Applied Bayesian Nonparametrics
Special Topics in Machine Learning
Brown University CSCI 2950-P, Fall 2011

Variational Inference for the Indian Buffet Process

Finale Doshi-Velez* — Engineering Department, Cambridge University, Cambridge, UK

Kurt T. Miller* — Computer Science Division, University of California, Berkeley, Berkeley, CA

Jurgen Van Gael* — Engineering Department, Cambridge University, Cambridge, UK

Yee Whye Teh — Gatsby Unit, University College London, London, UK

Soravit Beer Changpinyo

November 15, 2011

Main Contributions

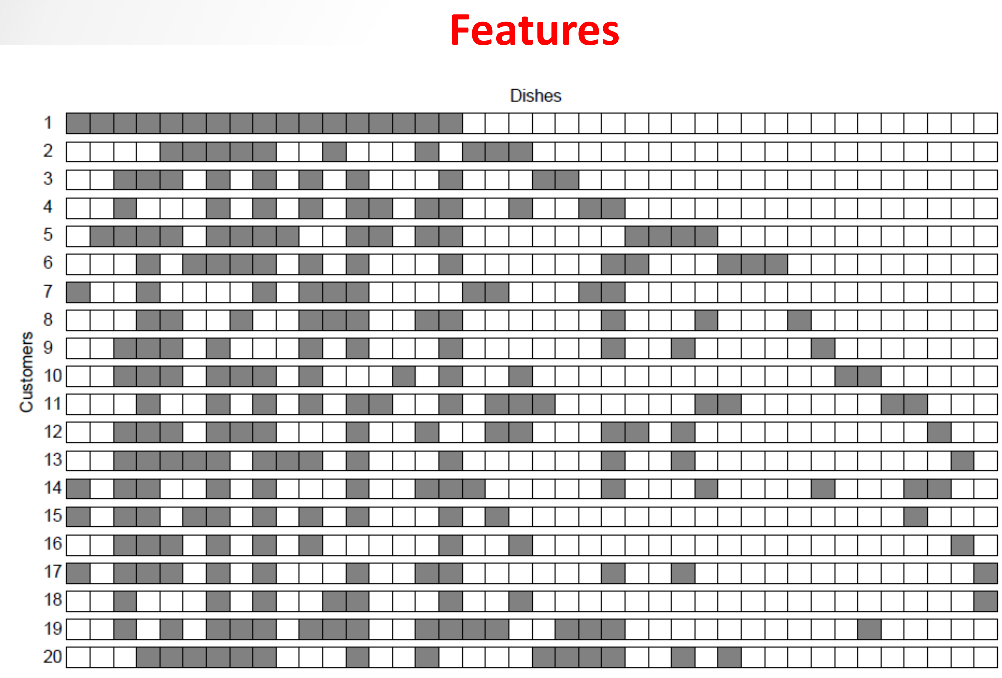
- **Deterministic alternative to samplers for inference in the Indian Buffet Process (IBP)**
- **Theoretical bounds on the truncation error**

Overview

- Backgrounds
  - The IBP and its stick-breaking construction
  - Latent feature model
  - Variational methods
- Variational Inference for the IBP
  - Finite variational approach
  - Infinite variational approach
- Truncation error
- Experiments

Indian Buffet Process

Griffiths and Ghahramani (2005)



A binary matrix generated by the Indian buffet process with $\alpha = 10$

The 1st customer tries Poisson(α) number of dishes

The ith customer

- takes dishes previously sampled with probability m_k/i
- tries Poisson(α/i) number of new dishes

m_k = the number of customers who previously tried dish k

Infinite Exchangeability :

the ordering of customers does not impact distribution

Indian Buffet Process

Griffiths and Ghahramani (2005)

Distribution on equivalence classes of binary matrices

$$p([Z]) = \frac{\alpha^K}{\prod_{h \in \{0,1\}^N \backslash \mathbf{0}} K_h!} \exp\{-\alpha H_N\} \prod_{k=1}^{K} \frac{(N - m_k)!(m_k - 1)!}{N!}$$

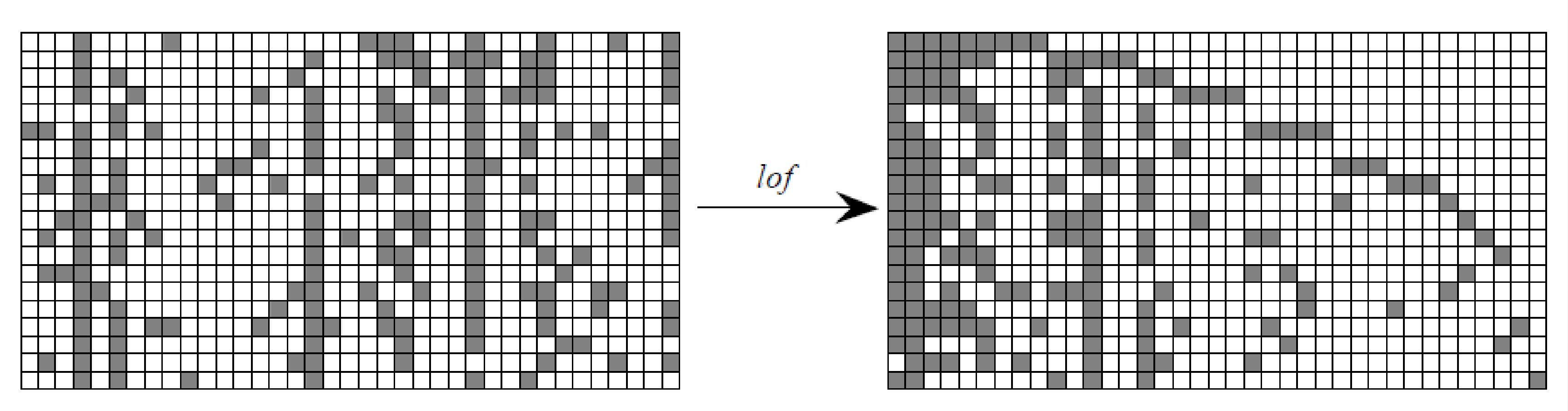
N = the number of rows
K = the number of nonzero columns
K_h = the number of history h among columns
m_k = the number of 1's in column k
H_N = the N^{th} harmonic number

α controls
the number of features per object AND the total number of features

Indian Buffet Process

Griffiths and Ghahramani (2005)

Define equivalence class by the left-ordered form



Indian Buffet Process: Stick-breaking construction

Teh et al. (2007)

$$v_1, v_2, \cdots \overset{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, 1)$$

$$\pi_1 = v_1$$

$$\pi_k = v_k \pi_{k-1} = \prod_{i=1}^{k} v_i$$

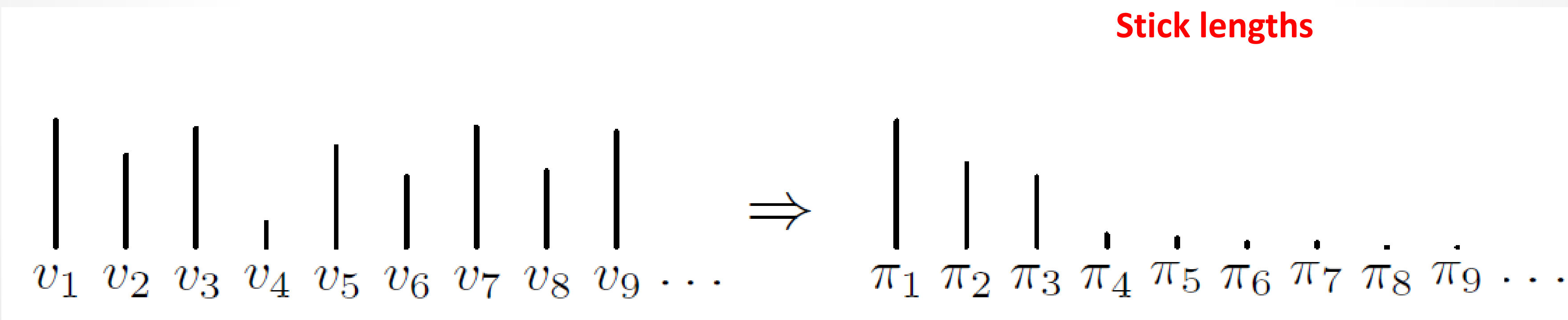


Figure from slides of Doshi-Velez et al.

Indian Buffet Process: Stick-breaking construction

Teh et al. (2007)

$$v_1, v_2, \cdots \overset{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, 1)$$

$$\pi_1 = v_1$$

$$\pi_k = v_k \pi_{k-1} = \prod_{i=1}^{k} v_i$$

$$z_{nk} \sim \text{Bernoulli}(\pi_k)$$

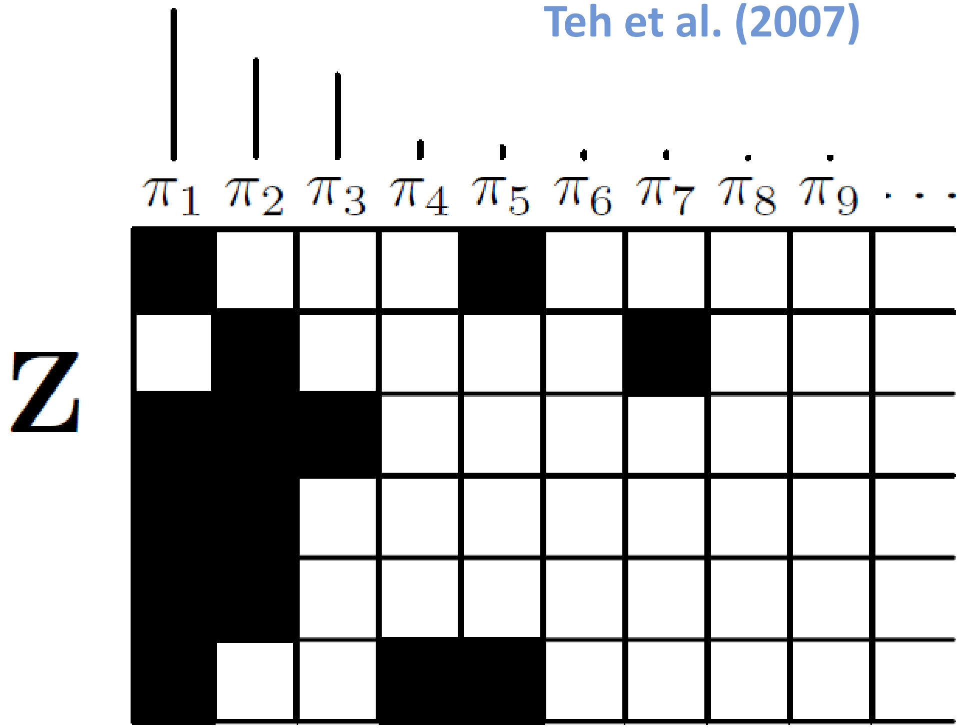


Figure from slides of Doshi-Velez et al.

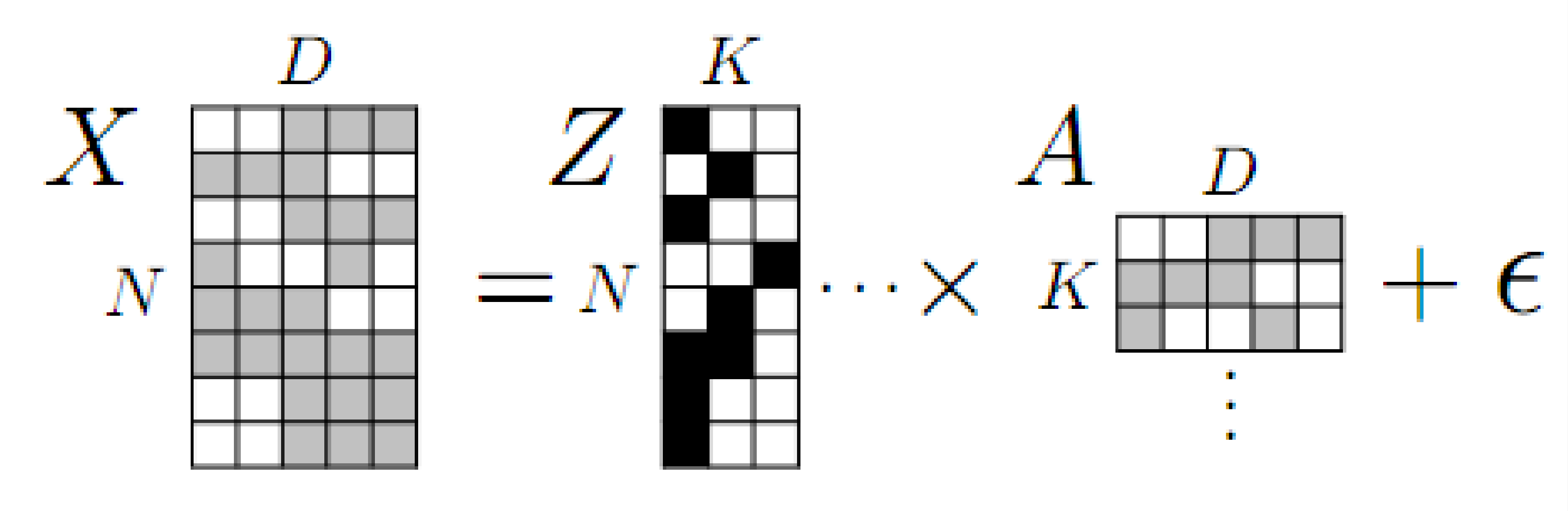
- **This representation will show up in the infinite variational approach**

Latent Feature Model

One example

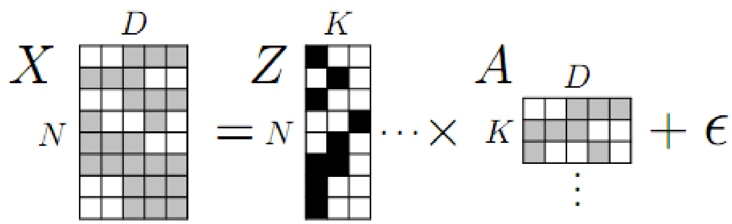
Each of the N objects has

D-dimensional observations **K-dimensional latent features**



Observed **Latent**

Latent Feature Model



Observed **Latent**

Assume Z and A are independent

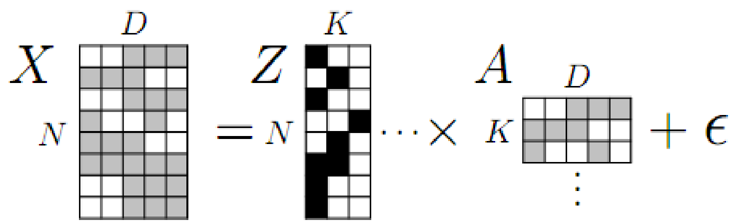
The posterior distribution of Z and A:

?

$$p(Z, A|X) \propto p(X|Z, A)p(Z)p(A)$$

Depend on applications

Latent Feature Model



Observed **Latent**

Assume Z and A are independent

The posterior distribution of Z and A:

$$p(Z, A|X) \propto p(X|Z, A)p(Z)p(A)$$

IBP

Depend on applications

Latent Feature Model:

Linear-Gaussian likelihood model

$$p(Z, A|X) \propto p(X|Z, A)p(Z)p(A)$$

$$p(X|Z, A) \sim \mathcal{N}(ZA, \sigma_n^2 I)$$
$$p(A) \sim \mathcal{N}(0, \sigma_A^2 I)$$
$$p(Z) \sim \text{IBP}(\alpha)$$

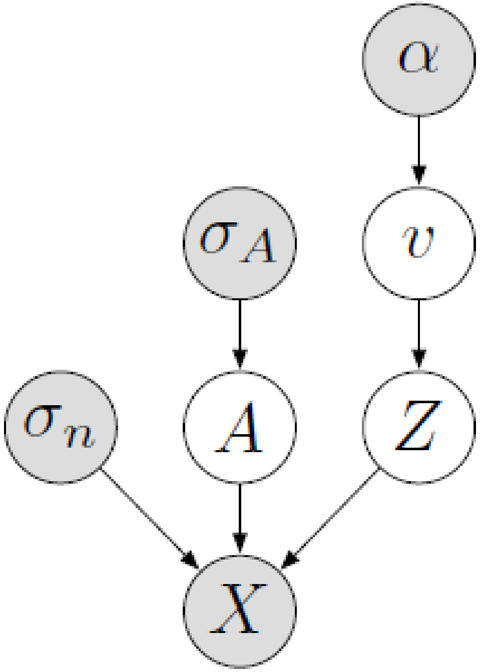
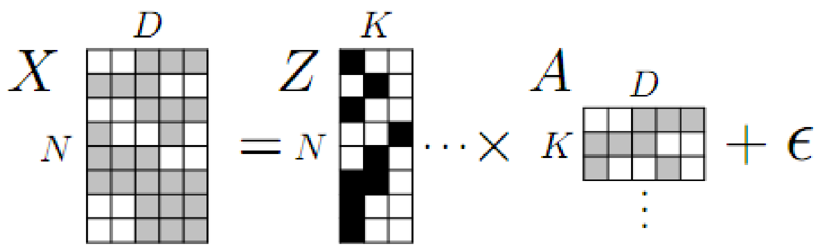


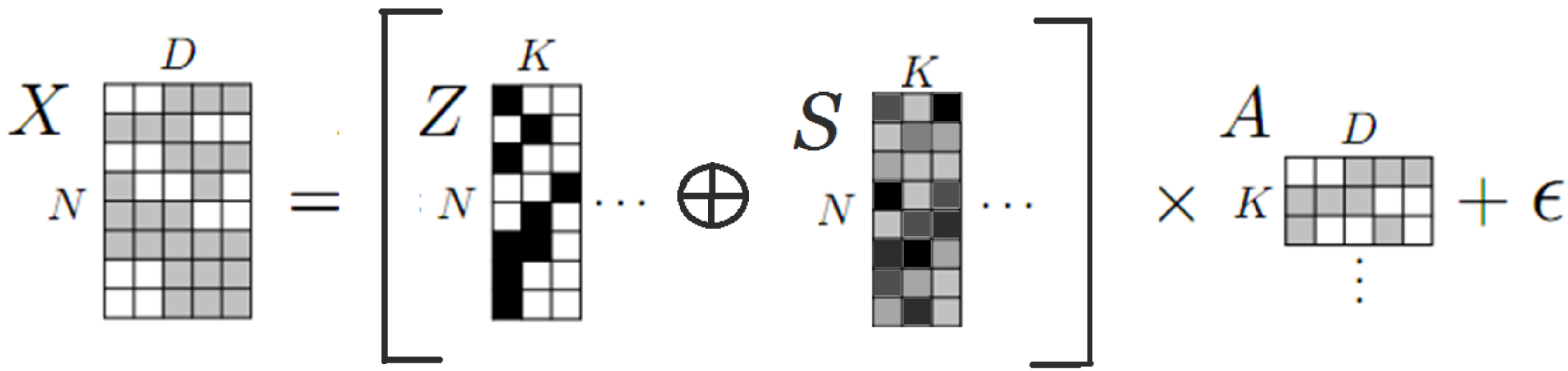
Figure from slides of Doshi-Velez et al.

Latent Feature Model:

Infinite Independent Component Analysis

Knowles and Ghahramani (2007)

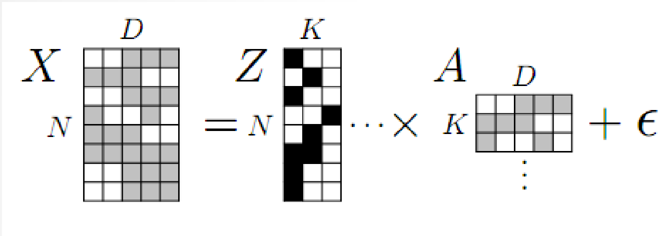


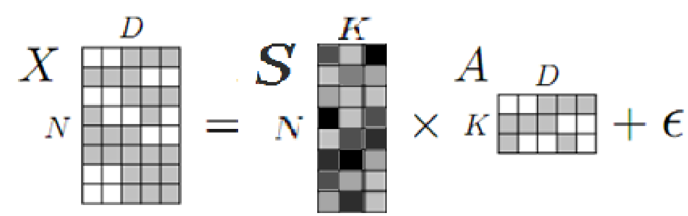


Latent Feature Model:

Infinite Independent Component Analysis

Knowles and Ghahramani (2007)



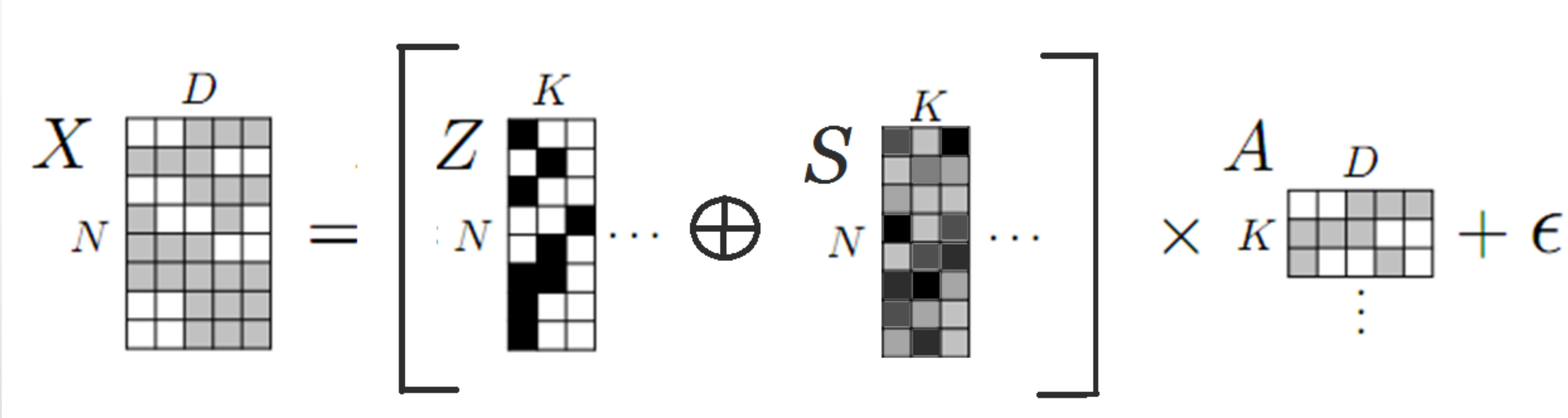


Independent Component Analysis (ICA)

Add signal S

Make sparse (Add Z)

Make infinite



Latent Feature Model:

Linear-Gaussian likelihood model

- This will be our main focus
- Will discuss the infinite ICA briefly

Goal: compute Z and A given X

$$p(Z, A|X) \propto p(X|Z, A)p(Z)p(A)$$

$$p(X|Z, A) \sim \mathcal{N}(ZA, \sigma_n^2 I)$$
$$p(A) \sim \mathcal{N}(0, \sigma_A^2 I)$$
$$p(Z) \sim \text{IBP}(\alpha)$$

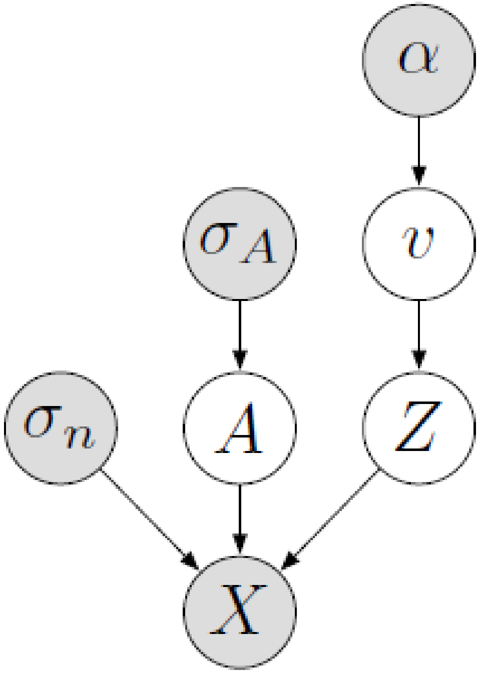


Figure from slides of Doshi-Velez et al.

Latent Feature Model:

Why intractable?

Goal: compute Z and A given X

$$p(Z, A|X) \propto p(X|Z, A)p(Z)p(A)$$

Problem

Even for finite K, we have 2^{NK} assignments for Z

$$p(X|Z, A) \sim \mathcal{N}(ZA, \sigma_n^2 I)$$
$$p(A) \sim \mathcal{N}(0, \sigma_A^2 I)$$
$$p(Z) \sim \mathrm{IBP}(\alpha)$$

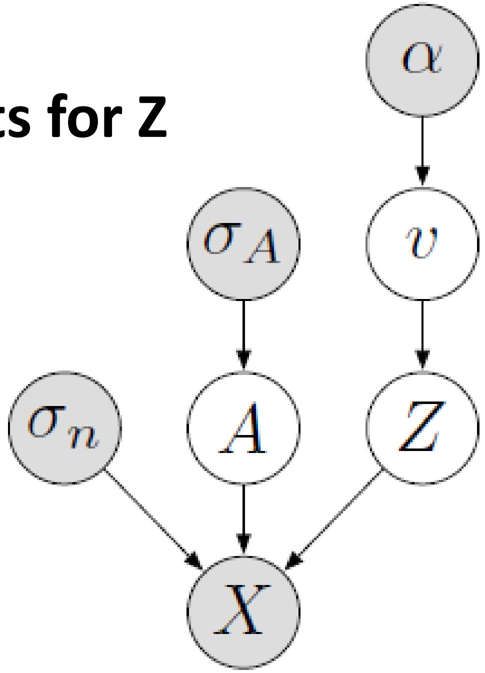


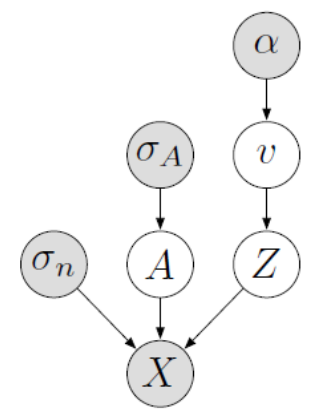
Figure from slides of Doshi-Velez et al.

Variational Methods: Big Picture

- **Approximate the true posterior with a *variational distribution* q from <u>tractable</u> family of distributions**
- **The variational distribution indexed by a set of *variational parameters***
- **Adjust the parameters to get the *tightest* lower bound possible**

Variational Inference:

KL-divergence



W = { π, Z, A }

$\theta = \{ \alpha, \sigma^2_A, \sigma^2_N \}$

ϑ = a vector of variational parameters

$$\log p(\mathrm{x}|\theta) \;\geq\; \mathbb{E}_q\left[\log p(\mathbf{W}, \mathrm{x}|\theta)\right] - \mathbb{E}_q\left[\log q(\mathbf{W})\right] \;=\; \mathcal{L}(q)$$

$$KL\left[q(\mathrm{w}|\vartheta) \,\|\, p(\mathrm{w}|\mathrm{x}, \theta)\right] \;=\; \log p(\mathrm{x}|\theta) - \mathcal{L}(q)$$

$$\arg\max_q \mathcal{L}(q) \Leftrightarrow \arg\min_q KL\left[q(\mathrm{w}|\vartheta) \,\|\, p(\mathrm{w}|\mathrm{x}, \theta)\right]$$

Variational Inference:
Choosing Variational Distribution

How to choose the variational distribution $q_\vartheta(w)$ such that the optimization of the bound is computationally tractable?

Typically, we break some dependencies between latent variables

Mean field variational approximations

Assume "fully factorized" variational distributions

$$q_\vartheta(w) = \prod_{m=1}^{M} q_{\vartheta_m}(w_m)$$

where $\vartheta = (\vartheta_1, \vartheta_2, \ldots, \vartheta_M)$

Mean Field Variational Inference:

Finite and Infinite Variational Approaches

$$q(\boldsymbol{W}) = q_{\boldsymbol{\tau}}(\boldsymbol{\pi}) q_{\phi}(\boldsymbol{A}) q_{\boldsymbol{\nu}}(\boldsymbol{Z})$$

Two approaches :

- **Finite variational approach**
- **Infinite variational approach**

Mean Field Variational Inference:

Finite and Infinite Variational Approaches

$$\arg\min_q KL\left[q(\mathrm{w}|\vartheta) \parallel p(\mathrm{w}|\mathrm{x}, \theta)\right]$$

	True posterior p	Variational distribution q $q(\boldsymbol{W}) = q_{\boldsymbol{\tau}}(\boldsymbol{\pi})q_{\boldsymbol{\phi}}(\boldsymbol{A})q_{\boldsymbol{\nu}}(\boldsymbol{Z})$
Finite	• Finite approximation p_K • beta-Bernoulli model Griffiths and Ghahramani (2005)	• Changes according to p • Straightforward computation of lower bound and parameter updates
Infinite	• Stay the same • Use stick-breaking representation	• Truncated stick-breaking construction • Deal with the distribution of v instead of π Blei and Jordan (2004)

Mean Field Variational Inference:

Finite Variational Approach

- **Beta-Bernoulli model – K finite (but large) truncation level**

$$
\begin{aligned}
\pi_k &\sim \text{Beta}(\alpha/K, 1) & \text{for } k \in \{1 \ldots K\}, \\
z_{nk} &\sim \text{Bernoulli}(\pi_k) & \text{for } k \in \{1 \ldots K\}, n \in \{1 \ldots N\}, \\
\boldsymbol{A}_{k\cdot} &\sim \text{Normal}(0, \sigma_A^2 I) & \text{for } k \in \{1 \ldots K\}, \\
\boldsymbol{X}_{n\cdot} &\sim \text{Normal}(\boldsymbol{Z}_{n\cdot}\boldsymbol{A}, \sigma_n^2 I) & \text{for } n \in \{1 \ldots N\},
\end{aligned}
$$

- **This is the model before we take K to ∞**

$$
p_K(\boldsymbol{W}, \boldsymbol{X}|\boldsymbol{\theta}) = \prod_{k=1}^{K}\left(p(\pi_k|\alpha)p(\boldsymbol{A}_{k\cdot}|\sigma_A^2 I)\prod_{n=1}^{N} p(z_{nk}|\pi_k)\right)\prod_{n=1}^{N} p(\boldsymbol{X}_{n\cdot}|\boldsymbol{Z}_{n\cdot}, \boldsymbol{A}, \sigma_n^2 I).
$$

- **Finite but still intractable**

Mean Field Variational Inference:

Finite Variational Approach

- **Beta-Bernoulli model – K finite (but large) truncation level**

$$\begin{aligned}
\pi_k &\sim \text{Beta}(\alpha/K, 1) && \text{for } k \in \{1 \dots K\}, \\
z_{nk} &\sim \text{Bernoulli}(\pi_k) && \text{for } k \in \{1 \dots K\}, n \in \{1 \dots N\}, \\
\boldsymbol{A}_{k\cdot} &\sim \text{Normal}(0, \sigma_A^2 I) && \text{for } k \in \{1 \dots K\}, \\
\boldsymbol{X}_{n\cdot} &\sim \text{Normal}(\boldsymbol{Z}_{n\cdot}\boldsymbol{A}, \sigma_n^2 I) && \text{for } n \in \{1 \dots N\},
\end{aligned}$$

- **Variational distribution** $q(\boldsymbol{W}) = q_{\boldsymbol{\tau}}(\boldsymbol{\pi}) q_{\boldsymbol{\phi}}(\boldsymbol{A}) q_{\boldsymbol{\nu}}(\boldsymbol{Z})$

$$\begin{aligned}
q_{\boldsymbol{\tau}_k}(\pi_k) &= \text{Beta}(\pi_k; \tau_{k1}, \tau_{k2}), \\
q_{\boldsymbol{\phi}_k}(\boldsymbol{A}_{k\cdot}) &= \text{Normal}(\boldsymbol{A}_{k\cdot}; \bar{\phi}_k, \boldsymbol{\Phi}_k), \\
q_{\nu_{nk}}(z_{nk}) &= \text{Bernoulli}(z_{nk}; \nu_{nk}).
\end{aligned}$$

Mean Field Variational Inference:

Finite Variational Approach

- **Variational distribution** $q(\boldsymbol{W}) = q_{\boldsymbol{\tau}}(\boldsymbol{\pi})q_{\boldsymbol{\phi}}(\boldsymbol{A})q_{\boldsymbol{\nu}}(\boldsymbol{Z})$

$$q_{\boldsymbol{\tau}_k}(\pi_k) = \text{Beta}(\pi_k; \tau_{k1}, \tau_{k2}),$$

$$q_{\boldsymbol{\phi}_k}(\boldsymbol{A}_{k\cdot}) = \text{Normal}(\boldsymbol{A}_{k\cdot}; \bar{\phi}_k, \boldsymbol{\Phi}_k),$$

$$q_{\nu_{nk}}(z_{nk}) = \text{Bernoulli}(z_{nk}; \nu_{nk}).$$

- **When the conditional distribution** $p(w_i \mid \mathbf{w}_{-i}, \mathbf{x}, \theta)$ **and variational distribution are both in exponential families, each step in coordinate ascent has a closed form solution (Beal, 2003; Wainwright and Jordan, 2008)**

Mean Field Variational Inference:

Infinite Variational Approach

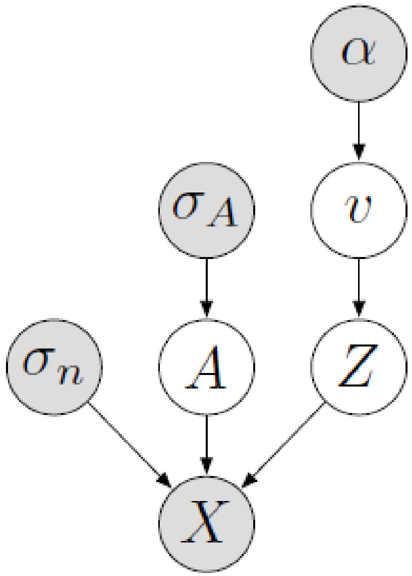
- **Consider the stick-breaking construction for the IBP**

$$
\begin{aligned}
v_k &\sim \text{Beta}(\alpha, 1) && \text{for } k \in \{1, \ldots, \infty\}, \\
\pi_k &= \prod_{i=1}^{k} v_i && \text{for } k \in \{1 \ldots \infty\}, \\
z_{nk} &\sim \text{Bernoulli}(\pi_k) && \text{for } k \in \{1 \ldots \infty\}, n \in \{1 \ldots N\}, \\
\boldsymbol{A}_{k\cdot} &\sim \text{Normal}(0, \sigma_A^2 I) && \text{for } k \in \{1 \ldots \infty\}, \\
\boldsymbol{X}_{n\cdot} &\sim \text{Normal}(\boldsymbol{Z}_{n\cdot}\boldsymbol{A}, \sigma_n^2 I) && \text{for } n \in \{1 \ldots N\}.
\end{aligned}
$$

Mean Field Variational Inference:

Infinite Variational Approach

- Consider the stick-breaking construction for the IBP
- Work on the distribution of v instead of π because $v_1, v_2, \ldots$ are independent while $\pi_1, \pi_2, \ldots$ are not



$$q(\boldsymbol{W}) = q_{\boldsymbol{\tau}}(\boldsymbol{v})q_{\boldsymbol{\phi}}(\boldsymbol{A})q_{\boldsymbol{\nu}}(\boldsymbol{Z})$$

Figure from slides of Doshi-Velez et al.

Mean Field Variational Inference:

Infinite Variational Approach

True distribution

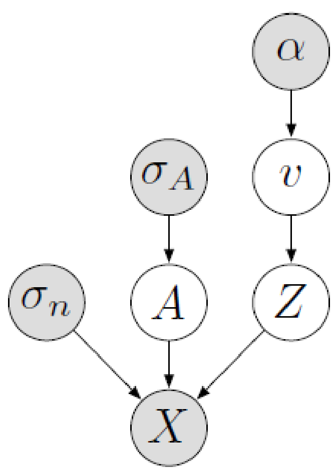
$$v_k \sim \text{Beta}(\alpha, 1) \qquad \text{for } k \in \{1, \dots, \infty\},$$

$$\pi_k = \prod_{i=1}^{k} v_i \qquad \text{for } k \in \{1 \dots \infty\},$$

$$z_{nk} \sim \text{Bernoulli}(\pi_k) \qquad \text{for } k \in \{1 \dots \infty\}, n \in \{1 \dots N\},$$

$$\boldsymbol{A}_{k\cdot} \sim \text{Normal}(0, \sigma_A^2 I) \qquad \text{for } k \in \{1 \dots \infty\},$$

$$\boldsymbol{X}_{n\cdot} \sim \text{Norn} \qquad \text{for } n \in \{1 \dots N\}.$$



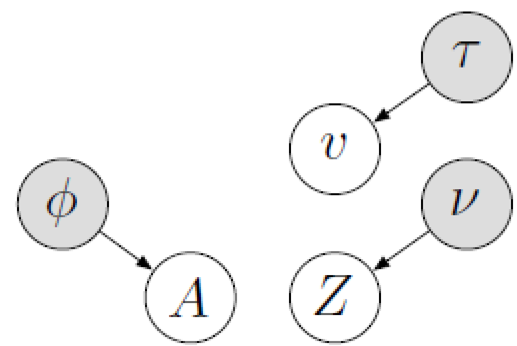
Variational distribution

$$q(\boldsymbol{W}) = q_{\boldsymbol{\tau}}(\boldsymbol{v}) q_{\boldsymbol{\phi}}(\boldsymbol{A}) q_{\boldsymbol{\nu}}(\boldsymbol{Z})$$

$$q_{\boldsymbol{\tau}_k}(v_k) = \text{Beta}(v_k; \tau_{k1}, \tau_{k2}),$$

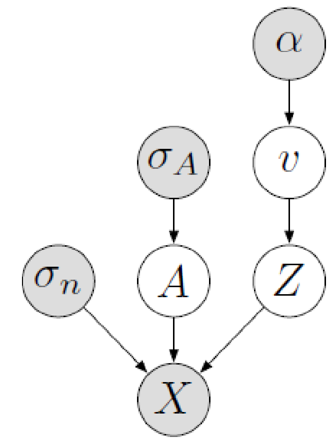
$$q_{\boldsymbol{\phi}_k}(\boldsymbol{A}_{k\cdot}) = \text{Normal}(\boldsymbol{A}_{k\cdot}; \bar{\phi}_k, \Phi_k),$$

$$q_{\nu_{nk}}(z_{nk}) = \text{Bernoulli}(Z_{nk}; \nu_{nk}).$$



Mean Field Variational Inference: Infinite Variational Approach:

Variational Lower Bound



$$\log p(\mathbf{x}|\theta) \geq \mathbb{E}_q\left[\log p(\mathbf{W}, \mathbf{x}|\theta)\right] - \mathbb{E}_q\left[\log q(\mathbf{W})\right]$$

$$\log p(\boldsymbol{X}|\boldsymbol{\theta}) \geq \sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{v}}\left[\log p(v_k|\alpha)\right] + \boxed{\sum_{k=1}^{K}\sum_{n=1}^{N} \mathbb{E}_{\boldsymbol{v},\boldsymbol{Z}}\left[\log p(Z_{nk}|\boldsymbol{v})\right]}$$
$$+ \sum_{k=1}^{K} \mathbb{E}_{\boldsymbol{A}}\left[\log p(\boldsymbol{A}_{k\cdot}|\sigma_A^2)\right] + \sum_{n=1}^{N} \mathbb{E}_{\boldsymbol{Z},\boldsymbol{A}}\left[\log p(\boldsymbol{X}_{n\cdot}|\boldsymbol{Z}, \boldsymbol{A}, \sigma_n^2)\right] + H[q]$$

Mean Field Variational Inference: Infinite Variational Approach:

Variational Lower Bound

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{v},\boldsymbol{Z}}\left[\log p(z_{nk}|\boldsymbol{v})\right] &= \mathbb{E}_{\boldsymbol{v},\boldsymbol{Z}}\left[\log p(z_{nk}=1|\boldsymbol{v})^{z_{nk}} p(z_{nk}=0|\boldsymbol{v})^{1-z_{nk}}\right] \\
&= \mathbb{E}_{\boldsymbol{Z}}\left[z_{nk}\right]\mathbb{E}_{\boldsymbol{v}}\left[\log \prod_{m=1}^{k} v_m\right] + \mathbb{E}_{\boldsymbol{Z}}\left[1-z_{nk}\right]\mathbb{E}_{\boldsymbol{v}}\left[\log\left(1-\prod_{m=1}^{k} v_m\right)\right] \\
&= \nu_{nk}\left(\sum_{m=1}^{k}\psi(\tau_{k2})-\psi(\tau_{k1}+\tau_{k2})\right) + (1-\nu_{nk})\mathbb{E}_{\boldsymbol{v}}\left[\log\left(1-\prod_{m=1}^{k} v_m\right)\right]
\end{aligned}
$$

- **Second line: definition of v**
- **Third line:**

$$E(\log X) = \psi(\alpha) - \psi(\alpha+\beta)$$

$$X \sim \mathrm{Beta}(\alpha,\beta)$$

Mean Field Variational Inference: Infinite Variational Approach:

Multinomial Lower Bound

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{v}}\left[\log\left(1-\prod_{m=1}^{k} v_m\right)\right] &= \mathbb{E}_{\boldsymbol{v}}\left[\log\left(\sum_{y=1}^{k}(1-v_y)\prod_{m=1}^{y-1} v_m\right)\right] \\
&= \mathbb{E}_{\boldsymbol{v}}\left[\log\left(\sum_{y=1}^{k} q_k(y)\frac{(1-v_y)\prod_{m=1}^{y-1} v_m}{q_k(y)}\right)\right] \\
&\geq \mathbb{E}_y\mathbb{E}_{\boldsymbol{v}}\left[\log(1-v_y)+\sum_{m=1}^{y-1}\log v_m\right]+H[q_k] \\
&= \mathbb{E}_y\left[\psi\left(\tau_{y2}\right)+\left(\sum_{m=1}^{y-1}\psi(\tau_{m1})\right)-\left(\sum_{m=1}^{y}\psi(\tau_{m1}+\tau_{m2})\right)\right]+H[q_k].
\end{aligned}$$

- **Third line: Jensen's inequality**
- **Last line:**

If $X \sim \text{Beta}(a,b)$ **then** $1-X \sim \text{Beta}(b,a)$

$\text{E}(\log X) = \psi(\alpha) - \psi(\alpha+\beta)$ **if** $X \sim \text{Beta}(\alpha,\beta)$

Mean Field Variational Inference: Infinite Variational Approach:

Multinomial Lower Bound

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{v}}\left[\log\left(1-\prod_{m=1}^{k} v_m\right)\right] &= \mathbb{E}_{\boldsymbol{v}}\left[\log\left(\sum_{y=1}^{k}(1-v_y)\prod_{m=1}^{y-1} v_m\right)\right] \\
&= \mathbb{E}_{\boldsymbol{v}}\left[\log\left(\sum_{y=1}^{k} q_k(y)\frac{(1-v_y)\prod_{m=1}^{y-1} v_m}{q_k(y)}\right)\right] \\
&\geq \mathbb{E}_y\mathbb{E}_{\boldsymbol{v}}\left[\log(1-v_y)+\sum_{m=1}^{y-1}\log v_m\right]+H[q_k] \\
&= \mathbb{E}_y\left[\psi(\tau_{y2})+\left(\sum_{m=1}^{y-1}\psi(\tau_{m1})\right)-\left(\sum_{m=1}^{y}\psi(\tau_{m1}+\tau_{m2})\right)\right]+H[q_k].
\end{aligned}
$$

- **Find q that maximizes this bound**

$$
q_{ki} \propto \exp\left(\psi(\tau_{i2})+\sum_{m=1}^{i-1}\psi(\tau_{m1})-\sum_{m=1}^{i}\psi(\tau_{m1}+\tau_{m2})\right)
$$

- **This bound gives closed form parameter updates for v**
- **Alternatively, we can use a Taylor series expansion of ln(1 - x), which gives a tighter bound but we will no longer have closed form updates**

Mean Field Variational Inference:

Infinite ICA

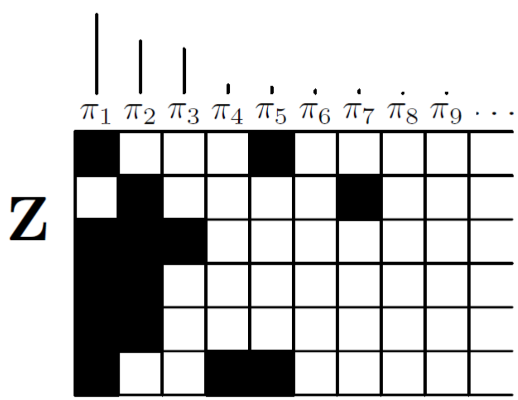
- Same techniques:

  Finite and infinite variational approaches
- Need to add variational updates for
  the signal matrix S
  - Use a Laplace approximation

Truncation Error

- For infinite variational approach
- Compare marginal distributions of the true IBP model and the truncated model
- Use Theorem 2 from Ishiwaran and James (2001)

$$
\begin{aligned}
\frac{1}{4}\int |m_K(\boldsymbol{X}) - m_\infty(\boldsymbol{X})| d\boldsymbol{X} &\leq \Pr(\exists k > K, n \text{ with } z_{nk} = 1) \\
&= 1 - \Pr\left(\text{all } z_{ik} = 0,\ i \in \{1, \dots, N\}, k > K\right) \\
&= 1 - \mathbb{E}\left[\left(\prod_{i=K+1}^{\infty} (1 - \pi_i)\right)^N\right] \\
&\leq 1 - \left(\mathbb{E}\left[\prod_{i=K+1}^{\infty} (1 - \pi_i)\right]\right)^N .
\end{aligned}
$$



Truncation Error

The Beta Process construction of the IBP (Thibaux and Jordan, 2007)

↓

Model $\pi_1, \pi_2, \ldots$ as a Poisson process on (0,1) with rate $\mu(x)dx = \alpha x^{-1}dx$
Model $\pi_{k+1}, \pi_{k+2}, \ldots$ as a Poisson process on $(0, \pi_k)$ with the same rate

↓

Apply the Levy-Khintchine formula:

the moment generating function of a Poisson process X with rate μ can be written as

$$\mathbb{E}[\exp(tf(X))] = \exp\left(\int (\exp(tf(y)) - 1)\,\mu(y)dy\right).$$

Truncation Error

Apply the Levy-Khintchine formula:

$$\mathbb{E}\left[\left(\prod_{i=K+1}^{\infty}(1-\pi_i)\right)\right] = \mathbb{E}\left[\exp\left(\sum_{i=K+1}^{\infty}\log(1-\pi_i)\right)\right]$$

$$\begin{aligned}\mathbb{E}\left[\exp\left(\sum_{i=K+1}^{\infty}\log(1-\pi_i)\right)\right] &= \mathbb{E}_{\pi_K}\left[\exp\left(\int_0^{\pi_K}(\exp(\log(1-x))-1)\,\mu(x)dx\right)\right] \\ &= \mathbb{E}_{\pi_K}[\exp(-\alpha\pi_K)].\end{aligned}$$

Apply Jensen's Inequality:

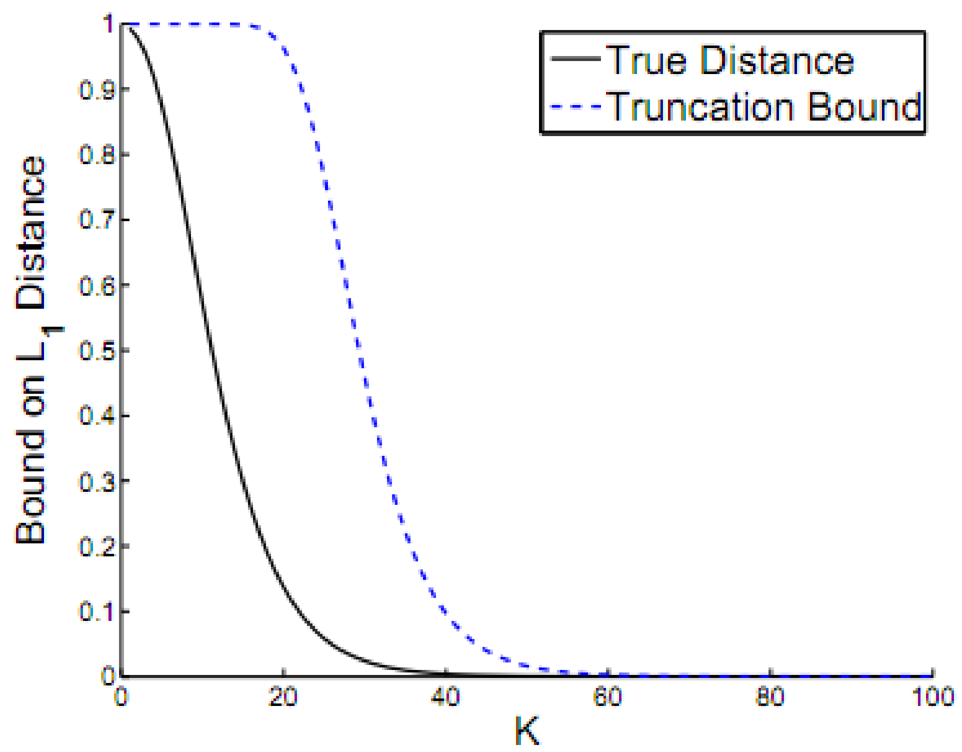
$$\begin{aligned}\mathbb{E}_{\pi_K}[\exp(-\alpha\pi_K)] &\geq \exp\left(\mathbb{E}_{\pi_k}[-\alpha\pi_K]\right) \\ &= \exp\left(-\alpha\left(\frac{\alpha}{1+\alpha}\right)^K\right)\end{aligned}$$

Truncation Error:
Final bound

$$\frac{1}{4}\int |m_K(X) - m_\infty(X)|dX \leq 1 - \exp\left(-N\alpha\left(\frac{\alpha}{1+\alpha}\right)^K\right)$$

Grows linearly with N
Decreases exponentially with K

Truncation Error



Truncation Error:

Heuristic bound using Taylor expansions

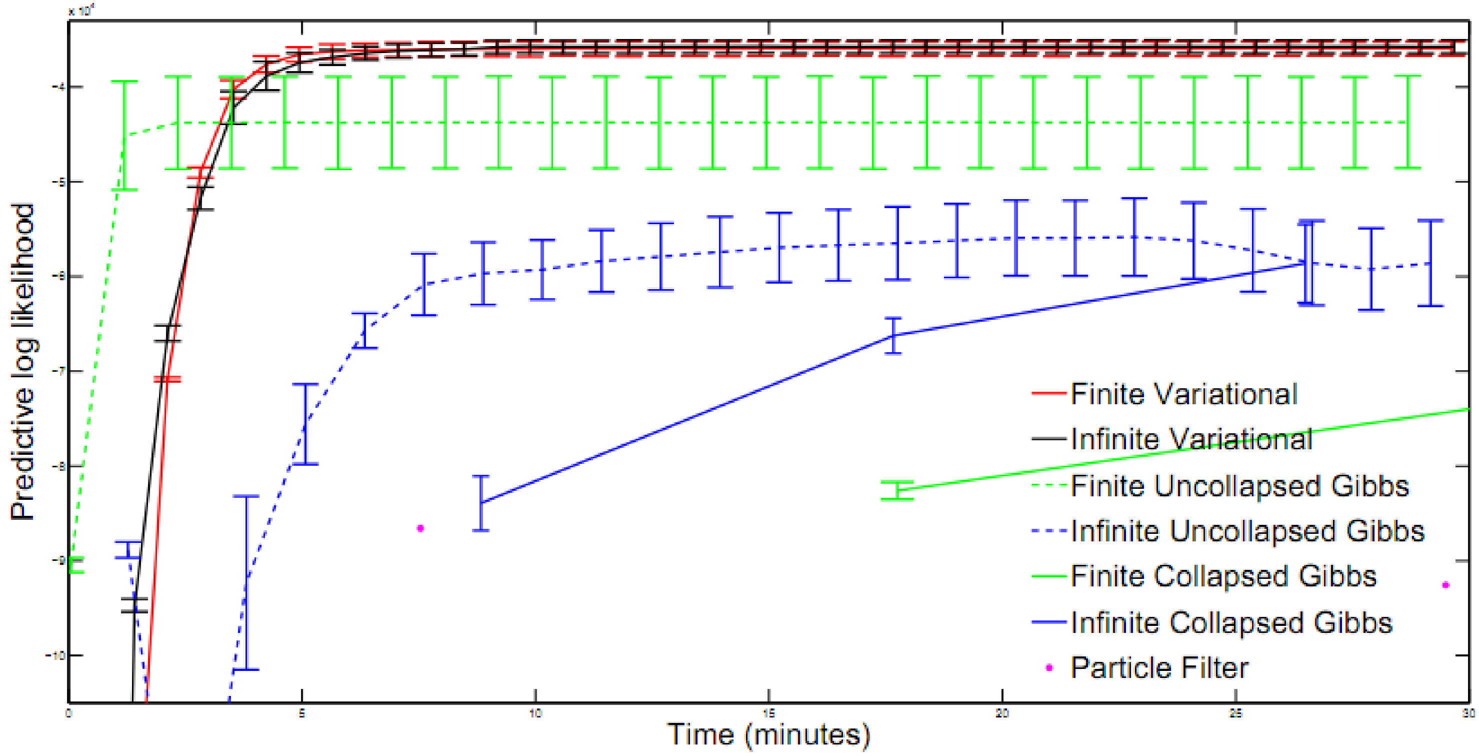
$$1 - \exp\left(-2N(\alpha+1)\left(\frac{\alpha}{\alpha+1}\right)^{K+1}\right)$$

Before:

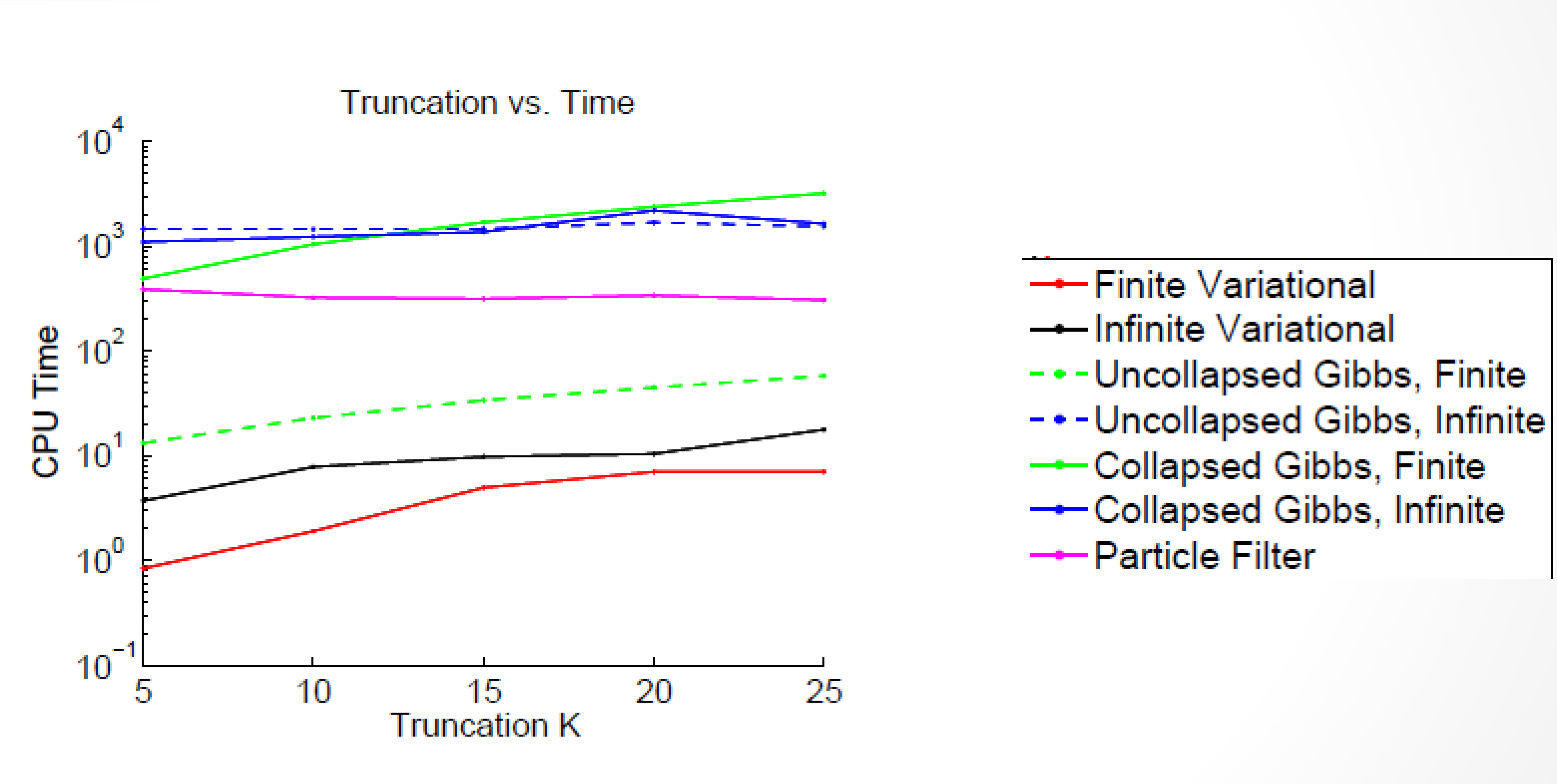
$$1 - \exp\left(-N\alpha\left(\frac{\alpha}{1+\alpha}\right)^{K}\right)$$

Test log-likelihoods on synthetic data

30-minute interval, N = 500, D = 500, K = 20, 5 initializations

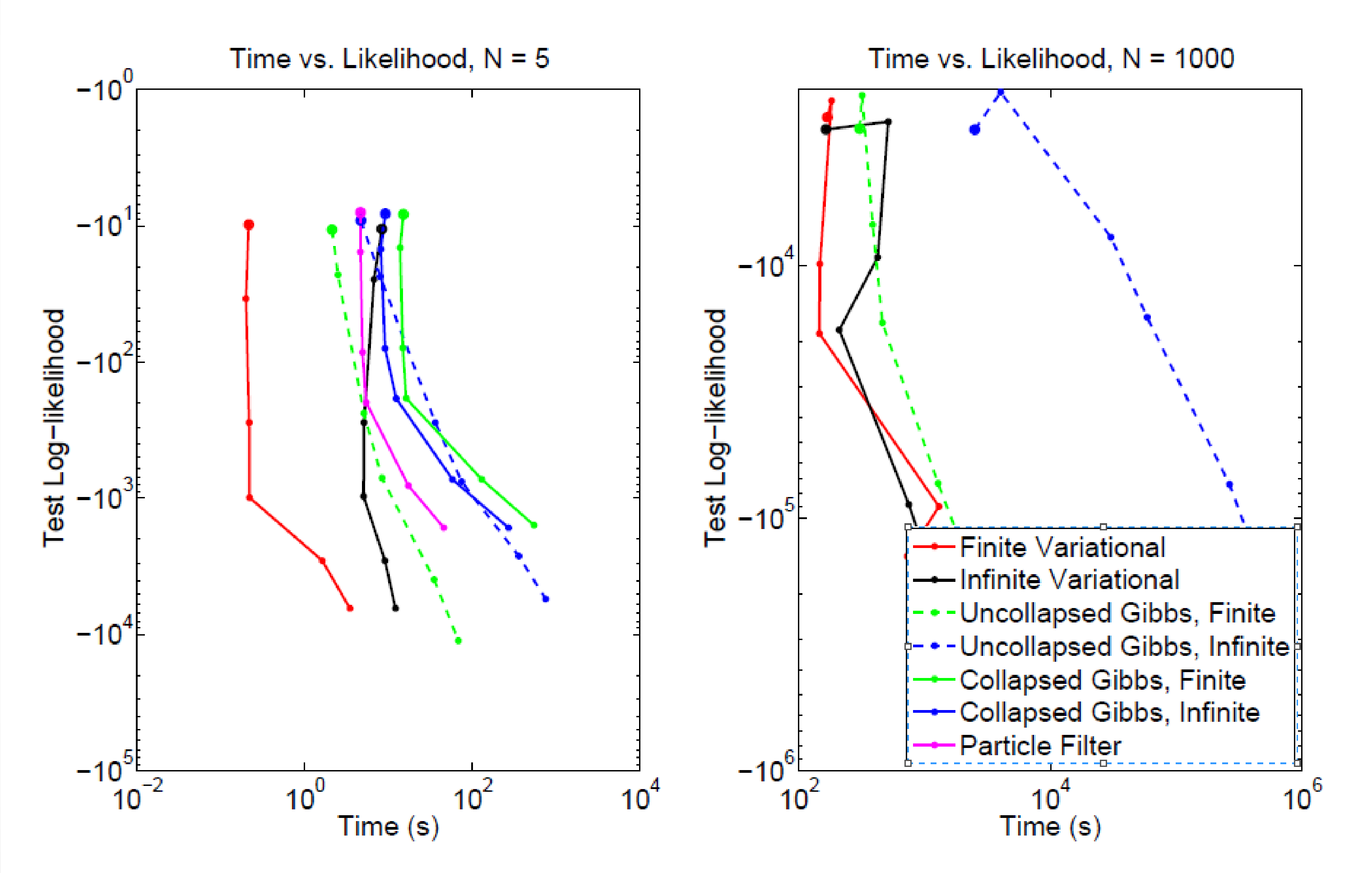


Truncation level vs. Time



Robustness to increasing dimension

D = 10, 50, 100, 500, 1000 (top to bottom), K = 20



The Effect of Dimensionality

Table 1: Running times in seconds and test log-likelihoods for the Yale Faces dataset.

Algorithm	K	Time	Test Log-Likelihood ($\times 10^6$)
Finite Gibbs	5	464.19	-2.250
	10	940.47	-2.246
	25	2973.7	-2.247
Finite Variational	5	163.24	-1.066
	10	767.1	-0.908
	25	10072	-0.746
Infinite Variational	5	176.62	-1.051
	10	632.53	-0.914
	25	19061	-0.750

Table 2: Running times in seconds and test log-likelihoods for the speech dataset.

Algorithm	K	Time	Test Log-Likelihood
Finite Gibbs	2	56	-0.7444
	5	120	-0.4220
	9	201	-0.4205
Infinite Gibbs	na	186	-0.4257
Finite Variational	2	2477	-0.8455
	5	8129	-0.5082
	9	8539	-0.4551
Infinite Variational	2	2702	-0.8810
	5	6065	-0.5000
	9	8491	-0.5486

References

- For a more detailed version of the paper, see this technical report http://mlg.eng.cam.ac.uk/pub/pdf/DosMilVanTeh09b.pdf