Lectures on Stochastic Stability

Sergey FOSS

Heriot-Watt University

Lecture 6

Coupling again: the Renovation Theory

1 Introduction

A large fraction of the applied probability literature is devoted to the study of existence and uniqueness of a stationary solution to a stochastic dynamical system, as well as convergence toward such a stationary solution. Examples abound in several application areas, such queueing theory, stochastic control, simulation algorithms, etc. Often, the model studied possesses a Markovian property, in which case several classical tools are available. In the absence of Markovian property, one has few tools to rely on, in general. The concept of *renovating event* was introduced by Borovkov (1978) in an attempt to produce general conditions for a strong type of convergence of a stochastic process, satisfying a stochastic recursion, to a stationary process. Other general conditions for existence/uniqueness questions can be found, e.g., in Anantharam and Konstantopoulos (1997,1999) and in Diaconis and Freedman (1999). The so-called *renovation theory* or *method of renovating events* has a flavor different from the aforementioned papers, in that it leaves quite a bit of freedom in the choice of a renovating event, which is what makes it often hard to apply: some ingenuity is required in constructing renovating events. Nevertheless, renovation theory has found several applications, especially in queueing-type problems [see, e.g., Brandt *et al.* (1990) and Baccelli and Brémaud (1994) for a variety of models and techniques in this area].

Renovation theory is stated for a discrete-time "stochastic recursive processes", i.e., random sequences $\{X_n\}$ defined by recursive relations of the form

$$X_{n+1} = f(X_n, \xi_{n+1}),$$

where f is appropriately measurable function, and $\{\xi_n\}$ a stationary random sequence. A standard example of such a recursion is a sequence of Keifer-Wolfowitz vectors in 2-server first-come-first-served queue which is defined by $\mathbf{W}_0 = (0, 0)$ and, for $n = 0, 1, \ldots$,

$$\mathbf{W}_{n+1} = R(\mathbf{W}_n + \mathbf{e}_1\sigma_n - \mathbf{i}t_n)^+$$

where $\mathbf{e}_1 = (1, 0)$, $\mathbf{i} = (1, 1)$, $R(x_1, x_2) = (x_{(1)}, x_{(2)})$ is the non-decreasing ordering, and $(x_1, x_2)^+ = (\max(0, x_1), \max(0, x_2))$. Here $\{t_n\}$ are inter-arrival and $\{\sigma_n\}$ are service times.

We take a fresh look at the renovation theory and formulate it for processes that do not necessarily obey stochastic recursions. We give a self-consistent overview of (extended) renovation theory, and strong coupling notions; second, we shed some light into the so-called *coupling from the past* property, which has drawn quite a bit of attention recently, especially in connection to the Propp-Wilson algorithm [see Propp and Wilson (1996)] for perfect simulation (alternative terminology:

exact sampling). We do this, by defining strong forward and backward coupling times. We also pay particular attention to a special type of backward coupling times, those that we call *verifiable times*: they form precisely the class of those times that can be simulated. Verifiable backward times exist, e.g., for irreducible Markov chains with finite state space (and hence exact sampling from the stationary distribution is possible), but they also exist in other models, such as the ones presented in the second part of the paper.

2 Strong coupling notions

We start with revisiting the notions of coupling (and coupling convergence) that we need. For general notions of coupling we refer to the monographs of Lindvall (1992) and Thorisson (2000). For the strong coupling notions of this paper, we refer to Borovkov and Foss (1992) and to Borovkov (1998).

Consider a sequence of random variables $\{X_n, n \in \mathbb{Z}\}$ defined on a probability space $(\Omega, \mathscr{F}, P)$ and taking values in another measurable space $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$. We study various ways according to which X couples with another stationary process $\{\tilde{X}_n, n \in \mathbb{Z}\}$. We push the stationarity structure into the probability space itself, by assuming the existence of a flow (i.e., a measurable bijection) $\theta : \Omega \to \Omega$ that leaves the probability measure P invariant, i.e., $\mathbf{P}(\theta^k A) = \mathbf{P}(A)$, for all $A \in \mathscr{F}$, and all $k \in \mathbb{Z}$. In this setup, a stationary process $\{\tilde{X}_n\}$ is, by definition, a θ-compatible process in the sense that $\tilde{X}_{n+1} = \tilde{X}_n \circ \theta$ for all $n \in \mathbb{Z}$. Likewise, a sequence of events $\{A_n\}$ is stationary iff their indicator functions $\{\mathbf{1}_{A_n}\}$ is stationary. Note that, in this case, $\mathbf{1}_{A_n} \circ \theta = \mathbf{1}_{\theta^{-1} A_n} = \mathbf{1}_{A_{n+1}}$ for all $n \in \mathbb{Z}$. In order to avoid technicalities, we assume that the σ-algebra $\mathscr{B}_{\mathscr{X}}$ is countably generated. The same assumption, without special notice, will be made for all σ-algebras below.

We next present three notions of coupling: *simple coupling*, *strong (forward) coupling* and *backward coupling*. To each of these three notions there corresponds a type of convergence. These are called *c-convergence*, *sc-convergence*, and *bc-convergence*, respectively. The definitions below are somewhat formal by choice: there is often a danger of confusion between these notions. To guide the reader, we first present an informal discussion. Simple coupling between two processes (one of which is usually stationary) refers to the fact that the two processes are a.s. identical, eventually. To define strong (forward) coupling, consider the family of processes that are derived from X "started from all possible initial states at time 0". To explain what the phrase in quotes means in a non-Markovian setup, place the origin of time at the negative index $-m$, and run the process forward till a random state at time 0 is reached: this is the process X^{-m} formally defined in (2). Strong coupling requires the existence of a finite random time $\sigma \geq 0$ such that all these processes are identical after σ. Backward coupling is–in a sense–the dual of strong coupling: instead of fixing the starting time (time 0) and waiting till the random time σ, we play a similar game with a random starting time (time $-\tau \leq 0$) and wait till coupling takes place at a fixed time (time 0). That is, backward coupling takes place if there is a finite random time $-\tau \leq 0$ such that all the processes started at times prior to $-\tau$ are coupled forever after time 0. The main theorem of this section (Theorem 1) says that strong (forward) coupling and backward coupling are equivalent, whereas an example (Example 1) shows that they are both strictly stronger than simple coupling.

We first consider simple coupling. Note that our definitions are more general than usual because we do not necessarily assume that the processes are solutions of stochastic recursions.

Definition 1 (simple coupling)**.**
1) The minimal coupling time between X and $\tilde{X}$ is defined by

$$\nu = \inf\{n \geq 0 : \ \forall\, k \geq n \ \ X_k = \tilde{X}_k\}.$$

2) More generally, a random variable ν' is said to be a coupling time between X and $\tilde{X}$ iff[1]

$$X_n = \tilde{X}_n, \quad \text{a.s. on } \{n \geq \nu'\}.$$

3) We say that X coupling-converges (or c-converges) to $\tilde{X}$ iff $\nu < \infty$, a.s., or, equivalently, iff $\nu' < \infty$, a.s., for some coupling time ν'.

Notice that the reason we call ν "minimal" is because (i) it is a coupling time, and (ii) any random variable ν' such that $\nu' \geq \nu$, a.s., is also a coupling time.

Proposition 1 (c-convergence criterion)**.** *X c-converges to $\tilde{X}$ iff*

$$\mathbf{P}(\liminf_{n\to\infty}\{X_n = \tilde{X}_n\}) = 1.$$

Proof. It follows from the equality

$$\{\nu < \infty\} = \bigcup_{n\geq 0}\bigcap_{k\geq n}\{X_k = \tilde{X}_k\},$$

the right hand side of which is the event $\liminf_{n\to\infty}\{X_n = \tilde{X}_n\}$. □

It is clear that c-convergence implies convergence in total variation, i.e.,

$$\lim_{n\to\infty}\sup_{B\in\mathscr{B}_{\mathscr{X}}^{\infty}} |\mathbf{P}((X_n, X_{n+1}, \ldots) \in B) - \mathbf{P}((\tilde{X}_n, \tilde{X}_{n+1}, \ldots) \in B)| = 0, \quad (1)$$

simply because the left hand side is dominated by $\mathbf{P}(\nu \geq n)$ for all n. In fact, the converse is also true, viz., (1) implies c-convergence (see Thorisson (2000), Theorem 9.4). Thus, c-convergence is a very strong notion of convergence, but not the strongest one that we are going to deal with in this paper.

The process $\tilde{X}$ in (1) will be referred to as the *stationary version* of X. Note that the terminology is slightly non-standard because, directly from the definition, if such a $\tilde{X}$ exists, it is automatically unique (due to coupling). The term is usually defined for stochastic recursive sequences (SRS). To avoid confusion, we talk about a *stationary solution* of an SRS, which may not be unique. See Section 4 for further discussion.

A comprehensive treatment of the notions of coupling, as well as the basic theorems and applications can be found in the paper of Borovkov and Foss (1992), for the special case of processes which form stochastic recursive sequences. For the purposes of our paper, we need to formulate some of these results beyond the SRS realm, and this is done below.

It is implicitly assumed above (see the definition of ν) that 0 is the "origin of time". This is, of course, totally arbitrary. We now introduce the notation

$$X_n^{-m} := X_{m+n}\circ\theta^{-m}, \quad m \geq 0,\ n \geq -m,$$

and consider the family of processes

$$X^{-m} := (X_0^{-m}, X_1^{-m}, \ldots), \quad m = 0, 1, \ldots \quad (2)$$

and the minimal coupling time $\sigma(m)$ of X^{-m} with $\tilde{X}$. The definition becomes clearer when X itself is a SRS (see Section 4 below).

[1]"B a.s. on A" means $\mathbf{P}(A - B) = 0$.

Definition 2 (strong coupling)**.**
1) The minimal strong coupling time between X *and* $\tilde{X}$ *is defined by*

$$\sigma = \sup_{m\geq 0} \sigma(m), \quad \textit{where}$$
$$\sigma(m) = \inf\{n \geq 0 : \ \forall\, k \geq n \ \ X_k^{-m} = \tilde{X}_k\}.$$

2) More generally, a random variable σ' *is said to be a strong coupling time (or sc-time) between* X *and* $\tilde{X}$ *iff*

$$\tilde{X}_n = X_n^0 = X_n^{-1} = X_n^{-2} = \cdots, \qquad \textit{a.s. on } \{n \geq \sigma'\}.$$

3) We say that $\{X_n\}$ *strong-coupling-converges (or sc-converges) to* $\{\tilde{X}_n\}$ *iff* $\sigma < \infty$*, a.s.*

Again, it is clear that the minimal strong coupling time σ is a strong coupling time, and that any σ' such that $\sigma' \geq \sigma$, a.s., is also a strong coupling time.

Even though strong coupling is formulated by means of two processes, X, and a stationary $\tilde{X}$, we will see that the latter is not needed in the definition.

Example 1 (Borovkov and Foss (1992))**.** We now give an example to show the difference between coupling and strong coupling. Let $\{\xi_n, n \in \mathbb{Z}\}$ be an i.i.d. sequence of random variables with values in $\mathbb{Z}_+$ such that $\mathbf{E}\xi_0 = \infty$. Let

$$X_n = (\xi_0 - n)^+, \quad \tilde{X}_n = 0, \quad n \in \mathbb{Z}.$$

The minimal coupling time between $(X_n, n \geq 0)$ and $(\tilde{X}_n, n \geq 0)$ is $\nu = \xi_0 < \infty$, a.s. Hence $\tilde{X}$ is the stationary version of X. Since

$$X_n^{-m} = X_{m+n} \circ \theta^{-m} = (\xi_{-m} - (m+n))^+,$$

the minimal coupling time between $(X_n^{-m}, n \geq 0)$ and $(\tilde{X}_n, n \geq 0)$ is $\sigma(m) = (\xi_{-m} - m)^+$. Hence the minimal *strong* coupling time between X and $\tilde{X}$ is $\sigma = \sup_{m\geq 0} \sigma(m)$. But $\mathbf{P}(\sigma \leq n) = \mathbf{P}(\forall m \geq 0 \ \ \xi_m - m \leq n) = \prod_{m\geq 0} \mathbf{P}(\xi_0 \leq m+n)$, and, since $\sum_{j\geq 0} \mathbf{P}(\xi_0 > j) = \infty$, we have that the latter infinite product is zero, i.e., $\sigma = +\infty$, a.s. So, even though X couples with $\tilde{X}$, it does not couple strongly.

Proposition 2 (sc-convergence criterion)**.** X *sc-converges to* $\tilde{X}$ *iff*

$$P\Big(\liminf_{n\to\infty} \bigcap_{m\geq 0} \{\tilde{X}_n = X_n^{-m}\} \Big) = 1.$$

Proof. It follows from the definition of σ that

$$\begin{aligned}
\{\sigma < \infty\} &= \bigcup_{n\geq 0} \{\sigma \leq n\} = \bigcup_{n\geq 0} \bigcap_{m\geq 0} \{\sigma(m) \leq n\} \\
&= \bigcup_{n\geq 0} \bigcap_{m\geq 0} \bigcap_{k\geq n} \{\tilde{X}_k = X_k^{-m}\} = \bigcup_{n\geq 0} \bigcap_{k\geq n} \bigcap_{m\geq 0} \{\tilde{X}_k = X_k^{-m}\} \\
&= \liminf_{n\to\infty} \bigcap_{m\geq 0} \{\tilde{X}_n = X_n^{-m}\},
\end{aligned}$$

and this proves the claim. □

The so-called backward coupling [see Foss (1983), Borovkov and Foss (1992) for this notion in the case of SRS] is introduced next. This does not require the stationary process $\tilde{X}$ for its definition. Rather, the stationary process is constructed once backward coupling takes place. Even though the notion appears to be quite strong, it is not infrequent in applications.

Definition 3 (backward coupling)**.**
1) The minimal backward coupling time for the random sequence $\{X_n, n \in \mathbb{Z}\}$ *is defined by*

$$\tau = \sup_{m \geq 0} \tau(m), \quad \text{where}$$
$$\tau(m) = \inf\{n \geq 0 : \ \forall\, k \geq 0 \ \ X_m^{-n} = X_m^{-(n+k)}\}.$$

2) More generally, we say that τ' *is a backward coupling time (or bc-time) for* X *iff*

$$\forall m \geq 0 \quad X_m^{-t} = X_m^{-(t+1)} = X_m^{-(t+2)} = \cdots, \quad \text{a.s. on } \{t \geq \tau'\}.$$

3) We say that $\{X_n\}$ *backward-coupling converges (or bc-converges) iff* $\tau < \infty$*, a.s.*

Note that τ is a backward coupling time and that any τ' such that $\tau' \geq \tau$, a.s., is a backward coupling time. We next present the equivalence theorem between backward and forward coupling.

Theorem 1 (coupling equivalence)**.** *Let* τ *be the minimal backward coupling time for* X*. There is a stationary process* $\tilde{X}$ *such that the strong coupling time* σ *between* X *and* $\tilde{X}$ *has the same distribution as* τ *on* $\mathbb{Z}_+ \cup \{+\infty\}$*. Furthermore, if* $\tau < \infty$ *a.s., then* $\tilde{X}$ *is the stationary version of* X*.*

Proof. Using the definition of τ, we write

$$\begin{aligned}\{\tau < \infty\} &= \bigcup_{n \geq 0} \{\tau \leq n\} = \bigcup_{n \geq 0} \bigcap_{m \geq 0} \{\tau(m) \leq n\} \\ &= \bigcup_{n \geq 0} \bigcap_{m \geq 0} \bigcap_{\ell \geq n} \{X_m^{-n} = X_m^{-\ell}\} = \bigcup_{n \geq 0} \bigcap_{\ell \geq n} \bigcap_{m \geq 0} \{X_m^{-n} = X_m^{-\ell}\}. \end{aligned} \tag{3}$$

Consider, as in (2), the process $X^{-n} = (X_0^{-n}, X_1^{-n}, X_2^{-n}, \ldots)$, with values in $\mathscr{X}^{\mathbb{Z}_+}$. Using this notation, (3) can be written as

$$\begin{aligned}\{\tau < \infty\} &= \bigcup_{n \geq 0} \bigcap_{\ell \geq n} \{X^{-n} = X^{-\ell}\} \\ &= \{\exists n \geq 0 \ X^{-n} = X^{-(n+1)} = X^{-(n+2)} = \cdots\}\end{aligned}$$

Thus, on the event $\{\tau < \infty\}$, the random sequence X^{-n} is equal to some fixed random element of $\mathscr{X}^{\mathbb{Z}_+}$ for all large n (it is eventually a constant sequence). Let $\tilde{X} = (\tilde{X}_0, \tilde{X}_1, \ldots)$ be this random element; it is defined on $\{\tau < \infty\}$. Let ∂ be an arbitrary fixed member of $\mathscr{X}^{\mathbb{Z}_+}$ and define $\tilde{X} \equiv \partial$ outside $\{\tau < \infty\}$. Since the event $\{\tau < \infty\}$ is a.s. invariant under θ^n, for all $n \in \mathbb{Z}$, we obtain that $\tilde{X}$ is a stationary process. Let σ be the strong coupling time between X and $\tilde{X}$. It is easy to see that, for all $n \geq 0$,

$$\{\sigma \circ \theta^{-n} \leq n\} = \bigcap_{\ell \geq n} \{\tilde{X} = X^{-\ell}\} = \{\tau \leq n\}. \tag{4}$$

Indeed, on one hand, from the definition of τ, we have

$$\{\tau \leq n\} = \bigcap_{\ell \geq n} \{X^{-n} = X^{-\ell}\}.$$

Now, using the $\tilde{X}$ we just defined we can write this as

$$\{\tau \leq n\} = \bigcap_{\ell \geq n} \{\tilde{X} = X^{-\ell}\}. \tag{5}$$

On the other hand, from the definition of σ (that is, the strong coupling time between X and $\tilde{X}$), we have:

$$\{\sigma \le n\} = \bigcap_{k\ge 0} \{\tilde{X}_{n+k} = X^0_{n+k} = X^{-1}_{n+k} = X^{-2}_{n+k} = \cdots\}.$$

Applying a shifting operation on both sides,

$$\{\sigma \circ \theta^{-n} \le n\} = \bigcap_{k\ge 0} \{\tilde{X}_k = X^n_k = X^{n-1}_k = X^{n-2}_k = \cdots\}. \tag{6}$$

The events on the right hand sides of (5) and (6) are identical. Hence (4) holds for all n, and thus $\mathbf{P}(\tau \le n) = \mathbf{P}(\sigma \circ \theta^n \le n) = \mathbf{P}(\sigma \le n)$, for all n. Finally, if $\mathbf{P}(\tau < \infty) = 1$ then $\mathbf{P}(\sigma < \infty) = 1$, and this means that X sc-converges to $\tilde{X}$. In particular, we have convergence in total variation, and so $\tilde{X}$ is the stationary version of X. □

Corollary 1. *The following are equivalent:*
1) X bc-converges.
2) $\lim_{n\to\infty} \mathbf{P}(\forall m \ge 0\ X^{-n}_m = X^{-(n+1)}_m = X^{-(n+2)}_m = \cdots) = 1$.
3) X sc-converges.
4) $\lim_{n\to\infty} \mathbf{P}(\forall k \ge 0\ X^0_{n+k} = X^{-1}_{n+k} = X^{-2}_{n+k} = \cdots) = 1$.

We can view any of the equivalent statements of Corollary 1 as an "intrinsic criterion" for the existence the stationary version of X.

Corollary 2. *Suppose that X bc-converges and let τ be the minimal backward coupling time. Let $\tilde{X}_0 = X_\tau \circ \theta^{-\tau}$. Then $\tilde{X}_n = \tilde{X}_0 \circ \theta^n$ is the stationary version of X. Furthermore, if τ' is any a.s. finite backward coupling time then $X_\tau \circ \theta^{-\tau} = X'_\tau \circ \theta^{-\tau'}$, a.s.*

Proof. Let $\tilde{X}$ be the stationary version of X. It follows, from the construction of $\tilde{X}$ in the proof of Theorem 1, that

$$(X^{-t}_0, X^{-t}_1, X^{-t}_2, \ldots) = (\tilde{X}_0, \tilde{X}_1, \tilde{X}_2, \ldots), \quad \text{a.s. on } \{t \ge \tau\}. \tag{7}$$

Thus, in particular, $\tilde{X}_0 = X^{-t}_0 = X_t \circ \theta^{-t}$, a.s. on $\{t \ge \tau\}$. Since $\mathbf{P}(\tau < \infty) = 1$, it follows that $\tilde{X}_0 = X_\tau \circ \theta^{-\tau}$, a.s. Now, if τ' is any backward coupling time, then (7) is true with τ' in place of τ; and if $\tau' < \infty$, a.s., then, as above, we conclude that $\tilde{X}_0 = X_{\tau'} \circ \theta^{-\tau'}$. □

3 The concept of verifiability and perfect simulation

One application of the theory is the simulation of stochastic systems. If we could sample the process at a bc-time, then would actually be simulating its stationary version. This is particularly useful in *Markov Chain Monte Carlo* applications. Recently, Propp and Wilson (1996) used the so-called *perfect simulation method* for the simulation of the invariant measure of a Markov chain. The method is actually based on sampling at a bc-time. To do so, however, one must be able to generate a bc-time from a finite history of the process. In general, this may not be possible because, even in the case when suitable renovation events can be found, they may depend on the entire history of the process.

We are thus led to the concept of a *verifiable time*. Its definition, given below, requires introducing a family of σ-fields $\{\mathscr{G}_{-j,m}, -j \le 0 \le m\}$, such that $\mathscr{G}_{-j,m}$ increases if j or m increases. We call this simply an increasing family of σ-fields. For fixed m, a *backwards stopping time* $\tau \ge 0$ with respect to $\mathscr{G}_{\cdot,m}$ means a stopping time with respect to the first index, i.e., $\{\tau \le j\} \in \mathscr{G}_{-j,m}$ for all $j \geq 0$. In this case, the σ-field $\mathscr{G}_{-\tau,m}$ contains all events A such that $A \cap \{\tau \leq j\} \in \mathscr{G}_{-j,m}$, for all $j \geq 0$.

Definition 4 (verifiable time)**.** *An a.s. finite nonnegative random time β is said to be verifiable with respect to an increasing family of σ-fields $\{\mathscr{G}_{-j,m}, -j \leq 0 \leq m\}$, if there exists a sequence of random times $\{\beta(m), m \geq 0\}$, with $\beta(m)$ being a backwards $\mathscr{G}_{\cdot,m}$–stopping time for all m, such that:*

(i) $\beta = \sup_{m \geq 0} \beta(m)$,

(ii) *For all $m \geq 0$, $X_m^{-n} = X_m^{-(n+i)}$ for all $i \geq 0$, a.s. on $\{n \geq \beta(m)\}$,*

(iii) *For all $m \geq 0$, the random variable $X_m^{-\beta(m)}$ is $\mathscr{G}_{-\beta(m),m}$–measurable.*

Some comments: First, observe that if β is any backwards coupling time, then it is always possible to find $\beta(m)$ such that (i) and (ii) above hold. The additional thing here is that the $\beta(m)$ are backwards stopping times with respect to some σ-fields, and condition (iii). Second, observe that any verifiable time is a backwards coupling time. This follows directly from (i), (ii) and Definition 3. Third, define

$$\beta_m = \max(\beta(0), \ldots, \beta(m))$$

and observe that

$$(X_0^{-t}, \ldots, X_m^{-t}) = (X_0^{-t-1}, \ldots, X_m^{-t-1}) = \cdots, \quad \text{a.s. on } \{t \geq \beta_m\}.$$

Thus, a.s. on $\{t \geq \beta_m\}$, the sequence $(X_0^{-t}, \ldots, X_m^{-t})$ does not change with t. Since it also converges, in total variation, to $(\tilde{X}_0, \ldots, \tilde{X}_m)$, where $\tilde{X}$ is the stationary version of X, it follows that

$$(X_0^{-t}, \ldots, X_m^{-t}) = (\tilde{X}_0, \ldots, \tilde{X}_m), \quad \text{a.s. on } \{t \geq \beta_m\}.$$

Therefore,

$$(X_0^{-\beta_m}, \ldots, X_m^{-\beta_m}) = (\tilde{X}_0, \ldots, \tilde{X}_m), \quad \text{a.s.}$$

Since $\beta_m \geq \beta(i)$, for each $0 \leq i \leq m$, we have $X_i^{-\beta_m} = X_i^{-\beta(i)}$, and this is $\mathscr{G}_{-\beta(i),i}$–measurable and so, a fortiori, $\mathscr{G}_{-\beta_m,m}$–measurable (the σ-fields are increasing). Thus, $(\tilde{X}_0, \ldots, \tilde{X}_m)$ is $\mathscr{G}_{-\beta_m,m}$–measurable. In other words, any finite-dimensional projection $(\tilde{X}_0, \ldots, \tilde{X}_m)$ of the stationary distribution can be "perfectly sampled". That is, in practice, $\{\mathscr{G}_{-j,m}\}$ contains our basic data (e.g., it measures the random numbers we are using), β_m is a stopping time, and $(\tilde{X}_0, \ldots, \tilde{X}_m)$ is measurable with respect to a stopped σ-field. This is what perfect sampling means, in an abstract setup, *without* reference to any Markovian structure.

Naturally, we would like to have a condition for verifiability. Here we present a sufficient condition for the case where renovating events of special structure exist. To prepare for the theorem below, consider a stochastic process $\{X_n, n \in \mathbb{Z}\}$ on $(\Omega, \mathscr{F}, P, \theta)$, the notation being that of Section 2. Let $\{\zeta_n = \zeta_0 \circ \theta^n, n \in \mathbb{Z}\}$ be a family of i.i.d. random variables. For fixed $\kappa \in \mathbb{Z}$, consider the increasing family of σ-fields

$$\mathscr{G}_{-j,m} := \sigma(\zeta_{-j-\kappa}, \ldots, \zeta_m).$$

Consider also a family $\{B_n, n \in \mathbb{Z}\}$ of Borel sets and introduce the events

$$A_{-j,m} := \{\zeta_{-j-\kappa} \in B_{-\kappa}, \ldots, \zeta_m \in B_{m+j}\}$$

$$A_0 := \bigcap_{m \geq 0} A_{0,m} = \{\zeta_{-\kappa} \in B_{-\kappa}, \ldots, \zeta_0 \in B_0, \ldots\}$$

$$A_n := \{\zeta_{n-\kappa} \in B_{-\kappa}, \ldots, \zeta_n \in B_0, \ldots\} = \theta^{-n} A_0.$$

Theorem 2 (verifiability criterion)**.** *With the notation just introduced, suppose* $\mathbf{P}(A_0) > 0$*. Suppose the* A_n *are renovating events for the process* X*,*
(???)
and that $X_m^{-i}\mathbf{1}_{A_{-j,m}}$ *is* $\mathscr{G}_{-j,m}$*–measurable, for all* $-i \le -j \le m$*. Then*

$$\beta := \inf\{n \ge 0 : \mathbf{1}_{A_{-n}} = 1\}$$

is a verifiable time with respect to the $\{\mathscr{G}_{-j,m}\}$.

Proof. We shall show that $\beta = \sup_{m\ge 0}\beta(m)$, for appropriately defined backwards $\mathscr{G}_{\cdot,m}$–stopping times $\beta(m)$ that satisfy the properties (i), (ii) and (iii) of Definition 4. Let

$$\beta(m) := \inf\{j \ge 0 : \mathbf{1}_{A_{-j,m}} = 1\}.$$

Since $A_{-j,m} \in \mathscr{G}_{-j,m}$, we immediately have that $\beta(m)$ is a backwards $\mathscr{G}_{\cdot,m}$–stopping time. Then

$$\beta(m) := \inf\{j \ge 0 : \zeta_{-j-\kappa} \in B_{-\kappa}, \ldots, \zeta_m \in B_{m+j}\}$$

is a.s. increasing in m, with

$$\sup_m \beta(m) := \inf\{j \ge 0 : \zeta_{-j-\kappa} \in B_{-\kappa}, \ldots\} = \inf\{j \ge 0 : \mathbf{1}_{A_{-j}} = 1\} = \beta.$$

Hence (i) of Def. 4 holds. We next use the fact that the A_n are renovating events. we have, for all $i \ge j$,

$$X_m^{-i}\mathbf{1}_{A_{-j}} = X_m^{-j}\mathbf{1}_{A_{-j}}, \quad \text{a.s.}$$

Since

$$A_{-j} = A_{-j,m} \cap \{\zeta_{m+1} \in B_{m+1+j}, \ldots\} =: A_{-j,m} \cap D_{j,m},$$

we have

$$X_m^{-i}\mathbf{1}_{A_{-j,m}}\mathbf{1}_{D_{j,m}} = X_m^{-j}\mathbf{1}_{A_{-j,m}}\mathbf{1}_{D_{j,m}}, \quad \text{a.s.}$$

By assumption, $X_m^{-i}\mathbf{1}_{A_{-j,m}}$ is $\mathscr{G}_{-j,m}$–measurable, for all $i \ge j$. By the independence between the ζ_n's, $D_{j,m}$ is independent of $\mathscr{G}_{-j,m}$. Hence, by Lemma 1 of the Appendix, we can cancel the $\mathbf{1}_{D_{j,m}}$ terms in the above equation to get

$$X_m^{-i}\mathbf{1}_{A_{-j,m}} = X_m^{-j}\mathbf{1}_{A_{-j,m}}, \quad \text{a.s.,}$$

for all $i \ge j$. Now,

$$\{\beta(m) = j\} \subseteq A_{-j,m}, \tag{8}$$

and so, by multiplying by $\mathbf{1}(\beta(m) = j)$ both sides, we obtain

$$X_m^{-i}\mathbf{1}(\beta(m) = j) = X_m^{-j}\mathbf{1}(\beta(m) = j), \quad \text{a.s.,}$$

for all $i \ge j$. By stationarity, $\beta(m) < \infty$, a.s., and so for all $\ell \ge 0$,

$$X_m^{-\beta(m)-\ell} = X_m^{-\beta(m)}, \quad a.s.$$

Hence (ii) of Def. 4 holds. Finally, to show that $X_m^{-\beta(m)}$ is $\mathscr{G}_{-\beta(m),m}$–measurable, we show that $X_m^{-j}\mathbf{1}(\beta(m) = j)$ is $\mathscr{G}_{-j,m}$–measurable. Using the inclusion (8) again, we write

$$X_m^{-j}\mathbf{1}(\beta(m) = j) = X_m^{-j}\mathbf{1}_{A_{-j,m}}\mathbf{1}(\beta(m) = j).$$

By assumption, $X_m^{-j}\mathbf{1}_{A_{-j,m}}$ is $\mathscr{G}_{-j,m}$–measurable, and so is $\mathbf{1}(\beta(m) = j)$. Hence (iii) of Def. 4 also holds. $\square$

A perfect simulation algorithm

In the remaining of this section, we describe a "perfect simulation algorithm", i.e., a method for drawing samples from the stationary version of a process. The setup is as in Theorem 2. For simplicity, we take $\kappa = 0$. That is, we assume that

$$A_0 = \{\zeta_0 \in B_0, \zeta_1 \in B_1, \ldots\}$$

has positive probability, and that the $A_n = \theta^{-n} A_0$ are renovating events for the process $\{X_n\}$. Recall that the $\{\zeta_n = \zeta_0 \circ \theta^n\}$ are i.i.d., and that $\mathscr{G}_{m,n} = \sigma(\zeta_m, \ldots, \zeta_n)$, $m \le n$. It was proved in Theorem 2 that the time $\beta = \inf\{n \ge 0 : \mathbf{1}_{A_{-n}} = 1\}$ is a bc-time which is verifiable with respect to the $\{\mathscr{G}_{m,n}\}$. This time is written as $\beta = \sup_{m \ge 0} \beta(m)$, where $\beta(m) = \inf\{j \ge 0 : \zeta_{-j} \in B_0, \ldots, \zeta_m \in B_{m+j}\}$. The algorithm uses $\beta(0)$ only. It is convenient to let

$$\begin{aligned} \nu_1 &:= \beta(0) = \inf\{j \ge 0 : \zeta_{-j} \in B_0, \ldots, \zeta_0 \in B_j\}, \\ \nu_{i+1} &:= \nu_i + \beta(0) \circ \theta^{-\nu_i}, \quad i \ge 1. \end{aligned}$$

In addition to the above, we are going to assume that

$$B_0 \subseteq B_1 \subseteq B_2 \subseteq \ldots$$

It is easy to see that this monotonicity assumption is responsible for the following

$$\nu_1 \circ \theta^{-j} \le \nu_1 - j, \quad \text{a.s. on } \{\nu_1 \ge j\}. \tag{9}$$

Owing to condition (ii) of Definition 4 we have

$$X_0^{-\nu_1} = \tilde{X}_0 = X_0^{-\nu_1 - i}, \quad \text{for any } i \ge 0.$$

That is, if we "start" the process at time $-\nu_1$, we have, at time 0, that X_0 is a.s. equal to the stationary $\tilde{X}_0$. Applying θ^{-j} at this equality we have $X_0^{-\nu_1} \circ \theta^{-j} = \tilde{X}_0 \circ \theta^{-j} = \tilde{X}_{-j}$. But $X_0^{-\nu_1} \circ \theta^{-j} = (X_{\nu_1} \circ \theta^{-\nu_1}) \circ \theta^{-j} = X_{-\nu_1 \circ \theta^{-j}} \circ \theta^{-\nu_1 \circ \theta^{-j} - j} = X_{-j}^{-j-\nu_1 \circ \theta^{-j}}$. That is,

$$X_{-j}^{-j-\nu_1 \circ \theta^{-j}} = \tilde{X}_{-j} = X_{-j}^{-j-\nu_1 \circ \theta^{-j} - i}, \quad \text{for any } i \ge 0. \tag{10}$$

But from (9), we have $\nu_1 \ge j + \nu_1 \circ \theta^{-j}$, if $\nu_1 \ge j$, and so, from (10),

$$X_{-j}^{-\nu_1} = \tilde{X}_{-j}, \quad \text{a.s. on } \{\nu_1 \ge j\}.$$

This means that if we start the process at $-\nu_1$, then its values on any window $[-j, 0]$ contained in $[-\nu_1, 0]$ match the values of its stationary version on the same window:

$$(X_{-j}^{-\nu_1}, \ldots, X_0^{-\nu_1}) = (\tilde{X}_{-j}, \ldots, \tilde{X}_0), \quad \text{a.s. on } \{\nu_1 \ge j\}. \tag{11}$$

It remains to show a measurability property of the vector (11) that we are simulating. By (iii) of Definition 4, we have that $X_0^{-\nu_1}$ is $\mathscr{G}_{-\nu_1,0}$–measurable. That is, if $\nu_1 = \ell$ then $\tilde{X}_0$ is a certain deterministic function of $\zeta_{-\ell}, \ldots, \zeta_0$. Thus, the functions h_ℓ are defined, for all $\ell \ge 0$, by the condition

$$X_0^{-\ell} = h_\ell(\zeta_{-\ell}, \ldots, \zeta_0), \quad \text{a.s. on } \{\nu_1 = \ell\},$$

or,

$$X_0^{-\nu_1} = h_{\nu_1}(\zeta_{-\nu_1}, \ldots, \zeta_0).$$

Hence for any $i \ge 0$,

$$X_{-i}^{-i-\nu_1 \circ \theta^{-i}} = X_0^{-\nu_1} \circ \theta^{-i} = h_{\nu_1 \circ \theta^{-i}}(\zeta_{-i-\nu_1 \circ \theta^{-i}}, \ldots, \zeta_{-i}).$$

But if $\nu_1 \geq j$, we have $\nu_1 \circ \theta^{-i} \leq \nu_1 - i$ for all $i \in [0, j]$, and so every component of $(X_0^{-\nu_1} \circ \theta^{-i}, 0 \leq i \leq j)$ is a deterministic function of $\zeta_0, \ldots, \zeta_{-\nu_1}$. Thus the vector appearing in (11) is a deterministic function of $\zeta_0, \ldots, \zeta_{-\nu_1}$, if $\nu_1 \geq j$. This is precisely the measurability property we need.

We now observe that, in (11), we can replace ν_1 by any ν_i:

$$(X_{-j}^{-\nu_i}, \ldots, X_0^{-\nu_i}) = (\tilde{X}_{-j}, \ldots, \tilde{X}_0), \quad \text{a.s. on } \{\nu_i \geq j\}, \quad i = 1, 2, \ldots$$

Hence if we want to simulate $(\tilde{X}_{-j}, \ldots, \tilde{X}_0)$ we search for an i such that $\nu_i \geq j$, and start the process from $-\nu_i$. It is now clear how to simulate the process on any window prior to 0.

To proceed forward, i.e., to simulate $\{\tilde{X}_n, n > 0\}$, consider first $\tilde{X}_1$. Note that

$$\begin{aligned}\tilde{X}_1 = \tilde{X}_0 \circ \theta &= h_{\nu_1}(\zeta_{-\nu_1}, \ldots, \zeta_0) \circ \theta \\ &= h_{\nu_1 \circ \theta}(\zeta_{-\nu_1 \circ \theta + 1}, \ldots, \zeta_1)\end{aligned}$$

Next note that $\nu_1 \circ \theta$ is either equal to 0, or to $\nu_1 + 1$, or to $\nu_2 + 1 = \nu_1 + \nu_1 \circ \theta^{-\nu_1} + 1$, etc. This follows from the definition of ν_1 and ν_i, as well as the monotonicity between the B_j. If $\nu_1 = 0$ (which is to say, $\zeta_1 \in B_0$) then $\tilde{X}_1 = h_0(\zeta_1)$. Otherwise, if $\zeta_1 \notin B_0$, but $\zeta_1 \in B_{\nu_1+1}$, then $\nu_1 \circ \theta = \nu_1 + 1$, and so $\tilde{X}_1 = h_{\nu_1+1}(\zeta_{-\nu_1}, \ldots, \zeta_1)$. Thus, for some finite (but random) j (defined from $\zeta_1 \in B_{\nu_j+1} \setminus B_{\nu_j}$), we have $\tilde{X}_1 = h_{\nu_j+1}(\zeta_{-\nu_j}, \ldots, \zeta_1)$. The algorithm proceeds similarly for $n > 1$.

The connection between perfect simulation and backward coupling was first studied by Foss and Tweedie (1998).

Weak verifiability

Suppose now that we drop the condition that $\mathbf{P}(A_0) > 0$, but only assume that

$$\beta(0) < \infty, \text{ a.s.}$$

Of course, this implies that $\beta(m) < \infty$, a.s., for all m. Here we can no longer assert that we have sc-convergence to a stationary version, but we can only assert existence in the sense described in the sequel. Indeed, simply the a.s. finiteness of $\beta(0)$ (and not of β) makes the perfect simulation algorithm described above realizable. The algorithm is shift-invariant, hence the process defined by it is stationary. One may call this process a stationary version of X. This becomes precise if $\{X_n\}$ itself is a stochastic recursive sequence, in the sense that the stationary process defined by the algorithm is also a stochastic recursive sequence with the same driver. (See Section 4.)

The construction of a stationary version, under the weaker hypothesis $\beta(0) < \infty$, a.s., is also studied by Comets *et al.* (2001), for a particular model. In that paper, it is shown that $\beta(0) < \infty$ a.s., iff

$$\sum_{n=1}^{\infty} \prod_{k=0}^{n} \mathbf{P}(\zeta_0 \in B_k) = \infty.$$

The latter condition is clearly weaker than $\mathbf{P}(A_0) > 0$. In Comets *et al.* (2001) it is shown that it is equivalent to the non-positive recurrence of a certain Markov chain, a realization which leads directly to the proof of this condition.

4 Strong coupling for stochastic recursive sequences

As in the previous section, let $(\Omega, \mathscr{F}, P, \theta)$ be a probability space with a P-preserving ergodic flow θ. Let $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$, $(\mathscr{Y}, \mathscr{B}_{\mathscr{Y}})$ be two measurable spaces. Let $\{\xi_n, n \in \mathbb{Z}\}$ be a stationary sequence of $\mathscr{Y}$-valued random variables. Let $f : \mathscr{X} \times \mathscr{Y} \to \mathscr{X}$ be a measurable function. A stochastic recursive sequence (SRS) $\{X_n, n \geq 0\}$ is defined as an $\mathscr{X}$-valued process that satisfies

$$X_{n+1} = f(X_n, \xi_n), \quad n \geq 0. \tag{12}$$

The pair $(f, \{\xi_n\})$ is referred to as the driver of the SRS X. The choice of 0 as the starting point is arbitrary.

A *stationary solution* $\{\tilde{X}_n\}$ of the stochastic recursion is a stationary sequence that satisfies the above recursion. Clearly, it can be assumed that $\tilde{X}_n$ is defined for all $n \in \mathbb{Z}$. There are examples that show that a stationary solution may exist but may not be unique. The classical such example is that of a two-server queue, which satisfies the so-called Kiefer-Wolfowitz recursion (see Brandt *et al.* (1990)). In this example, under natural stability conditions, there are infinitely many stationary solutions, one of which is "minimal" and another "maximal". One may define a particular solution, say $\{X_n^0\}$, to the two-server queue SRS by starting from the zero initial condition. Then X^0 sc-converges (under some conditions) to the minimal stationary solution. In our terminology then, we may say that the minimal stationary solution is the stationary version of X^0.

Stochastic recursive sequences are ubiquitous in applied probability modeling. For instance, a Markov chain with values in a countably generated measurable space can be expressed in the form of SRS with i.i.d. drivers.

The previous notions of coupling take a simpler form when stochastic recursive sequences are involved owing to the fact that if two SRS with the same driver agree at some n then they agree thereafter. We thus have the following modifications of the earlier theorems:

Proposition 3. *Let X, $\tilde{X}$ be SRS with the same driver $(f, \{\xi_n\})$, and assume that $\tilde{X}$ is stationary. Then*
(i) X c-converges to $\tilde{X}$ iff

$$\lim_{n\to\infty} \mathbf{P}(X_n = \tilde{X}_n) = 1.$$

(ii) X sc-converges to $\tilde{X}$ iff

$$\lim_{n\to\infty} \mathbf{P}(\tilde{X}_n = X_n = X_n^{-1} = X_n^{-2} = \cdots) = 1.$$

(iii) X bc-converges iff

$$\lim_{n\to\infty} \mathbf{P}(X_0^{-n} = X_0^{-(n+1)} = X_0^{-(n+2)} = \cdots) = 1.$$

The standard renovation theory (see Borovkov (1984,1998)) is formulated as follows. First, define renovation events:

Definition 5 (renovation event for SRS)**.** *Fix $n \in \mathbb{Z}$, $\ell \in \mathbb{Z}_+$ and a measurable function $g : \mathscr{Y}^{\ell+1} \to \mathscr{X}$. A set $R \in \mathscr{F}$ is called (n, ℓ, g)–renovating for the SRS X iff*

$$X_{n+\ell+1} = g(\xi_n, \ldots, \xi_{n+\ell}), \quad \textit{a.s. on } R. \tag{13}$$

An alternative terminology (Borovkov (1988)) is: R is a renovation event on the segment $[n, n+\ell]$.

We then have the following theorem.

Theorem 3 (renovation theorem for SRS)**.** *Fix $\ell \geq 0$ and $g : \mathscr{Y}^{\ell+1} \to \mathscr{X}$. Suppose that, for each $n \geq 0$, there exists a (n, ℓ, g)–renovating event R_n for X. Assume that $\{R_n, n \geq 0\}$ is stationary and ergodic, with $\mathbf{P}(R_0) > 0$. Then the SRS X bc-converges and its stationary version $\tilde{X}$ is an SRS with the same driver as X.*

Proof. For each $n \in \mathbb{Z}$, define $\hat{X}_{n,i}$, recursively on the index i, by

$$\begin{aligned} \hat{X}_{n,n+\ell+1} &= g(\xi_n, \ldots, \xi_{n+\ell}) \\ \hat{X}_{n,n+j+1} &= f(\hat{X}_{n,n+j}, \xi_{n+j}), \quad j \geq \ell+1, \end{aligned} \tag{14}$$

and observe that (13) implies that

$$\forall n \geq 0 \;\; \forall j \geq \ell+1 \;\; X_{n+j} = \hat{X}_{n,n+j}, \quad \text{a.s. on } R_n. \tag{15}$$

For $n \in \mathbb{Z}$, set $A_n := R_{n-\ell-1}$. Consider the stationary background

$$H_{n,p} := \hat{X}_{n-\ell-1,p}, \quad p \geq n.$$

Note that $H_{n,p} \circ \theta^k = H_{n+k,p+k}$ and rewrite (15) as

$$\forall n \geq \ell+1 \;\; \forall i \geq 0 \;\; X_{n+i} = H_{n,n+i}, \quad \text{a.s. on } A_n.$$

Since $\mathbf{P}(A_0) = \mathbf{P}(R_0) > 0$, there is a unique stationary version $\tilde{X}$, constructed by means of the bc-time

$$\gamma := \inf\{i \geq 0 : \; \mathbf{1}_{R_{-i-\ell-1}} = 1\}. \tag{16}$$

We have

$$\tilde{X}_n = X_{\gamma+n} \circ \theta^{-\gamma} = H_{-\gamma,n} = \hat{X}_{-\gamma-\ell-1,n}.$$

This $\tilde{X}_n$ can be defined for all $n \in \mathbb{Z}$. From this, and (14), we have that $\tilde{X}_{n+1} = f(\tilde{X}_n, \xi_n), \; n \in \mathbb{Z}$, i.e., $\tilde{X}$ has the same driver as X. □

It is useful to observe that, if R_n are (n, ℓ, g) renovating events for X with $\mathbf{P}(R_0) > 0$, then the stationary version $\tilde{X}$ satisfies $\tilde{X}_{-\gamma} = g(\xi_{-\gamma-\ell-1}, \ldots, \xi_{-\gamma-1})$, a.s., where γ is the bc-time defined in (16). More generally, if we consider the random set $\{j \in \mathbb{Z} : \; \mathbf{1}_{R_j} = 1\}$ (the set of renovation epochs), we have, for any α in this set, $\tilde{X}_\alpha = g(\xi_{\alpha-\ell-1}, \ldots, \xi_{\alpha-1})$, a.s.

Example 2. Consider 2-server queue with stationary ergodic driving sequences $\{(\sigma_n, t_n)\}$.

Example 3. Consider a Markov chain $\{X_n\}$ with values in a finite set S, having stationary transition probabilities $p_{i,j}, \; i,j \in S$. Assume that $[p_{i,j}]$ is irreducible and aperiodic. Although there is a unique invariant probability measure, whether X bc-converges to the stationary Markov chain $\tilde{X}$ depends on the realization of X on a particular probability space. We can achieve bc-convergence with a verifiable bc-time if we realize X as follows: Consider a sequence of i.i.d. random maps $\xi_n : S \to S$, $n \in \mathbb{Z}$ (and we write $\xi_n \xi_{n+1}$ to indicate composition). Represent each ξ_n as a vector $\xi_n = (\xi_n(i), i \in S)$, with *independent* components such that

$$\mathbf{P}(\xi_n(i) = j) = p_{i,j}, \quad i,j \in S.$$

Then the Markov chain is realized as an SRS by

$$X_{n+1} = \xi_n(X_n).$$

It is important to notice that the condition that the components of ξ_n be independent is not necessary for the Markovian property. It is only used as a means of constructing the process on a particular probability space, so that backwards coupling takes place. Now define

$$\beta = \inf\{n \geq 0 : \forall \, i,j \in S \quad \xi_0 \cdots \xi_{-n}(i) = \xi_0 \cdots \xi_{-n}(j)\}.$$

It can be seen that, under our assumptions, β is a bc-time for X, $\beta < \infty$, a.s., and β is verifiable. This bc-time is the one used by Propp and Wilson (1996) in their perfect simulation method for Markov chains. Indeed, the verifiability property of β allows recursive simulation of the random variable $\xi_0 \cdots \xi_{-\beta}(i)$ which (regardless of i) has the stationary distribution.

Example 4. Another interesting example is the process considered by Brémaud and Massoulié (1994) that has a "Markov-like" property with *random memory*. Consider a process $\{X_n, n \in \mathbb{Z}\}$ with values in a Polish space S and suppose that its transition kernel, defined by $\mathbf{P}(X_n \in \cdot \mid X_{n-1}, X_{n-2}, \ldots)$ is time-homogeneous [a similar setup is considered in the paper by Comets et al. (2001)], i.e. that there exists a kernel $\mu : K^\infty \times \mathscr{B}(K) \to [0,1]$ such that $\mathbf{P}(X_n \in B \mid X_{n-1}, X_{n-2}, \ldots) = \mu((X_{n-1}, X_{n-2}, \ldots); B)$, $B \in \mathscr{B}(K)$. This represents the dynamics of the process. In addition, assume that the dynamics does not depend on the whole past, but on a finite but random number of random variables from the past. It is also required that the random memory is "consistent" and that the minorization condition $\mu((X_{n-1}, X_{n-2}, \ldots), \cdot) \geq \varepsilon\nu(\cdot)$, where $\varepsilon \in (0,1)$ and ν a probability measure on $(K, \mathscr{B}(K))$, holds. See Brémaud and Massoulié (1994) for details. Then it is shown that renovation events do exist and that the process $\{X_n\}$ sc-converges to a stationary process that has the same dynamics μ.

Appendix: Auxiliary results

Lemma 1. *If Y_1, Y_2 and Z are three random variables such that Z is independent of (Y_1, Y_2), $\mathbf{P}(Z \neq 0) > 0$, and $Y_1 Z = Y_2 Z$, a.s., then $Y_1 = Y_2$, a.s.*

Proof. Since $\mathbf{P}(Y_1 Z = Y_2 Z) = 1$, we have

$$\mathbf{P}(Z \neq 0) = \mathbf{P}(Y_1 Z = Y_2 Z, Z \neq 0) = \mathbf{P}(Y_1 = Y_2, Z \neq 0) = \mathbf{P}(Y_1 = Y_2)\mathbf{P}(Z \neq 0),$$

where the last equality follows from independence. Since $\mathbf{P}(Z \neq 0) > 0$, we obtain the result $\mathbf{P}(Y_1 = Y_2) = 1$. □

Proposition 4. *Let $X_1, X_2, \ldots$ be a stationary-ergodic sequence of random variables with $\mathbf{E}X_1^+ < \infty$. Then*

$$\frac{1}{n} \max_{1 \leq i \leq n} X_i \to 0, \qquad \text{a.s. and in } L^1.$$

Proof. Without loss of generality, assume $X_n \geq 0$, a.s. Put $Y_n = \max_{1 \leq i \leq n} X_i$. Clearly,

$$Y_{n+k} \leq \max_{1 \leq i \leq n} X_i + \max_{n+1 \leq i \leq n+k} X_i = Y_n + Y_k \circ \theta^n.$$

Kingman's subadditive ergodic theorem shows that $Y_n/n \to c$, a.s., where $c \geq 0$. We will show that $c = 0$. If $c > 0$ then, for any $0 < \varepsilon < c/2$, there is k_0 such that $\mathbf{P}(Y_k/k > c + \varepsilon) < \varepsilon$ for all $k \geq k_0$. Fix $k \geq k_0$ and let $n = 2k$. We then have

$$\begin{aligned} \mathbf{P}(Y_n/n > 3c/4) &\leq \mathbf{P}(Y_n/n > (c+\varepsilon)/2) \\ &\leq \mathbf{P}(\max(Y_k/n, Y_k \circ \theta^k/n) > (c+\varepsilon)/2) \\ &\leq 2\mathbf{P}(Y_k/n > (c+\varepsilon)/2) \\ &= 2\mathbf{P}(Y_k/k > c+\varepsilon), \end{aligned}$$

which contradicts the a.s. convergence of Y_n/n to c. Hence $c = 0$. To show that $\mathbf{E}Y_n/n \to 0$ simply observe that the sequence Y_n/n is bounded by $S_n/n = (X_1 + \cdots + X_n)/n$, and since $S_n/n \to \mathbf{E}X_1$, a.s. and in L^1, it follows that $\{S_n/n\}$ is a uniformly integrable sequence, and thus so is $\{Y_n/n\}$. □

Proposition 5. *Let $X_1, X_2, \ldots$ be a stationary-ergodic sequence of random variables with $\mathbf{E}X_1 = 0$. Consider the stationary walk $S_n = X_1 + \cdots + X_n$, $n \geq 1$, with $S_0 = 0$. Put $M_n = \max_{0 \leq i \leq n} S_i$. Then $M_n/n \to 0$, a.s. and in L^1.*

Proof. Fix $\varepsilon > 0$. Let $M^* = \sup_{i \geq 0}(S_i - i\varepsilon)$. Then $M^* < \infty$, a.s. We have

$$M_n = \max_{0 \leq i \leq n} (S_i - i\varepsilon + i\varepsilon) \leq \max_{0 \leq i \leq n} (S_i - i\varepsilon) + n\varepsilon \leq M^* + n\varepsilon.$$

So, $M_n/n \leq M^*/n + \varepsilon$, a.s. This implies that $\limsup_{n\to\infty} M_n/n \leq \varepsilon$, a.s., and so $M_n/n \to 0$, a.s. Convergence in L^1 can be proved as follows. Let $b := \mathbf{E}X_1^+$. Define $S_n^{(+)} := X_1^+ + \cdots + X_n^+$. By the ergodic theorem, $S_n^{(+)}/n \to b$, a.s. and in L^1, and so $\{S_n^{(+)}/n\}$ is a uniformly integrable sequence. But $0 \leq M_n/n \leq S_n^{(+)}/n$. So $\{M_n/n\}$ is also uniformly integrable. □

References

[1] Anantharam, V. and Konstantopoulos, T. (1997) Stationary solutions of stochastic recursions describing discrete event systems. *Stoch. Proc. Appl.* **68**, 181-194.

[2] Anantharam, V. and Konstantopoulos, T. (1999) A correction and some additional remarks on: "stationary solutions of stochastic recursions describing discrete event systems". *Stoch. Proc. Appl.* **80**, 271-278.

[3] Baccelli, F. and Brémaud, P. (1994) *Elements of Queueing Theory.* Springer-Verlag, Berlin.

[4] Baccelli, F. and Foss S. (1995) On the saturation rule for the stability of queues. *J. Appl. Prob.* **32**, No.2, 494-507.

[5] Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

[6] Borovkov, A.A. (1978) Ergodicity and stability theorems for a class of stochastic equations and their applications. *Theory Prob. Appl.* **23**, 227-258.

[7] Borovkov, A.A. (1980) *Asymptotic Methods in Queueing Theory.* Nauka, Moscow (Russian edition). Revised English edition: Wiley, New York, 1984.

[8] Borovkov, A.A. (1998) *Ergodicity and Stability of Stochastic Processes.* Wiley, Chichester.

[9] Borovkov, A.A. and Foss, S.G. (1992) Stochastically recursive sequences and their generalizations. *Sib. Adv. Math.* **2**, n.1, 16-81.

[10] Borovkov, A.A. and Foss, S.G. (1994) Two ergodicity criteria for stochastically recursive sequences. *Acta Appl. Math.* **34**, 125-134.

[11] Brandt, A. Franken, P. and Lisek, B. (1990) *Stationary Stochastic Models.* Akademie-Verlag, Berlin.

[12] Brémaud, P. and Massoulié, L. (1994) Imbedded construction of stationary sequences and point processes with a random memory. *Queueing Systems* **17**, 213-234.

[13] Comets, F., Fernández, R. and Ferrari, P.A. (2001). Processes with long memory: regenerative construction and perfect simulation. Preprint.

[14] Diaconis, P. and Freedman, D. (1999) Iterated random functions. *SIAM Rev.* **41**, 45-76.

[15] Foss, S. (1983) On ergodicity conditions in multi-server queues. *Sib. Math. J.* **34**, n.6, 168-175.

[16] Foss, S. and Tweedie, R.L. (1998) Perfect simulation and backward coupling. *Stochastic Models* **14**, No. 1-2, 187-204.

[17] Lindvall, T. (1992, 2003) *Lectures on the Coupling Method.* Wiley, New York.

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

[19] Thorisson, H. (2000) *Coupling, Stationarity, and Regeneration.* Springer-Verlag, New York.