Neural Inverse CDF Sampling

Andrew Stirn: andrew.stirn@columbia.edu
Tony Jebara: jebara@cs.columbia.edu

Problem

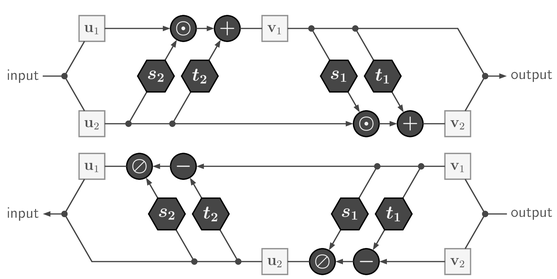
Thompson sampling (TS), a strategy to balance *exploration* and *exploitation* in multi-armed bandit problems, enjoys theoretical guarantees and strong empirical performance in a variety of domains. However, TS assumes causal relationships between selected actions and received rewards.

Contributions

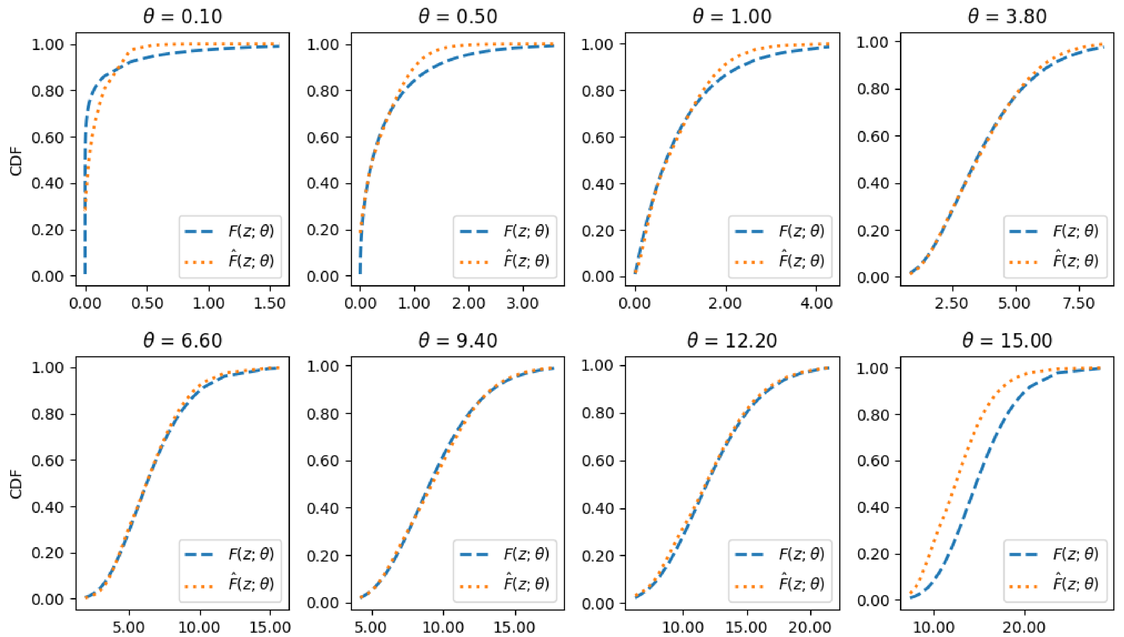
This article relaxes those assumptions by considering the *noncompliance* setting where a chosen action is not necessarily the implemented action. We formalize TS for both *observed noncompliance* and *latent noncompliance*. Further, we empirically demonstrate the performance benefit of our approach for a variety of environments with differing levels of stochastic noncompliance.

Invertible Neural Network

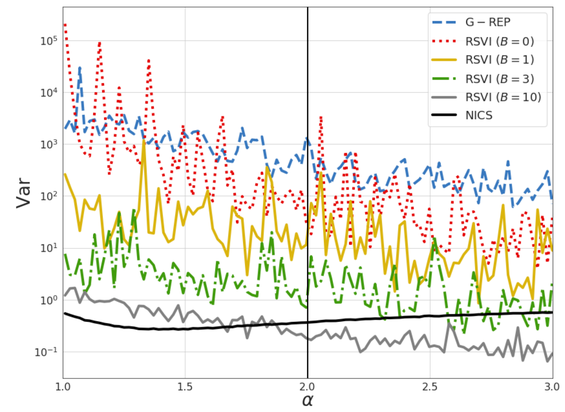
Affine coupling layer [1]



CDF Approximation for Gamma Distribution



Reparameterize



Results

Perplexity Scores

	Softmax	NGS
NVLDA	1099.96	1103.77
ProDLDA	???	???

References & Code

[1] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real NVP. *CoRR*, abs/1605.08803, 2016.

Source code available at:

`github.com/astirn/neural-inverse-cdf-sampling`