Distributed Markov chain Monte Carlo

Lawrence Murray

CSIRO Mathematics, Informatics and Statistics

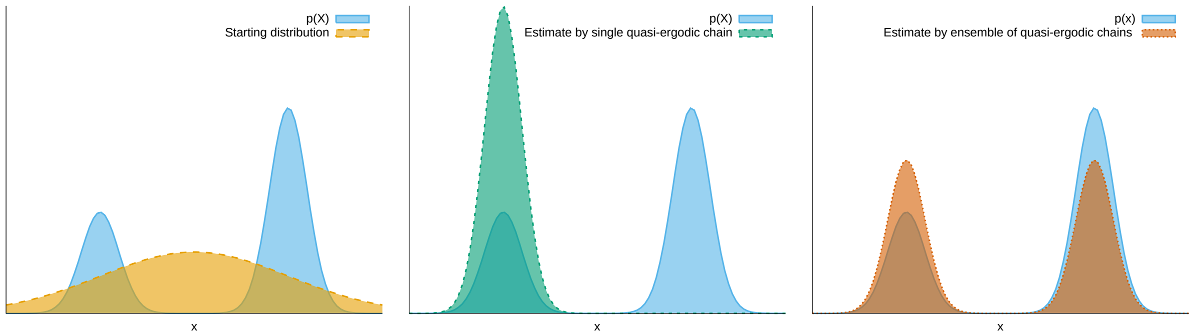
Motivation

- Bayesian inference in environmental models.
- Particle Markov chain Monte Carlo (PMCMC):
  - state-space model,
  - Metropolis-Hastings over $p(\boldsymbol{\Theta}|\mathbf{y}_{1:T})$,
  - use particle filter to estimate marginal likelihoods:

$$\int_{-\infty}^{\infty} p(\mathbf{y}_{1:T}, \mathbf{x}_{1:T}|\boldsymbol{\theta})\, d\mathbf{x}_{1:T}$$

- Particle filters executed on GPU, but evaluations still take several seconds, may require several minutes for larger models.
- Scale up to cluster level, one Markov chain per CPU-GPU pair.

Quasi-ergodicity and multiple chains



$p(X)$ is the target distribution, consisting of two isolated modes; **(left)** the starting distribution; **(centre)** typical posterior returned by a single quasi-ergodic chain; **(right)** typical posterior returned by multiple quasi-ergodic chains.

Convergence and multiple chains

If some portion ρ of steps, $0 < \rho \leq 1$ and typically up to $.5$, must be removed as burn-in from each chain, the maximum clock-time speedup through parallelisation is limited to $1/\rho$ (Amdahl's law).

Thus, a multiple-chain strategy must also reduce ρ as the number of chains increases in order to scale well.

Method

For each chain i, consider a proposal that mixes some *local* component $l_i(\boldsymbol{\theta}'_i|\boldsymbol{\theta}_i)$ with a *remote* or *global* component $R_i(\boldsymbol{\theta}'_i)$:

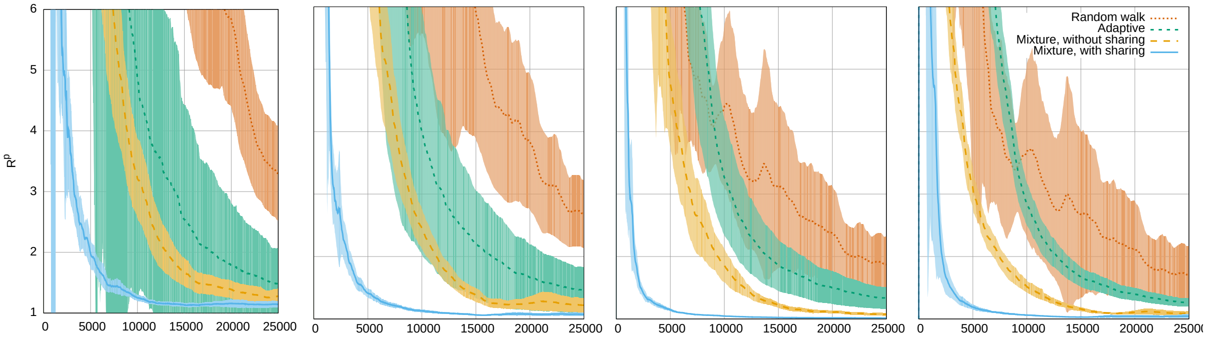
$$q_i(\boldsymbol{\theta}'_i) := (1-\alpha)l_i(\boldsymbol{\theta}'_i|\boldsymbol{\theta}_i) + \alpha R_i(\boldsymbol{\theta}'_i)\,,$$

$R_i(\cdot)$ can be constructed via some contributed component $r_j(\cdot)$ from each chain j. Consider:

$$R_i(\boldsymbol{\theta}'_i) \propto \max_{j=1}^{C} r_j(\boldsymbol{\theta}'_i)\,.$$

Importantly, $R_i(\cdot)$ can be adapted *asynchronously* as new information is received from other chains. Faults only deprive chains of timely adaptation, they do not impact correctness.

Early results



Evolution of the $\hat{R}^p$ statistic of Brooks & Gelman (1998) across steps for each method, with **(left to right)** 2, 4, 8 and 16 chains. Lines indicate mean across 20 runs, and shaded areas a half standard deviation either side.

CSIRO Mathematics, Informatics and Statistics

Lawrence Murray

Phone: +61 8 9333 6480
Email: lawrence.murray@csiro.au
Web: www.cmis.csiro.au

www.csiro.au

Thank you

Contact Us
Phone: 1300 363 400 or +61 3 9545 2176
Email: enquiries@csiro.au Web: www.csiro.au

CSIRO