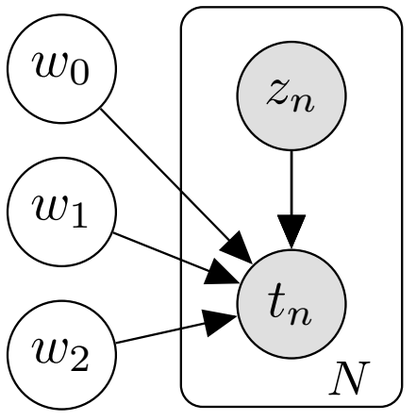
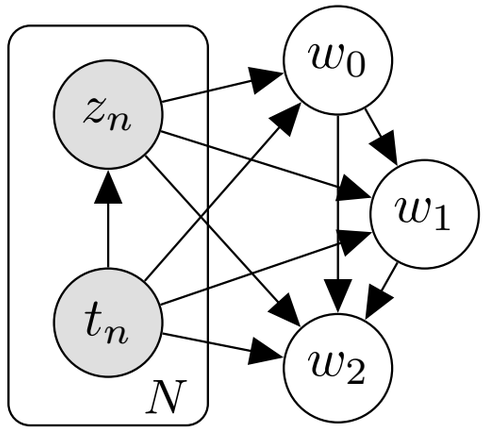
Inference Networks for Graphical Models

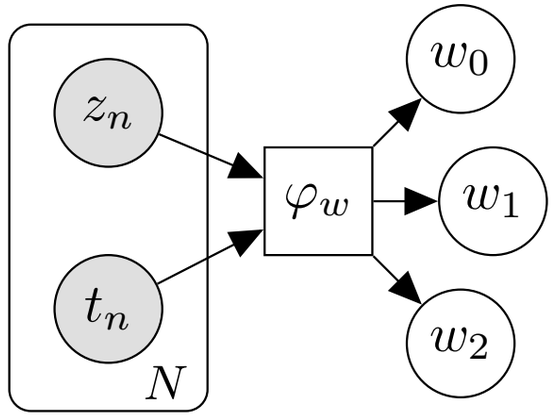
Brooks Paige and Frank Wood



A probabilistic model
generates data



An inverse model
generates latents



Can we **learn how to sample**
from the inverse model?

Idea: amortize inference by learning a map from data to target

Target density $\pi(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$, approximating family $q(\mathbf{x}|\lambda)$

Single dataset $\mathbf{y}$: $\underset{\lambda}{\operatorname{argmin}}\, D_{KL}(\pi||q_\lambda)$ ← fit λ to learn an importance sampling proposal

Averaging over
all possible datasets: $\lambda = \varphi(\eta, \mathbf{y})$ ← learn a mapping from arbitrary datasets to λ

$\underset{\eta}{\operatorname{argmin}}\, \mathbb{E}_{p(\mathbf{y})}\left[D_{KL}(\pi||q_{\varphi(\eta,\mathbf{y})})\right]$ …compiles away runtime costs of inference!