Statistics and Probability Letters 83 (2013) 2073–2076

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

# A note on formal constructions of sequential conditional couplings

Gareth O. Roberts [a], Jeffrey S. Rosenthal [b,*]

[a] Department of Statistics, University of Warwick, CV4 7AL, Coventry, UK

[b] Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3

ARTICLE INFO

*Article history:*
Received 28 March 2013
Received in revised form 13 May 2013
Accepted 13 May 2013
Available online 23 May 2013

*Keywords:*
Markov chain
Convergence rate
MCMC algorithm
Coupling constructions

ABSTRACT

We discuss a formal mathematical framework for certain coupling constructions via minorisation conditions, which are often used to prove bounds on convergence to stationarity of stochastic processes and Markov chain Monte Carlo algorithms.


## 1. Introduction

Bounds on the convergence of Markov chains and other stochastic processes to stationary distributions has become a very widely studied topic in recent years, motivated largely by applications to Markov chain Monte Carlo (MCMC) algorithms; see, for example, Gelfand and Smith (1990), Tierney (1994), Geyer (1992), Gilks et al. (1996), and Brooks et al. (2011). One common method of obtaining such bounds is through *coupling constructions*; see, for example, Thorisson (2000), Rosenthal (2002), and Roberts and Rosenthal (2004). Here, a second copy of a similar or identical process is constructed, jointly with the original process, and the probability of the two chains becoming (or remaining) equal is then examined and used.

Such coupling constructions are often presented in a somewhat informal and intuitive style, of the form "First construct one random variable as follows, then find a joint distribution for these two other random variables conditional on the first one, then conditionally construct a fourth random variable like this", etc. We believe this to be acceptable, and to lead to rigorously valid coupling constructions. However, we recently became aware that at least one mathematically minded reader is not comfortable with such informal descriptions. Thus, the purpose of this note is to provide a more formal mathematical version of common methods of constructing couplings of pairs of stochastic processes. We emphasise that none of the results presented here are particularly novel – they are simple consequences of the Kolmogorov extension theorem, the coupling inequality, and maximal couplings – but we hope that they will help clarify the application of coupling constructions to MCMC algorithms.

Coupling constructions are often used to bound the probability of two Markov chains with identical transition kernels *becoming* equal, thus bounding the total variation distance between them (see, for example, Thorisson, 2000, Rosenthal, 1995, and Rosenthal, 2002). Some other approaches instead bound the probability of two processes with different probability

* Corresponding author. Tel.: +1 416 978 4594; fax: +1 416 978 5133.
*E-mail addresses:* g.o.roberts@lancaster.ac.uk (G.O. Roberts), jeff@math.toronto.edu (J.S. Rosenthal).
*URL:* http://probability.ca/jeff/ (J.S. Rosenthal).

0167-7152/$ – see front matter 
http://dx.doi.org/10.1016/j.spl.2013.05.012

laws *staying* equal for all times up to time $N$. Both cases often involve conditional-type coupling constructions which are usually stated informally, but which can be stated formally if desired. Below, for concreteness, we concentrate on the second case, since the recent questions about coupling constructions originated there. However, similar methods can be used to "formalise" the coupling constructions in the first case too—and indeed in any situation in which couplings are constructed informally, one random variable at a time, in terms of various conditional distributions.

## 2. Statement of main result

Let $\{X_n\}_{n=0}^{\infty}$ and $\{X'_n\}_{n=0}^{\infty}$ be two different stochastic processes, defined possibly on different probability spaces, but taking values in the same Polish measurable state space $(\mathcal{X}, \mathcal{B})$ (e.g. on $\mathbf{R}^d$ with the Borel subsets). Let $F_n = (X_0, X_1, \ldots, X_n)$ and $F'_n = (X'_0, X'_1, \ldots, X'_n)$ be the two processes' histories up to time $n$. For $n \geq 1$, $A \in \mathcal{B}$, and a state history vector $s^{(n-1)} = (s_0, s_1, \ldots, s_{n-1}) \subseteq \mathcal{X}^n$, let $Q_n(A; s^{(n-1)}) = \mathbf{P}[X_n \in A \mid F_{n-1} = s^{(n-1)}]$ and $Q'_n(A; s^{(n-1)}) = \mathbf{P}[X'_n \in A \mid F'_{n-1} = s^{(n-1)}]$ be the regular conditional probability distributions. As a special case, when $n = 0$, let $F_{-1}$, $F'_{-1}$, and $s^{(-1)}$ each be the empty set, so that $Q_0(A; s^{(-1)}) = \mathbf{P}[X_0 \in A \mid \emptyset = \emptyset] = \mathbf{P}[X_0 \in A]$ and $Q'_0(A; s^{(-1)}) = \mathbf{P}[X'_0 \in A]$.

In terms of these definitions, a formal statement about sequential coupling constructions which attempt to keep the two processes equal is as follows.

**Theorem 1.** *Let $N$ be a non-negative integer. Suppose that, for each $0 \leq n \leq N$, there is $a_n \geq 0$ such that, for each state history vector $s^{(n-1)} = (s_0, s_1, \ldots, s_{n-1}) \subseteq \mathcal{X}^n$, either*

(i) $\sup_{A\in\mathcal{B}} |Q_n(A; s^{(n-1)}) - Q'_n(A; s^{(n-1)})| \leq a_n$; *or*
(ii) *there are random variables $W$ and $W'$, defined jointly on some probability measure space, each taking values in $\mathcal{X}$, which are measurable functions of $s^{(n-1)}$ (i.e., such that, if $W = W(s^{(n-1)})$ and $W' = W'(s^{(n-1)})$, then, for each $A \in \mathcal{B}$, the subsets $\{s^{(n-1)} : W(s^{(n-1)}) \in A\}$ and $\{s^{(n-1)} : W'(s^{(n-1)}) \in A\}$ are measurable subsets of $\mathcal{X}^n$), such that $\mathbf{P}[W \in A] = Q_n(A; s^{(n-1)})$ and $\mathbf{P}[W' \in A] = Q'_n(A; s^{(n-1)})$ for all $A \in \mathcal{B}$, and $\mathbf{P}[W = W'] \geq 1 - a_n$; or*
(iii) *there is a probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B})$ which is a measurable function of $s^{(n-1)}$ (i.e., such that, if $\nu(\cdot) = \nu_{s^{(n-1)}}(\cdot)$, then, for each $A \in \mathcal{B}$, the mapping $s^{(n-1)} \mapsto \nu_{s^{(n-1)}}(A)$ is a measurable function of $s^{(n-1)} \in \mathcal{X}^n$), such that $Q_n(A; s^{(n-1)}) \geq (1 - a_n)\,\nu(A)$ and $Q'_n(A; s^{(n-1)}) \geq (1 - a_n)\,\nu(A)$ for all $A \in \mathcal{B}$.*

*Then there exist random variables $\{\tilde{X}_n, \tilde{X}'_n\}_{n=0}^{N}$ defined jointly on some probability measure space, such that $\mathcal{L}(\tilde{X}_0, \ldots, \tilde{X}_N) = \mathcal{L}(X_0, \ldots, X_N)$, $\mathcal{L}(\tilde{X}'_0, \ldots, \tilde{X}'_N) = \mathcal{L}(X'_0, \ldots, X'_N)$, and, furthermore,*

$$\mathbf{P}(\tilde{X}_i = \tilde{X}'_i \text{ for } 0 \leq i \leq N) \geq 1 - a_0 - a_1 - a_2 - \cdots - a_N. \quad (*)$$

## 3. Background tools

In this section, we collect a few standard results that will be used to prove the above theorem.

**Proposition 2.** *Let $\rho$ and $\sigma$ be two probability measures on $(\mathcal{X}, \mathcal{B})$, and let $\epsilon \geq 0$. Then the following are equivalent.*

(i) $\sup_{A\in\mathcal{B}} |\rho(A) - \sigma(A)| \leq \epsilon$.
(ii) *There are jointly defined random variables $Y$ and $Z$ taking values on $(\mathcal{X}, \mathcal{B})$, such that $\mathbf{P}(Y \in A) = \rho(A)$ and $\mathbf{P}(Z \in A) = \sigma(A)$ for all $A \in \mathcal{B}$, and $\mathbf{P}[Y = Z] \geq 1 - \epsilon$.*
(iii) *There is a probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B})$ such that $\rho(A) \geq (1 - \epsilon)\,\nu(A)$ and $\sigma(A) \geq (1 - \epsilon)\,\nu(A)$ for all $A \in \mathcal{B}$.*

**Proof.** That (ii) implies (i) is the standard *coupling inequality*; see, for example, Thorisson (2000), or Eq. (13) of Roberts and Rosenthal (2004).

That (i) implies (ii) is a well-known property of couplings, corresponding to the existence of maximal couplings (see, for example, Griffeath, 1975 or Proposition 3(g) of Roberts and Rosenthal, 2004). Indeed, let $\eta(\cdot) = \rho(\cdot) + \sigma(\cdot)$ be a joint dominating measure, with corresponding Radon–Nikodym derivatives $g = \frac{d\rho}{d\eta}$ and $h = \frac{d\sigma}{d\eta}$, and let $m = \min(g, h)$. Then, let $a = \int_{\mathcal{X}} m\, d\eta$, $b = \int_{\mathcal{X}} (g - m)\, d\eta$, and $c = \int_{\mathcal{X}} (h - m)\, d\eta$. The statement is trivial if any of $a, b, c$ are zero, so assume that they are all positive. Then jointly construct independent random variables $R, U, V, I$ such that $R$ has density $m/a$, $U$ has density $(g - m)/b$, $V$ has density $(h - m)/c$, $\mathbf{P}[I = 1] = a$, and $\mathbf{P}[I = 0] = 1 - a$. Finally, let $Y = Z = R$ if $I = 1$, and let $Y = U$ and $Z = V$ if $I = 0$. Then $Y \sim \rho(\cdot)$ and $Z \sim \sigma(\cdot)$, and $U$ and $V$ have disjoint support, so $\mathbf{P}[U = V] = 0$. Hence, $\mathbf{P}[Y = Z] = \mathbf{P}[I = 1] = a$. But it is easily seen that $a = 1 - \sup_{A\in\mathcal{B}} |\rho(A) - \sigma(A)|$ (see, e.g., Roberts and Rosenthal, 2004, Proposition 3(f)), which gives the result.

That (iii) implies (i) follows by setting $\alpha(A) = \epsilon^{-1}[\rho(A) - (1 - \epsilon)\,\nu(A)]$ and $\beta(A) = \epsilon^{-1}[\sigma(A) - (1 - \epsilon)\,\nu(A)]$, so that $\alpha(\cdot)$ and $\beta(\cdot)$ are probability measures on $(\mathcal{X}, \mathcal{B})$, and $\rho(\cdot) = \epsilon\,\alpha(\cdot) + (1 - \epsilon)\,\nu(\cdot)$ and $\sigma(\cdot) = \epsilon\,\beta(\cdot) + (1 - \epsilon)\,\nu(\cdot)$, whence

$$|\rho(A) - \sigma(A)| = |\epsilon\,\alpha(A) + (1 - \epsilon)\,\nu(A) - \epsilon\,\beta(A) - (1 - \epsilon)\,\nu(A)| = \epsilon\,|\alpha(A) - \beta(A)| \leq \epsilon.$$

Finally, that (ii) implies (iii) follows by setting $\nu(A) = \frac{\mathbf{P}[Y\in A,\ Y=Z]}{\mathbf{P}[Y=Z]}$. □