Alternative Priors for Deep Generative Models

Eric Nalisnick
University of California, Irvine

In collaboration with

Padhraic Smyth

Outline

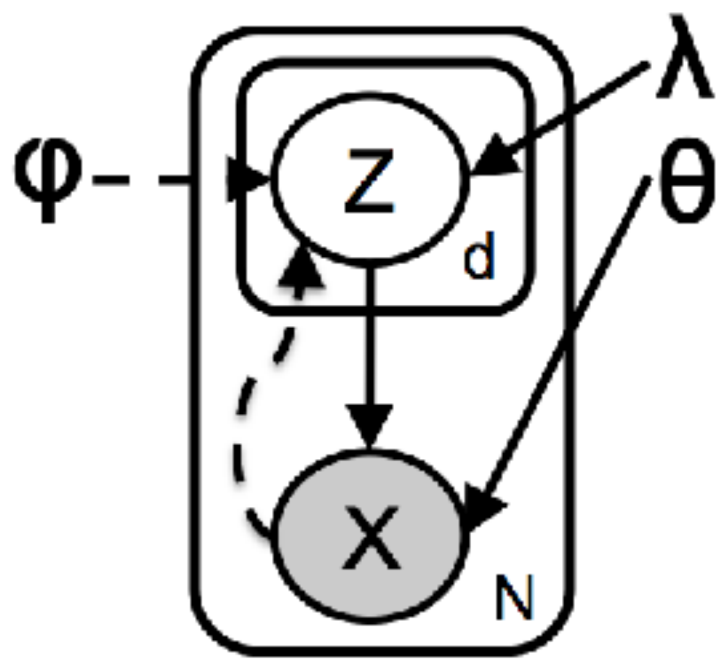
1. **Motivation:** Overview of research on deep generative models, and why we should consider priors different than the current ones.

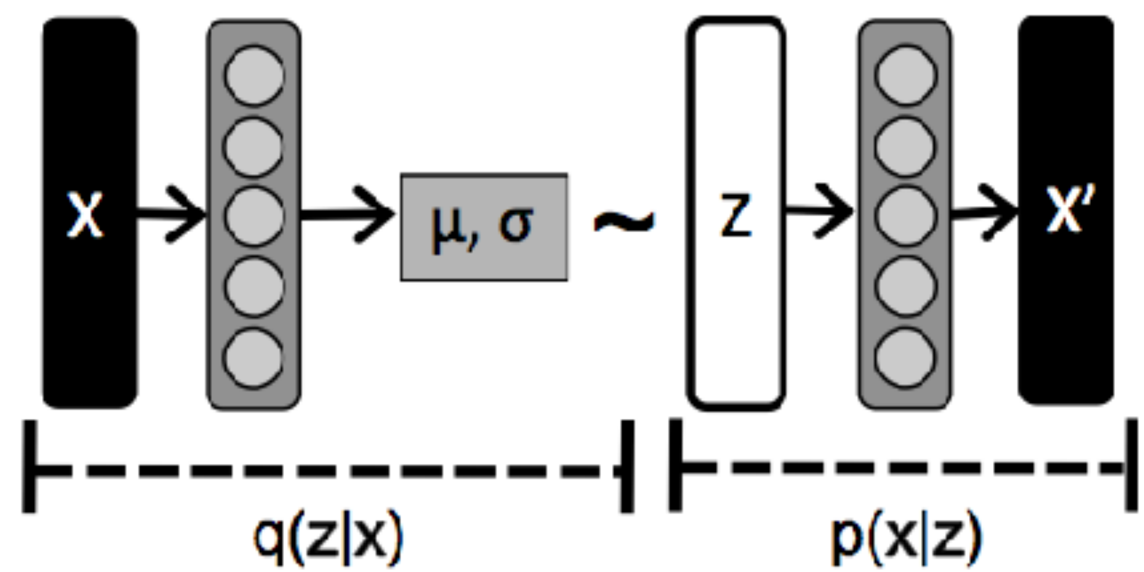
2. **Non-Parametric Priors:** Variational Autoencoders with infinite dimensional latent spaces via Dirichlet Process priors

3. **Objective Priors:** Black-box learning of invariant priors with an application to Variational Autoencoders.

The Variational Autoencoder (VAE)

(Kingma & Welling, 2014), (Rezende et al., 2014), (MacKay & Gibbs, 1997)





$$\log p_{\boldsymbol{\theta}}(\mathbf{x}_i) \geq \mathbb{E}_q[\log p_{\boldsymbol{\theta}}(\mathbf{x}_i|\mathbf{z}_i)] - KLD[q_{\boldsymbol{\phi}}(\mathbf{z}_i|\mathbf{x}_i)||p_{\boldsymbol{\lambda}}(\mathbf{z}_i)]$$

$$\approx \frac{1}{S}\sum_{s=1}^{S} \log p_{\boldsymbol{\theta}}(\mathbf{x}_i|\hat{\mathbf{z}}_{i,s}) - KLD[q_{\boldsymbol{\phi}}(\mathbf{z}_i|\mathbf{x}_i)||p_{\boldsymbol{\lambda}}(\mathbf{z}_i)]$$

$$q_{\boldsymbol{\phi}}(\mathbf{z}_i|\mathbf{x}_i) \;\approx\; p(\mathbf{z}_i|\mathbf{x}_i) \;\propto\; p_{\boldsymbol{\theta}}(\mathbf{x}_i|\mathbf{z}_i)\, p_{\boldsymbol{\lambda}}(\mathbf{z}_i)$$

$$q_{\phi}(\mathbf{z}_i|\mathbf{x}_i) \approx p(\mathbf{z}_i|\mathbf{x}_i) \propto p_{\boldsymbol{\theta}}(\mathbf{x}_i|\mathbf{z}_i)\, p_{\boldsymbol{\lambda}}(\mathbf{z}_i)$$

Inference Models
- Regression (Salimans & Knowles, 2014)
- Neural Networks (Kingma & Welling, 2014) (Rezende et al., 2014)
- Gaussian Processes (Tran et al., 2016)

Approximations via Transformation
- Normalizing Flows (Rezende & Mohamed, 2015)
- Hamiltonian Flow (Salimans et al, 2015)
- Inv. Auto-Regressive (Kingma et al., 2016)

Implicit Posterior Approximations
- Stein Particle Descent (Liu & Wang, 2016)
- Operator VI (Ranganath et al., 2016)
- Adversarial VB (Mescheder et al., 2017)

$$q_{\boldsymbol{\phi}}(\mathbf{z}_i|\mathbf{x}_i) \approx p(\mathbf{z}_i|\mathbf{x}_i) \propto p_{\boldsymbol{\theta}}(\mathbf{x}_i|\mathbf{z}_i)\, p_{\boldsymbol{\lambda}}(\mathbf{z}_i)$$

Inference Models
- Regression (Salimans & Knowles, 2014)
- Neural Networks (Kingma & Welling, 2014) (Rezende et al., 2014)
- Gaussian Processes (Tran et al., 2016)

Approximations via Transformation
- Normalizing Flows (Rezende & Mohamed, 2015)
- Hamiltonian Flow (Salimans et al, 2015)
- Inv. Auto-Regressive (Kingma et al., 2016)

Implicit Posterior Approximations
- Stein Particle Descent (Liu & Wang, 2016)
- Operator VI (Ranganath et al., 2016)
- Adversarial VB (Mescheder et al., 2017)

Direct Estimation of Model Evidence
- Importance Sampling (Burda et al., 2015)
- Random Projections (Grover & Ermon, 2016)

Other Divergence Measures
- Alpha (Hernandez-Lobato et al., 2016)
- Renyi (Li & Turner, 2016)
- Stein (Ranganath et al., 2016)

$$q_{\boldsymbol{\phi}}(\mathbf{z}_i|\mathbf{x}_i) \approx p(\mathbf{z}_i|\mathbf{x}_i) \propto p_{\boldsymbol{\theta}}(\mathbf{x}_i|\mathbf{z}_i)\, p_{\boldsymbol{\lambda}}(\mathbf{z}_i)$$

Inference Models
- Regression (Salimans & Knowles, 2014)
- Neural Networks (Kingma & Welling, 2014) (Rezende et al., 2014)
- Gaussian Processes (Tran et al., 2016)

Approximations via Transformation
- Normalizing Flows (Rezende & Mohamed, 2015)
- Hamiltonian Flow (Salimans et al, 2015)
- Inv. Auto-Regressive (Kingma et al., 2016)

Implicit Posterior Approximations
- Stein Particle Descent (Liu & Wang, 2016)
- Operator VI (Ranganath et al., 2016)
- Adversarial VB (Mescheder et al., 2017)

Direct Estimation of Model Evidence
- Importance Sampling (Burda et al., 2015)
- Random Projections (Grover & Ermon, 2016)

Other Divergence Measures
- Alpha (Hernandez-Lobato et al., 2016)
- Renyi (Li & Turner, 2016)
- Stein (Ranganath et al., 2016)

Empirical Priors
- Learned Auto-Regressive Prior (Chen et al., 2016)

“…to improve our variational bounds we should improve our priors and not just the encoder and decoder….perhaps we should investigate multimodal priors…”

M. Hoffman & M. Johnson. “ELBO Surgery”. NIPS 2016 Workshop on *Advances in Approx. Bayesian Inference*.

Part 1: Non-Parametric Priors

Publications

1. E. Nalisnick and P. Smyth. “Stick-Breaking Variational Autoencoders”. *ICLR 2017*.
2. E. Nalisnick, Lars Hertel, and P. Smyth. “Approximate Inference for Deep Latent Gaussian Mixtures”. NIPS 2016 Workshop on *Bayesian Deep Learning*.

The Dirichlet Process

$$G(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\zeta_k}$$

$$\pi_k = \begin{cases} v_1 \text{ if } k = 1 \\ v_k \prod_{j<k} (1 - v_j) \text{ for } k > 1 \end{cases} \qquad v_k \sim \text{Beta}(\alpha, \beta)$$

The Dirichlet Process

$$G(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\zeta_k}$$

$$\pi_k = \begin{cases} v_1 \text{ if } k = 1 \\ v_k \prod_{j<k} (1 - v_j) \text{ for } k > 1 \end{cases} \qquad v_k \sim \mathrm{Beta}(\alpha, \beta)$$

Two Obstacles:

1. Gradients through π_k
2. Gradients thought samples from $G(\cdot)$

Differentiating Through π_k

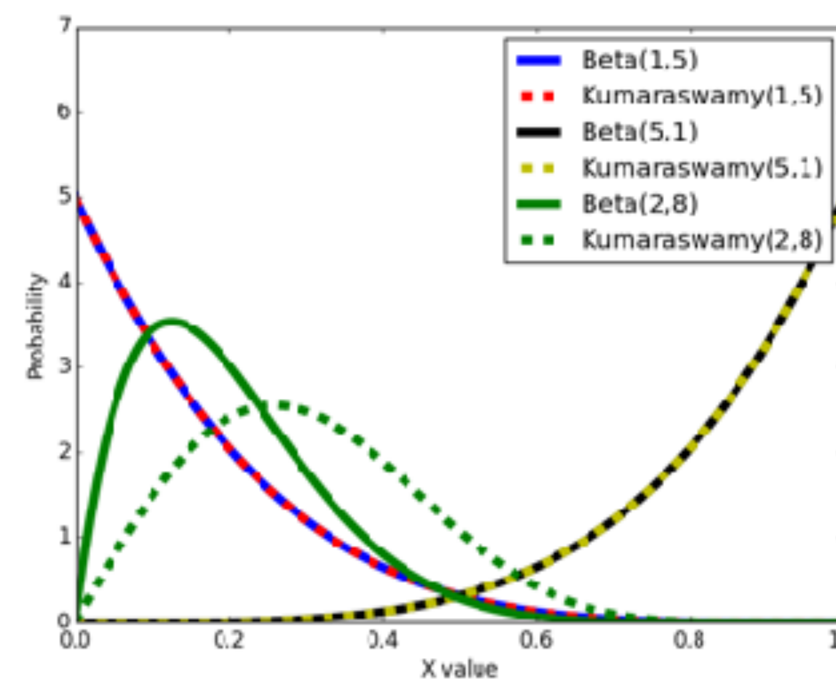
Obstacle: The Beta distribution does not have a non-centered parametrization (except in special cases)

Kumaraswamy Distribution: A Beta-like distribution with a closed-form inverse CDF. Use as variational posterior.

Poondi Kumaraswamy
(1930-1988)

$$\text{Kumaraswamy}(x; a, b) = abx^{a-1}(1 - x^a)^{b-1}$$

$$x \sim (1 - u^{\frac{1}{b}})^{\frac{1}{a}} \text{ where } u \sim \text{Uniform}(0, 1)$$



Stochastic Backpropagation through Mixtures

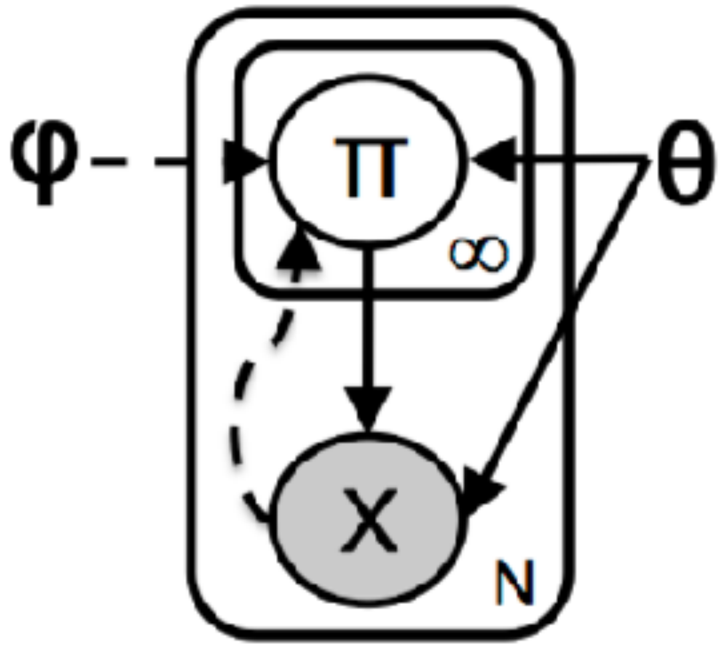
Obstacle: Not obvious how to use the *reparametrization trick* for samples from a mixture distribution.

Σ

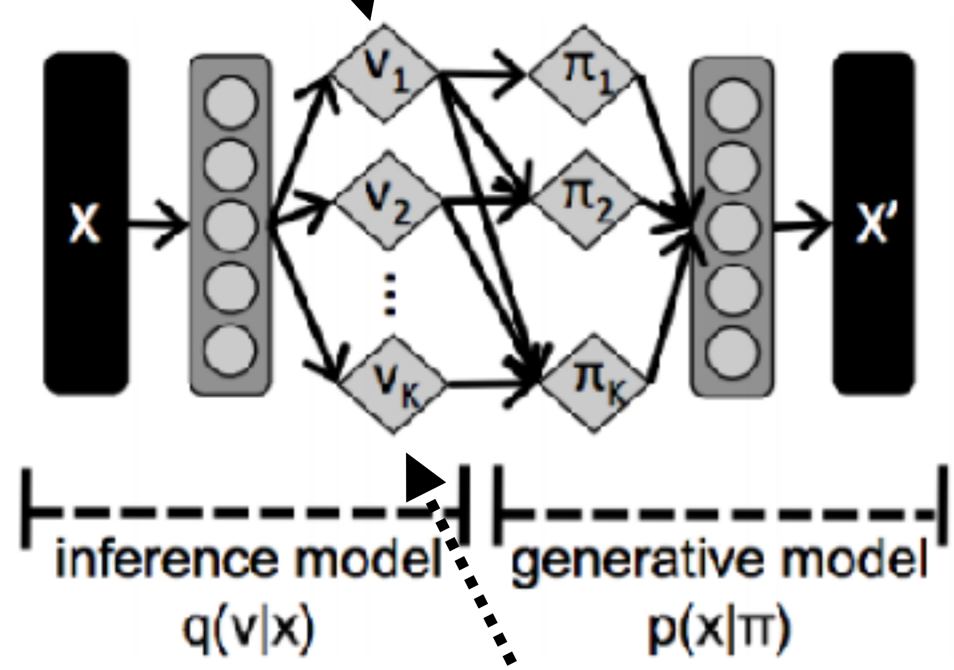
Brute Force Solution: Summing over all components results in a tractable ELBO but requires $O(K^S)$ decoder propagations.

MODEL #1: Stick-Breaking Variational Autoencoder

(Nalisnick & Smyth, 2017)



Kumaraswamy Samples



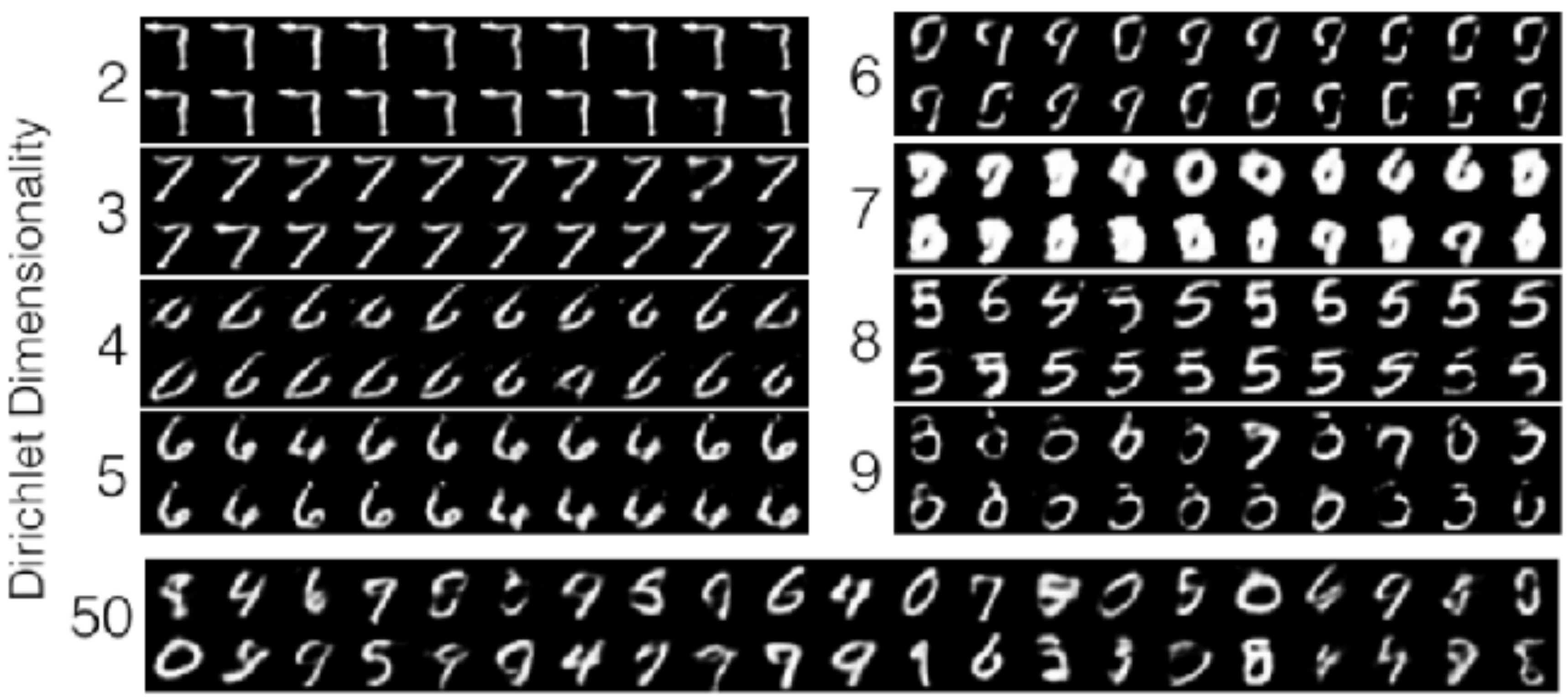
Truncated posterior;
not necessary but learns faster

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}, \phi; \mathbf{x}_i) = \frac{1}{S} \sum_{s=1}^{S} \log p_{\boldsymbol{\theta}}(\mathbf{x}_i | \boldsymbol{\pi}_{i,s}) - KL(q_{\boldsymbol{\phi}}(\boldsymbol{\pi}_i | \mathbf{x}_i) || p(\boldsymbol{\pi}_i; \boldsymbol{\alpha}_0))$$

Samples from Generative Model

(Nalisnick & Smyth, 2017)

Stick-Breaking VAE

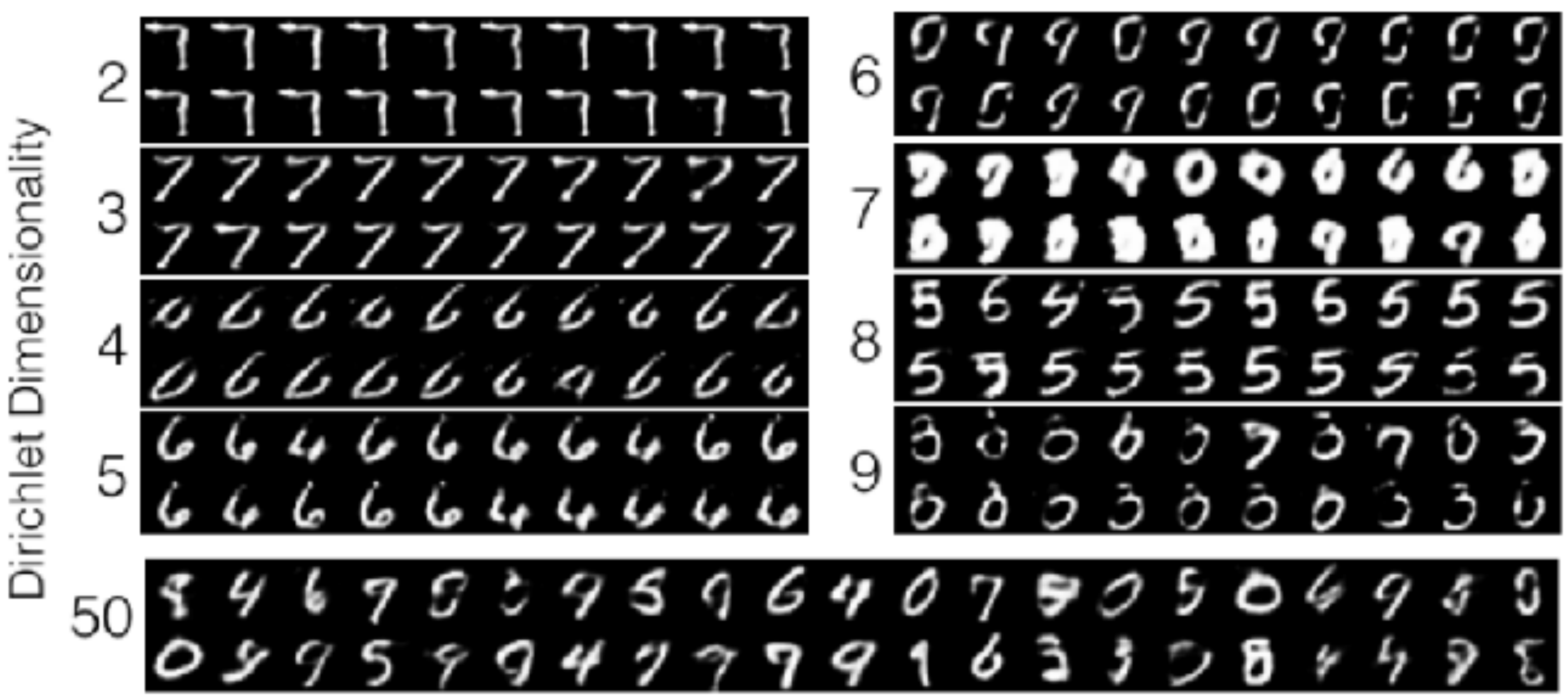


Truncation level of 50 dimensions, Beta(1,5) Prior

Samples from Generative Model

(Nalisnick & Smyth, 2017)

Stick-Breaking VAE



Truncation level of 50 dimensions, Beta(1,5) Prior

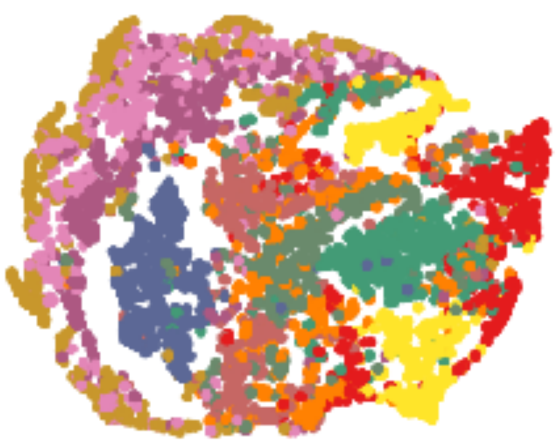
Gaussian VAE

50 dimensions, N(0,1) Prior

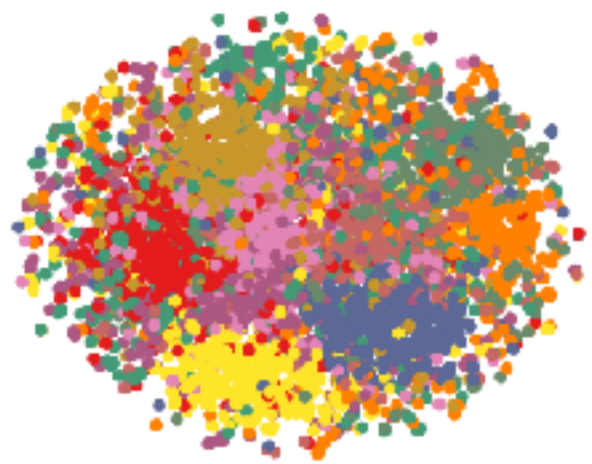
Quantitative Results for SB-VAE

(Nalisnick & Smyth, 2017)

Unsupervised



MNIST: Dirichlet
Latent Space (t-SNE)



MNIST: Gaussian
Latent Space (t-SNE)

	k=3	k=5	k=10
SB-VAE	9.34	8.65	8.90
Gauss-VAE	28.4	20.96	15.33
Raw Pixels	2.95	3.12	3.35

MNIST: kNN Classifier on Latent Space

	$-\log p(\mathbf{x}_i)$	
Model	**MNIST**	**MNIST+rot**
Gauss VAE	96.80	108.40
Kumar-SB VAE	98.01	112.33
Logit-SB VAE	99.48	114.09
Gamma-SB VAE	100.74	113.22

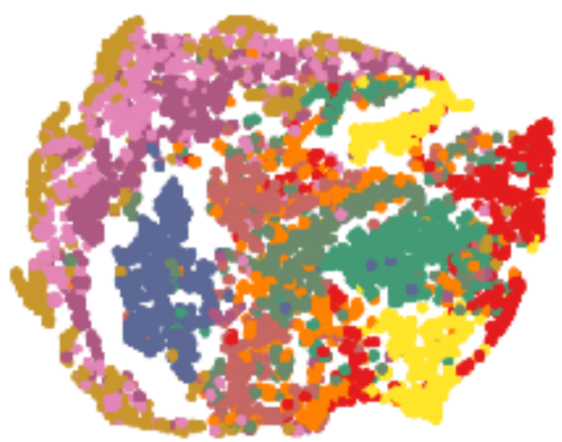
(Estimated) Marginal Likelihoods

Semi-
Supervised

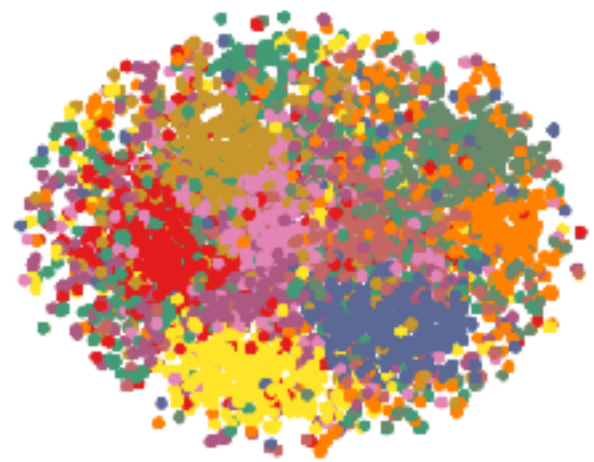
Quantitative Results for SB-VAE

(Nalisnick & Smyth, 2017)

Unsupervised



MNIST: Dirichlet
Latent Space (t-SNE)



MNIST: Gaussian
Latent Space (t-SNE)

	k=3	k=5	k=10
SB-VAE	9.34	8.65	8.90
Gauss-VAE	28.4	20.96	15.33
Raw Pixels	2.95	3.12	3.35

MNIST: kNN Classifier on Latent Space

	$-\log p(\mathbf{x}_i)$	
Model	**MNIST**	**MNIST+rot**
Gauss VAE	96.80	108.40
Kumar-SB VAE	98.01	112.33
Logit-SB VAE	99.48	114.09
Gamma-SB VAE	100.74	113.22

(Estimated) Marginal Likelihoods

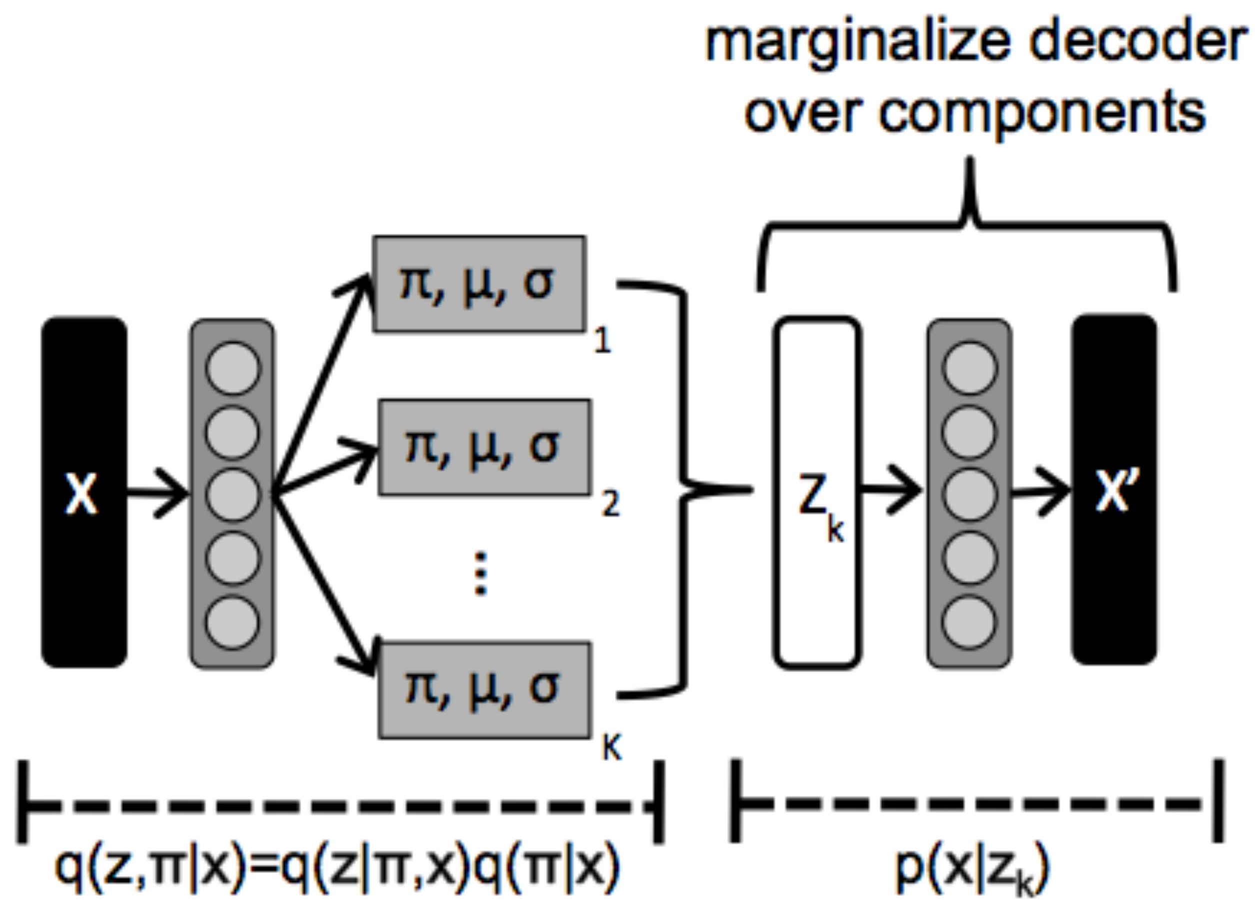
Semi-Supervised

Nonparametric version of (Kingma et al., NIPS 2014)'s M2 model

	MNIST (N=45,000)			**SVHN** (N=65,000)		
	10%	5%	1%	10%	5%	1%
SB-DGM	$4.86_{\pm.14}$	$5.29_{\pm.39}$	$\mathbf{7.34}_{\pm.47}$	$\mathbf{32.08}_{\pm4.00}$	$\mathbf{37.07}_{\pm5.22}$	$\mathbf{61.37}_{\pm3.60}$
Gauss-DGM	$\mathbf{3.95}_{\pm.15}$	$\mathbf{4.74}_{\pm.43}$	$11.55_{\pm2.28}$	$36.08_{\pm1.49}$	$48.75_{\pm1.47}$	$69.58_{\pm1.64}$
kNN	$6.13_{\pm.13}$	$7.66_{\pm.10}$	$15.27_{\pm.76}$	$64.81_{\pm.34}$	$68.94_{\pm.47}$	$76.64_{\pm.54}$

MODEL #2: Dirichlet Process Variational Autoencoder

(Nalisnick et al., 2016)



$$\mathcal{L}_{\mathrm{SGVB}} = \sum_k \mu_{\pi_k} [\frac{1}{S} \sum_s \log p_{\boldsymbol{\theta}}(\mathbf{x}_i|\hat{\mathbf{z}}_{i,k,s}) + \mathbb{E}_{q_k}[\log p(\mathbf{z}_i)]]$$

$$- \mathrm{KLD}[q(\boldsymbol{\pi}_k|\mathbf{x}_i)||p(\boldsymbol{\pi}_k)] - \frac{1}{S} \sum_s \log \sum_k \hat{\pi}_{i,k,s} q(\hat{\mathbf{z}}_{i,k,s}; \boldsymbol{\phi}_k)$$

Samples from Generative Model

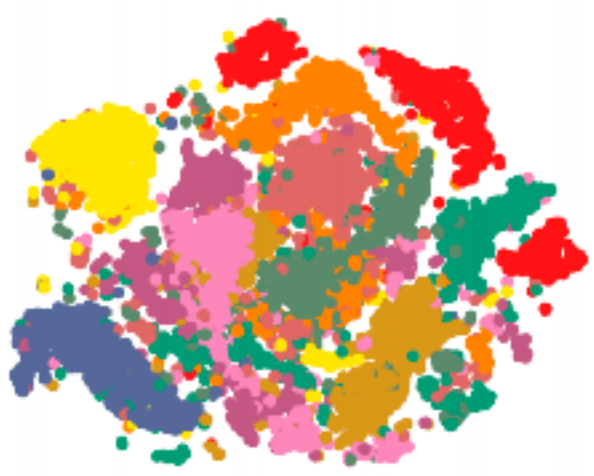
(Nalisnick et al., 2016)

MNIST Samples from
Component #1

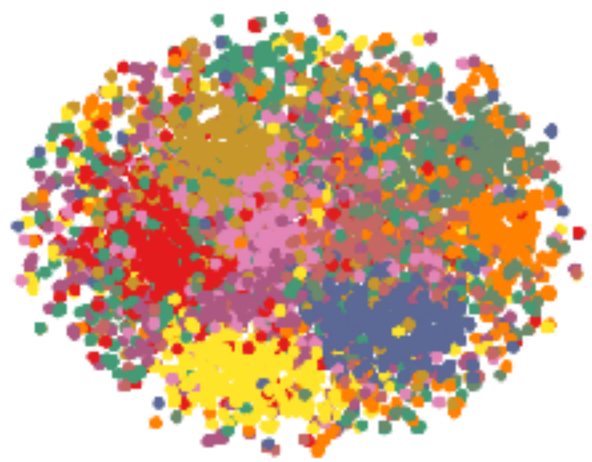
MNIST Samples from
Component #5

Quantitative Results for DP-VAE

(Nalisnick et al., 2016)



MNIST: Dirichlet Process
Latent Space (t-SNE)



MNIST: Gaussian
Latent Space (t-SNE)

	k=3	k=5	k=10
DLGMM	**9.14**	**8.38**	**8.42**
SB-VAE	9.34	8.65	8.90
Gauss-VAE	28.4	20.96	15.33

MNIST: kNN Classifier on Latent Space

	$-\log p_{\boldsymbol{\theta}}(\mathbf{x}_i)$ MNIST	$-\log p_{\boldsymbol{\theta}}(\mathbf{x}_i)$ OMNIGLOT
DLGMM (500d-3x25s)	**96.50**	123.50
DLDPMM (500d-17tx25s)	96.91	123.76
Gauss-VAE (500d-25s)	96.80	**119.18**
SB-VAE (500d-25t)	98.01	—

(Estimated) Marginal Likelihoods

Part 2: Objective Priors

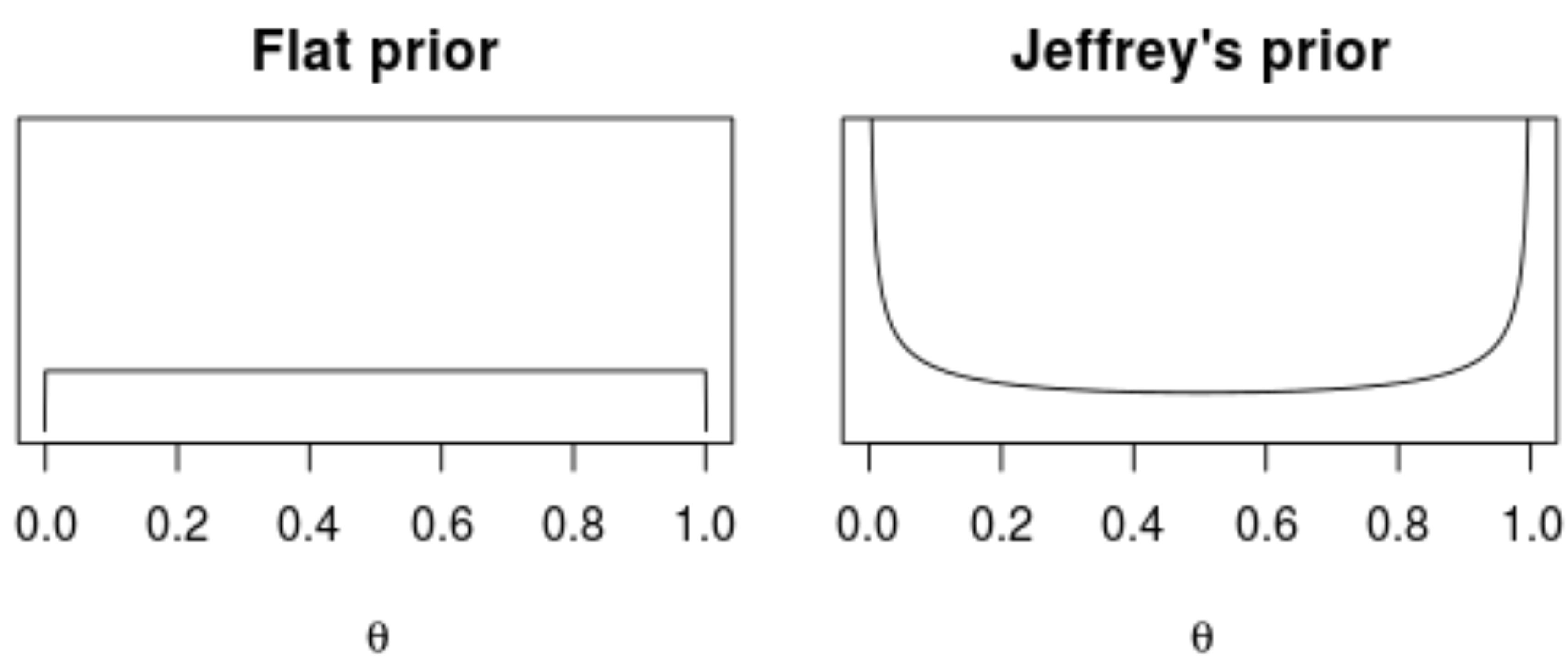
<u>Publications</u>

1. E. Nalisnick and P. Smyth. “Variational Reference Priors”. In submission to *Workshop Track, ICLR 2017*.

Objective Priors

Jeffreys Priors: Uninformative prior invariant to reparametrization. Represents a state of 'ignorance' in prior beliefs.

$$p^*(\boldsymbol{\theta}) \propto \sqrt{|\ \mathcal{F}(\boldsymbol{\theta})\ |}$$



Objective Priors

Reference Priors (Bernardo, 1979): Generalize the notion of an objective prior to the following definition:

$$p^*(\boldsymbol{\theta}) = \operatorname*{argmax}_{p(\boldsymbol{\theta})} I(\boldsymbol{\theta}, \mathcal{D})$$
$$= \operatorname*{argmax}_{p(\boldsymbol{\theta})} \int_{\mathcal{D}} p(\mathcal{D})\mathrm{KLD}[p(\boldsymbol{\theta}|\mathcal{D}) \;||\; p(\boldsymbol{\theta})]d\mathcal{D}.$$

They are also invariant to reparametrization, and equal to the Jeffreys prior in one-dimension. Hard to solve for analytically.

Variational Reference Priors

Can re-write the Reference prior objective as:

$$
\begin{aligned}
p^*(\boldsymbol{\theta}) &= \underset{p(\boldsymbol{\theta})}{\operatorname{argmax}} \int_{\mathcal{D}} p(\mathcal{D}) \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathcal{D}) \log \frac{p(\boldsymbol{\theta}|\mathcal{D})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta} d\mathcal{D} \\
&= \underset{p(\boldsymbol{\theta})}{\operatorname{argmax}} \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) \int_{\mathcal{D}} p(\mathcal{D}|\boldsymbol{\theta}) \log \frac{p(\mathcal{D}|\boldsymbol{\theta})}{p(\mathcal{D})} d\mathcal{D} d\boldsymbol{\theta} \\
&= \underset{p(\boldsymbol{\theta})}{\operatorname{argmax}} \mathbb{E}_{p(\boldsymbol{\theta})} KLD[p(\mathcal{D}|\boldsymbol{\theta}) \mid\mid p(\mathcal{D})].
\end{aligned}
$$

Variational Reference Priors

Can re-write the Reference prior objective as:

$$\begin{aligned} p^*(\boldsymbol{\theta}) &= \underset{p(\boldsymbol{\theta})}{\operatorname{argmax}} \int_{\mathcal{D}} p(\mathcal{D}) \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathcal{D}) \log \frac{p(\boldsymbol{\theta}|\mathcal{D})}{p(\boldsymbol{\theta})} d\boldsymbol{\theta} d\mathcal{D} \\ &= \underset{p(\boldsymbol{\theta})}{\operatorname{argmax}} \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}) \int_{\mathcal{D}} p(\mathcal{D}|\boldsymbol{\theta}) \log \frac{p(\mathcal{D}|\boldsymbol{\theta})}{p(\mathcal{D})} d\mathcal{D} d\boldsymbol{\theta} \\ &= \underset{p(\boldsymbol{\theta})}{\operatorname{argmax}} \mathbb{E}_{p(\boldsymbol{\theta})} KLD[p(\mathcal{D}|\boldsymbol{\theta}) \mid\mid p(\mathcal{D})]. \end{aligned}$$

Pick a functional family and optimize the parameters:

$$\boldsymbol{\lambda}^* = \underset{\boldsymbol{\lambda}}{\operatorname{argmax}} \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})} KLD[p(\mathcal{D}|\boldsymbol{\theta}) \mid\mid p(\mathcal{D})]$$

Variational Lowerbound

Marginal likelihood makes the objective intractable in most cases so we derive a lowerbound:

$$\mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}[\log p(\mathcal{D}|\boldsymbol{\theta})] - \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}\underline{[\log p(\mathcal{D})]}$$

Variational Lowerbound

Marginal likelihood makes the objective intractable in most cases so we derive a lowerbound:

$$\mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}[\log p(\mathcal{D}|\boldsymbol{\theta})] - \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}\underline{[\log p(\mathcal{D})]}$$

Using the Renyi bound (Li and Turner, 2016):

$$\log p(\mathcal{D}) \leq \frac{1}{1-\alpha} \log \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})} \left[p(\mathcal{D}|\boldsymbol{\theta})^{1-\alpha}\right] \quad \text{for} \quad \alpha < 0$$

Variational Lowerbound

Marginal likelihood makes the objective intractable in most cases so we derive a lowerbound:

$$\mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}[\log p(\mathcal{D}|\boldsymbol{\theta})] - \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}\underline{[\log p(\mathcal{D})]}$$

Using the Renyi bound (Li and Turner, 2016):

$$\log p(\mathcal{D}) \leq \frac{1}{1-\alpha} \log \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\left[p(\mathcal{D}|\boldsymbol{\theta})^{1-\alpha}\right] \quad \text{for} \quad \alpha < 0$$

Final variational objective:

$$\boxed{I(\boldsymbol{\theta}, \mathcal{D}) \geq \mathbb{E}_{p_{\boldsymbol{\lambda}}(\boldsymbol{\theta})}\mathbb{E}_{p(\mathcal{D}|\boldsymbol{\theta})}[\log p(\mathcal{D}|\boldsymbol{\theta}) - \max_{s} \log p(\mathcal{D}|\hat{\boldsymbol{\theta}}_s)]}$$

Types of Approximations

1. **Parametric:** Pick some known distribution with the proper support, preferably that can be sampled via a non-centered parametrization.

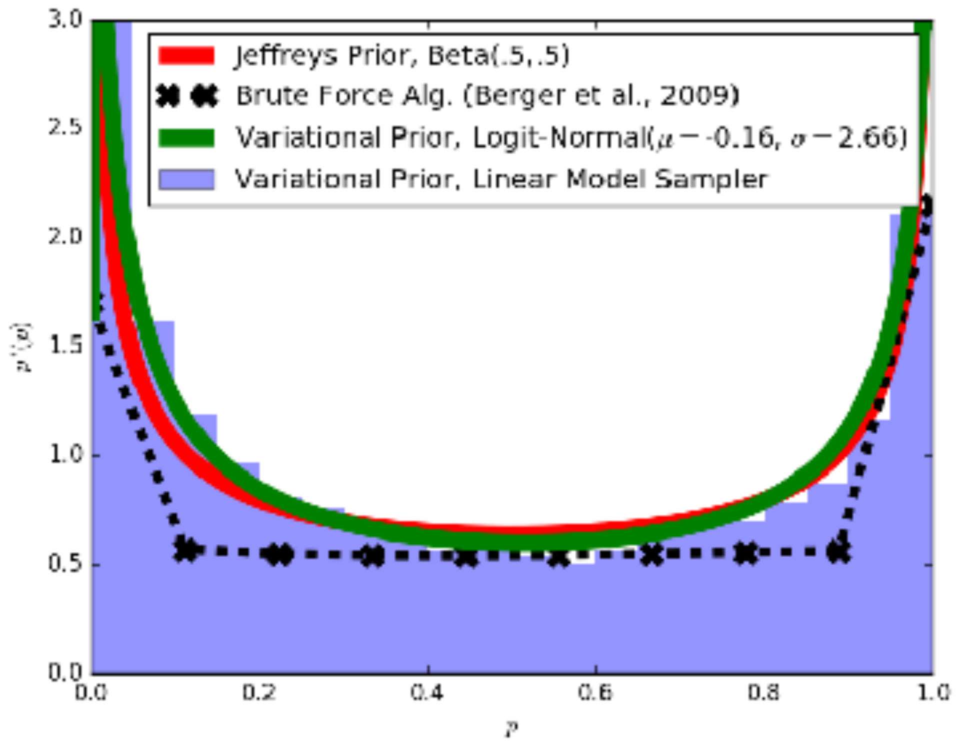
Types of Approximations

1. **Parametric:** Pick some known distribution with the proper support, preferably that can be sampled via a non-centered parametrization.

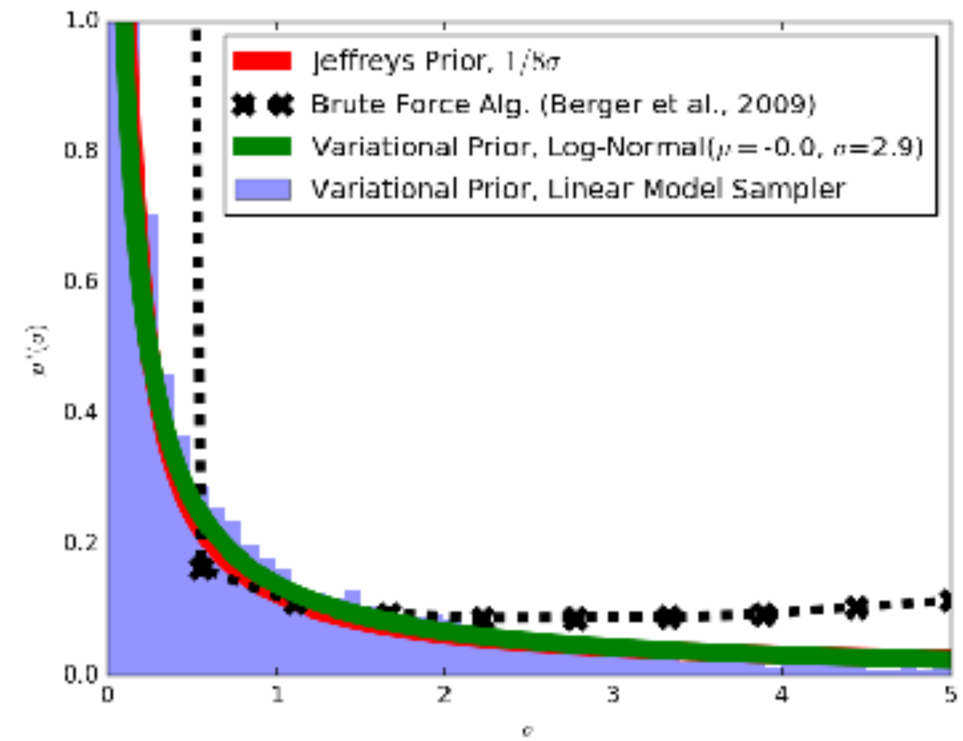
2. **Implicit Prior:** Notice that the variational objective doesn't require the prior be evaluated, just sampled from. Thus we can use an arbitrary transformation.

$$\hat{\boldsymbol{\theta}} = f(\boldsymbol{\lambda}, \hat{\boldsymbol{\epsilon}}) \text{ where } \boldsymbol{\epsilon} \sim p_0$$

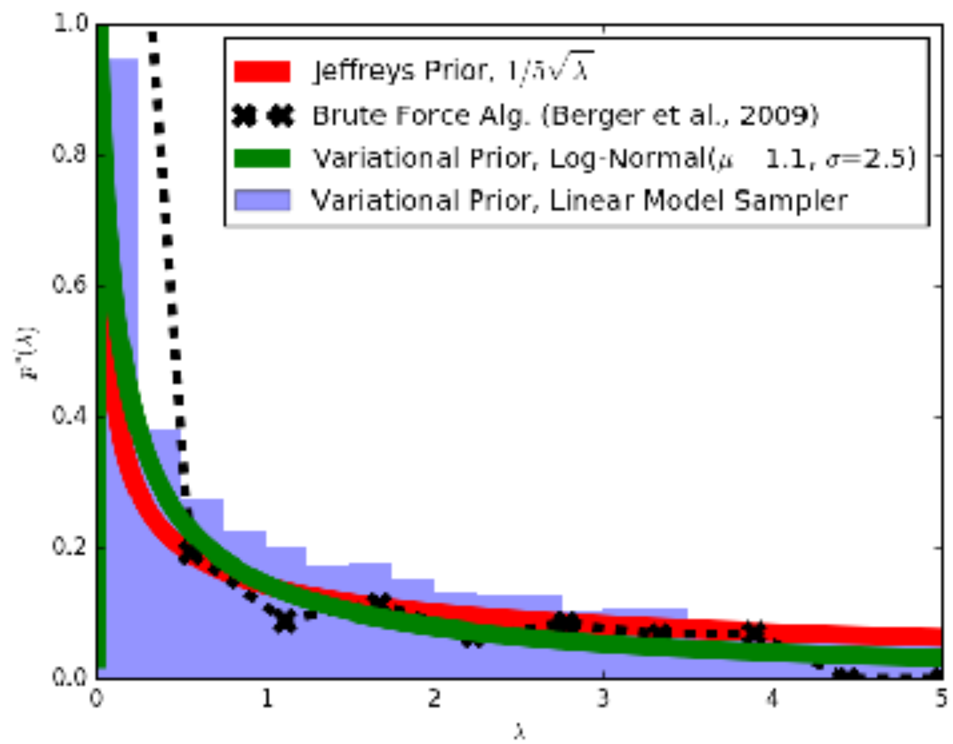
Recovering Jeffreys Priors



Bernoulli Parameter

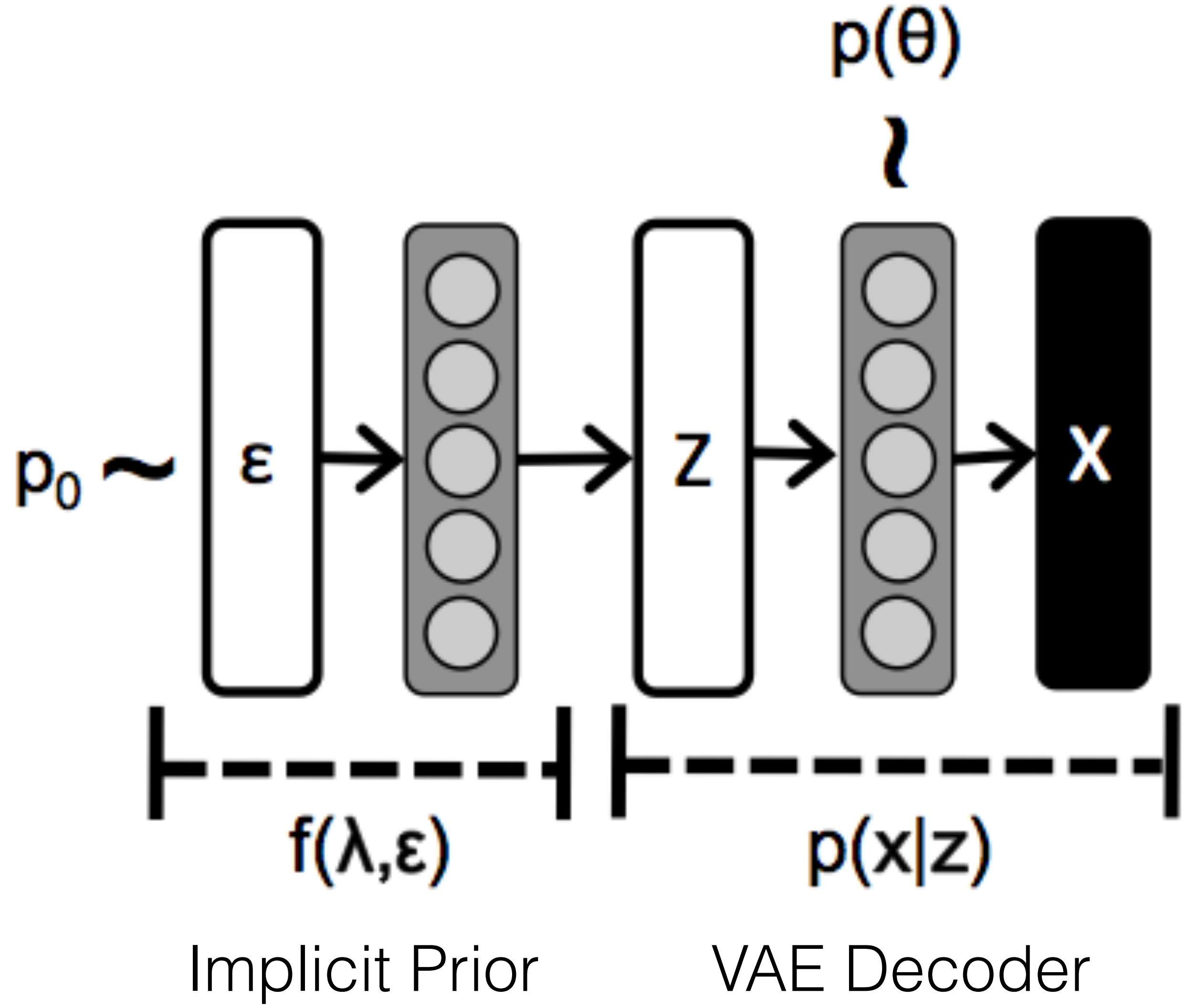


Gaussian Scale Parameter



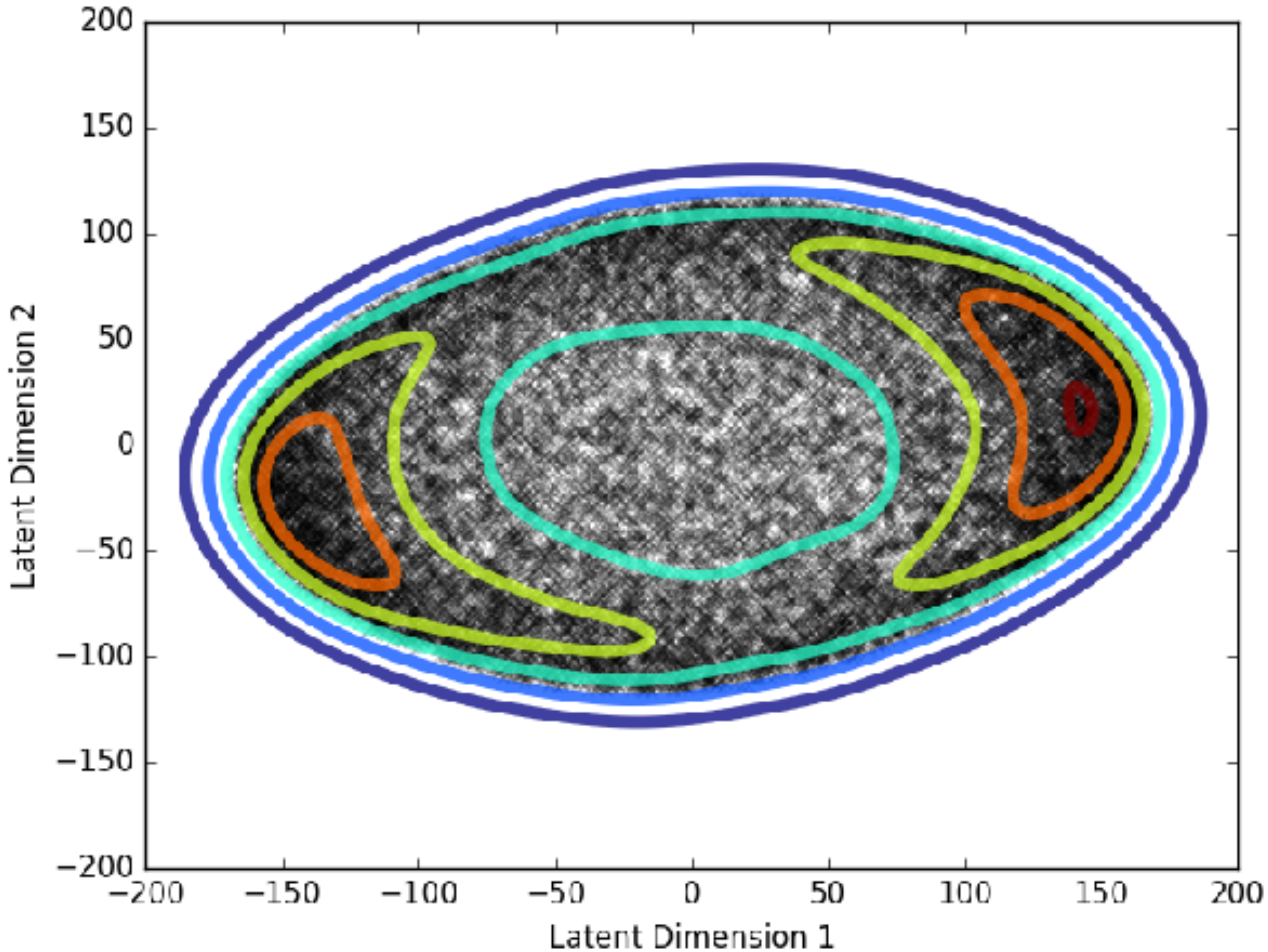
Poisson Parameter

Finding the VAE's Reference Prior



$$\mathcal{J}_{\text{RP-VAE}} = \log p(\hat{\mathbf{x}}_0 | \mathbf{z} = f(\boldsymbol{\lambda}, \hat{\epsilon_0})) - \max_s \log p(\hat{\mathbf{x}}_0 | \mathbf{z} = f(\boldsymbol{\lambda}, \hat{\boldsymbol{\epsilon}}_s))$$

Finding the VAE's Reference Prior



Conclusions

1 **Superior Latent Spaces with NP priors:** Seem to preserve class structure well, resulting in better discriminative properties. Their dynamic capacity naturally encodes factors of variation.

Conclusions

1. **Superior Latent Spaces with NP priors:** Seem to preserve class structure well, resulting in better discriminative properties. Their dynamic capacity naturally encodes factors of variation.

2. **Extra Computation:** Extra computation and the imposed model structure may be undesirable.

Conclusions

1. **Superior Latent Spaces with NP priors:** Seem to preserve class structure well, resulting in better discriminative properties. Their dynamic capacity naturally encodes factors of variation.

2. **Extra Computation:** Extra computation and the imposed model structure may be undesirable.

3. **Subjectivity in Priors:** Are arbitrary priors affecting our posteriors? Maybe. More analysis needed.

Future Work

1. **Stick-Breaking for Adaptive Computation:** Use stick-breaking to determine the number of computation steps, such as RNN recursions. Preliminary results suggest SOTA on language modeling.

Future Work

1 **Stick-Breaking for Adaptive Computation:** Use stick-breaking to determine the number of computation steps, such as RNN recursions. Preliminary results suggest SOTA on language modeling.

2 **Learning Robust Priors:** Can we *minimize* the mutual information objective to learn robust priors? Reasons to believe this would (approximately) encode Dropout as a Bayesian prior.

Papers and code at: `ics.uci.edu/~enalisni/`

ERIC NALISNICK

PhD Candidate
Computer Science
University of California, Irvine

Room 4228
Donald Bren Hall
Irvine, CA 92617
enalisni [at] uci [dot] edu

about resume code

I am a 4th year PhD candidate in the Computer Science Department at the University of California, Irvine. Padhraic Smyth is my advisor. My research interests reside in both the theory and application of Bayesian latent variable models, including neural networks. I've done research internships at Amazon, Microsoft Research / Bing, and Twitter Cortex. I graduated from Lehigh University (Bethlehem, PA) where I worked with Henry Baird.

PUBLICATIONS

PREPRINTS / WORKING PAPERS

Eric Nalisnick and Sachin Ravi. **Infinite Dimensional Word Embeddings.**

Eric Nalisnick and Padhraic Smyth. **Variational Reference Priors.**

CONFERENCE PUBLICATIONS

Eric Nalisnick and Padhraic Smyth. **Stick-Breaking Variational Autoencoders.** *International Conference on Learning Representations* (ICLR), Toulon, France, April 24-26 2017. [Code] [Supplemental Materials]

Eric Nalisnick, Bhaskar Mitra, Nick Craswell, and Rich Caruana. **Improving Document Ranking with Dual Word Embeddings.** *In Proceedings of the 25th World Wide Web Conference* (WWW), Short Paper, Montreal, Canada, April 11-15 2016.

Eric T. Nalisnick and Henry S. Baird. **Character-to-Character Sentiment Analysis in Shakespeare's Plays.** *In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (ACL), Short Paper, pages 479-83, Sofia, Bulgaria, August 4-9 2013. [Shakespeare Sentiment Explorer]

Eric T. Nalisnick and Henry S. Baird. **Extracting Sentiment Networks from Shakespeare's Plays.** *In Proceedings of the 12th International Conference on Document Analysis and Recognition* (ICDAR), pages 758-762, Washington, USA, August 25-28, 2013.

Thank you. Questions?