

The Indian Buffet Process

Eli Bingham¹ Matt Dickenson²

¹University of North Carolina

²Duke University

February 10, 2014

- 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

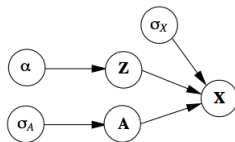


Figure: Griffiths and Ghahramani (2011) Figure 7

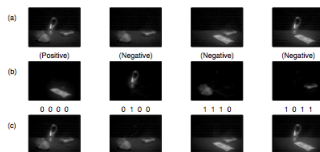


Figure: Griffiths and Ghahramani (2011) Figure 9

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 $\text{Poisson}(\alpha)$ 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 $\text{Poisson}(\frac{\alpha}{i})$ number of new dishes

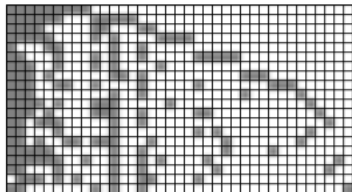


Figure: Griffiths and Ghahramani (2011) Figure 5

Background: Dirichlet Process

Finite version (Dirichlet distribution):

- Assignment of an object to a class is independent of all other assignments: $P(c|\theta) = \prod_{i=1}^N P(c_i|\theta) = \prod_{i=1}^N \theta_{c_i}$
- $\theta|\alpha \sim \text{Dirichlet}(\frac{\alpha}{K}, \dots, \frac{\alpha}{K})$ (if symmetric)
- $c_i|\theta \sim \text{Discrete}(\theta)$, 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 $K \rightarrow \infty$?

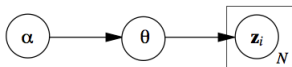


Figure: Griffiths and Ghahramani (2011) Figure 1

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 $\frac{\alpha}{i+\alpha-1}$

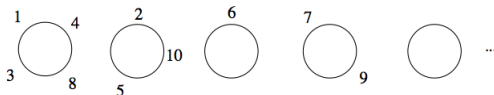


Figure: Griffiths and Ghahramani (2011) Figure 2

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

Beta Process

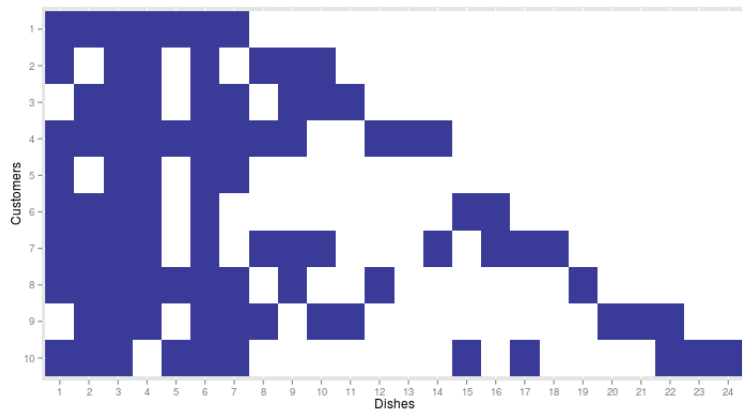
Indian Buffet Process

Stick-Breaking Construction of IBP

- 1 Recursively break (an initially unit-length) stick, breaking off a $\text{Beta}(\alpha, 1)$ 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.

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 $\text{Poisson}(\alpha)$ (by exchangeability)
- \mathbf{Z} remains sparse as $K \rightarrow \infty$: 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?”

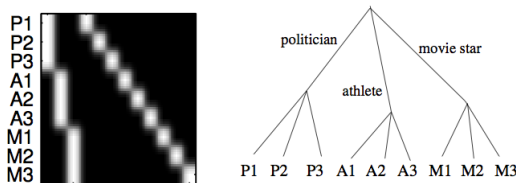


Figure: Görür, Jäkel, and Rasmussen (2006) Figure 3

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:

$$\begin{aligned}\mu_k &\sim \text{Beta}(\alpha, 1) \\ \pi_k &= \prod_{j=1}^k \mu_j \\ b_{m,k} &\sim \text{Bernoulli}(\pi_k)\end{aligned}$$

Extension 1: Topic Modeling

Focused topic model:

- ① for $k = 1, 2, \dots$
 - 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 $m = 1, \dots, M$
 - Sample binary vector b_m
 - Draw total number of words $n^{(m)} \sim \text{NB}(\sum_k b_{m,k} \phi_k, 1/2)$
 - Sample distribution over topics $\theta_m \sim \text{Dirichlet}(b_m \cdot \phi)$
 - For each word $w_{m,i}, i = 1, \dots, n^{(m)}$
 - 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

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

Limitations of IBP:

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

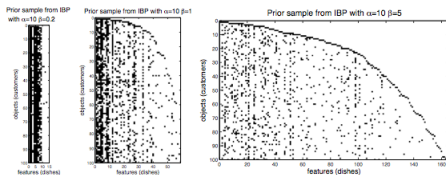


Figure: Griffiths and Ghahramani (2011) Figure 10