SOME MULTI-STEP COUPLING CONSTRUCTIONS FOR MARKOV CHAINS

L.A. BREYER AND G.O. ROBERTS

Abstract. We describe some old and new methods for coupling the flow of a discrete time Markov chain, provided its transition function is known. Examples including Hastings-Metropolis and Gibbs samplers are given.

1. Introduction

Coupling constructions have been widely successful in tackling the problem of determining when and how fast a Markov chain converges to its stationary distribution (Lindvall, 1992, Meyn and Tweedie, 1993, Rosenthal, 199?, Roberts and Tweedie, 199?). In this report, we shall present several new coupling constructions for Markov chains whose transition probabilities are known to within a constant factor.

Much of the success of coupling methods owes to the *coupling inequality*, which is usually applied in the following form: let X_t and X'_t be two Markov chains defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that both chains have the same transition probabilities, and that X'_t is started in the equilibrium distribution π (assumed to exist). Then

$$\sup_{A \in \mathcal{F}} |\mathbb{P}(X_t \in A) - \pi(A)| \leq \mathbb{P}(T > t), \tag{1}$$

where $T = \min\{s : X_s = X'_s\}$. Consequently, if $\mathbb{P}(T < \infty) = 1$, the law of X_t must converge in total variation to π, and we can give a rate of convergence if we study the distribution of T. Note that such arguments are purely probabilistic. The proof of (1) is so short that we repeat it here as a service to readers unfamiliar with it.

Date: January, 2000.

1991 *Mathematics Subject Classification.* Primary 60J, Secondary 60F.

Key words and phrases. Markov chains, Coupling Constructions, Metropolis-Hastings, Gibbs samplers.

Since the process X'_t is stationary, we have

$$\begin{aligned}\mathbb{P}(X_t \in A) - \pi(A) &= \mathbb{P}(X_t \in A) - \mathbb{P}(X'_t \in A) \\ &= \mathbb{P}(X_t \in A, T > t) - \mathbb{P}(X'_t \in A, T > t),\end{aligned}$$

as $X_t = X'_t$ on $\{T \leq t\}$. Consequently,

$$\begin{aligned}|\mathbb{P}(X_t \in A) - \pi(A)| &= |\mathbb{P}(X_t \in A \,|\, T > t) - \mathbb{P}(X'_t \in A \,|\, T > t)|\, \mathbb{P}(T > t) \\ &\leq \mathbb{P}(T > t)\end{aligned}$$

for any set A, for the distance between two numbers in $[0, 1]$ cannot be greater than one.

The problem we shall discuss in this report can be described as follows: We are given a computer program (sometimes known as a *stochastic recursive sequence*) for simulating a Markov chain with a specified set of transition probabilities from any initial position in the state space. By modifying this program slightly, we can ensure that two chains simulated from different initial locations have a positive chance of coupling in a finite time. The constructions we propose require little analytical knowledge about the Markov chain.

The report is organized as follows: Section 2 describes the need for coupling constructions, by showing the range of different path behaviours which can arise even when the transition probabilities are fixed. Section 3 recalls briefly the so-called splitting method, which is widely used and recommended as a generic coupling method for Markov chains. In Section 4, we describe how to couple paths using one or many independent proposals at each simulation step. Unlike splitting, this method does not require calculating minorizations explicitly before it can be used. Both methods are applicable whenever the n-step transition probabilities $P^n(x, dy)$ of a Markov chain are known. Since this is extremely unlikely when $n > 1$, these should essentially be thought of as one-step coupling constructions: two paths which are already sufficiently close can be coupled in one step.

The second part of this report (Sections 5 and onwards) discusses the problem of coupling Markov chain sample paths over many time steps. The couplings we propose are not Markov, but preserve marginal distributions and ergodic properties. Most importantly, they are based upon the method described in Section 4, which allows coupling with a minimum of explicit calculations.

2. Flows

Let a transition kernel $P(x, dy)$ be given on a sufficiently regular state space E. It is assumed that a Markov chain with transition probabilities P can be constructed as a flow. This usually proceeds as follows:

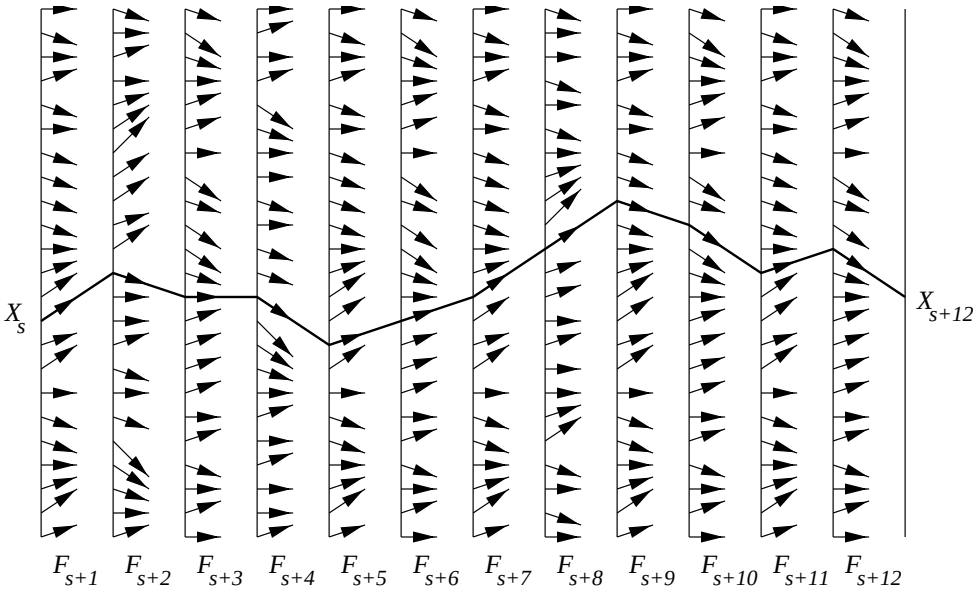

Figure 1. A Markov flow and a realization $X_t = \varphi_{s,t}(X_s)$.

Let $F : E \to E$ denote a random function (or computer program) which, given a point $x \in E$, produces a random variable $F(x)$ satisfying

$$\mathbb{P}\big(F(x) \in dy\big) = P(x, dy). \tag{2}$$

Since the function F is random, we may define an infinite sequence of independent, identically distributed functions $F_1, F_2, \dots$ and compose them, arriving at a (random) *flow*

$$\varphi_{s,t}(x) = F_t(F_{t-1}(\cdots F_{s+1}(x)\cdots)), \quad t \geq s. \tag{3}$$

Given any initial random variable $X_0 \in E$, we can now set $X_t = \varphi_{0,t}(X_0)$, $t = 1, 2, 3, \dots$ and this is then a realization of a Markov chain, initially distributed as X_0, and whose evolution is governed by P. We can also get a whole collection of Markov chains with transition probabilities P if we keep the realization of $\varphi_{0,t}$ fixed and vary X_0. These realizations are then clearly dependent.

As the functions $F_1, F_2, \ldots$ need only satisfy (2), there is considerable freedom in choosing them, and consequently a great deal of variety in the possible implementations of the flow $\varphi_{0,t}$. Consider the following example: let $E = \{1, \ldots, d\}$, $d > 3$, and consider the following transition kernel:

$$P(x, \{x \pm 1\}) = p, \quad P(x, \{x\}) = 1 - 2p, \quad P(x, \{y\}) = 0 \text{ otherwise}, \tag{4}$$

with the conventions $d + 1 = 1$ and $1 - 1 = d$ (that is , addition modulo d). There are several ways of constructing the associated Markov chain.

First way: Let $W \in \{-1, 0, +1\}$ be a random variable satisfying $\mathbb{P}(W = \pm 1) = p$ and $\mathbb{P}(W = 0) = 1 - 2p$. Then

$$F(x) = x + W \text{ satisfies } \mathbb{P}\big(F(x) = y\big) = P(x, \{y\}).$$

Defining φ_t as in (3), let us consider two Markov chains $X_t = \varphi_{0,t}(x)$ and $X'_t = \varphi_{0,t}(x')$. If $x \neq x'$, then we shall always have $X_t \neq X'_t$. Clearly, the paths must evolve in parallel, and $\mathbb{P}(T < \infty) = 0$.

Second way: Let $W_1, \ldots, W_n$ be independent random variables, each with the same distribution as W in the previous example. Clearly,

$$\bar{F}(x) = x + W_x \text{ satisfies } \mathbb{P}\big(\bar{F}(x) = y\big) = P(x, \{y\}).$$

Define $\bar{\varphi}_{s,t}$ as in (3) with F_t replaced by $\bar{F}_t$. Now a moment's thought reveals that the chains $\bar{X}_t = \bar{\varphi}_t(x)$ and $\bar{X}'_t = \bar{\varphi}_t(x')$ evolve independently until the first time $\bar{T}$ that $\bar{X}_t = \bar{X}'_t$ holds, after which they behave identically. A standard argument based on the Borel Cantelli lemma (Lindvall, 1992, p.??) shows that $\mathbb{P}(\bar{T} < \infty) = 1$.

Third way: The cases $\mathbb{P}(T < \infty) = 0$ and $\mathbb{P}(T < \infty) = 1$ are not the only two possibilities. If we allow the maps F_t to depend on each other, we can easily construct more complicated cases. Let $Z_0 = 1$, and $Z_{t+1} = Z_t + W_t$, where Z_t are independent with the same distribution as W in the first example. We set

$$\bar{\bar{F}}_t(x) = \begin{cases} x + W_t & \text{if } x = Z_t \text{ or } x = Z_t \pm 1, \\ x + W_{t,x} & \text{otherwise,} \end{cases}$$

where $W_{t,x}$ are also i.i.d. with the distribution of W, independent of everything else. Again, we can check that $\mathbb{P}\big(\bar{\bar{F}}(x) = y\big) = P(x, \{y\})$. We construct

the flow

$$\bar{\bar{\varphi}}_{s,t}(x) = \bar{\bar{F}}_t(\bar{\bar{F}}_{t-1}(\dots \bar{\bar{F}}_{s+1}(x)\dots)),$$

and note that if $x \neq x'$, $x \neq 1$, $x' \neq 1$, then the processes $\bar{\bar{X}}_t = \bar{\bar{\varphi}}_{0,t}(x)$ and $\bar{\bar{X}}'_t = \bar{\bar{\varphi}}_{0,t}(x')$ can either couple or not. If $x = 1$, $x' \neq 1$, they never couple.

We remark that in the examples above, the flows correspond to the same "Markov chain", i.e. the same transition probabilities, while the actual coupling behaviour is widely different. In all cases however, that is, for each of $X_t = \varphi_{0,t}(x)$, $\bar{X}_t = \bar{\varphi}_{0,t}(x)$, and $\bar{\bar{X}}_t = \bar{\bar{\varphi}}_{0,t}(x)$, the law of the process at time t converges to the uniform distribution on $E = \{1, \dots, d\}$, and this at the *same* rate.

Suppose now that d is sufficiently large, say $d = 100$. If $x, x' \in \{1, \dots, 100\}$ are chosen such that $|x - x'| > 1$, then the measures $P(x, \cdot)$ and $P(x', \cdot)$ are mutually singular.In particular, it is impossible to ensure that $F(x) = F(x')$ for *any* choice of random function F consistent with (2). In the worst case, we require 25 steps minimum to couple two initial points in the flow sufficiently far apart. It seems natural therefore to study coupling methods which operate over several time steps.

3. Coupling by Splitting

If there is to be any hope of applying coupling arguments to a chain given by a transition kernel, it is clear that we must be able to construct suitable flows (i.e. those such that $\mathbb{P}(T < \infty) > 0$) in a systematic way. One such method, and so far the only one to be widely successful, is the *splitting technique*, originally due to Nummelin (1978).

We assume that we can exhibit a *small set* $C \subset E$, that is, there exist $\epsilon > 0$, $n \geq 1$ and a probability measure μ on E such that

$$P^n(x, dy) \geq \epsilon\mu(dy), \quad x \in C. \tag{5}$$

Here $P^n(x, dy) = \int P(x, dx_1) \int P(x_1, dx_2) \dots P(x_{n-1}, dy)$ denotes the n-step transition probability kernel. If we write, for $x \in C$,

$$\begin{aligned} P^n(x, dy) &= \epsilon\mu(dy) + \big(P^n(x, dy) - \epsilon\mu(dy)\big) \\ &= \epsilon\mu(dy) + (1 - \epsilon)Q(x, dy), \end{aligned}$$

we have exhibited the n-step transition probability as a mixture of a kernel $Q(x, dy) = (1-\epsilon)^{-1}\big(P^n(x, dy) - \epsilon\mu(dy)\big)$ and a distribution μ independent of $x \in C$. The condition (5) ensures that Q is positive.

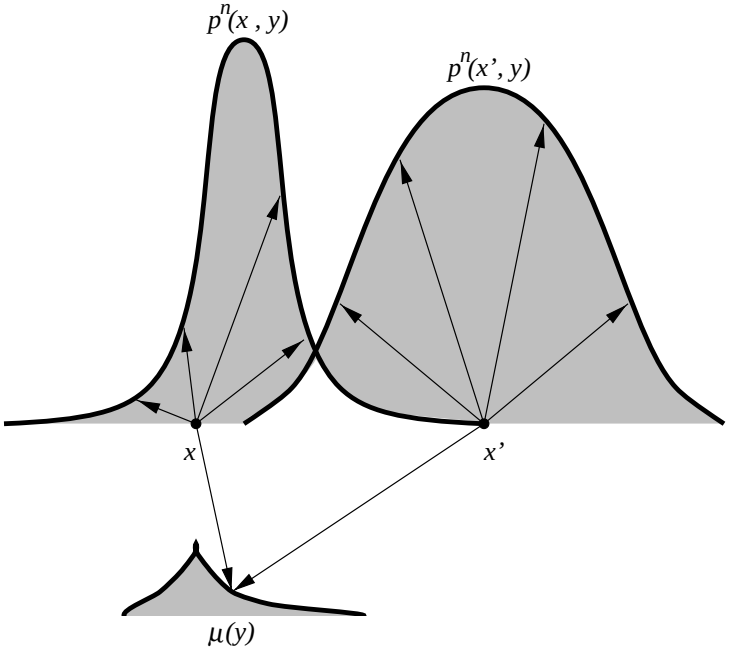

FIGURE 2. Splitting construction: with probability ϵ, both particles will select the same point whose law is μ. Otherwise, particle x chooses a point with law $q(x, y) = \big(p^n(x, y) - \mu(y)\big)/(1-\epsilon)$ and particle x' chooses a point with law $q(x', y) = \big(p^n(x', y) - \mu(y)\big)/(1-\epsilon)$.

Now suppose we simulate two Markov chains X_t and X'_t in any way we choose, as long as both chains can simultaneously enter C once in a while. When this happens at some time t_0 say, we modify the simulation as follows: with probability ϵ, generate $Z \sim \mu$ and set $X_{t_0+n} = X'_{t_0+n} = Z$. With the remaining $1-\epsilon$ probability, generate $X_{t_0+n} \sim Q(X_{t_0}, \cdot)$ and $X'_{t_0+n} \sim Q(X'_{t_0}, \cdot)$ (see Figure 3). We do not define the chains at the intermediate times $s = t_0+1, \ldots, t_0+n-1$. Clearly, at least with probability $\epsilon > 0$, we shall have $T = t_0 + n$ using this method.

While theoretically convenient, this construction does not describe how to simulate from Q. Fortunately, the following coupling method ("flow coupler") is easy to implement. We set $n = 1$ for simplicity, and because P^n for $n > 1$ is rarely known explicitly.

Gamma Coupler: Modifies an existing flow by operating on n-step transitions.

Requires: Explicit knowledge of the minorization condition (5) (with $n = 1$ for simplicity), including the Radon-Nikodym derivative

$$D(x, y) = \mu(dy)/P(x, dy),$$

and any flow φ_t built from maps F_t as in (3), say.

Returns: A flow

$$\Gamma(\varphi)_{s,t}(x) = \Gamma(F_t) \circ \cdots \circ \Gamma(F_{s+1})(x).$$

Method: For each t let Z_t be an independent random variable with distribution μ, and independently, let ξ_t be uniform on $[0, 1]$. We set

$$\Gamma(F_t)(x) = \begin{cases} Z_t & \text{if } x \in C \text{ and } D\big(x, F_t(x)\big) > \epsilon^{-1}\xi_t \\ F_t(x) & \text{otherwise.} \end{cases} \tag{6}$$

As output, the Gamma Coupler produces a modified flow $\Gamma(\varphi)_{s,t}$, and consequently a new family of Markov chains $X_t^\Gamma = \Gamma(\varphi)_{0,t}(x)$. It is important to check that these have the correct transition probabilities (i.e. P). The proof is short and well known, but we repeat it here for the convenience of the reader.

Lemma 1. *The Gamma Coupler couples flows correctly, i.e.*

$$\mathbb{P}\Big[\Gamma(F_t)(x) \in dy\Big] = P(x, dy).$$

Proof. Let $\varphi_{s,t}$ be a flow in the form (3), where each $F_s(x)$ has the law $P(x, \cdot)$. If f is any bounded, real valued function and $x \in C$, we have

$$\begin{aligned} \mathbb{E}\Big(f \circ \Gamma(F_s(x)\Big) - \int f(y)P(x, dy) \\ &= \mathbb{E}\Big(\big(f \circ Z - f \circ F_s(x)\big)1 \wedge \epsilon D(x, F_s(x))\Big) \\ &= \int (f(z) - f(y))P(x, dy) \wedge \epsilon\mu(dz). \end{aligned}$$

The last term is antisymmetric (equal to its negative) under the change of variables $y \leftrightarrow z$, and consequently it must be zero. Thus we see that $\Gamma(F_s)(x)$ has the law $P(x, \cdot)$. □

There is no guarantee in general that the chains $X_t^\Gamma = \Gamma(\varphi)_{0,t}(x)$ and $X'^\Gamma_t = \Gamma(\varphi)_{0,t}(x')$ will couple, unless we make further hypotheses upon the *original* flow $\varphi_{s,t}$. We will discuss this issue in a later section.

A simple generalization of the Gamma Coupler consists in taking a finite collection of small sets $C_1, \ldots, C_r$ say, each of which satisfies (5) with some ϵ_i and μ_i (but the same n). Then we take $Z_{t,i} \sim \mu_i$ define $D_i(x, y)$ accordingly, and replace (6) with

$$\Gamma(F_t)(x) = \begin{cases} Z_{t,i} & \text{if } x \in C_i \text{ and } D_i\big(x, F_t(x)\big) > \epsilon_i^{-1}\xi_t \\ F_t(x) & \text{otherwise.} \end{cases}$$

This version of the Gamma Coupler has been called the MultiGamma Coupler (Green and Murdoch, 1998).

4. Coupling with proposals

In this section, we describe another method for coupling chains by modifying an existing flow. To motivate the procedure, let us examine more closely the formula (6) for $x \in C$. It is not difficult to see that the recipe amounts to rejection sampling: we are attempting to simulate μ using random variables generated from $P(x, \cdot)$. This is possible since by (5), the latter distribution dominates μ.

The most important aspect of the Gamma Coupler for us is the following: given $F_t(x) \sim P(x, \cdot)$, we either keep this random variable, or replace it with Z_t according to a decision rule which yields $\Gamma(F_t)(x) \sim P(x, \cdot)$, that is, *preserves* the distribution $P(x, \cdot)$.

From this observation, it is clear now how to construct different coupling methods; for each x, define a Markov chain whose stationary distribution is $P(x, \cdot)$. This chain should propose a common value Y say, independent of x, which ensures coupling between any two points x and x' which both happen to accept Y.

In the coupling method described below, we shall assume that $P(x, dy)$ has a density $p(x, y)$ with respect to some σ-finite measure m on E. All other densities mentioned will also be with respect to m.

Independence Coupler: Modifies an existing flow to induce coupling, by operating on one step transitions.

Requires: One step transition probability density $p(x, y)$, proposal density $q(y)$ and any flow φ_t built from maps F_t as in (3), say.

Returns: A flow $\mathcal{C}_{\mathcal{I}}(\varphi)_{s,t} = \mathcal{C}_{Y_{t,1},\ldots,Y_{t,k}}(F_t) \circ \cdots \circ \mathcal{C}_{Y_{s+1,1},\ldots,Y_{s+1,k}}(F_{s+1})$.

Parameter: $k = 1, 2, 3, \ldots$ to regulate the "coupling propensity".

Method: Let $\mathcal{I} = (Y_{s,j} : 0 \le s \le t, 1 \le j \le k)$ be a vector of independent random variables with common density q, and let $\xi_{s,j}$ be corresponding independent uniforms on $[0, 1]$. Define

$$\mathcal{C}_{Y_{t,1},\ldots,Y_{t,k}}(F_t)(x) = \mathcal{C}_{Y_{t,k}}(\cdots \mathcal{C}_{Y_{t,1}}(F_t)(x) \cdots),$$

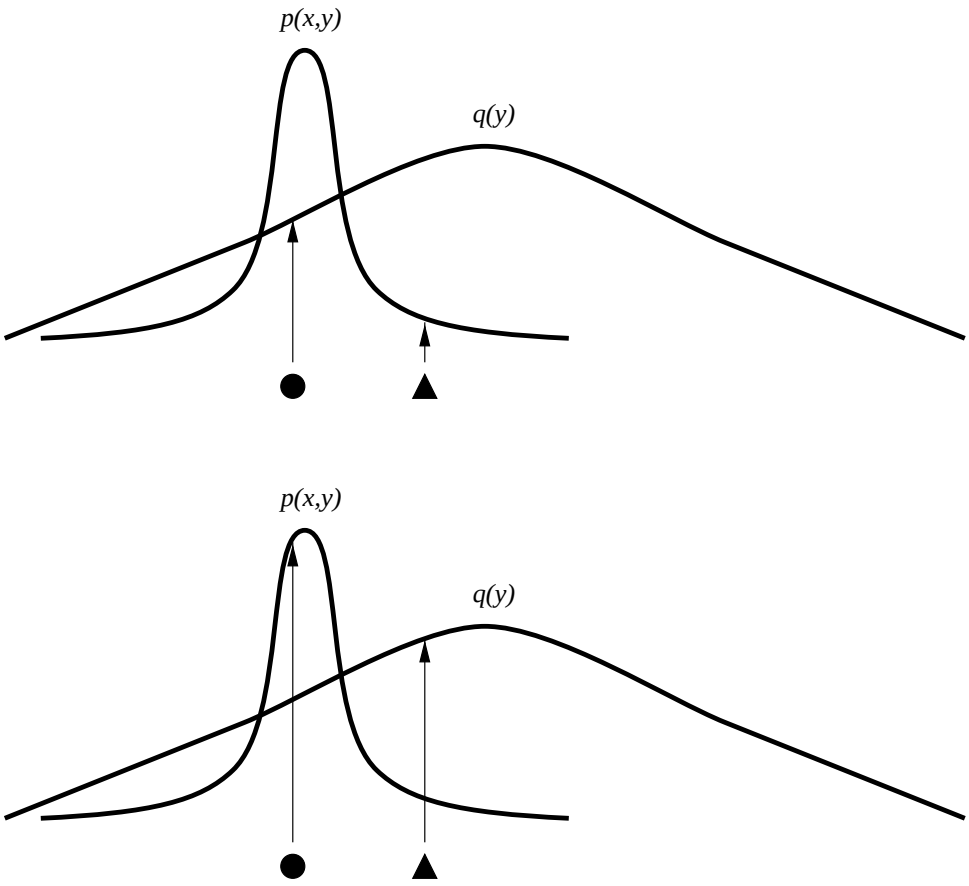

FIGURE 3. *Which is more likely, the top or the bottom picture?* If two independent points $F(x)$ and Y are given, the Independence Coupler chooses the jointly most likely provenance of each. Then it exchanges $F(x)$ and Y if necessary, and calls the result $\Gamma(F)(x)$.

where for any map F and random variable Y,

$$\mathcal{C}_Y(F)(x) = \begin{cases} Y & \text{if } \frac{p(x,Y)q(F(x))}{p(x,F(x))q(Y)} > \xi \\ F(x) & \text{otherwise.} \end{cases} \tag{7}$$

with the convention that $x/0 = 1$ and $0/0 = 0$. The variables $\xi_{s,j}$ are used in (7) only in conjunction with $Y_{s,j}$.

Figure 4 gives another interpretation of (7). To construct $\mathcal{C}_Y(F)(x)$, a choice is made between two independent random variables, namely $F(x)$ and Y. We measure the likelihood $p\big(x, F(x)\big)q\big(Y\big)$ that $F(x)$ has distribution $p(x,\cdot)$ and Y has distribution q against the likelihood $p(x,Y)q\big(F(x)\big)$ that Y has distribution $p(x,\cdot)$ and $F(x)$ has distribution q. In case the latter hypothesis is accepted, we set $\mathcal{C}_Y(F)(x) = Y$, otherwise $\mathcal{C}_Y(F)(x) = F(x)$.

Lemma 2. *The Independence Coupler couples flows correctly, i.e.*

$$\mathbb{P}(\mathcal{C}_Y(F)(x) \in dy) = P(x, dy).$$

Proof. See the proof in Breyer and Roberts (1999). □

A potential difficulty arises above when $P(x, dy)$ does not have a density (clearly this can only occur on an uncountable state space). A straightforward generalization of (7) is possible, if we replace the acceptance ratio by a Radon-Nikodym derivative. Alternatively, variations on the theme are also possible: in Breyer and Roberts (1999), it is shown how to couple the flow of a Metropolis-Hastings chain, whose transition probabilities are never absolutely continuous on an uncountable state space. See also the second example below.

Example: Random Walk on the circle. Let $E = \{e^{ix} : x \in \mathbb{R}\}$ denote the unit circle in the complex plane. We define a flow in the following way. Let $p : (-\pi, \pi] \to \mathbb{R}$ be a probability density and suppose that W_t are i.i.d. with density p. Then

$$F_t(e^{ix}) = e^{i(x+W_t)} \text{has t.f. given by } P\big(e^{ix}, d(e^{iy})\big) = \bar{p}(y-x)dy,$$

where $\bar{p} : \mathbb{R} \to \mathbb{R}$ is the periodic continuation of p. To apply the random coupling method, we take (say) a uniform proposal $q : (-\pi, \pi] \to \mathbb{R}$, $q(x) = (2\pi)^{-1}$. Then

$$\mathcal{C}_Y(F)(e^{ix}) = \begin{cases} e^{iY} & \text{if } \frac{\bar{p}(Y-x)}{\bar{p}(W)} > \xi, \\ e^{i(x+W)} & \text{otherwise.} \end{cases}$$

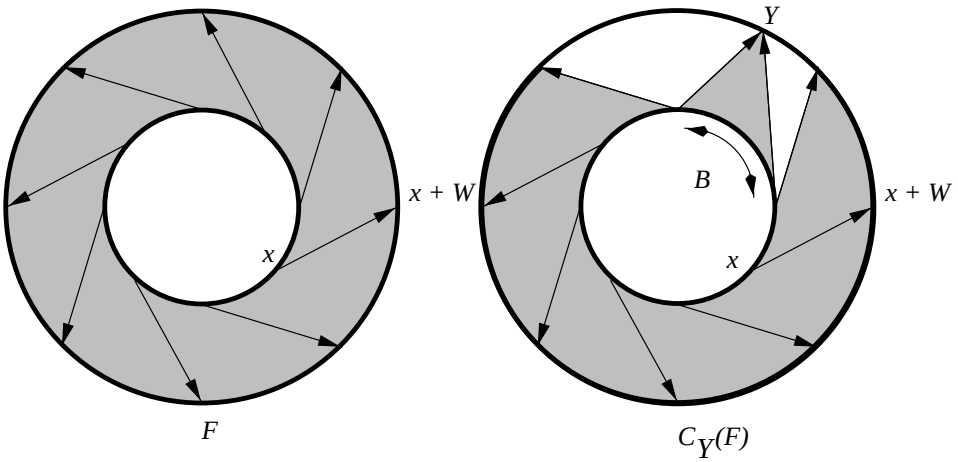

FIGURE 4. Effect of one iteration of the Independence Coupler on the circle map $F(e^{ix}) = e^{i(x+W)}$.

The effect produced is simple to describe. A random subset

$$B = \left\{ x : \bar{p}(Y - x) > \xi \bar{p}(W) \right\}$$

of the circle is created, and all points x within B now map to the common value Y. The map F outside of B is unchanged (see Figure 4).

Increasing the coupling propensity $k > 1$, we proceed as follows: since $\mathcal{C}_{Y_{t,1}}(F_t)(x) \sim P(x, \cdot)$ by construction, we apply (7) to this random function, giving a new random function $\mathcal{C}_{Y_{t,1},Y_{t,2}}(F_t) \sim P(x, \cdot)$. Since another point $Y_{t,2}$ would be proposed, a generally different section B' of the circle would accept. Thus the random function $\mathcal{C}_{Y_{t,1},Y_{t,2}}(F_t)$ would map a point x to either $Y_{t,1}$, $Y_{t,2}$ or $F_t(x)$. Iterating this procedure sufficiently, it is conceivable (and can be shown, see Breyer and Roberts, 1999) that eventually, $\mathcal{C}_{Y_{t,1},\ldots,Y_{t,k}}(F_t)(x)$ no longer depends on $F_t(x)$, but instead the function is built from a finite number of regions $B, B', \ldots$ and corresponding points $Y_{t,1}, Y_{t,2}, \ldots$. From the simulation point of view, this allows a continuum of

Markov chain paths $X_{t-1}(x) = \varphi_{0,t-1}(x)$, $x \in E$, to be replaced by a finite number $X_t(x) = \mathcal{C}_{Y_{t,1},\ldots,Y_{t,k}}(F_t)(X_{t-1}(x))$ in just one step!

Example: Metropolis-Hastings chains.

Let $q(x,z)$ be a kernel density and π a probability density on $E \subset \mathbb{R}^d$. A Hastings-Metropolis Markov chain is usually simulated by using a flow φ_t in the form (3), where

$$F_t(x) = \begin{cases} Z_t(x) & \text{if } \pi\big(Z_t(x)\big)q\big(Z_t(x),x\big) > \psi_t \pi(x) q\big(x, Z_t(x)\big) \\ x & \text{otherwise,} \end{cases} \tag{8}$$

where $Z_t(x)$ is a random variable with density $q(x,\cdot)$, and ψ_t is an independent uniform on $[0,1]$. The transition kernel P for this Markov chain can be written

$$P(x,dz) = \frac{1}{\pi(x)}\big(\pi(x)q(x,z) \wedge \pi(z)q(z,x)\big)dz + r(x)\delta_x(dz), \tag{9}$$

where $r(x)$, the *rejection probability*, is such that $P(x,E) = 1$, $a \wedge b$ means $\min(a,b)$ and we are assuming $\pi(x) > 0$ for simplicity.

Clearly, this transition function does not have a density with respect to any σ-finite reference measure, unless the state space E is countable. Nevertheless, we can still hope to couple the chain, since $P(x,dz)$ and $P(x',dz)$ are often partially absolutely continuous. For example, if $q(x,z)$ is continuous as a function of x, then this will be so if x and x' are sufficiently close.

To couple the flow φ_t constructed from the random functions F_t given in (8), we proceed as follows: *first*, we couple the proposal $Z_t(x)$: since $q(x,z)$ is a density, this poses no problems. Let $Y_{t,1}$ be an independent random variable with density b say, and $\xi_{t,1}$ and independent uniform on $[0,1]$. We set

$$\mathcal{C}_{Y_{t,1}}(Z_t)(x) = \begin{cases} Y_{t,1} & \text{if } q(x,Y_{t,1})b\big(Z_t(x)\big) > \xi_{t,1} q\big(x, Z_t(x)\big)b(Y_{t,1}) \\ Z_t(x) & \text{otherwise.} \end{cases}$$

Note that we still have $\mathbb{P}\big(\mathcal{C}_{Y_{t,1}}(Z_t)(x) \in dz\big) = q(x,z)dz$, so this is a valid way of generating a proposal for the Hastings Metropolis Markov chain. More generally,

we have $\mathbb{P}\big(\mathcal{C}_{Y_{t,1},\dots,Y_{t,k}}(Z_t)(x) \in dz\big) = q(x,z)dz$ for all $k \geq 1$. Now we simply set

$$\mathcal{C}_{\mathcal{Y}_t}(F_t)(x) = \begin{cases} \mathcal{C}_{\mathcal{Y}_t}(Z_t)(x) & \text{if } \pi\big(\mathcal{C}_{\mathcal{Y}_t}(Z_t)(x)\big)q\big(\mathcal{C}_{\mathcal{Y}_t}(Z_t)(x), x\big) > \psi_t \pi(x) q\big(x, \mathcal{C}_{\mathcal{Y}_t}(Z_t)(x)\big) \\ x & \text{otherwise,} \end{cases}$$

where $\mathcal{Y}_t = (Y_{t,1}, \dots, Y_{t,k})$ and we have $\mathbb{P}\big(F_t^{(n_0)}(x) \in dz\big) = P(x, dz)$ as in (9).

Example: Gibbs samplers

In $\mathbb{R}^d$, let $\pi(x_1, \dots, x_d)$ denote a probability density, and consider the conditionals

$$\pi_i(x_i \,|\, x_j, j \neq i) = \pi(x_1, \dots, x_d) / \int \pi(x_1, \dots, x_d) dx_i.$$

Associated with π_i is a kernel

$$P_i(x, dy) = \pi_i(y_i \,|\, x_j, j \neq i) dy_i \bigotimes_{j \neq i} \delta_{x_j}(dy_j),$$

which corresponds to the move $F_i(x_1, \dots, x_d) = (x_1, \dots, Z_i, \dots, x_d)$ where $Z_i \sim \pi_i(\cdot \,|\, x_j, j \neq i)$. Gibbs sampler flows consist in the random or periodic composition of the maps F_i where i ranges over all indices $1, \dots, d$. For simplicity, we shall consider only the case of periodic composition.

Note an important feature of Gibbsian flows: if $x \neq x'$ with at least two coordinates different, then $P_i(x, \cdot)$ and $P_i(x', \cdot)$ are always mutually singular. Consequently, *it is impossible to couple Gibbs sampler paths arising from x and x' in a single step.* Nevertheless, it is possible to couple Gibbs sampler flows over several simulation time steps, as follows:

A single complete sweep $F = F_d \circ \cdots \circ F_1$ through all coordinates has a transition density

$$p(x, y) = \prod_{i=1}^{d} \pi_i(y_i \,|\, y_1, \dots, y_{i-1}, x_{i+1}, \dots, x_d).$$

Hence if $Y = (Y_1, \dots, Y_d) \sim q(\cdot)$ independently of F, we define $\mathcal{C}_Y(F)$ exactly as in (7).

5. Catalytic Couplers

The methods presented in the first half of this report are able to couple two (or more) sample paths, by modifying the underlying flow $\varphi_{s,t}$ in such a way that

neighbouring paths can link up in a single simulation time step. One advantage is that the resulting flow (either $\Gamma(\varphi)_{s,t}$ or $\mathcal{C}_{\boldsymbol{Y}}(\varphi)_{s,t}$) remains Markovian. An intuitive pitfall is that in a given simulation, it is necessary to wait until two given paths come

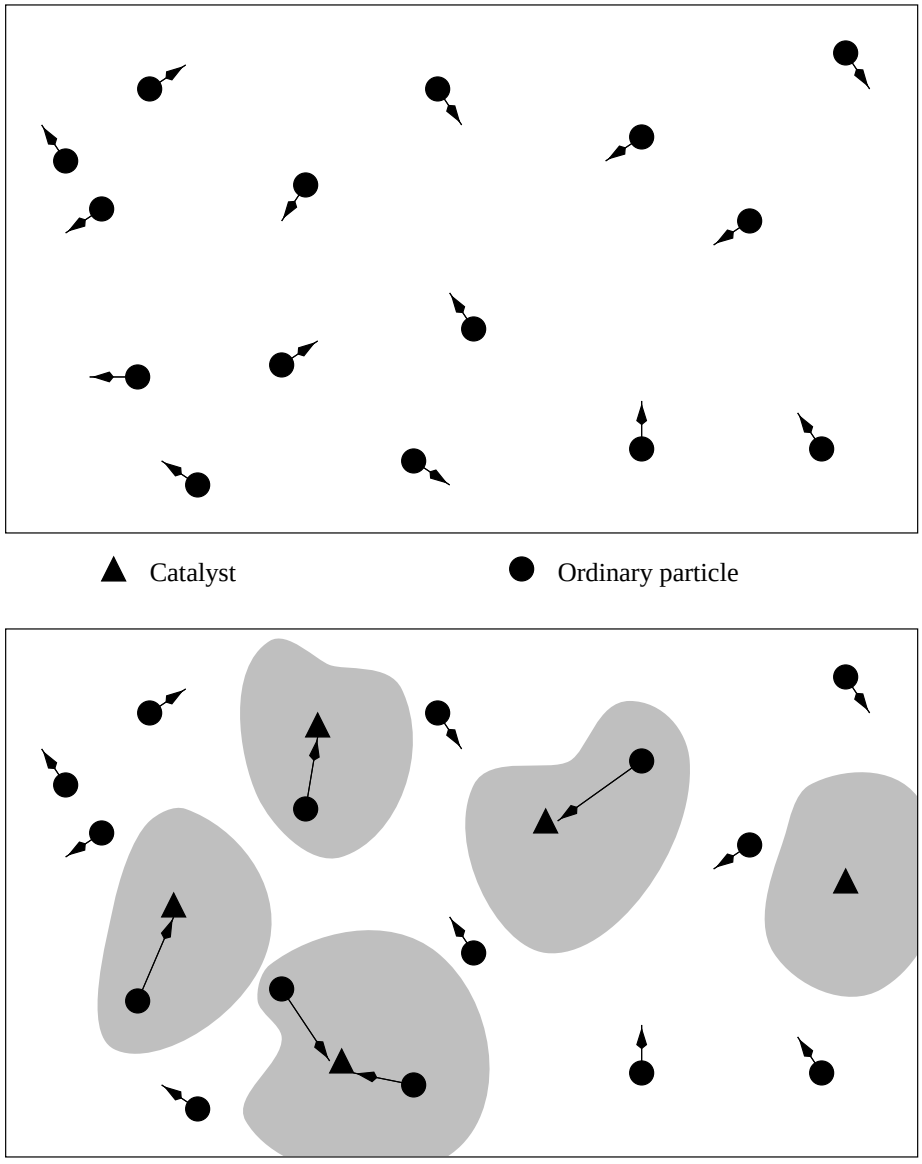

FIGURE 5. One step flow $F(x)$ without catalyst (top) and corresponding flow $\mathcal{C}_{Y^{(1)},\ldots,Y^{(5)}}(F)(x)$ with five catalysts. State space is two dimensional

close together before they have a chance of coupling. With the splitting technique, it is not immediately clear how this problem may be addressed, however in the case of the Independence Coupler, we can try giving the proposals $Y_{t,1}, Y_{t,2}, \dots$ some limited 'intelligence', thinking of them as processes in time $t = 0, 1, 2, \dots$. For each $j = 1, \dots, k$, we will call $Y_{t,j}$ a *catalyst*, since it helps two neighbouring paths to couple.

To emphasise the point, we shall also write $Y_t^{(j)}$ instead of $Y_{t,j}$. Here now is a definition:

Catalytic Coupler: Modifies an existing flow to facilitate coupling of n step transitions. If $\kappa(x, y) = p(x, y)$ below, we sometimes call this an *Autocatalytic Coupler*.

Requires: One step transition probability density $p(x, y)$, another transition density $\kappa(x, y)$ that we can sample from and any flow $\varphi_{s,t}$ built from maps F_t as in (3) say.

Parameter: $k = 1, 2, 3, \dots$ representing the number of catalysts.

Returns: A flow $\mathcal{C}_{Y^{(1)},\dots,Y^{(k)}}(\varphi)_{s,t} = \mathcal{C}_{Y_t^{(1)},\dots,Y_t^{(k)}}(F_t) \circ \cdots \circ \mathcal{C}_{Y_{s+1}^{(1)},\dots,Y_{s+1}^{(k)}}(F_{s+1})$.

Method: Independently of the flow $\varphi_{s,t}$ and of each other, let $Y_s^{(j)}$, $s = 0, 1, 2, \dots$ be Markov chains with transition density $\kappa(x, y)$ for $j = 1, \dots, k$. Let $\xi_s^{(j)}$ be associated uniforms on $[0, 1]$. We set

$$\mathcal{C}_{Y_t^{(1)},\dots,Y_t^{(k)}}(F_t) = \mathcal{C}_{Y_t^{(k)}}(\cdots \mathcal{C}_{Y_t^{(1)}}(F_t) \cdots),$$

and for any map F, $j = 1, \dots, k$, $s = 0, 1, 2, \dots$

$$\mathcal{C}_{Y_s^{(j)}}(F)(x) = \begin{cases} Y_s^{(j)} & \text{if } \frac{p(x,Y_s^{(j)})\kappa(Y_{s-1}^{(j)},F(x))}{p(x,F(x))\kappa(Y_{s-1}^{(j)},Y_s^{(j)})} > \xi_s^{(j)} \\ F(x) & \text{otherwise,} \end{cases} \tag{10}$$

using the conventions $x/0 = 1$ and $0/0 = 0$.

The Catalytic Coupler is a generalization of the Independence Coupler of Section 4, as can be seen by taking $\kappa(x, y) = q(y)$ and comparing (10) with (7).

The effect of the catalysts upon the flow is illustrated in Figure 5. At time zero, let us distribute a collection $(x_i : i = 1, \dots, m)$ of round particles throughout the state space E. Over time $t = 1, 2, \dots$, the particles evolve as dependent Markov chains $(X_t(x_i) = \varphi_{0,t}(x_i) : i = 1, \dots, m)$ with transition density $p(x, y)$, which may or may not result in some of the particles coupling successfully. Now apply the

Catalytic Coupler. We distribute k triangular "catalyst" particles at time zero, and allow them to evolve independently with transition density $\kappa(x, y)$ throughout E. When a round particle finds itself close enough to some triangular particle, it gets captured by the catalyst, and carried for a random time period, before being freed again. If two or more round particles are close to some triangular particle *at the same time*, all get captured and carried, but only one of these is ever freed again, as the others are now coupled to it. Eventually, only one round particle may survive. The path histories of the round particles are given by $(X_t = \mathcal{C}_{Y^{(1)},\ldots,Y^{(k)}}(\varphi)_{0,t}(x_i) : i = 1, \ldots, m$, while the path histories of the catalysts is given by $(Y_t^{(j)} : j = 1, \ldots, k)$.

We can interpret the density $\kappa(x, y)$ as labeling the "species" of catalyst that we are using. Different species κ have different abilities to carry particles for a long time, and also different ranges of action.

An interesting special case is when $p(x, y) = \kappa(x, y)$, that is, the catalyst particles are of the same type as the others. In this case, we will call the corresponding Flow Coupler *autocatalytic*. The important particularity of this coupler is that, once a catalyst captures a particle, it *never* frees it again. At best, it passes it to another catalyst to carry. Indeed, it is easy to check from (10) that

$$\mathcal{C}_{Y_t^{(j)}}(F)(Y_{t-1}^{(j)}) = Y_t^{(j)}, \tag{11}$$

irrespective of F. When $\kappa(x, y) \neq p(x, y)$, the relation (11) no longer holds with certainty.

6. The Autocatalytic Coupler

In this section, we briefly describe the case $k = 1$ of a single catalyst $Y_t = Y_t^{(1)}$ with $\kappa(x, y) = p(x, y)$. As stated in the previous section, this catalyst never frees captured particles. Instead, the ordinary particles in Figure 5 are swept up over time by the process Y_t until no independent ones are left. Since Y_t obeys the same transition probabilities $p(x, y)$, it 'sweeps in the right places'.

Proposition 3. *Let $\varphi_{s,t}$ be a flow associated with a positive recurrent, aperiodic transition density $p(x, y)$, and let Y_t be an independent catalyst. For any $x \in E$, let $X_t = \mathcal{C}_Y(\varphi)_{0,t}(x)$ and $T = \inf\{s : X_s = Y_s\}$ be the sweep up time of particle x. Then $\mathbb{P}(T < \infty) = 1$.*

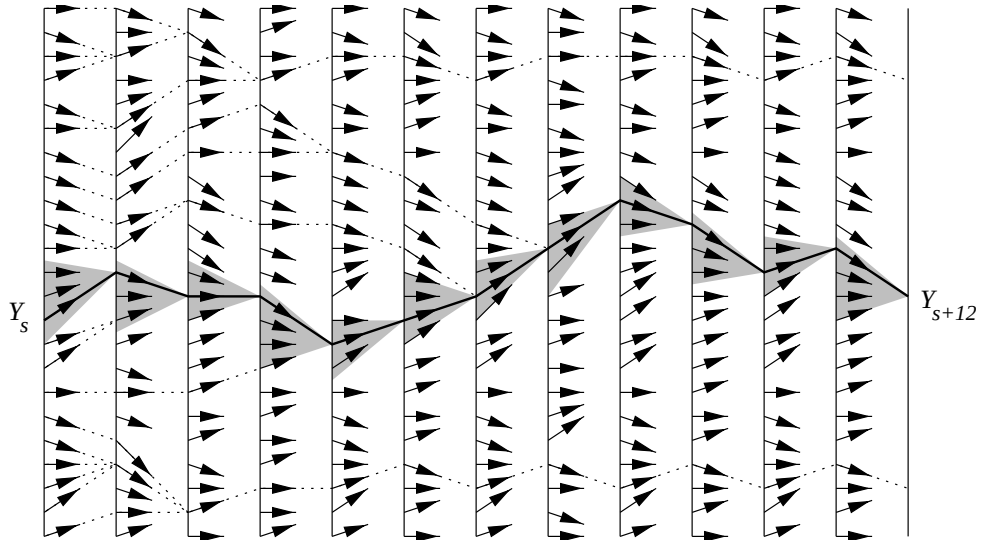

FIGURE 6. The flow in Figure 2 modified to include a single catalyst sweeping up particles.

Proof. Let C be a set such that $\pi(C) > 0$ and $y \mapsto \inf_{x \in C} p(x, y)$ is not identically zero nor infinite. We can thus find a set D with

$$\inf_{\substack{x_1, y_1 \in C \\ x_2, y_2 \in D}} \frac{p(x_1, y_2)p(y1, x_2)}{p(x_1, x_2)p(y_1, y_2)} > \epsilon \text{ say.}$$

Suppose that $\mathbb{P}(T = \infty) > 0$. By positive recurrence, aperiodicity and independence, both X_t and Y_t must enter C infinitely often on $\{T = \infty\}$. Each time this occurs, both processes may enter D immediately afterwards with some small probability $\alpha > 0$. Unpon entering D, coupling occurs independently with probability at least ϵ. This contradicts the assumption $T = \infty$. Hence $\mathbb{P}(T < \infty) = 1$ necessarily. □

As a consequence of introducing the catalyst Y_t into the flow, the process $X_t = \mathcal{C}_Y(\varphi)_{0,t}(x)$ is no longer Markov. This is however not a major issue since all asymptotic ergodicity properties are preserved.

Theorem 4. *Let $X_t = \mathcal{C}_Y(\varphi)_{0,t}(x)$, where $\varphi_{s,t}$ is defined as in* (3)*, and $\mathcal{C}_Y(\varphi)_{s,t}$ is the result of* (10) *with a single catalyst $Y_t = Y_t^{(1)}$. Then*

(1) $\mathbb{P}(X_s \in dy) = P^s(x, dy), s = 0, 1, 2, \ldots$

(2) *If $p(x,y)$ is positive recurrent, with invariant distribution π, then*

$$\lim_{t\to\infty} \frac{1}{t}\sum_{s=1}^{t} f(X_s) = \int f d\pi \quad \mathbb{P}-a.s.$$

for all integrable functions f.

(3) *If $p(x,y)$ admits a Central Limit Theorem, then so does the realization X_t.*

Proof. (1) We have, taking $q(y) = p(z,y)$ in Lemma 2, that

$$\begin{aligned}\mathbb{P}(X_1 \in dy) &= \int \mathbb{P}(X_1 \in dy \,|\, Y_0 = z)\mathbb{P}(Y_0 \in dz) \\ &= \int P(x,dy)\mathbb{P}(Y_0 \in dz) = P(x,dy).\end{aligned}$$

Hence by induction,

$$\begin{aligned}\mathbb{P}(X_{s+1} \in dy) &= \int \mathbb{P}(X_{s+1} \in dy \,|\, X_s = z_1, Y_{s-1} = z_2)\mathbb{P}(X_s \in dz_1 \,|\, Y_{s-1} \in z_2)\mathbb{P}(Y_{s-1} \in dz_2) \\ &= \int P(z_1,dy)P^s(x,dz_1)\mathbb{P}(Y_{s-1} \in dz_2) \\ &= P^{s+1}(x,dy).\end{aligned}$$

(2,3) Write

$$\begin{aligned}\sum_{s=1}^{t} f(X_s) &= \sum_{s=1}^{T} f(X_s) + \sum_{s=T+1}^{t} f(X_s) \\ &= \sum_{s=1}^{T}\big[f\circ\varphi_{0,s}(x) - f(Y_s)\big] + \sum_{s=1}^{t} f(Y_s).\end{aligned}$$

Since $\mathbb{P}(T < infty) = 1$, the first term on the right is a finite sum. We deduce immediately that

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} f(X_s) = \lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} f(Y_s) = \int f d\pi,$$

and similarly

$$\lim \frac{1}{\sqrt{t}}\sum_{s=1}^{t}\Big[f(X_s) - \int f d\pi\Big] = \lim \frac{1}{\sqrt{t}}\sum_{s=1}^{t}\Big[f(Y_s) - \int f d\pi\Big] = \mathcal{N}(0,\sigma^2),$$

where σ^2 is the asymptotic variance. □

These results generalize immediately by induction to the case of k independent catalysts, since $\mathcal{C}_{Y^{(1)},\ldots,Y^{(k)}}(\varphi)_{s,t} = \mathcal{C}_{Y^{(k)}}(\cdots\mathcal{C}_{Y^{(1)}}(\varphi)\cdots)_{s,t}$.

To end this section, we give an elementary proof of ergodicity for positive recurrent aperiodic Markov chains with a density.

Theorem 5. *Let X_t and Y_t be two independent Markov chains with positive recurrent, aperiodic transition density $p(x,y)$ and stationary distribution π. Assume that $Y_0 \sim \pi$, then*

$$\|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} \leq \mathbb{E}\prod_{s=0}^{t}\left(1 - \frac{p(X_s, Y_{s+1})p(Y_s, X_{s+1})}{p(X_s, X_{s+1})p(Y_s, Y_{s+1})}\right)_+, \tag{12}$$

and the right hand side tends to zero as t tends to infinity.

Proof. Without loss of generality, suppose that $X_t = \varphi_{0,t}(X_0)$ where the flow $\varphi_{s,t}$ is independent of Y. Both processes exist as part of the same flow $\mathcal{C}_Y(\varphi)_{s,t}$, since $Y_t = \mathcal{C}_Y(\varphi)_{0,t}(Y_0)$ and $X_t = \mathcal{C}_Y(\varphi)_{0,t}(X_0)$ on $\{T > t\}$, where $T = \inf\{s : X_s = Y_s\}$. Now obviously

$$\mathbb{P}(T > t) = \mathbb{P}(X_s \neq Y_s \text{ for all } s \leq t),$$

from which we deduce that

$$\begin{aligned}\mathbb{P}(T > t) &= \mathbb{P}\left(\frac{p(X_s, Y_{s+1})p(Y_s, X_{s+1})}{p(X_s, X_{s+1})p(Y_s, Y_{s+1})} \leq \xi_s \text{ for all } s \leq t\right) \\ &= \mathbb{E}\prod_{s=0}^{t}\left(1 - \frac{p(X_s, Y_{s+1})p(Y_s, X_{s+1})}{p(X_s, X_{s+1})p(Y_s, Y_{s+1})}\right)_+,\end{aligned}$$

and the coupling inequality (1) gives (12). □

7. Funnelweb Couplers

Recall that one motivation for introducing the autocatalytic coupler is to help the catalysts to 'find where to go'. Since the catalysts are independent, no concerted effort to bring particles together is therefore achieved. The idea of introducing independent catalysts into the flow can be taken a step further. Separate paths can be constructed in such a way that coalescence is achieved with certainty in a set number of simulation steps. To make this work, we build them from the other end: one final state is reached by several previous states, each in turn arrived at from several other states, etc. This naturally introduces a branching process. If we are not concerned about ensuring the coalescence of two specific initial states over time, we can build such a branching process in a straightforward manner.

Suppose that $p(x,y)$ has a stationary distribution density $\pi(x)$. The reverse density $\widetilde{p}(x,y) = p(y,x)\pi(y)/\pi(x)$ defines a Markov chain which, in equilibrium, can be interpreted using Bayes' theorem as the time reverse of a stationary chain

whose transitions are governed by $p(x,y)$. We may therefore construct plausible $p(x,y)$-paths of any length by generating $\widetilde{p}(x,y)$-paths backwards in time.

Fix a number n of simulation steps, and generate a single random variable $Y_{n,1}$. Let k be a desired number of offspring, and define random variables $Y_{n-1,1},\ldots,Y_{n-1,k}$ independently by $Y_{n-1,j} \sim \widetilde{p}(Y_{n,1},\cdot)$. Each of these variables in turn produces k offspring in the same manner, arriving at a grand total of k^2 offspring. We list this generation simply as $Y_{n-2,1},\ldots.Y_{n-2,k^2}$. Repeating the construction recursively, we

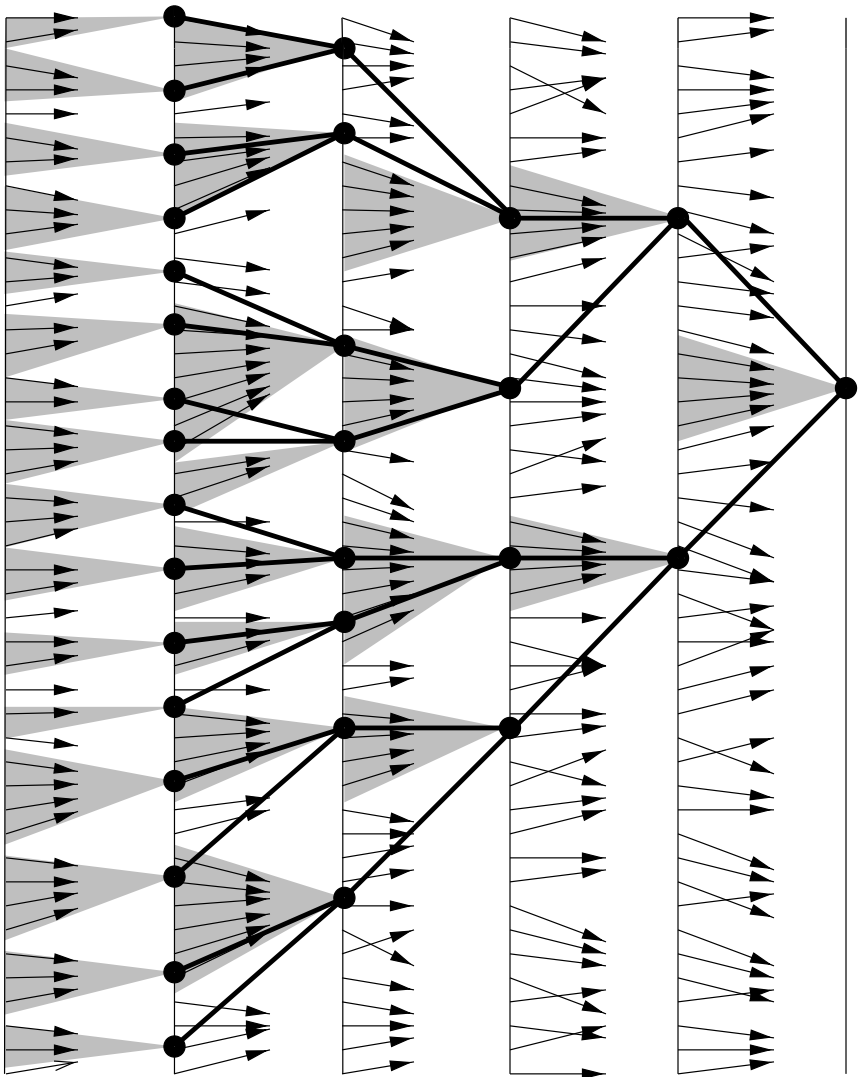

FIGURE 7. A Funnelweb Coupler with $k = 2$ and $n = 5$.

arrive eventually at the $(n-1)$th generation, consisting of k^{n-1} offspring denoted $Y_{1,1}, \dots, Y_{1,k^{n-1}}$.

Funnelweb Coupler: Modifies a section of flow $\varphi_{s,s+n}$ to induce coupling of exponentially many paths.

Requires: A section of a flow $\varphi_{s,s+n}$ built from maps $F_{s+1}, \dots, F_{s+n}$ as in (3). A 'forward' transition density $p(x,y)$ and a 'reverse' transition density $\widetilde{p}(x,y)$.

Parameter: $k = 2, 3, \dots$ representing the number of offspring per generation.

Returns: A flow $\mathcal{C}_{\mathcal{G}}(\varphi)_{s,s+n} = \mathcal{C}_{\mathcal{G}_0}(F_{s+n}) \circ \cdots \circ \mathcal{C}_{\mathcal{G}_{n-1}}(F_{s+1})$.

Method: Let $\mathcal{G}_0 = \{Y_{n,1}\}, \mathcal{G}_1 = \{Y_{n-1,1}, \dots, Y_{n-1,k}\}, \dots, \mathcal{G}_{n-1} = \{Y_{1,1}, \dots, Y_{1,k^{n-1}}\}$ represent the successive generations in a branching process with k offspring per generation such that, for any generation $j = 1, \dots, n-1$, each $Y \in \mathcal{G}_j$ has exactly one ancestor $Y^* \in \mathcal{G}_{j-1}$ such that $Y \sim \widetilde{p}(Y^*, \cdot)$ and given the previous generation, all offspring are independent. For $Y \in \mathcal{G}_j$ and F any map, we set

$$\mathcal{C}_Y(F)(x) = \begin{cases} Y & \text{if } \frac{p(x,Y)\widetilde{p}(Y^*,F(x))}{p(x,F(x))\widetilde{p}(Y^*,Y)} > \xi \\ F(x) & \text{otherwise,} \end{cases} \tag{13}$$

and $\mathcal{C}_{\mathcal{G}_j}(F) = \mathcal{C}_{Y_{j,1},\dots,Y_{j,kj}}(F) = \mathcal{C}_{Y_{j,kj}}(\cdots \mathcal{C}_{Y_{j,1}}(F) \cdots)$.

An illustration of the effect of the Funnelweb Coupler can be seen in Figure 7. It is clear that the processes $X_t = \mathcal{C}_{\mathcal{G}}(\varphi)_{0,t}(x)$ are not Markov chains at all. However, the marginal probabilities are still preserved, since each original map F_t from the original flow is modified by proposals independent of it (and of each other, given the previous generation). We prove this below.

Lemma 6. *Applying the Funnelweb Coupler to a section of flow $\varphi_{s,s+n}$ preserves the marginal distribution over n steps, i.e.*

$$\mathbb{P}\big(\mathcal{C}_{\mathcal{G}}(\varphi)_{s,s+n}(x) \in dy\big) = P^n(x, dy).$$

Proof. Clearly $\mathbb{P}\big(\mathcal{C}_{\mathcal{G}}(F_{s+n})(x) \in dy \,|\, \mathcal{G}_0\big) = P(x, dy)$ by Lemma 2 applied k times. We shall use induction with

$$\mathbb{P}(\mathcal{C}_{\mathcal{G}}(\varphi)_{s+n-k,s+n}(x) \,|\, \mathcal{G}_k, \dots, \mathcal{G}_0) = P^k(x, dy).$$

By hypothesis, given the generations $\mathcal{G}_k, \ldots, \mathcal{G}_0$, the variables $Y_{k+1,j}$ are independent of each other and of $F_{s+n-k-1}$. Hence Lemma 2 applied repeatedly gives

$$\mathbb{P}\big(\mathcal{C}_{\mathcal{G}_{k+1}}(F_{s+n-k-1})(x) \in dy \,|\, \mathcal{G}_k, \ldots, \mathcal{G}_0\big) = P(x, dy).$$

Moreover, given $\mathcal{G}_k, \ldots, \mathcal{G}_0$, the maps $\mathcal{C}_{\mathcal{G}_{k+1}}(F_{s+n-k-1})$ and $\mathcal{C}_{\mathcal{G}}(\varphi)_{s+n-k,s+n}$ are clearly independent, which gives

$$\begin{aligned}\mathbb{P}(\mathcal{C}_{\mathcal{G}}(\varphi)_{s+n-k-1,s+n}(x) \in dy) &= \int \mathbb{P}(\mathcal{C}_{\mathcal{G}}(\varphi)_{s+n-k,s+n}(z) \,|\, \mathcal{G}_k, \ldots, \mathcal{G}_0) P(x, dz) \\ &= \int P^k(z, dy) P(x, dz) = P^{k+1}(x, dy).\end{aligned}$$

This completes the induction and the proof. □

The Funnelweb Coupler allows the construction of compicated n-step flows whose transitions are governed by the n-step kernel $P^n(x, dy)$. Consequently, all asymptotic results (Laws of Large Numbers and Central LImit Theorems) apply to the Markov chains $X_{nt}(x) = \mathcal{C}_{\mathcal{G}^{(t)}}(\varphi)_{n(t-1),nt} \circ \cdots \circ \mathcal{C}_{\mathcal{G}^{(1)}}(\varphi)_{0,t}(x)$ where $x \in E$ and $\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(t)}$ are independent n-generation branching processes.

Note also that we cannot make the web '100% sticky': particles which have been captured by one generation may well be freed at the next simulation step. Intuitively, our choice $\widetilde{p}(x, y) = p(y, x)\pi(y)/\pi(x)$ ought to maximize the stickiness, because the paths proposed are typical (see Figure 7).

Acknowledgement. This research was supported by the European Union TMR network ERB-FMRX-CT96-0096 on "Spatial and Computational Statistics".

References

[1] Lindvall, T. (1992) *Lectures on the Coupling Method.* Wiley Interscience.

[2] Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability.*

[3] Nummelin, E. (1978) *what's his book called again?*

[4] Murdoch, D.J. and Green, P.J. (1998) *Exact Sampling from a Continuous State Space.*

Department of Mathematical Sciences,, Aalborg University,, Fredrik Bajers Vej 7E, DK-9220 Aalborg Ø,, Denmark

E-mail address: lbreyer@math.auc.dk

Department of Mathematics, Lancaster University, UK

E-mail address: g.o.roberts@lancaster.ac.uk