# Estimating Convergence of Markov chains with $L$-Lag Couplings

Niloy Biswas[1], Pierre E. Jacob[1], Paul Vanetti[2]

1: Harvard University 2: University of Oxford

## Motivation: Measure non-asymptotic bias of MCMC

- MCMC methods have **non-asymptotic bias**: they only reach a target distribution as the number of iterations goes to infinity.
- We introduce $L$-**lag couplings** to **generate computable, non-asymptotic upper bound estimates for the total variation and 1-Wasserstein distances** of general Markov chains to stationarity.
  - Total Variation Distance: e.g. histograms, credible intervals

$$d_{TV}(P,Q) = \sup_{\substack{h:|h|\leq 1 \\ X\sim P, Y\sim Q}} \big|\mathbb{E}[h(X)-h(Y)]\big| \quad (1)$$

  - 1–Wasserstein: e.g. all first moments

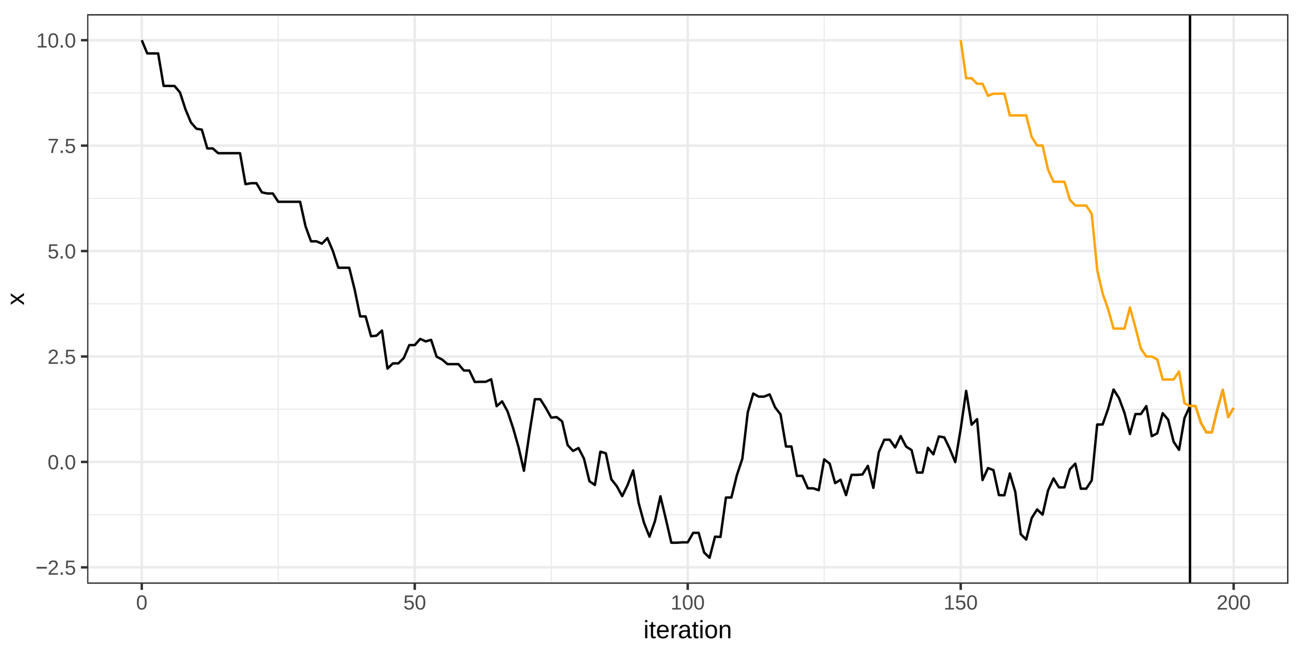
$$d_{\mathrm{W}}(P,Q) = \sup_{\substack{h:\|h(x)-h(y)\|\leq\|x-y\| \\ X\sim P, Y\sim Q}} \big|\mathbb{E}[h(X)-h(Y)]\big| \quad (2)$$

## What are $L$-Lag Couplings?

- A pair of Markov chains $(X_t, Y_t)_{t\geq 0}$ such that:
  - Same marginal distributions: $X_t \sim Y_t \sim \pi_t \; \forall t \geq 0$ with $\pi_t \overset{t\to\infty}{\Rightarrow} \pi$
  - $X_t$ and $Y_{t-L}$ meet **exactly** at time

$$\tau^{(L)} := \inf\{t > L : X_t = Y_{t-L}\}$$

  - Chains stay faithful after coupling: $X_t = Y_{t-L} \; \forall t \geq \tau^{(L)}$
- Example: 150-Lag Coupling of Random-Walk Metropolis–Hastings with start $\delta_{10}$ and target $\mathcal{N}(0,1)$




- Coupling algorithms for common MCMC methods available (e.g. RWMH, Gibbs samplers, HMC, Particle Gibbs)

## Main Theorem

**Theorem.** *Consider an $L$-lag coupling of chains $(X_t, Y_t)_{t\geq 0}$, where $X_t \overset{t\to\infty}{\Rightarrow} \pi$ and meeting time $\tau^{(L)} := \inf\{t > L : X_t = Y_{t-L}\}$ has sub-exponential tails. Then,*

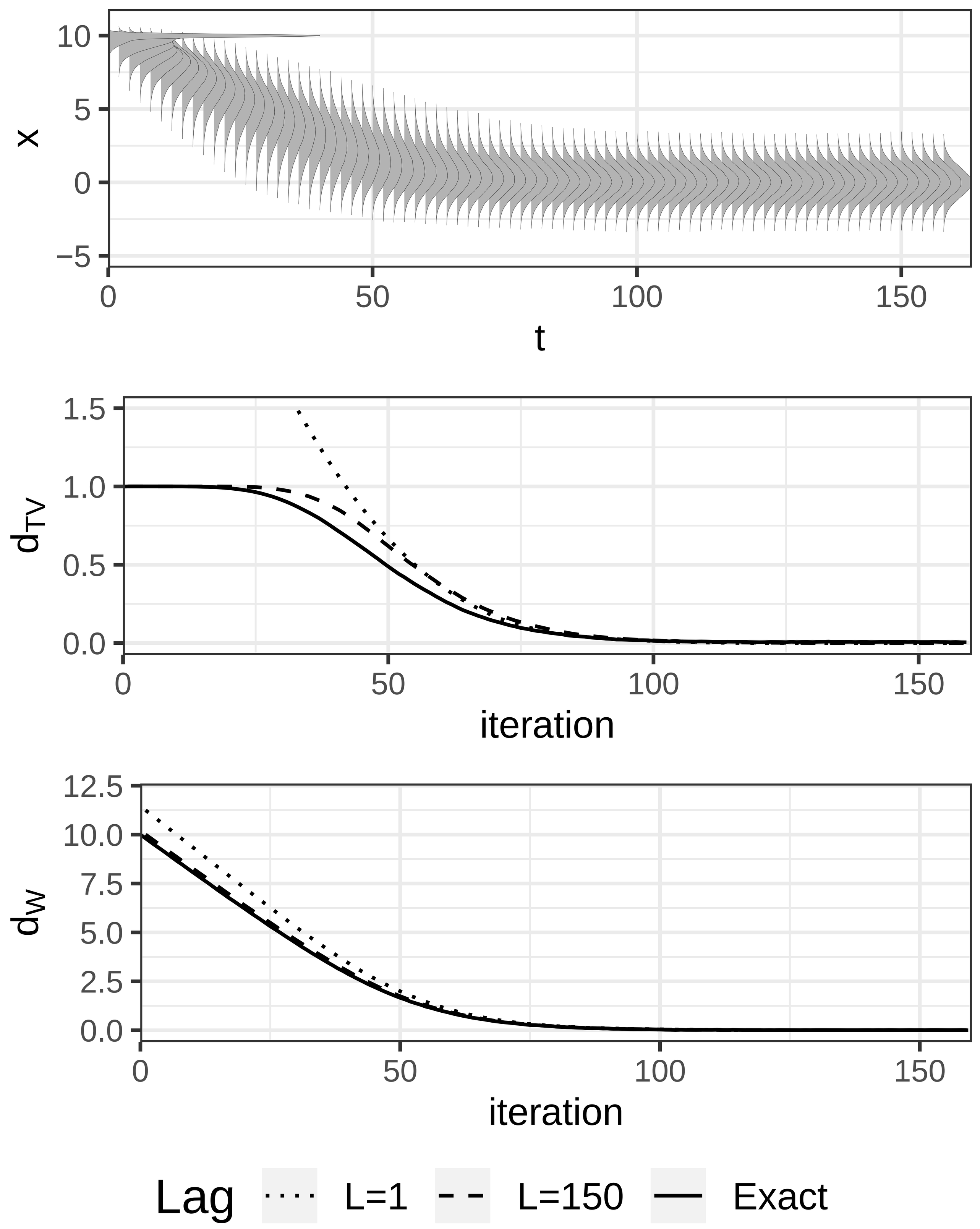
$$d_{TV}(\pi_t, \pi) \leq \mathbb{E}\Big[\max(0, \lceil \frac{\tau^{(L)} - L - t}{L} \rceil)\Big] \quad (3)$$

*Further assume for some $\eta > 0$, $(2+\eta)$-moments of chain $(X_t)_{t\geq 0}$ are uniformly bounded. Then,*

$$d_{W}(\pi_t, \pi) \leq \mathbb{E}\Big[\sum_{j=1}^{\lceil \frac{\tau^{(L)}-L-t}{L} \rceil} \|X_{t+jL} - Y_{t+(j-1)L}\|_1\Big] \quad (4)$$

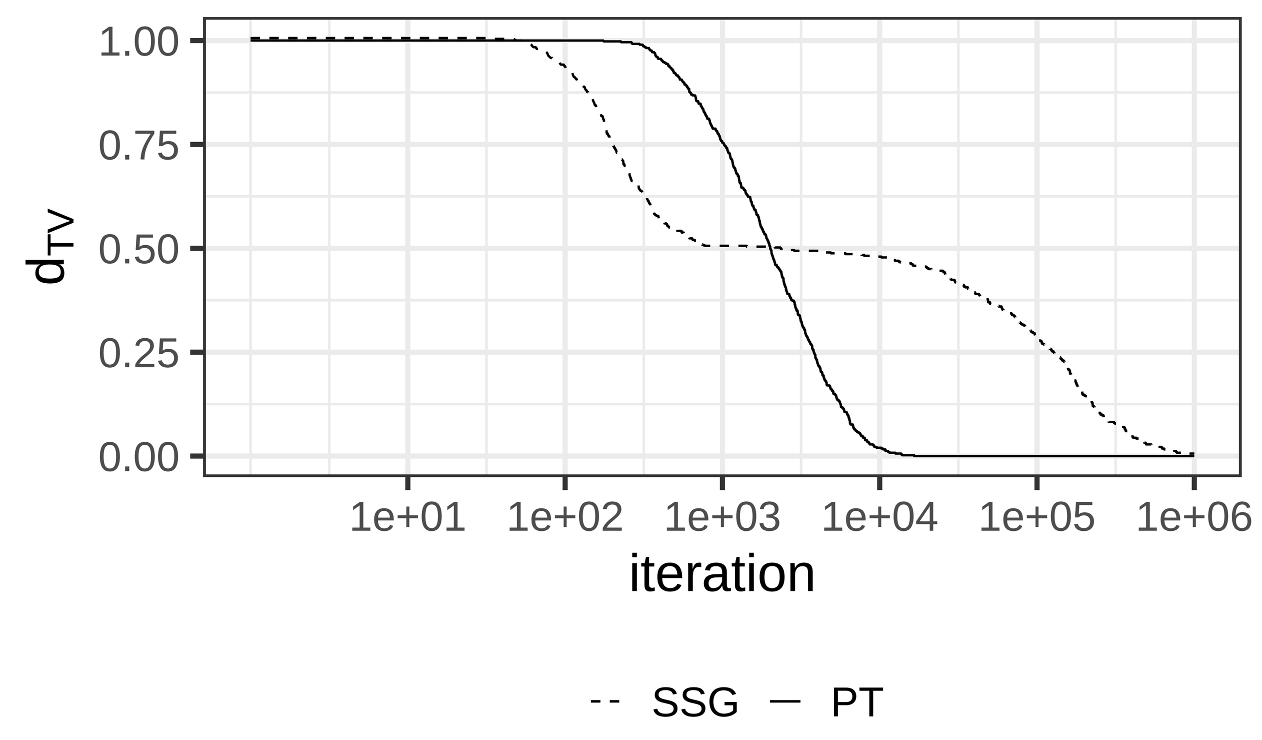
## Stylized Example

- Random-Walk Metropolis–Hastings: start $\delta_{10}$, target $\mathcal{N}(0,1)$



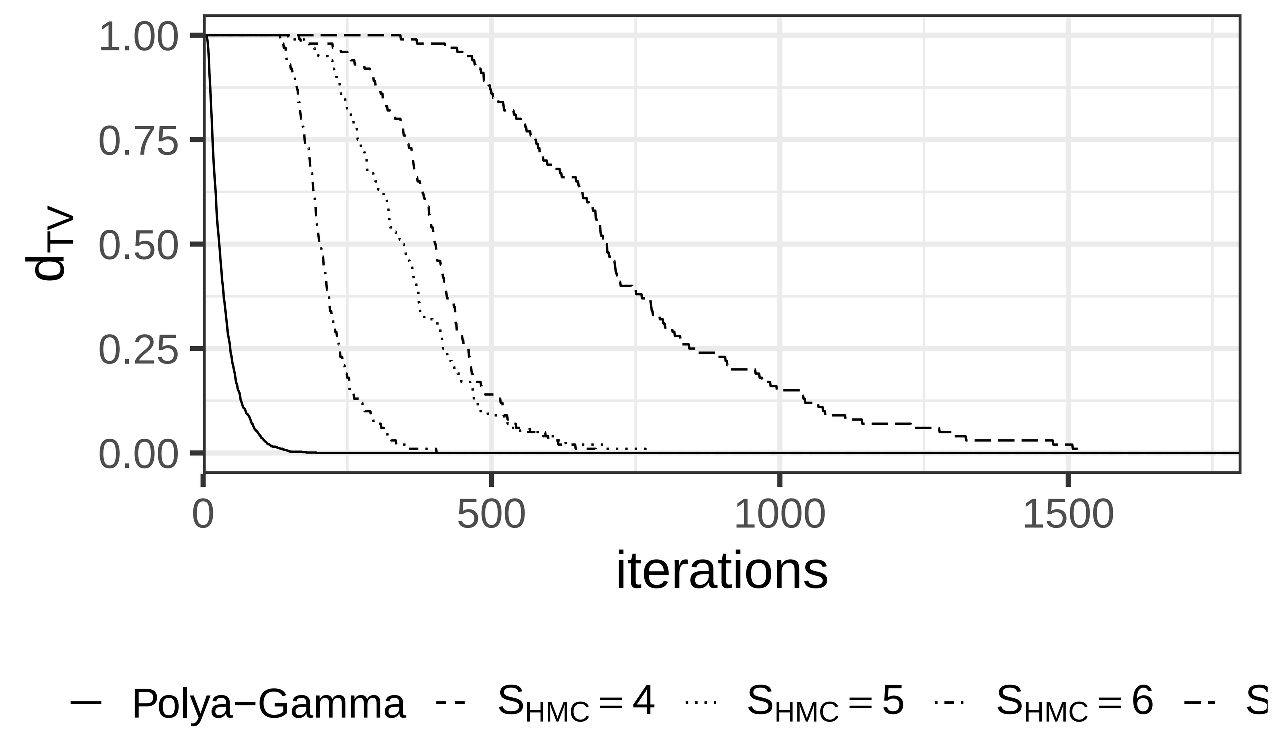

## Ising Model: Single-site Gibbs vs. Parallel Tempering

- Discrete state space: $\{-1,1\}^{32\times 32}$.
- Target: $\pi_\beta(x) \propto \exp(\beta \sum_{i\sim j} x_i x_j)$ for all $i \sim j$ neighboring sites.




## Bayesian Logistic Regression: HMC vs. Pólya-Gamma

- Sampling from the posterior:
  - Hamiltonian Monte Carlo (HMC) with parameters $\epsilon_{HMC}, S_{HMC}$
  - Parameter-free Pólya-Gamma (PG)




## References and Implementation


- Jacob, OLeary and Atchadé. Unbiased Markov chain Monte Carlo with couplings. *JRSS-B*, 2019.
- Heng and Jacob. Unbiased HMC with couplings. *Biometrika*, 2019.


- $L$-Lag Couplings Code: https://github.com/niloyb/LlagCouplings