– 428 –

NOTES ON PERFECT SIMULATION

by

W.S. Kendall

July 2004, revised February 2005

DEPARTMENT OF STATISTICS

UNIVERSITY OF WARWICK

– 428 –

NOTES ON PERFECT SIMULATION

by

W.S. Kendall

July 2004, revised February 2005

NOTES ON PERFECT SIMULATION

Wilfrid S. Kendall
Department of Statistics, University of Warwick

11th February 2005

This work was carried out while the author was
visiting the Institute for Mathematical Sciences,
National University of Singapore in March 2004.
The visit was supported by the Institute.

Version 2.19

Abstract

Perfect simulation refers to the art of converting suitable Markov Chain Monte Carlo algorithms into algorithms which return exact draws from the target distribution, instead of long-time approximations. The theoretical concepts underlying perfect simulation have a long history, but they were first drawn together to form a practical simulation technique in the groundbreaking paper of Propp and Wilson [78], which showed (for example) how to obtain exact draws from (for example) the critical Ising model. These lecture notes are organized around four main themes of perfect simulation: the original or classic Coupling From The Past algorithm (*CFTP*); variations which exploit regeneration ideas such as small-set or split-chain constructions from Markov chain theory (small-set *CFTP*); generalizations of *CFTP* which deal with non-monotonic and non-uniformly ergodic examples (dominated *CFTP*); and finally some theoretical complements.

Contents

Introduction

Perfect simulation refers to the art of converting suitable Markov Chain Monte Carlo (*MCMC*) algorithms into algorithms which return exact draws from the target distribution, instead of long-time approximations. The theoretical concepts underlying perfect simulation have a long history, but they were first drawn together to form a practical simulation technique in the ground-breaking paper of Propp and Wilson [78], which showed how (for example) to obtain exact draws from the critical Ising model on a finite lattice.

These notes derive from a series of four tutorial lectures given at the Institute for Mathematical Sciences, National University of Singapore, in March 2004, to an audience of PhD students and recent post-docs. The aim of the lectures was to introduce the collection of ideas which have developed around perfect simulation, since some of the audience might well have occasion to use the technique, and since in any case exposure to these ideas promotes useful lateral thinking about *MCMC*. I have tried to be rigorous, in the sense of avoiding mis-statements, but have not attempted to give a complete account. Some proofs and techniques are merely sketched, while some are omitted. In the actual lectures it was possible to illustrate many of the ideas using computer animations; this is not an option for printed notes, but in partial recompense I have included illustrations where possible. I have aimed the exposition at the level of a mathematically-trained graduate student; so examples are chosen to be illustrative rather than representative. It *is* possible to use *CFTP* for other than toy problems, but such examples would require detailed descriptions which would obscure intuition.

The lectures and these notes have alike been organized around four main themes of perfect simulation: the original or classic Coupling From The Past algorithm (*CFTP*) in §1; variations which exploit regeneration ideas such as small-set or split-chain constructions from Markov chain theory (small-set *CFTP*) in §2; generalizations of *CFTP* which deal with non-monotonic and non-uniformly ergodic examples (dominated *CFTP*) in §3; and finally in §4 some striking results relating *CFTP* to an apparently different algorithm due originally to Fill, as well as other theoretical complements.

The topic of perfect simulation is made up of a variety of interacting ideas rather than a single grand theory: more of an orchestra of complementary techniques than a virtuoso *prima donna* of a Big Theory. I hope that these notes will help to convey this variety, and to help others to engage with, be stimulated by, and contribute to the topic.

Useful reading

Here is a sample of useful resources concerning perfect simulation (particularly *CFTP*) and the underlying coupling ideas:

- Lindvall's introduction to coupling [62] (now available as [63]) should be required

reading for all applied probabilists, and lays invaluable foundations for an appreciation of coupling theory;

- Thorisson's monograph [92] gives a masterly exposition of the mathematical theory of coupling as well as a treatment of *CFTP* itself; Häggström's short undergraduate text [39] provides a most accessible introduction at the level of discrete Markov chains; finally, the monograph [71] of Møller and Waagepetersen provides significant material from the perspective of stochastic geometry (a major consumer of *CFTP*!).

- It is not possible in the short space afforded by these notes to be complete in assigning bibliographic credit, nor to give adequate coverage to the various applications of *CFTP*. The online bibliography developed by David Wilson should be the first port of call when seeking references to *CFTP*:

  `http://research.microsoft.com/~dbwilson/exact/`

- Finally, note that Wilson's online bibliography links to various useful tutorial essays; in particular we mention Casella *et. al.* [18], Dimakos [26] and Thönnes [90].

1 CFTP: the classic case

We begin this section with a brief indication of how *CFTP* fits in to the theory of *MCMC*. We then discuss one of the simplest possible examples of coupling (§1.1), before describing classic *CFTP* as applied to the doubly-reflecting random walk (§1.2). This serves as introduction to the fundamental theorem of *CFTP* (§1.3), which is further illustrated by two simple applications: to the dead leaves model (§1.4) and to the Ising model (§1.5). The section is completed by a rather less trivial application to point processes (§1.6) and a discussion of *CFTP* in space and time (§1.7), and finally a brief note on some historical and other complementary aspects of *CFTP* (§1.8).
MCMC arises in a number of different areas of mathematical science, with different emphases. (This makes interaction interesting and fruitful!) Here are some examples, several of which are discussed at length in other chapters in this volume:

- **Statistical mechanics.** Are there phase transition phenomena in specific infinite-dimensional systems? How do they behave?
- **Computer science.** Approximate counting problems can be solved in terms of algorithms which deliver approximately uniform random samples, which in turn can be solved using *MCMC*. In this area the key question is, how does the algorithm behave as the scale of the problem increases? Does the run-time increase exponentially, or polynomially?
- **Image analysis.** Given a noisy picture with some kind of geometric content: can we clean it up using modelling by spatial random fields? Can we identify significant features?
- **Statistics.**
  - **Bayesian.** Can we draw accurately (and *quickly* if possible!) from the posterior distribution on a space which may be low-dimensional but not at all symmetric?
  - **Frequentist.** What does the likelihood surface look like?

The paradigm for Markov chain Monte Carlo (*MCMC*) runs as follows. We want to understand the properties of a particular probability measure, which may be linked to a complicated state space, or may be specified in an indirect manner, or may in some other way be hard to deal with by explicit calculation. So we design a suitable Markov chain whose long-run equilibrium distribution is this probability measure. Sometimes this chain will arise naturally from the application context (if for example we are interested in the statistical equilibrium of a financial time series); sometimes it is suggested by the specification (if for example the probability measure is specified up to a normalization

factor as for the Ising model, so that we can use ideas of detailed balance and reversibility to design appropriate Markov chains). However the chain arises, we require that the target probability measure is the long-run equilibrium measure. We can then draw samples whose distribution is at least approximately the target probability measure, by running the chain for a time which is long enough for statistical equilibrium to be established at least approximately.
Thus the paradigm runs as follows:

- specify the target distribution indirectly;
- realize it as the equilibrium of a Markov chain;
- sample *approximately* from the target distribution by running the Markov chain for a long time (till it is near equilibrium).

A major question is, what is the length of the "burn-in" period, the period till the chain is near equilibrium? Options for answering this question are:

> Guess it or diagnose it from simulation output [15, 22];
>
> Or estimate it, analytically [25, 80, 83, 84], or empirically [47].

The question is, whether it is ever possible to do better than the above?
In a landmark paper, Propp and Wilson [78] showed how in in principle one can modify *MCMC* algorithms so that they deliver *exact* draws from the chain equilibrium distribution, at a price of random run-time length: the technique of *exact* or *perfect simulation*. Moreover they showed how such modifications can be constructed to provide exact draws in feasible computation time for interesting and non-trivial examples. Since then there has been a flood of work on the Propp-Wilson idea of *Coupling from the Past* (*CFTP*). In this lecture we will introduce *CFTP* by describing simple cases, which we will develop into examples of interest particularly in Bayesian statistics and stochastic geometry (the study of random patterns).
Before beginning this task, we should note there is another important issue to consider when undertaking *MCMC*: the best chains not only have short or at least manageable burn-in periods, but also *mix* rapidly (time series of [functions of] observations exhibit rapidly decaying correlation). Perfect simulation does not address this issue directly – though the challenge of devising modifications to ensure perfect simulation *may* suggest ways of improving the mixing rate.

1.1 Coupling and convergence: the binary switch

We commence by introducing the fundamental idea of *coupling*[a] (see [63, 92] for more on this large subject).
Consider the simplest possible case: a continuous-time Markov chain with just two states, which makes transitions from one state to the other at constant rate $1/\alpha$ (the *binary switch*). With care, we can simulate simultaneously from different starting points in such a manner that the two simulations *couple* (start to take the same values) from some random *coupling time* T onwards.

Algorithm 1.1 *Supply*

(a) a Poisson process (rate $1/\alpha$*) of* $0 \to 1$ *transitions,*

(b) independently a Poisson process (rate $1/\alpha$*) of* $1 \to 0$ *transitions.*

Use the transitions to build coupled processes X, Y *begun at* 0, 1 *(say). Do this as follows: each time a* $0 \to 1$ *transition appears, set* X *to* 1*. Each time a* $1 \to 0$ *transition appears, set* X *to* 0*. Do the same for* Y*. Clearly* X, Y *are (coupled) copies of the binary switch, coupling at the time* T *of the first Poisson incident, after which they evolve in lock-step.*

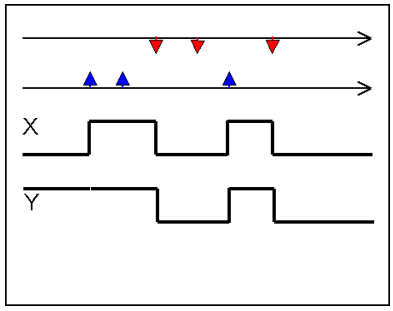


Figure 1: Coupled binary switches X and Y (lower two rows), driven by the same sequences of $0 \to 1$ and $1 \to 0$ transitions (upper two rows)

[a] Coupling, stochastic flows, also the notion of stochastic recursive sequences, arise in different parts of probability and stochastic analysis, but all express the same idea: one can realize the Markov chain of interest by a specific construction which allows one to compare different copies of the Markov chain begun at different starting points; the construction is not completely determined by the Markov chain but can be varied so long as single trajectories have the right distribution.

Then the classic *coupling inequality* argument shows

$$\begin{aligned} \mathrm{dist}_{\mathrm{TV}}(X_t, \pi) &= \sup_A \{\mathbb{P}[X_t \in A] - \pi(A)\} = \frac{1}{2}\sum_i |\mathbb{P}[X_t = i] - \pi_i| \\ &= \sup_A \{\mathbb{E}[\mathbb{I}[X_t \in A] - \mathbb{I}[X_t^* \in A]]\} \leq \sup_A \{\mathbb{E}[\mathbb{I}[X_t \in A \text{ but } X_t^* \neq X_t]]\} \\ &= \mathbb{E}[\mathbb{I}[X_t^* \neq X_t]] = \mathbb{P}[T > t] \quad (1) \end{aligned}$$

(with $\mathbb{I}[A]$ representing the indicator random variable for the event A), where

(1) π is the equilibrium distribution, so $\pi(0) = \pi(1) = 1/2$;

(2) X^* is a (coupled) copy of the Markov chain X started off in equilibrium (hence lying between X and Y, and continuing to do so if driven by the construction above);

(3) and $\mathrm{dist}_{\mathrm{TV}}$ is the total variation distance. (Note, this is a rather strong measure of distance from equilibrium; two real-valued random variables can almost surely have values very close together, and yet have maximum distance in total variation if one takes only rational values and the other takes only irrational values! Other kinds of coupling relate to more metric notions of distance.)

The coupling argument generalizes to arbitrary Markov chains:

(a) if we can couple a general Markov chain X to a version Y in statistical equilibrium, then such a coupling bounds the approach to equilibrium through Equation (1);

(b) if we allow *non-adapted* couplings then the bound is sharp [36, 38];

(c) however, non-adapted couplings can be very difficult to construct! *Co-adapted* couplings are typically easier to construct, and can supply usable bounds but in many cases these will not be sharp. (This point arises again in §4.4.)

Can we use such a coupling to *draw* from equilibrium? The binary switch example is deceptive: $X(T)$ is in equilibrium in the case of the binary switch, but not in general – a defect which becomes apparent even in one of the simplest imaginable generalizations, which we will now discuss.

1.2 Random walk *CFTP*

Consider the natural generalization of the above coupling, but applied (in discrete rather than continuous time) to the random walk X on $\{1, 2, \ldots, N\}$ which is *reflected* at the boundary points 1, N. Reflection here is implemented as follows: if the random walk

tries to move outside of the range $\{1, 2, \ldots, N\}$ then the relevant transition is simply disallowed (this is directly analogous to the way in which the binary switch behaves). We then obtain *synchronous* coupling (synchronously coupled random walks move up and down in parallel, except where prevented by barriers from moving in synchronization): the coupled random walks can only meet together at the barrier levels 1, N. Thus $X(T)$ cannot be a draw from equilibrium if $N > 2$.
The Propp-Wilson idea circumvents this problem by drawing on a well-established theme from ergodic theory: realize a Markov chain as a stochastic flow and evolve it not into the future but *from the past*! If we do this then we need to consider coupled realizations of the Markov chain started at all possible starting points. However if monotonicity is present then we need only focus on maximal and minimal processes, as for the binary switch in Section 1.1:

$X^{\text{lower},-n}$ begun at 1 at time $-n$,

$X^{\text{upper},-n}$ begun at N at time $-n$;

since the synchronous coupling arranges for these to *sandwich* all other realizations begun at time $-n$. We can therefore carry out an algorithm which is summarized informally below, and which is illustrated in Figure 2.

Algorithm 1.2

- *Run upper and lower processes from time $-n$.*
- *If the processes are coupled by time 0, return the common value.*
- *Otherwise, repeat but start at time $-2n$ (say), re-using randomness whenever possible.*

It is informative to consider a crude implementation of *CFTP* for this simple case, for example using the freely available statistical package `R` (see `http://cran.r-project.org/`). First define a list `innov` of innovations determining the evolution from time -2 to time 0.

```
innov <- 2*rbinom(2,1,1/2)-1
```

Now construct a function `cycle` which simulates maximal (`upper`) and minimal (`lower`) reflecting random walks on the state space $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ using `innov`. This function returns the common value if maximal and minimal processes coalesce; otherwise it returns `NA`.

```
cycle <- function (innov) {
 upper <- 10
 lower <- 1
```

```
for (i in 1:length(innov)) {
 upper <- min(max(upper+innov[i],1),10)
 lower <- min(max(lower+innov[i],1),10)
 }
if (upper!=lower) return(NA)
upper
}
```

If `cycle(innov)` returns `NA` (and clearly in this example it has to do so at least until the innovation length is sufficient to permit one of the maximal and minimal processes to cross from one end of the state space to the other) then further innovations must be inserted at the beginning of the `innov` vector, and `cycle(innov)` invoked again. This is conveniently packaged in a `while` loop.

```
while(is.na(cycle(innov)))
   innov <- c(2*rbinom(length(innov),1,1/2)-1, innov)
cycle(innov)
```

Of course `R` is not well-suited to this kind of algorithm, other than for purely illustrative purposes: much better results can be obtained using modern scripting languages such as *Python* (`http://www.python.org/`), particularly with *numarray* extensions (`http://www.stsci.edu/resources/software_hardware/numarray`). Figure 2 shows the effect of four cycles, resulting in a common value at time 0 on the third cycle.

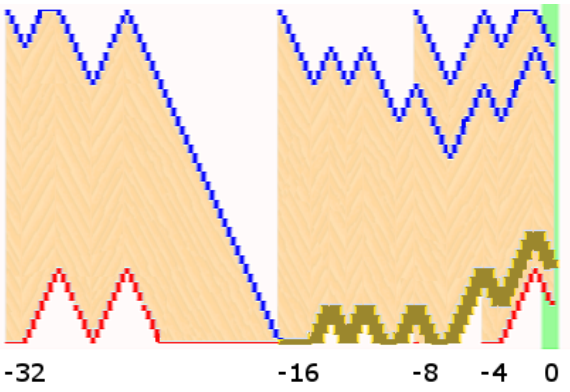


Figure 2: Classic *CFTP* for a reflecting random walk. Coalescence occurs at time -16 for this realization.

Various issues are illustrated in this figure:

- The algorithm extends the common path backwards into the past, not forwards into the future;
- One must use common randomness (in coupling random walks together) and *re-use* it (when coming from the past into a time interval within which random walks have already been simulated);
- One samples at time 0, *not* at the coupling time;
- There is a simple *rationale* for doubling the start-time $-n \to -2n$: this essentially represents a binary search for the coalescence time.

It is informative to consider what goes wrong if one deviates from this algorithm:

- suppose one runs the simulation into the future, not from the past, stopping (say) at a specified time t after coupling has first occurred. Since coupling occurs only at the boundaries, it can be shown in this case that the approximation to the equilibrium distribution is no better than if one had omitted the initial coupling phase completely!
- suppose one fails to re-use randomness. We expect this to bias towards cases in which coalescence occurs earlier in algorithmic time (since failure to re-use randomness would improve the chances of fast coalescence, essentially by permitting repeated attempts to coalesce over time intervals $[-2^k n, 0]$), and this is indeed the case;
- Sampling at coupling time instead of time 0 is obviously a bad idea; sampling an *independent* random walk at this time will still give a biased result.

1.3 The *CFTP* theorem

Morally the proof of classic *CFTP* is just 3 lines long. We express the coupling for X in terms of random *input-output maps* $F_{(-u,v]} : \mathcal{X} \to \mathcal{X}$, so $F_{(-n,t]}(x_0)$ is X_t begun at time $-n$ with the value $X_{-n} = x_0$.

Theorem 1.3 *[78] If* coalescence *is almost sure in Algorithm 1.2 (all inputs x_0 result in the same single output $F_{(-n,0]}(x_0)$ for large enough n) then* CFTP *samples from equilibrium.*

Proof: For each time-range $[-n, \infty)$ use the $F_{(-n,t]}$ to define

$$X_t^{-n} \quad = \quad F_{(-n,t]}(0) \qquad \text{for } -n \le t\,.$$

Finite coalescence time $-T$ is assumed. So

$$\begin{array}{rcll} X_0^{-n} & = & X_0^{-T} & \text{whenever } -n \leq -T\,; \\ \mathcal{L}\left(X_0^{-n}\right) & = & \mathcal{L}\left(X_n^0\right)\,. & \end{array}$$

If X converges to an equilibrium π in total variation dist_{TV} then

$$\text{dist}_{\text{TV}}(\mathcal{L}\left(X_0^{-T}\right),\pi) = \lim_n \text{dist}_{\text{TV}}(\mathcal{L}\left(X_0^{-n}\right),\pi) = \lim_n \text{dist}_{\text{TV}}(\mathcal{L}\left(X_n^0\right),\pi) = 0$$

hence the result. □

There is a crucial step in the classic proof of uniqueness and existence of long-run equilibrium for finite Markov chains which actually amounts to the assertion that coalescence is almost sure even for the *independent* coupling (chains evolve independently till they meet, then stick together). This is the step which argues that under aperiodicity and irreducibility there is an n such that all the n-step transition probabilities $p_{ij}^{(n)}$ are simultaneously positive.

Remark 1.4 *We are free to choose* any *"backwards random time" $-T$ so long as we can guarantee coalescence of $F_{(-T,0]}$. The binary search approach of random walk* CFTP *is deservedly popular, but there are alternatives: for example the "block-by-block" strategy of read-once* CFTP *(§2.6).*

Remark 1.5 *Monotonicity of the target process is* convenient *for* CFTP, *but not* essential. *Propp and Wilson [78, §3.2] use lattice theory to formalize the use of monotonicity. In §3.6 below we describe the* crossover trick *[51] for use in anti-monotonic situations.*

1.4 The falling leaves of Fontainebleau

A very visual and geometric application of *CFTP* in mathematical geology [55] was well-known to workers in the field well before the introduction of *CFTP* itself: it concerns the "dead-leaves" model, inspired by the falling leaves of Fontainebleau. The dead-leaves model describes a random mosaic as the limiting distribution of the random pattern obtained by allowing patterned tiles ("leaves") to fall at random on a window. Figure 3 shows the pattern beginning to build up. We can think of the "dead-leaves" process as a Markov chain with states which are elements of some "pattern space".
David Wilson has introduced the terminology *occlusion* CFTP for this kind of *CFTP*: the algorithm builds up the result piece-by-piece with no back-tracking, and the eventual perfect image is built up progressively, with each new portion "occluding" further developments in the corresponding region.

Figure 3: The falling leaves of Fontainebleau.

It is rather straightforward to make exact computations for the rate at which this chain attains equilibrium. However one can do better, very easily, by considering the pattern as it is built up, but from the perspective of looking up from underneath, rather than from on top looking down! Elementary probability arguments show, at any given time the pattern distribution is the same from either perspective. On the other hand the pattern viewed from below will stop changing as soon as complete coverage is attained; and it is then a simple matter to conclude that at that time (the time of complete occlusion) one obtains a draw from the required equilibrium distribution (this argument is actually close to that of the proof of Theorem 1.3: $F_{(-n,t]}$ now represents the superposition of random leaves falling over the period $(-n, t]$). Hence

Corollary 1.6 *Occlusion* CFTP *as described above delivers a sample from the dead leaves distribution.*

Example 1.7 *Consider the process of simulating* forwards *in time till the image is completely covered. This will result in bias.*[b]

Remark 1.8 *Web animations of perfect simulation for the dead leaves model can be found at* `http://www.warwick.ac.uk/go/wsk/abstracts/dead/`.

Remark 1.9 *Other examples of occlusion* CFTP *include the Aldous-Broder algorithm for generating random spanning trees [1, 14, 94, 97].*

[b] Hint: consider a field of view small enough for it to be covered completely by a single leaf: argue by comparison that the forwards simulation is relatively more likely to result in a pattern made up of just one large leaf!

1.5 Ising *CFTP*

Propp and Wilson [78] showed how to make exact draws from the critical Ising model on a finite lattice, using Sweeny's [88] single-bond heat-bath (Huber [46] has shown how to make this work for the full Swendsen-Wang algorithm). A simpler application uses the single-site heat-bath sampler to get exact draws from the sub-critical Ising model. Recall that the Ising model has probability mass function proportional to

$$\exp\left(\frac{J}{2}\sum\sum_{i\sim j}\sigma_i\sigma_j\right),$$

with spins $\sigma_i = \pm 1$, and indices i, j running through the nodes of a square lattice. Here J is the inverse temperature of the system, while $i \sim j$ denotes that sites i and j are neighbours. The heat bath algorithm updates nodes i (in systematic or in random order) according to the conditional distribution of σ_i given the remainder of the configuration. We can couple evolutions of the heat-bath algorithm in a way which is similar to our coupling of evolutions of the binary switch or the random walk: calculate the conditional probability p that $\sigma_i = -1$, and determine the update by drawing a Uniform$(0, 1)$ random variable U, setting $\sigma_i = +1$ if $U > p$.

The resulting coupling is monotonic, and so we can derive classic *CFTP* for the Ising model, by comparing maximal and minimal processes run from the past (the broad details of implementation are the same as for the case of the random walk *CFTP* illustrated in R code above). The heat-bath algorithm works well in the sub-critical case: however as parameters approach criticality so it takes progressively longer for coalescence to be attained. Figure 4 shows snapshots taken from the approach to coalescence for a systematic scan Gibbs sampler: snapshots of the upper process run along the top, the lower along the bottom, and the difference is indicated in the middle. Coalescence is nearly achieved in this run: according to the *CFTP* algorithm one must then re-wind back to an earlier start-time and re-run the coupled simulations, taking care to re-use randomness when available.

This *CFTP* algorithm adapts well to changes in the underlying graph structure, so long as the model remains ferromagnetic and phase transition phenomena are avoided.[c] For example consider the *conditioned* Ising model,[d] as used in image analysis applications. In Figure 5 we show the results when the Ising model is conditioned by a particular noisy image: the conditioning can be balanced off against strong interactions between sites, as could be predicted (of course) from theoretical considerations [58]. In this case coalescence is achieved already in the first run, though of course this is not guaranteed!

[c] Coding techniques will deal with the anti-ferromagnetic case for bi-partite graphs: as noted above in Remark 1.5 we can use the *crossover trick* to deal with other cases.

[d] Note: the statistician's conditioning $\equiv$ the physicist's external magnetic field!

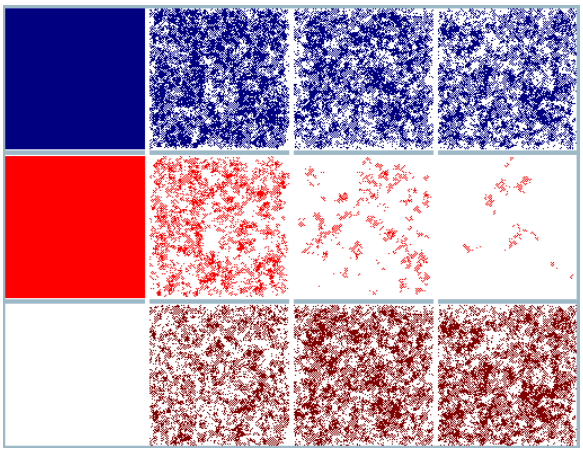


Figure 4: Classic *CFTP* for a sub-critical Ising model. Maximal and minimal processes occupy the upper and lower strips: the middle strip marks the sites at which maximal and minimal processes disagree. As agreement is not total at the end of the simulation (at time 0), it will be necessary to restart at an earlier time, re-using randomness when available.

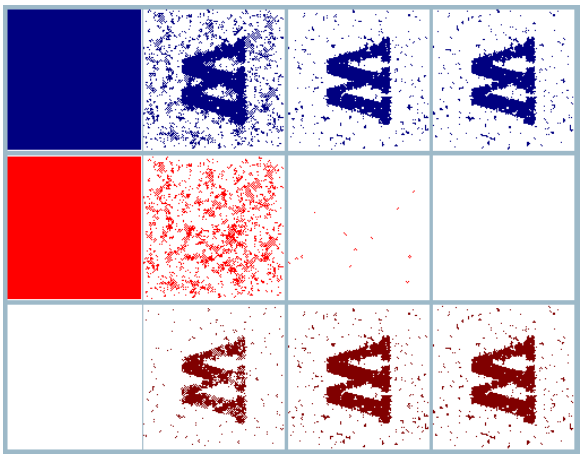


Figure 5: Classic *CFTP* for a conditioned Ising model. Maximal and minimal processes occupy the upper and lower strips: the middle strip marks the sites at which maximal and minimal processes disagree. Agreement is complete at the end of the simulation (at time 0), so the *CFTP* algorithm is then complete.

Remark 1.10 *Web animations of perfect simulations of conditioned Ising models can be found at*

`http://www.warwick.ac.uk/go/wsk/ising-animations/`.

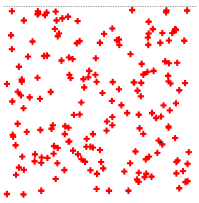
1.6 Point process *CFTP*

Classic *CFTP* is not limited to discrete models, as we have already seen in the case of the falling leaves model. We describe one further example: a perfect simulation procedure for attractive area-interaction point processes due to Häggström *et al.* [42].
The area-interaction point process[e] was proposed by Baddeley and Van Lieshout [7] as a model for random point patterns which can exhibit both clustering and repulsion. A succinct definition runs as follows: weight a Poisson process realization according to the area of the region of locations lying closer than r to some point of the pattern:

$$\text{pattern density} \quad \propto \quad \gamma^{-\text{area of region within distance } r \text{ of pattern}}. \tag{2}$$

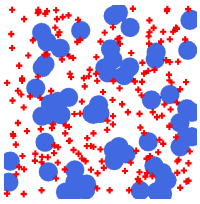
Remark 1.11 *If $\gamma > 1$ then the weighting favours patterns which group points close together (so as to reduce the area of the region); if $\gamma < 1$ then patterns are favoured which spread points away from each other.*
If $\gamma > 1$ (attractive case only!), then the above density is proportional to the probability that an independent Poisson process of suitable intensity places no points within distance r of the pattern.

Hence the area-interaction point process may be represented as the random pattern of red points generated by a point process of red and blue points, where red and blue points are distributed as Poisson patterns conditioned to be at least distance r from blue and red points respectively. This can be implemented as a (typically impracticable) rejection sampler: a more practical option is to use a Gibbs sampler, which is monotonic and so lends itself to *CFTP*.
Here is an illustrated step-by-step account of the Gibbs sampler.

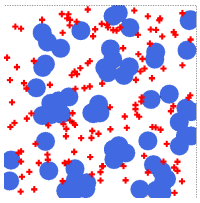


Construct a Poisson point process (centres of crosses).

[e] Known previously to physicists as the Widom-Rowlinson model [93].



Construct a new Poisson process (centres of discs), but censor all points of the new process such that a disc centred on the point overlaps centres of crosses.



Discard the old points formed from centres of crosses, construct a new Poisson process (centres of new crosses), but censor all points of the new process which would fall on a disc.

The Gibbs sampler cycles repeatedly through these last two steps, and the evolving pattern of cross centres converges in distribution to an attractive area-interaction point process; this is a consequence of the fact noted in Remark 1.11. Notice the duality between cross centres and centres of disks!

The *CFTP* construction is based on the observations

- there is a partial order for state pairs

  $$(\text{point pattern from cross centres}, \text{point pattern from disk centres})$$

  based on the order relationship $(\xi_1, \eta_1) \preceq (\xi_2, \eta_2)$ if $\xi_1 \subseteq \xi_2$ and $\eta_1 \supseteq \eta_2$;

- "highest" and "lowest" states under this ordering are $(\mathcal{X}, \emptyset)$ and $(\emptyset, \mathcal{X})$ where $\mathcal{X}$ is the full ground space: Note that these are "pseudo-states", and are never achieved by the target pattern itself!

The fact that there are highest and lowest pseudo-states is the key to the rather rapid rate of convergence exhibited by this algorithm (at least in parameter regions where there is no phase-transition effect): the underlying Markov chain is *uniformly ergodic* in a sense which we will make precise later (Definition 3.8), but which can be summarized by noting that convergence to equilibrium will always be at least as fast as convergence to equilibrium from one of the two extreme pseudo-states.
Issues of re-use of randomness can be dealt with by recording the entire new Poisson point pattern of disk centres or crosses introduced at each stage, and re-using this when appropriate.
Neither the Gibbs sampler nor the *CFTP* construction work for the non-attractive case. However we will see later (§3.4) how this may be overcome using a generalization of classic *CFTP*.

1.7 CFTP in space and time

When interactions are sufficiently weak (certainly weak enough that phase transitions cannot occur!) then the *CFTP* idea can be applied in space as well as time. In effect, one aims to capture a fragment of a virtual simulation in perfect equilibrium, for which the fragment is spatially limited as well as temporally limited. One does this by extending the simulation not only backwards in time, but also outwards in space [50]. In this case the binary search *rationale* must be modified according to the computational cost of extending the *CFTP* algorithm in *both* space *and* time. Interesting related theory is to be found in [41].

It is clear that phase transition phenomena will cause problems when we attempt to conduct *CFTP* in space as well as time; there will be a positive chance that the upper and the lower processes for the critical or supercritical Ising model simply do not coalesce at all if the spatial grid is being enlarged as well as extending the heat bath back in time. (Here of course is where several of the other chapters of this monograph start their story!) The *BFA* algorithm, described in §4.6, investigates further the relationship between this issue and percolation phenomena.

1.8 Some Complements

"The conceptual ingredients of *CFTP* were in the air" [79] for a long time before *CFTP* was formulated explicitly. For example consider:

- very early work by Kolmogorov [59][f] discusses chains begun in the indefinite past;
- the use of coupling between Markov chains by Doeblin [28];
- application of coupling from the past to study queueing and storage systems (to establish very general equilibrium theorems) in [48, 64];
- use of evolution of stochastic flows into the past not the future in [57, 61];
- the notion of stochastically recursive sequences (SRS) appearing in stochastic systems theory [34, 9, 10];
- the use of occlusion-type constructions to deliver samples from equilibrium in [1, 14];
- and genuine coupling from the past constructions used for theoretical ends in [91].

[f] Thanks to Thorisson [92] for this reference ....

However, it was not until Propp and Wilson [78] that everything was put together to show the then startling fact, that these ideas could produce exact draws from equilibrium for non-trivial Markov chains.

As §1.7 indicates, *CFTP* ideas can be applied in space not only in time. For another example, in [13] it is shown how to make perfect draws (without edge effects) from clustered random patterns with long-range interactions. Møller and Rasmussen [69] apply these ideas (together with notions of *domCFTP*– see §3.4 below) to self-exciting point processes.

Coupling and simulation are theoretical and implementational counterparts, with considerable twinning between techniques on the two areas. *CFTP* brings the two areas together in a very practical way. Other practical links are mentioned below in §2.7.

2 *CFTP* and regeneration

An early misconception about *CFTP* was that it could be applied only to monotonic Markov chains. We have already seen a mild counterexample: monotonicity is not particularly evident in the "dead leaves" model (though it can be forced into a monotonic framework using the notion of "region of occlusion"). More general treatments use ideas of regeneration, which we now introduce.
We begin by summarizing the theory of Markov chain small sets (§2.1), a theoretical discretization method which allows us to perform small-set *CFTP* for Markov chains on continuous state space (§2.2). We then survey variations on this theme: slice sampling (§2.3), the multi-shift sampler (§2.4), catalytic *CFTP* (§2.5), read-once *CFTP* (§2.6). These variations are all part of the tool-set for successful application of *CFTP* in practice. We conclude with a brief discussions of some more technical complements to small-set *CFTP* (§2.7).

2.1 Small sets

Suppose we desire to construct a coupling between two random variables X, Y yielding a maximal positive chance of $X = Y$ and otherwise not subject to any constraint. (This is related to the notion of *convergence stationnaire*, or "parking convergence", from stochastic process theory.) Clearly this coupling is relevant to *CFTP*, where we aspire to coalescence!
Given two overlapping probability densities f and g, we can implement such a coupling (X, Y) as follows:

- Compute $\alpha = \int (f \wedge g)(x)\, \mathrm{d}\, x$.
- With probability α return a draw of $X = Y$ from the density $(f \wedge g)/\alpha$.
- Otherwise draw X from $(f - f \wedge g)/(1 - \alpha)$ and Y from $(g - f \wedge g)/(1 - \alpha)$.

This is closely related to the method of *rejection sampling* in stochastic simulation.
From Doeblin's time onwards, probabilists have applied this to study Markov chain transition probability kernels:

Definition 2.1 ***Small set condition:*** *Let X be a Markov chain on a state space $\mathcal{X}$, transition kernel $p(x, \mathrm{d}\, y)$. The set $C \subseteq \mathcal{X}$ is a* small set of order[g] *k if for some probability measure ν and some $\alpha > 0$*

$$p^{(k)}(x, \mathrm{d}\, y) \quad \geq \quad \mathbb{I}\left[C\right](x) \times \alpha\, \nu(\mathrm{d}\, y)\,. \tag{3}$$

Here $\mathbb{I}\left[C\right](x)$ is the indicator function for the set C.

[g]The notion of the *order* of a small set is understated in most textbook treatments, but is important for the purposes of *CFTP*.

It is helpful to contemplate the simple example of a Markov chain on the unit interval whose transition density $p(x, y)$ is triangular with peak at x (Figure 6). Here the small set is the whole state space $\mathcal{X} = [0, 1]$, of order $k = 1$, $\alpha = 1/2$, and with probability measure ν given by the isoceles triangle density over $[0, 1]$.

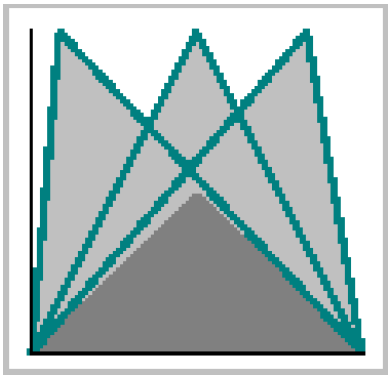


Figure 6: The triangular kernel for a Markov chain on the unit interval. The dark isoceles triangle corresponds to an unnormalized version of the density of ν for the small set property arising from application of Definition 2.1 to the entire interval.

Small sets are of major importance in the development of *CFTP*, so we spend a little time discussing their theory.
It is a central result that (non-trivial) small sets (possibly of arbitrarily high order) exist for any modestly regular Markov chain [66, 77]. Here "non-trivial" means, has a positive chance of being hit by the Markov chain started from a generic point. (We would need the language of ϕ-irreducibility to make this precise, which would take us too far afield. See either of the two references cited.)
The trouble with general Markov chains is that the chain may have zero probability of returning to *any* fixed starting point. However if there is a small set of order 1 then we may re-model the chain to fix this.

Theorem 2.2 *Let X be a Markov chain on a (non-discrete) state space $\mathcal{X}$, transition kernel $p(x, dy)$, with small set C of order 1. Then X can be represented using a new Markov chain on $\mathcal{X} \cup \{\boldsymbol{c}\}$, for $\boldsymbol{c}$ a regenerative pseudo-state.*

For details see [5, 76]. Higher-order small sets can be used if we are prepared to sub-sample the chain ....
Small sets (perhaps of higher order) can be used systematically to attack general state space theory using discrete state space methods [66].[h]

[h] Roberts and Rosenthal [82] also introduce "pseudo-small sets", which relate to coupling as small sets relate to *CFTP*.

A natural question is to ask whether we can go further, and use small sets to re-model the chain as an entirely discrete affair. The answer is almost affirmative!

Remark 2.3 *Non-trivial* small sets of order 1 *need not exist: however they do exist if (a) the kernel* $p(x, d\,y)$ *has a measurable density* and *(b) chain is sub-sampled at* even *times. (Both are needed: see the example in [54].)*

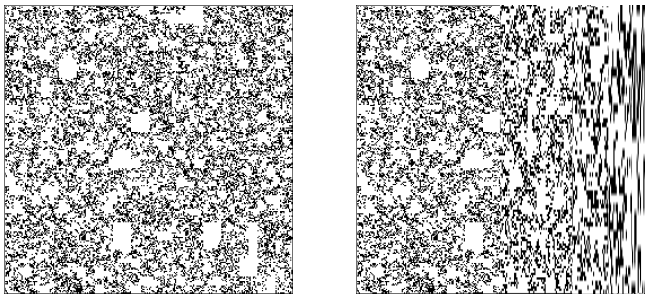


Figure 7: (a) A subset of the square which is free from non-null measurable rectangles; (b) A kernel which is free of small sets of order 1.

Figure 7 shows (a) a randomized construction of a measurable subset E of $[0,1]^2$, such that if $A \times B \subseteq E$ then $\mathrm{Leb}(A \times B) = 0$; (b) the support of a measurable function based on transformed replications of this subset which provides a transition density from $[0,1]$, to $[0,1]$, such that the density admits no (non-trivial) small sets.

Theorem 2.4 *[54] If the Markov chain has a measurable transition density* $p(x,y)$ *then the two-step density* $p^{(2)}(x,y)$ *can be expressed (non-uniquely) as a non-negative countable sum*

$$p^{(2)}(x,y) = \sum_i f_i(x) g_i(y)\,.$$

Proof (Sketch):
The key part of this proof is a mild variation on Egoroff's Theorem:
Let $p(x,y)$ be an integrable function on $[0,1]^2$. Then we can find subsets $A_\varepsilon \subset [0,1]$, increasing as ε decreases, such that

(a) for any fixed A_ε the "L^1-valued function" p_x is uniformly continuous on A_ε: for any $\eta > 0$ we can find $\delta > 0$ such that

$$\int_0^1 |p_x(z) - p_{x'}(z)|\,\mathrm{d}\,z < \eta$$

for $|x - x'| < \delta$ and x, $x' \in A_\varepsilon$.

(b) every point x in A_ε is of full relative density: as $u, v \to 0$ so

$$\text{Leb}([x-u, x+v] \cap A_\varepsilon)/(u+v) \to 1\,.$$

We can use this result to show that $p^{(2)}(x, y)$ has just enough near-continuity to supply a rich variety of small sets of order 2. □

This result can be used to construct a latent discrete Markov chain Y in even time which captures the time-dependence; the original chain X can be re-constructed using functions $X_{2n} = h(Y_n, Y_{n+1}, \varepsilon_n)$, where the ε_n are independent and identically distributed.

2.2 Murdoch-Green small-set *CFTP*

Green and Murdoch [37] showed how to use small sets to carry out *CFTP* when the state space is continuous with no helpful ordering.[i]

Example 2.5 *It is helpful to consider small-set* CFTP *in nearly the simplest possible case: recall the Markov chain triangular kernel on* $[0, 1]$ *illustrated above.*

At any time-step there is regeneration probability $1/2$ of drawing from the isoceles kernel $\nu(\mathrm{d}\, y)$; and we can couple all possible draws together so that if one uses the isoceles kernel then so do they all. Now small-set *CFTP* is easy to implement: start at time $-n$, at each step consider whether or not one has drawn from the isoceles kernel. There is no need to keep record of any draws until the first coupled draw from the isoceles kernel: from then on one can evolve the chain using the full kernel until time 0. If perchance one has failed to make a draw from the isoceles kernel, then one repeats the procedure from time $-2n$; however one must then take care to ensure that for steps from time $-n$ onwards one re-uses the common randomness, by drawing from the *residual kernel* obtained by renormalizing $p(x, \mathrm{d}\, y) - (1/2)\nu(\mathrm{d}\, y)$.
In more usual cases the regeneration probability will be drastically smaller than $1/2$: Green and Murdoch describe a "partitioned multi-gamma sampler", which carries out a more efficient small-set *CFTP* using a partition of the state space by small sets.

Example 2.6 *The result of small-set* CFTP *can be re-expressed as the composition of Geometrically many kernels (conditioned to avoid the small-set effect), with starting point randomized by small-set distribution* ν.

[i] Murdoch and Green use the term "multi-gamma sampler" instead of "small-set *CFTP*". This arises from Lindvall's [62] nomenclature for the kind of coupling described in §2.1. Why "gamma"? When asked, Lindvall explains this is because he had already used alpha and beta in his exposition ....

We use the notation of Definition 2.1, so that the small set which is the whole state space has associated regeneration probability α, and regeneration distribution $\nu(\mathrm{d}\, y)$. Then small-set *CFTP* gives the representation

$$\pi(\mathrm{d}\, x) \quad = \quad \sum_{r=1}^{\infty} \alpha(1-\alpha)^{r-1} \int \nu(\mathrm{d}\, y) \widetilde{p}^{(n)}(y, \mathrm{d}\, x)\,,$$

where $\widetilde{p}^{(n)}(y, \mathrm{d}\, x)$ is the n-step kernel corresponding to the kernel $\widetilde{p}(y, \mathrm{d}\, x)$ conditioned on no regeneration:

$$\widetilde{p}(y, \mathrm{d}\, x) \quad = \quad \frac{p(y, \mathrm{d}\, x) - \alpha\nu(\mathrm{d}\, x)}{1-\alpha}\,.$$

(See [6, 11, 43].)

2.3 Slice sampler *CFTP*

Consider the simple task of drawing from a one-dimensional density $f(x)$. (Note, this method is only interesting because it can be made to work in many dimensions ...) Suppose f is unimodal. We can define $\mathtt{g0}(y)$, $\mathtt{g1}(y)$ implicitly by: the requirement that $[\mathtt{g0}(y), \mathtt{g1}(y)]$ is the super-level set $\{x : f(x) \geq y\}$. The slice sampler alternates between drawing y uniformly from $[0, f(x)]$ and drawing x uniformly from $[\mathtt{g0}(y), \mathtt{g1}(y)]$ (see Figure 8).

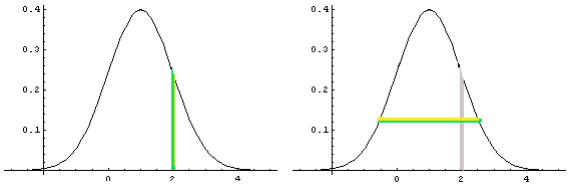


Figure 8: Slice sampler constructions.

There is rapid convergence (order of 530 iterations!) under a specific variation of log-concavity [81, Theorem 12].

Example 2.7 *Ideas of regeneration can be used to design a* perfect *slice sampler for the case of a unimodal density $f(x)$ with bounded support. (The "bounded support" condition can be lifted: see [67].)*

It is necessary to figure out how to make uniform choices for two versions of the process simultaneously, so as to preserve the partial ordering

$$(x, y) \quad \preceq \quad (x', y') \qquad \text{if} \quad f(x) \leq f(y)\,, \tag{4}$$

but *also* so as to have positive chances of coalescing. Figure 9 sketches out how to do this, exploiting the properties of the uniform distribution. The example is completed by determining how to extend the simulation backwards in time if coalescence has failed for the current cycle.

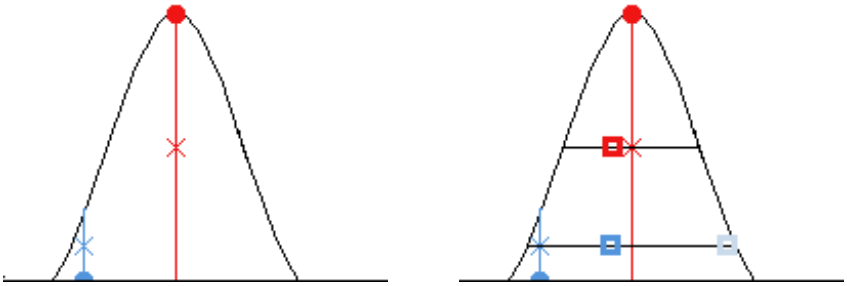


Figure 9: Perfect slice sampling: (a) first choose height for low point, next choose height for top point, making the two heights identical if top point height falls in possible range for low point height; then (b) first choose horizontal location for top point, next choose horizontal location for low point, making the two locations identical if low point location falls in possible range for top point location.

The technical issues in the above can all be resolved using the following idea:

(a) Given U a Uniform($[0,1]$) random variable, and $0 < \theta < 1$, we can draw V a Uniform($[0,\theta]$) random variable as follows: if $U \leq \theta$ then set $V = U$ otherwise draw V from the Uniform($[0,\theta]$) distribution. So we have arranged a coupling with $V \leq U$ and $\mathbb{P}[U = V] = \theta$.

(b) Now suppose we are given $V \leq U$ as above, and wish to draw W a Uniform($[0,\psi]$) random variable, with $\theta < \psi < 1$, such that $V \leq W \leq U$. If $U \leq \psi$ then set $U = W$, otherwise set $W = V$ with probability θ/ψ, otherwise draw W from the Uniform($(\theta,\psi]$) distribution. So we have arranged a coupling with $V \leq W \leq U$ and $\mathbb{P}[U = W] = \psi$, $\mathbb{P}[W = V] = \theta/\psi$.

2.4 Multi-shift sampler

The following question is a simple version of one which often arises in implementation of *CFTP*:

Question 2.8 *How can one draw X_x simultaneously from Uniform$[x, x+1)$ for all $x \in \mathbb{R}$, and couple the draws to coalesce to a countable number of different outcomes? [96].*

The answer is to use a random *unit span integer lattice*, $U+\mathbb{Z}$ where U is Uniform($[0,1)$). Then set X_x to be the unique point in the intersection $[x, x+1) \cap (U+\mathbb{Z})$.
Wilson [96] also considers more general distributions! For example, we can express a unimodal distribution as a mixture of uniform distributions, in a manner reminiscent of slice sampler ideas, in several ways, as illustrated in Figure 10. Once we have expressed the target distribution as a mixture of uniforms, say

$$\mathcal{L}(X) \quad = \quad \mathcal{L}(\text{Uniform}([-L,+R)))$$

for random L and R, then we can draw simultaneously from the location family of distributions $\mathcal{L}(X+x)$ by first drawing L, R, then constructing the random lattice $(U+\mathbb{Z}) \times (L+R)$, then finally setting X_x to be the unique point in the intersection $[x-L, x+R) \cap ((U+\mathbb{Z}) \times (L+R))$. The method also deals with multivariate and even multi-modal cases.

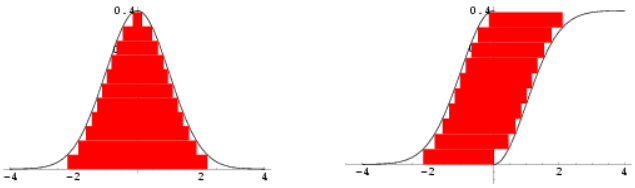


Figure 10: Two different ways of expressing a normal density as a mixture of uniform densities.

Corcoran and Schneider [20] carry this even further, showing how to couple draws from Uniform distributions with different ranges.

2.5 Catalytic *CFTP*

Breyer and Roberts [12] have devised an "automatic" variation on small-set *CFTP*: *catalytic* CFTP. The underlying idea is to perform simultaneous Metropolis-Hastings updates for all possible states, using a common Uniform$[0,1]$ random variable U to determine rejection or acceptance. For suitable proposal random fields Φ_x, it may be possible to identify when the input-output map $F_{(-t,0]}$ coalesces into a finite range; moreover the choice of construction for Φ_x can be varied from time point to time point. Simulations can be viewed at

`http://www.lbreyer.com/fcoupler.html`.

2.6 Read-once *CFTP*

Wilson [95] noted the following: one can build the input-output maps $F_{(-n,0]}$ of Theorem 1.3 from *i.i.d.* blocks

$$F_{(-nt,0]} \quad = \quad F_{(-t,0]} \circ \ldots \circ F_{(-(n-1)t,(-(n-2)t]} \circ F_{(-nt,-(n-1)t]} \, .$$

Let the blocking length t be chosen so that there is a positive chance of the map $B \stackrel{\mathcal{D}}{=} F_{(-t,0]}$ being coalescent. By a simple computation, the resulting *CFTP* procedure is identical to the following *forwards* procedure:

- Repeatedly draw independent realizations of B till a coalescent block is obtained; note coalesced output x.
- Repeatedly draw independent realizations of B; while these are *not* coalescent compute the update $x \leftarrow B(x)$.
- When a coalescent realization of B is obtained, return x *without* updating!

There are strong resonances with small-set *CFTP* (the possibility of coalescent B corresponds to the whole state space being a small set of order t), especially the representation discussed in Example 2.6, and with the dead leaves *CFTP* example (one can view Wilson's argument as involving a re-ordering in time).

Example 2.9 *The validity of read-once* CFTP *follows by establishing that the above forwards procedure produces a sequence of B maps which have the same distribution as would be produced by carrying out classic* CFTP*, but checking for coalescence block-by-block.*

The choice of the blocking length t is of course crucial! Wilson [95] explains how this can be done, in such a way as to be competitive with ordinary *CFTP*.
A feature of major importance of this method is that storage requirements are minimized: one needs only (a) to flag whether a block is coalescent, (b) to compute the output of a coalescent block, and (c) to track the current state of the chain as the blocks are produced.

2.7 Some Complements

We remark briefly on precursors to this idea. We have already noted the seminal nature of the split-chain construction [5, 76]. Regeneration ideas are not new to simulation, and have been used specifically in simulation for some time: see for example Asmussen *et al.* [4] and Mykland *et al.* [75].
The simple and direct construction of §2.2 is rather strongly limited by the need to find a regenerative probability measure ν for the whole state space (or a finite covering of the

state space by small sets, in the case of the partitioned version).[j] However it is a potentially important tool, whether in its original form or in the variants described above, when combined with other ideas such as the generalization of *CFTP* which we will describe in the next section: small-set coalescence can be an important component of other *CFTP* algorithms, especially when seeking to couple a pair of monotonic processes which otherwise will draw closer at only an exponential rate!

Murdoch and Rosenthal [73] use regenerative ideas to develop a useful perspective on how one might efficiently obtain repeated *CFTP* samples. Craiu and Meng [23] show how to achieve efficiency gains by making multiple *CFTP* runs using antithetic sampling. Meng [65] also suggests use of multistage sampling ideas to improve *CFTP* efficiency.

[j] But Hobert and Robert [43] use the idea to investigate convergence for *MCMC*.

3 Dominated *CFTP*

Up to this point all our *CFTP* methods and applications have applied only to Markov chains which are in some sense "bounded" (strictly speaking, uniformly ergodic in the sense of Definition 3.8). We now discuss how to lift this restriction.
We begin by considering some precursor notions from queueing theory (§3.1), and then define the important theoretical notions of uniform and geometric ergodicity (§3.2) and discuss their relationship with classic *CFTP* (§3.3). This leads straight to the idea of *dominated* CFTP (*domCFTP*), introduced in §3.4, which can apply to geometrically ergodic (and hence "unbounded") Markov chains. We describe this idea carefully in the simple but prototypical context of birth-death processes (§3.5) and present an application to point processes (§3.6). We conclude by describing a general theorem on the validity of *domCFTP* (§3.7).

3.1 Queues

Consider a $GI/G/1$ queue (intervals between customer arrivals are independent and identically distributed, as are the service times required by customers – though of course service time and inter-arrival time have different distributions). Lindley noticed a beautiful representation for waiting time W_n of customer n in terms of services S_n and inter-arrivals X_n, based on the observation that $S_n - X_{n+1}$ (if positive) is the extra amount of time for which customer $n+1$ must wait as compared to customer n.

Theorem 3.1 ***Lindley's equation:*** *Consider the waiting time identity for the $GI/G/1$ queue.*

$$\begin{aligned} W_{n+1} &= \max\{0, W_n + S_n - X_{n+1}\} = \max\{0, W_n + \eta_n\} \\ &= \max\{0, \eta_n, \eta_n + \eta_{n-1}, \ldots, \eta_n + \eta_{n-1} + \ldots + \eta_1\} \\ &\stackrel{\mathcal{D}}{=} \max\{0, \eta_1, \eta_1 + \eta_2, \ldots, \eta_1 + \eta_2 + \ldots + \eta_n\} \end{aligned}$$

and thus we obtain the steady-state expression

$$W_\infty \stackrel{\mathcal{D}}{=} \max\{0, \eta_1, \eta_1 + \eta_2, \ldots\}\,.$$

It is an exercise in classical probability theory (SLLN / CLT / random walks) to show that W_∞ will be finite if and only if $\mathbb{E}[\eta_i] < 0$ or $\eta_i \equiv 0$.

Remark 3.2 *Coupling and* CFTP *ideas enter into Theorem 3.1 at the crucial time-reversal step:*

$$\begin{aligned} \max\{0, \eta_n, \eta_n + \eta_{n-1}, \ldots, \eta_n + \eta_{n-1} + \ldots + \eta_1\} \quad &= \\ \stackrel{\mathcal{D}}{=} \quad \max\{0, \eta_1, \eta_1 + \eta_2, \ldots, \eta_1 + \eta_2 + \ldots + \eta_n\} & \end{aligned}$$

Compare Section 1.4 on falling leaves ....

Remark 3.3 *The problem about applying the* CFTP *Theorem 3.1 in this context is that we don't know* which $\eta_1 + \eta_2 + \ldots + \eta_n$ *attains the supremum* $\max\{0, \eta_1, \eta_1 + \eta_2, \ldots\}$.

The point of this remark is, we could use the above to make a draw from the equilibrium distribution W_∞, if only we could tell at which n the maximum is attained! Failing that, Theorem 3.1 suggests a simulation algorithm which approximates W_∞ from below by W_n for large n – the issue of choice of n corresponds to the burn-in decision for *MCMC*.. Notice that here we have a target distribution which is specified implicitly, as with the dead leaves example, rather than explicitly up to a normalizing constant.
Supposing we lose independence? Loynes [64] discovered a coupling application to queues with (for example) general *dependent* stationary inputs and associated service times, pre-figuring *CFTP*.

Theorem 3.4 *Suppose queue arrivals follow a* time-stationary point process *marked by service times, stretching back to time* $-\infty$. *Denote the arrivals and associated service times in* $(s, t]$ *by* $N_{s,t}$. *The impact of stationarity is that the distribution of the process* $\{N_{s,s+u} : u \geq 0\}$ *does not depend on the start-time* s. *Let* Q^{-T} *denote the behaviour of the queue observed from time* 0 *onwards if begun with* 0 *customers at time* $-T$. *The queue converges to statistical equilibrium if and only if*

$$\lim_{T\to\infty} Q^{-T} \quad \textit{exists almost surely.}$$

Remark 3.5 *Stoyan [87] develops this kind of idea. See also an application to storage problems in [48].*

Example 3.6 *It is informative to use simple* R *statistical package scripts and elementary calculations to investigation Lindley's equation and the Loynes coupling.*

3.2 Uniform and Geometric Ergodicity

There is a theoretical issue which forms a road block to the use of Lindley's representation in *CFTP*. Recall the notions of *uniform ergodicity* and *geometric ergodicity*.

Definition 3.7 *A Markov chain with transition kernel* $p(x, \cdot)$ *possesses* ***geometric ergodicity*** *if there are constants* $0 < \rho < 1$, $R(x) > 0$ *with*

$$\|\pi - p^{(n)}(x, \cdot)\|_{TV} \leq R(x)\rho^n$$

for all n, *for all starting points* x.

So a chain exhibits geometric ergodicity if equilibrium is approached at a geometric rate. Note that the geometric bound is moderated by a multiplicative factor depending on the chain's starting point. However ...

Definition 3.8 *A Markov chain with transition kernel $p(x,\cdot)$ possesses **uniform ergodicity** if there are constants $\rho \in (0,1)$, $R > 0$ not depending on the starting point x such that*

$$\|\pi - p^{(n)}(x,\cdot)\|_{TV} \leq R\rho^n$$

for all n, and uniformly in all starting points x.

So a chain exhibits uniform ergodicity if the geometric rate is not affected by the chain's starting point.

3.3 Classic *CFTP* and Uniform Ergodicity

Uniform ergodicity corresponds loosely to "virtually finite state space". However chains may still be uniformly ergodic even if the state space is far from finite: the Häggström *et al.* [42] chain in Section 1.6 is a good example of this.
On the other hand Lindley's theorem presents a class of examples which in general will not be uniformly ergodic. Think for example of the case of Uniform$[0,3]$ inter-arrival times, and service times deterministic and equal to 1: a queue of length n will then take at least n units of time to disperse completely, and this can be used as the basis of an argument to show failure of uniform ergodicity.
Foss and Tweedie [35] show that the (theoretical) possibility of classic *CFTP* is equivalent to uniform ergodicity. For classic *CFTP* needs *vertical coalescence*: every possible start from time $-T$ leads to the same result at time 0, and this implies a uniformly geometric rate of convergence to equilibrium (Figure 11).

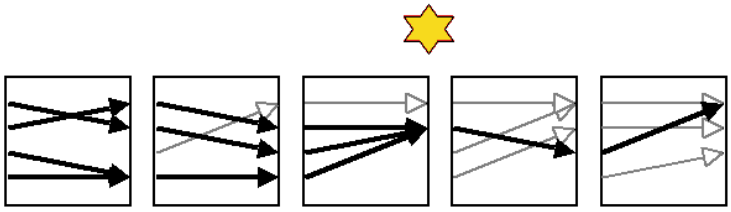


Figure 11: Vertical coalescence (starred) for a finite-state-space Markov chain, starting from a fixed initial time.

The converse, that uniform ergodicity implies the possibility *in principle* of classic *CFTP*, follows from small set theory.

Theorem 3.9 *(After Foss and Tweedie [35]) Suppose a Markov chain X on a general state space $\mathcal{X}$ has positive probability of hitting a specified small set C of order k, where the probability may depend on the starting position but is always positive.*[k] *Then X is uniformly ergodic if and only if classic* CFTP *is possible* in principle *(disregarding computational and implementation issues!).*

Proof (Outline):
It is clear from the *CFTP* construction that classic *CFTP* forces uniform ergodicity.
On the other hand, uniform ergodicity means we can choose n such that $p^{(n)}(x,\cdot)$ is close to equilibrium in total variation, uniformly in x. It follows that *in principle* we can design a split chain which has positive chance of applying regeneration every $n+k$ time steps, and this permits construction of small-set *CFTP* . For suppose C is the small set of order k as given in Definition 2.1, with $\pi(C) > 0$. Then for $\varepsilon \to 0$ as $n \to \infty$ uniformly in x,

$$\begin{aligned} p^{(n)}(x,C) &\geq (1-\varepsilon)\pi(C)\,, \\ p^{(n+k)}(x,\cdot) &\geq (1-\varepsilon)\alpha\pi(C)\nu(\cdot)\,. \end{aligned}$$

So we can apply small-set *CFTP* to the sub-sampled Markov chain X_{n+k}, $X_{2(n+k)}$, ...
□

Foss and Tweedie [35] also derive comparisons between moments of coalescent times and forward coupling times.
The practical obstacle here is that we will have to gain knowledge of $p^{(n)}(x,\cdot)$ to build the split chain. But in general we may expect $p^{(n)}(x,\cdot)$ to be less accessible than the equilibrium distribution itself!

3.4 Dominated *CFTP*

It follows from the very existence of *CFTP* constructions that all the chains discussed so far have been uniformly ergodic. Can we lift this uniform ergodicity requirement? *CFTP* almost works with *horizontal coalescence* as exhibited in the Lindley representation and illustrated in Figure 12: all sufficiently early starts from a specific location lead to the same result at time 0. But, as highlighted by the Lindley representation, the question is *how* one can identify when this has happened.

[k] This small-set condition is essentially a consequence of ϕ-irreducibility [77], which itself is implied by the much stronger condition of uniform ergodicity.

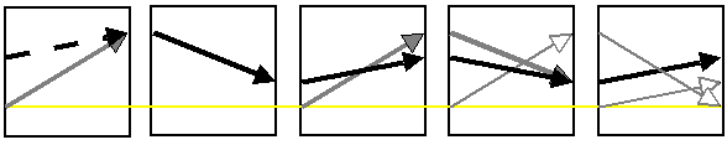


Figure 12: Horizontal coalescence starting from a fixed location: this will have occurred if all earlier starts from this location will also coalesce by time 0.

The idea of dominated *CFTP* (*domCFTP*) is as follows: generate target chains coupled to a *dominating process* for which equilibrium is known. Domination allows us to identify horizontal coalescence by checking starts from maxima given by the dominating process. We set this out in a formal definition. For the sake of clarity we consider a rather simple case, that of a discrete-time monotonic chain defined on $[0, \infty)$. (A much more general formulation, allowing for general state space and non-monotonicity, is given below in Theorem 3.12.)

Definition 3.10 *(*domCFTP*) Consider X, an ergodic Markov chain on $[0, \infty)$. Suppose it can be coupled as follows: for each $x \geq 0$, $-t < 0$ we can construct $X^{(x,-t)}$ to be X begun at x at time $-t$, such that if $s \geq -t$, $s \geq -u$, then*

$$X_s^{(x,-t)} \leq X_s^{(y,-u)} \qquad \textit{implies} \qquad X_{s+1}^{(x,-t)} \leq X_{s+1}^{(y,-u)} .$$

Suppose further we can build a dominating process *Y on $[0, \infty)$, which is stationary, defined for all time, and coupled to the $X^{(x,-t)}$ by*

$$X_s^{(x,-t)} \ \leq \ Y_s \qquad \textit{implies} \qquad X_{s+1}^{(x,-t)} \ \leq \ Y_{s+1}$$

whenever $s \geq -t$. Then the following algorithm delivers a perfect sample from the equilibrium distribution of X, so long as it terminates almost surely:

(1) Draw Y_0 from its equilibrium distribution;

(2) Simulate Y backwards in time *to time $-T$;*

(3) Set $y = Y_{-T}$, and simulate the upper-process *$X^{-T,y}$ and the* lower-process *$X^{-T,0}$ forwards in time to time 0 (*note: *these must be coupled to each other, to Y, and at later stages of the algorithm they must be coupled to other simulations of X at the same process time);*

(4) If $X_0^{-T,y} = X_0^{-T,0}$ then return their common value as a perfect draw from the equilibrium distribution of X. Otherwise extend the previous simulation of Y back to time $-2T$, update $-T$ to $-2T$, and repeat from step (3).

If we can make this work then *CFTP* can be applied to Markov chains which are merely geometrically ergodic [17, 51, 53, 55] or worse (geometric ergodicity $\neq$ *domCFTP*!). The issues are:

(a) can one draw from the equilibrium of Y?

(b) can one simulate Y backwards in time?

(c) can one simulate the upper- and lower-processes coupled both to other simulations of X and to Y as required in the definition?

(d) and, of course, will coalescence (which is to say, termination of the algorithm) occur almost surely, and will it occur with reasonable speed?

There is considerable freedom allowed in the choice of Y, so requirements (a), (b) are not hard to meet. The implementation of requirement (c) typically needs care; on the other hand (d) is typically half obvious (whether coalescence is almost sure) and half empirical (one investigates the speed by trying it out in practice!).

Theorem 3.11 *[51, 53] If coalescence is almost sure then* domCFTP *samples from equilibrium.*

Proof: For simplicity we continue to suppose the target process X is monotonic, and that it and the dominating process Y are non-negative.
Let $X^{\text{upper},-n}$, $X^{\text{lower},-n} = X^{-n}$ be versions of the target chain started at time $-n$ at $Y(-n), 0$ respectively. Let $-T$ be the latest time such that $X^{\text{upper},-T}(0) = X^{\text{lower},-T}(0) = X^{-T}(0)$ (so $-T$ is the *coalescence time*). Now argue as in Theorem 1.3 for classic *CFTP*: If X converges to an equilibrium π in total variation dist_{TV} then

$$\text{dist}_{\text{TV}}(\mathcal{L}\left(X_0^{-T}\right), \pi) = \lim_n \text{dist}_{\text{TV}}(\mathcal{L}\left(X_0^{-n}\right), \pi) = \lim_n \text{dist}_{\text{TV}}(\mathcal{L}\left(X_n^0\right), \pi) = 0$$

hence the result. □

Thus we can use realizations of the target process started from the dominating process to identify horizontal coalescence.

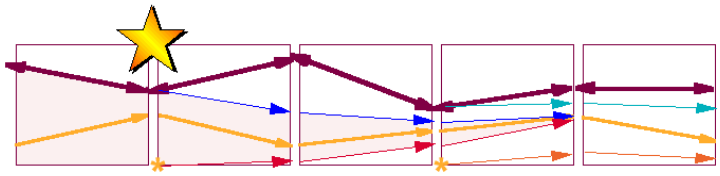


Figure 13: Dominated *CFTP*. Coalescence is assured at the starred time, since all previous starts below the dominating process are compelled to coalesce by time 0.

3.5 Non-linear birth-death processes

To illustrate *domCFTP* in detail, we describe a simple example taken from [49]. Consider a continuous-time non-linear birth-death process X, with transition rates

$$\begin{aligned} X \to X+1 \qquad & \text{at rate } \lambda_X\,, \\ X \to X-1 \qquad & \text{at rate } X\mu\,, \end{aligned}$$

for positive λ_X, μ. We suppose the birth rate λ_X is bounded above[l] by λ.
Of course it is possible to compute the equilibrium distribution using detailed balance. However here the object of the exercise is to construct a *domCFTP* method to draw exactly from the target distribution.
Note first that the non-linear birth-death process X can be bounded above, or *dominated* by, the linear birth-death process Y with transition rates

$$\begin{aligned} X \to X+1 \qquad & \text{at rate } \lambda\,, \\ X \to X-1 \qquad & \text{at rate } X\mu\,. \end{aligned}$$

Here *domination* means, if $0 \le X(0) \le Y(0)$ then we can construct coupled copies of X and Y such that the relationship $X \le Y$ is maintained for all time.
Indeed we can go further: given the process Y then for *any* x, $0 \le x \le Y(0)$, we can construct a copy X_x of X begun at x such that $0 \le X_a \le X_b \le Y$ for all time whenever $a \le b \le Y(0)$.
We do this as follows.
We construct X from Y by supposing, X can have a birth only if Y has a birth, and similarly for deaths.
Suppose to each birth incident and each death incident of Y there is attached an independent Uniform$[0,1]$ random *mark* U. So the construction of X is specified by determining for each Y incident whether or not this corresponds to an X incident.

[l] Monotonicity is required for λ_X in [49], which is unnecessarily restrictive!

- A birth incident $Y \to Y+1$ at time t with mark U is allowed to generate an X birth incident exactly when

$$U \quad \leq \quad \frac{\lambda_{X(t-)}}{\lambda}\,; \tag{5}$$

- A death incident $Y \to Y-1$ at time t with mark U is allowed to generate an X death incident exactly when

$$U \quad \leq \quad \frac{\mu X(t-)}{\mu Y(t-)} \quad = \quad \frac{X(t-)}{Y(t-)}\,. \tag{6}$$

It is apparent from $X(t-) \leq Y(t-)$ and the increasing nature of $\lambda_X \leq \lambda$ that the U-based criteria above use probabilities $\lambda_{X(t-)}/\lambda \leq 1$ and $X(t-)/Y(t-) \leq 1$ respectively. This permits an inductive argument, iterating through the birth and death incidents of Y, which shows $X \leq Y$ for all time, and which indeed also demonstrates $0 \leq X_a \leq X_b \leq Y$ if $0 \leq X_a(0) \leq X_b(0) \leq Y(0)$.
Now carry out the *CFTP* construction, but making starts at times $-n$, $-2n$, ...using a stationary realization of the dominating process, as in Definition 3.10, rather than the top-most state. To do this it is necessary to be able to

(1) draw from the equilibrium of the dominating process (easy here: detailed balance identifies the equilibrium distribution as Geometric);

(2) simulate the reversed process in equilibrium (easy here: by detailed balance the process is reversible).

The remaining requirements of Definition 3.10 are assured by the construction given above. An illustration of the result is given in Figure 14.
This example is rather trivial (in this case equilibrium is best simulated using a formula for the equilibrium distribution derived from considerations of detailed balance!). However similar examples can deal with cases where no formula for equilibrium is known (for example, perpetuities); moreover it is straightforward to introduce a spatial component to the birth-death process, which we discuss next.

3.6 Point processes

Dominated *CFTP* actually works on non-monotonic processes as well. For example, it can be applied to both attractive and repulsive area-interaction point processes [49, 51, 53]: using as target chain a *spatial* birth-and-death process, which give birth to points at a rate determined by the local interaction with pre-existent points, and which kills points at unit rate per point. This allows the use of *domCFTP* in a manner very similar to that of

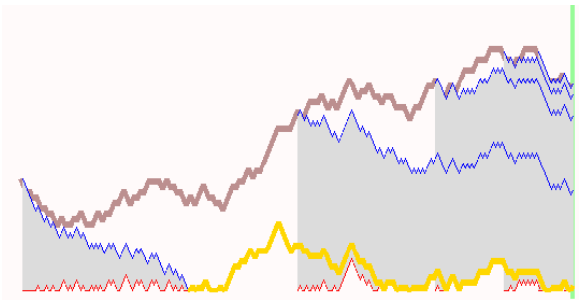


Figure 14: A birth-death example of *domCFTP*. Top curve is dominating process. Successive pairs of upper and lower processes generate shaded regions which sandwich earlier realizations: the earliest pair in the diagram produces coalescence.

Section 3.5, but with Uniform$[0,1]$ marks replaced by *geometric* marks which are Poisson clusters, as described in [49] and exploiting Exercise 1.11 as well as an analogue for the case of repulsion.[m] See Figure 15 for an illustration.
It is of interest in stochastic geometry that this expresses such point processes as explicit but highly *dependent thinnings* of Poisson point processes.
How exactly is the method of *domCFTP* (or indeed *CFTP* in general) adapted to non-monotonic cases? We can use the *crossover trick* [51], which we explain in the context of repulsive area interaction $\gamma < 1$. Create two chains to bound the target chain X above (by X^{upper}) and below (by X^{lower}). Cross over the rules for birth: a proposed point is born in X^{upper} if it would pass the test for X^{lower}, and *vice versa*. Then automatically

$$X^{\text{lower}} \quad \subseteq \quad X \quad \subseteq \quad X^{\text{upper}}$$

so *CFTP* can be applied. A general formulation for point processes is given in [53].
See also Huber's [45] use of "swap moves" in the context of bounding chains, which he uses to estimate a rapid-mixing regime. If the birth proposal is blocked by just one point, then replace the blocking point by the new proposal in a *swap*, with swap probability p_{swap} which we are free to choose; and adjust the bounding chains accordingly; this results in a provable speed-up of the *CFTP* algorithm.

3.7 A general theorem for *domCFTP*

Moving from *CFTP* to *domCFTP* suggests still further abstraction. This is helpful, for example, when considering *CFTP* for *conditioned* point processes as in [17]: it can be

[m] Sandeep Shah was the first to implement this, in his 2004 Warwick PhD thesis.

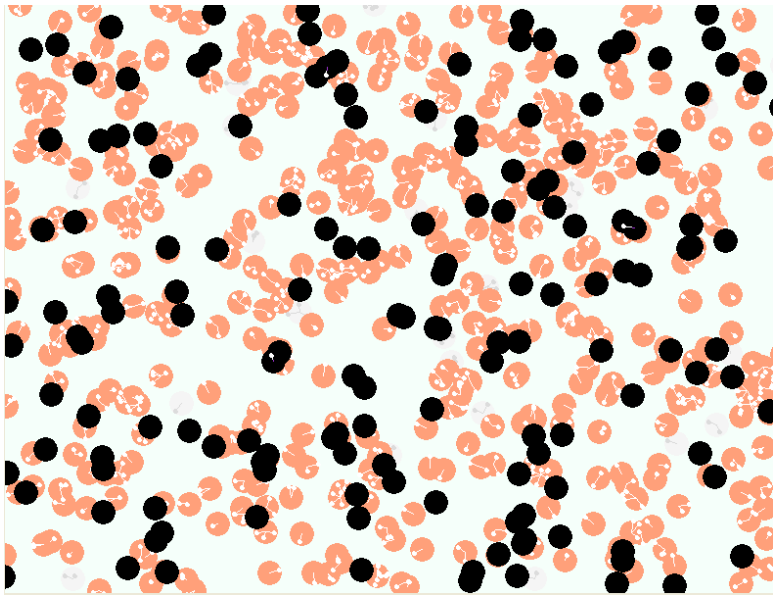


Figure 15: Dominated *CFTP* for attractive area-interaction point process with geometric marking using Poisson processes in disks. Dark disks are in both the lower process and the upper process; lighter disks are in the upper process (there are also some pale and ghostly disks corresponding to points which did not even get into the upper process!). Interaction is expressed by marking each disk with a Poisson cluster: on birth the cluster in a new disk must be covered by the current union of disks.

convenient to allow the upper- and lower-processes to move out of the conditioned state space.
Let X be a Markov chain on $\mathcal{X}$ which is ergodic and in statistical equilibrium.
Embed the state space $\mathcal{X}$ in a partially ordered space $(\mathcal{Y}, \preceq)$ so that $\mathcal{X}$ is at the *bottom* of $\mathcal{Y}$, in the sense that for any $y \in \mathcal{Y}$, $x \in \mathcal{X}$,

$$y \preceq x \qquad \text{implies} \qquad y = x\,.$$

We may then use the methods of Theorem 1.3 (*CFTP*) and Theorem 3.11 (*domCFTP*) to show:

Theorem 3.12 *Define a Markov chain Y on $\mathcal{Y}$ such that Y evolves as X after it hits $\mathcal{X}$; let $Y(-u,t)$ be the value at t of a version of Y begun at time $-u$,*

(a) of fixed initial distribution:
$\mathcal{L}\left(Y(-T,-T)\right) = \mathcal{L}\left(Y(0,0)\right)$, *and*

(b) obeying funnelling*:*
if $-v \le -u \le t$ *then* $Y(-v,t) \preceq Y(-u,t)$.

Suppose coalescence occurs: $\mathbb{P}\left[Y(-T,0) \in \mathcal{X}\right] \to 1$ *as* $T \to \infty$. *Then* $\lim Y(-T,0)$ *can be used for a* CFTP *draw from the equilibrium of* X.

3.8 Some Complements

Murdoch [74] points out a *MCMC* algorithm can be *forced* to become uniformly ergodic by altering the move to allow a small chance of an *independence sampler* move. This procedure forces the whole state-space to become small, and will be effective for suitable low-dimensional examples: however it is not clear how to implement it for point processes, for instance.
The crossover trick is generalized in [53], to cover cases where monotonicity is absent (see also Huber [46] on bounding chains); Häggström and Nelander [40] apply the trick to lattice systems.
Ambler and Silverman [2, 3] describe a practical method for implementing *domCFTP* for certain systems for which interaction is neither monotonic nor anti-monotonic, and apply this to generalized area-interaction point processes and thence to wavelet models with dependent coefficients. See also Holmes and Mallick [44], who apply *classic CFTP* to the case of models with independent coefficients.

4 Theory and connections

CFTP is not the only method of perfect simulation. Here we describe a different method, due to Fill. We begin by discussing a historical predecessor, Siegmund duality (§4.1); we use this to explain Fill's method (§4.2). We then describe a striking relationship recently introduced between Fill's method and *CFTP* (§4.3), which shows the first is actually a conditioned version of the second.
We then turn to questions of efficiency – whether *CFTP* can always be competitive with an idealized *MCMC* implementation which somehow just knows how long the burn-in period should be (§4.4). Finally we consider the link between *domCFTP* and geometric ergodicity (§4.5), and briefly present yet another variant on *CFTP*, the Backwards-Forwards Algorithm (§4.6), which has strong links to *domCFTP*.

4.1 Siegmund duality

An important alternative to *CFTP* makes fuller use of the notion of *time reversal*, as in the dead-leaves example, and Section 3.1 on queues. We begin with a beautiful duality.

Theorem 4.1 *(Siegmund duality) Suppose X is a process on $[0, \infty)$. When is there another process Y satisfying the following?*

$$\mathbb{P}\left[X_t \geq y | X_0 = x\right] \quad = \quad \mathbb{P}\left[Y_t \leq x | Y_0 = y\right] \tag{7}$$

Answer [86]: exactly when X is (suitably regular and) stochastically monotone*: $x \leq x'$ implies*

$$\mathbb{P}\left[X_t \geq y | X_0 = x\right] \quad \leq \quad \mathbb{P}\left[X_t \geq y | X_0 = x'\right] .$$

Proof (Outline):
Use Equation (7) to check monotonicity, and Fubini's Theorem to derive the Chapman-Kolmogorov equations. □

Remark 4.2 *If X is not stochastically monotone then Equation (7) will yield negative transition probabilities for Y!*

Remark 4.3 *It is a consequence of Equation (7) that Y is absorbed at 0, and X at ∞.*

Remark 4.4 ***Intuition:*** *think of the Siegmund dual this way. For fixed T, represent the $X^{(x)}$ begun at different x in a coupling as a monotonic stochastic flow (use stochastic monotonicity!) over the time interval $[0, T]$, and consider Y in terms of a kind of time-reversed dual flow Z coupled to X, using for example $Z^{(y)}_{t,0} = \inf\{u : X^{(u)}_t \geq y\}$ (see [19]).*

4.2 Fill's method

This beautiful idea grew into a method of simulation, and then a method of perfect simulation, *Fill's method* [31], which is an alternative to *CFTP*. It is based on the notion of a *strong uniform time* T [24] and associated notions of *set-valued duals*. Fill's method considered on its own is harder to explain than *CFTP*: we describe it in the simplest context of monotonicity, with state space the unit interval $[0, 1]$.
As described in Remark 4.4, we can view Siegmund duality in terms of a monotonic stochastic flow for $X_t^{(x)}$, and a time-reversed dual flow Z.

> Run X from the minimal state 0 at time 0, forwards to time T.
>
> Now run the coupled Siegmund dual Z from the maximal state 1 at time T *backwards in time* and *coupled to the previously realized path of* X, backwards to time 0.
>
> If $Z_{T,0}^{(1)} = 0$ (the minimal state) then return $X(T)$.
>
> Otherwise repeat the algorithm.

In fact Z is really a set-valued process: $Z_{T,0}^{(y)}$ represents the set of all initial values x which are mapped by the flow X to within the interval $[0, y]$ at time T. (This is the key to removing the monotonicity restriction.)
Despite its complexity, Fill's method has advantages too:

- it can provide *user-interruptibility*, subject to suitable implementation (perfect draws are not biased by censoring draws which take longer than a specified threshold);
- as we will see, it can be viewed as a conditional version of *CFTP*, and the conditioning can be used to speed up the algorithm.

4.3 *FMMR* and *CFTP*

Fill's method is at first sight quite different from *CFTP*. However Fill *et al.* [33] establish a profound and informative link with *CFTP–* for which reason it is now conventional to refer to Fill's method as the *FMMR method*. We explain this method using "blocks" as input-output maps for a chain, as in our description of Read-once *CFTP* in §2.6.
First recall that *CFTP* can be viewed in a curiously redundant fashion as follows:

> **Draw from equilibrium** $X(-T)$ and run forwards;
>
> continue to increase T until $X(0)$ is coalesced;
>
> **return** $X(0)$; note that by construction T and $X(0)$ are independent of $X(-T)$.

Figure 16 illustrates this construction, in which we perversely draw from the equilibrium (the very thing we are trying to achieve by this method), only to discard the draw in the course of the algorithm!

Figure 16: A representation of classic *CFTP* using a sequence of blocks.

Key observation: By construction, $X(0)$ and T are independent of $X(-T)$, so we can condition on their values!

Condition on a convenient $X(0)$;

Run X backwards to a fixed time $-T$;

Draw blocks conditioned on the X transitions;

If coalescence **then return** $X(-T)$ **else repeat**.

The construction is illustrated in Figure 17. Viewing this as a conditional form of the perverse representation of *CFTP* above, it follows that the returned value is a perfect draw from the desired equilibrium. Note that in this formulation there is no need to assume the target process X is at all monotonic. The set-valued dual flow Z is produced by the input-output maps. Note that the flow of time is reversed with respect to the description of Fill's method in Section 4.2.

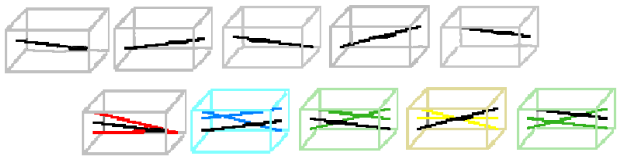


Figure 17: Illustrating *FMMR* using blocks. The top row represents the operation of running X backwards in time from a fixed starting point. The second row represents the procedure of extending this reversed path to a sequence of input-output blocks, for which one must then test coalescence.

This makes it apparent that there are gains to be obtained over *CFTP* by careful selection of the convenient $X(0)$. These gains can be dramatic! (See for example [27].)
It is natural to ask whether *FMMR* and *domCFTP* can somehow be combined.

Question 4.5 *Is there an effective* dominated *version of Fill's method?*

It is possible to devise such a combination in a rather straightforward manner, but implementation appears to lead to substantial theoretical difficulties in all but the most trivial of examples.

4.4 Efficiency and the price of perfection

How *efficient* might *CFTP* be? When there is strong enough monotonicity then useful bounds have been derived – even as early as the Propp-Wilson paper [78]. In the case of monotonic *CFTP* on a finite partially ordered space, Propp and Wilson [78] present a strong bound. Let ℓ be the longest *chain* in the space; let T^* be the coalescence time, let

$$\overline{d}(k) \quad = \quad \max_{x,y}\{P_x^{(k)} - P_y^{(k)}\}\,.$$

Then

$$\frac{\mathbb{P}\left[T^* > k\right]}{\ell} \quad \leq \quad \overline{d}(k) \quad \leq \quad \mathbb{P}\left[T^* > k\right]\,, \tag{8}$$

so *CFTP* is within a factor of being as good as possible.
In general *CFTP has* to involve coupling, usually co-adapted.[n] One expects co-adapted coupling to happen at some exponential rate, and convergence to equilibrium (in total variation norm dist_{TV}!) likewise. From the Coupling Inequality (1) we know that coupling cannot happen faster than convergence to equilibrium. But can it happen at a strictly slower rate? and for relatively simple Markov chains? Coupling can be used to find out about this coupling problem [16]! Here is a sketch of the argument.
Suppose we have the following asymptotics for a continuous-time Markov chain, holding large t:

$$|p_t(x_1, y) - p_t(x_2, y)| \quad \approx \quad c_2 \exp(-\mu_2 t)$$

while

$$\mathbb{P}\left[\tau > t | X(0) = (x_1, x_2)\right] \quad \approx \quad c \exp(-\mu t)\,.$$

(Such exponential rates are typical for many Markov chains.) The standard coupling argument then leads to

$$\begin{aligned}
&|p_t(x_1, y) - p_t(x_2, y)| \quad = \\
&= \quad |\mathbb{P}\left[X_1(t) = y | X_1(0) = x_1\right] - \mathbb{P}\left[X_2(t) = y | X_2(0) = x_2\right]| \; = \\
&|\mathbb{P}\left[X_1(t) = y | \tau > t, X_1(0) = x_1\right] - \mathbb{P}\left[X_2(t) = y | \tau > t, X_2(0) = x_2\right]| \\
&\qquad\qquad \times \mathbb{P}\left[\tau > t | X(0) = (x_1, x_2)\right]
\end{aligned}$$

[n] But Huber has an example which uses non-co-adapted coupling!

Now we proceed to a coupling of couplings! Let X^* be a independently coupled copy of X but transposed so as to begin at (x_2, x_1):

$$\begin{aligned}
&|\,\mathbb{P}\,[X_1(t) = y|\tau > t, X_1(0) = x_1] - \mathbb{P}\,[X_2(t) = y|\tau > t, X_2(0) = x_2]\,| \\
&= \quad |\,\mathbb{P}\,[X_1(t) = y|\tau > t, X(0) = (x_1, x_2)] - \\
&\qquad \mathbb{P}\,[X_1^*(t) = y|\tau^* > t, X^*(0) = (x_2, x_1)]\,| \\
&\leq \mathbb{P}\,[\sigma > t|\tau > t, \tau^* > t, X(0) = (x_1, x_2)] \quad (\approx c' \exp(-\mu' t))
\end{aligned}$$

for σ the time when X, X^* couple.
Thus $\mu_2 \geq \mu' + \mu$, with $\mu_2 > \mu$ if X, X^* couple at exponential rate μ'.

Remark 4.6 *So co-adapted coupling is strictly slower than convergence to equilibrium when a pair of co-adapted coupled chains can transpose before coupling (the opposite of monotonicity!).*

See [16] for more on this. Figure 18 presents a continuous-time Markov chain for which it may be shown that there are *no* co-adapted couplings which occur as fast as the approach to equilibrium: Cranston and Mountford [72] describes a still simpler example!

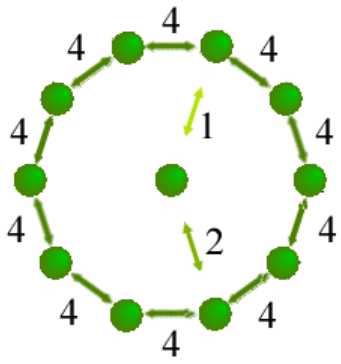


Figure 18: A Markov chain for which there is no efficient co-adapted coupling.

Remark 4.7 *[60] gives a computer-science type example involving graph-matchings: coupling becomes* much *slower than convergence to equilibrium as problem-size increases.*

Of course, this barrier may be overcome by using non-co-adapted couplings: it is an interesting question as to how far it is practical to do this in general.

4.5 Dominated *CFTP* and Foster-Lyapunov conditions

An application which mixes *domCFTP* and small-set *CFTP* is described in [21]. The upper envelope process must be formulated carefully: when the dominating process visits a small set, then one can attempt small-set coupling; however one must take care to ensure that the dominating process remains dominating when small-set coupling is attempted and fails!

There are similarities to Foster-Lyapunov conditions for assessing geometric ergodicity *etc* for Markov chains. Such conditions use a Lyapunov function Λ to deliver a controlled supermartingale off a small set.

We begin by discussing a Foster-Lyapunov condition for positive-recurrence.

Theorem 4.8 *[66] Positive-recurrence on a set C holds if C is a small set and one can find a constant $\beta > 0$, and a non-negative function Λ bounded on C such that for all $n > 0$*

$$\mathbb{E}\left[\Lambda(X_{n+1})|X_n\right] \quad \leq \quad \Lambda(X_n) - 1 + \beta\, \mathbb{I}\left[X_n \in C\right] . \tag{9}$$

Proof: Let N be the random time at which X first (re-)visits C. It suffices[o] to show $\mathbb{E}\left[N|X_0\right] < \Lambda(X_0) + \text{constant} < \infty$ (then use small-set regeneration together with the upper bound on the subsequent value of $\Lambda(X)$ provided by Inequality (9)).

By iteration of (9), we may deduce $\mathbb{E}\left[\Lambda(X_n)|X_0\right] < \infty$ for all n.

If $X_0 \notin C$ then (9) tells us $n \mapsto \Lambda(X_{n\wedge N}) + n \wedge N$ defines a nonnegative supermartingale ($\mathbb{I}\left[X_{(n\wedge N)} \in C\right] = 0$ if $n < N$) . Consequently

$$\mathbb{E}\left[N|X_0\right] \quad \leq \quad \mathbb{E}\left[\Lambda(X_N) + N|X_0\right] \quad \leq \quad \Lambda(X_0)\,.$$

If $X_0 \in C$ then the above can be used to show

$$\begin{aligned}
\mathbb{E}\left[N|X_0\right] &= \mathbb{E}\left[1 \times \mathbb{I}\left[X_1 \in C\right]|X_0\right] + \mathbb{E}\left[\mathbb{E}\left[N|X_1\right]\mathbb{I}\left[X_1 \notin C\right]|X_0\right] \\
&\leq \mathbb{P}\left[X_1 \in C|X_0\right] + \mathbb{E}\left[1 + \Lambda(X_1)|X_0\right] \\
&\leq \mathbb{P}\left[X_1 \in C|X_0\right] + \Lambda(X_0) + \beta
\end{aligned}$$

where the last step uses Inequality (9) applied when $\mathbb{I}\left[X_{n-1} \in C\right] = 1$. □

Now we consider a strengthened Foster-Lyapunov condition for geometric ergodicity.

Theorem 4.9 *[66] Geometric ergodicity holds if one can find a small set C, positive constants $\lambda < 1$, β, and a function $\Lambda \geq 1$ bounded on C such that*

$$\mathbb{E}\left[\Lambda(X_{n+1})|X_n\right] \quad \leq \quad \lambda\Lambda(X_n) + \beta\, \mathbb{I}\left[X_n \in C\right] . \tag{10}$$

[o] This martingale approach can be reformulated as an application of Dynkin's formula.

Proof: Define N as in Theorem 4.8.
Iterating (10), we may deduce $\mathbb{E}\left[\Lambda(X_n)|X_0\right] < \infty$ and more specifically we may infer that

$$n \quad \mapsto \quad \Lambda(X_{n\wedge N})/\lambda^{n\wedge N}$$

is a nonnegative supermartingale. Consequently

$$\mathbb{E}\left[\Lambda(X_N)/\lambda^N|X_0\right] \quad \leq \quad \Lambda(X_0)\,.$$

Using the facts that $\Lambda \geq 1$, $\lambda \in (0,1)$ and Markov's inequality we deduce

$$\mathbb{P}\left[N > n|X_0\right] \quad \leq \quad \lambda^n \Lambda(X_0)\,,$$

which delivers the required geometric ergodicity. □

It is tempting to try to define a dominating process using Λ, especially if one notes the work of Rosenthal [85] on quantitative convergence rates. The expectation inequality of supermartingale-type,

$$\mathbb{E}\left[\Lambda(X_{n+1})|X_n\right] \quad \leq \quad \lambda\Lambda(X_n) + \beta\,\mathbb{I}\left[X_n \in C\right]\,,$$

is enough to control the rate at which X visits C, but for *domCFTP* based on the ordering implied by Λ we require well-behaved *distributional* bounds on the families of distributions

$$\mathfrak{D}_x \quad = \quad \left\{\mathcal{L}\left(\Lambda(X_{n+1})|X_n\right) : X_n = u, \Lambda(u) \leq \Lambda(x)\right\},$$

and it is easy to construct badly behaved examples.

Example 4.10 *There is a Markov chain on $[0,\infty)$ which satisfies the conditions of Theorem 4.9, using $\Lambda(x) = x$ and $C = \{0\}$, but such that* any *Λ-dominating process for X (a process U for which $\Lambda(U_{n+1}) \geq \Lambda(X_{n+1})$ whenever $\Lambda(U_n) \geq \Lambda(X_n)$) must be transient! (see [52, §3].)*

However this issue can be circumvented using sub-sampling.

Theorem 4.11 *[52] If a Markov chain X is geometrically ergodic then it is possible to construct a dominating process based on a Foster-Lyapunov criterion (10), and hence to build a particular kind of* domCFTP *algorithm, for some sub-sampled version X_k, X_{2k}, ...*

Of course, just as for the Foss-Tweedie Theorem 3.9, this algorithm will not be practical! However it is of interest that the constructed dominating process is in some sense "universal": it can be chosen to be the work-load process for a $D/M/1$ queue with parameters depending only on the λ in the relevant Foster-Lyapunov condition (10).[p]

[p] It is natural to ask whether this result can be extended, by analogy with the Foss-Tweedie Theorem 3.9, to show equivalence between suitable *domCFTP* algorithms and some kind of ergodicity criterion: this is now being pursued at Warwick.

4.6 Backward-forward algorithm

A careful look at *domCFTP* for the area-interaction process, or generalizations to other point processes as described in [53], shows that the construction is as follows:

- build a space-time Poisson process of *free points*;
- convert free points into initial points for time-like line segments, hence constructing a space-time birth and death process;
- mark the free points independently;
- apply a causal thinning procedure in time order;
- *domCFTP* succeeds if the time-zero result (the set of points retained under thinning at time zero) of this thinning procedure stabilizes when thinning begins far enough back in time; apply a binary search procedure to capture a time early enough to ensure stabilization at time zero.

Fernández *et al.* [30] describe a variant of perfect simulation (*Backwards-Forwards Algorithm*, or *BFA*) which avoids the need to use binary search to iterate back through successive starts $-T$, $-2T$, $-4T$, ...

- Conduct a recursive *backwards sweep*, identifying all the free points (*ancestors*) which by thinning might conceivably influence subsequent points already identified as potentially influencing points in the region of interest;
- Work forwards through time in a *forwards sweep*,[q] thinning out ancestors to obtain the required perfect sample at time zero (assuming the previous backwards sweep has generated only finitely many ancestors).

Instead of coalescence, we now require sub-criticality of the *oriented percolation* implicit in the backwards sweep; computable conditions arise from standard branching process comparisons, and these conditions will generally apply under sufficiently weak interactions.
The *BFA* generalizes easily to deal with space windows of infinite volume processes (compare the "space-time *CFTP*" mentioned in Section 1.7). [r]
An enlightening theoretical example arises if one reformulates the Ising model using Peierls *contours* (lines separating ± 1 values). As is well known, these form a "non-interacting hard-core gas", interacting by perimeter exclusion, to which the Backwards-Forwards Algorithm may in principle be applied: see [29].

[q] The forwards sweep is deterministic given the free points and marks.

[r] Fernández *et. al.* point out [30], especially for infinite volume processes there is a "user-impatience" bias for which they estimate the effects.

Example 4.12 BFA *can be used to implement a perfect simulation of the Peierls contour model for low temperature Ising models.*

4.7 Some Complements

Thönnes [89] shows how to apply Fill's method to the Häggström *et al.* method (§1.6) for perfect simulation of the area-interaction point process. Møller and Schladitz [70] demonstrate its application to random fields, including anti-monotonic cases.

CFTP can also be used as a supplement to more empirical *MCMC*. We have already mentioned the use of small set *CFTP* by Hobert and Robert [43] (§2.7). Berthelsen and Møller [8] use *domCFTP* as a component in a more conventional *MCMC* approach to analysis of interacting point processes; [56] apply the idea of checking for coupling from maximal and minimal states so as to assure oneself that equilibrium has been achieved.

CFTP, *domCFTP*, *BFA*, and *FMMR* should not be supposed to exhaust the possibilities of perfect simulation! Recently Fill and Huber have introduced the *randomness recycler* [32], which allows a more systematic scan of the stochastic system to be simulated.

References

[1] D. J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3(4):450–465, 1990.

[2] G. K. Ambler and B. W. Silverman. Perfect simulation for bayesian wavelet thresholding with correlated coefficients. preprint, University of Bristol Department of Mathematics, 2004. Available at `http://www.maths.bris.ac.uk/~magka/research/papers/waveSUB.pdf`.

[3] G. K. Ambler and B. W. Silverman. Perfect simulation of spatial point processes using dominated coupling from the past with application to a multiscale area-interaction point process. preprint, University of Bristol Department of Mathematics, 2004. Available at `http://www.maths.bris.ac.uk/~magka/research/papers/att-rep.pdf`.

[4] S. Asmussen, P. W. Glynn, and H. Thorisson. Stationarity detection in the initial transient problem. *ACM Trans. Model. Comput. Simul.*, 2(2):130–157, 1992.

[5] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245:493–501, 1978.

[6] K. B. Athreya and O. Stenflo. Perfect sampling for Doeblin chains. *Sankhya*, 65(4):763–777, 2003.

[7] A. J. Baddeley and M. N. M. van Lieshout. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*, 47:601–619, 1995.

[8] K.K. Berthelsen and J. Møller. An efficient MCMC method for Bayesian point process models with intractable normalising constants. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Spatial point process modelling and its applications*, Publicacions de la Universitat Jaume I, pages 7–15, 2004.

[9] A. A. Borovkov. *Asymptotic methods in queuing theory*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1984.

[10] A. A. Borovkov and S. G. Foss. Stochastically recursive sequences and their generalizations. *Siberian Adv. Math.*, 2(1):16–81, 1992. Siberian Advances in Mathematics.

[11] L. A. Breyer and G. O. Roberts. Catalytic perfect simulation. *Methodol. Comput. Appl. Probab.*, 3(2):161–177, 2001.

[12] L. A. Breyer and G. O. Roberts. A new method for coupling random fields. *LMS J. Comput. Math.*, 5:77–94 (electronic), 2002.

[13] A. Brix and W. S. Kendall. Simulation of cluster point processes without edge effects. *Advances in Applied Probability*, 34(2):267–280, June 2002.

[14] A. Broder. Generating random spanning trees. In *Thirtieth Annual Symposium Foundations Computer Sci.*, pages 442-447, New York, 1989. IEEE.

[15] S.P. Brooks and G.O. Roberts. Assessing convergence of Markov Chain Monte Carlo algorithms. Technical report, University of Cambridge, May 1997.

[16] K. Burdzy and W. S. Kendall. Efficient Markovian couplings: examples and counterexamples. *The Annals of Applied Probability*, 10(2):362–409, May 2000.

[17] Y. Cai and W. S. Kendall. Perfect simulation for correlated Poisson random variables conditioned to be positive. *Statistics and Computing*, 12:229–243, July 2002.

[18] G. Casella, M. Lavine, and C. P. Robert. Explaining the perfect sampler. *Amer. Statist.*, 55(4):299–305, 2001.

[19] P. Clifford and A. Sudbury. A sample path proof of the duality for stochastically monotone Markov processes. *Ann. Probab.*, 13(2):558–565, 1985.

[20] J. N. Corcoran and U. Schneider. Shift and scale coupling methods for perfect simulation. *Probab. Engrg. Inform. Sci.*, 17(3):277–303, 2003.

[21] J. N. Corcoran and R. L. Tweedie. Perfect sampling of ergodic Harris chains. *The Annals of Applied Probability*, 11(2):438–451, 2001.

[22] M.K. Cowles and B.P. Carlin. Markov Chain Monte Carlo convergence diagnostics: A comparative review. Technical report, University of Minnesota, Div. of Biostatistics, USA, May 1995.

[23] R. V. Craiu and X.-L. Meng. Antithetic coupling for perfect sampling. In *Bayesian Methods with Applications to Science, Policy, and Official Statistics. Selected papers from ISBA 2000: the sixth world meeting of the International Society for Bayesian Analysis.*, pages 99–108, 2001.

[24] P. Diaconis and J. Fill. Strong stationary times via a new form of duality. *The Annals of Probability*, 18:1483–1522, 1990.

[25] Persi Diaconis and Daniel Stroock. Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61, 1991.

[26] X. K. Dimakos. A guide to exact simulation. *International Statistical Review*, 69(1):27–48, 2001.

[27] Robert P. Dobrow and J. A. Fill. Speeding up the FMMR perfect sampling algorithm: a case study revisited. *Random Structures and Algorithms*, 23(4):434–452, 2003.

[28] W. Doeblin. Sur les propriétés asymptotiques de mouvement régis par certains types de chaînes simples. *Bull. Math. Soc. Roum. Sci.*, 39(1,2):57–115,3–61, 1937.

[29] Roberto Fernández, Pablo A. Ferrari, and Nancy L. Garcia. Loss network representation of Peierls contours. *The Annals of Probability*, 29(2):902–937, 2001.

[30] Pablo A. Ferrari, Roberto Fernández, and Nancy L. Garcia. Perfect simulation for interacting point processes, loss networks and Ising models. *Stochastic Processes and Their Applications*, 102(1):63–88, 2002.

[31] J. A. Fill. An interruptible algorithm for exact sampling via Markov Chains. *The Annals of Applied Probability*, 8:131–162, 1998.

[32] J. A. Fill and M. Huber. The randomness recycler: a new technique for perfect sampling. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*, pages 503–511. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.

[33] J. A. Fill, M. Machida, D. J. Murdoch, and J. S. Rosenthal. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms*, 17(3-4):290–316, 2000.

[34] S. G. Foss. Ergodicity conditions in multiserver queueing systems with waiting time. *Sibirsk. Mat. Zh.*, 24(6):168–175, 1983.

[35] S. G. Foss and R. L. Tweedie. Perfect simulation and backward coupling. *Stochastic Models*, 14:187–203, 1998.

[36] Sheldon Goldstein. Maximal coupling. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete*, 46(2):193–204, 1978 / 1979.

[37] P. J. Green and D. J. Murdoch. Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion). In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 6*, pages 301–321. The Clarendon Press Oxford University Press, 1999. Presented as an invited paper at the 6th Valencia International Meeting on Bayesian Statistics, Alcossebre, Spain, June 1998.

[38] D. Griffeath. A maximal coupling for Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete*, 31:95–106, 1974 / 1975.

[39] O. Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.

[40] O. Häggström and K. Nelander. Exact sampling from anti-monotone systems. *statistica neerlandica*, 52(3):360–380, 1998.

[41] O. Häggström and J. E. Steif. Propp-Wilson algorithms and finitary codings for high noise Markov random fields. *Combin. Probab. Computing*, 9:425–439, 2000.

[42] O. Häggström, M. N. M. van Lieshout, and J. Møller. Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, 5(5):641–658, 1999. Was: Aalborg Mathematics Department Research Report R-96-2040.

[43] J. P. Hobert and C. P. Robert. A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling. *The Annals of Applied Probability*, 14(3):to appear, August 2004.

[44] C. C. Holmes and B. K. Mallick. Perfect simulation for Bayesian curve and surface fitting. preprint, Texas A&M University Department of Mathematics, 2003. Available at `http://stat.tamu.edu/~bmallick/papers/perf.ps`.

[45] M. Huber. The swap move: a tool for building better Markov chains. In *10th Annual Symposium on Discrete Algorithms*, 1999.

[46] M. Huber. A bounding chain for Swendsen-Wang. *Random Structures and Algorithms*, 22(1):43–59, 2003.

[47] V. E. Johnson. Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. *Journal of the American Statistical Association*, 91(433):154–166, 1996.

[48] W. S. Kendall. Coupling methods and the storage equation. *Journal of Applied Probability*, 20:436–441, 1983.

[49] W. S. Kendall. On some weighted Boolean models. In D. Jeulin, editor, *Advances in Theory and Applications of Random Sets*, pages 105–120, Singapore, 1997. World Scientific.

[50] W. S. Kendall. Perfect simulation for spatial point processes. In *Bulletin ISI, 51^{st} session proceedings, Istanbul (August 1997)*, volume 3, pages 163–166, 1997.

[51] W. S. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C. C. Heyde, editors, *Probability Towards 2000*, pages 218–234, New York, 1998. Springer-Verlag.

[52] W. S. Kendall. Geometric ergodicity and perfect simulation. *Electronic Communications in Probability*, 9:to appear, 2004.

[53] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered state spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, September 2000.

[54] W. S. Kendall and G. Montana. Small sets and Markov transition densities. *Stochastic Processes and Their Applications*, 99(2):177–194, May 2002.

[55] W. S. Kendall and E. Thönnes. Perfect simulation in stochastic geometry. *Pattern Recognition*, 32(9):1569–1586, 1999.

[56] W. S. Kendall and R. G. Wilson. Ising models and multiresolution quad-trees. *Advances in Applied Probability*, 35(1):96–122, March 2003.

[57] Yu. Kifer. *Ergodic theory of random transformations*, volume 10 of *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA, 1986.

[58] R. Kindermann and J. L. Snell. *Markov random fields and their applications*, volume 1 of *Contemporary Mathematics*. American Mathematical Society, Providence, R.I., 1980.

[59] A. N. Kolmogorov. Zur theorie der markoffschen ketten. *Math. Annalen*, 112:155–160, 1936.

[60] V. S. Anil Kumar and H. Ramesh. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms*, 18(1):1–17, 2001.

[61] G. Letac. A contraction principle for certain Markov chains and its applications. In *Random matrices and their applications (Brunswick, Maine, 1984)*, volume 50 of *Contemp. Math.*, pages 263–273. Amer. Math. Soc., Providence, RI, 1986.

[62] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1992. A Wiley-Interscience Publication.

[63] T. Lindvall. *Lectures on the coupling method*. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[64] R. Loynes. The stability of a queue with non-independent inter-arrival and service times. *Math. Proc. Camb. Phil. Soc.*, 58(3):497–520, 1962.

[65] X.-L. Meng. Towards a more general Propp-Wilson algorithm: multistage backward coupling. In *Monte Carlo methods (Toronto, ON, 1998)*, volume 26 of *Fields Inst. Commun.*, pages 85–93. Amer. Math. Soc., Providence, RI, 2000.

[66] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, 1993.

[67] A. Mira, J. Møller, and G. O. Roberts. Perfect slice samplers. *Journal of the Royal Statistical Society (Series B: Methodological)*, 63(3):593–606, 2001. See [68] for corrigendum.

[68] A. Mira, J. Møller, and G. O. Roberts. *Corrigendum:* Perfect slice samplers. *Journal of the Royal Statistical Society (Series B: Methodological)*, 64(3):581, 2002.

[69] J. Møller and J.G. Rasmussen. A note on a perfect simulation algorithm for marked Hawkes processes. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, editors, *Spatial point process modelling and its applications*, Publicacions de la Universitat Jaume I, pages 187–192, 2004.

[70] J. Møller and K. Schladitz. Extensions of Fill's algorithm for perfect simulation. *Journal of the Royal Statistical Society (Series B: Methodological)*, 61(4):955–969, 1999.

[71] J. Møller and R. Waagepetersen. *Statistical inference and simulation for spatial point processes*, volume 100 of *Monographs on Statistics and Applied Probability*. Chapman and Hall / CRC, Boca Raton, 2004.

[72] T. S. Mountford and M. Cranston. Efficient coupling on the circle. In *Game theory, optimal stopping, probability and statistics*, volume 35 of *IMS Lecture Notes Monogr. Ser.*, pages 191–203. Institute of Mathematical Statistics, Beachwood, OH, 2000.

[73] D. Murdoch and J. Rosenthal. Efficient use of exact samples. *Statistics and Computing*, 10:237–243, 1999.

[74] D. J. Murdoch. Exact sampling for Bayesian inference: unbounded state spaces. In *Monte Carlo methods (Toronto, ON, 1998)*, volume 26 of *Fields Inst. Commun.*, pages 111–121. Amer. Math. Soc., Providence, RI, 2000.

[75] P. Mykland, L. Tierney, and B. Yu. Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90(429):233–241, 1995.

[76] E. Nummelin. A splitting technique for Harris-recurrent chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete*, 43:309–318, 1978.

[77] E. Nummelin. *General irreducible Markov chains and nonnegative operators*. Cambridge University Press, Cambridge, 1984.

[78] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.

[79] J. G. Propp and D. B. Wilson. Coupling from the past: a user's guide. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 181–192. American Mathematical Society, Providence, RI, 1998.

[80] G. O. Roberts and N. G. Polson. On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society (Series B: Methodological)*, 56(2):377–384, 1994.

[81] G. O. Roberts and J. S. Rosenthal. Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society (Series B: Methodological)*, 61(3):643–660, 1999.

[82] G. O. Roberts and J. S. Rosenthal. Small and pseudo-small sets for Markov chains. *Stochastic Models*, 17(2):121–145, 2001.

[83] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin diffusions and their discrete approximations. *Bernoulli*, 2:341–364, 1996.

[84] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:96–110, 1996.

[85] J. S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Electronic Communications in Probability*, 7:no. 13, 123–128 (electronic), 2002.

[86] D. Siegmund. The equivalence of absorbing and reflecting barrier problems for stochastically monotone Markov processes. *The Annals of Probability*, 4(6):914–924, 1976.

[87] D. Stoyan. *Comparison methods for queues and other stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Chichester, 1983. Translation from the German edited by Daryl J. Daley.

[88] M. Sweeny. Monte Carlo study of weighted percolation clusters relevant to the Potts models. *Physical Review B*, 27(7):4445–4455, 1983.

[89] E. Thönnes. Perfect simulation of some point processes for the impatient user. *Advances in Applied Probability*, 31(1):69–87, 1999.

[90] E. Thönnes. A primer in perfect simulation. In K. R. Mecke and D. Stoyan, editors, *Statistical Physics and Spatial Statistics*, number 554 in Lecture Notes in Physics, pages 349 – 378. Springer-Verlag, 2000.

[91] H. Thorisson. Backward limits. *The Annals of Probability*, 16(2):914–924, 1988.

[92] H. Thorisson. *Coupling, stationarity, and regeneration*. Springer-Verlag, New York, 2000.

[93] B. Widom and J. S. Rowlinson. New model for the study of liquid-vapor phase transitions. *Journal of Chemical Physics*, 52:1670–1684, 1970.

[94] D. B. Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996)*, pages 296–303, New York, 1996. ACM.

[95] D. B. Wilson. How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms*, 16(1):85–113, 2000.

[96] D. B. Wilson. Layered Multishift Coupling for use in Perfect Sampling Algorithms (with a primer on CFTP). In N. Madras, editor, *Monte Carlo Methods*, volume 26 of *Fields Institute Communications*, pages 143–179. American Mathematical Society, 2000.

[97] D. B. Wilson and J. G. Propp. How to get an exact sample from a generic Markov chain and sample a random spanning tree from a directed graph, both within the cover time. In *Proceedings of the Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996)*, pages 448–457, New York, 1996. ACM.

Other University of Warwick Department of Statistics
Research Reports authored or co–authored by W.S. Kendall.

161: The Euclidean diffusion of shape.
162: Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence.
172: A spatial Markov property for nearest–neighbour Markov point processes.
181: Convexity and the hemisphere.
202: A remark on the proof of Itô's formula for C^2 functions of continuous semimartingales.
203: Computer algebra and stochastic calculus.
212: Convex geometry and nonconfluent Γ-martingales I: Tightness and strict convexity.
: The Propeller: a counterexample to a conjectured criterion for the existence of certain convex functions.
214: Convex Geometry and nonconfluent Γ-martingales II: Well–posedness and Γ-martingale convergence.
216: *(with E. Hsu)* Limiting angle of Brownian motion in certain two–dimensional Cartan–Hadamard manifolds.
217: Symbolic Itô calculus: an introduction.
218: *(with H. Huang)* Correction note to "Martingales on manifolds and harmonic maps."
222: *(with O.E. Barndorff-Nielsen and P.E. Jupp)* Stochastic calculus, statistical asymptotics, Taylor strings and phyla.
223: Symbolic Itô calculus: an overview.
231: The radial part of a Γ-martingale and a non-implosion theorem.
236: Computer algebra in probability and statistics.
237: Computer algebra and yoke geometry I: When is an expression a tensor?
238: *Itovsn3*: doing stochastic calculus with *Mathematica*.
239: On the empty cells of Poisson histograms.
244: *(with M. Cranston and P. March)* The radial part of Brownian motion II: Its life and times on the cut locus.
247: Brownian motion and computer algebra (Text of talk to BAAS Science Festival '92, Southampton Wednesday 26 August 1992, with screenshots of illustrative animations).
257: Brownian motion and partial differential equations: from the heat equation to harmonic maps (Special invited lecture, 49th session of the ISI, Firenze).
260: Probability, convexity, and harmonic maps II: Smoothness via probabilistic gradient inequalities.
261: *(with G. Ben Arous and M. Cranston)* Coupling constructions for hypoelliptic diffusions: Two examples.
280: *(with M. Cranston and Yu. Kifer)* Gromov's hyperbolicity and Picard's little theorem for harmonic maps.
292: Perfect Simulation for the Area-Interaction Point Process.
293: *(with A.J. Baddeley and M.N.M. van Lieshout)* Quermass-interaction processes.
295: On some weighted Boolean models.
296: A diffusion model for Bookstein triangle shape.
301: COMPUTER ALGEBRA: an encyclopaedia article.
308: Perfect Simulation for Spatial Point Processes.
319: Geometry, statistics, and shape.
321: From Stochastic Parallel Transport to Harmonic Maps.
323: *(with E. Thönnes)* Perfect Simulation in Stochastic Geometry.
325: *(with J.M. Corcuera)* Riemannian barycentres and geodesic convexity.
327: Symbolic Itô calculus in *AXIOM*: an ongoing story.
328: *Itovsn3* in *AXIOM*: modules, algebras and stochastic differentials.
331: *(with K. Burdzy)* Efficient Markovian couplings: examples and counterexamples.
333: Stochastic calculus in *Mathematica*: software and examples.
341: Stationary countable dense random sets.
347: *(with J. Møller)* Perfect Metropolis-Hastings simulation of locally stable point processes.
348: *(with J. Møller)* Perfect implementation of a Metropolis-Hastings simulation of Markov point processes
349: *(with Y. Cai)* Perfect simulation for correlated Poisson random variables conditioned to be positive.
350: *(with Y. Cai)* Perfect implementation of simulation for conditioned Boolean Model via correlated Poisson random variables.
353: *(with C.J. Price)* Zeros of Brownian Polynomials.
371: *(with G. Montana)* Small sets and Markov transition densities.
382: *(with A. Brix)* Simulation of cluster point processes without edge effects.
391: *(with E. Thönnes, A. Bhalerao, R.G. Wilson)* A Bayesian approach to inferring vascular tree structure from 2D imagery.
392: *(with A. Bhalerao, E. Thönnes, R.G. Wilson)* Inferring vascular tree structure from 2D and 3D imagery.
402: *(with R.G. Wilson)* Ising models and multiresolution quad-trees.

409: Constructing Markov chains for differentiable curves.
410: Simplification rules for ItoIntegral.
416: *(with C.J. Price)* Coupling iterated Kolmogorov diffusions.
427: Geometric Ergodicity and Perfect Simulation.
428: Notes on Perfect Simulation.

Also see the following related preprints

317: E. Thönnes: Perfect Simulation of some point processes for the impatient user.
334: M.N.M. van Lieshout and E. Thönnes: A Comparative Study on the Power of van Lieshout and Baddeley's J-function.
359: E. Thönnes: A Primer on Perfect Simulation.
366: J. Lund and E. Thönnes: Perfect Simulation for point processes given noisy observations.

If you want copies of any of these reports then please email your requests to the secretary using `<statistics@warwick.ac.uk>` (mail address: the Department of Statistics, University of Warwick, Coventry CV4 7AL, UK).