NOTES ON PERFECT SIMULATION

Wilfrid S. Kendall

Department of Statistics, University of Warwick

3rd March 2004

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

Version 1.28
## Contents

1. *CFTP*: the classic case ....................................................... 7  
2. *CFTP* and regeneration ..................................................... 35  
3. Dominated *CFTP* ............................................................... 51  
4. Theory and connections ...................................................... 77  

References ........................................................................... 103
Perfect (or exact) 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 underlying ideas of perfect simulation stretch back a long way, but they sprang into prominence as a result of a ground-breaking paper of Propp and Wilson (1996), which showed how to obtain exact draws from (for example) the critical Ising model.

In these lectures I will describe several themes of perfect simulation:

- **Lecture 1** the original or classic Coupling From The Past algorithm (*CFTP*);
- **Lecture 2** variations which exploit regeneration ideas: small-set or split-chain constructions from Markov chain theory (small-set *CFTP*);
- **Lecture 3** generalizations which deal with non-monotonic and non-uniformly ergodic examples (dominated *CFTP*); and
- **Lecture 4** striking results relating *CFTP* to an apparently different algorithm due originally to Fill, as well as other theoretical complements.

The talks (which will be organized approximately under the four headings above) will be illustrated with online animations of *CFTP* simulations and will describe related theoretical issues. The objective of these tutorial lectures is to give a sound and clear exposition of these themes, and hence to promote an interplay of ideas with practitioners of Markov Chain Monte Carlo.
Useful reading


- Online bibliography:
  http://research.microsoft.com/~dbwilson/exact/
Lecture 1

*CFTP*: the classic case

1.1 Coupling
1.2 Random walk *CFTP*
1.3 The *CFTP* theorem
1.4 Falling leaves
1.5 Ising *CFTP*
1.6 Point process *CFTP*
1.7 *CFTP* in space
1.8 Pre-history
Recall basic facts concerning Markov chain Monte Carlo (MCMC):

- Fundamental paradigm:
  - specify target distribution indirectly;
  - realize as equilibrium of Markov chain;
  - sample approximately from target distribution by running Markov chain for a long time (till near-equilibrium).

- How long is “burn-in” period?
  - Guess or diagnose (Cowles and Carlin 1995; Brooks and Roberts 1997),
  - estimate, whether analytically (Diaconis and Stroock 1991; Roberts and Polson 1994; Roberts and Tweedie 1996a; Roberts and Tweedie 1996b), or empirically (Johnson 1996),
  - or do better?
MCMC arises in a number of different areas of the mathematical sciences, with different emphases. (This makes interaction interesting and fruitful!) Here are some examples:

- **Statistical mechanics.** Infinite-dimensional systems: Are there phase transition phenomena? How do they behave?

- **Computer science.** Approximate counting: How does the algorithm behave as the scale of the problem increases? Does the run-time increase exponentially, or polynomially?

- **Image analysis.** Noisy geometric pictures: Can we clean up the pictures? Can we identify significant features?

- **Statistics.**
  - **Bayesian.** Can we draw accurately (and fast 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 Propp and Wilson (1996) CFTP idea:

- Design modified MCMC algorithms which deliver exact draws after randomly long run-time.
- Construct actual implementations for interesting cases.
- Approach is based on probabilistic coupling methods.
- Competitive with ordinary MCMC for amenable cases.

This is exact or perfect simulation.\(^1\)
Practical applications use a large number of different ideas, which will be introduced in these lectures.

\(^1\)Mark Huber suggests a useful taxonomy: divide exact sampling methods into (a) direct (for example, rejection sampling) and (b) perfect (for example, the Propp-Wilson approach to MCMC).
1.1. Coupling and convergence: the binary switch

We commence by introducing the fundamental idea of coupling (see Lindvall 2002, or my 2003 Durham symposium online notes for more on coupling). 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).

**Construction:** Supply

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

(b) independently *ditto* of $1 \rightarrow 0$ transitions.

Apply where applicable to build coupled processes $X, Y$ begun at 1, 2. **Clearly** $X, Y$ are (coupled) copies of the binary switch, coupling at the time $T$ of the first Poisson incident.
Then coupling shows

\[
dist_{TV}(X_t, \pi) = \sup_A \{\mathbb{P}[X_t \in A] - \pi(A)\} \leq \mathbb{P}[T \leq t]
\]

(1.1)

where

- \( \pi \) is the equilibrium distribution,
- \( \text{dist}_{TV} \) is total variation distance.

Diaconis (1996) reviews interesting features of convergence to equilibrium for random walks on the hypercube; the above chain is a basic building block of such random walks.
The coupling argument generalizes to arbitrary Markov chains:

- if we can couple a general Markov chain $X$ to a version $Y$ in statistical equilibrium, then Equation (1.1) bounds the approach to equilibrium;

- if we allow non-adapted couplings, the bound is sharp (see Griffeath 1975, Goldstein 1979);

- non-adapted couplings can be very difficult to construct! Co-adapted couplings can supply usable bounds but need not be sharp.

- Diaconis (1996) describes the cutoff phenomenon in simple contexts. Here is an interesting (but probably very hard) research question: how to relate CFTP to cutoff phenomena? (Useful hints in Propp and Wilson 1996, Wilson 2004.)

Can we use such a coupling to draw from equilibrium? The binary switch example is deceptive: $X(T)$ is in equilibrium in this case but in general will not be.
1.2. Random walk \textit{CFTP}

Apply coupling to random walk \(X\) \textit{reflected} in \(\{1, 2, \ldots, N\}\). Use \textit{synchronous} coupling (generalizes coupling above!): then \(X\) only couples at 1, \(N\). Thus \(X(T)\) cannot be a draw from equilibrium if \(N > 2\).

The Propp-Wilson idea draws on a well-established theme from ergodic theory: realize a Markov chain as a stochastic flow and evolve it not into the future but \textit{from the past}! However if we do this then we need to consider coupled realizations of the Markov chain started at all possible starting points. By monotonicity, we can focus on:

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

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

since these \textit{sandwich} all other realizations begun at time \(-n\).
• Run upper and lower processes from time $-n$.

• If coupling by time 0, return the common value.

• Otherwise, repeat but start at time $-2n$ (say).

Issues:

• Run from the past, not into the future;

• Re-use randomness;

• Sample at time 0, not at coupling time;

• There is a *rationale* for $n \to 2n$: binary search.
Exercise 1.1 Use simple R statistical package scripts and elementary calculations to investigate bias resulting from the following.

- Running into the future, not from the past, stopping (say) at the first time $t = 2^n$ after coupling has occurred;
- Failing to re-use randomness (we’d expect this to bias towards “earlier” coalescence);
- Sampling at coupling time instead of time 0 (obviously a bad idea; sampling an independent random walk at this time still gives a biased result).
1.3. The CFTP theorem

Morally the proof of classic CFTP is just 3 lines long. Express the coupling for $X$ in terms of random input-output maps $F_{(-u,v]}$.

**Theorem 1.2 (Propp and Wilson 1996):** If coalescence is almost sure then CFTP samples from equilibrium.

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

$$X_{t}^{-n} = F_{(-n,t]}(0) \quad \text{for } -n \leq t.$$

**Finite coalescence time $-T$ is assumed.** So

$$X_{0}^{-n} = X_{0}^{-T} \quad \text{whenever } -n \leq -T;$$

$$\mathcal{L}(X_{0}^{-n}) = \mathcal{L}(X_{n}^{0}).$$

If $X$ converges to an equilibrium $\pi$ in total variation $\text{dist}_{TV}$ then

$$\text{dist}_{TV}(\mathcal{L}(X_{0}^{-T}), \pi) = \lim_{n} \text{dist}_{TV}(\mathcal{L}(X_{0}^{-n}), \pi) = \lim_{n} \text{dist}_{TV}(\mathcal{L}(X_{n}^{0}), \pi) = 0$$

hence result. \qed
Remark 1.3  Note that we are free to choose any “backwards stopping 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: see for example the “block-by-block” strategy of read-once CFTP.

Remark 1.4  Note that monotonicity of the target process is convenient for CFTP, but not essential. Propp and Wilson (1996, §3.2) formalize use of monotonicity. Kendall (1998) describes a “crossover” trick for use in anti-monotonic situations; Kendall and Møller (2000) generalize the trick to cases where monotonicity is absent (see also Huber 2003); Häggström and Nelander (1998) apply the trick to lattice systems.
1.4. The falling leaves of Fontainebleau

Kendall and Thönnes (1999) describe a visual and geometric application of $CFTP$ in mathematical geology: this particular example being well-known to workers in the field previous to the introduction of $CFTP$ itself.

Occlusion $CFTP$ for the falling leaves of Fontainebleau.  
(The “dead leaves” model.)

Why “occlusion”? David Wilson’s terminology: this $CFTP$ algorithm builds up the result piece-by-piece with no back-tracking.

(Dead leaves in recent image analysis work: see Lee, Mumford, and Huang 2001, Gousseau and Roueff 2003.)
Corollary 1.5  Occlusion CFTP as described above delivers a sample from the dead leaves distribution.

Proof: Use the notation of Theorem 1.2: $F_{(-n,t]}$ now represents superposition of random leaves falling over the period $(-n,t]$. It suffices to note, general Markov chain theory\(^2\) can be combined with the apparatus of stochastic geometry to show the random pattern converges in distribution to a limiting random pattern. \(\square\)

Exercise 1.6  Show that bias results from simulating forwards in time till the image is completely covered.

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!

\(^2\)eg: regeneration when window is completely covered.
Remark 1.7  Web animations of perfect simulation for the dead leaves model can be found at
www.warwick.ac.uk/go/wsk/abstracts/dead/

Remark 1.8  Other examples of occlusion CFTP include the Aldous-Broder algorithm for generating random spanning trees (Broder 1989; Aldous 1990; Wilson and Propp 1996; Wilson 1996).
1.5. Ising CFTP


Classic CFTP for the Ising model (simple, sub-critical case).
Heat-bath dynamics run from past; compare maximal and minimal starts.

The simulation displays the results of a systematic scan Gibbs sampler. The left-most image is generated by the lower chain; the rightmost image by the upper chain, while the middle red image expresses the difference. The classic CFTP recipe is followed: after cycling through a time interval \([-T, 0]\), the simulation extends back to cycle through \([-2T, 0]\) unless the upper and lower chains coincide at time 0 (middle image is blank at end of cycle).

Details of the target Ising measure are configurable (neighbourhood structure varying between 4 nearest neighbours, 4+4 nearest and next-nearest neighbours, with a “symmetry” option fixing next-nearest interaction as \(1/\sqrt{2}\) of nearest interaction). Coalescence becomes very slow as parameters approach super-criticality.
We can also use this for the *conditioned* Ising model,\(^3\) as used in image analysis applications (see also Besag’s lectures).\(^4\)

Classic *CFTP* for the Ising model (super-critical case, conditioned). Heat-bath dynamics run from past; compare maximal and minimal starts.

Simulation display and options as before. Coalescence speed now depends on interaction between correlation parameters and conditioning, as could be predicted from theoretical considerations (Kindermann and Snell 1980).

\(^3\)Conditioning \(\equiv\) external magnetic field!

\(^4\)Note however, the Ising model is a very poor model for this particular (dithered) image!
Remark 1.9  Web animations of perfect simulations of conditioned Ising models (using simple images which are rather more suitable for 4-neighbour Ising models!) can be found at
www.warwick.ac.uk/go/wsk/ising-animations/
1.6. Point process *CFTP*

Classic *CFTP* is not limited to discrete models. We have already seen this in the case of the falling leaves model. Here is a further example: perfect simulation procedure for attractive area-interaction point processes (*Häggström, van Lieshout, and Møller 1999*).

Area-interaction point process: weight a Poisson process realization according to the area of the region closer than \( r \) to the pattern:

\[
\text{pattern density} \propto \gamma^{-\text{area of region within distance } r \text{ of pattern}}.
\]  

(1.2)

**Exercise 1.10** If \( \gamma > 1 \) (attractive case only!), show the above density is proportional to the probability that an independent Poisson process places no points within distance \( r \) of the pattern.

---

\(^5\) Also known as the *Widom and Rowlinson (1970)* model.
Hence the area-interaction point process may be represented as the pattern of red 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 (classic!) CFTP.

First we describe the Gibbs sampler, then we demonstrate the CFTP procedure using an animation.
Construct a Poisson point process (centres of red crosses).

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

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

Continue this procedure . . . (notice red/blue duality!)
Gibbs sampler CFTP for the attractive area-interaction point process as a marginal of a two-type soft-core repulsion point process.

The CFTP construction is based on the observations

1. there is a partial order for states (red pattern, blue pattern);

2. “highest” and “lowest” states are (“full” pattern, ∅) and (∅, “full” pattern):
   these are “pseudo-states”.

The fact that there are highest and lowest pseudo-states is the key to this algorithm’s quick convergence (at least when there is no phase-transition effect): the underlying Markov chain is uniformly ergodic.
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. See Kendall (1997b) for a description and Häggström and Steif (2000) for related theory, also §4.5 on the BFA algorithm.

In this case the binary search rationale guides how we should extend the *CFTP* algorithm in both space and time.
1.8. Pre-history of *CFTP*

The ideas of *CFTP* have been “in the air” (Jim Propp) for a long time. Elements can be found in:

- Kolmogorov (1936)\(^6\)
- Doeblin (1937)
- Loynes (1962), Kendall (1983)
- Letac (1986), Kifer (1986)
- Broder (1989), Aldous (1990)
- Thorisson (1988)

\(^6\)Thanks to Thorisson (2000) for this reference . . . .
For example consider Kolmogorov (1936):

Consider an inhomogeneous (finite state) Markov transition kernel $P_{ij}(s, t)$. Can it be represented as the kernel of a Markov chain begun in the indefinite past?

Yes!

Apply diagonalization arguments to select a convergent subsequence from the probability systems arising from progressively earlier and earlier starts:

$$\mathbb{P} \left[ X(0) = k \right] = Q_k(0)$$
For example consider Kolmogorov (1936):

Consider an inhomogeneous (finite state) Markov transition kernel $P_{ij}(s, t)$. Can it be represented as the kernel of a Markov chain begun in the indefinite past?

Yes!

Apply diagonalization arguments to select a convergent subsequence from the probability systems arising from progressively earlier and earlier starts:

$$\mathbb{P}[X(0) = k] = Q_k(0)$$
$$\mathbb{P}[X(-1) = j, X(0) = k] = Q_j(-1)P_{jk}(-1, 0)$$
For example consider Kolmogorov (1936):

Consider an inhomogeneous (finite state) Markov transition kernel $P_{ij}(s, t)$. Can it be represented as the kernel of a Markov chain begun in the indefinite past?

Yes!

Apply diagonalization arguments to select a convergent subsequence from the probability systems arising from progressively earlier and earlier starts:

$$
P[X(0) = k] = Q_k(0)$$

$$
P[X(-1) = j, X(0) = k] = Q_j(-1)P_{jk}(-1, 0)$$

$$
P[X(-2) = i, X(-1) = j, X(0) = k] = Q_i(-2)P_{ij}(-2, -1)P_{jk}(-1, 0)$$
For example consider Kolmogorov (1936):

Consider an inhomogeneous (finite state) Markov transition kernel $P_{ij}(s, t)$. Can it be represented as the kernel of a Markov chain begun in the indefinite past?

Yes!

Apply diagonalization arguments to select a convergent subsequence from the probability systems arising from progressively earlier and earlier starts:

$\mathbb{P} [X(0) = k] = Q_k(0)$
$\mathbb{P} [X(-1) = j, X(0) = k] = Q_j(-1)P_{jk}(-1, 0)$
$\mathbb{P} [X(-2) = i, X(-1) = j, X(0) = k] = Q_i(-2)P_{ij}(-2, -1)P_{jk}(-1, 0)$
\ldots
Lecture 2

**CFTP and regeneration**

2.1 Small sets
2.2 Small-set CFTP
2.3 Slice sampler CFTP
2.4 Multi-shift sampler
2.5 Catalytic CFTP
2.6 Read-once CFTP
2.7 Simulated tempering
2.8 Pre-history
2.1. Small sets

Suppose we want a coupling between two random variables $X, Y$ yielding a maximal positive chance of $X = Y$ and otherwise no constraint. (Related to notion of convergence stationnaire, or “parking convergence”, from stochastic process theory.) Clearly this 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 $\alpha = \int (f \land g)(x) \, dx$, and with probability $\alpha$ return a draw of $X = Y$ from the density $(f \land g)/\alpha$.
- Otherwise draw $X$ from $(f - f \land g)/(1 - \alpha)$ and $Y$ from $(g - f \land g)/(1 - \alpha)$.

This can be combined usefully with 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 $S$, transition kernel $p(x, dy)$. The set $C \subseteq S$ is a small set of order $k$ if for some probability measure $\nu$ and some $\alpha > 0$

$$p^{(k)}(x, dy) \geq \mathbb{1}[C](x) \times \alpha \nu(dy). \quad (2.1)$$

Here $\mathbb{1}[C](x)$ is the indicator function for the set $C$.

The animation displays evolution of a Markov chain on the unit interval whose transition density $p(x, y)$ is triangular with peak at $x$. Here the small set is the whole state space $[0, 1]$, of order $k = 1$, $\alpha = 1/2$, and $\nu$ has the isosceles triangle density over $[0, 1]$.

It is a central result that small sets (possibly of arbitrarily high order) exist for any modestly regular Markov chain (classical proof: see for example either of Nummelin 1984; Meyn and Tweedie 1993).
The trouble with general Markov chains is that the chain may have zero probability of returning to any 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 \( S \), transition kernel \( p(x, dy) \), with small set \( C \) of order 1. Then \( X \) can be represented using a new Markov chain on \( S \cup \{c\} \), for \( c \) a regenerative pseudo-state.

For details see Nummelin (1978), also Athreya and Ney (1978). 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. See Meyn and Tweedie (1993) also Roberts and Rosenthal (2001).¹

¹Roberts and Rosenthal (2001) also introduce “pseudo-small sets”, which relate to coupling as small sets relate to CFTP.
It is natural to ask whether we can go further, and use small sets to re-model the chain as an entirely discrete affair. However . . .

**Remark 2.3** Small sets of order 1 need not exist: however they do exist if (a) the kernel \( p(x, dy) \) has a measurable density and (b) chain is sub-sampled at even times. (Both are needed: example in Kendall and Montana 2002.)
Theorem 2.4 (Kendall and Montana 2002) 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: Key Lemma, 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 $\varepsilon$ decreases, such that

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

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

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

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

$$\text{Leb}([x - u, x + v] \cap A_\varepsilon)/(u + v) \to 1.$$
2.2. Murdoch-Green small-set CFTP

Green and Murdoch (1999) showed how to use small sets to carry out CFTP when the state space is continuous with no helpful ordering. Their prescription includes the use of a partition by several small sets, to speed up coalescence.

Small-set CFTP in nearly the simplest possible case: triangular kernel on \([0, 1]\).

**Exercise 2.5** Express Small-set CFTP as composition of Geometrically many kernels (conditioned to avoid small-set effect), with starting point randomized by small-set distribution \(\nu\).
2.3. **Slice sampler CFTP**

Task is to draw from density $f(x)$. Here is a one-dimensional example! (But note, this is only interesting because it can be made to work in many dimensions . . . ) Suppose $f$ unimodal. Define $g_0(y)$, $g_1(y)$ implicitly by: $[g_0(y), g_1(y)]$ is the super-level set $\{x : f(x) \geq y\}$. Alternate between drawing $y$ uniformly from $[0, f(x)]$ and drawing $x$ uniformly from $[g_0(y), g_1(y)]$.

![Graph of a unimodal distribution](image)

Roberts and Rosenthal (1999, Theorem 12) show rapid convergence (order of 530 iterations!) under a specific variation of log-concavity. Mira, Møller, and Roberts (2001) describe a perfect variant.
**Exercise 2.6** Design your own perfect slice sampler for the case when the unimodal density \( f(x) \) has bounded support.

**HINT:** You need to figure out how to make uniform choices for two versions of the process simultaneously, so as to preserve the partial ordering

\[
(x, y) \preceq (x', y') \quad \text{if} \quad f(x) \leq f(y),
\]  

but also so as to have positive chances of coalescing.
2.4. Multi-shift sampler

**Question 2.7** How to draw simultaneously from $\text{Uniform}(x, x + 1)$ for all $x \in \mathbb{R}$, and couple the draws?

**Answer:** Random unit span integer lattice (Wilson 2000b).

Wilson (2000b) also considers more general distributions! For example, we can express a normal distribution as a mixture of uniform distributions, and this allows us to do the same trick in several ways. Also deals with multivariate and even multi-modal cases.

The log-scale is useful when dealing with Gamma distributions. Corcoran and Schneider (2003) carry this even further.
2.5. Catalytic *CFTP*

Breyer and Roberts (2002) 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 \(\Phi_x\), 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 \(\Phi_x\) can be varied from time point to time point.

Simulations can be viewed at

www.lbreyer.com/fcoupler.html
2.6. Read-once CFTP

Wilson (2000a): build $F_{(-n,0]}$ of Theorem 1.2 from i.i.d. blocks

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

The blocking length $t$ is chosen so that there is a positive chance of the map $B = \mathcal{D} F_{(-t,0]}$ being coalescent. By a simple computation, the 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 $t$) and with dead leaves CFTP (one can view the argument as a re-ordering in time).

**Exercise 2.8** Prove the validity of read-once CFTP 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 only block-by-block. *Exercise 2.5 should be helpful!*

The choice of the blocking length $t$ is of course crucial!
2.7. Perfect simulated tempering

Recall that simulated tempering embeds the target distribution $\pi = \pi_0$ as one of a family $\pi_i : i \in I$ of distributions: for each $i \in I$ one devises a chain $X_i$ leaving $\pi_i$ invariant, and carries out simulated tempering by evolving according to chain $X_i$ while making occasional proposals to change $i$ to another value $i' \in I$. The distribution of the resulting chain is proportional to the target $\pi_0$ when restricted to $i = 0$.

Perfect simulated tempering (Møller and Nicholls 1999; Brooks et al. 2002) devises schemes in which the chain at $i = N$ (say) coalesces very fast, moreover it is possible to construct upper and lower $i$-chains after the manner of random walk CFTP. One can then carry out CFTP by evolving the upper and lower $i$-chains till (a) they coalesce at $i = N$, and (b) during the period of $i$-coalescence the $X_N$ chain itself coalesces.

This combines well with the ideas of read-once CFTP applied to the upper and lower $i$-chains (Brooks et al. 2002).
2.8. Pre-history of small-set $CFTP$

- Athreya and Ney (1978), Nummelin (1978)
- Asmussen, Glynn, and Thorisson (1992)
- Mykland, Tierney, and Yu (1995)
Lecture 3

Dominated *CFTP*

3.6 General theorem

3.1 Queues
3.2 Ergodicity
3.3 Birth and death
3.4 Perpetuities
3.5 Point processes
3.1. Queues

Consider a $GI/G/1$ queue. Lindley noticed a beautiful representation for waiting time $W_n$ of customer $n$ in terms of services $S_n$ and inter-arrivals $X_n$...

**Theorem 3.1 (Lindley’s equation)** Consider the $GI/G/1$ queue waiting time identity.

\[
W_{n+1} = \max\{0, W_n + S_n - X_{n+1}\} = \max\{0, W_n + \eta_n\} = \max\{0, \eta_n, \eta_n + \eta_{n-1}, \ldots, \eta_n + \eta_{n-1} + \ldots + \eta_1\} =_{D} \max\{0, \eta_1, \eta_1 + \eta_2, \ldots, \eta_1 + \eta_2 + \ldots + \eta_n\}
\]

and thus we obtain the steady-state expression

\[
W_{\infty} =_{D} \max\{0, \eta_1, \eta_1 + \eta_2, \ldots\}.
\]

If $\text{Var}[\eta_i] < \infty$ then SLLN/CLT/random walk theory shows $W_{\infty}$ will be finite if and only if $\mathbb{E}[\eta_i] < 0$ or $\eta_i \equiv 0$. 
Supposing we lose independence? Loynes (1962) discovered a coupling application to queues with (for example) general dependent stationary inputs and associated service times, pre-figuring CFTP.

**Theorem 3.2** Suppose queue arrivals follow a stationary point process stretching back to time \(-\infty\). Denote arrivals/associated service times in \((s, t]\) by \(N_{s,t}\) (stationarity: statistics of process \(\{N_{s,s+u} : u \geq 0\}\) do not depend on \(s\)). 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

\[
\lim_{T \to \infty} Q^T \quad \text{exists almost surely.}
\]

Remark 3.4  Coupling and CFTP ideas enter into Theorem 3.1 at the crucial time-reversal step:

\[
\max\{0, \eta_n, \eta_n + \eta_{n-1}, \ldots, \eta_n + \eta_{n-1} + \ldots + \eta_1\} = \\
= \mathcal{D} \max\{0, \eta_1, \eta_1 + \eta_2, \ldots, \eta_1 + \eta_2 + \ldots + \eta_n\}
\]

Compare Section 1.4 on falling leaves ….

Remark 3.5  The problem about applying Theorem 3.1 in CFTP is that we don’t know which \( \eta_1 + \eta_2 + \ldots + \eta_n \) attains the supremum \( \max\{0, \eta_1, \eta_1 + \eta_2, \ldots\} \).

Exercise 3.6  Use simple R statistical package scripts and elementary calculations to implement the Loynes coupling.
3.2. Uniform and Geometric Ergodicity

There is a theoretical issue road-blocking 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 is $R(x) > 0$, $0 < \rho < 1$ with

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

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

So a chain exhibits geometric ergodicity if equilibrium is approached at a geometric rate. Note that the geometric rate kicks in at a point which may be affected by the chain’s starting point. However . . .
Definition 3.8 A Markov chain with transition kernel $p(x, \cdot)$ possesses uniform ergodicity: there are $\rho \in (0, 1)$, $R > 0$ not depending on the starting point $x$ such that

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

for all $n$, uniformly for all starting points $x$.

So a chain exhibits uniform ergodicity if the geometric rate kicks in at a point at which is independent of the chain’s starting point.

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, van Lieshout, and Møller (1999) chain in Section 1.6 is a good example.
Foss and Tweedie (1998) show that the (theoretical) possibility of classic CFTP is equivalent (under modest regularity) to uniform ergodicity (rate of convergence to equilibrium does not depend on start-point). For classic CFTP needs vertical coalescence: every possible start from time $-T$ leads to same result at time 0, and it is easy to show that this implies a uniformly geometric rate of convergence to equilibrium.

Figure 3.1: Vertical coalescence
The converse, that uniform ergodicity implies classic CFTP under modest regularity is possible in principle, follows from small set theory. The required regularity is that from any starting point the chain should have positive chance of visiting a small set $C$ of some order $m$.\(^1\)

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 $n + m$ time steps, and this permits construction of small-set CFTP.

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!

---

\(^1\)Up to a null-set, this is a consequence of $\phi$-irreducibility (see for example Nummelin 1984).
It follows from the very existence of CFTP constructions, all our chains 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: “all sufficiently early starts from a specific location * lead to the same result at time 0”. But, as the Lindley representation highlights, the question is how to identify when this has happened?

Figure 3.2: Horizontal coalescence
Remark 3.9 **Dominated CFTP idea** (domCFTP): Generate target chains using 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.

If we can make this work then we can apply CFTP to Markov chains which are merely geometrically ergodic (Kendall 1998; Kendall and Møller 2000; Kendall and Thönnes 1999; Cai and Kendall 2002) or worse (geometric ergodicity ≠ domCFTP!).
Theorem 3.10 *(Kendall 1998; Kendall and Møller 2000): If coalescence is almost sure then domCFTP samples from equilibrium.*

**Proof:** For simplicity we suppose the target process $X$ 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 $0$, $Y(-n)$ 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 1.2 for classic CFTP: If $X$ converges to an equilibrium $\pi$ in total variation $\text{dist}_{TV}$ then

$$\text{dist}_{TV}(\mathcal{L}(X_{-T}^0), \pi) = \lim_{n} \text{dist}_{TV}(\mathcal{L}(X_{-n}^0), \pi) = \lim_{n} \text{dist}_{TV}(\mathcal{L}(X_{n}^0), \pi) = 0$$

hence result.$^2$

$^2$Note: the very last step makes implicit use of stationarity of $Y$!
Figure 3.3: Dominated CFTP

Thus we can use realizations of the target process started from the dominating process to identify horizontal coalescence.
3.3. Non-linear birth-death processes

To illustrate $domCFTP$ in detail, we describe a simple example taken from Kendall (1997a). Consider a continuous-time non-linear birth-death process $X$, with transition rates

$$X \rightarrow X + 1 \quad \text{at rate } \lambda_X,$$

$$X \rightarrow X - 1 \quad \text{at rate } X\mu,$$

for positive $\lambda_X, \mu$. We suppose the birth rate $\lambda_X$ is bounded above\(^3\) by $\lambda$.

Of course it is possible to compute the equilibrium distribution using detailed balance. However here we aim to construct a $domCFTP$ method to draw exactly from the target distribution.

\(^3\)Kendall (1997a) requires monotonicity for $\lambda_X$, which is unnecessarily restrictive!
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

\[
X \to X + 1 \quad \text{at rate } \lambda,
\]
\[
X \to X - 1 \quad \text{at rate } X\mu.
\]

Here *domination* means, if $0 \leq X(0) \leq Y(0)$ then we can construct coupled copies of $X$ and $Y$ such that the relationship $X \leq Y$ is maintained for all time.

Indeed we can go further: given the process $Y$ then for *any* $x$, $0 \leq x \leq Y(0)$, we can construct a copy $X^x$ of $X$ begun at $x$ such that $0 \leq X^a \leq X^b \leq Y$ for all time whenever $a \leq b \leq Y(0)$.

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.

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

$$U \leq \frac{\lambda_X(t-)}{\lambda}; \quad (3.1)$$

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

$$U \leq \frac{\mu_X(t-)}{\mu_Y(t-)} = \frac{X(t-)}{Y(t-)}. \quad (3.2)$$
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 $-n, -2n, \ldots$ using a stationary realization of the dominating process, 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).
3.4. A simple application of *domCFTP*: Perpetuities

Consider the simplest *perpetuity*, with distribution defined by

\[ \mathcal{L}(X) = \mathcal{L}(U(1 + X)) \]  \hspace{1cm} (3.3)

where \( U \) is a Uniform(0, 1) random variable which is independent of \( X \). Clearly \( X \) should have limiting distribution given by the infinite product

\[ U_1(1 + U_2(1 + U_3(1 + \ldots))) \]  \hspace{1cm} (3.4)

where the \( U_i \) are independent Uniform(0, 1) random variables, so \( U_i \) can be viewed as a random interest payment.

We want to know about this limiting distribution of \( X \): in effect we want to know about the equilibrium behaviour of the Markov chain whose update is

\[ X_{n+1} = U_{n+1}(1 + X_n) . \]
Monotonicity and CFTP ideas can be applied to the infinite product (3.4): certainly the limiting distribution satisfies

$$\mathcal{L}(X) \geq \lim_{n} \{U_1(1 + U_2(1 + U_3(1 + \ldots (1 + U_n \times 0) \ldots ))\}$$

and it is possible to establish equality using small sets and Foster-Lyapunov theory. However we are left with the dilemma already faced when considering the Lindley representation in the Section 3.1 on queues: we have horizontal coalescence here rather than vertical coalescence, so classical CFTP is inapplicable.
Here is how to apply domCFTP. By taking logs, we can produce an upper bound on the Markov chain $X_n$ which is a reflecting random walk:

\[
\log X_{n+1} = \log U_{n+1} + \log \left(1 + \frac{1}{X_n}\right) + \log X_n
\]

\[
\leq -V_{n+1} + \log \left(1 + \frac{1}{k}\right) + \log X_n \quad \text{if } X_n \geq k;
\]

where $V_{n+1}$ is an Exponential(1) random variable. Thus $\log X \leq Z$ if $Z$ is the reflecting random walk given by

\[
Z_{n+1} = \begin{cases} 
\alpha - V_{n+1} + Z_n & \text{if } \alpha - V_{n+1} + Z_n \geq \log k, \\
\alpha & \text{otherwise},
\end{cases}
\]

where $\alpha = \log(1 + 1/k)$. 

(3.5)
We need an equilibrium distribution for $Z$, which can be viewed as the workload process of a modest variation on a discrete $M/D/1$ queue. Actually it is easiest to bound $Z$ above by a reversible reflecting simple random walk $W$: we approximate $\alpha - V_{n+1}$ by a discrete random variable

$$W_{n+1} = \begin{cases} 
\alpha & \text{if } V_{n+1} \leq \alpha, \text{ with probability } 1 - e^{-\alpha}, \\
0 & \text{if } \alpha \leq V_{n+1} < 2\alpha, \text{ with probability } e^{-\alpha} - e^{-2\alpha}, \\
-\alpha & \text{if } 2\alpha \leq V_{n+1}, \text{ with probability } e^{-2\alpha},
\end{cases}$$

and note that the corresponding reflecting simple random walk $W$ on

$$\{n\alpha + \log k : n = 0, 1, 2, \ldots\}$$

dominates $Z$ (use coupling!), is positive-recurrent if $e^{-2\alpha} + e^{-\alpha} > 1$ (in which case $e^{\alpha}(e^{\alpha} - 1) < 1$), and then has equilibrium distribution which is Geometric:

$$\mathbb{P}[W = n\alpha + \log k] = (e^{\alpha}(e^{\alpha} - 1))^n (1 - e^{\alpha}(e^{\alpha} - 1)).$$

(3.6)
So we can draw from this Geometric distribution, then simulate the (reversible!) reflecting simple random walk $W$ backwards in time, and use $\exp(W)$ as dominating process. Calculations show that a choice of $k = 2.76796 \ldots$ is optimal, in the sense of minimizing the mean equilibrium value of $W$. There is still one problem left. If we apply the same value of $U_i$ to all possible starts lying below the dominating process, then we will not achieve coalescence. (We will get convergence at a geometric rate, which will eventually be absorbed by rounding error, but that is not the same thing!) It is necessary to use Wilson’s multi-shift trick.

A perfect perpetuity.
Note that the *approximate* algorithm, “run recurrence till initial conditions lost in floating point error” is *slower*!

See Devroye (2001) for a different (but still perfect!) approach.
3.5. Point processes

Dominated \textit{CFTP} works, for example, on both attractive and repulsive area-interaction point processes (Kendall 1998; Kendall and Møller 2000): using as target chain a \textit{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 \textit{domCFTP} in a manner very similar to that of Section 3.3, but with Uniform\([0, 1]\) marks replaced by marks which are Poisson clusters, as described in (Kendall 1997a).

Dominated \textit{CFTP} for attractive area-interaction point process with geometric marking using Poisson processes in disks.

It is of interest in stochastic geometry that this expresses such point processes as explicit but highly \textit{dependent thinnings} of Poisson point processes.
See also Huber (1999)'s notion of a “swap move”, 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 $p_{\text{swap}}$ which we are free to choose; this results in a provable speed-up of the CFTP algorithm.
3.6. A general theorem

Moving from CFTP to domCFTP suggests still further abstraction. This is convenient, for example, when considering CFTP for conditioned point processes as in Cai and Kendall (2002): it can be 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 exhibits equilibrium behaviour. Embed the state space $\mathcal{X}$ in a partially ordered space $(\mathcal{Y}, \preceq)$ so that $\mathcal{X}$ is at bottom of $\mathcal{Y}$, in the sense that for any $y \in \mathcal{Y}$, $x \in \mathcal{X}$,

$$y \preceq x \quad \text{implies} \quad y = x.$$
We may then use the methods of Theorem 1.2 (CFTP) and Theorem 3.10 (domCFTP) to show:

**Theorem 3.11** 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 $\mathcal{L}(Y(-T, -T)) = \mathcal{L}(Y(0, 0))$, and

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

Suppose coalescence occurs: $\mathbb{P}[Y(-T, 0) \in \mathcal{X}] \to 1$ as $T \to \infty$. Then $\lim Y(-T, 0)$ can be used for a CFTP draw from the equilibrium of $X$. 
Exercise 3.12 Murdoch (2000) 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 can be expressed in terms of the above general framework.
Lecture 4

Theory and connections

4.1 FMMR
4.2 Randomness Recycler
4.3 Efficiency
4.4 Foster-Lyapunov
4.5 BFA algorithm
4.1. Siegmund duality and FMMR

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?

$$
P [X_t \geq y | X_0 = x] = P [Y_t \leq x | Y_0 = y] \quad (4.1)
$$

Answer (*Siegmund 1976*): exactly when $X$ is (suitably regular and) stochastically monotone: $x \leq x'$ implies

$$
P [X_t \geq y | X_0 = x] \leq P [X_t \geq y | X_0 = x'] .
$$

**Proof (Outline):**

Use Equation (4.1) and Fubini’s Theorem to derive the Chapman-Kolmogorov equations.
Remark 4.2 If $X$ is not stochastically monotone then we get negative transition probabilities for $Y$!

Remark 4.3 Consequence:  
$Y$ is absorbed at 0, and $X$ at $\infty$.

Remark 4.4 Intuition: think of the Siegmund dual this way. Couple monotonically the $X^{(x)}$ begun at different $x$ (use stochastic monotonicity!), set $Y_t^{(y)} = x$ if $X_0^{(x)} = y$. 
This beautiful idea grew into a method of simulation, and then a method of perfect simulation called Fill’s method (Fill 1998) alternative to CFTP. Fill’s method considered on its own is harder to explain than CFTP.

On the other hand it has advantages too:

- it can provide user-interruptibility (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.
Fill’s method (Fill 1998; Thönnes 1999) is at first sight quite different from CFTP. It is based on the notion of a *strong uniform time* $T$ (Diaconis and Fill 1990) and is related to Siegmund duality. Fill et al. (2000) establish a profound link. We explain using “blocks” as input-output maps for a chain. 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 $X(-T)$. 
**Key observation:** By construction, $X(0)$ and $T$ are independent of $X(-T)$ so we can condition on them!

- 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.

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 Dobrow and Fill 2003.)
**Key observation:** By construction, \(X(0)\) and \(T\) are independent of \(X(-T)\) so we can condition on them!

- 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.

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 Dobrow and Fill 2003.)
Question 4.5 Is there a dominated version of Fill’s method?
4.2. Randomness Recycler

Fill and Huber (2000) introduce a quite different form of perfect simulation! Here is how they apply their Randomness Recycler algorithm to the problem of drawing a random independent subset $X$ of a graph $G$, weighted exponentially by number of points in $X$.

**Start:** $V = \emptyset$, $x \equiv 0$.  
**End:** $V = G$, $x$ indicates $X$ membership.

```
while $V \neq G$:
    $v \leftarrow choice (G - V)$  # Choose $v$ from $G \setminus V$
    $V.add (v)$
    if uniform $(0, 1) \leq 1/(1 + \alpha)$:
        $x[v] \leftarrow 0$  # Skip $v$ with prob $1/(1 + \alpha)$
    else:
        $x[v] \leftarrow 1$  # or tentatively include it ...
        $nbd \leftarrow []$  # ... iterate thro’ neighbours
        for $w \in neighbourhood (v)$:  # Valid?
            $nbd.append (w)$
            if $x[w] = 1$:  # If not valid ...
                $x[w] \leftarrow 0$  # ... remove all
                $x[v] \leftarrow 0$  # “contaminated” vertices
                $V \leftarrow V - [v] - nbd$
            break  # and move on
```
4.3. 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 Propp and Wilson (1996).

**Remark 4.6** In general CFTP has to involve coupling. One expects coupling to happen at some exponential rate, and convergence to equilibrium (in total variation norm $\text{dist}_{TV}$) likewise. From the Coupling Inequality (1.1) we know that coupling cannot happen faster than convergence to equilibrium.
In the case of monotonic CFTP on a finite partially ordered space, Propp and Wilson (1996) present a strong bound. Let $\ell$ be the longest chain in the space; let $T^*$ be the coalescence time, let

$$\bar{d}(k) = \max_{x,y} \{ P^{(k)}_x - P^{(k)}_y \}.$$ 

Then

$$\frac{\mathbb{P}[T^* > k]}{\ell} \leq \bar{d}(k) \leq \mathbb{P}[T^* > k],$$

(4.2)

so CFTP is within a factor of being as good as possible.

But can it happen slower? and for relatively simple Markov chains? Burdzy and Kendall (2000) show how to use coupling to find out about this coupling problem!
Coupling of couplings:...
Suppose $|p_t(x_1, y) - p_t(x_2, y)| \approx c_2 \exp(-\mu_2 t)$
while $\mathbb{P} [\tau > t | X(0) = (x_1, x_2)] \approx c \exp(-\mu t)$
Coupling of couplings:...
Suppose $|p_t(x_1, y) - p_t(x_2, y)| \approx c_2 \exp(-\mu_2 t)$
while $\mathbb{P}[\tau > t | X(0) = (x_1, x_2)] \approx c \exp(-\mu t)$

\[
|p_t(x_1, y) - p_t(x_2, y)| \leq \\
\leq |\mathbb{P}[X_1(t) = y | X_1(0) = x_1] - \mathbb{P}[X_2(t) = y | X_2(0) = x_2]| \leq \\
|\mathbb{P}[X_1(t) = y | \tau > t, X_1(0) = x_1] - \mathbb{P}[X_2(t) = y | \tau > t, X_2(0) = x_2]| \\
\times \mathbb{P}[\tau > t | X(0) = (x_1, x_2)]
\]
Coupling of couplings:...
Suppose $|p_t(x_1, y) - p_t(x_2, y)| \approx c_2 \exp(-\mu_2 t)$
while $\mathbb{P}[\tau > t|X(0) = (x_1, x_2)] \approx c \exp(-\mu t)$

$$|p_t(x_1, y) - p_t(x_2, y)| \leq \mathbb{P}[X_1(t) = y|X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|X_2(0) = x_2] \leq |\mathbb{P}[X_1(t) = y|\tau > t, X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|\tau > t, X_2(0) = x_2]| \times \mathbb{P}[\tau > t|X(0) = (x_1, x_2)]$$

Let $X^*$ be a coupled copy of $X$ but begun at $(x_2, x_1)$:

$$|\mathbb{P}[X_1(t) = y|\tau > t, X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|\tau > t, X_2(0) = x_2]| = |\mathbb{P}[X_1(t) = y|\tau > t, X(0) = (x_1, x_2)] - \mathbb{P}[X_1^*(t) = y|\tau > t, X^*(0) = (x_2, x_1)]|$$
Coupling of couplings:...

Suppose $|p_t(x_1, y) - p_t(x_2, y)| \approx c_2 \exp(-\mu_2 t)$
while $\mathbb{P}[\tau > t|X(0) = (x_1, x_2)] \approx c \exp(-\mu t)$

\[
|p_t(x_1, y) - p_t(x_2, y)| \leq |\mathbb{P}[X_1(t) = y|X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|X_2(0) = x_2]| \leq \\
|\mathbb{P}[X_1(t) = y|\tau > t, X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|\tau > t, X_2(0) = x_2]| \\
\times \mathbb{P}[\tau > t|X(0) = (x_1, x_2)]
\]

Let $X^*$ be a coupled copy of $X$ but begun at $(x_2, x_1)$:

\[
|\mathbb{P}[X_1(t) = y|\tau > t, X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|\tau > t, X_2(0) = x_2]| = |\mathbb{P}[X_1(t) = y|\tau > t, X(0) = (x_1, x_2)] - \\
\mathbb{P}[X^*_1(t) = y|\tau > t, X^*_1(0) = (x_2, x_1)]| \leq \mathbb{P}[\sigma > t|\tau > t, X(0) = (x_1, x_2)] \\
(\approx c' \exp(-\mu' t))
\]

for $\sigma$ the time when $X, X^*$ couple.
Thus $\mu_2 \geq \mu' + \mu$. 
Remark 4.7 So coupling is slower than convergence to equilibrium when coupled chains can swap round, the opposite of monotonicity! Burdzy and Kendall (2000) present heuristics to expand on this.

Here is a simple continuous-time Markov chain case.

Mountford and Cranston (2000) describe a still simpler example!
Remark 4.8 Kumar and Ramesh (2001) give a computer-science type example of graph-matchings: coupling becomes much slower than convergence to equilibrium as problem-size increases.

Remark 4.9 A different question about efficiency is, how best to use CFTP to obtain repeated samples? see Murdoch and Rosenthal (1999).
4.4. Dominated *CFTP* and Foster-Lyapunov conditions

Corcoran and Tweedie (2001) describe how to mix *domCFTP* and small-set *CFTP*. The upper envelope process must be formulated carefully . . . . When the dominating process visits a small set, then one can attempt small-set coupling. However the dominating process must be large enough to dominate instances 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 $V$ to deliver a controlled supermartingale off a small set.

We begin by discussing a Foster-Lyapunov condition for positive-recurrence.
Theorem 4.10 Meyn and Tweedie (1993) 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 \( V \) bounded on \( C \) such that for all \( n > 0 \)

\[
\mathbb{E}[V(X_{n+1})|X_n] \leq V(X_n) - 1 + \beta \mathbb{I}[X_n \in C]. \tag{4.3}
\]

Proof: Let \( N \) be the random time at which \( X \) first (re-)visits \( C \). It suffices\(^1\) to show \( \mathbb{E}[N|X_0] < V(X_0) + \text{constant} < \infty \) (then use small-set regeneration).

By iteration of (4.3), we deduce \( \mathbb{E}[V(X_n)|X_0] < \infty \) for all \( n \).

If \( X_0 \not\in C \) then (4.3) tells us \( n \mapsto V(X_{n\wedge N}) + n \wedge N \) defines a nonnegative supermartingale \( (\mathbb{I}[X_{(n\wedge N)} \in C] = 0 \text{ if } n < N) \). Consequently

\[
\mathbb{E}[N|X_0] \leq \mathbb{E}[V(X_N) + N|X_0] \leq V(X_0).
\]

\(^1\)This martingale approach can be reformulated as an application of Dynkin’s formula.
If $X_0 \in C$ then the above can be used to show

\[
\mathbb{E}[N|X_0] = \mathbb{E}[1 \times \mathbb{I}[X_1 \in C]|X_0] + \mathbb{E}[\mathbb{E}[N|X_1] \mathbb{I}[X_1 \notin C]|X_0] \\
\leq \mathbb{P}[X_1 \in C|X_0] + \mathbb{E}[1 + V(X_1)|X_0] \\
\leq \mathbb{P}[X_1 \in C|X_0] + V(X_0) + \beta
\]

where the last step uses Inequality (4.3) applied when

\[
\mathbb{I}[X_{n-1} \in C] = 1.
\]

\[\square\]

Now we consider a strengthened condition for geometric ergodicity.
Theorem 4.11  *Meyn and Tweedie (1993)* Geometric ergodicity holds if one can find a small set $C$, positive constants $\lambda < 1$, $\beta$, and a function $V \geq 1$ bounded on $C$ such that

$$E[V(X_{n+1})|X_n] \leq \lambda V(X_n) + \beta \mathbb{I}[X_n \in C]. \quad (4.4)$$

**Proof:** Define $N$ as in Theorem 4.10. Iterating (4.4), we deduce $E[V(X_n)|X_0] < \infty$ and more specifically

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

is a nonnegative supermartingale. Consequently

$$E\left[V(X_N)/\lambda^N|X_0\right] \leq V(X_0).$$

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

$$\mathbb{P}[N > n|X_0] \leq \lambda^n V(X_0),$$

which delivers the required geometric ergodicity.  \Box
Temptation: define dominating process using $V$. There is an interesting link – Rosenthal (2002) draws it even closer – but:

**Existence of Lyapunov function doesn’t ensure domCFTP**

The expectation inequality of supermartingale-type,

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

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

$$
\mathcal{D}_x = \{\mathcal{L}(V(X_{n+1}|X_n): X_n = u, V(u) \leq V(x)\},
$$

and it is easy to construct badly behaved examples. :-(
Exercise 4.12  Construct a Markov chain on \([0, \infty)\) which satisfies the conditions of Theorem 4.11, using \(V(x) = x\) and \(C = \{0\}\), but such that any \(V\)-dominating process for \(X\) (a process \(U\) for which \(V(U_{n+1}) \geq V(X_{n+1})\) whenever \(V(U_n) \geq V(X_n)\)) must be transient!
4.5. Backward-forward algorithm

A careful look at domCFTP for the area-interaction process, or generalizations to other point processes as described in Kendall and Møller (2000), shows that the construction is as follows:

- build space-time Poisson process of free points;
- convert free points into initial points for time-like line segments, hence space-time birth and death process;
- mark free points independently;
- apply causal thinning procedure in time order;
- domCFTP succeeds if the time-zero result of this thinning procedure stabilizes when thinning begins far enough back in time; apply binary search procedure to capture a time early enough to ensure stabilization at time zero.
Ferrari et al. (2002) describe a variant of perfect simulation (BFA, in full Backwards-Forwards Algorithm) which avoids the need to iterate back through successive starts \(-T, -2T, -4T\).

- Conduct recursive backwards sweep, identifying 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, thinning out ancestors to obtain required perfect sample at time zero (assuming backwards sweep generates only finitely many ancestors).

Instead of coalescence, we now require sub-criticality of the oriented percolation implicit in the backwards sweep; computable conditions for this can be obtained using 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).

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 *Fernández et al. (2001)*.  

**Exercise 4.13**  Use **BFA** to implement a perfect simulation of the Peierls contour model for low temperature Ising models.

---

[Ferrari et al. (2002)] point out, especially for infinite volume processes there is a “user-impatience” bias for which they estimate the effects.
References


This is a rich hypertext bibliography. Journals are linked to their homepages, and stable URL links (as provided for example by JSTOR or Project Euclid) have been added where known. Access to such URLs is not universal: in case of difficulty you should check whether you are registered (directly or indirectly) with the relevant provider. In the case of preprints, icons linking to homepage locations are inserted where available: note that these are probably less stable than journal links!


Cai, Y. and W. S. Kendall [2002, July]. Perfect simulation for correlated Poisson ran-


doi:


Häggström, O. and K. Nelander [1998]. Exact sampling from anti-monotone sys-
tems. *Statistica Neerlandica* 52(3), 360–380. doi


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


Wilson, D. B. and J. G. Propp [1996]. 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), New York, pp. 448–457. ACM.