Elect. Comm. in Probab. **9** (2004), 140–151

ELECTRONIC COMMUNICATIONS in PROBABILITY

GEOMETRIC ERGODICITY AND PERFECT SIMULATION

WILFRID S. KENDALL
Department of Statistics, University of Warwick, Coventgry CV4 7AL, UK
email: `w.s.kendall@warwick.ac.uk`

Submitted July 2, 2004, *accepted in final form* October 10, 2004

AMS 2000 Subject classification: 60J10, 65C05, 68U20
Keywords: CFTP, domCFTP, geometric Foster-Lyapunov condition, geometric ergodicity, Markov chain Monte Carlo, perfect simulation, uniform ergodicity

Abstract
This note extends the work of Foss and Tweedie (1998), who showed that availability of the classic Coupling from the Past (*CFTP*) algorithm of Propp and Wilson (1996) is essentially equivalent to uniform ergodicity for a Markov chain (see also Hobert and Robert 2004). In this note we show that all geometrically ergodic chains possess dominated *CFTP* algorithms (not necessarily practical!) which are rather closely connected to Foster-Lyapunov criteria. Hence geometric ergodicity implies dominated *CFTP*.

1 Introduction

Throughout this paper X will denote an aperiodic Harris-recurrent Markov chain on a measurable state space $\mathcal{X}$ which is a Polish space (the Polish condition is required in order to ensure existence of regular conditional probabilities). Let π denote the equilibrium probability distribution of X. Recall that X is said to be *geometrically ergodic* if it converges in total variation and at geometric rate to statistical equilibrium π, with multiplicative constant depending on the starting point:

$$\operatorname{dist}_{\mathrm{TV}}(\mathcal{L}(X_n), \pi) \quad \leq \quad V(X_0)\gamma^n \tag{1}$$

for some function $V : \mathcal{X} \to [1, \infty)$ and some rate $\gamma \in (0, 1)$. The chain X is said to be *uniformly ergodic* if the function V can be chosen to be constant.
We also recall the notion of a small set:

Definition 1 *A subset $C \subseteq \mathcal{X}$ is a* small set (of order k) *for the Markov chain X if there is a* minorization condition*: for $\beta \in (0, 1)$, and probability measure ν,*

$$\mathbb{P}[X_k \in E \mid X_0 = x] \quad \geq \quad \beta\, \mathbb{I}[x \in C] \times \nu(E) \quad \textit{for all measurable } E \subseteq \mathcal{X}\,. \tag{2}$$

Results are often stated in terms of the more general notion of *petite sets*; however for ψ-irreducible aperiodic chains the two notions are equivalent (Meyn and Tweedie 1993, Theorem 5.5.7).

Foss and Tweedie (1998) use small set theory to show that the condition of uniform ergodicity for such X is *equivalent* to the existence of a Coupling from the Past algorithm (based on X) in the sense of Propp and Wilson (1996). This *classic CFTP* algorithm delivers a perfect sample from the equilibrium distribution of X. The key to the Foss and Tweedie argument is to remark that in case of uniform ergodicity the entire state space is small. Sub-sampling the process X if necessary (to reduce the order of the small set to 1), one can then devise a classic *CFTP* algorithm which is actually of the form introduced by Murdoch and Green (1998) as the *multigamma coupler*. Hobert and Robert (2004) develop the Foss and Tweedie argument to produce approximations to deal with *burn-in* (time till approximate equilibrium) in the geometrically ergodic case.

The Foss and Tweedie result might be thought to delimit and constrain the possible range of applicability of *CFTP*. However it is also possible to sample perfectly from the equilibrium of some strictly geometrically ergodic chains using a generalization: namely *dominated* CFTP (*domCFTP*) as introduced in Kendall (1998), Kendall and Møller (2000), Cai and Kendall (2002). In this note we show that this is generic: geometric ergodicity implies the existence of a special form of *domCFTP* algorithm adapted to the geometric ergodicity in question. Recent expositions of quantitative convergence rate estimation depend heavily on small sets and their relatives (see for example Rosenthal 2002), so this piece of *CFTP* theory connects to quantitative convergence theory in a rather satisfying way.

To describe this special form of *domCFTP*, we must first introduce the notion of a Foster-Lyapunov condition. Geometric ergodicity for our X is equivalent to a *geometric Foster-Lyapunov condition* involving recurrence on small sets (this can be extracted from Meyn and Tweedie 1993, Theorem 16.0.1):

$$\mathbb{E}\left[\Lambda(X_{n+1}) \mid X_n = x\right] \quad \leq \quad \alpha\Lambda(x) + b\,\mathbb{I}\left[X_n \in C\right] , \tag{3}$$

for some $\alpha \in (0,1)$ and $b > 0$, some small set C, and a drift function $\Lambda : \mathcal{X} \to [1,\infty)$ which is bounded on C. Note that $\alpha + b \geq 1$ is required, as is $\Lambda|_{C^c} \geq \alpha^{-1}$, since we impose $\Lambda \geq 1$.

Now the moment condition (3) implies that every sub-level set $\{x \in \mathcal{X} : \Lambda(x) \leq c\}$ is small (as indeed do weaker conditions; Meyn and Tweedie 1993, Theorem 14.2.3).

It is convenient to isolate the notion of a *pseudo-drift function* as Λ in Equation (3).

Definition 2 *A* (Foster-Lyapunov) pseudo-drift function *for a Markov chain state space $\mathcal{X}$ is a measurable function*

$$\Lambda : \mathcal{X} \to [1,\infty)$$

such that sub-level sets $\{x \in \mathcal{X} : \Lambda(x) \leq \lambda\}$ are small for all $\lambda \geq 1$.

Thus a pseudo-drift function has the properties of a Foster-Lyapunov drift function but is not endowed with a specific moment condition.

Now we can define the special form of *domCFTP* which we require, which is adapted to a specified pseudo-drift function.

Definition 3 *Suppose that Λ is a pseudo-drift function for an Harris-recurrent Markov chain X. We say the stationary ergodic random process Y on $[1,\infty)$ is a* dominating process for X based on the pseudo-drift function Λ *(with* threshold h *and* coalescence probability ε*) if it is coupled co-adaptively to realizations of $X^{x,-t}$ (the Markov chain X begun at x at time $-t$) as follows:*

(a) for all $x \in \mathcal{X}$, $n > 0$, and $-t \leq 0$, almost surely

$$\Lambda(X^{x,-t}_{-t+n}) \quad \leq \quad Y_{-t+n} \qquad \Rightarrow \qquad \Lambda(X^{x,-t}_{-t+n+1}) \quad \leq \quad Y_{-t+n+1}\,; \tag{4}$$

(b) *moreover if* $Y_m \leq h$ *for some fixed* m *then the probability of* coalescence *(conditional on past of* Y *and past potential coalescence events) is at least* ε*, where coalescence means that the set*

$$\left\{X_{m+1}^{x,-t} \;:\; \textit{such that } -t \leq m \textit{ and } \Lambda(X_m^{x,-t}) \leq Y_m\right\}$$

is a singleton set (inequality is used in $-t \leq m$ *rather than equality as this is a condition on coupling of* $X^{x,-t}$ *for all* $-t \leq m$*);*

(c) *and finally,* $\mathbb{P}[Y_n \leq h]$ *must be positive.*

Suppose Y is a dominating process for X based on the pseudo-drift function Λ. The following *domCFTP* algorithm then yields a draw from the equilibrium distribution of X.

Algorithm 4

Simulate Y *backwards in equilibrium till the most recent* $T < 0$ *for which* $Y_T \leq h$*;*

while coalescence does not occur at time T*:*

extend Y *backwards till the most recent* $S < T$ *for which* $Y_S \leq h$*;*

set $T \leftarrow S$*;*

simulate the coupled X *forwards from time* $T+1$*, starting with the unique state produced by the coalescence event at time* T *(conditioned on* Y*);*

return X_0 *as a perfect draw from equilibrium.*

This algorithm terminates almost surely as a consequence of the conditions imposed in Definition 3.
Practical implementation considerations are: (1) can one draw from the equilibrium of Y? (2) can one simulate Y backwards in equilibrium? (3) can one couple the dominated target processes $X^{x,-t}$ with Y so as to ensure the possibility of regeneration? (4) can one determine when this regeneration has occurred? and, of course, (5) will the algorithm not run too slowly?
The simplest kind of ordinary small-set *CFTP*, as in Murdoch and Green (1998), is recovered from this Algorithm by taking $Y \equiv h$, and requiring the whole state-space to be contained in $\{x : \Lambda(x) \leq h\}$ and hence small. In actual constructions, care must be taken to ensure that Y dominates a coupled collection of X for which coalescence is possible as specified in Definition 3(b) (see the treatment of *CFTP* for Harris chains in Corcoran and Tweedie 2001).
The proof that this algorithm returns a perfect draw from the equilibrium distribution of X is an easy variation on the usual *domCFTP* argument, found at varying levels of generality in Kendall 1998; Kendall and Møller 2000; Cai and Kendall 2002. The key is to observe that Algorithm 4 reconstructs a coalesced trajectory which may be viewed as produced by the Markov chain begun at time $-\infty$ at some specified state x such that $\Lambda(x) \leq h$: the proof is then an exercise in making this heuristic precise.
The Foss and Tweedie (1998) argument, and the fact that the geometric Foster-Lyapunov condition (3) would certainly produce a dominating process if the expectation inequality was replaced by a stochastic domination, together suggest our main result, to be proved in Section 2:

Theorem 5 *If* X *is a geometrically ergodic Markov chain, and* Λ *is a pseudo-drift function for* X *which is derived from some geometric Foster-Lyapunov condition, then there is an almost surely terminating* domCFTP *algorithm for* X *(possibly subject to sub-sampling) using* *a dominating process based on the pseudo-drift function Λ, as in Algorithm 4 and Definition 3.*

As in the case of the Foss and Tweedie (1998) result, this algorithm need not be at all practical!

2 Proof of Theorem 5

We begin with a lemma concerning the effect of sub-sampling on the geometric Foster-Lyapunov condition.

Lemma 6 *Suppose X satisfies a geometric Foster-Lyapunov condition: for some $\alpha < 1$, some pseudo-drift function Λ, and small set $C = \{x \in \mathcal{X} : \Lambda(x) \leq c\}$.*

$$\mathbb{E}\left[\Lambda(X_{n+1}) \mid X_n = x\right] \quad \leq \quad \alpha\Lambda(x) + b\,\mathbb{I}\left[\Lambda(X_n) \leq c\right] . \tag{5}$$

Under k-sub-sampling we obtain a similar condition but with different constants:

$$\mathbb{E}\left[\Lambda(X_{n+k}) \mid X_n = x\right] \quad \leq \quad \alpha^{k-1}\Lambda(x) + b'\,\mathbb{I}\left[\Lambda(X_n) \leq c'\right] , \tag{6}$$

and also, if $k \geq 2$,

$$\mathbb{E}\left[\Lambda(X_{n+k}) \mid X_n = x\right] \quad \leq \quad \alpha\Lambda(x) + b''\,\mathbb{I}\left[\Lambda(X_n) \leq c''\right] . \tag{7}$$

Moreover $b' = b/(1-\alpha)$, $c\prime = b/(\alpha^{k-1}(1-\alpha)^2)$ may be chosen not to depend on c, and $b'' = b/(1-\alpha)$, $c'' = b/(\alpha(1-\alpha)^2)$ may be chosen to depend neither on c nor on $k \geq 2$.

We are able to choose b', c', b'', c'' not to depend on c because we have allowed generous sub-sampling (*i.e.*: k-sub-sampling to change α to α^{k-1}).

Proof: Iterating Equation (5),

$$\begin{aligned}
\mathbb{E}\left[\Lambda(X_{n+k}) \mid X_n = x\right] \quad &\leq \quad \alpha^k\Lambda(x) + \sum_{j=1}^{k} \alpha^{j-1} b\,\mathbb{E}\left[\mathbb{I}\left[\Lambda(X_{n+k-j}) \leq c\right] \mid X_n = x\right] \\
&\leq \quad \alpha^k\Lambda(x) + \frac{b}{1-\alpha} \\
&= \quad \alpha^{k-1}\Lambda(x) - \alpha^{k-1}(1-\alpha)\Lambda(x) + \frac{b}{1-\alpha} \\
&\leq \quad \begin{cases} \alpha^{k-1}\Lambda(x) & \text{if } \Lambda(x) > \frac{b}{\alpha^{k-1}(1-\alpha)^2} , \\ \alpha^{k-1}\Lambda(x) + b/(1-\alpha) & \text{otherwise.} \end{cases}
\end{aligned}$$

Hence we may choose $b' = b/(1-\alpha)$, $c' = b/(\alpha^{k-1}(1-\alpha)^2)$. Alternatively

$$\begin{aligned}
\mathbb{E}\left[\Lambda(X_{n+k}) \mid X_n = x\right] \quad &\leq \quad \alpha\Lambda(x) - \alpha(1-\alpha^{k-1})\Lambda(x) + \frac{b}{1-\alpha} \\
&\leq \quad \begin{cases} \alpha\Lambda(x) & \text{if } \Lambda(x) > \frac{b}{\alpha(1-\alpha)(1-\alpha^{k-1})} , \\ \alpha\Lambda(x) + b/(1-\alpha) & \text{otherwise.} \end{cases}
\end{aligned}$$

Hence we may choose $b'' = b/(1-\alpha)$, $c'' = b/(\alpha(1-\alpha)^2)$ if $k \geq 2$. □

Proof (of Theorem 5):
We first construct the dominating process.
Consider Markov's inequality applied to the geometric Foster-Lyapunov inequality (3). Any dominating process Y must satisfy the stochastic domination (4) described in Definition 3. Consequently, in default of further distributional information about $\mathbb{P}[\Lambda(X_{n+1})|X_n = x]$, if Y is to be a dominating process based on the pseudo-drift function Λ then we need Y to be stationary ergodic but also to satisfy

$$\mathbb{P}[Y_{n+1} \geq \alpha z y \mid Y_n = z] \quad \geq \quad \sup_{x:\Lambda(x)\leq z} \frac{\mathbb{E}[\Lambda(X_{n+1}) \mid X_n = x]}{\alpha z y}\,. \tag{8}$$

Now if $C \subseteq \{x \in \mathcal{X} : \Lambda(x) \leq c\}$ then

$$\begin{aligned} \sup_{x:\Lambda(x)\leq z} \frac{\mathbb{E}[\Lambda(X_{n+1}) \mid X_n = x]}{\alpha z y} \quad &\leq \quad \sup_{x:\Lambda(x)\leq z} \frac{\alpha\Lambda(x) + b\,\mathbb{I}[x : \Lambda(x) \leq c]}{\alpha z y} \\ \leq \quad \sup_{x:\Lambda(x)\leq z} \frac{\alpha\Lambda(x)}{\alpha z y} \quad &\leq \quad \frac{1}{y} \qquad \text{so long as } z \geq c + \frac{b}{\alpha}\,. \end{aligned}$$

Consequently Y is a possible candidate for a dominating process based on the pseudo-drift function Λ if

$$\mathbb{P}[Y_{n+1} \geq \alpha z y \mid Y_n = z] \quad = \quad \begin{cases} 1/y & \text{if } z \geq c + \frac{b}{\alpha}\,, \\ 1 & \text{otherwise.} \end{cases} \tag{9}$$

If we define U by $Y = (c + b/\alpha)\exp(U)$ (so U is a *log-dominating process*) then U is the system workload of a $D/M/1$ queue, sampled at arrivals, with arrivals every $\log(1/\alpha)$ units of time, and service times being independent and of unit Exponential distribution. The process U is a random walk with reflection (of Skorokhod type) at 0: as its jump distribution is Exponential(1) $- \log(1/\alpha)$ we may deduce it is positive-recurrent if and only if $\alpha < e^{-1}$.
In case $e^{-1} < \alpha < 1$, U and $Y = (c + b/\alpha)\exp(U)$ fail to be positive-recurrent. However the same construction will work if we use Equation (6) of Lemma 6 to justify sub-sampling X with a sampling period k large enough to ensure a geometric Foster-Lyapunov condition (3) using Λ as pseudo-drift but with α replaced by $\alpha^{k-1} < e^{-1}$, and amending b to b', c to c' as in Inequality (6).
Thus without loss of generality we may assume $\alpha < e^{-1}$, and so this Y can be run in statistical equilibrium, and thus qualifies as least partly as a dominating process for the purposes of Theorem 5. In the sequel we assume moreover that further sub-sampling has been carried out based on Equation (7), to ensure that the following small set is of order 1:

$$\{x \in \mathcal{X} \,:\, \Lambda(x) \leq h\} \qquad \text{for} \qquad h = \max\left\{c + \frac{b}{\alpha}, \frac{b}{\alpha(1-\alpha)}\left(1 + \frac{1}{1-\alpha}\right)\right\}\,. \tag{10}$$

Here the level $h \geq c + b/\alpha$ is fixed so as to ensure $h = c'' + b''/(1-\alpha)$ with b'', c'' given as in Equation (7); thus h supplies a stable threshold for geometric Foster-Lyapunov conditions, even allowing for further sub-sampling if required. Note in particular that $Y = (c+b/\alpha)\exp(U)$ is able to sink below h, since $h \geq c + b/\alpha$ and the system workload U can reach zero.
To fulfil the requirements on a dominating process given in Definition 3, we need to construct a coupling between Y and the target process X expressed in terms of a random flow of independent maps $F_{-t+n+1} : \mathcal{X} \to \mathcal{X}$:

$$X^{x,-t}_{-t+n+1} \quad = \quad F_{-t+n+1}(X^{x,-t}_{-t+n})$$

satisfying the distributional requirement that $X^{x,-t}$ should evolve as the Markov chain X, the domination requirement expressed by the implication (4), and also the regeneration requirement that with probability ε the set

$$\{F_n(u) \ : \ \text{such that } \Lambda(u) \leq h\}$$

should be a singleton set. The well-known link between stochastic domination and coupling can be applied together with the arguments preceding Equation (9) to show that we can couple the various $X^{x,-t}$ with Y co-adaptively in this manner so that the implication (4) holds: note that here and here alone we use the Polish space nature of $\mathcal{X}$, which allows us to construct the couplings by use of regular conditional probability distributions for the various $X^{x,-t}$ conditioned on the $\Lambda(X^{x,-t})$. Thus all that is required is to show that this stochastic domination coupling can be modified to allow for regeneration.

The small set condition for $\{x \in \mathcal{X} : \Lambda(x) \leq h\}$ means there is a probability measure ν and a scalar $\beta \in (0,1)$ such that for all Borel sets $B \subseteq [1,\infty)$, whenever $\Lambda(x) \leq h$,

$$\mathbb{P}\left[\Lambda(X_{n+1}) \in B \mid X_n = x\right] \quad \geq \quad \beta\nu(B)\,. \tag{11}$$

Moreover the stochastic domination which has been arranged in the course of defining Y means that for all real u, whenever $\Lambda(x) \leq y$,

$$\mathbb{P}\left[\Lambda(X_{n+1}) > u \mid X_n = x\right] \quad \leq \quad \mathbb{P}\left[Y > u \mid Y = y\right]\,. \tag{12}$$

We can couple in order to arrange for regeneration if we can identify a probability measure $\widetilde{\nu}$, defined solely in terms of ν and the dominating jump distribution $\mathbb{P}[Y \geq u \mid Y = y]$, such that for all real u

$$\begin{aligned} \mathbb{P}\left[\Lambda(X_{n+1}) > u \mid X_n = x\right] - \beta\nu((u,\infty)) \quad &\leq \quad \mathbb{P}\left[Y > u \mid Y = y\right] - \beta\widetilde{\nu}((u,\infty)) \\ \nu((u,\infty)) \quad &\leq \quad \widetilde{\nu}((u,\infty)) \end{aligned}$$

and moreover

$$\mathbb{P}\left[Y_{n+1} \in B \mid Y_n = y\right] \quad \geq \quad \beta\widetilde{\nu}(B)\,.$$

For then at each step we may determine whether or not regeneration has occurred (with probability β); under regeneration we use stochastic domination to couple ν to $\widetilde{\nu}$; otherwise we use stochastic domination to couple the residuals.

Results to this effect may be gleaned from Roberts and Rosenthal (2001): for the sake of explicit exposition we state and prove an interior lemma.

Lemma 7 *Suppose U, V are two random variables defined on $[1,\infty)$ such that*

(a) The distribution $\mathcal{L}(U)$ is stochastically dominated by the distribution $\mathcal{L}(V)$:

$$\mathbb{P}[U > u] \quad \leq \quad \mathbb{P}[V > u] \qquad \textit{for all real } U\,; \tag{13}$$

(b) U satisfies a minorization condition: for some $\beta \in (0,1)$ and probability measure ν: $B \subseteq [1,\infty)$,

$$\mathbb{P}[U \in B] \quad \geq \quad \beta\nu(B) \qquad \textit{for all Borel sets } B \subseteq [1,\infty)\,. \tag{14}$$

Then there is a probability measure μ stochastically dominating ν and such that $\beta\mu$ is minorized by $\mathcal{L}(V)$. Moreover μ depends only on $\beta\nu$ and $\mathcal{L}(V)$.

Proof (of Lemma 7):
Subtract the measure $\beta\nu((u,\infty))$ from both sides of Inequality (13) representing the stochastic domination $\mathcal{L}(U) \preceq \mathcal{L}(V)$. By the minorization condition (14) the resulting left-hand-side is nonnegtive. Thus for all real u

$$0 \quad \leq \quad \mathbb{P}[U > u] - \beta\nu((u,\infty)) \quad \leq \quad \mathbb{P}[V > u] - \beta\nu((u,\infty))$$

Now $\mathcal{L}(U) - \beta\nu$ is a nonnegative measure (because of the minorization condition (14)). Consequently $\mathbb{P}[U > u] - \beta\nu((u,\infty))$ must be non-increasing in u and so we may reduce the right-hand side by minimizing over $w \leq u$:

$$\begin{aligned}\mathbb{P}[U > u] - \beta\nu((u,\infty)) \quad &\leq \quad \inf_{w\leq u}\{\mathbb{P}[V > w] - \beta\nu((w,\infty))\} \\ &= \quad \mathbb{P}[V > u] - \beta\mu((u,\infty))\end{aligned}$$

where μ is the potentially *signed* measure defined by

$$\beta\mu([1,u]) \quad = \quad \mathbb{P}[V \leq u] - \sup_{w\leq u}\{\mathbb{P}[V \leq w] - \beta\nu([1,w))\} \; .$$

In fact μ is a probability measure on $[1,\infty)$. Both $\mu(\{1\}) = \nu(\{1\})$ and $\mu([1,\infty)) = 1$ follow from considering $u = 1$, $u \to \infty$. Now we show μ is nonnegative:

$$\begin{aligned}&\beta\mu((u,u+u']) - \mathbb{P}[u < V \leq u+u'] \\ &\quad = \quad -\sup_{w\leq u+u'}\{\mathbb{P}[V \leq w] - \beta\nu([1,w))\} + \sup_{w\leq u}\{\mathbb{P}[V \leq w] - \beta\nu([1,w))\} \; .\end{aligned}$$

If the first supremum were to be attained at $w \leq u$ then the two suprema would cancel. If the first supremum were to be attained at $w' \in [u, u+u']$ then

$$\begin{aligned}&\beta\mu((u,u+u']) - \mathbb{P}[u < V \leq u+u'] \\ &\quad = \quad -\mathbb{P}[V \leq w'] + \beta\nu([1,w')) + \sup_{w\leq u}\{\mathbb{P}[V \leq w] - \beta\nu([1,w))\} \\ &\quad \geq \quad -\mathbb{P}[V \leq w'] + \beta\nu([1,w')) + \mathbb{P}[V \leq u] - \beta\nu([1,u)\end{aligned}$$

and hence

$$\beta\mu((u,u+u']) \quad \geq \quad \mathbb{P}[w' < V \leq u+u'] + \beta\nu([u,w')) \quad \geq \quad 0\,.$$

So we can deduce $\beta\mu$ is in fact a nonnegative measure.
On the other hand

$$\begin{aligned}&\beta\mu((u,u+u']) - \mathbb{P}[u < V \leq u+u'] \\ &\quad = -\sup_{w\leq u+u'}\{\mathbb{P}[V \leq w] - \beta\nu([1,w))\} + \sup_{w\leq u}\{\mathbb{P}[V \leq w] - \beta\nu([1,w))\} \;\leq\; 0\,,\end{aligned}$$

hence

$$0 \quad \leq \quad \beta\mu((u,u+u']) \quad \leq \quad \mathbb{P}[u < V \leq u+u']\,, \tag{15}$$

so $\beta\mu$ is absolutely continuous with respect to $\mathcal{L}(V)$ and indeed we can deduce

$$\beta\,\mathrm{d}\,\mu(u) \quad = \quad \mathbb{I}[\mathbb{P}[V > \cdot] - \beta\nu((\cdot,\infty)) \text{ hits current minimum at } u]\,\mathrm{d}\,\mathbb{P}[V \leq u] \; . \tag{16}$$

The minorization of $\beta\mu$ by $\mathcal{L}(V)$ follows from this argument: dependence only on $\beta\nu$ and $\mathcal{L}(V)$ from construction; finally, stochastic domination of $\beta\nu$ from

$$\begin{aligned} \beta\mu((u,\infty)) &= \mathbb{P}[V>u] - \inf_{w\le u}\{\mathbb{P}[V>w] - \beta\nu((w,\infty))\} \\ &= \sup_{w\le u}\{\beta\nu((w,\infty)) - \mathbb{P}[w<V\le u]\} \quad \ge \quad \beta\nu((u,\infty)) \,. \end{aligned}$$

□

Now use Lemma 7 to couple $\mathcal{L}(X_{n+1} \mid X_n = x)$ to $\mathcal{L}(Y_{n+1} \mid Y_n = y)$ whenever $\Lambda(x) \le y$ in a way which implements stochastic domination and ensures all the X_{n+1} regenerate simultaneously whenever $Y \le h$. This concludes the proof of Theorem 5. □

Note that the algorithm requires us to be able to draw from the equilibrium distribution of Y and to simulate its time-reversed equilibrium dual. Up to an additive constant $\log(Y)$ is the workload of a $D/M/1$ queue. This queue is amenable to exact calculations, so these simulation tasks are easy to implement (specializing the theory of the $G/M/1$ queue as discussed, for example, in Grimmett and Stirzaker 1992, ch. 11). However in general we do *not* expect this "universal dominating process" to lead to practical *domCFTP* algorithms! The difficulty in application will arise in determining whether or not regeneration has occurred as in Algorithm 4. This will be difficult especially if sub-sampling has been applied, since then one will need detailed knowledge of convolutions of the probability kernel for X (potentially a harder problem than sampling from equilibrium!).
Of course, in practice one uses different dominating processes better adapted to the problem at hand. For example an $M/D/1$ queue serves as a good log-dominating process for perpetuity-type problems and gives very rapid *domCFTP* algorithms indeed, especially when combined with other perfect simulation ideas such as multishift *CFTP* (Wilson 2000b), read-once *CFTP* (Wilson 2000a), or one-shot coupling (Roberts and Rosenthal 2002).
Finally note that, in cases when $\alpha \in [e^{-1}, 1)$ or when the small set $\{x \in \mathcal{X} : \Lambda(x) \le h\}$ is of order greater than 1, we are forced to work with coupling constructions that are effectively *non-co-adapted* (that is, sub-sampling means that target transitions X_{mk} to X_{mk+1} depend on sequences $Y_{mk}, Y_{mk+1}, \ldots, Y_{mk+k}$). The potential improvements gained by working with non-adapted couplings are already known not only to theory (the non-co-adapted filling couplings of Griffeath 1975; Goldstein 1979; and the efficiency considerations of Burdzy and Kendall 2000) but also to practitioners (Huber 2004: non-Markovian techniques in *CFTP*; Hayes and Vigoda 2003: non-Markovian conventional MCMC for random sampling of colorings).

3 Counter-example

We complete this note by describing a counter-example to show that the use of sub-sampling in the construction of Theorem 5 is essential.

Proposition 8 *There is a Markov chain X satisfying a Foster-Lyapunov condition with drift function Λ (and hence itself geometrically ergodic), such that without use of sub-sampling any dominating process Y based on Λ will fail to be positive-recurrent.*

Proof: We begin by choosing a sequence of disjoint measurable sets S_1, S_2, ..., subsets of $[1,\infty)$ such that each set places positive measure in every non-empty open set. We assert and prove the possibility of this by using an interior lemma:

Lemma 9 *One can construct a measurable partition S_1, S_2, ... of $[1,\infty)$,*

$$S_1 \sqcup S_2 \sqcup S_3 \sqcup \ldots \quad = \quad [1,\infty)\,,$$

with the property $Leb(S_i \cap (u,v)) > 0$ for all $0 < u < v < \infty$, all $i \in \{1,2,\ldots\}$.

Proof (of Lemma 9):
Enumerate the rational numbers in $[0,1)$ by $0 = \tilde{q}_0$, $\tilde{q}_1$, $\tilde{q}_2$, ... Choose $\alpha < 1/2$, and define

$$A_0 \quad = \quad \bigcup_{k=1}^{\infty} \bigcup_{n=0}^{\infty} \left[\tilde{q}_n + k, \tilde{q}_n + k + \alpha 2^{-n}\right]\,.$$

Then for each $k \geq 1$

$$\alpha \quad \leq \quad \mathrm{Leb}\left(A_0 \cap [k, k+1)\right) \quad \leq \quad 2\alpha\,.$$

Continue by defining a sequence of nested subsets $A_r \subset A_{r-1}$ by

$$A_r \quad = \quad \bigcup_{k=1}^{\infty} \bigcup_{n=0}^{\infty} \left[\frac{\tilde{q}_n + k}{2^r}, \frac{\tilde{q}_n + k}{2^r} + \frac{\alpha}{4^r} 2^{-n}\right]\,, \tag{17}$$

satisfying

$$\frac{\alpha}{4^r} \quad \leq \quad \mathrm{Leb}\left(A_r \cap \left[\frac{k}{2^r}, \frac{k+1}{2^r}\right)\right) \quad \leq \quad \frac{2\alpha}{4^r}\,. \tag{18}$$

Thus the measurable shell $B_r = A_r \setminus A_{r+1}$ places mass of at least $\frac{\alpha}{2\times 4^r}$ in each interval $\left[\frac{k}{2^r}, \frac{k+1}{2^r}\right)$.
It follows that if S is defined by

$$S \quad = \quad \bigcup_{s=1}^{\infty} \left(A_{r_s} \setminus A_{r_s+1}\right)$$

then $\mathrm{Leb}(S \cap U) > 0$ for every open set $U \subset [1,\infty)$. The desired disjoint sequence S_1, S_2, ... is obtained by considering a countably infinite family of disjoint increasing subsequences of the natural numbers. ☐

We revert to the proof of Proposition 8.
The Markov chain X is constructed on state space $[1,\infty)$, with pseudo-drift function $\Lambda(x) \equiv x$. We begin by fixing $\alpha \in (e^{-1}, 1)$, and set $C = [1, \alpha^{-1}]$. The set C will be the small set for the Foster-Lyapunov condition. Choose a measurable partition $S_1 \sqcup S_2 \sqcup S_3 \sqcup \ldots = [1,\infty)$ as in Lemma 9. Enumerate the rational numbers in $[1,\infty)$ by q_1, q_2, ...
We define the transition kernel $p(x,\cdot)$ of X on $[1,\infty)$ as follows:

For $x \in [1, \alpha^{-1}]$, set

$$p(x, \mathrm{d}\, y) \quad = \quad \exp(-(y-1))\,\mathrm{d}\, y \quad \text{for } y \geq 1\,,$$

so that if $X_n \in C$ then $X_{n+1} - 1$ has a unit rate Exponential distribution. Then:

C is a small set for X of order 1 (in fact it will be a regenerative atom!);

if $X_n \in C$ then $\mathbb{E}\left[X_{n+1}\right] = 2$;

if X has positive chance of visiting state 1 then the whole state space $[1,\infty)$ will be maximally Leb-irreducible.

For $x > \alpha^{-1}$ and $x \in S_i$, set

$$p(x, \mathrm{d}\, y) \quad = \quad \left(1 - \frac{\alpha}{q_i}\right)\delta_0(\mathrm{d}\, y) + \frac{\alpha}{q_i}\delta_{q_i x}(\mathrm{d}\, y)\,.$$

Note that, because we are using the identity pseudo-drift function $\Lambda(x) \equiv x$,

if $x \notin C$ then $\mathbb{E}\left[\Lambda(X_{n+1}) \mid X_n = x\right] = \mathbb{E}\left[X_{n+1} \mid X_n = x\right] = \alpha x$;
if $x \notin C$ then $\mathbb{P}\left[X_{n+1} = 1 \mid X_n = x\right] > 0$.

Thus X satisfies a geometric Foster-Lyapunov condition based on drift function Λ and small set C, and so is geometrically ergodic.
Suppose Y is a dominating process for X based on the identity function $\Lambda(x) \equiv x$. This means it must be possible to couple Y and X such that, if $\Lambda(X_n) = X_n \leq Y_n$ then $\Lambda(X_{n+1}) = X_{n+1} \leq Y_{n+1}$. This can be achieved if and only if

$$\mathbb{P}\left[X_{n+1} \geq z \mid X_n = u\right] \quad \leq \quad \mathbb{P}\left[Y_{n+1} \geq z \mid Y_n = x\right]$$

for all $z \geq 1$, and Lebesgue-almost all $u < x$. Therefore we require of such Y that

$$\begin{aligned}
\mathbb{P}\left[Y_{n+1} \geq \alpha x y \mid Y_n = x\right] \quad &\geq \quad \operatorname*{ess\,sup}_{u<x} \left\{\mathbb{P}\left[X_{n+1} \geq \alpha x y \mid X_n = u\right]\right\} \\
= \quad &\sup_i \operatorname{ess\,sup} \left\{\frac{\alpha}{q_i} \;:\; \alpha^{-1} < u < x, u \in S_i, q_i u > \alpha x y\right\} \\
= \quad &\sup_i \left\{\frac{\alpha}{q_i} \;:\; q_i > \alpha y\right\} \quad = \quad \frac{1}{y}\,,
\end{aligned}$$

using Markov's inequality, then the construction of the kernel of X, then the measure-density of the S_i.
So such a Markov chain Y must also (at least when above level α^{-1}) dominate $\exp(Z)$, where Z is a random walk with jump distribution Exponential(1) + $\log(\alpha)$. Hence it will fail to be positive-recurrent on the small set C when $\alpha \geq e^{-1}$. $\square$

There may exist some subtle re-ordering to provide *domCFTP* for such a chain based on a different pseudo-drift function; however the above lemma shows that *domCFTP* must fail for dominating processes for X based on the pseudo-drift function Λ.

4 Conclusion

We have shown that geometric ergodicity (more strictly, a geometric Foster-Lyapunov condition) implies the existence of a special kind of *domCFTP* algorithm. The algorithm is not expected to be practical: however it connects perfect simulation firmly with more theoretical convergence results in the spirit of the Foss and Tweedie (1998) equivalence between classic *CFTP* and uniform ergodicity. Note also that the "universal dominating process", the sub-critical $\exp(D/M/1)$ so derived, is itself geometrically ergodic.
It is natural to ask whether other kinds of ergodicity (for example, polynomial ergodicity) can also be related to perfect simulation in this way; this is now being pursued by Stephen Connor as part of his PhD research at Warwick.

References

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

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

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

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

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

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

Grimmett, G. R. and D. R. Stirzaker [1992]. *Probability and random processes* (Second ed.). New York: The Clarendon Press Oxford University Press.

Hayes, T. and E. Vigoda [2003]. A non-Markovian coupling for randomly sampling colorings. Preprint, University of Chicago Department of Computer Science., PDF file available.

Hobert, J. P. and C. P. Robert [2004]. A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling. *The Annals of Applied Probability 14*(3), 1295–1305.

Huber, M. [2004, March]. Time dependent update functions for perfect sampling. Conference presentation at IMS meeting on Monte Carlo Markov chain methods in Singapore, Department of Mathematics, Duke University.

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

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

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

Murdoch, D. J. and P. J. Green [1998]. Exact sampling from a continuous state space. *Scandinavian Journal of Statistics Theory and Applications 25*, 483–502.

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

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

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

Roberts, G. O. and J. S. Rosenthal [2002]. One-shot coupling for certain stochastic recursive sequences. *Stochastic Processes and Their Applications 99*(2), 195–208.

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

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.

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