MAD Mix: flow family for discrete distributions

8

goal: approximate *multivariate* discrete posterior $p(x),\ x \in \mathbb{N}^M$

solution: construct a MixFlow based on the MAD map

$$q_N = \frac{1}{N} \sum_{n=0}^{N-1} T_{\mathrm{MAD}\,\sharp}^{n} q_0$$

i.i.d. sampling by evaluating MAD map

$$(X_0, U_0) \sim q_0, \quad K \sim \mathrm{Unif}\{0, \ldots, N-1\}$$

$$\implies T_{\mathrm{MAD}}^{K}(X_0, U_0) \sim q_N$$

density through mixed change of variables

$$q_N(x,u) = \frac{1}{N}\sum_{n=0}^{N-1} \frac{q_0(T_{\text{MAD}}^{-n}(x,u))}{\prod_{j=0}^{n} J_{\text{c}}(T_{\text{MAD}}^{-j}(x,u))}$$

product of prb ratios

problem: in practice also have real-valued latent variables

solution: can combine this idea with continuous MixFlows see paper for details

tl;dr you retain m.p., invertibility, etc.

MAD Mix: flow family for discrete distributions

goal: approximate *multivariate* discrete posterior $p(x),\ x \in \mathbb{N}^M$

solution: construct a MixFlow based on the MAD map

$$q_N = \frac{1}{N}\sum_{n=0}^{N-1} T^n_{\mathrm{MAD}\,\sharp} q_0$$

i.i.d. sampling by evaluating MAD map

$$(X_0, U_0) \sim q_0, \quad K \sim \mathrm{Unif}\{0, \dots, N-1\}$$
$$\implies T^K_{\mathrm{MAD}}(X_0, U_0) \sim q_N$$

problem: in practice also have real-valued latent variables

solution: can combine this idea with continuous MixFlows

see paper for details

tl;dr you retain m.p., invertibility, etc.

density through mixed change of variables

$$q_N(x,u) = \frac{1}{N}\sum_{n=0}^{N-1} \frac{q_0(T^{-n}_{\mathrm{MAD}}(x,u))}{\prod_{j=0}^{n} J_c(T^{-j}_{\mathrm{MAD}}(x,u))}$$

product of prb ratios

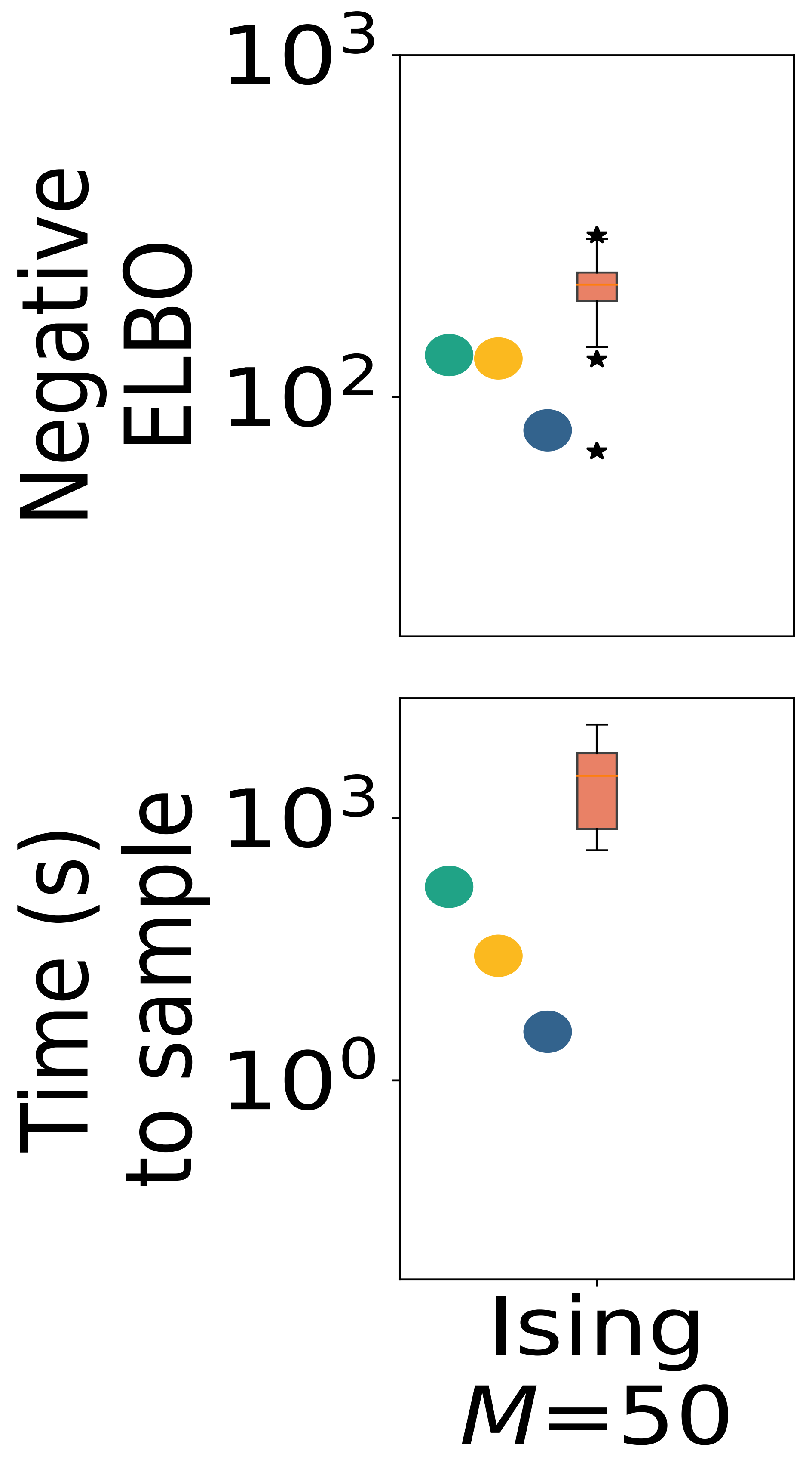
Quantitative results, real-world data

$\pi(x)$
Mad Mix (ours)
Gibbs
Mean field
Concrete
Dequantization

"distance" to p

lower is better

CPU time



purely discrete synthetic (d=2^{50})

setup: $N \approx 500$, $\xi = \pi/16$ for MAD Mix; 5K iterations for Gibbs (+20K burn-in); wide architecture search for continuous-embedding flows (concrete & dequantization)