Paul Vanetti and **Arnaud Doucet** (*University of Oxford*)
Bias correction for MCMC is a long-standing problem, for which this paper represents a breakthrough. We propose two generalizations to derive new unbiased estimators.

First we propose the *lagged estimator*, obtained by starting chain Y after L steps of the chain X and coupling those two chains such that $X_t = Y_{t-L}$ for all $t \geq \tau$ where τ is the meeting time. For some N, k such that $N \geq k + L$, we can then exploit the identity

$$\frac{1}{L} \sum_{n=N-L+1}^{N} \mathbb{E}[h(X_n)] = \frac{1}{L} \left\{ \sum_{i=k}^{k+L-1} \mathbb{E}[h(X_i)] + \sum_{j=k+L}^{N} \{\mathbb{E}[h(X_j)] - \mathbb{E}[h(Y_{j-L})]\} \right\}. \tag{1}$$

Lags greater than one were used in [BJV19] to improve probability metric bounds. The estimator (1) is similar to the time-averaged estimator in [JOA20] when $L = m - k$, but the bias correction term incurred is not inflated by a large coefficient. Our empirical results (Table 1) in a simple scenario suggest that (1) can outperform the time-averaged estimator at a similar computational cost.

For L large enough, where X_L is approximately stationary, the bias correction term of (1) can be interpreted as a removal of the burn-in, representing the difference between a stationary chain and the first iterations of a new chain. This motivates our second innovation, which is to use the X chain of each pair of coupled chains as the Y chain for another pair.

If we simulate R pairs $(X^{(r)}, Y^{(r)})$, using the chain $Y_i^{(r)} = X_i^{(r+1)}$ for $r \in \{1, \ldots, R-1\}$, and for $r = R$ we take a novel chain (and not the X for any other pair), then averaging the estimates over the R pairs yields an estimator in which the negative bias correction terms for the first $R - 1$ pairs cancel with the first samples from the next processor. Thus the estimator is equivalent to that obtained by a single long chain with a lag choice of RL; see Figure 1 for an illustration.

If we had used $Y_i^{(R)} = X_i^{(1)}$ for the Rth pair, the resulting scheme would very closely match the parallel implementation of circular coupling [Nea99]. The framework provided in this discussion gives a new interpretation to this scheme, which was designed to provide "states that all have close to the equilibrium distribution". A natural question to ask is whether it yields unbiased estimates and, if this is not the case, whether it can be modified to achieve exact unbiasedness.

References

[BJV19] Niloy Biswas, Pierre E Jacob, and Paul Vanetti. Estimating convergence of Markov chains with L-lag couplings. In *Advances in Neural Information Processing Systems*, pages 7389–7399, 2019.

k	m	σ	k	L	σ
1	10	430	1	9	66.7
10	100	34.2	10	90	11.9
100	1000	0.119	100	900	0.119

Table 1: Standard deviation for (left) the time-averaged estimator (with lag 1) and (right) the lagged estimator (with $L = m - k$). The underlying MCMC chain is a Metropolis random walk with target $\pi(x) = \mathcal{N}(x; 0, 1)$, initialization $\pi_0(x_0) = \mathcal{N}(x_0; 0, 5^2)$, proposal $q(x, x^*) = \mathcal{N}(x^*; x, 1)$, and test function $h(x) = x^2$. σ is the standard deviation of a single estimate.

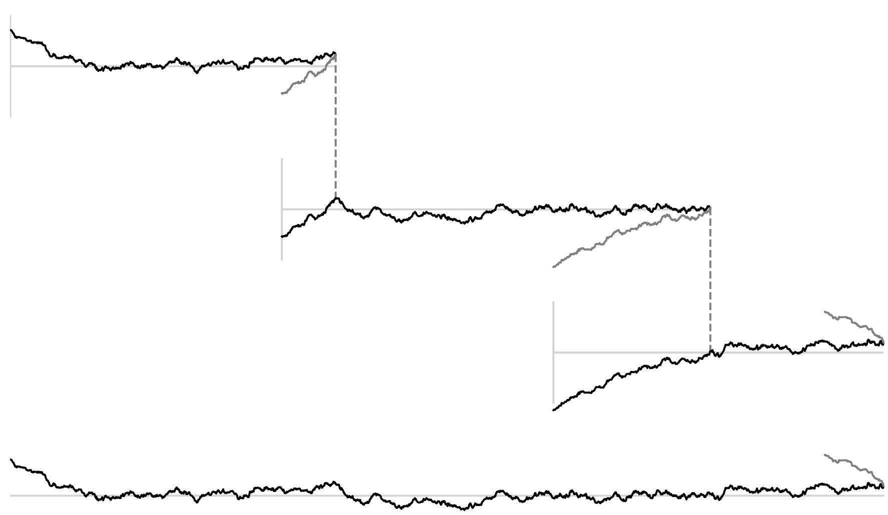

Figure 1: Sharing chains between pairs. The top three plots are three pairs; the X chain of the second is used as the Y chain for the first, and similarly the X of the third for the Y of the second. The final plot is the aggregate chain. The black lines represent positive contributions to the total estimate, while the grey lines represent negative contributions.

[JOA20] Pierre E Jacob, John O'Leary, and Yves Atchadé. Unbiased Markov chain Monte Carlo with coupling. *Journal of the Royal Statistical Society Series B*, 2020.

[Nea99] Radford M Neal. Circularly-coupled Markov chain sampling. *arXiv preprint arXiv:1711.04399*, 1999.