Negative dependence tightens variational bounds

Pierre-Alexandre Mattei, Jes Frellsen

Inria

UNIVERSITÉ CÔTE D'AZUR

DTU Technical University of Denmark

Motivation: Importance-Weighted Autoencoders (IWAEs) aka Monte Carlo objectives

- **Latent-variable model** $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z})$
- "Easy to optimise" **variational lower bound** based on importance sampling:

$$\mathcal{L}_K = \mathbb{E}\left[\log\left(\frac{1}{K}\sum_{k=1}^{K} w_k\right)\right] \leq \log p(\mathbf{x}), \text{ with } w_k = \frac{p(\mathbf{x}|\mathbf{z}_k)p(\mathbf{z}_k)}{q(\mathbf{z}_k|\mathbf{x})}$$

- A more accurate MC estimate $\frac{1}{K}\sum_{k=1}^{K} w_k$ should (?) lead to a tighter variational bound
    - But **what does "more accurate" means?**
    - **Do variance-reduction techniques lead to tighter bounds?**

Burda, Grosse, and Salakhutdinov, *Importance Weighted Autencoders*, ICLR 2016

Do variance-reduction techniques leads to tighter bounds?

- There are two “simple” ways to reduce the variance of $\frac{1}{K}\sum_{k=1}^{K} w_k$
  - **Increase K**, which **provably tightens the bound** (Burda et al., 2016) and works wonders in practice
  - Use **negatively dependent weigths w_k** which was explored recently by Huang et al. (2019), Ren et al. (2019), and Wu et al. (2019). It **works very well in practice but lacks non-asymptotic theoretical justification**.
  - **Warning: variance reduction $\not\Longrightarrow$ tighter bound in general!**

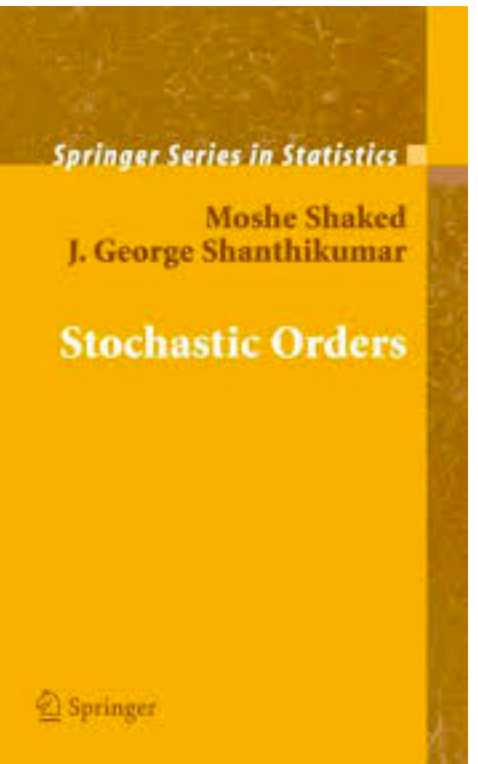
Burda, Grosse, and Salakhutdinov, *Importance Weighted Autencoders*, ICLR 2016
Huang, Sankaran, Dhekane, Lacoste, and Courville, *Hierarchical Importance Weighted Autencoders* ICML 2019
Ren, Zhao, and Ermon, *Adaptive Antithetic Sampling for Variance Reduction*, ICML 2019
Wu, Goodman, and Ermon, *Differentiable Antithetic Sampling for Variance Reduction in Stochastic Variational Inference*, AISTATS 2019

Negative variance tightens variational bounds

- We show, in a precise sense, that **the more negatively dependent the weights, the tighter the bound**
- We quantify negative dependence using the notion of **supermodular order** $\preceq_{\mathrm{SM}}$

Theorem (negative dependence tightens the bound). *For all pairs Q_1, Q_2 of probability distributions over $\mathbb{R}^K$,*

$$Q_1 \preceq_{\mathrm{SM}} Q_2 \implies \mathcal{L}_K(Q_1) \geq \mathcal{L}_K(Q_2).$$



Shaked and Shanthikumar, *Stochastic Orders,* Springer, 2007