Transport Reversible Jump Proposals

L. Davies[1,3] R. Salomone[2,3] M. Sutton[1,3] C. Drovandi[1,3]

[1]School of Mathematical Sciences – Queensland University of Technology

[2]School of Computer Science – Queensland University of Technology

[3]Center for Data Science – Queensland University of Technology

Motivation

The problem of interest is sampling probability distribution π on

$$\mathcal{X} = \bigcup_{k\in\mathcal{K}} (\{k\} \times \Theta_k), \tag{1}$$

with *parameters* $\boldsymbol{\theta}_k \in \Theta_k \subseteq \mathbb{R}^{n_k}$ and *model index* (or indicator) $k \in \mathcal{K}$. We want to make inference on the joint distribution (or conditional factorization) $\pi(k, \boldsymbol{\theta}_k) = \pi(k)\pi(\boldsymbol{\theta}_k|k)$. When data $\boldsymbol{y}$ is introduced, this becomes $\pi(k, \boldsymbol{\theta}_k|\boldsymbol{y}) = \pi(k|\boldsymbol{y})\pi(\boldsymbol{\theta}_k|k, \boldsymbol{y})$.

Contributions

I. A new class of RJMCMC proposals, called *transport reversible jump* (TRJ) proposals are developed that have desirable properties (see Proposition 1).
II. Efficacy of the proposed approach is demonstrated in numerical studies on challenging examples using *approximate* transport maps.
III. A modified version of the model probability estimator of [1] is applied to assess the quality proposals.
IV. An alternative "all-in-one" approach to training approximate TMs using *conditional normalizing flows* is explored (see paper).

Reversible Jump Markov Chain Monte Carlo

We want to propose from point $\boldsymbol{x}$ to point $\boldsymbol{x}'$, noting $\boldsymbol{\theta}_k$, $\boldsymbol{\theta}'_{k'}$ have dimensions n_k, $n_{k'}$ respectively. Following [2], we:

- Require dimensions match: introduce auxiliary variables $\boldsymbol{u}_k \in \mathcal{U}_{k,k'} \subseteq \mathbb{R}^{w_k}$ and $\boldsymbol{u}_{k'} \in \mathcal{U}_{k',k} \subseteq \mathbb{R}^{w_{k'}}$ such that $n_k + w_k = n_{k'} + w_{k'}$.
- Choose a diffeomorphism e. $\boldsymbol{\theta}_{k'}, \boldsymbol{u}_{k'} = h_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{u}_k)$.

A (simplified) RJMCMC Algorithm when $n_{k'} > n_k$ is:

1. Propose model index $k' \sim j_k(\,\cdot\,)$
2. Propose auxiliary variables $\boldsymbol{u}_k \sim g_{k,k'}(\,\cdot\,)$
3. Accept with probability

$$\alpha(\boldsymbol{x}, \boldsymbol{x}') = 1 \wedge \frac{\pi(\boldsymbol{x}')j_{k'}(k)g_{k',k}(\boldsymbol{u}'_{k'})}{\pi(\boldsymbol{x})j_k(k')g_{k,k'}(\boldsymbol{u}_k)}\left|J_{h_{k,k'}}(\boldsymbol{\theta}_k, \boldsymbol{u}_k)\right|. \tag{2}$$

Transport Maps and Normalizing Flows

A function $T : \mathbb{R}^n \to \mathbb{R}^n$ is called a *transport map* (TM) from distribution $\mu_{\boldsymbol{\theta}}$ to distribution $\mu_{\boldsymbol{Z}}$ if $\mu_{\boldsymbol{Z}} = T\sharp\mu_{\boldsymbol{\theta}}$, i.e. $\mu_{\boldsymbol{Z}}$ is the *pushforward* of $\mu_{\boldsymbol{\theta}}$ using the measurable function T.

Let $\{T_{\boldsymbol{\psi}}\}$ be a family of *diffeomorphisms* with domain on the support of some arbitrary *base* distribution $\mu_{\boldsymbol{Z}}$. Then, for fixed parameters $\boldsymbol{\psi}$, the PDF of the random vector $\boldsymbol{\vartheta} = T_{\boldsymbol{\psi}}(\boldsymbol{Z})$ is

$$\mu_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}; \boldsymbol{\psi}) = \mu_z(T^{-1}_{\boldsymbol{\psi}}(\boldsymbol{\vartheta}))|J_{T^{-1}_{\boldsymbol{\psi}}}(\boldsymbol{\vartheta})|, \;\; \boldsymbol{\vartheta} \in \mathbb{R}^n. \tag{3}$$

Distributions $\mu_{\boldsymbol{\vartheta}}$ are *flow-based models*, where $\{T_{\boldsymbol{\psi}}\}$ are the *normalizing flows* [5]. With finite samples $\boldsymbol{s} \sim \pi$, we obtain an *approximate* TM $\hat{T}$ via density estimation, minimising the KLD from $\{\boldsymbol{s}\}$ to $\mu_{\boldsymbol{\vartheta}}$.

Proposed Method: Transport Reversible Jump

First, let ν be a univariate closed-form distribution e.g. standard normal ϕ. For each model $k \in \mathcal{M}$:

- Define reference distributions $\otimes_{n_k}\nu$.
- Train flows on samples $\boldsymbol{s}_k \sim \pi(\boldsymbol{\theta}_k|k)$ to obtain $\hat{T}_k$.

Then, a transdimensional proposal where $n_{k'} > n_k$ is

$$\begin{aligned} \boldsymbol{z}_k &\leftarrow T_k(\boldsymbol{\theta}_k), \\ \boldsymbol{z}'_{k'} &\leftarrow \bar{h}_{k,k'}(\boldsymbol{z}_k, \boldsymbol{u}_k), \\ \boldsymbol{\theta}_{k'} &\leftarrow T^{-1}_{k'}(\boldsymbol{z}'_{k'}), \end{aligned} \tag{4}$$

where $\bar{h}_{k,k'}$ is a *volume-preserving* diffeomorphism on $\otimes_{n_k}\nu$.

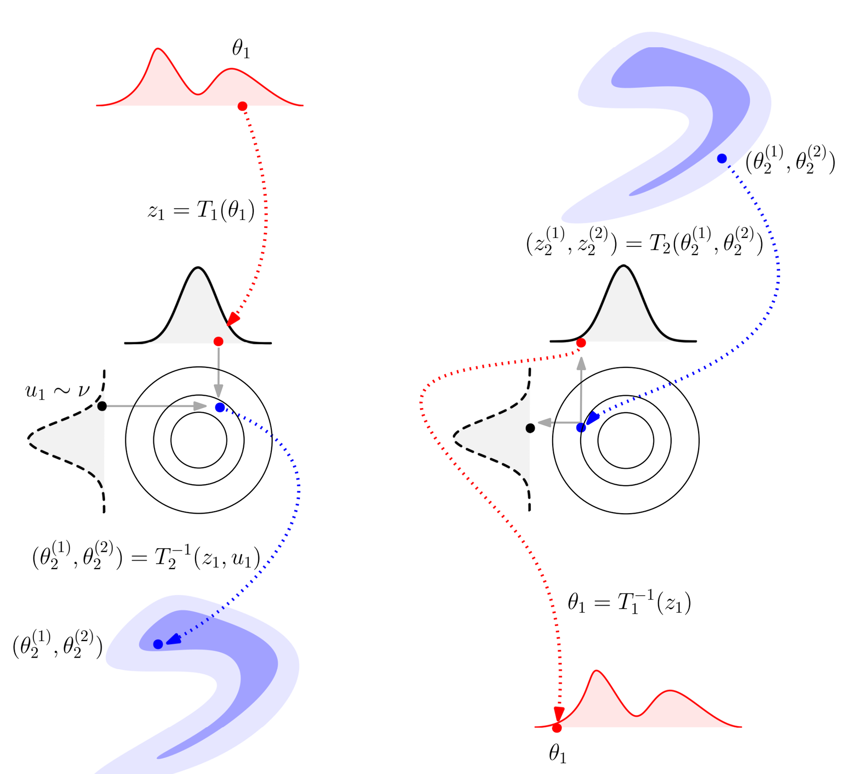

Figure 1. Illustration of the proposal class. Here, the reference ν is Gaussian. The diffeomorphisms ($\bar{h}_{k,k'}$) on the reference distributions concatenate or extract coordinates as required.

Proposition: Exact Transport RJMCMC

Suppose that RJMCMC proposals are of the form described in (4), and for each $k \in \mathcal{K}$, satisfy $T_k\sharp\pi_k = \otimes_{n_k}\nu$. Then, (2) reduces to

$$\alpha(\boldsymbol{x}, \boldsymbol{x}') = 1 \wedge \frac{\pi(k')}{\pi(k)}\frac{j_{k'}(k)}{j_k(k')}. \tag{5}$$

Corollary

Provided the conditions of Proposition 1 are satisfied, choosing $\{j_k\}$ such that

$$\pi(k')j_{k'}(k) = \pi(k)j_k(k'), \quad \forall k, k' \in \mathcal{K}, \tag{6}$$

leads to a rejection-free proposal.

Illustrative Example: 1D 2D Sinh-Arcsinh Target

Using the inverse sinh-arcsinh transformation of [3]

$$S_{\boldsymbol{\epsilon},\boldsymbol{\delta}}(\cdot) = \sinh(\boldsymbol{\delta}^{-1} \odot (\sinh^{-1}(\cdot) + \boldsymbol{\epsilon})), \;\; \boldsymbol{\epsilon} \in \mathbb{R}^n, \boldsymbol{\delta} \in \mathbb{R}^n_+.$$

For $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_n, \mathrm{I}_{n\times n})$ and lower triangular $n \times n$ matrix L, the *exact* (or "perfect") transport is

$$T(\boldsymbol{Z}) = S_{\boldsymbol{\epsilon},\boldsymbol{\delta}}(\mathrm{L}\boldsymbol{Z}), \;\text{ i.e. } T^{-1}(\cdot) = \mathrm{L}^{-1}S^{-1}_{\boldsymbol{\epsilon},\boldsymbol{\delta}}(\cdot), \tag{7}$$

for chosen reference distributions ϕ_n, $n_k = k$. The PDF for $\boldsymbol{\theta} = T(\boldsymbol{Z})$ takes the form in (3). The target of interest for this example, where $\boldsymbol{\theta}_1 = (\theta^{(1)}_1)$ and $\boldsymbol{\theta}_2 = (\theta^{(1)}_2, \theta^{(2)}_2)$, is

$$\pi(k, \boldsymbol{\theta}_k) = \begin{cases} \frac{1}{4}p_{\epsilon_1,\delta_1,1}(\boldsymbol{\theta}_1), & k = 1, \\ \frac{3}{4}p_{\boldsymbol{\epsilon}_2,\boldsymbol{\delta}_2,\mathrm{L}}(\boldsymbol{\theta}_2), & k = 2, \end{cases} \tag{8}$$

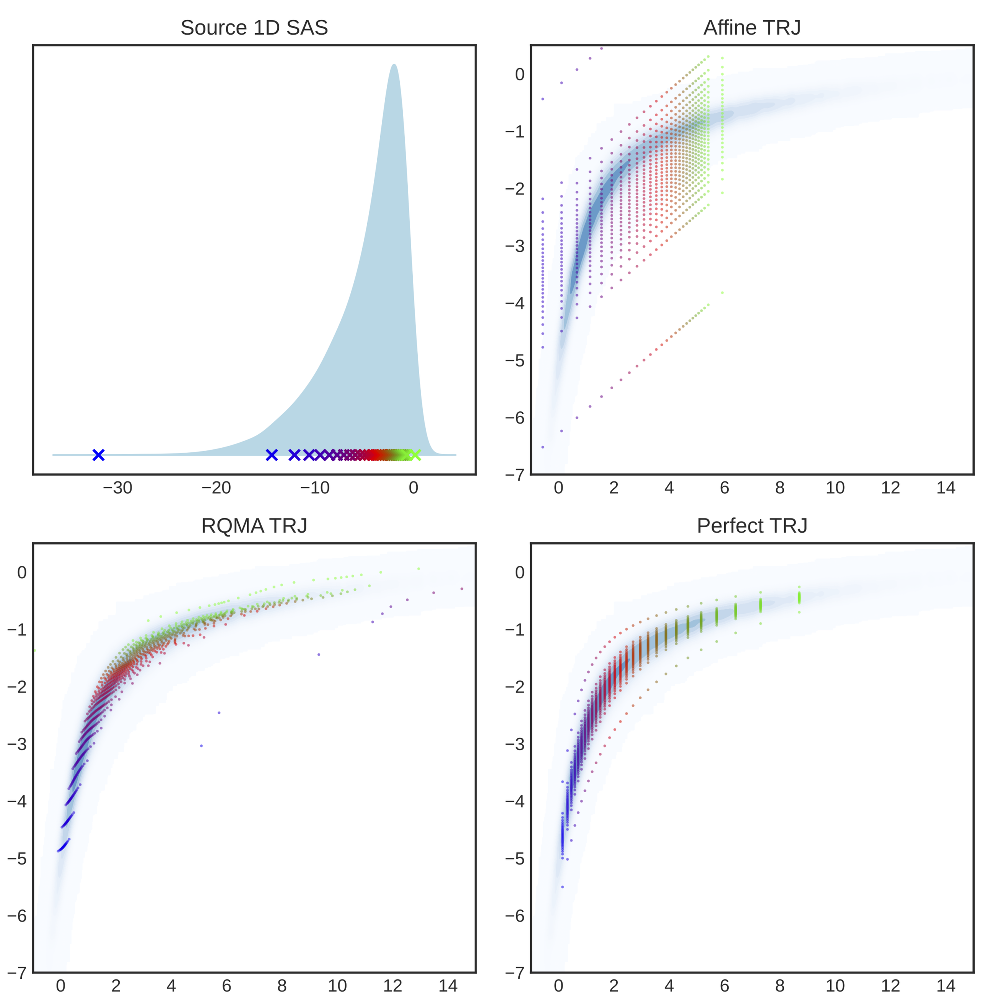

Figure 2. Systematic draws from conditional target $\pi(x_1|k=1)$ of (8) are transported from $(1, \theta_1) \in \mathcal{K} \times \mathbb{R}^1$ (*top left*) to $(2, (\theta_1, \theta_2)) \in \mathcal{K} \times \mathbb{R}^2$ via TRJ proposals using: *Top right* Approximate affine, *Bottom left* Approximate RQMA-NF, *Bottom right* Perfect TM. The auxilliary variables in the proposals are also drawn systematically (30 for each point in the source distribution).

Modified Bartolucci Bridge Sampling Estimator

For an RJMCMC chain, [1] showed that the Bayes factor $B_{k,k'}$ (ratio of marginal likelihoods) is estimated via

$$\hat{B}_{k,k'} = \frac{N^{-1}_{k'}\sum_{i=1}^{N_{k'}}\alpha'_i}{N^{-1}_k\sum_{i=1}^{N_k}\alpha_i}, \tag{9}$$

where $N_{k'}$ and N_k are the number of proposed moves from model k' to k, and from k to k', respectively. Assuming uniform prior model probabilities, estimates of posterior model probabilities are

$$\hat{\pi}(k) = \hat{B}^{-1}_{j,k}\left(1 + \sum_{i\in\mathcal{K}\setminus\{j\}}\hat{B}_{i,j}\right)^{-1}, \;\text{ for arbitrary } j \in \mathcal{K}. \tag{10}$$

The **Modified Bartolucci Estimator** (MBE) adopts the above for proposals from *samples* of the conditional targets.

References

[1] Francesco Bartolucci, Luisa Scaccia, and Antonietta Mira. Efficient Bayes factor estimation from the reversible jump output. *Biometrika*, 93(1):41–52, 2006.

[2] Peter J Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[3] M. C. Jones and Arthur Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96(4):761–780, December 2009.

[4] Hedibert Freitas Lopes and Mike West. Bayesian model assessment in factor analysis. *Statistica Sinica*, 14(1):41–67, 2004.

[5] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1530–1538. PMLR, June 2015.

Bayesian Factor Analysis Example

Task is to select a model for monthly exchange rates of six currencies relative to the British pound, spanning January 1975 to December 1986, denoted as $\boldsymbol{y}_i \in \mathbb{R}^6$ for $i = 1, ..., 143$, of the random vector $\boldsymbol{Y}$. Assume $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{0}_6, \boldsymbol{\Sigma})$, where k is no. factors, $\boldsymbol{\Sigma} = \boldsymbol{\beta}_k\boldsymbol{\beta}_k^\top + \boldsymbol{\Lambda}$, Λ is a 6×6 positive diagonal matrix, and $\boldsymbol{\beta}_k$ is a $6 \times k$ lower-triangular matrix with a positive diagonal. Following [4], for each $\boldsymbol{\beta}_k = [\beta_{ij}]$ with $i = 1, \ldots, 6$, $j = 1, \ldots, k$, the priors are

$$\begin{aligned} \beta_{ij} &\sim \mathcal{N}(0, 1), \;\; i < j \\ \beta_{ii} &\sim \mathcal{N}_+(0, 1), \\ \Lambda_{ii} &\sim \mathcal{IG}(1.1, 0.05), \end{aligned} \tag{11}$$

Posterior probability of $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\Lambda})$ for $k = 2, 3$ via Bayes' Thm

$$\pi(k, \boldsymbol{\theta}_k|\boldsymbol{y}) \propto p(k)p(\boldsymbol{\beta}_k|k)p(\boldsymbol{\Lambda})\prod_{i=1}^{143}\phi_{\boldsymbol{\beta}\boldsymbol{\beta}^\top+\boldsymbol{\Lambda}}(\boldsymbol{y}_i). \tag{12}$$

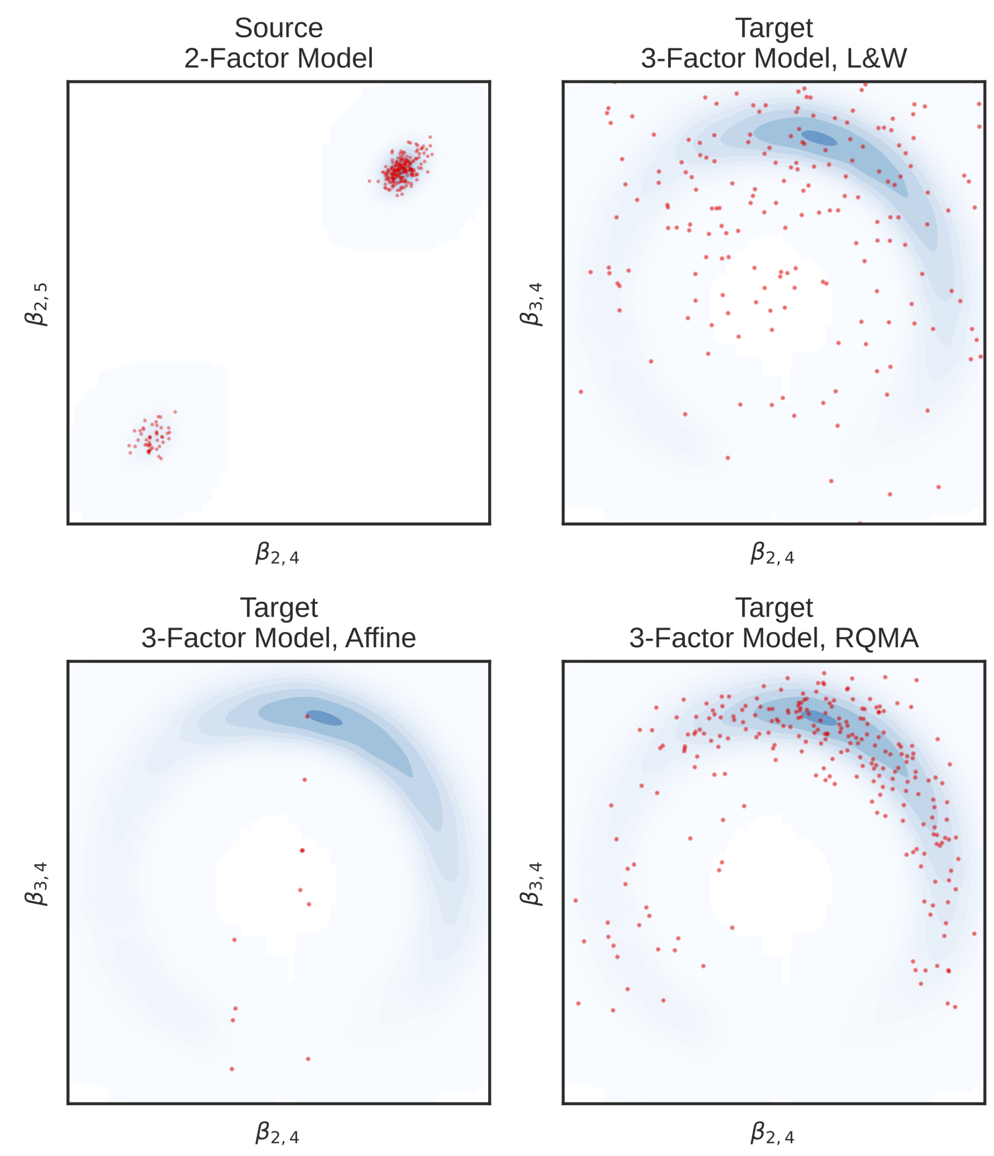

Figure 3. A visualization (using selected bivariate plots) of the proposal from points on the 2-factor model (*top-left*) to proposed points on the 3-factor model for each proposal type: [4] independence RJMCMC proposal (*top right*); TRJ with Affine map (*bottom left*); TRJ with RQMA-NF map (*bottom right*).

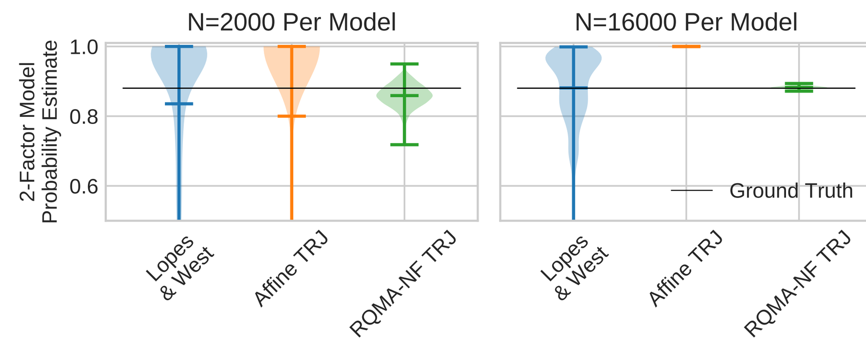

Figure 4. Violin plot showing the variability of the 2-factor model probability estimates in the case where only the 2-factor and 3-factor models are compared. Model probability estimates are obtained via the MBE. Ground truth is estimated via extended individual SMC runs ($N = 5 \cdot 10^4$).

Example: Robust Regression Variable Selection

Please see the paper (QR code below) for a detailed description.

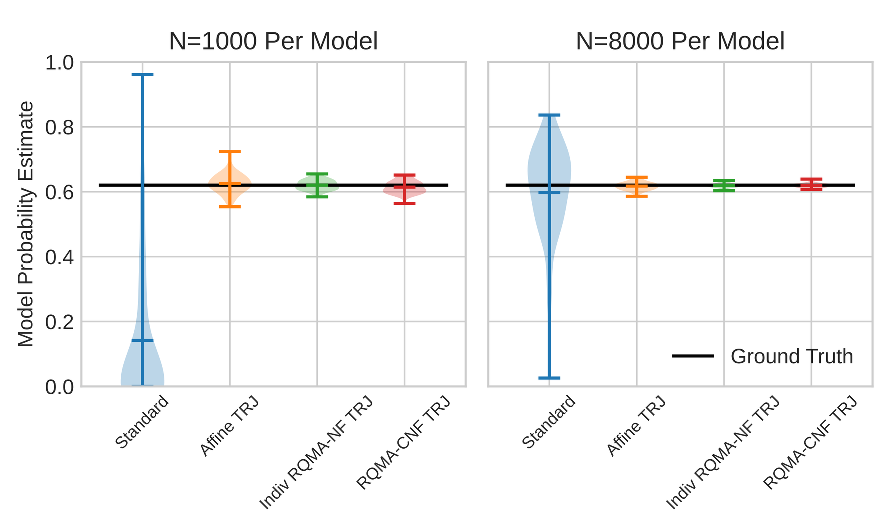

Figure 5. Violin plot showing the variability of the $k = (1, 1, 1, 1)$ model probability estimate for each proposal type using the MBE vs ground truth individual SMC ($N = 5 \cdot 10^4$). Individual SMC with $N = 1000, 8000$ particles sampled conditional targets split into training/test samples for a total of 80 passes.

Conclusions and Future Research Avenues

- Theoretically rejection-free approach for RJMCMC.
- Proposals using approximate TMs outperform baseline.
- Caveat of requiring samples to train the approximate TMs.
- Conditional NF approach to reduce training time (see paper).
- Efforts are justified in expensive-likelihood scenarios.
- MBE benchmark performance justifies further investigation.

Pre-print is available via: