Divide and Couple: Using Monte Carlo Variational Objectives for Posterior Approximation

Justin Domke and Daniel Sheldon

University of Massachusetts Amherst

Overview

- Variational inference gives both a *lower-bound* on the log-likelihood and an *approximate posterior*.
- Easy to get other lower-bounds. Do they also give approximate posteriors?
- This work: A general theory connecting likelihood bounds to posterior approximations.

Take $p(\boldsymbol{z}, \boldsymbol{x})$ with $\boldsymbol{x}$ fixed.

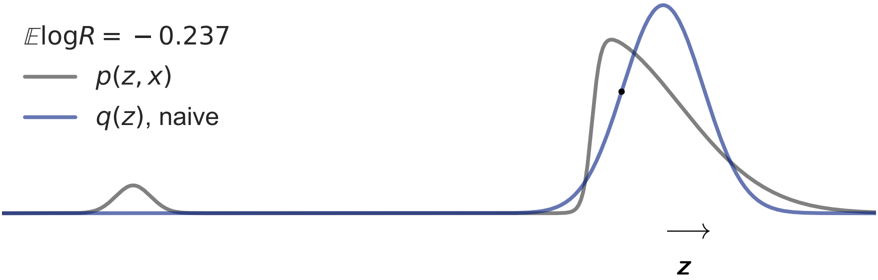
Take $p(\boldsymbol{z}, \boldsymbol{x})$ with $\boldsymbol{x}$ fixed.

Observation: If $\mathbb{E}\, R = p(\boldsymbol{x})$, then $\mathbb{E} \log R \leq \log p(\boldsymbol{x})$.

Take $p(\boldsymbol{z}, \boldsymbol{x})$ with $\boldsymbol{x}$ fixed.

Observation: If $\mathbb{E}\, R = p(\boldsymbol{x})$, then $\mathbb{E} \log R \leq \log p(\boldsymbol{x})$.

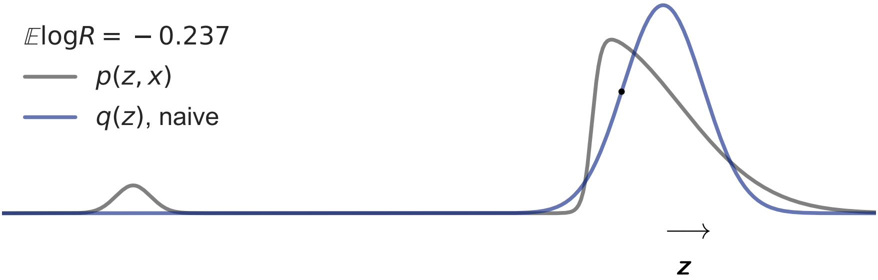
Example: Take $R = \frac{p(\boldsymbol{x},\boldsymbol{z})}{q(\boldsymbol{z})}$ for z $\sim q$ Gaussian, optimize q.



Take $p(\boldsymbol{z}, \boldsymbol{x})$ with $\boldsymbol{x}$ fixed.

Observation: If $\mathbb{E}\, R = p(\boldsymbol{x})$, then $\mathbb{E} \log R \leq \log p(\boldsymbol{x})$.

Example: Take $R = \frac{p(\boldsymbol{x}, \boldsymbol{z})}{q(\boldsymbol{z})}$ for z $\sim q$ Gaussian, optimize q.



Take $p(\boldsymbol{z}, \boldsymbol{x})$ with $\boldsymbol{x}$ fixed.

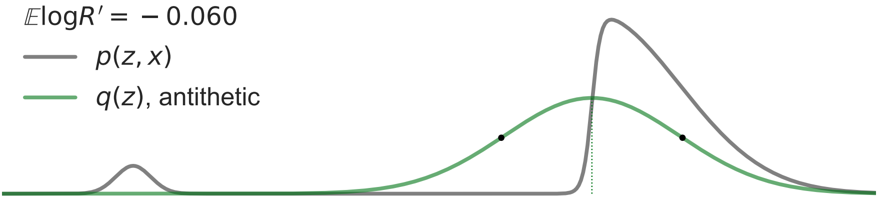
Observation: If $\mathbb{E}\, R = p(\boldsymbol{x})$, then $\mathbb{E} \log R \leq \log p(\boldsymbol{x})$.

Example: Take $R = \frac{p(\boldsymbol{x},\boldsymbol{z})}{q(\boldsymbol{z})}$ for z $\sim q$ Gaussian, optimize q.

Decomposition: $KL\left(q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})\right) = \log p(\boldsymbol{x}) - \mathbb{E} \log R$.

- Likelihood bound: ✓
- Posterior approximation: ✓

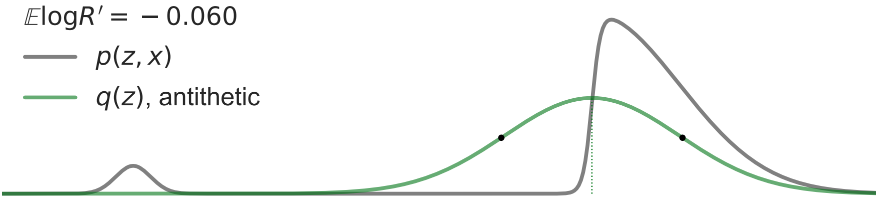
Recent work: Better Monte Carlo estimators R.



Recent work: Better Monte Carlo estimators R.

Antithetic Sampling: Let $T(\mathbf{z})$ "flip" $\mathbf{z}$ around mean of q.

$$R = \frac{1}{2}\left(\frac{p(\mathbf{z}, \mathbf{x}) + p(T(\mathbf{z}), \mathbf{x})}{q(\mathbf{z})}\right)$$

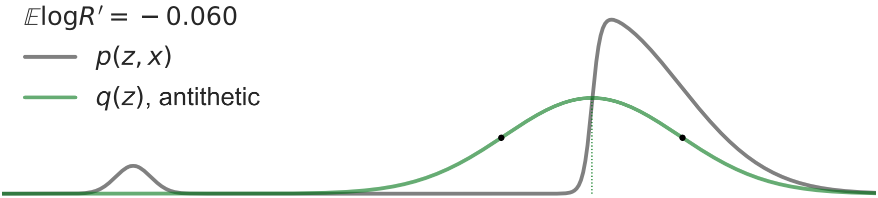


Recent work: Better Monte Carlo estimators R.

Antithetic Sampling: Let $T(\boldsymbol{z})$ "flip" $\boldsymbol{z}$ around mean of q.

$$R = \frac{1}{2}\left(\frac{p(\boldsymbol{z},\boldsymbol{x}) + p(T(\boldsymbol{z}),\boldsymbol{x})}{q(\boldsymbol{z})}\right)$$

- Likelihood bound: ✓
- Posterior approximation: ×



Recent work: Better Monte Carlo estimators R.

Antithetic Sampling: Let $T(\mathbf{z})$ "flip" $\mathbf{z}$ around mean of q.

$$R = \frac{1}{2}\left(\frac{p(\mathbf{z},\mathbf{x}) + p(T(\mathbf{z}),\mathbf{x})}{q(\mathbf{z})}\right)$$

- Likelihood bound: ✓
- Posterior approximation: ×

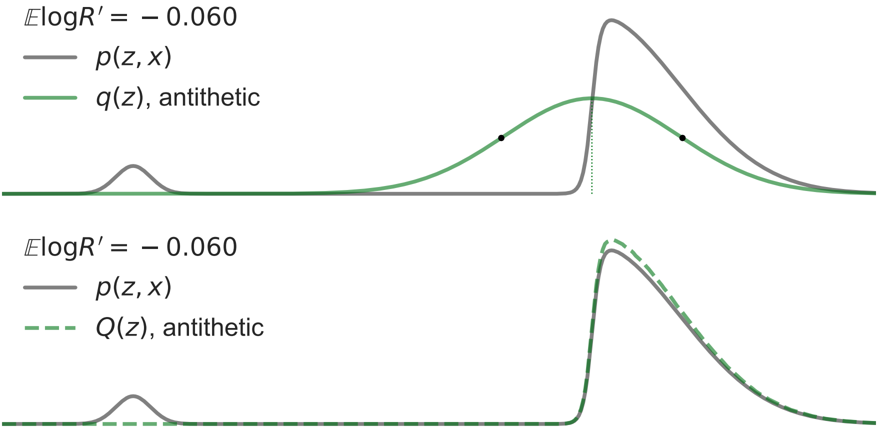
This paper: Is some *other* distribution close to p?

Contribution of this paper: Given estimator with $\mathbb{E}\, R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL\left(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})\right) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$

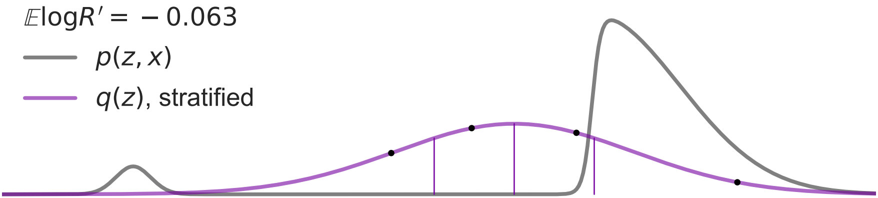
Contribution of this paper: Given estimator with $\mathbb{E}\, R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL\left(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})\right) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$



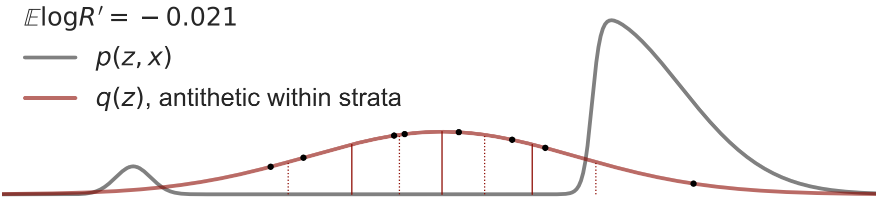
Contribution of this paper: Given estimator with $\mathbb{E}\,R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL\left(Q(\boldsymbol{z})\|p(\boldsymbol{z}|\boldsymbol{x})\right) \leq \log p(\boldsymbol{x}) - \mathbb{E}\log R.$$



Contribution of this paper: Given estimator with $\mathbb{E}\, R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$



Contribution of this paper: Given estimator with $\mathbb{E}\, R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$

How?

Contribution of this paper: Given estimator with $\mathbb{E}\, R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$

Unbiased estimator: Where is $\boldsymbol{z}$?

$$\mathop{\mathbb{E}}_{\boldsymbol{\omega}} R(\boldsymbol{\omega}) = p(\boldsymbol{x})$$

Contribution of this paper: Given estimator with $\mathbb{E} R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$

Unbiased estimator: Where is $\boldsymbol{z}$?

$$\mathbb{E}_{\boldsymbol{\omega}} R(\omega) = p(\boldsymbol{x})$$

We suggest: Need a *coupling*:

$$\mathbb{E}_{\boldsymbol{\omega}} R(\omega) \underbrace{a(\boldsymbol{z}|\boldsymbol{\omega})}_{\text{coupling}} = p(\boldsymbol{z}, \boldsymbol{x})$$

Contribution of this paper: Given estimator with $\mathbb{E}\,R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL\left(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})\right) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$

Unbiased estimator: Where is $\boldsymbol{z}$?

$$\mathop{\mathbb{E}}_{\boldsymbol{\omega}} R(\omega) = p(\boldsymbol{x})$$

We suggest: Need a *coupling*:

$$\mathop{\mathbb{E}}_{\boldsymbol{\omega}} R(\omega) \underbrace{a(\boldsymbol{z}|\boldsymbol{\omega})}_{\text{coupling}} = p(\boldsymbol{z}, \boldsymbol{x})$$

Then, exist augmented distributions s.t.

$$KL\left(Q(\boldsymbol{z}, \boldsymbol{\omega}) \| p(\boldsymbol{z}, \boldsymbol{\omega}|\boldsymbol{x})\right) = \log p(\boldsymbol{x}) - \mathbb{E} \log R$$

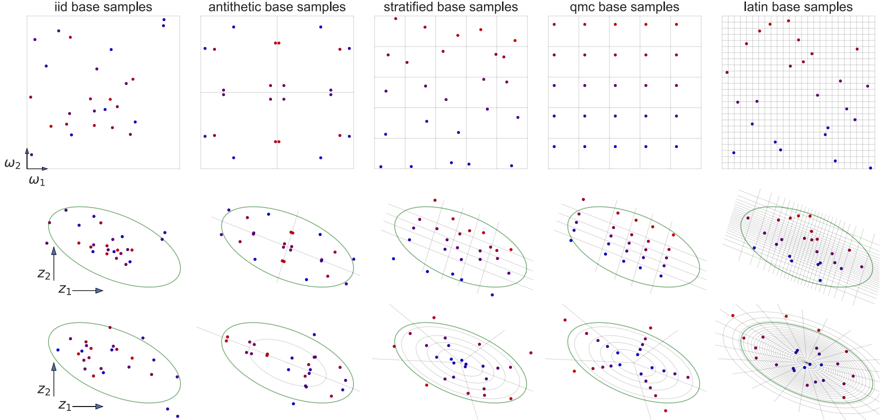
Contribution of this paper: Given estimator with $\mathbb{E}\,R = p(\boldsymbol{x})$, we show how to construct $Q(\boldsymbol{z})$ such that

$$KL(Q(\boldsymbol{z}) \| p(\boldsymbol{z}|\boldsymbol{x})) \leq \log p(\boldsymbol{x}) - \mathbb{E} \log R.$$

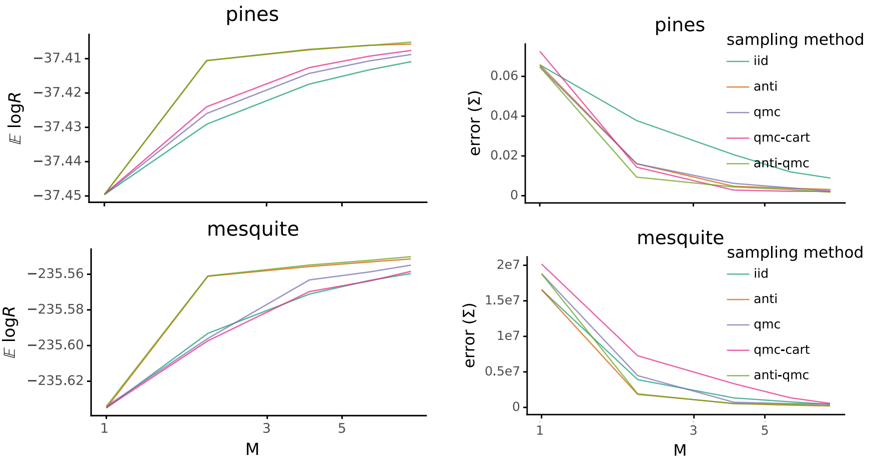
Summary:

- Tightening a bound $\log p(\boldsymbol{x}) - \mathbb{E} \log R$ is equivalent to VI in an *augmented state space* $(\boldsymbol{\omega}, \boldsymbol{z})$.
- To sample from $Q(\boldsymbol{z})$ draw ω then $\mathrm{z} \sim a(\mathrm{z}|\omega)$.
- Paper gives couplings for:
  - Antithetic sampling
  - Stratified sampling
  - Quasi Monte Carlo
  - Latin hypercube sampling
  - Arbitrary recursive combinations of above

Implementation: Different sampling methods with Gaussian q.



Experiments confirm: Better likelihood bounds ⇔ better posteriors



Poster: Tue Dec 10th, 5:30-7:30pm @ East Exhibition Hall B + C #166