Markov chains

Jonathan L.F. King
University of Florida, Gainesville FL 32611-2082, USA
squash@ufl.edu
Webpage http://squash.1gainesville.com/
5 June, 2018 (at *03:21*)

ABSTRACT: Markov chains, neither the 1-step nor the multi-step, are stable under finite-block codes.

Geometric preliminaries.

In a real vectorspace $\mathbf{V}$, say that

$$\dagger: \qquad \sum_{j=1}^{N} \alpha_j \mathbf{v}_j \quad (\text{with each } \alpha_j \in \mathbb{R})$$

is a ***linear combination*** (*lin.comb*) of vectors (points) $\mathbf{v}_1, \ldots, \mathbf{v}_N$. If, further, these scalars satisfy

$$\ddagger: \qquad \alpha_1 + \alpha_2 + \cdots + \alpha_N = 1,$$

then we call (†) a ***weighted average*** of the points. Finally, if (‡) *and* each $\alpha_j \geqslant 0$, then we call (†) a ***convex average*** of the points.

Given a set $S \subset \mathbf{V}$ of points, we define three supersets

$$\mathrm{Spn}(S) \supset \mathrm{AffSpn}(S) \supset \mathrm{Hull}(S).$$

The ***span*** is the set of all lin.combs (†), as $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ ranges over *all finite* subsets of S. The ***affine span*** is the set of all (†) satisfying (‡), whereas the ***hull*** is the smaller set of all convex averages. Thus $\mathrm{Spn}(S)$ is the smallest *subspace* (that includes S) whereas $\mathrm{AffSpn}(S)$ is the smallest *affine-space* and $\mathrm{Hull}(S)$ is the smallest *convex set*.

A point $\mathbf{w} \in C$ is an "***extreme point*** of a convex set C" if: *Whenever we write* $\mathbf{w} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$ *as a convex average* (of points $\mathbf{v}_1, \mathbf{v}_2 \in C$), *then necessarily* $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{w}$. A non-void set $C \subset \mathbf{V}$ is an N-dimensional ***simplex*** (an "N-simplex") if we can write it as

$$C = \mathrm{Hull}(\mathbf{w}_1, \ldots, \mathbf{w}_{N+1})$$

where no $\mathbf{w}_j$ is in the affine-span of the others. Equivalently, C has precisely $N{+}1$ extreme-pts, and $\mathrm{Dim}(C) = N$.

Existence of an invariant vector

Fix a posint $\mathfrak{D}$. Let $\mathbb{P} = \mathbb{P}^{\mathfrak{D}-1}$ be the simplex of probability vectors $\mathbf{v} \in \mathbb{R}^{\mathfrak{D}}$. Fix a $\mathfrak{D}{\times}\mathfrak{D}$ ***(column)-stochastic*** matrix M; each column is a prob.vec. Let $M{:}\mathbb{P}\circlearrowleft$ denote the map $\mathbf{v} \mapsto \mathsf{M}\mathbf{v}$ for a column-vector $\mathbf{v}$.

1: Perron-Frobenius Theorem (weak version). *There exists a fixpt* $\boldsymbol{\sigma} \in \mathbb{P}$, *i.e a column vector* $\boldsymbol{\sigma}$ *with* $\mathsf{M}\boldsymbol{\sigma} = \boldsymbol{\sigma}$. ♢

Proof (Brouwer fixed-pt). Function $M()$ is cts in, say, the $\mathbb{L}^1$-topology. Since $\mathbb{P}$ is homeomorphic with the $[\mathfrak{D}-1]$-disk, Brouwer applies to yield a fixed-point $\boldsymbol{\sigma} \in \mathbb{P}$. ♦

Proof (Cesàro averages). Fix a vector $\mathbf{v} \in \mathbb{P}$. Let

$$\mathbf{v}_N := \mathbb{A}_N(\mathbf{v}) := \frac{1}{N}\sum_{j\in[0..N)} \mathsf{M}^j\mathbf{v}.$$

Since $\mathbb{P}$ is cpt there is a $\boldsymbol{\sigma} \in \mathbb{P}$ and increasing seq $\vec{N}$ with $\mathbf{v}_{N_k} \xrightarrow{k\to\infty} \boldsymbol{\sigma}$. By cty of $M()$, then,

$$\mathsf{M}\boldsymbol{\sigma} = \mathsf{M}\cdot\lim_{k\to\infty}\mathbf{v}_{N_k} = \lim_{k\to\infty}\mathsf{M}\cdot\mathbf{v}_{N_k}$$

And observe that

$$\mathsf{M}\cdot\mathbf{v}_N = \frac{1}{N}\sum_{j\in[1..N]}\mathsf{M}^j\mathbf{v} = \frac{1}{N}[\mathsf{M}^N\mathbf{v} - \mathsf{M}^0\mathbf{v}] + \mathbb{A}_N(\mathbf{v}).$$

Sending $k \to \infty$ sends $N \to \infty$, so $\frac{1}{N}[\mathsf{M}^N\mathbf{v} - \mathsf{M}^0\mathbf{v}]$ goes to $\mathbf{0}$. Thus $\mathsf{M}\boldsymbol{\sigma} = \boldsymbol{\sigma}$. ♦

Exer. E1. Prove that the original full seq. $(\mathbf{v}_n)_1^\infty$ converges to $\boldsymbol{\sigma}$.

Fix ε and use $\mathbf{a} \approx \mathbf{b}$ to mean $\|\mathbf{a} - \mathbf{b}\| \leqslant \varepsilon$. ISTShow

$$\limsup_n \|\boldsymbol{\sigma} - \mathbf{v}_n\| \leqslant 3\varepsilon.$$

To this end, WLOG 7 is large enough that $\mathbb{A}_7(\mathbf{v}) \approx \boldsymbol{\sigma}$. Since M is a contraction and commutes with $\mathbb{A}_7$,

$$\mathbb{A}_7(\mathsf{M}^k\mathbf{v}) \approx \mathsf{M}^k\boldsymbol{\sigma} = \boldsymbol{\sigma}.$$

For each posint L, then,

$$\boldsymbol{\sigma} \approx \frac{1}{L}\sum_{\ell\in[0..L)}\mathbb{A}_7(\mathsf{M}^{7\ell}\mathbf{v}) \overset{\text{note}}{=\!=} \mathbb{A}_{7L}(\mathbf{v}).$$

(Rest left as exercise.) □

Defn. Let $\mathbf{v} \geqslant 0$ mean that each component $v_i \geqslant 0$; ditto for ">". (Same convention for matrices.) Use $\|\mathbf{v}\| := \sum_1^{\mathfrak{D}} |v_i|$ for the $\mathbb{L}^1$-norm. Note that

2: *If* $\mathbf{v} \geqslant 0$ *then* $\|\mathsf{M}\mathbf{v}\| = \|\mathbf{v}\|$,

since M is col-stochastic.

Computing its operator-norm on $\mathbb{L}^1(\mathbb{R}^{\mathfrak{D}})$,

$$\|\mathsf{M}\|_{\mathrm{op}} = 1.$$

□

3: Perron-Frobenius Theorem (stronger). *Suppose that α is positive, where*

$$\dagger: \qquad \alpha := \alpha(\mathsf{M}) := \operatorname*{Min}_{i,j\in[1..\mathfrak{D}]} \mathsf{M}_{i,j}\,.$$

Then M is a $[1-\alpha]$ contraction-mapping on $\mathbb{P}$, there is a unique fixed-pt $\boldsymbol{\sigma} \in \mathbb{P}$, and $\mathsf{M}^n\mathbf{v} \to \boldsymbol{\sigma}$ for each $\mathbf{v} \in \mathbb{P}$. Indeed, $\|\boldsymbol{\sigma} - \mathsf{M}^n\mathbf{v}\| \leqslant 2\cdot[1-\alpha]^n$. ♢

Proof. For $\mathbf{u},\mathbf{w} \in \mathbb{P}$, our objective is

$$\|\mathsf{M}\mathbf{u} - \mathsf{M}\mathbf{w}\| \overset{?}{\leqslant} \|\mathbf{u}-\mathbf{w}\|\cdot[1-\alpha]\,.$$

WLOG $\mathbf{u} \neq \mathbf{w}$. WLOG $\mathbf{u}$ and $\mathbf{w}$ have *disjoint supports*. (Let $\mathbf{v} := \mathbf{u}-\mathbf{w}$, decompose into pos/neg parts $\mathbf{v} = \mathbf{v}^+ - \mathbf{v}^-$, rename to $\mathbf{u}-\mathbf{w}$, having scaled to make these $\mathbf{u}$ and $\mathbf{w}$ probability vectors.) So now $\|\mathbf{u}-\mathbf{w}\| = 2$ and our goal becomes

$$\ddagger: \qquad \|\mathsf{M}\mathbf{u} - \mathsf{M}\mathbf{w}\| \overset{?}{\leqslant} 2 - 2\alpha\,.$$

From ($\dagger$), each of $\mathbf{u}' := \mathsf{M}\mathbf{u}$ and $\mathbf{w}' := \mathsf{M}\mathbf{w}$ dominates $\alpha\cdot\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$. So for each index $j \in [1..\mathfrak{D}]$,

$$|\mathbf{u}'_j - \mathbf{w}'_j| \leqslant \mathbf{u}'_j + \mathbf{w}'_j - 2\alpha\,.$$

Summing over j yields ($\ddagger$). ♦

4: Perron-Frobenius Corollary. *Suppose, for some posint K, that $\mathsf{M}^K > 0$. Then M has a unique fixedpt $\boldsymbol{\sigma} \in \mathbb{P}$. Further, $\exists\beta < 1$ so that:*

$$\forall\mathbf{v}\in\mathbb{P}, \forall n \geqslant K: \quad \left\|\mathsf{M}^n\mathbf{v} - \boldsymbol{\sigma}\right\| \leqslant \beta^n\,. \qquad ♢$$

Proof. Let $\boldsymbol{\sigma}$ be the fixed-pt under M^K. Then $\mathsf{M}\boldsymbol{\sigma} = \lim_n \mathsf{M}\cdot[\mathsf{M}^K]^n\boldsymbol{\sigma} = \lim_n [\mathsf{M}^K]^n\cdot\mathsf{M}\boldsymbol{\sigma}$. And this latter is $\boldsymbol{\sigma}$, since *every* vector converges to $\boldsymbol{\sigma}$ under powers of M^K. Etc. ♦

Remark. The ***transition graph*** for M has vertices $[1..\mathfrak{D}]$. This digraph, G, has an edge from j to i IFF entry $[\mathsf{M}]_{i,j}$ is positive. The meaning of $[\mathsf{M}^3]_{i,j}$ is the probability, having started in state j, of being in state i after *exactly* 3 steps. (There may exist several 3-step paths from j to i.)

G is strongly connected IFF $\forall i,j \ \exists k$ *with* $[\mathsf{M}^k]_{i,j} > 0$. I.e, IFF $\operatorname{Max}_{k\in[1..\mathfrak{D})} [\mathsf{M}^k]_{i,j}$ is positive, for each i,j. □

5: Frobenius Thm. *Suppose* $\operatorname{Gcd}(L_1,\ldots,L_N) = 1$. *Then there exists K so that*

$$L_1\mathbb{N} + L_2\mathbb{N} + \cdots + L_N\mathbb{N} \supset [K..\infty)\,. \qquad ♢$$

I.e, the non-negative linear combinations include an infinite interval. *Proof.* Exercise.

If G is strongly connected and the Gcd of all G-cycles is 1, then we say that G is ***tight*** (std: irreducible and aperiodic).

A (directed) loop in G is a ***simple loop*** if it repeats no vertex.

6: Theorem. *TFAEquivalent.*

a: *There exists a posint K with $\mathsf{M}^K > 0$.*

b: *G is tight.*

c: *$\exists K$ so that $\forall k \geqslant K$: $\mathsf{M}^k > 0$.* ♢

Proof a⇒b. There is a K-path from each state to each other, so certainly G is str. connected. WLOG G has ⩾2 states. Pick a state and a nbr $A \to B$. By hyp., we have paths $B \rightsquigarrow B$ and $B \rightsquigarrow A$, each of length K. Concatenating the latter with $A \to B$ gives a loop of length $K+1$. And $\operatorname{Gcd}(K, K+1) = 1$. ♦

Proof b⇒c. Let $L_1,\ldots,L_N$ denote the simple loops and also their lengths. Since G is finite, ISTFix two states A, B and show $\forall_{\text{large}}\, k$ that there is a k-path from A to B. Let p_j be a path from A to some state S_j in L_j. Let π be a path $A \rightsquigarrow B$.

Arbitrary natnums $\vec{n} := (n_1, \ldots, n_N)$ give rise to this path going from A to B: Go from A to S_1, circle the L_1 loop n_1 times, then return to state A. Now go to S_2, etc. Finally, after returning to A from S_N, follow our path from $A \rightsquigarrow B$. This total path has length

$$T + 2\sum_j n_j \cdot L_j .$$

where T is $\text{Len}(\pi) + 2\sum_j \text{Len}(p_j)$. Now the Frobenius thm (5) finishes the proof. ♦

Courtesy (6) and (3), a tight G has a *unique* stationary measure (invariant vector); agree to call it $\boldsymbol{\sigma}_{\mathsf{G}}$ or $\boldsymbol{\sigma}_{\mathsf{M}}$.

For a str.conn G, let $\text{CycGcd}(\mathsf{G})$ be the Gcd of the (lengths of) the simple loops in G (hence, of *all* the loops in G).

For a posint Q, let $\mathsf{G}^{(Q)}$ be the digraph of M^Q.

7: Theorem. *Take a str. connected digraph G (use M for its matrix) and let $L_1, \ldots, L_N$ denote the simple-loop lengths. Let $Q := \text{Gcd}(L_1, \ldots, L_N)$.*

Then $\mathsf{G}^{(Q)}$ has precisely Q many strongly connected components,

$$\dagger: \quad \mathsf{G}^{(Q)} = \mathsf{H}_0 \sqcup \mathsf{H}_1 \sqcup \ldots \sqcup \mathsf{H}_{Q-1} .$$

Each H has simple-loop lengths $\frac{L_1}{Q}, \ldots, \frac{L_N}{Q}$ (and possible others) and thus is tight. The components (†) can have been numbered so that M carries each H_j to $\mathsf{H}_{j\oplus 1}$ (addition mod Q). I.e, each state in H_j goes, under M, to a $\mathsf{H}_{j\oplus 1}$-state.

The original G has a unique invariant measure. It is

$$8: \quad \boldsymbol{\sigma}_{\mathsf{G}} = \frac{1}{Q} \cdot [\boldsymbol{\sigma}_0 + \boldsymbol{\sigma}_1 + \cdots + \boldsymbol{\sigma}_{Q-1}] ,$$

where $\boldsymbol{\sigma}_j$ denotes the M^Q-invariant measure $\boldsymbol{\sigma}_{\mathsf{H}_j}$. Moreover, M carries each measure $\boldsymbol{\sigma}_j$ to the next in circular order. That is, $\mathsf{M}\boldsymbol{\sigma}_j = \boldsymbol{\sigma}_{j\oplus 1}$, where we view $\boldsymbol{\sigma}_j$ as a col-vector whose non-zero entries are on the states of H_j. ♢

Proof. WLOG suppose $Q = 6$. Distinguish a state $S \in \mathsf{G}$. For a state B, suppose π_1, π_2 are paths $S \rightsquigarrow B$. Concatenate each with a particular $B \rightsquigarrow S$. Now we have two loops, so their lengths must be be congruent mod 6. Thus

$$\text{Len}(\pi_1) \equiv_6 \text{Len}(\pi_2) .$$

Thus we can label each state $B \in \mathsf{G}$ by either "0",…,"5" modulo its distance from S. The states with label j are the vertices of H_j.

Consider a G-loop L and some state $A \in L$; suppose its label is 4. Going along the loop, then, the next five states are $4\oplus 1, 4\oplus 2, 4\oplus 3, 4\oplus 4$ and $4\oplus 5$. So *all* 6 labels occur on L. In H_4, then, our state A lies in a loop of length $\frac{L}{6}$. (Note that a non-simple loop in G might give rise to a simple loop in H_4.)

Lastly, suppose μ is an M-invariant measure on the states of G. We need to show, for $j = 0, 1, \ldots, 5$, that the restriction

$$*: \quad \mu\rfloor_{\mathsf{H}_j} = \frac{1}{6} \cdot \boldsymbol{\sigma}_j .$$

But M carries H_j to $\mathsf{H}_{j\oplus 1}$, so μ must give mass$=\frac{1}{6}$ to each H component. And μ is invariant under M^6, whence $(*)$, since each H_j has a *unique* M^6-invariant measure. ♦

Reversibility

Use $\text{Inv}(\mathsf{G})$ for the set of M-invariant measures on G. Let $\text{Rev}(\mathsf{G})$ be the set of infinitely reversible measures; those measures μ_0 so that there exists probability measures μ_j with $\mathsf{M}\mu_j = \mu_{j-1}$.

Evidently $\text{Inv}(\mathsf{G})$ and $\text{Rev}(\mathsf{G})$ are convex subsets of $\mathbb{P}$.

Defn. A vertex S of a str.conn digraph G is ***robust*** if each of its descendents is an ancestor. If S is not robust then say it is ***leaky***; this, since probability on S can leak-out to a descendent which is not an ancestor, and thus this probability can never get back to S.

Use $\mathrm{Core}(\mathsf{G})$ to denote the subgraph of robust vertices (and the directed-edges between them). Decomposing this into str.connected components

$$\mathrm{Core}(\mathsf{G}) \;=\; \mathsf{C}_1 \sqcup \mathsf{C}_2 \sqcup \ldots \sqcup \mathsf{C}_{N+1}$$

we will call the "str.conn decomposition of G's core". □

The next lemma explores a state S which is not robust.

9: Leakage Lemma. *Suppose we have an edge $S \to B$ with*

$$B \notin \mathcal{A} \;:=\; \mathrm{Ancestor}(S)$$

Then $\exists\, L$ posint and $\varepsilon > 0$ s.t for each $\mu \in \mathbb{P}$:

$$\forall n \geqslant L: \quad [\mathsf{M}^n \mu](\mathcal{A}) \;\leqslant\; [1-\varepsilon]^n\,.$$

In particular, each reversible μ is supported on $\mathrm{Core}(\mathsf{G})$. ◊

Proof. Fix an edge $S \xrightarrow{\lambda} S'$ to a non-ancestor S'. For each $B \in \mathcal{A}$, let L_B be the length of a path $B \rightsquigarrow S$, and let τ_B be the product of the transition-probs along this path. Let $N := {}^{\#}\mathcal{A}$ and

$$L \;:=\; \operatorname*{Max}_B L_B \quad \text{and} \quad \tau \;:=\; \operatorname*{Min}_B \tau_B\,.$$

Now consider a measure which puts total-mass $\mathbf{m}$ on $\mathcal{A}$. It must put mass $\frac{\mathbf{m}}{N}$ on at least one state of $\mathcal{A}$; say B. Thus in τ_B many steps, mass $\frac{\mathbf{m}}{N} \cdot \tau_B$ will arrive at S in τ_B. The upshot?

In each L steps, a mass of $\delta := \frac{\mathbf{m}}{N} \cdot \tau \cdot \lambda$,

leaks from $\mathcal{A}$, never to return. In particular, a reversible measure μ must have all its support in $\mathrm{Core}(\mathsf{G})$. ♦

A coded Markov need not even be a generalized Markov process

(11Mar2002: I typed this from a printed copy from 29Mar1985. I edited it slightly.) Below, "process" means "stationary process". A process is "generalized Markov" if it is n-step Markov for some n.

Goal. I exhibit a three-state ergodic Markov process, alphabet $(\mathsf{b}_0, \mathsf{b}_1, \mathsf{c})$, coded by means of a length-one code to a two-state process. The code simply lumps the two b-states into a single "superstate" S. The coded alphabet is thus $(\{\mathsf{b}_0, \mathsf{b}_1\}, \mathsf{c})$, which I denote as (S, c).

Construction. Choose probabilities

$$0 \;<\; p_0 \;<\; p_1 \;<\; 1$$

and let $q_j := 1 - p_j$. State b_j goes to itself with prob. p_j, and goes to c with prob. q_j. Finally, state c goes to each b_j with probability $\frac{1}{2}$.

Evidently there is a unique stationary probability distribution of probabilities Y_0, Y_1, Y_{c} on states $\mathsf{b}_0, \mathsf{b}_1, \mathsf{c}$. Easily $Y_0, Y_1, Y_{\mathsf{c}} > 0$.

For a sequence $x_1 x_2 x_3 \ldots x_N$ of letters, let $\mathsf{P}(x_1 x_2 \ldots x_N)$ denote the probability that x_1 occurs followed by x_2, etc. Let x^n abbr. n consecutive letters x.

Proof. To show that the (S, c)-process is not generalized Markov, we compute the probability that the process produces c at the next step, conditioned on it having just produced $n{+}1$ consecutive letters S.

First note that

$$\mathsf{P}(\mathsf{S}^{n+1}) \;=\; \mathsf{P}(\mathsf{b}_0 {\mathsf{b}_0}^n) + \mathsf{P}(\mathsf{b}_1 {\mathsf{b}_1}^n)\,,$$

since to go between the b-states one must leave S. Hence

10: $$\mathsf{P}(\mathsf{S}^{n+1}) \;=\; Y_0 \cdot {p_0}^n \;+\; Y_1 \cdot {p_1}^n\,.$$

Similarly,

$$\textbf{11:}\quad \mathsf{P}(\mathtt{S}^{n+1}\mathtt{c}) \;=\; Y_0 \cdot {p_0}^n \cdot q_0 \,+\, Y_1 \cdot {p_1}^n \cdot q_1\,.$$

Consequently, the conditional probability equals

$$*\textbf{:}\quad \begin{aligned} \mathsf{P}(\mathtt{c} \mid \mathtt{S}^{n+1}) \;&=\; \frac{\mathsf{P}(\mathtt{S}^{n+1}\mathtt{c})}{\mathsf{P}(\mathtt{S}^{n+1})} \\ &=\; \frac{Y_0 \cdot {p_0}^n \cdot q_0 \,+\, Y_1 \cdot {p_1}^n \cdot q_1}{Y_0 \cdot {p_0}^n \,+\, Y_1 \cdot {p_1}^n}\,. \end{aligned}$$

Now divide top and bottom by ${p_1}^n$, then send $n \nearrow \infty$. Since $p_0/p_1 < 1$, we conclude that

$$\textbf{12:}\quad \lim_{n\to\infty} \mathsf{P}(\mathtt{c} \mid \mathtt{S}^{n+1}) \;=\; q_1\,.$$

Eventual constancy. Were $(\mathtt{S}, \mathtt{c})$ some n-step Markov process, then $n \mapsto \mathsf{P}(\mathtt{c} \mid \mathtt{S}^{n+1})$ would be eventually-constant. But, by cross multiplying in $(*)$, an equality $\mathsf{P}(\mathtt{c} \mid \mathtt{S}^{n+1}) = q_1$ would imply that

$$Y_0 \cdot {p_0}^n \cdot q_0 \;=\; Y_0 \cdot {p_0}^n \cdot q_1\,.$$

Yet $Y_0 \cdot {p_0}^n$ is not zero, so $q_0 = q_1$. Hence $p_1 = p_0$. ◊

Filename: `Problems/Probability/markov-chains.JK.latex`
As of: *Tuesday 15May2018.* Typeset: *5Jun2018 at 03:21.*