Simulation of Maximal Couplings for Finite State Markov Chains

Iddo Ben-Ari (University of Connecticut)[1]

Coin Flippers 2026

[1]Code implementation first written by MCREU22 students Alan Boles (curr. PhD candidate U of WA), Scott McIntye, (curr. PhD candidate at Duke U), and Bronson Zhou (curr. PhD candidate at Yale U)

Outline

- Coupling for RVs and MCs.
- Maximal Couplings of MCs.
- Algorithm for Maximal Couplings.
- Examples.

Coupling and Total Variation

Definition 1 (Coupling of Random Variables)

Let ν_1, ν_2 be probability measures on a countable set S. A coupling of ν_1 and ν_2 is an $S \times S$-valued random vector $\mathbf{X} = (X_1, X_2)$ with $P(X_k \in \cdot) = \nu_k,\ \ k = 1, 2$.

Definition 2 (Total Variation)

Let ν_1, ν_2 be probability measures on S. Define the total variation distance

$$d_{TV}(\nu_1, \nu_2) = \max_A(\nu_1(A) - \nu_2(A)) = \sum_{j:\nu_1(j)>\nu_2(j)} (\nu_1(j) - \nu_2(j))$$

$$= 1 - \sum_j \min(\nu_1(j), \nu_2(j)) = \frac{1}{2}\sum_j |\nu_1(j) - \nu_2(j)|.$$

Proposition 1

1. *If* $\mathbf{X}$ *is a coupling of* ν_1, ν_2*, then* $d_{TV}(\nu_1, \nu_2) \leq P(X_1 \neq X_2)$*.*
2. *There exists a coupling attaining the equality. A coupling with this property is called maximal.*

Proof

1. $d_{TV}(\nu_1, \nu_2) = 1 - \sum_j \underbrace{\min(\nu_1(j), \nu_2(j))}_{\geq P(X_1 = j = X_2)} \leq 1 - \sum_j P(X_1 = j = X_2) = P(X_1 \neq X_2)$
2. From above, any coupling with the property $P(X_1 = j = X_2) = \min(\nu_1(j), \nu_2(j)),\ j \in S$ will be maximal. Finding one is immediate from a decomposition of measures:

$$m = \underbrace{\min(\nu_1, \nu_2)}_{\text{coupling measure}},\ \underbrace{a_k = \nu_k - m}_{\text{excess measure}},\ k = 1, 2.$$

• Sample $C \sim$ Bernoulli$(m(S))$, and M, Z_1, Z_2 independent of C with respective distributions proportional to m, a_1, a_2 (set to zero if the respective measure is zero).
• The equality is achieved by the pair $X_k = CM + (1 - C)Z_k,\ k = 1, 2.$

Proposition 1

1. *If* $\mathbf{X}$ *is a coupling of* ν_1, ν_2, *then* $d_{TV}(\nu_1, \nu_2) \leq P(X_1 \neq X_2)$.
2. *There exists a coupling attaining the equality. A coupling with this property is called maximal.*

Proof

1. $d_{TV}(\nu_1, \nu_2) = 1 - \sum_j \underbrace{\min(\nu_1(j), \nu_2(j))}_{\geq P(X_1 = j = X_2)} \leq 1 - \sum_j P(X_1 = j = X_2) = P(X_1 \neq X_2)$
2. From above, any coupling with the property $P(X_1 = j = X_2) = \min(\nu_1(j), \nu_2(j)),\ j \in S$ will be maximal. Finding one is immediate from a decomposition of measures:

$$\underbrace{m = \min(\nu_1, \nu_2)}_{\text{coupling measure}},\ \underbrace{a_k = \nu_k - m}_{\text{excess measure}},\ k = 1, 2.$$

• Flip coins to simulate ~~Sample~~ $C \sim \text{Bernoulli}(m(S))$, and M, Z_1, Z_2 independent of C with respective distributions proportional to m, a_1, a_2 (set to zero if the respective measure is zero).
• The equality is achieved by the pair $X_k = CM + (1 - C)Z_k,\ k = 1, 2.$

Coupling of MCs

Setup

- p, a TF on a countable state space S, and μ_1, μ_2, probability distributions on S.
- (p, μ_k)-MC: a MC $\mathbf{X}_k = (\mathbf{X}_k(t) : t \in \mathbb{Z}_+)$ with TF p and initial distribution μ_k.

Definition 3 (Coupling of MCs)

- *A (p, μ_1, μ_2)-coupling: An $S \times S$ process $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) = (X(t) = (X_1(t), X_2(t)) : t \in \mathbb{Z}_+)$, where $\mathbf{X}_k$ a (p, μ_k)-MC, $k = 1, 2$.*
- *Coupling time: $T = \inf\{t \in \mathbb{Z}_+ : X_1(t) = X_2(t)\}$.*
- *Assumption: couplings are sticky, $X_1(T + t) = X_2(T + t)$ for $t \in \mathbb{Z}_+$.*

Example 1

p is SSRW on $\mathbb{Z}$, $\mu_1 = \delta_0, \mu_2 = \delta_{2k}$ for some $k \in \mathbb{N}$.

- *Independent coupling: $\mathbf{X}_2$ is independent of $\mathbf{X}_1$.*
- *Reflection coupling: increments of $\mathbf{X}_2$ are opposite to those of $\mathbf{X}_1$.*

The Coupling Connection

Let $\mathbf{X}_k$ be a (p,μ_k)-MC. Write p_k^t for the distribution of $X_k(t)$:

$$p_k^t(j) = P(X_k(t)=j) = \sum_i \mu_k(i)p^t(i,j),\ j\in S.$$

Proposition 2 (Coupling Inequality)

If $\mathbf{X}$ a-(p,μ_1,μ_2) coupling, then

$$d_{TV}(p_1^t,p_2^t)\le P(T>t),\ t\in\mathbb{Z}_+.$$

The coupling is called maximal if equality holds for all $t\in\mathbb{Z}_+$.

Theorem 1 (Griffeath '74, Pitman '76, Lindvall '92, etc.)

There exists a maximal (p,μ_1,μ_2)-coupling.

Comment

- Difficulty comes from the fact that we need to maximally couple all distributions simultaneously: the initial distributions, the propagation of their excesses, etc.
- There is a zoo of various couplings. Some depend on particular structure and geometry (e.g., reflection, monotone), while others are more general (e.g., independent, one-step greedy, global maximal, and shift couplings).

Example: SSRW

Setup

- p is SSRW on $\mathbb{Z}$, $\mu_1 = \delta_0$, $\mu_2 = \delta_{2k}$.
- $\mathbf{D} = (D(t) = X_2(t) - X_1(t) : t \in \mathbb{Z}_+)$ difference process, $D(0) = 2k$.
- Then $T = \inf\{t \in \mathbb{Z}_+ : D(t) = 0\}$.

Reflection Coupling

- $X_2(t) = 2k - X_1(t)$.
- $\mathbf{D}$ is RW with increments $+2, -2$ with resp. prob. $\frac{1}{2}, \frac{1}{2}$.

Independent Coupling

- $\mathbf{D}$ is RW with increments $+2, 0, -2$ with resp. prob. $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, a "lazy" (slowed-down) version of $\mathbf{D}$ for Reflection.

Conclusions

- $T_{indpendent}$ stoch. dominates $T_{reflection}$ $\Rightarrow$ Independent is not maximal.
- Reflection is maximal, as by symmetry:
  1. T is the hitting time of the midpoint k by both $\mathbf{X}_1, \mathbf{X}_2$.
  2. $p_1^t(j) > p_2^t(j)$ if and only if $j < k$, and so

$$d_{TV}(p_1^t, p_2^t) = E[\underbrace{\mathbf{1}_{\{X_1(t)<k)\}}}_{\supseteq\{T>t\}} - \underbrace{\mathbf{1}_{\{X_2(t)<k\}}}_{=0 \text{ on } \{T>t\}}, T > t] = P(T > t).$$

Example: 3-State Symmetric Chain

$$p = \frac{1}{6}\begin{pmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \mu_1 = \delta_1, \mu_2 = \delta_2.$$

"Greedy Coupling"

Maximize meeting probability for every step.

- Conditioned on $\{X_1(t) \neq X_2(t)\}$, the probability of $\{X_1(t+1) = X_2(t+1)\}$ is at most the sum of minima of the reps. rows $= \frac{1}{3} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$, for any pair of states.
- Therefore $P(T > t) = 3^{-t}$. Maximal?
- But e. values of p are $1, \pm\frac{1}{\sqrt{12}}$ (calculate the traces of p, p^2), and so (spectral theorem)
$$d_{TV}(p_1^t, p_2^t) = \Theta((\frac{1}{\sqrt{12}})^t) = 3^{-t}\Theta((\frac{\sqrt{3}}{2})^t). \text{ Oops.}$$
- Where's the catch?

Example: 3-State Symmetric, Continued

Explanation

- Coupling is for distributions, not states.
- Distributions are convex combinations of rows, not necessarily rows...
- Applying greedy to the propagated distributions results in the following diagram:

$$\delta_1, \delta_2 \xrightarrow{1/3} \underbrace{\frac{1}{2}(\delta_1 + \delta_2), \delta_3}_{\text{normalized excess}} \xrightarrow{3/12} (\delta_2, \delta_1) \longrightarrow \cdots$$

resulting in the probability $\frac{1}{3} \times \frac{3}{12} = \frac{1}{12}$ of not coupling in any two consecutive steps, achieving maximality.

Additional discussion

- A Markovian coupling is one in which the component processes are MCs with respect to the joint filtration. The discussion is a proof that there does not exist a maximal Markovian coupling in this case.
- Here, however, one can construct a MC which is a maximal coupling.

Maximal couplings and what's in this work

Pitman's construction

- Key observation: maximality determines the distribution of $(X_1(T), T)$.
- Sample $(X_1(T), T)$. Conditioned on that, sample $\mathbf{X}_1$ and $\mathbf{X}_2$ backward in time, independently as some specific time inhomogeneous Markov chains.

This work

- The mathematical object presented here is identical to Pitman's construction. The differences are in details of the construction, resulting in a working algorithm for computer simulation.
- Maximal couplings require iterative computations that introduce complexity. These calculations also provide the total variation, so maximal couplings are not efficient method for probabilistically estimating the total variation.

Measure Decomposition: The Bookkeeping Slide

Adaptive Decomposition

Idea. Maximally couple the initial distributions (Prop. 1-2), propagate the excess, repeat...
Define three sequences of sub-probability measures on S indexed by $t \in \mathbb{Z}_+$:

$$m_0 = \underbrace{\min(\mu_1, \mu_2)}_{\text{max coupling at time 0}} \quad \text{and } a_{k,0} = \underbrace{\mu_k - m_0}_{\text{excess}}, \ k = 1, 2.$$

Continue inductively, through propagation of the excess measures:

$$m_{t+1} = \underbrace{\min(p(a_{1,t}, \cdot), p(a_{2,t}, \cdot))}_{\text{max coupling at time t+1}} \qquad a_{k,t+1} = \underbrace{p(a_{k,t}, \cdot) - m_{t+1}}_{\text{excess}}.$$

Linearization

The non-linear decomposition can be linearized through absolute continuity.

- $a_{k,0}(j) \leq \mu_k(j) \ \Rightarrow \ \exists d_{k,0}(j) \in [0, 1], \ a_{k,0} = d_{k,0}(j)\mu_k(j).$
- $a_{k,t+1}(j) \leq p(a_{k,t}, j) \ \Rightarrow \ \exists d_{k,t+1}(j) \in [0, 1] \ a_{k,t+1}(j) = p(a_{k,t}, j)d_{k,t+1}(j).$

Thus, letting

$$q_{k,t}(i, j) = p(i, j)\underbrace{d_{k,t}(j)}_{\text{coin flip!}} \text{ and } \bar{q}_{k,t} = p(i, j)(1 - d_{k,t}(j)), \ t \in \mathbb{N}, \ i, j \in S$$

we have

	Excess	Coupling
Initiation	$a_{k,0} = d_{k,0}\mu_k$	$m_{k,0} = (1 - d_{k,0})\mu_{k,0}$
Linear propagation	$a_{k,t+1} = q_{k,t+1}(a_{k,t}, \cdot)$	$m_{t+1} = \bar{q}_{k,t+1}(a_{k,t}, \cdot)$

Path Measures

The decomposition can be lifted to path measures:

Path Measure	$t-$th marginal
$a_k(\mathbf{i}) = \alpha_{k,0}(i_0)\prod_{0\le s\le t-1} q_{k,s+1}(i_s, i_{s+1}),\ \ t\in\mathbb{Z}_+,\ \ \mathbf{i}\in S^{t+1}$	$\sum_{i_0,\dots,i_{t-1}} a_k(\mathbf{i}) = a_{k,t}(i_t)$
$\bar{a}_{k,t}(\mathbf{i}) = \begin{cases} m_0(i_0) & t=0 \\ a_k(i_0,\dots,i_{t-1})\bar{q}_{k,t}(i_{t-1},i_t) & t\in\mathbb{N}\end{cases}$	$\sum_{i_0,\dots,i_{t-1}} \bar{a}_{k,t}(\mathbf{i}) = m_t(i_t).$

If $m_t(i_t) > 0$ one can define a probability measure $\bar{a}_{k,t,i_t}$ on S^{t+1} through

$$\bar{a}_{k,t,i_t}(\mathbf{i}) = \frac{\bar{a}_{k,t}(\mathbf{i})}{m_t(i_t)}.$$

Key Observation. Linear propagation implies $\bar{a}_{k,t,i_t}$ can be sampled step by step, in reverse.

Lemma 3 (Reverse Chain)

For $t\in\mathbb{N}$,

$$\bar{a}_{k,t,i_t}(\mathbf{i}) = \overleftarrow{\bar{q}}(i_t, i_{t-1})\overleftarrow{q}_{k,t-1}(i_{t-1}, i_{t-2})\cdots\overleftarrow{q}_{k,1}(i_1, i_0),$$

where $\overleftarrow{\bar{q}}_{k,s}$, $\overleftarrow{q}_{k,s}$ are TF given by

$$\overleftarrow{\bar{q}}_{k,s}(i,j) = \frac{a_{k,s-1}(j)\bar{q}_{k,s}(j,i)}{m_s(i)} \text{ and } \overleftarrow{q}_{k,s}(i,j) = \frac{a_{k,s-1}(j)\bar{q}_{k,s}(j,i)}{a_{k,s}(i)}.$$

Some Notation

In the next two definitions we assume that **X** is a (p, μ_1, μ_2)-coupling.

Definition 4 (Single Process Coupling Measures)

The pre- and up-to-coupling coupling measures of the k-th marginal, $\mathbb{PC}_k, \mathbb{UC}_k$ are the family of measures given by

$$\mathbb{PC}_k(\mathbf{i}) = P(X_k([0,t]) = \mathbf{i} \ , T > t)$$
$$\mathbb{UC}_k(\mathbf{i}) = P(X_k([0,t]) = \mathbf{i}, T = t), \ \ t \in \mathbb{Z}_+, \mathbf{i} = (i_0, \ldots, i_t) \in S^{t+1}.$$

Definition 5 (Multiplicative Measures)

Suppose $(\nu_t : t \in \mathbb{Z}_+)$ is a family of sub-probability measures with ν_t a measure on S^{t+1}. The family is k-multiplicative if there exist constants $d_{k,\cdot}(\cdot) \in [0,1]$ so that

$$\nu_t(\mathbf{i}) = \prod_{0 \le s \le t} d_{k,s}(i_s) P(X_k([0,t]) = \mathbf{i}), \ \ t \in \mathbb{Z}_+, \ \ \mathbf{i} = (i_0, \ldots, i_t) \in S^{t+1}.$$

Comment

The family of path measures $(a_k(\cdot)|_{S^{t+1}} : t \in \mathbb{Z}_+)$ is multiplicative.

Maximal Couplings: Properties

Qualitative:

Lemma 4

Let $\mathbf{X}$ *be a* (p, μ_1, μ_2)*-coupling. The coupling is maximal if and only if, for all* $t \in \mathbb{Z}_+$ *with* $P(T > t) > 0$*,*

$$P(X_1(t) \in \cdot \,|T > t) \perp P(X_2(t) \in \cdot \,|T > t).$$

Quantitative:

Proposition 5

Let $\mathbf{X}$ *be a maximal a-*(p, μ_1, μ_2)*-coupling.*

- *Then* $\begin{aligned} P(X_k(t) = i, T > t) &= \alpha_{k,t}(i) \\ P(X_k(t) = i, T = t) &= m_t(i) \end{aligned}$
- *If, in addition,* $\mathbb{PC}_k$ *is k-multiplicative,* $\mathbf{X}$ *then for* $t \in \mathbb{N}$ *and* $\mathbf{i} \in S^{t+1}$

$$\mathbb{PC}_k(\mathbf{i}) = \alpha_{k,0}(i_0) \prod_{s=0}^{t-1} q_{k,s+1}(i_s, i_{s+1})$$

$$\begin{aligned} \mathbb{UC}_k(\mathbf{i}) &= \mathbb{PC}_k(i_0, \ldots, i_{t-1}) \bar{q}_{k,t}(i_{t-1}, i_t) \\ &= \bar{a}_{k,t,i_t}(\mathbf{i}) m_t(i_t). \end{aligned}$$

Maximal Coupling Construction

This is our version for Pitman's construction.

Theorem 2

1. *Sample* $\mathbf{X}_1$ *as a* (p, μ_1)*-MC.*
2. *Sample* T*: for* $t \in \mathbb{Z}_+$*, sample a Bernoulli* B_t *with* $P(B_t = 1) = 1 - d_{k,t}(X_1(t))$ *independently of everything else. Let* T *be the time of the first success,*

$$T = \inf\{t \in \mathbb{Z}_+ : B_t = 1\}.$$

3. *Set* $X_2(T + t) = X_1(T + t),\ t \in \mathbb{Z}_+$*.*
4. *Sample* $\mathbf{X}_2([0, T])$*:*
    - *If* $0 < T < \infty$*, sample* $X_2([0, T])$ *in reverse according to* $\bar{\alpha}_{2,T,X(T)}$ *through Lemma 3.*
    - *If* $T = \infty$*, sample* $\mathbf{X}_2$ *as the inhomogeneous chain of time with initial distribution* $\frac{a_{2,0}b_0}{a_\infty}$ *and transition from time* s *to time* $s+1$ *given by* $\frac{1}{b_s(i_s)} q_{2,s+1}(i_s, i_{s+1}) b_{s+1}(i_{s+1})$*, where* $a_\infty = \lim_{t\to\infty} a_{2,t}(S)$ *and* $b_s(i_s) = \lim_{t\to\infty} \prod_{r=s}^{t} q_r(i_r, i_{r+1})$*.*

Then $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ *is a maximal* (p, μ_1, μ_2)*-coupling. Moreover, the pre-coupling measures* $\mathbb{PC}_{1,2}$ *are* $1, 2$*-multiplicative.*

Comments

- Pitman does not address the case $T = \infty$. He first samples $(X(T), T)$ and obtains both $\mathbf{X}_1$ and $\mathbf{X}_2$ in reverse, as in the first item of Step 4. This is great as a mathematical proof, but ineffective for simulation.
- Step 4 can be skipped if primary interest is in $\mathbb{PC}_k$.

Algorithmic Implementation

Maximal Coupling Sampler

```
t ← 0                                                   ▷ Sample X_1
X_1(0) ~ μ_1
while Uniform[0,1] < d_{1,0}(X_1(t)) do                 ▷ Continue until coupling time
    X_1(t+1) ~ p(X_1(t), ·)
    t ← t + 1
end while
X_2(t) ← X_1(t)                                         ▷ Sample conditioned X_2 backwards
if t > 0 then
    X_{t-1} ~ ←q̄_t(X_2(t), ·)
    t ← t − 1
    while t > 0 do
        Y_{t-1} ~ ←q_t(X_2(t), ·)
        t ← t − 1
    end while
end if
```

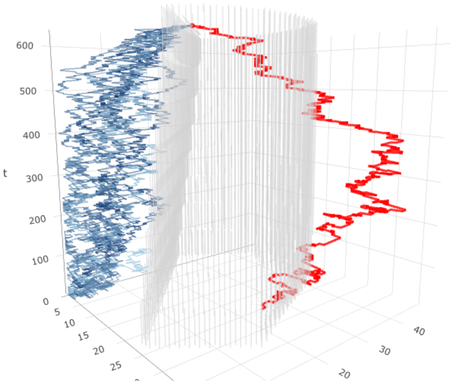
Comments

- All parameters used are calculated and stored globally during the sampling of $\mathbf{X}_1$ for reuse in future iterations, resulting in a speedup in successive samples.
- This algorithm is designed for the case $\lim_{t\to\infty} d_{TV}(p_1^t, p_2^t) = 0$. If the limit is positive, with that probability the sampling of $\mathbf{X}_1$ will not terminate.

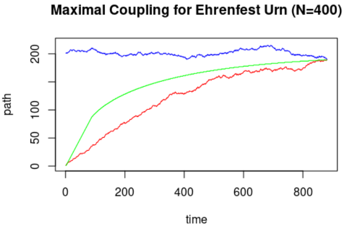
Example: RW on Grid

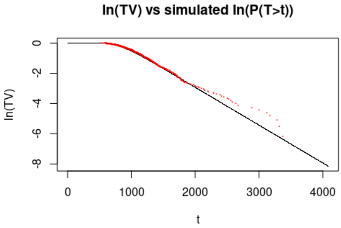
RW on Square Grid

- p is nearest neighbor RW on a 50×50 grid, μ_1 and μ_2 are concentrated at the center and corner, resp.
- Time represented by z-axis.
- $\mathbf{X}_1$ starts at the center and sampled until T.
- Conditioned on $(X_1(T), T)$, 10 independent copies of $\mathbf{X}_2$ are sampled in reverse.
- Gray points represent states where maximal coupling can occur.
- The distribution of $\mathbf{X}_2$ is quite different from that of a RW.



Example: The Ehrenfest Urn





Left image:

- $N = 400$, lazy model: $p(i,j) = \frac{1}{2}\begin{cases} \frac{N-i}{N} & j = i+1,\ 0 \le i \le N-1 \\ 1 & j = i,\ 0 \le i \le N \\ \frac{i}{N} & j = i-1,\ 1 \le i \le N. \end{cases}$
- $\mathbf{X}_1$ has initial dist. $\mu_1 = \delta_1$.
- $\mathbf{X}_2$ has initial dist, μ_2, $\mathrm{Bin}(N, 0.5)$ (stationary for p).
- Green curve represents states where max coupling must occur.

Right image:

- Simulation of 500 maximal couplings with same parameters as before.
- Black curve: $\ln d_{TV}(p^t_{\mu_1}, p^t_{\mu_2})$.
- red scatter: ln of the empirical tail of T.

Flipping out...

TЧenk Yon!