A Stick-Breaking Construction of the Beta Process

John Paisley, Aimee Zaas, Christopher W. Woods,
Geoffrey S. Ginsburg, Lawrence Carin

ICML 2010

CSCI 2950-P, Fall 2011
Michael Bryant

Outline

- Review/description of beta process
- A stick-breaking construction of beta distribution
- A stick-breaking construction of the BP
- Derivation
- Inference
- Experiments

Why?

1) Stick-breaking constructions are “fully Bayesian”
2) No marginalization needed
3) Leads the way to:
    - new inference algorithms
    - generalizations (think nCRP)
    - more parameterizations (two-parameter IBP
4) Because it's cool

Related constructions of the beta process

- *Stick-breaking construction for the Indian buffet process*
  - specifically for the IBP, limited to one-parameter case

- *Hierarchical beta processes and the Indian buffet process*
  - no “true” stick-breaking as in Sethuraman's construction

- *Indian buffet processes with power-law behavior* *(Teh & Görür, 2009)*
  - much more measure-theoretic; exists only in theory

- *Beta processes, stick-breaking, and power laws*
  - similar, but cleaner and provides a Pitman-Yor extension

The beta process

Infinite number of coin tossing probabilities

$$\begin{aligned} H_K &= \sum_{k=1}^{K} \pi_k \delta_{\theta_k} \\ \pi_k &\overset{iid}{\sim} \text{Beta}\left(\frac{\alpha\gamma}{K}, \alpha(1-\frac{\gamma}{K})\right) \\ \theta_k &\overset{iid}{\sim} \frac{1}{\gamma} H_0 \\ & K \to \infty,\, H_K \to H \\ & H \sim \text{BP}(\alpha H_0) \end{aligned}$$

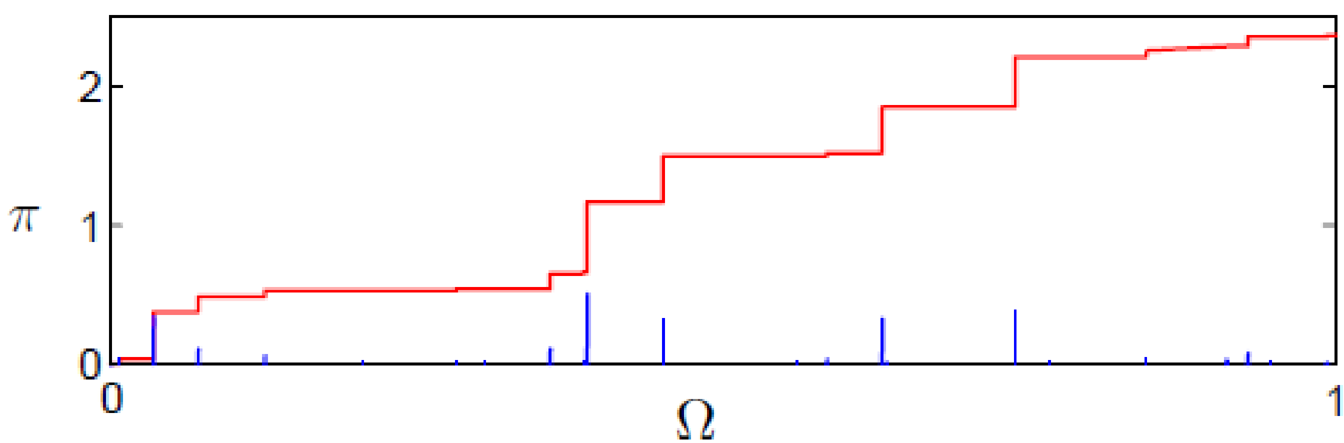


Figure from Thibaux & Jordan, 2007

Alternate construction of the beta distribution

From Sethuraman (1994)

To draw $\pi \sim \mathrm{Beta}(a, b)$:

$$\begin{aligned}
\pi &= \sum_{i=1}^{\infty} V_i \prod_{j=1}^{i-1} (1 - V_j) \mathbb{I}(Y_i = 1) \\
V_i &\overset{iid}{\sim} \mathrm{Beta}(1, a + b) \\
Y_i &\overset{iid}{\sim} \mathrm{Bernoulli}\left(\frac{a}{a + b}\right)
\end{aligned}$$

The Dirichlet process

$$\begin{aligned}
G &\sim \mathrm{DP}(\alpha H) \\
G &= \sum_{i=1}^{\infty} V_i \prod_{j=1}^{i-1} (1 - V_j) \delta_{\theta_i} \\
V_i &\overset{iid}{\sim} \mathrm{Beta}(1, \alpha) \\
\theta_i &\overset{iid}{\sim} H
\end{aligned}$$

Stick-breaking construction of the beta process (1)

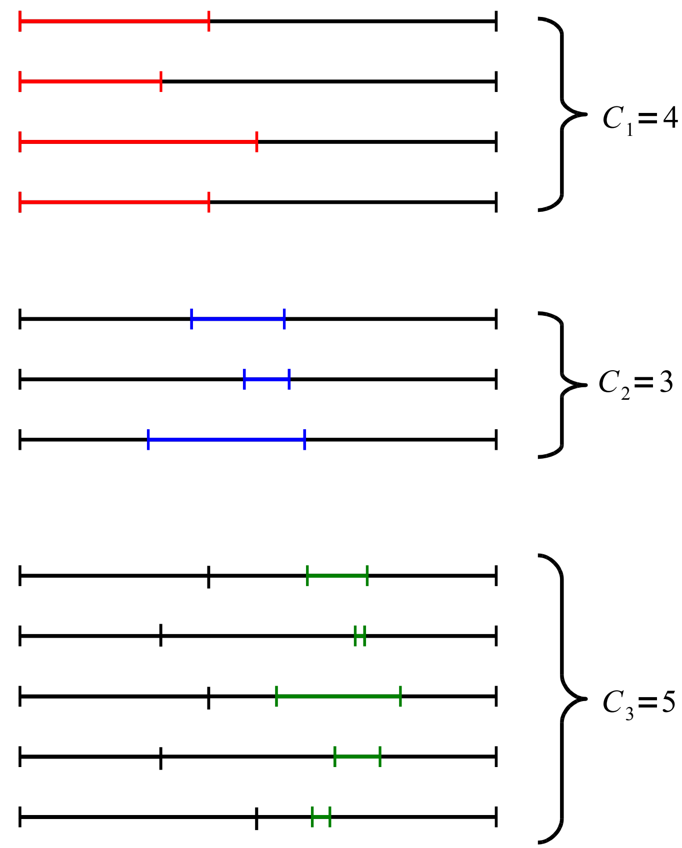
$$
\begin{aligned}
H &= \sum_{i=1}^{\infty}\sum_{j=1}^{C_i} V_{ij}^{(i)} \prod_{\ell=1}^{i-1}(1 - V_{ij}^{(\ell)})\delta_{\theta_{ij}} \\
C_i &\overset{iid}{\sim} \text{Poisson}(\gamma) \\
V_{ij}^{(\ell)} &\overset{iid}{\sim} \text{Beta}(1, \alpha) \\
\theta_{ij} &\overset{iid}{\sim} \frac{1}{\gamma} H_0
\end{aligned}
$$

$$\Downarrow$$

$$H \sim \text{BP}(\alpha H_0)$$

Stick-breaking construction of the beta process (2)

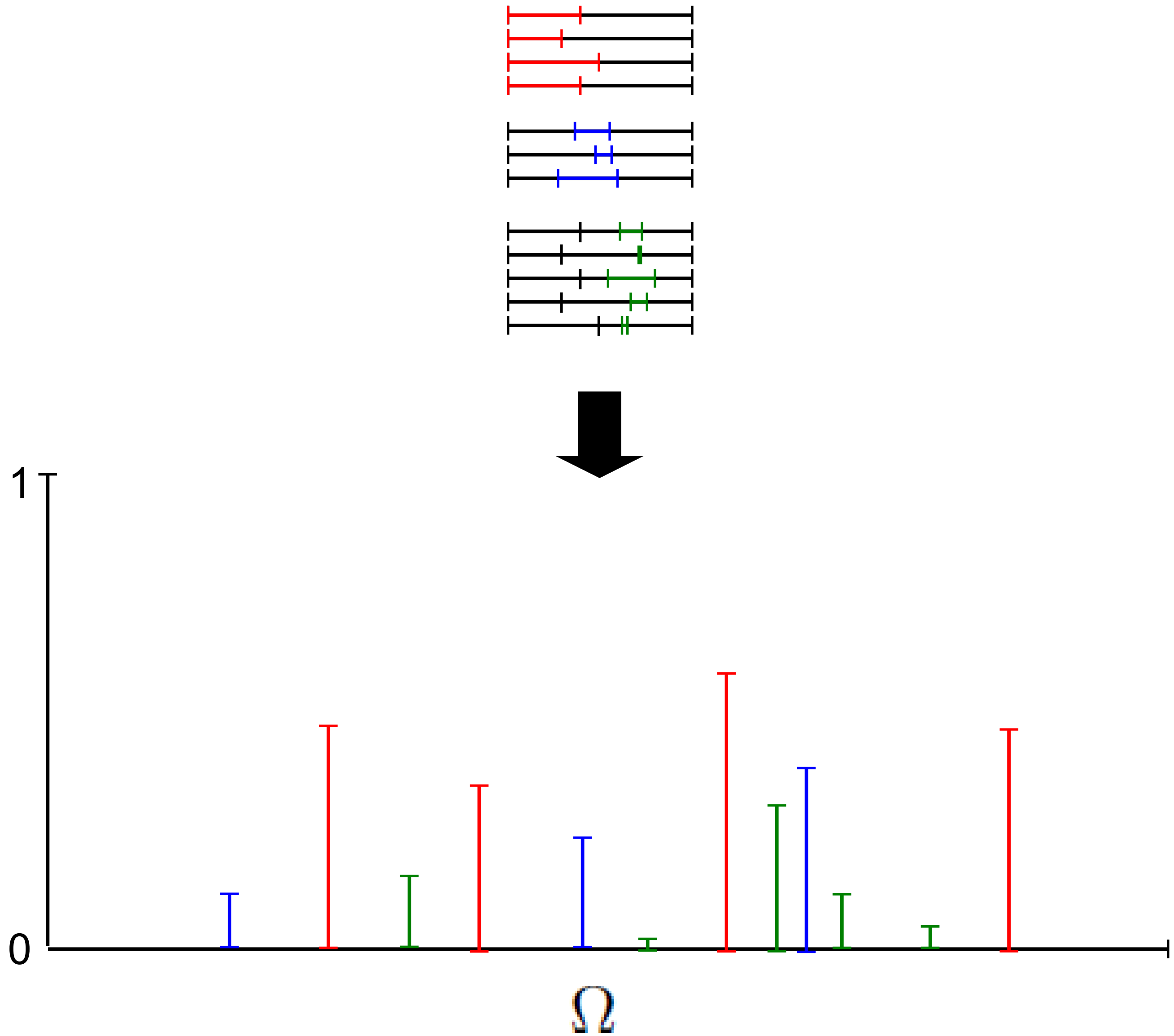
$$H = \sum_{j=1}^{C_1} V_{1,j}^{(1)} \delta_{\theta_{1,j}} + \sum_{j=1}^{C_2} V_{2,j}^{(2)} (1 - V_{2,j}^{(1)}) \delta_{\theta_{2,j}} + \sum_{j=1}^{C_3} V_{3,j}^{(3)} (1 - V_{3,j}^{(2)})(1 - V_{3,j}^{(1)}) \delta_{\theta_{3,j}} + \cdots$$



Expected weight of an atom in round *i*

$\alpha^{(i-1)}/(1+\alpha)^i$

Stick-breaking construction of the beta process (3)



Derivation (1)

Beta process

$$
\begin{aligned}
H_K &= \sum_{k=1}^{K} \pi_k \delta_{\theta_k} \\
\pi_k &\overset{iid}{\sim} \mathrm{Beta}\left(\overset{\text{a}}{\frac{\alpha\gamma}{K}}, \overset{\text{b}}{\alpha(1 - \frac{\gamma}{K})}\right) \\
\theta_k &\overset{iid}{\sim} \frac{1}{\gamma} H_0
\end{aligned}
$$

Beta process (stick-broken)

$$
\begin{aligned}
H &= \sum_{i=1}^{\infty} \sum_{j=1}^{C_i} V_{ij}^{(i)} \prod_{\ell=1}^{i-1} (1 - V_{ij}^{(\ell)}) \delta_{\theta_{ij}} \\
C_i &\overset{iid}{\sim} \mathrm{Poisson}(\gamma) \\
V_{ij}^{(\ell)} &\overset{iid}{\sim} \mathrm{Beta}(1, \alpha) \\
\theta_{ij} &\overset{iid}{\sim} \frac{1}{\gamma} H_0
\end{aligned}
$$

$$
\begin{aligned}
\pi_k &= \sum_{l=1}^{\infty} \hat{V}_{kl} \prod_{m=1}^{l-1} (1 - \hat{V}_{km}) \mathbb{I}(\hat{Y}_{kl} = 1) \\
\hat{V}_{kl} &\overset{iid}{\sim} \mathrm{Beta}(1, \alpha) \longleftarrow \text{a + b} \\
\hat{Y}_{kl} &\overset{iid}{\sim} \mathrm{Bernoulli}\left(\frac{\gamma}{K}\right) \longleftarrow \text{a / (a + b)}
\end{aligned}
$$

Alternate beta representation

Derivation (2)

Procedure for constructing the limit of $\boldsymbol{\pi}^{(K)}$

$$\hat{\boldsymbol{V}} \in (0,1)^{K \times K} \qquad\qquad \hat{\boldsymbol{Y}} \in \{0,1\}^{K \times K}$$

$$\hat{V}_{kl} \overset{iid}{\sim} \mathrm{Beta}(1,\alpha) \qquad\qquad \hat{Y}_{kl} \overset{iid}{\sim} \mathrm{Bernoulli}\left(\frac{\gamma}{K}\right)$$

$$\begin{bmatrix} 0.56 & 0.25 & 0.65 & \cdots \\ 0.12 & 0.73 & & \\ 0.39 & & \ddots & \\ \vdots & & & \end{bmatrix} + \begin{bmatrix} 0 & 1 & 1 & \cdots \\ 1 & 0 & & \\ 0 & & \ddots & \\ \vdots & & & \end{bmatrix} \Rightarrow \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \end{bmatrix}$$

Instead of constructing all of $\hat{\boldsymbol{Y}}$, draw only the indices $\{(k,l) : \hat{Y}_{kl} = 1\}$

Derivation (3)

$$\left.\begin{bmatrix} 0 & 1 & 1 & \cdots \\ 1 & 0 & & \\ 0 & & \ddots & \\ \vdots & & & \end{bmatrix}\right\} \sum_{k=1}^{K} \hat{Y}_{ki} \overset{iid}{\sim} \text{Poisson}(\gamma)$$

$$\text{Poisson}(\gamma) = \lim_{K \to \infty} \text{Binomial}\left(K, \frac{\gamma}{K}\right)$$

New procedure:
1) Draw number of non-zero locations ~ Poisson(gamma)
2) Draw indices of non-zero locations uniformly from {1, …, K}
3) Draw atoms associated with locations
4) Re-index the locations and atoms

Inference (1)

Rewrite *H* in terms of a round indicator; use MCMC methods.

Round indicator

$$d_k := 1 + \sum_{i=1}^{\infty} \mathbb{I}\left(\sum_{j=1}^{i} C_j < k\right)$$

$d_k = r$ means k^{th} atom overall is drawn in round *r*

Rewritten beta process

$$\begin{aligned} H \mid \{d_k\}_{k=1}^{\infty} &= \sum_{k=1}^{\infty} V_{k,d_k} \prod_{j=1}^{d_k - 1} (1 - V_{kj}) \delta_{\theta_k} \\ V_{kj} &\overset{iid}{\sim} \mathrm{Beta}(1, \alpha) \\ \theta_k &\overset{iid}{\sim} \frac{1}{\gamma} H_0 \end{aligned}$$

Sample:

d_k - round indicators

α - concentration parameter

γ - mass parameter

z_{nk} - components of the binary matrix

Inference (2)

Sample round indicators:

$$p\left(d_k = i \mid \overset{1}{\{d_l\}_{l=1}^{k-1}}, \{z_{nk}\}_{n=1}^{N}, \alpha, \gamma\right) \propto p(\overset{2}{\{z_{nk}\}_{n=1}^{N}} \mid d_k = i, \alpha) p(d_k = i \mid \{d_l\}_{l=1}^{k-1}, \gamma)$$

Sample mass parameter:

$$p\left(\gamma \mid \{d_i\}_{i=1}^{K}, \{z_n\}_{n=1}^{N}, \alpha\right)$$

Sample concentration parameter:

$$p(\alpha \mid \overset{3}{\{z_n\}_1^N}, \{d_k\}_1^K) \propto \prod_{k=1}^{K} p(\{z_{nk}\}_1^N \mid \overset{4}{\alpha}, \{d_k\}_1^K) p(\alpha)$$

Sample components of the binary vectors:

$$p(z_{nk} = 1 \mid \alpha, d_k, Z_{\text{prev}}) = \frac{\int_{(0,1)^{d_k}} p(\overset{5}{z_{nk} = 1} \mid \vec{V}) p(Z_{\text{prev}} \mid \vec{V}) p(\vec{V} \mid \alpha, d_k)\, d\vec{V}}{\int_{(0,1)^{d_k}} p(Z_{\text{prev}} \mid \vec{V}) p(\vec{V} \mid \alpha, d_k)\, d\vec{V}}$$

Inference (3)

Their inference algorithm relies heavily on Monte Carlo integration to approximate intractable or computationally expensive integrals.

An alternative: variational inference! (Paisley ICML 2011), a supplemental reading for today.

$$
\begin{aligned}
H &= \sum_{k=1}^{\infty} V_k \mathrm{e}^{-T_k} \delta_{\omega_k}, \\
V_k &\stackrel{iid}{\sim} \mathrm{Beta}(1, \alpha), \\
T_k &\sim \mathrm{Gamma}(d_k - 1, \alpha), \\
\sum_{k=1}^{\infty} \mathbf{1}_{d_k}(r) &\stackrel{iid}{\sim} \mathrm{Poisson}(\gamma), \quad r \in \mathbb{N}_+, \\
\omega_k &\stackrel{iid}{\sim} \frac{1}{\gamma} H_0.
\end{aligned}
$$

Inference (2)

Given $C_1, C_2, \ldots$ and with a conjugate prior $\gamma \sim \mathrm{Gamma}(a, b)$, we can sample from the posterior of γ.

The posterior of alpha is obtained by again integrating out the stick-breaking random variables *V:*

$$p(\alpha | \{z_n\}_1^N, \{d_k\}_1^K) \propto \prod_{k=1}^{K} p(\{z_{nk}\}_1^N | \alpha, \{d_k\}_1^K) p(\alpha)$$

which is approximated by Monte Carlo integration at a discretized set of points

Inference (3)

Finally, in the latent factor models, sample the binary matrix *Z* using:

$$p(z_{nk} = 1|\alpha, d_k, Z_{\text{prev}}) = \int_{(0,1)^{d_k}} p(z_{nk} = 1|\vec{V})p(\vec{V}|\alpha, d_k, Z_{\text{prev}})\; d\vec{V}$$

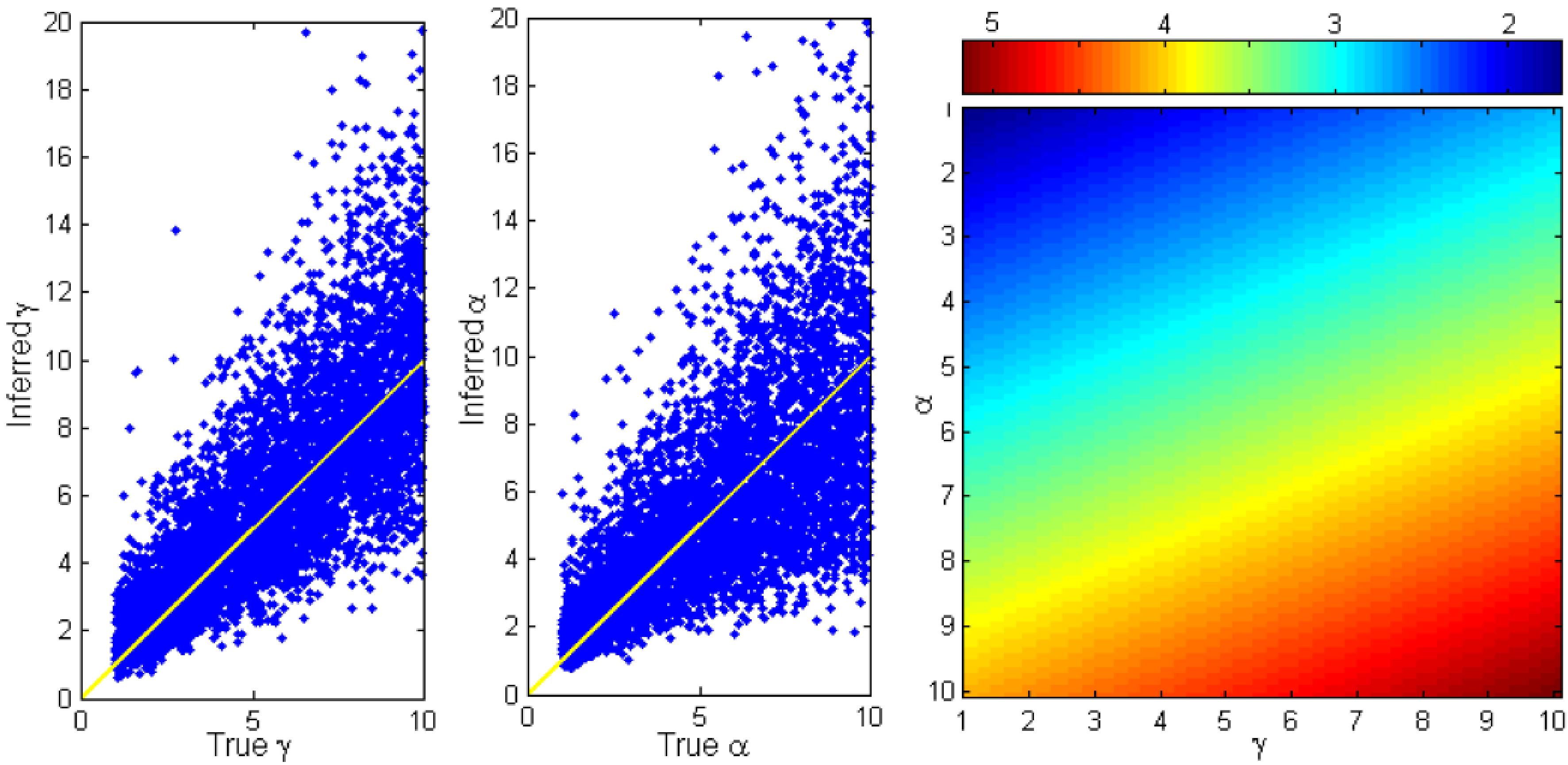
$$= \frac{\int_{(0,1)^{d_k}} p(z_{nk} = 1|\vec{V})p(Z_{\text{prev}}|\vec{V})p(\vec{V}|\alpha, d_k)\; d\vec{V}}{\int_{(0,1)^{d_k}} p(Z_{\text{prev}}|\vec{V})p(\vec{V}|\alpha, d_k)\; d\vec{V}}$$

Again use the Monte Carlo integration from before.

Experiments (synthetic data)

- Generate $\pi^{(K)}$ for $K = 100{,}000$

- Sample $\{z_n\}_{n=1}^{1000}$ from a Bernoulli process with these underlying probabilities



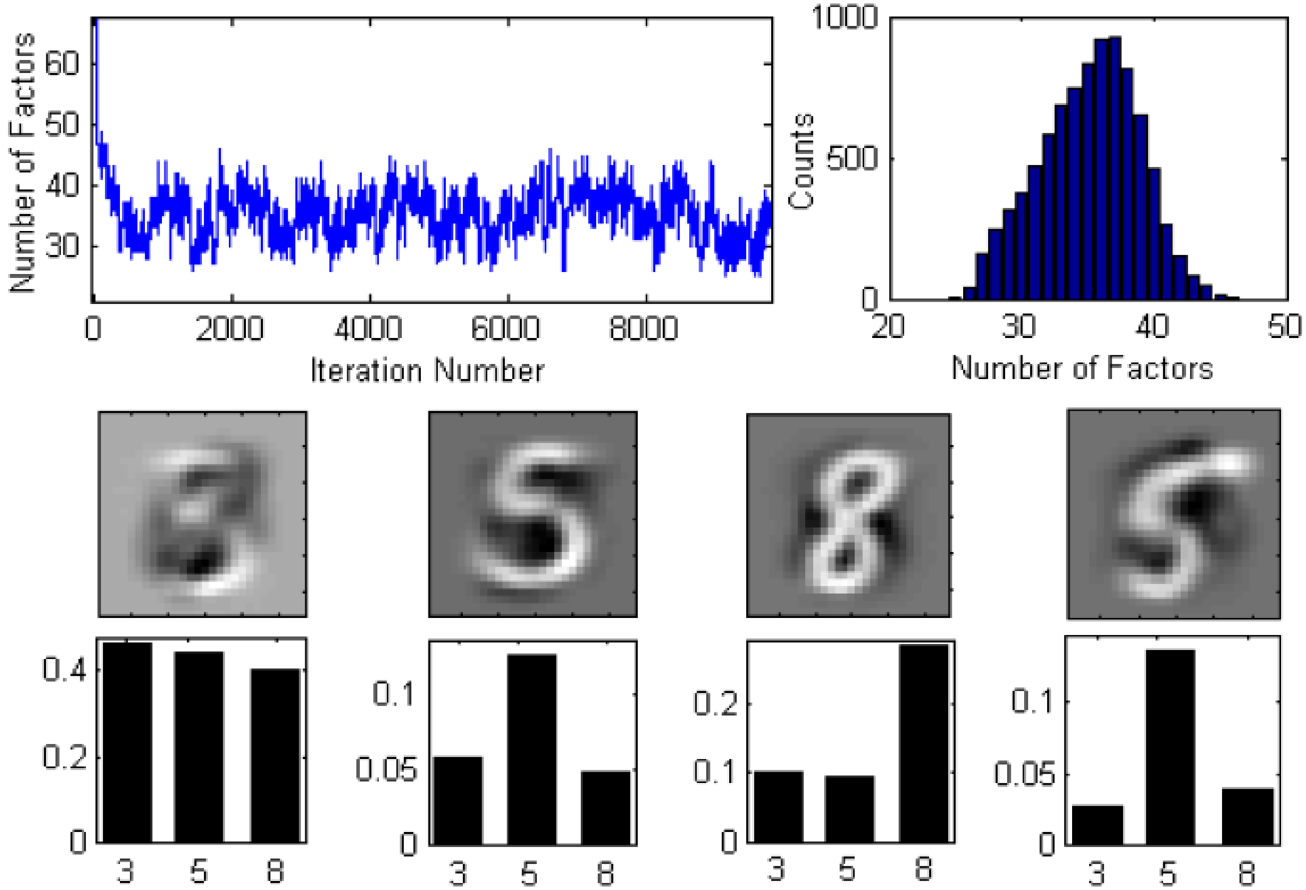
(a) (b) (c)

Experiments (MNIST)

- Latent factor model (Griffiths & Ghahramani, 2005)

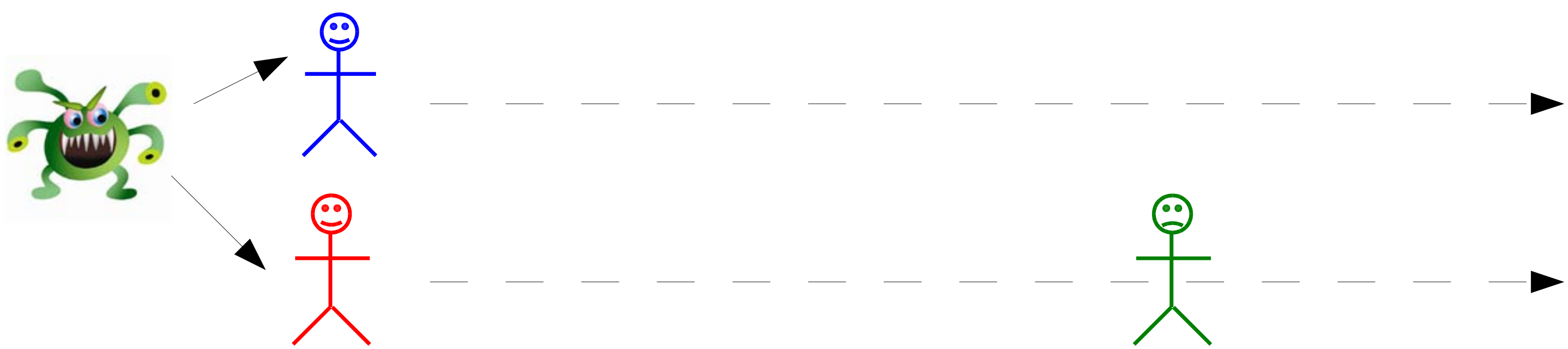
$$X = \Phi(W \circ Z) + E$$

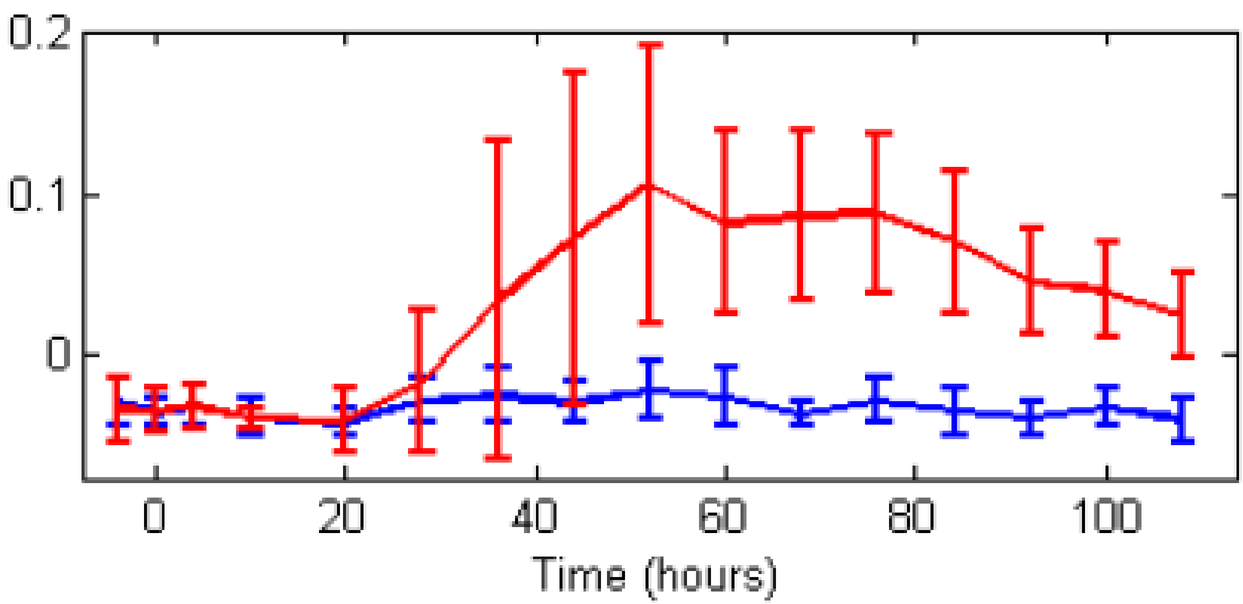
- Joint MCMC and variational inference method



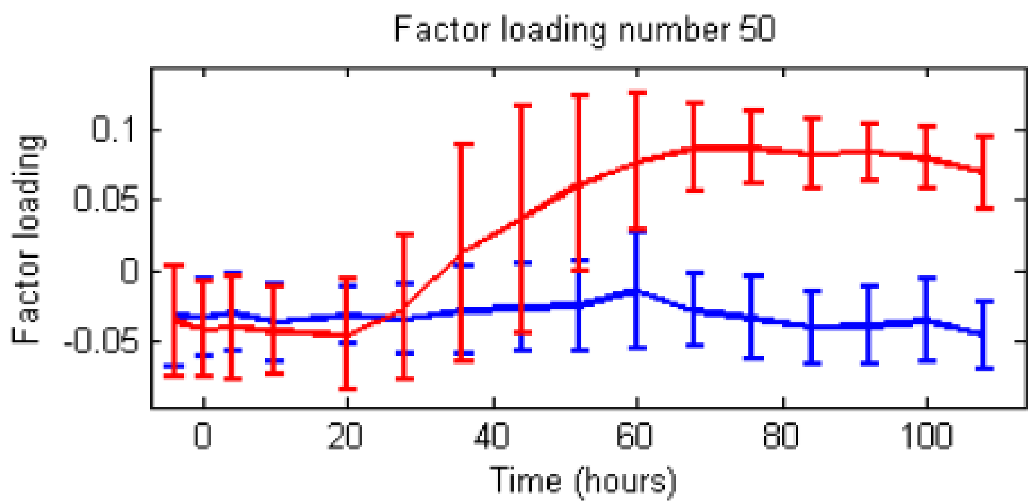
Experiments ("viral challenge" data)

Same latent factor model as for MNIST digits.



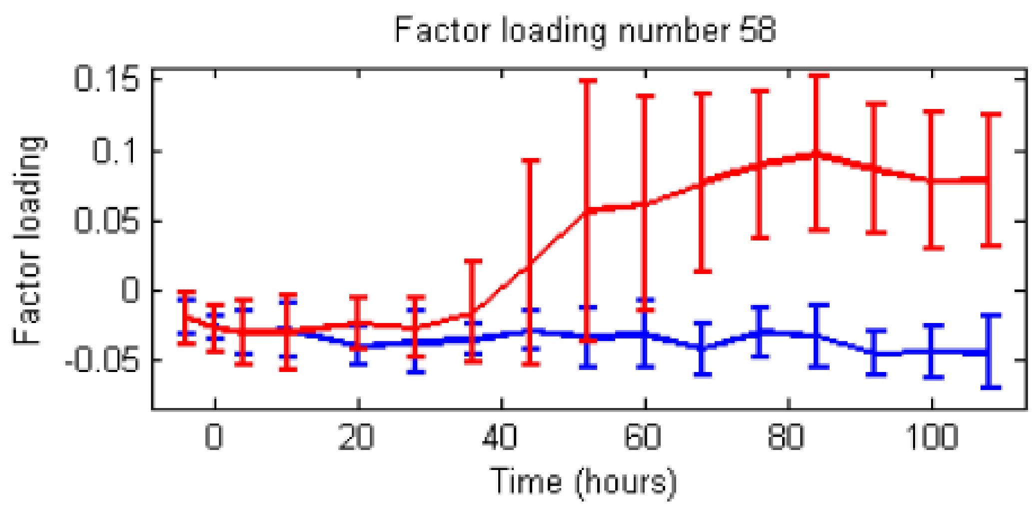


Experiments ("viral challenge" data)



Top 5 genes associated with this factor:

'RSAD2'
'IFIT1'
'IFI44L'
'IFI44'
'OAS3'



Top 5 genes associated with this factor:

'IFI27'
'PI3'
'C1QB'
'OTOF'
'SIGLEC1'

Questions?