

Uniwersytet Wrocławski
Wydział Matematyki i Informatyki
Instytut Matematyczny
specjalność teoretyczna

Bartosz Sójka

**Explanation of connection between Hopf
algebras and Markov chains**

Praca licencjacka
napisana pod kierunkiem
prof. dr. hab. Dariusza Buraczewskiego

Wrocław 2018

Abstract

In [DPR14] Persi Diaconis, Amy Pang and Arun Ram described how to use Hopf algebras for study Markov chains. As it involves ideas from quite different branches of mathematics it could be hard to grasp a concept if someone is not familiar with them. The point of this paper is to describe some of their results in a more step-by-step, simplified way, so that they could be accessible for third year students who have passed probability and abstract algebra courses. I will focus on the example of shuffling cards by inverse riffle shuffle method. Structure will be as follows: firstly there will be introduction to both Hopf algebras and Markov chains, then there will be explanation how to describe a Markov chain with a Hopf algebra, finally I will describe how to find left eigenbasis and right eigenbasis of Markov chain associated with riffle shuffling using Hopf algebras.

1 Markov chains

Finite Markov chain is a random process on the finite set of states such that the probability of being in some state in the moment $n + 1$ depends only on in which state one was in the moment n . Now we will put this more formally.

Let $S = \{s_1, \dots, s_k\}$. The sequence of random variables (X_0, X_1, \dots) with values in S is a Markov chain with state space S if for all $n \in \mathbb{N}$, for all $s_{i_0}, s_{i_1}, \dots, s_{i_{n+1}} \in S$ such that

$$\mathbb{P}(X_0 = s_{i_0}, \dots, X_n = s_{i_n}) > 0$$

following condition (called Markov property) holds:

$$\mathbb{P}(X_{n+1} = s_{i_{n+1}} \mid X_0 = s_{i_0}, \dots, X_n = s_{i_n}) = \mathbb{P}(X_{n+1} = s_{i_{n+1}} \mid X_n = s_{i_n}). \quad (1)$$

It states that for all $s_i, s_j \in S$ the probability of moving from the state s_i to the state s_j is the same no matter what states $s_{i_0}, \dots, s_{i_{n-1}}$ were visited before.

For the Markov chain (X_0, X_1, \dots) the $|S| \times |S|$ matrix $K_{i,j} = \mathbb{P}(X_{n+1} = s_j \mid X_n = s_i)$ is called the transition matrix. We will sometimes write $K(s_i, s_j)$ instead of $K_{i,j}$. Note that the sum of any row is equal to 1 since it is the sum of probabilities of moving somewhere from s_i . Now the $K_{i,j}^n$ is the chance of moving from s_i to s_j in n steps.

Markov chains can be also viewed as random walks on the directed, labeled graphs, where states are vertices and edge's label is the probability of moving from one vertex to another.

Card shuffling can be viewed as a Markov chain on all possible arrangements of the cards in the deck with $K(x, y)$ equal to probability of going from arrangement x to arrangement y in one shuffle.

More extensive introduction can be found in [LPW17].

Cořtam cořtam stationary distribution.

2 Hopf algebras

2.1 Tensor products

First we will introduce tensor product of the vector spaces. Let V, W be vector spaces over the field K . Let Z be a vector space with basis $V \times W$. Note, that we

are taking entire $V \times W$ as a basis of Z not just a basis of $V \times W$. Consequently every non-zero element of Z has unique representation in the form $\sum_{i=1}^n \alpha_i(v_i, w_i)$.

Let \simeq be the smallest equivalence relation on Z satisfying:

For all $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, $k \in K$

$$\begin{aligned}(v, w_1) + (v, w_2) &\simeq (v, w_1 + w_2), \\ (v_1, w) + (v_2, w) &\simeq (v_1 + v_2, w), \\ k(v, w) &\simeq (kv, w), \\ k(v, w) &\simeq (v, kw).\end{aligned}$$

Since for all $z_1, z_2, z_3, z_4 \in Z$, all $k \in K$

$$\begin{aligned}z_1 \simeq z_2 \wedge z_3 \simeq z_4 &\implies z_1 + z_3 \simeq z_2 + z_4 \text{ and} \\ z_1 \simeq z_2 &\implies kz_1 \simeq kz_2,\end{aligned}$$

we treat Z/\simeq as a vector space with operations

$$\begin{aligned}[z_1]_{/\simeq} + [z_2]_{/\simeq} &:= [z_1 + z_2]_{/\simeq}, \\ k[z_1]_{/\simeq} &:= [kz_1]_{/\simeq}.\end{aligned}$$

We denote equivalence class $[(v, w)]_{/\simeq}$ as $v \otimes w$. The tensor product $V \otimes W := Z/\simeq$. Note, that in $V \otimes W$ there are vectors that can not be written as $v \otimes w$ for any v, w . However every $z \in V \otimes W$ can be written in as $z = \sum_{i=1}^n v_i \otimes w_i$ for some $v_1, \dots, v_n \in V$, $w_1, \dots, w_n \in W$. (More detailed explanation of this fact and the following example will come in the Observation 1..)

For example take $v_1, \dots, v_n, w_1, \dots, w_n$ such that they are lineary independent in corresponding spaces. Then take $\sum_{i=1}^n (v_i, w_i)$. There are no v, w such

that $[(v, w)]_{/\simeq} = \left[\sum_{i=1}^n (v_i, w_i) \right]_{/\simeq}$. Thus for the element $\left[\sum_{i=1}^n (v_i, w_i) \right]_{/\simeq}$ of

$V \otimes W$ there are no v, w such that $v \otimes w = \left[\sum_{i=1}^n (v_i, w_i) \right]_{/\simeq}$. However, since

$$\left[\sum_{i=1}^n (v_i, w_i) \right]_{/\simeq} = \sum_{i=1}^n [(v_i, w_i)]_{/\simeq} \text{ it can be written as } \sum_{i=1}^n v_i \otimes w_i.$$

Now we will make some futher observations on how $V \otimes W$ looks like.

Observation 1. *If $\{b_i\}_{i \in I}$, $\{c_j\}_{j \in J}$ are bases of, respectively, V and W , then $\{b_i \otimes c_j : i \in I, j \in J\}$ is the basis of $V \otimes W$.*

Proof. Let $z = \sum_{i=1}^n \alpha_i(v_i, w_i)$ be an arbitraly non-zero element of Z . We will

show that $[z]_{/\simeq}$ has representation as $\sum_{i=1}^m \beta_i [(b_i, c_i)]_{/\simeq} \left(= \sum_{i=1}^m \beta_i (b_i \otimes c_i) \right)$.

$$\begin{aligned}
[z]_{/\simeq} &= \left[\sum_{i=1}^n \alpha_i (v_i, w_i) \right]_{/\simeq} = \sum_{i=1}^n \alpha_i [(v_i, w_i)]_{/\simeq} = \\
&= \sum_{i=1}^n \alpha_i \left[\left(\sum_{j=1}^{l_1} \gamma_{i,j} b_{i,j}, \sum_{k=1}^{l_2} \gamma_{i,k} c_{i,k} \right) \right]_{/\simeq} = \sum_{i=1}^n \alpha_i \left[\sum_{j=1}^{l_1} \gamma_{i,j} \left(b_{i,j}, \sum_{k=1}^{l_2} \gamma_{i,k} c_{i,k} \right) \right]_{/\simeq} = \\
&= \sum_{i=1}^n \alpha_i \left[\sum_{j=1}^{l_1} \gamma_{i,j} \left(\sum_{k=1}^{l_2} \gamma_{i,k} (b_{i,j}, c_{i,k}) \right) \right]_{/\simeq} = \sum_{i=1}^n \alpha_i \left[\sum_{\substack{1 \leq j \leq l_1 \\ 1 \leq k \leq l_2}} \gamma_{i,j} \gamma_{i,k} (b_{i,j}, c_{i,k}) \right]_{/\simeq} = \\
&= \sum_{i=1}^n \alpha_i \left(\sum_{\substack{1 \leq j \leq l_1 \\ 1 \leq k \leq l_2}} \gamma_{i,j} \gamma_{i,k} [(b_{i,j}, c_{i,k})]_{/\simeq} \right) = \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq l_1 \\ 1 \leq k \leq l_2}} \alpha_i \gamma_{i,j} \gamma_{i,k} [(b_{i,j}, c_{i,k})]_{/\simeq}
\end{aligned}$$

Thus $\{b_i \otimes c_j : i \in I, j \in J\}$ spans $V \otimes W$. To prove linear independence we can observe that if $\sum_{i=1}^m \alpha_i [(v_i, w_i)]_{/\simeq} = 0$ then either v_1, \dots, v_n or w_1, \dots, w_n have to be lineary dependent. It can't occur if v_1, \dots, v_n and w_1, \dots, w_n are from the bases of V and W .

This observation also justifies recently cited fact and the example. \square

Observation 2. If V and W are finite dimentional and $\dim(V) = n$, $\dim(W) = m$, then $\dim(V \otimes W) = nm$.

Proof. The proof is imediate from the Observation 1.. Since if $\{b_i\}_{i \in I}$, $\{c_j\}_{j \in J}$ are bases of, respectively, V and W and $\dim(V) = n$ and $\dim(W) = m$, then $|\{b_i \otimes c_j : i \in I, j \in J\}| = nm$ \square

Observation 3. $V \otimes W$ is a vector space of elements in the shape of $\sum_{i=1}^n v_i \otimes w_i$ with operations on them defined such that for all $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, $k \in K$ there hold

$$\begin{aligned}
v_1 \otimes w + v_2 \otimes w &= (v_1 + v_2) \otimes w, \\
v \otimes w_1 + v \otimes w_2 &= v \otimes (w_1 + w_2), \\
k(v \otimes w) &= (kv) \otimes w = v \otimes (kw).
\end{aligned}$$

Proof. This observation is just recall of the definition. \square

Observation 4. For vector spaces U, V, W over the field K there is an natural isomorphism between $(U \otimes V) \otimes W$ and $U \otimes (V \otimes W)$ therefore there is no ambiguity in writing $U \otimes V \otimes W$ or a product of any greater number of vector spaces in that way. (Also we will write " $u \otimes v \otimes w$ " for some of their elements.) Form of elements, operations on them and structure of that vector spaces are

fully analogous to described above (in respect to all "coordinates" in terms like $u \otimes v \otimes w$ and so on). So the space $U \otimes V \otimes W$ has elements of shape $\sum_{i=1}^n u_i \otimes v_i \otimes w_i$ (each for some $u_1, \dots, u_n \in U$, $v_1, \dots, v_n \in V$, $w_1, \dots, w_n \in W$) and for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, $k \in K$ there hold

$$\begin{aligned} u_1 \otimes v \otimes w + u_2 \otimes v \otimes w &= (u_1 + u_2) \otimes v \otimes w, \\ u \otimes v_1 \otimes w + u \otimes v_2 \otimes w &= u \otimes (v_1 + v_2) \otimes w, \\ u \otimes v \otimes w_1 + u \otimes v \otimes w_2 &= u \otimes v \otimes (w_1 + w_2), \\ k(u \otimes v \otimes w) &= (ku) \otimes v \otimes w = u \otimes (kv) \otimes w = u \otimes v \otimes (kw). \end{aligned}$$

Proof. Left for the reader. \square

Observation 5. If V is a vector space over K , then all elements of $K \otimes V$ ($V \otimes K$) can be expressed in form $1 \otimes v$ ($v \otimes 1$) and there are natural isomorphisms $\varphi_L : K \otimes V \rightarrow V$, ($\varphi_R : V \otimes K \rightarrow V$) given by

$$\begin{aligned} \varphi_L(k \otimes v) &= kv, \\ \varphi_R(v \otimes k) &= kv. \end{aligned}$$

Proof. An arbitrary element of $K \otimes V$ has form $\sum_{i=1}^n k_i \otimes v_i$ but

$$\sum_{i=1}^n k_i \otimes v_i = \sum_{i=1}^n 1 \otimes k_i v_i = 1 \otimes \sum_{i=1}^n k_i v_i.$$

φ_L is linear (left for the reader) and is bijection because for all $v, v_1, v_2 \in V$

$$\varphi(1 \otimes v) = v$$

and

$$\begin{aligned} 1 \otimes v_1 = 1 \otimes v_2 &\iff 1 \otimes v_1 - 1 \otimes v_2 = 0 \iff \\ 1 \otimes (v_1 - v_2) = 0 &\iff v_1 - v_2 = 0 \iff v_1 = v_2. \end{aligned}$$

The proof for $V \otimes K$ and φ_R is analogous. In the later sections we will use notations of φ_L and φ_R for those isomorphism for any space. \square

Remark. Thanks to Observation 3. there is no ambiguity in writing $kv \otimes w$. I hope that this third observation will also help in understanding what tensor product is and what is not. It will be good to keep it in mind when we will be intensively dealing with it in a combinatorial way in the following sections.

2.2 Hopf algebras

Now there will be full definition of a Hopf algebra. Although it is quite long and involves definition of ??? operations, I decided to put it in a consistent fragment, due to believe that thanks to that it will be a better reference.

If reader will feel lost in this section it is recommended to read it in parallel to the section ??? where many examples are provided or treat it just as a reference

when formal definition will be needed. Another reason of arranging text like that (and possibility of treating this section just as a reference), is that for most of the time we will not be using full structure of a Hopf algebra. Nethertheless it is good to see the full shape of what we are dealing with. So now will come full definition nonetheless we will try to explain it piece by piece.

Remark. Let K be a field. In following section k , if not steted otherwise, will denotes an arbitrary element from this field. If not stated otherwise, all vector spaces will be over K and all tensor products will be taken over K .

Definition 1. A **K -algebra** is a vector space \mathcal{H} with additional associative, linear operation $m : \mathcal{H} \otimes \mathcal{H} \rightarrow \mathcal{H}$ called multiplication and linear map $u : K \rightarrow \mathcal{H}$ called unit such that for all $a \in \mathcal{H}$

$$m(u(1) \otimes a) = m(a \otimes u(1)) = a.$$

Explanation. Operation m defines on \mathcal{H} a structure of a unitary ring by setting the ring multiplication (let it be denoted as \cdot) as $a \cdot b = m(a \otimes b)$. The identity element of that ring multiplication is then $u(1)$. (We will be calling $u(1)$ also an identity element of multiplication m in K -algebra \mathcal{H} or the 1 in the \mathcal{H} , sometimes denoted $1_{\mathcal{H}}$)

Proof. The fact, that m is associative means that for all $a_1, a_2, a_3 \in \mathcal{H}$

$$m(m(a_1 \otimes a_2) \otimes a_3) = m(a_1 \otimes m(a_2 \otimes a_3)).$$

That implies that

$$(a \cdot b) \cdot c = m(m(a \otimes b) \otimes c) = m(a \otimes m(b \otimes c)) = a \cdot (b \cdot c).$$

So \cdot is proper ring mutiplication. Recalling the definition of u we can write that

$$u(1) \cdot a = a \cdot u(1) = a$$

So indeed it is an identity element of that ring. As u is linear map it can be seen as natural insertion of a field K into an algebra \mathcal{H} that maps 1_K to $1_{\mathcal{H}}$ (1 from the K to the identity element of multiplication in \mathcal{H}) and extends lineary. Given that we can observe that for all $a \in \mathcal{H}$, all $k \in K$, a multiplied by $u(k)$ (no matter if form the left or right) is exactly the ka (an element of vector space \mathcal{H}). So we can think about $u[K]$ as a copy of K in \mathcal{H} .

Remark. Let U, V, W, Z be a vector spaces over field K . Soon we will start to using notation $\varphi \otimes \psi : U \otimes V \rightarrow W \otimes Z$ which, for φ, ψ such that $\varphi : U \rightarrow W$, $\psi : V \rightarrow Z$, means a linear map that for all $u \in U$, $v \in V$ satisfies:

$$(\varphi \otimes \psi)(u \otimes v) = \varphi(u) \otimes \psi(v).$$

By linearity for element of shape $\sum_{i=1}^n u_i \otimes v_i$ it will take form:

$$(\varphi \otimes \psi)\left(\sum_{i=1}^n u \otimes v\right) = \sum_{i=1}^n \varphi(u) \otimes \psi(v).$$

I , if not stated otherwise, will be an identity in the adequate space.
 T , if not stated otherwise, will be the twist map $T : V \otimes W \rightarrow W \otimes V$, which is linear map such that for any $v \otimes w \in V \otimes W$

$$T(v \otimes w) = w \otimes v.$$

Throught this paper we will omit the "o" symbol for composition of functions where will be no risk of confusion and we will write $\varphi\psi(x)$ instead of $(\varphi \circ \psi)(x)$.

Definition 2. A *K -coalgebra* is a vector space \mathcal{H} with additional coassociative, linear operation $\Delta : \mathcal{H} \rightarrow \mathcal{H} \otimes \mathcal{H}$ called comultiplication and a linear map $\varepsilon : \mathcal{H} \rightarrow K$ called counit such that for all $a \in \mathcal{H}$

$$\begin{aligned} (\varepsilon \otimes I)\Delta(a) &= 1 \otimes a \text{ and} \\ (I \otimes \varepsilon)\Delta(a) &= a \otimes 1. \end{aligned}$$

Note that properties of an unit from a K -algebra also can be written in that manner as:

$$\begin{aligned} m(u \otimes I)(1 \otimes a) &= a \text{ and} \\ m(I \otimes u)(a \otimes 1) &= a \end{aligned}$$

means exactly what was in the definition of u .

Explanation. We will introduce a notation called Sweedler notation [Swe69] which will be very usefull for writing coproducts. As for all $a \in \mathcal{H}$ we have

$$\Delta(a) = \sum_{i=1}^n a_{1,i} \otimes a_{2,i}, \text{ we will write}$$

$$\Delta(a) = \sum a_1 \otimes a_2.$$

This notation surpresses the index "i". Somewhere there can be also encountered an interjaced notation $\Delta(a) = \sum_{(a)} a_{(1)} \otimes a_{(2)}$.

In many cases comultiplication can be seen as a sum of possible decomposition of an element into elements "smaller" in some sens. For example, later it will come out that exactly the comultiplication will be the operation that will model the process of cutting the deck of cards into pieces in riffle shuffle. In examples that we will working with (graded, connected Hopf algebras) comultiplication will represent some kind of natural decomposition in the more general way. What does it mean in the strict sense will be presented in Definition 5. when we will be introducing graded Hopf Algebras.
 Examples.

The coassociativity of Δ means, that $(\Delta \otimes I)\Delta = (I \otimes \Delta)\Delta$. In Sweedler notation it can be written as

$$\forall_{a \in \mathcal{H}} \sum \Delta(a_1) \otimes a_2 = \sum a_1 \otimes \Delta(a_2)$$

or in more expand form as

$$\forall_{a \in \mathcal{H}} \sum a_{11} \otimes a_{12} \otimes a_2 = \sum a_1 \otimes a_{21} \otimes a_{22}. \quad (2)$$

Because of these equalities, terms from (2) can be written as $\sum a_1 \otimes a_2 \otimes a_3$ without ambiguity.

We can also define

$$\Delta_2 := (\Delta \otimes I)\Delta$$

Now, for all $a \in \mathcal{H}$ there will be an equality

$$\Delta_2(a) = \sum a_1 \otimes a_2 \otimes a_3$$

which can be viewed as a sum of possible decompositions of a into three parts. In this point of view we can informally say that coassociativity of Δ means that Δ represent decomposition such that, when did twice, probabilities of possible outcomes are the same no matter which set of part (a_1 or a_2) have been tooked in the second iteration. It will be put more precise in the section 3. where we will present connection between Markov chains and Hopf algebras. Now we will take it a step futher:

Let C will be a coalgebra with comultiplication Δ and counit ε . We will recurrently define the sequence of maps $(\Delta_n)_{n \geq 0}$, such that $\Delta_n : C \rightarrow \underbrace{C \otimes \cdots \otimes C}_{n \text{ times}}$ as follows:

$$\begin{aligned} \Delta_1 &:= \Delta, \\ \Delta_n &:= (\Delta \otimes \underbrace{I \otimes \cdots \otimes I}_{n \text{ times}}) \Delta_{n-1}. \end{aligned}$$

Which can be seen as composed iterations of Δ .

By induction it can be proved that for all $n \geq 2, i \in \{1, \dots, n-1\}, m \in \{0, \dots, n-i\}$ we have

$$\Delta_n = (I^m \otimes \Delta_i \otimes I^{n-i-m}) \Delta_{n-i},$$

where I^m means $\overbrace{I \otimes \cdots \otimes I}^{m \text{ times}}$.

The proof can be found in [DNR00] (Proposition 1.1.7 and Lemma 1.1.10, sites 5-7). This formula is a generalization of a coassociativity. Thanks to that we can write

$$\Delta_n(a) = \sum a_1 \otimes \cdots \otimes a_n$$

with no ambiguity.

Interpretation is an extension of that described in the previous paragraph for $n = 2$. Now we are just decomposing a to n parts and probabilities of outcomes do not depend on which factors we are applying Δ in the next steps.

The counit property written in Sweedler notation takes form

$$\begin{aligned} \sum \varepsilon(a_1) \otimes a_2 &= 1 \otimes a, \\ \sum a_1 \otimes \varepsilon(a_2) &= a \otimes 1. \end{aligned}$$

Applying on both sides isomorphisms φ_L and φ_R from Observation 5. respectively we get

$$\begin{aligned} \sum \varepsilon(a_1) a_2 &= a, \\ \sum a_1 \varepsilon(a_2) &= a. \end{aligned}$$

References

- [DNR00] Sorin Dascalescu, Constantin Nastasescu, and Serban Raianu. *Hopf algebras: an introduction*. Pure and Applied Mathematics. CRC Press, 1 edition, 2000.
- [DPR14] Persi Diaconis, C. Y. Amy Pang, and Arun Ram. Hopf algebras and markov chains: two examples and a theory. *Journal of Algebraic Combinatorics*, 39(3):527–585, May 2014.
- [LPW17] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. AMS MBK #107. American Mathematical Society, 2017.