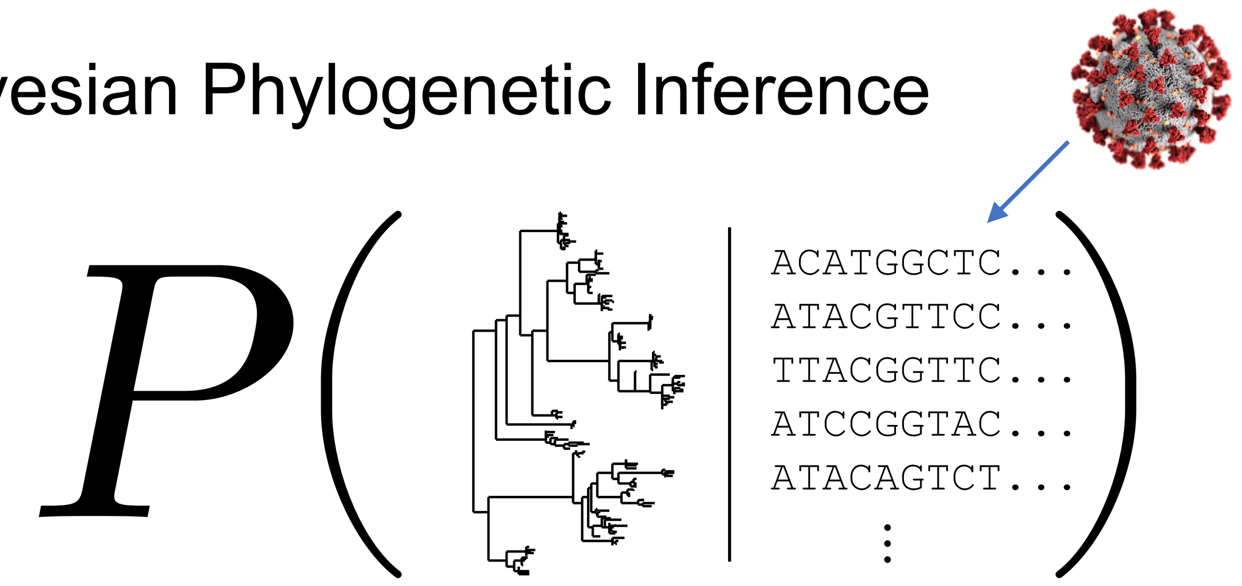
Improved Variational Bayesian Phylogenetic Inference with Normalizing Flows

Cheng Zhang, Peking University
Email: chengzhang@math.pku.edu.cn

Bayesian Phylogenetic Inference



Bayesian approaches for reconstructing the evolution history (e.g., phylogenetic trees) from molecular sequence data (e.g., DNA, RNA or protein) is to estimate the following phylogenetic posterior

$$p(\tau, q \mid Y) \propto p(Y \mid \tau, q) \cdot p(\tau, q)$$

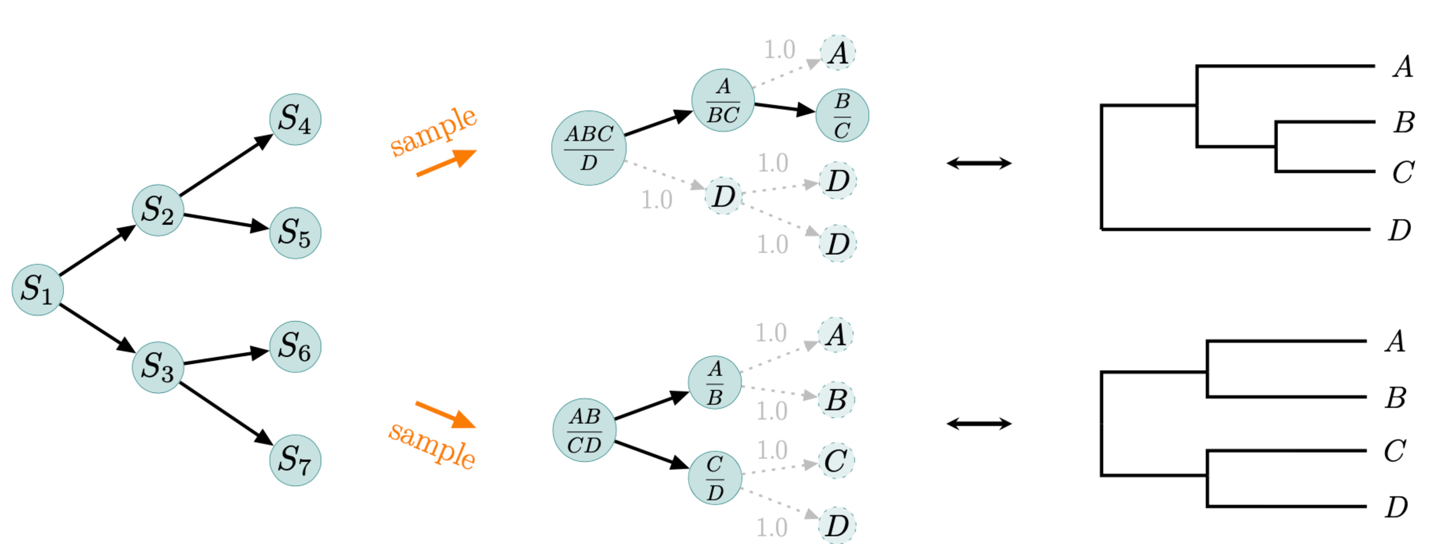
Unfortunately, traditional random walk MCMC methods do not scale.

Variational Bayesian Phylogenetic Inference (VBPI) is a recently proposed general variational framework for Bayesian phylogenetic inference that outperforms MCMC methods with greatly enhanced computation efficiency. However, the diagonal lognormal branch length approximation might be too simple for real data posteriors.

Based on carefully designed ***permutation equivariant normalizing flows*** for the non-Euclidean branch length space across tree topologies, we propose **VBPI-NF** as a first step to empower phylogenetic posterior estimation with deep learning techniques that

- Significantly improve the branch length approximations.
- Can provide additional amortization benefit.

Subsplit Bayesian Networks



Tree probability estimation

- *Rooted Tree*:

$$p_{\mathrm{sbn}}(T=\tau) = p(S_1 = s_1)\prod_{i>1} p(S_i = s_i \mid S_{\pi_i} = s_{\pi_i})$$

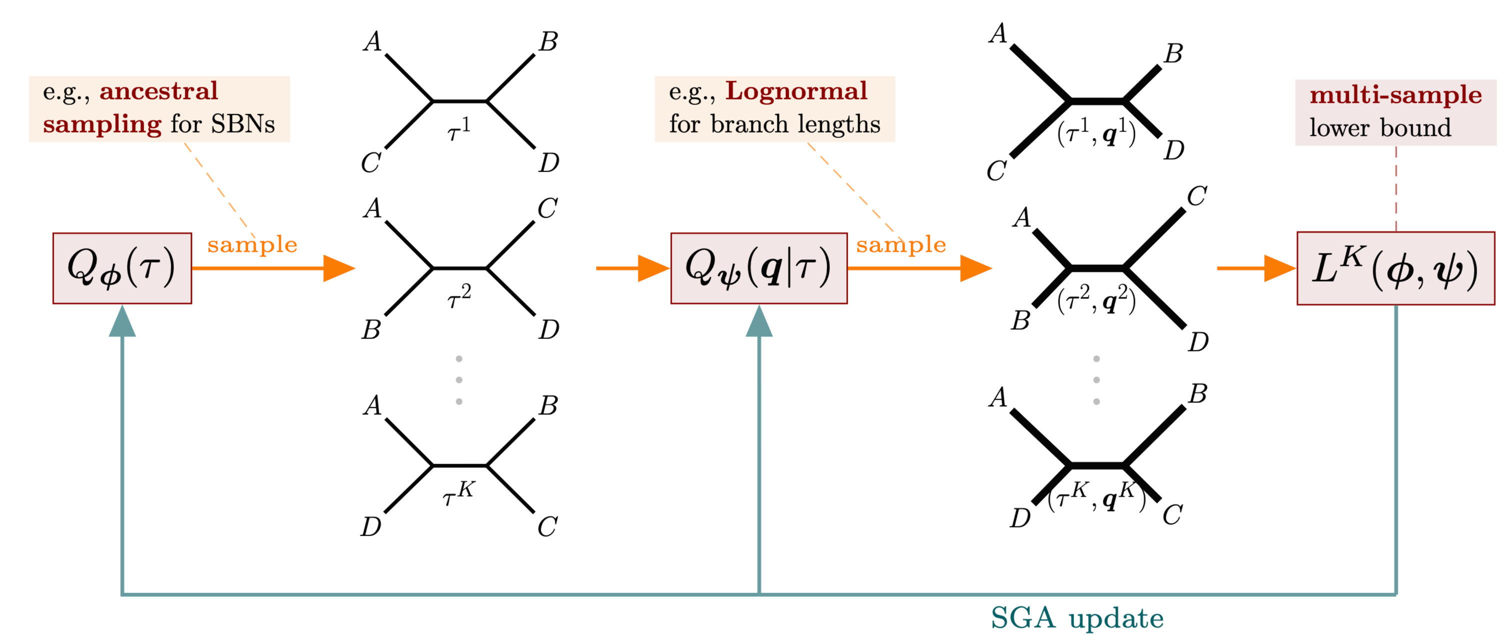
- *Unrooted Tree*:

$$p_{\mathrm{sbn}}(T^{\mathrm{u}}=\tau) = \sum_{s_1 \sim \tau} p(S_1 = s_1)\prod_{i>1} p(S_i = s_i \mid S_{\pi_i} = s_{\pi_i})$$

Tree sampling

- *Rooted Tree*: *ancestral sampling*
- *Unrooted Tree*: *ancestral sampling + root deletion*

Variational Bayesian Phylogenetic Inference



Training Objective. A multi-sample lower bound is used that facilitates exploration in the tree space

$$L^K(\phi,\psi) = \mathbb{E}_{Q_{\phi,\psi}(\tau^{1:K}, q^{1:K})} \log\left(\frac{1}{K}\sum_{i=1}^{K}\frac{p(Y\mid\tau^i,q^i)p(\tau^i,q^i)}{Q_\phi(\tau^i)Q_\psi(q^i\mid\tau^i)}\right) \le \log p(Y)$$

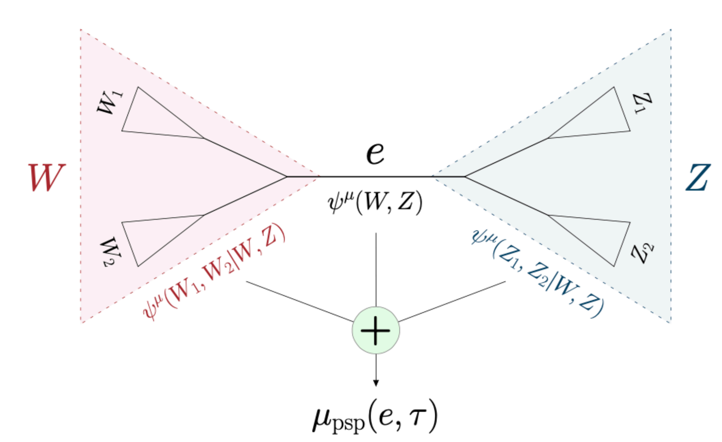
Branch Length Approximation. Currently, VBPI still uses simple diagonal Lognormal distribution for branch length approximation

$$Q_\psi(q\mid\tau) = \prod_{e\in E(\tau)} p^{\mathrm{Lognormal}}\left(q_e \mid \mu(e,\tau), \sigma(e,\tau)\right)$$

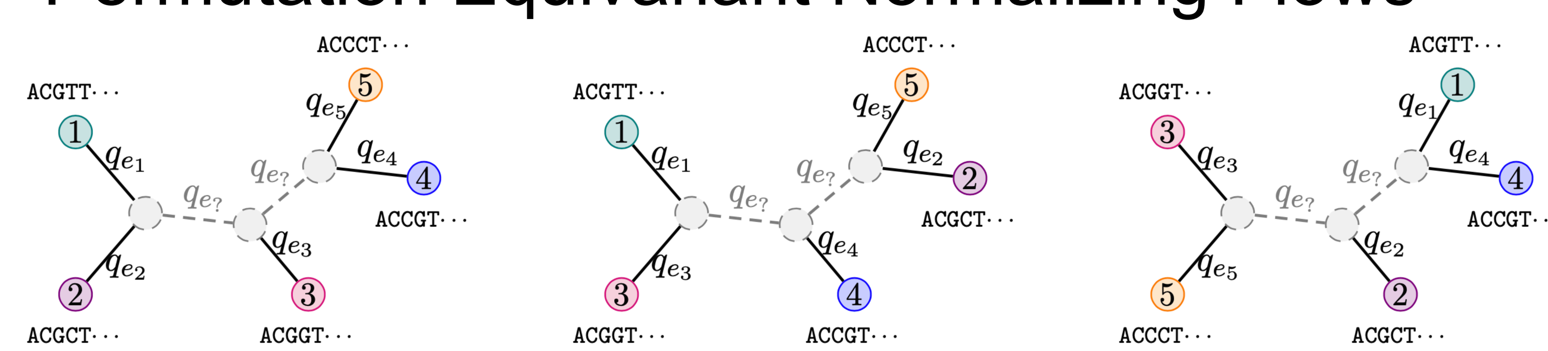
Structured Amortization via primary subsplit pairs (PSP).

$$\mu_{\mathrm{psp}}(e,\tau) = \psi^{\mu}_{e/\tau} + \sum_{s\in e/\!/\tau}\psi^{\mu}_{s}$$

$$\sigma_{\mathrm{psp}}(e,\tau) = \psi^{\sigma}_{e/\tau} + \sum_{s\in e/\!/\tau}\psi^{\sigma}_{s}$$



Permutation Equivariant Normalizing Flows



Inconsistent alignments of branch length vectors across tree topologies

Permutation Equivariant Planar Flows

$$z_e = \tilde{q}_e + \gamma_e a\left(\sum_{e'\in E(\tau)} w'_e \tilde{q}'_e + b\right), \quad e \in E(\tau)$$

Permutation Equivariant RealNVP

$$z_e = \tilde{q}_e,\ e\in S^c. \quad z_e = \tilde{q}_e \exp(\alpha_e(\tilde{q}_{S^c})) + \beta_e(\tilde{q}_{S^c}),\ e \in S$$

where

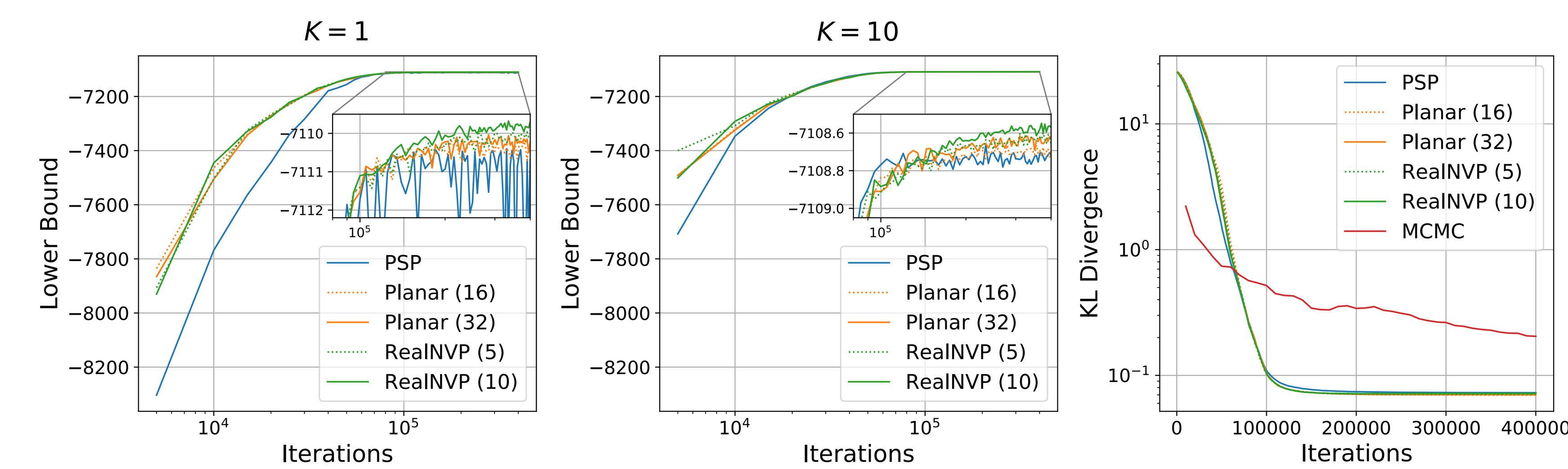
$$\begin{bmatrix}\alpha_e(\tilde{q}_{S^c})\\ \beta_e(\tilde{q}_{S^c})\end{bmatrix} = \begin{bmatrix}(w_e^{\alpha})^T\\ (w_e^{\beta})^T\end{bmatrix}\rho\left(\sum_{e'\in S^c}\tilde{q}_{e'}w_{e'} + b\right) + \begin{bmatrix}b_e^{\alpha}\\ b_e^{\beta}\end{bmatrix}$$

Structured amortization via PSP is used for all parameters.

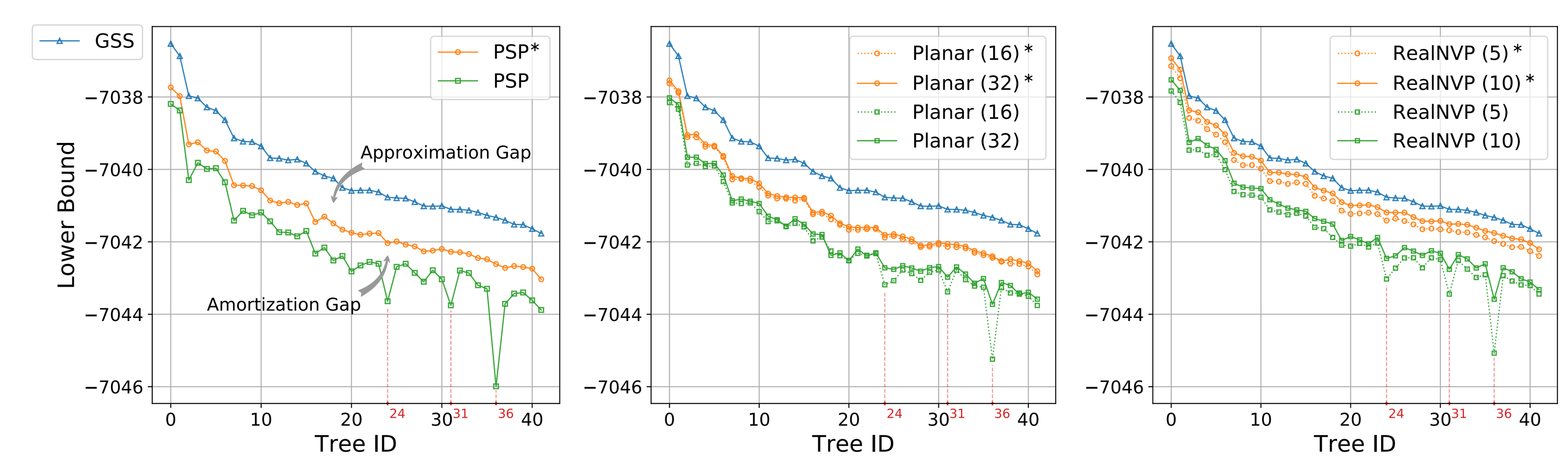
Experiments

	Data set	DS1	DS2	DS3	DS4	DS5	DS6	DS7	DS8
	# Taxa	27	29	36	41	50	50	59	64
	# Sites	1949	2520	1812	1137	378	1133	1824	1008
LB (K=1)	PSP	-7111.23(1.04)	-26369.63(0.69)	-33736.60(0.33)	-13332.37(0.54)	-8218.35(0.20)	-6729.27(0.50)	-37335.15(0.11)	-8655.48(0.38)
	Planar(16)	-7110.33(0.16)	-26368.80(0.27)	-33736.14(0.14)	-13331.92(0.11)	-8217.98(0.13)	-6728.89(0.18)	-37334.78(0.11)	-8655.15(0.17)
	Planar(32)	-7110.22(0.17)	-26368.69(0.23)	-33736.02(0.21)	-13331.73(0.12)	-8217.90(0.14)	-6728.68(0.19)	-37334.60(0.12)	-8654.97(0.16)
	RealNVP(5)	-7110.12(0.13)	-26368.75(0.24)	-33735.86(0.10)	-13331.71(0.11)	-8217.80(0.14)	-6728.54(0.15)	-37334.44(0.11)	-8654.62(0.13)
	RealNVP(10)	**-7109.80(0.11)**	**-26368.59(0.23)**	**-33735.81(0.12)**	**-13331.39(0.08)**	**-8217.56(0.12)**	**-6728.04(0.14)**	**-37333.94(0.09)**	**-8654.02(0.12)**
LB (K=10)	PSP	-7108.73(0.02)	-26367.88(0.02)	-33735.29(0.02)	-13330.34(0.03)	-8215.57(0.04)	-6725.48(0.04)	-37332.69(0.03)	-8651.88(0.04)
	Planar(16)	-7108.70(0.02)	-26367.80(0.01)	-33735.21(0.01)	-13330.28(0.02)	-8215.44(0.04)	-6725.42(0.04)	-37332.50(0.03)	-8651.80(0.04)
	Planar(32)	-7108.64(0.02)	-26367.77(0.01)	-33735.17(0.01)	-13330.22(0.02)	-8215.37(0.03)	-6725.32(0.04)	-37332.43(0.03)	-8651.72(0.04)
	RealNVP(5)	-7108.63(0.02)	-26367.77(0.01)	-33735.18(0.01)	-13330.22(0.02)	-8215.36(0.03)	-6725.33(0.04)	-37332.42(0.03)	-8651.62(0.04)
	RealNVP(10)	**-7108.58(0.02)**	**-26367.75(0.01)**	**-33735.16(0.01)**	**-13330.16(0.02)**	**-8215.29(0.03)**	**-6725.18(0.04)**	**-37332.30(0.02)**	**-8651.41(0.03)**
ML	PSP	-7108.39(0.18)	-26367.71(0.08)	-33735.09(0.10)	-13329.93(0.21)	-8214.44(0.48)	-6724.13(0.48)	-37331.92(0.32)	-8650.12(0.58)
	Planar(16)	-7108.39(0.15)	-26367.70(0.07)	-33735.09(0.07)	-13329.93(0.17)	-8214.49(0.42)	-6724.25(0.45)	-37331.91(0.26)	-8650.42(0.52)
	Planar(32)	-7108.40(0.14)	-26367.70(0.06)	**-33735.09(0.05)**	-13329.93(0.16)	-8214.50(0.38)	-6724.19(0.44)	-37331.93(0.23)	-8650.40(0.50)
	RealNVP(5)	-7108.40(0.14)	**-26367.71(0.04)**	-33735.09(0.06)	-13329.92(0.16)	-8214.50(0.38)	-6724.28(0.39)	**-37331.92(0.22)**	-8650.46(0.44)
	RealNVP(10)	**-7108.39(0.11)**	**-26367.71(0.04)**	**-33735.09(0.05)**	**-13329.92(0.13)**	-8214.51(0.36)	**-6724.25(0.37)**	**-37331.90(0.22)**	**-8650.42(0.41)**
	SS	-7108.42(0.18)	-26367.57(0.48)	-33735.44(0.50)	-13330.06(0.54)	**-8214.51(0.28)**	-6724.07(0.86)	-37332.76(2.42)	-8649.88(1.75)

Lower bounds and marginal likelihood estimation



Computational complexity and convergence behaviors



Inference gaps on tree topologies

Gap	PSP		Planar (16)		Planar (32)		RealNVP (5)		RealNVP (10)	
	Tree 36	All	Tree 36	All	Tree 36	All	Tree 36	All	Tree 36	All
Approximation	1.29	1.21	1.12	1.08	1.07	1.02	0.65	0.62	**0.43**	**0.40**
Amortization	3.37	0.84	2.80	0.82	**1.33**	**0.72**	3.10	0.98	1.83	0.93
Inference	4.66	2.05	3.92	1.90	2.40	1.74	3.75	1.60	**2.26**	**1.33**

References

D. Rezende and S. Mohamed (2015). "Variational inference with normalizing flows.". In: International Conference on Machine Learning, 2015, pp. 1530–1538.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio (2017). "Density estimation with real nvp.". In: International Conference on Learning Representations, 2017.

Cheng Zhang and Frederick A. Matsen IV (2018). "Generalizing tree probability estimation via Bayesian networks.". In: Advances in Neural Information Processing Systems 31, pp. 1444–1453.

Cheng Zhang and Frederick A. Matsen IV (2019). "Variational Bayesian phylogenetic inference.". In: International Conference on Learning Representations, 2019.