Bayesian Nonparametrics II

Indian Buffet Process

Sarah M Brown

Electrical and Computer Engineering
Northeastern University

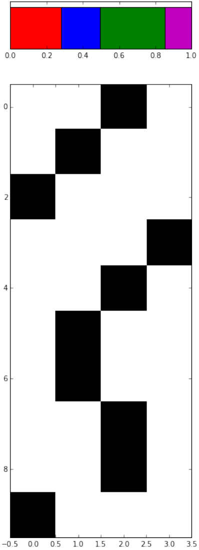
Summary

- Reviewed Gaussian Mixture Modeling
- GEM distribution is an infinite extension of the Dirichlet
- DPMM is a generative process using the GEM on cluster priors
- Stick-Breaking is a representation of the GEM or Dirichlet prior
- (mulitvariate) Poyla Urn is a representation of categorical marginals with Beta (or Dirichlet) prior
- Hoppe-Urn is a finite representation of the marginal with GEM prior
- CRP is a finite representation of the marginal with GEM prior

Motivating Example

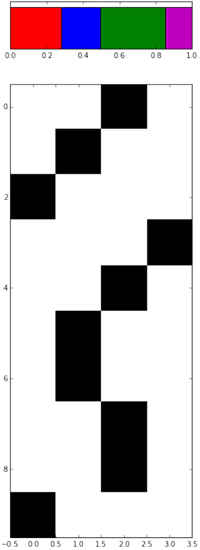
Many images each with some subset of 4 objects

From Clustering to Latent Feature Allocation

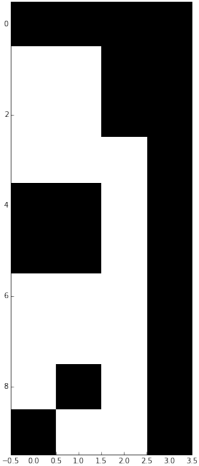


- Write cluster assignments as a binary matrix: $Z_{i,k} = 1$ if sample i belongs to cluster k

From Clustering to Latent Feature Allocation



- Write cluster assignments as a binary matrix: $Z_{i,k} = 1$ if sample i belongs to cluster k
- what if samples could belong to multiple latent groups?



Finite Latent Feature Allocation

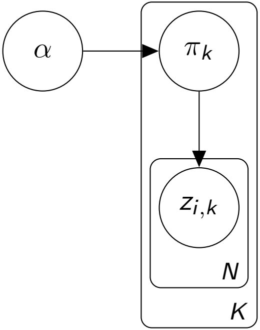
$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad (1)$$

$$z_{i,k} | \pi_k \sim \text{Ber}\left(\pi_k\right) \qquad (2)$$

Finite Latent Feature Allocation

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad (1)$$

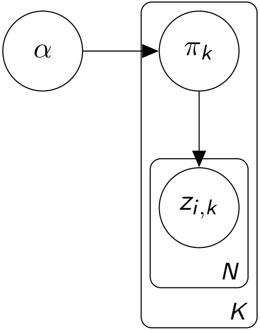
$$z_{i,k} | \pi_k \sim \text{Ber}(\pi_k) \qquad (2)$$



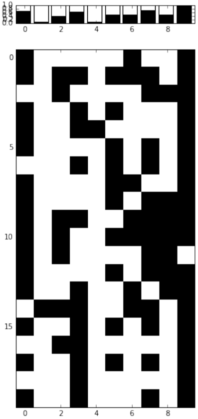
Finite Latent Feature Allocation

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \quad (1)$$

$$z_{i,k}|\pi_k \sim \text{Ber}(\pi_k) \quad (2)$$



$K = 10, N = 20, \alpha = 8$



Marginal on Z

for finite K

Model:

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right)$$
$$z_{i,k} | \pi_k \sim \text{Ber}(\pi_k)$$

Marginal on Z

for finite K

Model:

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right)$$
$$z_{i,k}|\pi_k \sim \text{Ber}\left(\pi_k\right)$$

Recall:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$
$$\Gamma(m) = (m-1)! m \in \mathcal{Z}$$
$$\Gamma(x) = x\Gamma(x-1) x > 0$$

Marginal on Z

for finite K

Model:

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right)$$

$$z_{i,k}|\pi_k \sim \text{Ber}\left(\pi_k\right)$$

Recall:

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(m) = (m-1)! m \in \mathcal{Z}$$

$$\Gamma(x) = x\Gamma(x-1) x > 0$$

So:

$$\begin{aligned}
P(Z) &= \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} p(z_{i,k}|\pi_k) \right) p(\pi_k) d\pi_k \\
&= \prod_{k=1}^{K} \frac{\mathrm{B}(n_k + \frac{\alpha}{K}, N - n_k + 1)}{\mathrm{B}(\frac{\alpha}{K}, 1)} \\
&= \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}
\end{aligned}$$

Marginal on Z

for finite K

Model:

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right)$$

$$z_{i,k}|\pi_k \sim \text{Ber}\left(\pi_k\right)$$

Recall:

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(m) = (m-1)! m \in \mathcal{Z}$$

$$\Gamma(x) = x\Gamma(x-1) x > 0$$

So:

$$P(Z) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} p(z_{i,k}|\pi_k) \right) p(\pi_k) d\pi_k$$

$$= \prod_{k=1}^{K} \frac{\mathrm{B}(n_k + \frac{\alpha}{K}, N - n_k + 1)}{\mathrm{B}(\frac{\alpha}{K}, 1)}$$

$$= \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- Follows from Beta-Binomial Conjugacy

Marginal on Z

for finite K

Model:

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right)$$
$$z_{i,k}|\pi_k \sim \text{Ber}\left(\pi_k\right)$$

Recall:

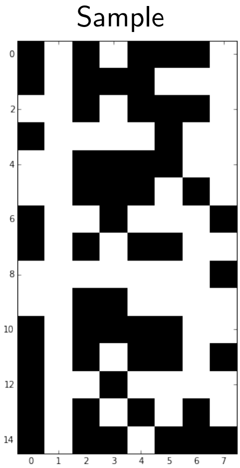
$$\text{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$
$$\Gamma(m) = (m-1)! m \in \mathcal{Z}$$
$$\Gamma(x) = x\Gamma(x-1) x > 0$$

So:

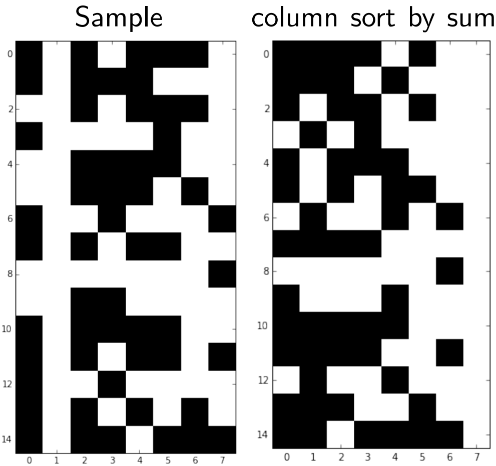
$$P(Z) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} p(z_{i,k}|\pi_k)\right) p(\pi_k) d\pi_k$$
$$= \prod_{k=1}^{K} \frac{\text{B}(n_k + \frac{\alpha}{K}, N - n_k + 1)}{\text{B}(\frac{\alpha}{K}, 1)}$$
$$= \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- Follows from Beta-Binomial Conjugacy
- Exchangeable, depends only on $n_k = \sum_{i=1}^{N} z_{i,k}$

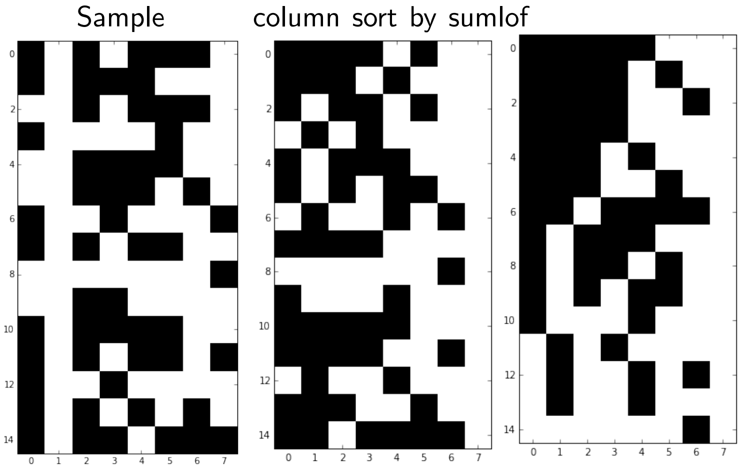
Left Ordered Form



Left Ordered Form



Left Ordered Form



Left Ordered Form

Properties:

- many to one mapping

Left Ordered Form

Properties:

- many to one mapping
- every Z has a unique lof

Left Ordered Form

Properties:

- many to one mapping
- every Z has a unique lof
- Can define an equivalence X and Y are lof equivalent if $lof(X) = lof(Y)$

Left Ordered Form

Properties:

- many to one mapping
- every Z has a unique lof
- Can define an equivalence X and Y are lof equivalent if $lof(X) = lof(Y)$
- Uses history: feature k at sample i is $(z_{1,k}, \ldots, z_{(i-1),k}$
- K_h is the number of features with history h

Left Ordered Form

Properties:

- many to one mapping
- every Z has a unique lof
- Can define an equivalence X and Y are lof equivalent if $lof(X) = lof(Y)$
- Uses history: feature k at sample i is $(z_{1,k}, \ldots, z_{(i-1),k}$
- K_h is the number of features with history h

New marginal:

$$P([Z]) = \sum_{Z \in [Z]} P(Z)$$
$$= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

Left Ordered Form

Properties:

- many to one mapping
- every Z has a unique lof
- Can define an equivalence X and Y are lof equivalent if $lof(X) = lof(Y)$
- Uses history: feature k at sample i is $(z_{1,k}, \ldots, z_{(i-1),k}$
- K_h is the number of features with history h

New marginal:

$$
\begin{aligned}
P([Z]) &= \sum_{Z \in [Z]} P(Z) \\
&= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}
\end{aligned}
$$

Marginal on Z

$K \to \infty$

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k} | \pi_k \sim \text{Ber}(\pi_k)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

Marginal on Z

$K \to \infty$

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k} | \pi_k \sim \text{Ber}(\pi_k)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- For $i = 1$, the chance of each feature k is independent
  $p(z_{1,k} = 1|\alpha) = \int \text{Ber}(\pi_k)\, \text{Beta}\left(\frac{\alpha}{K}, 1\right) = \frac{\alpha}{K}$

Marginal on Z

$K \to \infty$

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k} | \pi_k \sim \text{Ber}(\pi_k)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- For $i = 1$, the chance of each feature k is independent $p(z_{1,k} = 1|\alpha) = \int \text{Ber}(\pi_k)\,\text{Beta}\left(\frac{\alpha}{K}, 1\right) = \frac{\alpha}{K}$
- Let $K_1 = \sum_{k=1}^{K} z_{1,k}$ then $p(K_1|\alpha) = \text{Binomial}\left(\frac{\alpha}{K}, K\right)$

Marginal on Z

$K \to \infty$

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k}|\pi_k \sim \text{Ber}(\pi_k)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- For $i = 1$, the chance of each feature k is independent $p(z_{1,k} = 1|\alpha) = \int \text{Ber}(\pi_k)\,\text{Beta}\left(\frac{\alpha}{K}, 1\right) = \frac{\alpha}{K}$
- Let $K_1 = \sum_{k=1}^{K} z_{1,k}$ then $p(K_1|\alpha) = \text{Binomial}\left(\frac{\alpha}{K}, K\right)$
- $\lim_{K\to\infty} p(K_1|\alpha) = \text{Poisson}(\alpha)$

Marginal on Z

$K \to \infty$

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k} | \pi_k \sim \text{Ber}(\pi_k)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- For $i = 1$, the chance of each feature k is independent $p(z_{1,k} = 1|\alpha) = \int \text{Ber}(\pi_k)\, \text{Beta}\left(\frac{\alpha}{K}, 1\right) = \frac{\alpha}{K}$
- Let $K_1 = \sum_{k=1}^{K} z_{1,k}$ then $p(K_1|\alpha) = \text{Binomial}\left(\frac{\alpha}{K}, K\right)$
- $\lim_{K\to\infty} p(K_1|\alpha) = \text{Poisson}(\alpha)$

Subsequent, i

- Let $n_{<i,k} = \sum_{j=1}^{i-1} z_{j,k}$

Marginal on Z

$K \to \infty$

$$\pi_k|\alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k}|\pi_k \sim \text{Ber}\left(\pi_k\right)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- For $i = 1$, the chance of each feature k is independent $p(z_{1,k} = 1|\alpha) = \int \text{Ber}\left(\pi_k\right) \text{Beta}\left(\frac{\alpha}{K}, 1\right) = \frac{\alpha}{K}$
- Let $K_1 = \sum_{k=1}^{K} z_{1,k}$ then $p(K_1|\alpha) = \text{Binomial}\left(\frac{\alpha}{K}, K\right)$
- $\lim_{K\to\infty} p(K_1|\alpha) = \text{Poisson}\left(\alpha\right)$

Subsequent, i

- Let $n_{<i,k} = \sum_{j=1}^{i-1} z_{j,k}$
- for a previously used k, $p(z_{i,k} = 1) = \frac{\frac{\alpha}{K} + n_{<i,k}}{\frac{\alpha}{K} + 1 - i - 1} \to \frac{n_{<i,k}}{i}$

Marginal on Z

$K \to \infty$

$$\pi_k | \alpha \sim \text{Beta}\left(\frac{\alpha}{K}, 1\right) \qquad z_{i,k} | \pi_k \sim \text{Ber}(\pi_k)$$

$$P([Z]) = \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(n_k + \frac{\alpha}{K})\Gamma(N - n_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

- For $i = 1$, the chance of each feature k is independent $p(z_{1,k} = 1|\alpha) = \int \text{Ber}(\pi_k)\, \text{Beta}\left(\frac{\alpha}{K}, 1\right) = \frac{\alpha}{K}$
- Let $K_1 = \sum_{k=1}^{K} z_{1,k}$ then $p(K_1|\alpha) = \text{Binomial}\left(\frac{\alpha}{K}, K\right)$
- $\lim_{K\to\infty} p(K_1|\alpha) = \text{Poisson}(\alpha)$

Subsequent, i

- Let $n_{<i,k} = \sum_{j=1}^{i-1} z_{j,k}$
- for a previously used k, $p(z_{i,k} = 1) = \frac{\frac{\alpha}{K} + n_{<i,k}}{\frac{\alpha}{K} + 1 - i - 1} \to \frac{n_{<i,k}}{i}$
- Also, Poisson $\left(\frac{\alpha}{i}\right)$ new features

Indian Buffet Process

sampling scheme for marginal of $z_{i,k}|\alpha$

First Customer: Sample Poisson $\left(\frac{\alpha}{i}\right)$ dishes

Indian Buffet Process

sampling scheme for marginal of $z_{i,k}|\alpha$

First Customer: Sample Poisson $\left(\frac{\alpha}{i}\right)$ dishes
Each subsequent customer, i:

- Sample previously samples dishes by popularity $p(z_{i,k} = \frac{n_{<i,k}}{i}$

Indian Buffet Process

sampling scheme for marginal of $z_{i,k}|\alpha$

First Customer: Sample Poisson $\left(\frac{\alpha}{i}\right)$ dishes
Each subsequent customer, i:

- Sample previously samples dishes by popularity $p(z_{i,k} = \frac{n_{<i,k}}{i}$
- Sample Poisson $\left(\frac{\alpha}{i}\right)$ new dishes poisson distribution

Indian Buffet Process

sampling scheme for marginal of $z_{i,k}|\alpha$

First Customer: Sample Poisson $\left(\frac{\alpha}{i}\right)$ dishes
Each subsequent customer, i:

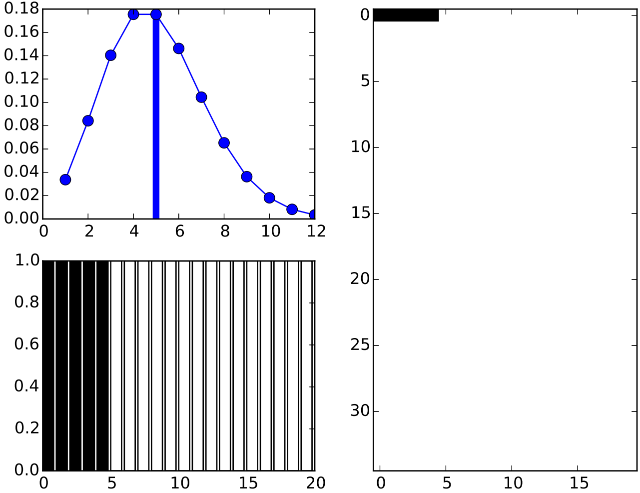
- Sample previously samples dishes by popularity $p(z_{i,k} = \frac{n_{<i,k}}{i}$
- Sample Poisson $\left(\frac{\alpha}{i}\right)$ new dishes poisson distribution

Properties:

- Effective dimension, $K_+ \sim$ Poisson $\left(\alpha \sum_{i=1}^{N} \frac{1}{i}\right)$
- Number of dishes sampled by each customer is Poisson (α) by exchangeability

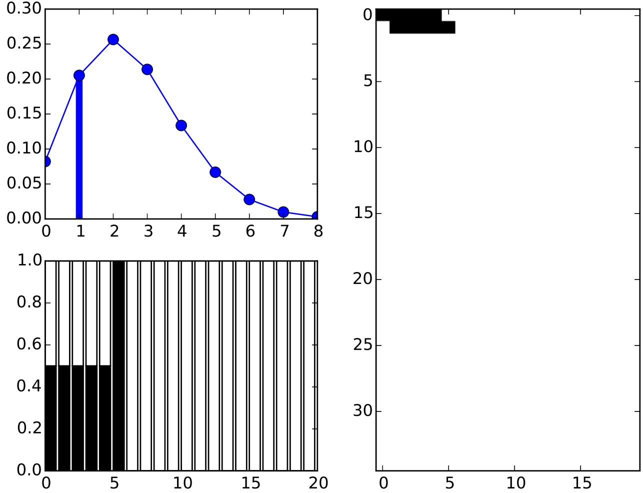
IBP Sampling

$\alpha = 5$



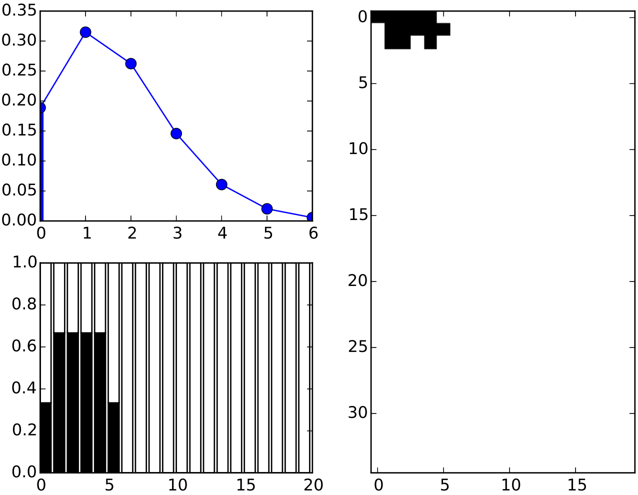
IBP Sampling

$\alpha = 5$



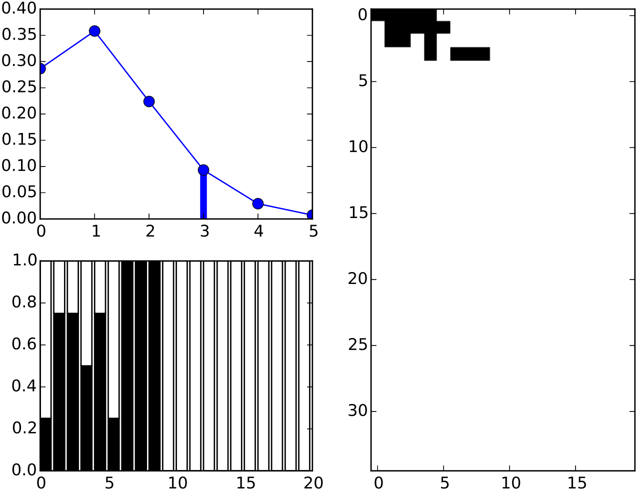
IBP Sampling

$\alpha = 5$



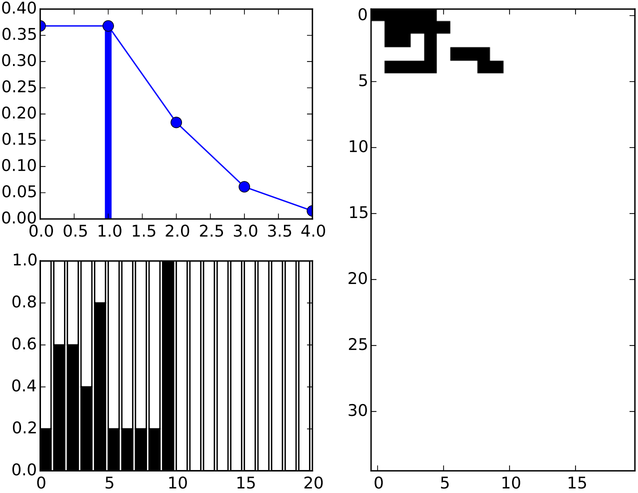
IBP Sampling

$\alpha = 5$



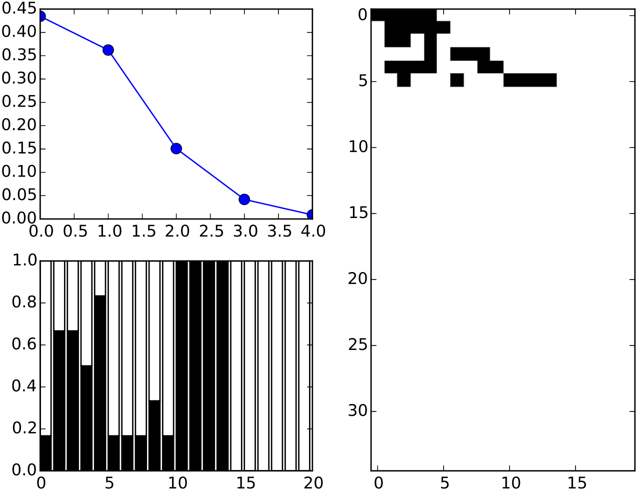
IBP Sampling

$\alpha = 5$



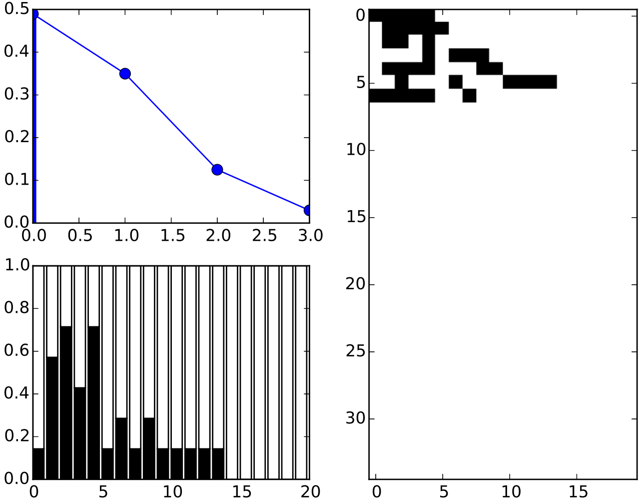
IBP Sampling

$\alpha = 5$



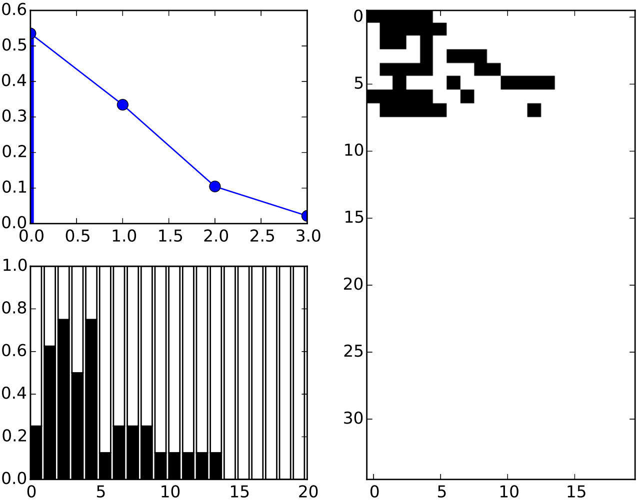
IBP Sampling

$\alpha = 5$



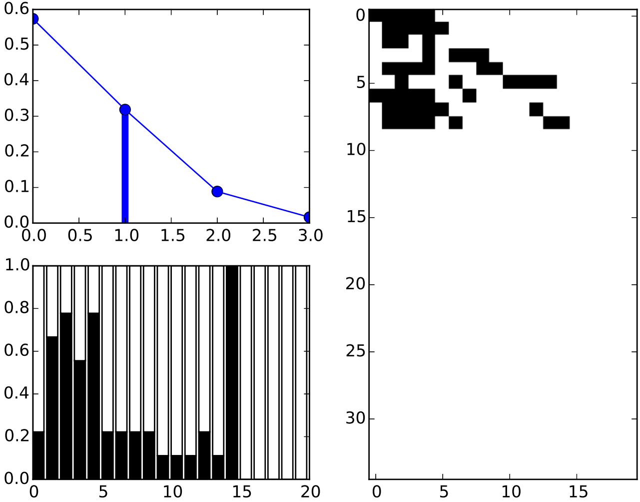
IBP Sampling

$\alpha = 5$



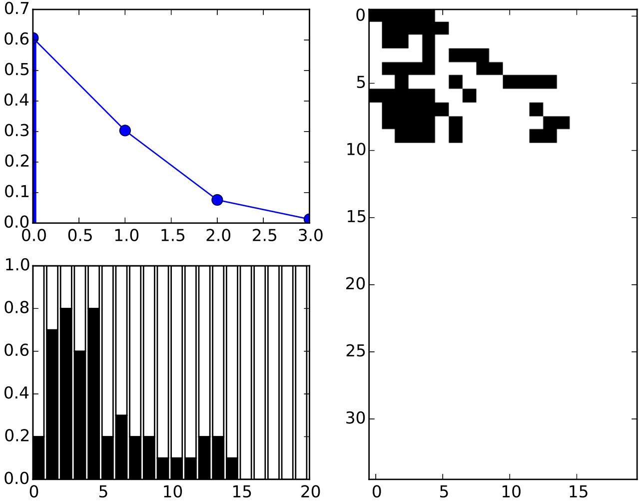
IBP Sampling

$\alpha = 5$



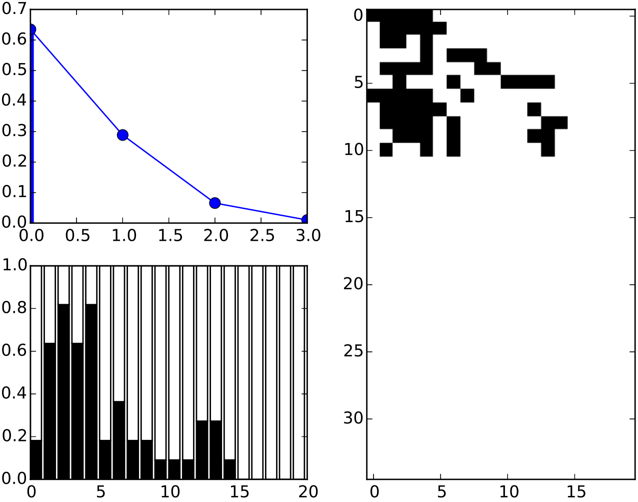
IBP Sampling

$\alpha = 5$



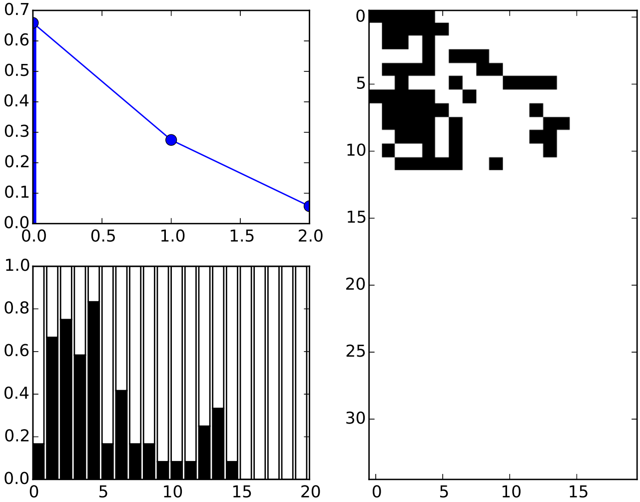
IBP Sampling

$\alpha = 5$



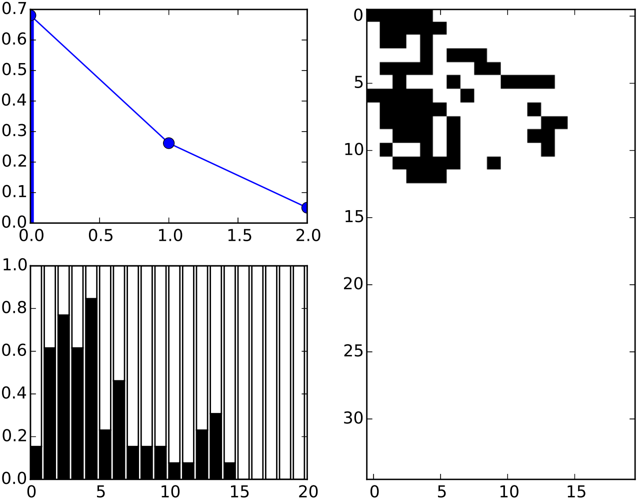
IBP Sampling

$\alpha = 5$



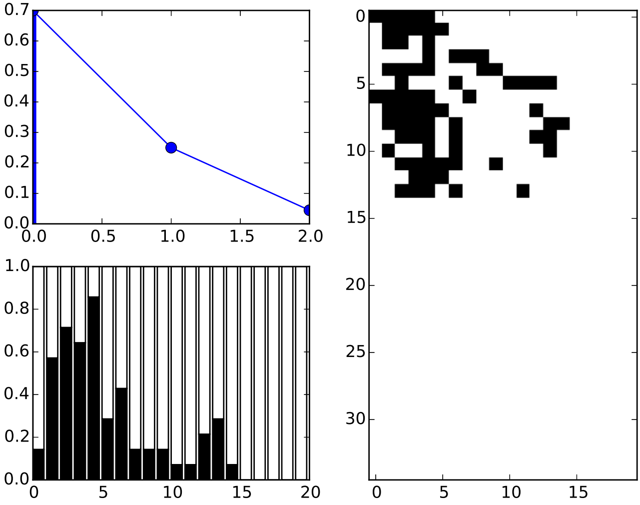
IBP Sampling

$\alpha = 5$



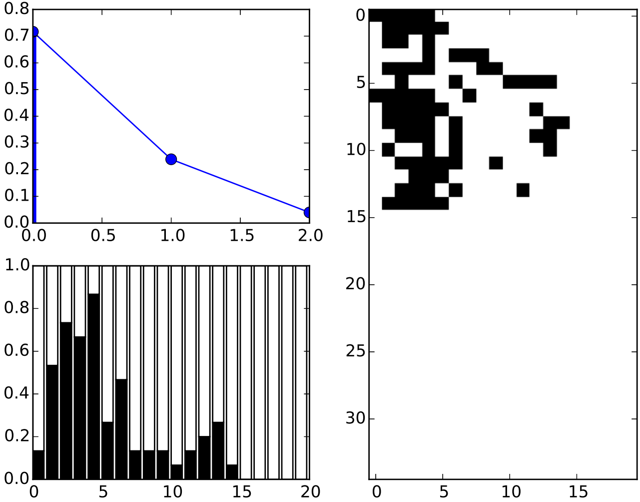
IBP Sampling

$\alpha = 5$



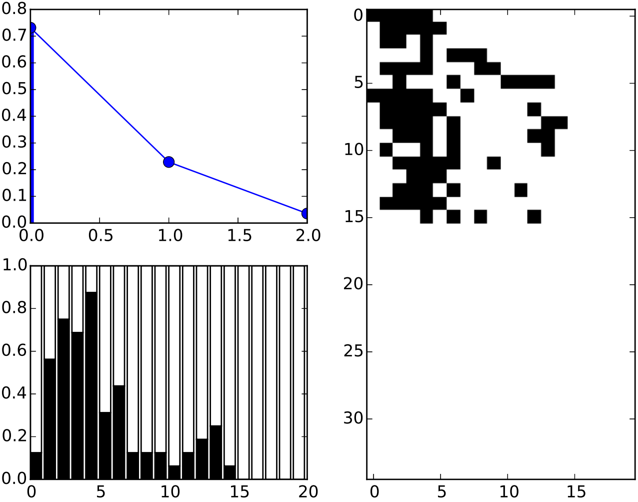
IBP Sampling

$\alpha = 5$



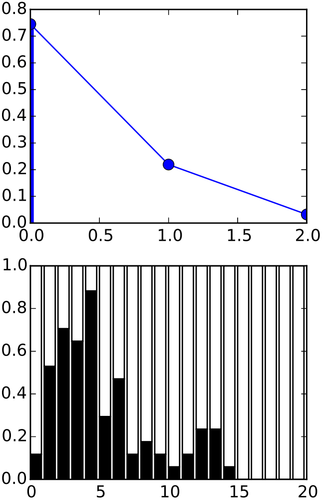
IBP Sampling

$\alpha = 5$



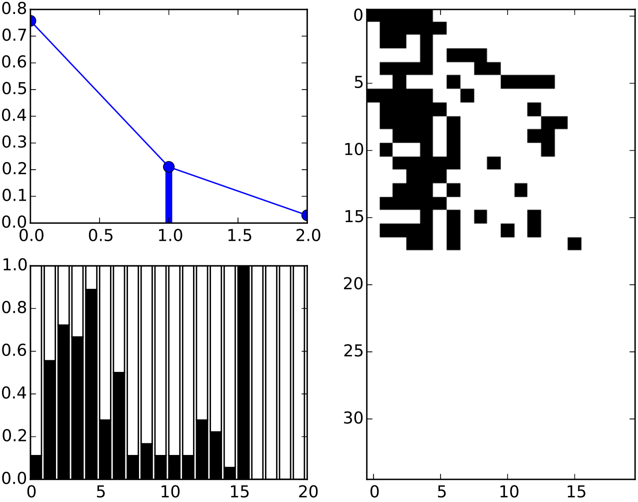
IBP Sampling

$\alpha = 5$



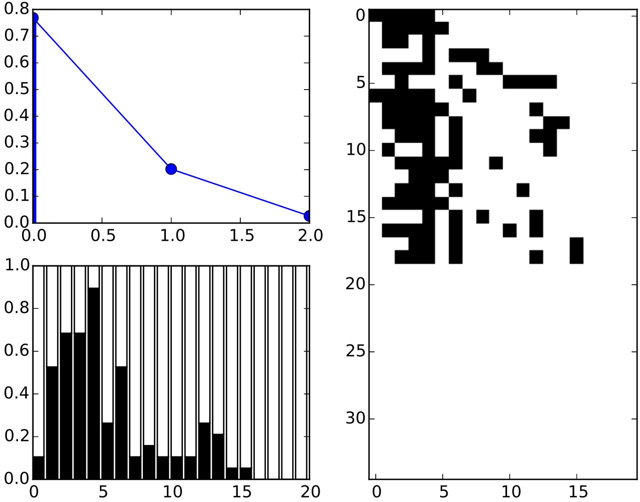
IBP Sampling

$\alpha = 5$



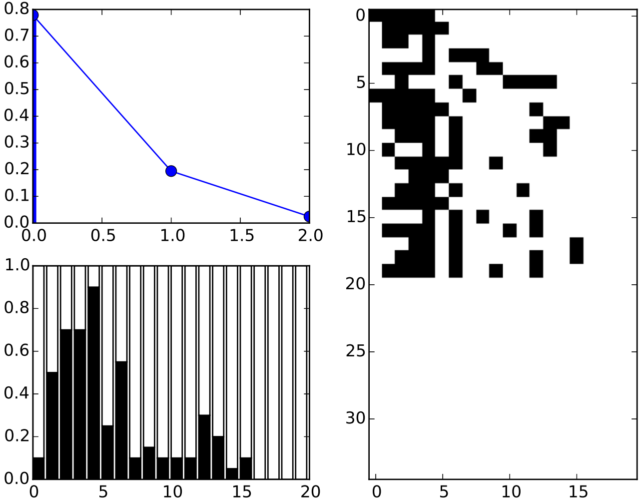
IBP Sampling

$\alpha = 5$



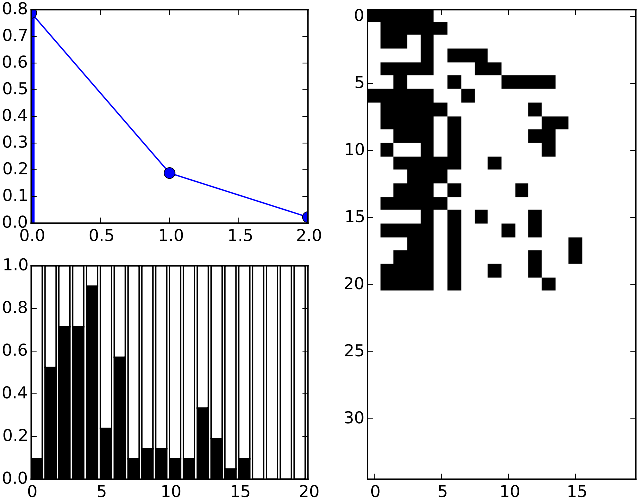
IBP Sampling

$\alpha = 5$



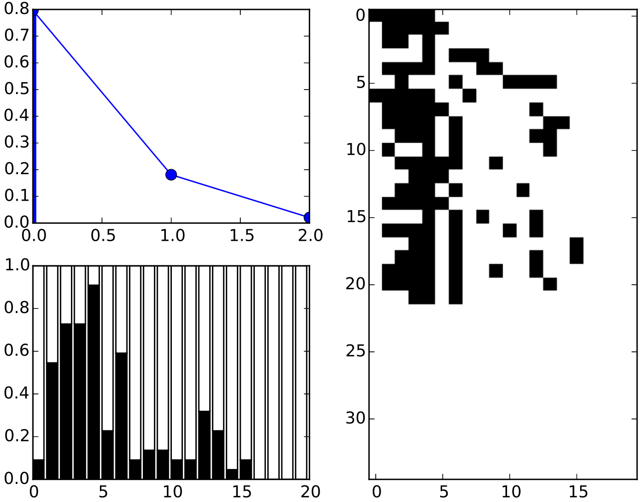
IBP Sampling

$\alpha = 5$



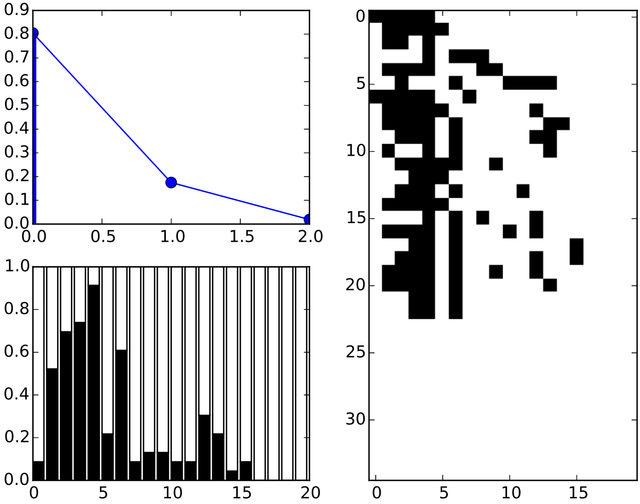
IBP Sampling

$\alpha = 5$



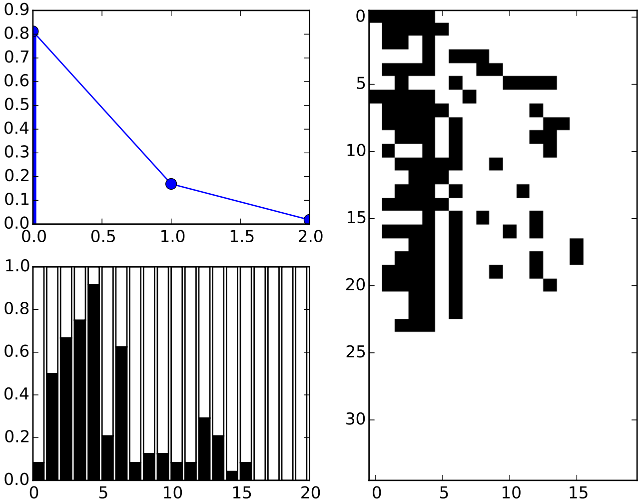
IBP Sampling

$\alpha = 5$



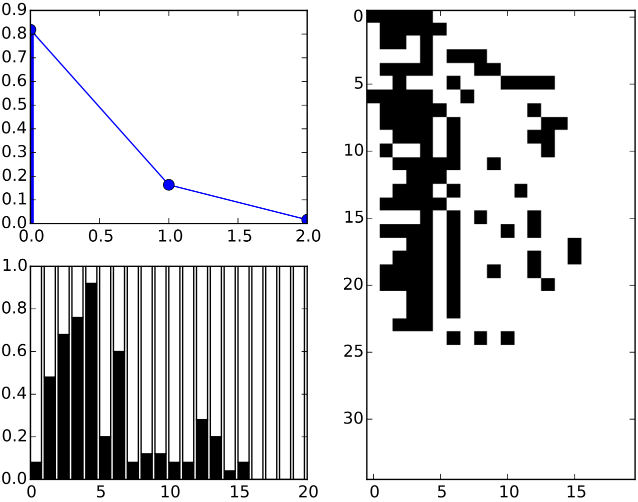
IBP Sampling

$\alpha = 5$



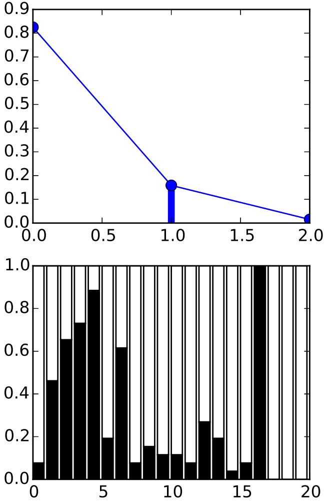
IBP Sampling

$\alpha = 5$



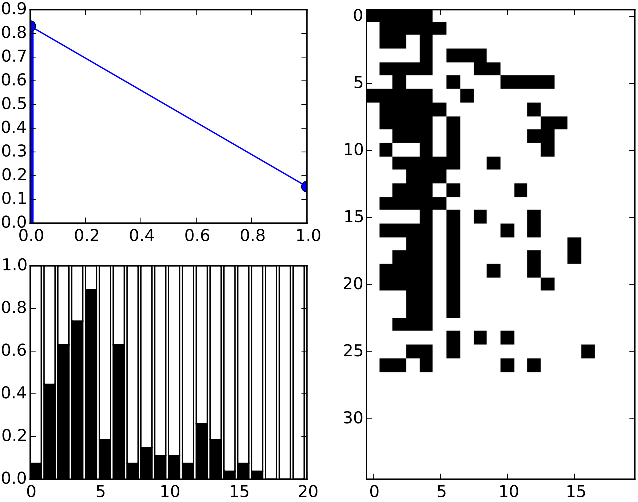
IBP Sampling

$\alpha = 5$



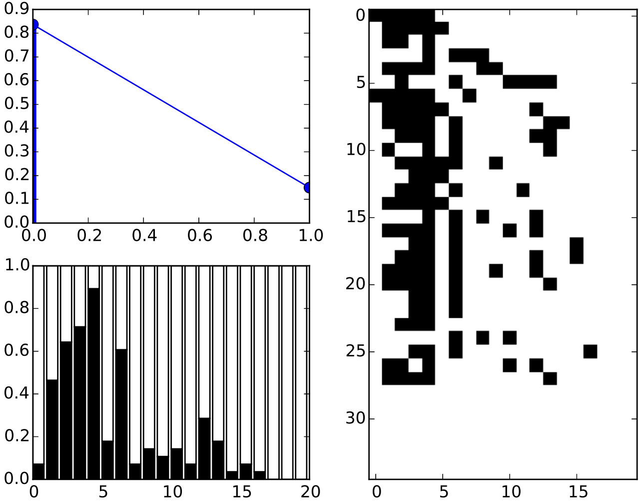
IBP Sampling

$\alpha = 5$



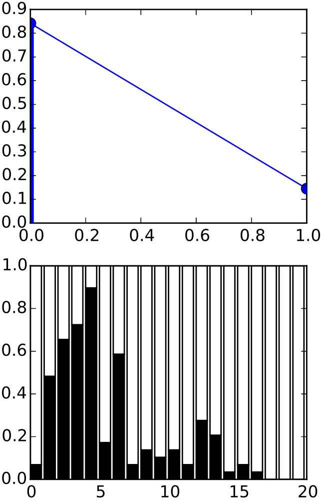
IBP Sampling

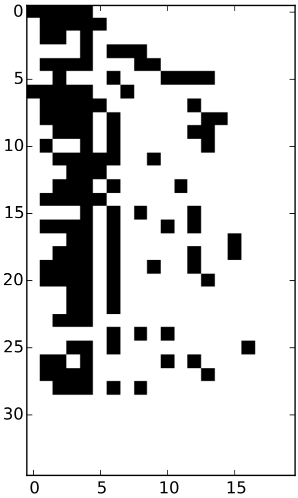
$\alpha = 5$



IBP Sampling

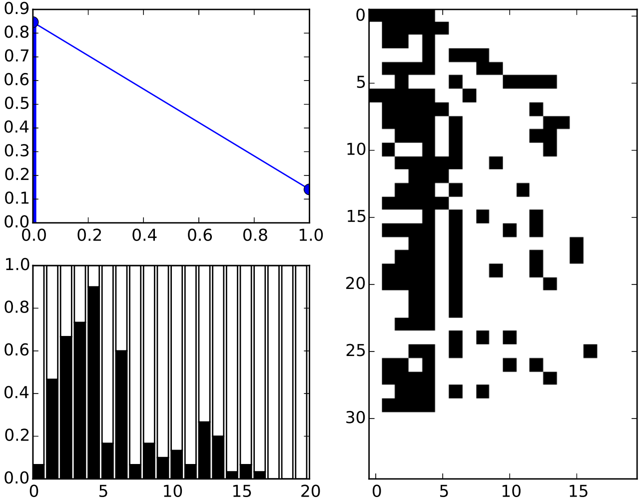
$\alpha = 5$





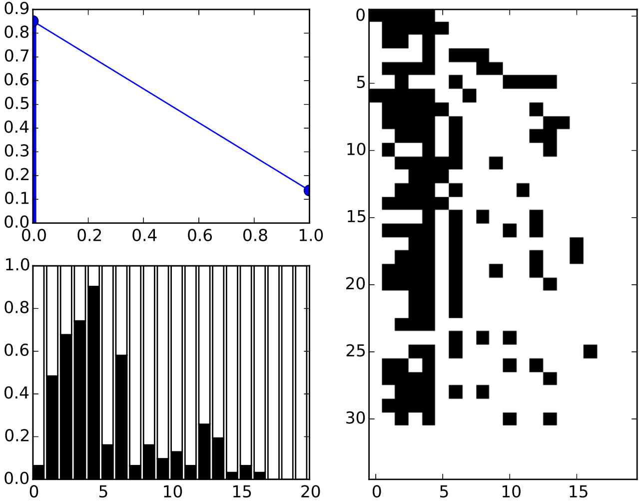
IBP Sampling

$\alpha = 5$



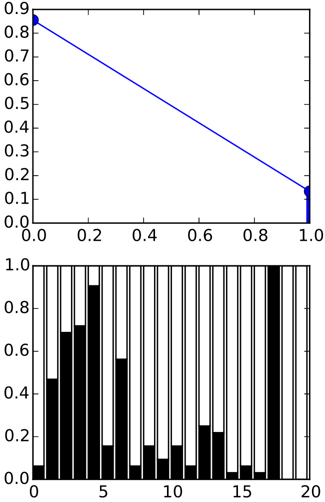
IBP Sampling

$\alpha = 5$



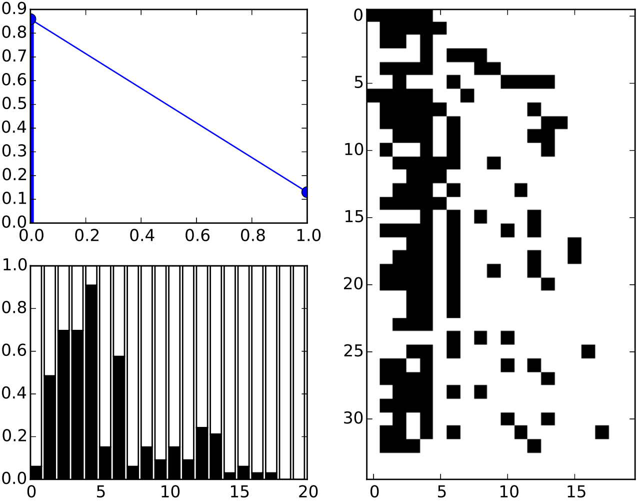
IBP Sampling

$\alpha = 5$



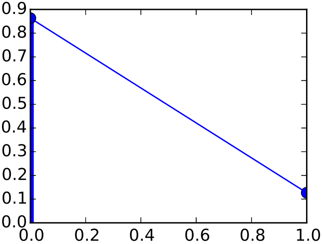
IBP Sampling

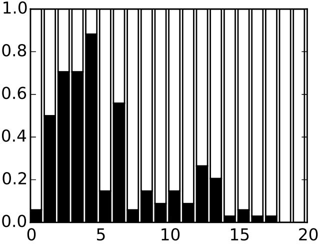
$\alpha = 5$

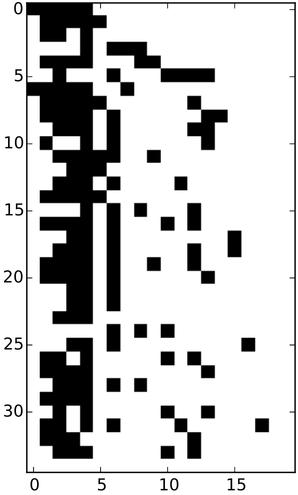


IBP Sampling

$\alpha = 5$

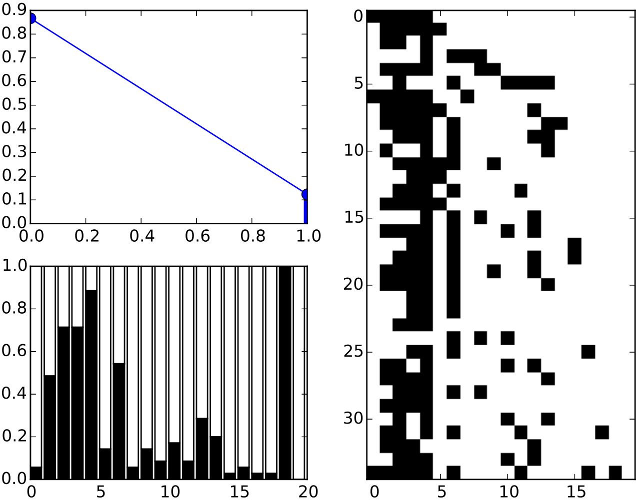






IBP Sampling

$\alpha = 5$



Gibbs Sampler

To sample, we need: $P(z_{i,k} = 1 | Z_{-i,k})$

Finite: $P(z_{i,k} = 1 | Z_{-i,k}) = \frac{n_{-i,k} + \frac{\alpha}{K}}{N + \frac{\alpha}{K}}$

Infinite: (by limit or IBP) $P(z_{i,k} = 1 | Z_{-i,k}) = \frac{n_{-i,k}}{N}$ new features: Poisson $\left(\frac{\alpha}{N}\right)$

Algorithm for $Z \sim P(Z)$:

- start with arbitrary binary matrix
- iterate through rows:
  - if $m_{-i,k} > 0$ set $z_{i,k} = 1$ by above
  - else, delete column k
  - add Poisson $\left(\frac{\alpha}{N}\right)$ new features

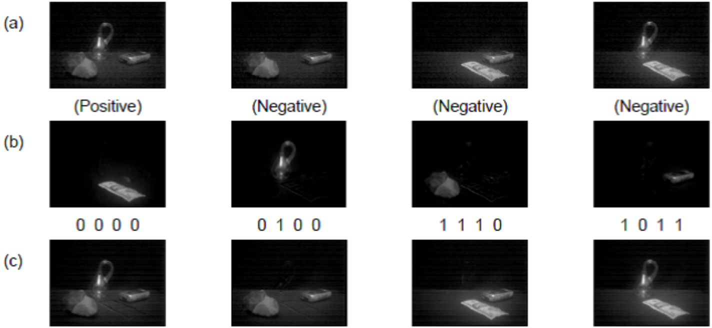
This converges to a matrix drawn from $P(Z)$

Sampling the Posterior

The real target is $P(Z|X)$
Full conditional: $P(z_{i,k} = 1|Z_{-i,k}, X) \propto P(X|Z)P(z_{i,k} = 1|Z_{-i,k})$
Algorithm:

- start with arbitrary binary matrix
- iterate through rows:
  - if $m_{-i,k} > 0$ set $z_{i,k} = 1$ *incorporating the likelihood*
  - else, delete column k
  - add new columns with prior Poisson $\left(\frac{\alpha}{N}\right)$ and $P(X|Z)$ likeilihood

Example Application



(a) 4 sample images from 100 (b) posterior mean of the weights of the four most frequent features, with signs (c) reconstructions of images in (a) from model with codes

Summary

- Latent feature allocation allows each sample to belong to multiple groups
- Beta prior on bernouli draws, to construct a binary matrix
- Indian Buffet Process is a generative process for the matrix marginal
- IBP yields a Gibbs Sampler
- (note) There is a stick breaking scheme... it yields variational inference

Conclusion

Bayesian nonparametrics allow distributions without *fixed* parameters

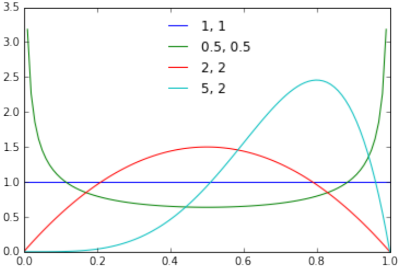
Food Metaphors explain the marginals of the categorical (CRP) or Bernouli (IBP) distributions

Food Metaphors yield Gibbs Samplers

Stick breaking metaphors yield variational inference

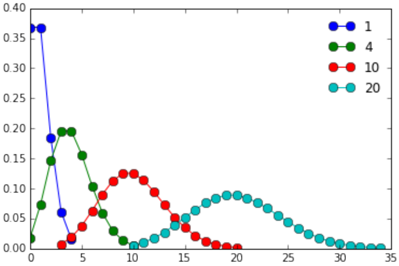
Beta Distribution

$$\text{Beta}\left(\rho|\alpha\right) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\rho^{\alpha_1 - 1}(1-\rho)^{\alpha_2 - 1}$$



Poisson Distribution

$$\text{Poisson}\,(k|\lambda) = \frac{\lambda^k}{k!} \exp -\lambda$$



Binomial

$$p(\sum_{k=1}^{K} z_{1,k} = k) = \binom{K}{k} \frac{\alpha}{K}^{k} (1 - \frac{\alpha}{K})^{K-k}$$

marginal