

THE PROBLEM

Variational Bayesian inference frequently relies on Monte Carlo (MC) integration when intractable expectations appear in the evidence lower bound (ELBO). Variational inference evaluates these expectations wrt the proposed posterior approximations. In many cases, such as in Variational Auto-Encoders, we maximize the ELBO via gradient methods, however, to do so in conjunction MC integration requires that samples from the posterior approximations are differentiable wrt their respective parameters. Closed-form “reparameterization tricks” unfortunately only exist for a subset of distribution families (e.g. Gaussian).

INVERSE CDF SAMPLING

For continuous univariate distribution, inverse CDF sampling allows us to reparameterize a sample from a Uniform distribution to the target distribution:

$$\epsilon \sim \text{Uniform}(0, 1)$$

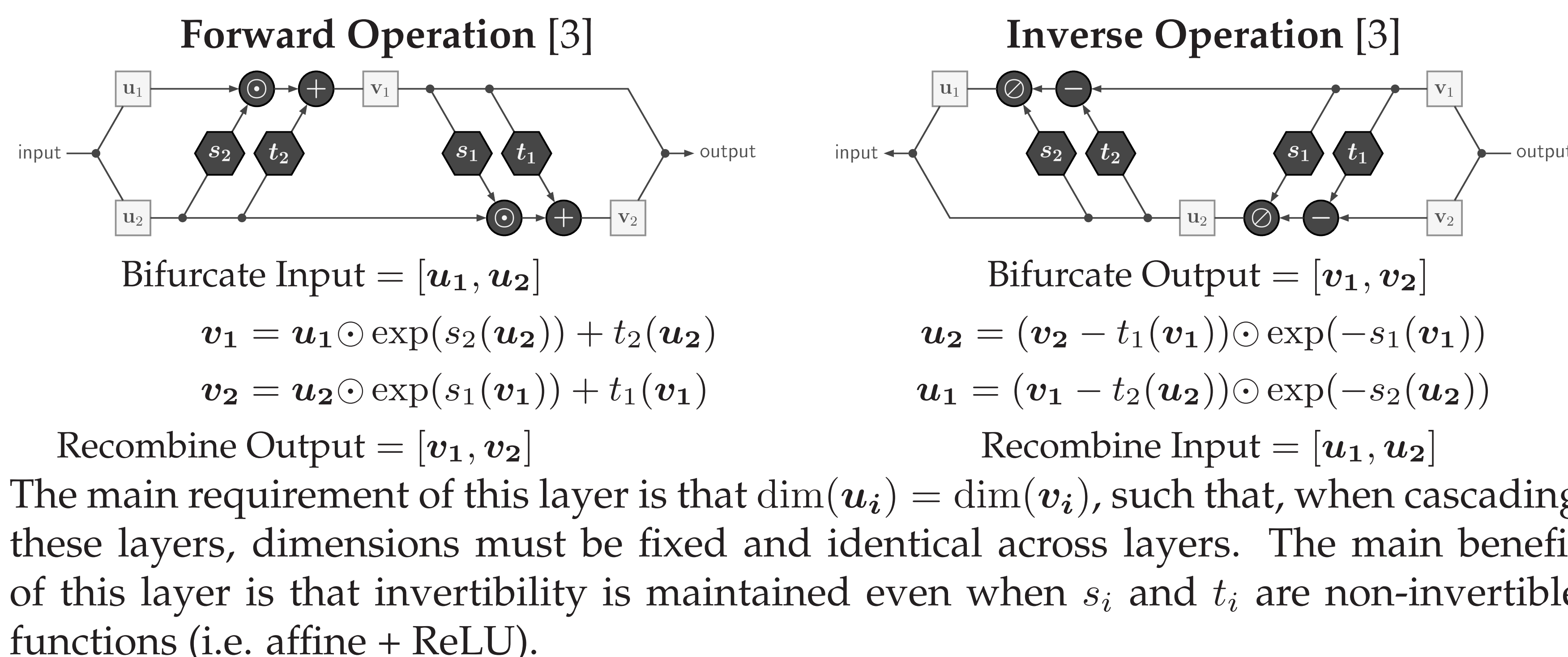
$$z = F^{-1}(\epsilon; \theta)$$

Unfortunately, not all continuous univariate distributions have a closed form quantile function, F^{-1} . In these cases, $\frac{\partial z}{\partial \theta}$ does not exist, prohibiting gradient methods.

OUR CONTRIBUTIONS

We leverage invertible neural network (INN) layers [1, 2] to approximate continuous univariate CDF's, allowing for inverse CDF sampling while ensuring the existence of $\frac{\partial z}{\partial \theta}$. Because the INN blocks are analytically invertible, our approximation maintains *monotonicity* and *one-to-oneness* ensuring: $z = \hat{F}^{-1}(\hat{F}(z; \theta); \theta)$.

INVERTIBLE NEURAL NETWORK: AFFINE COUPLING LAYER



INN MODIFICATIONS

Replacing $\exp(\cdot)$ with $b(\cdot)$, improves numerical stability ($b < e$) and maintains invertibility. Let $f_i(x) = \text{elu}(W_{f_i} \cdot x + b_{f_i})$, where $f \in \{s, t\}$, and W_{f_i} and b_{f_i} are the outputs of multi-layer perceptrons operating on the distribution parameters θ . Let loss be $|F(z_i; \theta_n) - \hat{F}(z_i; \theta_n)|$.

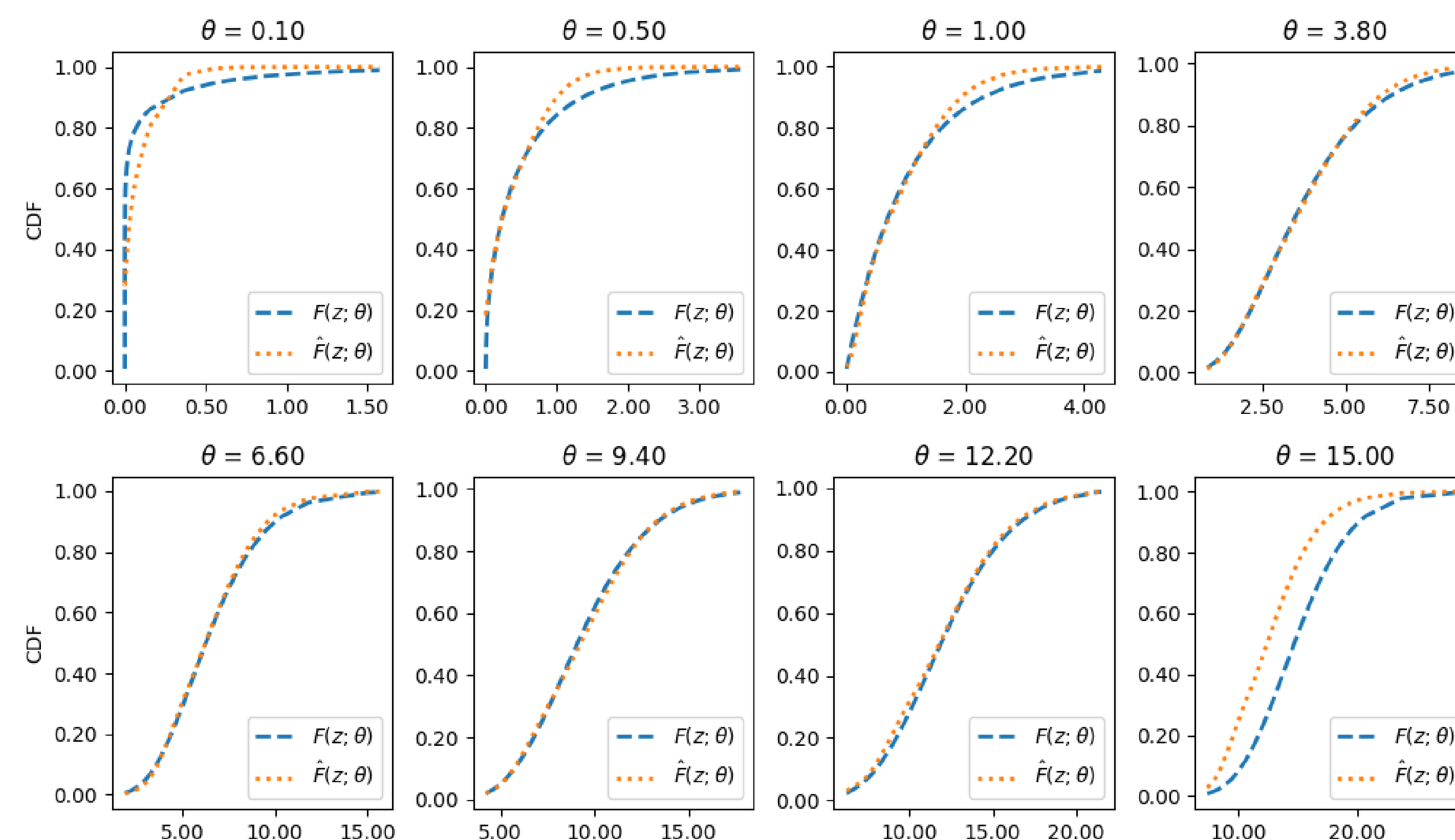
GAMMA CDF TRAINING BATCH

```

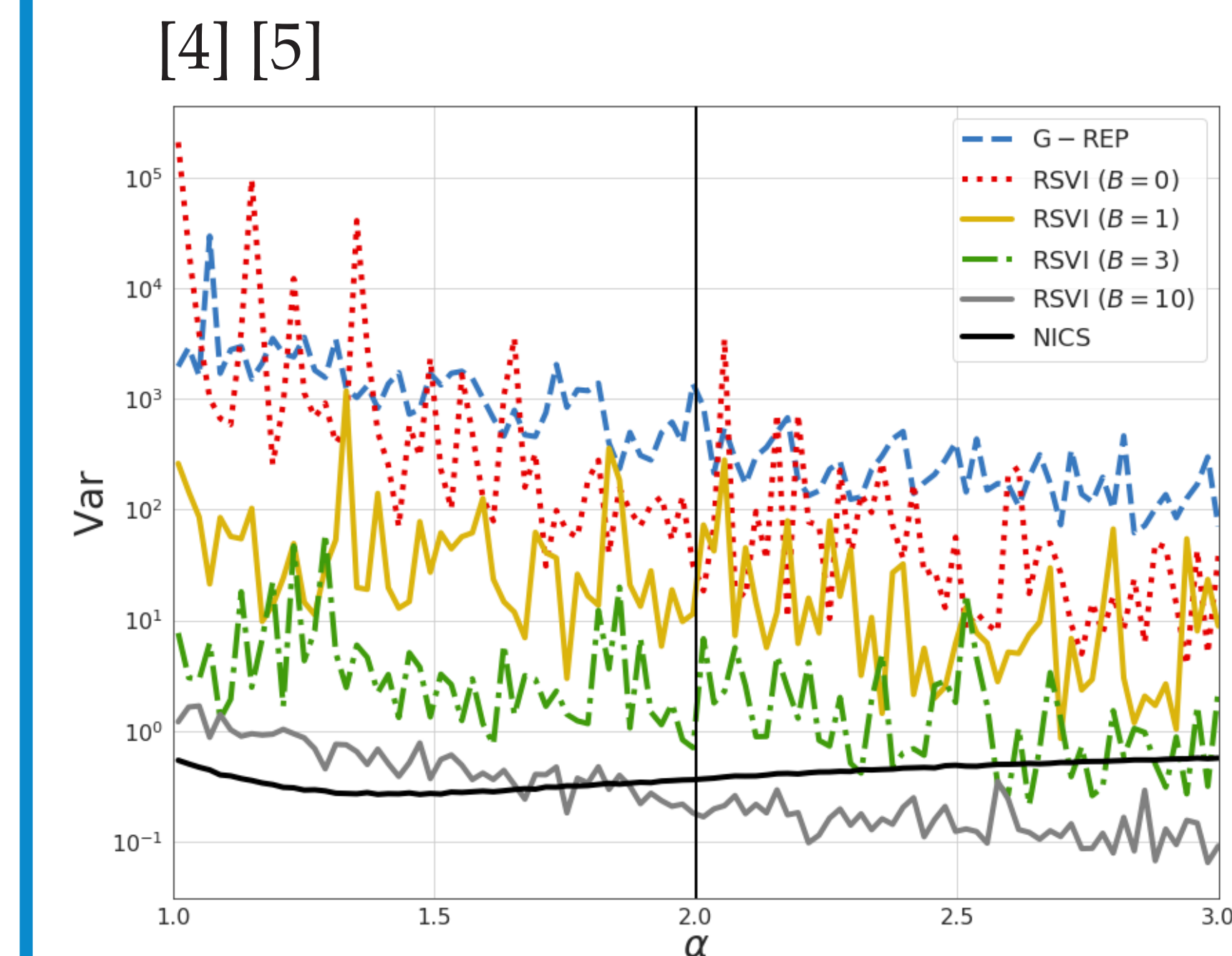
Initialize batch  $\mathcal{B} = \{\emptyset\}$ 
for  $n \in \{1, \dots, N_{\text{batch}}\}$  do
  Sample:  $\theta_n \stackrel{iid}{\sim} p(\theta)$ 
  for  $i \in \{1, \dots, N_z\}$  do
    Sample:  $z_i \stackrel{iid}{\sim} \text{Gamma}(\theta_n, 1)$ 
    Append:  $\mathcal{B} := \mathcal{B} \cup (z_i, \theta_n, F(z_i; \theta_n))$ 
  end for
end for

```

CDF APPROXIMATION FOR GAMMA($\theta, 1$) DISTRIBUTION: $\theta \in (0, 15]$



DIRICHLET APPROXIMATION



RESULTS

[6] [7]
Perplexity Scores

| | Softmax | NGS |
|---------|---------|---------|
| NVLDA | 1099.96 | 1103.77 |
| ProDLDA | 1161.69 | 1099.94 |

REFERENCES & CODE

- [1] L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using real NVP,” *CoRR*, vol. abs/1605.08803, 2016. [Online]. Available: <http://arxiv.org/abs/1605.08803>
- [2] L. Ardizzone, J. Kruse, C. Rother, and U. K  the, “Analyzing inverse problems with invertible neural networks,” in *International Conference on Learning Representations*, 2019.
- [3] V. L. L. Heidelberg, “Analyzing inverse problems with invertible neural networks,” 2018. [Online]. Available: <https://hci.iwr.uni-heidelberg.de/vislearn/inverse-problems>
- [4] C. A. Naesseth, F. J. R. Ruiz, S. W. Linderman, and D. M. Blei, “Reparameterization gradients through acceptance-rejection sampling algorithms,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.
- [5] B. Lab, “Ars reparameterization,” 2017. [Online]. Available: <https://github.com/blei-lab/ars-reparameterization/>
- [6] C. S. Akash Srivastava, “Autoencoding variational inference for topic models,” in *International Conference on Learning Representations*, 2017.
- [7] A. Srivastava, “Autoencoding variational inference for topic models,” 2017. [Online]. Available: https://github.com/akashgit/autoencoding_vi_for_topic_models

Source code available at: github.com/astirn/neural-inverse-cdf-sampling