# Efficient Use of Exact Samples

Duncan Murdoch

Queen's University, Canada

University of Western Ontario

Jeffrey Rosenthal

University of Toronto

June 2, 1998

# Introduction

Propp and Wilson (1996) described a protocol called *coupling from the past* (CFTP) for exact sampling from the steady-state distribution of a Markov chain Monte Carlo (MCMC) process.

- a past time is identified from which the paths of coupled Markov chains starting at every possible state would have coalesced into a single value by the present time

- this value is a sample from the steady-state distribution.

Producing an exact sample typically requires a large computational effort, but yields only one perfect sample. This paper addresses the question:

**How can we make computationally efficient use of the sample paths in the CFTP algorithm?**

## Coupling from the Past (CFTP)

We wish to sample from a probability distribution $\pi(\cdot)$ on a state space $\mathcal{X}$, and have an ergodic Markov chain $P(x, \cdot)$ having $\pi$ as its stationary distribution.

**Start coupled Markov chains from all possible starting points, from arbitrarily far backwards in time, and get them all to coalesce to a single value at time $t = 0$.**

Random variables:

$$
\begin{aligned}
N_{x,t} &= \text{the next value a chain in state } x \text{ will take} \\
N_{x,t}^{s} &= \text{the value a chain will take at time } s \text{ if it starts} \\
&\phantom{=} \text{ at } x \text{ at time } t \\
G_t &= \text{a compatible chain running in stationarity}
\end{aligned}
$$

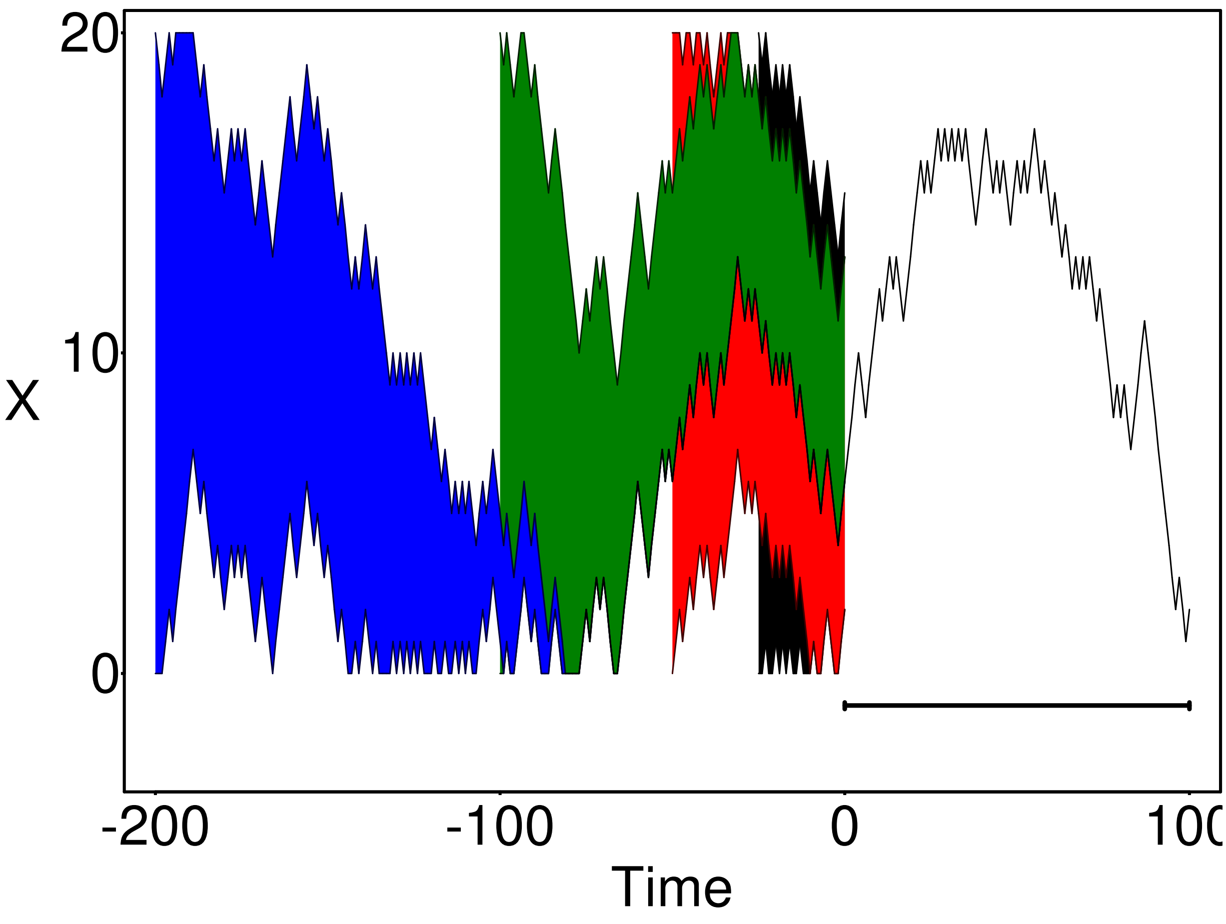
Values $N_{x,t_1}$ and $N_{y,t_2}$ are independent for $t_1 \neq t_2$.

CFTP: Find a time $-T$ such that $N_{x,-T}^{0}$ does not depend on $x$. Then $N_{x,-T}^{0} = G_0 \sim \pi(\cdot)$.

Only a (random) finite number of steps need to be calculated, to get an exact sample from $\pi$!

There are many tricks (monotone coupling, dominated CFTP, gamma coupling, etc.) to aid in finding $T$ and calculating $N_{x,t}$.

## Illustration of CFTP




Underlined (positive time) area is simulated using a standard MCMC method. Earlier draws use CFTP with doubling from an initial guess of -25 to find $T$.

## Tours using CFTP

**Repeated CFTP (RCFTP):**

1. Run ordinary CFTP to obtain a sample $X \sim \pi(\cdot)$.

2. For some fixed value $T_0$, set $X_0 = X$, and run any convenient implementation of the Markov chain forwards in time for an additional $T_0$ steps.

3. The $T_0 + 1$ values observed represent one tour.

4. Repeat independently.

Tours are IID; each value is an exact sample.

**Forward Coupling (FC):**

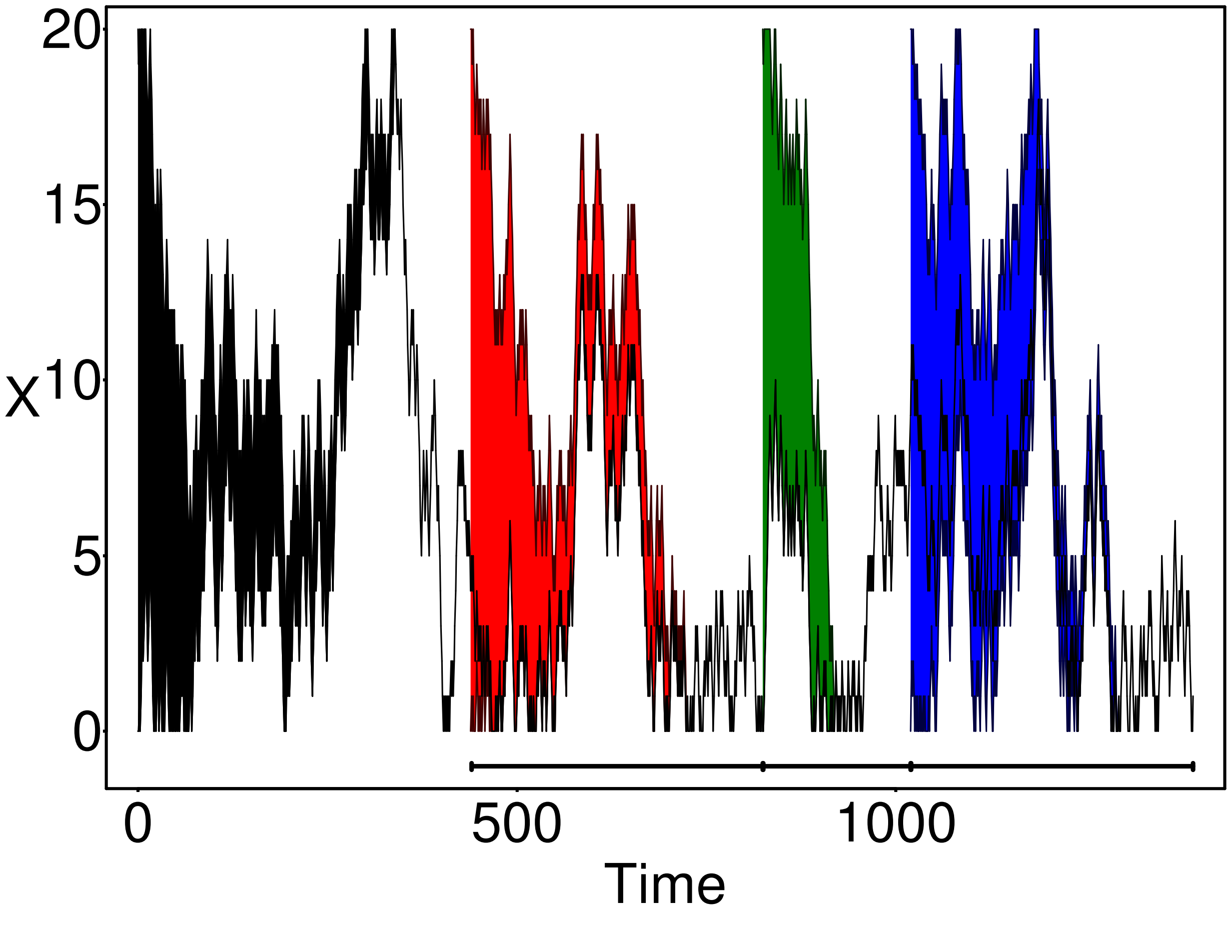
1. Run the chain *forwards* in time, from time 0 onward, until all paths have coalesced.

2. Keep track of one particular path. Use it as one tour.

3. Repeat (using the coalescing value of one run as the starting value of the next run, but not counting it twice).

Tours are identically distributed, but not independent. Non-adjacent tours are independent ("MA(1)" tours). Draws are not from $\pi(\cdot)$.

## Illustration of FC



## More tours...

### Concatenated CFTP (CCFTP):

1. Run ordinary CFTP to obtain a sample $X \sim \pi(\cdot)$.
2. Run it again to obtain a sample $Y \sim \pi(\cdot)$, a value of $T$, and a realization of $N_{x,t}$ for $t \in -T, \ldots, -1$.
3. The first tour of length $T$ consists of the values in the path from $X$ to $Y$ using $N_{x,t}$ to define the transitions, counting $Y$ but not $X$.
4. Repeat using $Y$ in place of $X$ to generate subsequent tours.

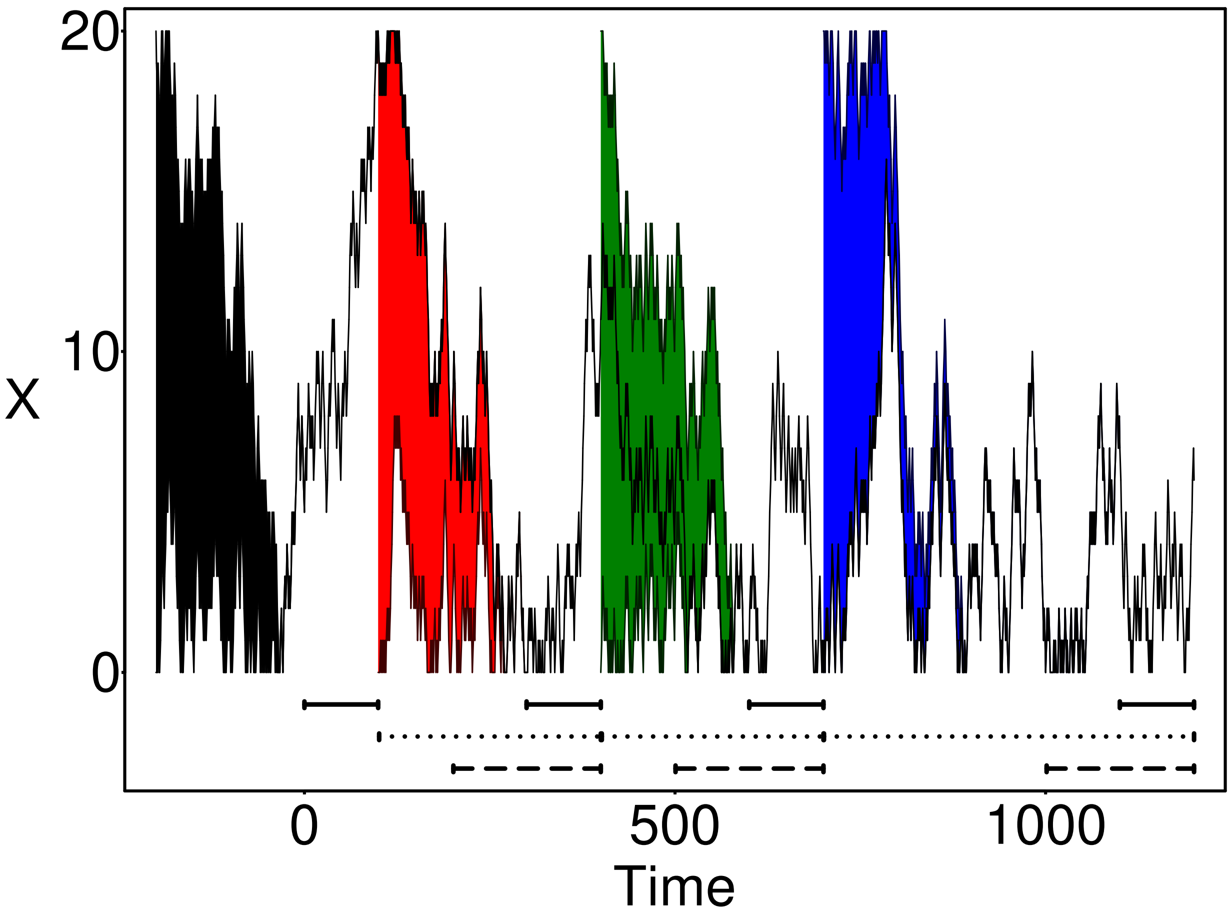
MA(1) tours; elements are stationary draws.

### Guarantee Time CFTP (GTCFTP):

1. Fix a value $T_g > 0$ to use as an initial guess for $T$ in the CFTP algorithm.
2. Run ordinary CFTP to obtain a sample $X \sim \pi(\cdot)$.
3. Run it again to obtain a sample $Y \sim \pi(\cdot)$, a value of $T \geq T_g$, and a realization of $N_{x,t}$ for $t \in -T, \ldots, -1$.
4. The first tour of length $T$ consists of the final $T_g$ values in the path from $X$ to $Y$ using $N_{x,t}$ to define the transitions.
5. Repeat using $Y$ in place of $X$ to generate subsequent tours.

MA(1) tours; elements are stationary draws.

## Illustration of RCFTP, CCFTP and GTCFTP




Solid line: values for RCFTP

Dotted line: values for CCFTP

Dashed line: values for GTCFTP

## Estimation

In studying $X \sim \pi(\cdot)$, we are often interested in $\pi(g) = \mathbf{E}\,(g(X))$.

The usual MCMC estimator: average over the chain:

$$\widehat{\pi}(g) = (1/T)\sum_{i=1}^{T} g(X_i)$$

CFTP gives us IID samples from $\pi(\cdot)$. Should we just average those?

Mykland *et al.* (JASA, 1995) broke up the chain into IID tours, which *regenerated* when they returned to a fixed state. Can we use tours?

We have values in perfect stationarity. Can we do more?

$$\widetilde{\pi}(g) = (1/N)\sum_{i=1}^{N} \frac{\sum_{j=1}^{T_i} g(X_{ij})}{T_i}$$

is biased for regeneration tours. Can we use CFTP to make it unbiased?

## Comparison

| | RCFTP | FC | CCFTP | GTCFTP |
|---|---|---|---|---|
| Direction | $-$ | $+$ | $-$ | $-$ |
| # of samples | $T_0 + 1$ | $T + T_0$ | $T^* + T_0$ | $T_g + T_0$ |
| $\widetilde{\pi}(g)$ biased | No | Yes | Yes | No |
| Tours | IID | MA(1) | MA(1) | MA(1) |

Notes:

1. Bias of $\widetilde{\pi}(g)$ in CCFTP comes from dependence between tour values and tour length.

2. When $T$ is found by testing all values, $T$ and $T^*$ have the same distribution.

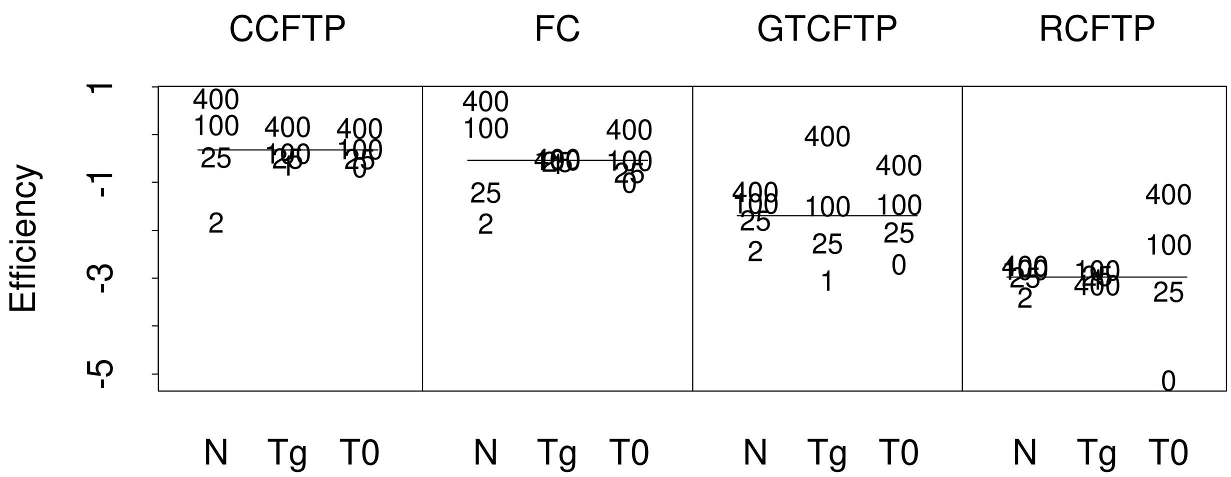
3. When the search for $T$ is done by doubling test values, $T$ is about twice as big as $T^*$.

4. All the CFTP methods cost about the same to evaluate; FC is easier.

5. With some models, simple MCMC steps are faster to evaluate than coupled ones.

## Example

- Random walk on $0, \ldots, 20$.
- $N_{x,t} = \max[0, \min(20, x \pm 1)]$ gives *monotonicity*: paths never cross.
- Simulate only states 0 and 20; others are sandwiched between.
- $\pi(X = x) = 1/21$ for $x = 0, \ldots, 20$.
- Simulated 1000 runs with 4 values of $N$, $T_g$ and $T_0$.




Efficiency = $\log_2$(Precision/Cost)

## Conclusions

1. Naive use of one value per CFTP run is not recommended.

2. To get unbiased IID tours, use RCFTP with large $T_0$. Use $N$ large enough to get a reasonable variance estimate from an IID sample (e.g. $N > 30$). Use $T_g$ large enough to avoid long search.

3. To get unbiased MA(1) tours, use GTCFTP. Use large enough $T_g$ to avoid many doublings, decrease the variance by using large $T_0$.

4. In other cases, use FC (or Mykland *et al.* methods). Use large $N$, so that tour length variability doesn't make estimates unstable.

For more on CFTP and exact sampling, see David Wilson's web page
`http://dimacs.rutgers.edu/~dbwilson/exact.html/`.

For our full paper, see
`http://www.mast.queensu.ca/~dmurdoch`.