theta(\xi) = \theta_k \text{ where } \xi \in (\mu_k, \mu_{k-1})
\]

\[
\xi_i \sim \text{Uniform}[0, 1]
\]

- [0, 1]\{\mu_1, \mu_2, \ldots\} is an infinite collection of intervals with total length 1.
- \(\theta(\xi)\) is piecewise constant on each interval \((\mu_k, \mu_{k-1})\).
- Item \(i\) belongs to cluster \(k\) if \(\xi_i \in (\mu_k, \mu_{k-1})\), and has parameter \(\theta(\xi_i)\).
Relational Data

**Relational data** consists of observations of the *relationships* among objects, rather than of the objects themselves.

Examples:

<table>
<thead>
<tr>
<th>objects</th>
<th>relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>proteins</td>
<td>interactions</td>
</tr>
<tr>
<td>neurons</td>
<td>connectivity</td>
</tr>
<tr>
<td>webpages</td>
<td>hyper-links</td>
</tr>
<tr>
<td>users, movies</td>
<td>ratings</td>
</tr>
<tr>
<td>authors, papers</td>
<td>authorship</td>
</tr>
</tbody>
</table>

This talk: relations among two (possibly distinct) sets of objects. Can be visualized as matrices (arrays).
Infinite Relational Model

The IRM is a biclustering model, with rows and columns clustered according to DP.

\[
\begin{align*}
\{\mu_1 > \mu_2 > \cdots \} & \sim \text{Poisson}(\alpha \mu^{-1}, (0, 1)) \\
\{\nu_1 > \nu_2 > \cdots \} & \sim \text{Poisson}(\beta \nu^{-1}, (0, 1)) \\
\theta_{kl} & \sim H \\
\theta(\xi, \eta) &= \theta_{kl}
\end{align*}
\]

where \( \xi \in (\mu_k, \mu_{k-1}) \), \( \eta \in (\nu_l, \nu_{l-1}) \)

\[
\begin{align*}
\xi_i & \sim \text{Uniform}[0, 1] \\
\eta_j & \sim \text{Uniform}[0, 1]
\end{align*}
\]

- \( \theta(\xi, \eta) \) is a piecewise constant function on \([0, 1]^2\).
- Parameter for \( R_{i,j} \) is \( \theta(\xi_i, \eta_j) \).
- Linear time per gibbs sweep complexity [FIXME cite Vikash].
Goal: Self-consistent, multiresolution (ideally hierarchical) relational modeling

Not satisfied by IRM, coalescent, annotated hierarchies

Approach: Build infinitely exchangeable stochastic process for relational data
Exchangeable Sequences

An infinitely exchangeable sequence of variables $X_1, X_2, \ldots$ is invariant to permutations:

$$p(X_{1:n} = x_{1:n}) = p(X_{1:n} = x_{\sigma(1:n)})$$

for all $n \geq 1$ and $\sigma \in S_n$.

Examples: measurements, images, iid samples from density or generative model.

**Theorem (de Finetti)**

There is a latent variable $\theta$ such that:

$$p(X_{1:n}) = \int p(\theta) \prod_{i=1}^{n} p(X_i|\theta) \, d\theta$$
Exchangeability in Relational Data

A relation $[R_{i,j}]_{i,j \geq 1}$ is **separately exchangeable** if it is invariant to separate permutations on rows and columns:

$$p(R_{1:n,1:m}) = p(R_{\sigma(1:n),\pi(1:m)})$$

for all $n, m \geq 1$, $\sigma \in S_n$, $\pi \in S_m$.

It is **jointly exchangeable** if:

$$p(R_{1:n,1:n}) = p(R_{\sigma(1:n),\sigma(1:n)})$$

for all $n \geq 1$, $\sigma \in S_n$.

- Separately exchangeable relations are jointly exchangeable, but not vice versa.
Aldous-Hoover Representation

**Theorem**

A separately exchangeable relation $R$ has the following representation:

\[ \xi_i = \text{latent representation of } i^{th} \text{ row} \]
\[ \eta_j = \text{latent representation of } j^{th} \text{ column} \]
\[ \theta = \text{latent representation of relation} \]

such that

\[
p(R_{1:n,1:m}) = \int p(\theta) \prod_{i=1}^{n} p(\xi_i) \prod_{j=1}^{m} p(\eta_j) \cdot \prod_{i,j} p(R_{i,j}|\theta, \xi_i, \eta_j) \, d\xi_{1:n} \, d\eta_{1:m} \, d\theta.
\]
Theorem

A jointly exchangeable relation $R$ has the following representation:

$$
\begin{align*}
\xi_i &= \text{latent representation of } i^{\text{th}} \text{ row/column} \\
\theta &= \text{latent representation of relation}
\end{align*}
$$

such that

$$
p(R_{1:n,1:n}) = \int p(\theta) \prod_{i=1}^{n} p(\xi_i) \prod_{i \leq j} p(R_{i,j}, R_{j,i} | \theta, \xi_i, \xi_j) \, d\xi_{1:n} \, d\theta
$$

where $p(R_{i,j}, R_{j,i} | \theta, \xi_i, \xi_j) = p(R_{j,i}, R_{i,j} | \theta, \xi_j, \xi_i)$.
Properties of the Mondrian Process

- It is a multi-dimensional generalization of the Poisson process on simply connected spaces.
- It is self-consistent wrt to restriction and slicing.
- Consistency means that it can be “extended” to an infinite-dimensional object.
- Consistency also means that the conditional Mondrian is also of nice form.
- Different base spaces yield different models with different inductive biases.
The Mondrian Process on the Unit Square

The **Mondrian process** *hierarchically clusters* relations. We can think of the Mondrian as *recursively cutting* $[0, 1]^2$ into two halves, using only horizontal and vertical cuts.
Generating Poisson Processes

Consider a Poisson process over interval \((a, b)\) with varying rate \(\mu\) of finite total mass \(\mu_0 = \mu((a, b))\).

- Split \((a, b)\) into tiny intervals, each interval \(dx\) draws a Bernoulli with mean \(\mu(dx)\).
- Draw \(n \sim \text{Poisson}(\mu_0)\), draw \(x_i \sim \mu/\mu_0\) for \(i = 1, \ldots, n\).
- If constant rate: \(x_0 = a\); draw \(x_i \sim x_{i-1} + \text{Exp}(\mu)\) until \(x_i > b\).

Break-and-Branch: \(\text{Poisson}(\lambda, \mu, (a, b))\)

1. Start with \(\lambda = 1\).
2. Draw \(E \sim \text{Exp}(\mu_0)\) where \(\mu_0 = \mu((a, b))\), and let \(\lambda' = \lambda - E\).
3. If \(\lambda' < 0\) stop, else:
4. Draw \(x \sim \mu/\mu_0\).
5. Return \(\{x\} \cup \text{Poisson}(\lambda', \mu, (a, x)) \cup \text{Poisson}(\lambda', \mu, (x, b))\).
Generating Poisson Processes

Consider a Poisson process over interval \((a, b)\) with varying rate \(\mu\) of finite total mass \(\mu_0 = \mu((a, b))\).

- Split \((a, b)\) into tiny intervals, each interval \(dx\) draws a Bernoulli with mean \(\mu(dx)\).
- Draw \(n \sim \text{Poisson}(\mu_0)\), draw \(x_i \sim \mu/\mu_0\) for \(i = 1, \ldots, n\).
- If constant rate: \(x_0 = a\); draw \(x_i \sim x_{i-1} + \text{Exp}(\mu)\) until \(x_i > b\).

**Break-and-Branch:** Poisson\((\lambda, \mu, (a, b))\)

1. Start with \(\lambda = 1\).
2. Draw \(E \sim \text{Exp}(\mu_0)\) where \(\mu_0 = \mu((a, b))\), and let \(\lambda' = \lambda - E\).
3. If \(\lambda' < 0\) stop, else:
4. Draw \(x \sim \mu/\mu_0\).
5. Return \(\{x\} \cup \text{Poisson}(\lambda', \mu, (a, x)) \cup \text{Poisson}(\lambda', \mu, (x, b))\).
The Mondrian Process

\[ \text{Mondrian}(\lambda, \mu_{1:D}, \Theta_{1:D}) \]

1. Start with \( \lambda = 1 \).
2. Draw \( E \sim \text{Exp}(\sum_d \mu_d(\Theta_d)) \), and let \( \lambda' = \lambda - E \).
3. If \( \lambda' < 0 \) return \( \langle \Theta_{1:D} \rangle \), else:
4. Draw \( d \) with \( p(d) \propto \mu_d(\Theta_d) \).
5. Draw \( x \sim \mu_d/\mu_d(\Theta_d) \).
6. Return
   \[ \langle \lambda', d, x, \text{Mondrian}(\lambda', \mu_{1:D}, \Theta_{1:D}^{\leq dX}), \text{Mondrian}(\lambda', \mu_{1:D}, \Theta_{1:D}^{> dX}) \rangle. \]
Modelling a Mondrian with a Mondrian
Modelling Diplomatic Relations

food/animals crude materials
minerals/fuels basic goods
diplomats

UK JAPAN SPAIN USA YUGOSWITZ FINLA ISRAE EGYPT ALGER CZECH SYRIA PAKIS NEWZE THAIL INDON CHINA ARGEN ECUAD MADAG BRAZL YUGOS INDON USA ISRAE ECUAD ALGER
The Mondrian Process for Hierarchical Relational Modeling

- Sample a n-coalescent from Kingman’s prior.
- Run the Mondrian Process on the resulting tree-structured, simply connected space.
- Self consistent prior on partitions, by restriction self-consistency of Mondrian and coalescent infinite exchangeability
- Inference: write coalescent, Mondrian in Church. Use generic MH.
Hierarchical Relational Modeling
Conclusions

- **Challenges:**
  - Representation: Marginalized (as in DP-Poisson to stick-breaking to CRP)
  - Inference: Exploit marginalized representations; alternatives?

- **Contributions:**
  - Defined Mondrian: simply connected generalization of Poisson
    - Unit square for FIXME
    - Randomly sampled coalescent for hierarchical relational modeling
  - Connected classical representation theorems to relational modeling
  - MCMC inference: special purpose, and Church generic
