ARMS: Antithetic-REINFORCE-Multi-Sample Gradient for Binary Variables

Alek Dimitriev
Mingyuan Zhou

University of Texas at Austin, McCombs School of Business

ICML
International Conference On Machine Learning

Table of Contents

Monte Carlo objectives

Goal: expectation based objectives of the form $\mathbb{E}_{p_\phi(x)}[f(x)]$

Can be found in many areas:

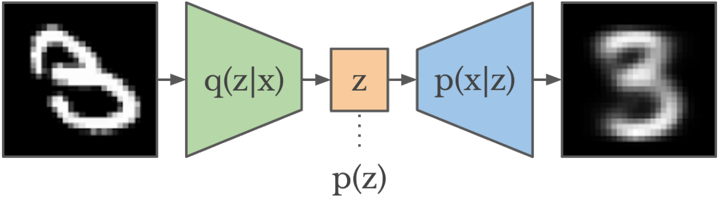
- Variational inference: $\mathbb{E}_{q_\phi(z|x)}[\ln p(x,z) - \ln q_\phi(z|x)]$
- Reinforcement learning: $\mathbb{E}_{\pi_\phi(x)}[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)]$
- Finance (options pricing): $\mathbb{E}_{p_{s_0}(s_T)}[e^{-\gamma T} \max\{s_T - K, 0\}]$
- Operations research (discrete queuing): $\mathbb{E}_{p_\phi(y_{1:T})}\left[\frac{\sum_{t=1}^{T} L_t(y_{1:t})}{\tau(y_{1:T})}\right]$
- Experimental design: $\mathbb{E}_{p_\phi(y)}[\mathbb{1}_{y<y_{\text{best}}}]$
- Et al. SDEs, GANs, bandits and online learning, econometrics, instrumental variables, counterfactual reasoning, ...

Discrete variational autoencoder

- Maximize evidence lower bound (ELBO):

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\ln p_\theta(\boldsymbol{z}, \boldsymbol{x}) - \ln q_\phi(\boldsymbol{z}|\boldsymbol{x})] \leq \ln p(\boldsymbol{x})$$

- Equivalent to minimizing *KL* divergence $\text{KL}(q(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z}|\boldsymbol{x})$
- $q(\boldsymbol{z}|\boldsymbol{x})$ is called an encoder, usually deep neural network $\boldsymbol{x} \to \phi$
- $p(\boldsymbol{x}|\boldsymbol{z})$ is a decoder, also neural network
- $p(\boldsymbol{z})$ is a prior distribution
- In a discrete VAE $q(\boldsymbol{z}|\boldsymbol{x}) = \prod_{d=1}^{D} \text{Bern}(z_d|\sigma(\phi_d))$, with logits ϕ



Monte Carlo gradients

- Why Monte Carlo gradients?
- Most expectations are too complicated to integrate
- Why the score function (REINFORCE) gradient?
- Most general, works with almost any distribution:

$$\nabla_\phi \mathbb{E}_{p_\phi(x)}[f(x)] = \mathbb{E}_{p_\phi(x)}[f(x)\nabla_\phi \ln p_\phi(x)]$$

- Unbiased: $E[g_{\mathrm{RF}}(x)] = E[f(x)\nabla_\phi \ln p_\phi(x)] = \nabla_\phi$

Table of Contents

Estimators for binary variables

- General REINFORCE: $f(z)\nabla \ln p_\phi(z)$. What if z is binary?
- Let $b \sim \text{Bern}(p)$, $p = \sigma(\phi)$.
- REINFORCE: $g_{\text{RF}} = f(b)(b - p)$.
- What if we have n independent samples?
- LOORF (Leave One Out REINFORCE):

$$g_{\text{LOORF}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(f(b_i) - \frac{1}{n}\sum_{j=1}^{n} f(b_j)\right)(b_i - p)$$

Antithetic estimators for binary variables

- What if we want to use antithetic (negatively correlated) pairs?
- If $u \sim \text{Unif}(0,1)$, then $\mathbb{1}_{u<p} \sim \text{Bern}(p)$.
- ARM: $g_{\text{ARM}} = \left(f(\mathbb{1}_{u<p}) - f(\mathbb{1}_{u>1-p})\right)\left(\frac{1}{2} - u\right)$.
- Let $b = \mathbb{1}_{u<p},\ b' = \mathbb{1}_{u>1-p}$.
- DisARM/U2G noticed randomness can be integrated out:

$$g_{\text{DisARM}} = \tfrac{1}{2}\big(f(b) - f(b')\big)(b - b')\max(p, 1-p)$$

Copulas

- n-dimensional or multivariate probability distribution with uniform marginals.
- $\boldsymbol{u} = (u_1, ..., u_n) \sim \mathcal{C}_n$, such that $\forall i : u_i \sim \text{Unif}(0, 1)$.
- How to create a copula?
- Start with some multivariate distribution $\mathcal{M}$.
- Calculate all marginal CDFs: $\forall i : F_{x_i}(x)$.
- Apply them to any sample: $\boldsymbol{x} = (x_1, ..., x_n) \sim \mathcal{M}$.
- $\boldsymbol{u} = (u_1, .., u_n)$, where $u_i = F_{x_i}(x_i)$, is a copula sample.

Table of Contents

Approach

- n independent samples (LOORF) perform better than $n/2$ antithetic pairs (DisARM)
- Is it possible to combine the two approaches?
- Yes! If we can:
- Sample n antithetic variables
- Debias the estimator

Similarities

- LOORF for $n = 2$? $b, b' \sim \text{Bern}(p)$

$$g_{\text{2-LOORF}} = \frac{1}{2}\Big(f(b) - f(b')\Big)(b - b')$$

- DisARM for $b = \mathbb{1}_{u<p}$, $b' = \mathbb{1}_{(1-u)<p}$

$$g_{\text{DisARM}} = \frac{1}{2}\Big(f(b) - f(b')\Big)(b - b')\max(p, 1 - p)$$

- Two extremes: no correlation $\longleftrightarrow$ minimal correlation (antithetic).
- Can we generalize to an arbitrarily dependent Bernoulli pair? Yes!

Antithetic Reinforce Two Sample (ARTS) Estimator

- Let $(b, b') \sim \mathcal{B}_2(p)$ denote a sample from a bivariate Bernoulli distribution with correlation $\rho = \text{corr}(b, b')$.
- An unbiased estimator is:

$$g_{\text{ARTS}} = \frac{g_{\text{2LOORF}}}{1-\rho} = \frac{1}{2}\Big(f(b) - f(b')\Big)(b - b')\frac{1}{1-\rho}$$

- If $\rho = 0$, we obtain two sample LOORF.
- What is the lowest possible correlation for a Bernoulli pair b, b'?
- $\rho = -\min(\frac{p}{1-p}, \frac{1-p}{p})$, which results in DisARM.

How to go from two to n samples?

- Key observation: LOORF is exactly the same as averaging all $\binom{n}{2}$ pairs!

$$g_{\text{LOORF}}(b_1, ..., b_n) = \frac{1}{n}\sum_{i=1}^{n}\left(f(b_i) - \frac{1}{n}\sum_{j=1}^{n} f(b_j)\right)\nabla_\phi \ln p(b_i)$$
$$= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{2}\Big(f(b_i) - f(b_j)\Big)\Big(\nabla_\phi \ln p(b_i) - \nabla_\phi \ln p(b_j)\Big)$$
$$= \frac{1}{n(n-1)}\sum_{i\neq j} g_{\text{2-LOORF}}(b_i, b_j),$$

- Idea: what if the debiasing term is identical for all pairs?

ARMS

- Assume we have n Bernoulli variables, with $\rho = \text{corr}(b_i, b_j)$, $\forall i \neq j$.
- An unbiased gradient is:

$$g_{\text{ARMS}} = \frac{g_{\text{LOORF}}}{1-\rho} = \frac{1}{n-1}\sum_{i=1}^{n}\left(f(b_i) - \frac{1}{n}\sum_{j=1}^{n} f(b_j)\right)\frac{b_i - p}{1-\rho}$$

- Last hurdle: how to sample n antithetic Bernoulli?
- If we have n antithetic uniform variables.
- Then $b_i = \mathbb{1}_{u_i < p}$ are antithetic Bernoulli.
- How to obtain n antithetic uniform variables? Copulas!
- (Sidenote: must also be able to calculate *rho*)

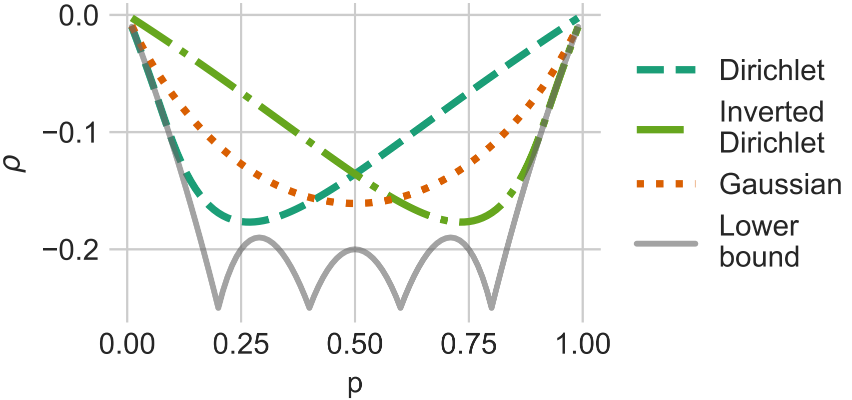
Antithetic Gaussian copula

- Gaussian copula: numerical marginal and bivariate CDFs.
- Sample $(x_1, ..., x_n) \sim \mathcal{N}(0, \Sigma)$, with $\Sigma_{ii} = 1$, $\Sigma_{ij} = -1/(n-1)$.
- Let $u_i = \Phi(x_i)$, where $\Phi(x)$ is the standard Gaussian CDF.
- $(u_1, ..., u_n)$ is a copula with pairwise correlation close to to ρ_{min}.
- The correlation between to Bernoulli variables is: $\rho = \frac{\mathbb{E}[b_i b_j] - p^2}{p(1-p)}$
- $E[b_i b_j] = P(b_i = b_j = 1) = P(u_i < p, u_j < p) = \Phi(p, p)$.

Antithetic Dirichlet copula

- When $\alpha = 1$, we have both a analytical marginal and bivariate CDF.
- Sample $(d_1, ..., d_n) \sim \text{Dir}(1, ..., 1)$.
- Can alternatively do the following:
- 1. $v_i \sim \text{Unif}(0, 1),\ i = 1 \dots n$
- 2. $d_i = \ln(v_i)/\sum_{j=1}^{n} \ln(v_j)$
- $u_i = 1 - (1 - d_i)^{n-1}$
- $\rho = \frac{\max(0, 2(1-p)^{\frac{1}{n-1}} - 1)^{n-1} - (1-p)^2}{p(1-p)}$

How good are the correlations?



Putting it all together

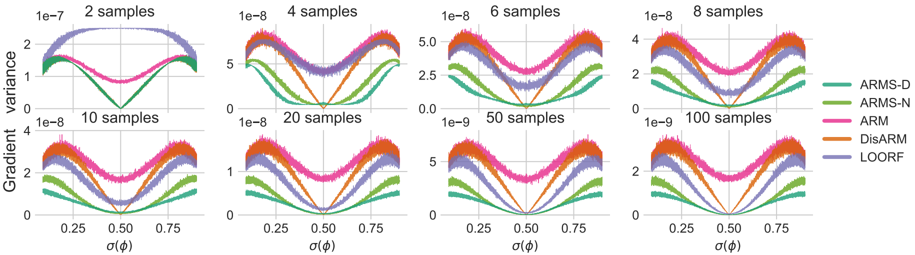
1. Sample n antithetic uniform variables (either Dirichlet or Gaussian)
2. Transform to Bernoulli $b_i = \mathbb{1}_{u_i < p}$ and calculate ρ
3. Obtain an unbiased estimator ARMS:

$$g_{\text{ARMS}} = \frac{1}{n-1} \sum_{i=1}^{n} \left(f(b_i) - \frac{1}{n} \sum_{j=1}^{n} f(b_j) \right) \frac{b_i - p}{1 - \rho}$$

Table of Contents

Toy example

- Maximize: $\mathcal{E}(\phi) = \mathbb{E}_b[(b - 0.499)^2], \quad b \sim \text{Bern}(\sigma(\phi))$
- Below: variance (all already unbiased) of each gradient as the function is maximized from $p_{init} = 0.05$ to $p_{end} = 0.95$.



Training ELBO

Dataset	Model	Samples	ARMS-D	ARMS-N	LOORF	DisARM	RELAX
Dynamic MNIST	Linear	4	-112.13 ± 0.10	**-111.96 ± 0.09**	-112.32 ± 0.04	-113.26 ± 0.05	-112.98 ± 0.25
		6	-111.03 ± 0.02	**-110.89 ± 0.07**	-110.99 ± 0.07	-112.11 ± 0.03	-111.46 ± 0.06
		8	**-110.30 ± 0.04**	-110.62 ± 0.06	-110.42 ± 0.04	-111.78 ± 0.07	-110.58 ± 0.01
		10	**-110.08 ± 0.05**	-110.14 ± 0.09	-110.17 ± 0.04	-111.08 ± 0.11	-110.17 ± 0.09
	Nonlinr	4	**-98.65 ± 0.16**	-98.97 ± 0.13	**-98.62 ± 0.05**	-100.45 ± 0.16	-100.52 ± 0.08
		6	-98.53 ± 0.13	**-97.87 ± 0.01**	-98.14 ± 0.18	-99.28 ± 0.11	-99.17 ± 0.17
		8	**-97.90 ± 0.12**	**-97.89 ± 0.10**	-98.14 ± 0.21	-98.69 ± 0.21	-98.80 ± 0.02
		10	-97.64 ± 0.06	**-97.32 ± 0.11**	-97.50 ± 0.29	-98.62 ± 0.12	-98.69 ± 0.07
Fashion MNIST	Linear	4	**-252.56 ± 0.11**	-252.69 ± 0.06	-252.71 ± 0.09	-254.02 ± 0.05	-253.53 ± 0.06
		6	-251.94 ± 0.13	**-251.73 ± 0.05**	-252.03 ± 0.08	-252.97 ± 0.06	-252.31 ± 0.14
		8	**-251.32 ± 0.11**	**-251.11 ± 0.23**	-251.41 ± 0.10	-252.57 ± 0.05	-251.36 ± 0.08
		10	-251.29 ± 0.02	**-251.08 ± 0.08**	-251.26 ± 0.03	-251.75 ± 0.21	**-251.16 ± 0.06**
	Nonlinr	4	**-235.65 ± 0.12**	-235.75 ± 0.06	-235.80 ± 0.07	-236.54 ± 0.06	-236.77 ± 0.03
		6	-235.47 ± 0.19	**-235.36 ± 0.08**	-235.70 ± 0.13	-235.94 ± 0.05	-236.20 ± 0.25
		8	-235.41 ± 0.10	**-235.19 ± 0.14**	-235.40 ± 0.13	-235.62 ± 0.16	-235.70 ± 0.14
		10	**-235.18 ± 0.11**	-235.32 ± 0.05	-235.59 ± 0.01	-235.60 ± 0.09	-235.46 ± 0.19
Omniglot	Linear	4	**-118.25 ± 0.08**	**-118.27 ± 0.05**	-118.41 ± 0.07	-119.24 ± 0.17	-118.75 ± 0.08
		6	**-117.62 ± 0.01**	**-117.62 ± 0.04**	-117.75 ± 0.08	-118.47 ± 0.12	-117.90 ± 0.03
		8	**-117.60 ± 0.05**	-117.66 ± 0.12	-117.74 ± 0.10	-118.41 ± 0.10	-117.71 ± 0.02
		10	**-117.03 ± 0.09**	**-116.99 ± 0.04**	-117.21 ± 0.08	-117.70 ± 0.01	-117.13 ± 0.05
	Nonlinr	4	-112.09 ± 0.27	**-112.03 ± 0.12**	-112.20 ± 0.26	-113.24 ± 0.16	-114.08 ± 0.35
		6	-111.50 ± 0.06	**-111.39 ± 0.10**	**-111.26 ± 0.15**	-112.30 ± 0.05	-113.71 ± 0.13
		8	**-110.91 ± 0.04**	-111.01 ± 0.06	**-110.85 ± 0.35**	-111.82 ± 0.09	-113.64 ± 0.10
		10	**-110.66 ± 0.05**	-110.79 ± 0.26	-110.79 ± 0.20	-111.33 ± 0.19	-114.00 ± 0.10

Conclusion and future work

- ARMS generalizes n iid samples (LOORF) and two antithetic samples (DisARM)
- Future work: extension to categorical variables, IWAE bound with antithetic samples, ...

References

Dong, Zhe, Andriy Mnih, and George Tucker. "DisARM: An Antithetic Gradient Estimator for Binary Latent Variables". In: *Advances in Neural Information Processing Systems 33*. 2020.

Kool, Wouter, Herke van Hoof, and Max Welling. "Buy 4 REINFORCE Samples, Get a Baseline for Free!" In: *Workshop, Deep Reinforcement Learning Meets Structured Prediction, ICLR*. 2019.

Mohamed, Shakir et al. "Monte Carlo Gradient Estimation in Machine Learning". In: *J. Mach. Learn. Res.* 21 (2020), 132:1–132:62.

Salimans, Tim and David A Knowles. "On using control variates with stochastic approximation for variational bayes and its connection to stochastic linear regression". In: *arXiv preprint arXiv:1401.1022* (2014).

Williams, Ronald J. "Simple statistical gradient-following algorithms for connectionist reinforcement learning". In: *Machine learning* 8.3-4 (1992), pp. 229–256.

Yin, Mingzhang et al. "Probabilistic Best Subset Selection by Gradient-Based Optimization". In: *arXiv e-prints* (2020).

Thank You!
Questions Welcome!