The Indian Buffet Process

Eli Bingham[1] Matt Dickenson[2]

[1]University of North Carolina

[2]Duke University

February 10, 2014

Outline

1. Introduction
2. Dirichlet and Chinese Restaurant Processes
3. Beta and Indian Buffet Processes
4. Gibbs sampling
5. Demonstration/Visualization
6. Applications: Choice Behavior and Collaborative Filtering
7. Extensions: Topic Models and Cascading IBP
8. Discussion

Example

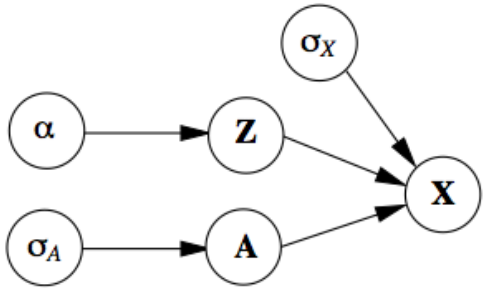


Figure: Griffiths and Ghahramani (2011) Figure 7

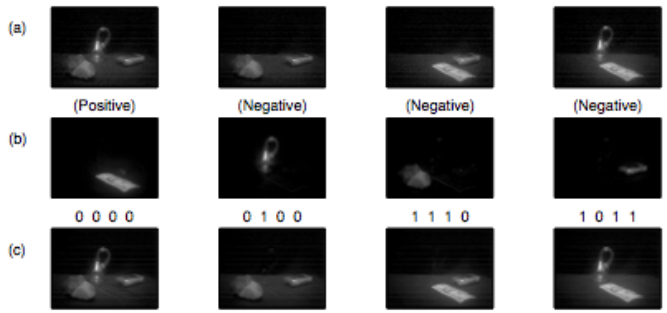


Figure: Griffiths and Ghahramani (2011) Figure 9

Motivation

When and why would we use IBP?

- As a prior on sparse binary matrices with a countably infinite number of columns

Introduction

Indian Buffet Process:

1. N customers enter (in sequence) a buffet restaurant with an infinite number of dishes
2. First customer fills her plate with Poisson(α) number of dishes
3. i^{th} customer samples dishes in proportion to their popularity, with probability $\frac{m_k}{i}$, where m_k is the number of previous customers who sampled dish k
4. i^{th} customer then samples a Poisson($\frac{\alpha}{i}$) number of new dishes

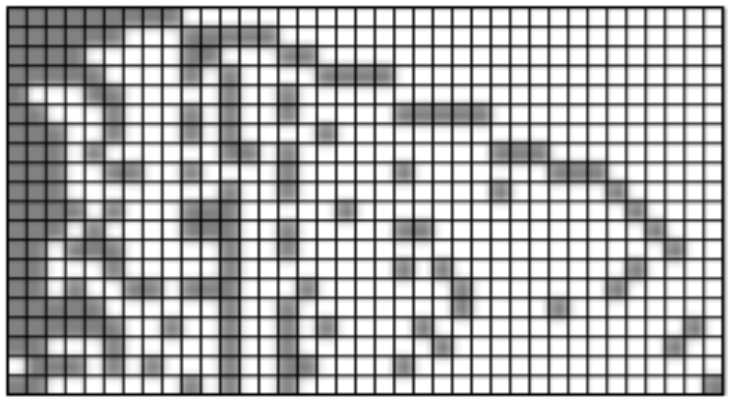


Figure: Griffiths and Ghahramani (2011) Figure 5

Background: Dirichlet Process

Finite version (Dirichlet distribution):

- Assignment of an object to a class is independent of all other assignments: $P(c|\theta) = \prod_{i=1}^{N} P(c_i|\theta) = \prod_{i=1}^{N} \theta_{c_i}$
- $\theta|\alpha \sim \text{Dirichlet}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})$ (if symettric)
- $c_i|\theta \sim \text{Discrete}(\theta)$, where Discrete : Bernoulli :: Multinomial : Binomial

Integrating out θ: $P(c) = \frac{\prod_{k=1}^{K} \Gamma(m_k + \frac{\alpha}{K})}{\Gamma(\frac{\alpha}{K})^K} \frac{\Gamma(\alpha)}{\Gamma(N+\alpha)}$

What happens as $K \to \infty$?

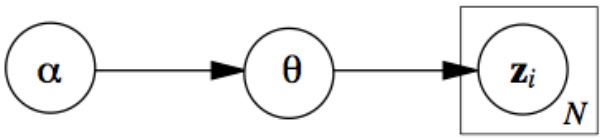


Figure: Griffiths and Ghahramani (2011) Figure 1

Background: Chinese Restaurant Process

1. N customers enter (in sequence) a restaurant with an infinite number of tables, each with infinite seating
2. First customer sits at first table with probability $\frac{\alpha}{\alpha} = 1$
3. i^{th} customer sits at the k^{th} table with probability $\frac{m_k}{i+\alpha-1}$, where m_k is the number of previous customers who sat at table k, or a new table with probability $\frac{\alpha}{i+\alpha-1}$

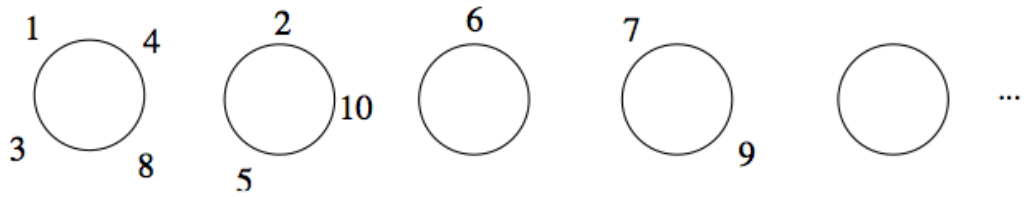


Figure: Griffiths and Ghahramani (2011) Figure 2

Limitation: each object (customer) can only belong to one class (table).

Beta Process

Indian Buffet Process

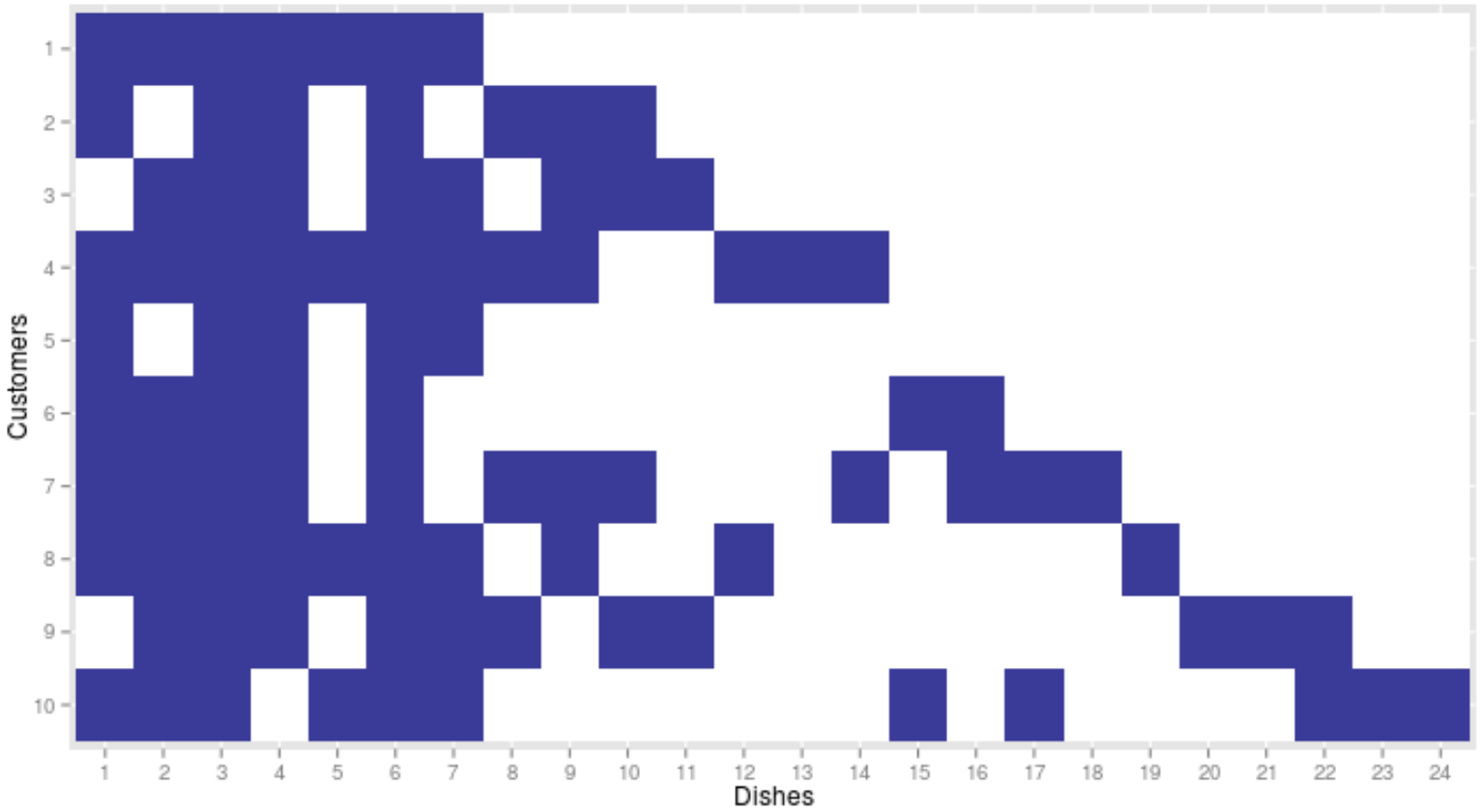
Stick-Breaking Construction of IBP

1. Recursively break (an initially unit-length) stick, breaking off a Beta($\alpha, 1$) portion at each step
2. Let each portion of the "stick", π_k represent the probability of each feature (sorted from largest to smallest)

This helps to show the relation between the Dirichlet process and the IBP. The stick-breaking construction is also useful for defining inference algorithms.

Demo

mcdickenson.shinyapps.io/ibp-demo



Properties of the Resulting Distribution

- The "effective" dimension $K_+ \sim \text{Poisson}(\alpha H_N)$
- The number of dishes on each customer's plate is distributed Poisson(α) (by exchangeability)
- $\mathbf{Z}$ remains sparse as $K \to \infty$: effective dimensions of $\mathbf{Z}$ are $N \times K_+$, and the expected number of entries is $N\alpha$

Inference by Gibbs Sampling

Variational Inference

Application 1: Choice Behavior

"A Choice Model with Infinitely Many Latent Features"
(Görür, Jäkel, and Rasmussen, ICML 2006)

- Customers compare items (e.g. cell phones) based on the (binary) features of each; more features are better
- Number of features is potentially infinite and ordering is not important, so IBP is used
- Celebrity example: "With whom would you prefer to spend an hour of conversation?"

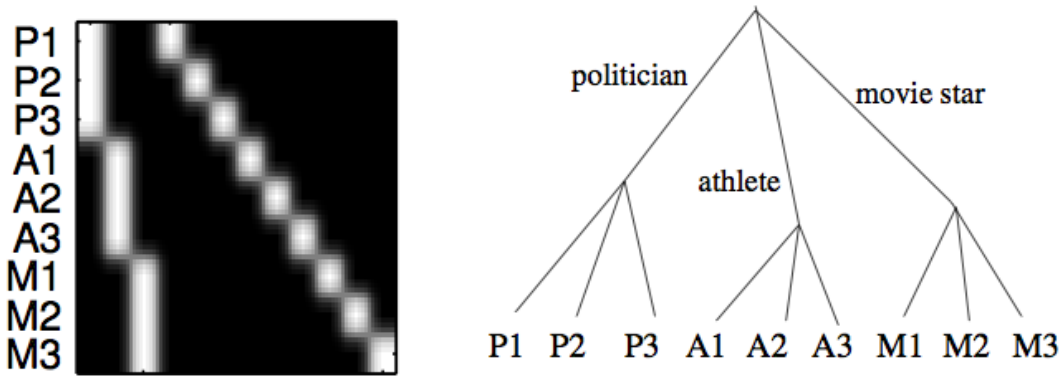


Figure: Görür, Jäkel, and Rasmussen (2006) Figure 3

Application 2: Collaborative Filtering

Extension 1: Topic Modeling

"The IBP Compound Dirichlet Process
and its Application to Focused Topic Modeling"
Williamson, Wang, Heller, and Blei (2010)

Stick-breaking construction:

$$\mu_k \sim \text{Beta}(\alpha, 1)$$

$$\pi_k = \prod_{j=1}^{k} \mu_j$$

$$b_{m,k} \sim \text{Bernoulli}(\pi_k)$$

Extension 1: Topic Modeling

"The IBP Compound Dirichlet Process
and its Application to Focused Topic Modeling"
Williamson, Wang, Heller, and Blei (2010)

Focused topic model:

1. for $k = 1, 2, \ldots$
   - Sample stick length π_k
   - Sample relative mass $\phi_k \sim \text{Gamma}(\gamma, 1)$
   - Draw topic distribution over words: $\beta_k \sim \text{Dirichlet}(\eta)$
2. for $m = 1, \ldots, M$
   - Sample binary vector b_m
   - Draw total number of words $n^{(m)} \sim NB(\sum_k b_{m,k}\phi_k, 1/2)$
   - Sample distribution over topics $\theta_m \sim \text{Dirichlet}(b_m \cdot \phi)$
   - For each word $w_{m,i}, i = 1, \ldots, n^{(m)}$
     1. Draw topic index $z_{m,i} \sim \text{Discrete}(\theta_m)$
     2. Draw word $w_{m,i} \sim \text{Discrete}(\beta_{z_{m_i}})$

Extension 1: Topic Modeling

"The IBP Compound Dirichlet Process
and its Application to Focused Topic Modeling"
Williamson, Wang, Heller, and Blei (2010)

An advantage of the focused topic model is that it separates the global topic proportions from the distribution over topics within a topic. A rare topic within the corpus can be dominant within a document (e.g. baseball), and a frequent topic can be a small proportion of many documents.

Extension 2: Collaborative Filtering

Discussion

Limitations of IBP:

1. Coupling of average number of features α and total number of features $N\alpha$ (can be overcome with a two-parameter generalization)
2. Computationally complex, can be time-consuming

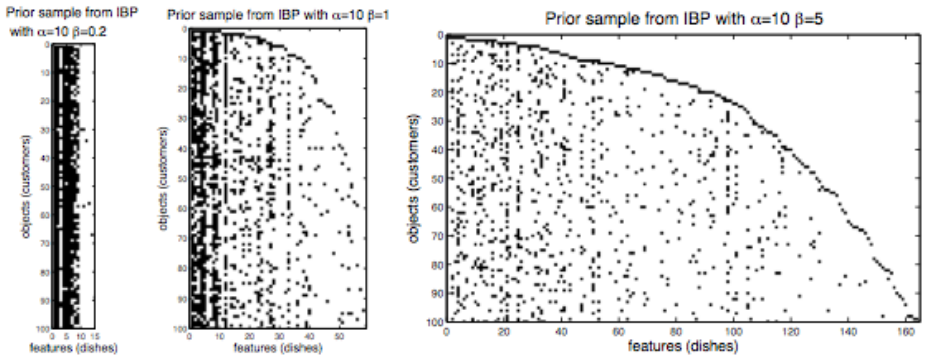


Figure: Griffiths and Ghahramani (2011) Figure 10