Nested Variational Inference

Heiko Zimmermann[1] **Hao Wu[2]** **Babak Esmaeili[1]** **Jan-Willem van de Meent[1,2]**

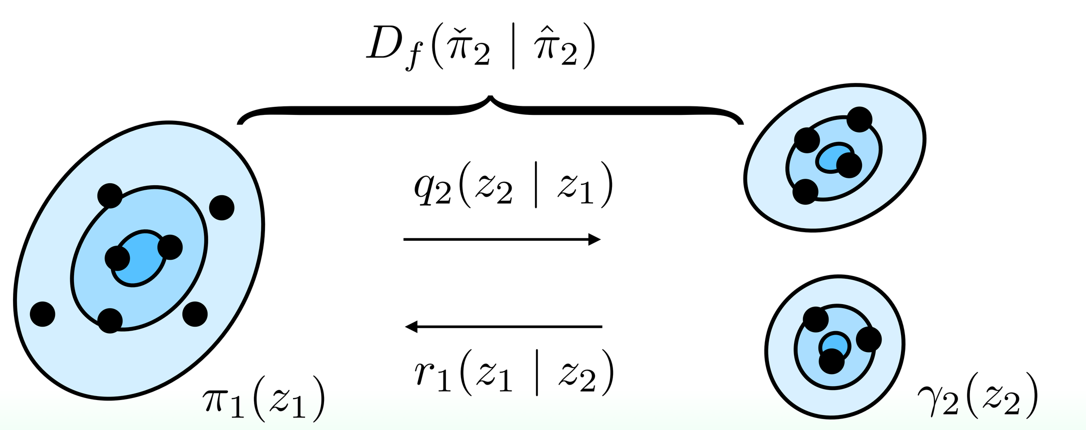
[1]Amsterdam Machine Learning Lab, University of Amsterdam, Amsterdam, Netherlands
[1]{h.zimmermann, b.esmaeili, j.w.vandemeent}@uva.nl

[2]Khoury College of Computer Sciences, Northeastern University, Boston, USA
[2]{wu.hao10, j.vandemeent}@northeastern.edu

1

Shortcomings of standard VI

- Potentially hard one-step learning problem
- For mode-covering divergences, the variance of weights and gradient estimates is high due to initially bad samples
- For mode-seeking divergences, variational distribution might not capture all modes



$z_1 \sim \pi_1$

$z_2 \sim q_2(\cdot \mid z_1)$

$$w = \frac{\gamma_2(z_2) r_1(z_1 \mid z_2)}{\pi_1(z_1) q_2(z_2 \mid z_1)}$$

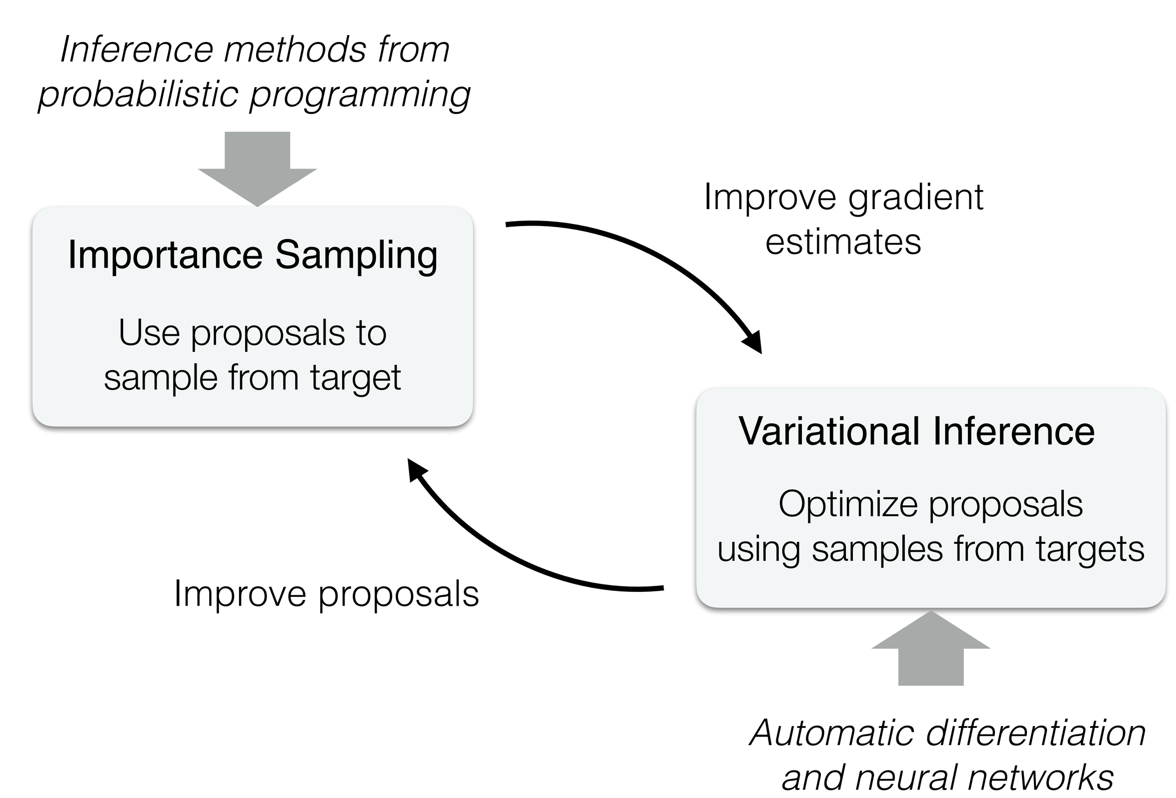
$$\pi_2(z_2) = Z_2^{-1} \gamma_2(z_2)$$

$$\hat{\pi}_2(z_1, z_2) = \pi_1(z_1) q_2(z_2 \mid z_1)$$

$$\check{\pi}_2(z_1, z_2) = \pi_2(z_2) r_1(z_1 \mid z_2)$$

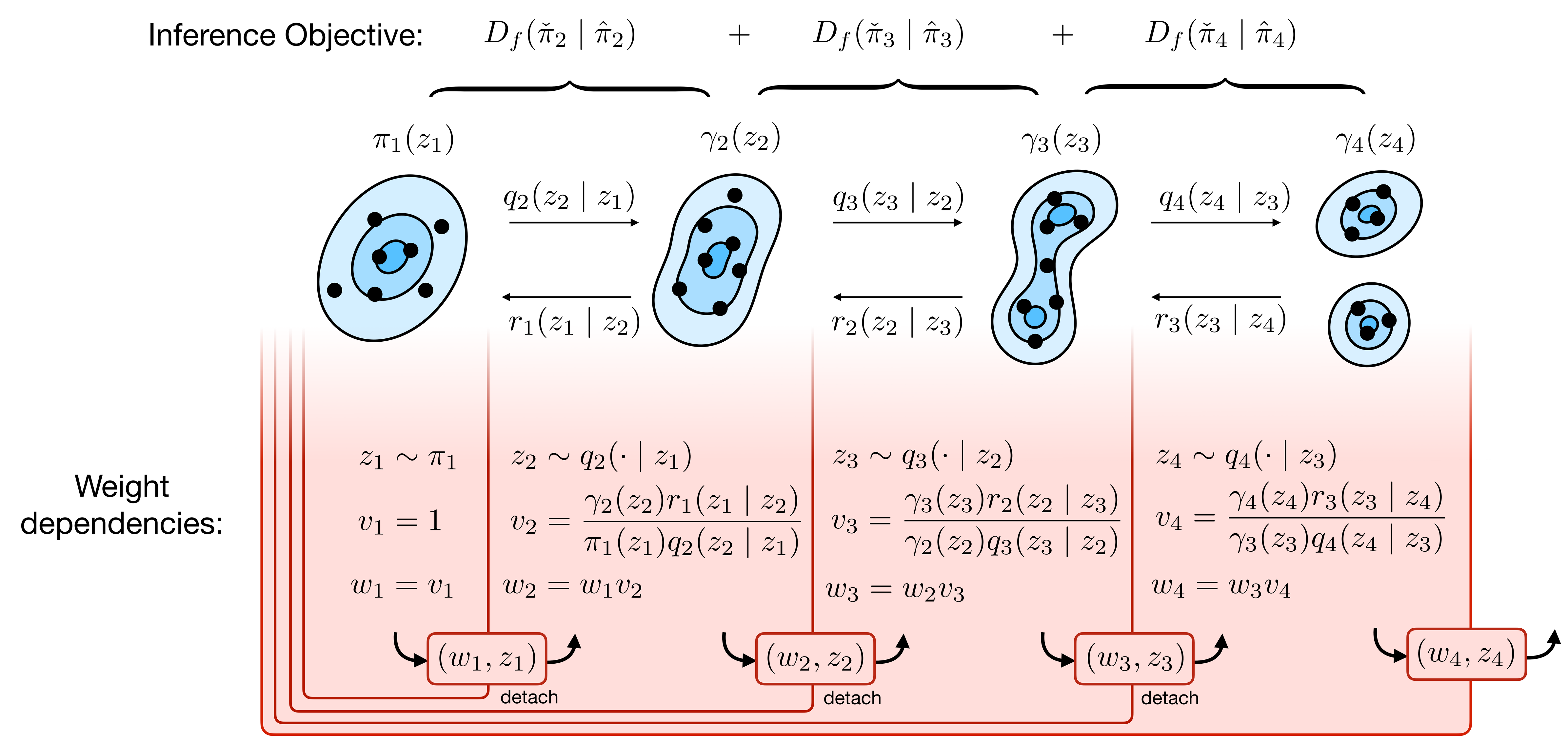
Combining Importance Sampling with VI

- Break problem into sequence of easier problems
- Initial samples are good enough to make progress on early stages; Optimizing earlier stages improves samples for later stages
- Samples/Optimization can be guided by choice of intermediate distributions



2

Nested Variational Inference



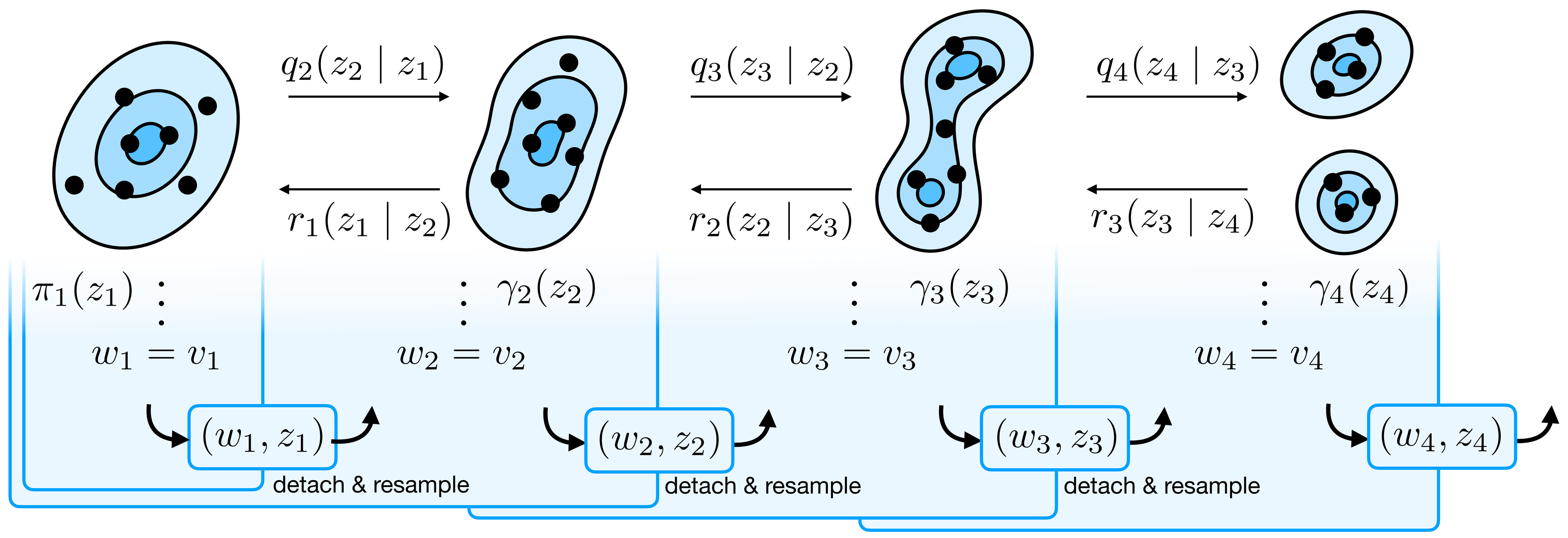
3

Proper Weighting and Resampling

Proper weighting *Let π be a probability density. For some constant $c > 0$, a random pair $(w, z) \sim \Pi$ is properly weighted (p.w.) for an unnormalized probability density $\gamma \equiv Z\pi$ if $w \geq 0$ and for all measurable functions g*

$$\mathbb{E}_{w,z \sim \Pi}\left[w\, g(z)\right] = c \int dz\ \gamma(z)\, g(z) = cZ\mathbb{E}_{z \sim \pi}\left[g(z)\right].$$

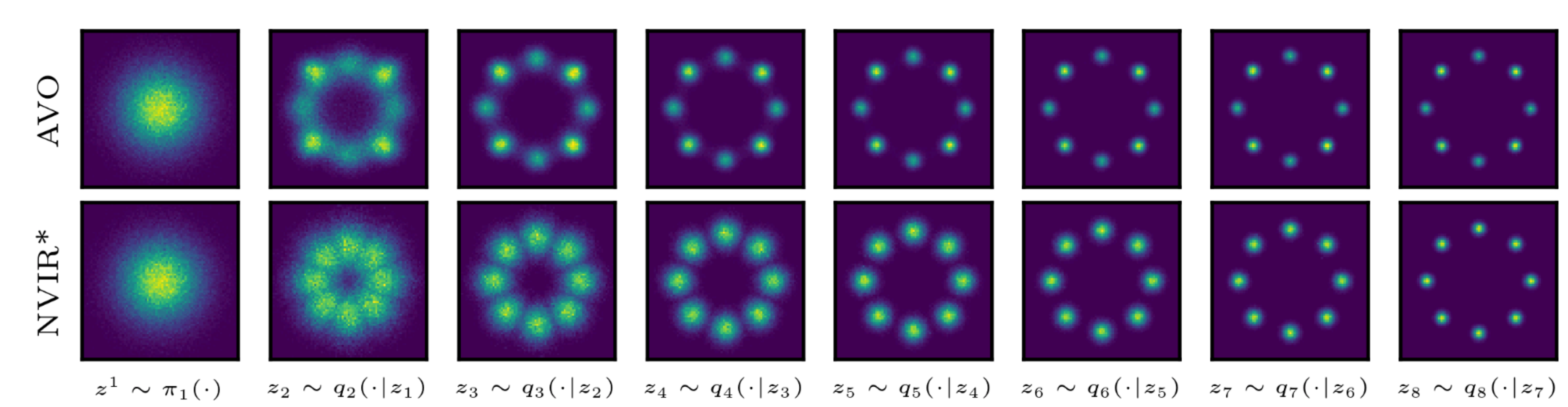
We can propose from any sampler as long as it produces properly weighted samples for the proposal density of interest. **Resampling preserves proper weighting.**



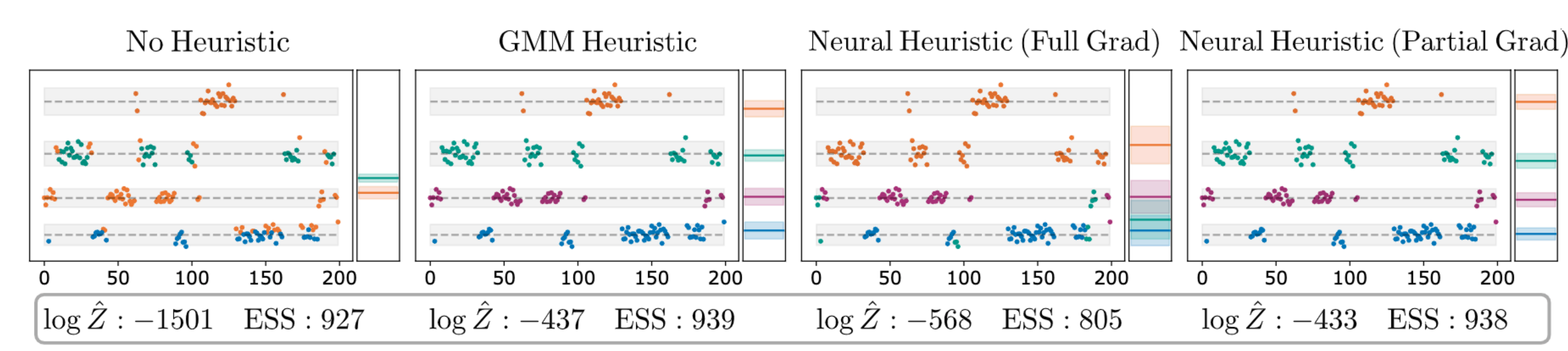
4

Experiments

Training Annealed Samplers

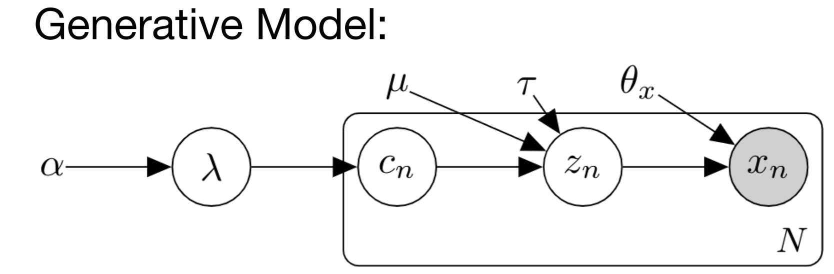


Time Series Model



Deep Generative Model

Generative Model:



Inference Model: