Coupling From the Past: Why Doubling With Reuse is Essential

LP

Abstract

We describe two natural-looking but *wrong* variants of Propp–Wilson [PW96] coupling from the past (CFTP) and explain why each one produces biased samples. We test all variants against the q-area measure on up-right lattice paths, whose exact distribution is given by q-binomial coefficients, and verify that only the correct backward-doubling protocol reproduces it.

1 Setup

Let Φ be a random map on a finite state space Ω whose iteration defines an ergodic Markov chain with stationary distribution π. *Coupling from the past* (CFTP) draws an exact sample from π by the following protocol.

Correct CFTP (backward doubling with reuse).

1. Fix a sequence of independent random maps $\Phi_{-1}, \Phi_{-2}, \Phi_{-3}, \ldots$, generated lazily.
2. Set $T = 1$. Apply Φ_{-1} to *every* starting state $x \in \Omega$ and check whether all chains land on the same state at time 0.
3. If not, double: set $T \leftarrow 2T$, generate the new maps $\Phi_{-T}, \ldots, \Phi_{-T/2-1}$ (fresh), but **reuse** the maps $\Phi_{-T/2}, \ldots, \Phi_{-1}$ already generated. Run the composition $\Phi_{-1} \circ \cdots \circ \Phi_{-T}$ on all of Ω.
4. Repeat until coalescence. Output the common value.

The **reuse** of randomness for the later portion of the time interval is the critical ingredient. We now describe two variants that omit it and show why each one fails.

2 Wrong variant 1: forward coupling

Forward coupling. Start all chains at time 0 from every state in Ω, apply the *same* random map at each step $\Phi_1, \Phi_2, \ldots$, and stop at the first (random) time τ when all chains have coalesced. Output the common state X_τ.

This looks symmetric with CFTP but the output distribution is wrong. The coalescence time τ is a **stopping time** that depends on the trajectory. The state X_τ is the outcome at a data-dependent time, not at a fixed time run from the infinite past. In particular, X_τ is biased

toward states that act as "attractors" for the coupling — states that many other states tend to map into.

Concretely, the forward-coupling output distribution is

$$\Pr[\text{output} = x] \;=\; \sum_{t=1}^{\infty} \Pr[\tau = t,\; X_\tau = x] \tag{2.1}$$

and there is no reason for (2.1) to equal $\pi(x)$. The correct CFTP avoids this because the target time (time 0) is fixed in advance and independent of when coalescence is detected in the backward direction.

3 Wrong variant 2: early coalescence output

Early coalescence output. Run the correct backward-doubling protocol with reuse, but during the composition $\Phi_{-1} \circ \cdots \circ \Phi_{-T}$, check for coalescence at *every* intermediate time step. As soon as the top and bottom chains agree (say after applying maps $\Phi_{-T}, \ldots, \Phi_{-T+k}$, corresponding to time $-T+k$), output the coalesced state **at that intermediate time** instead of continuing to time 0.

Once the chains agree, they remain identical under the remaining maps $\Phi_{-T+k+1}, \ldots, \Phi_{-1}$, so stopping early looks harmless. But the coalesced state at time $-T+k$ is **not** the same as the state at time 0 — the remaining maps still act:

$$\Phi_{-1} \circ \cdots \circ \Phi_{-T+k+1}(x) \;\neq\; x \quad \text{in general.} \tag{3.1}$$

By outputting x instead, we sample at a data-dependent time (the coalescence time within the window), which introduces the same stopping-time bias as forward coupling. Early coalescence is more likely when the maps happen to be strongly contracting, and such maps push toward "attracting" states — so the early-output distribution overweights them.

4 Test case: q-area measure on lattice paths

An up-right lattice path from $(0,0)$ to (N,M) is a sequence of N right-steps and M up-steps. The *area* $a(p)$ of a path p is the number of unit squares below it (equivalently, the sum of the height at each right-step). The **q-area measure** assigns

$$\pi_q(p) \;=\; \frac{q^{a(p)}}{\binom{N+M}{M}_q} \tag{4.1}$$

where the normalizing constant is the q-binomial coefficient.

Markov chain and monotone coupling

Encode a path as a binary string $s \in \{0,1\}^{N+M}$ (0 = right, 1 = up). At each step, pick a uniformly random adjacent pair (s_i, s_{i+1}). If $s_i \neq s_{i+1}$, apply the *heat bath* rule:

$$(s_i, s_{i+1}) \;\leftarrow\; \begin{cases} (1,0) & \text{with probability } \frac{q}{1+q} \quad (\text{"higher" path}), \\ (0,1) & \text{with probability } \frac{1}{1+q} \quad (\text{"lower" path}). \end{cases}$$

This satisfies detailed balance with respect to π_q. The update is **monotone**: using a shared uniform $u \in [0,1]$ for the heat bath, if the top chain is above the bottom chain before the update, it remains above afterwards. So CFTP requires tracking only the maximal path ($UU\cdots UR\cdots RR$) and the minimal path ($RR\cdots RU\cdots UU$).

Numerical experiment

The accompanying script `cftp_check.py` implements three sampling protocols for paths from $(0,0)$ to $(3,3)$ with $q = 0.3$ (there are $\binom{6}{3} = 20$ paths):

1. **Correct CFTP.** Backward doubling with reuse of random maps. At each doubling step, the composition $\Phi_{-1} \circ \cdots \circ \Phi_{-T}$ is applied to both the maximal path $UUU\,RRR$ and the minimal path $RRRUUU$. Coalescence is checked only at time 0 (after all maps have been applied). The common state at time 0 is the output.

2. **Forward coupling (wrong).** Both extreme paths start at time 0 and receive the same random updates $\Phi_1, \Phi_2, \ldots$ going *forward*. The algorithm stops at the first time τ when the two paths agree, and outputs the common state X_τ.

3. **Early coalescence output (wrong).** Same backward-doubling protocol as (1), with the same reused maps, but during the composition $\Phi_{-T}, \ldots, \Phi_{-1}$ the algorithm checks for coalescence at every intermediate step. When the two paths first agree (at some time $-T + k < 0$), the coalesced state at that moment is output *without applying the remaining maps to time* 0.

4. **Early coalescence every 5 steps (wrong).** Same as (3), but coalescence is checked only every 5th step of the composition (plus at time 0). Checking less frequently reduces the bias compared to checking at every step, since the output state has been pushed closer to time 0 on average — but the bias does not vanish.

The script draws 50 000 samples from each method and compares the empirical frequencies against the exact q-area probabilities via total variation distance.

The correct CFTP matches the exact distribution up to sampling noise. All wrong variants show systematic deviations: forward coupling and early-output-every-step dramatically overweight "middle" paths (the coalescence attractors). The every-5-steps variant has the same qualitative bias but reduced in magnitude — checking less frequently pushes the output closer to time 0, partially attenuating the stopping-time effect.

Remark 4.1. The q-area distribution is a natural test case because: (a) the exact probabilities are known in closed form, (b) the state space is small enough to compare full distributions, and (c) the monotone coupling makes CFTP efficient.

References

[PW96] J. Propp and D. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Struct. Algor. **9** (1996), no. 1-2, 223–252. ↑1