Estimating Gradients for Discrete Distributions by Sampling Without Replacement

Wouter Kool, Herke van Hoof & Max Welling

International Conference on Learning Representations (ICLR) 2020

ORTEC OPTIMIZE YOUR WORLD

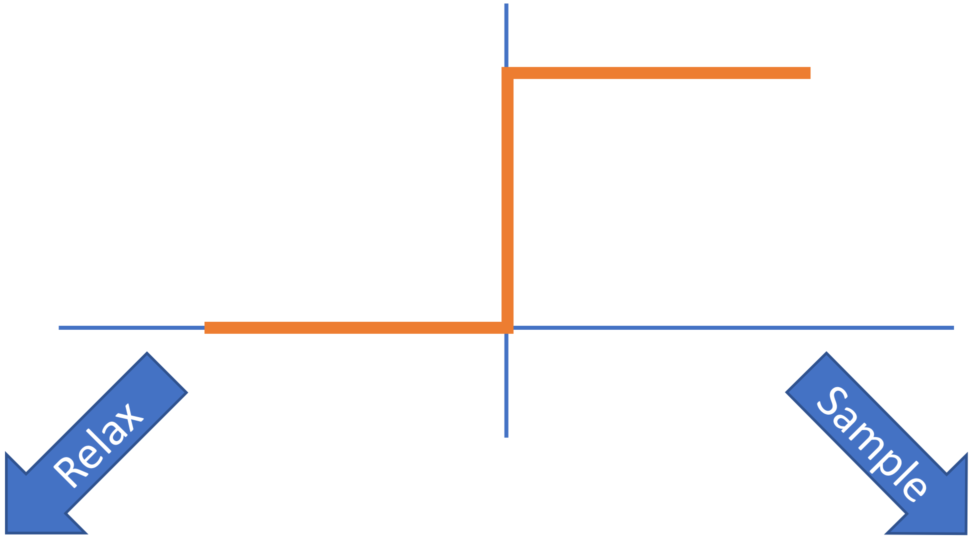
University of Amsterdam

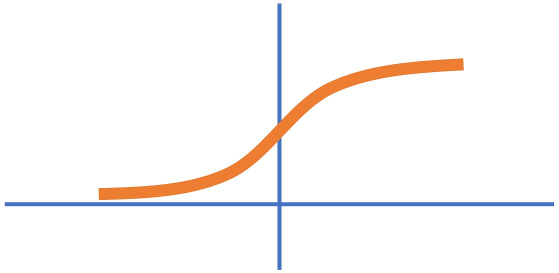
AMLAB Amsterdam Machine Learning Lab

Problems of discrete nature

- Reinforcement Learning
- Machine Translation / Image Captioning
- Discrete Latent Variable Modelling
- (Hard) Attention

Gradient of discrete operation





Biased

“REINFORCE”

$$\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(x)\, f(x)$$

High variance

REINFORCE

$$\nabla_{\boldsymbol{\theta}} E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(x)}[\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(x)\, f(x)]$$

REINFORCE

$$\nabla_{\boldsymbol{\theta}} E_{p_{\boldsymbol{\theta}}(x)}[f(x)] \approx \nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(x)\, f(x)$$

REINFORCE with multiple samples

$$\nabla_{\boldsymbol{\theta}} E_{p_{\boldsymbol{\theta}}(x)}[f(x)] \approx \frac{1}{k} \sum_{i=1}^{k} \nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(x_i)\, f(x_i)$$

REINFORCE with baseline

$$\nabla_{\boldsymbol{\theta}} E_{p_{\boldsymbol{\theta}}(x)}[f(x)] \approx \frac{1}{k} \sum_{i=1}^{k} \nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(x_i) \left(f(x_i) - \underbrace{\frac{\sum_{j \neq i} f(x_j)}{k-1}}_{\text{Baseline}} \right)$$

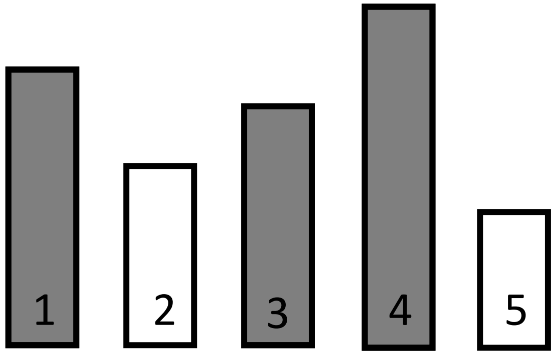
Mnih & Rezende, 2016

Sampling *without* replacement

Since duplicate samples are uninformative!

*In a deterministic setting

Sampling without replacement



$$B = (3,4,1)$$

$$p(B) = p(b_1) \times \frac{p(b_2)}{1 - p(b_1)} \times \frac{p(b_3)}{1 - p(b_1) - p(b_2)}$$

Ordered samples without replacement

$$p(B) = \prod_{i=1}^{k} \frac{p(b_i)}{1 - \sum_{j<i} p(b_j)}$$

Sequence $B = (3,4,1)$

Unordered samples without replacement

$$p(B) = \prod_{i=1}^{k} \frac{p(b_i)}{1 - \sum_{j<i} p\left(b_j\right)}$$

Set $S = \{1,3,4\}$

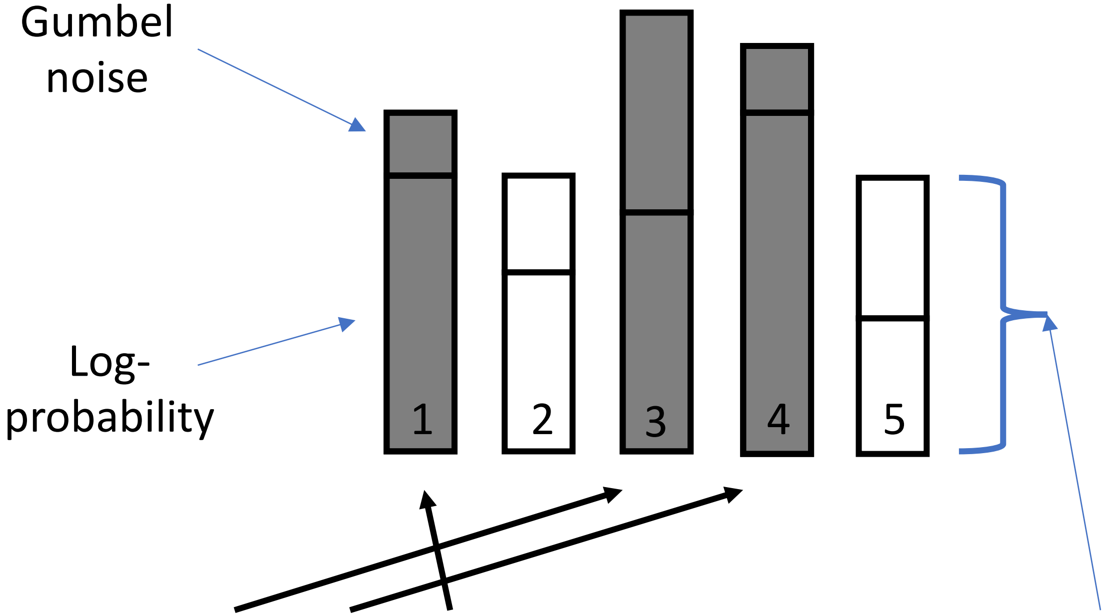
Unordered samples without replacement

$$p(S) = \sum_{B \in \mathcal{B}(S)} p(B) = \sum_{B \in \mathcal{B}(S)} \prod_{i=1}^{k} \frac{p(b_i)}{1 - \sum_{j<i} p(b_j)}$$

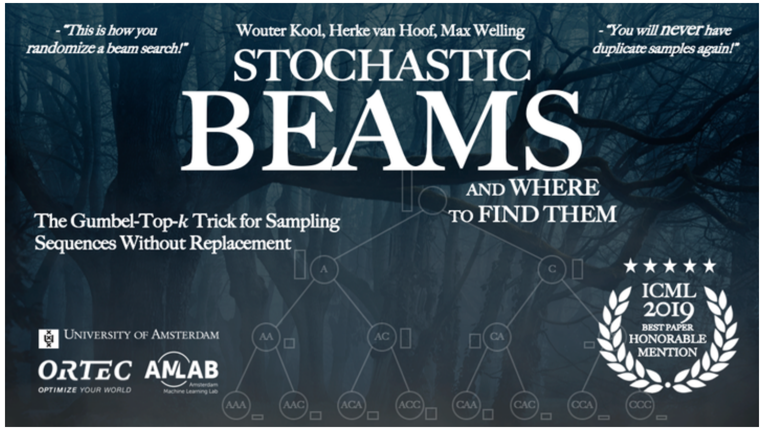
Set $S = \{1,3,4\}$

Sum over $k!$
permutations

Gumbel-Top-k sampling



$$B = b_1, \ldots, b_k = \underset{i}{\arg \operatorname{top} k} \, G_{\phi_i}$$



https://arxiv.org/abs/1903.06059

http://www.jmlr.org/papers/v21/19-985.html

$$B = (3,4,1)$$
$$S = \{1,3,4\}$$

Back to our problem

$$\nabla_{\boldsymbol{\theta}} E_{p_{\boldsymbol{\theta}}(x)}[f(x)]$$

Estimating the expectation

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)]$$

The single sample estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(B)}[f(b_1)]$$

Separating the expectation

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(B|S)}[f(b_1)]\right]$$

Set of unordered samples

Conditional distribution of their order

Separating the expectation

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)]\right]$$

Rao-Blackwellizing the estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)]\right]$$

$$E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)] = \sum_{s \in S} P(b_1 = s|S)\, f(s)$$

Rao-Blackwellizing the estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)]\right]$$

$$E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)] = \sum_{s \in S} P(b_1 = s|S) f(s)$$

$$P(b_1 = s|S) = \frac{P(S|b_1 = s)P(b_1 = s)}{P(S)}$$

Rao-Blackwellizing the estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)]\right]$$

$$E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)] = \sum_{s \in S} P(b_1 = s|S) f(s)$$

$$P(b_1 = s|S) = \underbrace{\frac{P(S|b_1 = s)}{P(S)}}_{\text{Leave-one-out ratio } R(S,s)} \underbrace{P(b_1 = s)}_{p_{\boldsymbol{\theta}}(s)}$$

Rao-Blackwellizing the estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)]\right]$$

$$E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)] = \sum_{s \in S} P(b_1 = s|S) f(s)$$

$$P(b_1 = s|S) = R(S, s) p_{\boldsymbol{\theta}}(s)$$

Rao-Blackwellizing the estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\left[E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)]\right]$$

$$E_{p_{\boldsymbol{\theta}}(b_1|S)}[f(b_1)] = \sum_{s \in S} R(S, s) p_{\boldsymbol{\theta}}(s) f(s)$$

Rao-Blackwellizing the estimator

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(S)}\Bigg[\underbrace{\sum_{s \in S} R(S,s)p_{\boldsymbol{\theta}}(s)f(s)}_{\text{Unordered set estimator}}\Bigg]$$

Murthy 1957

Combining with REINFORCE

$$E_{p_{\boldsymbol{\theta}}(x)}[f(x)]$$

$$= E_{p_{\boldsymbol{\theta}}(S)}\left[\sum\nolimits_{s \in S} R(S,s)p_{\boldsymbol{\theta}}(s)f(s)\right]$$

Combining with REINFORCE

$$E_{p_{\boldsymbol{\theta}}(x)}[\nabla_\theta \log p_{\boldsymbol{\theta}}(s)\, f(x)]$$

$$= E_{p_{\boldsymbol{\theta}}(S)}\left[\sum_{s \in S} R(S,s) p_{\boldsymbol{\theta}}(s) \nabla_\theta \log p_{\boldsymbol{\theta}}(s)\, f(s)\right]$$

Combining with REINFORCE

$$\nabla_\theta E_{p_{\boldsymbol{\theta}}(x)}[f(x)] = E_{p_{\boldsymbol{\theta}}(x)}[\nabla_\theta \log p_{\boldsymbol{\theta}}(s)\, f(x)]$$

$$= E_{p_{\boldsymbol{\theta}}(S)}\left[\sum_{s \in S} R(S,s) \underbrace{p_{\boldsymbol{\theta}}(s) \nabla_\theta \log p_{\boldsymbol{\theta}}(s)}_{\nabla_\theta p_{\boldsymbol{\theta}}(s)} f(s)\right]$$

Combining with REINFORCE

$$\nabla_\theta E_{p_{\boldsymbol{\theta}}(x)}[f(x)]$$

$$= E_{p_{\boldsymbol{\theta}}(S)}\left[\underbrace{\sum_{s \in S} R(S,s) \nabla_\theta p_{\boldsymbol{\theta}}(s) f(s)}_{\text{Unordered set policy gradient estimator}}\right]$$

Include a baseline

Second order
leave-one-out
ratio

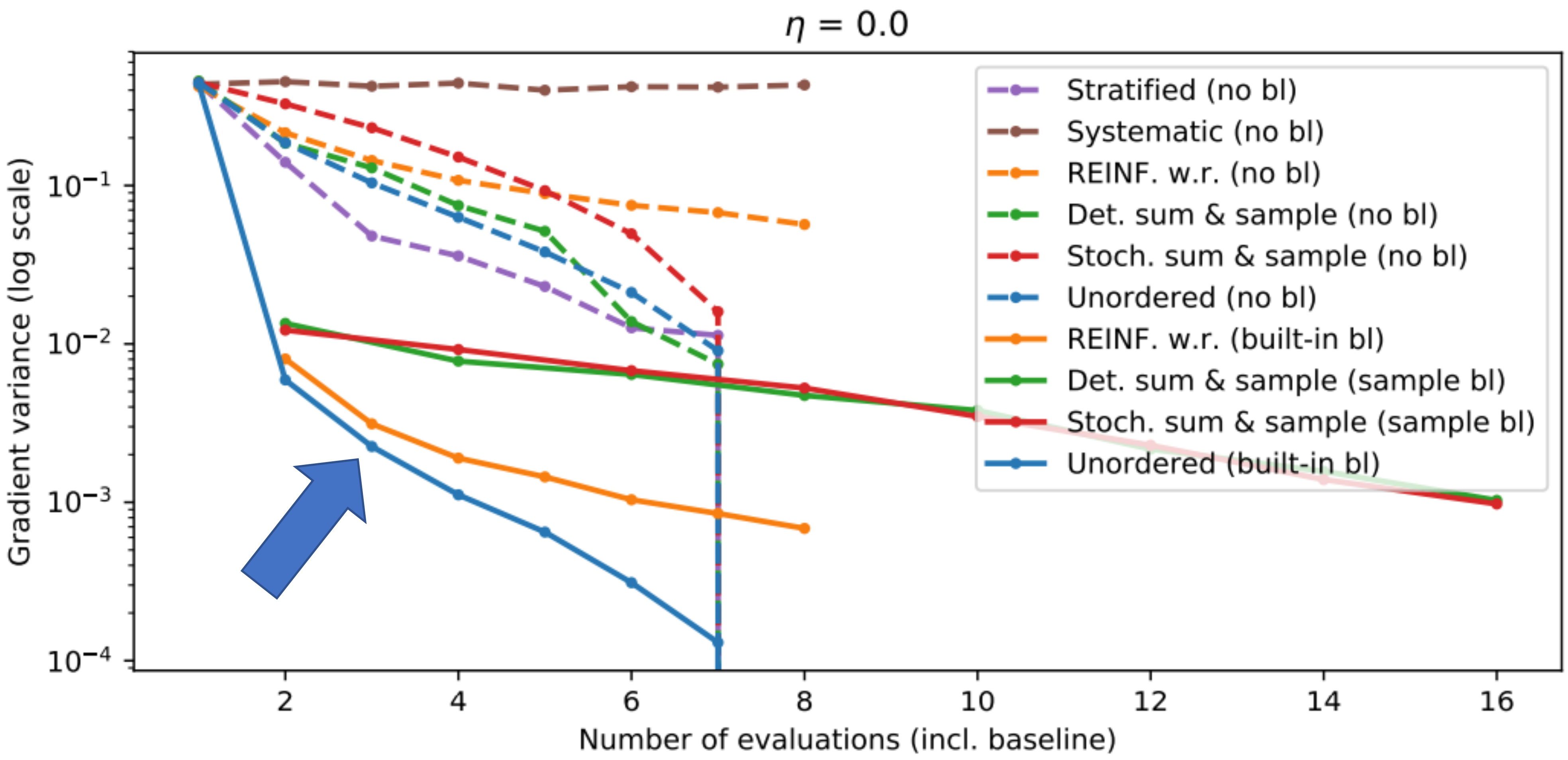
$$\nabla_{\theta} E_{p_{\boldsymbol{\theta}}(x)}[f(x)]$$

$$= E_{p_{\boldsymbol{\theta}}(S)}\left[\sum_{s \in S} R(S,s) \nabla_{\theta} p_{\boldsymbol{\theta}}(s) \left(f(s) - \underbrace{\sum_{s' \in S} R^{\setminus s}(S,s') p_{\boldsymbol{\theta}}(s') f(s')}_{\text{'Baseline'}} \right)\right]$$

Unbiased!

Experiments

Bernoulli gradient variance



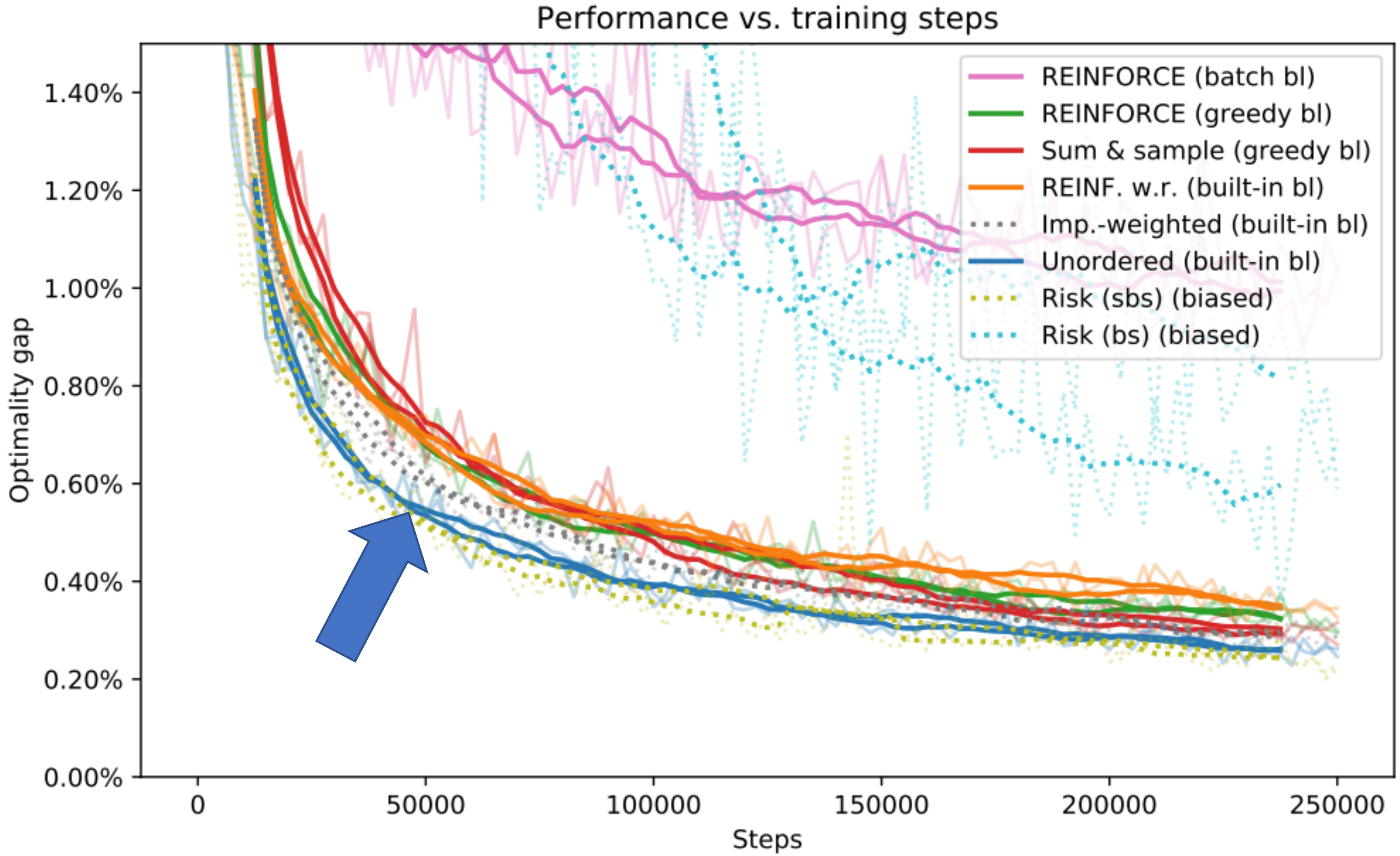
(a) High entropy ($\eta = 0$)

Categorical Variational Auto-Encoder (grad. var.)

Table 1: VAE gradient log-variance of different unbiased estimators with $k = 4$ samples.

Domain	ARSM	RELAX	REINFORCE (no bl)	REINFORCE (sample bl)	Sum & sample (no bl)	Sum & sample (sample bl)	REINF. w.r. (built-in bl)	Unordered (built-in bl)
Small 10^2	13.45	11.67	11.52	7.49	**6.29**	**6.29**	6.65	**6.29**
Large 10^{20}	15.55	15.86	13.81	8.48	13.77	8.44	**7.06**	**7.05**

Travelling Salesman Problem



Take away

The unordered set estimator

- Low-variance
- Unbiased
- Alternative to Gumbel-Softmax

End of story