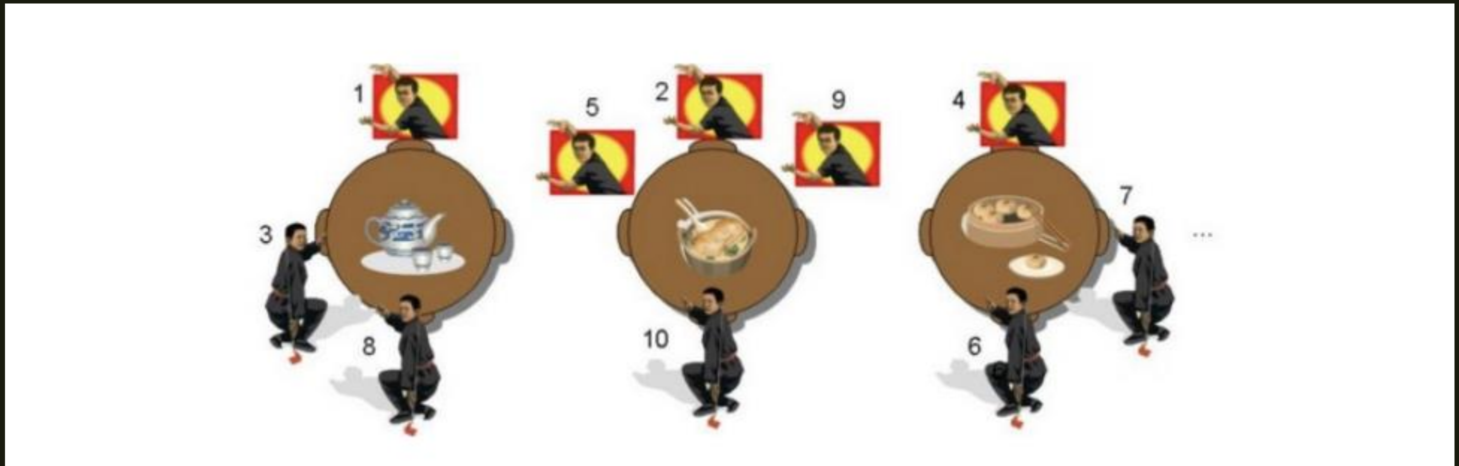
Chinese Restaurant Process,
Stick Breaking Process,
and Dirichlet Process Mixture Model

•Inspiring example

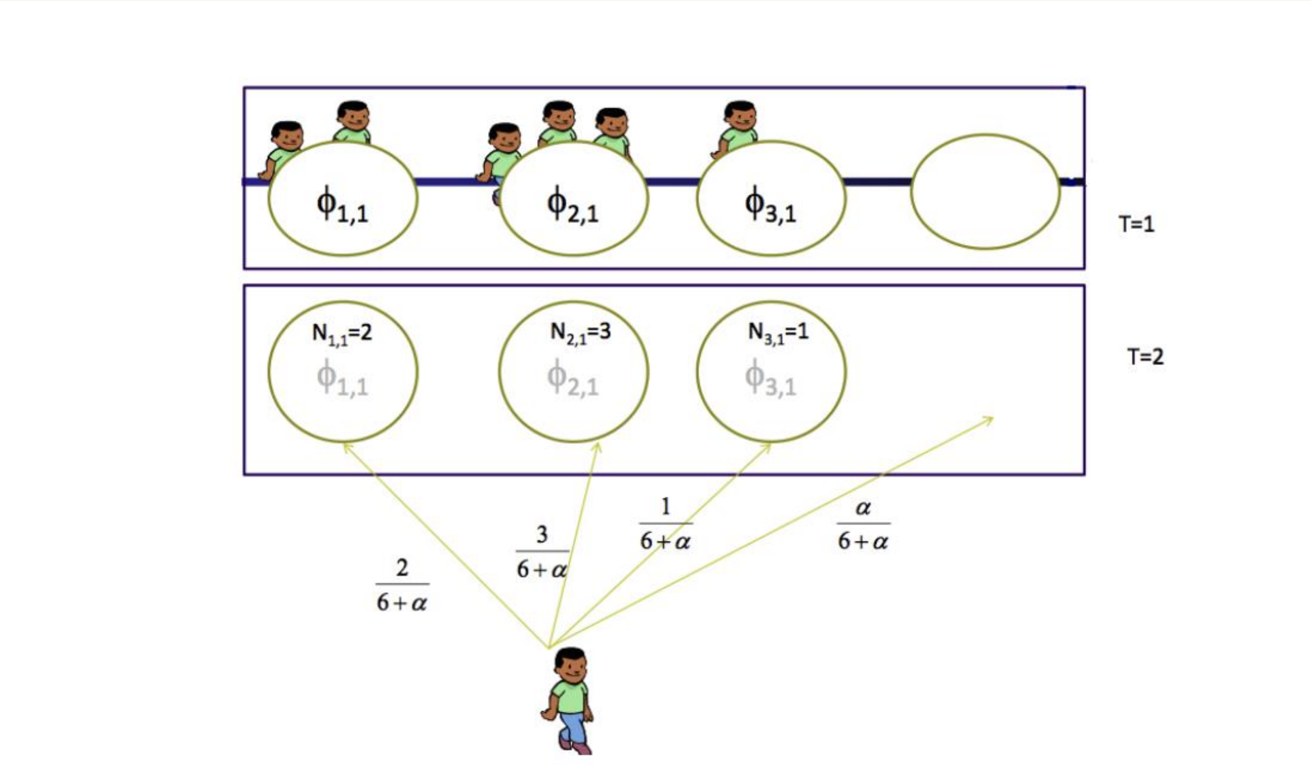
Suppose we asked one hundred people which number was the most unlucky. Of those people, fifty said ‘13’, forty said ‘4’, and ten said ‘87’. These individual differences are unlikely to be caused by chance. Instead, what we appear to have are three distinct groups of people. How should we model these data?

Possible approach:

1. All partitions equally likely
2. Fixed number of groups
3. Prior over the number of groups
4. Model the probabilities associated with each group



CHINESE RESTAURANT PROCESS



Exchangeability

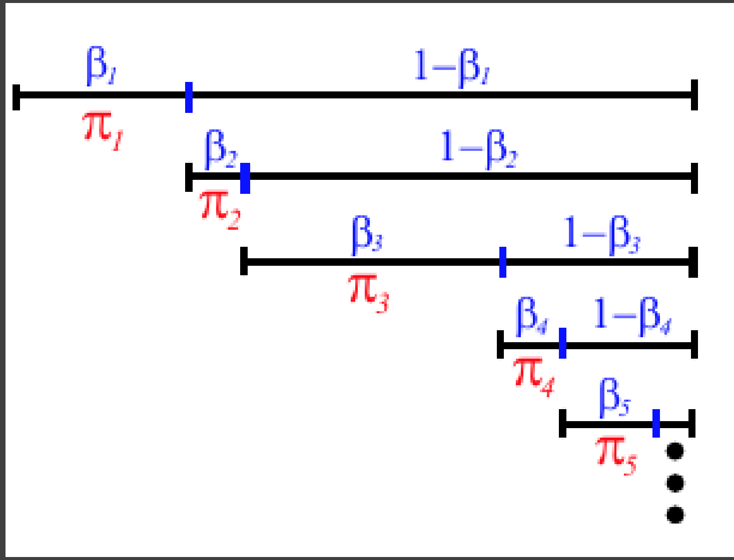
Probability of…

Join an existing table $\frac{c}{n-1+\alpha}$

Join a new table $\frac{\alpha}{n-1+\alpha}$

simulation

Stick Breaking Process



Define an infinite sequence of beta random variables $\beta_k \sim Beta(1, \alpha_0)$, and define an infinite sequence of mixing proportions as $\pi_1 = \beta_1, \pi_k = \beta_k \Pi_{l=1}^{k-1}(1-\beta_l), for\ k = 2, 3, \ldots$

Differences between CRP and Stick Breaking Process

Similarities:

- Both the CRP and SBP are used to model probability distributions over partitions of a set of objects.
- Both are used in Bayesian nonparametric models, which allows for flexibility in the number of clusters or components in the mixture model.
- Both are based on the idea of assigning weights to each cluster or component in the mixture model.

Differences:

- The CRP is a discrete distribution, while the SBP is a continuous distribution. In the CRP, the weights are drawn from a discrete probability distribution, while in the SBP, the weights are drawn from a continuous beta distribution.
- The CRP assigns objects to clusters with probabilities proportional to the number of objects already assigned to each cluster, while the SBP assigns weights to each cluster with probabilities proportional to the product of all the weights assigned to previous clusters.
- The CRP is often used in models such as clustering, where the number of clusters is unknown and can grow with the data. The SBP is often used in models such as topic modeling, where the number of topics is typically fixed and known ahead of time.

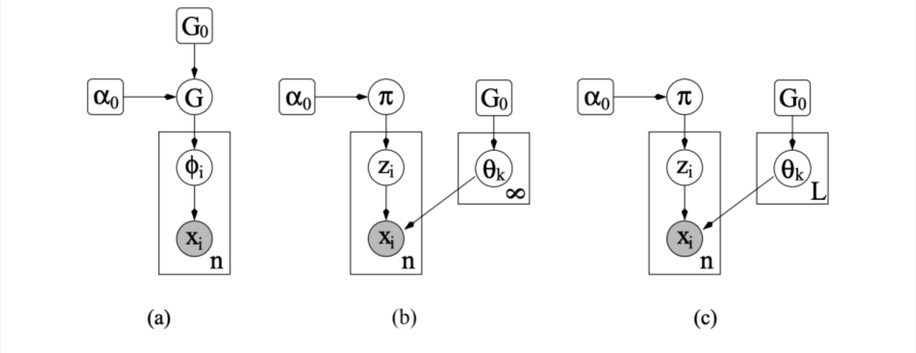
Dirichlet Process Mixture Model

We write G ~ DP(α_0, G_0) if G is a random probability measure distributed according to a DP.

G_0 is called the base measure of G (also known as the base distribution), and α_0 is the concentration parameter.

DP can be used in the mixture model setting in this way:

Suppose we have a set of data X = ($x_1, \dots, x_n$) and assume exchangeability. Given a draw G ~ DP(α_0, G_0) , independently draw n *latent factors* from G: $\phi_i \sim G$. Then, for each $i = 1, \dots, n$, draw $x_i \sim F(\phi_i)$, for any arbitrary distribution F.



SBP as a Special Case of DP

$$\pi \mid \alpha_0 \sim Stick(\alpha_0)$$
$$z_i|\pi \sim \pi$$
$$\theta_k|G_0 \sim G_0$$
$$x_i|z_i, (\theta_k)_{k=1}^{\infty} \sim F(\theta_{z_i})$$

Here $G = \Sigma_{k=1}^{\infty} \pi_k \delta_{\theta_k}$ and $\phi_i = \theta_{z_i}$

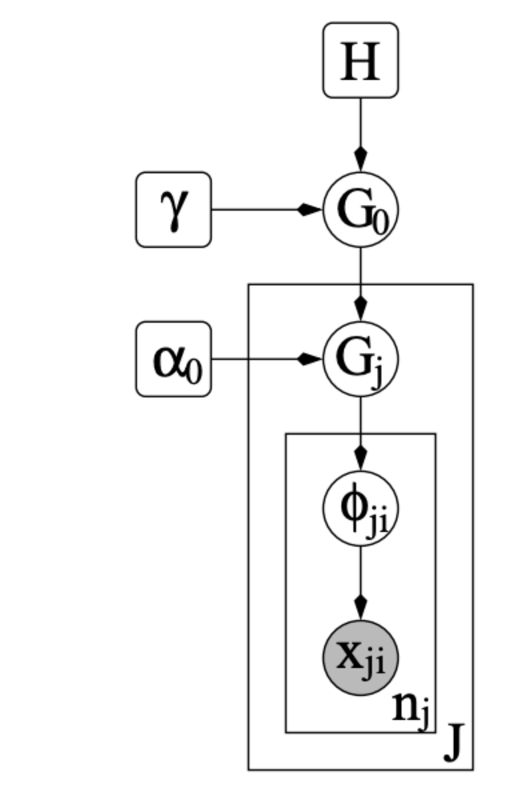
Gibbs Sampling in CRP and DPMM

•In the context of the Chinese Restaurant Process (CRP), Gibbs sampling can be used to sample from the posterior distribution of the clustering of the data points. In the CRP, data points are assigned to clusters based on a set of probabilities that depend on the existing cluster assignments and a concentration parameter. Gibbs sampling can be used to iteratively sample from the posterior distribution of the cluster assignments, given the current assignments and the data.

Gibbs Sampling in CRP and DPMM

•In the context of the Dirichlet Process Mixture Model (DPMM), Gibbs sampling can be used to sample from the posterior distribution of the mixture components. The DPMM is a probabilistic model that allows for infinite-dimensional mixtures, where the number of mixture components is not fixed but rather determined by the data. Gibbs sampling can be used to iteratively sample from the posterior distribution of the mixture assignments and the parameters of the mixture components, given the current assignments and the data.

HIERARCHICAL DIRICHLET PROCESSES



Applications

CRP:

- Clustering
- topic modeling
- image segmentation

SBP:

- Topic modeling
- time series modeling

REFERENCE

Teh, Yee, et al. "Sharing clusters among related groups: Hierarchical Dirichlet processes." *Advances in neural information processing systems* 17 (2004).

Neal, Radford M. "Markov chain sampling methods for Dirichlet process mixture models." *Journal of computational and graphical statistics* 9.2 (2000): 249-265.

https://www.stat.berkeley.edu/~jordan/653.pdf

Sethuraman, Jayaram. "A CONSTRUCTIVE DEFINITION OF DIRICHLET PRIORS." *Statistica Sinica*, vol. 4, no. 2, 1994, pp. 639–50. *JSTOR*, http://www.jstor.org/stable/24305538. Accessed 19 Mar. 2023.

Paisley, John. "A Tutorial on the Dirichlet Process for Engineers Technical Report." *Department of Electrical & Computer Engineering Duke University, Durham, NC* (2018): 1-23.

https://topicmodels.west.uni-koblenz.de/ckling/tmt/crp.html?parameters=1&dp=1#

THANK YOU!