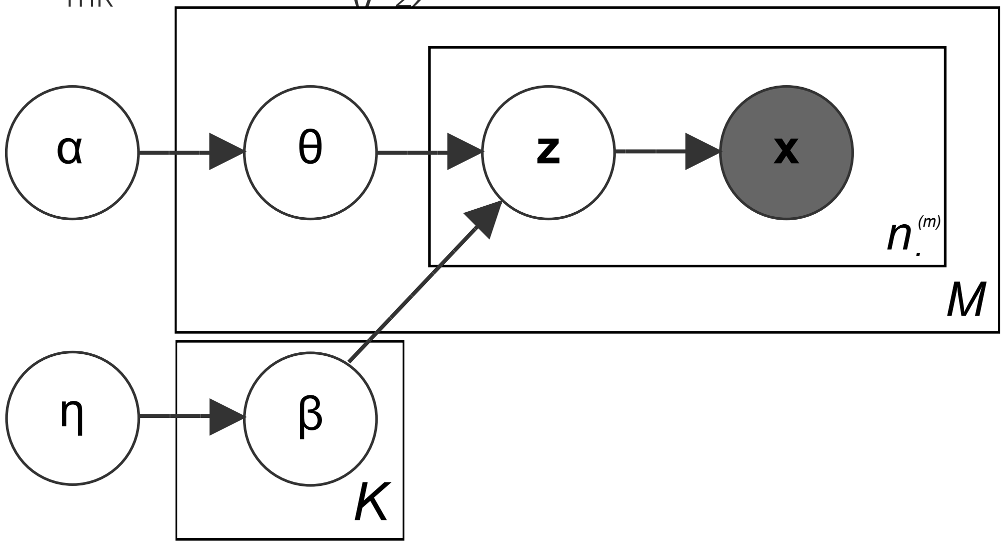
Topic models

- Topic models describe documents using a distribution over features.
- Each feature is a distribution over words
- Each document is represented as a collection of words (usually unordered – “bag of words” assumption).
- The words within a document are distributed according to a document-specific mixture model
  - Each word in a document is associated with a feature.
- The features are shared between documents.
- The features learned tend to give high probability to semantically related words – “topics”

Latent Dirichlet allocation

- For each topic $k=1,\ldots,K$
  - Sample a distribution over words, $\beta \sim \text{Dir}(\eta_1,\ldots, \eta_V)$
- For each document $m=1,\ldots,M$
  - Sample a distribution over topics, $\theta_m \sim \text{Dir}(\alpha_1,\ldots,\alpha_K)$
  - For each word $n=1,\ldots,N_m$
    - Sample a topic $z_{mn} \sim \text{Discrete}(\theta_m)$
    - Sample a word $w_{mk} \sim \text{Discrete}(\beta_z)$



Blei et al, 2002

Constructing a topic model with infinitely many topics

- LDA: Each distribution is associated with a distribution over K topics.
- Problem: How to choose the number of topics?
- Solution:
  - Infinitely many topics!
  - Replace the Dirichlet distribution over topics with a Dirichlet process!
- Problem: We want to make sure the topics are *shared* between documents

Sharing topics

- In LDA, we have M independent samples from a Dirichlet distribution.
- The weights are different, but the topics are fixed to be the same.
- If we replace the Dirichlet distributions with Dirichlet processes, each atom of each Dirichlet process will pick a topic *independently* of the other topics.

Sharing topics

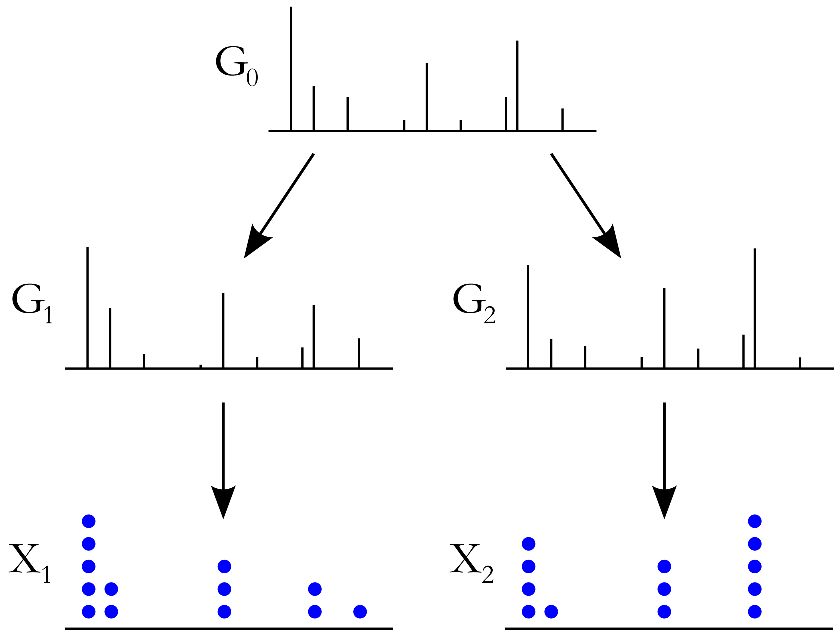
- Because the base measure is *continuous*, we have zero probability of picking the same topic twice.
- If we want to pick the same topic twice, we need to use a *discrete* base measure.
- For example, if we chose the base measure to be $H = \sum_{k=1}^{K} \alpha_k \delta_{\beta_k}$, then we would have LDA again.

- We want there to be an infinite number of topics, so we want an *infinite, discrete* base measure.
- We want the location of the topics to be random, so we want an *infinite, discrete, random* base measure.

Hierarchical Dirichlet Process (Teh et al, 2006)

- Solution: Sample the base measure from a Dirichlet process!

$$G_0 \sim \mathrm{DP}(\gamma, H)$$
$$G_m \sim \mathrm{DP}(\alpha, G_0)$$

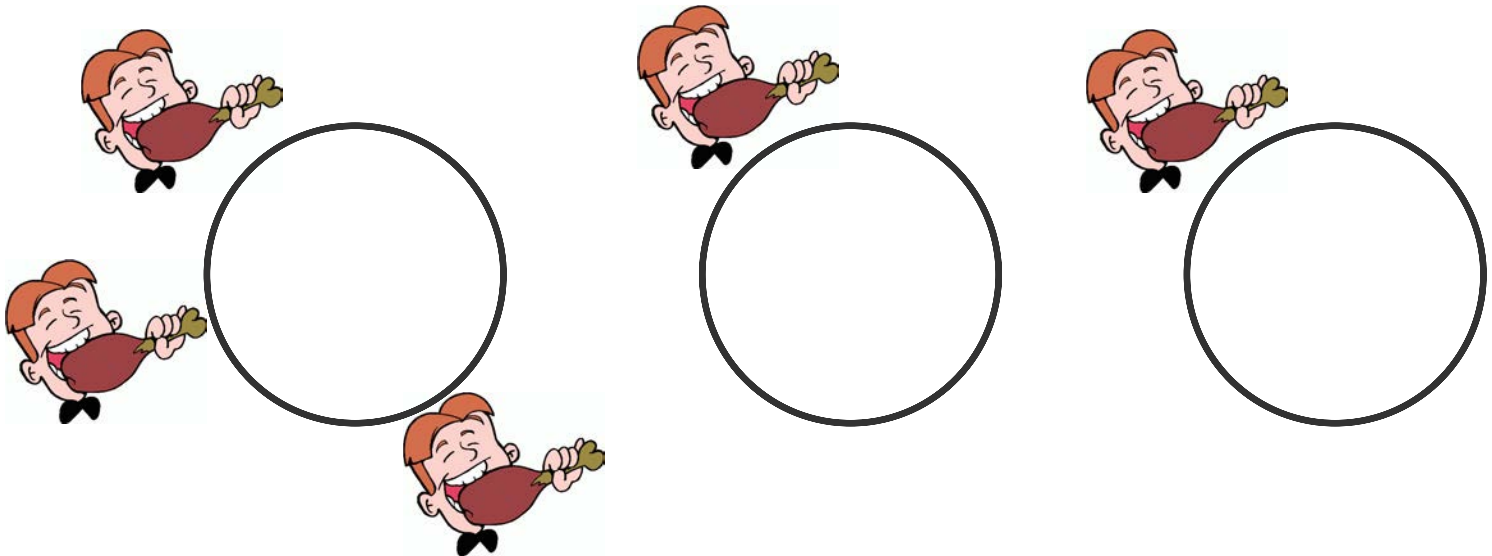


Chinese restaurant franchise

- Imagine a *franchise* of restaurants, serving an infinitely large, global menu.
- Each table in each restaurant orders a single dish.
- Let n_{rt} be the number of customers in restaurant r sitting at table t.
- Let m_{rd} be the number of tables in restaurant r serving dish d.
- Let $m_{.d}$ be the number of tables, across *all* restaurants, serving dish d.

Chinese restaurant franchise

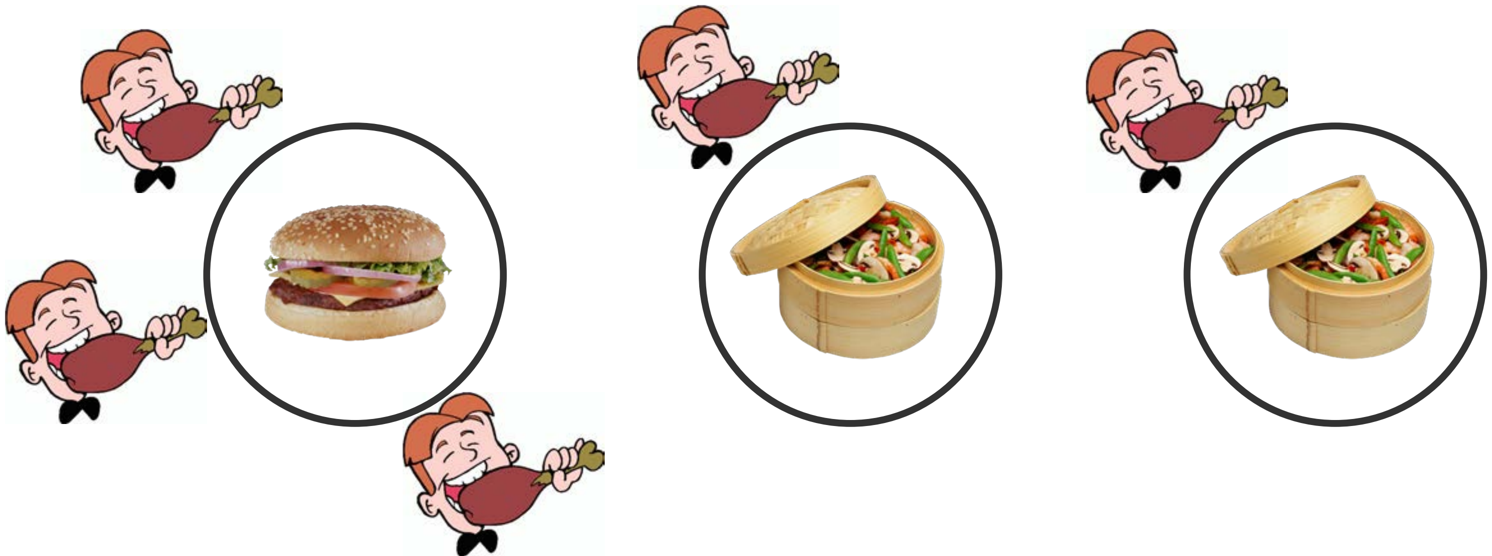
- Customers enter the restaurants, and sit at tables according to the Chinese restaurant process
  - The first customer enters a restaurant, and picks a table.
  - The n^{th} customer enters the restaurant. He sits at an existing table with probability $m_k/(n-1+\alpha)$, where m_k is the number of people sat at table k. He starts a new table with probability $\alpha/(n-1+\alpha)$.



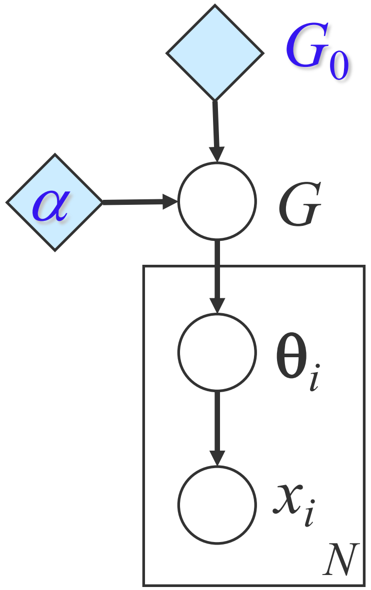
Chinese restaurant franchise

- Each *table* in each restaurant picks a *dish*, with probability proportional to the number of times it has been served across *all* restaurants.

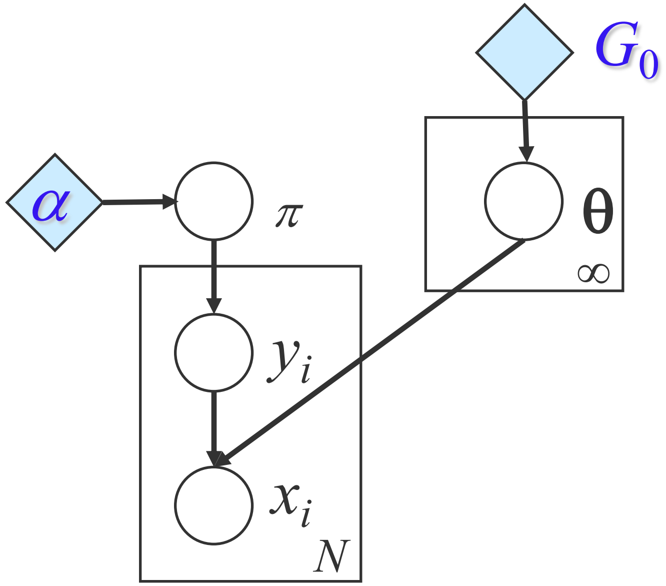
$$p(\text{table } t \text{ chooses dish } d|\text{previous tables}) = \begin{cases} \frac{m_d}{T+\gamma} & \text{for an existing table} \\ \frac{\gamma}{T+\gamma} & \text{for a new table} \end{cases}$$



Recall: Graphical Model Representations of DP

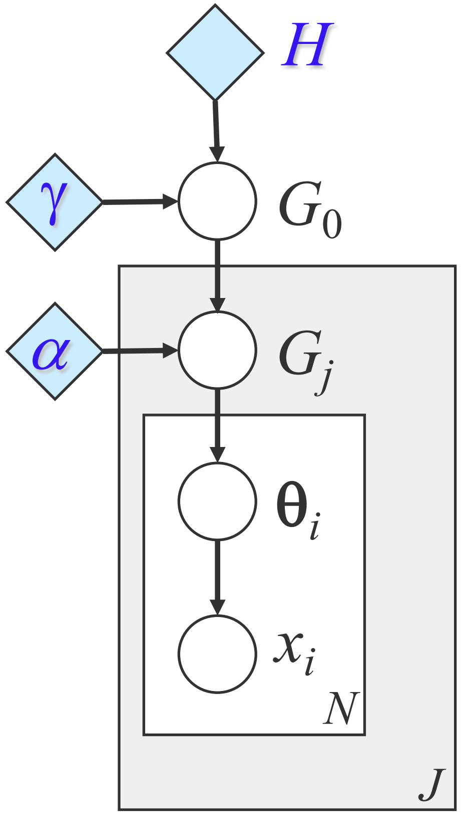


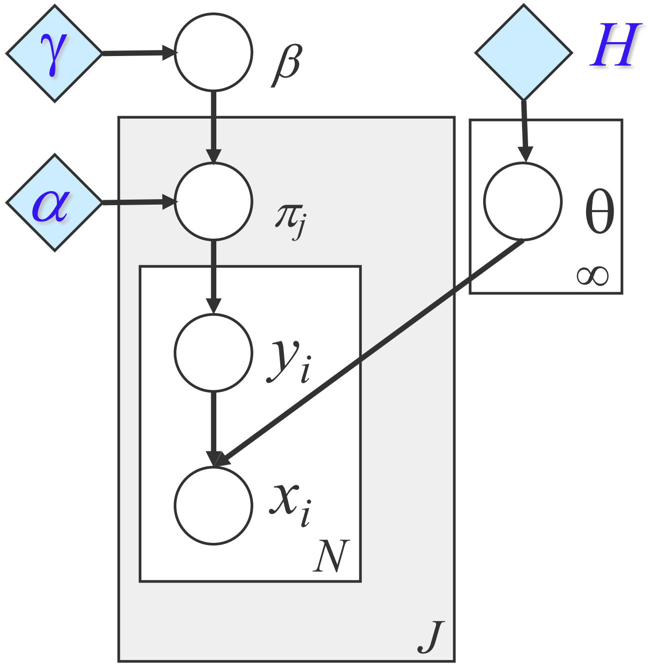
The Pólya urn construction



The Stick-breaking construction

Hierarchical DP Mixture





$\theta_k \sim H$

$\beta = \text{Stick}(\gamma), G_0 = \sum_{k=1}^{\infty} \beta_k \delta(\theta_k)$

$\pi_j = \text{Stick}(\alpha, \beta), G_j = \sum_{k=1}^{\infty} \pi_k \delta(\theta_k)$

Stick(α, β):

$\pi'_{jk} \sim \text{Beta}\left(\alpha\beta_k, \alpha\left(1 - \sum_{l=1}^{k} \beta_l\right)\right), \quad \pi_{jk} = \pi'_{jk} \prod_{l=1}^{k-1} \left(1 - \pi'_{jl}\right).$

An infinite topic model

- Restaurants = documents; dishes = topics.

- Let H be a V-dimensional Dirichlet distribution, so a sample from H is a distribution over a vocabulary of V words.
- Sample a global distribution over topics,

$$G_0 := \sum_{k=1}^{\infty} \pi_k \delta_{\beta_k} \sim \mathrm{DP}(\alpha, H)$$

- For each document $m=1,\dots,M$
  - Sample a distribution over topics, $G_m \sim DP(\gamma, G_0)$.
  - For each word $n=1,\dots,N_m$
    - Sample a topic $\phi_{mn} \sim \text{Discrete}(G_0)$.
    - Sample a word $w_{mk} \sim \text{Discrete}(\phi_{mn})$.