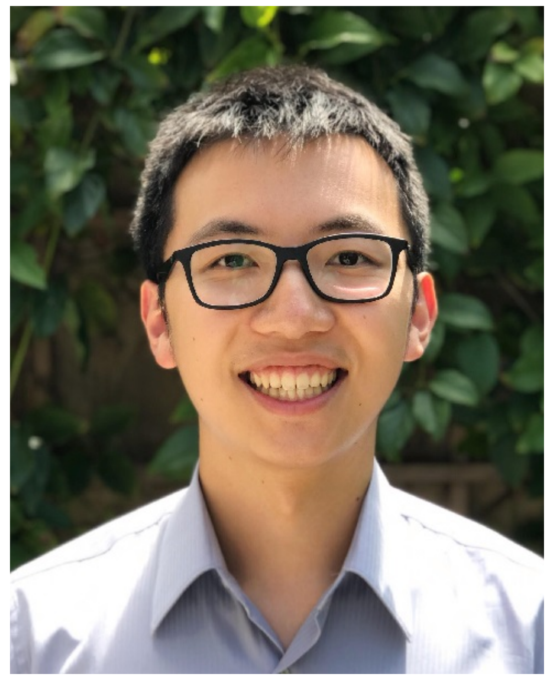
Probabilistic Circuits for Variational Inference in Discrete Graphical Models

NeurIPS 2020

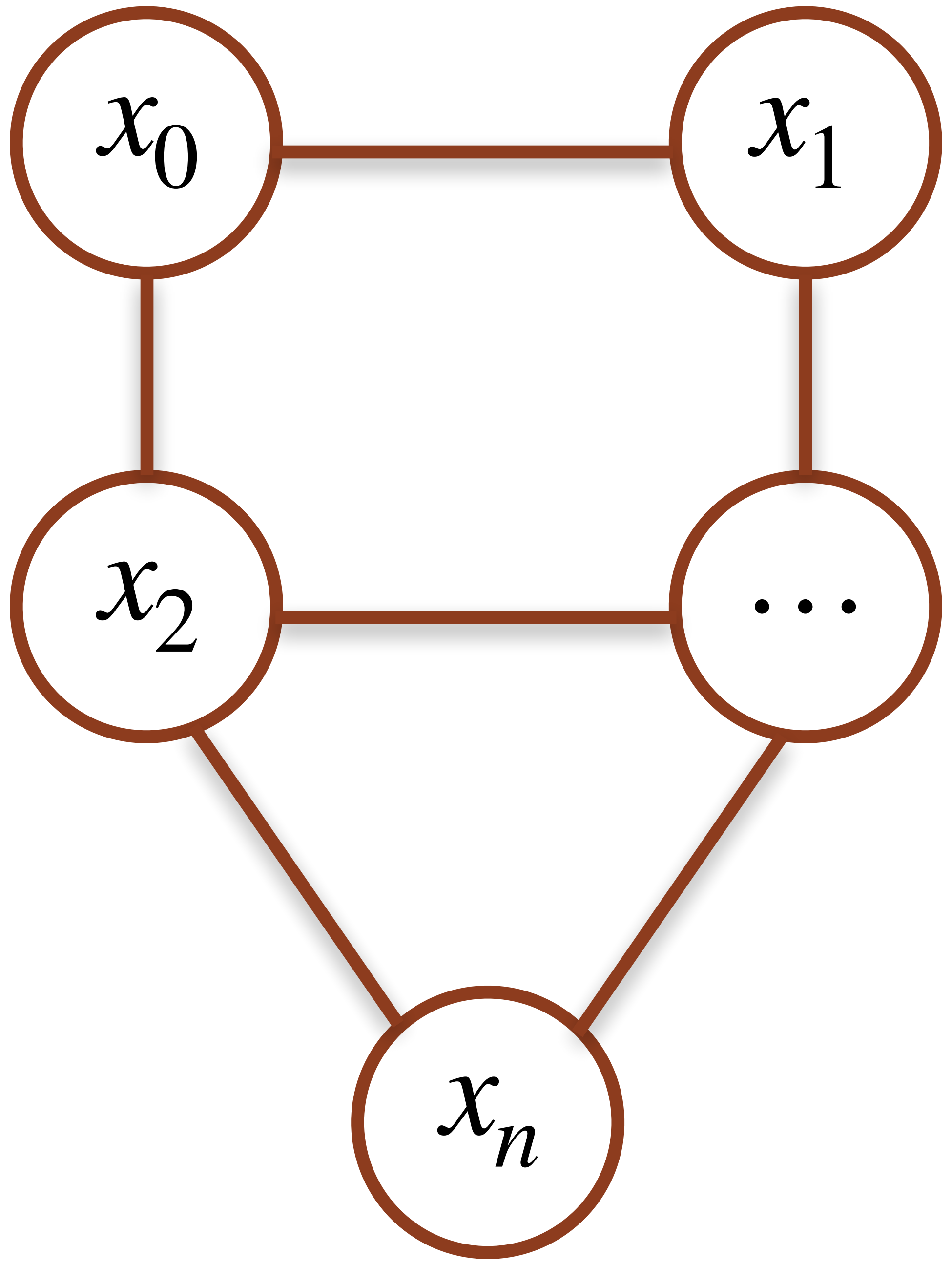


Andy Shih

Stefano Ermon

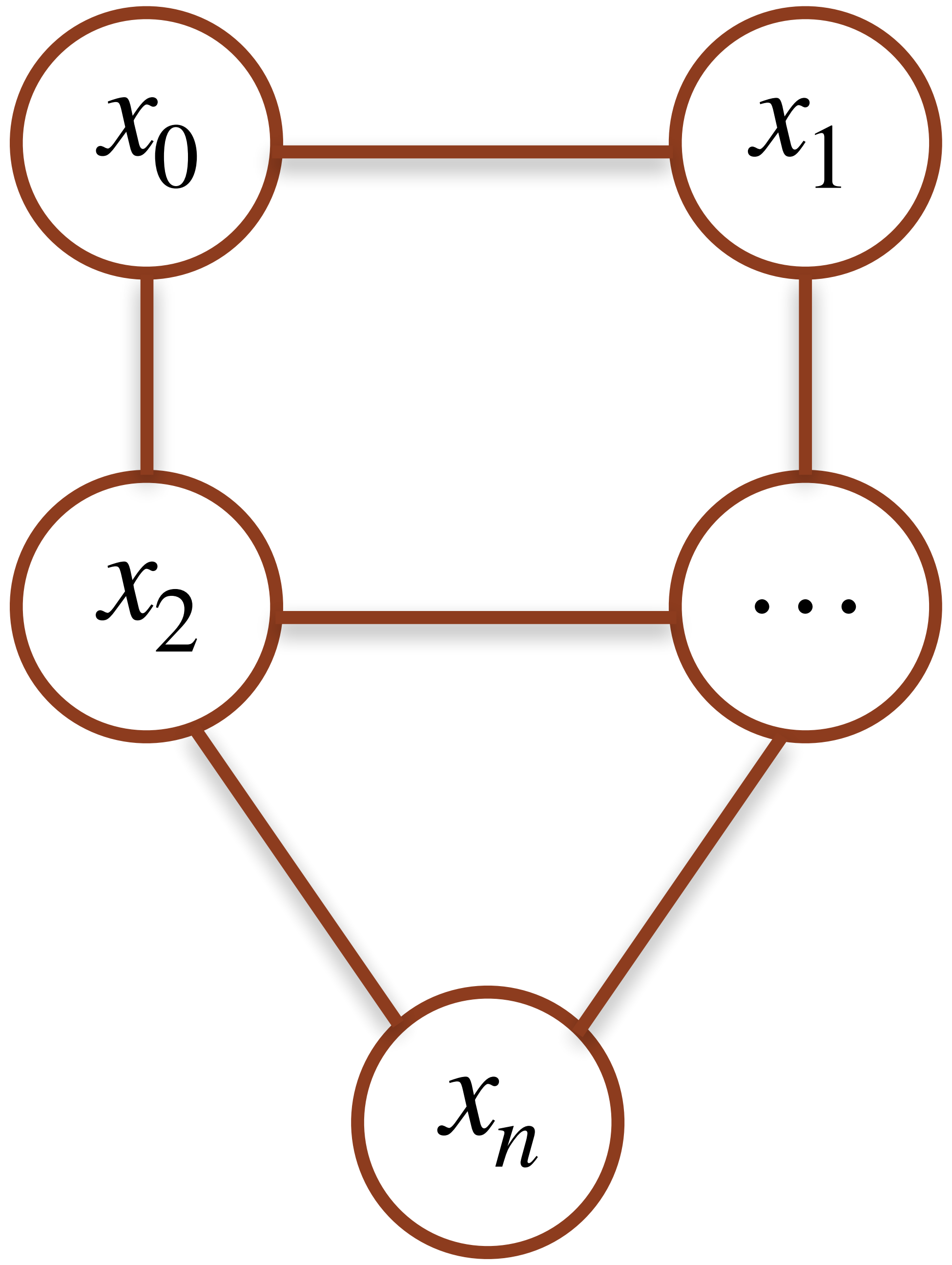
Variational Inference

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$



Variational Inference

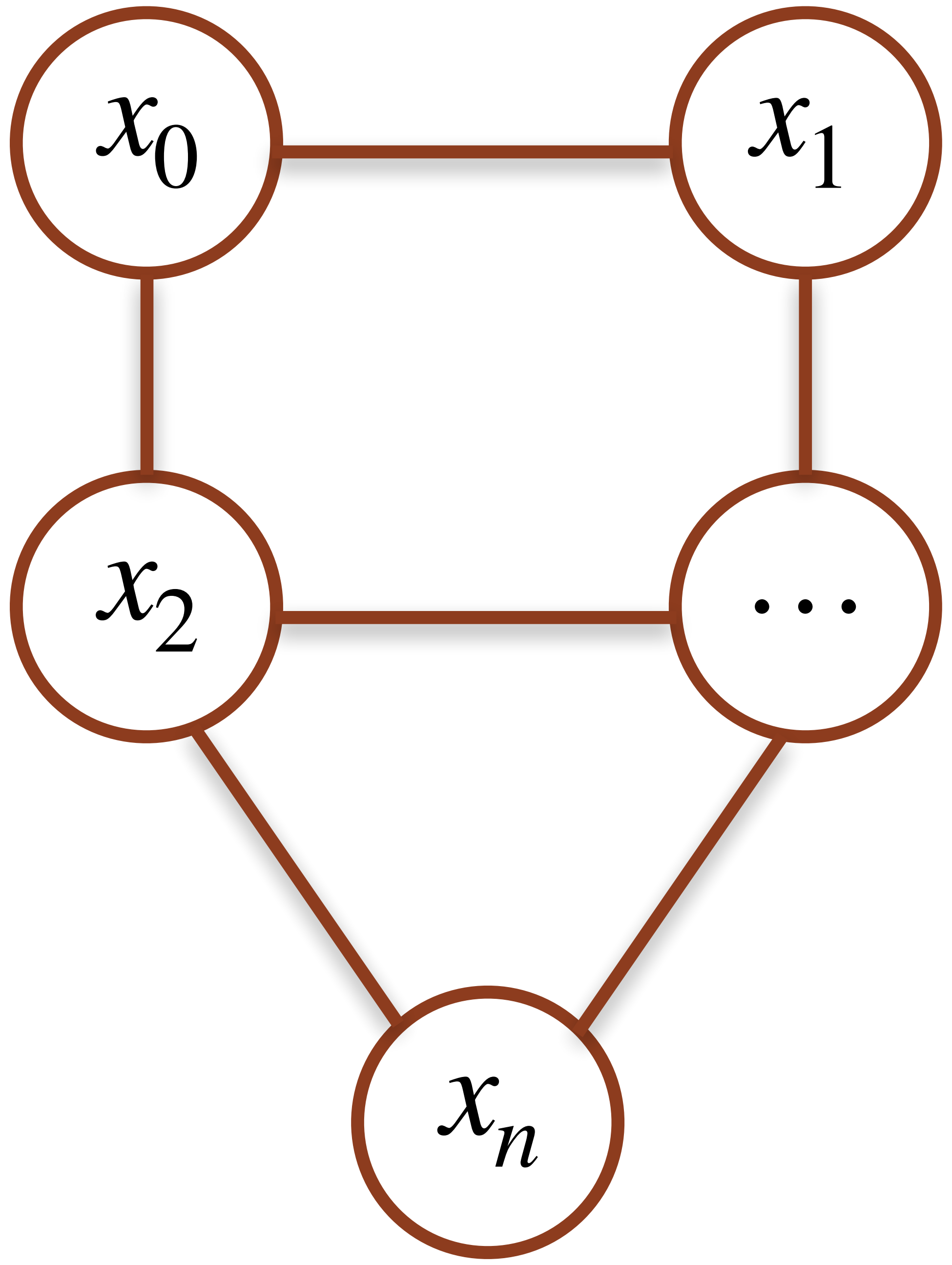
$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$



Variational Inference

$$\boxed{\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]}$$

Any choice of q gives a lower bound



Variational Inference

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Any choice of q gives a lower bound

Choice of q

Analytic optimization:

- mean field
- structured mean field

Variational Inference

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Any choice of q gives a lower bound

Choice of q

Analytic optimization:

- mean field
- structured mean field

Stochastic optimization:

- neural networks

Variational Inference

$$\log Z \geq \mathbb{E}_{\mathbf{x} \sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

sample

Choice of q

Analytic optimization:

- mean field
- structured mean field

Stochastic optimization:

- neural networks

Variational Inference

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

sample

Choice of q

Analytic optimization:

- mean field
- structured mean field

Stochastic optimization:

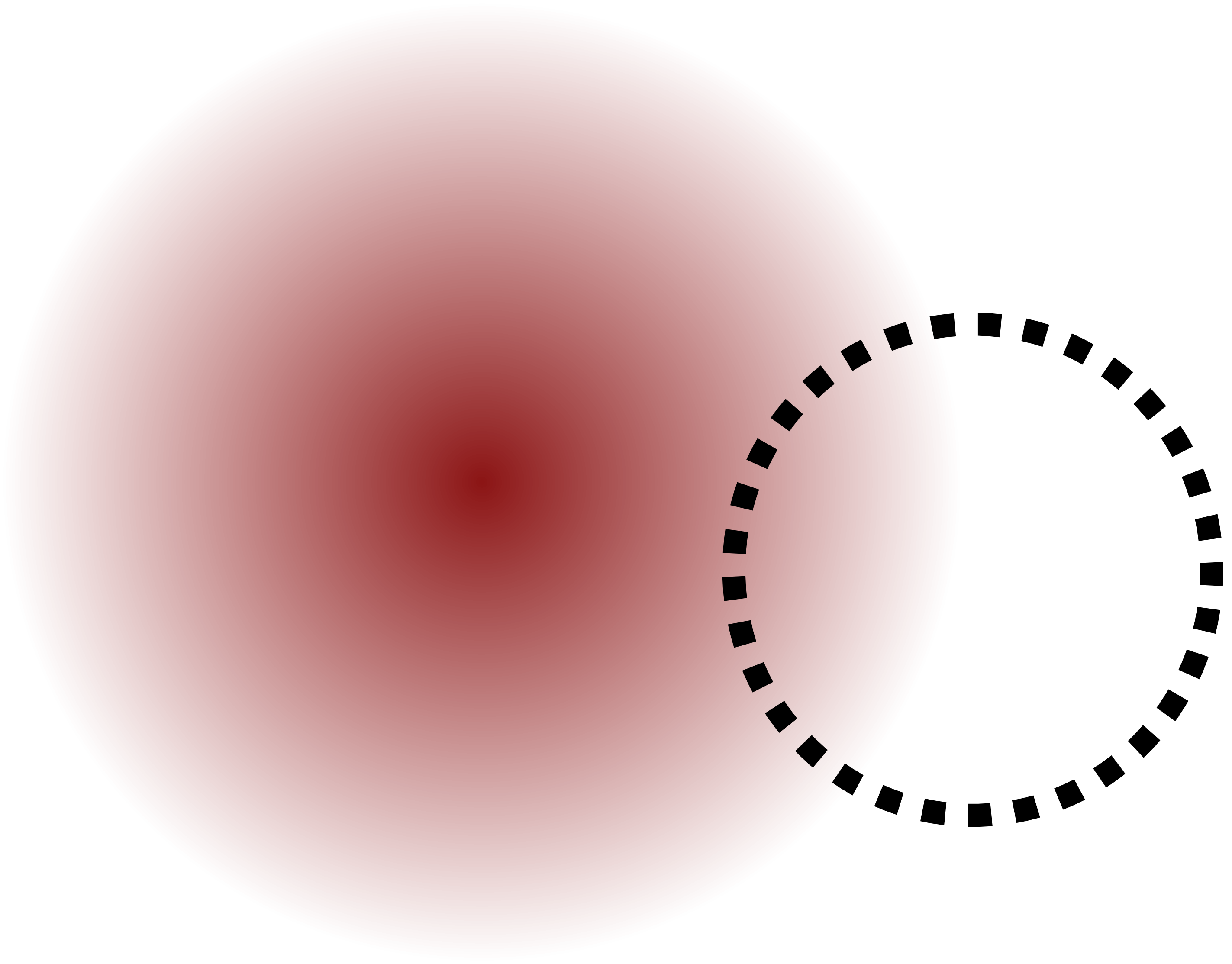
- neural networks

continuous

discrete

[Zhang 2017]

Sampling — Continuous Settings

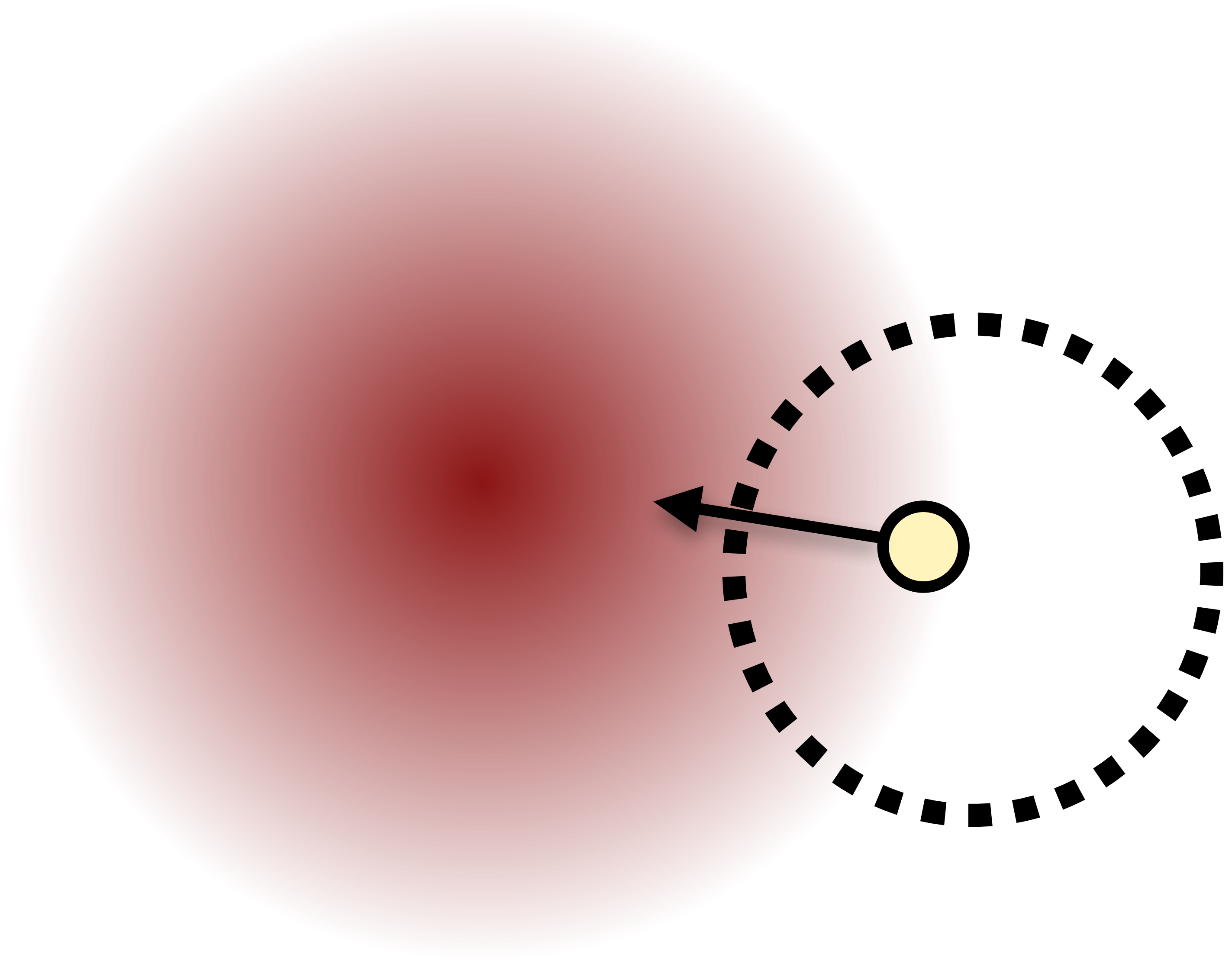


$$\log Z \geq \mathbb{E}_{\mathbf{x} \sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Proposal distribution q

Target distribution p

Sampling — Continuous Settings



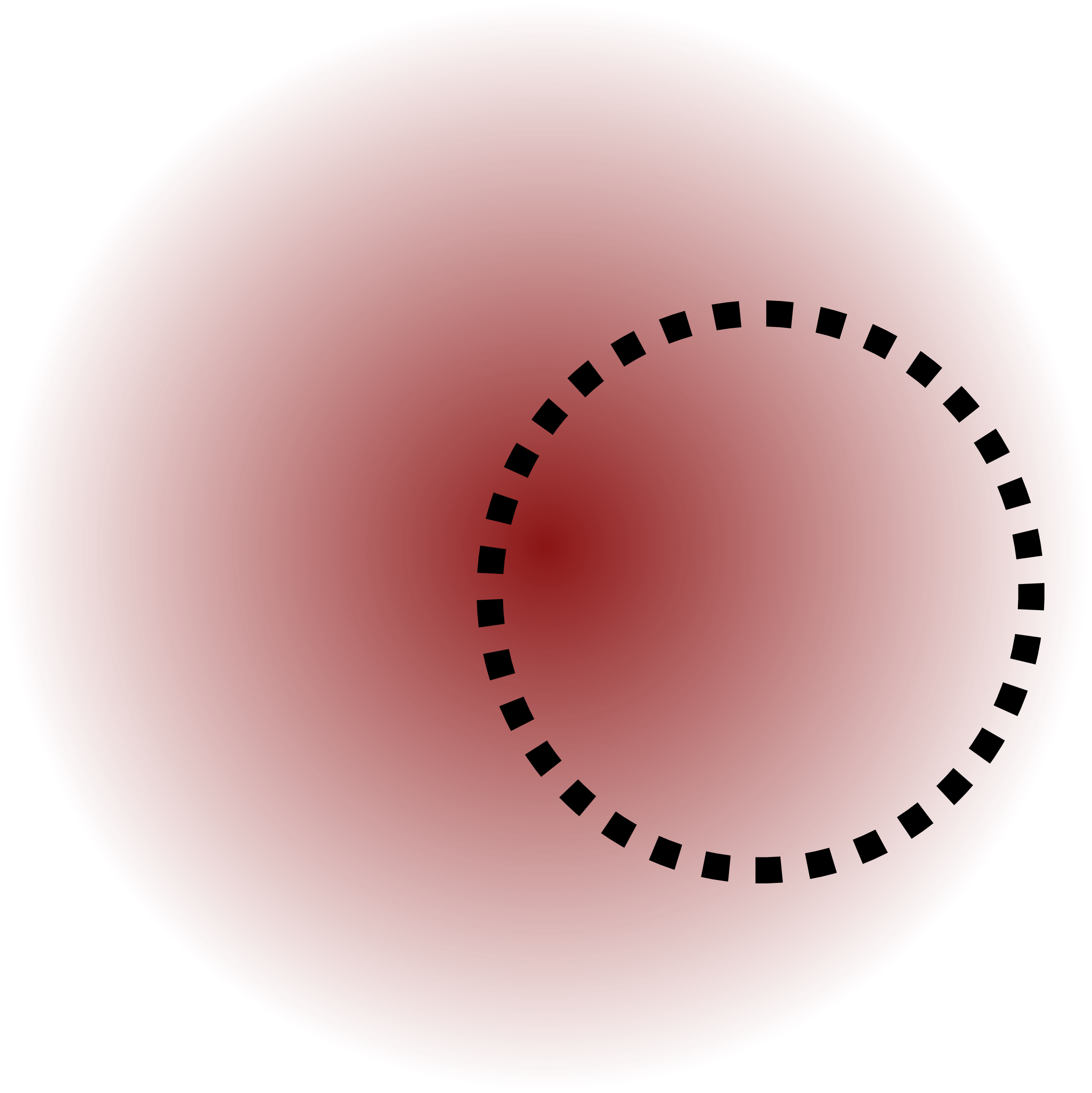
draw a sample

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Proposal distribution q

Target distribution p

Sampling — Continuous Settings

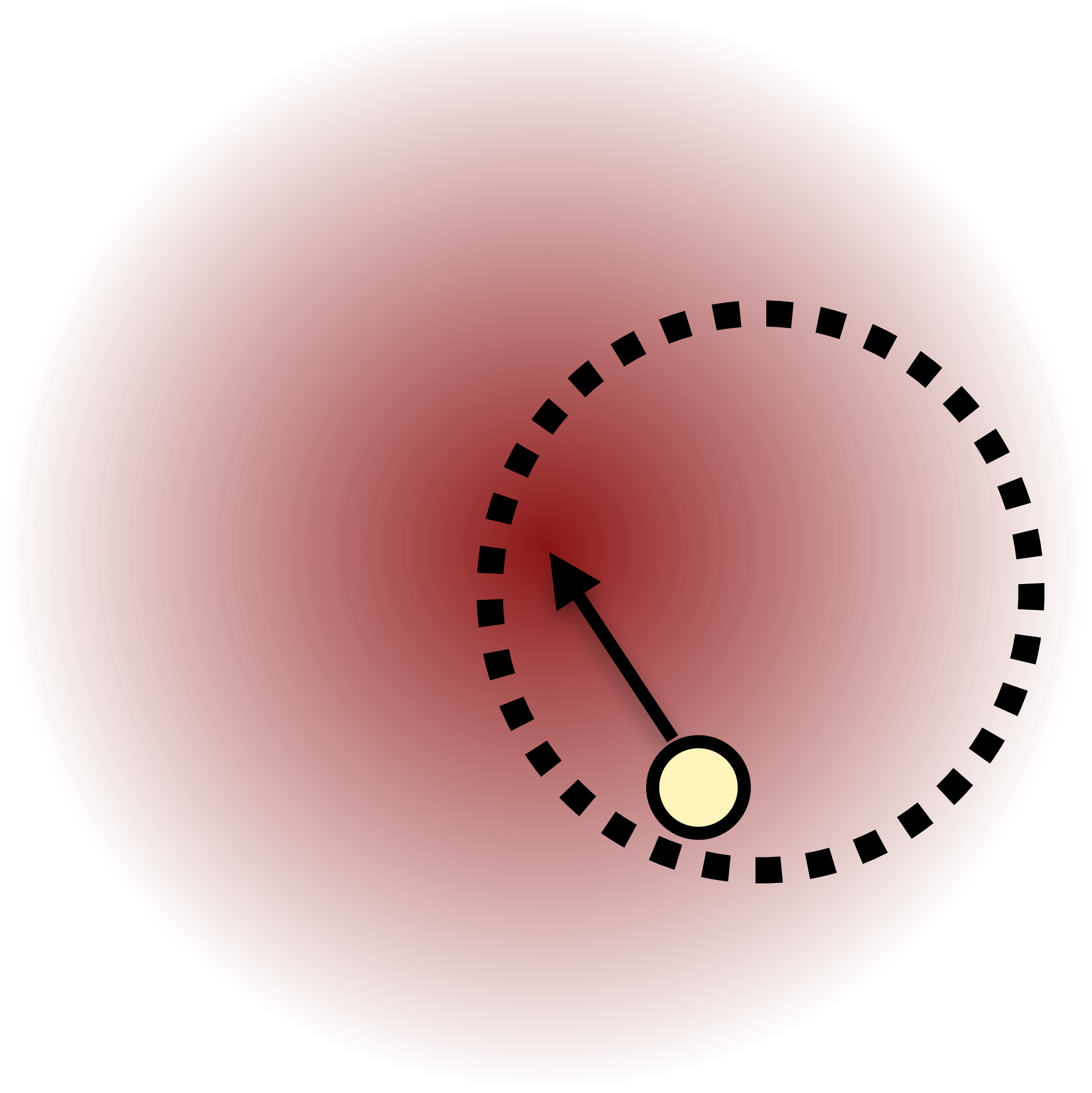


Proposal distribution q

Target distribution p

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Sampling — Continuous Settings



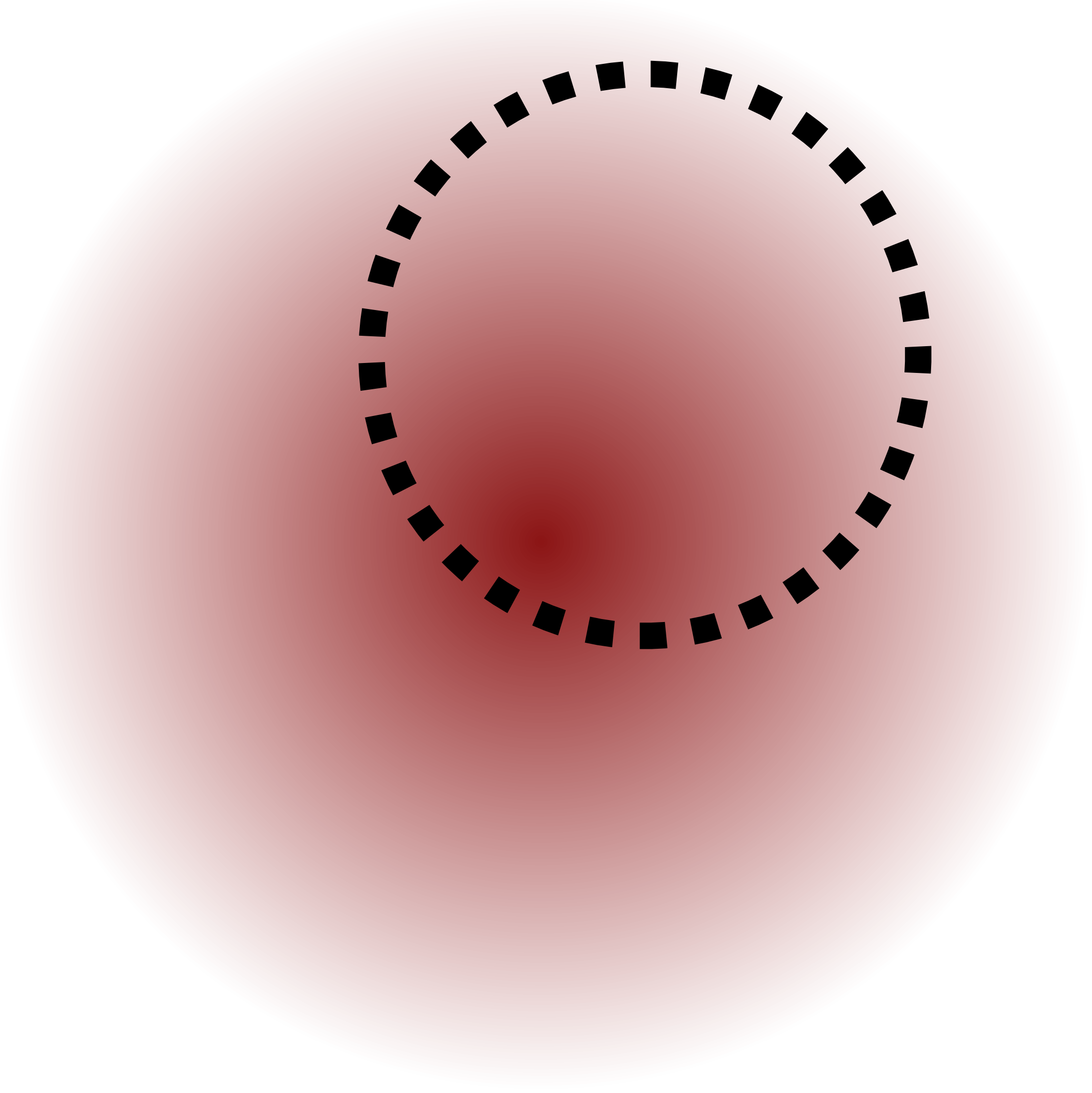
draw a sample

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Proposal distribution q

Target distribution p

Sampling — Continuous Settings

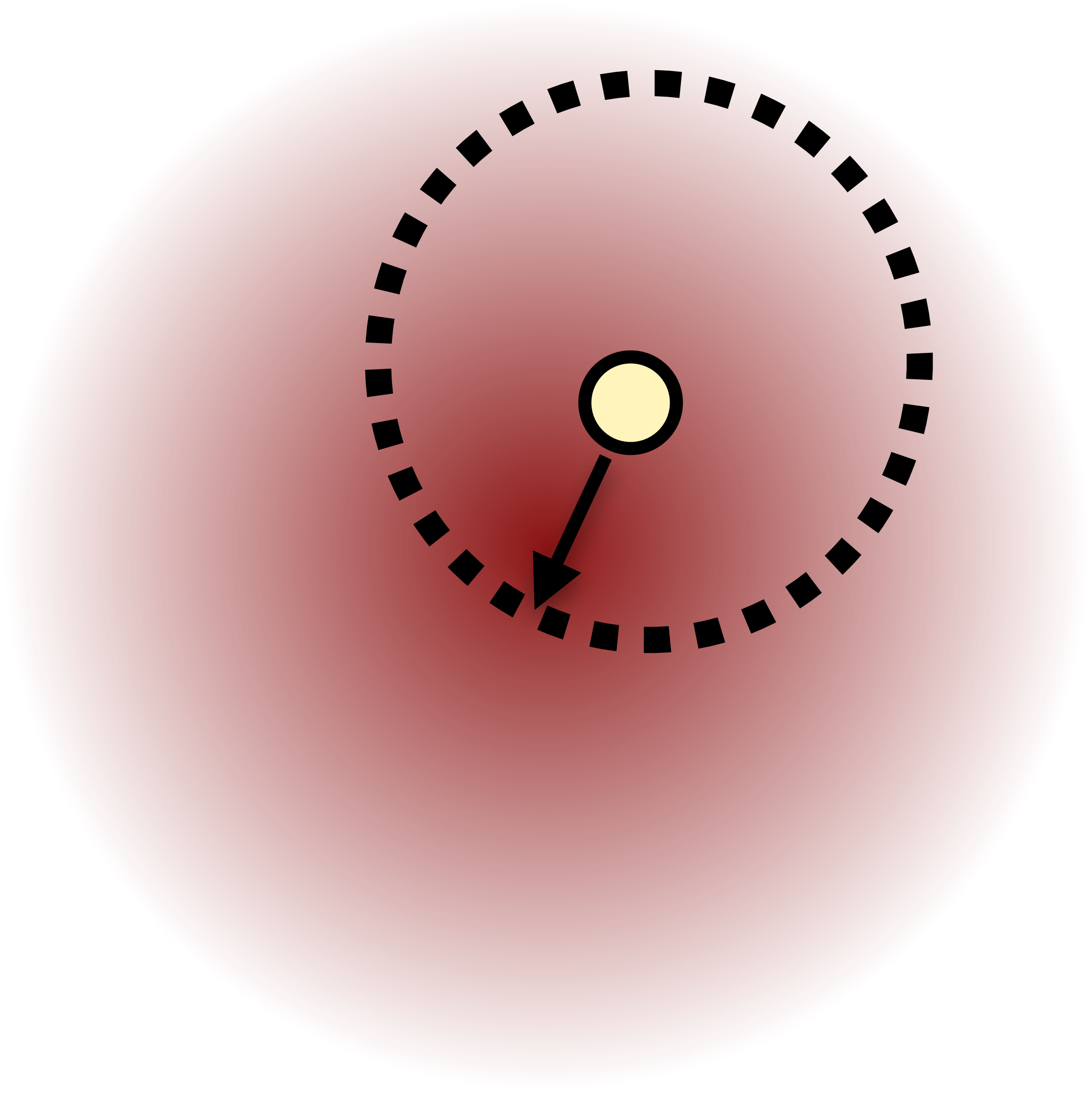


$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Proposal distribution q

Target distribution p

Sampling — Continuous Settings



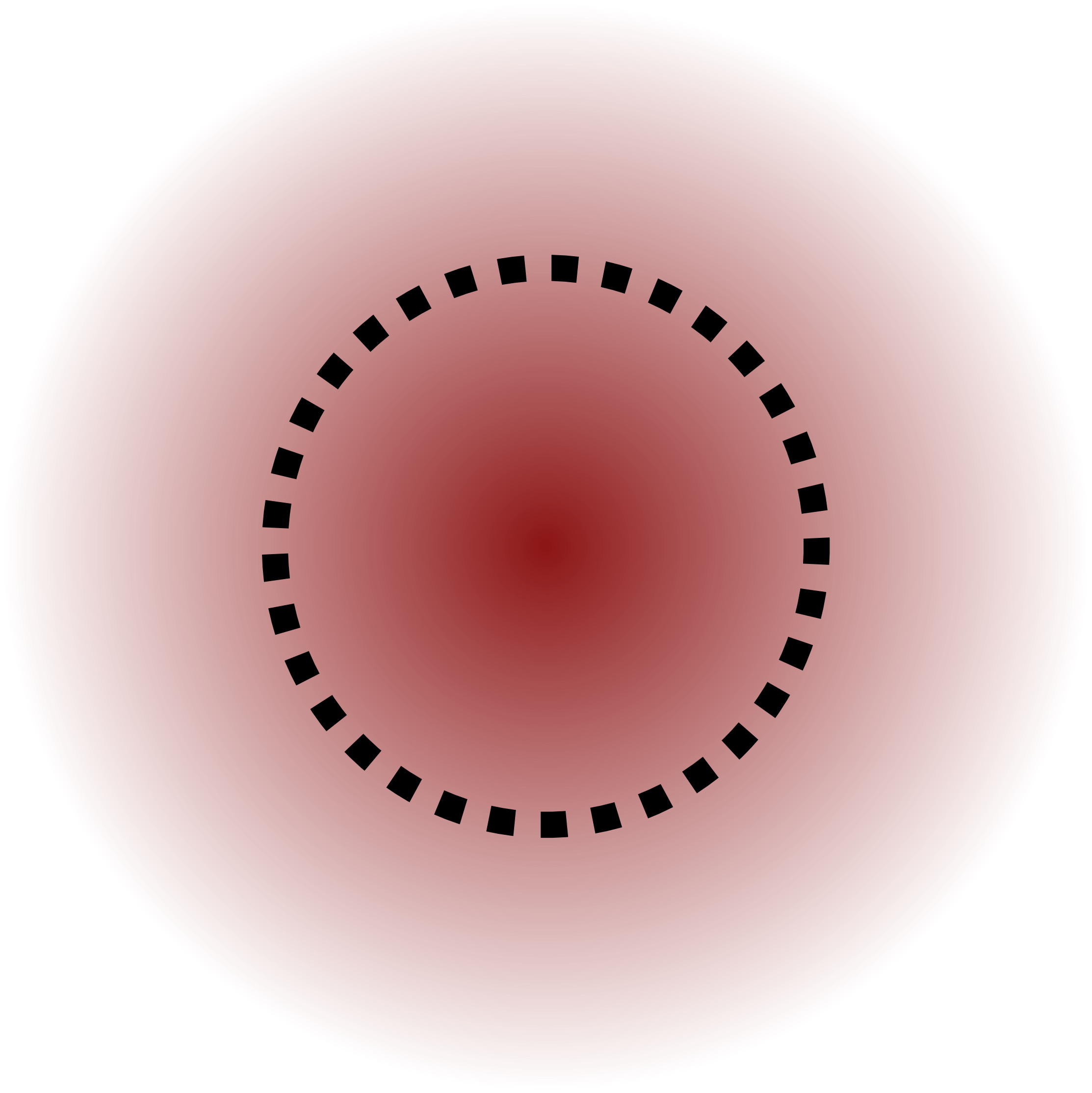
draw a sample

$$\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Proposal distribution q

Target distribution p

Sampling — Continuous Settings

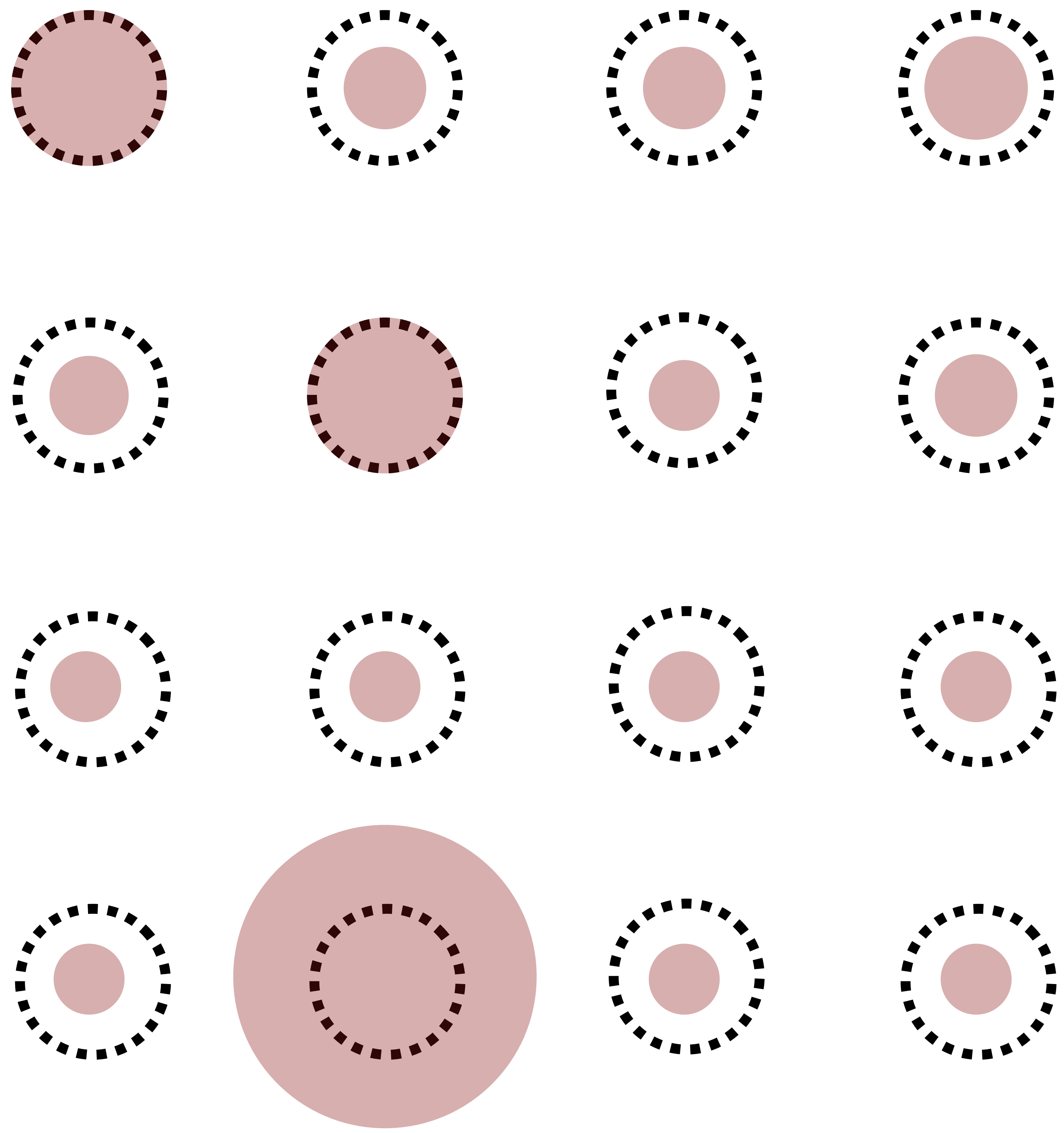


Proposal distribution q

Target distribution p

$$\boxed{\log Z \geq \mathbb{E}_{\mathbf{x} \sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]}$$

Sampling — Discrete Settings

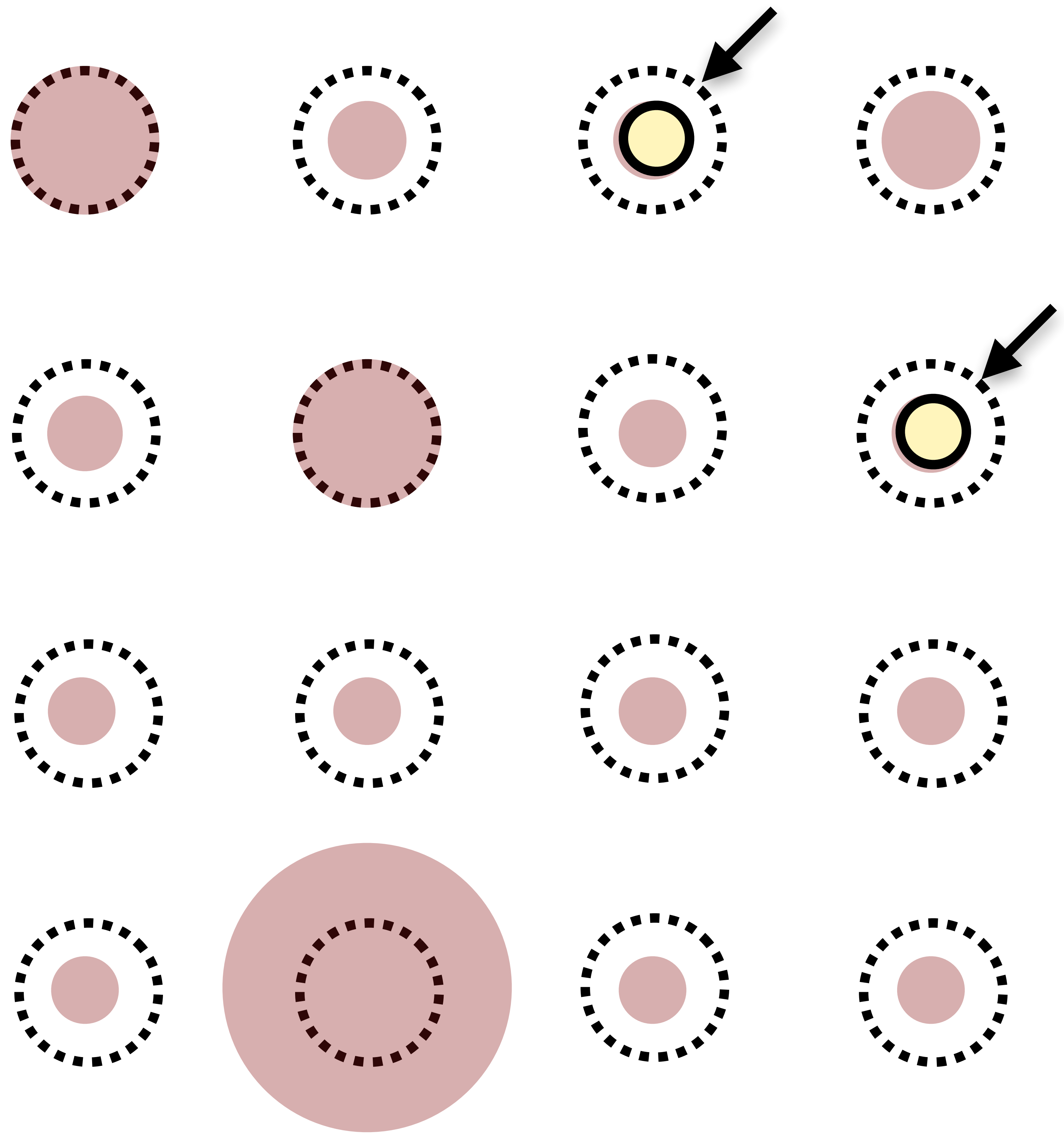


Each circle is a point in discrete space
Larger circle = high probability mass

Proposal distribution q

Target distribution p

Sampling — Discrete Settings

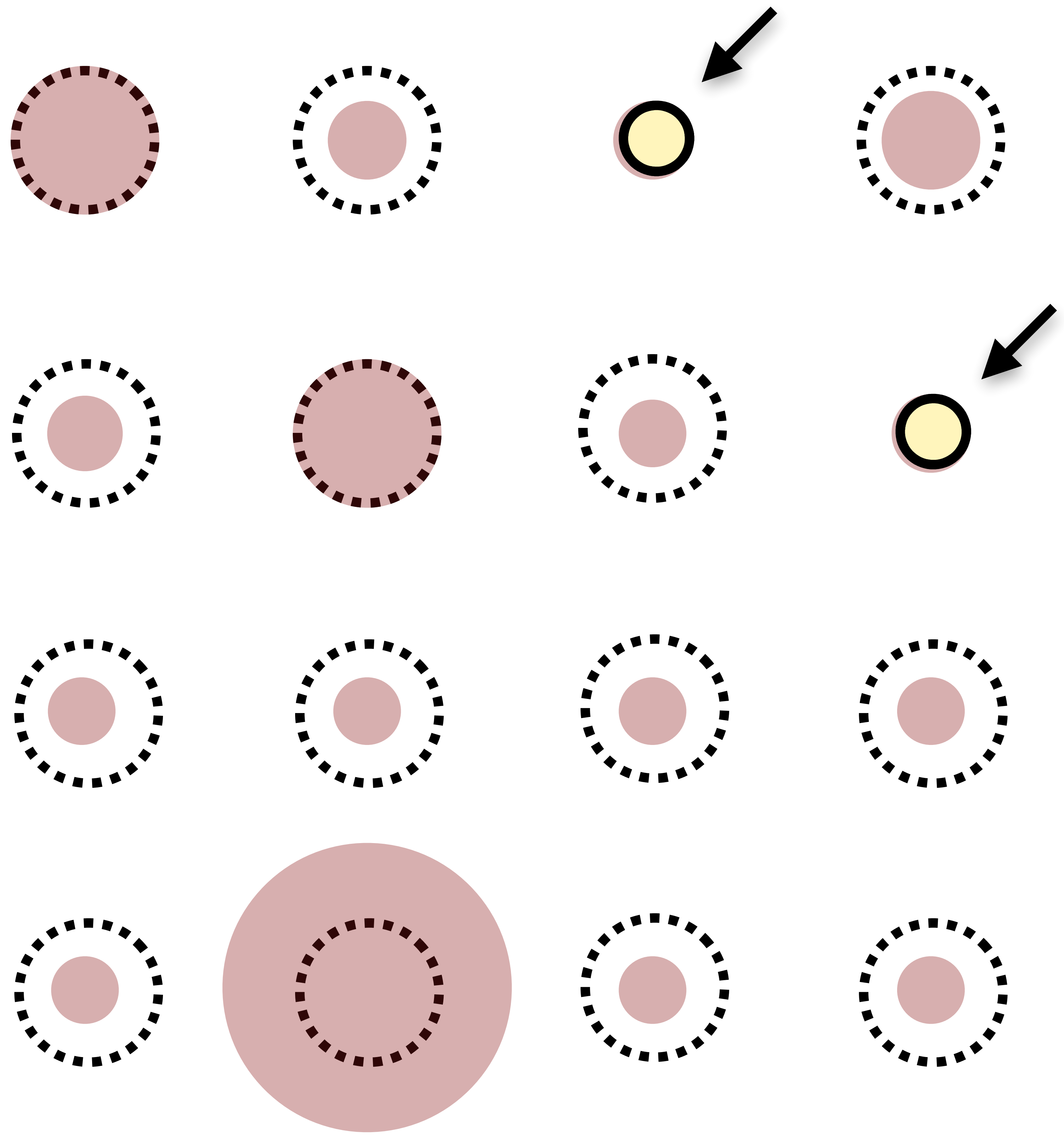


Each circle is a point in discrete space
Larger circle = high probability mass

Proposal distribution q

Target distribution p

Sampling — Discrete Settings

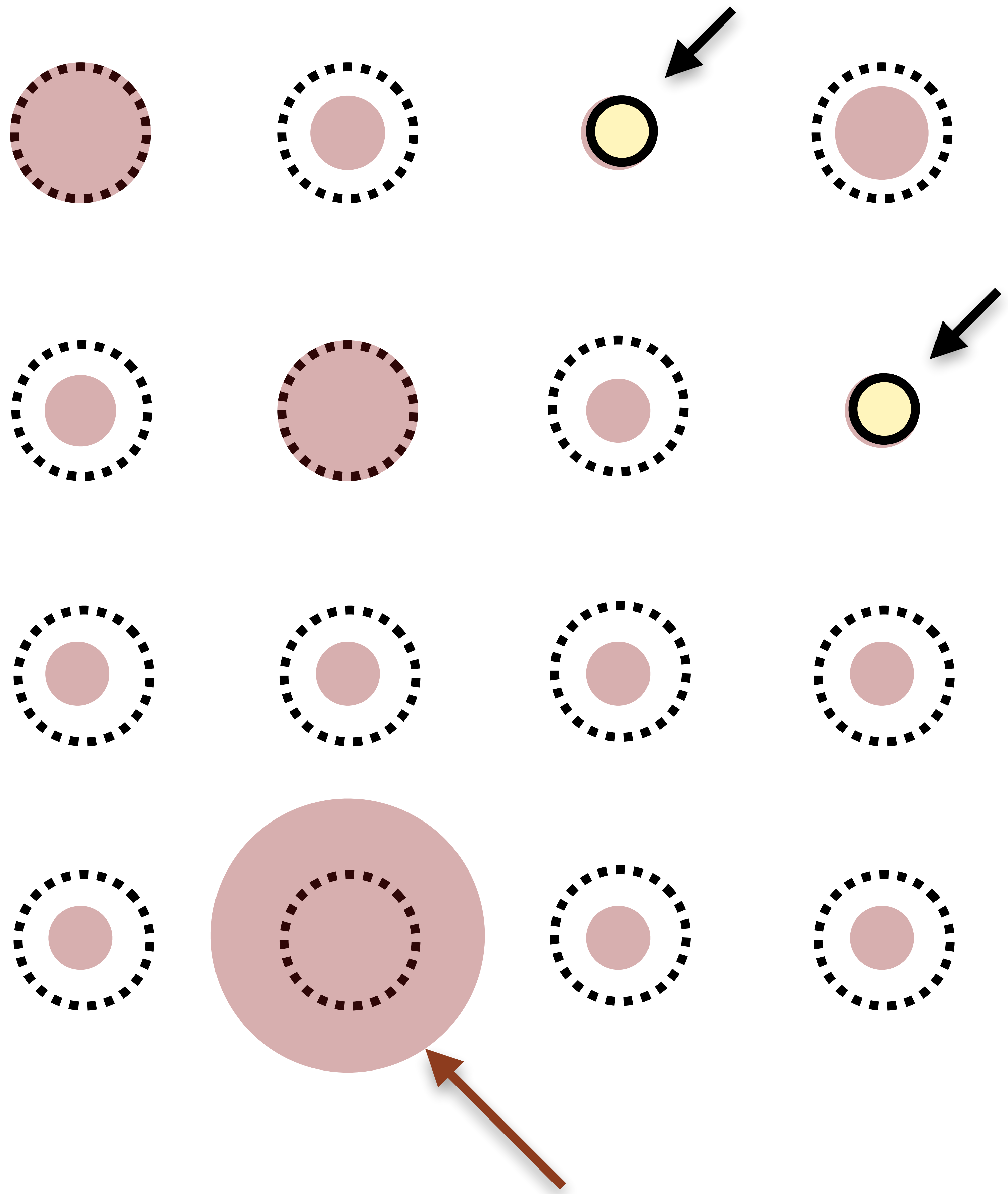


Each circle is a point in discrete space
Larger circle = high probability mass

Proposal distribution q

Target distribution p

Sampling — Discrete Settings



missing out on the big mode

Each circle is a point in discrete space
Larger circle = high probability mass

- No information "around" the samples
- Very high variance

Cannot easily optimize!

Discrete Settings

Choice of q

Analytic optimization:

- mean field
- structured mean field

Stochastic optimization:

- neural networks

Discrete Settings

Avoid sampling

Choice of q

Analytic optimization:

- mean field
- structured mean field

~~Stochastic~~ optimization:

- ~~neural~~ networks

Discrete Settings

Choice of q

Avoid sampling

Expressive distribution

Analytic optimization:

- ~~mean field~~
- ~~structured mean field~~

~~Stochastic optimization:~~

- ~~neural networks~~

Discrete Settings

Avoid sampling

Expressive distribution

Choice of q

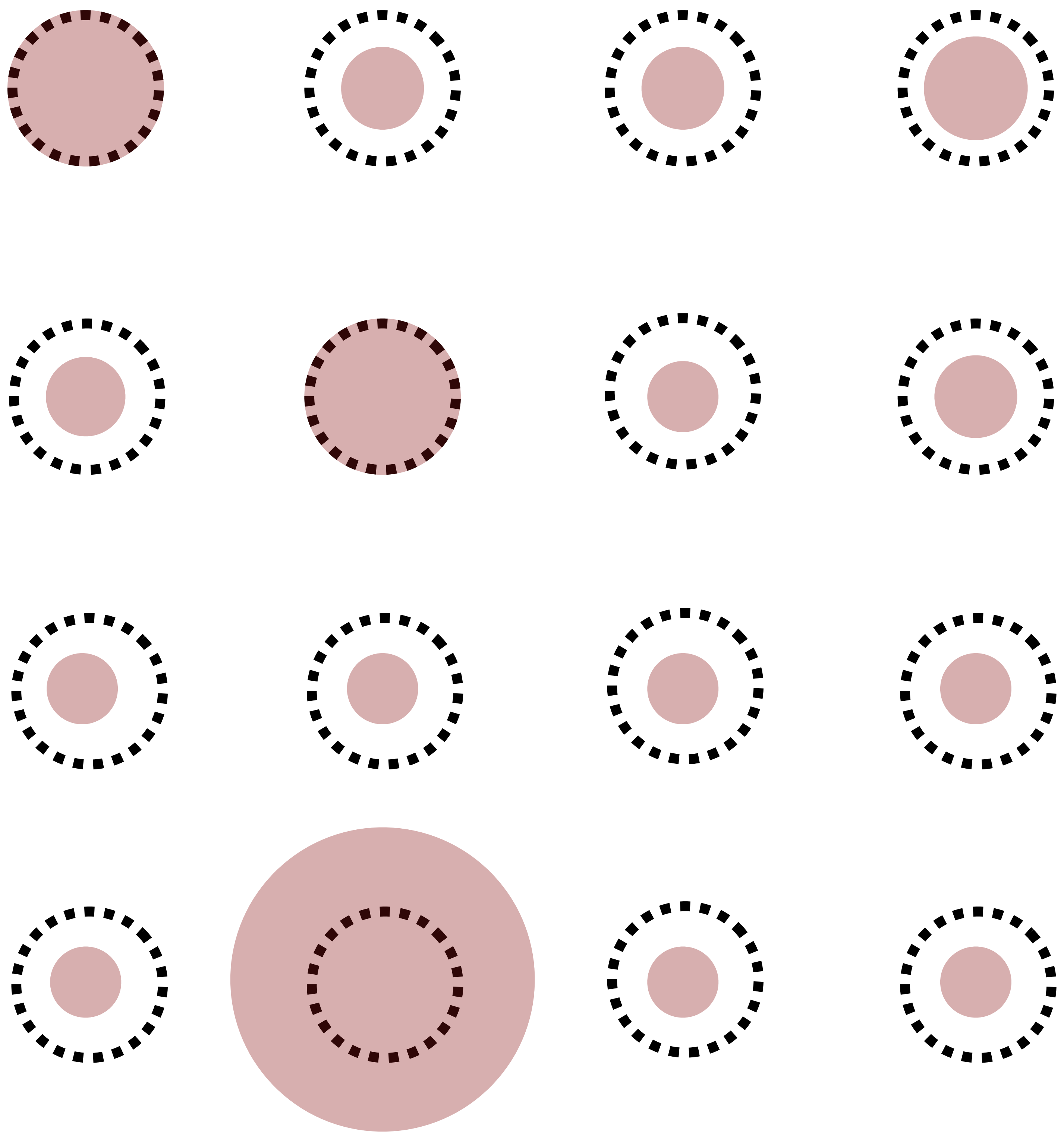
Analytic optimization:

- ~~mean field~~
- ~~structured mean field~~
- sum product networks

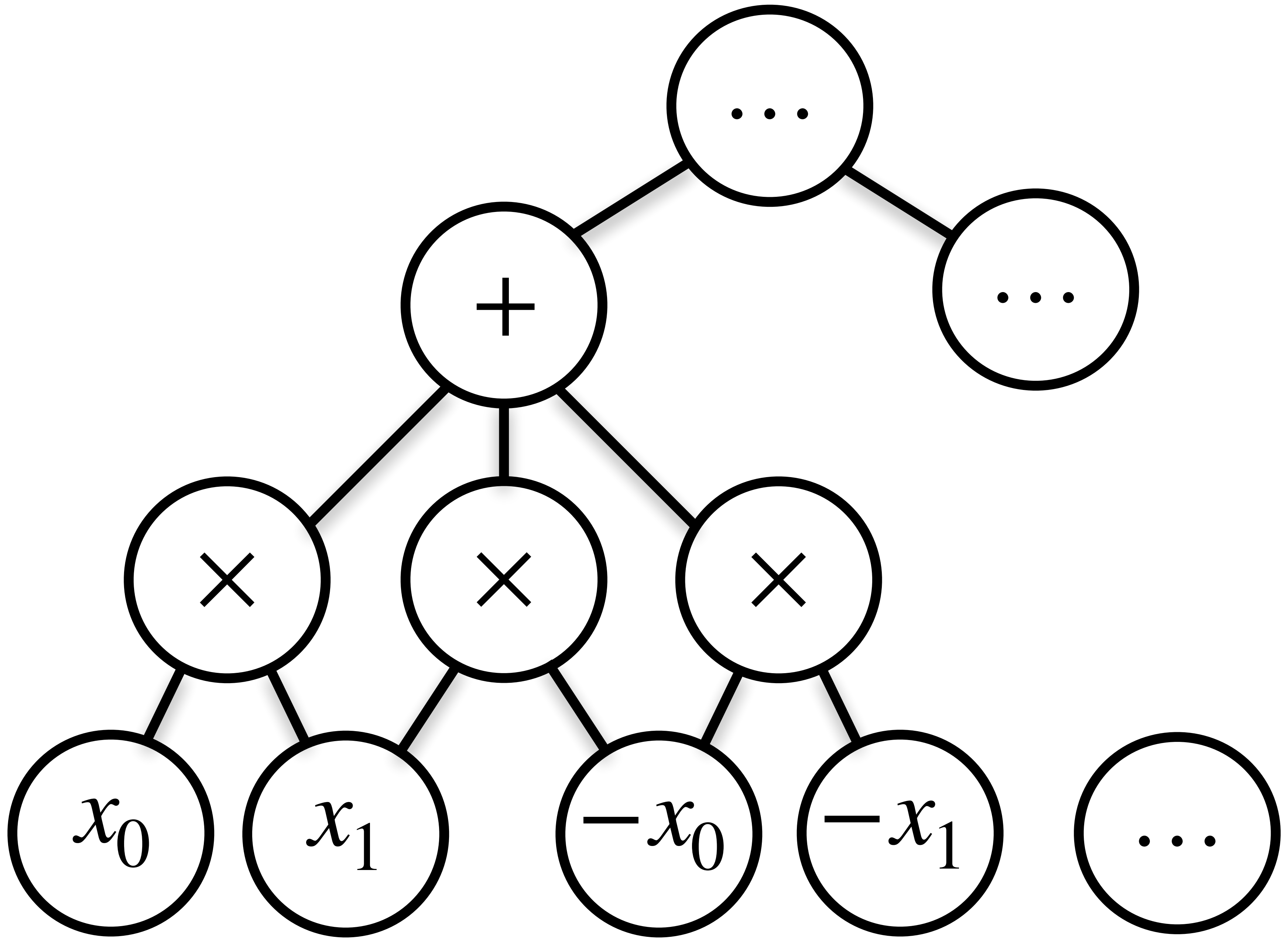
~~Stochastic~~ optimization:

- ~~neural~~ networks

Sum Product Networks

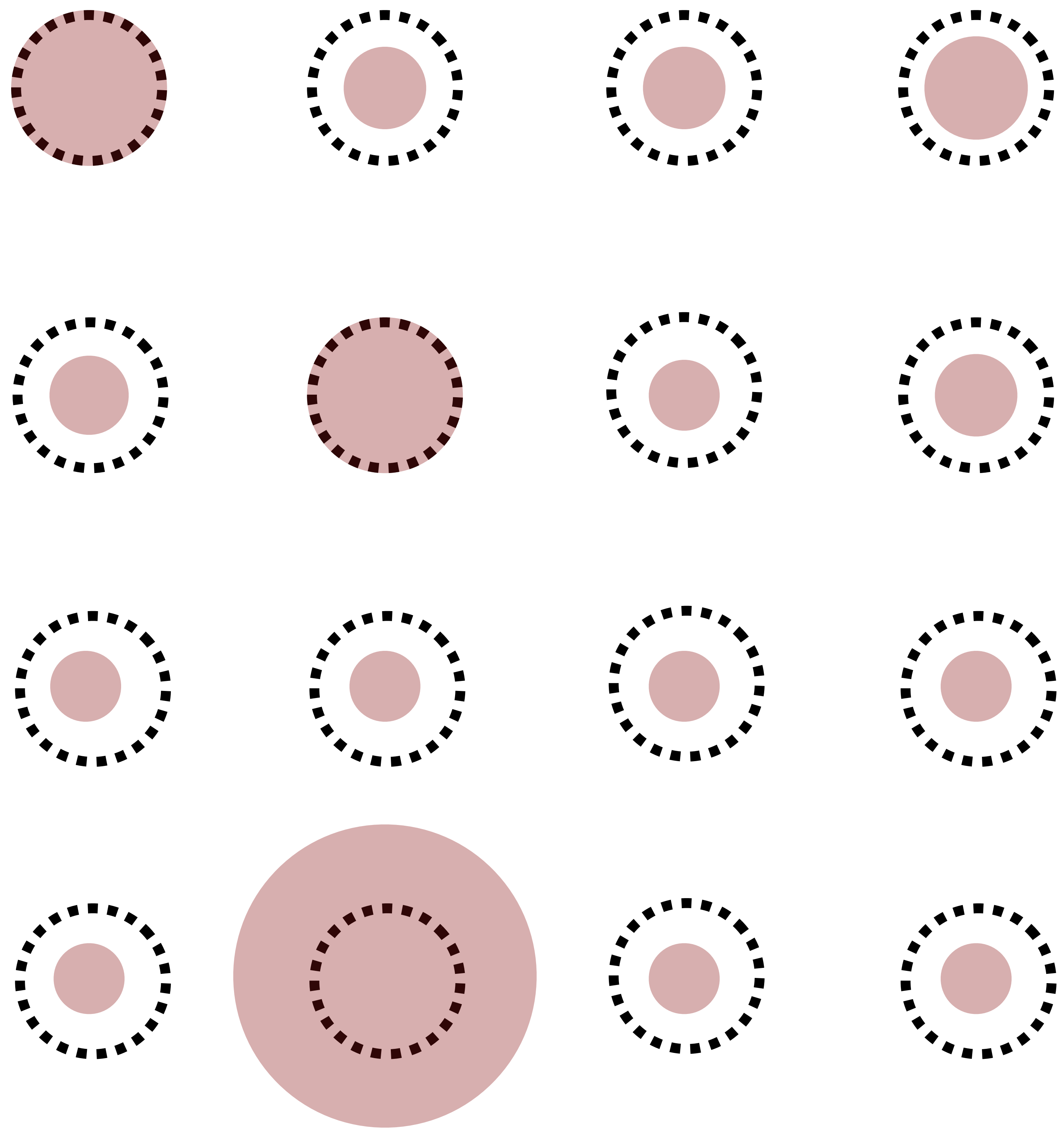


Proposal distribution q



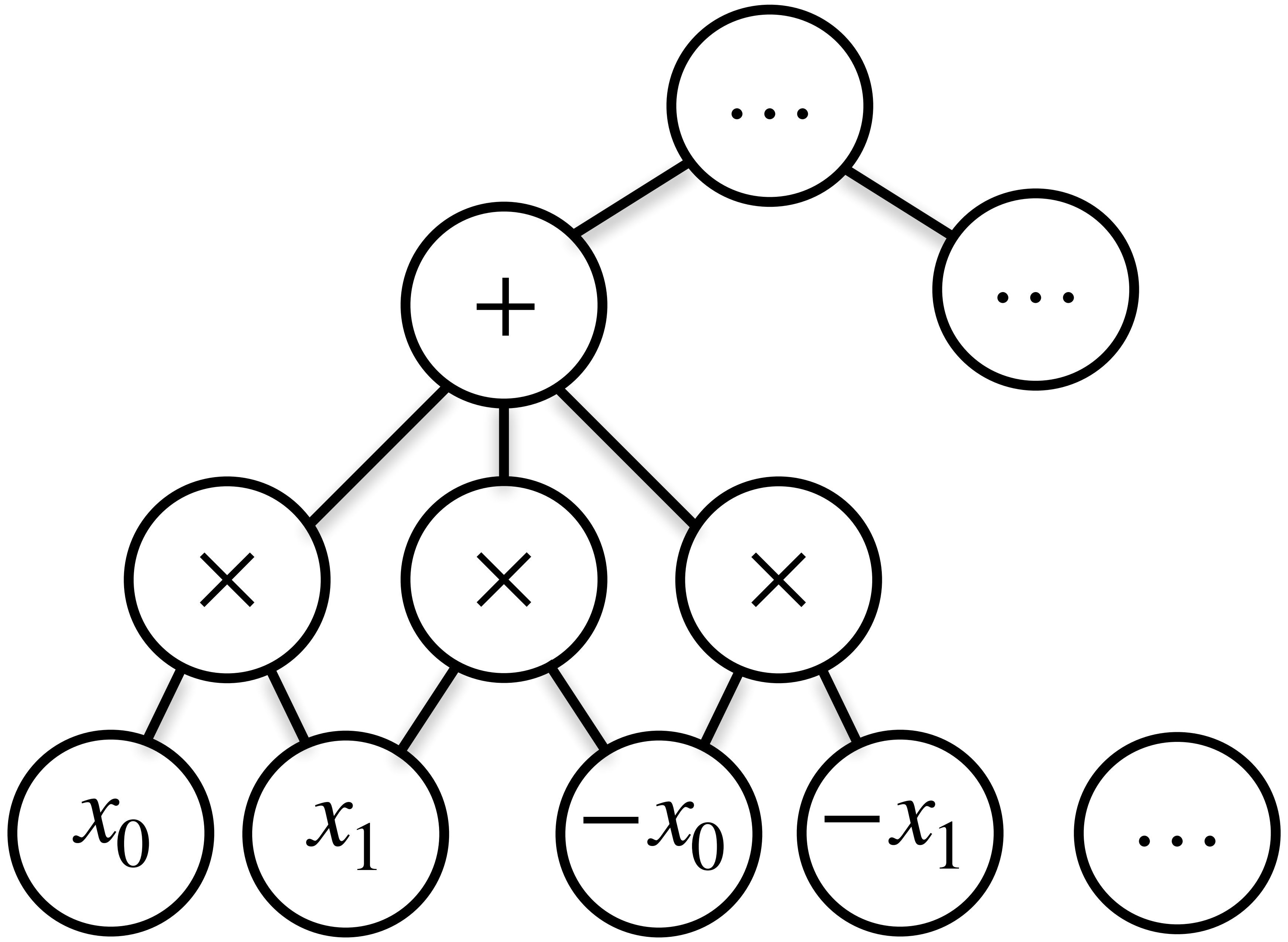
$$\log Z \geq \mathbb{E}_{\mathbf{x} \sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]$$

Sum Product Networks



exact gradients

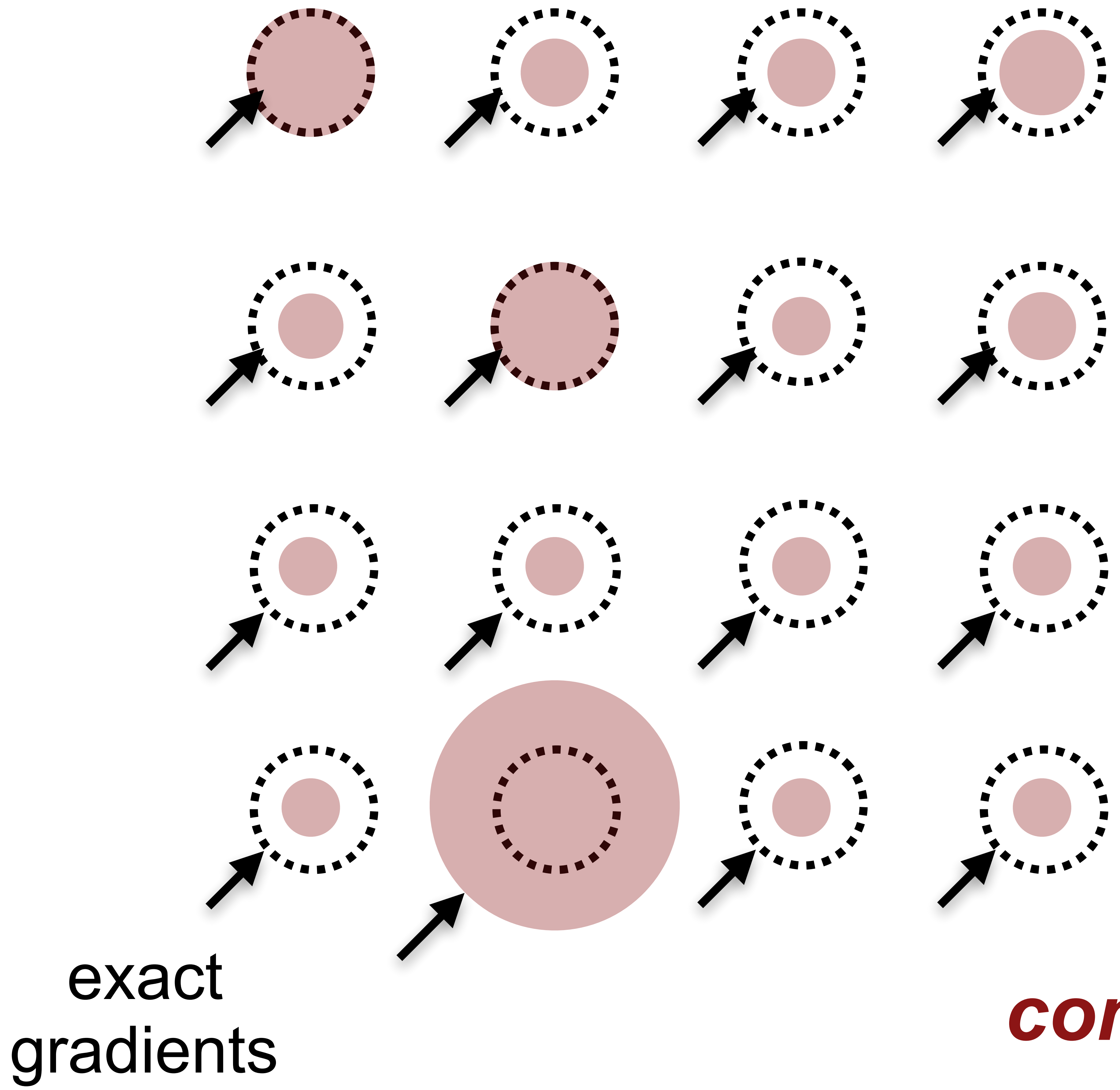
Proposal distribution q



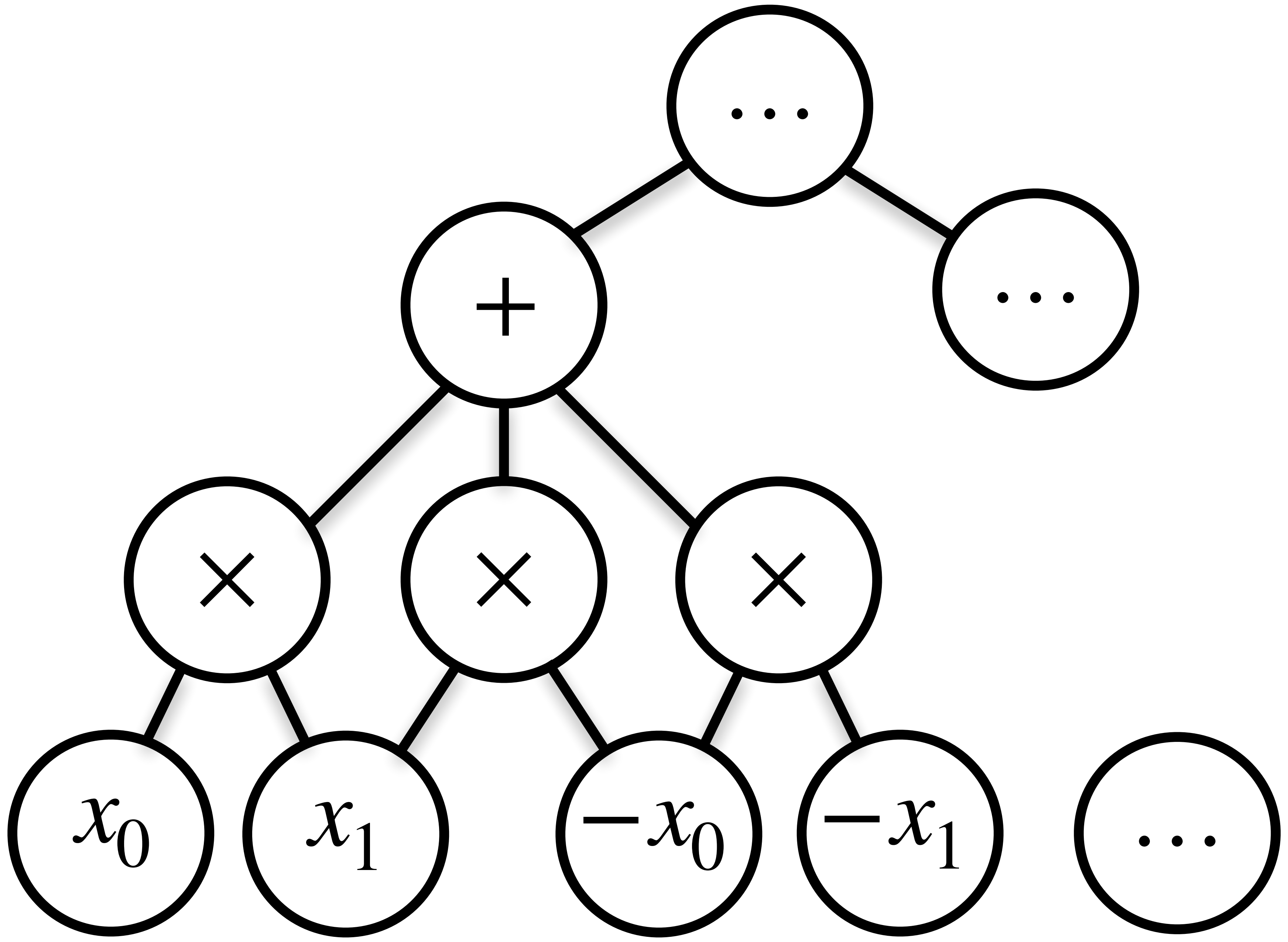
compute analytically!

$$\boxed{\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]}$$

Sum Product Networks



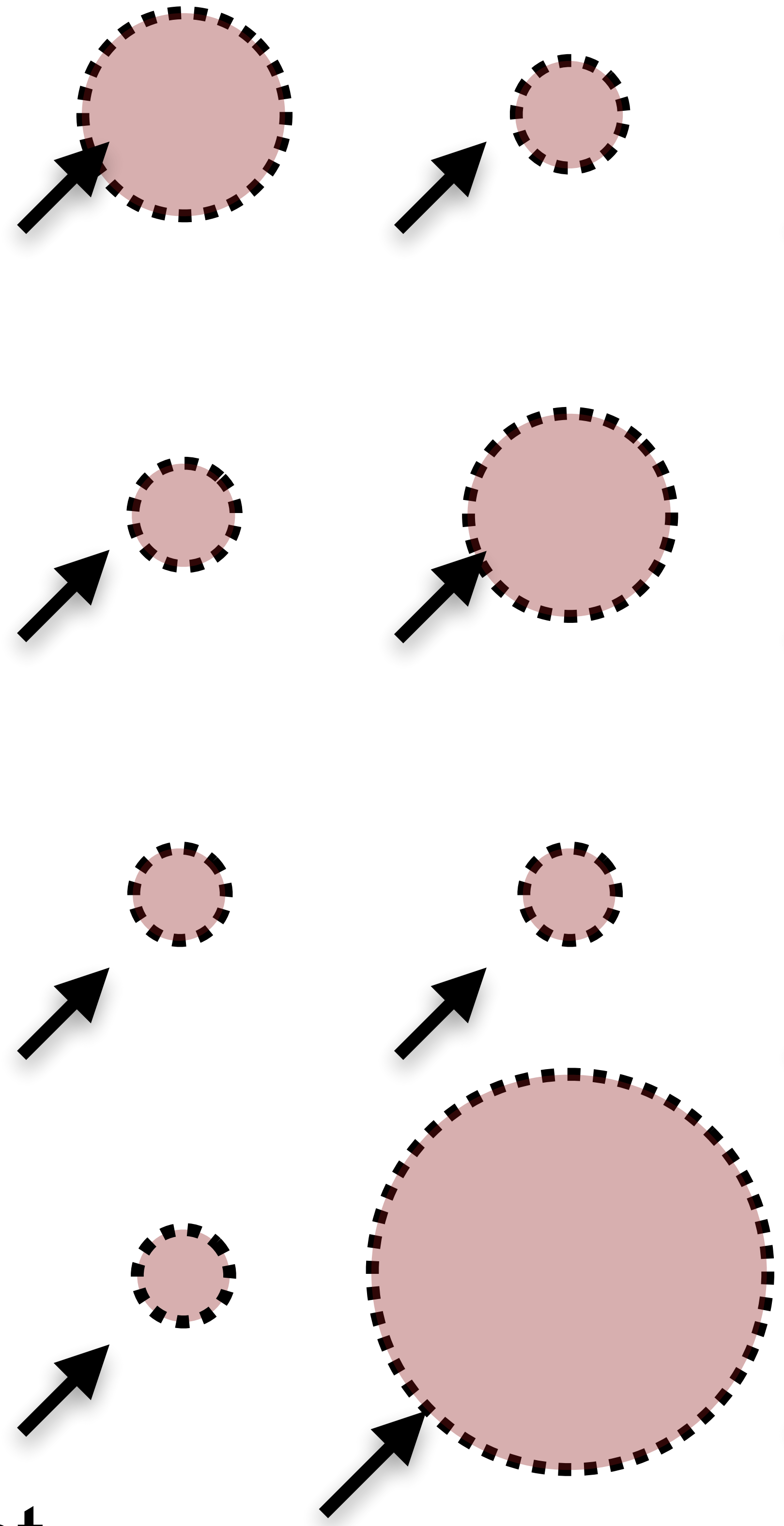
Proposal distribution q



compute analytically!

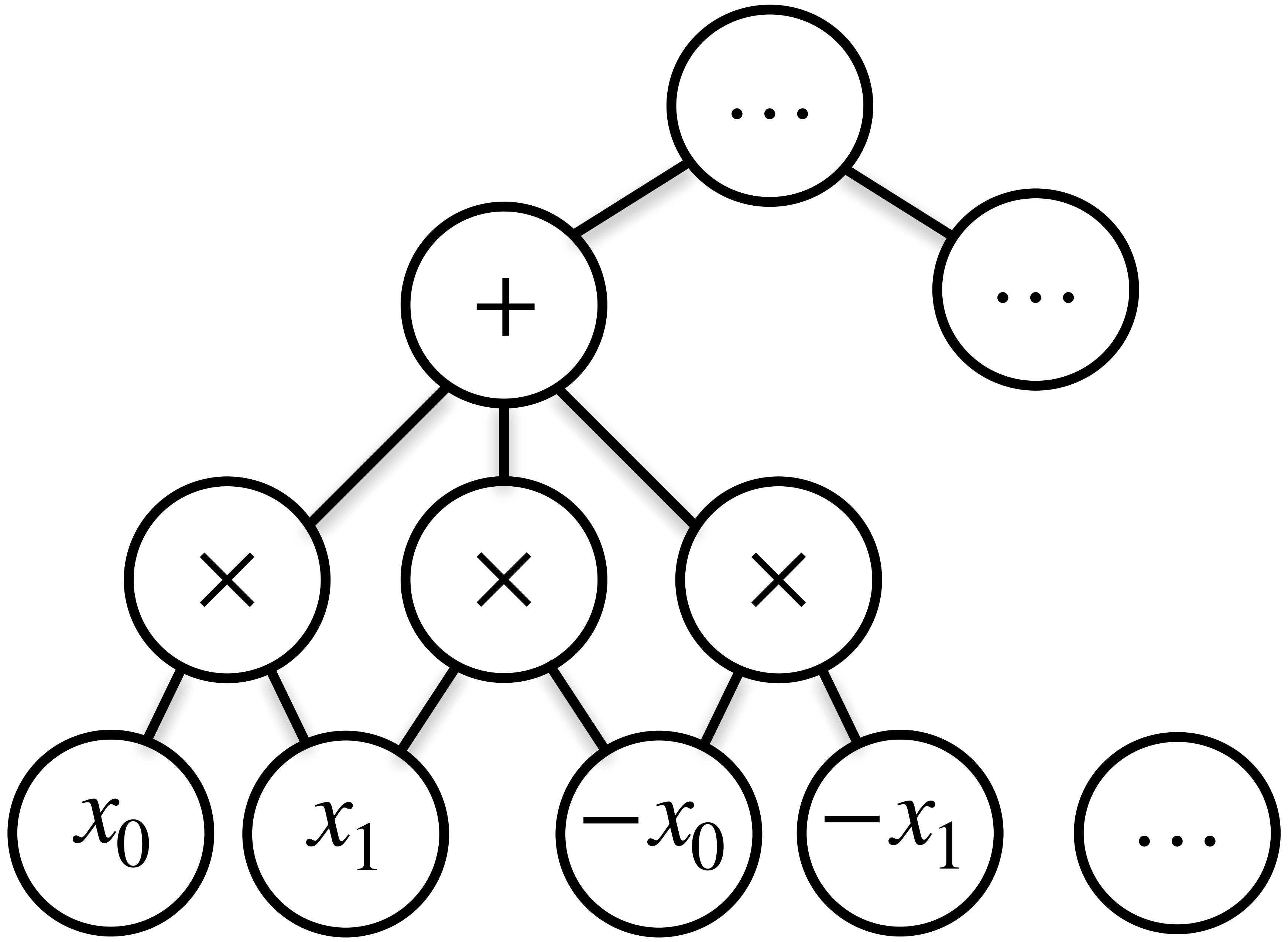
$$\boxed{\log Z \geq \mathbb{E}_{\mathbf{x} \sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]}$$

Sum Product Networks



exact
gradients

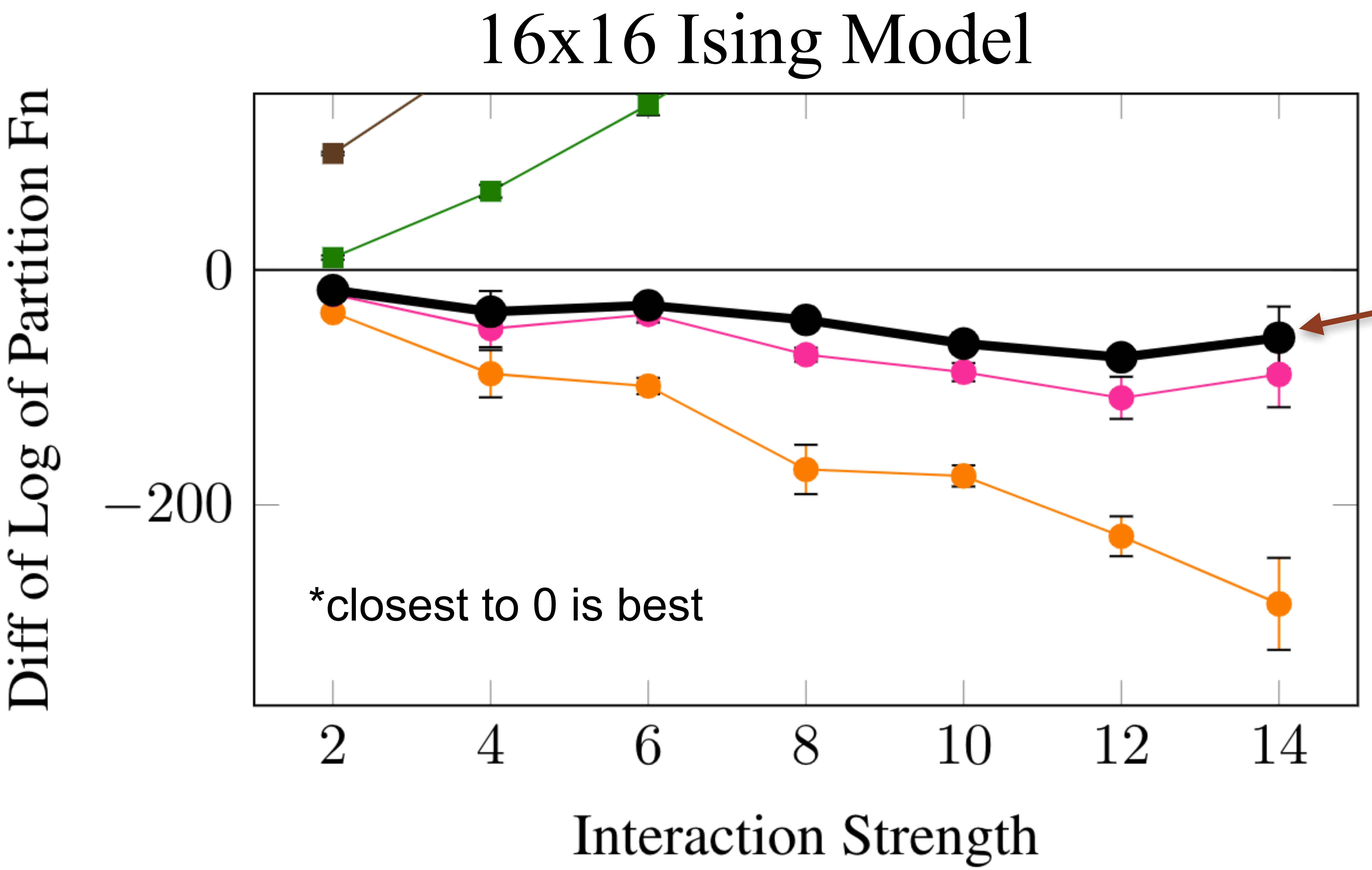
Proposal distribution q



compute analytically!

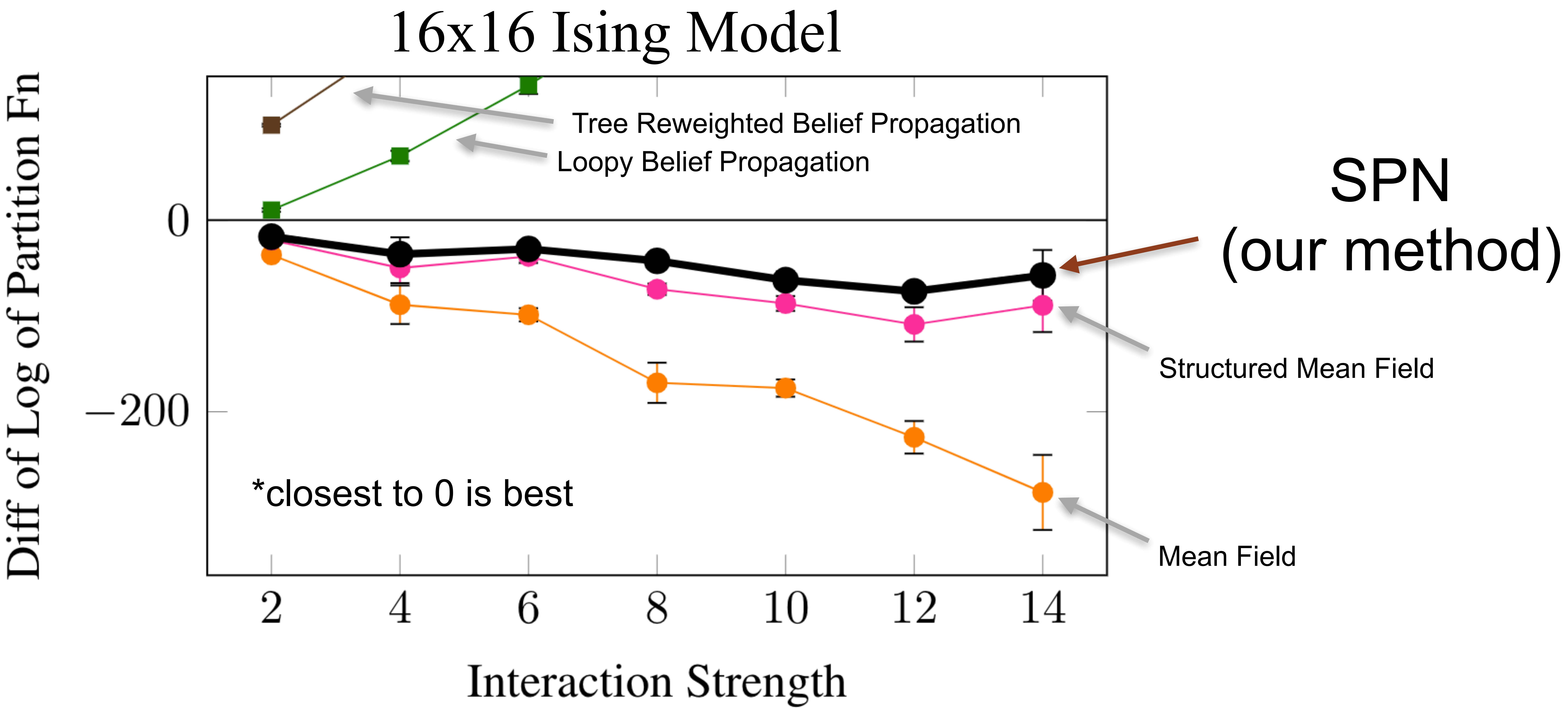
$$\boxed{\log Z \geq \mathbb{E}_{\mathbf{x}\sim q}[\log p(\mathbf{x}) - \log q(\mathbf{x})]}$$

Experiments



SPN
(our method)

Experiments



Summary

Discrete settings: sampling ✗

Summary

Discrete settings: sampling

Probabilistic Circuits (e.g. Sum Product Networks)

- Expressive family of distributions!
- Can compute gradients analytically — *no sampling!*

Thanks!

andyshih@cs.stanford.edu