**On Efficiency of Markovian Couplings**  
Mason DiCicco, joint with Ewan Harlow (U. Wisconsin) and Micheal Dotzel (U. Missouri)  
Project Advisor: Iddo Ben-Ari, Mathematics  
† Support from: UConn Trimble Family SURF award, NSA Markov Chains REU 2018

---

**Markov Chains (MCs)**

A Random process $X = (X_t)_{t \in \mathbb{Z}^+}$ taking values in a finite set $\mathcal{S}$ with the transition function

$$P(X_{t+1} = j|X_t = i) = P(X_t = j|X_{t-1} = i) = p(i,j) \quad \text{for all} \quad i,j \in \mathcal{S}.$$  

Definition 1. Let $\mu_1$ and $\mu_2$ be two probability distributions on $\mathcal{S}$. The total variation distance between the distribution of $X_t$ under $\mu_1$ and the distribution of $X_t$ under $\mu_2$ is defined as

$$d_{TV}(\mu_1, \mu_2) = \max_{i,j \in \mathcal{S}} |\mu_1(i) - \mu_2(j)|.$$  

**Ergodic Theorem for MCs**

A transition function $p$ is

- **Irreducible** if for any two states $i, j$, there exists a $t > 0$ such that $p^t(i,j) > 0$. This means that it is possible to get from any state to any other state using only transitions of positive probability.
- **Aperiodic** if for every state $i$, $\gcd\{t : p^t(i,i) > 0\} = 1$.

**The Method of Coupling**

Definition 2. Let $p$ be a transition function on a state space $\mathcal{S}$. An $\mathcal{S} \times \mathcal{S}$-valued process $(X,Y)$ is called a coupling for $p$ if each marginal process $X = (X_t)_{t \in \mathbb{Z}^+}$ and $Y = (Y_t)_{t \in \mathbb{Z}^+}$ is a MC with transition function $p$.

Suppose $(X,Y)$ is a coupling. We define the meeting time $\tau$:

$$\tau = \inf \{t \in \mathbb{Z}^+ : X_t = Y_t\}.$$  

Lemma 1 (Aldous’ inequality). Suppose $(X,Y)$ is a coupling with $(X_0, Y_0) = (i,j)$. Then

$$d_{TV}(X,Y) \leq \tau > 0.$$  

Thus, coupling is a probabilistic technique for obtaining upper bounds on $d_{TV}(X,Y)$ through real distribution of certain random variables.

**Greedy Couplings**

Definition 3. We call a coupling $(X,Y)$ greedy Markovian if it satisfies the following two properties:

- **Markovian**. It is a MC: $(X_{t+1}, Y_{t+1})$ is a deterministic function of $(X_t, Y_t)$ and a Uniform-$[0,1]$ random variable, independent of $(X_0, Y_0), \ldots, (X_t, Y_t)$.
- **Greedy**. The probability that $Y_{t+1} = X_{t+1}$, conditioned on $X_t, Y_t = (x,y)$ is maximized: it is equal to

$$\sum_j \min\{P_X(X_{t+1} = j), P_Y(X_{t+1} = j)\} = 1 - d_{TV}(X,Y).$$  

**Our results**

Consider a symmetric TF of the form

$$p = \begin{bmatrix} a & b & c \\ b & c & d \\ c & d & e \end{bmatrix}$$

We analyze all greedy Markovian couplings for $p$.

The diagram below represents the coupling from the state $(1,2)$ for the coupled process.

**Total Variation Distance**

Definition 1. Let $\mu_1$ and $\mu_2$ be two probability distributions on $\mathcal{S}$. The total variation distance between the distribution of $X_t$ under $\mu_1$ and the distribution of $X_t$ under $\mu_2$ is defined as

$$d_{TV}(\mu_1, \mu_2) = \max_{i,j \in \mathcal{S}} |\mu_1(i) - \mu_2(j)|.$$  

If, say, $\mu_1 = \delta_i$ for some state $i$, we write $d_{TV}(i, \mu_2)$ for $d_{TV}(\delta_i, \mu_2)$, etc.

A transition function $p$ is

- Irreducible if for any two states $i, j$, there exists a $t > 0$ such that $p^t(i,j) > 0$. This means that it is possible to get from any state to any other state using only transitions of positive probability.
- Aperiodic if for every state $i$, $\gcd\{t : p^t(i,i) > 0\} = 1$.

**Theorem 1** (Ergodic Theorem for MCs).

1. If $p$ is irreducible, then it possesses a unique stationary distribution $\pi$: $P(X_t = i) = \pi(i)$ for all $i$, or, equivalently, $\pi P = \pi$.
2. If $p$ is irreducible and aperiodic, there exists a constant $\rho \in [0,1)$ such that

$$\max_{i,j} d_{TV}(\delta_i, \mu_j) \leq \max_{i,j} d_{TV}(\delta_i, \delta_j) \leq \rho.$$  

**Note:** From linear algebra, the smallest $\rho$ satisfying (2) is the norm of the subdominant eigenvalue for $p$.

**Theorem 2.** Let $p$ be a symmetric $3 \times 3$ irreducible transition function. Let

$$M_{1,2} = \{k : P_k i > P_k a\}.$$  

Then a greedy Markovian coupling is efficient if and only if

$$M_{2,1} = M_{1,2} = M_{1,1}. $$  

1. The diagram corresponds to the particular case $a < c$ and $b < c$.
2. The area of the diagram is 1, representing the probability of all possible configurations for the coupled process in the next step.
3. The time spent on the left has respective areas $a$ and $b$ representing the meeting in the next step at $1$ and $2$, respectively.
4. The area of the remaining two regions represent the probability of not meeting in the next step. This is split into two regions, because
   (a) The leftover probability from the transition $2 \rightarrow 1$ is $b - c$.
   (b) The leftover probability from the transition $2 \rightarrow 2$ is $c - b$.
   (c) The leftover probability from transitions $1 \rightarrow 3$ is only from the transition $1 \rightarrow 3$, and is equal to $\gamma - \delta = (8 - a) + (c - a)$.

Summarizing:

\[
\begin{array}{c|ccc}
(\text{node}) & 1 & 2 & 3 \\
\hline
1 & a & b & c \\
2 & c & d & e \\
\end{array}
\]

- Item 4(c) captures a feature which is key to the analysis: if no meeting occurs, one of the chains necessarily transitions to some state determined by structure of the TF.
- Through reduction by symmetry and exhaustive examination of the remaining cases, the following result was obtained: